CONFORMATIONAL CHARACTERIZATION AND CLASSIFICATION OF INTRINSICALLY DISORDERED AMYLOID-β (1-42) THROUGH MOLECULAR DYNAMICS SIMULATIONS

by

Simiao Lu

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
October 2022

In the memory of my grandparents, Rui and Cheng.

Table of Contents

List of Tables

List of Figures

Abstract

Intrinsically disordered proteins (IDPs) are a group of proteins that lack the ability to fold into a well-defined 3D structure. The free energy landscapes of these proteins are assumed to have many competing low-energy states leading to an absence of a single tertiary structure. Characterizing their conformational spaces can be difficult by using standard experimental and computational techniques. Amyloid-β (Aβ) peptide, a prototypic IDP, aggregates into fibrils that are implicated in the pathogenesis of Alzheimers disease (AD). This thesis focuses on defining conformational states of the Aβ42 monomer using molecular dynamics (MD) simulations to better understand conformational changes of IDPs. The beginning of the amyloid aggregation includes a conformational transition from α- to β-dominated form of the Aβ42 monomer. We conducted simulations starting from one helical structure and one monomeric unit of the β-sheet fibril in aqueous solution. We observed that simulations from both directions would converge to a collection of conformational states during the α-to-β transition. MD simulations were performed in aqueous solution starting from monomeric systems based on various atomic structures. By merging structural ensembles into the same trajectory, the conformational space of the Aβ42 monomer was summarized using the principal component analysis. This ensemble suggests possible paths between configurational states. In the absence of a single fixed structure, I classified these heterogeneous conformers using alternative and novel approaches. Combining these, I have re-defined key structural elements of the monomeric form of Aβ42 tertiary structures. MD simulations of the Aβ42 monomer in ethanol-water cosolvents were implemented successively as a function of ethanol composition, mimicking the change of environment polarity. I demonstrated that the monomeric form reverts to an extended α-helix in a low polarity environment. Also, the α-to-β transition could be reversible by altering the solvent polarity. Observations and analyses on Aβ42 interacting with ethanol suggest similar behaviors of the peptide as when it binds with lipid membranes. Significantly, the central methodology of this thesis could be applicable for characterizing and categorizing structures of other IDPs.

List of Abbreviations Used

IDP Intrinsically disordered protein

PDB Protein data bank

NMR Nuclear magnetic resonance

NOE Nuclear Overhauser effect

$^{\mathbf{3}}$**J** Three bond scalar coupling

SAXS Small-angle X-Ray scattering

cryo-EM Cryo-electron microscopy

MD Molecular dynamics

AD Alzheimer's disease

APP Amyloid precursor protein

IDR Intrinsically disordered region

Aβ Amyloid-β

ss-NMR Solid-state nuclear magnetic resonance

TEM Transmission electron microscopy

AFM Atomic force microscopy

GPU Graphical processing unit

REMD Replica exchange molecular dynamics

PCA Principal component analysis

dPCA Dihedral principal component analysis

HFIP Hexafluoroisopropanol

TFE 2,2,2-trifluoroethanol

SAMD Simulated annealing molecular dynamics

QM Quantum mechanics

MM Molecular mechanics

GROMOS GROningen MOlecular Simulation

PBC Period boundary conditions

GROMACS GROningen MAchine for Chemical Simulations

PC Principal component

cPCA Cartesian-coordinates principal component analysis

C_α-PCA C_α-distance based principal component analysis

con-PCA Contact-based principal component analysis

CD Circular dichroism

GPA Generalized procrustes analysis

RMSD Root-mean-square deviation

FEL Free energy landscape

DPC Discretizing principal component

CS/ES Collapsed state/Extended state

R_g Radius of gyration

SASA Solvent accessible surface area

NTR N-terminal region

CHC Central hydrophobic cluster

HTR Hydrophilic turn region

CTR C-terminal region

VMD Visual molecular dynamics

DSSP The dictionary of protein secondary structure

PCC Pearson correlation coefficient

HD Hamming distance

HB Hydrogen bond

SB Salt bridge

DMSO Dimethyl sulfoxide

RMSF Root-mean-square fluctuation

SD Standard deviation

Acknowledgements

First, I would like to thank my supervisor, Dr. Christian Blouin, who guided me during the challenges of graduate school. His example as a caring, patient and generous person gives me courage to carry on and makes me want to be a better person.

I thank my committee members Dr. Kathryn Vanya Ewart, Dr. Stephen Bearne and Dr. Robert Beiko for their advice, patience and pushing me forward.

I thank the past and present members of both my lab and Beiko lab, for helping me with my work and spending time with me. Specially, I would like to thank Dr. Jose Sergio Hleap and Alex Safatli for their help, advice, and so many great memories we share.

I thank my husband, Long, who leads me through the valley of darkness with light of hope and helps me become open-minded, strong and independent.

I thank my parents and in-laws for their long-term support.

Chapter 1

Introduction

Proteins are macromolecules that are able to fold into well-defined tertiary structures that exhibit specific functions. However, a large number of natural occurring proteins can carry out biological roles without forming a well-folded structure [1, 2]. These proteins could appear fully or partially disordered. The discovery of these intrinsically disordered proteins (IDPs) has challenged the traditional sequence-structure-function paradigm [3–5]. Their importance is due to their high frequency in proteomes. Forty-four percent of human protein-coding genes contain encode disordered segments whose length is larger than 30 amino acids [6]. Importantly, IDPs not only play a central role in cellular signaling and regulatory networks [7, 8], but also are associated with various human disorders [5]. Some IDPs possess diverse states containing transient structural elements [9], or exist as molten-globules showing a compact structure with some secondary structure content [2]. Despite the fact that these dynamic motions and large-scale conformational changes are critical elements of their function, we still do not fully understand how dynamics and conformational changes link a protein's shape and its function [10, 11].

IDPs do not adopt stable secondary or tertiary structures. Their high flexibility inherently encoded in their primary amino acid sequence, thereby IDPs exist as ensembles of many different metastable conformations. The timescale of the structural interconversion is much faster than the transition from the folded to unfolded state of a globular protein (Figure 1.1). The unique features of IDPs, such as structural heterogeneity, distinct response to the environmental conditions, multifunctionality, and structural plasticity make it challenging for their structural determination with traditional techniques.

X-ray crystallography has been the most important protein structure determination technique. 90% of 140,000 structures in the Protein Data Bank (PDB) are determined using X-ray crystallography. However, high resolution models are usually

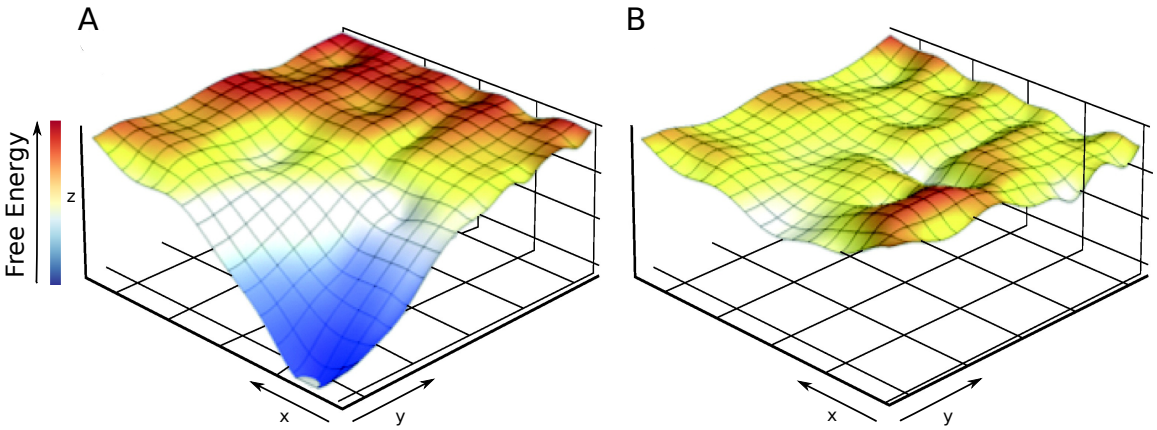

Figure 1.1: Schematics for free energy landscape of a globular protein (A) and an intrinsically disordered protein (B), where the xy plane represents the conformational space can be sampled of the corresponding peptide and z axis measures the relative free energies of its conformations. A globular protein has a *funnel-shaped* global energy minimum whereas an IDP has multiple local energy minima separated by small barriers. This figure is adapted from Fisher and Stultz [22], published in *Current Option in Structural Biology* with permission from ©2011 Elsevier.

only available for endpoints of dynamic processes or structures captured during the protein motion and conformational changes. This technique provides indirect information to roughly measure the protein flexibility accompaning by atomic displacement parameters or B-factors [11–15]. Nuclear magnetic resonance (NMR) spectroscopy is the main experimental technique that allows the investigation of IDPs at the atomic level. It also provides "indirect dynamic data" [14] but a richer source of information including nuclear overhauser effects (NOEs), chemical shifts, three bond scalar coupling (^{3}J), residual dipolar coupling, paramagnetic relaxation enhancements, *etc* [16]. It is not trivial to transform these observables into 3D models, thus, a combination with theoretical methods is beneficial [17, 18]. Other techniques such as small-angle X-ray scattering (SAXS) [19], hydrogen-deuterium exchange mass spectrometry [20], and recently cryo-electron microscopy (cryo-EM) [21] contribute to studies of structural determination and conformational dynamics of IDPs. Similar to X-ray crystallography and NMR, these methods are unable to give information about individual conformations in the ensemble but yield statistical averages of an entire ensemble.

The intrinsic difficulties of experimental techniques to provide direct information on protein dynamics have encouraged the development of simulation approaches. For the past decade, computational methods have proven to be crucial to study dynamics of IDPs [17, 23]. Molecular dynamics (MD) simulates a set of interacting atoms of a given molecular system over the evolution of time by solving Newton's equation of motion. With recent advances in computational power and algorithms, the MD method has made it possible to study larger biological systems at time scales from nanoseconds, microseconds to milliseconds [24, 25]. In this thesis, we are investigating large-scale conformational changes and structures of the intrinsically disordered amyloid-β (Aβ) peptide using MD combined with the clustering method, principal component analysis and other bioinformatic tools (see Chapter 2).

A large-scale conformational change is a significant conformational transition of a protein structure over a cellular process, such as ligand binding or protein folding in the response to a environmental change [26]. In this context, a large-scale conformational change indicates a series of drastic geometric changes of Aβ42 related to the *misfolding* and aggregation. In the *misfolding*, this change refers to a

conformational transition from a soluble membrane-attached α-helix structure to β-rich aggregation-prone-state in the extracellular environment. In the progression of amyloid aggregation, misfolded forms undergo further conformational changes and self-assemble into insoluble amyloid fibrils (see Section 1.3.2). Using MD to simulate IDP (such as Aβ) produces a trajectory including snapshots of conformations, many with large structural dissimilarities. To deal with such dissimilarity, a protocol of MD trajectory analysis is introduced in Chapter 3.

In the rest of this chapter, a brief summary of amyloid cascade hypothesis will be presented (Section 1.2). The hypothesis emphasizes the central role of Aβ in explaining the etiology and pathogenesis of Alzheimer's disease (AD). Then, a review of Aβ including amyloid precursor protein (APP), amyloid aggregation models, its suggested monomeric, oligomeric and fibrillar structures from experimental observations is described (see Section 1.3). In Section 1.4, a brief introduction of recent computational techniques and bioinformatic approaches applied in structural biology is presented. Lastly, a review of MD studies of monomeric Aβ folding in the published literature is presented in Section 1.5.

1.1 Intrinsically Disordered Proteins and Human Diseases

Many proteins are associated with neurodegeneration [9, 27], diabetes [28], cardiovascular diseases [29], genetic diseases [30, 31] and other amyloidosis [32, 33] are either intrinsically disordered or contain long intrinsically disordered regions (IDRs). Approximately 70% of human cancer-related proteins are predicted to contain long unstructured regions [34, 35]. Some of these disorders involve a conversion of the functional state of specific proteins into an aggregate state that accumulate as fibrils. These misfolded proteins deposit as amyloid fibrils and are called amyloidogenic proteins. Generally, the amyloidogenic proteins can be divided in two types: 1) prion proteins that present a well-defined structure with partially disordered region(s) on the molecule; and 2) proteins that show changes over the entire chain such as Amyloid-β (Aβ), tau, α-synclein and huntingtin protein. This behaviour of *misfolding* and aggregation of IDRs/IDPs are especially common in neurodegenerative diseases such as Alzheimer's [36, 37], Parkinson's [38] and Huntington's disease [39].

1.2 Amyloid-β Peptide and Alzheimer's Disease

Alzheimer's disease (AD) is one of the most prevalent neurodegenerative disease that causes progressive loss of cognitive and functional abilities. Many hypotheses about AD have been developed, among which the amyloid cascade hypothesis is widely accepted. This hypothesis was first proposed by Hardy and Higgins [40] which proposed Aβ as the central figure in the overall disease mechanism. They also suggested that the mis-metabolism of amyloid precursor protein (APP) (see Section 1.3.1) and Aβ accumulation were the primary events in AD. The full hypothesis has been described as a cascade of Aβ deposition, tau phosphorylation, neurofibrillary tangles formation, neuronal death and dementia. Many preclinical and clinial studies provided steady experimental data to increasingly support the hypothesis. Today, the presence of Aβ deposits or senile plaques on the hippocampus and the overlying cortical regions is considered as one of the definitive features of AD [41, 42]. The current strategies for AD treatment based on the amyloid hypothesis are mainly targeting the inhibition of Aβ aggregation [43].

Although many studies tried to confirm Aβ's central role in AD pathogenesis, the exact mechanisms are still unclear [36, 44, 45]. A huge number of studies showed inconsistent evidence [46] against the amyloid cascade hypothesis, which denies the direct correlation between Aβ accumulation and neuronal loss and cognitive decline [47]. The amyloid hypothesis was the most tested one for AD. Percentage of clinical trials based on this hypothesis reached to 22.3% in 2019 [43]. Sadly, there is still no significant progress on clinical trials related to amyloid hypothesis to the present day. The reasons for these controversial results are still under debate [48, 49]. One of the explanations is that this hypothesis ignores physiological functions of Aβ. The presence of the peptide is throughout the lifespan, and it has been found in all vertebrates examined up to 2018 [50]. Also, its molecular sequence shows a high degree of conservation. A variety of physiological roles of Aβ peptide have been proposed including antimicrobial properties [51], helping recovery from brain injury [52], prevention of blood-brain barrier leakage [53] *etc.* Further details can be seen in current reviews [50, 54]. Some valid evidence demonstrated the biological significance of Aβ peptide and its possible involvement in the protection and repair of central nervous system, which makes the relationship between the peptide and AD even

more complicated.

1.3 Amyloid-β Peptide

As Aβ peptides are the major components of senile plaques and the complicated correlation between the peptide and AD, it is essential to understand their structure and biochemical properties at the molecular level. Aβ, a 39- to 43- residue peptide, is produced by the sequential proteolytic cleavages of a much larger protein, the amyloid precursor protein (APP) [55]. The most abundant species are Aβ40 and Aβ42, and the latter one is considered to be more neurotoxic. The sequences of Aβ40 and Aβ42 are shown in Figure 1.2. The peptide has an intra-membrane region ranging from G29 to V40 or A42 with all hydrophobic residues, which also reduces its solubility. Histidine 14 shows significance in formation and stabilization of a loop region in the Aβ42 fibrilization [56]. Six glycine residues of the chain have been reported to be crucial for its chemical and physical properties [56, 57]. Particular regions on the 42-residue peptide ((^{16}KLVFFA^{21} and ^{37}GGVVIA^{42}) are described as two key amyloid-forming segments [58]. Despite the large number of studies on Aβ fragments that contribute further complex but sometimes inconsistent data on its properties, a more accurate and detailed model of the full-length peptide is needed.

1.3.1 Amyloid Precursor Protein

APP is a 770-residue post-translationally modified transmembrane protein with a large extracellular domain and a short cytoplasmic domain. It contains several conserved motifs in both the extracellular region and intracellular domain. Interestingly, the Aβ sequence is not conserved and unique to APP. Two pathways have been described in the reaction of proteolytic cleavage of APP by various enzymes (see Figure 1.2). In the dominant non-amyloidogenic pathway, α-secretase cleaves in the middle of the segment containing Aβ of the extracellular domain, releasing the N-terminal fragment into the medium. The C-terminal fragment that remains anchored to the membrane, is further cleaved by a γ-secretase complex inside the transmembrane domain, producing APP intracellular domain and Aβ fragments (Aβ(17-40), Aβ(17-42)). In the amyloidogenic pathway, APP is cleaved by a β-secretase in the extracellular region at a specific position followed by a second cleavage within the transmembrane

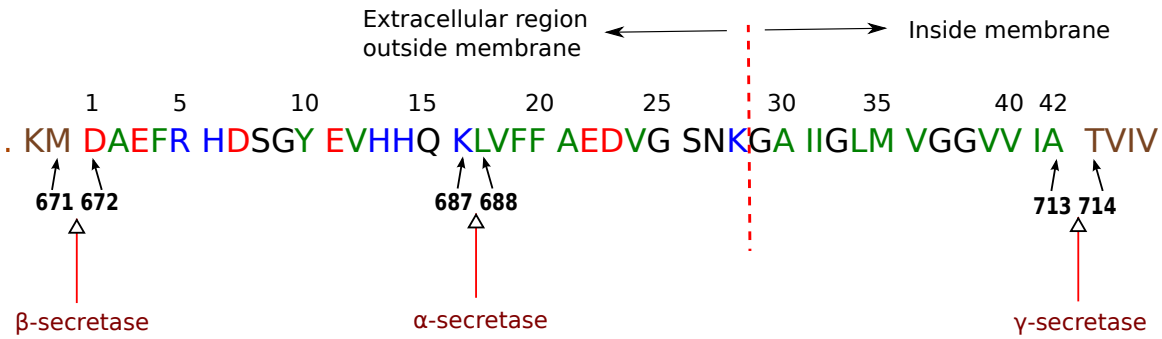

Figure 1.2: The amino acids sequence of Aβ (upper numbers) and the section of the APP (lower numbers). Negatively and positive charged residues are red and blue, hydrophobic residues are shown in green and other residues are in black. Residues that are out of the regions of Aβ are coloured in brown.

domain by γ-secretase, yielding a mixture of Aβ components. Among them, Aβ40 and Aβ42 are the dominant species. Compared to Aβ40, Aβ42 is less soluble and has higher propensity to self-aggregation.

1.3.2 Amyloid Aggregation

Although proteins fold into 3D structures intrinsically, the process is complex and susceptible to errors [59]. Generally, misfolded or partially folded proteins are degraded by quality control systems (Figure 1.3) such as the proteasome and autophagy [60,61]. Dysfunction of this system or overloading caused by the protein aggregation leads to protein *misfolding* and further fibrilization in turn. This phenomenon is much more widespread in medicine and biology. Some amyloid aggregates were reported to deliver physiological functions in particular biologicial systems in bacteria, plants and mammals [62]. Our interests in this area arise from considerations that understanding protein *misfolding* and aggregation may (1) help to elucidate the physiochemical features of proteins folding; and (2) is expected to shed light on the pathogenesis of neurodegenerative diseases.

Two mechanisms are usually proposed for amyloid aggregation: nucleated polymerization [63, 64] and nucleated conformational conversion [65]. Their common features on aggregation pathways are described in three phases: a lag phase, an exponential growth phase and a plateau regime (see Figure 1.3). The formation of a stable oligomeric nucleus that forms from misfolded monomers occurs during the lag phase. To be specific, the *misfolding* of Aβ usually refers to a conformational transition from membrane-bound α-helix structure to an aggregation-prone-state with largely β character [66]. Chapter 3 focuses on investigating this mechanistic pathway of conformational conversion at a molecular level using MD. The kinetics of the lag phase depends on the monomer concentration and the presence of aggregation nuclei. The formation of the nuclei is thermodynamically unfavourable and thus a rate-limiting step. The assembling rate starts to increase exponentially once stable oligomeric nuclei are generated. The system enters the saturation phase if the monomer concentration becomes significantly low. Addition of nuclei seeds in the rate-limiting step can shorten the lag time and activate the production of mature fibrils.

Previous proposed kinetics models [67–69] represent the aggregation process with

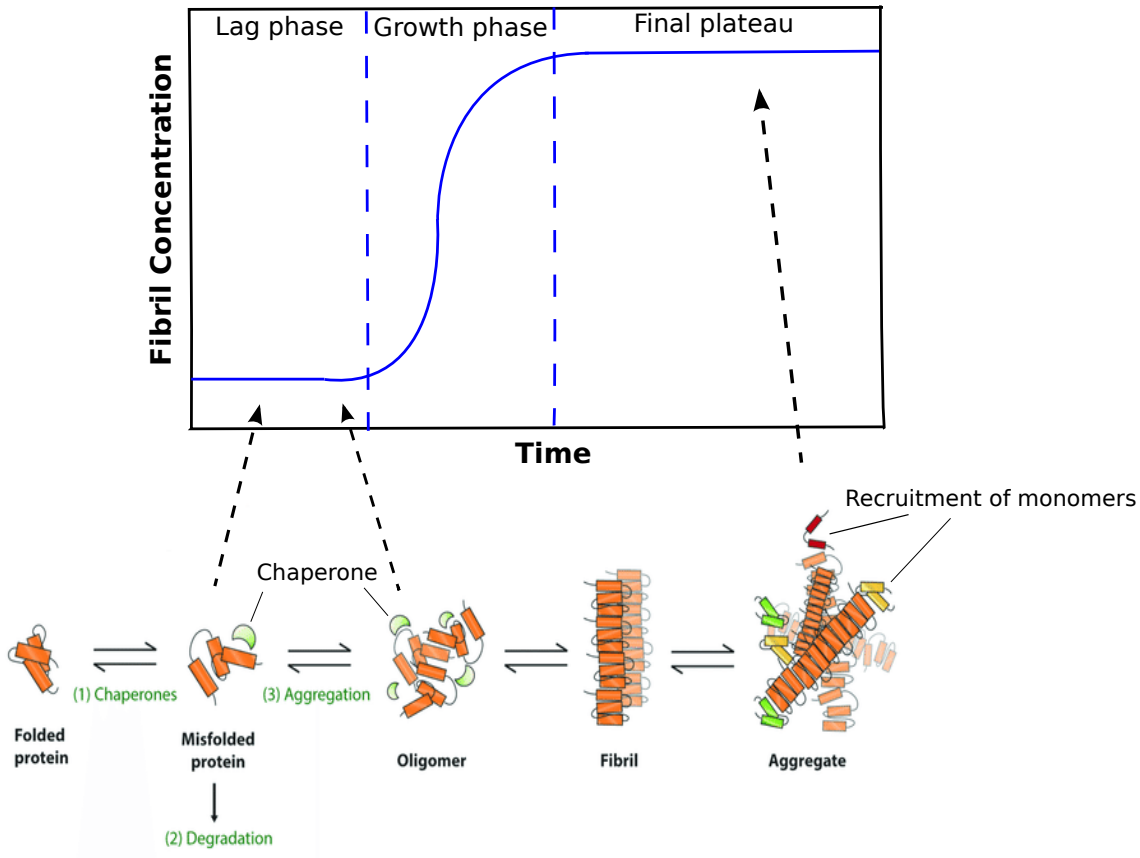

Figure 1.3: Proposed mechanism of amyloid aggregation includes three phases: a lag phase starts with formation of oligomers from misfolded monomers, a fibril elongation phase and a final plateau with highly ordered structured fibrils formed. A misfolded protein can be (1) refolded or (2) degraded, or (3) undergo aggregation [76].

a sigmiodal curve to model the observed growth of fibrils. These models provide a molecular-level description including primary nucleation, secondary nucleation and elongation. Primary nucleation involves only monomers in solution [70] or the attachment of monomers on the foreign surface [71–73]. The nucleus is formed as a secondary nucleation on the surface of the existing aggregates. This can be treated as an *induced-fit* reaction that is catalyzed by the parent aggregate seed with the same type of monomeric building block. Sometimes, a fragmentation step is also included in these kinetics models, where fibrils break into pieces exposing more elongation ends [74, 75].

It is challenging to study the amyloid aggregation mechanism for a few reasons. First, the formation of oligomeric nuclei limits the rates of fibrilization. Furthermore, off-pathway non-fibrillar aggregates can be formed competing with on-pathway nucleated polymerization [77, 78], which is common in experimental studies. Also, amyloid proteins are sensitive to different conditions (such as solution or surface system), leading to slow primary nucleation.

1.3.3 The Definition of Aβ Misfolding

As previously stated, the amyloid cascade hypothesis overshadows the evidence that Aβ serves several important biologcal functions. It is necessary to discuss the definition of Aβ *misfolding* in the context of this work. Many results show that Aβ has antimicrobial properties. The monomer may capture and perforate microbes with its hairpin loop, while oligomers and aggregates may behave as an extracellular trap to immobile microbes. Fibrils could insert into cell membrane of microbes to create passages which allow solutes through the membrane, leading to the death of microbes [79]. Once the soluble monomer is released to the solution, in each stages of the following conformational changes plus self-aggregation, a monomer or aggregated species may serve as a particular function. Considering this, Aβ may not misfold but fold into a variety of functional states with different physiological purposes. It is reasonable to assume that the whole aggregation process described in the nucleated polymerization model or the nucleated conformational conversion model can be treated as a complete folding progression of the peptide with multiple biologicial functions. In the rest of this chapter, *misfolding* will be mentioned to indicate the initial stages of

a series conformational changes in the aggregation models. More details and a more appropriated definition will be discussed in chapter 3.

1.3.4 Structure of the Aβ Peptide Monomer

NMR and MD simulations are two major techniques used for structural elucidation of Aβ monomers. Neither Aβ40 nor Aβ42 crystallizes thus there is no X-ray structure of the monomer in any form. In a membrane-mimicking environment, both of the peptides show high proportions of α-helix conformation on the chain. Aβ40 displays mainly an α-helix conformation over the chain with a loop between G25-N27 (PDB ID: 1BA4) [80] (Figure 1.4A). The N-terminus is unstructured ranging from D1-H14. Aβ42 adopts two α-helix regions over residues S8-V25 and L28-V38, separated by a β-turn (PDB ID: 1IYT) [81] (Figure 1.4C). In water-containing solutions, there is a loss of helical content on both peptides. Aβ40 forms a 3_{10}-helix between H13-D23 in a complete aqueous environment with no structures neither on N- nor C-terminus (PDB ID: 2LFM) [82] (Figure 1.4B). There is a decreased proportion of helix structure of Aβ42 in 70% aqueous solution (PDB ID: 1Z0Q) (Figure 1.3.4D). The region between W10-D23 remains in α-helix, while the portion between L34-G38 contains helical structure in some degree. The turn region mentioned in the structure solved in membrane-mimicking environment remains present [83].

1.3.5 Structure of Aβ Oligomers

Aβ oligomers are widely regarded as the neurotoxic and pathogenic form of Aβ [84–87]. They appear as transient species that undergo conformational conversion from their monomeric precursors to more massive and stable fibrils. Their heterogeneity in structure and size [88–90] and low kinetic stability make it difficult to elucidate atomic-level structures and assembling pathways. Several structural models of Aβ oligomers have been proposed based on high-resolution X-ray crystallography and NMR data. Tay *et al.* [91], Yu *et al.* [92] and Amhed *et al.* [89] agree in a general sense with predictions of oligomer structures of Aβ42 with mixed parallel and anti-parallel β-sheet structure using NMR, which is different from reported fibril structure containing only a parallel β-sheet structure [93]. However, there are also

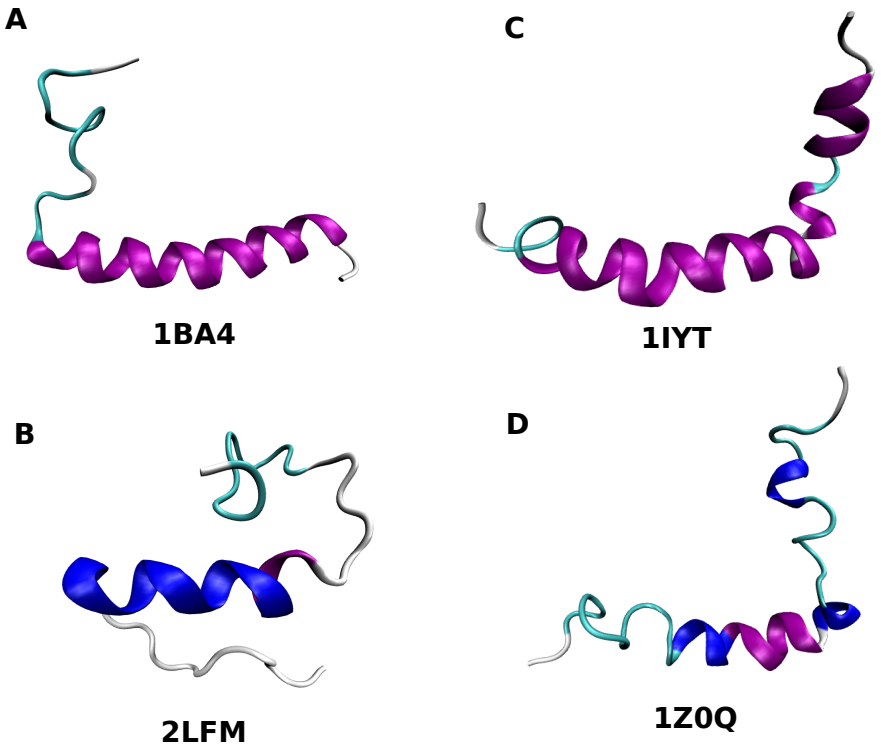

Figure 1.4: A list of structures of full-length Aβ: A) Aβ40 monomer structure in water-micelle like environment (PDB: 1BA4); B) Aβ40 monomer structure in pure water (PDB: 2LFM); C) Aβ42 monomer structure (PDB: 1IYT) in a water/HFIP mixture of 20:80 (v/v); D) Aβ42 monomer structure (PDB: 1Z0Q) in a water/HFIP mixture of 70:30 (v/v). All structures are shown in cartoon representation using Visual Molecular Dynamics (VMD) software version 1.9.4a37 [96]. The peptides are shown in the orientation of N-terminus on the left and C-terminus on the right. Secondary structure are shown in different colours, in which α-helix is in purple, 3_{10}-helix is blue, turn is cyan, and coil is white.

inconsistency between their results. Yu *et al* predicts the parallel intermolecular arrangement between β-strands formed on the region of L34 and V40, while Tay *et al.* indicates the intermolecular contacts are between F19-I31. Amhed *et al.* proposed a model for pentameric disc-shape oligomers without in-register parallel β-sheets. Some oligomeric structures of Aβ fragments are also reported. Pham *et al.* [94] presented the crystallography structure of oligomeric assemblies formed by the Aβ (15-23) peptides. Spencer *et al.* [95] proposed oligomeric formation of Aβ(17-36) using X-ray crystallography, including the forms of trimers and other higher-order oligomers.

1.3.6 Aβ Fibrils

Amyloid fibrils are the most abundant Aβ aggregates in AD brain tissues. They are insoluble and non-crystalline, which makes them incompatible with X-ray crystallography and solution NMR studies. *In vitro*, other techniques such as X-ray diffraction [97], solid-state NMR (ss-NMR) [98], cryo-EM [99], transmission electron microscopy (TEM) [100], atomic force microscopy (AFM) [101] and MD simulations [102] were used to study Aβ fibril structures. These have provided on both structural insights and information on the mechanisms of stacking and elongation pathways.

Early X-ray diffraction studies first found that amyloid fibrils contain cross-β structures, which is the fundamental property for the future development of structural models [103–105]. In ss-NMR and other studies [106, 107], preliminary model of Aβ40 fibrils were described as a *U-shaped* or hairpin conformation containing separated in-register parallel β-sheets. A hydrophobic core is formed between two β-sheets. Four stabilizing factors support this topology; 1) hydrogen bonding between the backbone amide groups of adjacent chains, 2) Van der Waal interactions between top and bottom β-sheets within the hydrophobic core, 3) increased entropy of water molecules from the hydrophobic effect around the β-sheets, and 4) intramolecular salt bridge between D23 and L28.

Different Shapes of Aβ Fibrils

Three main types of fibril structures for Aβ42 are *U-shaped*, *S-shaped* and the most recently published *LS-shaped* structures. By using hydrogen/deuterium-exchange

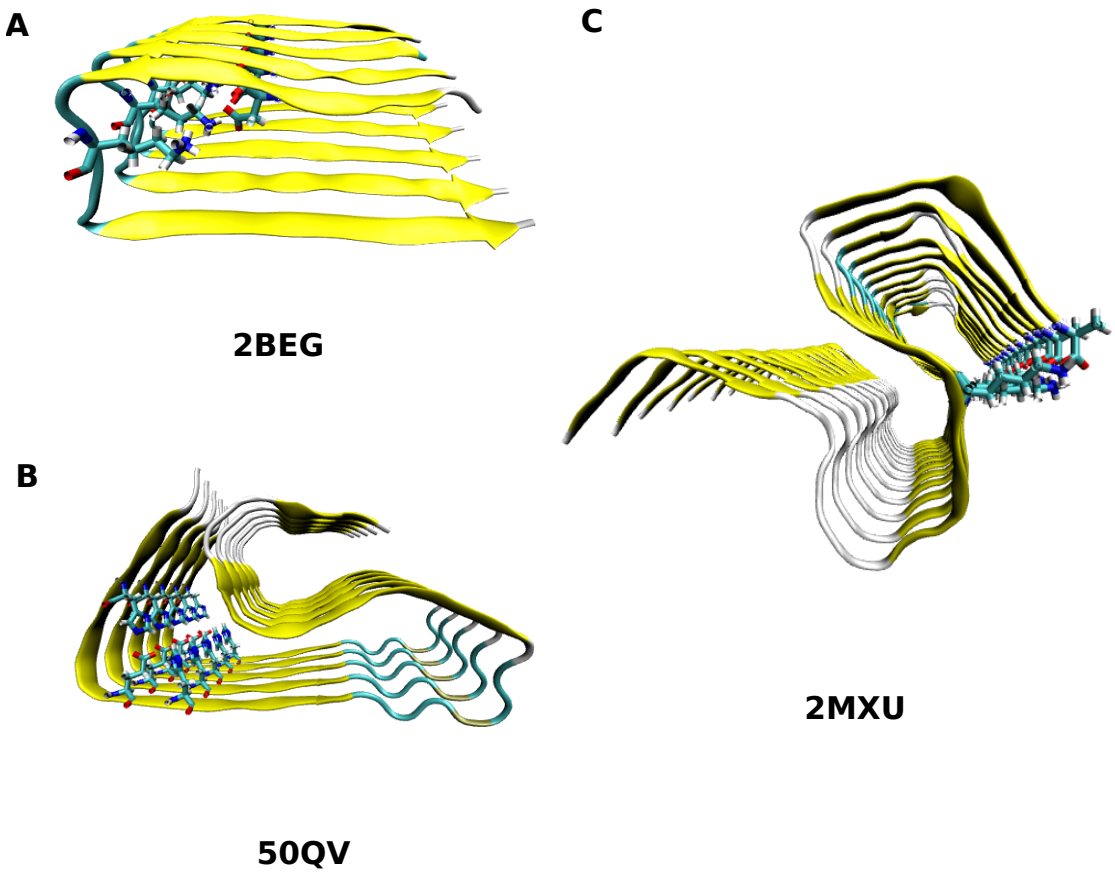

Figure 1.5: Examples of three main types of Aβ fibril structures, which are (A) *U-shaped* (PDB: 2BEG), (B) *S-shaped* (PDB: 2MXU) and (C) *LS-shaped* (PDB: 5OQV) structures. All structures are shown in cartoon representation using VMD [96]. Secondary structures are shown in colours, in which extended-β is in yellow, bridged-β is in tan, and coil is in white. Corresponding salt bridges are shown in bond representation.

NMR, Lührs *et al.* [93] reported a *U-shaped* topology containing two intermolecular, parallel, in-register β-sheets in regions of V18-S26 and I31-A42. Unlike the Aβ40 fibril structure, Aβ42 fibrils are stabilized by intermolecular domain swapping side-chain interactions. This includes an intermolecular salt bridge between D23 and L28 (Figure 1.5A). Three recent NMR studies [108–110] have revealed an *S-shaped* conformation of Aβ42 fibrils with three intramolecular β-sheets (β1(12-18), β2(24-33) and β3(36-40)) and an intramolecular salt bridge formed between L28 and A42 carboxyl terminus (Figure 1.5B). The *LS-shaped* structure was determined by cryo-EM combined with ss-NMR, in which the N-terminus is *L-shaped* and C-terminus is *S-shaped* [58]. The subunit conformation is stabilized by three hydrophobic clusters: 1) A2, V36, F4 and L34, 2) L17, I31 and F19, and 3) A30, I32, M35 and V40. It contains three salt bridges between D1-L28, D7 and R5, and E11 and H6 and H13, particularly the last one is considered to stabilize the kink in the N-terminus of the β-sheet around Y10 (Figure 1.5C).

Different Arrangement of the Aβ Fibril Interface

The basic propagating unit of Aβ40 could be a dimer with 2-fold symmetry or a trimer with 3-fold symmetry, while Aβ42 fibril usually propagates over a 2-fold symmetric dimer structure. In the 2-fold symmetry, Aβ fibrils contain two symmetric strands that form separate β-sheets in a double layered cross-β motifs. Both *U-shaped* and *S-shaped* fibrils could be elongated on the basis on a planar 2-fold symmetric unit. Paravastu *et al.* [107] elucidated that residue M35 plays an important role in stabilizing the 2-fold symmetric Aβ40 fibril protofilament via the interaction along and across the fibril axis (Figure 1.6A). Colvin *et al.* [108] and Wälti [109] observed different *S-shaped* Aβ42 fibril structures, however, both of them arrange in the 2-fold symmetry. In the first structure, intermolecular contacts between Q15, L17, L34 and M35 are considered to maintain the 2-fold symmetry (Figure 1.6B). The dimeric system of the second structure is stabilized by two hydrophobic cores involves M35 and either Q15 or L17 (Figure 1.6C). In the *LS-shaped* Aβ42 fibril structure, salt bridges that are formed between D1 and L28 from different units stabilize its 2-fold symmetry (Figure 1.6D). Different from the *S-shaped* 2-fold symmetric unit, the *LS-shaped* system is not a true dimer. The dimeric interface is not planar but has a staggered

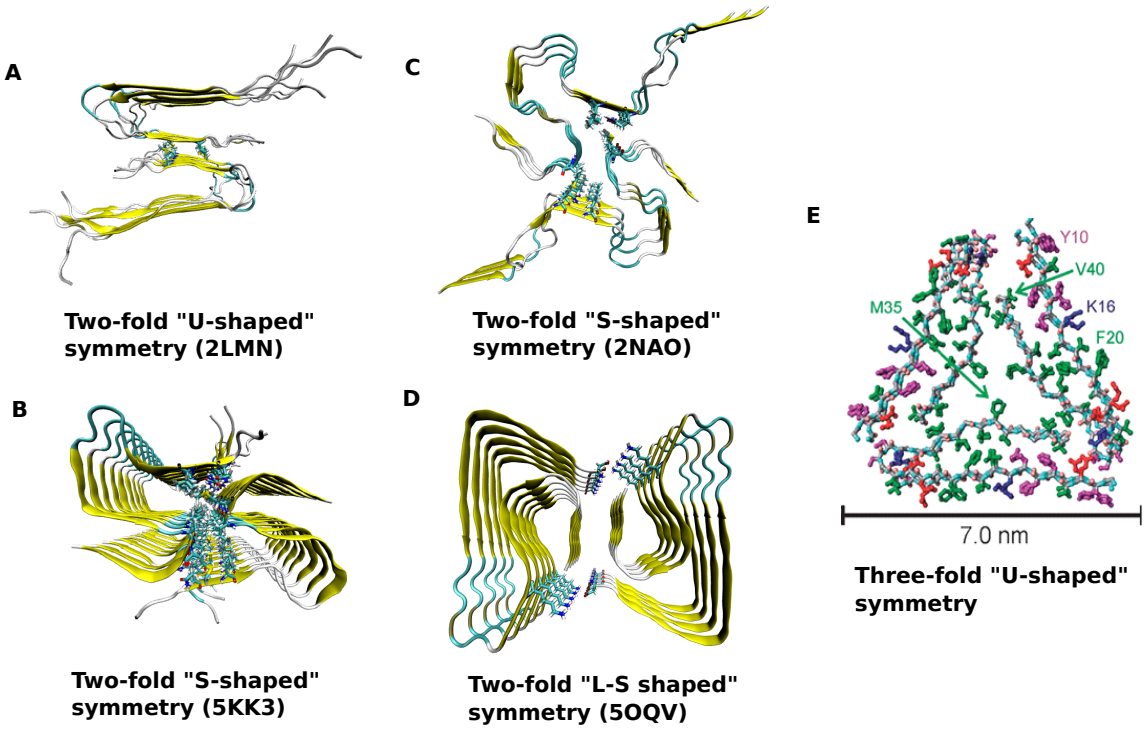

Figure 1.6: Fibril structures that were solved *in vitro*, which are (A) 2-fold symmetry of *U-shaped* (PDB: 2LMN), (B) 2-fold symmetry of *S-shaped* (PDB: 5KK3), (C) another 2-fold symmetry of *S-shaped* (PDB: 2NAO), (D) 2-fold symmetry of *LS-shaped* (PDB: 5OQV) and (E) a proposed structural model of 3-fold *U-shaped*. All structures are shown in cartoon representation using VMD [96]. Secondary structures are shown in colours, in which extended-β is in yellow, bridged-β is in tan, and coil is in white. Corresponding key residue contacts are shown in bond representation.

arrangement. One chain within the dimer rises along the fibril axis from the N- to the C-terminus, generating grooves and curbs at the interface of the fibril ends. During the elongation, each monomer binds alternatively to the curb, where two identical binding sites would be seen. A 3-fold symmetric *U-shaped* Aβ40 fibril unit consists of three β-strands that form separate β-sheets in a triangular cross-β motif. Paravastu *et al.* [107] elucidated that this model contains an interior channel induced by G33 and G37, and M35 sidechains are observed to point to the center of the fibril (Figure 1.6E). Both in the 2-fold and 3-fold symmetry, the protofilaments could aggregate along the fibril growth direction and have a helical symmetry along the axis.

Structure of Aβ Fibrils in Brain Tissues

As Aβ peptides aggregate into polymorphic fibrils *in vitro* and various forms of oligomers and protofibrillar aggregates under different conditions, it is important to clarify which structure develops in brain tissues. Lu *et al.* [111] investigated Aβ40 fibril structures derived from brain tissues of two AD patients using ss-NMR and electron microscopy. Surprisingly, experimental data indicated a single predominant Aβ40 fibril structure in different regions of the cerebral cortex from each patient. However, one structure showed significant differences from the other. The structure model based on Aβ40 fibril from patient I developed (Figure 1.7A), which implied the protofilaments in 3-fold symmetry. Conformational features in this structure include a twist at region of F19-D23 and a kink at G33 that allows sidechains of I32 and L34 to establish intermolecular interactions with different sets of Aβ40 molecules. Key intermolecular residue contacts include F4-V24, R5-V24, D7-S26, S8-V24, A30-V40 and I32-V39. Compared to the 3-fold symmetric Aβ40 fibril structure *in vitro*, this structure is apparently different and more complex. A recent cryo-EM study [112] reported discovery of multiple species of Aβ fibrils (Figure 1.7B and 1.7C) that were derived from brain tissues of three AD patients with a low ratio of Aβ42. All the brain-derived fibrils are right-hand twisted and contain a similarly internal structure of protofilaments. Their subunits fold in a C-shape that is in a different manner compared to previous Aβ fibril structures identified *in vitro*.

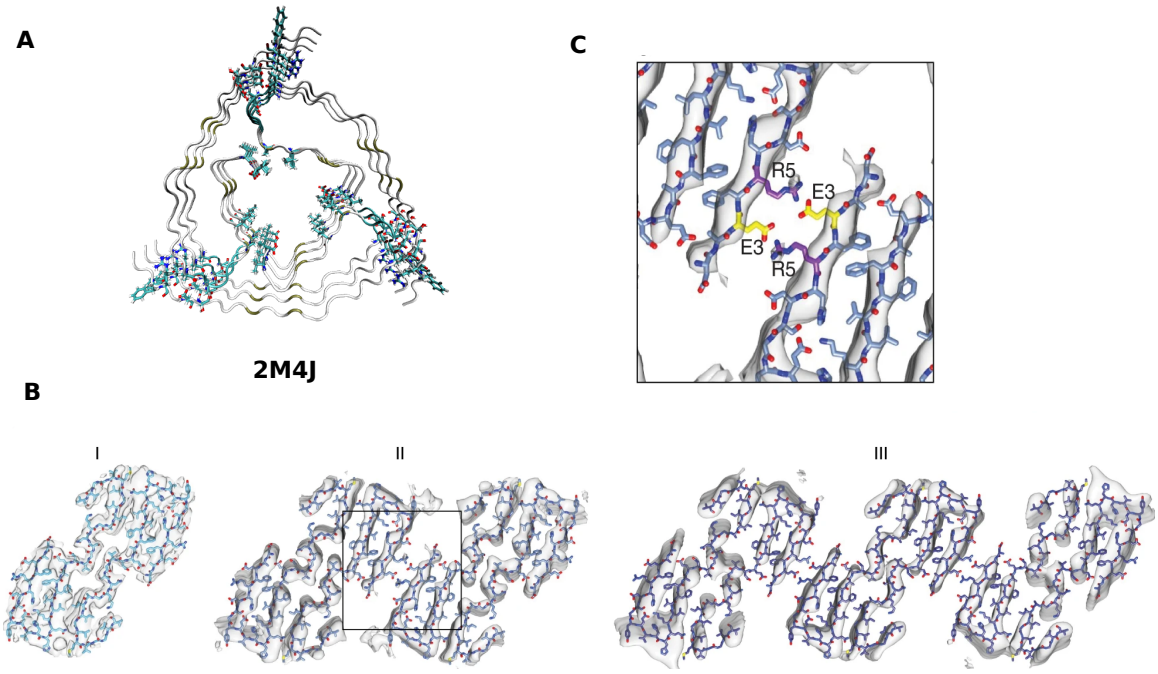

Figure 1.7: Two fibril structures that were observed *in vivo*, which are (A) 3-fold symmetry of Aβ40 fibril (PDB: 2M4J), (B) Cross-sectional density of fibril morphologies I-III (gray) superimposed with the molecular model obtained with morphology I [112], (C) A close look between the protofilament interface in morphology II [112]. Structure (A) is shown in cartoon representation using VMD [96]. Secondary structures are shown in colours, in which extended-β is in yellow, bridged-β is in tan, and coil is in white. Corresponding key residue contacts are shown in the bond representation.

1.4 Molecular Dynamics

Molecular dynamics is a computer simulation method to explore conformational space for macromolecules. It is a powerful tool for predicting structures of systems that are hard to study with experimental techniques. The method has also been applied to study other fundamental problems of in biochemistry, including protein folding, protein-ligand binding and protein self-aggregation. Early studies on Aβ structures focused on exploring the structural role of short Aβ segments. This was due to the high computational demands to sample the conformational ensemble of the full-length of Aβ peptide. Coarse-grained and atomistic models were used to investigate the amyloid aggregation phenomenon because of the long timescale of the process. With advance in fine-tuning calculations, parallelization, and widely availability of graphical processing units (GPUs), the performance of MD simulations have been largely improved. It is now possible to study the full-length Aβ peptide using all-atom MD simulations in explicit solvent at various temperatures and pH conditions. Furthermore, aggregation studies of around 20 amyloidogenic peptides by the means of explicit solvent atomistic simulations can reach to micro- or even millisecond scales. This work simulated a total of 15.2 μs of the Aβ42 monomer in different environments.

1.4.1 Challenges and Limitations of MD Simulations on Conformational Space Exploration of the Aβ Peptide and Conformational Transition in Amyloid Aggregation

It is challenging to explore the conformational space of monomeric and oligomeric Aβ using MD simulations. First, it is not guaranteed that the broad conformational space of the Aβ42 could be thoroughly explored via a single simulation [113]. Enhanced sampling methods such as replica exchange molecular dynamics (REMD) will provide new insights into the structural transitions of the full-length Aβ peptide. Also, these simulations do not reproduce accurate and heterogeneous physiological conditions. The heterogeneity of monomers and oligomers and polymorphism of fibril structures are dependent on the conditions. *In vivo*, amyloid formation is affected by both intracellular and extracellular environments, such as oxidative stress, fluctuations in temperature, pH, and metal ion concentrations [114]. Third, the conformational

conversion of Aβ operates on a huge range of length- and timescale [115], which is still challenging for current MD simulation ability. For example, the aggregation rate of the initial stages of oligomerization is on a second-level timescale. Lastly, the accuracy of force fields for disordered proteins is less validated. The accuracy of classical MD simulations exploring protein structure and dynamics depends on the accuracy of the force field selected including parameters for protein, water, ions and other molecules. Modern force fields have been refined on well-folded proteins, however, those for IDPs are not well optimized due to the difficulty in obtaining experimental data and models for these sytems. Previous MD studies reported diverse structural features of the Aβ peptide because of different selections of structural input, applied force fields, and water models. Several benchmarking studies of force fields for Aβ peptide or IDPs are provided elsewhere [116–118].

1.4.2 Analysis of MD simulations

Many MD trajectory analysis softwares [119–121] were introduced for the last decade. These softwares take advantage of techniques in math, statistics and computer science to process large data sets, such as interactive computing [122], clustering methods [123], dimensionality reduction [124], network analysis [125], machine learning [126] and visualization [127]. However, very few trajectory analyses were developed specially for IDPs. In this thesis, we are focusing on capturing the major structural characteristics in the folding of Aβ and discovering possible common structural features. We designed a protocol using clustering methods, principal component analysis, and network analysis.

1.5 Recent Progress on MD studies Related to Aβ Monomer Structures

In this work, we use all-atom MD to study the full-length Aβ monomer (namely Aβ40 and Aβ42) structures. MD studies of oligomer structures, fibril structures and aggregation mechanisms will not be discussed, but some recent papers are listed here [128–136]. Both classic MD and REMD are applied to investigate the monomer structural features in different conditions (solvents, temperature and pH). REMD is an enhanced sampling method, where several copies (replicas) of the system are simulated in parallel using MD simulations at different temperature. The number of

replicas depends on the total number of atoms present in the system including the solvent molecules, ions and other molecules. Thus, the simulation can easily become computationally expensive.

To understand the conformational conversion of Aβ monomers in *misfolding*, classic MD simulations were performed both in the aqueous and membrane environment. Luttmann *et al.* [137] observed the formation of a turn region at A21-G33 and high flexibility of the N-terminus in the aqueous environment. Xu *et al.* [138] studied Aβ40 using all-atom MD in both aqueous and biomembrane environment. A helix to coil transition was observed in the full conformational change of α-helix-β-sheet in an aqueous solution. Four glycines (G25, G29, G33 and G37) were considered to play important roles to convert the helix-conformation to β-sheet via a β-rich intermediate. In the work combining REMD and dihedral principal component analysis (dPCA), Xu *et al.* proposed the structural characteristics of the Aβ42 in an aqueous solution. The chain contains 60%-80% coils, 10%-20% β-strands and a significantly low content of helices (<10%).

The structures were compared under different environmental conditions. For example, Flock *et al.* [139] researched on the *misfolding* of the peptide in the aqueous solution at different temperature and pH, suggesting that at least one of the helices of the membrane-bound structure was rapidly converted into a coil or β-strands. The influence of various solvents on Aβ42 conformation was discussed by Yang *et al.* [140]. Helix regions on the chain are maintained in the low polarity solvents including hexafluoroisopropanol (HFIP) and 2,2,2-trifluoroethanol (TFE), while an α-helix to β-sheet conformational change was observed in aqueous solution.

The dynamic behaviour of two termini of the peptide was also of interest to researchers. Valerio *et al.* [141] proposed the importance of hydrophobicity, flexibility and mobility of N-terminus for the monomer to remain in a misfolded form. Another MD study revealed that the first 10 residues of N-terminus for both Aβ40 and Aβ42 are unstructured. A reduction of electrostatic repulsion between the two terminal regions aids the formation of β-sheet of Aβ42 [142]. In an REMD study, Miyashita *et al.* [143] discovered distinct preferences of the termini in different environments that C-terminal region favours membrane and N-terminal region prefers aqueous solution.

The structural comparison between Aβ40 and Aβ42 were discussed in many

REMD studies. The two extra residues on the C-terminus of Aβ42 are considered to be responsible for a higher propensity for aggregation. This was confirmed by Yang *et al.* based on the observation of higher stability of the β-structures of Aβ42 compared to that of Aβ40. Another extended REMD study performed by Sgourakis *et al.* [144] provided an explanation that the C-terminus of Aβ42 is more structured than that of Aβ40. In a recent study, Song *et al.* [145] revealed the heterogeneous conformation ensemble of Aβ40 and Aβ42 at equilibrium by performing all-atom MD in water. Two major conformations, collapsed and extended, were observed that both adopted around 35% β-strand and about 60% unstructured coils.

Other studies also provided interesting observations on the monomer structures under different circumstances. Rosenman *et al.* [116] explored the structural properties of Aβ40 using three force fields in water, which showed similar results. There was also a proposed acceleration mechanism of the oligomer formation at the interface in a combination study of MD simulations and NMR experiments [146]. They observed that a stable attachment of the Aβ40 peptide on the hydrophilic/hydrophobic interface. MD studies related to mutations on the peptide were also examined through all-atom MD simulations. A detailed review is provided elsewhere [147].

Based on various MD studies of full-length Aβ monomer structures, some common features can be summarized. First, there is a conformational conversion of α-helix to β-sheet observed in the *misfolding* process. The N-terminus is usually observed to be coils, whilst the C-terminus is more structured in an aqueous solution. The two termini prefer membrane and aqueous environment respectively. The C-terminus plays an essential role for the stability of the β-structure in water. To conclude, the peptide chain could be treated as four regions, which are the unstructured N-terminus (D1-Y10), a helical region (Y10-A21), a turn region (around A21-A30), and a β-structured C-terminus (A30-A42).

1.6 Thesis Objectives and Organization

Proteins are fascinating molecular machines that are able to perform operations with other biomolecules for basic cellular functions in the requirement of converting chemical energy into mechanical work. To serve as protein machines, it was long assumed that a protein is necessary to fold properly into a specific 3D structure. The fact

that IDPs without unique structures play crucial roles in biological processes challenges our understanding of protein science. Misfolding diseases including numerous neurodegenerative diseases and amyloidoses are considered to associated with the aggregation of specific IDPs from their soluble functional monomeric states into stable, highly ordered, filamentous amyloid fibrils.

Alzheimer's disease is a debilitating neurodegenerative disorder without a known clear mechanistic pathology or effective treatment. The most widely accepted amyloid cascade hypothesis and various studies based on this hypothesis theory suggest that there is a complex relationship between the accumulation of amyloid fibrils and the disease. Many questions have been raised regarding the validity of this hypothesis since there are more than 200 failed clinical trials over the past three decades of drugs targeting this peptide. The molecular dynamics and structures of the peptide itself as well as the process of aggregation remain unclear. Large-scale conformational changes, structural heterogeneity and environmental sentitivity all contribute to the complexity of this system.

Out of our interests of unique conformational behaviors of IDPs in various environmental conditions, we use Aβ42 as the model to study IDPs via MD simulations. First, we simulate the molecular mechanism of *misfolding*. Although the observation of a conformational conversion was reported from both experimental and computational studies, there is a lack of explicit definition of *misfolding* and description of the complete process in molecular details. Due to limited computational power and the absence of a parent fibril, we designed a simulation model coined as a *double-ended search* in Chapter 3 to recover the *misfolding* process. On the other hand, MD simulations generate huge data sets so that a proper trajectory analysis specially for IDPs with large-scale conformational changes (Aβ42 in this case) is required. I also proposed an analysing protocol, combining network analysis, dimensionality reduction, clustering method and self-designed visualization to deal with the data from classic and simulated annealing (SA) MD simulations. The results showed the presence of interconversion between collapsed and extended conformations and provided identification of possible metastable species in the *misfolding* events. A summarized mechanistic pathway is provided and key structural features are compared with previously published results.

Considering the existence of heterogeneous monomer and fibril subunit structures, we carried out MD simulations starting from two monomeric 42-residue structures and 5 fibril subunits to sample the conformational space of the monomeric Aβ42. To enhance sampling, we also performed replicas of several original MD simulations and applied simulated annealing for two of the starting structures. According to a total of 14 MD simulations, we observed the existence of consistent structural features through all trajectories. In Chapter 4, we proposed that there are common geometric features in Aβ42 even though the soluble peptide is intrinsically disordered. Principal component analysis (PCA) was employed to visualize conformational landscapes and free energy maps. By simulating the monomers in Chapter 3, we realized that the dynamic conformational space of the Aβ42 peptide is large. Instead of creating the landscape on one simulation, we combined all simulations to extend to a global view of the conformational space. We also took advantage of multiple self-designed methods for characterizing this space. Identified common structural features in water are concluded in a comprehensive visualization on a proposed structural model.

In Chapter 5, we tested the environmental sensitivity of the Aβ42 monomer by adding ethanol to the aqueous environment. In the modelling within the mixed solvent, we observed different dynamic behaviors of the peptide on the ethanol/water interface than that in pure water. The adsorption kinetics, dynamics and stability were discussed, which prompts further investigations of Aβ42 interacting with the fluid or membrane interface.

Chapter 2

Methods: Molecular Modelling to Molecular Simulation

2.1 Molecular Mechanics

One of the major challenges for studying chemical systems using computational models is the accurate description of the interaction between atoms and molecules. In principle, quantum mechanics (QM) can be utilized to predict the time evolution of a molecular system by solving the time-dependent Schödinger's equation with no input of empirical parameters. It is generally applicable to achieve first principle in the description of the motion of atomic nuclei and surrounding electrons [148]. In reality, this equation have never been solved exactly for any chemical systems except for the hydrogen atoms. Although many QM-based methods have been developed [149–152], the prohibitive computational costs dramatically limits the size of the system that can be modeled.

Molecular mechanics (MM) allows the modeling of large biomolecular complexes. It describes a molecular system using principles of classic mechanics, where a molecule is treated as a collection of point masses (nuclei) connected by spings (bonds). This method calculates the structure and energy of molecules based on nulear motions because it is assumed that the optimum distribution of eletrons can be found once the positions of the nuclei are located. Under this assumption, nuclear motions such as vibrations and rotations are treated separately from electrons. The theoretical basis of this assumption is rooted in QM via the Born-Oppenheimer's approximation that nulei are much heavier than electrons so that the mass of electrons is considered negligible small. With this simplification, the potential energy surface of a molecular system described by the potential energy function in MM is in its electronic ground state. As electronics are not treated, MM is unable to describe the formation and cleavage of covalent chemical bonds.

2.2 Force Fields for Biomolecules

The potential energy function of MM is a combination of simple potential formulations such as bond lengths, bond angle, non-bonded interactions. All of the contants in these potential energy terms are obtained from data determined by QM calculations and physical measurements of small molecules. A set of mathmetical expressions together with empirical parameters is referred to a force field that provides a versatile and efficient description of chemical systems. In the past half century, huge efforts were devoted to build up databases of classes of molecules. A variety of force fields have been developed in the modeling of macromolecules (see section 2.2.4). The general form of the potential energy function employed in most MM force fields is,

$$E_{total} = E_{bonds} + E_{angles} + E_{dihedrals} + E_{improper} + E_{nb} + [special\ terms]. \tag{2.1}$$

2.2.1 Bonded Interactions

The first four terms define the energy contribution from intramolecular interactions, which represent the covalent bond stretching (E_{bonds}), bond angle bending E_{angles}), dihedrals or bond torsion ($E_{dihedrals}$) and the planar dihedral angles terms ($E_{improper}$), respectively. Considering that bonds are described as springs in a molecule, bond stretching and bending are evaluated as harmonic potentials centered on equilibrium values for lengths and angles by using Hooke's law. Thus, E_{bonds} and E_{angles} can be expressed as,

$$E_{bonds} = \sum_{bonds} k_b(b - b_0)^2 \tag{2.2}$$

$$E_{angles} = \sum_{angles} k_\theta(\theta - \theta_0)^2. \tag{2.3}$$

In equation 2.2 and 2.3, b and θ represent the current values of bond lengths and bond angles, while constants b_0 and θ_0 are the reference bond length and bond angle. The parameters k_b and k_θ are the equilibrium force constants for bonds and angles, respectively. The energetic contribution from both potential terms are calculated by summing up all bonds and all angles in the system. The third sum of the expression

refers to the torsion angle potential function. It models the presence of steric barriers between atoms separated by 3 covalent bonds (1, 4 pairs), thus, it is important for describing molecular conformations. Torsional angle rotations differ from bond stretching and bending as its potential is periodic through a $360°/n$ rotation. The equation takes the form of,

$$E_{dihedrals} = \sum_{dihedrals} \sum_{n} k_{\chi}(1 + \cos(n\chi - \delta), \tag{2.4}$$

where χ is the dihedral angle, three parameters k_{χ}, n, and δ are the height of the rotational barrier, the multiplicity or the periodicity of the rotational barrier and the reference angle (or phase angle). The value of n is usually between 1 and 4, and two or more multiplicities are often used for a give torsion angle to improve the accuracy in predicting the change in energy as the torsion angle rotation can be large. The last bonded interaction term refers to "improper" torsions or out-of-plane-bending, which function as correction factors for out-of-plane-deviations, for example, to keep aromatic rings planar.

2.2.2 Non-Bonded Interactions

The energy term represents the non-bonded interactions in the potential energy function including predominantly van der Waals interaction energy and electrostatic interaction energy. As it is assumed that a force field developed based on a small set of molecules can be applied to predict a broader set of molecules with similar chemical groups, the accuracy of modeling these interactions is particularly essential to determine the transferability of a force field. Non-bonded interactions act only between atoms between molecules or separated by at least three bonds, and are pairwise additive. The total energic contribution from the non-bonded forces accounts for the sum of all interacting atomic pairs in a system. The potential function of van der Waals interactions is implemented as Lennard-Jones potential,

$$E_{LJ}(r) = (\frac{r_{min}}{r})^{12} - 2(\frac{r_{min}}{r})^{6}, \tag{2.5}$$

where r_{min} is the most favorable distance between atoms. The first term of the Equation 2.5 represents a repulsive force at the distance where the electronic shells

overlap between an atomic pair. The second term represents the attractive force, also refers to the dispersion force, which arises from fluctuations in the charge distribution in the electron cloud. $E_{LJ}(r)$ is dependent on the distance r of the atoms. Both of the effects become significant as r decreases, and decay to zero rapidly as r increases.

The electrostatic energy is the function describes the differences in the charge distribution within a molecule, which accounts for the interations of polar and charge groups. The energy contribution for the potential is modeled using Coulombic potentials,

$$E_{Coul}(r) = \frac{q_i q_j}{\epsilon r}, \tag{2.6}$$

where q_i and q_j are the point charges of atom i and j, and ϵ represents the dielectric constant. Coumloub potential is computed for all partial and fully charged atomic pairs in the system. The accuracy of the electrostatic term is dependent on the "correct" assignment of charges on each atom. However, it is a crude approximation to consider atoms as fixed point charge since electrons delocalized around nuclei and bonds.

2.2.3 Special Terms

Based on the general form of the potential energy function, a variety of special terms can be taken into account. This differentiation in forms of equations can affect the choice of force field and parameters for the systems of interest. For example, a cross term that reflects the coupling between adjacent bonds can be found in the form of $E_{cross}(b_1, b_2) = K_{b_1,b_2}(b_1 - b_{1,0})(b_2 - b_{2,0})$. Other cross terms of bond-angle, bond-torsion and angle-torsion can be introduced in a similar fashion. These energy terms serve as corrections due to the intramolecular energy. These improve the accuracy of reproduction of experimental measurements for model compounds. Hydrogen bonding is an important type of electrostatic interaction and a powerful driving force for the protein folding. Some early versions of force fields employed an extra term to improve the accuracy of the hydrogen bonding energy [153]. However, it is well demonstrated that hydrogen bonding can be usually reproduced by an adequate choice of van der Waals parameters and electrostatic potential terms [153, 154].

2.2.4 GROMOS Force Field

In the need of extended scope of research, force fields have been developed to model complex phenomena such as polarization [155–157], and to be applicable to a wider range of environments such as lipid membranes [158], solid-state interfaces and metals. Examples of force fields that have been designed for biomolecular simulations are CHARMM [159–161], AMBER [162–164], GROMOS [165–169], and OPLS [170,171]. To improve the accuracy on diverse systems, some of these force fields have specialized versions (e.g AMBER-99 [172], AMBER-03 [173], CHARMM-22 [161], CHARMM-36 [174]; GROMOS-96 53A6 [169], GROMOS 54A7 [175]). GROMOS 54A7 was used to perform MD simulations in this work.

The GROMOS (GROningen MOlecular Simulation) force field and the associated biomolecular simulation package was developed by W. F. van Gunsteren and coworkers (at Harvard University (USA), the University of Groningen (The Netherlands) and the Swiss Federal Institution of Technology (Switzerland)) since 1978. The central idea for the design and parameterization of GROMOS force field is to reach a balance between an accurate description of the interaction energy as function of conformation and a simplicity of the potential energy function. The initial parameters of bonded interactions were obtained from crystallographic and spectroscopic data for small molecules. Non-bonded interaction parameters were obtained from crystallographic data and atomic polarizabilities and then refined to improve agreement with experimental data. The parameterization of non-bonded interaction was further improved by the application of statistical-mechanical approaches. Combination of experimental and *ab initio* quantum-chemical data were used for the development of parameters. With the increased computational power and in the accuracy of methods, the second generation force fields such as GROMOS 45A4 allowed the calculation of liquid possible and condensed-phase thermodynamic data in the parameterization procedure. The subsequent generation of the GROMOS force field parameter sets 53A5 and 53A6 was reparameterized against free enthalpies of hydration and solvation for the purpose of simulating solvation effects and partition properties in the biomolecular processes. Another aim in this reparameterization was to allow simulations of proteins and lipid membranes using one force field. However, an underestimated stability

of short α-helix was diagnosed in the validation of the force field, leading to a further reparameterization of torsional angle terms against a large set of high-resolution crystal structures for version 54A7. Modification in 54A7 refined secondary structure elements stability, while retaining the agreement with other experimental data.

2.3 Molecular Dynamics

MD uses an algorithm to simulate time evolution of a set of interacting atoms under the influence of a specific force field by solving the Newton's equation of motion. QM and MM give a mathematical description of chemical phenomena, while MD serves as a tool to implement experiments, enabling us to observe biomolecular processes and providing coarse information about microscopic states.

The fundamental elements for a MD simulation are (1) the forces on the particles which can be calculated from the potential energy of interacting atoms and (2) solving the equation of motion for the description of the system dynamics. In a system composed of N interacting atoms, the potential energy is approximated by a MM force field described above. The force on an atom i, at time t, is given by

$$\vec{F_i} = -\nabla_i E = -\frac{\partial V}{\partial \vec{r_i}}, i = 1, 2, 3, ..., N, \tag{2.7}$$

where E is the potential energy function, and the variable r_i represents spatial coodinates of the atom i. According to the Newton's second law of motion, the acceleration $\vec{a_i}$ of the atom i is calculated from

$$\vec{a_i}(t) = \frac{\vec{F_i}}{m_i}, \tag{2.8}$$

where m_i is the atom mass. The velocity of the atom i at a subsequent time step $\vec{v_i}(t + \delta t)$ (usually on the order of femtosecond (fs)) is calculated from

$$\vec{v_i}(t + \delta t) = \vec{v_i}(t) + \int_t^{t+\delta t} \vec{a_i}(t)dt = \vec{v_i}(t) + \vec{a_i} \cdot \delta t \tag{2.9}$$

Then the position of atom i at time $t + \delta t$ is given by

$$\vec{x_i}(t + \delta t) = \vec{x_i}(t) + \vec{v_i}(t)\delta t + \frac{\vec{a_i}(t)\delta t^2}{2} \tag{2.10}$$

To describe the evolution of the system, the velocities and positions of sets of interacting atoms (Equations 2.9 and 2.10) are approximated numerically at every time step using a time-intergration algorithm. Various schemes of intergration are available, such as the leapfrog [176] and the velocity algorithms [177]. A small intergration time step δt ranging from 1 to 4 fs is typically chosen to maintain numerical stability. During the time interval, forces acting on each atom is assumed to be constant. Once the positions and velocities are computed, atoms are moved to new positions and subsquent interatomic forces are updated based on the new coordinates. This process is repeated to obtain coordinates of the system at future time steps.

2.3.1 Temperature and Pressure

Temperature is a measure of the average kinetic energy of the particles in a system. The instantaneous temperature of a macroscopic system can be calculated using the kinetic energy of every atom. According to the equipartition theorem in a (classical) many-body system at the thermal equilibrium, the average kinetic energy (K) per degree of freedom is related to temperature (T). This relation is expressed as

$$K = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}m_i v_i^2 = \frac{3}{2}k_B T, \tag{2.11}$$

where m_i and v_i are the mass and velocity of the atom i, N is the number of atoms and k_B is Boltzmann constant. This equation is used as an operational definition of the temperature in MD. Temperature of a system is determined by the total kinetic energy normalized by the number of degrees of freedom:

$$T(t) = \frac{1}{3}N\sum_{i=1}^{N}\frac{m_i v_i^2(t)}{k_B} \tag{2.12}$$

Pressure P in a MD simulation is evaluated via an emsemble average of the microscopic or instantaneous pressure $\mathcal{P}$ [178]. In a system of N particles in a volume V, the microscopic pressure is expressed as

$$\mathcal{P} = \frac{1}{V}\left(\frac{1}{3}\sum_i m_i \mathbf{v}_i \cdot \mathbf{v}_i + \frac{1}{3}\sum_i \mathbf{r}_i \cdot \mathbf{f}_i\right), \tag{2.13}$$

where m_i is the mass, r_i is the position, v_i is the velocity and f_i is the force acting on particle i. Then, the macroscopic pressure P can be simply obtained as $P = \langle \mathcal{P} \rangle$, where the angular brackets imply a time or statistical average over the appropriate ensemble. For the system with pairwise interaction, the pressure is commonly computed by taking the average of an intantaneous pressure function, that is

$$P = \left\langle \frac{Nk_BT}{V} + \frac{1}{3} \sum_i \sum_{j<i} \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} \right\rangle, \tag{2.14}$$

where T is the temperature, $\mathbf{r}_{ij}$ is an intermolecular vector between a molecular pair, and $\mathbf{f}_{ij}$ is the corresponding intermolecular force. The first and second terms in Equation 2.14 represent the kinetic energy of the particles and the residual contribution arising from the inter-particle interactions, respectively.

2.3.2 Statistical Ensemble and Averages

MD simulations generate information at the microscopic level (such as positions, velocities, individual kinetic and potential energies) of a macroscopic system. Statistical mechanics connects microscopic properties of a system to its macroscopic thermodynamics quantities such as temperature, chemical potential and free energy. In a physical measurement, macroscopic observables are determined from an average over all possible microstates of a system, which is referred as the ensemble average. In a MD simulation, a time-trajectory of positions of atoms in the system is generated, thus, average quantities can be evaluated by performing time averaging along the trajectory in phase-space.

MD simulations can be carried out either in the canonical (NVT) ensemble, which is characterized by a constant particle number (N), volume (V) and temperature (T) or in the isothermal-isobaric (NPT) ensemble, where the number of particles and temperature are fixed values, while the volume is replaced by the average pressure (P). These thermodynamic variables can be treated as control parameters that specify the conditions under which a MD simulation is performed. Simulations are typically carried out in the NPT ensemble since it is the closest to typical experimental conditions.

The probability for the canonical ensemble to find a system in the microstate i,

with an energy E_i is given by the Boltzmann distribution,

$$p_i = \frac{1}{Z} e^{\frac{-E_i}{kT}}, \tag{2.15}$$

in which

$$Z = \sum_i e \frac{-E_i}{kT}. \tag{2.16}$$

Equation 2.16 is the partition function of the system from which any thermodynamic variable can be derived, such as the Gibbs free energy (G) for a system in the NPT ensemble, and the Helmholtz free energy (A) for a system in the NVT ensemble. The relative free energy ΔG is a central thermodynamic quantity in the field of biochemistry. It measures the difference between two states of interest, indicating whether a reversible biochemical process will occur spontaneously. For any two states 0 and 1, ΔG can be computed via their respective probability p_0 and p_1,

$$\Delta G = -k_B T \ln \frac{p_0}{p_1}. \tag{2.17}$$

If a system is allowed to evolve for long enough time, the system will eventually visit all possible states. Thus, the time average of any of its macroscopic property is equivalent to its ensemble average. This is called the ergodic hypothesis which is one of the fundamental assumptions in statistical mechanics [179]. In practice, this means that if a fixed duration MD simulation explores all relevant regions of the phase space using a feasible amount of computer resources, statistically-meaningful experimental quantities can be estimated from this simulation.

2.3.3 Advanced Sampling Techniques and Simulated Annealing

A simulation timescale of at least micro/millosecond is required for a MD simulation to adequately sample a potential energy surface, which is computationally expensive. Also, transitions between conformations that are seperated by high-energy barriers appear to be rare events in a classical MD simulation. Considering that biomolecular systems generally have rugged energy landscapes, many important states may not be visited by a typical length of simulation. It is more challenging to explore the conformational space of a disordered protein due to the existence of a large number

of conformational states in its ensemble [180]. To enhance the ability of MD sampling and enable extended timescales, many advanced sampling techniques such as REMD, metadynamics and simulated annealing (SA) are used for biomolecular simulations. A brief review of SAMD is provided below. Detailed reviews of other enhanced sampling methods are available elsewhere [181, 182].

The SA algorithm [183] developed based on the Monte Carlo algorithm is a technique to find the minimum energy configuration of a system. The method depends on an artificial temperature that decreases during the simulation. The algorithm usually starts the simulation at high temperature to overcome energy barriers followed by gradually cooling to reach low energy regimes. It is common to combine the SA simulation with an extended classical simulation at room temperature for the final refinement.

2.3.4 Periodic Boundary Conditions

MD simulations are usually performed on finite size of systems containing a few thousand and sometimes up to millions of particles. However, the behaviour of finite systems is distinct from that of infinite systems. In a finite system, a non-negligible fraction of particles in a MD simulation are located in the vicinity of the boundaries of the simulation cell (or "box"). In order to simulate bulk systems, it is essential to choose boundary conditions that mimics the presence of an infinite bulk surrounding. Periodic boundary conditions (PBC) are a set of equations applied on a simulation cell to mimic an infinite lattice of replicated cells. With PBC, only the particles in the central box are modeled. When a particle leaves the simulation region, an image particle will re-enter the box on the opposite side. This way, the surface effects are eliminated and the system can be treatred as a bulk system. To avoid counting duplicated interactions from multiple adjacent image cells, the minimum image criterion is applied that among all images of particles, only the pairs with the closest distances are considered as interacting pairs.

2.3.5 Preparing a Biomolecular System for MD Simulation Using GROMACS

GROMACS (GROningen MAchine for Chemical Simulations) is one of the widely used open-source software package. It primarily aims to perform dynamic simulations of biomolecules, which supports for different force fields and offers a toolkit for preparation, calculation and trajectory analysis. A workflow is provided (Figure 2.1) to illustrate a typical MD protocol.

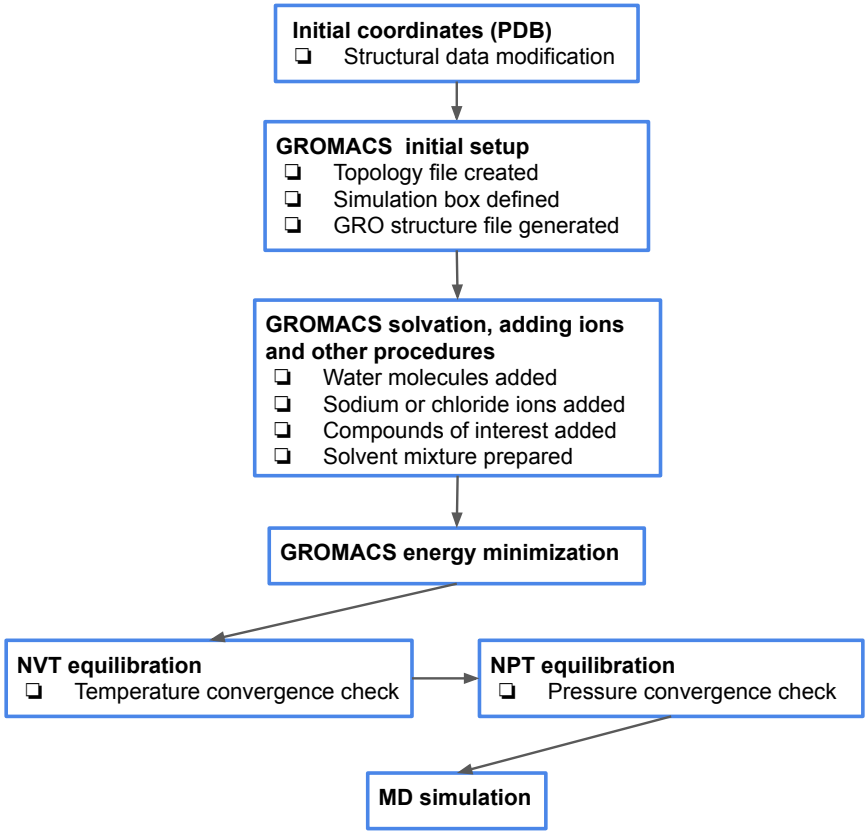

Figure 2.1: The basic GROMACS workflow including additional procedures due to various experimental purposes.

The initial geometric input of a known system (either from crystallography or NMR) that can be prepared as Carterian coordinates or internal coordinates is usually extracted from the PDB (https://www.rcsb.org/). This structural data may require a modification. For example, missing residues are modelled into the protein or a monomeric structure is abstracted from its polymer system. In the case of a crystal structure, the information of hydrogen atoms is added to the model. A topology is

prepared based on this input, which contains all the information required to describe the system for the purposes of simulation including atom masses, bond lengths, bond angles and charges. To create this topology, a force field and water model is then selected. The next step is to define the shape and size of a simulation box.

To closely mimic *in vitro* conditions, the protein molecule is solvated with water molecules and other compounds such as ligands, metal ions, and salt are added to the simulation box. In Chapter 5, structures of Aβ monomers and oligomers are prepared in various concentrations (v/v) of ethanol in water, wherein the number of ethanol molecules under a certain concentration is calculated according to the total volume of the generated simulation box. After inserting the target number of ethanol molecules, water molecules are added to fill the rest volume of the box. Sodium or chloride ions are automatically added to neutralize the charge of the system.

Energy minimization is usually performed on the system to remove any unusual geometry that could have been introduced during the system preparation. Prior to performing the production simulation, two stages of equilibration (NVT and NPT ensemble) are applied to remove inherent bias in the initial structure in terms of geometry and solvent orientation. An equilibration phase is conducted as a short simulation to allow solvent molecules to relax around the protein and for the system to reach thermal equilibrium. Typically, 50-100 ps of equilibration is sufficient to reach a plateau at the intended value.

2.4 Analysis of MD Trajectories

MD simulations of an IDP such as Aβ result in geometric data sets in high dimensionality due to the complex dynamics of and transient conformations explored by the protein [184]. Some properties are hidden in the complexity of the data and difficult to be characterized. Two statistical techniques, a clustering method (or community detection algorithm) and PCA are used to resolve Aβ42‘s structural properties in this work. Clustering groups members of the ensemble with similar topologies, while PCA projects members of the ensemble into a maximally informative low dimensional space.

2.4.1 Clustering with Community Detection in Networks

Definitions of Graph, Network, and Community Structure

In mathematics, a graph consists of a set of vertices that are connected by edges [185]. It provides a convenient way to represent various kinds of mathematical objects. Graph theory is the study of graphs that allows us to model pairwise relations between objects. This concept is further applied to network theory that takes advantage of graph representation to model and analyze the real-world structures [185]. Such applications include a number of fields, and of particular interest here is the implementation in biological sciences, such as protein-protein interaction networks [186], gene regulatory networks [187] and metabolic networks [188]. One of the most relevant properties of real-world networks is the community structure or clustering that the vertices are organized into communities or clusters, with higher density of within-group edges than between-group edges [189]. Generally, a community or cluster can be treated as an independent compartment of the structure, containing a group of vertices sharing common features [190,191]. In this thesis, a conformation in the sampled ensemble is abstracted as a vertice and a weighted edge represents a measure of similarity. Two similar structures are connected by an edge with a highly weighted degree, while dissimilar structures are weakly or not connected at all. Then, the detection of the community structure depends on pairwise geometric similarities.

Community Detection Algorithms and Modularity

Identifying community structure is an analytic strategy for understanding different structures of networks [192]. Community detection or clustering refers to the idenfication of community structure in a network. A detailed review on different community detection algorithms, benchmarking and comparison is provided elsewhere [193]. The modularity-based algorithm proposed by Newman and Girvan [192] is briefly described here.

A graph partition is the procedure to divide vertices into clusters, such that each vertex belongs to one cluster. To evaluate the "goodness" of a partition, Newman and Girvan proposed a quality function, the modularity index Q, to quantitatively rank partitions based on their scores given by the quality function:

$$Q = \sum_i (e_{ii} - {a_i}^2), \tag{2.18}$$

where e_{ii} is the fraction of edges that fall within the group, while $a_i = \sum_j e_{ij}$ is the fraction of edges that connect to vertices in community i. Then, all the edges that link between vertices without regard for their belonging communities, e_{ij}, equals to $a_i a_j$. A large positive value of Q is expected to indicate a good partition, thus, executing modularity becomes a problem of modularity maximization. In a network of n nodes and m edges, the computational cost of this algorithm is $O(m^2 n)$.

Newman proposed to use greedy optimization to maximize the modularity index Q. Later on, Clauset *et al.* [194] (*i.e.*, fast-greedy) provided a better approach based on the greedy optimization that lowers the order of cost to $O(n \log n^2)$. In this algorithm, the modularity index is re-defined as:

$$Q = (\frac{1}{2m}) \sum_{vw} (A_{vw} - (\frac{k_v k_w}{2m}) \delta(C_v, C_w)), \tag{2.19}$$

where A_{vw} is the weight of the edge between two vertices v and w, while $k_v = \sum_v A_{vw}$ and $k_w = \sum_v A_{vw}$ represent the weighted degree of vertices v and w, defining as the summary of edge weights that has vertex v and w respectively as an endpoint. C_v and C_w are the communities that vertices v and w belong to. The δ is a binary function where $\delta(C_v, C_w)$ equals to 1 if both v and w are in the same community, and 0 otherwise. This modularity index Q ranges between -1 and 1, and describes the fraction of the within-community edge weights against the expected fraction of the same edges from a randomized weighted network regardless of the partition. Therefore, a positive value of modularity indicates there are more connections within the community than one would expect by random chance [195]. In this work, this fast-greedy approach [194] was utilized for community detection and we assume that the maximal Q community assignment is the true clustering of these data.

2.4.2 Principal Component Analysis of MD Simulations

An MD trajectory for a system of N atoms can be treated as an ordered set of $3N$-dimensional vectors. To deal with these data, I reduce the description of the highly

correlated molecular motion of $3N$ atomic coordinates to a smaller set to explain a phenomena of interest.

Principal component analysis (PCA) [196] is a multivariate technique used to systematically reduce the number of dimensions required to describe protein dynamics through a decomposition process. To be specific, a multivariate data set (such as high-dimensional MD data) has p correlated variables, $\mathbf{r} = (r_1, r_2, ..., r_p)$. PCA reduces its dimensionality by finding combinations of the data set based on variances to produce a transformed set of variables, $\mathbf{x} = (x_1, x_2, ..., x_p)$, that are uncorrelated. The indices of r_i are called the principal components (PCs). The calculated number of PCs equals to the number of p original variables. Usually, only a few PCs will account for most of the variation from $\mathbf{x}$. Thus, the dimensionality reduction can be achieved by choosing the value of m that is much less than p, where m is the number of PCs that are necessary to involve the majority of the variation in the data set.

The linear transformation of the Cartersian coordinates of the structure to PCs starts with the construction of the $p \times p$ covariance matrix. It is important to realize that index selection (such as all atoms, all-nonhydrogen atoms or all C_α) in the analysis biases PCA to extract information including large-scale motions. It was reported that a selection of all-atom in the analysis for localized events may fail to discover the localized motions [197]. In the case of a molecule with N atoms and selection of all C_α, the covariance matrix is described as,

$$\sigma_{ij} = \langle (r_i - \langle r_i \rangle)(r_j - \langle r_j \rangle) \rangle \,, \tag{2.20}$$

where $r_1, ..., r_{3N}$ are the Cartesian coordinates of all C_α in the system. The angular bracket represents the average over all sampled conformations. Diagonalization of the covariance matrix yields in $3N$ eigenvectors ($\mathbf{v^i}$) and eigenvalues (λ_i) which describe the modes of the collective motion and their corresponding amplitudes. By sorting the eigenvectors in the decreasing order based on their corresponding eigenvalues, the observed motions are filtered from the largest to the smallest spatial scales. The PCs,

$$V_i = \mathbf{v^i} \cdot \mathbf{r} \tag{2.21}$$

are the projections of the data $r = (r_1, ..., r_{3N})^T$ onto the eigenvectors. Assuming

that large-scale motions are along the first few PC modes and the small-scale motions are along the rest PC modes, the largest part of the system's fluctuations can be described in a few PC modes having the greatest variances. Usually, the first two or three PCs are sufficient to account for enough of the variance to describe the most important motions of a protein, or the "essential dynamics" [198, 199]. Furthermore, the first two or three PCs may serve as reaction coordinate to represent the conformational space of a biomolecular system (see Chapter 3). Specifically, this resulting low-dimensional representation of the dynamics $\mathbf{x}$ is often utilized to construct a free energy landscape (see Chapter 3 and 4), via,

$$\Delta G(\mathbf{x}) = -k_B T \ln P(\mathbf{x}), \tag{2.22}$$

where k_B is the Boltzmann constant, and P is the probability distruction of the MD data along $\mathbf{x}$. Characterizing the minima (which represents the conformational states of the system) and connection between the minima (which represents the barriers) reveals the metastable and transition states of the system, which allows us to identify pathways of their kinetics in a biomolecular process [200–206]. This approach has been useful in the study of structural dynamics of protein folding and for exploring aggregation pathways of IDPs [207–209].

Besides Cartesian coordinates (cPCA), PCA of MD data can also be performed on internal motions of the system, such as backbone dihedral angles PCA (dPCA), C_α-distance-based PCA (C_αPCA) and contact-based PCA (conPCA). A detailed discussion on using Cartesian coordinates versus internal coordinates to perform PCA of a MD trajectory is available elsewhere [201, 210]. PCA based on internal coordinates may provide better separation of internal and overal motion. It can also miss relevant motions that correspond to major transitions in Cartesian space [211]. Many components may be required in using dPCA, which leads to a high dimensionality of the reaction coordinate to generate conformational space or energy landscape [201]. Cartesian coordinates PCA is, in general, convenient to use for the visualization of the molecular structure. In the following chapters, cPCA is used to generated conformational space explored by the MD simulations.

Chapter 3

Capturing the Large-Scale Events of the Aβ42 Misfolding from a Double-Ended Molecular Dynamics Sampling

Contributions: Simiao Lu conducted the research and wrote the chapter. Simiao Lu, Jose Sergio Hleap and Christian Blouin designed the experiment and analyzed the data. Simiao Lu, Jose Sergio Hleap and Alex Safatli wrote the codes for trajectory analysis. Christian Blouin provided editorial input and guidance.

3.1 Introduction

Alzheimer‘s disease is a pathological neurodegenerative condition that is characterized by the aggregation of the Aβ protein into extracellular senile plaques and the formation of intracellular neurofibrillary tangles. Aβ peptides exist in lengths varying from 39 to 43 residues. They are produced by the intramembrane proteolytic cleavage of the APP in the amyloidogenic pathway by β- and γ-secretases. Among these, Aβ42 is found to aggregate much faster and be more abundant in the brains of AD patients [212]. In the past two decades, Aβ oligomers have been reported to possibly be the primary neurotoxic agents rather than its monomers or fibrils [213–215]. Oligomers are produced via the rate-limiting primary nucleation process of the amyloid aggregation that monomers spontaneously self-associate into the oligomeric nuclei. The formation of new aggregates is mediated by the addition of monomers to the template fibrillar species in the secondary nucleation phase. This monomer to oligomer transition is thought to be initiated by the *misfolding* of the monomer to an aggregation-prone conformation [216, 217]. As a result, both types of the nucleation required a detailed investigation at the molecular level.

Compared to well-elucidated aggregation pathways of the Aβ peptide [218–220], our understanding of the nature, structure, and dynamics of its *misfolding* process is limited. It is difficult to study the *misfolding* mechanism with experimental methods since the peptide aggregates rapidly and their fibrils are insoluble in water. This encourages the application of simulation strategies, the MD simulation technique in particular, to fill this gap by investigating these molecular events. Experiments reported various oligomeric forms of the Aβ42 containing primary β-sheet-rich structure, and their β-sheet contents increase with higher-order oligomers [221]. NMR studies suggested that Aβ fragments and monomers employed compact coil structure or transient secondary structures in solution [222–227]. In the simulation the *misfolding* process of the Aβ42, it is expected to observe a conformational conversion of the peptide into an intermediate, partially folded conformation that shows a propensity for the formation of β-strands [228]. To identify this kind of conformation, we proposed a simulation protocol called *double-ended search*, where two sets of MD simulations are performed from opposite ends of a monomer-to-oligomer/fibril pathway towards a soluble monomeric intermediate state. To be specific, one MD simulation

starts from the membrane-bound structure of the Aβ42, and the other is simulated from one subunit of the fibril structure. It is expected that the former will mimic the structural transition from an α-helical to a β-rich intermediate state A. In a fibril structure, the stabilization of a subunit usually comes from the support from adjacent units. By using a single subunit as the starting structure representing the other end of *misfolding*, the sampling is expected to explore structures from fibril to the intermediate state B, sharing some common topological characters with A. Therefore, the goal is to reach convergence between MD simulations from two directions, adopting structural ensembles that fulfill the above description of the intermediate state(s). With the convergence, the route of the *misfolding* pathway can be explored. This *double-ended search* lowers the computational expense by simulating along a specific pathway, and allows us to search for the structure of aggregation-prone state without *a priori* knowledge. Considering that the conformational space of monomeric Aβ can be broad, with limited sampling ability, two SAMD simulations were also performed starting from both ends. In the high temperature environment of an SAMD, the system has an increased possibility to cross high energy barriers and explore a larger conformational space. A slow cooling phase (400 ns in this case) was applied to allow for a thorough exploration for the surrounding energy landscape since quick cooling may trap the system in a sub-optimal local energy well [183, 229].

Diverse conformational ensembles are generated from MD simulations of monomeric Aβ in solution. One obstacle in interpreting these dynamical datasets is to find a simplified representation to summarize transient structures. Here, an MD trajectory is treated as a graph network made of n samples, where each sample structure is abstracted as a node and each pair of nodes are connected an edge weighted by the pair's similarity. This graph-based representation reduces complex 3D geometric data into an abstraction retaining only connectivity information (*i.e.*, structural similarity in this case). Such a technique has been applied in trajectory analysis of IDPs to deal with charactization of large-scale conformational changes [184, 230]. A community detection method (or clustering method) is used to compare these different conformations to partition structures into "communities", based on their structural similarity. A general definition of a community is a group of nodes that share more correlation within than with structures outside [231]. In this case, clustering classifies

similar configurations into the same community. As MD sampling progresses according to a Boltzmann distribution, lower energy substates will be more populated than higher energy ones, leading to different sizes of communities. In general, clustering gives a refined view of how an MD simulation samples particular energy wells.

Another obstacle is to decode the high dimensionality of the geometric dataset of MD trajectories. To deal with this, PCA, a multivariate statistical technique, is carried out in the Cartesian space (cPCA) to extract essential motions obtained from MD simulations [198]. Here, we focus on the two most explanatory dimensions as reaction coordinates for the conformational space construction. Although this simplifies the representation of the conformational space, it is sufficient to reveal the direction of ensemble-encoded conformational changes. On this PC space, MD trajectory data are shown as compact and well separated subspaces. The resulting communities from actual clustering are then projected onto this 2D space. An agreement between the formation of subspaces and community structure is expected since this application of PCA is considered as a useful means for visual community validation [232].

In this chapter, we explored the *misfolding* pathway of Aβ42 monomer using classical and SAMD simulations. We implemented the *double-ended search* simulation protocol, where MD simulations were performed from the opposite ends of a monomer-to-oligomer/fibril pathway. The secondary structure profiles and ranges of β-strands on the peptide in our simulations are in agreement with circular dichroism (CD) estimations and NMR measurements respectively. Our simulation data showed that the secondary structural transition may initiate from both terminal regions of the Aβ42. Fast and frequent structural interconversions between collapsed and extended states were observed on β-dominated monomeric structures from the current simulations. PCA was used as a tool to construct the free energy landscape of monomeric Aβ42, showing a possible pathway between the two ends of a simulated *misfolding*. We further used a systematic approach to characterize major conformational states of the whole space and intermediates along the pathway. Finally, we used a novel method to show transitional motions between major states by discretizing along the first two principal components, resulting in series of average structures that capture the large-scale folding events in the *misfolding* process. Combining our results and the previous knowledge, we propose a possible molecular description of the *misfolding*

mechanism.

3.2 Methods

3.2.1 System Preparation

Two initial structures for the ends of the *misfolding* pathway were defined as the membrane-bound, two-helix Aβ42 monomer (Sα-np), and the hairpin-shaped (or U-shaped) subunit of Aβ42 fibril (S2β) (Table 3.1, Figure 3.1). The membrane-bound conformation was based on the NMR structure of Aβ42 peptide (PDB ID: 1IYT) [81]) solved in the non-polar solution that mimics the lipid environment of membranes. This structure is boomerang-shaped and consists of two helices, over S8-V24 and K28-G38, separated by a 3-residue kink. The hairpin-shape subunit conformation was extracted from the *U-shaped* Aβ42 fibril structure (PDB ID: 2BEG [93]) that was solved in aqueous solution. One subunit structure presents a *hairpin* shape that consists two β-strands connected by a loop region of N27-A30, where the β1 region ranges from L17-S26 and the β2 region includes I31-A42. The disordered region (D1-K16) of the N-terminus was added manually in PyMol [233] to create a full length of Aβ42 (Table 3.1).

Table 3.1: Summary of MD Simulation.

PDB ID	Modification[a]	Starting Structure Label	Classical Simulation	Time (μs)(T)[b]	Annealing Simulation	Time (μs)(T)
1IYT[c]	-	Sα-np[d]	MDα	1.2 (300 K)	MDα-ann	0.4 (475 K - 300 K) 0.4 (300 K)
2BEG[e]	Spliced subunit, N-terminus added	S2β[f]	MDβ	1.2 (300 K)	MDβ-ann	0.4 (475 K - 300 K) 0.4 (300 K)

[a] Modification made for starting structures preparation [b] Temperature [c] [81]
[d] α stands for the helix-kink-helix structure, and np represents that it was solved in the non-polar environment. [e] [93] [f] 2β stands for its structure of 2 β-strands.

MD simulations starting from Sα-np and S2β were respectively named as MDα and MDβ (Table 3.1, Figure 3.1). The expected simulating directions were: (1) forward conversion – from the α-helix structure to an intermediate state of β-strand dominated species and (2) reverse conversion – from the a single fibril subunit back to

the intermediate state. Two SA experiments were labeled as MDα-ann and MDβ-ann, respectively (Table 3.1, Figure 3.1).

3.2.2 Molecular Dynamics Setup of the Monomeric Aβ42

All simulations were performed with GROMACS 5.1.4 package [234], applying GROMOS 54a7 force field [175] on the peptide and solvent molecules in an SPC/E [235] water box. The following parameters were used for all simulations in this study unless otherwise noted. The leapfrog Verlet integration algorithm was used with an integration time step of 2 fs. A rhombic dodecahedron box was used with periodic boundary conditions. Long range electrostatics interactions were calculated by the Particle Mesh Ewald summation with a Fourier spacing of 0.12 nm and a cubic interpolation order [236]. Coulomb and van der Waals cut-off distances were both set to 1.0 nm. Verlet cut-off scheme [237] was used to reach high performance when computing non-bonded interactions. A maximum force less than $100\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$ was obtained for both systems at the end of the energy minimization. A 100 ps NVT simulation was conducted at 300 K. The LINCS method [238] was used to restrain all bonds for an integration step of 2 fs. The protein and the solvent were coupled separately to a modified Berendsen thermostat called *V-rescale* [239]. Then, a 100 ps NPT simulation was performed at 300 K to generated the initial structure for subsequent production run. The Parrinello-Rahman barostat [240, 241] at 1.0 bar with a compressibility of 4.5×10^{-5} bar was used for pressure coupling. All parameters in the production simulations were set to the same as the NPT equilibration. The coordinates were saved every 20 ps for a total similation time of 1.2 μs.

3.2.3 Simulated Annealing of the Monomeric Aβ42

An 800-ns SA experiment was executed with the SPC/E water model for each starting structure. For each simulation, energy minimization and equilibration was achieved with the same settings in classical MD simulations described above before SA. Short-range electrostatics and van der Waals were treated with a 1.0-nm cutoff. Long-range electrostatics were treated via the particle-mesh Ewald algorithm with a 0.16-nm Fourier spacing and a cubic interpolation. The systems were started at 475 K and gradually cooled down to 300 K linearly for 400 ns at a cooling rate of 0.1 K per

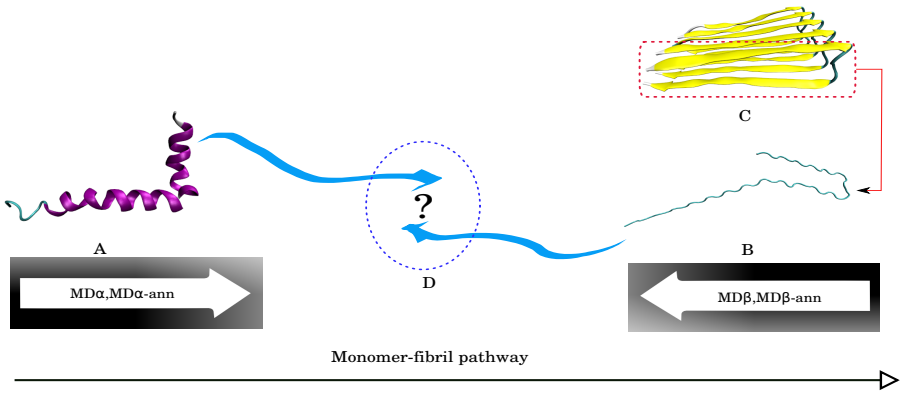

Figure 3.1: The *double-ended search* for the Aβ *misfolding*, where A, B, C and D represent Sα-np, S2β, Aβ42 fibril and a collection of intermedate states.

120 ps. This initial temperature (*i.e.*, 475 K) was selected after a few tests. SA MD simulations with this temperature allow structural rearrangment of the peptide within 100 ns. While systems were cooled down, Berendsen weak-coupling algorithm [242] was used with a reference temperature of 300 K. The systems were kept at room temperature after SA for 400 ns by using the V-rescale ensemble and Parrinello-Rahman barostat. The second half of both simulations were performed with the settings in classical MD simulations described above.

3.2.4 Trajectory Analysis Protocol

Clustering and Principal Component Analysis

Here we develop a trajectory protocol to characterize conformational ensemble into groups of structures with unique structural features. The procedures are listed below:

1. MD simulations starting from both ends are combined. Abstract each configuration as a shape using landmarks (Section Definition of landmark).

2. Align all structures in the combined trajectory (Section Structural alignment).

3. Create an all-versus-all similarity matrix (Section Similariy matrix).

4. Generate a network where each sample in the trajectory is a node and every pair of samples is connected by a weighted edge (Section Graph abstraction). The values of the weights are extracted from the similarity matrix.

5. Apply the clustering method to partition the defined graph to clusters based on the structural similarity information (Section Community structure (or clustering optimization)). Define each inferred community with its major structural characteristics by calculating the average structure.

6. Apply principal component analysis of Cartesian coordinates (cPCA) to construct the conformational space formed by PC1 and PC2. Project the clustering results onto the PC space to show the community structure (Section Principal component analysis).

7. Construct the free energy landscape based on cPCA of the combined trajectory (Section Free energy landscape).

Definition of landmark– A landmark is a point of correspondence on a shape object (protein structure in the MD trajectory in this case) that matches between and within the population [243]. Each residue in the structure was assigned by a landmark point. Here, a landmark is defined as the centroid position (x, y, z) of the residue, *i.e.*, the average Cartesian coordinates in three dimensions by calculating over all the heavy atoms of the residue.

Structural alignment– Generalized procrustes analysis (GPA) is a statistical shape analysis to provide optimized superimposition of two or multiple configurations [244]. In this method, a set of shapes are superimposed by translating, rotating and scaling one shape towards the other, where a shape is described by a set of homologous landmark points. A partial GPA [245,246] was used to align the structures to remove the interference of rotations and translations. In this approach, each structure is abstracted as a shape by assigning landmarks to residues [247]. Each configuration of landmarks are aligned to the initial estimated mean shape by translating and rotating. A new calculated mean shape is taken for further registration, involving only rotation. This process is iterated until the mean shape does not change significantly. The implementation of the method was based on the R package *shapes* [248].

Input data– Each MD trajectory was sampled for every 20 ps. All 4 MD simulations were added into a single file, resulting in a combined trajectory of 200,000 frames in the total simulation time of 4 μs. GPA was performed over all 200,000 frames in the combined trajectory. The first 100 ns of trajectory data in each simulation was discarded due to large structural changes. In the consideration of efficiency, 1800 aligned frames (per 20 ns) were selected as input data for clustering and construction of conformational space using PCA.

Similarity matrix– A pairwise distance matrix is created over all samples of the combined trajectory. This distance is represented by a similarity metric that is computed as per the equation $R = \frac{1}{RMSD + 0.00001}$ (R-score), where RMSD (root-mean-square-deviation of atomic positions) was calculated at C_α atoms of each residue. The added term to the denominator is to avoid calculation errors in the unlikely case of two identical structures due to the limited precision of the coordinate system. This R-score is a measure of similarity between samples that a larger value of the score indicates higher shared structural similarity of them.

Graph abstraction– Assume that in a trajectory of n samples, each sample is abstracted as a node. An edge is connected between each pair of nodes. Let $G = (N, M)$ be a complete and undirected graph, where N is the set of nodes representing all samples, and M is a set of edges derived from the similarity matrix. Each R-score pre-calculated in the matrix M is used to assign an edge weight to the corresponding sample pair.

Community structure (or clustering optimization)– Within the defined graph, the community structure is evaluated using a modularity-based algorithm *fast-greedy* approach [194] due to its efficiency for detecting clusters. Communites in this case are defined based on structural (vertices on the graph) similarity, which means that groups of similar structures should be interconnected with higher weighted edges. The community detection is resolved by finding a graph partitioning that maximizes the modularity index (Q) [195] (Section 2.4.1). The number of communities is determined by the optimization procedure with no requirement of *a priori* knowledge about this information. The output is a series of community memberships in the order of snapshot generation in the trajectory. Each trajectory sample obtains a membership number that indicates the community to which it belongs. Structures within a single community are considered to contain distinguishing and common structural features. In formula, the community structure S is represented as a set of subsets $C = C_1, ..., C_k$, such that $S = \bigcup_{i=1}^{k} C_i$ and $C_i \cap C_j = \emptyset$ for $i \neq j$. Consequently, any instance in S belongs to exactly one and only one subset. An average structure is computed and displayed in *cartoon* mode (using PyMOL ver. 1.4.1 [233]) to represent the common structural characteristics owned by a community.

Principal Component Analysis of MD Simulations– PCA was performed on a data matrix converted from the Cartesian coordinates of structures in the trajectory (*i.e.*, cPCA). In the data matrix, each row is a structure and each column corresponds to coordinates (x, y, z) of centroids of all residues in the structures, *i.e.*, 42×3 columns in a row in this case. A linear transformation is used in the data matrix to generate a $3N \times 3N$ covariance matrix C. Diagonalization of the covariance matrix generates $3N$ eigenvectors (v^i)and eigenvalues (λ_i). The eigenvectors give a vectorial description of each component of the collective motion with indication of the

direction of the motion, while an eigenvalue for the corresponding eigenvector represents the amount of contribution of this particular component of the motion. Thus, a time-independent low-dimensional conformational space of Aβ42 monomer can be constructed by projecting the trajectory with the first two PCs. In this work, cPCA was performed by using the pca function provided in the package Scikit-learn [249] in Python 2.7 [250]. Visulization of inferred communities determined by the modularity optimization was achieved by projecting the membership vector on the 2D cPCA space.

Free energy landscape– On the basis of cPCA, the free energy landscape (FEL) was constructed via

$$\Delta G(\mathbf{r}) = -k_B T[\ln P(\mathbf{r}) - \ln P_{max}], \tag{3.1}$$

where P is the probability distribution of the molecular system along the coordinate $\mathbf{r}$. Here, $\mathbf{r}$ is defined as count of points in each unit cell over the total number of points, where the size of each cells equals to $n \times n$ that is resulted from the pixelation of the 2D cPCA space. P_{max} denotes the maximum of $P(\mathbf{r})$, which is subtrated to ensure that $\Delta G = 0$ is the lowest free energy minimum.

Discretizing Principal Component Axes

As the first few PCs are expected to represent the essential motions in the trajectory, it is plausible to hypothesize that the variance displacement along a PC coordinate corresponds to a specific series of motion changes. To explore this intuitive meaning of PC, on a constructed 2D PC space, the first two components were discretized (DPC) into bins with an equal size (r) along the axes. Each bin area is represented by an average structure (S_n). The set of average structures captures the major conformational changes in a time-independent manner. Because r is an arbitrary value, the number of discretized bin area i varies along the axis.

For example, Figure 3.2 shows the cPCA space that was generated based on the combined trajectory of all four MD simulations, where the PC1 axis ranges between $-105.4, 140.65$. In the case of $r = 50$, then $i = 11$, which means a total of 11 average structures is calculated. With abstraction of the most covariance (25.17% in this case) shown in PC1, the largest-scale of conformational changes of the Aβ42 monomer from

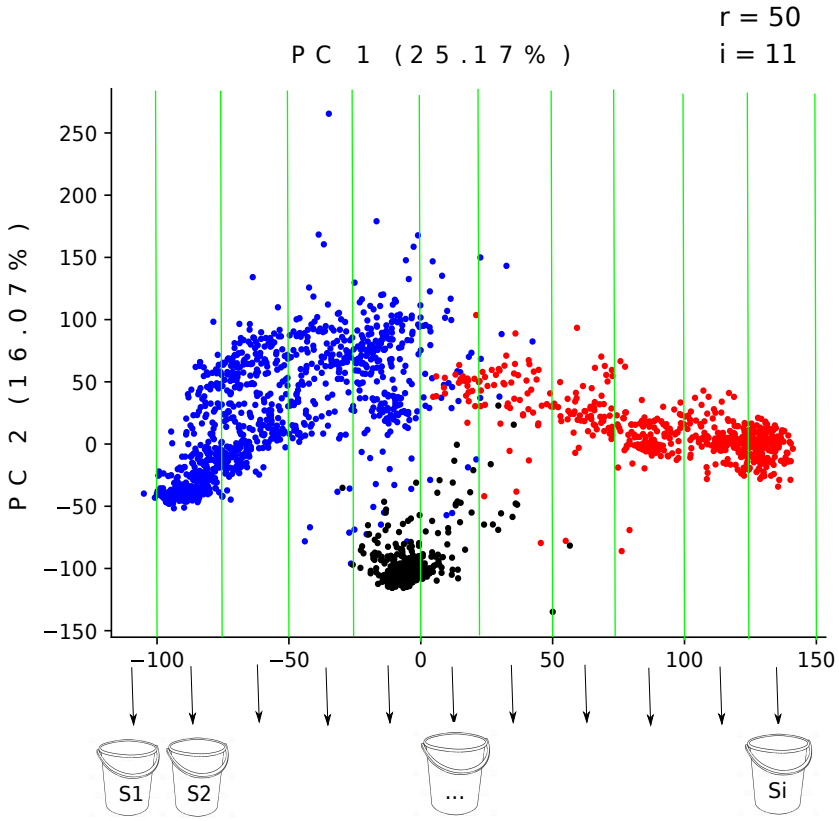

Figure 3.2: An example of binning through PC1 axis resulting in 11 pieces in the discretizing size of 50.

the sampled area can be represented by these 11 structures instead of examination of the whole geometric data set. To include more variances and visualize the changes of smaller motions, one can discretize lower PC modes by using the same method. In this way, a decreasing scale of motions may be separated by discretizing from the highest to the lowest PC modes. DPC can be useful to abstract major motions from the complex dynamics of an IDP without the consideration of timescale.

3.3 Results

3.3.1 Validation of the Conformational Ensembles

Two sets of simulations were generated for both helix-starting (*i.e.*, Sα-np) and hairpin-starting (*i.e.*, S2β) structures, each containing a 1.2 μs classical MD simulation and a 800 ns SA simulation. Validation of the samplings was made through comparison of the back-calculated three-bond J-coupling constant and chemical shifts with the experimental measurements. Description of the method and the results of the comparison are provided in Sections 4.2.3 and 4.3.1 of Chapter 4.

3.3.2 Secondary Structure Transition in Misfolding

The simulation data on the monomeric form of Aβ suggested large-scale conformational changes and an inherent propensity to form β-rich structures in water. The DSSP (the dictionary of protein secondary structure) [251] algorithm was carried out on the simulation data to determine the secondary structure content on Aβ monomers. For each MD simulation, the proportion of helical and β structures in each sample was evaluated over time (Figure 3.3). β-strands and coil/bend/turn are the dominant secondary structure elements for all simulations, regardless of the initial structure. The helix-starting structure undergoes a quick conformational transition in water, whose occupancy of β-strand and coil/bend/turn are $\sim 30\%$ and $\sim 60\%$, respectively (Figure 3.3A). This result is consistent with CD estimations [252, 253] and other measures from other computational work [145]. Higher occupancy of β-strands was shown on the peptide in MDβ, which averages $\sim 57\%$ (Figure 3.3C). In the SA simulations, high temperature environment triggers vigorous structural dynamics of Aβ, which can be seen from the significant fluctuation of β-content. Until they cool fully to the

room temperature, there is a relatively stable conformation with $\sim 40\%$ and $\sim 63\%$ of β-strand-content displayed on the peptide in MDα-ann and MDβ-ann (Figure 3.3B-D), resectively. Despite the effect of high temperature, Aβ42 monomer in the hairpin-starting conformation comprises higher β-strand-content than the monomer in helix-starting structure in an aqueous solution. The high percentage of β-structure transformation was also reported from previous computational studies [145, 254]. Another study also demonstrated the strong correlation between the high β-propensity and aggregation rate of Aβ peptide [255].

A significant structural transition of Aβ peptide from α-helix to β-sheet is observed in MDα (Figure 3.3A). The helical content drops from $\sim 67\%$ to $\sim 10\%$ at 0-200 ns and a further decreasing was observed at $\sim$ 980 ns. The retaining region of helices sites at E22-I31, which usually displays as the U-turn shape on the Aβ40 and Aβ42 protofilaments [93, 106, 256]. In contrast, no structural transition from β-sheet to α-helix was observed in MDβ in an aqueous solution (Figure 3.3C). A rapid conformational transition was observed in SA simulation starting from Sα-np. The helical content drops sharply at 475 K and remains less than 1% for most of the simulation. This implies that high temperature induces the rate-limiting primary nucleation process of amyloid aggregation, which is consistent with a recent study suggesting high temperature may accelerate the formation of β-sheet on monomeric Aβ [257].

The conformational transition to β-rich structure starts from the flexible termini. In MDα, the starting structure Sα-np shows a two-helix conformation at regions β1 (L17-S26) and β2 (I31-A42) of S2β (Figure 3.1A). The β1 region loses its helical content at $\sim$ 10 ns, while helix structures over the β2 regions last until 110 ns. Two termini (D1-F4 (a1) and G38-A42 (a2)) form two β-strands and move towards each other in the first 200 ns (Figure 3.4A1). An antiparallel β-sheet is formed between two termini, acting as the interlocking teeth of a zipper, then collapsed into a core with the middle of the chain (Figure 3.4A2). The structural stability improves due to this formation of hydrophobic cluster, leading to a slower transtion to β-structure (Figure 3.4A4). A transient helical region forms between A21-I32 during 210-476 ns (Figure 3.4A3, A4). This is similar to Reddy *et al.* in that F19-V24 displays a propensity to form α-helix [216]. Ball *et al.* also observed the existing helical conformation on

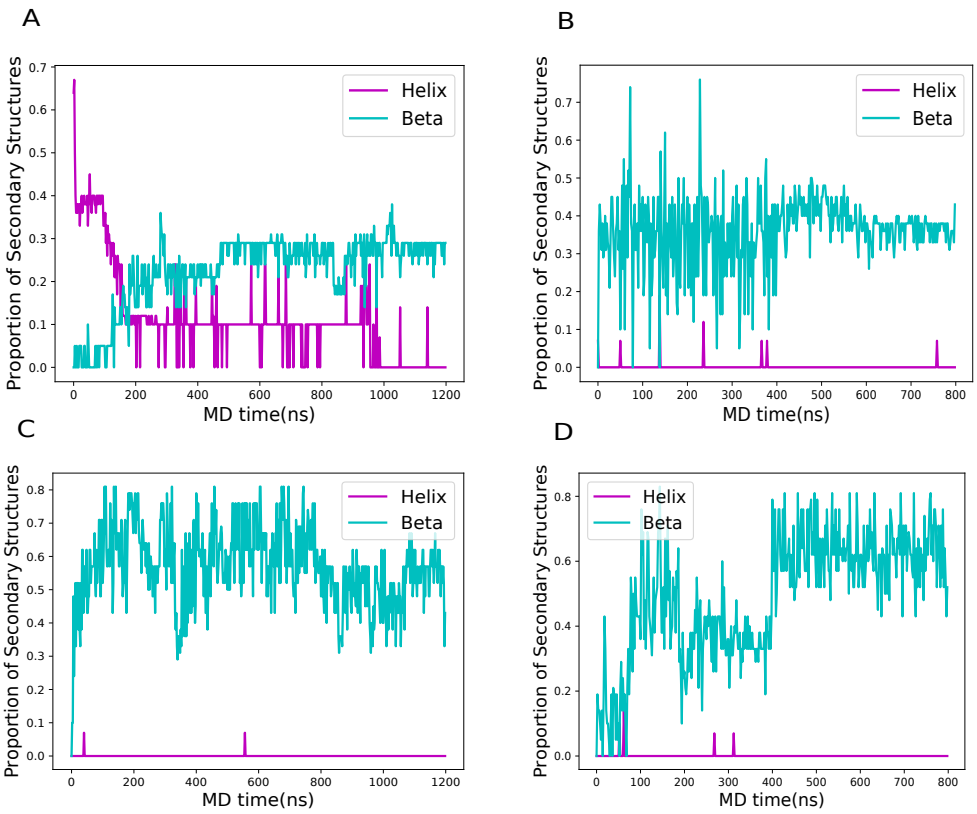

Figure 3.3: The proportion change of helix and β in every sample in MDα (A), MDα-ann (B), MDβ (C) and MDβ-ann (D). The purple line shows the change of helical elements (α-helix, 3_{10}- helix and π-helix) in the trajectory. The cyan line shows the change of β elements (isolated β-bridge and extended strand) in the trajectory.

monomeric Aβ42, however, on different regions which are H14-L17 and I32-M35 [227]. The packed conformation of antiparallel β-sheet and the helix region is maintained until the end of the simulation.

3.3.3 Conformational Interconversion between Collapsed and Extended States

Two dominant conformational states, collapsed and extended states (CS and ES), were observed in MDβ, MDα-ann and MDβ-ann. Previous studies [145,258] reported similar observation on conformational equilibrium of monomeric Aβ40 and Aβ42 between the two states. Also, they provided detailed description of CS and ES of Aβ monomers. Based on these, a CS is described as a highly dynamic globule with its hydrophobic residues collapsed into a core, while polar and charged residues at the surface are facing to the solvent. An ES displays a long, more stretched shape that the main stablizing factor is the hydrogen bonds between β-strands instead of hydrophobic clustering. Representative conformations of the two states were shown for each MD simulation in Figure 3.4. Hydrophobic clusters of collapsed conformations are higlighted, where the most frequently observed hydrophobic residues forming the hydrophobic core include F4, Y10, V12, L17, V18, F19, F20, L34, M35 and V36. This result is in better agreement with the observation from Song *et al.* (*i.e.*, F4, V18, F20, A21, I31, L34 and V36) [145]) and from Bossie *et al.* (*i.e.*, L17-A21) [259], but some differences from Ball *et al.* that hydrophobic cluster is formed between K16-A21 and G29-V36 [227].

Between three simulations, CS and ES exhibit conformational heterogeneity in ranges of β-strands and loops. Interestingly, a haipin-like conformation is observed in the beginning of MDα-ann due to the formation of a β-sheet around the C-terminal region. Two to four moving ranges of β-strands on the Aβ42 were observed from our simulations (Figure 3.4), which matches the results from Liu *et al.* [260] and Ball *et al.* [261]. Among them, three β-strands formed in both MDβ and MDα-ann, locating at E3-S8 (b1)/ K16-A21 (b2)/L34-V39 (b3), and E3-H6 (c1)/V12-D23 (c2)/G33-I41 (c3), respectively (Figure 3.4B, C). For MDβ-ann, four β-strands are shown on the chain, involving F4-D7 (d1)/E11-H14 (d2)/L17-D23 (d3)/L34-V40 (d4) (Figure 3.4D). In contrast to *U-shaped* Aβ fibril [93], b2 and c2 are formed at locations closer

to the N-terminus side comparing to β1 (L17-S26). Structures simulated from MDβ-ann adopted two strands d2 and d3 around the region of β1. Regions of b3, c3, d4 and β2 (I31-A42) are located around the same area. All structures from the MD simulations display wider loop regions (E22-G33, V24-I32 and V24-G33 for MDβ, MDα-ann, and MDβ-ann) in comparison with Sα-np and S2β.

Interconversions between collapsed and extended conformations were observed in MDβ, MDα-ann and MDβ-ann (Figure 3.4B-D), which matches the observation from a previous computational work [262]. Our results revealed two types of the conversion between two states resulting from distinct β-sheet arrangements. One type shows a bending motion in the middle of the double-hairpin structure (Figure 3.4B, D), leading to an open/close folding event. The other conversion is between the ES of a single-hairpin-shape at the C-terminus with a long and mobile N-terminus (Figure 3.4C1), and the CS of a second β-sheet formed on the side of the single hairpin that is between β-strands on two termini (Figure 3.4C2). In this CS, the C-terminus orientation adjusts its orientation to increase the possibility of hydrophobic interactions between a loop region (G9-Q15) and itself, forming a dynamic core (Figure 3.4C2, C3). Another stabilizing factor in this conformation is the salt bridge established between D23 and K28. This is in agreement with NMR measurements in the *U-shaped* fibril structure [93]. To show the frequency of the interconversion, the solvent accessible surface area (SASA) of the residues that are most involved in the hydrophobic collapses was computed for all samples versus trajectory time (Figure 3.5A-D), where CS and ES conformations are observed below 7.0 nm^2 and above 8.0 nm^2. Typical conformations of two states mentioned in Figure 3.4 are labeled. Compact conformations of the Aβ42 are observed to be favorable in solution in ranges of 300 K to 475 K (Figure 3.5E). High temperature increases its structural dynamics, leading to higher frequency of the conformational conversion, and larger probability to stay in ES (Figure 3.5C, D). Compared to frequent interconversions in SA simulations, the open/close folding event in MDβ showed a slow and gradual transitions between states (Figure 3.5B). Simulations starting from Sα-np suggested the preference of staying in a collapsed conformation, where the probability of CS and ES are $\sim$ 89.0%/$\sim$ 0.3% and $\sim$ 88.0%/$\sim$ 2.8% for MDα and MDα-ann- respectively. Lower ratios of CS/ES (*i.e.*,

higher probability staying at ES) are shown in simulations starting from hairpin-starting-structure. These are $\sim 51.0\%/\sim 15.0\%$ and $\sim 47.0\%/\sim 21.8\%$ for MDβ and MDβ-ann, respectively.

3.3.4 Construction of a Connected Conformational Space Based on the Combined Trajectory

PCA was carried out to construct conformational space of Aβ42 monomer from Cartesian coordinates. Two classical MD simulations (CT1), classical MD simulations and first half of two SA simulations (CT2), and all simulations (CT3) (Table 3.2) were projected onto the plane formed by the first two components (Figure 3.6A, B and C). Each simulation with the first 100 ns discarded were combined into a single file as the input for PCA. To identify the possible *misfolding* pathway between two end-states, part of the simulation data before steady states were included in the projection. Each point in the landscape represent an individual structure. The gradient colors indicates a third coordinate of simulation time.

Table 3.2: Summary of combined trajectories.

Combined Trajectory	**CT1**	**CT2**	**CT3**
Simulations	MDα	MDα	MDα
	MDβ	MDβ	MDβ
		MDα-ann (cooling)[a]	MDα-ann (all)[b]
		MDβ-ann (cooling)	MDβ-ann (all)

[a] the cooling phase of SA simulations (or first 400 ns)
[b] the whole SA simulation (800 ns)

The projection of CT1 revealed a bi-partition of the conformational landscape that structures from each MD simulation are located mainly on the positive and negative side of the PC1 axis (Figure 3.6A). The division indicated that conformations sampled from MDα and MDβ share less common structural characteristics along the mean axes (Figure 3.6A) since PCA was able to separate the conformational states with distinct features into different locations on the plane without *a priori* knowledge of memberships. The exploration space sampled by MDβ occupied a broad domain on the landscape plot. This may be explained by the occurrence of a large-scale conformational change that was simulated in MDβ, while the helix-starting-structure experienced dynamics due to the early formation of a compact structure (see Section

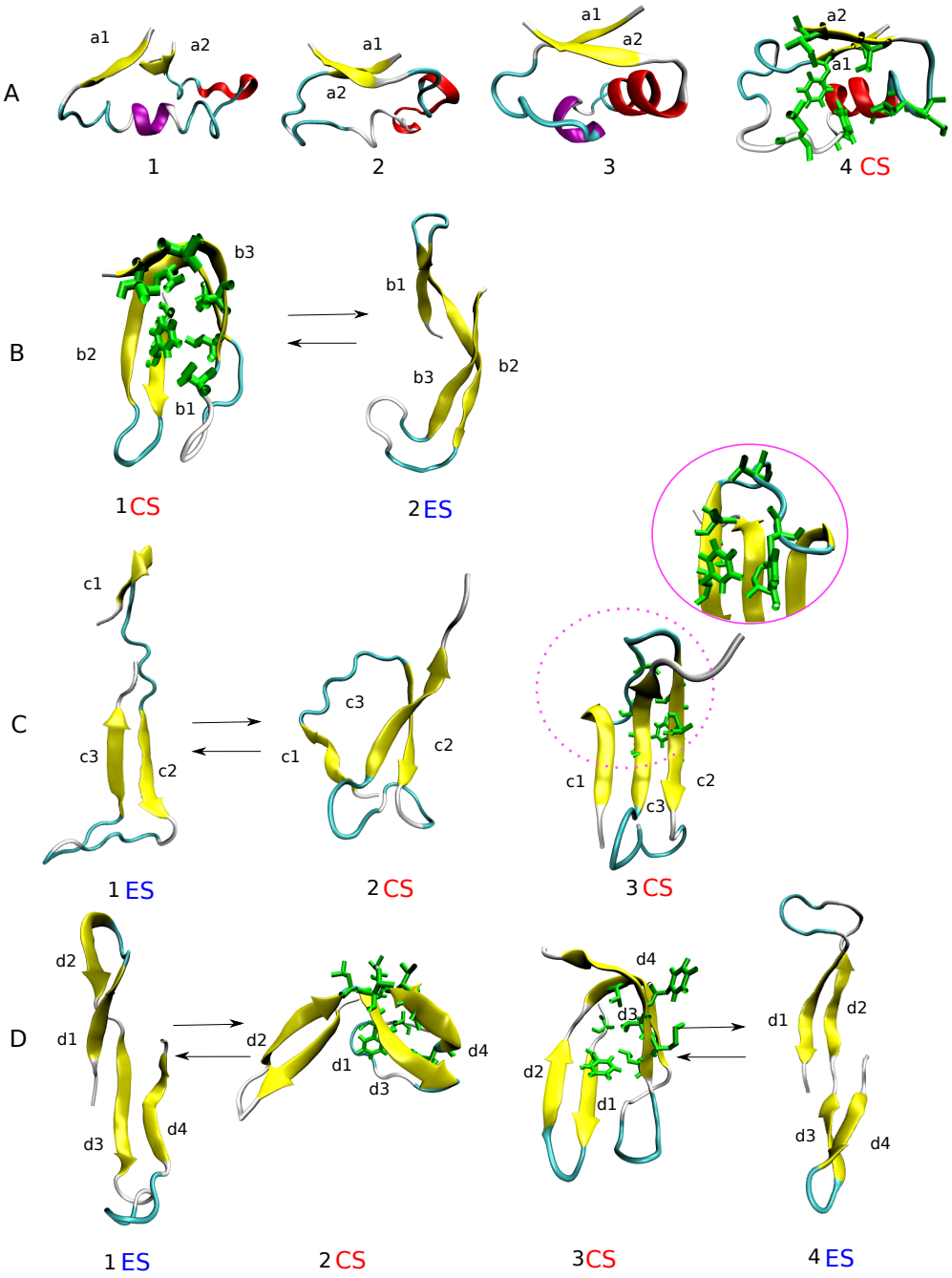

Figure 3.4: Selected snapshots and typical collapsed conformation (CS) and extended conformation (ES) for simulation A) MDα; B) MDβ; C) MDα-ann and D) MDβ-ann. Color codes on the peptide indicate the secondary structure elements which are α-helix in purple, π-helix in red, extended-β in yellow, turn in cyan and coil in white. Hydrophobic residues, including F4, Y10, V12, L17, V18, F19, F20, L34, M35 and V36, that are frequently observed in the hydrophobic clustering are displayed in bond representation. Arrows indicate the occurrence of interconversion between CS (red) and ES (blue) in the corresponding simulations. All structures are placed with the N-terminus on the left and the C-terminus on the right.

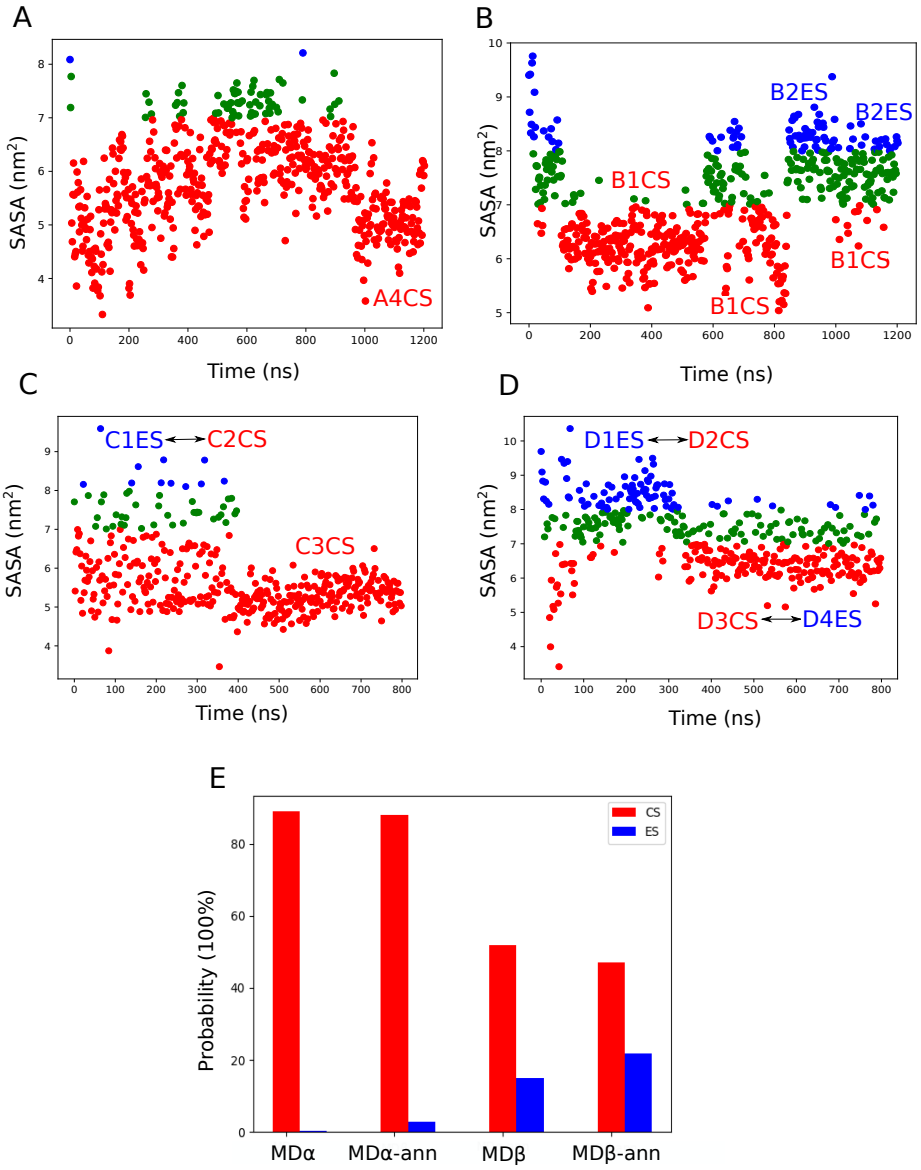

Figure 3.5: The apprearance of collapsed (CS) and extended conformations states (ES) in Aβ42 along time in MD simulation A) MDα; B) MDβ; C) MDα-ann and D) MDβ-ann. The solvent accessible surface area (SASA) is calculated based on the most frequently observed hydrophbic residues in the hydrophobic clusters of collapsed conformations (*i.e.*, F4, Y10, F19, F20, L34, M35 and V36). The collapsed conformation is defined as SASA below 7.0 nm^2 and colored in red, while the extended conformation is defined as SASA above 8.0 nm^2 and colored in blue. Green dots represent transient conformations between the two states. Labels of letters and a number refer to structures in Figure 3.4. Arrows indicate the occurrence of interconversion between CS (red) and ES (blue) in the corresponding simulations. E) The probability as a percentage of observing collapsed/extended conformations in each simulation.

3.3.2 and 3.3.3 for details). The combined 2.4 μs sampling is insufficient to cover the full *misfolding* pathway.

The sampled space is connected between two ends by adding SA simulations into the dataset. Figure 3.6B shows the projection of combined trajectory in the addition of the first half (cooling phase) of MDα-ann and MDβ-ann (CT2). With the addition, a subspace is shown on the blank area between MDα and MDβ, but closer to the side of negative values along PC1 axis. This insertion bridges the disconnected space without altering shapes of the PCA plot for MDα and MDβ. A 3-armed-shaped conformational landscape is created by adding the rest parts of the SAMDs (CT3) (Figure 3.6C). One effect of this addition is that the structural ensemble in MDα-ann is pushed into a third direction forming a condensed area, while structures in MDβ-ann stay at relatively the same location in the projection. The ranges of two PC coordinates extends with the additions, while the summed variances of two PC modes decreases from 64.08% to 48.10% and then to 41.24%. Our results show that cPCA on CT2 gives a visual representation of the *misfolding* pathway of monomeric Aβ42 from the *double-ended search*. On the other hand, cPCA of CT3 displayed the existence of a third conformational state that is separated from the pathway. In this application of dimensional reduction, increasing amount of geometric input leads to less variance included in the first two PCs, however, cPCA managed to decode the complex geometric information and revealed three subspaces of distinct orientated conformational ensemble.

By performing cPCA of combined trajectories, free energy landscapes were constructed in $-ln(density)$ of k_BT units for identifying ensemble of similar structures, where the cPCA space is pixelated by the value of $n = 3$ for CT1 (Figure 3.6D) and $n = 10$ for CT2 and CT3 (Figure 3.6E, F). The cPCA landscape of CT1 exhibits two minima, while in the cPCA of CT2, intermediate states appear that connect two independent conformational states in the addition of SA simulations. Three major minima and at least five intermediate states were characterized in the cPCA landscape of CT3, where high energy barriers were identified on the conjunction area between three minima. This minimal dimensionality of an FEL determined by the first few PCs was reported to be valuble in describing protein folding dynamics [209]. A recent computational study, exploring dynamics of Aβ(16-35) peptide, took advantage

of this approach and with clustering methods revealed its structural interconversion between multiple characterized metastable states [263].

Samples from MDα and MDβ were assigned into two communities in Figure 3.6G, which is consistent with the disconnected sampled area of CT1 generated by PCA. Clustering on CT2 also gives two communities, where structures of MDα-ann, MDβ and MDβ-ann were arranged into the same community (cb1). This result is also supported from the projection of CT2 using PCA that the samples from first half of SA simulations show up closer to the MDβ subspace. Above this, a third community (cc3) was identified from clustering of CT3 that is populated with samples mainly from the second half of MDα-ann (Figure 3.6F). This result is in agreement with the three dominant states identified from FEL of CT3. It seems that this application of modularity based on RMSD focuses on characterizing major conformational states, while intermediate states are merged into an adjacent minimum. The projection of CT3 on the plane formed by the first three PC modes is also provided in Appendix A.

3.3.5 Characterization of Conformational States along the Misfolding Pathway

For each inferred community from clustering, an average structure is calculated to visualize their main structural features (Figure 3.7B). The structures of the closest RMSD against the geometric coordinates of average structures are also listed (Figure 3.7B). Community cc1 is represented by collapsed double *U-shaped* conformation in the formation of two β-sheets which is similar to structure B1 and D3 in Figure 3.4B, D. Representative of cc2 is a compact joint-ends conformation that is configurational similar to A4 (Figure 3.4A). The newly formed helices on the C-terminal is shown in the average structure, indicating its importance in formation of the hydrophobic core. A collapsed conformation for community cc3 (*i.e.*, as3) shares a similar topology to C3 (Figure 3.4C). In this conformation, three major factors contribute to its stabilization: hydrogen bonds between three β-strands, the hydrophobic cluster formed between C-terminus and β-strands, and a salt bridge (D23-K28). All three structures are compact due to a preference of Aβ42 monomers to stay in a collapsed state in

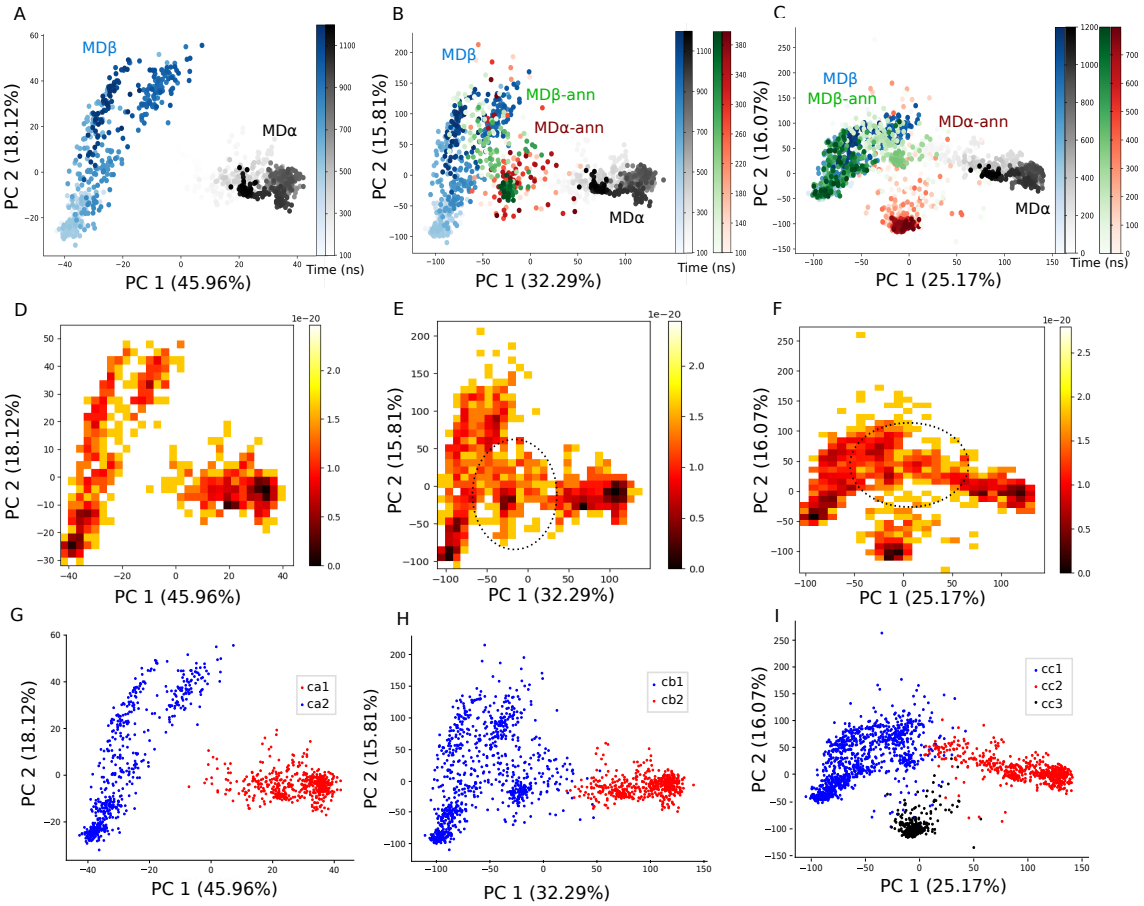

Figure 3.6: The projection of CT1 (A), CT2 (B) and CT3 (C) on the plane formed by the first two PC modes, and the involved MD simulations are labeled in colors. Color bars show the timescale in corresponding trajectories in ns that the darker color of a node indicate that the represented structure was sampled later in the original MD simulation. D, E and F show the free energy landscapes (in k_BT units) obtained from CT1, CT2 and CT3. Black circles indicate the connection between MD simulations sampling from two ends. G, H and I show the clustering result on each combined trajectory, where color codes indicate different inferred communities.

an aqueous solution. According to SASA analysis in Section 3.3.3, collapsed conformations possesses higher stability with the formation of the hydrophobic cluster, resulting in its longer residency time at both simulations, which agrees with results from Song *et al.* [145].

The energy landscape of CT3 (all simulations) was characterized by 8 conformational states from a visual inspection. They were labeled from 1 to 8 in the order of a decreasing population size (Figure 3.8A). The first 3 states account for $\sim$ 20%, $\sim$ 13%, and $\sim$ 13% of the total population, respectively. These are significantly higher proportions than other states. Thus, these 3 states are considered to be the major conformational states of the full sampling space. Their structural characteristics represent the most important conformational ensemble sampled from MDβ (and MDβ-ann), MDα, and MDα-ann, respectively. Compared with average structures computed from the clustering result, their dominant structural features match with structure 1-3. This agreement confirms the ability of the clustering method to capture major conformational characteristics from an MD trajectory. The population sizes of states 4-6 are relatively the same, which are between $\sim$ 4%-6%. Among them, state 4 was characterized from MDα, while both states 5 and 6 were from MDβ. States 7 and 8 comprise the lowest population ($\sim$ 2% each) that are located between two ends of the *misfolding* pathway.

Representative conformations labeled were computed for the 8 lowest free-energy basins (Figure 3.8B). States 1, 5, 6, and 7 correspond to metastable conformational states involved in the structural interconversion observed in MDβ and MDβ-ann that are separated by small free energy barriers. Among them, structure 1 is in collapsed shape; 5 and 6 are extended conformations; while 7 is the transitional structure between them. It is interesting to notice that the C-terminus of structure 5 and 6 form β-sheets with the middle part of peptide on the opposite side. On the other hand, a single energy well (basin 3) represents the collapsed conformations in MDα-ann was characterized. The conformational space is connected from the 2D cPCA representation, thereby only part of molecular details is shown. A longer time of MD simulation initiating from helix-starting-structure or SA simulation starting at a lower temperature (<475 K) is required to complete the mechanism in describing further loss of helical content on the peptide.

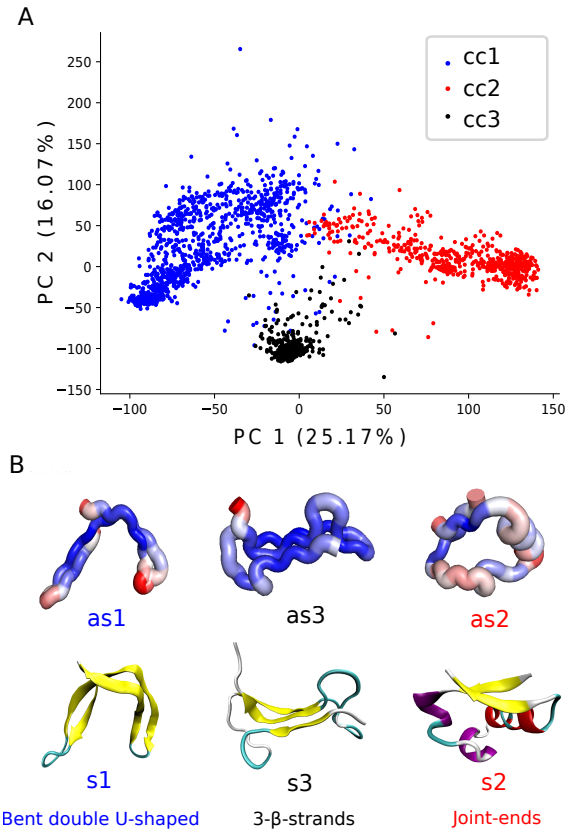

Figure 3.7: Average structures and their corresponding closest frames in trajectories (B) in the representations of community detected of CT3 (A). The average structures (as1, as2 and as3) were generated using PyMol in *putty* mode display. The thickness of the chain in the representation and also the color range of red-white-purple suggest the extent of RMSD changes among all structures in a community. Snapshots of closest structures (s1, s2 and s3) to as1, as2 and as3 and shown below. All structures are presented with the N-terminus on the left and the C-terminus on the right.

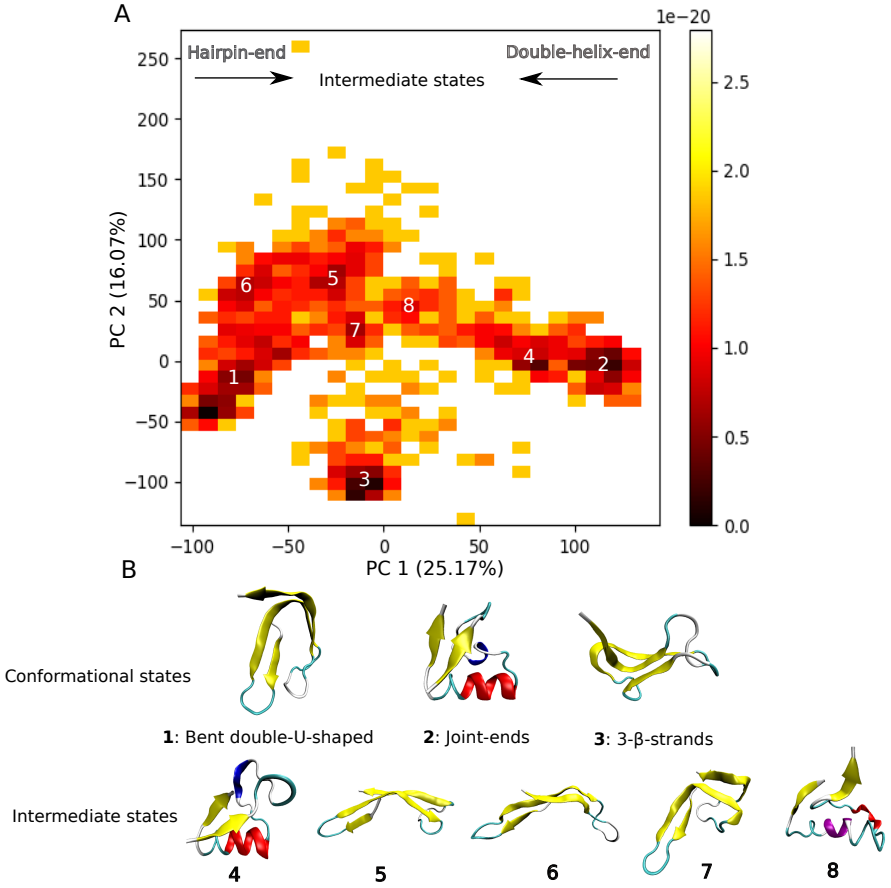

Figure 3.8: (A) Free energy landscape (in k_BT units) of Aβ42 monomer obtained from CT3, shown as a function of principal components PC1 and PC2. The numbers from 1 to 8 label conformational states of the system. States 1-3 are major conformational states, while 4-8 are low-populated intermediate states. (B) Representative structures for free-energy basins defined as $-\ln(density)$ in k_BT units. All structures are presented with the N-terminus on the left and the C-terminus on the right.

3.3.6 Exploring the Misfolding Pathway of the Aβ42 monomer Using Dscretizing Principal Component

Using PCA and clustering, we characterized dominant conformational states; however, transitional structures between these states were not considered in the above approaches. Characterizing transitional conformations could provide useful molecular information about the Aβ monomer folding events. To do this, DPC is performed by binning structures within tiles defined by a grid in the first two dimensions of PCA, resulting in series of average structures. The application of this method ordinates structural features across simulations in a time-independent manner. In other words, one could visualize the coordination space in a simplified and ordered way.

To focus on the *misfolding* pathway, the DPC method is firstly applied on cPCA of CT2 (see Section 3.3.4). Along PC1 (Figure 3.9 CT2), the binning spans from the hairpin-starting-structure end to the helix-starting-structure end. The resulted average structures described the structural transitions from the joint-ends shape to double *U-shaped* conformation (a10-a1) via three steps: (1) in the form of joint-ends structure, there is a gradual loss of helical content especially around the C-terminus; (2) an extended conformation is formed with the complete loss of helices on the peptide; and (3) the double *U-shaped* conformation is generated through the bending motion of the extended conformation. The same method was performed along PC1 on the cPCA projection of CT3 (Figure 3.9 CT3). From Sa1 to Sa4, the open/close folding event is shown from the double-U-shape to the extended conformation. Then, Sa4 to Sa6 display the transtion from the extended to the 3-β-strands structure due to the interactions between two termini. Sa7 to Sa13 describe the conformational formation of joint-ends structure. It is obvious that there is a discontinuity of structural transformation from Sa6 to Sa7. From the PCA projection of CT3, the conformational state of 3-β-strands appear on the location where is in the middle of the pathway along PC1. When binning through this area along PC1, the involvement of the major state dilutes the density of the transitional structures showing transformation from the helix-retaining to the β-rich conformations. Therefore, it is more appropriate to use cPCA of CT2 to display the molecular mechanism of the *misfolding* pathway using DPC.

We also applied this method along PC2 based on cPCA projection of CT2 and

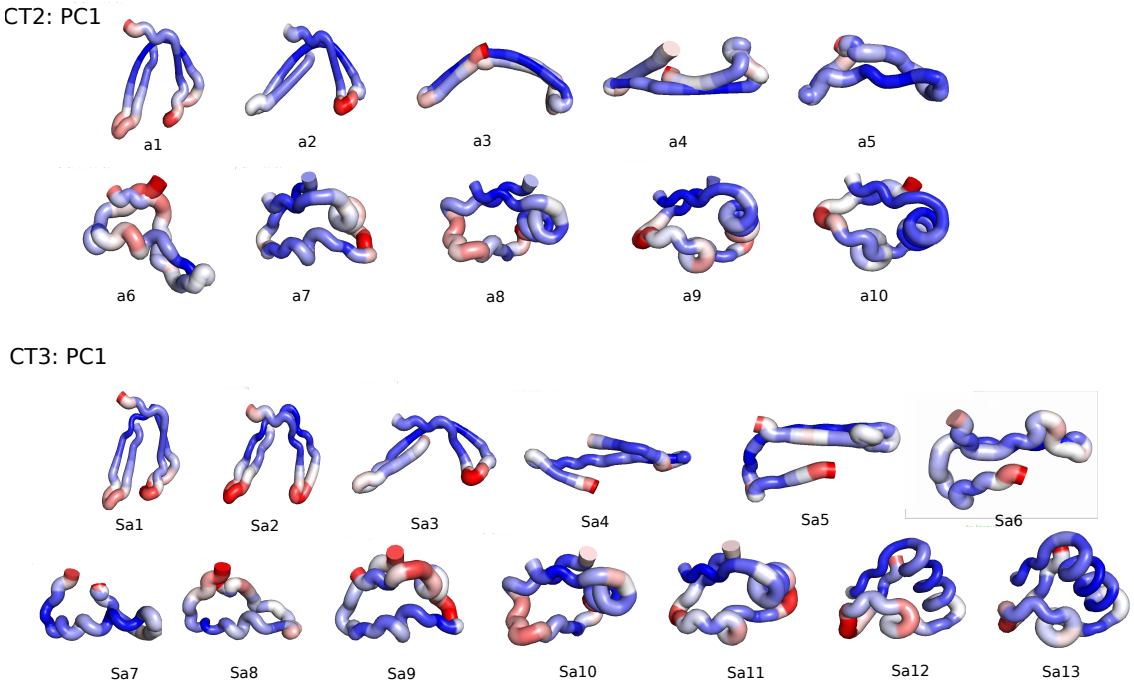

Figure 3.9: Two series of average structures are obtained from DPC along PC1 based on CT2 and CT3, respectively. The thickness and the color range of red-white-purple of the average structures suggest the extent of RMSD changes among structures in the corresponding binning area. All average structures are presented with the N-terminus on the left and the C-terminus on the right.

CT3, respectively. Both of them concentrate on describing the open/close folding event that mainly separates the collapsed and extended conformations on the negative and positive coordinates of PC2 (see Appendix B). As PCA has the ability to abstract the scale of conformational changes in a decreasing order via PC1, PC2, and *etc.*, DPC on the first principal component show the largest-scale motion in the transition from one collapsed conformation to another through extended shapes.

3.4 Discussion

We carried out a *double-ended search* simulation protocol for predicting the route of the *misfolding* pathway of monomeric Aβ42 in water. A membrane-bound structure and one subunit of fibril were selected to represent two ends of a monomer-to-oligomer/fibril pathway and used as starting structures for MD simulations. Distinct conformational ensembles were characterized from different simulations. By performing the MD simulation on the helix-starting structure (or membrane-bound structure), the α-to-β conformational transition was found to start from the two terminal regions. Aβ42 was demonstrated to be more neurotoxic than Aβ40 due to its higher aggregation propensity [218, 264]. It was proposed that this tendency may be related to the reduced flexibility of the C-terminus [264]. It was hypothesized that the Aβ42 has more structure C-terminus with the two additional hydrophobic residues (I41 and A42) compared to the shorter alloform (*i.e.*, Aβ40), whihch was observed in REMD simulations [145, 225]. A recent study reported the last two residues of Aβ42 give a lower barrier on the energy landscape to the primary nucleation pathway compared to the states of Aβ40 [265]. From our results, the conformational transitions on the termini lead to the formation of an antiparallel β-sheet, then trigger the helix-starting-structure to collapse into a state in higher kinetic stability. The transition rate slows down and maintains after the peptide's collapse (Figure 3.3A). Taking all this into consideration, we propose that the conformational transition to β-structure on the terminal regions of Aβ42 give rise to exhibit a kinetically long-preserved intermediate state having a distinct surface shape of the hydrophobic cluster. In an aqueous solution, monomers at such state could interact with each other or the oligomers through the surfaces of their hydrophbic clusters in the nucleation process. Since various collapsed conformations were discovered from different

experiments [144, 145, 227, 261, 266, 267], their unique characteristics may determine the aggregated states, leading to distinct oligomeric and fibrillar conformations [68].

The obtained ensemble from the present simulations revealed two dominant conformational states, collapsed and extended conformations, which was first proposed by Song *et al.* [145]. Structural interconversions were discovered between these two states from MD simulations starting from the hairpin-structure or at high temperatures. The major contributing factor to the dynamics of interconversion is thought to be the formation of the hydrophobic core in the collapsed conformations, where their loop/turn regions usually play an important role in this formation (Figure 3.4). The preference of the peptide to form a core matches the observations from previous studies [145, 258]. The higher probability of Aβ42 monomer staying at collapsed conformation may result in higher chances for the nucleation staring from this conformation. Apart from these, no extended shapes were discovered in MDα. This may indicate that the structural interconversion will only occur after a complete loss of helices. The obtained extended conformations exhibit higher β-content, forming sheets that are the featured structure in oligomeric/fibrillar states of Aβ42. Then, it is proposed that collapsed conformations could initiate the primary nucleation phase, while extended shapes may involve more in the stage of secondary nucleation. Some critical characteristics may be maintained in the further formation of higher-level aggregates such as the β-sheet arrangment.

A conformational equilibium between collapsed and extended double hairpin structure was characterized from the SAMD simulation starting from the *U-shaped* conformation. Intrachain β-hairpin structure of the monomeric Aβ42 is critical for Aβ oligomerization due to such motif extends the hydrophobic surface that exposes to the solvent [89, 92, 268]. In contrast, the hydrophilic N-terminus of the Aβ42 was reported to be primarily unstructured in NMR fibril structures [93, 109, 110]. Observation of a double-hairpin conformation of Aβ42 monomer was reported by Das *et al.* [269] that the formation of a second hairpin is attributed to the hydrophobic clustering between the N-terminus (D1-R5) and the central region (K16-A21). They illustrated an increased population of double hairpin structure in monomers similar to those reported in wild-type Aβ42 oligomers can be achieved by the A2V mutation. A significant population of extended double-hairpin motif were formed under the high

temperature environment from SAMD simulation, while the equilibrium shifts to the collapsed states after returning to the room temperature condition. Topologically, this recurrent interconversion is caused by the loop region between two hairpins bringing loosely packed hydrophobic core. Such equilibrium provides varied probability of visiting both states that enhances the chance of each conformation to follow different aggregation pathway [258, 270].

PCA allows us visualize a diverse conformational space from the combined trajectory data. Performing SA simulations are finding shared states on the 2D cPCA surfaces, suggesting that there exists a pathway from the two ends of the *misfolding*. Our results suggest that the *double-ended search* simulation protocol could be applied to recover a biochemical pathway by simulating from opposite conformational ends of the pathway. By using free energy landscape and clustering, three discrete conformational states were characterized on the whole space. Their agreement in characterized conformational ensemble exhibits the ability of trajectory modularity (or clustering optimization) to capture large-scale conformational changes from folding events of IDPs and distribute samples into well-defined communities/minima that own common structural features, but with distinct characteristics from each other. Besides lowest energy basins, intermediate states were characterized on the free energy landscape along the *misfolding* pathway. However, energy barriers between them may be much higher than the values shown on the calculated landscape since some of the intermedate structures were sampled from SA simulations. Thus, transtions between these states may be slow at physiological temperatures. This is also the phenomenon we observed from MDβ that conversions between collapsed and extended conformations were less frequent than in SA simulations (Figure 3.5B comparing with D). The separate state diverged from the pathway (Figure 3.7community cc3) exhibits very different conformations due to its unique β-sheets arrangements, but similar ranges of β-strands. It is suggested that Aβ42 could be treated as a peptide chain of high flexibility but with frequent emergence of stable low-level structures on the same regions of residues (see Chapter 4 for details).

We were able to capture a major range of motions by discretizing PCs. PCA is a powerful tool to organize geometric covariance in such a way that common features appearing at different time points can be arranged in adjacent space. This particular

way of organizing sampled structures permits identification of a series of continuous major structural motions from a MD trajectory for equilibria visualized in a time-independent manner. Average structures calculated from discrete groups along the first two princial components suggest that the conformational changes follow two large-scale transitions: the secondary structure transition and the interconversion. Based on these, we propose the *misfolding* mechanism that includes at least 5 stages. (1) First, conformational transition starts on the termini or C-terminus. A previous CD study elucidated that a primary nucleation pathway may be initiated from the Aβ C-terminus [271]. The kinetics of transition on this region may strongly affect the nucleation rate. (2) A kinetically stable collapsed conformation is then formed, and the primary nucleation may begin via the face-to-face stacking mode of hydrophobic clusters owned by collapsed structures. The amphiphilic nature of Aβ42 encourages their preferential to maximize the hydrophobic force between the interpeptide hydrophobic patches [226, 272]. (3) The monomers undergo further transition within the collapsed shape until a complete loss of helices occurs. It has been suggested that water molecules are depleted between monomers when hydrophobic attraction brings them close enough [226]. Formation of inter-sheet H-bonds becomes essential at this stage since hydrogen bonds are stronger than hydrophobic interactions. (4) Higher β-content on the peptide leads to various conformations in different β-sheet arrangments. (5) At least 2 types of structural interconversions can be shown in this stage: firstly, collapsed and extended conformations with the same β-sheets arrangement are converted to each other, and secondly, one collapsed conformation could transform to another via extended shapes, leading to the structure of different hydrophobic collapse or an alloform with a different type of β-sheet arrangment. At the last stage, both primary and secondary nucleation processes can occur since the convertion between collapsed and extended shapes opens up a diversity of oligomers and fibrils assembly pathways. It is noted here that the *conformational selection* mechanism may only be applicable in the secondary nucleation [273].

3.5 Conclusions

By performing MD simulations following a *double-ended search* simulation protocol, we were able to predict the route for a plausible *misfolding* pathway starting from

a helical structure and progressing to a hairpin structure. Discrete conformational ensembles were characterized by construction of free energy landscapes using PCA and trajectory modularity along the pathway. The later approach takes advantage of network representation of trajectory structures and community detection on the established network system. Both methods reach an agreement in capturing three major conformational states, showing distinct structural characteristics. Our results suggest the important role of both terminal regions of Aβ42 in triggering conformational transition and forming a compact fold that may initiate the primary nucleation. Furthermore, structural interconversions between collapsed and extended conformations were observed for monomeric structures with higher β-content, showing higher aggregation propensity that may involved in both primary and secondary nucleation processes. Along the complete *misfolding* pathway shown on cPCA space, we applied discretizing PCs following the first two principal components, uncovering transitioning motions between populated ensemble. Taken together, we proposed a possible mechanism for *misfolding* and early stage nucleation pathways including five distinct and sequential structural stages.

Chapter 4

Defining the Homogeneous Segments from the Hetergeneous Tertiary Structure Ensemble of the Aβ (1-42) Peptide

Authors' contributions: Simiao Lu conducted the research and wrote the chapter. Simiao Lu and Christian Blouin designed the experiment and analyzed the data. Simiao Lu, Jose Sergio Hleap and Alex Safatli wrote the codes for trajectory analysis. Christian Blouin provided editorial input and guidance.

4.1 Introduction

Intrinsically disordered proteins (IDPs), polypeptides that lack stable secondary and tertiary structure, comprise 25-30% of protein-encoding sequences found in the human genome. They are involved in many essential biological processes and human diseases. Aβ peptide, one of the most studied IDPs, is the primary constituent of the senile plaque that is widely accepted as a pathological hallmark of Alzheimer's disease. The formation of the senile plaque originates from the appearance of the misfolded Aβ peptide. The complex molecular pathways involve self-assembly events from a misfolded monomer to an insoluble fibrillar deposit made of layered cross β-sheets. Among these aggregates of Aβ, oligomers have been determined as the most neurotoxic species due to their influence in inducing membrane disorder in the extracellular space. The presence of intracellular Aβ oligomers can provoke events of cellular damage including elevated endoplasmic reticulum stress, mitochondria damage, altered proteolysis, *etc* [274–276]. However, characterization of these oligomeric species challenge the classical structural biology in the aqueous media due to their metastability and polymorphism [277]. Although the application of experimental techniques such as NMR [278] and cryo-EM [112] provided insights in understanding the fibril formation in different external conditions, development of analytical/computational tools are still required for an accurate description of statistical properties of the Aβ peptide [17].

Alternative theories have been proposed to describe the structures of monomeric Aβ species in solution at the early stage of the amyloid aggregation. An early study suggested that the amyloid fibrilization may occur by joining two monomers with the topology of two β-strands on the most hydrophobic regions (L17-A21 and A30-V40), separated by a turn in the central hydrophilic region (E22-G29) [279]. Urbanc *et al.* suggested the stacking pattern of an Aβ42 oligomer is formed via intermolecular interactions between I31-A42 C-terminal regions and K16-A21 [280]. Kirkitadze *et al.* and Fezoui *et al.* observed intermediary helical states in the initial lag phase of the aggregation [252, 281]. Based on this, another mechanism was hypothesized that on-pathway helices-containing oligomers, assembled based on partially helical monomers, could transform into β-sheet-containing oligomers and fibrils. A more recent MD study observed the occurrence of the helix-retaining region (V12-G25) on the monomeric Aβ42 in water [282]. They further elucidated that the molecular details of the association between two α-helical monomers is via hydrophobic interactions between central hydrophobic domains. An alternative mechanism was proposed from a recent NMR study that the early-stage intermediate of an Aβ aggregate is formed between monomers in irregular structures without high occupancy of β-content through predominanly hydrophobic driven self-interaction [226]. The *conformational selection* mechanism was also widely accepted for the understanding of the early aggregation. A computational study mimicked the Aβ oligomerization by performing the protein-protein docking between a NMR-determined ring-shaped fibril structure and a characterized monomeric conformation with the same geometry. This result implied that the polymorphism of oligomers could stem from the self-stacking based on various monomeric conformers containing distinctive structural characteristics [270, 283, 284]. From these studies, a logical assumption is raised that many amyloidogenic pathways could exist along the aggregation process starting from self-interaction between Aβ monomers with the particular aggregation-prone structure, leading to structural heterogeneity in oligomeric and fibrillar states through their specific stacking patterns. Therefore, characterizing the structural diversity of the Aβ42 monomer in the aqueous environment is important for understanding the molecular mechanism

of the amyloid aggregation and their association with the AD pathogenesis. A typical NMR-determined helix-kink-helix conformation of the Aβ42 monomer (PDB ID: 1IYT), solved in a non-polar solvent, comprises four distinct regions: a charged N-terminal region (NTR; residues 1-15), a central hydrophobic cluster (CHC, residues 16-21), a hydrophilic turn region (HTR, residues 22-29), and a mostly hydrophobic C-terminal region (CTR; residues 30-42). Previous computational studies revealed the existence of common local conformations on Aβ alloforms in the aqueous medium (Figure 4.1) [43,116,144,145,225,227,258,261,261,266,269,285–289]. When projecting the regions onto an alignment of per-residue secondary structure preferences of the Aβ42 in water from previous simulations, particular characteristics on the similar locations are found on the peptide even through the two-helix conformation was solved in a different environment. This inspires us to hypothesize that the Aβ monomer could be considered as the molten globule IDP that predominantly shows compact states but largely disordered architecture with frequent emergence of stable low-level structures on the same regions of residues. We refer to these regions as structured segments or segments in the following context. If the specific ranges of structural segments can be identified, it is possible to categorize Aβ42's various conformations into architectures based on their interacting patterns between segments.

It is difficult to explore the vast, flat and weakly structured energy landscape of an IDP. By choosing multiple relevant starting configurations, rather than a single starting structure, one could achieve more thorough sampling with affordable runtime. This strategy was applied in other studies for more exploration of protein conformational space using REMD [290, 291]. Previous computational studies characterized alternative conformational states with modified initial structures by manually adding the missing residues to NMR structures of Aβ fragments [116, 145, 262]. Inspired by these studies, we prepared the starting configurations of MD simulations not only based on NMR structures of the Aβ42 monomer but also on subunits taken from various protofibrils solved in NMR and cryo-EM studies. Solid-state NMR (ss-NMR) measurements revealed the formation of two parallel, in-register β-strands between residues around CHC and CTR, separated by the HTR (PDB ID: 2BEG). Further studies reported structural polymorphism within Aβ fibrils, such as a triple β-motif (PDB ID: 2MXU) or a *U-shaped* fold containing 4 β-strands (PDB ID: 5KK3). While

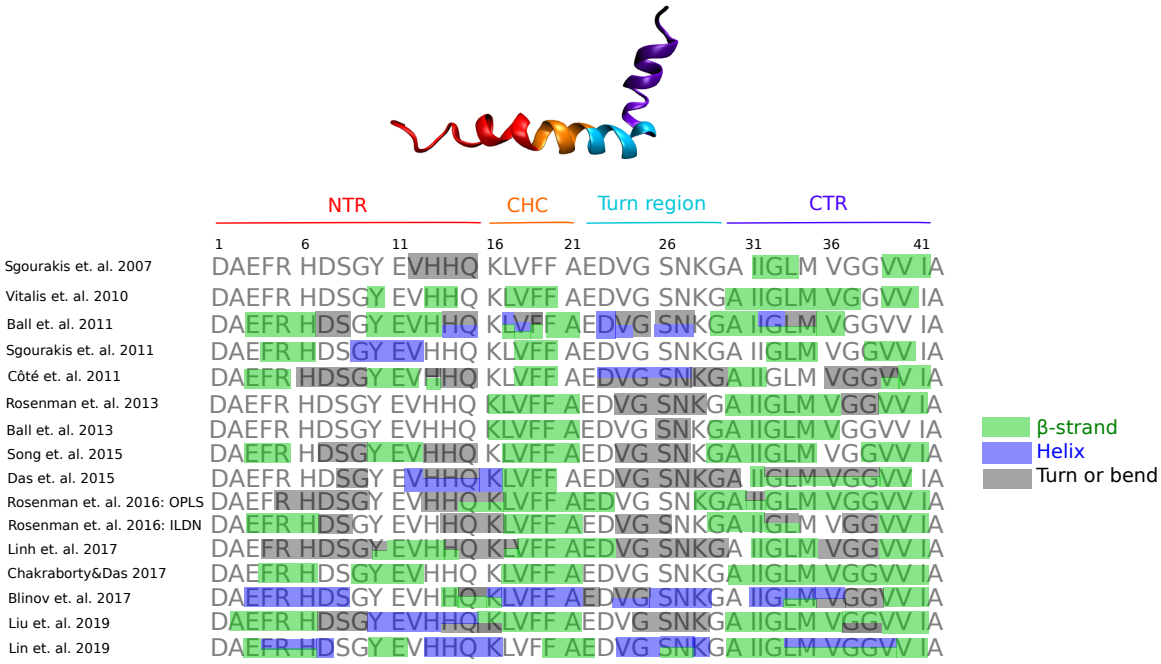

Figure 4.1: Per-residue secondary structure preferences of the monomeric Aβ42 in an aqueous solution are shown from previous simulations, indicating similar local conformations in solution. Divided regions on the peptide are named as NTR, CHC, a turn region and CHC based on a two-helix NMR solved conformation in a non-polar environment (PDB ID: 1IYT).

residues around CHC and CTR are the major contributors to the fibrillar structural stability, the NTR appeared to be mostly disordered from these experiments. However, recent experimental studies suggested that the NTR is an important stabilizing factor to help form more compact fibrillar folds. For example, an NMR and cryo-EM study determined the strucuture of an alternative polymorphic fibril, showing a double-horseshoe-like cross-β-sheet with the NTR partially structured (PDB ID: 2NAO). A following cryo-EM experiment also observed that the NTR formed β-strand conformation in the fibril structure, resulting in an overall *LS-shaped* backbone alignments (PDB ID: 5OQV). In this study, configurations of the monomeric unit fetched from fibril structures mentioned above and two NMR Aβ42 monomeric structures (PDB ID: 1IYT and 1Z0Q) were used as initial structures to perform MD simulations. To obtain plenty of Aβ42 topologies for the precise segmentation and geometric classification, we also used simulated annealing (SA) and replicas for MD samplings. An SA simulation allows for effective global exploration on a rugged energy landscape by starting with a high temperature and cooling the system gradually. A replica refers to a repeated simulation of the same system with identical parameters, while the only change is a randomly generated initial velocity according to a Maxwell distribution.

In the current study, we tried to classify the heterogeneous tertiary structure ensembles of the monomeric Aβ42 in an aqueous solution based on its secondary structure arrangements. We implemented 15 classical and SAMD simulations using the force field of GROMOS96 54a7 in an aggregated simulation time of 8.4 μs starting from 7 different conformations of the Aβ42 monomer. Our simulation data showed good consistency with the three-bond-J-coupling constants and chemical shifts reported in previous experimental work. The secondary structure profiles and ranges of β-strands on the peptide from the present simulations were in agreement with CD estimations and NMR measurements, respectively. In the analysis of local structures, A-hinge-B-hinge-C arrangements were observed across all structural ensembles, where A, B, and C are structured segments on the peptide determined from the secondary structure analysis. Tertiary structure analysis using the frequency contact matrix also indicated that the peptide adopted mainly three regions, in agreement with A, B, and C in lengths. Therefore, various conformations of the Aβ peptide can be described

and categorized by intra-region interactions, resulting in four contact patterns. Our results suggested that there exists consistent local conformational segments on the intrinsically disordered Aβ42 peptide. These segments are used as building blocks for the peptide's topological construction, giving some frequently emerging motifs coupling with tertiary-level flexibility. Our classification of Aβ monomeric topologies helps us discretize conformational variability of the peptide and matches these structures to different assembly pathways in the early stage of the Aβ42 aggregation proposed from previous studies in the literature.

4.2 Methods

4.2.1 System Setup

Seven starting structures in three conformations, helix, U-shape, and S-shape were downloaded from the PDB [292] (www.rcsb.org) and prepared for MD simulations. Descriptions on Aβ42 monomeric and fibrillar conformations were provided in Sections 1.3.4 and 1.3.6. The helix-starting conformations (Helix-np and Helix-pl) were fetched from two NMR structures of Aβ42 monomers solved in non-polar (PDB ID: 1IYT) and polar (PDB ID: 1Z0Q) solvents, respectively (see Figure 4.2). The U-shaped-starting conformation was obtained from one of the subunits of the NMR solved Aβ42 fibril structure (PDB ID: 2BEG) (See Figure 4.2). Four *S-shaped*-starting structures (*i.e.*, S-shape1, S-shape2, S-shape3, and S-shape4) were prepared by slicing subunits from three ss-NMR (PDB ID: 2MXU, 5KK3, and 2NAO) and one cryo-EM (PDB ID: 5OQV) Aβ42 fibril structures. Detailed information including methods, structural features, and experimental conditions of the original protein structures were summarized in Table 4.1. Missing residues of the N-terminus were modelled into each monomer taken from the fibril structure by using PyMol [233] to create the desired length of 42-residue. Each prepared monomeric structure was placed in the center of a rhombic dodecahedron box with box edge 1.0 nm away from the peptide. Each dodecahedron box was solvated with water molecules using extended simple point charge (SPC/E) model and neutralized by counterions (Cl^- and Na^+).

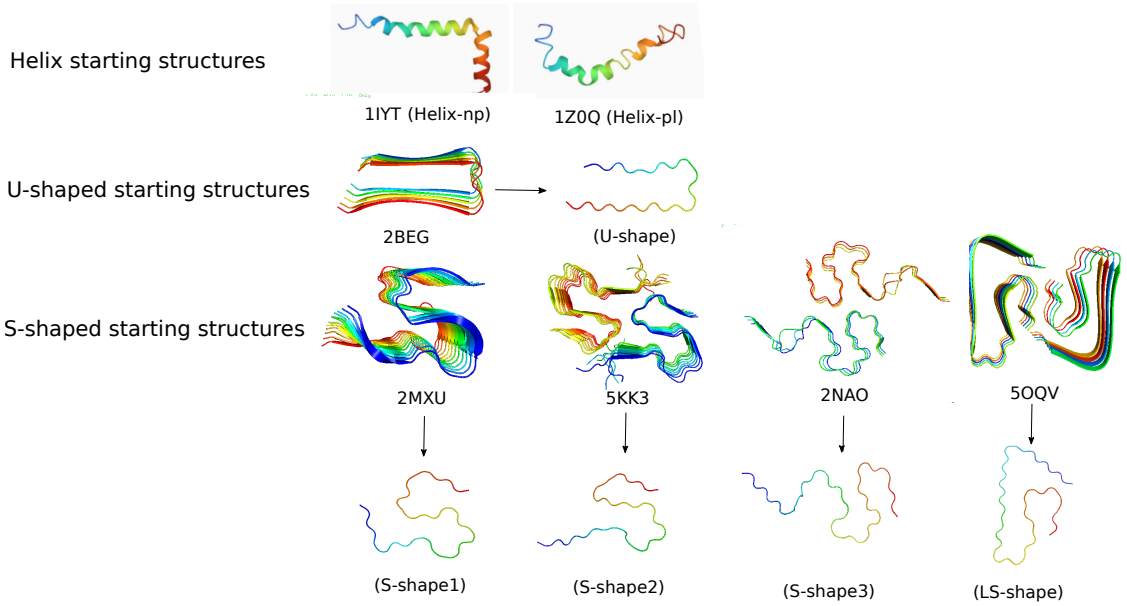

Figure 4.2: Starting structures setup for MD simulations, including 7 structures in 3 conformations.

4.2.2 MD Simulations of Aβ Monomers

All simulations of 7 Aβ monomeric systems (see Table 4.2) were carried out with GROMACS 5.1.4 package [234], applying the GROMOS96 54a7 force field on the peptides and solvents [175]. Classical and SAMD simulations (MD1, MD4, MD6, and MD9) implemented in Chapter 3, starting from Helix-np and U-shape, were involved in the following trajectory analysis protocol. Each of these simulations were duplicated for 400 ns with different initial velocities to assess reproducibility (MD2, MD3, MD5, MD7, MD8, and MD10). Related simulation details were discussed in the Method section of Chapter 3.

Simulations starting from Helix-pl and 4 *S-shaped* structures were performed for 400 ns, respectively (MD11-MD15). After solvation and addition of ions, the potential energy of each system was minimized using the steepest descent algorithm for 500 steps. The leapfrog Verlet integration algorithm was used with an integration time step of 2 fs. Periodic boundary conditions were implemented in x, y, and z directions. The long range electrostatics interactions were calculated using the Particle Mesh Ewald summation with a Fourier grid spacing of 0.12 nm (MD12-MD15) or 0.14 nm (MD11) [236]. Coulomb and van der Waals cut-off distances were both set to 1.0 nm. The Verlet cut-off scheme [237] was used to reach high performance when computing non-bonded interactions. A maximum force less than $100\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$ was obtained for both systems at the end of the energy minimization. NVT ensemble was conducted for 100 ps. The LINCS method [238] was used to restrain all bonds for an integration step of 2 fs. The protein and the solvent were coupled separately to a modified Berendsen thermostat called *V-rescale* [239] at 300 K. NPT ensemble was then conducted for 100 ps as well. Pressure coupling was requested in this phase to the Parrinello-Rahman barostat [240, 241] at 1.0 bar with a compressibility of 4.5×10^{-5} bar. In the production run, all parameters were set to the same step of equilibration and the coordinates of the system were saved every 20 ps.

Table 4.1: Summary of Simulation Systems.

PDB ID	Method	Monomeric or Subunit Length	Structure	Experimental Condition	Modification[a]	Starting Structure Label
1IYT[b]	*CD, NMR*	*1-42*	helix-kink-helix (Boomerang)	HFIP[c]/H_2O 80:20 (v/v)	-	Helix-np
2BEG[d]	*NMR*	*18-42*	parallel, in-register β-sheets	10 mM Tris DCl, 150 mM sodium chloride,D_2O	Spliced subunit, N-terminus added	U-shape
1Z0Q[e]	*NMR*	*1-42*	double-helix, less helical contents on N-terminus	HFIP/H_2O 30:70 (v/v)	-	Helix-pl
2MXU[f]	*ss-NMR*	*11-42*	triple-β-motif	10 mM sodium phosphate, H_2O	Spliced subunit, N-terminus added	S-shape1
5KK3[g]	*ss-NMR*	*15-42*	four β-strands in a S-shaped	20 mM sodium phosphate, 0.2 mM EDTA,[g] 0.02% sodium azide, H_2O	Spliced subunit, N-terminus added	S-shape2
2NAO[h]	*ss-NMR, EM*	*1-42*	double-horseshoe-like cross-β-sheet	100 mM sodium chloride, 100 mM sodium phosphate, 100 uM zinc chloride, 95% H_2O/5% D_2O	Spliced subunit	S-shape3
5OQV[i]	*ss-NMR, cryo-EM*	*1-42*	LS-shaped	30% v/v ACN,[j] 0.1% (v/v) TFA,[k] H_2O	Spliced subunit	S-shape4 (or LS-shape)

[a] Modification made for starting structures preparation [b] [81] [c] hexafluoroisopropanol [d] [93] [e] [83] [f] [110] [g] [108] [h] [109] [i] acetonitril [j] trifluoroacetic acid [k] [58]

Table 4.2: Summary of MD simulations.

Starting Structure	MD Type	Time (ns)	Temperature (K)	Experiment	Simulation Label
Helix-np	classic	1200	300	-	MD1-1, MD1-2, MD1-3
	classic	400	300	replicate	MD2
	classic	400	300	replicate	MD3
	simuated annealing	800	475 - 300	-	MD4-1, MD4-2
	simuated annealing	400	475 - 300	annealing replicate	MD5
U-shape	classic	1200	300	-	MD6-1, MD6-2, MD6-3
	classic	400	300	replicate	MD7
	classic	400	300	replicate	MD8
	simulated annealing	800	475 - 300	-	MD9-1, MD9-2
	simulated annealing	400	475 - 300	annealing replicate	MD10
Helix-pl	classic	400	300	-	MD11
S-shape1	classic	400	300	-	MD12
S-shape2	classic	400	300	-	MD13
S-shape3	classic	400	300	-	MD14
S-shape4 (or LS-shape)	classic	400	300	-	MD15

4.2.3 Prediction of NMR Observables from Simulations

Predicted J-coupling constants ($^3J_{H^N H^A}$) and chemical shifts from the MD trajectories were quantitatively compared with corresponding experimental data for validations of the obtained ensembles. Chemical shifts of C_α, C_β, N, and H_α were calculated from simulation coordinates, whereas J-couplings were back-calculated from the simulation coordinates using the Karplus equation:

$$J = A\cos^2(\phi - 60^\circ) + B\cos(\phi - 60^\circ) + C \tag{4.1}$$

where, A, B, and C are semi-empirically determined parameters and ϕ is the peptide dihedral angle. The STRIDE [293] program was used to predict the ϕ angles for each residue of Aβ42 at every selected snapshot from each MD simulation and the combined trajectory. Five sets of Karplus parameters [225, 266, 294, 295] (Table 4.3) were used for computational derived J-couplings in comparison with previously published data [225, 226, 266, 296].

Table 4.3: Different sets of Karplus parameters used for J-coupling back-calculations.

	A	B	C
Vuister and Bax (1993)	6.51	-1.76	1.60
Sgourakis *et al.* (2007)	7.70	-1.90	0.06
Vögeli *et al.* (2007)	8.40	-1.36	0.33
Rosenman *et al.* (2013)[a]	6.92	-1.62	0.55
Rosenman *et al.* (2013)[b]	6.88	-6.50	-3.53

[a] for Aβ-M35ox(S), where S stands for l-methionine-(S)-sulfoxide [b] for Aβ42

Both RMSD and Pearson correlation coefficient (PCC) were used to examine the correlations of these observables between experiments and the MD ensemble average, and PCC is defined as:

$$PCC = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)\sigma_x\sigma_y} \tag{4.2}$$

where x and y are data setes in size n with the standard deviation σ_x, and σ_y, respectively.

4.2.4 Trajectory Analysis Protocol

Input Data

Each MD trajectory was sampled for every 20 ps with the first 100 ns discarded due to large structural changes. All 15 MD trajectories were combined into a single input file, resulting in a combined trajectory of 270,000 frames in the total simulation time of 8.4 μs. All structures were aligned using generalized procrustes analysis (GPA) [245, 246] to the estimated mean structure by rotating and translating, while scaling is not considered in this case. The mean structure was re-calculated and updated in each iteration of the alignment until the convergence when the mean structure does not change significantly within an iteration. The process of this alignment was implemented based on the R package *shapes* [248]. The aligned frames are used as the geometric input for validation of NMR observables, secondary structure analysis, RMSD changes against simulation time, and identification of contact patterns. For the efficiency, 2700 aligned frames (every 20 ns) were selected as the input data for the clustering and constructing the monomeric Aβ42 conformational space.

Secondary Structure Logo

A sequence logo is a graphical method for displaying patterns in DNA/RNA or protein sequence conservation, providing assistance in discovering and analysing these patterns [297]. Based on this method, we designed a secondary structure logo for presenting consensus and diversity of secondary structure elements on a protein chain. Secondary structure profiles were collected from the combined trajectory using the DSSP algorithm [251] for each structure. Each residue in the sequences was recoded to a one-letter secondary structure classification. The secondary structure logo was created from the collection of recoded sequences using a web server named WebLogo [298].

Aggregated Start-Stop Matrix

A start-stop matrix is defined for recording regions with continuous and consistent signals of α-helical and β-strand structures on a protein structure. This matrix is a

symmetric binary matrix, in which its columns and rows are residue numbers. Considering a MD trajectory of n structures s_1,..., s_n, a start-stop matrix was created for each structure as A_1,..., A_n. Within a matrix A_x, if an ordered segment was detected between residues i and j on structure s_x, element $A[i, j]$ returned 1, otherwise, it returned 0, where $1 \leq i \leq 42$ and $1 \leq j \leq 42$ in this case. The aggregated start-stop matrix A_{agg} is the sum of all start-stop matrices and expressed as:

$$A_{agg} = \sum_{x=1}^{n} A_x. \tag{4.3}$$

A higher value of an element in A_{agg} indicates higher conservation of an ordered segment of the protein in its conformational dynamics.

Clustering Analysis Using Different Similarity Criteria

In a graph representation, let each structure in the trajectory be abstracted as a node. The graph was fully connected with weighted edges. The weight was computed via a similarity metric, $\frac{1}{RMSD + 0.00001}$ (R-score), where RMSD was calculated at C_α atoms of each residue. The addition to the denominator was to avoid calculation errors in the unlikely case of two identical structures given the level of precision of the coordinate system. Community structure was detected using a modularity-based algorithm called fast-greedy approach [194] on the defined graph, yielding inferred communities or clusters of similar structures. An average structure was calculated and displayed in the *cartoon* mode (using PyMOL ver. 1.4.1 [233]) to represent the common structural characteristics owned by a community. Detailed description on this clustering analysis is provided in the Method section of Chapter 3.

Clustering analysis based on inter-residue contacts was also carried out for comparing with the assessed community structure using the R-score. Instead of using a set of atomic coordinates, the 3D structure of a protein can also be described by a set of spatial interactions, such as residue-residue contacts. Assume that structures s_i and s_j in the MD trajectory are presented by contact matrices M_i and M_j. To evaluate the similarity between them, M_i and M_j were converted into a 1D entry of integers, named as string S_i and S_j, respectively. A Hamming distance ($HD_{S_{ij}}$ [299]

was calculated between S_i and S_j, which measures the minimum number of substitutions required to change S_i to S_j. The equation $\frac{1}{HD_{S_{ij}}}$ (H-score) was used as a second structural similarity criterion instead of the R-score for the community detection on the defined graph, where a larger value of the score indicates a higher shared structural similarity. Other procedures in this clustering analysis are the same as the demonstration above.

Principal Component Analysis

This section is identical to the paragraph Principal Component Analysis of MD Simulations in Section 3.2.4 of Chapter 3.

PCA on the combined trajectory PCA was performed on a data matrix converted from the Cartesian coordinates of structures in the combined trajectory. In the data matrix, each row is a structure and each column corresponds to coordinates (x, y, z) of centroids of all residues in the structures, *i.e.*, 42×3 columns. A linear transformation is used in the data matrix to generate a $3N \times 3N$ covariance matrix C. Diagonalization of the covariance matrix generates $3N$ eigenvectors (v^i) and eigenvalues (λ_i). The eigenvectors gives a vectorial description of each component of the collective motion with indication of the direction of the motion, while a eigenvalue for the corresponding eigenvector represent the amount of contribution of this particular component of the motion. Thus, a time-independent low-dimensional conformational space of Aβ42 monomer can be constructed by projecting the trajectory with the first two PCs. In this work, PCA was performed by using the pca function provided in the package Scikit-learn [249] in Python 2.7 [250]. Visualization of inferred communities determined by the modularity optimization was achieved by projecting the membership vector on the 2D PCA space.

Construction of the free energy landscape On the basis of cPCA, the free energy landscape (FEL) was constructed via,

$$\Delta G(\mathbf{r}) = -k_B T[\ln P(\mathbf{r}) - \ln P_{max}] \tag{4.4}$$

where P is the probability distribution of the molecular system along the coordinate $\mathbf{r}$. Here, $\mathbf{r}$ is defined as the number of points in each unit cell over the total number of points, where the size of each cells equals to $n \times n$ that is resulted from

the pixelation of the 2D cPCA space. P_{max} denotes the maximum of $\mathbf{r}$, which is substrated to ensure that $\Delta G = 0$ is the lowest free energy minimum. Here, the 2D PCA space is pixelated by the value of $n = 10$.

Definition of Contacts

The 3D structure of a protein can be expressed as a contact map, in which residue-residue contacts (or contacts) are pairs of spatially close residues. The definition of contacts are mainly using C_α-C_α or C_β-C_β distances with a threshold at 7 or 8 $\mathring{A}$ [300, 301]. However, the accuracy of contacts predictions to the true contacts are relatively low. Here, we define that a contact between two residues exists if the distance between them is equal or less than 4.5 $\mathring{A}$ for all atoms. Thus, a contact map/matrix of a protein structure is a symmetric binary matrix, where rows and columns are residue numbers. The matrix is in the form:

$$M = \begin{pmatrix} a_{11} & \dots & a_{1n} \\ a_{21} & \dots & a_{2n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}$$

where $n = 42$ in this case, and each element equals to either 1 in the case of existing contact between corresponded residues, or 0 indicating no contacts detected. Such an approach is more accurate with the considerations of the distances between every pair of atoms in two side chains, which is chosen according to the results of previous studies [201, 302].

Frequency Contact Matrix

A contact matrix describes the 3D protein structure focusing on its spatial interactions between residues. The residues frequently involved in such interactions or contacts make great contribution to the structural stability at the tertiary-level. To identify these residues and their contacts, a frequency contact matrix was generated from the combined trajectory. Considering a MD trajectory of n structures, s_1,..., s_n, a binary contact matrix was created for each structure as M_1,..., M_n according to the contact definition (Section 4.2.4). Then, a frequency contact matrix is expressed as:

$$M_{freq} = \frac{\sum_{x=1}^{n} M_x}{n}. \tag{4.5}$$

The contact frequency $M_{freq}[i, j]$ between residues i and j equals to the aggregated number of such contact divided by n, in which $1 \leq i \leq 42$ and $1 \leq j \leq 42$ in this case. The value of $M_{freq}[i, j]$ ranges between 0 to 1, where 0 means that there is no contact between corresponded residues across all structures, and 1 indicates the constantly existence of such contact by definition from the present simulation data. The objective of this approach is to visualize the frequency of contacts across an ensemble of simulations. A previous study gave a similar definition of *contact probability* [303]. This application in characterizing the interatomic interaction of Aβ fibril models provided useful information of their assembly patterns [304].

Definitions of Salt Bridge (SB)

A rigorous definition of SB was firstly adopted based on Kumar and Nussinov's paper [305] for the consideration of SBs in good geometries, where a SB is formed when (1) the distance of the oppositely charged functional group centroids is less than 4.0 $\mathring{A}$, and (2) at least one pair of side-chain carboxyl oxygen atoms and side-chain nitrogen atoms between the functional groups is within a 4.0 $\mathring{A}$ distance. Such a restrictive definition may underestimate the SB population due to the ignorance of a major type of SB forming from the backside of arginine and SB flexibility of the Aβ peptide [306]. The application of this restrictive definition resulted in overall low populations of SBs (see Appendix F Figure F.1). Thus, the definition of Barlow and Thornton [307] was then used to detect all possible SBs with an acidic residue defined as interacting with a basic residue if any N-O atom pair is within a 4.0 $\mathring{A}$ cutoff. Only results obtained from the second definition is discussed in the current chapter.

Frequency Hydrogen Bond Matrix

The STRIDE method [293] was employed to assign backbone hydrogen bonds (HBs) from the present simulation data, where the hydrogen bond energy was calculated using an empirical energy function [308].

A backbone hydrogen bond matrix was used to present the intramolecular hydrogen bonding pattern of a protein structure. In a MD trajectory of n structures s_1,..., s_n, a backbone hydrogen bond matrix was generated for each structure as HB_1,..., HB_n, where rows and columns represent the donor and acceptor residue indices, respectively. Within a matrix H_x, if a HB was determined between the donor residue i and the acceptor residue j on the structure s_x, element $HB[i, j]$ returns 1, otherwise, it returns 0. Here, $1 \leq i \leq 42$ and $1 \leq j \leq 42$. The frequency hydrogen bond matrix HB_{freq} was generated based on the aggregated hydrogen bonding matrices and expressed as:

$$HB_{freq} = \frac{\sum_{x=1}^{n} HB_x}{n}. \tag{4.6}$$

The value of an element $HB_{freq}[i, j]$ ranges between 0 to 1, where 0 means that there is no HB between corresponded residues i and j across all structures, and 1 indicates the constantly presence of such HB in the current simulations.

4.3 Results

4.3.1 Comparison of Calculated $^3J_{HNHA}$ and Chemical Shifts to Experimental Values

NMR spectroscopy is a well-suited experimental technology to study structures and dynamics of IDPs *in vitro* and *in vivo*. Typically, chemical shifts, residual dipolar couplings, J-couplings, and NOEs are used to investigate the propensities of secondary structures [309]. Validation of the simulation samplings was made by comparing the back-calculated $^3J_{HNHA}$ values and chemical shifts for NMR measurements. We performed a total of 15 classical and SAMD simulations of length 0.4-1.2 μs that start from 7 monomeric structures of Aβ42, yielding various transient states in distinct conformations. To validate sampled conformational ensemble from different MD simulations, we characterized 13 most populated transient states by applying community detection method (or clustering) and PCA on the combined trajectory over all production ensembles.

J-coupling is a scalar splitting between two correlated magnetic nuclei mediated by

the intervening electrons through bonds, which generally display a strong dependence on the molecular geometry. The $^{3}J_{HNHA}$, a three-bond J-coupling, correlates with secondary structure in polypeptides. The measurement of $^{3}J_{HNHA}$ is described as a function of backbone angle ϕ in the empirically parameterized Karplus equation, which not only takes accounts of the dependence of the coupling on the intervening torsion angle but the impact by electronegativity of substituents. Thus, the values of these couplings can be back-calculated based on the ensemble average of the dihedral angles derived from the simulation data. The magnitude of $^{3}J_{HNHA}$ provides a quick estimation of secondary structure content that low values (< 5.5 Hz) indicate α-helix, high values (> 8.0 Hz) are correlated with β-sheet, and in the range of 5.5-8.0 Hz is associated with random coil or a mixed ensemble of states [310].

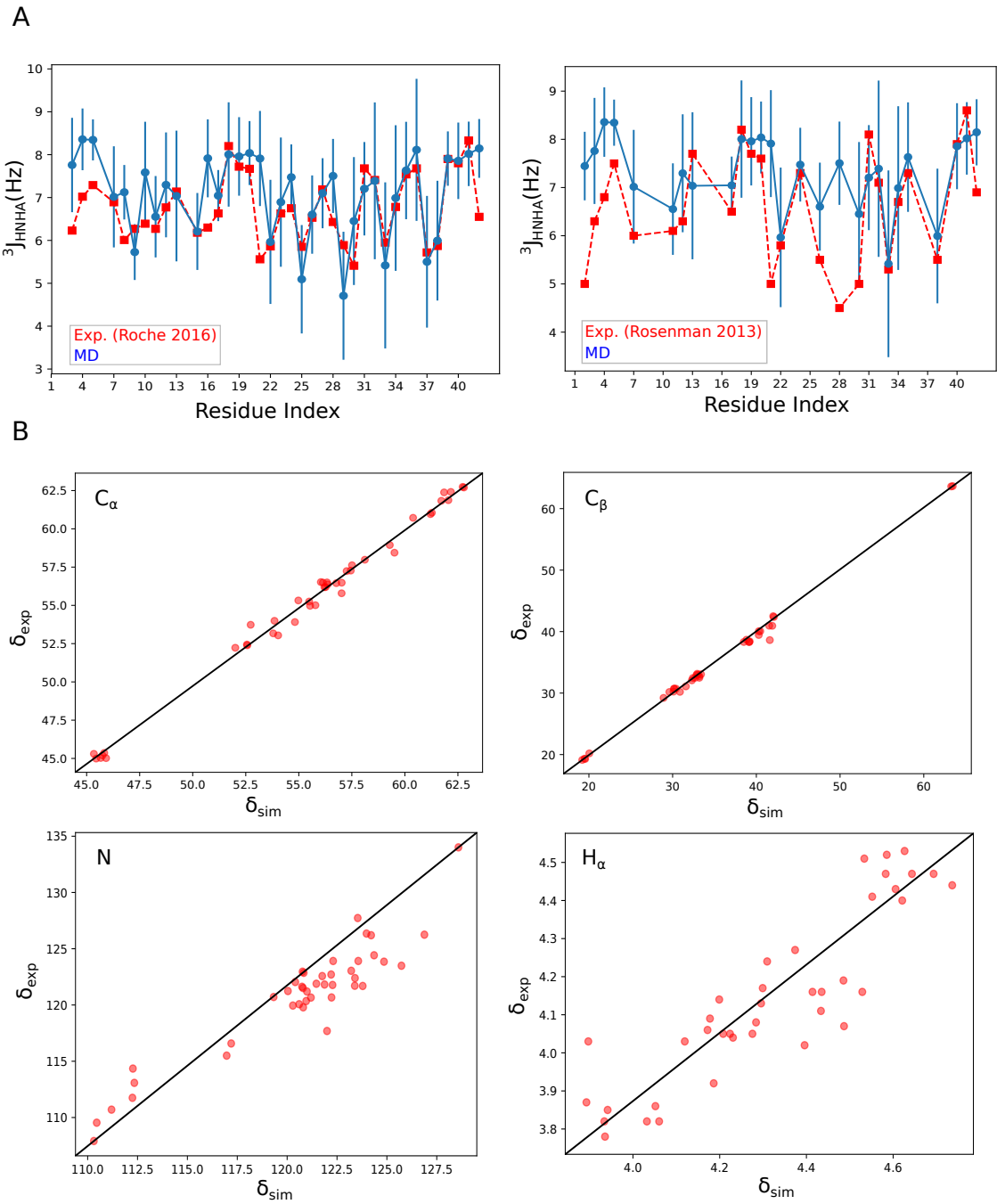

Figure 4.3: (A) The back-calculated $^3J_{HNHA}$ constants from the coordinates of the present simulations compared with NMR measurements from Roche *et al.* [226]. The simulation data were plotted as blue circles, and the error bars denote the standard deviations among characterized conformational states from 15 MD trajectories. (B) Correlation of predicted chemical shifts from the Aβ42 production ensemble compared to experimentally determined values from Wälti *et al.* [311].

Our simulations showed a generally good consistency with a recent and earlier reported experimental $^3J_{HNHA}$ values [226, 266]. Different parameter sets were inserted into the Karplus equation for examining the sensitivity of back-calculated J-coupling constants (see Section 4.2.3). The best fitting set, Vuister and Bax's Karplus equation parameter set [294], was determined by a lower RMSD (0.818 and 1.196 Hz) and higher PCC (0.648 and 0.529) (Table 4.4). Our results had less or comparable agreement with experimental data in contrast with previous computational studies using other force fields [144,145,225,227,254]. As observed in Figure 4.3A on the left, residues E3-R5, S8, Q15 at the flexible N-terminus, F20 at the hydrophobic patch, D23, N27, G29 at the turn region and C-terminus A42 exhibited higher values and the largest deviations from the experimental data. In the comparison with experimental measurements from Rosenman *et al.* [266], the N-terminus region (A2-R5, D7, and E11-V12) F20, S26, K28, A30, and A42 displayed the most differences. Combining these, the high flexibility over the region of F20-A30 with multiple charges may cause the disagreement between our result and other experimental data. The bias towards high values on the N-terminus region and A42 may indicate a possible over-sampling of β-conformations when using the GROMOS96 54a7 force field. Small differences of PCC values were detected by altering different parameter sets listed in Section 4.2.3. It was also suggested by Rosenman [312] that the Vuister and Bax parameters of the Karplus equation may not suitable for the system of Aβ monomer (Aβ40 in their case). The segments of E3-R5, V18-A21, M35-V36, and V39-A42 showed the largest $^3J_{HNHA}$ values, which indicates a higher tendency to form β-conformations. This generally matches observations from a recent computational study [254]. Large standard deviations of computed $^3J_{HNHA}$ values from the simulation data were observed as various conformational states were characterized from simulations with different starting point. This observation was also reported from a previous computational study using the GROMOS force field [145]. These variations may imply the conformational plasticity of Aβ42 in the aqueous medium within the microsecond-scale simulations. Rosenman *et al.* suggested that long timescale simulations are required for convergence in conformational sampling of Aβ monomers, and further improvement in the consistency of properties comparison with the experimental ensemble.

Table 4.4: PCC and RMSD of calculated $^3J_{HNHA}$ and chemical shifts for the present simulations compared to experimentally and computationally determined values.

NMR Observables		RMSD(Hz/ppm)	PCC
$^3J_{HNHA}$	Roche 2016	0.818	0.648
	Rosenman 2013	1.196	0.529
Chemical shifts	C_α	0.501	0.996
	C_β	0.677	0.998
	N	1.753	0.938
	H_α	0.204	0.894

Chemical shifts are the most information-rich data obtained by NMR spectroscopy in the structural study of IDPs. The observed chemical shifts of IDPs are the average of interconverting conformers in the timescale up to milliseconds since IDPs exist as interconverting conformers in solution which is also observed for Aβ monomers from our result (see Chapter 3 Section 3.3.3) and previous computational studies [145,258]. In this work, chemical shifts are predicted from the simulation data using a hybrid chemical shift calculator SHIFTX2 [313] that combines a structure-based shift prediction algorithm (SHIFTX+ [314]) and a sequence-based alignment method (SHIFTY+ [315]). These computational programs were paramterized based on a training set of ordered proteins, thus, they may not be competent for the structural study of IDPs. Nevertheless, SHIFTX2 (or SHIFTX+) and SPARTA+ [316] are widely utilized and seem to be feasible in the previously relevant computational studies [227, 254, 312, 317].

Predicted chemical shifts yield a generally good agreement with the experimental data. Figure 4.3B depicts the correlation between calculated chemical shifts from simulated ensemble with NMR measurements from Wälti *et al.* [311] for C_α, C_β, N, and H_α. The performance of consistency was evaluated in terms of RMSD and PCC (Table 4.4). The primary PCCs of C_α, C_β and N are of the similar magnitudes as the values reported in SHIFTX2 [313], with the exception of H_α. Compared with RMSD or PCC cited in previous computational studies simulating Aβ using other force fields [227,254,312], our results displayed a better or comparable performance. Taken together, the good agreement with NMR observables indicated that GROMOS96 54a7 produces a reliable prediction of the conformational shape of the Aβ42 monomer. Earlier studies suggested that both alloforms may adopt similar secondary structure content in an aqueous solution [145, 226, 254, 312]. We then compared previously

reported $^{3}J_{HNHA}$ and chemical shifts of the Aβ40 and Aβ42 with the present results. Although the soluble Aβ42 was described as close to random coil in conformation according to the previously cited literatures for experimental measurements [226, 311], significant local conformations were observed on the characterized simulation ensemble. This controversial result, simutaneously showing a high consistency with NMR observables, may indicate that the average of many simulation ensembles of the Aβ monomer in solution is close to a random coil.

4.3.2 Secondary Structure

Distinct morphologies of the Aβ42 monomer in solution were noticed by simulating from different starting points. A diversity of local structural features were assigned using the DSSP algorithm [251]. Table 4.5 summarizes the average occupancies of helices and β-structures of the Aβ42 in MD simulations from different types of initial structures. Among them, we observed $\sim 8.7\%$ and $\sim 24.3\%$ of helices and β-content on the monomer from MD1 (1.2 μs simulation starting from Helix-np), which is in high consistency with the CD spectroscopy results [252, 253]. Higher helical content were identified from replicated but shorter MD simulations (MD2 and MD3) starting from the membrane-bound structure (Helix-np), which were $\sim 22.9\%$ and $\sim 25.4\%$, respectively. A decreasing proportion of helices on the peptide is expected with longer simulation time in the aqueous medium. We observed $\sim 13.5\%$ and $\sim 23.0\%$ of β-content of Aβ42 in MD2 and MD3, indicating alternative local conformational features possessed in the sampled ensemble from different replicated MD simulations with randomly generated initial velocities. The NMR structure Helix-pl (PDB ID: 1Z0Q) was solved in a polar solution (30:70 HFIP[1]/H_2O), showing significantly retained helices on the N-terminus. The structural ensemble obtained from the 400-ns simulation starting from this Helix-pl (*i.e.,* MD14) still showed a high proportion of helices in an aqueous solution, which was $\sim$ 37.5%. Simulations (MD6-8 and MD12-15) starting from sliced subunits (*i.e.,* the *U-shaped* and *S-shaped* monomers) exhibited a preference of β-structure in the secondary structure compositions with a low or none helical content. A richer β-content of the monomer was observed in MD6 with longer simulation time. Comparing among simulations with starting

[1]Hexafluoroisopropanol

configurations taken from different fibril structures, our result suggested that the *U-shaped* conformation may exhibit higher propensity of β-structure formation within the same simulation time. High temperature condition of SA simulations increased the efficiency of forming β-structure on the peptide, which was detailedly discussed in the Chapter 3. Also, previous and present simulation data implied an increasing proportion of β-content on the peptide with longer simulation time in the aqueous environment [145]. Thus, higher occupancies of β characters are expected in longer MD simulations starting from the Helix-pl and S-shapes.

Table 4.5: Proportions of the average secondary structure content of the Aβ42 for MD simulations starting from initial structures in different shapes.

Starting Structure	MD Simulation (μs)	Helix	Beta
Helix-np	MD1 (1.2)	0.087	0.243
	MD2 (0.4)	0.229	0.135
	MD3 (0.4)	0.254	0.230
Helix-np (ann)[a]	MD4 (0.8)	0.001	0.362
	MD5 (0.4)	0.001	0.532
U-shape	MD6 (1.2)	0.000	0.576
	MD7 (0.4)	0.000	0.343
	MD8 (0.4)	0.000	0.259
U-shape (ann)	MD9 (0.8)	0.000	0.531
	MD10 (0.4)	0.000	0.438
Helix-pl	MD11 (0.4)	0.375	0.053
S-shape	MD12 (0.4)	0.010	0.144
	MD13 (0.4)	0.000	0.234
	MD14 (0.4)	0.004	0.178
	MD15 (0.4)	0.000	0.174

[a] SAMD simulation

The two most stated hypotheses on the formation of an Aβ oligomer are based on the β-sheet stacking topology [129, 265, 268, 318] and hydrophobic interaction between compact helical monomers. Considering this, characterizing local conformational features of Aβ peptide becomes one of the most fundamental tasks in understanding the monomer-to-oligomer structural evolution. We observed the emergence of β-conformation on similar locations of the Aβ42 monomer by comparing the most abundant secondary structure profiles over residues with the results of previous computational studies. To accomplish this, we generated a secondary structure logo by

aligning the secondary structure profiles collected from each sample in the combined trajectory (Figure 4.4). Each column of the logo shows the relative frequency of the secondary structure element detected at per-residue across all trajectory samples in a decreasing order. The total height of the column (measured in the unit of bits) represents the information content, which is a value related to the conservation of secondary structure at this location. The scale of the vertical axis was adjusted due to the varied information content of the input. As the most predominant local conformational element is placed at the top of every residue position in the logo, three structured segments were identified with the highest propensity of forming β-strands, locating at E3-R5, Q15-A21, and I31-V40. Per-residue secondary structure preference across various computational studies were summarized in an alignment and compared with the logo. This alignment is identical to the one shown earlier in the chapter [43, 116, 144, 145, 225, 227, 258, 261, 261, 266, 269, 285–289]. Common regions of β-conformation were recognized based on the alignment placing at E3-R5, K16-A21, A30-V36, and V39-A41, which generally agrees with our result. Regions in-between structured segments were defined here as hinges due to their preference in staying at coil/turn/bend. The ranges of hinges roughly match the turn/bend regions reported from previous studies mentioned above. Also, an outstanding signal of retaining helices was showed around region of Y10-F19 mainly from the simulation starting from the Helix-pl (MD11), which agrees with the observation from an earlier study [83]. Combining these, we proposed that the average ensemble of the Aβ42 monomer in the aqueous environment could be seen as conformations containing three structured segments connected by two hinge regions. Hinges fluctuate in lengths on the peptide and provide tertiary-level flexibility, leading to heterogeneity in loop, β-arrangements, and the overall topology. This concept could be used to demonstrate and analyse the conformational variety of the intrinsically disordered Aβ42 from current and previous studies [145, 227, 287].

The fluctuating ranges of these segments contribute to the complex dynamics of Aβ42 monomer in solution, which allows the peptide to easily interchange between conformational states (see Section 3.3.3). It is important to understand the secondary structure dynamics prior to characterizing the varied architectures of the monomeric Aβ42. To do it, we created a zero matrix with rows and columns defined as the

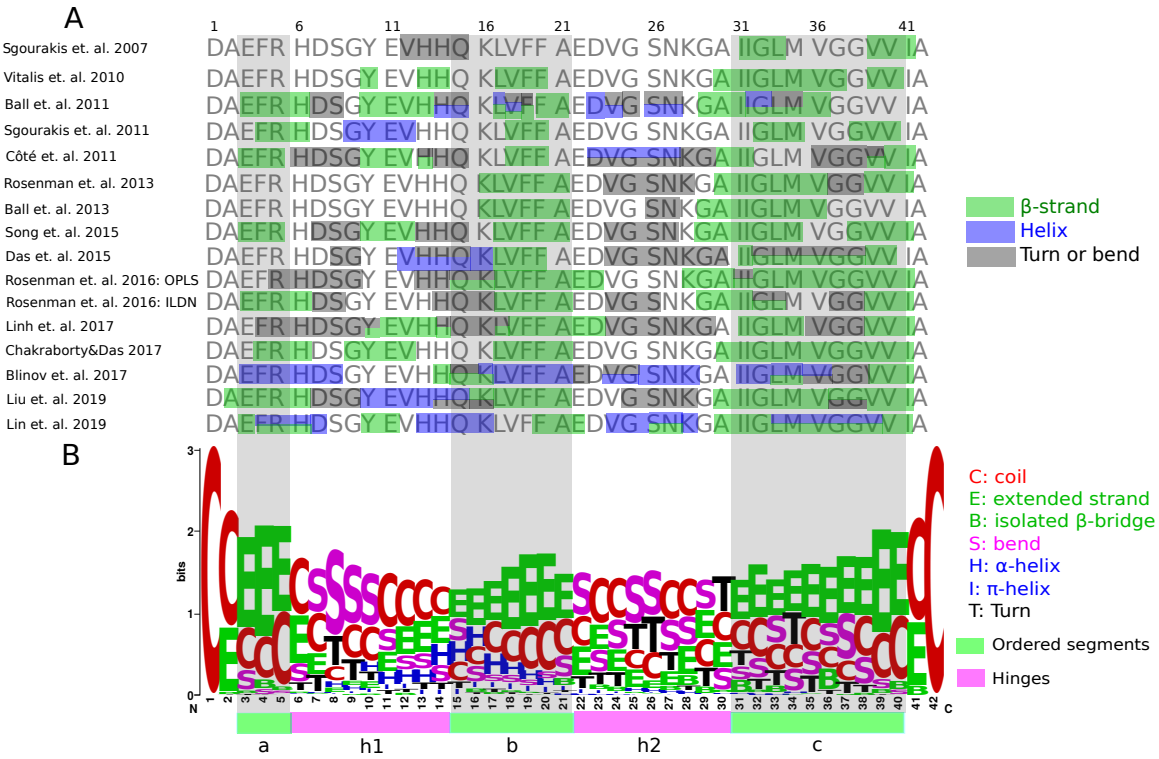

Figure 4.4: Summary of per-residue secondary structure preferences from previous simulations of the monomeric Aβ42 or wild-type Aβ42 and comparison with secondary structure analysis from present simulations. The secondary structure logo obtained from the combined trajectory provided a direct visualization of structured segments and hinges of the A*beta*42 monomer.

start and stop residue indices in order to label the first and last residue numbers of these segments. For each sample in the combined trajectory, we detected segments containing α-helix or β-strands and recorded their spans in the aggregated start-stop matrix (Figure 4.5). For example, a β-strand is locating at region E3-R5 on the structure i, then cell $a_{3,5}$ is updated by the addition of 1. The color variation of blue-white-red indicates the magnitude of the count in a cell of the matrix.

Figure 4.5A depicts the resulted start-stop matrix over all trajectory structures, where the original plot is shown on the left of panel B. This particular way of display allowed us to discover four main features of the monomeric Aβ42's secondary structure distribution that could be hard to be obtained from a regular secondary structure analysis. First, the graphic layout presents three triangle shapes A, B, and C in A2-S8, S8-K28, and N27-I41 (that were coloured in red, yellow, and green), suggesting that the monomeric Aβ42 prefers to form structured segments in these regions. In constrast with the matrices generated accounting just for β-strand, and α-helix in Figure 4.5B, the three-triangle shape of the aggregated matrix originated from the formation of β characters in these regions. The helical content appeared mainly in the region H6/Q15-V12/K28 (D), and E22/A30-S26/M35 (E), which overlaps with the region B, and occupies the area between the region B and C. Second, the coloured ranges of A, B, and C in Figure 4.5A indicated their maximum scope of a segment that could be formed around the corresponding area. These three regions were intersected by 1-2 residues, suggesting that the span of a segment is inversely proportional to the range(s) of its adjacent hinge(s). Here, we proposed that a typical pattern of the monomeric Aβ's secondary structure distribution in the aqueous solution is A-hinge-B-hinge-C. Third, amino acids that were frequently involved in a segment were captured by the start-stop matrix. We then defined those shorter regions showing significant local conformations in G9-E22, N27-I31, and G33-I41 as patch B1, C1, and C2. To be specific, B1 occupied two segmental regions in Y10-F19, and Q15-A21 that were also identified using the secondary structure logo. Region C could be divided into C1 and C2 connected by a short loop, forming a motif-like structure. Lastly, some residues on the peptide showed low propensities in defining the termini of a segment, such as S8-G9, V24-N27, G29, I32-G33, and G38. These locations included five out of six glycine residues in Aβ42 sequence, and this phenomenon will be discussed later

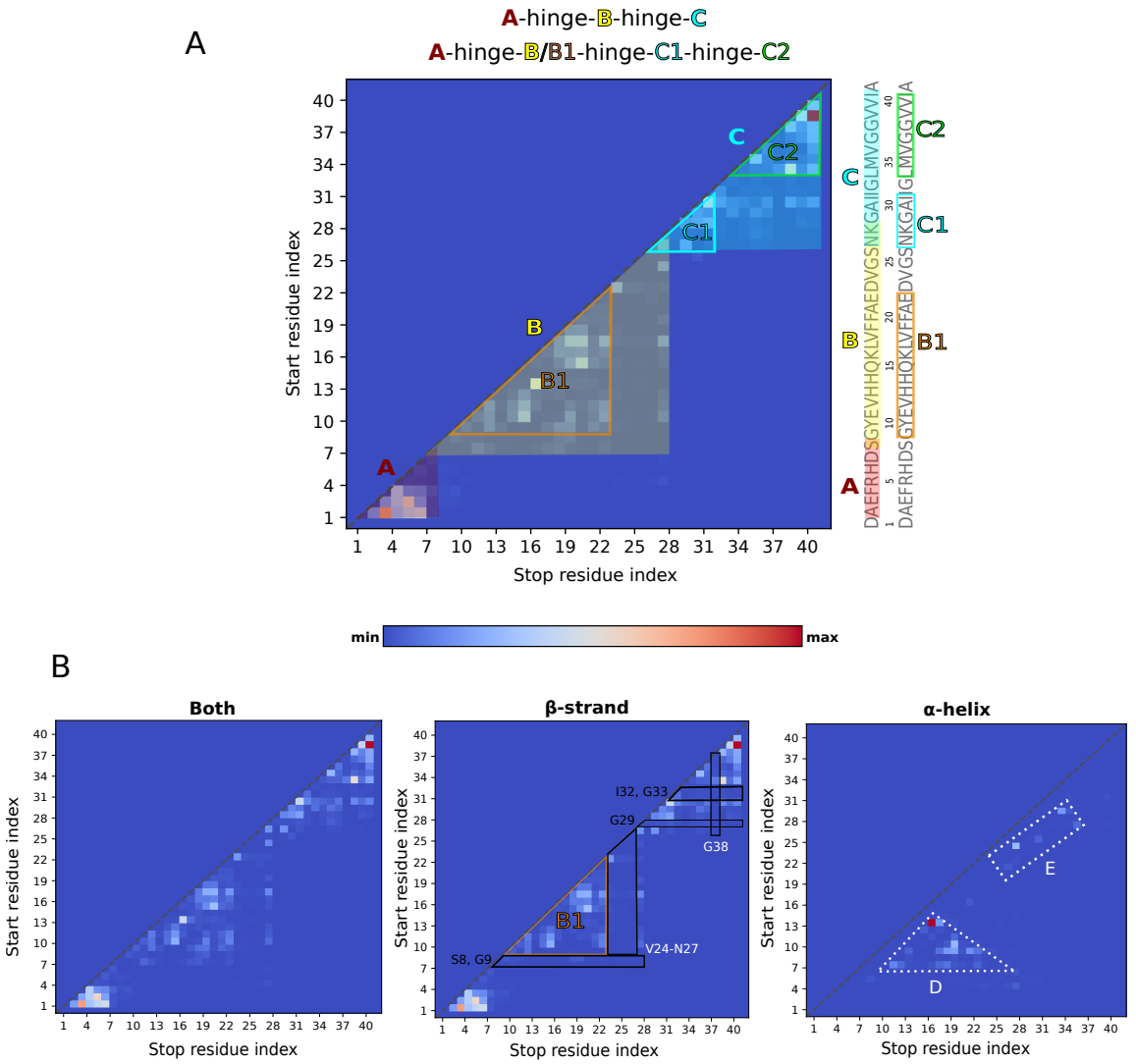

Figure 4.5: A regularly local conformational pattern of the Aβ42 monomer in solution could be shown as A-hinge-B-hinge-C or A-hinge-B/B1-hinge-C1-hinge-C2. (A) The start-stop matrix generated from the combined trajectory showing segments of α-helix and β-strand over all structures, where the rows and columns are the residue indices of the first and last residues of detected segments. B1, C1, and C2 are sub-regions under B and C with the strongest signal of showing α-helix and β-strand. The ranges of A-C and sub-regions were labelled on the amino acid sequences of the Aβ42 on the side. (B) Plot of start-stop matrices generated for both α-helix and β-strand segments (the same matrix as A), β-strand segments, and α-helix segments, respectively. Black trapezoids on the middle panel indicated low preference of forming a segment starting from S8-G9, G29, and I32-G33 (labelled in black) and ending with V24-N27 and G38 (labelled in white). D and E highlighted the two regions with the most frequent occurrence of α-helix on the peptide.

in the chapter. Combining these, the local structure of the monomeric Aβ42 could also be displayed as an A-hinge-B/B1-hinge-C1-hinge-C2 pattern.

4.3.3 Characterization of Structural Emsembles from MD Simulations Using Different Starting Structures of Monomeric Aβ42

MD trajectories were produced via (1) various starting conformations, (2) controlling heating and cooling of the system (SAMD), and (3) replicas of a certain type of a MD simulation. A total of 15 MD trajectories with varied simulation lengths were merged into one dataset. PCA was carried out based on the Cartesian coordinates of the combined trajectory (cPCA) to identify one or several ensembles of similar structures along MD trajectories (Figure 3.6A). Four of the simulations, classical and SA simulations from the helix-np and *U-shaped* conformations, were discussed and analysed in Chapter 3. The involvement of these data aimed to generate more broad conformational space of the monomeric Aβ42 in solution, and for the comparison between previously and currently projected energy landscapes.

PCA results projected onto a 2D space illustrated relations among MD simulations of the monomeric Aβ42 using different simulating strategies. The proportion of variance explained by the first few PC modes was relatively low with the inclusion of all simulations, which accounted for 31.9% of PC 1 and PC2 and 41.1% of the first three PCs (see Figure 3.6)A). On the projection, starting points (or starting conformations) emerged and diffused at various locations. It was suggested that one long MD simulation may only be able to sample at the local neighbourhood around the starting point [319–321]. Our results also showed that the monomeric system in one simulation tends to maintain at a similar phase of sampling after 200 ns even though the Aβ peptide is intrinsically disordered in an aqueous solution. It is then uncertain whether a fundamentally conformational change will occur in a longer simulation time (see Appendix D). To understand the dynamical behaviour of the Aβ42 peptide through its large-scale conformational changes, we concentrated on collecting samples of the MD trajectory displaying the most conformational variability of the peptide. Since one long simulation could get trapped in a local minimum, performing replicas of this simulation allows for continuously sampling of the remaining solution space [320]. Replicas of different initial velocies could produce different trajectories

leading to better coverage of one or multiple free energy basins around the current local state. Figure 3.6B presents the free energy landscape as a function of the first two principal components of Cartesian coordinates. On the left, replicas from helix-np expanded the local neighbourhood search significantly in comparison with the previous landscape projection generated in Chapter 3 (see Section 3.3.4). On the other hand, another transient state exhibiting as a subspace on the right bottom corner was sampled from the replica simulation starting from the *U-shaped* conformation.

SA simulations plus their replicas permit local and global search for energetically favourable transient states. Our results showed that the sampled ensemble obtained from classical MD simulations of helix-np and *U-shaped* conformations were differentiated by positive and negative values along the PC1 axis on the projection. This was also noticed in the previous projection that local states sampled from two long MD simulations of these two initial structures were disconnected (see Section 3.3.4) [2]. Due to the ruggedness of the energy landscape, local minima are usually separated by high energy barriers. Although replicas of these two simulations extended the explored solution space to some extent, a narrow gap still appeared between states. We conducted SA simulations starting from both configurations with an initially high temperature setting in order to overcome the barriers between states for a rapid solution space exploration. Each SA simulation in this experiment was performed with a 400-ns cooling-down phase from 475 K to 300 K followed by another 400-ns sampling at the room temperature (see Section 3.2.3). The cooling-down phase of SA simulations from the helix-np filled in the gap between two states on the 2D cPCA space, thereby a conformational transition pathway was identified from the helix-np towards the *U-shaped* Aβ42 on the energy landscape (Figure 3.6B). Compared with previously calculated pathway (see Section 3.3.4), higher coverage of the surrounding solution space along the pathway was achieved in the usage of both SA and replicas. Also, an off-pathway conformational state was recognized through the after-cooling-down phase sampling of SA simulations from both sides.

RMSD was unable to discriminate between very different conformations of the Aβ42 monomer. Recall that discrete subspaces generated on the cPCA space can

[2] In Chapter 3, structures from simulations starting from Helix-np and *U-shaped* are located on the opposite sides of the PC1 axis.

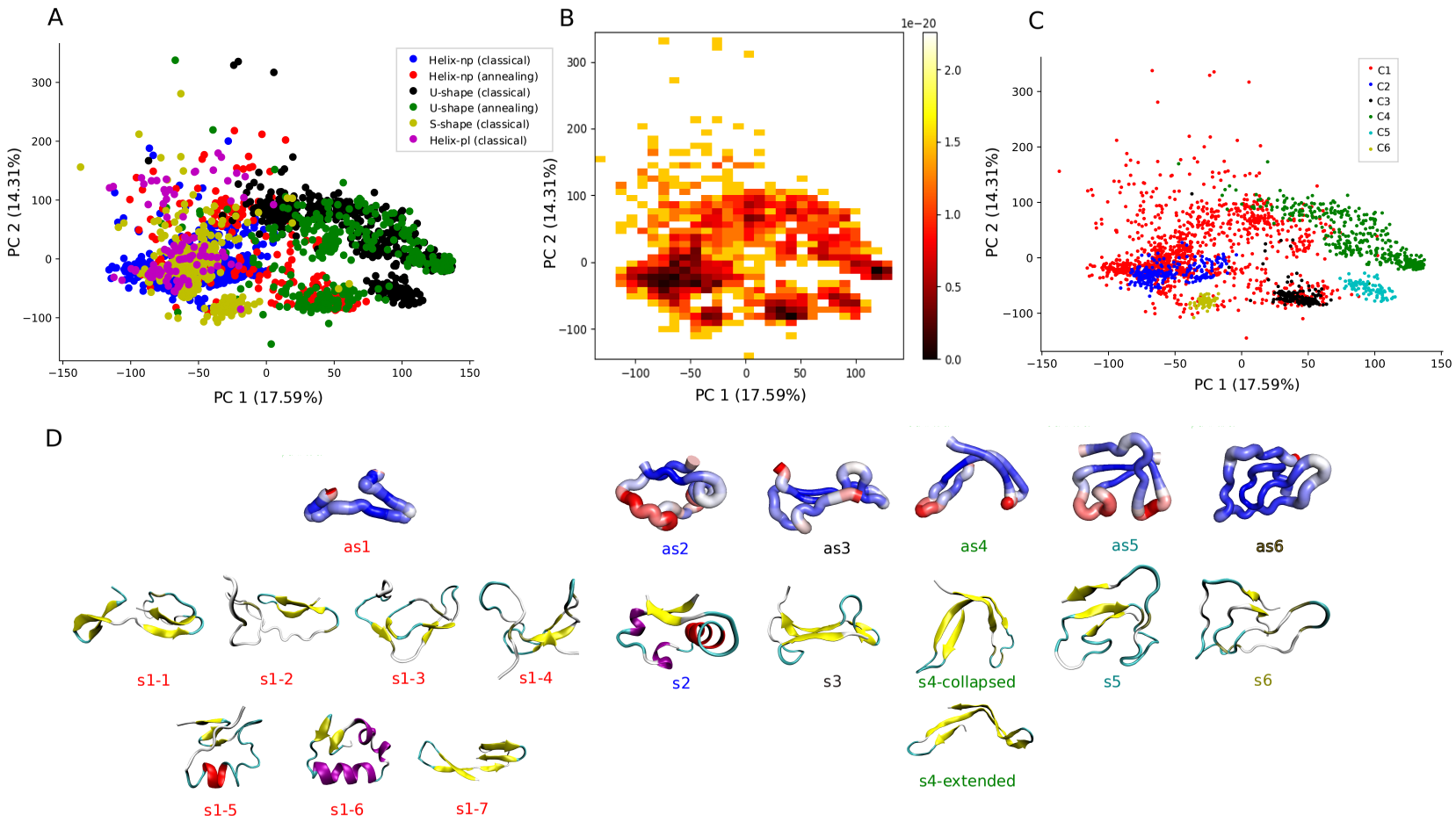

Figure 4.6: (A) Projection of the combined trajectory on the plane formed by the first two principal components (PC1 and PC2). The involved MD simulations starting from different initial structures were labeled in different colors; (B) The free energy landscape was constructed in $-ln(density)$ of k_BT units based on the current cPCA projection, where the 2D space was pixelated by the value of $n = 10$; (C) Six identified clusters of similar configurations were indicated in colors; (D) Average structures (as1, as2, and as3) were generated using PyMol in *putty* mode display. The thickness of the chain in the representation and also the color range of red-white-purple suggests the extent of RMSD changes among all structures in a community. Snapshots of closest structures to the averages were shown below. All structures were placed with the N-terminus on the left and the C-terminus on the right.

be seen as conformational states identified from the MD sampling. Thus, an agreement between this partition and determined community structure is expected for the characterization of obtained structural ensembles. The community detection method *fast-greedy* [194] was firstly applied on the combined trajectory using R-score as the similarity metric (see Appendix C). The computed community structure presented 3 diffused network clusters that did not match the separation of PCA subspaces. Such loosely defined communities may be resulted from the definition of the similarity metric since one of the major drawbacks of RMSD is that it is more sensitive to the local structure than the global topology when the value is large. For example, the RMSD of two protein structures may return a high value when both of them share a similar global topology as one of the loop region has different length or their C-terminal regions show alternative orientations. This situation cannot be distinguished from the case when the two structures have significantly different architectures. We obtained diverse conformers in a wide range of geometric variance by simulating the Aβ42 monomer from various initial structures. Because of this, a distance metric with higher sensitivity and specificity is required that different levels of pairwise structural dissimilarity (eg. local structure vs global topology) can be output as distant numbers. H-score was defined based on the concept of residue-residue contact and utilized as a new similarity criterion for the network partitioning (see Section 4.2.4 for detailed definition). Instead of representing a protein structure via a set of residue centroids coordinates, a contact map discretizes the protein configuration at the level of its potential intramolecular interactions among residues, leading to a reduced effect from local flexibility in proteins due to the definition of contacts. Then, the force behind the formation of network clusters becomes the grouping of structures with a certain type of contact pattern.

A general agreement has been reached between the clustering result based on the H-score and the 2D cPCA space shape. Communities C2 to C6 capture distinct folds from more than half of the combined trajectory data including ensembles obtained from simulations starting with almost all types of initial structures except for Helix-pl (Figure 4.6C). Table 4.6 lists the proportions of structures belonging to simulations from different species of starting structures for each inferred community. An average structure was calculated to indicate the unique structural features shared within each

cluster (Figure 4.6D). Configurations have the smallest RMSD against their averages were selected to represent the conformational characteristics for the further analysis.

Table 4.6: Proportions of containing structures from simulations of different initial structures for each cluster.

Proportions (%)	C1	C2	C3	C4	C5	C6
Helix-np	21.8	100	-	-	-	-
Helix-np (ann)[a]	12.4	-	100	0.5	-	-
U-shape	9.7	-	-	74.6	100	-
U-shape (ann)	17.8	-	-	24.9	-	-
Helix-pl	8.9	-	-	-	-	-
S-shape	29.5	-	-	-	-	100

[a] SAMD simulation

Transient or less-folded structures were grouped into a big cluster. Community C1 contained a collection of configurations with much larger H-score distribution than other communities. Its highly geometric variance was reflected by the magnitude of C1 over the whole projection. Figure 4.6C and Figure E.2 in Appendix E show that this community comprised of one or more low-dimensional subspaces besides C2 and a sparsely populated region connecting to all the other inferred communities. More and better-defined communities might be needed to locate the positions of possible basins within this region so that the covered terrain can be described more properly. Table 4.6 reveals that community C1 included configurations from simulations of all types of initial structures, thereby it is highly probable that C1 was a mixture of transition and meta-stable states. This phenomenon was also reported from a previous clustering study based on MD simulations of disordered proteins [230]. Nevertheless, such loosely grouping still captured some common conformational characteristics owned by the containing structures. We compared its average structure with seven selected structures (s1-1 to s1-7) that are closest to the average geometric coordinates from different simulations. All the topologies contained the intramolecular interactions of the middle of the chain with both termini, yet there were a lack of contacts between the termini. In contrast, C2-C6 featured structures with a higher contact volume between the termini, allowing the peptide to collapse into a tightened core. To verify this, we calculated the average radius of gyration (R_g) for each cluster with standard

deviation. A reference R_g can be predicted for Aβ42 if it folds into a fully collapsed, globular state using the empirical equation proposed by Kolinski and Skolnick [322], which is:

$$\langle R_g \rangle = 0.22n^{0.38}, \tag{4.7}$$

where n is the number of residues. Then a R_g of 0.91 nm is expected for a well-defined cluster of collapsed state. Table 4.7 demonstrates that the average R_g of C2, C3, C5, and C6 range between 0.9-1.0 nm, which is in agreement with the R_g of compact ensemble of Aβ42 reported from previous simulations [117, 261, 312, 323, 324]. As hypothesized, C1 has the largest distribution of structure sizes that many transient states of extended conformations from all simulations were grouped into this very community. C4 was another mixing state including the relevant extended and collapsed conformations on the both sides of an interconverting equilibrium (Figure 4.6D).

Table 4.7: Radius of gyration (R_g) of each identified community is reported in its mean values with standard deviations, maximum and minimum values.

Cluster	C1	C2	C3	C4	C5	C6
Mean (nm)	1.16	0.97	1.09	1.21	0.97	1.10
SD[a](nm)	0.19	0.05	0.05	0.17	0.04	0.01
Maximum (nm)	2.45	1.13	1.45	1.78	1.06	1.14
Minimum (nm)	0.88	0.89	1.02	0.99	0.89	1.07

[a] standard deviation

4.3.4 Tertiary Structure

Contact Patterns

A frequency contact matrix was computed based on the average-ensemble contact frequencies over all trajectory data, illustrating formation of local contacts all over the chain. The matrix presents a characteristic contact pattern of tertiary structure shared by the sampled Aβ42 populations, suggesting that their structures adopted broadly three regions which are D1-S8, G9-S26, and N27-A42 (Figure 4.7a), in agreement with the observation from Rosenman *et al.* [116]. Widths of these regions were

amply equivalent to those of segments detected from the secondary structure analysis which are A (A2-S8), B (S8-K28), and C (N27-I41). This confirms our hypothesis that the structured segments of the Aβ42 peptide could be treated as the basic building blocks of its various conformations.

The frequency contact matrix is further segmented by regions A, B, and C in Figure 4.7b that are highlighted in primary colors red, yellow, and cyan. Contact probability signals shown within forming contact blocks in secondary colors orange, silver, and green indicate the presence of intra-region interactions between regions that are described as contact pairs AB, AC, and BC, respectively. For simplification, both AC1 and AC2 are counted as AC, and same as BC that comprises of BC1 and BC2. A forth pair CC (or C1C2) is considered for the case of the region C in contact with itself. Figure 4.7b depicts different contact pairs in monomeric structures with matching colors that were used and described in the frequency contact matrix. The matrix showed that frequencies of significant contacts are mostly below 60%, which raises the question that if these four contact pairs always appear simultaneously in every sampled structure. To explore this, contact patterns with different combinations of contact pairs were calculated.

We classified all configurations from simulations into four types of architectures based on the four contact pairs. Considering AB, AC, BC, and C1C2, possible contact patterns are $2^4 - 1$ in total, where the combinations could be the presence of one pair, two pairs, three pairs, and four pairs of contacts. Four contact patterns AB-BC, AB-BC-AC, AB-BC-CC, and AB-BC-AC-CC were found from the present simulation data that account for 4.7%, 28.9%, 13.2%, and 53.2%, respectively (Figure 4.8b). Recall that two dominant conformational states of the Aβ42 peptide, the collapsed and extended states, were observed from our conformational samplings (see Section 3.3.3) and previous studies [145, 258]. To explore the correlation between a certain contact pattern and the formation of collapsed or extended conformations, we calculated the mean solvent accessible surface areas (SASA) of most frequently observed hydrophobic residues in the hydrophobic clusters of collapsed conformations (*i.e.* F4, Y10, F19, F20, L34, M35, and V36) for each contact pattern, where collapsed and extended conformations are defined as SASA below 7.0 nm^2, and above 8.0 nm^2, respectively (Figure 4.8c). AB-BC-AC and AB-BC-AC-CC patterns contained most

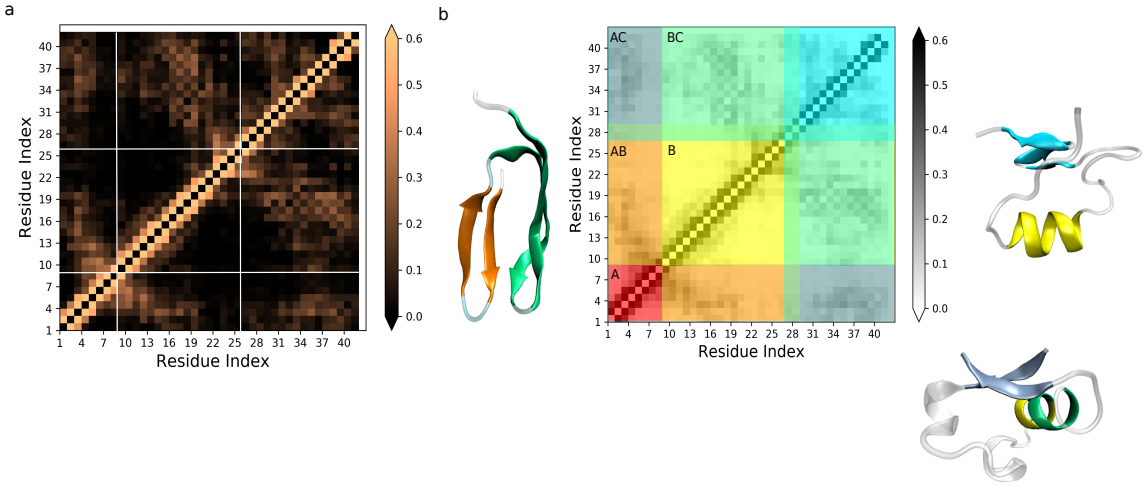

Figure 4.7: (a) Frequency contact map illustrates an all-atom average-ensemble contact probability over all trajectory data. Invalid squares above, on and below the diagonal that represent residue i in contact with itself and adjacent residues $(i+1$ and $(i+2)$ are not considered in the contact analysis. The multi-domain pattern was demarcated with white lines, indicating intramolecular interactions between regions D1-S8, G9-S26, and N27-A42. (b) The same frequency contact matrix was segregated by regions A, B, and C determined from the secondary structure analysis. Contacts shown within the resulted blocks of orange, silver, green, and blue (black dash square) indicate existence of contact pairs AB, AC, BC, and C1C2 respectively. For each contact pair, its represented intra-region interaction are displayed in structures with matching colors.

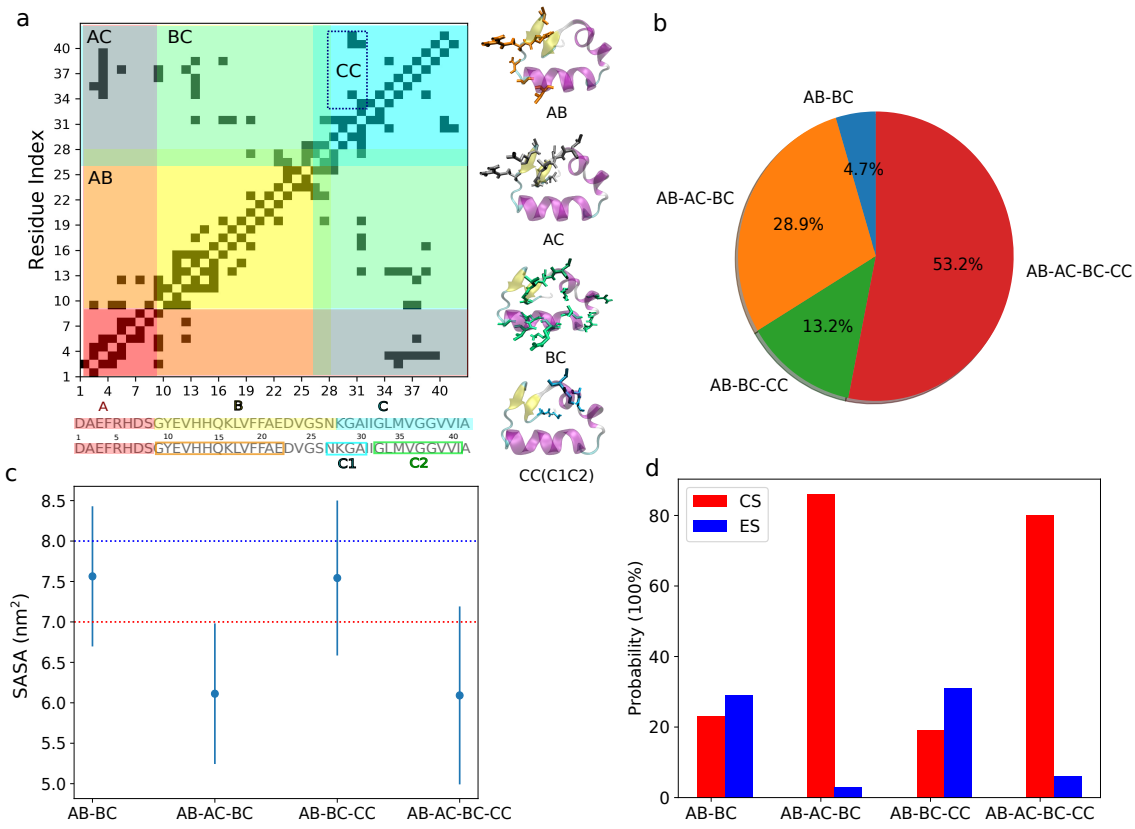

Figure 4.8: Obtained structural ensembles were classified based on contact patterns. (a) An example contact map of an Aβ42 monomeric structure (S_a) with regions A, B, and C highlighted in primary colors red, yellow, and cyan, forming contact blocks in secondary colors. Existence of contacts within each block indicated the presence of contact pairs AB, AC, BC, and CC(C1C2) on the S_a; (b) Populations of structures in different contact patterns; (c) Solvent accessible surface area of the hydrophobic residues that were frequently observed in the hydrophobic clustering of collapsed conformations; (d) Probabilities of the Aβ42 peptide staying in collapsed and extended conformations in each contact pattern.

of the collapsed conformations, while structures of AB-BC and AB-BC-CC showed as partially folded or extended topologies. Figure 4.8d also provides the probabilities of collapsed and extended conformations in each contact pattern.

Each contact pair functions as a particular structural component in the folding of the Aβ42 monomer. Representative structures for identified communities were examined and classified into matching categories of contact patterns under collapsed and extended/transient conformations (Figure 4.9). Figure 4.8a gives an example contact map of one representative structure with segmented contact blocks highlighted. The corresponding intramolecular interactions were shown in the structure with bond representation of matching colors. Besides each structure in Figure 4.9, the population of its represented community or conformation over all sampled structures was provided. By analysing these topologies using the contact pattern, we concluded the specific function of each contact pair as a building block in the formation of the monomeric architecture from the low-level structural conservation to high-level conformational plasticity of the Aβ42 peptide. A high preference of staying in collapsed conformations was observed from present simulations. One of the crucial stabilizing factor in the formation of a compact structure is the hydrophobic clustering. Contact pair BC existed in all structures, indicating the potential of forming a hydrophobic core in all states of the Aβ42 monomer since the half of the peptide on the C-terminal side is mostly constructed by hydrophobic residues. Contact pair AB was also discovered in all contact patterns, however, fairly low contact content were shown in some topological types (*e.g*, s1-3, s1-6, s3, and s6 in Figure 4.9). This phenomenon suggests that the N-terminus possesses high flexibility while in contact with the central region, which also explains why such region was observed disordered in previously solved fibril structures [93, 108, 110]. On the other hand, AC is related to the formation of globular architecture of the Aβ42 peptide because contact patterns that were lack of AC contacts (*e.g*, AB-BC and AB-BC-CC) accounted for mostly partially unfolded structures. The interconversion between collapsed and extended conformations accounts for the formation and loss of AC contacts (*e.g*, s4-collapsed and s4-extended in Figure 4.9). This discovery supports the importance of two extra C-terminal residues of the Aβ42 in the contribution of its structural stabilization. We also observed that the β-sheet formed between termini stabilized the collapsed

states by diminishing the overall conformational entropy (*e.g*, s1-4, s1-6, s2, s3, s5, and s6 in Figure 4.9). Contact pair CC usually occurs in the transient structures between the collapsed and extended states, suggesting its main function in temporarily stabilizing the C-terminus via the hydrophobic clustering (*e.g*, s1-1, s1-2, s1-5, s1-6, and s1-7 in Figure 4.9). Two advantages of this type of hydrophobic clustering are: (1) diminishing the overall conformational entropy via contacting with the region B forming the hairpin motif; and (2) providing structural flexibility on the first half of the peptide to some extent by reducing the contacts with the region A or B. Also, among the representative structures, CC contacts formed within the s1-2, s1-5, and s1-7 showed significantly high β-content that may enhance the aggregation propensity via the C-terminal hydrophobic attraction between monomers.

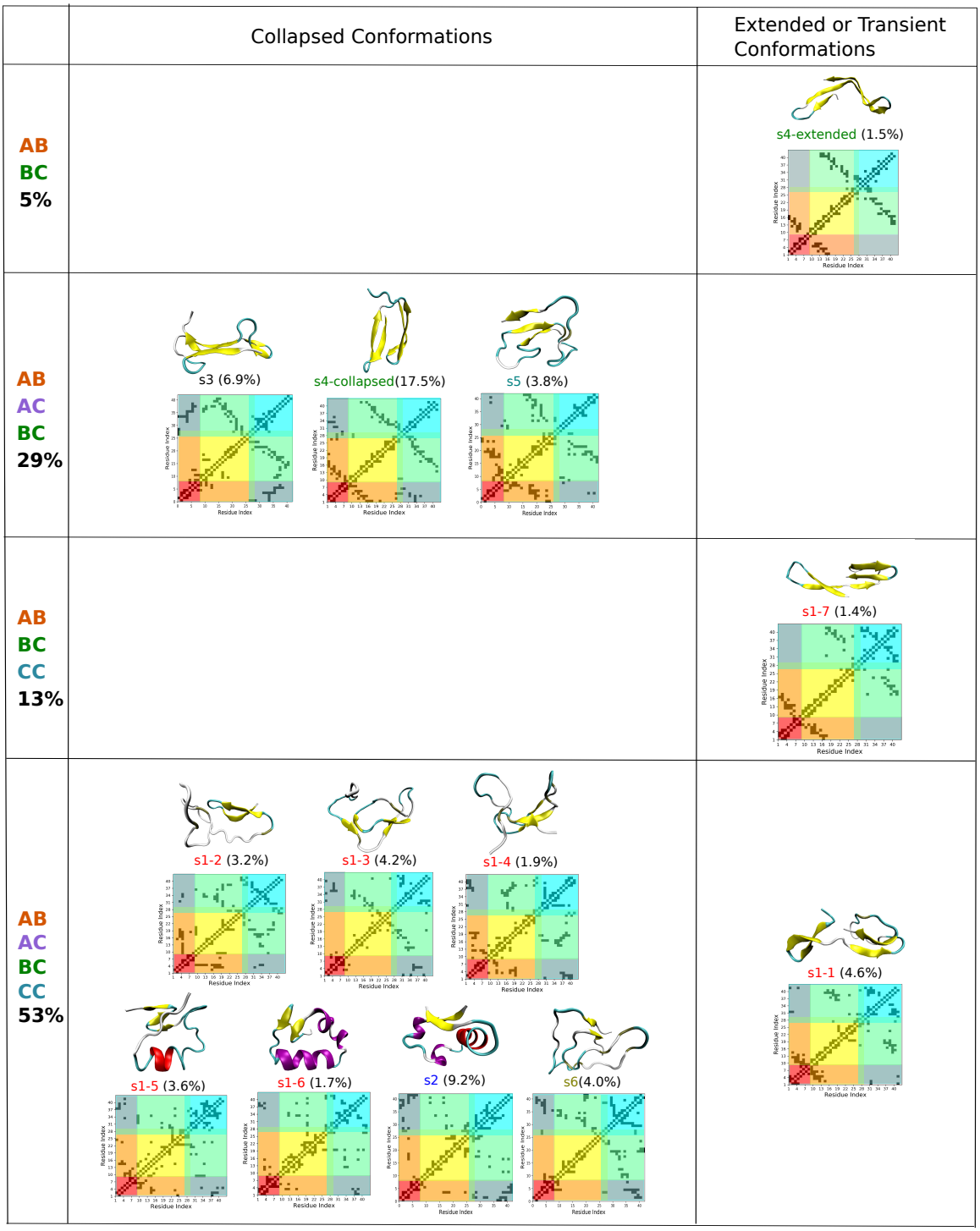

Figure 4.9: Continued on the following page.

Figure 4.9: Classification of representative structures of communities according to their contact patterns and degrees of folding. The latter is defined by SASA of the most frequently observed hydrophobic residues in hydrophobic clusters of collapsed conformations. Percentages of s2 to s6 were computed based on their represented community sizes over all production ensembles. The percentages of structures in community c1 were calculated via a similarity metric q. It is defined that a structure s was considered to be similar enough to a representative structure (s1-1 to s1-7) if the difference of their contact maps was less than 8.5%. This difference cutoff is an empirical value evaluated by the Hamming distance. Maps below the structures illustrated the contacts for residue centroids, wherein regions A, B, and C were highlighted. All structures were presented with the N-terminus on the left and the C-terminus on the right.

Electrostatic Interactions

The featured contact pattern shared by the equilibrated Aβ42 populations (Figure 4.10a left panel) demonstrates high frequencies of intramolecular interacting between three portions which agree with the lengths of regions determined from the secondary structure analysis. This suggests that the formation of significant contacts in Aβ42 monomeric structure may primarily arise from atomic hydrogen bonding although the defined contacts discriminate the chemical natures of atomic interactions (*i.e.*, electrostatics). Heterogeneity in β-sheet arrangements and loops were observed to be the major factors for the polymorphism of the Aβ42 monomeric structure in this experiment. Since most of HBs between backbone atoms are within a helix, sheet or turn, a local hydrogen bond matrix of Aβ42 ensembles could be used to represent its topologies in a graphically characteristic pattern. The right panel of Figure 4.10a elucidates the occupancies of backbone HBs over all production ensembles, which also exhibits a similar pattern compared to the frequency contact matrix as expected. Such a pattern shows an approximate symmetry along a shifted anti-diagonal, indicating that the presence of helices and short loops all over the chain. The matrix was further demarcated by white lines according to spans of region A, B, and C, then the segmented hydrogen bonding blocks were described in the concept of HB contact pairs (which are the same as contact pairs).

The comparisons between these frequency contact matrices and hydrogen bonding matrices in Figure 4.10b reveal that Aβ42 ensemble in a certain contact pattern also shared a matching characteristic pattern of hydrogen bonding. The former matrices

were used to discretize the spatial interactions in a level in order to uncover the building blocks of Aβ42 folds from the simulations, while the later ones provided specific information on the presence of significant motifs such as the hairpin formed between region B and C. Differences in the backbone hydrogen bonding patterns represent significant structural features among contact pattern populations. HBs of A21-L34 and F19-V36 account for high occupancies, which 30.7% in AB-BC and 51.6% in AB-BC-CC, respectively. Both HBs were involved in the formation of the β-sheet between regions B and C, wherein their donors and acceptors were exchangeable. In contrast, structures of AB-BC-AC and AB-BC-AC-CC patterns showed higher conformational variety due to lower contribution of HBs in the formation of β-sheets but loops in the presence of the contact pair AC.

Besides β-sheets, another stabilizing factor in the formation of Aβ fibrils is the salt bridge, wherein intra-chain SBs stabilize the turn region connecting two β-strands, while inter-chain SBs function as linkers in the complementary packing between layers or protofilaments [58, 93, 102, 106, 108–110, 325, 326]. NMR studies and a combined NMR/cryo-EM approach suggested different types of SBs in various polymorphs of Aβ fibrils, among these, SBs to K28 appear to the most essential ones. For example, the *U-shaped* model exhibits a β-strand-turn-β-strand motif, reinforced by a D23-K28 SB between the two strands [102]. A K28-A42 SB formed mainly via an intramolecular contact stabilizes the triple-β-sheet motif in the *U-shaped* Aβ42 fibril [108–110]. Inter-chain SB of D1-K28 was observed to establish the interactions between the protofilaments in the *LS-shaped* fibril structure [58]. From the simulations starting with monomeric units of these fibrils, we calculated the frequencies of intra-chain SBs especially for the ones discovered in the original polymorphs and compared the favoured SBs in different contact pattern populations.

R5 plays an important role in orienting N-terminal residues and stabilizing the kink between the region A and B. Figure 4.12A depicts the populations of all 18 SBs formed by three positively and six negatively charged residues for the structural ensemble of different contact patterns with adopting Barlow and Thornton‘s definition described in Section 4.2.4. To investigate if SBs have any effect on the establishment of contact network in Aβ42 configurations, we highlighted the SB matrix indicating

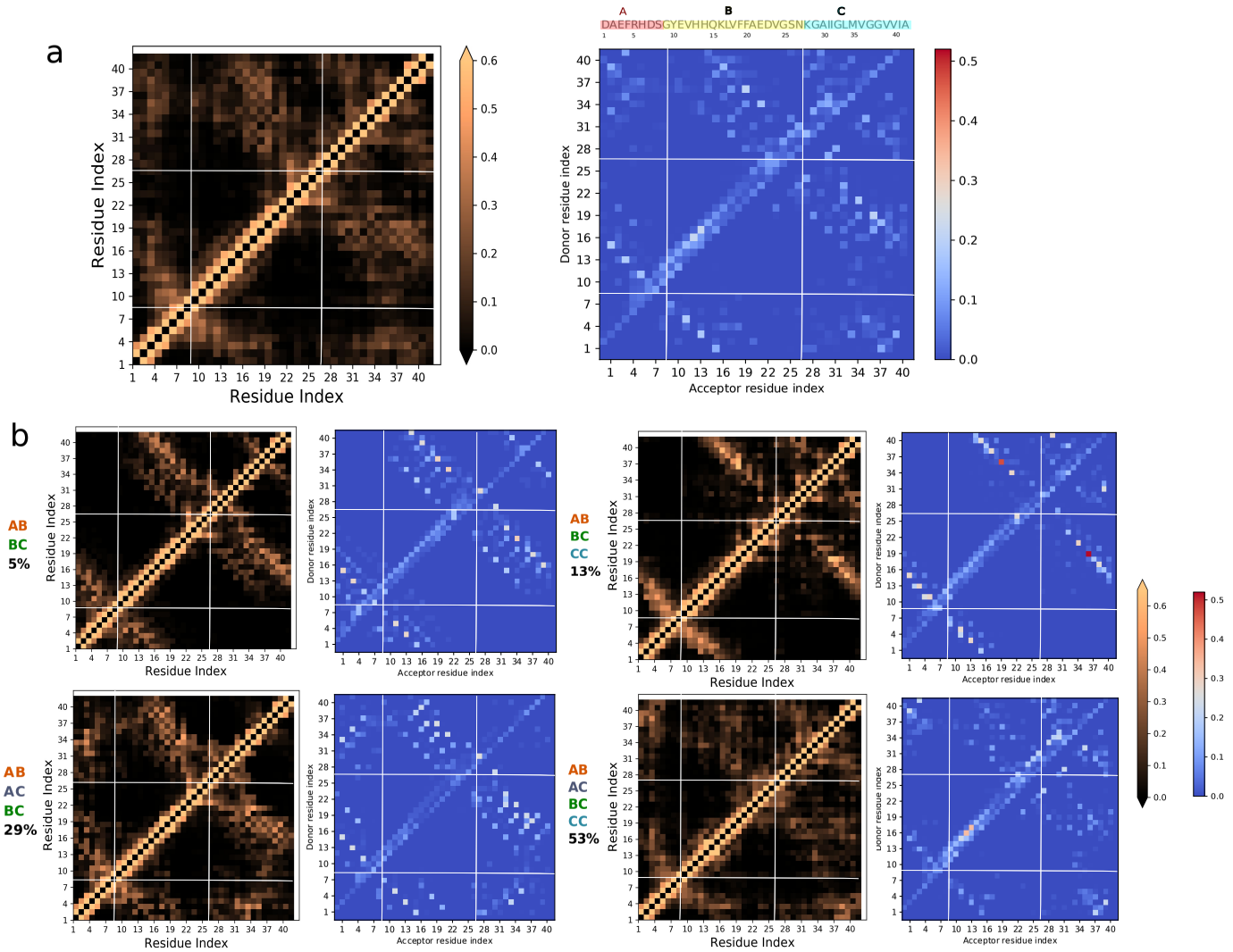

Figure 4.10: Plots of matrices illustrating the probability of backbone HBs over (a) all production ensembles of the Aβ42 and (b) structures in contact patterns of AB-BC, AB-BC-CC, AB-BC-AC, and AB-BC-AC-CC respectively. Each matrix was segmented into signal blocks by white lines. There was a shifted diagonal in each matrix due to the presence of helical structures in each contact pattern and low probability of HBs formed between adjacent or close residues across the chain.

the corresponding contact pairs demonstrated in Section 4.3.4. Two SBs to R5, E3-R5 and R5-E11, showed occupancies up to 15.6% and 9.0%, respectively. Direct observation of E3-R5 SB in community c3 ensemble (Figure 4.11B) suggested its function in adjusting the E3-H6 β-strand to a range of angles in order to form a sheet with the central region. This adjustment arises from the flexibility of the side-chain-side-chain contact, leading to a variety of SB distances. Such a SB was observed in a cryo-EM structure of Aβ40 fibrils *in vivo* that was formed between E3 and R5 from adjacent protofilaments, which requires the β-sheets of both protofilaments to orient their interface to expose the charged residues [112]. The presence of the SB R5-E11 was displayed on a collapsed conformation from community c4 (Figure 4.11B), driving the formation of the a loop region between D7 to E11. A similar SB formed between E11 and H6 was found to stabilize the kink around Y10 in the *LS-shaped* Aβ42 fibril structure [58].

The most reported D23-K28 SB contributes in the formation of the cross-β structures of fibrils [93, 102], however, a low propensity (*i.e.*, 5.1%) was observed from the MD trajectories. Similar observations were demonstrated using the same definition of SB in the previous simulations performed with different force fields [116]. An early computational study suggested that the significant motif of *U-shaped* amyloid fibril structure including the pre-formed D23-K28 contact and an intact VGSN turn is uncommon in monomeric conformations [327]. The formation of this SB in a stable structure requires sufficient peptide concentration to overcome the large barrier for the desolvation of D23 and K28. Figure 4.11B reveals that the D23-K28 SB stabilizes the hydrophilic turn region between the E22 and G29 [226], which further implicates that the formation of the C-terminal hairpin depends on the stability of the D23-K28 SB. Another SB formed between E22 and K28 (*i.e.*, 2.9%) acted the same in the collapsed conformations of the Aβ42 such as the s1-3 and s1-4. This is reasonable because of the preference of the region E22-A30 exhibiting turn/bend/coil structure shown on the secondary structure logo. Thus, residues E22/D23 and K28 on the edge of the hinge formed a zipper-like motif, prompting the hydrophobic packing or intra-sheet formation between the region B and C.

We further calculated populations of SBs among structural ensembles in different contact patterns. SBs to R5 showed highest frequencies in all patterns (Figure 4.12),

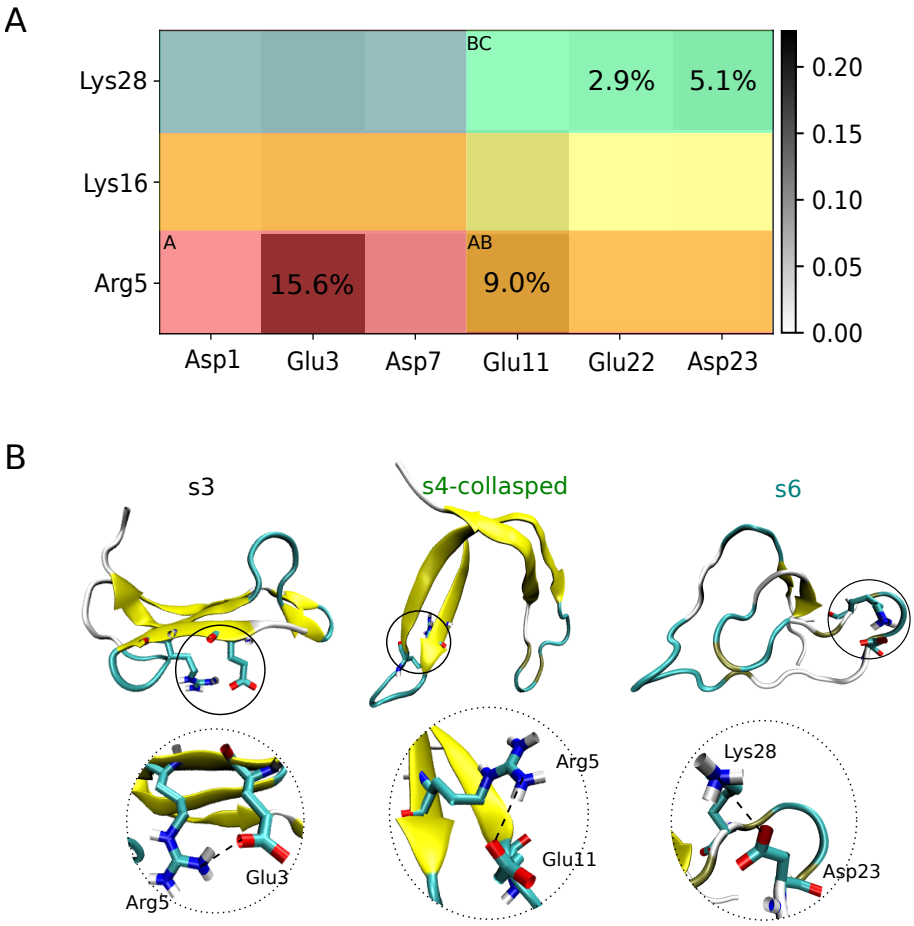

Figure 4.11: Populations of all 18 SBs formed by 3 positively charged and 6 negatively charged residues for (A) all production ensembles using the Barlow and Thornton's definition described in Section 4.2.4. Major SBs with relatively high populations were reported in proportions. Original plot was generated white-gray-black. Regions A, B, and C were colored onto the matrix with red, yellow, and cyan; (B) Representative structures for the three most populated SBs were shown in the order of E3-R5, R5-E11, and D23-K28 from left to right.

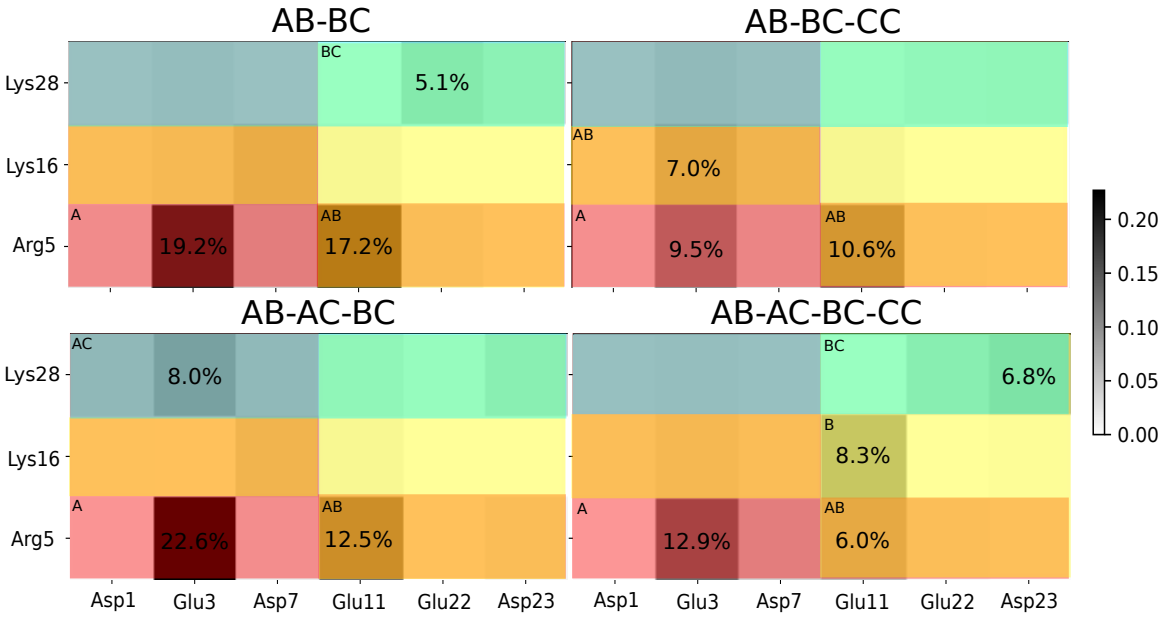

Figure 4.12: Populations of all 18 SBs formed by 3 positively charged and 6 negatively charged residues over sampled structures in different contact patterns (A-D) using the Barlow and Thornton's definition described in Section 4.2.4. Major SBs with relatively high populations were reported in proportions. The original plot was generated white-gray-black. Regions A, B, and C were colored onto the matrix with red, yellow, and cyan.

indicating the primary effect of SBs on ordering and stabilizing the N-terminus based on the present simulations. Figure 4.12 also illustrates the emergence of distinct SBs in Aβ42 topologies if different contact patterns. An E22-K28 SB showed an occupancy of 5.1% in AB-BC, implying its minor contribution in stabilizing turn region between the region B and C. In contrast, no SB related to the BC contact pair appeared in AB-BC-CC but another AB type SB E3-K16 (*i.e.*, 7.0%). The last charged amino acid K28 is the first residue of the region C and the rest of the region (G29-A42) is basically composed of hydrophobic residues. The lack of the SB in the BC type may be explained by that the contact network formed within the region C diminishes the overall conformational entropy via the hydrophobic clustering or intra-regional sheet formation on the C-terminus. Contact patterns AB-BC-AC and AB-BC-AC-CC included almost all collapsed conformations from the sampled ensembles with complex contact networks. In AB-BC-AC, an E3-K28 SB had a propensity of 7.0% that enhances the intra-regional contact binding between the region A and C. Such a type of SB was not found in the AB-BC-AC-CC ensemble, however, an occupancy of 6.8% D23-K28 SB was observed in supporting the stabilization of the C-terminus beyond CC contacts.

4.3.5 Characterization of Structural Dynamics of Monomeric Aβ42

DPC method (see Section 3.2.4 for details) was applied to interpret the Aβ42 dynamics especially for its conformational interconvertion between collapsed and extended structures in solution. PCA was used to ordinate all structures from multiple simulations along the axes, which characterizes the structural variation of Aβ42's major motions. We visualized the nature of the first two PCs by binning structures according to these axes and computing average structures that capture the continuum between extended and collapsed conformations.

By discretizing along PC1, the topology of the Aβ42 monomer evolved from (a) a1-a5: several collapsed conformations with CC contacts and close termini (*e.g* s1-2, s1-5, and s6), (b) a6-a7: an extended structure (*e.g* s1-7), (c) a8: a collapsed conformation similar to s3, (d) a9-a10: the extended state of s4 to (e) the collapsed state of s4 cotaining a double-hairpin motif. By using the concept of contact patterns, this structural evolution can also be described as following: (a) AB-BC-AC-CC, (b)

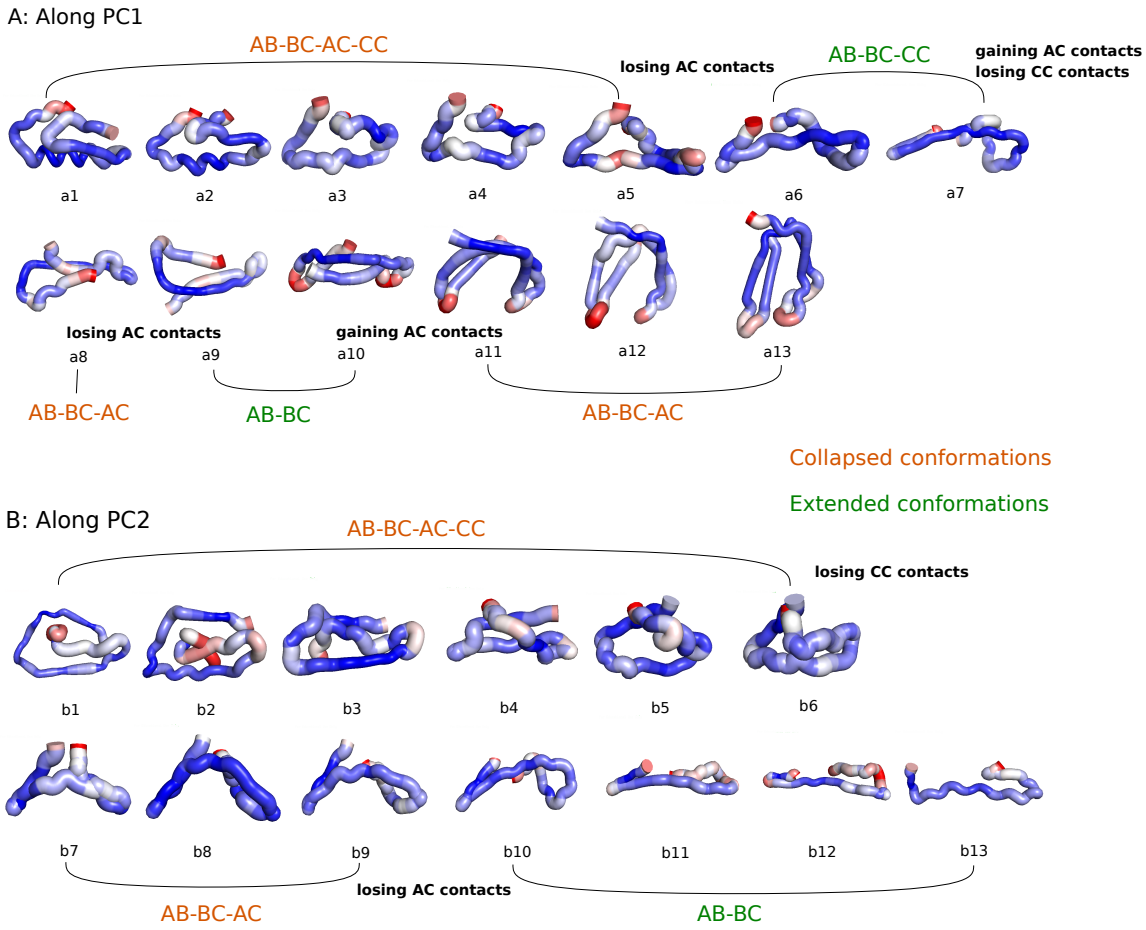

Figure 4.13: Two series of average structures were obtained via applying the DPC along (A) PC1 and (B) PC2 based on the combined trajectory in the binning size of 25, respectively. Binning pieces over 200 (scores) along PC2 were discarded. Contact patterns in collapsed and extended shapes were colored in orange and green, respectively. The thickness and color range of red-white-purple of average structures suggested the extent of RMSD changes among structures in the corresponding binning area. All average structures were presented with the N-terminus on the left and the C-terminus on the right.

AB-BC-CC, (c) AB-BC-AC, (d) AB-BC, and (e) AB-BC-AC, in which (a), (c), and (e) are collapsed conformations while (b) and (d) are extended ones. Conversion between a collapsed and extended conformation was thought to be related to the presence and absence of AC contacts. Based on our sampling, two major ways of establishing AC contacts were discovered, which were via (1) parallel or antiparallel β-sheet formed between termini, *i.e.*, (a), (b) to (c), and (2) hydrophobic clustering between AB and BC β-sheets implying an *open-closed* movement, *i.e.*, (d) to (e). Thus, it is logical to assume that these collapsed states could be connected with each other on the energy landscape via transient extended states.

On the other hand, the result obtained from discretizing along PC2 displayed a different path for this collapsed-extended conversion. The architecture of the Aβ monomer shifted from (e) b1-b6: the AB-BC-AC-CC pattern (*e.g* s6 and s2), (f) b7-b9: the AB-BC-AC pattern (*e.g* between s4-collapsed and s4-extended) to (g) b10-b13: the AB-BC pattern (*e.g* s4-extended). Specially, we observed a conformational transformation from one collapsed state (e) directly to another (f) through the loss of CC contacts while forming AC contacts between the termini. Region C or the C-terminus can diminish the overall conformational entropy via contact with the region B forming the hairpin motif or maintain the structural flexibility to some extent by clustering with itself to reduce contacting with the central region. From one aspect, this supports our speculation raised based on the clustering result that structures of the AB-BC-AC-CC contact pattern showed less compactness in topologies. These structures could be transient states bridging to other metastable states on the energy landscape.

4.4 Discussion and Conclusions

Self-association of the Aβ peptide naturally yielding high order assemblies is widely accepted as the signature of AD. Polymorphism of their structures lies in the formation of varied structural motifs such as the hairpin motif within the monomers that affects the packing manners of the further formed multimers. Properly characterizing the Aβ monomeric structures that comprise such motifs becomes essential in understanding the peptide aggregation process. A rough comparison of monomeric Aβ42 secondary structures characterized through NMR (PDB ID: 1IYT) and computer

simulations (listed in the introduction) indicated the possible existence of structural segments of α-helix or β-strand on the similar locations of the peptide in different environments or simulated using different force fields. To investigate this, we performed classical and SAMD simulations by initiating from different monomeric conformers of the full length Aβ42 with replicas while comparing calculated J-couplings and chemical shifts with their experimental counterparts.

Secondary structural character determined from analysis of these trajectories (*i.e.*, secondary structure logo) were consistent with previous CD estimations and recent computational studies with the highest propensity of forming β-strands locating at homogeneous regions. Further characterization of the Aβ42 ensemble with consistent pieces of β-structures using start-stop matrix revealed the frequent emergence of these segments even though these conformers were kinetically unstable. Our simulations suggested that the Aβ42 peptide adopts structures of three β-dominated regions at E3-R5, Q15-A21 and I31-V40 with hinges between. Experimental data suggested the importance of the β-strand formation at Q15-A21 within the Aβ42 fibril nucleus [**?**, 56, 58]. Comparison of the maximum ranges of these regions which are A2-S8, S8-K28 and N27-I41 revealed that the lengths of the hinges could range between zero to ten or more residues. Such mobility was caused by a structural change from a defined secondary structure to a loop that simultaneously affected the lengths of adjacent segments and further the peptide topology. Herein, glycines are commonly known as secondary structure breakers due to their flexibility, locating at the termini of secondary structures or within a loop [328]. In this experiment, glycines were not only defining the end of these segments but also were key residues for the formation of Aβ fibrils. To be specific, G9, G25 and G29 served as secondary structure breakers, separating the secondary structure core within regions A, B, and C from loops between. The hydrophobic C-terminal region of the Aβ42 peptide contains a G25-XXX-G29-XXX-G33-XXX-G37 motif that is commonly found in the transmembrane α-helices, termed a *glycine zipper* [329]. This motif can form a parallel, in-register β-sheet by placing two glycines to the same face, creating the surface notches or grooves for sheet-to-sheet packing [330]. Mutational analyses addressed the importance of G33 and G37 that link to the aggregation dynamics of the peptide [56, 331, 332]. For exmaple, one of the studies increased the hydrophobicity at

positions of G29 and G33 by substituting the glycine to alanine and isoleucine. Aβ42 G33A and G33I peptides showed a preference of forming higher order of oligomers, whereas G29 substututions have no effect [332]. G25 and G29 of the motif seem to only function as flexible hinges. The start-stop matrix displayed the possibility of the region C to form a β-strand or a small motif comprising C1 and C2 with a kink at I32-G33. This indicates that G33 could both induce the aggregation propensity and provide the backbone flexibility based on different Aβ42's monomeric conformations.

Analysis of the frequency contact matrix suggested a constant set of intramolecular interactions between three regions based on the equilibrated Aβ42 population, matching identified structural segments in these ranges. Such an agreement supports our hypothesis that the heterogeneous architectures of the Aβ42 monomer can be described by the adopted secondary structure arrangements *i.e.*, region A, B, and C. Clustering on the obtained ensemble captured several distinct conformations that agreed with previously identified signature motif of amyloid oligomers, *i.e.*, the β-pleated structure [145, 266], the turn-strand-turn-strand-turn-strand motif [283] and the hairpin motif between the central region and the C-terminus of the peptide [89]. We further characterized these structures using regions A, B, and C as building blocks, giving four contact patterns AB-BC, AB-BC-AC, AB-BC-CC, and AB-BC-AC-CC. Herein, we noticed that AC contacts helped reduce solvent exposure of hydrophobic residues by forming a more compact structure. We thus further classified the collapsed and extended states of various Aβ42 monomeric conformations into a few types using the concept of contact patterns.

The hydrogen bond matrix and frequency contact matrix generated for the ensemble of each contact pattern showed a similar characteristic pattern. This suggested that the heterogeneous tertiary structure ensemble of Aβ42 peptide are driven by secondary structure arrangements (*i.e.*, structured segments). On the foundation of ordered segments, the Aβ42 topology is built up via intra-regional interaction, forming specific contact pairs. Each contact pair can supply three different functions in contributing the topological construction depending on the locations of segments. These functions are β-sheet formation, forming a loop to provide a certain orientation between regions and backbone flexibility, and the hydrophobic clustering. For example, AB contacts in representative structures s3 of AB-BC-AC pattern in Figure

4.9 were involved in a loop region to cause sheet formation between the region B and C, and simultaneously establishing hydrophobic core with the region A. On the other hand, s3 plus the other two structures of the same pattern (*i.e.*, s4-collapsed and s5) showed that there was no direction-specificity in monomeric conformations between these regions. To be specific, by displaying the corresponding structures in the aspect of direct intra-regional interactions, s3, s4-collapsed and s5 successively give A-C-B, A-B-C, and C-A-B topologies, indicating the strong ability of the peptide in conformational adaptability and plasticity. Another thing is that most of the collapsed but less kinetically stable structures were in AB-BC-AC-CC pattern, which also possessed the greatest population of the sampled ensemble. It is suggested here that these transient structures are on the pathway between collapsed and extended states, and their flexibility is attributed to CC contacts causing less contribution in the overall stabilization. Also, all helical structures were found with AB-BC-AC-CC pattern with the cross interaction between the termini motif.

Featured SBs in fibrillar architectures such as K28-A42 in *S-shaped* structures [108, 110] and D1-K28, R5-D7 in the *LS-shaped* fibril were not characterized from the present or earlier simulated ensembles. Low occupancies of E22/D23-K28 SBs were found in the structural ensemble of the AB-BC and AB-BC-AC-CC pattern, respectively, that contributes in stabilizing *U-shaped* protein aggregates [102]. Various types of SBs identified from the ensemble were generally in low percentage, which agrees with the earlier simulation data [116, 286]. Unlike those SBs found in fibrils, our simulations showed that the most seen E3-R5 and R5-E11 SBs in monomers function in orientating the N-terminus for the formation of β-strands. This supports the viewpoint that the stabilization of the N-terminal domain of each monomeric unit play an important role in the amyloid oligomerization and fibrilization [333,334]. These SBs in fibrils were also found absent in oligomers [335], which may be attributed to the distinct aggregation mechanism between the primary and secondary nucleation processes. It is proposed that the monomer-to-oligomer process follows a *conformational selection* mechanism that the peptide oligomerization may be dominantly driven by hydrophobic interactions due to the preference of monomers staying at compact hydrophobic collapse with more ordered structures in aqueous solution.

Subunits could undergo further geometric changes with an increasing β-content during the oligomerization process since fibril subunits carry more extended topologies. Within the fibrillar architecture, the β-sheet stacking and SBs become more essential for the overall stabilization. The secondary nucleation was suggested to be catalyzed by the existing Aβ fibril lateral surfaces. One mechanism hypothesized in a recent computational study was that the monomer was recruited through an adsorption to the fibril surface [336]. Herein, the helical structure showed higher propensity for the adsorption and underwent a conformational extension while on the surface, which suggests an *induced-fit* mechanism of the secondary nucleation. Lastly, a few SBs showed up among the ensembles of different contact patterns, which can be explained via their distinct ways of geometric construction.

In this study, we defined three regions on the Aβ42 peptide as the basic building units to describe its various collapsed and extended conformations. We provide a brief summary of applied analyses assisting us to identify the basic building units for the Aβ42 topological construction. Figure 4.14 combines the secondary structure logo, start-stop matrix, and frequency contact matrix to show that (1) the highest propensity of β-strands (ordered segments) and hinges are characterized over the structural ensemble; (2) the consistent secondary structure arrangement divides the peptide into three regions; and (3) outstanding signals of intra-regional interactions also reveal that the peptide adopts three regions which share the same lengths with regions determined from (2). Combined these, we defined four contact pairs between segments which are AB, BC, AC and C1C2, in which AB and BC are the fundamental motifs in the Aβ42 monomer topological construction while AC and C1C2 contacts are related to the formation of preferred collapsed conformations. Four combinations of contact pairs (*i.e.*, contact patterns) were identified from our simulation data, in which β-dominanted structures in AB-BC-AC pattern possessed more steady topologies due to the well-formed hydrophobic core and many interactions between regions.

Our simulations starting from different initial structures were recovering a spectrum of structures composed of collapsed and extended conformations of the Aβ42 monomer via hopping between structures that are driven by specific sequence elements. The flexibility of the peptide is attributed to the conversion between collapsed and extended conformations, specifically, the gain and loss of AC contacts. Extended

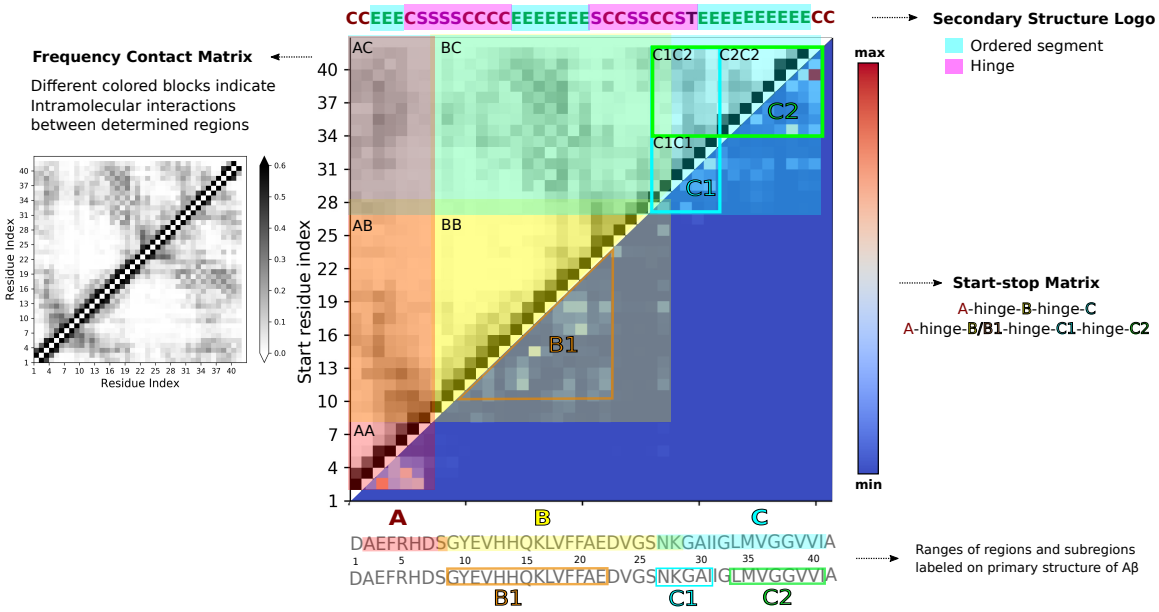

Figure 4.14: Synthesis of analyses in identifying consistent local conformations and describing heterogeneous tertiary structures of the Aβ42 ensemble using ordered segments as building blocks, resulting in categorization of peptide topologies in contact patterns.

conformations were in much lower occupancies compared to the collapsed ones, however, they are important transient states bridging between the same or different compact structures. We utilized the DPC method to discretize such complex dynamics into the continuum of the soluble Aβ42 monomer structure from the extended to collapsed states although it is unreasonable to expect any singular simulation to perform this transition simply. Combining all these results, we conclude that the folding of the Aβ42 monomer can be categorized into a few patterns although the peptide is intrinsically disordered. Also, various architectures built up based on different contact patterns indicate molecular mechanisms of many self-assembly pathways for Aβ42 in solution.

Chapter 5

Response of the Aβ42 Structure to Binary Solvent System and Solvent Polarity Changes

Contributions: Simiao Lu conducted the research and wrote the chapter. Christian Blouin provided editorial input and guidance.

5.1 Introduction

The previous two chapters described simulations that examined the metastability, interconversion and polymorphism of the Aβ42 peptide. As discussed in the first chapter, distinct oligomeric and fibrillar Aβ structures can exist depending on the enviromental conditions. The landscape of an IDP such as Aβ lacks energetic bias between local minima. Different environmental factors can modify the landscape by lowering energy minima and/or raising energy barriers. As Aβ is mainly a hydrophobic peptide, I investigate in this work the effect of solvent polarity on monomeric Aβ42 structure. We apply the methodology described in Chapters 3 and 4 to Aβ42 in ethanol-water mixtures to compare the structural behavior with those obtained in other binary solvents and lipid bilayers.

5.1.1 Solvent Effect on the Tertiary Structure Formation of the Aβ Monomer

The Aβ peptide is amphiphilic: it contains a hydrophilic N-terminus, a hydrophobic central region, and a hydrophobic C-terminus. Each of these segments prefers certain conformations in the aqueous solution. Chapter 4 classifies heterogeneous monomeric conformations of Aβ42 into a few topological types based on these segments. However, we know that environmental changes influence the propensity of a peptide to particular secondary structures. Examples of environmental factors could be pH, temperature, concentration, and the presence of metal ions or small

molecules [83,337–340]. The physiochemical properties of Aβ segments modulate aggregation behavior by affecting the assembly pathway in the nucleation and elongation processes via *conformational selection* mechanism [44, 273, 283, 341]. Depending on the surrounding conditions, various forms of Aβ may co-exist *in vivo* [264]. Also, distinct fibril structures were derived *in vivo* that were different from fibrils formed *in vitro* [111, 112, 264]. Although numerous structural analyses of Aβ responding to changing environmental factors were provided from experimental and computational studies, we have limited knowledge on the behavior of Aβ in the response to solvent polarity changes.

It is thought that Aβ remains predominantly in the α-helical structure in a membrane environment (or apolar organic solvent) [81], while it exhibits a tendency to collaspe into conformations with mainly β-strand and unstructured coil in the aqueous solution [145, 226]. The α-to-β transition in water is discussed in Chapter 3. Fluorinated alcohols, particularly HFIP[1] and TFE[2], are usually used as organic solvents in setting up Aβ aggregation experiments *in vitro* [342, 343]. Under a range of concentrations, fluorinated alcohols form solvent clusters that could reduce solvent polarity around the solvated peptide/protein. Tomaselli *et al.* suggested that conformational transition (or *misfolding*) of Aβ42 occurs when the water content is higher than 80% (v/v) in an HFIP/water mixture [83]. Also, this conformational transition can be reversible by modulating the HFIP composition. Pachahara *et al.* observed that 20% and 50% HFIP/water mixtures favor the α-helical conformation on Aβ alloforms which further prevents fibril formation [344]. Conformational variation of Aβ monomer was also reported in TFE/water mixtures where the helix content was gradually increased to about 80% with the 20% TFE composition [252]. DMSO[3] is another solvent commonly used to prepare Aβ stocks. Unlike fluorinated alcohols, DMSO is presumed to maintain the monomeric state of the peptide in any dilution with water [345]. To the best of our knowledge, few computational studies have focused on the influence of solvent on Aβ conformations and aggregation pathways. One early study performed 20-ns MD simulations with the GROMOS96 43a1 force field to explore the conformational preferences of Aβ42 in HFIP, TFE, DMSO, and

[1] hexafluoroisopropanol
[2] trifluoroethanol
[3] dimethyl sulfoxide

water respectively [140]. Both fluorinated alcohols promoted the helix propensity on the C-terminus, while C-terminal residues of the peptide adopted a random coil structure in DMSO. A more recent computational investigation simulated the conformational changes of Aβ42 in DMSO, revealing decreasing numbers of hydrogen bonds between the central zone and C-terminus compared with topologies observed in simulated aqueous environments [346].

5.1.2 Structure of Ethanol-Water Solutions

The ethanol-water binary solution exhibits several structural and dynamic anomalies, such as partial molar volume, diffusion coeffcient, viscosity, excess entropy, compressibility, *etc.* Although controvertial results were reported according to different studies [347–349], the existence of ethanol-rich clusters in the binary mixture has been widely accepted. Franks and Ives first hypothesized that the anomalous behaviors arise from structural transformations in ethanol-water systems at low concentrations [350]. They proposed the *iceberg* model by describing it as the formation of a low entropy cage of water with strong hydrogen bonds around hydrophobic headgroups of ethanol molecules in the binary solvent. This idea is supported by various experimental and theoretical studies, while different perspectives were introduced for the understanding of this phenomenon [347, 351–354].

It is a widespread view that alcohol and water must be mixed homogeneously at the molecular level. However, experimental studies reported the co-existence of ethanol- and water-clusters at various concentrations. In a low frequency Raman spectroscopy study, Nishi *et al.* observed a local structural evolution of the solution at 0.2 mole fraction ($\approx$ 35% vol) [355]. They indicated that the ethanol- and water-aggregates are too weak to cause microscopic phase separation. A later neutron diffraction analysis revealed the segregation between methanol and water, which was due to incomplete mixing at the molecular level [347]. Guo *et al.* elucidated that most methanol and water molecules exist as the structures similar to the ones found in pure liquids according to an X-ray emission spectropscopy experiment [356]. Further evidence using mass spectrometric analysis has helped clarify that the microscopic phase separation occurs between a wide ratio range of 10-90% volume fraction of ethanol [357].

Other experimental and theoretical studies suggested that aberrant thermodynamic properties are associated with the formation of clathrate-like structures in the binary solution as a result of hydrophobic hydration [358–362]. The combination of hydrophobic interactions between ethanol molecules and hydrogen bonding between the hydroxyl groups of water and alcohol molecules may act synergistically to drive each other on the formation of microheterogenerous clusters in the mixtures. Cipiciani and coworkers suggested a significant change of the mixture properties at mole fraction $\chi = 0.055$ ($\approx 15.8\%$ ethanol by vol) [363]. With mass spectrometric analysis, the formation of ethanol clusters was observed at 0.07 mole fraction ($\approx 20\%$ vol) [364]. Dolenko *et al.* calculated the enthalpies of formation/weakening of hydrogen bonds for aqueous ethanol solutions of different compositions [365]. With other supporting data, they confirmed the existence of clathrate-like structures in mixtures at around 20% ethanol (vol).

Computational studies using MD simulations were conducted in an attempt to capture the structure of the ethanol-water mixture and its low concentration limits. Banerjee *et al.* suggested that abnormal properties of binary mixtures are due to the sudden appearence of a bicontinuous phase at the concentration range of 17-26% ethanol (vol). Noskov *et al.* performed simulations for the structural characterization of the mixtures at various compositions [353]. They found that there is a transition from a complete percolating network of hydrogen bonded water molecules at low concentration to a nonpercolating hydrogen bonded network at high concentrations. Ghoufi and coworkers also observed the transition of the hydrogen bonded network at $\chi = 0.5$ mole fraction ($\approx 76.4\%$ ethanol vol) [354]. Another computational study suggested that the hydrophobic association of ethanol molecules is attributed to a brittle hydration shell [366]. A recent study showed the anomalous behavior of physical and dynamic properties of the ethanol-water mixture at a composition range ($\chi = 0.10 - 0.15 \approx$ 26.4-36.3% ethanol vol). A change of hydrogen bonding pattern of ethanol is found at similar concentrations [367].

5.1.3 Objectives

In this chapter, we explore the IDP sensitivity to solvent polarity. We perform MD simulations on Aβ monomeric system by dissolving the peptide in a series of

ethanol/water mixtures. We examine how the structural and dynamic properties of the Aβ42 monomer change as a function of the ethanol composition. Our choice of solvent mixture is based on the considerations of three perspectives. First, ethanol lowers polarity without phase separation, and it is usually used as a model to inspect the balance of hydrophobic interactions and hydrogen bonds in the hydration of proteins. Second, low to moderate consumption of alcoholic spirits may protect the brains from aging and even reduces the development of AD [368,369]. This discovery inspired studies to explore the effect of ethanol on Aβ aggregation. Combining experimental analyses *in vitro* and computational simulations have shown that ethanol reduces the toxicity of Aβ by altering its structural stability [370]. New evidence revealed that 8.3% (vol) ethanol composition affects the Aβ pentamer stability [371]. However, the molecular mechanisms of this association are largely unknown (to our knowledge). Lastly, there is a lack of studies on molecular description on folding events of IDPs (*i.e.,* Aβ in this case) in the aqueous solution of ethanol. Understanding the solvent polarity on amyloid fibrillization requires characterizing the structural features of the monomer as it transitions from a polar to an apolar environment. Experimental techniques such as SAXS and CD [372–375] provide important information on the protein structural properties in the presence of ethanol. These studies revealed that the ethanol-induced conformational changes of proteins at various concentrations arise because of the anomalous behaviors described above.

In the present study, we report the behavior of Aβ42 in the ethanol-water binary solutions. To compare with previous results derived from aqueous solution, the system of α-helical Aβ42 (PDBID: 1IYT) is simulated in the mixtures using the same force field and water model combination as in Chapters 3 and 4. Here, our focuses are monitoring the structural response of the peptide to the varying solution polarity and determining how the monomer and ethanol affect each other. According to this secondary and tertiary structure analysis, we found that the α-to-β transition can be reversible depending on the solvent polarity. We also discuss our results in terms of the structural stability of Aβ42 and its interactions with the ethanol cluster at various ethanol concentrations. We observed the insertion of the Aβ42 C-terminus into the ethanol aggregate when the water content was below 50%. As part of the transmembrane protein, Aβ42 is initially embedded into the lipid bilayer before

proteolytic cleavage. Understanding the Aβ-membrane interaction after the peptide is produced is crucial for exploring the mechanisms of interfacial folding and the initial stage of self-aggregation. These results suggested that environmental polarity could be one of the key factors determining how Aβ and the membrane bind to each other.

5.2 Method

5.2.1 Molecular Simulations of the Aβ Monomer in Binary Mixtures

The NMR structure of Aβ peptide, solved in an apolar environment (PDB ID: 1IYT [81], was used as the starting structure and obtained from the Protein Data Bank database [292] (www.rcsb.org). For system set-up, the monomeric structure was modified using the PDBnet module [376], a Python library that allows the users to manipulate MD trajectories, as well as perform various analyses. MD simulations were performed using the GROMACS software package [234] and GROMOS96 54a7 force field for the peptide and mixed solvent [175]. For each simulation, the extended simple point charge model (SPC/E) was used to simulate water. The coordinates and topology of ethanol were generated in the GROMOS format using the PRODRG online server [377]. The partial charges and torsional parameters of the topology file were manually adjusted [378]. Inserted ethanol molecules were treated as united atoms with the GROMOS96 54a7 force field, *i.e.*, the full atomistic details have been retained except for the hydrogen atoms attached to the carbon atoms. Dodecahedron simulation boxes were generated containing ethanol and water in a series of compositions (5%, 10%, 15%, 20%, 25%, 30%, and 50% ethanol (vol)). Amino acids were deprotonated, and then the prepared protein system was solvated into the simulation box filled with the ethanol-water binary solution. Each simulation box was created by using periodic boundary conditions, where the minimum distance between peptide and the box edge was 1.0 nm. MD simulations of the ethanol-water binary solutions without the protein system were also performed for comparison. More details of these simulations including box sizes are summarized in Table 5.1.

The following parameters were used for all simulations in this study unless otherwise noted. The monomeric Aβ42 structure was neutralized by adding counterions (Cl^- and Na^+) before equilibration. The neutralized system was minimized in the

pure water or mixed solvent environment by the steepest descent algorithm for 500 steps to relax the protein-solvent system. The leapfrog Verlet integration algorithm was used with an integration time step of 2 fs. Periodic boundary conditions are implemented in x, y, and z directions. The long range electrostatic interactions were calculated using the Particle Mesh Ewald summation with a Fourier-spacing of 0.1 nm and an interpolation order of 4 [236]. Coulomb and van der Waals cut-off distances were both set to 1.0 nm. The Verlet cut-off scheme [237] was used to reach high performance when computing non-bonded interactions. A maximum force less than $100\,\mathrm{kJ\,mol^{-1}\,nm^{-1}}$ was obtained for both systems at the end of the energy minimization. A NVT simulation was conducted for 100 ps. The LINCS method [238] was used to restrain all bonds of the solute and ethanol and water molecules for an integration step of 2 fs. The protein and the solvent (water or ethanol/water) were coupled separately to a modified Berendsen thermostat called *V-rescale* [239] at 300 K. Then, a NPT simulation was performed for 100 ps to generate the initial structure for the production simulation. Pressure coupling was requested in this phase to the Parrinello-Rahman barostat [240, 241] at 1.0 bar with a compressibility of 4.5×10^{-5} bar. In the production run, all parameters were set to the same step of equilibration and the coordinates of the system were saved for every 20 ps. For the simulations of binary solutions, energy minimization and equilibration at constant temperature and volume were performed. A production run for 20 ns at constant pressure and temperature was then carried out after equilibration.

Table 5.1: Proportions of the average secondary structure content of Aβ42 for MD simulations starting from initial structures in different shapes.

EtOH/H_2O v/v (%)	Simulation Box Volume, after Solvation (nm^3)	Number of Molecules (EtOH/H_2O)	MD Simulation (μs)
EtOH/H_2O Binary Mixture			
10	272.480	280/8147	0.02
20	272.480	563/7459	0.02
30	272.480	844/6793	0.02
Aβ42 (PDBID: 1IYT) in EtOH-H_2O Solutions			
0	266.882	0/8696	0.8 + 0.4 (replica)[a]
5	265.548	88/8323	0.4
10	272.076	229/7957	0.8 + 0.4 (replica)
15	277.517	422/7544	0.4 + 0.4 (replica)
20	283.317	563/7182	0.8 + 0.4 (replica)
25	293.408	704/6855	0.4
30	294.342	792/6585	0.8 + 0.4 (replica)
50	325.545	1350/5413	0.4

[a] A repeated MD simulation of identical atomic coordinates and parameters with randomly created initial velocity

5.2.2 Trajectory Analysis Protocol

Ensemble Analysis

To characterize the ensemble of Aβ monomer sampled in different solvent mixtures, we clustered the obtained structures and abstracted the essential motions as per Chapters 3 and 4. Detailed procedures are provided in Section 3.2.4 and Section 4.2.4.

We used the concept of contact pattern defined in Chapter 4 Section 4.3.4 to classify the Aβ42 ensembles obtained from simulations performed in 5%, 10%, 15%, 20%, 25%, 30% and 50% v/v ethanol/water mixtures into a few types of architectures. A brief summary of the procedures are provided as follows. We first identified the range of ordered segments using the secondary structure logo and start-stop matrix (see Chapter 4, Sections 4.2.4 and 4.2.4 for details). An ordered segment was defined to contain no less than four residues that carried consistent α-helix or β-strand character on a certain structure S_n. For example, a pattern of three-segment separated by two hinges was detected over the Aβ42 monomer structure in an aqueous solution, where segments are defined as A, B, and C. Also, tertiary structure analysis (frequency contact matrix) shows a characteristic pattern of intramolecular contacts, suggesting that

monomeric structures break down into three regions separated by two hinges. The robustness of this classification stems from the agreement of the locations between secondary structure segments and regions of intramolecular interactions. Thus, these segments could be considered as building blocks for the construction of Aβ42 tertiary structure. Then, we tested the existence of intra-region contacts between segments by generating the contact map for each structure in a simulation (see Section 4.2.4), resulting in contact pairs (*e.g.*, AB, BC, AC). Lastly, the possibilities of combinations of contact pairs were calculated over the obtained ensemble (*e.g.*, AB-BC and AB-BC-AC), and each combination was called a contact pattern of Aβ42 peptide.

5.3 Results

Volume fraction is used for the ethanol/water mixtures in the following text (*i.e.*, % ethanol).

5.3.1 Secondary Structure of the Aβ42 Monomer in Ethanol-Water Mixtures

Mixtures of water and ethanol significantly decelerate the α-to-β transition of Aβ42 even at low concentrations. Figure 5.1 includes the annotated content of all DSSP secondary structural elements for each simulation shown in the format of a sequence logo. Each logo is comprised of stacks of letters, with the height of each letter proportional to the contribution to the information content of this secondary structure assignment at this position over the obtained ensemble. The overall height of each column (y-axis) is measured in bits and adjusted based on the information content of the input. The initial conformation (or the NMR structure) with two helical regions (S8-G25 and K28-G38) connected by a β-turn was formed in an 80% HFIP/water v/v mixture that mimics the lipid membrane surroundings [81].

A following study tested the conformational transition of Aβ42 from an apolar to a polar environment by increasing the water content in the mixtures [83]. They claimed the occurrance of an α-to-β transition when the amount of water was greater than 80%. In contrast, no significant β-character of Aβ42 was found in the concentration range of 5-50% ethanol/water mixtures. Low percentages of β-bridge emerge at a few locations of terminal residues in 5% and 10% mixtures, indicating the structural

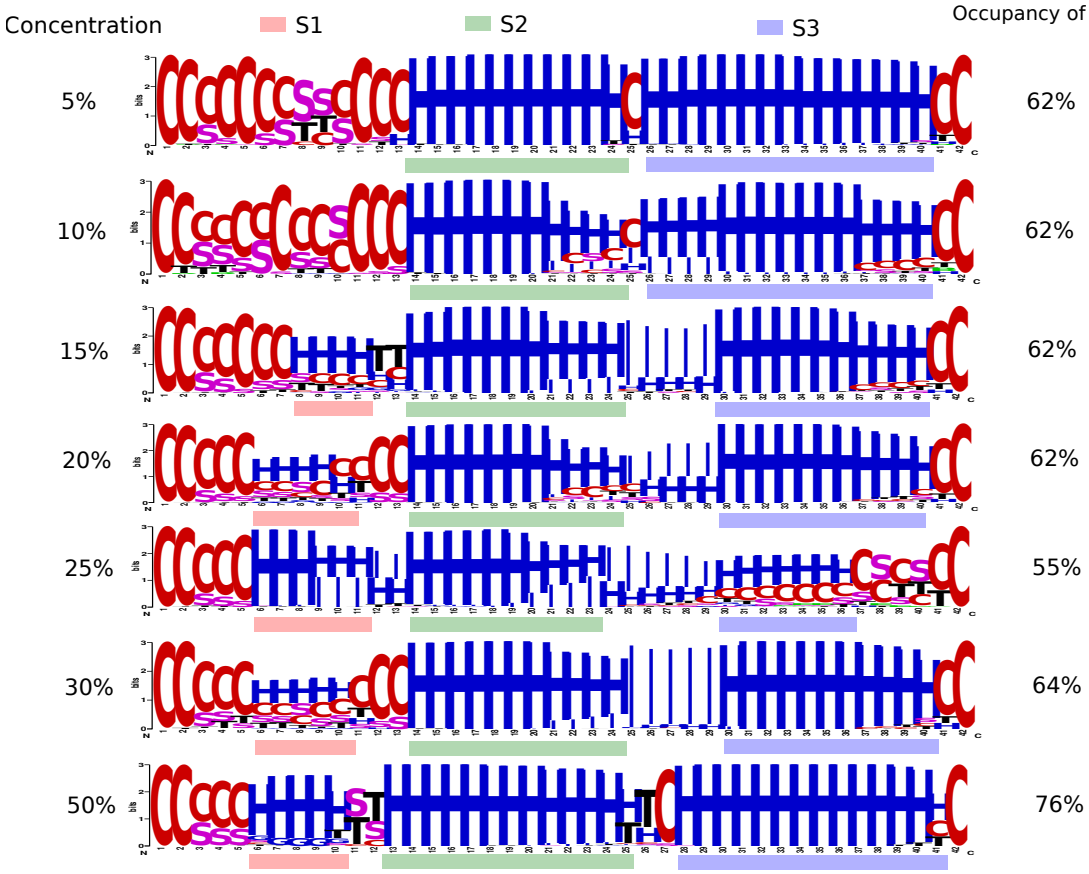

Figure 5.1: A list of secondary structure logos generated from obtained Aβ42 structural ensembles simulated in 5%, 10%, 15%, 20%, 25%, 30% and 50% (vol) aqueous ethanol solutions. In a logo, C, S, T, H, I and B represent for secondary structure elements of coil, bend, turn α-helix, π-helix and β-bridge, respectively. S1, S2 and S3 indicate segments showing highest propensity of forming α-helix according to the corresponding logo. The α-helix content was calculated for each simulation.

transition could start from two termini. Our simulation data showed a structural conversion of N-terminal residues S8-H13 from helices to coil at low concentrations (*i.e.*, 5-10%), while the helical region on the C-terminus was retained. Instead, the aforementioned investigation suggested that the C-terminal helix was lost first with preserved N-terminal helicity with increasing polarity in the medium [83].

The Aβ42 monomer prefers the conformation of three helical segments in the low polarity environment created by the addition of ethanol. Our sampling revealed a relatively stable secondary structure arrangement on the Aβ42 monomer in the concentration range of 15-50% ethanol. This local organization consisted of three helical segments S1, S2, and S3 (*i.e.*, H6-Y10, H14-V24 and A30-V40) that are connected by two hinges locating at V12-H13 and G25-G29 (see Figure 5.1). The long helicity in the N-terminal part of the initial conformation evolved to two segments of α-helix (S1 and S2) separated by a hinge around V12-H13 when the ethanol content was raised above 15%. Local conformational stability of S3 was observed across all ensembles simulated in varied mixtures. High conformational flexibility around the β-turn (G25-S26) was noticed in the mixtures below 50% ethanol composition, in agreement with the previous observation of Aβ42 in 30% HFIP/water solvent [83].

A low polarity environment promotes peptide rigidity by affecting the structural behavior of flexible residues such as histidines and glycines. Histidines (H13 and H14) and glycines (G25 and G29) act as segment breakers in the formation of the conformation of three helical segments. Previous structural studies using experimental and computational techniques suggested that H13 and H14 are part of a turn in Aβ42 fibrils and support fibrillar stability [147, 379, 380]. Our results show that the adjacent histidines could play different roles in the local arrangement. To be specific, H13 contributes to the formation of a turn motif, while H14 stabilizes the turn by maintaining secondary structural rigidity. A recent work using electron paramagnatic resonance spectroscopy showed that H13 and H14 function differently in the Aβ nucleation and elongation processes [56]. On the other hand, the G25XXXG29 motif of the peptide shows π-helical character according to the DSSP definition in the concentration range of 15-30%. This motif was considered to provide backbone flexibility and induce contacts between the central and C-terminal regions (see Section 4.4 for

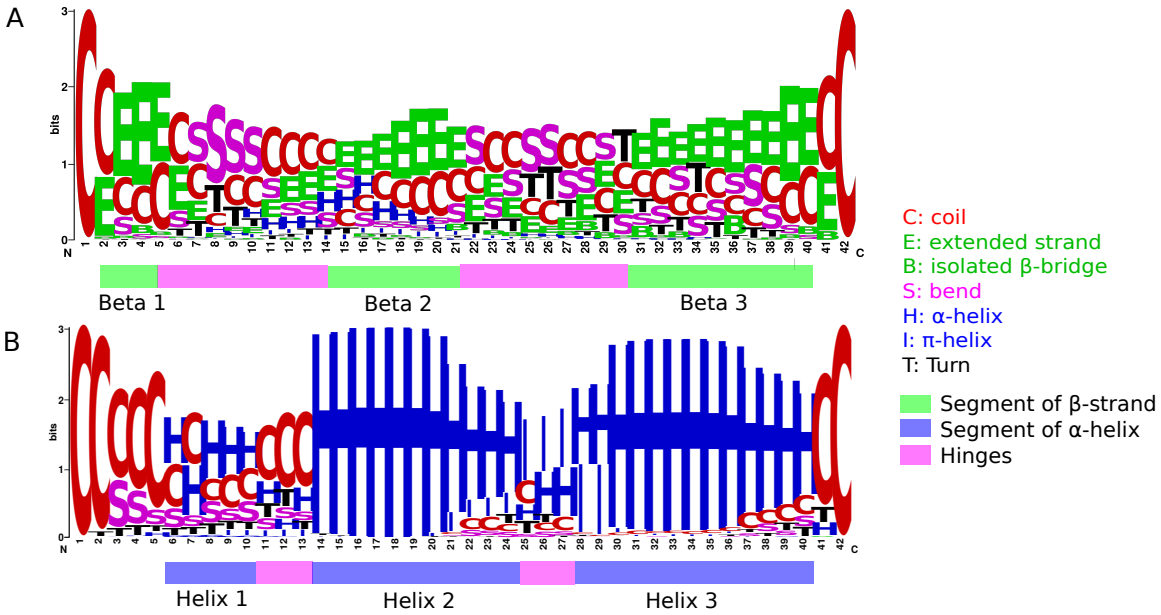

Figure 5.2: Comparison between secondary structure logos generated from simulation data of (A) Aβ42 monomer in an aqueous solution starting from different conformers and (B) Aβ42 structure in various ethanol-water mixtures. Detailed description of logo (A) is provided in Section 4.3.2. Logo (B) shows the frequencies of all DSSP secondary structure over all ensemble populations obtained from simulations in 5%, 10%, 15%, 20%, 25%, 30% and 50% (vol) aqueous ethanol solutions. Segments of β-strand, α-helix, and hinges are shown in green, blue, and pink, respectively.

details). The presence of a π-helix implies the distortion of α-helix and increasing entropy over the region, although DSSP version 3.0.0 may overestimate the occurence of π-helices [381, 382]. An increasing α-helix occupancy was observed when the ethanol content reached 50% (Figure 5.1, logo at 50%). Histidines and glycines play important roles in the formation of longer helical segments on the peptide, contributing to the enhanced the overall rigidity. This result reveals the relation between fluctuating residues and flexibility/rigidity of Aβ42 in response to the environmental polarity.

Localization of residues on the central and C-terminal regions of Aβ42 seed β-strands in a polar environment, whereas a location migration might occur in the transtion of N-terminal helix to β conformation. Figure 5.2B depicts the logo exhibiting the content of secondary structure elements that was calculated over all trajectory data simulated in binary solvents. This logo was then compared with the one generated for Aβ42 obtained for ensembles in water (see Figure 5.2A and Section 4.3.2 for the display and generation details). Aβ42 peptide adopts wider hinges in a polar medium, promoting changes in long-range contacts and flexibilities. In other words, the solvent polarity effect on the local conformation of Aβ42 not only includes the α-to-β transition but also the range and degree of the secondary structural conservation. Additionally, a migration of the N-terminal segment was recognized in the structural transition. Helicity at H6-Y10 transform into a turn from less polar to polar medium, whereas high β-propensity emerges at E3-R5. These observations are believed to be attributed to the co-existence of hydrophobic and hydrophilic interactions in the aqueous mixtures of ethanol, which will be discussed in more details in Section 5.3.3.

5.3.2 Segmentation and Tertiary Structure of the Aβ42 Monomer in Ethanol-Water Mixtures

A classification methodology, as described in Chapter 4, was applied to categorize the obtained ensembles of Aβ42 in various ethanol-water mixtures into different structural classes according to the interactions between local conformational regions. The local conformational regions were determined through a start-stop matrix, with a region defined as a certain range of residues on the peptide frequently forming a segment or regular secondary structure of α-helix or β-strand. A start-stop matrix was computed

based on the combined trajectory of all simulations, recording segment locations for each snapshot conformation. Detail for the generation of a start-stop matrix were described in Section 4.2.4.

The current matrix shown in Figure 5.3A (lower triangle) suggests that the Aβ42 monomer could be segmented into one to three structural elements, traversing from H6-S8, H13-H14 or G25-A30 to D7-E11, V18-S26 or L34-V40. To classify the variable-length segmentation on the peptide, regions a, b, and c were defined to include residues H6-V12, H13-S26, and G25-V40, respectively, representing all three-segment conformations. With these defined regions, structures containing one or two segments could be categorized into one of 11 possible groups. Obtained ensembles yielded two dominant segmentation types b,c and a,b,c with a combined occupancy of 89.2% (see Figure 5.3B). The occupancy of segmentation types involved longer segments, such as a+b, b+c, and a+b+c are low because the existence of the π-helix is not included as part of a helical segment. Noncanonical helical conformations of π-helix characterized from the present simulation data showed an increased bending hinge flexibility, leading to a potential to form a helical kink. The extent of the hinge bending was observed to be controlled by the overall topology and interactions with the solvent molecules.

The monomeric Aβ42 structure was described by the arrangement of two or three helical segments in the range 5-50% ethanol. To investigate the solvent composition effect on the segmentation of the Aβ42, we determined the segmentation groups with the highest occupancies for each ensemble (see Figure 5.3C and D). The result suggests that the Aβ42 monomer was frequently occupied with the two-segment arrangement b,c at low concentrations (*i.e.*, 5% and 10%), whereas the three-segment type a,b,c dominates when the concentration reaches 50%. The peptide adopts both the b,c and a,b,c types with similar occupancies in the range of 15-30% ethanol. One exception is that the monomeric structure also populates with the one-segment conformation a+b at the ethanol concentration of 25%. A gradual decreasing content of helices on the C-terminal residues of the monomer was observed, with the segmentation type transforming from a+b, c to a+b between 350-400 ns.

The formation of a structural motif on the Aβ42 peptide in ethanol-water mixtures was also represented by a contact pair. The ensemble-averaged contact probabilities

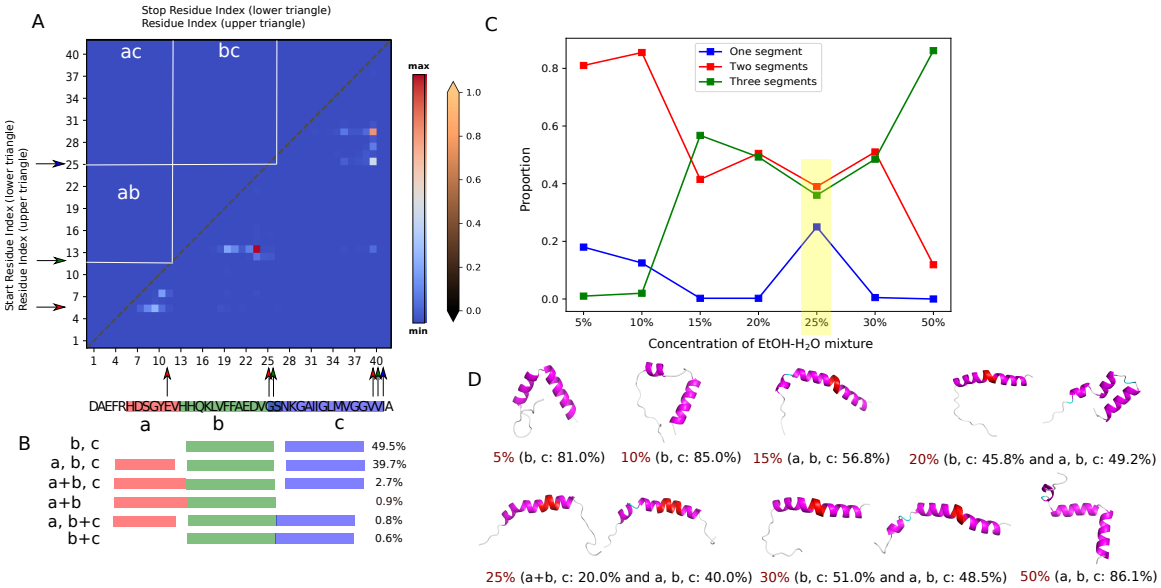

Figure 5.3: Various ordered segments on the Aβ42 monomer in 5%, 10%, 15%, 20%, 25%, 30% and 50% ethanol/water (v/v) mixtures. (A) (lower triangle) The start-stop matrix illustrating emergence of helical segments over ensemble populations obtained from various simulations in different ethanol-water mixtures. Arrows in matching color along the y-axis and x-axis indicate the cutting positions for the segmentation. (upper triangle) The frequency contact matrix illustrates the probability of contact for all side-chain atoms, where two residues are defined to be in contact if the distance between each possible pair of atoms in two side chains is equal or less than 4.5$\mathring{A}$. The multi-domain pattern is demarcated with white lines, indicating intramolecular interactions between region pairs. These regions are assigned on the primary structure as a, b and c according to the start-stop matrix, in the range of H6-V12, H13-S26 and G25-V40, respectively. (B) Six populated segmentation types of Aβ42 monomer in the binary solvent are listed in the decreasing order of population. Segments in different spans are expressed based on a, b, and c, where the ones divided by a hinge are expressed in the combination of letters and a comma and the ones located in connected regions are annotated by letters with a plus symbol. (C) Frequencies of segmentation arrangments were computed for the structural ensemble obtained from each simulation. Selected snapshot structures with highest-frequency segmentation among the population in each simulation are shown in (D), generated by PyMOL [233]. The color codes indicate the secondary structure elements that are α-helix in purple, π-helix in red, turn in cyan, and coil in white. All structures are presented with the N-terminus on the left and the C-terminus on the right.

(Figure 5.3A, upper triangle) were calculated from frames of simulations under various conditions, revealing low frequencies of long-range contacts between three parts of Aβ42 that match regions a, b, and c determined from the segmentation classification. The map is segregated into contact blocks based on this three-region pattern, giving three contact pairs: ab, ac, and bc. Contact map also illustrates the probability of contact within residues of region a (*i.e.*, contact pair aa). These contacts exist mainly due to the frequent formation of salt bridges between N-terminal charged residues. The contact pair aa is not considered in the topological classification, instead, salt bridges on the peptide will be discussed later. The teritiary structure of Aβ42 can be described by the combination of these contact pairs, termed as a contact pattern. Given three contact pairs, possible contact patterns are $2^3 - 1$ in total, where the combinations could be the presence of one pair, two pairs, and three pairs of contacts. Two regions are recognized to be in contact by satisfying two conditions: (1) the distance of any pairwise residues r_x and r_y from these regions is below the defined threshold, and (2) r_x and r_y do not belong to the same segment based on the DSSP secondary structure assignment for the consideration of conformations containing long helical segment located at connecting regions (*e.g.*, b+c).

The structural diversity and the compactness of the Aβ42 monomer is composition and polarity dependent in low concentrations of ethanol. A low polarity environment promotes structural homogeneity of the peptide, resulting in fewer topological patterns compared with those in the polar medium. To distinguish from the annotations of segmentation types, capital letters are used for the contact pattern identification. Aβ42 peptide gave three dominant contact patterns: AB, BC, and AB-BC in the ethanol-water mixtures. These topological patterns depended on secondary structure segmentation, each of which corresponds to two out of six segmentation types listed in Figure 5.3B. Conformations of AB pattern contained b+c segment plus a helical or random coil N-terminus, whereas the ones of BC pattern possess a N-terminal segment a+b with a disordered C-terminus. The latter pattern was populated in the structural ensemble from 25% ethanol. The monomer in the binary solvent had the highest population of AB-BC pattern, indicating its high propensity of breaking into three parts, *i.e.*, segmentation type of a,b,c or b,c. Ensembles obtained in low

Table 5.2: Population proportion of contact patterns for each simulation.

Concentration (% v/v)	Contact Pattern (%)							
	AB	AC	BC	AB-BC	AB-AC	AC-BC	AB-AC-BC	None
5	15.0	0.0	0.0	72.0	3.0	0.0	**10.0**	0.0
10	10.2	0.0	0.0	70.3	0.7	0.0	**18.8**	0.0
15	1.3	0.0	4.8	94.0	0.0	0.0	0.0	0.0
20	4.5	0.0	0.0	92.7	0.0	0.0	2.8	0.0
25	14.0	0.0	**39.0**	44.0	0.0	0.0	0.0	3.0
30	0.5	0.0	0.0	99.5	0.0	0.0	0.0	0.0
50	0.0	0.0	0.0	100.0	0.0	0.0	0.0	0.0

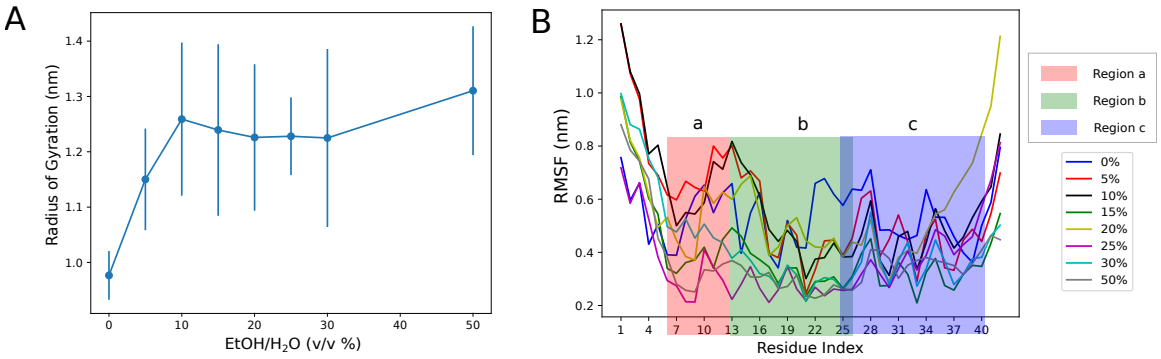

Figure 5.4: (A) Mean value of radius of gyration calculated on the ensemble obtained from each simulation with standard deviation. (B) Root Mean Square Fluctuation (RMSF) of atomic positions averaged on each protein residue is plotted against the residue number for each simulation, where region a, b, and c are highlighted in red, green, and blue, respectively.

ethanol content (*i.e.*, 5-10%) sampled minor population of compact structures involving interactions between termini, yielding AB-AC and AB-AC-BC patterns. The addition of ethanol reduced the structural complexity of Aβ42 such that the peptide presented only one pattern AB-BC when the water content dropped to 50%. This could be related to a shifted preference of the peptide from staying in collapsed to more extended state when lowing the solvent polarity. Figure 5.4A shows that the average radius of gyration (R_g) value elevates significantly when the ethanol content rises to 10%, then a plateau is reached in the concentration range of 10-30% ethanol. As discussed in Chapter 4, an R_g of $\sim$0.9 nm is expected for Aβ42 if it collapses into a completely folded conformation (see Equation 4.7). The mean values of R_g reside between 1.22-1.26 nm with a larger variance than the values predicted for structures simulated in aqueous solution. This indicates that Aβ42 tends to pack into low R_g conformations when dissolving into the binary solvent, however, these MD ensembles contain many extended populations. Considering the reducing structural heterogeneity, ethanol may enhance the conformational stability of extended conformations by decreasing the contact volume between regions. The plateau may imply that ethanol composition becomes a minor factor in affecting the peptide shapes and sizes when it rises to 10% and above.

To further examine the effect of ethanol on the regional flexibility of Aβ42, we computed the residue-based root mean square fluctations (RMSF) of the corresponding backbones with respect to the NMR structure (*i.e.*, the initial structure) over each ensemble (see Figure 5.4B). As the NMR structure was obtained in a membrane-mimicking environment, this plot also compares the local dyanmics of the peptide in various ethanol-water mixtures to that in the aqueous solution. Overall, high ethanol content restricted the conformational freedom of all regions of Aβ42. Residues in between regions (E11-K16 and E22-K28) exhibited higher flexibility than those in other parts in pure water and low ethanol content. In contrast, polarity changes showed less effect on structural dynamics of hydrophobic components including a C-terminal glycine zipper (*i.e.*, L17-A21, G29-G33 and V36-V40). This observation conveys the importance of the intrincially disordered Aβ in maintaining structural rigidity to some extent, coordinating with other disordered parts for environmental adaption. Inter- and intra-molecular interactions between these components have

been considered fundamental in influencing the aggregation pathways [102, 113, 266].

5.3.3 Adsorption of the Helical Aβ42 to an Ethanol-Water Interface

Our simulation revealed that ethanol molecules cluster into aggregate(s) in the binary solvent mixtures in the absence of Aβ42. To test the existence of ethanol clustering, we first performed 100-ns MD simulations for ethanol-water binary mixtures at concentrations of 10% 20%, and 30% ethanol (v/v). Small clusters of ethanol molecules were formed preferentially with dispersed ones surrounded by water molecules in 10% ethanol content. One main ethanol aggregate was formed with fewer sparse molecules nearby at 20%, and 30% ethanol concentrations. According to the concept of clathrate hydrates, water molecules crystallize into cage-like structures around the suitable sized guest solutes [383,384]. A previous MD study revealed that the ethanol guest is hosted by 24 water molecules via hydrogen bonds, forming a configuration containing both hexagonal and pentagonal faces [385]. In comparison, the structure of one ethanol-rich cluster was favored in 30% mixture, resulting in phase separation at the molecular level. This could be explained by there being an insufficient number of water molecules to provide clathrate cavities for all ethanol molecules [363]. Generally, these observations with specific mixing ratios agree with previous experimental results [357, 363].

The presence of Aβ42 affects the ethanol clustering especially at low concentrations. The description below is for the simulation of Aβ42 in 10% ethanol-water mixture. The C-terminal residues of Aβ42 tended to make a small number of hydrophobic contacts first with the alkyl groups of ethanol molecules, forming a quasi-stable state. At this early stage, most of ethanol molecules were spread over the simulation box and the Aβ42 underwent conformational changes by losing N-terminal and G25-G29 regional helices. The formation of stable ethanol aggregates in the simulation with a 10% ethanol-water mixture without the peptide required $\sim$3 ns. The existence of a peptide-ethanol quasi-state slightly slowed the kinetics of ethanol clustering ($\approx$ 6 ns). Over the course of the simulation, there was a combination of rapid increase in peptide-ethanol and ethanol-ethanol contact formation. A polarity difference caused by the initial peptide-ethanol interactions drove further ethanol clustering, leading to less dispersed molecules compared to the simulation for the 10% ethanol-water binary

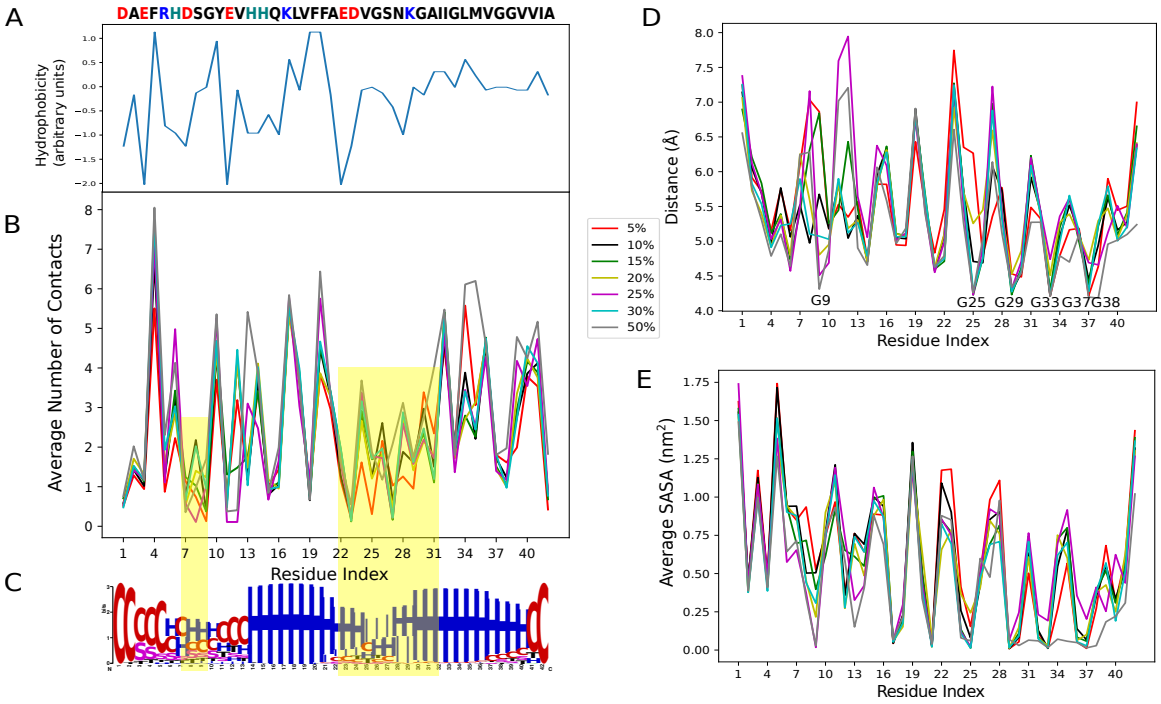

Figure 5.5: Ethanol interaction of Aβ42 peptide in 5%, 10%, 15%, 20%, 25%, 30%, and 50% ethanol/water (v/v) mixtures. (A) Primary sequence of the Aβ42 peptide with basic and acidic amino acided indicated in blue and red, respectively. Hydrophobicity of the peptide is plotted against the residue index, with values according to the hydrophobicity scale of Wimley and White [386] for proteins at a membrane interface. (B) Average number of contacts on per-residue location to the ethanol surface over each trajectory ensemble. Regions with the lowest average number of contacts with ethanol molecules are highlighted with yellow rectangles. (C) Aggregated secondary structure logo over all obtained ensembles in ethanol-water mixtures, which is identical to Figure 5.2B. (D) Average C_α distance to the spatially closest ethanol molecule for each residue of Aβ42 over the production ensemble obtained from each simulation. Residues that are closest to the ethanol surface are all glycines and labeled specifically. (E) Average solvent accessible surface area (SASA) of each residue over the production ensemble of Aβ simulations in various mixtures, where the SASA of a certain Aβ42 conformation means its exposing area to the aqueous subphase.

solvent. Upon fully binding to the ethanol surface, the thermodynamic unfolding of Aβ42 was limited.

The Aβ42 monomer adopted an amphiphilic helical structure at the ethanol-water interface, stabilized by a balance of electrostatic and hydrophobic interactions. In a study of Aβ40 interacting with micelles, the peptide secondary structure and aggregation behavior was strongly affected by the interacting surfactant molecules [387]. Similar structural behavior was observed for the Aβ42 peptide upon adsorption. Assuming that ethanol molecules cluster to an aggregate in the micro-state, an interface with a hydrophobicity difference is created between water and ethanol molecules. The amphiphilic Aβ42 that contains the varying physiochemical properties of different components (see Figure 5.5A), partitions residues with high and low hydrophobicity towards ethanol and water subphases respectively via its helical structure. Large hydrophobic residues of Aβ42 establish high contact volume with the ethanol surface, whereas polar regions such as E22-A30, exhibit the least average contacts, suggesting that these regions are more solvated by water than other parts of the peptide (see Figure 5.5B). Also, the helical Aβ42 adsorbs to the ethanol surface by using the best space advantage. The monomer places itself closest to the ethanol aggregate's exterior by exposing the small-sized nonpolar amino acids (*i.e.*, glycines and alanines) to the surface (Figure 5.5D). This adsorption behavior helps the monomer preserve helicity on its hydrophobic patches, biasing the α-helix to β-sheet transition. It also lowers the opportunity for intramolecular interaction between the core and C-terminus of the monomer (forming a hairpin motif), thereby effectively inhibiting the aggregation process.

The adsorption dynamics of Aβ42 to the ethanol-water interface is correlated to solvent polarity. In this experiment, the solvent polarity was controlled by altering the number of ethanol and water molecules added to the simulation box. Thus, high ethanol content or low polarity solvent results in the formation of relatively large ethanol aggregates. The stabilization of Aβ42 depends on the interfacial network formed between solvent molecules. Conversely, Aβ42 adsorption has an impact on the interfacial hydrophobic/hydrophilic ratio, leading to different ethanol aggregate surface shape. Figure 5.6 shows the Aβ42 configuration after interfacial adsorption in various aqueous ethanol solutions. At a low ethanol concentration (see Figure 5.6A),

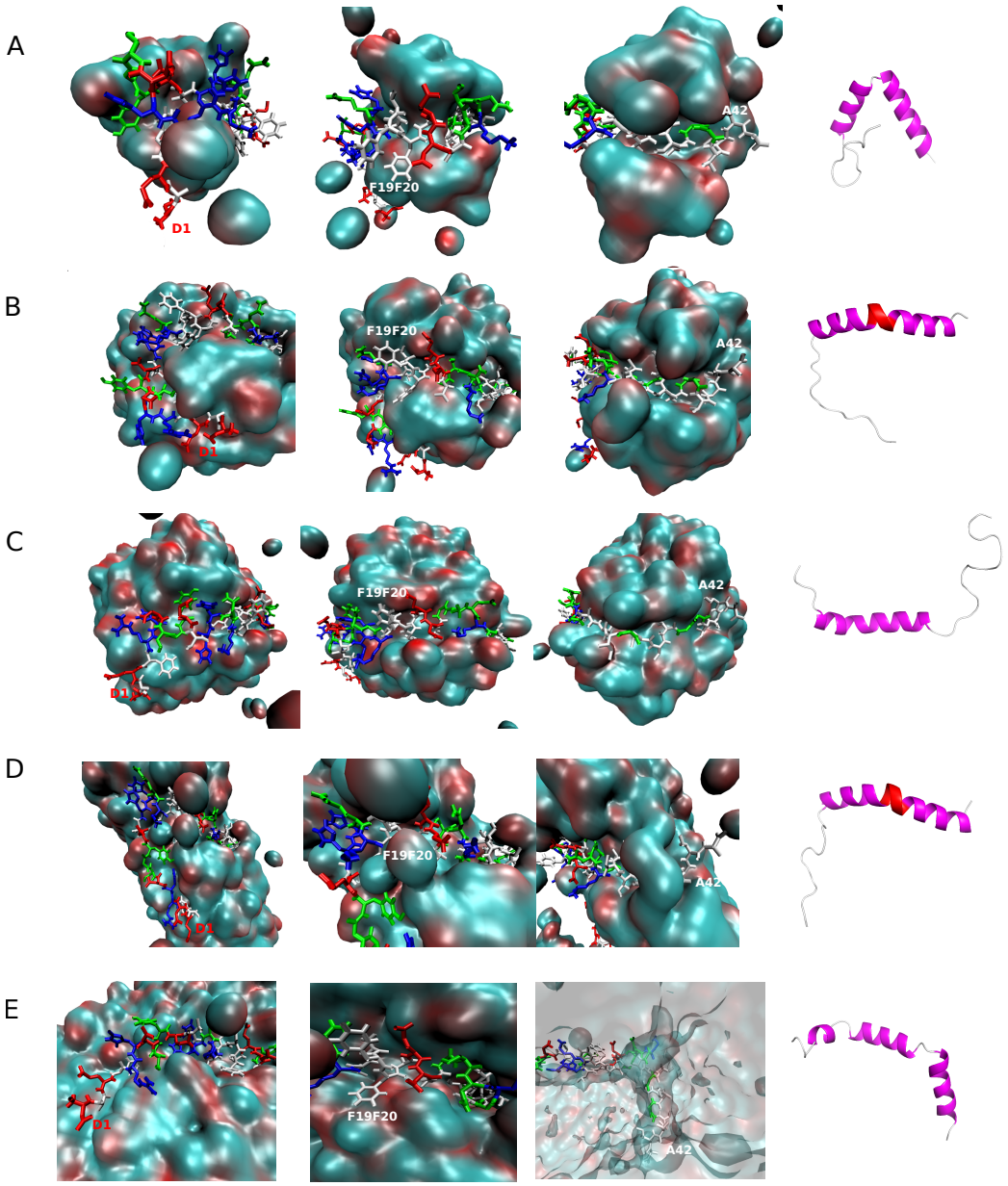

Figure 5.6: Behavior of Aβ42 peptide at the ethanol-water interface in A) 5%, B) 20%, C) 25%, D) 30%, and E) 50% ethanol-water mixtures. Representative conformations of Aβ42 monomer interacting with the ethanol aggregate are displayed. Protein is shown in stick representation and its positively charged, negatively charged, polar, and nonpolar residues are blue, red, green, and white, respectively. Ethanol aggregates are shown in surface representation with the ethyl and hydroxyl groups colored cyan, and red, respectively. Each conformation is displayed in the order of N-terminal, central, and C-terminal regions from left to right. Subplots of Aβ42-ethanol structures were rendered using VMD [96]. Ribbon representation of the same protein structure for each case is exhibited at the end of each row with the N-terminus placed on the left. Secondary structure strips for each centroid, as calculated by DSSP [390], are also shown, where α-helix, turn and coil are purple, cyan, and white, respectively. Subplots of only the protein structure were rendered using PyMol [233].

the peptide structure rearranged to maximize its interacting area with the interface and exposed its hydrophobic residues towards the ethanol subphase. However, it seems that ethanol molecules preferred self-clustering when the ethanol concentration was increased, resulting in higher SASA of Aβ42 hydrophobic patches to the water subphase. Figure 5.6A-C shows that the immersion depth of helical Aβ42 in the ethanol aggregate became shallow when the ethanol content changed from 5% to 25%. The extent of exposure of Aβ42 to the polar subphase was also calculated and is depicted in Figure 5.5E. The increasing solidity of the ethanol aggregate in the concentration range of 5-25% ethanol allows highly solvent accessible to water on the hydrophobic C-terminal region G29-A42. This promoted more extended conformations of Aβ42 and its structural transition to random coil (see Figure 5.6A-C). This trend was reversed in the concentration range of 30-50% ethanol that the solvent accessibility for C-terminal residues is decreased significantly (see Figure 5.5E) due to the insertion of these regional helices into the aggregate (see Figure 5.6D-E). This result was consistent with previous studies of the embedding Aβ42 in membranes wherein C-terminal residues of the peptide adopted transmembrane helical structure [388,389]. Lastly, we noted that the G25XXXG29 motif plays an important role in the adsorption dynamics, interfacial structural rearrangments, and aggregate insertion of Aβ42. The π-helix conformation formed around this region provided backbone flexibility for the peptide's bending movement with the changing shape of the aggregate.

5.3.4 Salt Bridges

Hydrophobic contacts among β-strands and salt bridges (SB) are the two major contributors to the structural stabilization of the Aβ42 monomer in an aqueous solution. Due to the difficulty in forming β-structure over the peptide in the ethanol-water binary solvent, the generation of intra-peptide SBs becomes critical for its conformational stability. Recall in Section 5.3.2 that E11-K16 and E22-K28 are the first two components that undergoes secondary structural change due to their high dynamics. This benefits the structural segmentation, thereby promoting the folding event of the peptide by increasing the chances for intra-segmental contacts. Nevertheless, Aβ42 requires helical characteristics for the surface binding or insertion into the ethanol

aggregate. SBs in this case support the preferential three-segment conformation in two ways. First, these SBs, such as E11-K16 and E22/D23-K28, control the bending movement over the loop regions to a certain degree to increase its adsorbed areas to the ethanol-water interface (see Figure 5.7A). Second, Figure 5.7A-C revealed that the formation of D3-R5, R5-D7, E11-K16, and D23-K28 SBs on Aβ42 could function as zippers to prevent the loss or distortion of helices in a low polarity solvent.

The formation of SBs on Aβ42 in an ethanol-water mixture was correlated to its adsorption behavior. Compared with overall low SB populations observed from the sampled ensembles in water (Section 4.3.4 Chapter 4), high concentrations of ethanol promote the formation of a diverse number of SBs of Aβ42 peptide due to the lack of overall intramolecular contacts. As mentioned in Section 5.3.3, the destabilization of Aβ42 conformation around 20-30% ethanol concentration was observed due to the increasing solidity of the ethanol aggregate, exposing the peptide C-terminus towards the polar subphase. Formation of SB E22/D23-K28 on the Aβ42 monomer at this concentration range appeared to conpenstate for the increasing entropy on the C-terminal region (see Figure 5.8). High occupancies of E3-R5, R5-D7 and D11-K16 SBs were identified at the concentrations of 5%, 10%, 15% and 50% ethanol. Sufficient hydrophobic contact volume between Aβ42 and the ethanol surface were estabilished that retained the helicity on the hydrophobic patches of the peptide. Therefore, only SBs related to N-terminal stability were populated among these structural ensembles.

5.3.5 Conformational Variation of the Aβ42 Monomer in Ethanol-Water Mixtures

Decreasing solvent polarity prevents the original α-helix on the Aβ42 monomer from breaking up. To analyze the conformational variation of the Aβ42 monomer in different ethanol-water mixtures, we first merged the trajectory data generated under each solvent mixture into a single geometric input file with the first 200 ns discarded for each simulation. Principal component analysis was carried out over the Cartesian coordinates of the combined trajectory (cPCA). The result was projected onto a 2D plane formed by the first two principal components (PC), which involved an aggregated 56.73% of the geometric variances (see Figure 5.9A). Ensembles obtained from simulations showed a trend along PC 1 (containing 43.50% variances) with increasing

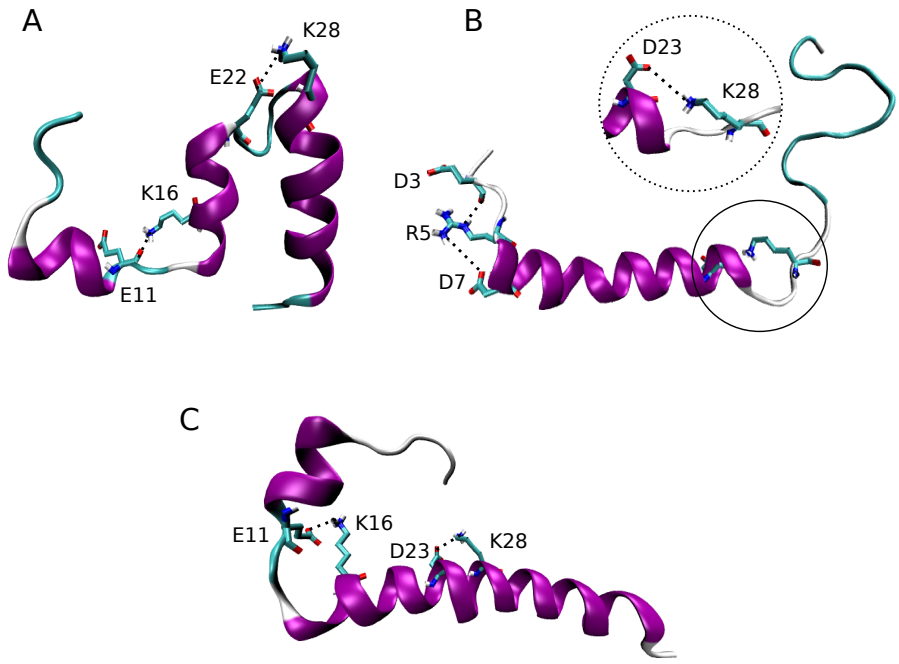

Figure 5.7: Featured SBs are shown on a selected configuration from the Aβ42 ensemble sampled at the concentration of (A) 20%, (B) 25%, and (C) 30% (v/v) ethanol-water mixture respectively, where residues are displayed in stick representation. The backbone is colored based on the secondary structure elements, in which α-helix, turn, and coil are purple, cyan, and white, respectively. All structures are presented with the N-terminus on the left and the C-terminus on the right.

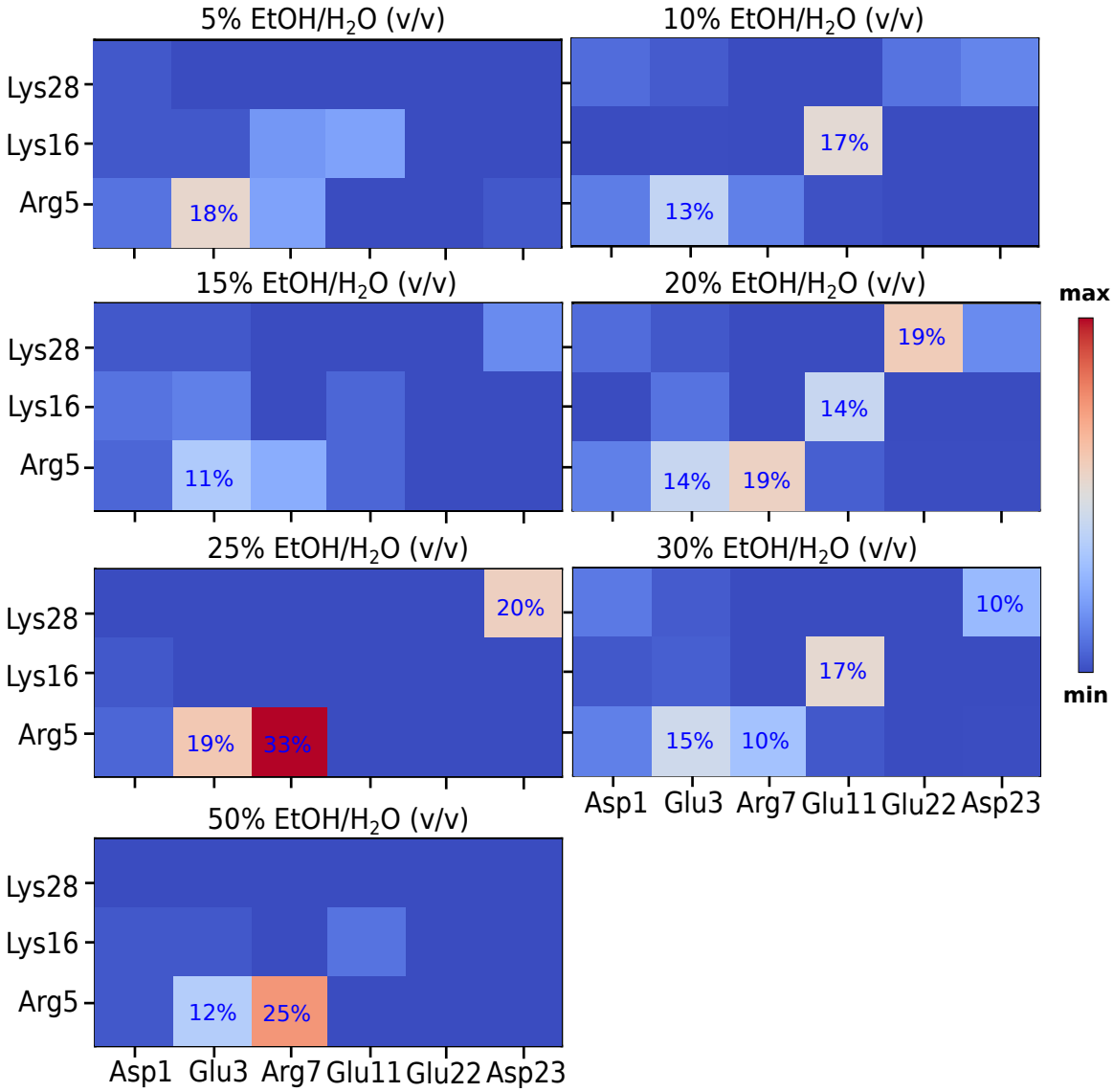

Figure 5.8: Populations of all 18 SBs formed by 3 positively charged and 6 negatively charged residues for the structural ensembles sampled at different ethanol compositions using Barlow and Thornton's definition described in Section 4.2.4. Major SBs with populations over 10% are reported in proportions.

ethanol concentrations. Figure 5.9B and C further depict this trend by comparing the mean values of PC 1 with the standard deviation (SD) and the probability density of PC 1 over the ensemble obtained under each condition. On the other hand, the expression of the major motion within each simulation is distributed over similar ranges along the PC 2 axis (containing 13.23% variances), except for structures simulated in the pure water. That the smallest SD and narrowest distribution of PC 1 coordinates were observed for Aβ42 monomer structures simulated in water suggests that this condition afforded relative steady dynamics as compared with other conditions. Large conformational fluctuations for sampled ensembles in ethanol concentrations of 10%, 15%, and 20% were observed, which could be attributed to the changing shape of the ethanol aggregate surface. A plateau of mean PC 1 coordinates is reached at 15-20% (v/v) ethanol/water, indicating that no prominent conformational differences existed between the ensembles at the current and higher concentrations of ethanol. Figure 5.4 shows a similar trend on the degree of conformational expansion of Aβ42. Combining these results, the α-to-β transition of the Aβ monomer could be significantly hindered when the water content reaches 85% and lower. A slightly higher ethanol composition limit was reported from the previous NMR study using the HFIP-water mixture at the HFIP concentration of 20% (v/v) or less [83].

Ethanol inhibits Aβ42 self-association by promoting the extended shape of the peptide on the ethanol-water interface. With the ordination of all ensembles in the PCA projection, the DPC method (see Section 3.2.4 for details) was utilized to show how topological changes of the Aβ42 monomer correlated with the increasing ethanol composition by binning along the PC 1 axis. This yielded in a series of average structures that captured the continuum of an *open-close* progression on the conformation across the population of all ensembles in the decreasing order of the ethanol/water concentration (see Figure 5.9D). This progression included the decline of helical content especially on the C-terminus, a decreasing distance between termini, and the emergence of a hinge region firstly on the N-terminal side (a3-a5) and then on the C-terminus (a6-a13). A loss of helical structure at the N-terminus was observed in all ensemble populations, inducing an overall entropy on the extended conformations of Aβ42 at relatively high concentrations of ethanol (a1-a9). These results suggest that (1) the α-to-β transition may be reversed by changing the composition of water and

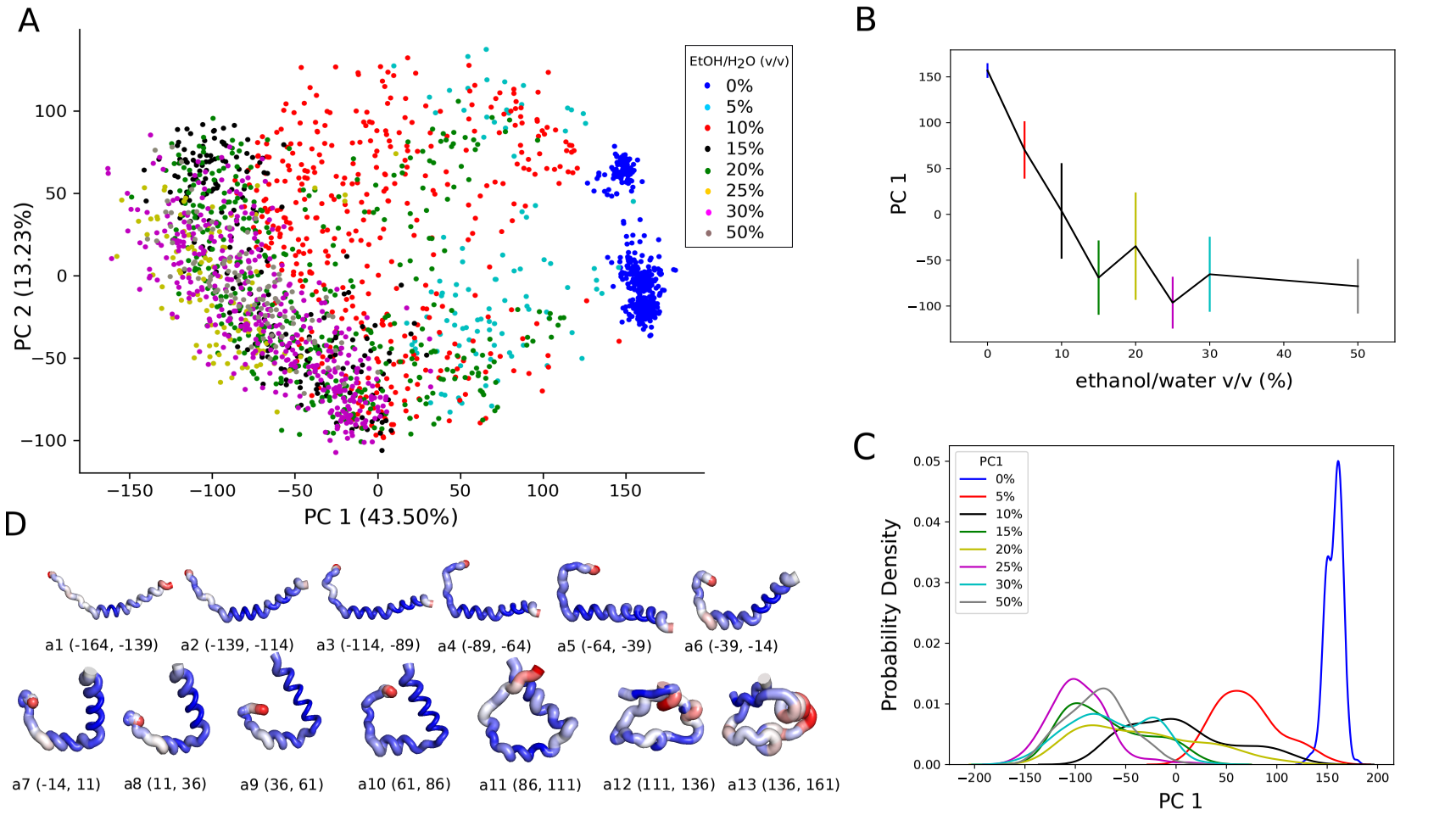

Figure 5.9: PCA of Aβ42 monomer trajectory data obtained in various concentration of ethanol-water mixtures. (A) Projection of the combined trajectory data on a 2D plane formed by PC1 and PC2; (B) mean PC 1 values of ensembles simulated in each ethanol/water mixture with standard deviations; (C) the distribution plot of the ensembles obtained from each ethanol-water mixture along the PC1 axis using the kernel density estimation; (D) A series of average structures (a1-a13) generated by using the Discretizing PC (DPC) method along PC 1 based on the combined trajectory data with a binning size of 25. The thickness and the color range of red-white-purple of the average structures suggest the extent of RMSD changes among structures in the corresponding binning area. All average structures are presented with the N-terminus on the left and the C-terminus on the right.

ethanol, in agreement with a previous NMR study [83]; and (2) ethanol may reduce the Aβ aggregation due to the absence of or low amount of β-structure formation at high ethanol concentration. Another explanation could be that a concentrated ethanol-water mixture prevents the hydrophobic clustering within the peptide, which further blocks the hydrophobic interactions between collapsed monomers in the primary nucleation process. An earlier experimental study revealed that ethanol prevented Aβ dimerization *in vitro* [370].

5.4 Discussion

By dissolving Aβ42 monomer in the ethanol-water mixture, its structural and dynamic properties change as a function of the ethanol composition. Various simulations performed in this study have shown that the kinetics and direction of the Aβ42 α-to-β transition could be manipulated by altering the solvent polarity. Combining analyses of secondary structure character, structural segmentation, and contact pattern classification, a *misfolding* mechanism may be proposed as follows. The loss of helicity on the peptide starts at the N-terminus (around S8-H13) and two hinge regions, resulting in the three-segment arrangement. Therefore, the mobility and secondary structure content of the Aβ42 N-terminal region could serve as the determining region governing the aggregation rate and fibrillar stability. Loop regions play an essential role in providing backbone flexibility so that the peptide folds into compact configurations via hydrphobic clustering between the central and C-terminal regions. A further secondary structural transition on the Aβ42 monomer may occur before and after its recruitment to the oligomer or fibril. A detailed description of the last stage was provided in Chapter 3. Histidines and glycines are key residues for defining the termini of segments, effectively separating ordered and disordered components on the peptide. In this way, a limited number of segmentation types can be formed on the peptide, thereby the oligomeric and fibrillar structures are subject to the combination of secondary structural segments of their subunits. In addition, high propensity of Aβ42 to adopt helical conformations in low polarity solvent. Such phenomenon was reported in previous studies dissolving the peptide into other binary solvents such as HFIP-water [81,391]. Different helical configurations of Aβ42 are favored, depending on the solvent conditions. This further affects the α-to-β transition

rate and formation of fibrillar and non-fibrillar aggregates [344]. Therefore, the fact of heterogenerous structures of Aβ oligomers and the fibril obtained from different experimental studies *in vitro* is expected. Future studies could focus on comparing Aβ42 structural differences obtained under different conditions, where the effect of multiple factors such as solvent polarity, temperature, and pH could be examined.

Ethanol molecules could cluster into multiple small ones or a single big aggregate at the molecular level according to our simulations. Previous SANS study also discovered the ethanol clustering when dissolving lysozyme into the binary solvent [392]. Aβ42 adopts a monolayer amphiphilic helical structure on the ethanol-water interface. Our data have shown that the peptide adsorbs to the ethanol surface when ethanol content is less than 25%, while an insertion of Aβ42 C-terminal residues (G29-A42) into the ethanol aggregate was observed at ethanol concentrations above 25%. Both cases were demonstrated in previous studies of Aβ species interactions with different lipid environments [344, 389, 393–398]. The kinetics of ethanol clustering especially at 25-30% has a major influence on the helical content of Aβ42 C-terminus. Insufficient contact formation between the peptide and ethanol molecules may result in over-exposure of Aβ42 hydrophobic patches to the polar subphase, triggering a loss of helicity. In this case, the peptide tends to stay on the ethanol surface rather than embedding its C-terminus into the aggregate. We conclude that the occurrance of Aβ42 adsorption or insertion to the ethanol aggregate seems to be dependent on the solvent polarity and the helical content of the Aβ peptide. This point has also been raised from a previous computational study of Aβ-membrane interactions wherein the degree of Aβ42 embedded to the membrane is strongly correlated to the cholesterol composition of the lipid membrane and the helical content retained on the peptide [389].

As experimental data on IDPs at a fluid interface are scarce, we compare the adsorption behavior of Aβ42 to interficial properties of globular protein adsorption to an oil-water interface [399], as well as to globular protein lysozyme [392] and amyloid-like aggregation of β-lactoglobulin [375] in an ethanol-water solution. Combining these literature reports with our results, we summarize features of Aβ42 adsorption to ethanol-water interface to provide insights for further studies associated with IDPs binding and aggregating at the fluid or membrane interface. First, in the presence

of a protein in the aqueous ethanol solution, ethanol molecules can either bind with the protein or form clusters in the bulk. Second, the structural response of Aβ42 to the change of solvent polarity is attributed to a sensitive balance of hydrophobic and polar interactions in the binary solvent. Third, increasing ethanol content or decreasing solvent polarity drives structural rearrangement on the peptide including higher helical compositions, larger bending angles around the hinge regions, and a greater extent of exposing hydrophobic patches to the polar subphase.

The inhibition of Aβ aggregation in a low polarity medium can be accounted for by several possible explanations. One is that low polarity solvent promotes helical conformations of Aβ42. The driving force for oligomerization was suggested to be the interpeptide association via a central hydrophobic cluster. Assembly of β-sheets between interpeptide hydrophobic components is required for polymeric stability since hydrogen bonds are stronger than hydrophobic interactions [272, 400]. Deceleration and prevention of the α-to-β transition results in difficulty at the early stage of amyloid nucleation. Another reason could be the sheltering of C-terminal residues of Aβ42 from the polar subphase. The protection of the Aβ42 C-terminal region not only removes the hydrophobic attraction force to some extent but also restricts conformational freedom of the peptide to orient itself for oligomerization. The formation of the *U-shaped* monomer was suggested to be aggregation prone due to its configurational benefit in estabilishing more hydrophobic and favorable electrostatic interactions (*i.e.*, salt bridges and β-sheet hydrogen bonds) during fibrillization. Hydrophobic interactions of ethanol molecules around the protein significantly lower the chances of contact between the central and C-terminal regions. The last element is attributed to the positioning of GXXXG motif within the ethanol aggregate in the insertion of Aβ42. It was demonstrated that the flat surface provided by the glycine residues can be regarded as a framework for a wide variety of specific interactions for dimerization [401]. Our simulation of Aβ42 at 50% ethanol content has shown that the G25XXXG29 motif helps place the Aβ42 central helices on the aggregate surface whereas the C-terminal helices are embeded. Such fixed helix orientation of the central region would limit its hydrophobic interaction with another monomer.

5.5 Conclusions

Aβ42 has a fascinating sequence design such that in polar solvent the peptide collapses into a micelle-like structure while in low polarity or nonpolar solvents the peptide forms a monolayer amphiphilic helix structure. The conformational variation of Aβ42 in binary solvents with different ethanol content has been shown. A low polarity environment restricts the peptide conformational freedom by first increasing the local conformational conservation and minimizing the hinge ranges. The solvent polarity has a direct effect on the fluctuating locations and lengths of helices. We have observed a rigid C-terminal helical structure due to its hydrophobic interactions with the ethanol molecules. The stabilization of C-terminal residues also hinders the Aβ42 aggregation, including inhibition of the conformational transtion and formation of the aggregation prone *U-shaped* structure. Alteration of ethanol concentration or solvent polarity could control the α-to-β transition rate and also reverse the process. By manipulating the surrounding conditions in the preparation Aβ42, distinctive monomeric structures could be formed, leading to various aggregation pathways.

Our adsorption analysis suggests that (1) ethanol molecules cluster at the molecular level; (2) Aβ42 adsorbs to the ethanol/water interface and forms a monolayer amphiphilic helix structure; (3) the peptide has the ability to modify the surface of ethanol aggregates and even insert into the ethanol layer at high concentration; and (4) the structural behavior of the peptide in the binary mixture is composition-dependent. As part of the transmembrane protein, Aβ42 is initially embedded into the lipid bilayer before proteolytic cleavage. Understanding Aβ42-membrane interaction after the peptide is produced is crucial for exploring the mechanisms of interfacial folding and the initial stage of self-aggregation. These results suggest that environmental polarity could be one of the key factors determining how Aβ42 and membranes interact.

Chapter 6

Concluding Remarks and Future Work

6.1 Concluding Remarks

Intrinsically disordered proteins (IDPs) possess some fascinating biophysical features that are far beyond the rules of classical structure-function paradigm. Not only the concept of intrinsic disorder but also their active involvement in various biological processes and association with human diseases inspire us to expand our knowledge of these special members of the protein world. The primary objective of my research was to characterize conformational states of intrinsically disordered Aβ42 in different environments. To link the experimentally known conformations of Aβ42, exploration of the vast conformational space was achieved by performing microsecond simulations of the peptide monomer starting from different experimentally solved structures. We learned that Aβ42 contains ordered and disordered segments that are not completely independent. These parts have the ability to sense the environment and respond to each other's change. According to this theory, I have classified its tertiary structures into a few types based on the ordered segments. This work contributes to the knowledge of some common behaviors shared with other IDPs. Also, a few designed conformational characterization strategies could be applicable to structural and dynamic studies of other globular and disordered proteins.

Beginning in Chapter 3, a specific α-to-β transition or *misfolding* pathway of Aβ42 was examined by performing systematic simulations originated from defined opposite conformation ends towards each other‘s direction in an aqueous solution. My results indicated that Aβ42 frequently interconverts between collapsed and extended conformations along the transition route. Extended states are intermediate states bridging between different collapsed states on the free energy landscape of Aβ42. Due to the flatness of Aβ‘s energy landscape (also for other IDPs), a certain folded configuration could easily convert into another, revealing the conformational heterogeneity, complexity of *misfolding* and aggregation pathway network and high structural plasticity

of the peptide. This chapter was also featured by the use of principal component analysis (PCA) in a way to break down and simplify the folding event into a series of continuous motions.

Although Aβ species carry high flexbility in the aqueous medium, I noticed that some parts of the peptide actually show homogeneous characters especially in secondary structure content. Chapter 4 aimed to identify these character and categorize topologies based on the rigid segments into more specific classes rather than just collapsed and extended conformations. Simulations initiated from different starting points were carried out and combined into the same sampling space in the utilization of PCA. By defining conformational states on the joint landscape, classifying the conformations, and discretizing motions between these conformations, I was able to organize complex dynamics from currently explored space into the continuum of open/close folding events.

To test the robustness of previously applied approaches, I executed the aforementioned analyses on the sampled ensembles of Aβ42 in various ethanol-water mixtures in Chapter 5. Another purpose of this chapter was to examine the structural sensitivity of Aβ42 to the change of a single environmental factor which is solvent polarity in this case. The polarity alteration has a direct impact on the reaction rate and directions of the α-to-β transition. Surprisely, I observed that ethanol molecules cluster into different sizes of aggregate in the binary mixture depending on the concentration and the presence of protein. Aβ42 adsorbs to the ethanol surface at low concentrations and embeds its C-terminus inside the aggregate when the ethanol content rises to 50%. These results share consistency with the observation of the peptide behaviors in different lipid environments. Moreover, I summarized the effect of low polarity on inhibiting the Aβ aggregation.

Taken together, the results in Chapters 3 to 5 have demonstrated the applicability of MD simulations in investigating structural features of IDPs for specific reaction pathway and responses to effect of different environmental factors.

6.2 Future Directions

It would be interesting to further explore the questions addressed in this thesis using different modeling techniques in order to verify the results reported here. However, the

findings do correspond well with previously published work. A comparison between my results and previously reported experimental data in Chapter 4 has shown a generally good consistency. Since there is no force field specifically designed for IDPs, the assessment of other reliable force fields such as CHARMM22 and OPLS-AA with currently used GROMOS96 54a7 is necessary for the accuracy of structural sampling. For the simulations testing the peptide reaction to environmental changes such as solvent polarity, different sampling techniques including combinations of force fields with other water models are needed for evalution. Also, other advanced sampling methods such as replica-exchange molecular dynamics (REMD) should be employed for a wider conformational space exploration in the future experiments.

6.2.1 Aβ42 Oligomers in the Aqueous Solution

Over two decades ago, emergence of studies proposed that small soluble oligomers of Aβ can cause neurotoxicity rather than the mature fibrils [213, 402, 403]. One hypothesized mechanism involves the ability of oligomeric species to penetrate into the membrane and form unregulated ion channels in brain cells, leading to disturbances in neuronal Ca^{2+} homeostasis, alteration of synaptic information processing and utimately inducing neurotoxicity [404, 405]. The original idea is attributed to the significant correlation between intra-neuronal accumulation of Aβ and elevated level of the calcium in the cortex of Alzheimer's disease patients. Previous simulation data revealed that the trimer to pentamer species of both Aβ alloforms are prone to form pores that is sufficiently large for the access of water molecules and Ca^{2+} ions [406, 407].

Initial systems of these simulations can be employed by extracting a single monomer from the *U-shaped* Aβ42 full length fibril structure that was determined with the combined NMR and EM data *in vivo* (PDB ID: 2NAO) [109]. The simulation box size for dimeric and trimeric systems is enlarged by increasing the minimum distance between solute and the box boundry to 2.0 nm. To compare with previous simulations of monomers, the same combination of force field and water model could be used. Two 400-ns equilibration MD simulations have been performed for the dimeric and trimeric system, respectively. However, more replicas with longer simulation time are needed for the sufficient exploration of energy landscape. REMD is preferred in the

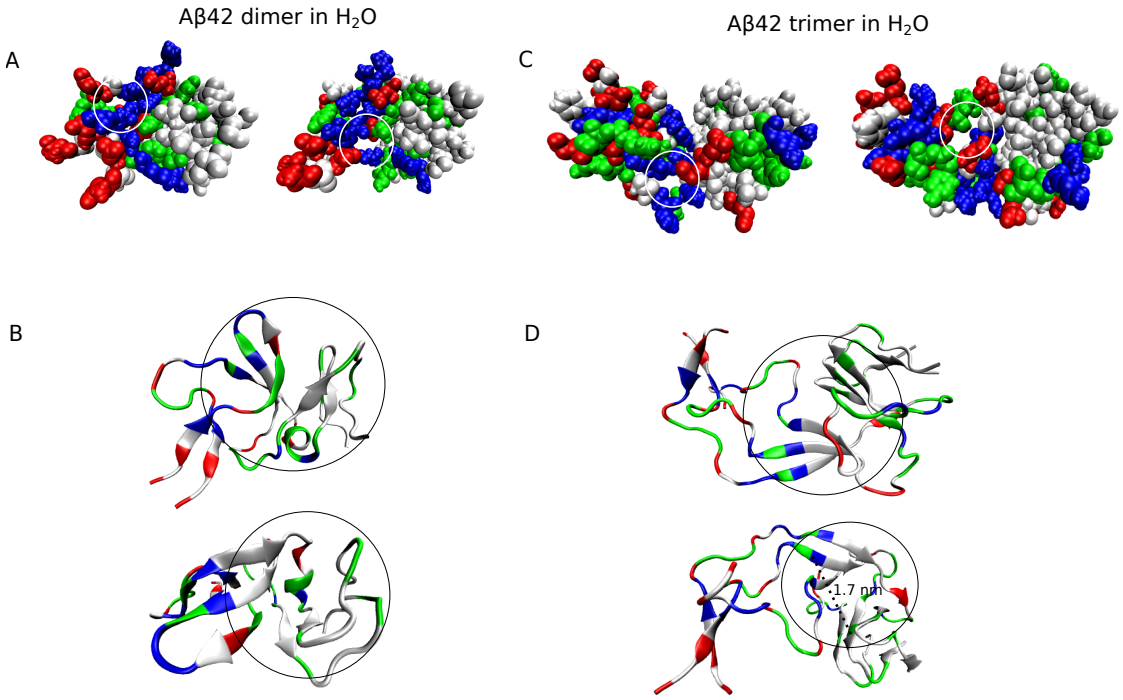

Figure 6.1: Selected snapshots of porous conformations of (A) Aβ42 dimer and (C) trimer in solution are shown in the representations of Van der Waals surface using Visual Molecular Dynamics (VMD) software version 1.9.4a37 [96]. Locations of pores on the structures are indicated in white circles. β-Barrel-like motif are implied in black cirles on (B) Aβ42 dimer and (C) trimer representative structures, where both conformations are displayed from the front (upper) and top (lower) view. Basic, acidic, polar and non-polar residues are blue, red, green, and white, respectively.

consideration of sampling efficiency.

A preliminary result has shown that Aβ42 trimer can form porous conformations that is water- and ion-permeable in solution. Figure 6.1A and C depict the formation of pores on Aβ42 dimer and trimer topologies from the present simulation data even in the absence of lipid bilayer. We utilized the MOLEonline 2018 update [408] to determine the formation and calculate the size of pores in the obtained oligomer configurations. Only a few dimeric configurations with pores were recognized, however, the size of these channels are too narrow for water molecules and Ca^2+ ions to pass through. Porous conformations were observed more frequently on trimeric configurations and usually formed between peptides via charged and polar residues sitting between β11 and β2. The narrowest region of these channels are at the average of 1.7-2.0 $\mathring{A}$, which is sizely large enough for water molecules and Ca^2+ ions (about 1.4 $\mathring{A}$) to travel through. This result is consistent with the observation from Nguyen *et al.* [407] and Voelker *et al.* [406].

Both Aβ42 dimer and trimer form partially open barrel-like motif in water. β-Barrels are common structures found in outer membrane of Gram-negative bacteria and in the mitochondria membrane. Earlier studies tried to link the occurrence of a barrel structure in Aβ oligomers with their formation of an ion channel in a lipid bilayer [409, 410]. To investigate this, a number of computational studies modelling different Aβ42 assemblies were conducted and revealed the emergence of barrel motif on the oligomeric topologies [407,411,412]. Our simulation data showed the formation of partially open barrel-shaped Aβ42 dimer and trimer configurations in solution (Figure 6.1B and D), respectively. Different from previouly observed barrel-shaped Aβ oligomer structures containing six to eight antiparallel β-strands [407, 412], the barrel-like motifs within dimer and trimer are made by the mixing of parallel and antiparallel β-strands mainly between hydrophobic residues located on regions of β2 and β4. Specifically, the dimer barrel is composed of five β-strands including an intramolecular β-turn-β motif, while the trimer forms more complete cylinder-shape containing six strands with an average diameter of 1.7 nm. The size of the β-barrel in the trimer is larger than the observation from Nguyen *et al.* (which is about 1.2 nm) [407]. Such a difference could be caused by the use of different simulation protocol *i.e.*, selection of force field, sampling length and initial conformation. Since

only a partially closed barrel-like structure were found from our sampled ensemble. It is emphasized here that the β-barrel motif in our simulation is separated from the location of porous conformation, which matches one case demonstrated by Nguyen *et al.* [407].

6.2.2 Oligomerization of the Aβ42 in the Presence of Lipid Membranes

A previous computational study suggested that the fragment Aβ(25-35) interacts with the model membranes in three stages including adsorption, nucleation on the surface and penetration of the oligomeric form [413]. To examine the molecular mechanisms underlying these processes on the full-length peptide, simulations on different stages could be conducted. First, the helical conformation of Aβ42 (PDBID: 1IYT) is used as the starting conformation. The whole system would be then built build with the helical peptide inserted with its longitudinal axis parallel to the membrane axis. An appropriate lipid bilayer model and the composition of cholestrol would need to be determined before the simulation set-up. Multiple simulations with different cholestrol concentration could be tested with the presence of Aβ42 since membrane polarity could affect the peptide surface behavior [389]. A initial simulation time set to around 3 μs for equilibration would be appropriate. Another monomer would be added into the simulation box at this point on the membrane surface to test for possible dimerization. The formation of trimer is expected in the next step simulation since a previous computational study demonstrated the emergence of porous conformation on the Aβ42 trimer [407].

Appendix A

3D PCA (Chapter 3)

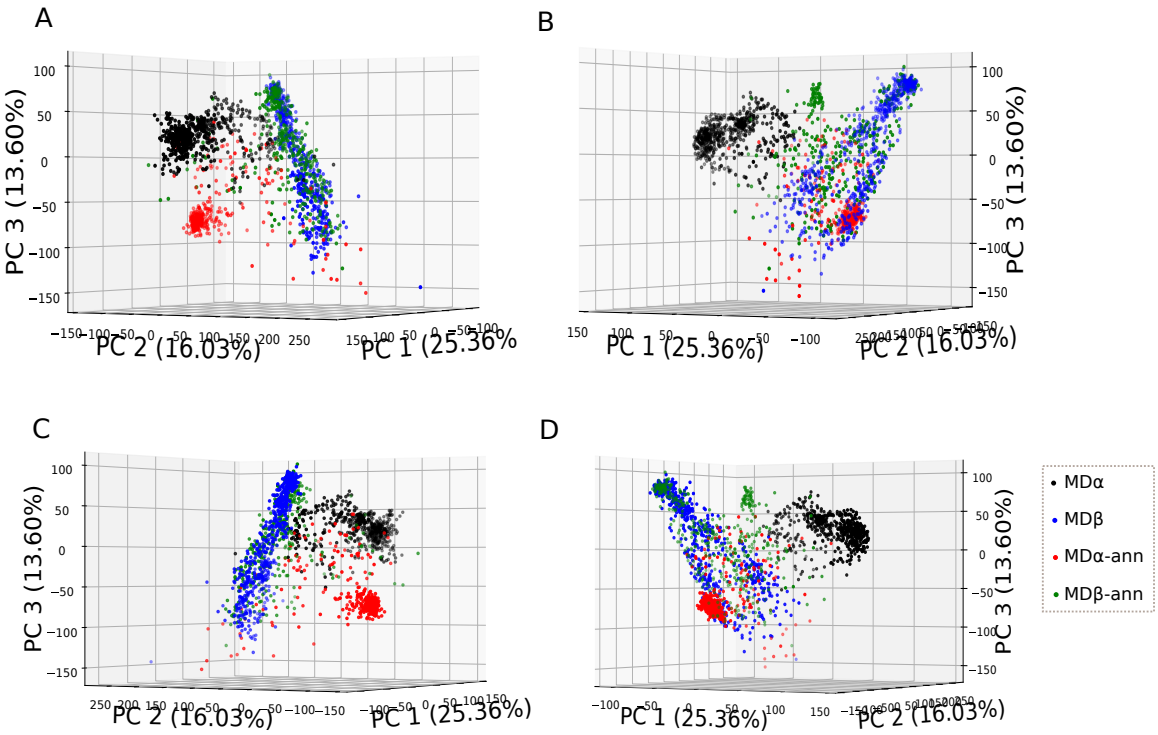

Figure A.1: Projection of all simulations (CT3) in a 3D space formed by the first three principal components. The 3D plot rotates about the z-axis in the couterclockwise by an angle A) 60°, B) 120°, C) 210° and D) 300°.

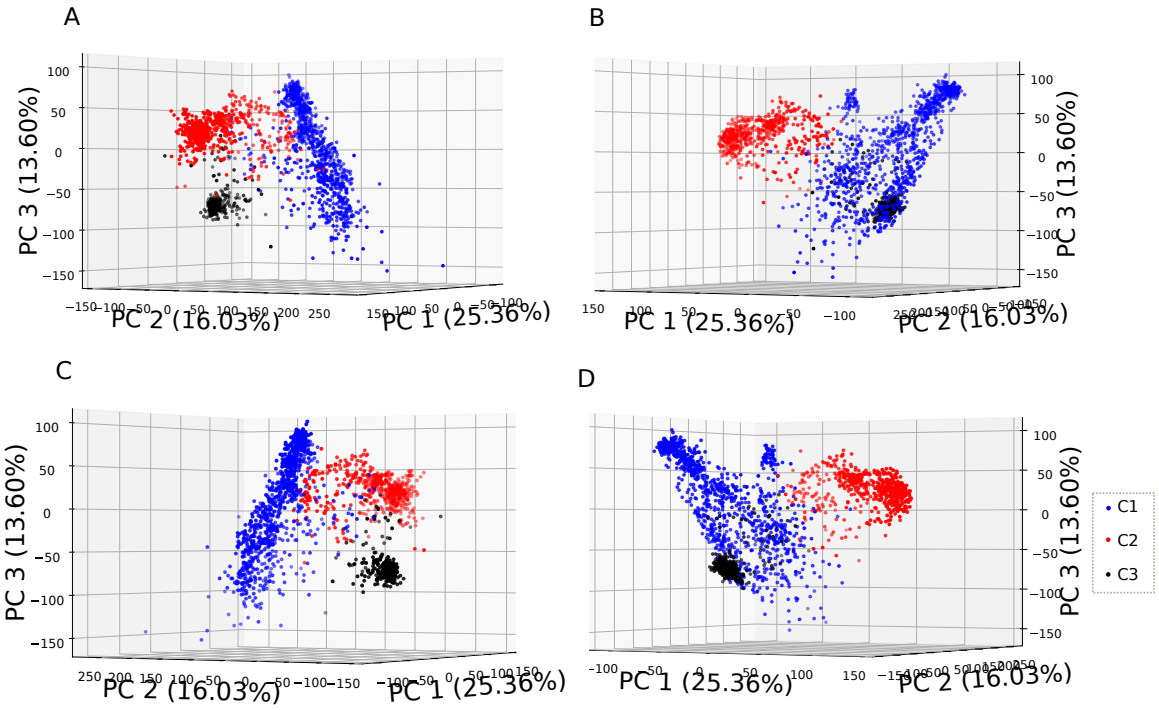

Figure A.2: Same projection as above and the cPCA space is colored by the formation of clusters. The 3D plot rotates about the z-axis in the couterclockwise by an angle A) 60°, B) 120°, C) 210° and D) 300°.

Appendix B

Discretizing along PC2 coordinate based on the cPCA of CT2 and CT3

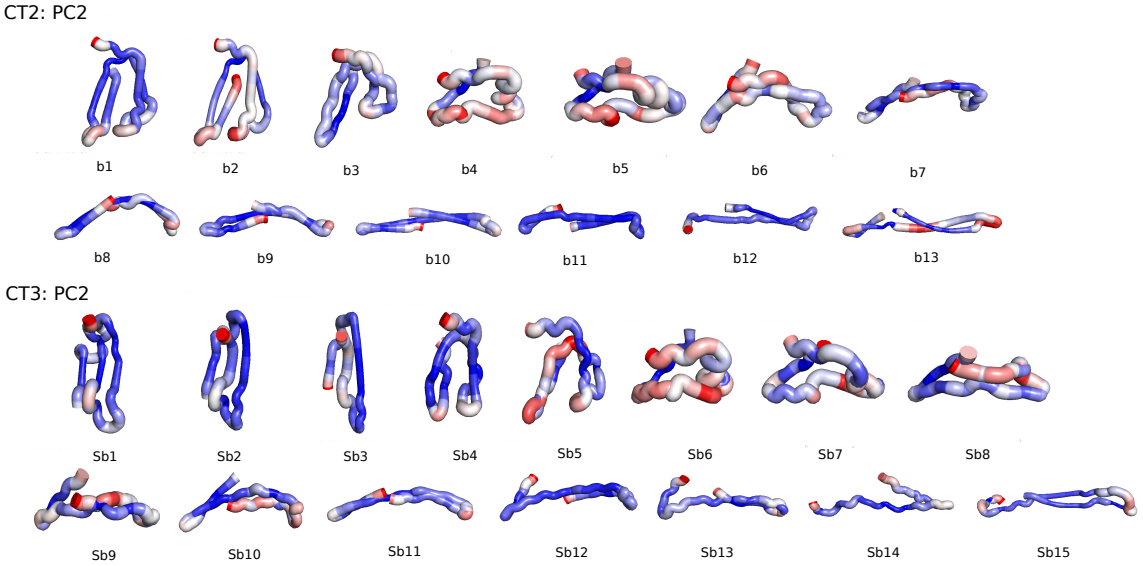

Figure B.1: Two series of average structures are obtained from DPC along PC1 and PC2 based on CT3. The thickness and the color range of red-white-purple of the average structures suggest the extent of RMSD changes among structures in the corresponding binning area.

Appendix C

Clustering on the combined trajectory via R-score

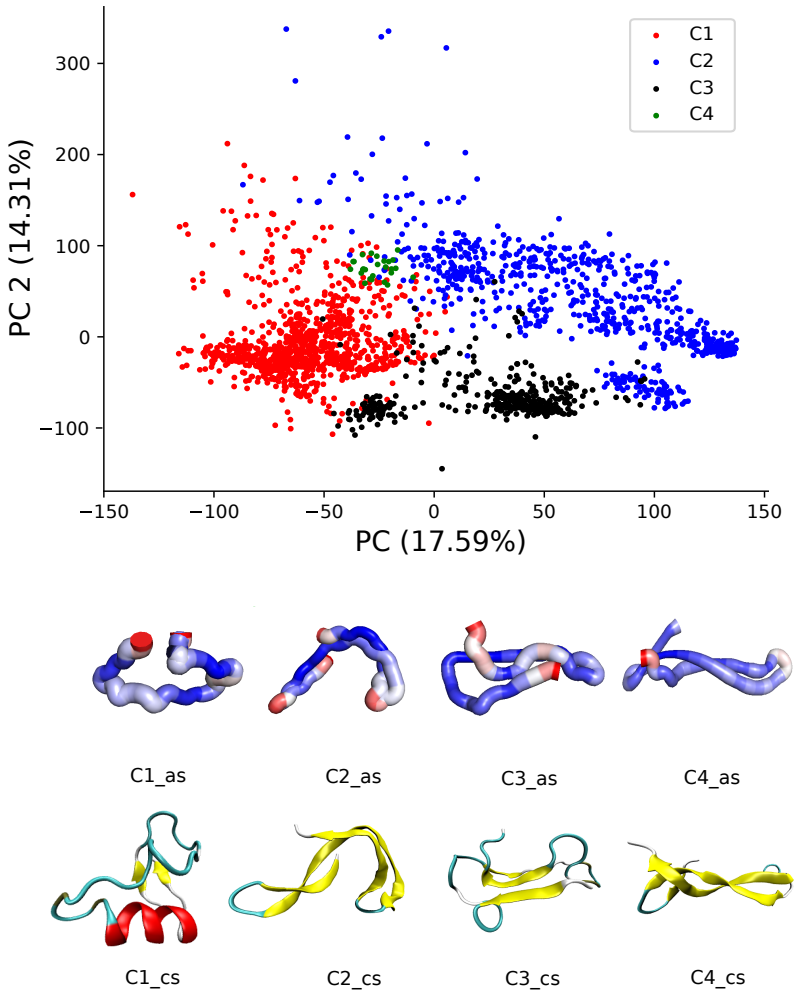

Figure C.1: Community structure detected from the clustering result on the combined trajectory using R-score. Averages structures (`C1_as` to `C4_as`) were generated for the representation of structural characteristics for each cluster. The thickness and the color range of red-white-purple on the average structures suggest the extent of RMSD changes among all structures in the corresponding community. Representive structures that have smallest RMSD to each average are also shown.

Appendix D

Structural changes of Aβ42 in RMSD against simulation time in MD simulations

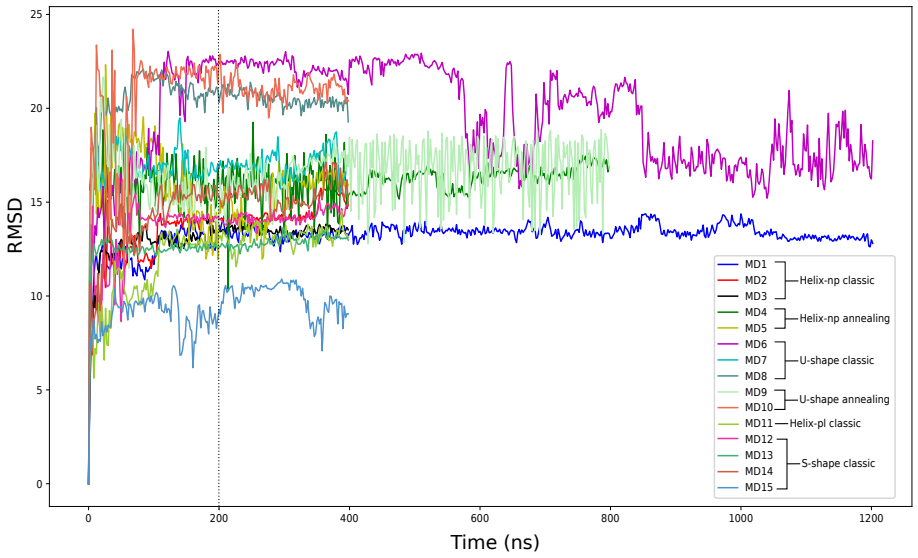

Figure D.1: Evolution of the C_α-RMSD with respect to the corresponding Aβ42 starting structure of the MD simulations.

Appendix E

3D PCA (Chapter 4)

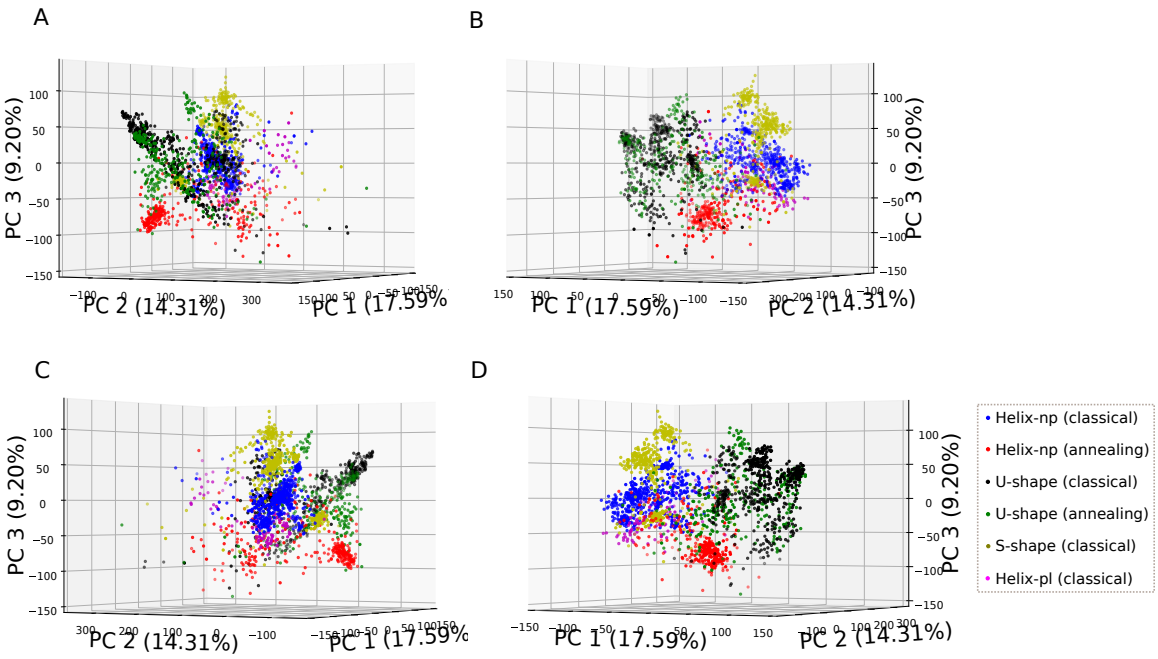

Figure E.1: Projection of the combined trajectory including 15 MD simulation data in a 3D space formed by the first three principal components. The 3D plot rotates about the z-axis in the couterclockwise by an angle A) 60°, B) 120°, C) 210° and D) 300°.

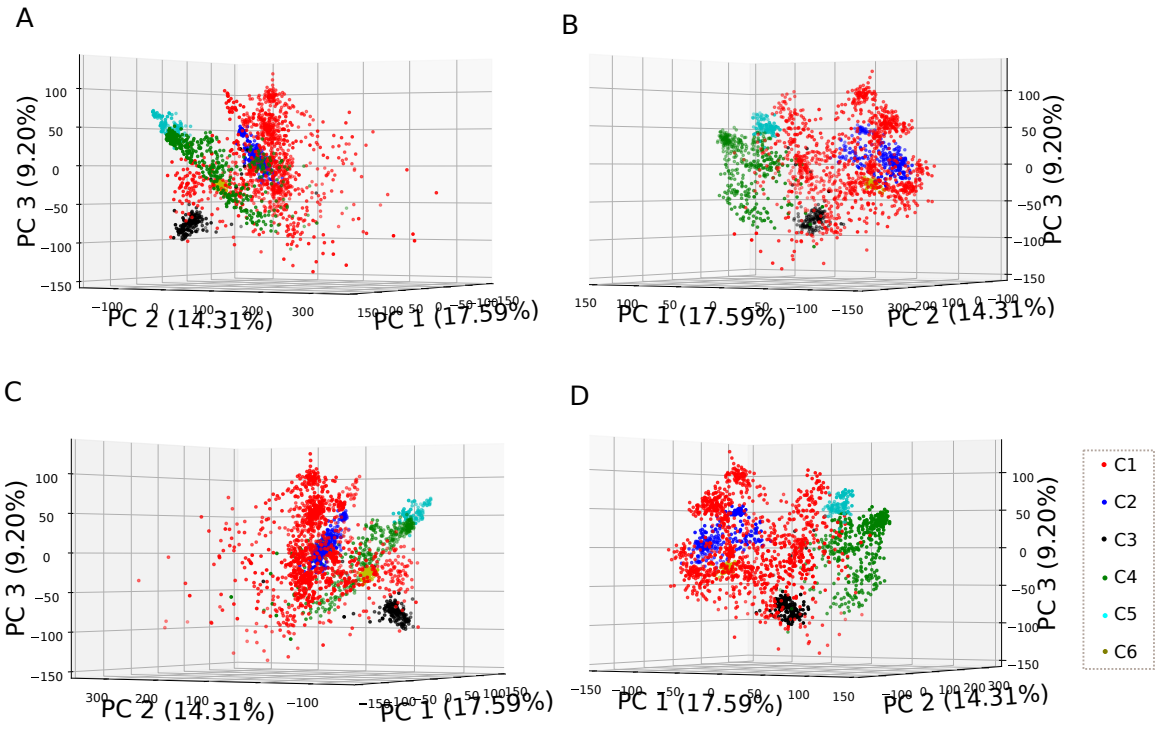

Figure E.2: Same projection as above and the cPCA space is colored by the formation of clusters. The 3D plot rotates about the z-axis in the couterclockwise by an angle A) 60°, B) 120°, C) 210° and D) 300°.

Appendix F

Salt bridges populations in different contact patterns (Chapter 4)

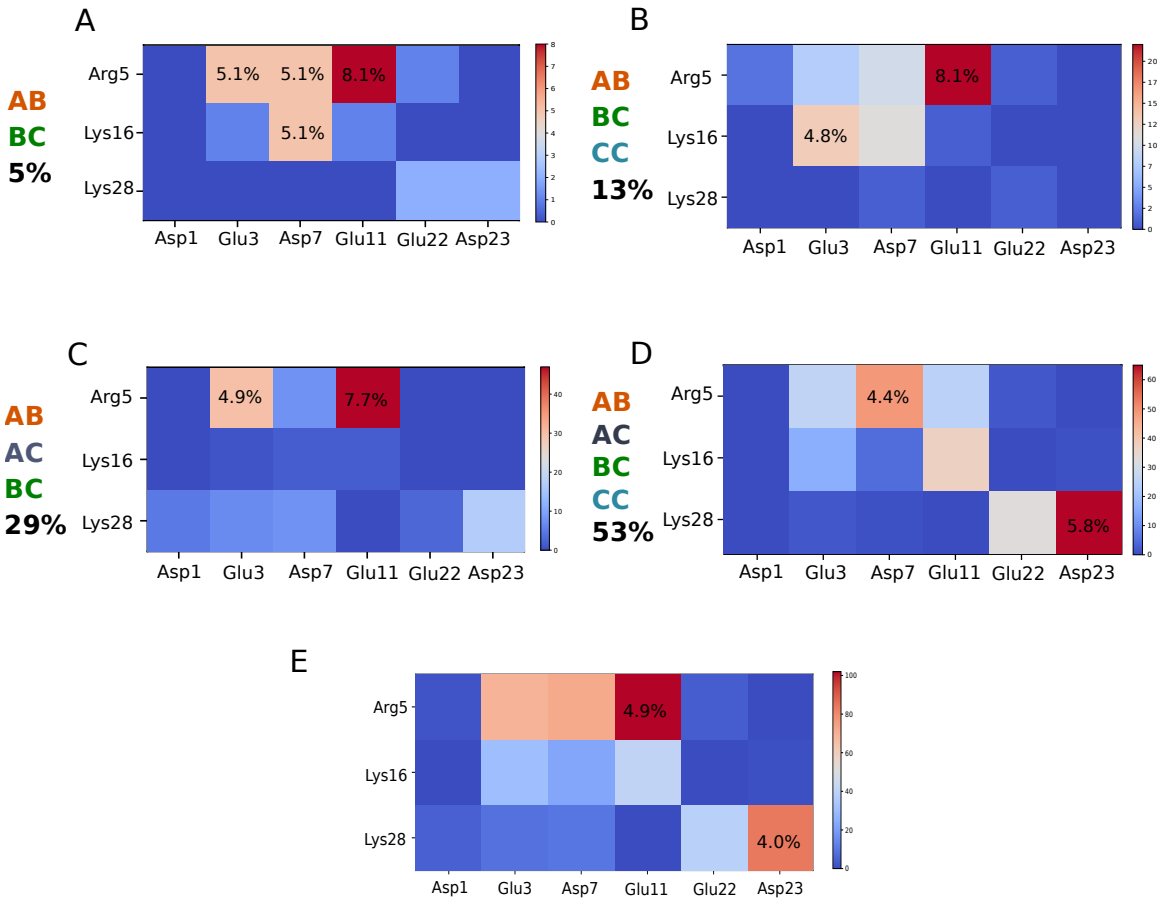

Figure F.1: Populations of all 18 salt bridges formed by 3 positively charged residues and 6 negatively charged residues, calculated for structural ensemble of different contact patterns (A-D) and over all structures (E) using the restrictive definition described by Kumar and Nussinov's paper (see Section 4.2.4 for details). Populations that exceed 4% in different situations are reported.

Bibliography

[1] Peter E Wright and H Jane Dyson. Intrinsically unstructured proteins: reassessing the protein structure-function paradigm. *Journal of Molecular Biology*, 293(2):321–331, 1999.

[2] A Keith Dunker, J David Lawson, Celeste J Brown, Ryan M Williams, Pedro Romero, Jeong S Oh, Christopher J Oldfield, Andrew M Campen, Catherine M Ratliff, Kerry W Hipps, et al. Intrinsically disordered protein. *Journal of Molecular Graphics and Modelling*, 19(1):26–59, 2001.

[3] Vladimir N Uversky. Intrinsically disordered proteins and their mysterious(meta) physics. *Frontiers in Physics*, 7:10, 2019.

[4] Avner Schlessinger, Christian Schaefer, Esmeralda Vicedo, Markus Schmidberger, Marco Punta, and Burkhard Rost. Protein disordera breakthrough invention of evolution? *Current Opinion in Structural Biology*, 21(3):412–418, 2011.

[5] M Madan Babu. The contribution of intrinsically disordered regions to protein function, cellular complexity, and human disease. *Biochemical Society Transactions*, 44(5):1185–1200, 2016.

[6] Matt E Oates, Pedro Romero, Takashi Ishida, Mohamed Ghalwash, Marcin J Mizianty, Bin Xue, Zsuzsanna Dosztányi, Vladimir N Uversky, Zoran Obradovic, Lukasz Kurgan, et al. D2p2: database of disordered protein predictions. *Nucleic Acids Research*, 41(D1):D508–D516, 2013.

[7] Peter E Wright and H Jane Dyson. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell biology*, 16(1):18–29, 2015.

[8] H Jane Dyson and Peter E Wright. Intrinsically unstructured proteins and their functions. *Nature Reviews Molecular Cell biology*, 6(3):197–208, 2005.

[9] Vladimir N Uversky. Intrinsically disordered proteins and their (disordered) proteomes in neurodegenerative disorders. *Frontiers in Aging Neuroscience*, 7:18, 2015.

[10] Mark Benhaim, Kelly K Lee, and Miklos Guttman. Tracking higher order protein structure by hydrogen-deuterium exchange mass spectrometry. *Protein and Peptide Letters*, 26(1):16–26, 2019.

[11] Modesto Orozco. A theoretical view of protein dynamics. *Chemical Society Reviews*, 43(14):5051–5066, 2014.

[12] MJE Sternberg, DEP Grace, and DC Phillips. Dynamic information from protein crystallography: an analysis of temperature factors from refinement of the hen egg-white lysozyme structure. *Journal of Molecular Biology*, 130(3):231–253, 1979.

[13] Bertram Terence Martin Willis and Arthur William Pryor. *Thermal vibrations in crystallography*, volume 50. Cambridge University Press Cambridge, 1975.

[14] Detlef Reichert, Tatiana Zinkevich, Kay Saalwächter, and Alexey Krushelnitsky. The relation of the x-ray b-factor to protein dynamics: insights from recent dynamic solid-state nmr data. *Journal of Biomolecular Structure and Dynamics*, 30(6):617–627, 2012.

[15] Dennis Sherwood and Jon Cooper. *Crystals, X-rays and proteins: comprehensive protein crystallography*. OUP Oxford, 2010.

[16] Robert Konrat. Nmr contributions to structural dynamics studies of intrinsically disordered proteins. *Journal of Magnetic Resonance*, 241:74–85, 2014.

[17] Supriyo Bhattacharya and Xingcheng Lin. Recent advances in computational protocols addressing intrinsically disordered proteins. *Biomolecules*, 9(4):146, 2019.

[18] Santiago Esteban-Martín, Robert Bryn Fenwick, and Xavier Salvatella. Synergistic use of nmr and md simulations to study the structural heterogeneity of proteins. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 2(3):466–478, 2012.

[19] Alexey G Kikhney and Dmitri I Svergun. A practical guide to small angle x-ray scattering (saxs) of flexible and intrinsically disordered proteins. *FEBS Letters*, 589(19PartA):2570–2577, 2015.

[20] Edgar A Hodge, Mark A Benhaim, and Kelly K Lee. Bridging protein structure, dynamics, and function using hydrogen/deuterium-exchange mass spectrometry. *Protein Science*, 2019.

[21] Emeka Nwanochie and Vladimir N Uversky. Structure determination by single-particle cryo-electron microscopy: Only the sky (and intrinsic disorder) is the limit. *International Journal of Molecular Sciences*, 20(17):4186, 2019.

[22] Charles K Fisher and Collin M Stultz. Constructing ensembles for intrinsically disordered proteins. *Current Opinion in Structural Biology*, 21(3):426–431, 2011.

[23] Ashutosh Srivastava, Tetsuro Nagai, Arpita Srivastava, Osamu Miyashita, and Florence Tama. Role of computational methods in going beyond x-ray crystallography to explore protein structure and dynamics. *International Journal of Molecular Sciences*, 19(11):3401, 2018.

[24] Frank Noé. Beating the millisecond barrier in molecular dynamics simulations. *Biophysical Journal*, 108(2):228, 2015.

[25] Kresten Lindorff-Larsen, Paul Maragakis, Stefano Piana, and David E Shaw. Picosecond to millisecond structural dynamics in human ubiquitin. *The Journal of Physical Chemistry B*, 120(33):8313–8320, 2016.

[26] Barry J Grant, Alemayehu A Gorfe, and J Andrew McCammon. Large conformational changes in proteins: signaling and other functions. *Current Opinion in Structural Biology*, 20(2):142–147, 2010.

[27] Vladimir N Uversky. The triple power of d (3): protein intrinsic disorder in degenerative diseases. *Front Bioscience (Landmark Ed)*, 19:181–258, 2014.

[28] Zhihua Du and Vladimir N Uversky. A comprehensive survey of the roles of highly disordered proteins in type 2 diabetes. *International Journal of Molecular Sciences*, 18(10):2010, 2017.

[29] Yugong Cheng, Tanguy LeGall, Christopher J Oldfield, A Keith Dunker, and Vladimir N Uversky. Abundance of intrinsic disorder in protein associated with cardiovascular disease. *Biochemistry*, 45(35):10448–10460, 2006.

[30] Vladimir Vacic, Phineus RL Markwick, Christopher J Oldfield, Xiaoyue Zhao, Chad Haynes, Vladimir N Uversky, and Lilia M Iakoucheva. Disease-associated mutations disrupt functionally important regions of intrinsic protein disorder. *PLoS Computational Biology*, 8(10), 2012.

[31] Uros Midic, Christopher J Oldfield, A Keith Dunker, Zoran Obradovic, and Vladimir N Uversky. Protein disorder in the human diseasome: unfoldomics of human genetic diseases. *Bmc Genomics*, 10(S1):S12, 2009.

[32] Kelty R Baker and Lawrence Rice. The amyloidoses: clinical features, diagnosis and treatment. *Methodist DeBakey Cardiovascular Journal*, 8(3):3, 2012.

[33] Laura M Dember. Amyloidosis-associated kidney disease. *Journal of the American Society of Nephrology*, 17(12):3458–3471, 2006.

[34] Matt E Oates, Pedro Romero, Takashi Ishida, Mohamed Ghalwash, Marcin J Mizianty, Bin Xue, Zsuzsanna Dosztanyi, Vladimir N Uversky, Zoran Obradovic, Lukasz Kurgan, et al. D2p2: database of disordered protein predictions. *Nucleic Acids Research*, 41(D1):D508–D516, 2012.

[35] Lilia M Iakoucheva, Celeste J Brown, J David Lawson, Zoran Obradović, and A Keith Dunker. Intrinsic disorder in cell-signaling and cancer-associated proteins. *Journal of Molecular Biology*, 323(3):573–584, 2002.

[36] Saeed Sadigh-Eteghad, Babak Sabermarouf, Alireza Majdi, Mahnaz Talebi, Mehdi Farhoudi, and Javad Mahmoudi. Amyloid-beta: a crucial factor in alzheimer's disease. *Medical Principles and Practice*, 24(1):1–10, 2015.

[37] Fong Ping Chong, Khuen Yen Ng, Rhun Yian Koh, and Soi Moi Chye. Tau proteins and tauopathies in alzheimers disease. *Cellular and Molecular Neurobiology*, 38(5):965–980, 2018.

[38] Michel Goedert, Ross Jakes, and Maria Grazia Spillantini. The synucleinopathies: twenty years on. *Journal of Parkinson's Disease*, 7(s1):S51–S69, 2017.

[39] Zhiqiang Zheng and Marc I Diamond. Huntington disease and the huntingtin protein. In *Progress in Molecular Biology and Translational Science*, volume 107, pages 189–214. Elsevier, 2012.

[40] John A Hardy and Gerald A Higgins. Alzheimer's disease: the amyloid cascade hypothesis. *Science*, 256(5054):184–186, 1992.

[41] Geert De Meyer, Fred Shapiro, Hugo Vanderstichele, Eugeen Vanmechelen, Sebastiaan Engelborghs, Peter Paul De Deyn, Els Coart, Oskar Hansson, Lennart Minthon, Henrik Zetterberg, et al. Diagnosis-independent alzheimer disease biomarker signature in cognitively normal elderly people. *Archives of Neurology*, 67(8):949–956, 2010.

[42] Eric Karran, Marc Mercken, and Bart De Strooper. The amyloid cascade hypothesis for alzheimer's disease: an appraisal for the development of therapeutics. *Nature Reviews Drug Discovery*, 10(9):698–712, 2011.

[43] Pei-Pei Liu, Yi Xie, Xiao-Yan Meng, and Jian-Sheng Kang. History and progress of hypotheses and clinical trials for alzheimers disease. *Signal Transduction and Targeted Therapy*, 4(1):1–22, 2019.

[44] E Hubin, NAJ Van Nuland, K Broersen, and K Pauwels. Transient dynamics of aβ contribute to toxicity in alzheimers disease. *Cellular and Molecular Life Sciences*, 71(18):3507–3521, 2014.

[45] Daniela Puzzo and Ottavio Arancio. Amyloid-β peptide: Dr. jekyll or mr. hyde? *Journal of Alzheimer's Disease*, 33(s1):S111–S120, 2013.

[46] Gary P Morris, Ian A Clark, and Bryce Vissel. Inconsistencies and controversies surrounding the amyloid hypothesis of alzheimer's disease. *Acta Neuropathologica Communications*, 2(1):135, 2014.

[47] Roberta Ricciarelli and Ernesto Fedele. The amyloid cascade hypothesis in alzheimer's disease: it's time to change our mind. *Current Neuropharmacology*, 15(6):926–935, 2017.

[48] Dennis J Selkoe. Resolving controversies on the path to alzheimer's therapeutics. *Nature Medicine*, 17(9):1060, 2011.

[49] Li-Kai Huang, Shu-Ping Chao, and Chaur-Jong Hu. Clinical trials of new drugs for alzheimer disease. *Journal of Biomedical Science*, 27(1):1–13, 2020.

[50] Holly M Brothers, Maya L Gosztyla, and Stephen R Robinson. The physiological roles of amyloid-β peptide hint at new ways to treat alzheimer's disease. *Frontiers in Aging Neuroscience*, 10:118, 2018.

[51] Maya L Gosztyla, Holly M Brothers, and Stephen R Robinson. Alzheimers amyloid-β is an antimicrobial peptide: a review of the evidence. *Journal of Alzheimer's Disease*, 62(4):1495–1506, 2018.

[52] Sergio Bagnato, Maria Andriolo, Cristina Boccagni, and Giuseppe Galardi. Prolonged changes in amyloid-β metabolism after a severe traumatic brain injury. *Neuroreport*, 28(5):250–252, 2017.

[53] Glenda M Bishop and Stephen R Robinson. The amyloid hypothesis: let sleeping dogmas lie? *Neurobiology of Aging*, 23(6):1101–1105, 2002.

[54] John E Morley, SA Farr, AD Nguyen, and F Xu. What is the physiological function of amyloid-beta protein?, 2019.

[55] Dennis J Selkoe. Translating cell biology into therapeutic advances in alzheimer's disease. *Nature*, 399(6738):A23–A31, 1999.

[56] Frederick Hsu, Giovanna Park, and Zhefeng Guo. Key residues for the formation of aβ42 amyloid fibrils. *ACS Omega*, 3(7):8401–8407, 2018.

[57] Arvi Rauk. The chemistry of alzheimers disease. *Chemical Society Reviews*, 38(9):2698–2715, 2009.

[58] Lothar Gremer, Daniel Schölzel, Carla Schenk, Elke Reinartz, Jörg Labahn, Raimond BG Ravelli, Markus Tusche, Carmen Lopez-Iglesias, Wolfgang Hoyer, Henrike Heise, et al. Fibril structure of amyloid-β (1–42) by cryo–electron microscopy. *Science*, 358(6359):116–119, 2017.

[59] Ken A Dill and Justin L MacCallum. The protein-folding problem, 50 years on. *Science*, 338(6110):1042–1046, 2012.

[60] Aaron Ciechanover and Yong Tae Kwon. Protein quality control by molecular chaperones in neurodegeneration. *Frontiers in Neuroscience*, 11:185, 2017.

[61] Ingo Amm, Thomas Sommer, and Dieter H Wolf. Protein quality control and elimination of protein waste: the role of the ubiquitin–proteasome system. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research*, 1843(1):182–196, 2014.

[62] Massimo Stefani. Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1739(1):5–25, 2004.

[63] Samuel IA Cohen, Michele Vendruscolo, Christopher M Dobson, and Tuomas PJ Knowles. Nucleated polymerization with secondary pathways. iii. equilibrium behavior and oligomer populations. *The Journal of Chemical Physics*, 135(6):08B612, 2011.

[64] Eri Chatani and Naoki Yamamoto. Recent progress on understanding the mechanisms of amyloid nucleation. *Biophysical Reviews*, 10(2):527–534, 2018.

[65] John S Schreck and Jian-Min Yuan. A kinetic study of amyloid formation: fibril growth and length distributions. *The Journal of Physical Chemistry B*, 117(21):6574–6583, 2013.

[66] Feng Ding, Jose M Borreguero, Sergey V Buldyrey, H Eugene Stanley, and Nikolay V Dokholyan. Mechanism for the α-helix to β-hairpin transition. *Proteins: Structure, Function, and Bioinformatics*, 53(2):220–228, 2003.

[67] Sara Linse. Mechanism of amyloid protein aggregation and the role of inhibitors. *Pure and Applied Chemistry*, 91(2):211–229, 2019.

[68] Sara Linse. Monomer-dependent secondary nucleation in amyloid formation. *Biophysical Reviews*, 9(4):329–338, 2017.

[69] Chuang-Chung Lee, Arpan Nayak, Ananthakrishnan Sethuraman, Georges Belfort, and Gregory J McRae. A three-stage kinetic model of amyloid fibrillation. *Biophysical Journal*, 92(10):3448–3458, 2007.

[70] Fabrizio Chiti and Christopher M Dobson. Protein misfolding, amyloid formation, and human disease: a summary of progress over the last decade. *Annual Review of Biochemistry*, 86:27–68, 2017.

[71] Céline Galvagnion, Alexander K Buell, Georg Meisl, Thomas CT Michaels, Michele Vendruscolo, Tuomas PJ Knowles, and Christopher M Dobson. Lipid vesicles trigger α-synuclein aggregation by stimulating primary nucleation. *Nature Chemical Biology*, 11(3):229, 2015.

[72] Robert Vacha, Sara Linse, and Mikael Lund. Surface effects on aggregation kinetics of amyloidogenic peptides. *Journal of the American Chemical Society*, 136(33):11776–11782, 2014.

[73] Silvia Campioni, Guillaume Carret, Sophia Jordens, Lucrece Nicoud, Raffaele Mezzenga, and Roland Riek. The presence of an air–water interface affects formation and elongation of α-synuclein fibrils. *Journal of the American Chemical Society*, 136(7):2866–2875, 2014.

[74] Sofia Giorgetti, Claudio Greco, Paolo Tortora, and Francesco Antonio Aprile. Targeting amyloid aggregation: an overview of strategies and mechanisms. *International Journal of Molecular Sciences*, 19(9):2677, 2018.

[75] Tuomas PJ Knowles, Christopher A Waudby, Glyn L Devlin, Samuel IA Cohen, Adriano Aguzzi, Michele Vendruscolo, Eugene M Terentjev, Mark E Welland, and Christopher M Dobson. An analytical solution to the kinetics of breakable filament assembly. *Science*, 326(5959):1533–1537, 2009.

[76] Esther Stroo, Mandy Koopman, Ellen AA Nollen, and Alejandro Mata-Cabana. Cellular regulation of amyloid formation in aging and disease. *Frontiers in Neuroscience*, 11:64, 2017.

[77] Rosa Crespo, Eva Villar-Alvarez, Pablo Taboada, Fernando A Rocha, Ana M Damas, and Pedro M Martins. What can the kinetics of amyloid fibril formation tell about off-pathway aggregation? *Journal of Biological Chemistry*, 291(4):2018–2032, 2016.

[78] Evan T Powers and David L Powers. Mechanisms of protein fibril formation: nucleated polymerization with competing off-pathway aggregation. *Biophysical Journal*, 94(2):379–391, 2008.

[79] Bruce L Kagan, Hyunbum Jang, Ricardo Capone, Fernando Teran Arce, Srinivasan Ramachandran, Ratnesh Lal, and Ruth Nussinov. Antimicrobial properties of amyloid peptides. *Molecular Pharmaceutics*, 9(4):708–717, 2012.

[80] Murray Coles, Wendy Bicknell, Andrew A Watson, David P Fairlie, and David J Craik. Solution structure of amyloid β-peptide (1- 40) in a water- micelle environment. is the membrane-spanning domain where we think it is? *Biochemistry*, 37(31):11064–11077, 1998.

[81] Orlando Crescenzi, Simona Tomaselli, Remo Guerrini, Severo Salvadori, Anna M D'Ursi, Piero Andrea Temussi, and Delia Picone. Solution structure of the alzheimer amyloid β-peptide (1–42) in an apolar microenvironment: Similarity with a virus fusion domain. *European Journal of Biochemistry*, 269(22):5642–5648, 2002.

[82] Subramanian Vivekanandan, Jeffrey R Brender, Shirley Y Lee, and Ayyalusamy Ramamoorthy. A partially folded structure of amyloid-beta (1–40) in an aqueous environment. *Biochemical and Biophysical Research Communications*, 411(2):312–316, 2011.

[83] Simona Tomaselli, Veronica Esposito, Paolo Vangone, Nico AJ van Nuland, Alexandre MJJ Bonvin, Remo Guerrini, Teodorico Tancredi, Piero A Temussi, and Delia Picone. The α-to-β conformational transition of alzheimer's aβ-(1–42) peptide in aqueous media is reversible: a step by step conformational analysis suggests the location of β conformation seeding. *ChemBioChem*, 7(2):257–267, 2006.

[84] Rakez Kayed and Cristian A Lasagna-Reeves. Molecular mechanisms of amyloid oligomers toxicity. *Journal of Alzheimer's Disease*, 33(s1):S67–S78, 2013.

[85] Masafumi Sakono and Tamotsu Zako. Amyloid oligomers: formation and toxicity of aβ oligomers. *The FEBS Journal*, 277(6):1348–1358, 2010.

[86] Fabrizio Chiti and Christopher M Dobson. Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, 75:333–366, 2006.

[87] Yan He, Mei-Mei Zheng, Yan Ma, Xiao-Juan Han, Xue-Qiang Ma, Chuan-Qiang Qu, and Yi-Feng Du. Soluble oligomers and fibrillar species of amyloid β-peptide differentially affect cognitive functions and hippocampal inflammatory response. *Biochemical and Biophysical Research Communications*, 429(3-4):125–130, 2012.

[88] Summer L Bernstein, Nicholas F Dupuis, Noel D Lazo, Thomas Wyttenbach, Margaret M Condron, Gal Bitan, David B Teplow, Joan-Emma Shea, Brandon T Ruotolo, Carol V Robinson, et al. Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of alzheimer's disease. *Nature Chemistry*, 1(4):326, 2009.

[89] Mahiuddin Ahmed, Judianne Davis, Darryl Aucoin, Takeshi Sato, Shivani Ahuja, Saburo Aimoto, James I Elliott, William E Van Nostrand, and Steven O Smith. Structural conversion of neurotoxic amyloid-β 1–42 oligomers to fibrils. *Nature Structural and Molecular Biology*, 17(5):561, 2010.

[90] Gal Bitan, Sabrina S Vollers, and David B Teplow. Elucidation of primary structure elements controlling early amyloid β-protein oligomerization. *Journal of Biological Chemistry*, 278(37):34882–34889, 2003.

[91] William M Tay, Danting Huang, Terrone L Rosenberry, and Anant K Paravastu. The alzheimer's amyloid-β (1–42) peptide forms off-pathway oligomers and fibrils that are distinguished structurally by intermolecular organization. *Journal of Molecular Biology*, 425(14):2494–2508, 2013.

[92] Liping Yu, Rohinton Edalji, John E Harlan, Thomas F Holzman, Ana Pereda Lopez, Boris Labkovsky, Heinz Hillen, Stefan Barghorn, Ulrich Ebert, Paul L Richardson, et al. Structural characterization of a soluble amyloid β-peptide oligomer. *Biochemistry*, 48(9):1870–1877, 2009.

[93] Thorsten Lührs, Christiane Ritter, Marc Adrian, Dominique Riek-Loher, Bernd Bohrmann, Heinz Döbeli, David Schubert, and Roland Riek. 3d structure of alzheimer's amyloid-β (1–42) fibrils. *Proceedings of the National Academy of Sciences*, 102(48):17342–17347, 2005.

[94] Johnny D Pham, Nicholas Chim, Celia W Goulding, and James S Nowick. Structures of oligomers of a peptide from β-amyloid. *Journal of the American Chemical Society*, 135(33):12460–12467, 2013.

[95] Ryan K Spencer, Hao Li, and James S Nowick. X-ray crystallographic structures of trimers and higher-order oligomeric assemblies of a peptide derived from aβ17–36. *Journal of the American Chemical Society*, 136(15):5595–5598, 2014.

[96] William Humphrey, Andrew Dalke, and Klaus Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[97] Robin Roychaudhuri, Mingfeng Yang, Minako M Hoshi, and David B Teplow. Amyloid β-protein assembly and alzheimer disease. *Journal of Biological Chemistry*, 284(8):4749–4753, 2009.

[98] Holger A Scheidt, Isabel Morgado, Sven Rothemund, and Daniel Huster. Dynamics of amyloid β fibrils revealed by solid-state nmr. *Journal of Biological Chemistry*, 287(3):2017–2021, 2012.

[99] Matthias Schmidt, Alexis Rohou, Keren Lasker, Jay K Yadav, Cordelia Schiene-Fischer, Marcus Fändrich, and Nikolaus Grigorieff. Peptide dimer structure in an aβ (1–42) fibril visualized with cryo-em. *Proceedings of the National Academy of Sciences*, 112(38):11858–11863, 2015.

[100] Valerie L Anderson and Watt W Webb. Transmission electron microscopy characterization of fluorescently labelled amyloid β 1-40 and α-synuclein aggregates. *BMC Biotechnology*, 11(1):125, 2011.

[101] Ashok Parbhu, Hai Lin, Julian Thimm, and Ratneshwar Lal. Imaging real-time aggregation of amyloid beta protein (1–42) by atomic force microscopy. *Peptides*, 23(7):1265–1270, 2002.

[102] Workalemahu M Berhanu and Ulrich HE Hansmann. Structure and dynamics of amyloid-β segmental polymorphisms. *PLoS One*, 7(7), 2012.

[103] Hideyo Inouye, Paul E Fraser, and Daniel A Kirschner. Structure of beta-crystallite assemblies formed by alzheimer beta-amyloid protein analogues: analysis by x-ray diffraction. *Biophysical Journal*, 64(2):502, 1993.

[104] Margaret Sunde, Louise C Serpell, Mark Bartlam, Paul E Fraser, Mark B Pepys, and Colin CF Blake. Common core structure of amyloid fibrils by synchrotron x-ray diffraction. *Journal of Molecular Biology*, 273(3):729–739, 1997.

[105] Daniel A Kirschner, Carmela Abraham, and Dennis J Selkoe. X-ray diffraction from intraneuronal paired helical filaments and extraneuronal amyloid fibers in alzheimer disease indicates cross-beta conformation. *Proceedings of the National Academy of Sciences*, 83(2):503–507, 1986.

[106] Aneta T Petkova, Yoshitaka Ishii, John J Balbach, Oleg N Antzutkin, Richard D Leapman, Frank Delaglio, and Robert Tycko. A structural model for alzheimer's β-amyloid fibrils based on experimental constraints from solid state nmr. *Proceedings of the National Academy of Sciences*, 99(26):16742–16747, 2002.

[107] Anant K Paravastu, Richard D Leapman, Wai-Ming Yau, and Robert Tycko. Molecular structural basis for polymorphism in alzheimer's β-amyloid fibrils. *Proceedings of the National Academy of Sciences*, 105(47):18349–18354, 2008.

[108] Michael T Colvin, Robert Silvers, Qing Zhe Ni, Thach V Can, Ivan Sergeyev, Melanie Rosay, Kevin J Donovan, Brian Michael, Joseph Wall, Sara Linse, et al. Atomic resolution structure of monomorphic aβ42 amyloid fibrils. *Journal of the American Chemical Society*, 138(30):9663–9674, 2016.

[109] Marielle Aulikki Wälti, Francesco Ravotti, Hiromi Arai, Charles G Glabe, Joseph S Wall, Anja Böckmann, Peter Güntert, Beat H Meier, and Roland Riek. Atomic-resolution structure of a disease-relevant aβ (1–42) amyloid fibril. *Proceedings of the National Academy of Sciences*, 113(34):E4976–E4984, 2016.

[110] Yiling Xiao, Buyong Ma, Dan McElheny, Sudhakar Parthasarathy, Fei Long, Minako Hoshi, Ruth Nussinov, and Yoshitaka Ishii. Aβ (1–42) fibril structure illuminates self-recognition and replication of amyloid in alzheimer's disease. *Nature Structural and Molecular Biology*, 22(6):499–505, 2015.

[111] Jun-Xia Lu, Wei Qiang, Wai-Ming Yau, Charles D Schwieters, Stephen C Meredith, and Robert Tycko. Molecular structure of β-amyloid fibrils in alzheimers disease brain tissue. *Cell*, 154(6):1257–1268, 2013.

[112] Marius Kollmer, William Close, Leonie Funk, Jay Rasmussen, Aref Bsoul, Angelika Schierhorn, Matthias Schmidt, Christina J Sigurdson, Mathias Jucker, and Marcus Fändrich. Cryo-em structure and polymorphism of aβ amyloid fibrils purified from alzheimers brain tissue. *Nature Communications*, 10(1):1–8, 2019.

[113] Bogdan Barz and Brigita Urbanc. Dimer formation enhances structural differences between amyloid β-protein (1–40) and (1–42): an explicit-solvent molecular dynamics study. *PloS One*, 7(4), 2012.

[114] Michael C Owen, David Gnutt, Mimi Gao, Sebastian KTS Wärmländer, Jüri Jarvet, Astrid Gräslund, Roland Winter, Simon Ebbinghaus, and Birgit Strodel. Effects of in vivo conditions on amyloid aggregation. *Chemical Society Reviews*, 48(14):3946–3996, 2019.

[115] Martin Wolff, Dmitry Unuchek, Bo Zhang, Valentin Gordeliy, Dieter Willbold, and Luitgard Nagel-Steger. Amyloid β oligomeric species present in the lag phase of amyloid formation. *PloS One*, 10(5), 2015.

[116] David J Rosenman, Chunyu Wang, and Angel E García. Characterization of aβ monomers through the convergence of ensemble properties among simulations with multiple force fields. *The Journal of Physical Chemistry B*, 120(2):259–277, 2016.

[117] Martín Carballo-Pacheco and Birgit Strodel. Comparison of force fields for alzheimer's aβ42: A case study for intrinsically disordered proteins. *Protein Science*, 26(2):174–185, 2017.

[118] Jing Huang and Alexander D MacKerell Jr. Force field development and simulations of intrinsically disordered proteins. *Current Opinion in Structural Biology*, 48:40–48, 2018.

[119] David K Brown, David L Penkler, Olivier Sheik Amamuddy, Caroline Ross, Ali Rana Atilgan, Canan Atilgan, and Özlem Tastan Bishop. Md-task: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics*, 33(17):2768–2771, 2017.

[120] Lars Skjærven, Shashank Jariwala, Xin-Qiu Yao, and Barry J Grant. Online interactive analysis of protein structure ensembles with bio3d-web. *Bioinformatics*, 32(22):3510–3512, 2016.

[121] Robert T McGibbon, Kyle A Beauchamp, Matthew P Harrigan, Christoph Klein, Jason M Swails, Carlos X Hernández, Christian R Schwantes, Lee-Ping Wang, Thomas J Lane, and Vijay S Pande. Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophysical Journal*, 109(8):1528–1532, 2015.

[122] Fernando Pérez and Brian E Granger. Ipython: a system for interactive scientific computing. *Computing in Science and Engineering*, 9(3):21–29, 2007.

[123] Andrew E Torda and Wilfred F van Gunsteren. Algorithms for clustering molecular dynamics configurations. *Journal of Computational Chemistry*, 15(12):1331–1340, 1994.

[124] Gareth Aneurin Tribello and Piero Gasparotto. Using dimensionality reduction to analyze protein trajectories. *Frontiers in Molecular Biosciences*, 6:46, 2019.

[125] Broto Chakrabarty, Varun Naganathan, Kanak Garg, Yash Agarwal, and Nita Parekh. Naps update: network analysis of molecular dynamics data and protein–nucleic acid complexes. *Nucleic Acids Research*, 47(W1):W462–W470, 2019.

[126] Razieh Karamzadeh, Mohammad Hossein Karimi-Jafari, Ali Sharifi-Zarchi, Hamidreza Chitsaz, Ghasem Hosseini Salekdeh, and Ali Akbar Moosavi-Movahedi. Machine learning and network analysis of molecular dynamics trajectories reveal two chains of red/ox-specific residue interactions in human protein disulfide isomerase. *Scientific Reports*, 7(1):1–11, 2017.

[127] Thomas J Lane, Diwakar Shukla, Kyle A Beauchamp, and Vijay S Pande. To milliseconds and beyond: challenges in the simulation of protein folding. *Current Opinion in Structural Biology*, 23(1):58–65, 2013.

[128] Son Tung Ngo, Duc Toan Truong, Nguyen Minh Tam, and Minh Tho Nguyen. Egcg inhibits the oligomerization of amyloid beta (16-22) hexamer: theoretical studies. *Journal of Molecular Graphics and Modelling*, 76:1–10, 2017.

[129] Banafsheh Mehrazma and Arvi Rauk. Exploring amyloid-β dimer structure using molecular dynamics simulations. *The Journal of Physical Chemistry A*, 123(22):4658–4670, 2019.

[130] Justin A Lemkul and David R Bevan. Aggregation of alzheimers amyloid β-peptide in biological membranes: A molecular dynamics study. *Biochemistry*, 52(29):4971–4980, 2013.

[131] Son Tung Ngo, Huynh Minh Hung, Duc Toan Truong, and Minh Tho Nguyen. Replica exchange molecular dynamics study of the truncated amyloid beta (11–40) trimer in solution. *Physical Chemistry Chemical Physics*, 19(3):1909–1919, 2017.

[132] Anne M Brown and David R Bevan. Molecular dynamics simulations of amyloid β-peptide (1-42): Tetramer formation and membrane interactions. *Biophysical Journal*, 111(5):937–949, 2016.

[133] Hisashi Okumura and Satoru G Itoh. Structural and fluctuational difference between two ends of aβ amyloid fibril: Md simulations predict only one end has open conformations. *Scientific Reports*, 6:38422, 2016.

[134] Mingyan Dong, Wei Zhao, Dingkun Hu, Hongqi Ai, and Baotao Kang. N-terminus binding preference for either tanshinone or analogue in both inhibition of amyloid aggregation and disaggregation of preformed amyloid fibrils toward introducing a kind of novel anti-alzheimer compounds. *ACS Chemical Neuroscience*, 8(7):1577–1588, 2017.

[135] Wenhui Xi, Wenhua Wang, Gabrielle Abbott, and Ulrich HE Hansmann. Stability of a recently found triple-β-stranded aβ1–42 fibril motif. *The Journal of Physical Chemistry B*, 120(20):4548–4557, 2016.

[136] Philippe Derreumaux. *Alzheimer's disease: insights into low molecular weight and cytotoxic aggregates from in vitro and computer experiments: molecular basis of amyloid-beta protein aggregation and fibril formation*, volume 7. World Scientific, 2013.

[137] Edgar Luttmann and Gregor Fels. All-atom molecular dynamics studies of the full-length β-amyloid peptides. *Chemical Physics*, 323(1):138–147, 2006.

[138] Yechun Xu, Jianhua Shen, Xiaomin Luo, Weiliang Zhu, Kaixian Chen, Jianpeng Ma, and Hualiang Jiang. Conformational transition of amyloid β-peptide. *Proceedings of the National Academy of Sciences*, 102(15):5403–5407, 2005.

[139] Dagmar Flöck, Stefano Colacino, Giorgio Colombo, and Alfredo Di Nola. Misfolding of the amyloid β-protein: A molecular dynamics study. *Proteins: Structure, Function, and Bioinformatics*, 62(1):183–192, 2006.

[140] Cao Yang, Jinyu Li, Yu Li, and Xiaolei Zhu. The effect of solvents on the conformations of amyloid β-peptide (1–42) studied by molecular dynamics simulation. *Journal of Molecular Structure: Theorem*, 895(1-3):1–8, 2009.

[141] Mariacristina Valerio, Alfredo Colosimo, Filippo Conti, Alessandro Giuliani, Alessandro Grottesi, Cesare Manetti, and Joseph P Zbilut. Early events in protein aggregation: molecular flexibility and hydrophobicity/charge interaction in amyloid peptides as studied by molecular dynamics simulations. *Proteins: Structure, Function, and Bioinformatics*, 58(1):110–118, 2005.

[142] Olujide O Olubiyi and Birgit Strodel. Structures of the amyloid β-peptides aβ1–40 and aβ1–42 as influenced by ph and a d-peptide. *The Journal of Physical Chemistry B*, 116(10):3280–3291, 2012.

[143] Naoyuki Miyashita, John E Straub, and D Thirumalai. Structures of β-amyloid peptide 1- 40, 1- 42, and 1- 55- the 672- 726 fragment of app in a membrane environment with implications for interactions with γ-secretase. *Journal of the American Chemical Society*, 131(49):17843–17852, 2009.

[144] Nikolaos G Sgourakis, Myrna Merced-Serrano, Christos Boutsidis, Petros Drineas, Zheming Du, Chunyu Wang, and Angel E Garcia. Atomic-level characterization of the ensemble of the aβ (1–42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms. *Journal of Molecular Biology*, 405(2):570–583, 2011.

[145] Wanling Song, Yuanyuan Wang, Jacques-Philippe Colletier, Huaiyu Yang, and Yechun Xu. Varied probability of staying collapsed/extended at the conformational equilibrium of monomeric aβ 40 and aβ 42. *Scientific Reports*, 5:11024, 2015.

[146] Satoru G Itoh, Maho Yagi-Utsumi, Koichi Kato, and Hisashi Okumura. Effects of a hydrophilic/hydrophobic interface on amyloid-β peptides studied by molecular dynamics simulations and nmr experiments. *The Journal of Physical Chemistry B*, 123(1):160–169, 2018.

[147] Linh Tran and Tâp Ha-Duong. Exploring the alzheimer amyloid-β peptide conformational ensemble: A review of molecular dynamics approaches. *Peptides*, 69:86–91, 2015.

[148] Erwin Schrödinger. An undulatory theory of the mechanics of atoms and molecules. *Physical Review*, 28(6):1049, 1926.

[149] Robert G Parr. Density functional theory of atoms and molecules. In *Horizons of quantum chemistry*, pages 5–15. Springer, 1980.

[150] Attila Szabo and Neil S Ostlund. *Modern quantum chemistry: introduction to advanced electronic structure theory*. Courier Corporation, 2012.

[151] Trygve Helgaker, Poul Jorgensen, and Jeppe Olsen. *Molecular electronic-structure theory*. John Wiley & Sons, 2014.

[152] Wolfram Koch and Max C Holthausen. *A chemist's guide to density functional theory*. John Wiley & Sons, 2015.

[153] Alexander D MacKerell Jr. Empirical force fields for biological macromolecules: overview and issues. *Journal of Computational Chemistry*, 25(13):1584–1604, 2004.

[154] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules j. am. chem. soc. 1995, 117, 5179-5197. *Journal of the American Chemical Society*, 118(9):2309–2309, 1996.

[155] Arieh Warshel, Mitsunori Kato, and Andrei V Pisliakov. Polarizable force fields: history, test cases, and prospects. *Journal of Chemical Theory and Computation*, 3(6):2034–2045, 2007.

[156] Christopher M Baker. Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 5(2):241–254, 2015.

[157] Justin A Lemkul, Jing Huang, Benoît Roux, and Alexander D MacKerell Jr. An empirical polarizable force field based on the classical drude oscillator model: development history and recent applications. *Chemical Reviews*, 116(9):4983–5013, 2016.

[158] Alexander P Lyubartsev and Alexander L Rabinovich. Force field development for lipid membrane simulations. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1858(10):2483–2497, 2016.

[159] Bernard R Brooks, Robert E Bruccoleri, Barry D Olafson, David J States, S a Swaminathan, and Martin Karplus. Charmm: a program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry*, 4(2):187–217, 1983.

[160] Alexander D MacKerell Jr, Joanna Wiorkiewicz-Kuczera, and Martin Karplus. An all-atom empirical energy function for the simulation of nucleic acids. *Journal of the American Chemical society*, 117(48):11946–11975, 1995.

[161] Alex D MacKerell Jr, Donald Bashford, MLDR Bellott, Roland Leslie Dunbrack Jr, Jeffrey D Evanseck, Martin J Field, Stefan Fischer, Jiali Gao, H Guo,

Sookhee Ha, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *The Journal of Physical Chemistry B*, 102(18):3586–3616, 1998.

[162] Paul K Weiner and Peter A Kollman. Amber: Assisted model building with energy refinement. a general program for modeling molecules and their interactions. *Journal of Computational Chemistry*, 2(3):287–303, 1981.

[163] Wendy D Cornell, Piotr Cieplak, Christopher I Bayly, Ian R Gould, Kenneth M Merz, David M Ferguson, David C Spellmeyer, Thomas Fox, James W Caldwell, and Peter A Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *Journal of the American Chemical Society*, 117(19):5179–5197, 1995.

[164] David A Pearlman, David A Case, James W Caldwell, Wilson S Ross, Thomas E Cheatham III, Steve DeBolt, David Ferguson, George Seibel, and Peter Kollman. Amber, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Computer Physics Communications*, 91(1-3):1–41, 1995.

[165] Wilfred F van Gunsteren and Herman JC Berendsen. Groningen molecular simulation (gromos) library manual. *Biomos, Groningen*, 24(682704):13, 1987.

[166] Wilfred F van Gunsteren, SR Billeter, AA Eising, PH Hünenberger, PKHC Krüger, AE Mark, WRP Scott, and IG Tironi. Biomolecular simulation: the gromos96 manual and user guide. *Vdf Hochschulverlag AG an der ETH Zürich, Zürich*, 86, 1996.

[167] Xavier Daura, Alan E Mark, and Wilfred F Van Gunsteren. Parametrization of aliphatic chn united atoms of gromos96 force field. *Journal of Computational Chemistry*, 19(5):535–547, 1998.

[168] Lukas D Schuler, Xavier Daura, and Wilfred F Van Gunsteren. An improved gromos96 force field for aliphatic hydrocarbons in the condensed phase. *Journal of Computational Chemistry*, 22(11):1205–1218, 2001.

[169] Chris Oostenbrink, Alessandra Villa, Alan E Mark, and Wilfred F Van Gunsteren. A biomolecular force field based on the free enthalpy of hydration and solvation: the gromos force-field parameter sets 53a5 and 53a6. *Journal of Computational Chemistry*, 25(13):1656–1676, 2004.

[170] William L Jorgensen and Julian Tirado-Rives. The opls [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6):1657–1666, 1988.

[171] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[172] Junmei Wang, Piotr Cieplak, and Peter A Kollman. How well does a restrained electrostatic potential (resp) model perform in calculating conformational energies of organic and biological molecules? *Journal of Computational Chemistry*, 21(12):1049–1074, 2000.

[173] Yong Duan, Chun Wu, Shibasish Chowdhury, Mathew C Lee, Guoming Xiong, Wei Zhang, Rong Yang, Piotr Cieplak, Ray Luo, Taisung Lee, et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, 24(16):1999–2012, 2003.

[174] Jing Huang and Alexander D MacKerell Jr. Charmm36 all-atom additive protein force field: Validation based on comparison to nmr data. *Journal of Computational Chemistry*, 34(25):2135–2145, 2013.

[175] Nathan Schmid, Andreas P Eichenberger, Alexandra Choutko, Sereina Riniker, Moritz Winger, Alan E Mark, and Wilfred F van Gunsteren. Definition and testing of the gromos force-field versions 54a7 and 54b7. *European Biophysics Journal*, 40(7):843–856, 2011.

[176] Roger Williams Hockney, SP Goel, and JW Eastwood. Quiet high-resolution computer models of a plasma. *Journal of Computational Physics*, 14(2):148–158, 1974.

[177] William C Swope, Hans C Andersen, Peter H Berens, and Kent R Wilson. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics*, 76(1):637–649, 1982.

[178] Michael P Allen and Dominic J Tildesley. *Computer simulation of liquids*. Oxford university press, 2017.

[179] Douglas Poland. Statistical mechanics (mcquarrie, donald a.), 1977.

[180] Sarah Rauscher and Régis Pomès. Molecular simulations of protein disorder. *Biochemistry and Cell Biology*, 88(2):269–290, 2010.

[181] Rafael C Bernardi, Marcelo CR Melo, and Klaus Schulten. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochimica et Biophysica Acta (BBA)-General Subjects*, 1850(5):872–877, 2015.

[182] Yi Isaac Yang, Qiang Shao, Jun Zhang, Lijiang Yang, and Yi Qin Gao. Enhanced sampling in molecular dynamics. *The Journal of Chemical Physics*, 151(7):070902, 2019.

[183] Scott Kirkpatrick, C Daniel Gelatt, and Mario P Vecchi. Optimization by simulated annealing. *Science*, 220(4598):671–680, 1983.

[184] Gianmarc Grazioli, Rachel Wagner Martin, and Carter Tribley Butts. Comparative exploratory analysis of intrinsically disordered protein dynamics using machine learning and network analytic methods. *Frontiers in Molecular Biosciences*, 6:42, 2019.

[185] Rishi Pal Singh. Application of graph theory in computer science and engineering. *International Journal of Computer Applications*, 104(1), 2014.

[186] Anna CF Lewis, Nick S Jones, Mason A Porter, and Charlotte M Deane. The function of communities in protein interaction networks at multiple scales. *BMC Systems Biology*, 4(1):100, 2010.

[187] Dennis M Wilkinson and Bernardo A Huberman. A method for finding communities of related genes. *Proceedings of the National Academy of Sciences*, 101(suppl 1):5241–5248, 2004.

[188] Roger Guimera and Luis A Nunes Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005.

[189] Mark EJ Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69(6):066133, 2004.

[190] Santo Fortunato. Community detection in graphs. *Physics Reports*, 486(3-5):75–174, 2010.

[191] Bo Yang, Dayou Liu, and Jiming Liu. Discovering communities from social networks: Methodologies and applications. In *Handbook of Social Network Technologies and Applications*, pages 331–346. Springer, 2010.

[192] Mark EJ Newman and Michelle Girvan. Finding and evaluating community structure in networks. *Physical Review E*, 69(2):026113, 2004.

[193] Muhammad Aqib Javed, Muhammad Shahzad Younis, Siddique Latif, Junaid Qadir, and Adeel Baig. Community detection in networks: A multidisciplinary review. *Journal of Network and Computer Applications*, 108:87–111, 2018.

[194] Aaron Clauset, Mark EJ Newman, and Cristopher Moore. Finding community structure in very large networks. *Physical Review E*, 70(6):066111, 2004.

[195] Mark EJ Newman. Analysis of weighted networks. *Physical Review E*, 70(5):056131, 2004.

[196] DW Goodall. Objective methods for the classification of vegetation. iii. an essay in the use of factor analysis. *Australian Journal of Botany*, 2(3):304–324, 1954.

[197] Brita G Schulze and Jeffrey D Evanseck. Cooperative role of arg45 and his64 in the spectroscopic a3 state of carbonmonoxy myoglobin: Molecular dynamics simulations, multivariate analysis, and quantum mechanical computations. *Journal of the American Chemical Society*, 121(27):6444–6454, 1999.

[198] Andrea Amadei, Antonius BM Linssen, and Herman JC Berendsen. Essential dynamics of proteins. *Proteins: Structure, Function, and Bioinformatics*, 17(4):412–425, 1993.

[199] Toshiko Ichiye and Martin Karplus. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins: Structure, Function, and Bioinformatics*, 11(3):205–217, 1991.

[200] Laura Riccardi, Phuong H Nguyen, and Gerhard Stock. Free-energy landscape of rna hairpins constructed via dihedral angle principal component analysis. *The Journal of Physical Chemistry B*, 113(52):16660–16668, 2009.

[201] Matthias Ernst, Florian Sittel, and Gerhard Stock. Contact-and distance-based principal component analysis of protein dynamics. *The Journal of Chemical Physics*, 143(24):12B640_1, 2015.

[202] Alexandros Altis, Moritz Otten, Phuong H Nguyen, Rainer Hegger, and Gerhard Stock. Construction of the free energy landscape of biomolecules via dihedral angle principal component analysis. *The Journal of Chemical Physics*, 128(24):06B620, 2008.

[203] José Nelson Onuchic, Zaida Luthey-Schulten, and Peter G Wolynes. Theory of protein folding: the energy landscape perspective. *Annual Review of Physical Chemistry*, 48(1):545–600, 1997.

[204] Ken A Dill and Hue Sun Chan. From levinthal to pathways to funnels. *Nature Structural Biology*, 4(1):10–19, 1997.

[205] Keith D Ball, R Stephen Berry, Ralph E Kunz, Feng-Yin Li, Ana Proykova, and David J Wales. From topographies to dynamics on multidimensional potential energy surfaces of atomic clusters. *Science*, 271(5251):963–966, 1996.

[206] Martin Gruebele. Protein folding: the free energy surface. *Current Opinion in Structural Biology*, 12(2):161–168, 2002.

[207] Abhinav Jain and Gerhard Stock. Hierarchical folding free energy landscape of hp35 revealed by most probable path clustering. *The Journal of Physical Chemistry B*, 118(28):7750–7760, 2014.

[208] Laura Riccardi, Phuong H Nguyen, and Gerhard Stock. Construction of the free energy landscape of peptide aggregation from molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 8(4):1471–1479, 2012.

[209] Gia G Maisuradze, Adam Liwo, and Harold A Scheraga. Relation between free energy landscapes of proteins and dynamics. *Journal of Chemical Theory and Computation*, 6(2):583–595, 2010.

[210] Florian Sittel, Abhinav Jain, and Gerhard Stock. Principal component analysis of molecular dynamics: On the use of cartesian vs. internal coordinates. *The Journal of Chemical Physics*, 141(1):07B605_1, 2014.

[211] Steven Hayward and Bert L De Groot. Normal modes and essential dynamics. In *Molecular Modeling of Proteins*, pages 89–106. Springer, 2008.

[212] Hiroaki Fukumoto, Asano Asami-Odaka, Nobuhiro Suzuki, Hiroyuki Shimada, Yasuo Ihara, and Takeshi Iwatsubo. Amyloid beta protein deposition in normal aging has the same characteristics as that in alzheimer's disease. predominance of a beta 42 (43) and association of a beta 40 with cored plaques. *The American Journal of Pathology*, 148(1):259, 1996.

[213] Iryna Benilova, Eric Karran, and Bart De Strooper. The toxic aβ oligomer and alzheimer's disease: an emperor in need of clothes. *Nature Neuroscience*, 15(3):349–357, 2012.

[214] Rakez Kayed, Elizabeth Head, Jennifer L Thompson, Theresa M McIntire, Saskia C Milton, Carl W Cotman, and Charles G Glabe. Common structure of soluble amyloid oligomers implies common mechanism of pathogenesis. *Science*, 300(5618):486–489, 2003.

[215] Christian Haass and Dennis J Selkoe. Soluble protein oligomers in neurodegeneration: lessons from the alzheimer's amyloid β-peptide. *Nature Reviews Molecular Cell Biology*, 8(2):101–112, 2007.

[216] Govardhan Reddy, John E Straub, and Devarajan Thirumalai. Influence of preformed asp23- lys28 salt bridge on the conformational fluctuations of monomers and dimers of aβ peptides with implications for rates of fibril formation. *The Journal of Physical Chemistry B*, 113(4):1162–1172, 2009.

[217] Simon Mitternacht, Iskra Staneva, Torleif Härd, and Anders Irbäck. Comparing the folding free-energy landscapes of aβ42 variants with different aggregation properties. *Proteins: Structure, Function, and Bioinformatics*, 78(12):2600–2608, 2010.

[218] Peter Niraj Nirmalraj, Jonathan List, Shayon Battacharya, Geoffrey Howe, Liang Xu, Damien Thompson, and Michael Mayer. Complete aggregation pathway of amyloid β (1-40) and (1-42) resolved on an atomically clean interface. *Science Advances*, 6(15):eaaz6014, 2020.

[219] Ann Tiiman, Jekaterina Krishtal, Peep Palumaa, and Vello Tõugu. In vitro fibrillization of alzheimers amyloid-β peptide (1-42). *AIP Advances*, 5(9):092401, 2015.

[220] Gal Bitan, Marina D Kirkitadze, Aleksey Lomakin, Sabrina S Vollers, George B Benedek, and David B Teplow. Amyloid β-protein (aβ) assembly: Aβ40 and aβ42 oligomerize through distinct pathways. *Proceedings of the National Academy of Sciences*, 100(1):330–335, 2003.

[221] Leonid Breydo and Vladimir N Uversky. Structural, morphological, and functional diversity of amyloid oligomers. *FEBS Letters*, 589(19):2640–2648, 2015.

[222] Christian Haass, Michael G Schlossmacher, Albert Y Hung, Carmen Vigo-Pelfrey, Angela Mellon, Beth L Ostaszewski, Ivan Lieberburg, Edward H Koo, Dale Schenk, David B Teplow, et al. Amyloid β-peptide is produced by cultured cells during normal metabolism. *Nature*, 359(6393):322–325, 1992.

[223] S Zhang, K Iwata, MJ Lachenmann, JW Peng, S Li, ER Stimson, Y-A Lu, AM Felix, JE Maggio, and JP Lee. The alzheimer's peptide aβ adopts a collapsed coil structure in water. *Journal of Structural Biology*, 130(2-3):130–141, 2000.

[224] Mingfeng Yang and David B Teplow. Amyloid β-protein monomer folding: free-energy surfaces reveal alloform-specific differences. *Journal of Molecular Biology*, 384(2):450–464, 2008.

[225] Nikolaos G Sgourakis, Yilin Yan, Scott A McCallum, Chunyu Wang, and Angel E Garcia. The alzheimers peptides aβ40 and 42 adopt distinct conformations in water: a combined md/nmr study. *Journal of Molecular Biology*, 368(5):1448–1457, 2007.

[226] Julien Roche, Yang Shen, Jung Ho Lee, Jinfa Ying, and Ad Bax. Monomeric aβ1–40 and aβ1–42 peptides in solution adopt very similar ramachandran map distributions that closely resemble random coil. *Biochemistry*, 55(5):762–775, 2016.

[227] K Aurelia Ball, Aaron H Phillips, David E Wemmer, and Teresa Head-Gordon. Differences in β-strand populations of monomeric aβ40 and aβ42. *Biophysical Journal*, 104(12):2714–2724, 2013.

[228] Holger Wille, Lyudmyla Dorosh, Sara Amidian, Gerold Schmitt-Ulms, and Maria Stepanova. Combining molecular dynamics simulations and experimental analyses in protein misfolding. In *Advances in Protein Chemistry and Structural Biology*, volume 118, pages 33–110. Elsevier, 2019.

[229] Darryl D Humphreys, Richard A Friesner, and Bruce J Berne. Simulated annealing of a protein in a continuum solvent by multiple-time-step molecular dynamics. *The Journal of Physical Chemistry*, 99(26):10674–10685, 1995.

[230] Joshua L Phillips, Michael E Colvin, and Shawn Newsam. Validating clustering of molecular dynamics simulations using polymer models. *BMC Bioinformatics*, 12(1):1–23, 2011.

[231] Santo Fortunato and Claudio Castellano. Community structure in graphs. *ArXiv Preprint ArXiv:0712.2716*, 2007.

[232] Bert L de Groot, Xavier Daura, Alan E Mark, and Helmut Grubmüller. Essential dynamics of reversible peptide folding: memory-free conformational dynamics governed by internal hydrogen bonds. *Journal of Molecular Biology*, 309(1):299–313, 2001.

[233] Warren Lyford DeLano. Pymol, 2002.

[234] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C Smith, Berk Hess, and Erik Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[235] HJC Berendsen, JR Grigera, and TP Straatsma. The missing term in effective pair potentials. *Journal of Physical Chemistry*, 91(24):6269–6271, 1987.

[236] Tom Darden, Darrin York, and Lee Pedersen. Particle mesh ewald: An n log (n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[237] Szilárd Páll and Berk Hess. A flexible algorithm for calculating pair interactions on simd architectures. *Computer Physics Communications*, 184(12):2641–2650, 2013.

[238] Berk Hess, Henk Bekker, Herman JC Berendsen, and Johannes GEM Fraaije. Lincs: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[239] Giovanni Bussi, Davide Donadio, and Michele Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.

[240] Shuichi Nosé and ML Klein. Constant pressure molecular dynamics for molecular systems. *Molecular Physics*, 50(5):1055–1076, 1983.

[241] Michele Parrinello and Aneesur Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12):7182–7190, 1981.

[242] Herman JC Berendsen, JPM van Postma, Wilfred F van Gunsteren, ARHJ DiNola, and Jan R Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[243] Ian L Dryden and Kanti V Mardia. Statistical shape analysis: Wiley series in probability and statistics, 1998.

[244] John C Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.

[245] J. T. Kent. *New directions in shape analysis. In The art of statistical science. A tribute to GS Watson (K. V. Mardia, ed.).* Chichester (UK) Wiley, 1992.

[246] I. L. Dryden and K. V. Mardia. *Statistical Shape Analysis, with Applications in R. Second Edition.* Wiley, Chichester, 2016.

[247] Jose Hleap. Comparative quantitative genetics of protein structures: A composite approach to protein structure evolution. 2017.

[248] I. L. Dryden. `shapes` *package*. R Foundation for Statistical Computing, Vienna, Austria, 2016. Contributed package, Version 1.1-13.

[249] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[250] Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

[251] Wolfgang Kabsch and Christian Sander. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers*, 22(12):2577–2637, 1983.

[252] Youcef Fezoui and David B Teplow. Kinetic studies of amyloid β-protein fibril assembly: differential effects of α-helix stabilization. *Journal of Biological Chemistry*, 277(40):36948–36954, 2002.

[253] Kenjiro Ono, Margaret M Condron, and David B Teplow. Structure–neurotoxicity relationships of amyloid β-protein oligomers. *Proceedings of the National Academy of Sciences*, 106(35):14745–14750, 2009.

[254] Ziye Liu, Fan Jiang, and Yun-Dong Wu. Significantly different contact patterns between aβ40 and aβ42 monomers involving the n-terminal region. *Chemical Biology & Drug Design*, 94(3):1615–1625, 2019.

[255] Tran Thi Minh Thu, Nguyen Truong Co, Ly Anh Tu, and Mai Suan Li. Aggregation rate of amyloid beta peptide is controlled by beta-content in monomeric state. *The Journal of Chemical Physics*, 150(22):225101, 2019.

[256] Matthias Schmidt, Carsten Sachse, Walter Richter, Chen Xu, Marcus Fändrich, and Nikolaus Grigorieff. Comparison of alzheimer aβ (1–40) and aβ (1–42) amyloid fibrils reveals similar protofilament structures. *Proceedings of the National Academy of Sciences*, 106(47):19813–19818, 2009.

[257] Kai-Fang Huang, Shu-Heng Dong, Su-Su Zhong, Hao Li, and Li-Li Duan. Study on the amyloid aβ42 with accelerated molecular dynamics simulations. *Communications in Theoretical Physics*, 71(9):1121, 2019.

[258] Andreas Vitalis and Amedeo Caflisch. Micelle-like architecture of the monomer ensemble of alzheimers amyloid-β peptide in aqueous solution and its implications for aβ aggregation. *Journal of Molecular Biology*, 403(1):148–165, 2010.

[259] Fabrizio Bossis and Luigi L Palese. Amyloid beta (1–42) in aqueous environments: effects of ionic strength and e22q (dutch) mutation. *Biochimica et Biophysica Acta (BBA)-Proteins and Proteomics*, 1834(12):2486–2493, 2013.

[260] Chengqiang Liu, Wei Zhao, Xiaofeng Xing, Hu Shi, Baotao Kang, Haiying Liu, Ping Li, and Hongqi Ai. An original monomer sampling from a ready-made aβ42 nmr fibril suggests a turn-β-strand synergetic seeding mechanism. *ChemPhysChem*, 20(12):1649–1660, 2019.

[261] K Aurelia Ball, Aaron H Phillips, Paul S Nerenberg, Nicolas L Fawzi, David E Wemmer, and Teresa Head-Gordon. Homogeneous and heterogeneous tertiary structure ensembles of amyloid-β peptides. *Biochemistry*, 50(35):7612–7628, 2011.

[262] Prabir Khatua, Jaya C Jose, Neelanjana Sengupta, and Sanjoy Bandyopadhyay. Conformational features of the aβ 42 peptide monomer and its interaction with the surrounding solvent. *Physical Chemistry Chemical Physics*, 18(43):30144–30159, 2016.

[263] Nelson A Alves and Rafael B Frigori. Structural interconversion in alzheimers amyloid-β (16–35) peptide in an aqueous solution. *The Journal of Physical Chemistry B*, 122(6):1869–1875, 2018.

[264] Guo-fang Chen, Ting-hai Xu, Yan Yan, Yu-ren Zhou, Yi Jiang, Karsten Melcher, and H Eric Xu. Amyloid beta: structure, biology and structure-based therapeutic development. *Acta Pharmacologica Sinica*, 38(9):1205–1235, 2017.

[265] Weihua Zheng, Min-Yeh Tsai, and Peter G Wolynes. Comparing the aggregation free energy landscapes of amyloid beta (1–42) and amyloid beta (1–40). *Journal of the American Chemical Society*, 139(46):16666–16676, 2017.

[266] David J Rosenman, Christopher R Connors, Wen Chen, Chunyu Wang, and Angel E García. Aβ monomers transiently sample oligomer and fibril-like configurations: Ensemble characterization using a combined md/nmr approach. *Journal of Molecular Biology*, 425(18):3338–3359, 2013.

[267] Song-Ho Chong, Mirae Park, and Sihyun Ham. Structural and thermodynamic characteristics that seed aggregation of amyloid-β protein in water. *Journal of Chemical Theory and Computation*, 8(2):724–734, 2012.

[268] Danting Huang, Maxwell I Zimmerman, Patricia K Martin, A Jeremy Nix, Terrone L Rosenberry, and Anant K Paravastu. Antiparallel β-sheet structure within the c-terminal region of 42-residue alzheimer's amyloid-β peptides when they form 150-kda oligomers. *Journal of Molecular Biology*, 427(13):2319–2328, 2015.

[269] Payel Das, Brian Murray, and Georges Belfort. Alzheimers protective a2t mutation changes the conformational landscape of the aβ1–42 monomer differently than does the a2v mutation. *Biophysical Journal*, 108(3):738–747, 2015.

[270] Qin Qiao, Gregory R Bowman, and Xuhui Huang. Dynamics of an intrinsically disordered protein reveal metastable conformations that potentially seed aggregation. *Journal of the American Chemical Society*, 135(43):16092–16101, 2013.

[271] Pinaki Misra, Ravindra Kodali, Saketh Chemuru, Karunakar Kar, and Ronald Wetzel. Rapid α-oligomer formation mediated by the aβ c terminus initiates an amyloid assembly pathway. *Nature Communications*, 7(1):1–15, 2016.

[272] D Thirumalai, Govardhan Reddy, and John E Straub. Role of water in protein aggregation and amyloid polymorphism. *Accounts of Chemical Research*, 45(1):83–92, 2012.

[273] Samuel IA Cohen, Sara Linse, Leila M Luheshi, Erik Hellstrand, Duncan A White, Luke Rajah, Daniel E Otzen, Michele Vendruscolo, Christopher M Dobson, and Tuomas PJ Knowles. Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proceedings of the National Academy of Sciences*, 110(24):9758–9763, 2013.

[274] Tomohiro Umeda, Takami Tomiyama, Naomi Sakama, Saya Tanaka, Mary P Lambert, William L Klein, and Hiroshi Mori. Intraneuronal amyloid β oligomers cause cell death via endoplasmic reticulum stress, endosomal/lysosomal leakage, and mitochondrial dysfunction in vivo. *Journal of Neuroscience Research*, 89(7):1031–1042, 2011.

[275] R Resende, E Ferreiro, C Pereira, and C Resende De Oliveira. Neurotoxic effect of oligomeric and fibrillar species of amyloid-beta peptide 1-42: involvement of endoplasmic reticulum calcium release in oligomer-induced cell death. *Neuroscience*, 155(3):725–737, 2008.

[276] Dong-Hyung Cho, Tomohiro Nakamura, Jianguo Fang, Piotr Cieplak, Adam Godzik, Zezong Gu, and Stuart A Lipton. S-nitrosylation of drp1 mediates β-amyloid-related mitochondrial fission and neuronal injury. *Science*, 324(5923):102–105, 2009.

[277] Shayon Bhattacharya, Liang Xu, and Damien Thompson. Revisiting the earliest signatures of amyloidogenesis: Roadmaps emerging from computational modeling and experiment. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(4):e1359, 2018.

[278] Eric B Gibbs, Erik C Cook, and Scott A Showalter. Application of nmr to studies of intrinsically disordered proteins. *Archives of Biochemistry and Biophysics*, 628:57–70, 2017.

[279] Noel D Lazo, Marianne A Grant, Margaret C Condron, Alan C Rigby, and David B Teplow. On the nucleation of amyloid β-protein monomer folding. *Protein Science*, 14(6):1581–1596, 2005.

[280] B Urbanc, M Betnel, L Cruz, G Bitan, and DB Teplow. Elucidation of amyloid β-protein oligomerization mechanisms: discrete molecular dynamics study. *Journal of the American Chemical Society*, 132(12):4266–4280, 2010.

[281] Marina D Kirkitadze, Margaret M Condron, and David B Teplow. Identification and characterization of key kinetic intermediates in amyloid β-protein fibrillogenesis. *Journal of Molecular Biology*, 312(5):1103–1119, 2001.

[282] Shayon Bhattacharya, Liang Xu, and Damien Thompson. Molecular simulations reveal terminal group mediated stabilization of helical conformers in both amyloid-β42 and α-synuclein. *ACS Chemical Neuroscience*, 10(6):2830–2842, 2019.

[283] Linh Tran, Nathalie Basdevant, Chantal Prévost, and Tâp Ha-Duong. Structure of ring-shaped aβ 42 oligomers determined by conformational selection. *Scientific Reports*, 6:21429, 2016.

[284] Bogdan Barz, Qinghua Liao, and Birgit Strodel. Pathways of amyloid-β aggregation depend on oligomer shape. *Journal of the American Chemical Society*, 140(1):319–327, 2018.

[285] Sébastien Côté, Philippe Derreumaux, and Normand Mousseau. Distinct morphologies for amyloid beta protein monomer: Aβ1–40, aβ1–42, and aβ1–40 (d23n). *Journal of Chemical Theory and Computation*, 7(8):2584–2592, 2011.

[286] Nguyen Hoang Linh, Tran Thi Minh Thu, LyAnh Tu, Chin-Kun Hu, and Mai Suan Li. Impact of mutations at c-terminus on structures and dynamics of aβ40 and aβ42: a molecular simulation study. *The Journal of Physical Chemistry B*, 121(17):4341–4354, 2017.

[287] Srirupa Chakraborty and Payel Das. Emergence of alternative structures in amyloid beta 1-42 monomeric landscape by n-terminal hexapeptide amyloid inhibitors. *Scientific Reports*, 7(1):1–12, 2017.

[288] Nikolay Blinov, Massih Khorvash, David S Wishart, Neil R Cashman, and Andriy Kovalenko. Initial structural models of the aβ42 dimer from replica exchange molecular dynamics simulations. *ACS Omega*, 2(11):7621–7636, 2017.

[289] Yuxi Lin, Haeri Im, Sihyun Ham, et al. Characterizing the structural and thermodynamic properties of aβ42 and aβ40. *Biochemical and Biophysical Research Communications*, 510(3):442–448, 2019.

[290] Asim Okur, Daniel R Roe, Guanglei Cui, Viktor Hornak, and Carlos Simmerling. Improving convergence of replica-exchange simulations through coupling to a high-temperature structure reservoir. *Journal of Chemical Theory and Computation*, 3(2):557–568, 2007.

[291] Asim Okur, Lauren Wickstrom, Melinda Layten, Raphäel Geney, Kun Song, Viktor Hornak, and Carlos Simmerling. Improved efficiency of replica exchange simulations through use of a hybrid explicit/implicit solvation model. *Journal of Chemical Theory and Computation*, 2(2):420–433, 2006.

[292] Helen M Berman, John Westbrook, Zukang Feng, Gary Gilliland, Talapady N Bhat, Helge Weissig, Ilya N Shindyalov, and Philip E Bourne. The protein data bank. *Nucleic Acids Research*, 28(1):235–242, 2000.

[293] Dmitrij Frishman and Patrick Argos. Knowledge-based protein secondary structure assignment. *Proteins: Structure, Function, and Bioinformatics*, 23(4):566–579, 1995.

[294] Geerten W Vuister and AD Bax. Quantitative j correlation: a new approach for measuring homonuclear three-bond j (hnh. alpha.) coupling constants in 15n-enriched proteins. *Journal of the American Chemical Society*, 115(17):7772–7777, 1993.

[295] Beat Vögeli, Jinfa Ying, Alexander Grishaev, and Ad Bax. Limits on variations in protein backbone dynamics from precise measurements of scalar couplings. *Journal of the American Chemical Society*, 129(30):9377–9385, 2007.

[296] Yilin Yan, Scott A McCallum, and Chunyu Wang. M35 oxidation induces aβ40-like structural and dynamical changes in aβ42. *Journal of the American Chemical Society*, 130(16):5394–5395, 2008.

[297] Thomas D Schneider and R Michael Stephens. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Research*, 18(20):6097–6100, 1990.

[298] Gavin E Crooks, Gary Hon, John-Marc Chandonia, and Steven E Brenner. Weblogo: a sequence logo generator. *Genome Research*, 14(6):1188–1190, 2004.

[299] Richard W Hamming. Error detecting and error correcting codes. *The Bell System Technical Journal*, 29(2):147–160, 1950.

[300] Badri Adhikari and Jianlin Cheng. Protein residue contacts and prediction methods. In *Data Mining Techniques for the Life Sciences*, pages 463–476. Springer, 2016.

[301] Chao Yuan, Hao Chen, and Daisuke Kihara. Effective inter-residue contact definitions for accurate protein fold recognition. *BMC Bioinformatics*, 13(1):1–13, 2012.

[302] Jose Sergio Hleap, Edward Susko, and Christian Blouin. Defining structural and evolutionary modules in proteins: a community detection approach to explore sub-domain architecture. *BMC Structural Biology*, 13(1):1, 2013.

[303] Marco A Deriu, Gianvito Grasso, Ginevra Licandro, Andrea Danani, Diego Gallo, Jack A Tuszynski, and Umberto Morbiducci. Investigation of the josephin domain protein-protein interaction by molecular dynamics. *PLoS One*, 9(9):e108677, 2014.

[304] Gianvito Grasso, Martina Rebella, Stefano Muscat, Umberto Morbiducci, Jack Tuszynski, Andrea Danani, and Marco A Deriu. Conformational dynamics and stability of u-shaped and s-shaped amyloid β assemblies. *International Journal of Molecular Sciences*, 19(2):571, 2018.

[305] Sandeep Kumar and Ruth Nussinov. Salt bridge stability in monomeric proteins. *Journal of Molecular Biology*, 293(5):1241–1255, 1999.

[306] Jason E Donald, Daniel W Kulp, and William F DeGrado. Salt bridges: geometrically specific, designable interactions. *Proteins: Structure, Function, and Bioinformatics*, 79(3):898–915, 2011.

[307] David J Barlow and JM Thornton. Ion-pairs in proteins. *Journal of molecular biology*, 168(4):867–885, 1983.

[308] Douglas F Sticke, Leonard G Presta, Ken A Dill, and George D Rose. Hydrogen bonding in globular proteins. *Journal of Molecular Biology*, 226(4):1143–1159, 1992.

[309] Simone Kosol, Sara Contreras-Martos, Cesyen Cedeño, and Peter Tompa. Structural characterization of intrinsically disordered proteins by nmr spectroscopy. *Molecules*, 18(9):10802–10828, 2013.

[310] Ravi Pratap Barnwal, Ashok K Rout, Kandala VR Chary, and Hanudatta S Atreya. Rapid measurement of 3 j (hn–h α) and 3 j (n–h β) coupling constants in polypeptides. *Journal of Biomolecular NMR*, 39(4):259–263, 2007.

[311] Marielle Aulikki Wälti, Julien Orts, Beat Vögeli, Silvia Campioni, and Roland Riek. Solution nmr studies of recombinant aβ (1–42): From the presence of a micellar entity to residual β-sheet structure in the soluble species. *ChemBioChem*, 16(4):659–669, 2015.

[312] David Jacob Rosenman. *Biophysical characterization of intrinsically disordered peptides through molecular dynamics simulations and solution nuclear magnetic resonance spectroscopy*. Rensselaer Polytechnic Institute, 2015.

[313] Beomsoo Han, Yifeng Liu, Simon W Ginzinger, and David S Wishart. Shiftx2: significantly improved protein chemical shift prediction. *Journal of Biomolecular NMR*, 50(1):43, 2011.

[314] Stephen Neal, Alex M Nip, Haiyan Zhang, and David S Wishart. Rapid and accurate calculation of protein 1 h, 13 c and 15 n chemical shifts. *Journal of Biomolecular NMR*, 26(3):215–240, 2003.

[315] David S Wishart, M Scott Watson, Robert F Boyko, and Brian D Sykes. Automated 1 h and 13 c chemical shift prediction using the biomagresbank. *Journal of Biomolecular NMR*, 10(4):329–336, 1997.

[316] Yang Shen and Ad Bax. Sparta+: a modest improvement in empirical nmr chemical shift prediction by means of an artificial neural network. *Journal of Biomolecular NMR*, 48(1):13–22, 2010.

[317] Theodoros K Karamanos, Arnout P Kalverda, Gary S Thompson, and Sheena E Radford. Mechanisms of amyloid formation revealed by solution nmr. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 88:86–104, 2015.

[318] Sébastien Santini, Guanghong Wei, Normand Mousseau, and Philippe Derreumaux. Pathway complexity of alzheimer's β-amyloid aβ16-22 peptide assembly. *Structure*, 12(7):1245–1255, 2004.

[319] Christina Bergonzo, Niel M Henriksen, Daniel R Roe, Jason M Swails, Adrian E Roitberg, and Thomas E Cheatham III. Multidimensional replica exchange molecular dynamics yields a converged ensemble of an rna tetranucleotide. *Journal of Chemical Theory and Computation*, 10(1):492–499, 2014.

[320] Bernhard Knapp, Luis Ospina, and Charlotte M Deane. Avoiding false positive conclusions in molecular simulation: the importance of replicas. *Journal of Chemical Theory and Computation*, 14(12):6127–6138, 2018.

[321] Kingsley Y Wu, David Doan, Marco Medrano, and Chia-en A Chang. Modeling structural interconversion in alzheimers amyloid beta peptide with classical and intrinsically disordered protein force fields. *Journal of Biomolecular Structure and Dynamics*, pages 1–18, 2021.

[322] Andrzej Kolinski and Jeffrey Skolnick. Monte carlo simulations of protein folding. i. lattice model and interaction scheme. *Proteins: Structure, Function, and Bioinformatics*, 18(4):338–352, 1994.

[323] Andrij Baumketner, Summer L Bernstein, Thomas Wyttenbach, Gal Bitan, David B Teplow, Michael T Bowers, and Joan-Emma Shea. Amyloid β-protein monomer structure: A computational and experimental study. *Protein Science*, 15(3):420–428, 2006.

[324] Duilio F Raffa and Arvi Rauk. Molecular dynamics study of the beta amyloid peptide of alzheimer's disease and its divalent copper complexes. *The Journal of Physical Chemistry B*, 111(14):3789–3799, 2007.

[325] Bogdan Barz, Alexander K Buell, and Soumav Nath. Compact fibril-like structure of amyloid β-peptide (1–42) monomers. *Chemical Communications*, 57(7):947–950, 2021.

[326] Aneta T Petkova, Wai-Ming Yau, and Robert Tycko. Experimental constraints on quaternary structure in alzheimer's β-amyloid fibrils. *Biochemistry*, 45(2):498–512, 2006.

[327] Bogdan Tarus, John E Straub, and D Thirumalai. Dynamics of asp23- lys28 salt-bridge formation in aβ10-35 monomers. *Journal of the American Chemical Society*, 128(50):16159–16168, 2006.

[328] Kenichiro Imai and Shigeki Mitaku. Mechanisms of secondary structure breakers in soluble proteins. *Biophysics*, 1:55–65, 2005.

[329] Sanguk Kim, Tae-Joon Jeon, Amit Oberai, Duan Yang, Jacob J Schmidt, and James U Bowie. Transmembrane glycine zippers: physiological and pathological roles in membrane proteins. *Proceedings of the National Academy of Sciences*, 102(40):14278–14283, 2005.

[330] Wei Liu, Evan Crocker, Wenyi Zhang, James I Elliott, Burkhard Luy, Huilin Li, Saburo Aimoto, and Steven O Smith. Structural role of glycine in amyloid fibrils formed from transmembrane α-helices. *Biochemistry*, 44(9):3591–3597, 2005.

[331] Anja Harmeier, Christian Wozny, Benjamin R Rost, Lisa-Marie Munter, Haiqing Hua, Oleg Georgiev, Michael Beyermann, Peter W Hildebrand, Christoph Weise, Walter Schaffner, et al. Role of amyloid-β glycine 33 in oligomerization, toxicity, and neuronal plasticity. *Journal of Neuroscience*, 29(23):7582–7590, 2009.

[332] Karen E Marshall, Devkee M Vadukul, Liza Dahal, Alina Theisen, Milena W Fowler, Youssra Al-Hilaly, Lenzie Ford, György Kemenes, Iain J Day, Kevin Staras, et al. A critical role for the self-assembly of amyloid-β1-42 in neurodegeneration. *Scientific Reports*, 6(1):1–13, 2016.

[333] Brian Murray, Bhanushee Sharma, and Georges Belfort. N-terminal hypothesis for alzheimers disease, 2017.

[334] Christian A Söldner, Heinrich Sticht, and Anselm HC Horn. Role of the n-terminus for the stability of an amyloid-β fibril with three-fold symmetry. *PLoS One*, 12(10):e0186347, 2017.

[335] Bappaditya Chandra, Debanjan Bhowmik, Barun Kumar Maity, Kaustubh R Mote, Debabrata Dhara, Ravindra Venkatramani, Sudipta Maiti, and Perunthiruthy K Madhu. Major reaction coordinates linking transient amyloid-β oligomers to fibrils measured at atomic level. *Biophysical Journal*, 113(4):805–816, 2017.

[336] Kohei Noda, Yuhei Tachi, and Yuko Okamoto. Structural characteristics of monomeric aβ42 on fibril in the early stage of secondary nucleation process. *ACS Chemical Neuroscience*, 11(19):2989–2998, 2020.

[337] Mariacristina Valerio, Fernando Porcelli, Joseph P Zbilut, Alessandro Giuliani, Cesare Manetti, and Filippo Conti. ph effects on the conformational preferences of amyloid β-peptide (1–40) in hfip aqueous solution by nmr spectroscopy. *ChemMedChem: Chemistry Enabling Drug Discovery*, 3(5):833–843, 2008.

[338] Shan-Yang Lin, Horng-Lun Chu, and Yen-Shan Wei. Secondary conformations and temperature effect on structural transformation of amyloid β (1–28),(1–40) and (1–42) peptides. *Journal of Biomolecular Structure and Dynamics*, 20(4):595–601, 2003.

[339] Cecilia Wallin, Jinghui Luo, Jüri Jarvet, Sebastian KTS Wärmländer, and Astrid Gräslund. The amyloid-β peptide in amyloid formation processes: Interactions with blood proteins and naturally occurring metal ions. *Israel Journal of Chemistry*, 57(7-8):674–685, 2017.

[340] Andrew J Doig and Philippe Derreumaux. Inhibition of protein aggregation and amyloid formation by small molecules. *Current Opinion in Structural Biology*, 30:50–56, 2015.

[341] Yifat Miller, Buyong Ma, and Ruth Nussinov. Polymorphism in alzheimer aβ amyloid organization reflects conformational selection in a rugged energy landscape. *Chemical Reviews*, 110(8):4820–4838, 2010.

[342] Sanjai Kumar Pachahara, Nitin Chaudhary, Chilukuri Subbalakshmi, and Ramakrishnan Nagaraj. Hexafluoroisopropanol induces self-assembly of β-amyloid peptides into highly ordered nanostructures. *Journal of Peptide Science*, 18(4):233–241, 2012.

[343] Bo Zhang-Haagen, Ralf Biehl, Luitgard Nagel-Steger, Aurel Radulescu, Dieter Richter, and Dieter Willbold. Monomeric amyloid beta peptide in hexafluoroisopropanol detected by small angle neutron scattering. *PLoS One*, 11(2):e0150267, 2016.

[344] Sanjai Kumar Pachahara, Harikrishna Adicherla, and Ramakrishnan Nagaraj. Self-assembly of aβ40, aβ42 and aβ43 peptides in aqueous mixtures of fluorinated alcohols. *PLoS One*, 10(8):e0136567, 2015.

[345] Chih-Lung Shen and Regina M Murphy. Solvent effects on self-assembly of beta-amyloid peptide. *Biophysical Journal*, 69(2):640–651, 1995.

[346] Myeongsang Lee, Hyun Joon Chang, Jung Yeon Park, Joonha Shin, Jong Woo Park, Jee Woo Choi, Jae In Kim, and Sungsoo Na. Conformational changes of aβ (1–42) monomers in different solvents. *Journal of Molecular Graphics and Modelling*, 65:8–14, 2016.

[347] S Dixit, Jason Crain, WCK Poon, John L Finney, and Alan K Soper. Molecular segregation observed in a concentrated alcohol–water solution. *Nature*, 416(6883):829–832, 2002.

[348] Keiko Nishikawa and Takao Iijima. Small-angle x-ray scattering study of fluctuations in ethanol and water mixtures. *The Journal of Physical Chemistry*, 97(41):10824–10828, 1993.

[349] Takaaki Sato, Akio Chiba, and Ryusuke Nozaki. Dynamical aspects of mixing schemes in ethanol–water mixtures in terms of the excess partial molar activation free energy, enthalpy, and entropy of the dielectric relaxation process. *The Journal of Chemical Physics*, 110(5):2508–2521, 1999.

[350] F t Franks and DJG Ives. The structural properties of alcohol–water mixtures. *Quarterly Reviews, Chemical Society*, 20(1):1–44, 1966.

[351] J Fidler and PM Rodger. Solvation structure around aqueous alcohols. *The Journal of Physical Chemistry B*, 103(36):7695–7703, 1999.

[352] Akihiro Wakisaka and Kazuo Matsuura. Microheterogeneity of ethanol–water binary mixtures observed at the cluster level. *Journal of Molecular Liquids*, 129(1-2):25–32, 2006.

[353] Sergei Yu Noskov, Guillaume Lamoureux, and Benoît Roux. Molecular dynamics study of hydration in ethanol- water mixtures using a polarizable force field. *The Journal of Physical Chemistry B*, 109(14):6705–6713, 2005.

[354] Aziz Ghoufi, Franck Artzner, and Patrice Malfreyt. Physical properties and hydrogen-bonding network of water–ethanol mixtures from molecular dynamics simulations. *The Journal of Physical Chemistry B*, 120(4):793–802, 2016.

[355] K Egashira and N Nishi. Low-frequency raman spectroscopy of ethanol- water binary solution: evidence for self-association of solute and solvent molecules. *The Journal of Physical Chemistry B*, 102(21):4054–4057, 1998.

[356] J-H Guo, Yi Luo, Andreas Augustsson, Stepan Kashtanov, J-E Rubensson, David K Shuh, Hans Ågren, and Joseph Nordgren. Molecular structure of alcohol-water mixtures. *Physical Review Letters*, 91(15):157401, 2003.

[357] Akihiro Wakisaka and Takahiro Ohki. Phase separation of water–alcohol binary mixtures induced by the microheterogeneity. *Faraday Discussions*, 129:231–245, 2005.

[358] Marija Nedić, Tobias N Wassermann, René Wugt Larsen, and Martin A Suhm. A combined raman-and infrared jet study of mixed methanol–water and ethanol–water clusters. *Physical Chemistry Chemical Physics*, 13(31):14050–14063, 2011.

[359] Ruoyu Li, Carmine DAgostino, James McGregor, Michael D Mantle, J Axel Zeitler, and Lynn F Gladden. Mesoscopic structuring and dynamics of alcohol/water solutions probed by terahertz time-domain spectroscopy and pulsed field gradient nuclear magnetic resonance. *The Journal of Physical Chemistry B*, 118(34):10156–10166, 2014.

[360] Cuijuan Zhang and Xiaoning Yang. Molecular dynamics simulation of ethanol/water mixtures for structure and diffusion properties. *Fluid Phase Equilibria*, 231(1):1–10, 2005.

[361] SSN Murthy. Detailed study of ice clathrate relaxation: evidence for the existence of clathrate structures in some water- alcohol mixtures. *The Journal of Physical Chemistry A*, 103(40):7927–7937, 1999.

[362] Sergey Burikov, Tatiana Dolenko, Svetlana Patsaeva, Yuriy Starokurov, and Viktor Yuzhakov. Raman and ir spectroscopy research on hydrogen bonding in water–ethanol systems. *Molecular Physics*, 108(18):2427–2436, 2010.

[363] A Cipiciani, G Onori, and G Savelli. Structural properties of water-ethanol mixtures: a correlation with the formation of micellar aggregates. *Chemical Physics Letters*, 143(5):505–509, 1988.

[364] A Wakisaka, S Komatsu, and Y Usui. Solute-solvent and solvent-solvent interactions evaluated through clusters isolated from solutions: Preferential solvation in water-alcohol mixtures. *Journal of Molecular Liquids*, 90(1-3):175–184, 2001.

[365] Tatiana A Dolenko, Sergey A Burikov, Sergey A Dolenko, Alexander O Efitorov, Ivan V Plastinin, Viktor I Yuzhakov, and Svetlana V Patsaeva. Raman spectroscopy of water–ethanol solutions: the estimation of hydrogen bonding energy and the appearance of clathrate-like structures in solutions. *The Journal of Physical Chemistry A*, 119(44):10806–10815, 2015.

[366] Ming-Liang Tan, Benjamin T Miller, Jerez Te, Joseph R Cendagorta, Bernard R Brooks, and Toshiko Ichiye. Hydrophobic hydration and the anomalous partial

molar volumes in ethanol-water mixtures. *The Journal of Chemical Physics*, 142(6):064501, 2015.

[367] Ritaban Halder and Biman Jana. Unravelling the composition-dependent anomalies of pair hydrophobicity in water–ethanol binary mixtures. *The Journal of Physical Chemistry B*, 122(26):6801–6809, 2018.

[368] Jürgen Rehm, Omer SM Hasan, Sandra E Black, Kevin D Shield, and Michaël Schwarzinger. Alcohol use and dementia: a systematic scoping review. *Alzheimer's Research & Therapy*, 11(1):1–11, 2019.

[369] Shuju Bai, Wenbo Wang, Zhiwei Zhang, Mengyao Li, Zehan Chen, Jiuqiao Wang, Yanlin Zhao, Lu An, Yuxiang Wang, Shu Xing, et al. Ethanol alleviates amyloid-β-induced toxicity in an alzheimers disease model of caenorhabiditis elegans. *Frontiers in Aging Neuroscience*, 13, 2021.

[370] David Ormeno, Fernando Romero, Julio López-Fenner, Ándres Avila, Ataulfo Martínez-Torres, and Jorge Parodi. Ethanol reduces amyloid aggregation in vitro and prevents toxicity in cell lines. *Archives of Medical Research*, 44(1):1–7, 2013.

[371] Atsuya Matsui, Jean-Pierre Bellier, Takeshi Kanai, Hiroki Satooka, Akio Nakanishi, Tsukasa Terada, Takafumi Ishibe, Yoshiaki Nakamura, Hiroyasu Taguchi, Nobuyasu Naruse, et al. The effect of ethanol on disassembly of amyloid-β1-42 pentamer revealed by atomic force microscopy and gel electrophoresis. *International Journal of Molecular Sciences*, 23(2):889, 2022.

[372] Sophia Jordens, Jozef Adamcik, Idit Amar-Yuli, and Raffaele Mezzenga. Disassembly and reassembly of amyloid fibrils in water- ethanol mixtures. *Biomacromolecules*, 12(1):187–193, 2011.

[373] Dzhigangir A Faizullin, Tatiana A Konnova, Thomas Haertlé, and Yuriy F Zuev. Secondary structure and colloidal stability of beta-casein in microheterogeneous water-ethanol solutions. *Food Hydrocolloids*, 63:349–355, 2017.

[374] Partha Pyne, Debasish Das Mahanta, Himanshu Gohil, SS Prabhu, and Rajib Kumar Mitra. Correlating solvation with conformational pathways of proteins in alcohol–water mixtures: a thz spectroscopic insight. *Physical Chemistry Chemical Physics*, 23(32):17536–17544, 2021.

[375] Jil J Kayser, Philipp Arnold, Anja Steffen-Heins, Karin Schwarz, and Julia K Keppler. Functional ethanol-induced fibrils: Influence of solvents and temperature on amyloid-like aggregation of beta-lactoglobulin. *Journal of Food Engineering*, 270:109764, 2020.

[376] Christian Blouin, Alex Safatli, and Jose Sergio Hleap. Pdbnet module.

[377] Alexander W Schüttelkopf and Daan MF Van Aalten. Prodrg: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallographica Section D: Biological Crystallography*, 60(8):1355–1363, 2004.

[378] Justin A Lemkul, William J Allen, and David R Bevan. Practical considerations for building gromos-compatible small-molecule topologies. *Journal of Chemical Information and Modeling*, 50(12):2221–2235, 2010.

[379] Lei Gu, Joyce Tran, Lin Jiang, and Zhefeng Guo. A new structural model of alzheimers aβ42 fibrils based on electron paramagnetic resonance data and rosetta modeling. *Journal of Structural Biology*, 194(1):61–67, 2016.

[380] Kristoffer Brännström, Tohidul Islam, Linda Sandblad, and Anders Olofsson. The role of histidines in amyloid β fibril assembly. *Febs Letters*, 591(8):1167–1175, 2017.

[381] Jan Zacharias and Ernst-Walter Knapp. Protein secondary structure classification revisited: processing dssp information with pssc. *Journal of Chemical Information and Modeling*, 54(7):2166–2179, 2014.

[382] Richard B Cooley, Daniel J Arp, and P Andrew Karplus. Evolutionary origin of a secondary structure: π-helices as cryptic but widespread insertional variations of α-helices that enhance protein functionality. *Journal of Molecular Biology*, 404(2):232–246, 2010.

[383] Werner F Kuhs and Werner Kuhs. *Physics and Chemistry of Ice*. Number 311. Royal Society of Chemistry, 2007.

[384] Felix Franks. Water a comprehensive treatise: volume 4: aqueous solutions of amphiphiles and macromolecules. 2013.

[385] Saman Alavi, Ryo Ohmura, and John A Ripmeester. A molecular dynamics study of ethanol–water hydrogen bonding in binary structure i clathrate hydrate with co2. *The Journal of Chemical Physics*, 134(5):054702, 2011.

[386] William C Wimley and Stephen H White. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nature Structural Biology*, 3(10):842–848, 1996.

[387] Nicklas Österlund, Yashraj S Kulkarni, Agata D Misiaszek, Cecilia Wallin, Dennis M Krüger, Qinghua Liao, Farshid Mashayekhy Rad, Jüri Jarvet, Birgit Strodel, Sebastian KTS Wärmländer, et al. Amyloid-β peptide interactions with amphiphilic surfactants: Electrostatic and hydrophobic effects. *ACS Chemical Neuroscience*, 9(7):1680–1692, 2018.

[388] Matthew A Barrett, Richard J Alsop, Thomas Hauß, and Maikel C Rheinstädter. The position of aβ22-40 and aβ1-42 in anionic lipid membranes containing cholesterol. *Membranes*, 5(4):824–843, 2015.

[389] Nicoletta Liguori, Paul S Nerenberg, and Teresa Head-Gordon. Embedding aβ42 in heterogeneous membranes depends on cholesterol asymmetries. *Biophysical Journal*, 105(4):899–910, 2013.

[390] W Kabsch and C Sander. Dssp: definition of secondary structure of proteins given a set of 3d coordinates. *Biopolymers*, 22:2577–2637, 1983.

[391] Guanghong Wei and Joan-Emma Shea. Effects of solvent on the structure of the alzheimer amyloid-β (25–35) peptide. *Biophysical Journal*, 91(5):1638–1647, 2006.

[392] Maria Grazia Ortore, Paolo Mariani, Flavio Carsughi, Stefania Cinelli, Giuseppe Onori, José Teixeira, and Francesco Spinozzi. Preferential solvation of lysozyme in water/ethanol mixtures. *The Journal of Chemical Physics*, 135(24):12B620, 2011.

[393] Charles H Davis and Max L Berkowitz. Structure of the amyloid-β (1- 42) monomer absorbed to model phospholipid bilayers: a molecular dynamics study. *The Journal of Physical Chemistry B*, 113(43):14480–14486, 2009.

[394] Hasna Ahyayauch, Aritz B Garcia-Arribas, Massimo E Masserini, Sergio Pantano, Felix M Goni, and Alicia Alonso. β-amyloid (1–42) peptide adsorbs but does not insert into ganglioside-containing phospholipid membranes in the liquid-disordered state: modelling and experimental studies. *International Journal of Biological Macromolecules*, 164:2651–2658, 2020.

[395] Yan Lu, Xiao-Feng Shi, Freddie R Salsbury Jr, and Philippe Derreumaux. Influence of electric field on the amyloid-β (29-42) peptides embedded in a membrane bilayer. *The Journal of Chemical Physics*, 148(4):045105, 2018.

[396] Chetan Poojari, Andreas Kukol, and Birgit Strodel. How the amyloid-β peptide and membranes affect each other: an extensive simulation study. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1828(2):327–339, 2013.

[397] Thomas L Williams and Louise C Serpell. Membrane and surface interactions of alzheimers aβ peptide–insights into the mechanism of cytotoxicity. *The FEBS Journal*, 278(20):3905–3917, 2011.

[398] Hasna Ahyayauch, Michal Raab, Jon V Busto, Nagore Andraka, José-Luis R Arrondo, Massimo Masserini, Igor Tvaroska, and Félix M Goñi. Binding of β-amyloid (1–42) peptide to negatively charged phospholipid membranes in the liquid-ordered state: modeling and experimental studies. *Biophysical Journal*, 103(3):453–463, 2012.

[399] Jotam Bergfreund, Pascal Bertsch, and Peter Fischer. Adsorption of proteins to fluid interfaces: Role of the hydrophobic subphase. *Journal of Colloid and Interface Science*, 584:411–417, 2021.

[400] Raimon Sabaté and Joan Estelrich. Evidence of the existence of micelles in the fibrillogenesis of β-amyloid peptide. *The Journal of Physical Chemistry B*, 109(21):11027–11032, 2005.

[401] William P Russ and Donald M Engelman. The gxxxg motif: a framework for transmembrane helix-helix association. *Journal of Molecular Biology*, 296(3):911–919, 2000.

[402] Monica M Pallitto, Jyothi Ghanta, Peter Heinzelman, Laura L Kiessling, and Regina M Murphy. Recognition sequence design for peptidyl modulators of β-amyloid aggregation and toxicity. *Biochemistry*, 38(12):3570–3578, 1999.

[403] Charles G Glabe and Rakez Kayed. Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis. *Neurology*, 66(1 suppl 1):S74–S78, 2006.

[404] Ratnesh Lal, Hai Lin, and Arjan P Quist. Amyloid beta ion channel: 3d structure and relevance to amyloid channel paradigm. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 1768(8):1966–1975, 2007.

[405] Najeeb A Shirwany, Daniel Payette, Jun Xie, and Qing Guo. The amyloid beta ion channel hypothesis of alzheimers disease. *Neuropsychiatric Disease and Treatment*, 3(5):597, 2007.

[406] Matthew J Voelker, Bogdan Barz, and Brigita Urbanc. Fully atomistic aβ40 and aβ42 oligomers in water: Observation of porelike conformations. *Journal of Chemical Theory and Computation*, 13(9):4567–4583, 2017.

[407] Hoang Linh Nguyen, Huynh Quang Linh, Paolo Matteini, Giovanni La Penna, and Mai Suan Li. Emergence of barrel motif in amyloid-β trimer: A computational study. *The Journal of Physical Chemistry B*, 124(47):10617–10631, 2020.

[408] Lukáš Pravda, David Sehnal, Dominik Toušek, Veronika Navrátilová, Václav Bazgier, Karel Berka, Radka Svobodová Vařeková, Jaroslav Koča, and Michal Otyepka. Moleonline: a web-based tool for analyzing channels, tunnels and pores (2018 update). *Nucleic Acids Research*, 46(W1):W368–W373, 2018.

[409] Hyunbum Jang, Fernando Teran Arce, Srinivasan Ramachandran, Bruce L Kagan, Ratnesh Lal, and Ruth Nussinov. Disordered amyloidogenic peptides may insert into the membrane and assemble into common cyclic structural motifs. *Chemical Society Reviews*, 43(19):6750–6764, 2014.

[410] Hyunbum Jang, Fernando Teran Arce, Srinivasan Ramachandran, Ricardo Capone, Ratnesh Lal, and Ruth Nussinov. β-barrel topology of alzheimer's β-amyloid ion channels. *Journal of Molecular Biology*, 404(5):917–934, 2010.

[411] Thanh D Do, Nichole E LaPointe, Rebecca Nelson, Pascal Krotee, Eric Y Hayden, Brittany Ulrich, Sarah Quan, Stuart C Feinstein, David B Teplow, David Eisenberg, et al. Amyloid β-protein c-terminal fragments: Formation of cylindrins and β-barrels. *Journal of the American Chemical Society*, 138(2):549–557, 2016.

[412] Wenhui Xi, Elliott K Vanderford, and Ulrich HE Hansmann. Out-of-register aβ42 assemblies as models for neurotoxic oligomers and fibrils. *Journal of Chemical Theory and Computation*, 14(2):1099–1110, 2018.

[413] Andreia Cuco, Ana Paula Serro, José Paulo Farinha, Benilde Saramago, and Amélia Gonçalves da Silva. Interaction of the alzheimer aβ (25–35) peptide segment with model membranes. *Colloids and Surfaces B: Biointerfaces*, 141:10–18, 2016.