Lattice model for amyloid peptides: OPEP force field parametrization and applications to the nucleus size of Alzheimer’s peptide

Thèse dirigée par: Pr. Philippe DERREUMAUX and Dr. Phuong NGUYEN

Soutenue le 20, Septembre, 2016
Acknowledgements

First, I would like to express my sincere gratitude to my supervisors Prof. Philippe DERREUMAUX and Dr. Phuong NGUYEN for the continuous support of my Ph.D studies and related research, for their patience, motivation, and immense knowledge. They always gave me constructive feedbacks throughout my PhD, particularly during those times where I felt defeated by my research. Their guidance helped me in all the time of research, and writing of this thesis. I could not have imagined having a better supervisor and mentor for my Ph.D studies.

Second, I would like to thank all members of my thesis jury: Prof. Tap HA-DUONG from university Paris Sud, Dr. Human REZAEI from INRA for spending their time to read and evaluated my thesis, Dr. Pierre TUFFERY from university Paris 7, Dr. Yasmine CHEBARO from Strasbourg for their insightful comments and encouragement, but also for the hard question which incited me to widen my research from various perspectives.

My sincere thanks also go to our collaborator from the institute of Physics, Polish Academy of Sciences, Warsaw, Poland, Pr. Mai Suan Li for introducing and teaching me about protein lattice model and his useful discussion during my PhD period. Without his precious support it would not be possible to conduct this research.

It was a grateful opportunity to work at the Laboratory of Theoretical Biochemistry where I learned much during my time here, and met many great researchers and students. I would like to thank all people in the lab, especially my fellow lab-mates including Marina Katava, Dario De Vecchis, Mara Chircotto, Aixiao Li, Gilles Lamothe for the stimulating discussions, and for all the time we have shared in the last three years. Many thanks go to Victoria, secretary of our lab, who always support me in the administrations and Geoffrey Letessier who helps me information technologies. My project could not be finish without their help.
Last but not the least, I would like to thank my beloved family: my parents, my sister, my brother-in-law and my nephews for their unconditional love, for supporting me spiritually throughout writing this thesis and my life in general from oversea. They always give me inspirations and encouragements into my life. Without them, I will be unable to deal with all of the storms which every living day gives me.
Abstract

The neurodegenerative Alzheimer’s disease (AD) is affecting more than 40 million people worldwide and is linked to the aggregation of the amyloid-β proteins of 40/42 amino acids. Despite many experimental and theoretical studies, the mechanism by which amyloid fibrils form and the 3D structures of the early toxic species in aqueous solution remain to be determined. In this thesis, I studied the structures of the early formed oligomers of the amyloid-β peptide and the critical nucleus size of two amyloid-β peptide fragments using either coarse-grained or all-atom simulations. First, at the coarse-grained level, I developed a lattice model for amyloid protein, which allows us to study the nucleus sizes of two experimentally well-characterized peptide fragments $A\beta_{16−22}$ and $A\beta_{37−42}$ of the Alzheimer’s peptide $A\beta_{1−42}$. After presenting a comprehensive OPEP force-field parameterization using an on-lattice protein model with Monte Carlo simulations and atomistic simulations, I determined the nucleus sizes of the two fragments. My results show that the nucleation number is 10 chains for $A\beta_{16−22}$ and larger than 20 chains for $A\beta_{37−42}$. This knowledge is important to help design more effective drugs against AD. Second, I investigated the structures of the dimer $A\beta_{1−40}$ using extensive atomistic REMD simulations. This study provides insights into the equilibrium structure of the $A\beta_{1−40}$ dimer in aqueous solution, opening a new avenue for a comprehensive understanding of the impact of pathogenic and protective mutations in early-stage Alzheimer’s disease on a molecular level.

Résumé

La maladie d’Alzheimer touche plus de 40 millions de personnes dans le monde et résulte de l’agrégation du peptide beta-amyloïde de 40/42 résidus. En dépit de nombreuses études expérimentales et théoriques, le mécanisme de formation des fibres et des plaques n’est pas élucidé, et les structures des espèces les plus toxiques restent à déterminer. Dans cette thèse, je me suis intéressée à deux aspects. (1) La détermination du noyau de nucléation ($N^*$) de deux
fragments $A\beta_{16-22}$ and $A\beta_{37-42}$. Mon approche consiste à déterminer les paramètres OPEP du dimère $A\beta_{16-22}$ en comparant des simulations Monte Carlo sur réseau et des dynamiques moléculaires atomiques par échange de répliques. Les paramètres fonctionnant aussi sur le trimère $A\beta_{16-22}$ et les dimères et trimères $A\beta_{37-42}$, j’ai étudié la surface d’énergie libre des décamères et mes simulations montrent que $N^*$ est de 10 chaînes pour $A\beta_{16-22}$ et est supérieure à 20 chaînes pour $A\beta_{37-42}$. (2) J’ai ensuite étudié les structures du dimère $A\beta_{1-40}$ par simulations de dynamique moléculaire atomistique par échanges de répliques. Cette étude, qui fournit les conformations d’équilibre du dimère $A\beta_{1-40}$ en solution aqueuse, ouvre des perspectives pour une compréhension de l’impact des mutations pathogènes et protectrices au niveau moléculaire.
## Contents

1 Introduction .......................................................... 1
   1.1 Protein misfolding, aggregation and diseases .................... 1
   1.2 Alzheimer’s disease and the $\beta$ - amyloid peptide .......... 3
   1.3 Amyloid Fibril structures ....................................... 6
   1.4 Structures of the $A\beta_{1-40}/A\beta_{1-42}$ monomers and dimers .. 8
      1.4.1 From experiments ........................................... 8
      1.4.2 From simulations ........................................... 8
   1.5 Mechanisms for amyloid fibril formation and nucleation numbers . 11
   1.6 Goals of the thesis ............................................. 14

2 Methodology .......................................................... 16
   2.1 The protein model ................................................ 16
      2.1.1 All - atom model ........................................... 16
      2.1.2 Coarse - grained models ................................... 17
         2.1.2.1 The OPEP force field ................................. 17
         2.1.2.2 Lattice model ......................................... 19
   2.2 Molecular Dynamics (MD) simulations ............................ 21
      2.2.1 Thermostat .................................................. 22
      2.2.2 Barostat .................................................... 23
      2.2.3 Periodic Boundary condition (PBC) .......................... 23
   2.3 Monte Carlo (MC) simulations .................................... 23
   2.4 Enhance sampling - replica exchange MD and MC simulations ...... 24
   2.5 Data analysis .................................................... 25
      2.5.1 Specific heat ................................................ 25
      2.5.2 Common quantities ......................................... 26
         2.5.2.1 Root mean square deviation (RMSD) ................... 26
         2.5.2.2 Radius of gyration ($R_g$) ............................ 26
         2.5.2.3 Number of H-bonds .................................. 26
2.5.2.4 Number of Contacts ........................................ 26
2.5.2.5 Secondary structure ......................................... 27
2.5.2.6 Populations of antiparallel and parallel conformations . 27
2.5.2.7 Order parameter ............................................. 27
2.5.2.8 Binding free energy ......................................... 27
2.5.3 Clustering methods ........................................... 28
2.5.4 Principal components analysis ................................ 28
2.5.5 Identification of the structure of oligomers .................. 28
2.5.5.1 Intra - molecular state .................................... 29
2.5.5.2 Inter - molecular state .................................... 29
2.5.5.3 Overall state ............................................... 30

3 Lattice model for amyloid peptides: OPEP force field parametrization and applications to the nucleus size of Alzheimer’s peptides 31
3.1 Introduction ....................................................... 33
3.2 Systems and force field parametrization strategy ............. 34
3.3 Simulation details ................................................ 36
  3.3.1 Lattice simulations .......................................... 36
  3.3.2 All-atom simulations ........................................ 38
3.4 Results ............................................................ 39
  3.4.1 Sampling convergence ........................................ 39
  3.4.2 The need for the force field parametrization .............. 42
  3.4.3 Force field parametrization for \( A\beta_{16-22} \) dimer .......... 44
  3.4.4 Force field validation for \( A\beta_{16-22} \) trimer, \( A\beta_{37-42} \) dimer, and \( A\beta_{37-42} \) trimer .......... 47
  3.4.5 Applications: Free energy landscape of \( A\beta_{16-22} \) and \( A\beta_{37-42} \) decamer .............. 51
  3.4.6 Nucleus size for \( A\beta_{16-22} \) and \( A\beta_{37-42} \) peptides .......... 54
3.5 Discussions ...................................................... 55
3.6 Conclusions ...................................................... 60

4 Results from all-atom model: Structures of \( A\beta_{1-40} \) dimer Wild-Type 62
4.1 Introduction ....................................................... 64
4.2 Simulation details ................................................ 65
4.3 Data analysis ...................................................... 67
  4.3.1 Secondary structure .......................................... 67
  4.3.2 \( \beta \)-hairpin ................................................. 67
5 Coarse-grained and all-atom simulations towards the early and late steps of amyloid fibril formation

5.1 Paper

Conclusion

Bibliography
Chapter 1

Introduction

1.1 Protein misfolding, aggregation and diseases

Proteins, which convert into well-defined folded compact structures, called native states, are complex, yet elegant, machines fine-tuned by evolution to properly fulfill a variety of tasks in the crowded cellular environment. Yet, with aging, some proteins assemble into harmful amyloid aggregates associated with neurodegenerative diseases, such as Alzheimer’s disease (AD), which presents a complex and costly challenge to our society. It has been shown that amyloid formation depends on a) environmental factors which include concentration of protein, and type of solvent, salt, metal ions, pH of medium, temperature and pressure and b) sequence properties involving hydrophobicity, polarity and β-sheet secondary structure content [1]. It has even been postulated that amyloid formation is a generic property of any polypeptide [2].

Since the first reported case of brain amyloid plaque formation in 1854 [3], the accumulation of abnormally folded protein deposits has been recognized as a key histopathological feature of approximately 50 debilitating human diseases with a multitude of disparate symptoms (Table 1.1). All these diseases are associated with the misfolding of normally soluble, functional peptides and proteins of different amino acid sequences and their subsequent conversions into amyloid fibrils [4]. The majority of these amyloid-derived diseases are essentially sporadic (i.e., do not result from amino acid mutations) [5], and are associated with extracellular amyloid deposits or intracellular inclusions with amyloid characteristics.

Among all neurodegenerative diseases, Alzheimer’s disease (AD) has recently been labelled as ”the twenty-first century plague”, since it is the most common cause of dementia worldwide, accounting for 60-80% of all dementia cases [6, 7]. Age is the major risk factor of AD, and AD is the 6th leading cause of death in the US [7]. According to a recent report [6], there are more than 46.8 million people worldwide affected with AD in 2015.

According to a recent report [6], there are more than 46.8 million people worldwide affected with AD in 2015.
<table>
<thead>
<tr>
<th>Disease</th>
<th>Aggregating protein or peptide</th>
<th>Polypeptide length (number of residues)</th>
<th>Structure of protein or peptide</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Neurodegenerative diseases</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alzheimer’s disease</td>
<td>Amyloid-β peptide</td>
<td>37 - 43</td>
<td>Intrinsically disordered</td>
</tr>
<tr>
<td>Spongiform encephalopathies</td>
<td>Prion protein or its fragments</td>
<td>230</td>
<td>Intrinsically disordered and α-helical</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>α-synuclein</td>
<td>140</td>
<td>Intrinsically disordered</td>
</tr>
<tr>
<td>Amyotrophic lateral sclerosis</td>
<td>Superoxide dismutase 1</td>
<td>153</td>
<td>β-sheet and Ig-like</td>
</tr>
<tr>
<td>Huntington’s disease</td>
<td>Huntingtin fragments</td>
<td>Variable</td>
<td>Mostly intrinsically disordered</td>
</tr>
<tr>
<td>Familial amyloidotic polyneuropathy</td>
<td>Transthyretin mutants</td>
<td>127</td>
<td>β-sheet</td>
</tr>
<tr>
<td><strong>Non-neuropathic systemic amyloidosis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amyloid light chain (AL) amyloidosis</td>
<td>Immunoglobulin (Ig) light chains or its fragments</td>
<td>90</td>
<td>β-sheet and Ig-like</td>
</tr>
<tr>
<td>AmyloidA(AA) amyloidosis</td>
<td>Serum amyloid A1 protein fragments</td>
<td>76-104</td>
<td>α-helical and unknown fold</td>
</tr>
<tr>
<td>Senile systemic amyloidosis</td>
<td>Wild-type transthyretin</td>
<td>127</td>
<td>β-sheet</td>
</tr>
<tr>
<td>Haemodialysis-related amyloidosis</td>
<td>β-microglobulin</td>
<td>99</td>
<td>β-sheet and Ig-like</td>
</tr>
<tr>
<td>Lysozyme amyloidosis</td>
<td>Lysozyme mutants</td>
<td>130</td>
<td>α-helical and β-sheet</td>
</tr>
<tr>
<td><strong>Non-neuropathic localized amyloidosis</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Apolipoprotein A1(Apo A-1)amyloidosis</td>
<td>ApoA-1 fragments</td>
<td>80-93</td>
<td>Intrinsically disordered</td>
</tr>
<tr>
<td>TypeII diabetes</td>
<td>Amylin</td>
<td>37</td>
<td>Intrinsically disordered</td>
</tr>
<tr>
<td>Injection-localized amyloidosis</td>
<td>Insulin</td>
<td>21 and 30</td>
<td>α-helical and insulin-like</td>
</tr>
</tbody>
</table>

Table 1.1: A selection of human diseases associated with extracellular amyloid deposits or intracellular inclusions with amyloid-like characteristics. See Ref.[4, 5] for a full description of the approximately 50 human protein misfolding diseases and their associated proteins.
An estimation in 2015 shows that there will be over 9.9 million new cases of dementia every year, implying one new case every 3 seconds. The number of Alzheimer’s patients will be doubled every 20 years, reaching 74.7 millions in 2030 and 131.5 millions in 2050 (Fig 1.1). The new estimates are 12-13% higher than those made for the World Alzheimer report 2009. In France, more than 1,175,000 people are suffering with Alzheimers disease, and more than 200,000 people are diagnosed with Alzheimers or another dementia each year [8].

Alzheimer’s disease has a huge economic impact [6]. The global costs of dementia have increased from US$ 604 billion in 2010 to US $ 818 billion in 2015. If dementia care were a country, it would be the worlds 18th largest economy, more than the market values of companies such as Apple (US$ 742 billion), Google (US$ 368 billions) and Exxon (US$ 357 billions). The enormous medical implications of AD have motivated intense research efforts for more than twenty years to understand amyloid fibril formation, the origin of toxicity so as to design more efficient drugs [5, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18].

1.2 Alzheimer’s disease and the β - amyloid peptide

Alzheimer’s disease (AD), a progressive irreversible neurodegenerative disorder with loss of cortical and subcortical neurons and marked atrophy, is characterized by the intracellular formation and accumulation of neurofibrillar tangles formed from filaments of microtubue-associated highly phosphorylated tau proteins and the deposition of insoluble extracellular amyloid plaques made of the amyloid-beta peptide [17, 19]. AD progressively affects
Figure 1.2: Metabolic processing of amyloid precursor protein (APP) by secretases leads to Aβ-peptides of varying length. The cleavage sites for α-, β-, and γ-secretase are noted. Known familial mutations are clustered in regions of the APP sequence that correspond to the N- or C-terminal regions of the Aβ protein. A third critical cluster of mutations is found on the N-terminal side of the VGSN turn region, a structural motif observed in the Aβ-protein in aqueous solution, membrane-mimicking solutions, micellular environments, and solid-state fibrils [24].

normal brain functions such as memory, judgement and cognition and result in the failure of important cellular processes [20]. The major constituents of the extracellular plaques consist of amyloid β (Aβ) peptides of 39-43 amino acids [18]. Aβ peptide results from the proteolytic cleavages of the transmembrane amyloid precursor protein (APP), located in chromosome 21, in the lipid-rich microdomains (lipid rafts) of endosomes and the plasma membrane [21, 22] by the β and γ-secretases [23]. There are many familial AD-associated mutations in Aβ, except at its C-terminus spanning residues 30-42 (Figure 1.2).

In human brain, Aβ appears in two predominant forms with different COOH-termini, Aβ1-40 and Aβ1-42. The Aβ1-42 wild-type (WT) sequence is DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIIGLMVGVVIA and has therefore two additional IA residues compared to Aβ1-40. Many studies have demonstrated that Aβ1-42 forms insoluble amyloid fibrils more rapidly than Aβ1-40 and has a higher toxicity [25, 26]. Increases in the ratio of Aβ1-42 production relative to Aβ1-40 were also measured in the plasma of patients diagnosed with FAD (familial Alzheimer’s disease) [27] and observed in transgenic mice mod-
els [28]. Finally, many truncated variants, such as $A\beta_{1-30}$ and $A\beta_{1-26}$, and $A\beta$ peptides with proteolytic removal of D1 and A2 and subsequent cyclization of E3 to a pyroglutamate, have been detected by mass spectrometry in human AD brains. Despite continuous debate, there is strong evidence that an imbalance between production and clearance of $A\beta_{1-40}/A\beta_{1-42}$ and related $A\beta_{1-40}$ proteins (either truncated or post-translational modified) plays a key role in initiating AD.

There are five drugs currently available for AD, however, they are only effective for 6-12 months and for half of the patients with milder forms of Alzheimer’s. Though targeting BACE1, modulating the response of the innate immune system, interfering on apolipoprotein E4 and other components of cholesterol metabolism, and regulating endosomal vesicle recycling are possible avenues, controlling $A\beta$ self-assembly with inhibitors is considered as one of the most promising solutions to delay the onset or stop the progression of AD [29]. Despite many in vitro and in vivo studies, drug after drug have failed to slow down the progression of Alzheimer’s disease in human trials [18, 30]. The challenge arises first from the intrinsic disorder structure nature of the human wild-type (WT) $A\beta$ monomer in aqueous solution. $A\beta_{1-42}$ WT sequence has two hydrophobic patches L17-A21 (CHC) and A30-A42 (C-terminus) and two hydrophilic patches E22-G29 (loop region) and D1-K16 (N-terminus). The challenge also comes from the lack of high-resolution structures and formation/dissociation rates of the low molecular weight $A\beta_{1-40}/A\beta_{1-42}$ oligomers, including dimers, which are believed to be the most critical players in the pathology, and for these oligomers, we have at hand low-resolution structural data. Also the experimental sigmoidal kinetics of amyloid formation, which is the result of a linear combination of microscopic reactions involving primary classical and secondary nucleation (fragmentation and surface-dependent lateral) processes, is very sensitive to the experimental conditions and the sequences, with some mutations enhancing or reducing fibrillogenesis and toxicity, and overall we know little on the topology and size of the primary nucleus. The complexity also comes from the heterogeneity of the conformations of each populated oligomer, their high propensity to aggregate and the sensitivity of the process to pH, temperature, concentration, the inherent flexibility, so the mechanisms leading to toxicity and amyloid fibril formation are largely unknown. Finally, while dimers, trimers, and 12-mers ($A\beta^*56$) [31] are considered as the most critical players the pathology of AD [32], larger aggregates and fragmentation events are toxic as well. In what follows, I review what is known about the final $A\beta$ fibril products and the early $A\beta_{1-40}/A\beta_{1-42}$ monomer and dimer structures from the most recent experiments and simulations. Soluble $A\beta$ dimers are the smallest toxic species in AD [31], and when they are isolated from Alzheimer cortex, they directly induce tau hyperphosphorylation and neuritic degeneration [33].
1.3 Amyloid Fibril structures

A wide range of experimental methods is currently being used or developed to improve our understanding of the amyloid oligomers and fibril formation [5, 13, 18] such as solid state nuclear magnetic resonance (ss-NMR), circular dichroism (CD), X-ray crystallography, ion mobility mass spectrometry (IM-MS), electron microscopy (EM), cryo-electron microscopy (cryoEM), transmission electron microscopy (TEM), atomic force microscopy (AFM), Fourier transform infrared spectroscopy (FTIR), X-rays diffraction, dynamics light scattering (DLS), electron paramagnetic resonance (EPR), hyperfine sublevel correlation (HYSCORE), and electron-nuclear doubled resonance (ENDOR) [34, 35, 36, 37].

There is no single experimental technique able to fully provide structural information of Aβ fibrils due to their insolubilities. Thus, our structural knowledge of Aβ fibrils is obtained from the combination of various experimental techniques. While fibril dimensions (nanometer length scale) have been investigated by electron microscopy (EM) and atomic force microscopy (AFM) [38, 39, 40], fibril mass has been measured by scanning transmission electron microscopy (TEM), and more recently, tilted beam TEM [41]. Secondary structure (mostly β-strand) can be detected by CD and FTIR [42, 43]. The cross-β architecture, in which Aβ molecules assemble into β-sheets with β-strands oriented perpendicular to the long fibril’s axis, has been found in X-ray fiber diffraction studies [44, 45, 46, 47]. The β-sheet structure also has been confirmed by the binding of β-sheet specific dyes such as thioflavin-T and Congo red [42].

One of the most important experimental techniques to study structures of fibrils is ss-NMR, which provide the most atomic-level detail of Aβ amyloid fibrils [48]. Complementary information has also been obtained from electron paramagnetic resonance (EPR) [49, 50, 51]. Exploiting ss-NMR technique, together with EM and EPR, Tycko et al. investigated the structures of the Aβ1−40 oligomers, protofibrils and fibrils [52, 53]. They showed that the structure of Aβ1−40 is polymorphic and displays the in-register parallel β-sheet organization as found by many other fibril-forming systems. They proposed molecular models for specific Aβ1−40 polymorphs with 2-fold and 3-fold symmetry about the long fibril axis (Fig 1.3). Interestingly, the fibrils formed by the D23N mutant of the Aβ1−40 peptide were found to form both straight fibrils with the parallel cross-β structure and more curved and shorter fibrils with an antiparallel cross-β structure, as shown in Figure 1.4. Interestingly, the antiparallel beta-sheet structure of the Aβ1−40 D23N fibril is metastable and can slowly convert to the parallel organization [54]. It turns out that the detailed molecular structures depend on the experimental conditions. A final comment is that, while fibrils of synthetic Aβ1−40/Aβ1−42 peptides display perfect U-shaped forms

6
Figure 1.3: Structural models for Aβ₁−₄₀ fibrils developed from solid state NMR and electron microscopy data. Models with 3-fold (a) and 2-fold (b) symmetry correspond to twisted fibrils and to individual filaments within striated ribbon fibrils, respectively [52].

Figure 1.4: Schematic comparison of antiparallel (left) and parallel (right) cross-β structures that can be constructed from similar U-shaped peptides. Figure has been taken from Ref. [52].
with beta-strands spanning the CHC and the C-terminus, and the N-terminus disordered, fibrils of AD-brain derived $\text{A}\beta_{1-40}$ peptides show deformed U-shaped states and, remarkably, the structure varies from one patient to another [52, 53]. Overall, this high degree of polymorphism, which arises from many physical factors and persists in vitro and in brain tissues, is correlated to different phenotypes and is rather bad news for drug design because one drug may be efficient for one patient but not for another.

1.4 Structures of the $\text{A}\beta_{1-40}/\text{A}\beta_{1-42}$ monomers and dimers

1.4.1 From experiments

There is strong NMR evidence that the $\text{A}\beta_{1-40}$ and $\text{A}\beta_{1-42}$ monomers are intrinsically disordered, with a $\beta$-strand character at the CHC core and the C-terminus and a propensity for turns at specific positions [55]. Using photo-induced cross-linking of unmodified proteins and various preparation methods, CD analysis suggested that the $\beta$-strand varies from 12% to 25%, while the $\alpha$-helix content varies from 3% to 9% at 295K and pH 7.5 on day 0, so for monomers and small oligomers [56]. Using $^{15}N$ spin NMR relaxation, Yan et al. reported that the side chain methyl groups at C-terminus are more rigid in $\text{A}\beta_{1-42}$ than in $\text{A}\beta_{1-40}$ [57]. Another NMR study conducted at much higher salt concentration showed the $\text{A}\beta_{1-40}$ monomer forms a 3$_{10}$ helix from H13 to D23 with the N- and C-termini disordered [58].

We do not have any NMR information on the dimers and very little information has emerged from other experimental studies because the $\text{A}\beta_{1-40}/\text{A}\beta_{1-42}$ dimer is aggregation-prone and exists in equilibrium with monomers and higher-order oligomers. However, two ion mobility mass spectrometry (IM-MS) studies provided a cross collision section (CCS) of 1142 and 1245 Å$^2$ for $\text{A}\beta_{1-40}$ dimer, and detected the coexistence of two alternative structural forms (compact and extended) in a pool of low-order $\text{A}\beta_{1-40}$ oligomers. Finally, Fourier Transform Infrared spectroscopy and solid-state NMR experiments indicate that small oligomers have some antiparallel beta-sheet structure rather than the parallel beta-structure as observed in amyloid fibrils [18, 59].

1.4.2 From simulations

All-atom replica exchange molecular dynamics (REMD) studies have shown that the $\text{A}\beta_{1-40}$ and $\text{A}\beta_{1-42}$ monomers can form transient beta hairpins using the OPLS-AA/L force field and the TIP3P water model. These metastable structures of $\text{A}\beta_{1-40}$ and $\text{A}\beta_{1-42}$ monomers are described in Figure 1.5 [60].
As seen in Figure 1.5, prevalent in Aβ1−42 (above panel) is a beta-hairpin defined by hydrogen bonds between V18/G33 and F20/I31, whereas Aβ1−40 (below panel) has a dominant bonding register specified by hydrogen bonds between V18/M35 and F20/G33, presenting a shift of two residues for the C-terminal partner. Interestingly, the predicted β-hairpin for the Aβ1−40 matched the register of the affibody-trapped monomeric Aβ1−40 hairpin proposed by Hoyer et al. [61].

Other REMD studies for Aβ using OPLS-AA/TIP3P and AMBER99sb/TIP4P-Ew, published by Sgourakis et al., predicted that Aβ was mostly flexible, but possessed some structured fragments, in particular, β-hairpins populating the C-terminus in Aβ1−42, but not Aβ1−40 [62, 63]. Using a different approach, called multiple-reservoir replica exchange (MRRE), Ball and coworkers carried out 100 ns simulations with AMBER99sb/TIP4P-Ew to determine structure of Aβ1−42. The results showed that Aβ1−42 was mostly disordered, with considerable α-helix in regions Y10-F19 and E22-N27 and little to not β structure [64]. Recently, Zhu et al. performed a 100ns/replica REMD simulation using AMBER99sb/TIP3P and reported that Aβ1−42 monomer forms contacts between L17-A21 and I31-V36 with a transient turn in the region D23-N26, consistent with a quasi-hairpin-like conformation [65].

Aβ monomers were also studied by REMD with the six-bead coarse-grained OPEP model in the implicit solvent [66] and Monte Carlo simulated annealing with the all-atom
PROFASI force field [67]. Results obtained from OPEP simulations showed that Aβ ensembles are mostly turn/coil, but possess substantial β-sheet propensity in the N-terminus [68]. The PROFASI simulation displayed that Aβ1−42 has high beta-strand structure all over the amino acids.

Overall, although the most recent all-atom studies in both explicit and implicit solvent have shown some consistent description of the features of the Aβ ensemble, most structural identifications of the Aβ1−40 and Aβ1−42 peptides remain highly divergent, and these variations may come from different simulation conditions, extend of sampling, or trajectory analysis.

Similar to monomers, the structures of Aβ1−40 and Aβ1−42 dimers have been studied by means of all-atom and coarse-grained simulations, but none were performed using all-atom models in explicit solvent and exhaustive sampling techniques. The most recent simulations include 900 ns/replica Hamiltonian replica exchange molecular dynamics (REMD) with the six-bead CG OPEP force field, all-atom REMD with a solvent-accessible surface area implicit solvent [69], discontinuous MD with a four-bead CG model [70], and all-atom MD simulations starting from CG DMD structures.

Using the six-bead CG OPEPv3 model with implicit solvent [66], the structures of Aβ1−40 and Aβ1−42 were determined by Hamiltonian replica exchange MD (HT-REMD) simulations [71]. The results of this study showed that both alloforms populate mostly turn/random coil conformations with a β-sheet propensity at the C-terminal region higher than that in monomers. While both dimers of Aβ1−40 and Aβ1−42 are characterized by CHC/CHC, CHC/C-terminal region, and C-terminal region/ C-terminal region interpeptide hydrophobic contacts, the Aβ1−42 dimer has a higher probability to form β-strand at the CHC and in the C-terminal region than the Aβ1−40. Furthermore, the free energy surface of the Aβ1−42 dimer is also broader and more complex than that of the Aβ1−40 dimer.

Extensive DMD simulations coupled to a four-bead CG model found that the Aβ dimer conformations are collapsed and disordered with a small content of beta-strands linked by loops and turns [70]. The Aβ1−42 dimer has a higher propensity of beta-sheets at the CHC and C-terminal region than the Aβ1−40 dimer. Aβ1−40 dimer formation is mainly driven by intermolecular interactions between the CHC regions, while the C-terminal region plays a significant role for Aβ1−42.

Finally, fifty nanosecond MD stability simulations with OPLS-AA/ TIP3P or SPC/E starting from the dominant DMD-obtained CG structures confirm the main DMD results and enable a precise analysis of secondary structures, salt bridges, and free energy landscapes [72]. Overall, the free energy landscape of Aβ1−42 is much more complex than that of Aβ1−40 [18].
1.5 Mechanisms for amyloid fibril formation and nucleation numbers

Recently, the classical and modern theories of linear growth phenomena have been used to study the role of growth and nucleation process to be defined through experimental in vitro time courses of amyloid formation [80, 81].

As seen in figure 1.6, amyloid formation can occur through different pathways: primary and secondary pathways. While the primary nucleation pathways results in the formation of aggregates from interactions solely between soluble monomers (Figure 1.6a), the secondary nucleation pathways involve either monomer-independent or secondary monomer-dependent secondary events. Monomer-independent secondary nucleation process, like fragmentation, generate new aggregates at a rate depending only on the level of aggregates, whereas monomer-dependent secondary pathways, like surface-catalysed nucleation, create aggregates at a rate that depends on the concentrations of both monomeric proteins and existing aggregates. The reaction constants are labeled $n_c$, $n_2$, and $k_2$, respectively, and the nucleation reaction orders with respect to the monomeric peptide are denoted $n_c$ and $n_2$ [80].
of effective growth rates or lag phases. The aggregation processes can be classified into three categories: (i) nucleation and fragmentation processes which increase the number of aggregates, (ii) growth process that leads to increase in the size of existing aggregates and (iii) degradation processes to decrease the size and/or the disappearance of aggregates.

Amyloid fibril formation is commonly monitored in vitro by optical methods such as fluorescence. Figure 1.7 displays the kinetics of amyloid formation from different sorts of nucleations, including the primary, the secondary and the fragmentation nucleations. The integrated rate laws, obtained from filament growth processes that appear under the action of various microscopic processes, generally follow the sigmoidal functions (Figure 1.7.b, blue line) that are characterized by a lag time $\tau_{lag}$ and a maximal growth rate ($r_{max}$, red line). The behavior that a lag phase occurs before a rapid growth phase, is a property of nucleated polymerization; the growth phase is followed by a plateau phase as the peptides are converted to fibrils. Additionally, the conversion of a protein molecule from its soluble state into the amyloid form can take place by nucleation, as well as by templating or seeding from existing fibrils. Figure 1.7.c reflects the comparison of integrated rate laws with experimental kinetic measurements and enables the relative importance of specific microscopic processes to be tested. In this example, the aggregation kinetics for increasing concentrations of amyloid-$\beta$ peptide are compared with the integrated rate laws which consists of primary nucleation (left), fragmentation (centre) and monomer-dependent secondary nucleation. The data show that the secondary nucleation is the majority under these conditions.

The nucleation of amyloid fibril is a process linked to the generation of nanoscale fibrils or protofilaments [61]. Below the nucleus size $N^*$, the nanofibril tends to dissolve rather than growth, and above the nucleus size, the systems can growth irreversibly into a mature fibril [82]. The critical nucleus cannot be determined directly as it exists only transiently, and determining the nucleation number remains a challenge.

Using fluorescence correlation spectroscopy, one recent experimental estimate of the size of the critical nucleus for $\text{A}\beta_{1-40}$/$\text{A}\beta_{1-42}$ peptides is $N^* > 29$. Using classical nucleation theory, the nucleus size may be considered as a turnover point of the free energy plotted as the function of the number of chains. Using two CG models, it was found that $N^* = 10 - 11$ for $\text{A}\beta_{1-40}$ [18]. Computationally, $N^*$ can also be estimated from the concentration dependence of the lag phase time [83] and it was found using a simplified model and Langevin dynamics simulations that $N^*$ can vary from 4 to 35 for linear peptides (i.e. for fibril states without U-shape), depending on the energy difference between the amyloid-competent and amyloid-protected minima [84]. Using atomistic classical nucleation theory,
Figure 1.7: Kinetic of amyloid formation. (a): different nucleation pathways; (b): Integrated rate laws obtained for filament growth processes that occur under the action of these various microscopic processes commonly take the form of sigmoidal functions (blue line), which are characterized by a lag time ($\tau_{\text{lag}}$) and a maximal growth rate ($r_{\text{max}}$; red line). (c): Comparison of such integrated rate laws with experimental kinetic measurements enables the relative importance of specific microscopic processes to be tested. In this example the aggregation kinetics for increasing concentrations (coloured lines) of the amyloid-$\beta$ peptide are compared with integrated rate laws that contain primary nucleation (left), fragmentation (centre) and monomer-dependent secondary nucleation (right). $n_c$ and $n_2$ indicate the sizes of the critical nuclei for primary and secondary nucleation, respectively; $k_n$ and $k_-$ indicate the rates of primary nucleation and fragmentation, respectively; and $m$ and $M$ denote the concentrations of the monomers and the fibril mass, correspondingly [4].
it was found that $N^*$ is 15 for $\text{A} \beta_{1-40}$ at a protein concentration of 120$\mu$M, but variation in the supersaturation of the phase can cause $N^*$ to increase to 50 [85].

1.6 Goals of the thesis

My thesis was aimed at two goals. While the experimental sigmoidal kinetics of amyloid formation can be described by a combination of microscopic reactions involving primary classical and secondary nucleation (fragmentation and surface-dependent lateral) processes, overall we know little on the topology and size of the primary nucleus, which is known to be sequence- and experimental condition-dependent. These factors range from salt and metal concentration, energy landscape of the monomer and population of the monomeric aggregation-prone state, shear flows to the supersaturation of the protein solution. In addition, using a mesoscopic model and a CG model, it has been shown that increasing the total side-chain hydrophobicity switches the fibrillization mechanism from one- to two-step nucleation, where in the one-step nucleation, the beta-sheet—enriched nucleus forms directly from the solution, and in the two-step nucleation, soluble monomers first assemble into disordered oligomers, which subsequently convert into a beta-sheet nucleus [86].

Clearly all-atom simulations in explicit solvent (even with metadynamics) are not appropriate to determine the nucleus size which can vary between 7 and 40 peptides. On the other hand, coarse-grained protein lattice models approximate atomistic details and keep the essential interactions. They are, therefore, suitable for capturing generic features of amyloid formation at low computational cost. However, as our aim is to study the critical nucleus sizes of two experimentally well-characterized peptide fragments $\text{A} \beta_{16-22}$ and $\text{A} \beta_{37-42}$ of the full length $\text{A} \beta_{1-42}$ Alzheimer's peptide, it is important that simulations with the lattice model reproduce all-atom simulations. To this end, we present a comprehensive force field parameterization based on the OPEP force field for an on-lattice protein model developed by Frenkel et al., which incorporates explicitly the formation of hydrogen bonds and directions of side-chains. Using the same parameters and REMC simulations, we show that the nucleation number is on the order of 10 chains for $\text{A} \beta_{16-22}$ and larger than 20 chains for $\text{A} \beta_{37-42}$ [87]. This knowledge has important implications to help design more efficient drugs against Alzheimer's disease [88, 89].

The second goal was to characterize the intrinsic disorder of the $\text{A} \beta_{1-40}$ dimer by means of extensive replica exchange molecular dynamics (REMD) simulations. We know that the inherent flexibility of $\text{A} \beta_{1-40}$ has thus far precluded its structural characterization by all
experimental methods. All previous simulations on the dimer were performed using all-atom or CG models and implicit solvent models. Attacking the dimer is justified because it is the smallest toxic species in Alzheimer’s disease [31], and when they are isolated from Alzheimer cortex, they directly induce Tau hyperphosphorylation, resulting neuritic degeneration [33]. Understanding the structural and dynamic properties of \( A\beta_{1-40} \) dimer is thus very important to design drugs, and can open a new avenue for a comprehensive understanding of the impact of pathogenic and protective mutations in early-stage Alzheimer disease at a molecular level [90].

The structure of my thesis is organized as follows. In chapter 2, I introduce the methodology and the techniques used to attack the two problems, including the coarse-grained protein lattice model and the parallel tempering Monte Carlo protocol and the all-atom replica exchange molecular dynamics simulations, as well as the data analysis tools for both lattice and all-atom cases. In chapter 3 and 4, I present the results of the OPEP coarse-grained simulations aimed at determining the nucleus sizes of two amyloid peptides, and of the simulations aimed at determined the structures of the \( A\beta_{1-40} \) dimer. Chapter 5 consists on a recent review on the early and late steps of amyloid fibril formation by means of coarse-grained and all-atom simulations. Finally, I summarize the main results and findings of my work and discuss on-going or possible perspectives.

Overall, this thesis has led to 3 publications: 1, Bogdan Tarus, Thanh T. Tran, Jessica Nasica-Labouze, Fabio Sterpone, Phuong H. Nguyen, and Philippe Derreumaux, "Structures of the Alzheimer’s Wild-Type \( A\beta_{1-40} \) Dimer from Atomistic Simulations", The Journal of Physical Chemistry B, 119, 10478-10487, 2015, DOI: 10.1021/acs.jpcb.5b05593 (July 2015).


3, Mara Chiricotto, Thanh Thuy Tran, Phuong H. Nguyen, Simone Melchionna, Fabio Sterpone, Philippe Derreumaux, "Coarse-grained and all-atom simulations towards the early and late steps of amyloid", Israel Journal of Chemistry, in press (June 2016) (M.C. and T.T. Tran contributing equally).
Chapter 2

Methodology

In this chapter, all methods and techniques used in this thesis are briefly introduced, including coarse-grained lattice and all-atom models, Molecular Dynamics (MD) and Monte Carlo simulations methods and the data analysis tools.

2.1 The protein model

2.1.1 All-atom model

In an all-atom model, the interactions between atoms are usually represented by an empirical potential energy of the form:

$$U(\vec{r}) = \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2$$

$$+ \sum_{\text{dihedrals}} V_n \left( \frac{1}{2} \left( 1 + \cos(n\chi - \delta_n) \right) \right)$$

$$+ \sum_{\text{impropers}} K_{\text{imp}} (\varphi - \varphi_0)^2$$

$$+ \sum_{\text{nonbonded}, i<j} \varepsilon_{ij} \left[ \left( \frac{r_{ij}^0}{r_{ij}} \right)^{12} - 2 \left( \frac{r_{ij}^0}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (2.1)$$

Here, the first two terms describe the short-range interactions, where the bonds and angles are represented by a simple harmonic expression [91, 92], and the dihedral energies are modeled by a Fourier expansion. The last sum accounts for the intermolecular or nonbonded interactions, where the van der Waals interaction is treated by a 6-12 potential and the electrostatic interactions are modeled by a Coulomb potential based on atom-centred point charges. For a discussion on various force-field parametrization procedures and development, see, for example Refs. [93, 94, 95, 96]. In this work, we used the AMBER-f99SB-ILDN [97] and CHARMM22* [98] force-fields.
2.1.2 Coarse-grained models

2.1.2.1 The OPEP force field

The OPEP (Optimized Potential for efficient protein structure Prediction) coarse-grained model represents each amino acid by six centres of force: the side-chain is represented by an unique bead located at the centre of mass of nonhydrogen atoms while atomic resolution is reserved for the backbone that includes N, HN, Cα, C and O atoms. An exception is Proline whose side-chain is represented by all heavy atoms. The disulfide (S-S) bonds can be treated as two non-bonded beads or described at an atomic level using local terms. The implicit solvent OPEP function is expressed as a sum of local (bond-lengths, bond angles, impropers torsion of the side chains and the amide bonds, backbone torsions), non-local and two-body and four-body hydrogen-bonding interactions:

$$ U = U_{\text{local}} + U_{\text{non-local}} + U_{\text{HB}}, \quad (2.2) $$

where the first term, $U_{\text{local}}$, indicates all the local interactions that are related to the molecular connectivity, the second term, $U_{\text{nonlocal}}$, covers all the non-bonded interaction terms and the last one, $U_{\text{HB}}$, accounts for the hydrogen-bond potential of backbone atoms. The local contributions are given by the following interactions:

$$ U_{\text{local}} = \omega_b \sum_{\text{bonds}} K_b (b - b_0)^2 + \omega_\theta \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2 $$

$$ + \omega_\phi \sum_{\text{torsions}} K_\phi (1 + \cos(n\phi - \phi_0)) + \omega_{\phi,\psi} (\sum_{\phi} U_{\phi} + \sum_{\psi} U_{\psi}), \quad (2.3) $$

where each contribution $\omega_x$ represents an optimized weighting factor. The first three terms are bond stretching, angle bending, and torsion energies. The last term adds harmonic biases reconfine the dihedrals within the Ramachandran limits [66]:

$$ U_{\phi} = K_{\phi\psi}(\phi - \phi_0)^2 $$

$$ U_{\psi} = K_{\phi\psi}(\psi - \psi_0)^2, \quad (2.4) $$

where $\phi_0 = \phi$ if the torsion is found in the interval $[\phi_{\text{lower}}, \phi_{\text{upper}}]$, otherwise $\phi_0 = \min(\phi - \phi_{\text{lower}}, \phi_{\text{upper}})$. The limits for the $\phi$ angles are $\phi_{\text{lower}} = -160^\circ$ and $\phi_{\text{upper}} = -60^\circ$ while for the $\psi$ are $\psi_{\text{lower}} = -60^\circ$ and $\psi_{\text{upper}} = 160^\circ$.

The non-bonded interactions consist of the terms:

$$ U_{\text{non-local}} = \omega_{1,4} \sum_{1,4} U_{\text{VdW}}^1 $$

$$ + \omega_{1,4} \sum_{M',M} U_{\text{VdW}}^1 + \sum_{M',C_a} U_{\text{VdW}}^1 + \sum_{M',S_c} U_{\text{VdW}}^1 + \sum_{C_a,S_c} U_{\text{VdW}}^1 $$

$$ + \omega_{C_a,C_a} \sum_{C_a,C_a} U_{\text{VdW}}^2 + \omega_{S_c,S_c} \sum_{S_c,S_c} U_{\text{VdW}}^2 $$

17
where $M'$ denotes all the backbone atoms except $C_\alpha$. The first short-range term accounts for all 1-4 interactions. The second block of terms contains long-range interactions with respect to sequence position involving main chain atoms $M'$ and mixed interactions ($M'$, $S_c$), ($M'$, $C_\alpha$), ($C_\alpha$, $S_c$) that share the same weighing factor $\omega_{1>4}$ in the functional form $U_{\text{VdW}}^1$. The last two terms stands for interactions involving ($C_\alpha$, $C_\alpha$) and ($S_c$, $S_c$) pairs.

The van der Walls (VdW) potential in OPEP has two distinct forms denote with $U_{\text{VdW}}^1$ and $U_{\text{VdW}}^2$ that depend on the centre of forces involved. The first type $U_{\text{VdW}}^1$ is a standard 12-6 potential:

$$U_{\text{VdW}}^1 = \varepsilon_{\text{ij}} \left( \frac{r_{\text{ij}}^6}{r_{\text{ij}}^{12}} - 2 \left( \frac{r_{\text{ij}}^6}{r_{\text{ij}}^{10}} \right)^6 \right), \quad (2.6)$$

where $r_{\text{ij}}$ is the distance between particles $i$ and $j$, $r_{\text{ij}}^0 = (r_{\text{ij}}^1 + r_{\text{ij}}^2)/2$ with $r_{\text{ij}}^1$ and $r_{\text{ij}}^2$ are the van der Waals radius of particles $i$ and $j$, and $\varepsilon_{\text{ij}}$ is the coupling constant. The second potential function $U_{\text{VdW}}^2$ reads as follows:

$$U_{\text{VdW}}^2 = U_{\text{AR}} - \varepsilon_{\text{ij}} \left( \frac{r_{\text{ij}}^0}{r_{\text{ij}}} \right)^8 H(-\varepsilon_{\text{ij}}) \quad (2.7)$$

where the repulsive final term holds for $\varepsilon_{\text{ij}} < 0$ while the attractive/repulsive interactions are described by the potential term:

$$U_{\text{AR}} = \varepsilon_{\text{ij}} \left[ \left( \frac{G(r_{\text{ij}}^0)}{r_{\text{ij}}} \right)^6 e^{-2r_{\text{ij}}} + A_0 \tanh \left[ 2(r_{\text{ij}} - r_{\text{ij}}^0 - 0.5) \right] - 1 \right] H(\varepsilon_{\text{ij}}), \quad (2.8)$$

where $G(r_{\text{ij}}^0)$ is determined by imposing the value of the potential at a specific location $r_{\text{ij}}^0$. The $H(\varepsilon)$ denotes the Heavyside function, and equal to 1 for $\varepsilon \geq 0$ and 0 otherwise. For $r_{\text{ij}} > r_{\text{ij}}^0$ the potential is controlled by the second part of eqn (2.8), which parameters ensure a steeper profile with respect to a 12-6 Lennard-Jones potential.

The final contribution $U_{\text{HB}}$ consists of two-body ($U_{\text{HB}}^{(2)}$) and a four-body ($U_{\text{HB}}^{(4)}$):

$$U_{\text{HB}} = U_{\text{HB}}^{(2)} + U_{\text{HB}}^{(4)} \quad (2.9)$$

The two-body term ($U_{\text{HB}}^{(2)}$) consists of short and long range HBs with respect to the position along the sequence:

$$U_{\text{HB}}^{(2)} = \sum_{i,j,i,j+4} \omega_{1>4}^{\text{HB}} \varepsilon_{1,4}^{\text{HB}}(r_{\text{ij}}) \mu(\alpha_{\text{ij}}) + \sum_{i,j,i+4} \omega_{1>4}^{\text{HB}} \varepsilon_{1,4}^{\text{HB}}(r_{\text{ij}}) \varepsilon_{1,4}^{\text{HB}}(r_{\text{ij}}) \mu(\alpha_{\text{ij}}) \quad (2.10)$$

where

$$\mu(r_{\text{ij}}) = 5 \left( \frac{\sigma}{r_{\text{ij}}} \right)^{12} - 6 \left( \frac{\sigma}{r_{\text{ij}}} \right)^{10} \quad (2.11)$$

$$\varepsilon(\alpha_{\text{ij}}) = \begin{cases} \cos^2(\alpha_{\text{ij}}), & \alpha_{\text{ij}} > 90 \\ 0, & \text{otherwise} \end{cases} \quad (2.12)$$
The sum runs over all residues separated by \( j \geq i + 4 \), \( r_{ij} \) is the O...H distance between the carbonyl oxygen and the amide hydrogen in the backbone, \( \alpha_{ij} \) is the angle \( \text{N}\hat{H}\text{O} \) and \( \sigma \) is the equilibrium distance of the HB.

The four-body term takes the form of sum of weighted products of Gaussian functions, each monitoring the existence of an HB on the basis of distance criteria. The Gaussian functions are computed for each possible HB pairs and they give a contribution to the four-body term only if tight conditions on sequence-separation are verified:

\[
U^{(4)}_{\text{HB}} = \omega_{\alpha}^{\text{HB}} \sum \epsilon_{\alpha}^{\text{HB}} \exp\left(-\left(r_{ij} - \sigma\right)^2/2\right) \exp\left(-\left(r_{kl} - \sigma\right)^2/2\right) \Delta(ijkl) \\
+ \omega_{\beta}^{\text{HB}} \sum \epsilon_{\beta}^{\text{HB}} \exp\left(-\left(r_{ij} - \sigma\right)^2/2\right) \exp\left(-\left(r_{kl} - \sigma\right)^2/2\right) \Delta'(ijkl) \quad (2.13)
\]

where

\[
\Delta(ijkl) = \begin{cases} 
1 & \text{if } (k,l) = (i+1, j+1) \\
0 & \text{otherwise} 
\end{cases} \quad (2.14)
\]

\[
\Delta'(ijkl) = \begin{cases} 
1 & \text{if } (k,l) = (i+2, j-2) \text{ or } (i+2, j+2) \\
0 & \text{otherwise}
\end{cases} \quad (2.15)
\]

These conditions help stabilize \( \alpha \) - helices and \( \beta \) - sheets, independently of the \((\phi, \varphi)\) dihedral angles, but also any segment satisfying the conditions on \((ijkl)\).

In the last version of OPEP (version 5) [99], the well depth value at the minimum \( \epsilon^0 = 3.89 \text{ kcal mol}^{-1} \) for the Ile-Ile contact and \( \epsilon^0 = 4.05 \text{ kcal mol}^{-1} \) for a Lys-Glu salt-bridge. The \( \epsilon^0 \) value, at the minimum, of an intramolecular H-bond is 3.3 and 2.7 kcal mol\(^{-1}\) for \((i, i+4)\) and \((i, j \geq i+5)\) interactions and 2.7 kcal mol\(^{-1}\) for an intermolecular H-bond. The value of \( \epsilon^0 \) of four-body H-bond terms are 1.4 and 3.6 kcal mol\(^{-1}\) for \( \alpha \) - helices and \( \beta \) - sheets.

### 2.1.2.2 Lattice model

In this thesis, we employed the lattice model developed by Abel et al. [100]. Here, a residue \( i \)th is represented by a bead located on a corner of a simple cubic lattice with position \( \vec{p}_i \). Its side chain is described by a unit vector \( \hat{d}_i \), which is not allowed to point in the same direction as the backbone. Thus, only four side-chain directions are allowed for a central residue, and five side-chain directions are possible for the terminus residues. A residue only has contact with its nearest neighbour which is not a sequence in the chain. A contact \( C_{ij} \) between two residues \( i \) an \( j \) is defined as:

\[
C_{ij} = \begin{cases} 
1 & \text{if } |\vec{p}_i - \vec{p}_j| = 1 \text{ and } |i - j| > 1 \\
0 & \text{otherwise}
\end{cases} \quad (2.16)
\]
The potential energy function consists of the following terms:

\[ E = E_{aa}(\varepsilon_{aa}) + E_{state}(\varepsilon_{ss}) + E_{hb}(\varepsilon_{hb}) + E_{steric}(\varepsilon_s) + E_{solvent}(\varepsilon_{wa}) \] (2.17)

Here, \( E_{aa} \) represents the total energy of the pairwise interactions \( \varepsilon_{aa} \) between two residues. The second term, \( E_{state} \) is the total state energy. The third term \( E_{hb} \) is the sum of the hydrogen bond (H-bond) energies \( \varepsilon_{hb} \). The fourth term, \( E_{steric} \), describes the total energy of the steric penalty \( \varepsilon_s \) to prevent clashes between side-chain and backbone atoms. The last term \( E_{solvent} \) describes the total energy of the pairwise interactions \( \varepsilon_{wa} \) between amino acids and solvent molecules.

**Interactions between amino acids**

In the present model, the total potential energy of pairwise interactions between amino acids, \( E_{aa} \), depends both on the positions on the lattice and side chain directions of the two residues and is given by:

\[ E_{aa} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} C_{i,j} K_{i,j} \varepsilon_{a_i,a_j} \] (2.18)

where \( K_{i,j} \) indicates whether two residues \( i \) and \( j \) are allowed to interact. That is they are in contact, if the side chain vectors are oriented in the same direction or pointed toward with each other. The strengths of the pairwise interaction energies between different types of amino acids \( (a_i, a_j) \) are provided by the force fields such as Thirumalai-Betancourt [101], Abeln [100, 102, 103] and OPEP [104] force fields.

**State energy**

The secondary structure is modeled by two possible states, coil and strand for each residue. Here, we assign a zero energy value for the coil state and a negative value \( \varepsilon_{ss} \) for the strand state.

**H-bond energy**

The total energy of the H-bonds for a configuration is defined as:

\[ E_{hb} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \varepsilon_{hb} H_{i,j} C_{i,j} \] (2.19)

where \( \varepsilon_{hb} \) represents the potential energy per H-bond and \( H_{i,j} \) denotes whether or not a H-bond between residue \( i \) and \( j \) exists. A H-bond is formed if two residues are in contact, side chains are oriented in the same direction and both residues are in "strand" states.
Steric hindrance penalty
Using a steric hindrance energy term, we prevent consecutive side chains being oriented in the same direction. In real protein structures such conformations are blocked as a result of steric hindrance from clashes between side chain and backbone atoms; such detail is not included in this model. To prevent these conformations, we use:

\[
E_{\text{steric}} = \sum_{i=1}^{N} \epsilon_s S_i \tag{2.20}
\]

where \(\epsilon_s\) is the energy penalty for steric hindrance and \(S_i\) denotes whether or not residue \(i\) is in a state that causes steric hindrance:

\[
S_i = \begin{cases} 
1 & \text{if } \hat{d}_{i-1} = \hat{d}_i \text{ or } \hat{d}_{i+1} = \hat{d}_i \\
0 & \text{otherwise} 
\end{cases} \tag{2.21}
\]

Solvent interaction.
A solvent molecule is modeled by empty lattice sites. The total solvation energy, \(E_{\text{solv}}\) is given by the following expression:

\[
E_{\text{solv}} = \frac{1}{2} \sum_{\text{solv}} \sum_{i=1}^{N} \epsilon_{a_i,\text{solv}} K_{i,\text{solv}} \tag{2.22}
\]

Here, \(\epsilon_{a_i,\text{solv}}\) is the strength between the amino acid type \(a_i\) and solvent; \(K_{i,\text{solv}}\) indicates whether or not an interaction between the solvent and amino acid type \(i\) appears, the residue must be in contact with a solvent site and its side chain direction towards the solvent.

2.2 Molecular Dynamics (MD) simulations
For a \(N\)-particle-system with a position vector \(\vec{r}^N = \{\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N\}\), with the potential \(U(\vec{r}^N)\), the force acting upon \(i\)th atom can be computed as the negative gradient of the energy (vector first derivatives), with respect to atomic displacements as show in the following equation:

\[
\vec{f}_i = -\nabla_{\vec{r}_i} U(\vec{r}^N) \tag{2.23}
\]

In MD simulations the time evolution of a set of interacting particles is followed via the solution of Newton’s equations of motion, eqn[2.24], where \(\vec{r}_i(t) = (x_i(t), y_i(t), z_i(t))\) is the position vector of \(i\)th particle and \(\vec{f}_i\) is the force acting upon \(i\)th particle at time \(t\) and \(m_i\) is the mass of the particle.

\[
\vec{f}_i = m_i \frac{d^2 \vec{r}_i(t)}{dt^2} \tag{2.24}
\]
Integrating the Equations of Motion

The positions and velocities after an interval time $\delta t$ can be estimated by numerically solving the Newton equation based on the positions, velocities and forces at time $t$. Repeating the integrations, the time evolution of the system can be observed. Algorithms have been developed to integrate the equation with the aim computational efficiency and high accuracy. One very important criterion for a good algorithm is the conservation of energy, for both short time and long time means. Below are the two most widely used-algorithms.

**Verlet algorithm**
The position of atom $i$ at time $t + \delta t$ is estimated as [105].

$$\vec{r}_i(t + \delta t) \approx 2\vec{r}_i(t) - \vec{r}_i(t - \delta t) + \vec{f}_i(t) m_i \delta t^2. \quad (2.25)$$

The error estimated in position is of order $O(\delta t^4)$, where $\delta t$ is the time step in Molecular Dynamics scheme. Verlet algorithm does not use the velocity to calculate the new position, in contrast, it requires previous position.

**Velocity Verlet algorithm**
The position and velocity of the atom $i$ at time $t + \delta t$ are:

$$\vec{r}_i(t + \delta t) = \vec{r}_i(t) + \vec{v}_i(t) \delta t + \frac{\vec{f}_i(t)}{m} \delta t^2, \quad (2.26)$$

$$\vec{v}_i(t + \delta t) = \vec{v}_i(t) + \frac{\vec{f}_i(t + \delta t) + \vec{f}_i(t)}{2m} \delta t, \quad (2.27)$$

In this algorithm, the usual choice for initialising the velocities is to generate Maxwell-Boltzmann distribution, which is a normal distribution with a mean corresponding to a given temperature for the sample.

### 2.2.1 Thermostat

To mimic the laboratory condition where the temperature fluctuates around a constant value, thermostat is necessary for MD simulation. Among several thermostat algorithms, the velocity - rescale thermostat algorithm for sampling the canonical ensemble [106] is widely used to maintain the system temperature at a desired value. In this method, the
velocities of all the particles are rescaled $v_{\text{rescale}} = \alpha v$ by a properly chosen random factor $\alpha$:

$$
\alpha^2 = e^{-\delta t/\tau} + \frac{\bar{K}}{N_f K} \left( 1 - e^{-\delta t/\tau} \right) \left( R_1^2 + \sum_{i=2}^{N_f} R_i^2 \right) + 2 e^{-\delta t/2\tau} \sqrt{\frac{\bar{K}}{N_f K} \left( 1 - e^{-\delta t/\tau} \right) R_1},
$$

(2.28)

where $\bar{K} = N_f/2\beta$ is the average kinetic energy at the target temperature, $K$ is the instantaneous kinetic energy, $N_f$ is the number of degrees of freedom, $\beta$ is the inverse temperature and $R_i$ is a random number from a Gaussian distribution with unitary variance.

### 2.2.2 Barostat

Several algorithms have been developed to maintain the pressure in MD simulation at desired value. Here, we used the Berendsen method [107] where the system pressure is set to a desired value by the scaling factor (for each dimension):

$$
\mu = \left[ 1 - \frac{\Delta t}{\tau_p} \left( P_0 - P(t) \right) \right]^{1/3},
$$

(2.29)

where $P_0$ is the desired value of pressure, $P(t)$ is the instantaneous pressure, and $\tau_p$ the coupling time constant for the pressure scaling.

### 2.2.3 Periodic Boundary condition (PBC)

Periodic boundary condition is very often used for condensed phase simulations with implicit and explicit solvent in Molecular Dynamics and Monte Carlo methods. With periodic boundary condition, the volume consisting of $N$ particles is considered as the primitive cell of an infinite periodic lattice of identical cells. Consequently, a given particle now interacts with all other particles in the infinite periodic system, that means, all other particles in the same periodic cell and all particles (including its own periodic image) interact with all other cells [108].

### 2.3 Monte Carlo (MC) simulations

MC method is based on the importance sampling algorithm [109], in which a random walk is constructed in such a way that the probability of visiting a particular point $\vec{r}^N = \{ \vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N \}$ is proportional to the Boltzmann factor $\exp[-\beta U(\vec{r}^N)]$, where $\vec{r}^N$ stands for the coordinates of all $N$ particles, and $U$ is the potential energy of the systems. There are many ways to construct such a random walk and they are crucial to the accuracy of the obtained results. In the approach introduced by Metropolis et al. [109], the following
scheme is proposed:
1. Select a particle at random, and calculate system’s energy \( U(\mathbf{r}^N) \).
2. Give the particle a random displacement, \( \mathbf{r}'^N = \mathbf{r}^N + \Delta \), and calculate system’s new energy \( U(\mathbf{r}'^N) \).
3. Accept the move from \( \mathbf{r}^N \) to \( \mathbf{r}'^N \) with probability
   \[
   \text{acc}(o \mapsto n) = \min\left(1, \exp\{-\beta [U(\mathbf{r}'^N) - U(\mathbf{r}^N)]\}\right) 
   \]
   (2.30)

Exploiting the Metropolis criterion, the canonical ensemble (NVT) is sampled. According to the Metropolis scheme, in order to decide whether to accept or reject the trial move, a random number, indicates by \( \text{Ran} f \), from a uniform distribution in the interval \([0,1]\), is generated. The trial move is accepted if \( \text{Ran} f < \text{acc}(o \mapsto n) \) and rejected, otherwise.

Clearly, it is very important that our random-number generator does indeed generate numbers uniformly in the interval \([0,1]\); otherwise the Monte Carlo sampling will be biased. The quality of random-number generators should never be taken for granted. Assumed that the ergodic hypothesis holds, ensemble average of an observable \( A(\mathbf{r}^N) \) can be estimated from the sampled trajectory of the systems:
\[
\langle A \rangle \approx \frac{1}{L} \sum_{i=1}^{L} A(\mathbf{r}^N), 
\]
(2.31)
where \( L \) is the total number of MC steps.

2.4 Enhance sampling - replica exchange MD and MC simulations

The method constructs a generalized ensemble from \( M \) noninteracting trajectories with temperatures \( T_m (m = 1, \ldots, M) \). A state of this generalized ensemble is characterized by \( X = \{\ldots, x_m^{(i)}, \ldots\} \), where \( x_m^{(i)} \) represents the coordinates \( q_m^{(i)} \) and momenta \( p_m^{(i)} \) of all atoms of the \( i \)th replica at temperature \( T_m \). The algorithm consists of two consecutive steps: (1) independent constant-temperature simulations of each replica, and (2) exchange of the trajectories (i.e., the coordinates) between the replicas (defined by the temperature) according to a Metropolis-like criterion. Considering, for example, the exchange \( X = \{\ldots, x_m^{(i)}, \ldots, x_n^{(j)}, \ldots\} \leftrightarrow X' = \{\ldots, x_m^{(j)}, \ldots, x_n^{(i)}, \ldots\} \), the criterion is given by
\[
W(x_m^{(i)}, x_n^{(j)}) = \begin{cases} \exp(-\Delta) & \text{if } \Delta > 0 \\ 1 & \text{if } \Delta \leq 0 \end{cases} 
\]
(2.32)
\[
\Delta = [U(q^{(j)}) - U(q^{(i)})]([\beta_i - \beta_j)], 
\]
(2.33)
where $\beta_i = 1/(k_BT_i)$; $U(q^{(j)})$ and $U(q^{(i)})$ are the potential energies of the $j$th and $i$th replica, respectively. After the exchange, the momenta $p^{(i)}$ and $p^{(j)}$ of the replicas are given by

$$p'^{(i)} = \sqrt{\frac{T_n}{T_m}} p^{(i)}, \quad p'^{(j)} = \sqrt{\frac{T_m}{T_n}} p^{(j)}$$ (2.34)

Iterating steps (1) and (2), the trajectories of the generalized ensemble perform a random walk in temperature space. This facilitates an efficient conformational sampling of the energy landscape of the system, even in the presence of multiple local minima. To obtain the canonical average of any physical quantity at a specific temperature, one performs a time average over all trajectory segments of the replica corresponding to this temperature.

To achieve the optimal performance of the algorithm for a given system, the temperatures of the replicas are chosen such (1) that the lowest temperature is small enough to sufficiently sample the states of low energy, (2) that the highest temperature is large enough to overcome energy barriers of the system, and (3) that the acceptance ratio criterion is sufficiently high. Unlike replica-exchange Molecular Dynamics simulation which considers the momenta and position of all the particles in the system, replica-exchange Monte Carlo simulation only takes into account the positions of the particles. A replica-exchange is an "unphysical" movement, so one cannot conclude about dynamics. In other words, using replica-exchange molecular dynamics, one only really carries out a form of sampling, but not "true" molecular dynamics [110].

### 2.5 Data analysis

#### 2.5.1 Specific heat

Heat capacity at constant volume reflects the fluctuation of energy in the system and it is one of the important properties to describe the first order phase transition. It allows to determine the transition temperature of the system and can be calculated as the following equation:

$$C_v = \frac{1}{k_BT^2} ( < E^2 > - < E >^2 )$$ (2.35)

where $k_B$ is the Boltzmann constant, $T$ is temperature and $E$ is the energy of the system, $< ... >$ denotes for the average value.
2.5.2 Common quantities

2.5.2.1 Root mean square deviation (RMSD)

A root mean square deviation is a strong indicator of conformational changes of a protein, or the similarity between two structures. It is the measure of the average distance between the atoms (usually the backbone atoms) of superimposed proteins. Given two sets of $N$-atom-conformations $v$ and $w$ in the Cartesian coordinates, the RMSD is defined as follows [111]:

$$\text{RMSD}(v, w) = \sqrt{\frac{1}{N} \sum_{i=1}^{N} || v_i - w_i ||^2}$$  \hspace{1cm} (2.36)

2.5.2.2 Radius of gyration ($R_g$)

The radius of gyration describes the dimensions, or the compactness of a molecule, and is given by:

$$R_g = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (r_i - \overline{r})^2},$$  \hspace{1cm} (2.37)

where $r_i$ is the position of atom $i$th, $\overline{r}$ is the centre of mass and $N$ is the number of atoms in the molecule.

2.5.2.3 Number of H-bonds

A H-bond in an all-atom conformation is considered to form when the acceptor-donor distance is not more than 0.35 nm, and the acceptor-donor-hydrogen angle is not more than $35^\circ$, while in lattice configuration, a H-bond is formed if two residues are in contact, side chains is oriented in the same direction and both residues are in ”strand” state.

2.5.2.4 Number of Contacts

In all-atom model, a $C_\alpha - C_\alpha$ contact is defined if the distance is $< 0.5$nm. Side-chain-side-chain (SC - SC) and backbone - backbone (BB - BB) contacts are defined when the minimum distance between their heavy atoms is $< 0.45$nm. A salt-bridge (SB) between two charged side chains is formed if the distance between two specific atoms is $< 0.46$nm. In lattice simulations, two residues are in contact if they are separated by a lattice spacing $a = 0.3$nm.
2.5.2.5 Secondary structure

The secondary structures of an all-atom conformation are calculated using STRIDE [112]. The eight STRIDE structures are grouped into four structures: \( \beta = \) extended + bridge, \( \alpha = \) \( \alpha \)-helix + \( 3_{10} \)-helix, turn = turn + bend, and coil = coil + \( \pi \)-helix.

2.5.2.6 Populations of antiparallel and parallel conformations

For both lattice and all-atom oligomers, a parallel or antiparallel \( \beta \)-sheet of size \( n \) (called \( n \)-stranded \( \beta \)-sheet) is defined as a set of \( n \) \( \beta \)-strands connected consecutively by at least two backbone H-bonds, and the cosine of the angle between two end-to-end vectors of two consecutive hydrogen bonded strands is \( \cos(\theta) \geq 0.7 \) or \( \cos(\theta) \leq 0.7 \), respectively. Here, a \( \beta \)-strand is defined if it consists of at least three consecutive residues in the \( \beta \) structure.

2.5.2.7 Order parameter

In order to characterize the order of the system, we used the nematic order parameter \( P_2 \) [113]. In terms of the unit vector \( u_i \) linking N- and C-termini for the \( i \)-th peptide, \( P_2 \) is

\[
P_2 = \sum_{i=1}^{N} \frac{r_{NC}^i}{L_i} p_2^0,
\]

where \( p_2^0 = \frac{1}{2N} \sum_{i=1}^{N} \frac{3}{2} (u_i \cdot d)^2 - \frac{1}{2} \), where \( d \) (the director) is a unit vector defining the preferred direction of alignment of the oligomer, \( N \) is the number of molecules, \( r_{NC}^i \) is the end-to-end vector that connects two \( C_\alpha \) atoms from the termini of the \( i \)-th peptide. The end-to-end distance in the fully stretched state \( L_i = (N_i - 1)a \), where \( N_i \) is a number of amino acids in \( i \)-th monomer and \( a \) is the distance between two consecutive \( C_\alpha \) atoms.

2.5.2.8 Binding free energy

The binding free energy is estimated using the Poisson-Boltzmann Surface Area (PBSA) method [114]:

\[
\Delta G_{\text{bind}} = \Delta G_{\text{elec}} + \Delta G_{\text{nonpol}} = \Delta G_{\text{elec}} + \gamma \Delta MS.
\]  

The electrostatic component, \( \Delta G_{\text{elec}} \), is calculated by solving the Poisson-Boltzmann equation implemented in the CHARMM package [95] on the solvated dimer and isolated monomers:

\[
\Delta G_{\text{elec}} = \Delta G_{\text{elec,dimer}} - (\Delta G_{\text{elec,}A\beta_1} + \Delta G_{\text{elec,}A\beta_2})
\]

The dielectric constant of the peptide is set to 4, and that of water, to 80. The nonpolar component of the binding free energy, \( \Delta G_{\text{nonpolar}} \), is calculating by multiplying the molecular surface buried at the interface between the two peptides, \( \Delta MS \), by the surface tension.
constant $\gamma = 0.0072$ kcal/(mol.Å$^2$) [115] where $\Delta MS$ is estimated as:

$$\Delta MS = SASA_{dimer} - (SASA_{A\beta 1} + SASA_{A\beta 2})$$ (2.41)

### 2.5.3 Clustering methods

To identify the dominant structure exploring during the simulation, it is necessary to group the sampled conformations corresponding to their similarity. This type of approach, called clustering, has proven very powerful technique in identifying important conformations, and evaluating kinetic rates between different geometries. We have mainly carried out clustering analysis using the Hartigan-Wong k-means algorithm [116] as implemented in the R program [117].

### 2.5.4 Principal components analysis

To reduce the dimensionality of the high-dimensional of the simulation data, one can employ the principal component analysis method, also called quasiharmonic analysis or essential dynamics method [118, 119, 120, 121]. Considering the dynamics of $N$-particle-system, the basic idea is that the correlated internal motions are represented by the covariance matrix:

$$\sigma_{i,j} = \left\langle \left( q_i - \langle q_i \rangle \right) \left( q_j - \langle q_j \rangle \right) \right\rangle,$$ (2.42)

where $q_1, ..., q_{3N}$ are the mass-weighted Cartesian coordinates of the molecule and $\langle ... \rangle$ indicates the average over all sampled conformations. The PCA represents a linear transformation that diagonalizes the covariance matrix and thus removes the instantaneous linear correlations among the variables. Ordering the eigenvalues of the transformation decreasingly, it has been shown that a large part of the system fluctuations can be described in terms of only a few PCA eigenvectors or principal components [118, 119, 120, 121]. In our project, we used both PCA and dihedral principal component analysis (dPCA) [122] to analysis the output from simulations.

### 2.5.5 Identification of the structure of oligomers

To characterize the oligomer states in our all-atom simulations, we applied the general method proposed by Nguyen et al. [123]. The basic idea of this method is that overall molecular structures of oligomers are the product of intramolecular and intermolecular states which are briefly described below.
2.5.5.1 *Intra - molecular state*

Supposing that the oligomer is simulated for \( N_s \) steps, and consisted of \( N_c \) chemically identical chains, each chain has \( N_r \) residues.

The intramolecular states of the oligomers are characterized in term of the structures of single monomers. Let \[ \left[ \phi_i^m(t_n), \psi_i^m(t_n) \right] \quad (m = 1, ..., N_r, i = 1, ..., N_c) \] be the backbone dihedral angles of the residue \( m^h \) of the chain \( i^h \), measured at time \( t_n (n = 1, ..., N_s) \). These angles are mapped to new variables \[ \left[ \phi_i^m(t_n), \psi_i^m(t_n) \right] \rightarrow \left[ \Phi_i^m(t_j), \Psi_i^m(t_j) \right], \quad j = (n - 1)N_c(N_c - 1)/2 + i. \] These variables can be considered as the angles of a monomer simulated alone for \( N_c \times N_s \) steps. All possible intramolecular states of all chains in the oligomer are fully described according to the knowledge of the set of the angles \[ \left[ \Phi_i^m(t_j), \Psi_i^m(t_j) \right] \]. In other words, the intramolecular states of the oligomer can be obtained in term of the structures of single-chains, therefore avoiding the degeneracy problem. These structures of single chains are obtained through the dihedral principal component analysis (dPCA) [122]. The covariance matrix is defined as:

\[
\sigma_{ab} = \left\langle \left( q_a - \langle q_a \rangle \right) \left( q_b - \langle q_b \rangle \right) \right\rangle, \tag{2.43}
\]

where \( q_c = \cos(\alpha_c) \) and \( q_{c+1} = \sin(\alpha_c) \). Here, \( \alpha \in \{ \Phi^m, \Psi^m \}, m = 1, ... 2 \times N_r \) and \( \langle ... \rangle \) indicates the average over all sampled conformation of the single-chain trajectory of length \( N_c \). The diagonalization of the covariance matrix results in the principal components which are used to construct the free energy landscape (FEL), \( F_{\text{single}} \), of a single monomer. The FEL spanned by the first \( p \) principal components \( V = (V_1, ..., V_p) \) is given by \( F_{\text{single}}(V) = -k_B T \ln \left( \frac{P(V)}{P_{\text{max}}} \right) \), where \( P(V) \) is the probability distribution obtained from a histogram of the MD data and \( P_{\text{max}} \) denoted its maximum, which is devised to ensure that \( \Delta G = 0 \) for the lowest free energy minimum.

Let us call \( \tilde{s}_k^i \), \( (k = 1, ..., n_{\text{state}}) \), the single-molecule states on \( F_{\text{single}}(V_1, ..., V_p) \). An intramolecular state, \( S_{\text{intra}}^m (m \text{ denotes state’s index}) \), of the oligomer is completely characterized by \( n_{\text{state}} \) single-molecule state \( \tilde{s}_k^i \). In principle, there are \( (n_{\text{state}})^{N_c} \) possible combinations of \( \tilde{s}_k^i \) states that form states \( S_{\text{intra}}^m \); however, in practice, not all these states are sampled during the simulation due to the degeneracy in those states.

2.5.5.2 *Inter - molecular state*

There are \( i = 1, ..., N_c \times (N_c - 1)/2 \) pair-chains (or double-chains) and each double-chain composed \( m = 1, ..., N_r \times N_r \) distances between the inter-chain centres of mass of the side-chains. The intermolecular states of the oligomer are characterized in terms of double-molecule states that described the position-orientation arrangements of any two chains.
Considering \( [d^m_i(t_n)] \) as the distance of the \( m^{th} \) inter-chain side-chain pair of the \( i^{th} \) double-chain measured at time \( t_n, (n = 1, \ldots, N_s) \). All these distances are mapped to \([d^m_i(t_n)] \mapsto [D_m(t_j)] \), \( m = 1, \ldots, N_r \times N_r, j = (n - 1) \times N_c \times (N_c - 1)/2 + i \). This new representation can be imagined as the distances between the inter-chain centres of mass of the side-chains for two peptides simulated alone during \( N_c \times (N_c - 1) \times N_s/2 \) steps.

All positions and orientation of all the chains in the oligomer are fully described by a knowledge of the set \( D_m(t_j) \). That is, the intermolecular states of the oligomer are obtained in terms of combinations of the structures of double-chains, thereby avoiding the degeneracy. These structures are obtained through the principal components analysis (PCA) of the inverse distances \( 1/D_m(t_j) \) [204]. The PCA are carried out by computing the covariance matrix

\[
\sigma_{ab} = \left\langle \left( D^a - \langle D^a \rangle \right) \left( D^b - \langle D^b \rangle \right) \right\rangle, \tag{2.44}
\]

where \( a, b = 1, \ldots, N_r \times N_r \) and \( \langle \ldots \rangle \) denotes the average over all sampled double-chains of a virtual trajectory of length \( N_c \times (N_c - 1) \times N_s/2 \). By diagonalizing \( \sigma \), we receive \( N_r \times N_r \) eigenvectors \( \mathbf{v}^{(l)} = \{ \nu_{nl} \} \) (\( \nu_{nl} \) is the \( n^{th} \) component of the \( l^{th} \) eigenvectors), and eigenvalue \( \lambda_n \), which are rank-ordered in descending order, i.e., \( \lambda_1 \) represents the largest eigenvalue. The eigenvectors and eigenvalues of \( \sigma \) yield the modes of collective motion and their amplitudes. The \( l^{th} \) principal component is defined as \( V_l = \mathbf{v}^{(l)} \cdot \mathbf{q} \). The FEL explored by the first few principal components \( V = (V_1, \ldots, V_p) \) is given by \( F_{\text{double}}(V) = -k_B T \left[ \ln \left( \frac{P(V)}{P_{\text{max}}} \right) \right] \), where \( P(V) \) is the probability distribution obtained from a histogram of MD data.

Considering the state \( s^k_d (k = 1, \ldots, n_{\text{state}}) \) as the double-molecular states on \( F_{\text{double}}(V_1, \ldots, V_p) \). An intermolecular state, \( S^l_{\text{inter}} \) (1 indicated the state’s index), of the oligomer is absolutely characterized by \( N_c \times (N_c - 1)/2 \) double-molecular state \( s^k_d \), since the positions and orientations of the double-molecules states are known from the trajectories. In principle, there are \( M = n_{\text{state}}^{N_c \times (N_c - 1)/2} \) possible combinations of the \( s^k_d \) states which formed states \( S^m_{\text{inter}} \), nevertheless, in practice, there are degenerated states, so the number of combinations states is much less than \( M \).

### 2.5.5.3 Overall state

After constructing the intramolecular states \( S^m_{\text{intra}} \) and the intermolecular states \( S^l_{\text{inter}} \), the overall states of the oligomer are simply calculated by the product basis \( S^n_{\text{overall}} = S^m_{\text{intra}} \times S^l_{\text{inter}}, \) where \( n \) denotes the index of the state.
Chapter 3

Lattice model for amyloid peptides: OPEP force field parametrization and applications to the nucleus size of Alzheimer’s peptides

Sumary

The neurodegenerative Alzheimer’s disease is affecting more than 40 million people worldwide and is linked to the aggregation of the amyloid-$\beta$ ($A\beta$) proteins of 40/42 amino acids. Despite extensive experimental studies, the mechanism of formation of amyloid fibrils and plaques is still unclear. To complement experiments, computational studies based on all-atom simulations are very often used. They are however limited in the sampling problem for small systems and short time scales (ns-$\mu$s).

Protein lattice models approximate atomistic details and keep the essential interactions, but their force field is not optimal. In this work, we present a comprehensive OPEP force-field parameterization for two representative peptide fragments $A\beta_{16-22}$ and $A\beta_{37-42}$ of the Alzheimer’s peptide $A\beta_{1-42}$ using an on-lattice protein model, which incorporates explicitly the formation of hydrogen bonds and directions of side-chains. Our bottom-up approach starts with the determination of the best force-field for the $A\beta_{16-22}$ dimer by fitting its equilibrium parallel and anti-parallel beta-sheet populations obtained from the lattice simulations to that of all-atom simulations. Interestingly, the calibrated force-field is transferable to the trimer $A\beta_{16-22}$ as well as $A\beta_{37-42}$ dimer and trimer. Encouraged by this finding, we study the free energy landscapes of the two decamers. Our results show that the
dominant structure of the $\text{A} \beta_{16-22}$ decamers matches the microcrystal structure. Pushing the simulations for aggregates between 4-mer and 12-mer suggests a nucleus size for fibril formation of 10 chains. In contrast, the decamer $\text{A} \beta_{37-42}$ is largely disorder with mixed by parallel and antiparallel chains, suggesting that the nucleus size is much larger than 10 peptides. Our refined force field coupled to this on-lattice model should provide useful insights into the critical nucleation number associated with neurodegenerative diseases. The contents of this chapter have been published in our recent paper: Thanh Thuy Tran, Phuong H. Nguyen, and Philippe Derreumaux, "Lattice model for amyloid peptides: OPEP force field parametrization and applications to the nucleus size of Alzheimer's peptides", The Journal of Chemical Physics 144, 205103 (2016); doi: 10.1063/1.4951739 (May 2016) [87].

Résumé du Chapitre 3: Un modèle réseau pour les peptides amyloïdes: paramétrization OPEP et application à la maladie d’Alzheimer.

La maladie d’Alzheimer touche plus de 40 millions de personnes dans le monde et résulte de l’agrégation du peptide beta-amyloïde de 40/42 résidus. En dépit de nombreuses études expérimentales, le mécanisme de formation des fibres et des plaques n’est pas élucidé. Les simulations atomiques étant limitées à des petits systèmes et des échelles de temps entre la ns et la $\mu$s, nous présentons un modèle réseau basé sur le champ de forces OPEP pour deux peptides modèles $\text{A} \beta_{16-22}$ et $\text{A} \beta_{37-42}$ du peptide beta-amyloïde.

Notre approche consiste à déterminer les paramètres OPEP du dimère $\text{A} \beta_{16-22}$ en comparant les simulations Monte Carlo sur réseau et des dynamiques moléculaires atomiques par échange de réplicas. Il s’avère que ces paramètres fonctionnent aussi sur le trimère $\text{A} \beta_{16-22}$ et les dimères et trimères $\text{A} \beta_{37-42}$, nous permettant d’étudier la surface d’énergie libre des décamères. Nos simulations montrent que la structure dominante du 10-mère $\text{A} \beta_{16-22}$ est celle observée expérimentalement et suggèrent un noyau de nucléation de 10 chaînes. En revanche, la structure dominante du 10-mère $\text{A} \beta_{37-42}$ est désordonnée, suggérant un noyau de nucléation beaucoup plus grand que 10 chaînes.

Le contenu de ce chapitre a été publié: Thanh Thuy Tran, Phuong H. Nguyen, and Philippe Derreumaux, "Lattice model for amyloid peptides: OPEP force field parametrization and applications to the nucleus size of Alzheimer’s peptides", The Journal of Chemical Physics 144, 205103 (2016); doi: 10.1063/1.4951739 (May 2016) [87].
3.1 Introduction

A number of proteins and peptides have been found to aggregate into oligomers and then insoluble amyloid fibrils composed of cross $\beta$-sheets stabilized by backbone hydrogen bonds (H-bond) [124, 125]. In human brains, fibrils are associated with several neurodegenerative diseases, such as Alzheimer’s and Parkinson’s disease. Thus, it is important to understand the structure and aggregation mechanism [17, 18] to design more efficient inhibitors of aggregation and toxicity [126, 127]. From the experimental side, a wide range of methods is currently being developed in order to improve our understanding of the amyloid fibril formation [5, 13]. For example, the self-assembly process is studied in details by monitoring the formation of inter-residue contacts using fluorescence methods [128], photoinduced cross linking [129], and of $\beta$-sheet aggregates using isotope-edited infrared spectroscopy [130], atomic force spectroscopy [131]. From the theoretical side, the large majority of all-atom MD simulations are performed for small oligomers, and there are only a few all-atom simulations of large oligomers with sizes ranging from 10 to 20-mers and time scales of 50-500 ns [132, 133, 134, 135]. This is because, at least partly, the all-atom simulations are too expensive to capture the length and time scales associated with fibril formation. Several coarse-grained off-lattice models have been developed and employed to explore the early steps of aggregation and the mechanism of the fibril formation [136, 83, 137, 138, 139, 140, 66, 144, 145, 141, 142, 146, 147, 148, 149, 150, 151, 152]. Yet, such off-lattice models are generally still too expensive to study large systems. As a remedy, several coarse-grained on-lattice models have been developed [153, 154, 155, 156]. Among these, the recent developed model of Abeln and coworkers is very promising. This model incorporates explicitly the formation of hydrogen bonds, directions of side-chain, and uses the information-based pairwise interaction between the twenty amino acids, allowing one to capture the aggregation and amyloid formation of large proteins at low computational cost [100, 102]. Our primary aim is to employ and develop further this lattice model by incorporating the well-developed coarse-grained off-lattice force-field OPEP [104] into the model, and parametrizing the missing parameters, and finally to determine the critical nucleus size $n^*$ of fibril formation. As the first step in this direction, we consider two representative and well-studied peptide segments $\text{A}_1\beta_{16-22}$ and $\text{A}_2\beta_{37-42}$ of the full length $\text{A}_{1-42}$ Alzheimer’s peptide. These two peptides have been extensively employed as prototypical systems in the amyloid protein research field. The crystal structures show that $\text{A}_1\beta_{16-22}$ peptide forms fibril consisting of anti-parallel $\beta$-sheets with anti-parallel $\beta$-strands within individual sheet, while the $\text{A}_2\beta_{37-42}$ peptide forms anti-parallel $\beta$-sheets with parallel $\beta$-strands within each sheet [125, 157, 158]. Guided by these structural features, one way
to estimate $n^*$ is to follow the populations of the parallel and anti-parallel structures as a function of the number of peptide chains. With this in mind, our bottom-up approach starts with the force field parameterization for the dimers and this is done by comparing the populations of the parallel and anti-parallel conformations obtained from lattice and all-atom simulations. The obtained force-field is then validated for the trimers, and finally used to study the free energy landscapes of decamers.

### 3.2 Systems and force field parametrization strategy

In this work, we consider the dimers, trimers and decamers formed by the peptide segments $\text{A}\beta_{16-22}$ (Lys$^+$-Leu-Val-Phe-Phe-Ala-Glu$^-$) and $\text{A}\beta_{37-42}$ (Gly$^+$-Gly-Val-Val-Ile-Ala) of the full length $\text{A}\beta_{1-42}$ Alzheimer’s peptide. Representative structure of these peptides in the lattice representation is shown in Fig. 3.1. The pH is set to 7 with the N-terminus is treated as NH$_3^+$ and the C-terminus as CO$_2^-$.

Let’s us denote $\varepsilon_{aa}$ for the contact energies between two inner residues, $\varepsilon_{+-}$ between oppositely charged termini, $\varepsilon_{++}$, $\varepsilon_{--}$ between like-charged termini, and $\varepsilon_{+a}$, $\varepsilon_{-a}$ between charged termini and inner residues. For the sake of simplicity we assume that $\varepsilon_{++} = \varepsilon_{--} \equiv \varepsilon_{\pm\pm}$ and $\varepsilon_{+a} = \varepsilon_{-a} \equiv \varepsilon_{\pm a}$. The contact energies $\varepsilon_{aa}$ are described by the OPEP force-field,
### Table 3.1: The OPEP force field of pairwise interaction matrix between two residues of the \( \text{A}\beta_{16-22} \) (above) and \( \text{A}\beta_{37-42} \) (below) peptides. The values are normalized with respect to the maximal interactions Phe-Phe for \( \text{A}\beta_{16-22} \) and Val-Ile for \( \text{A}\beta_{37-42} \). The remaining interactions between charged residues and inner residues \( \varepsilon_{\pm a} \), between like-charged termini \( \varepsilon_{\pm\pm} \), and between oppositely charged termini \( \varepsilon_{\pm-} \) have to be optimized. The values are in the dimensionless \( k_B T \) unit.

<table>
<thead>
<tr>
<th></th>
<th>Lys⁺</th>
<th>Leu</th>
<th>Val</th>
<th>Phe</th>
<th>Phe</th>
<th>Ala</th>
<th>Glu⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lys⁺</td>
<td>( \varepsilon_{++} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{+-} )</td>
</tr>
<tr>
<td>Leu</td>
<td>( \varepsilon_{\pm a} )</td>
<td>-0.988</td>
<td>-0.769</td>
<td>-0.951</td>
<td>-0.951</td>
<td>-0.369</td>
<td>( \varepsilon_{\pm a} )</td>
</tr>
<tr>
<td>Val</td>
<td>( \varepsilon_{\pm a} )</td>
<td>-0.769</td>
<td>-0.709</td>
<td>-0.817</td>
<td>-0.817</td>
<td>-0.356</td>
<td>( \varepsilon_{\pm a} )</td>
</tr>
<tr>
<td>Phe</td>
<td>( \varepsilon_{\pm a} )</td>
<td>-0.951</td>
<td>-0.817</td>
<td>-1.000</td>
<td>-1.000</td>
<td>-0.393</td>
<td>( \varepsilon_{\pm a} )</td>
</tr>
<tr>
<td>Phe</td>
<td>( \varepsilon_{\pm a} )</td>
<td>-0.951</td>
<td>-0.817</td>
<td>-1.000</td>
<td>-1.000</td>
<td>-0.393</td>
<td>( \varepsilon_{\pm a} )</td>
</tr>
<tr>
<td>Ala</td>
<td>( \varepsilon_{\pm a} )</td>
<td>-0.369</td>
<td>-0.356</td>
<td>-0.393</td>
<td>-0.393</td>
<td>-0.178</td>
<td>( \varepsilon_{\pm a} )</td>
</tr>
<tr>
<td>Glu⁻</td>
<td>( \varepsilon_{+-} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{-} )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Gly⁺</th>
<th>Gly</th>
<th>Val</th>
<th>Val</th>
<th>Ile</th>
<th>Ala⁻</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gly⁺</td>
<td>( \varepsilon_{++} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{+-} )</td>
</tr>
<tr>
<td>Gly</td>
<td>( \varepsilon_{\pm a} )</td>
<td>-0.296</td>
<td>0.036</td>
<td>0.036</td>
<td>0.201</td>
<td>( \varepsilon_{\pm a} )</td>
</tr>
<tr>
<td>Val</td>
<td>( \varepsilon_{\pm a} )</td>
<td>0.036</td>
<td>-0.898</td>
<td>-0.898</td>
<td>-1.000</td>
<td>( \varepsilon_{\pm a} )</td>
</tr>
<tr>
<td>Val</td>
<td>( \varepsilon_{\pm a} )</td>
<td>0.036</td>
<td>-0.898</td>
<td>-0.898</td>
<td>-1.000</td>
<td>( \varepsilon_{\pm a} )</td>
</tr>
<tr>
<td>Ile</td>
<td>( \varepsilon_{\pm a} )</td>
<td>0.201</td>
<td>-1.000</td>
<td>-1.000</td>
<td>-0.902</td>
<td>( \varepsilon_{\pm a} )</td>
</tr>
<tr>
<td>Ala⁻</td>
<td>( \varepsilon_{+-} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{\pm a} )</td>
<td>( \varepsilon_{-} )</td>
</tr>
</tbody>
</table>

Table 3.1: The OPEP force field of pairwise interaction matrix between two residues of the \( \text{A}\beta_{16-22} \) (above) and \( \text{A}\beta_{37-42} \) (below) peptides. The values are normalized with respect to the maximal interactions Phe-Phe for \( \text{A}\beta_{16-22} \) and Val-Ile for \( \text{A}\beta_{37-42} \). The remaining interactions between charged residues and inner residues \( \varepsilon_{\pm a} \), between like-charged termini \( \varepsilon_{\pm\pm} \), and between oppositely charged termini \( \varepsilon_{\pm-} \) have to be optimized. The values are in the dimensionless \( k_B T \) unit.
which has been developed and tested for single and amyloid proteins and shown to be very
efficient \cite{99}. The values are listed in Tab. 3.1.

To facilitate a comparison with the other force fields such as Abeln force-field \cite{100} and
the hydrophobic-polar (HP) model \cite{82, 156, 159, 160}, these $\varepsilon_{aa}$ values were normalized
with respect to the maximal interaction pair, that is Phe-Phe of the $A\beta_{16-22}$ and Ile-Val of the $A\beta_{37-42}$ peptides, respectively. As mention in chapter 2, following the work of Abeln
et al., the numerical values of the H-bond and steric interactions (in the dimensionless $k_B T$
unit) are given by \cite{100, 102, 103}:

$$
\varepsilon_{hb} = -0.5, \quad \varepsilon_s = 0.55.
$$

Unless indicated, the $E_{\text{solvent}}$ term is set to zero. The remain parameters ($\varepsilon_{+-}$, $\varepsilon_{\pm\pm}$, $\varepsilon_{\pm a}$, $\varepsilon_{ss}$) are determined from the intervals

$$
\varepsilon_{+-} \in [-1.4, \cdots, 0], \quad \varepsilon_{\pm\pm} \in [0, \cdots, 0.7],
\varepsilon_{ss} \in [-0.3, \cdots, 0], \quad \varepsilon_{\pm a} \in [0, \cdots, 0.7],
$$

(3.1)

by fitting the equilibrium conformational distributions obtained from the lattice model sim-
ulations to those of all-atom simulations. In principle, one has to vary these four variables simultaneously in the parameterization process. However, with the scanning step of 0.1 in
the above intervals, we have $15 \times 8 \times 8 \times 4 = 3840$ parameter sets, and the calculation is too expensive. Thus, at every step, we vary only one parameter while fix the other values. We
note that the maximal absolute values in the above intervals are chosen empirically: $\varepsilon_{+-} = -1.4$ is larger than the hydrophobic interactions $\varepsilon_{aa}$ to favor the anti-parallel configurations,
$\varepsilon_{\pm\pm} = 0.7$ is larger than the OPEP averaged interactions $\langle \varepsilon_{aa} \rangle = 0.26$ between like-charged residues, and $\varepsilon_{\pm a} = 0.7$ is larger than the OPEP averaged interactions $\langle \varepsilon_{aa} \rangle = 0.11$ between charged - hydrophobic residues. The strand state energy $\varepsilon_{ss} = -0.3$ is strong enough to bias the peptides to the extended conformations.

3.3 Simulation details

3.3.1 Lattice simulations

The initial structures of the $A\beta_{16-22}$ and $A\beta_{37-42}$ dimers, trimers,...and decamers are pre-
pared by translating the monomers, which are in the fully extended state, along the x-axis
by lattice constants, resulting in $n$-stranded $\beta$-sheets ($n = 2, 3, 4, \ldots, 10$). The volume of
the simulation box is $24 \times 24 \times 24$ lattice points for the monomers, $30 \times 30 \times 30$ for the
dimers, $35 \times 35 \times 35$ for the trimers, $38 \times 38 \times 38$ for the 4mers, $41 \times 41 \times 41$ for the
5mers, 43 × 43 × 43 for the 6mers, 46 × 46 × 46 for the 7mers, 48 × 48 × 48 for the 8mers, 50 × 50 × 50 for the 9mers, and 51 × 51 × 51 for the decamers, resulting in the same molecular concentration of 4.5mM for all systems.

To simulate the properties of protein system with lattice model, we used Monte Carlo simulation algorithm where trial moves are accepted according to the Metropolis rule,

\[ P_{\text{acc}} = \min\left\{ 1, \exp\left( \frac{-\Delta E}{k_B T} \right) \right\}, \]

where \( T \) is the simulation temperature, \( k_B \) is the Boltzmann constant and \( -\Delta E \) is the energy difference between the new and the old configuration of the system. The trial moves include internal moves: end move, corner flip, crank shaft, point rotation, side chain orientation, change residue state of a chain, and rigid body moves: rotation and translation between two chains. The corner flip describes a rotation of 180 degrees of a given particle about the line joining its neighbour along the chain. The crankshaft move involves a rotation by 90 degrees of two consecutive particles. A point rotation is a turn, around a randomly chosen pivot particle, of the whole section starting from the pivot particle and going to the end of the chain. For details, readers are referred to Refs.[103, 161].

In addition, the state of each residue may be altered by a local move from strand to coil, and vice versa. The transition from coil to strand is only occurred if the following conditions are satisfied:
1. There is no turn in the backbone at residue \( i \).
2. Side chains of the sequential neighbour are oriented into the opposite direction, if the neighbouring residues are in the strand state.

We should recall that when making the Monte Carlo move to change the state of a residue, since any move is accepted according to the criterion defined in the equation (2.32), the potential energy of hydrogen bonds is taken into account. Residues in the strand state are not allowed to change their backbone configuration or their side chain direction. With these moves we expect to have a good balance between cooperative moves and single particle moves, as well as an efficient sampling of the compact configuration of the protein, which are crucial for the study of the equilibrium properties of the native state.

At each iteration a single local trial move is performed and a global trial move (including point rotations) is performed with the probability \( (P_{\text{global}} = 0.1) \). In all simulations, the program of Abeln et al. is employed [162]. We only modify the code to take into account the state energy of residues, and the four-body H-bond interaction.

Starting from these initial structures, the simulations are carried out in the NVT ensemble at a high temperature of \( T^* = 0.5 \) (in the dimensionless \( k_B T \) unit) for \( 10^4 \) Monte Carlo
(MC) steps. The final structures are used for subsequent replica exchange MC (REMC) simulations. We use 24 replicas for the dimers, trimers and 120 replicas for the decamers, covering the temperature range of $T^* = 0.15 - 0.4$. The simulations are run with $10^{10}$ MC steps for the dimers, trimers and $5 \times 10^{10}$ MC steps for the decamers. The exchange between two neighboring temperatures is attempted every 1000 MC steps, configurations are save every 100 MC steps and the first $10^5$ steps are excluded in the analyses.

### 3.3.2 All-atom simulations

In order to simulate the features of protein systems with an all-atom model, we used replica exchange Molecular Dynamics simulations with various force fields such as AMBER-f99SB-ILDN [97], CHARMM22* [98]. The GROMACS program [163] was used to performed simulations. The simulation systems were the $\text{A}_{16-22}$ and $\text{A}_{37-42}$ with the number of chains from 2 to 3. The initial all-atom conformations of the $\text{A}_{16-22}$ and $\text{A}_{37-42}$ monomers are extracted from the structures of the $\text{A}_{10-35}$ and $\text{A}_{17-42}$ peptides available in the Protein Data Bank (ID: 1hz3 and 2BEG, respectively) [164, 165].

The initial conformations of the dimers and trimers are obtained by replicating the individual monomers in random orientations. We use the AMBER-f99SB-ILDN force field [97] to model the peptides and TIP3P water model to describe the solvent. Starting from the above constructed structure of each system a short simulation of 1 ns is carried out in NPT ensemble followed by another 1 ns NVT simulation. The last structures are taken as the initial structures for replica-exchange molecular dynamics (REMD) simulation. The REMD simulations are carried out with 48 replicas for the dimers and 64 replicas for the trimers, varying from 290 K to 466 K, using the temperature method developed by Patriksson and van der Spoel [166, 167]. Exchanges between neighboring replicas are attempted every 2 ps, leading to an average acceptance ratio of 25%. Each replica is run for 200 ns, data is save every 2 ps and the first 50 ns is excluded in the analyses. The GROMACS program [163] (version:4.6.5) is used to perform the simulations. The bond lengths with hydrogen atoms are fixed with the SHAKE algorithm [168] and the equations of motion are integrated with a time step of 2 fs using the velocity Verlet integrator [105]. The electrostatic interactions are calculated using the particle mesh Ewald method and a cutoff of 1.1 nm [169]. A cutoff of 1.2 nm is used for the Van der Waals interactions. The nonbonded pair lists are updated every 10 fs. Temperatures are controlled by the velocity-rescaling thermostat, developed by Bussi-Donadio-Parinello and found to sample the canonical ensemble [106].
3.4 Results

In the following we present our force field parametrization procedure for the lattice model of the Aβ_{16–22} dimer and compare its structures with those obtained from all-atom simulations. Then, we show that the best parameters can be transferred to the Aβ_{16–22} trimer, the dimer and trimer Aβ_{37–42}. Eventually we applied the force-field to exploring the free energy landscape of Aβ_{16–22} and Aβ_{37–42} decamer as well as their nucleus sizes.

3.4.1 Sampling convergence

The convergence of the OPEP REMC simulations was checked by comparing results obtained using 0.5×10^{10} and 10^{10} MC steps for the dimers, trimers, 2.5×10^{10} and 5×10^{10} MC steps for the decamers. We found that the results are virtually identical as shown in Fig. 3.2 to Fig. 3.7 of this thesis. These figures show, as examples, the distribution of various quantities including the order parameter $P_2$, the total number of inter-chain H-bonds, $N_{H\text{-bond}}$, and the potential energy, $E^*$ obtained from the simulations using the best set of parameters shown in Eq. 3.1 for all systems. As seen, the results obtained using the half and full trajectories are essentially identical in all systems, indicating well-converged simulations.

![Figure 3.2: Sampling convergence of the Aβ_{16–22} dimer REMC simulation. Shown are the distributions of various quantities including the order parameter $P_2$, the total number of inter-chain H-bonds, $N_{H\text{-bond}}$, and the potential energy $E^*$ at room temperature (just below the melting temperature of the monomer) using 0.5×10^{10} (black) and 10^{10} (red) MC steps.](image-url)
Figure 3.3: Sampling convergence of the Aβ_{16-22} trimer REMC simulation. Shown are the distributions of various quantities including the order parameter $P_2$, the total number of inter-chain H-bonds, $N_{H\text{-}bond}$, and the potential energy $E^*$ at room temperature (just below the melting temperature of the monomer) using $0.5 \times 10^{10}$ (black) and $10^{10}$ (red) MC steps.

Figure 3.4: Sampling convergence of the Aβ_{16-22} decamer REMC simulation. Shown are the distributions of various quantities including the order parameter $P_2$, the total number of inter-chain H-bonds, $N_{H\text{-}bond}$, and the potential energy $E^*$ at room temperature (just below the melting temperature of the monomer) using $2.5 \times 10^{10}$ (black) and $5 \times 10^{10}$ (red) MC steps.
Figure 3.5: Sampling convergence of the Aβ<sub>37−42</sub> dimer REMC simulation. Shown are the distributions of various quantities including the order parameter $P_2$, the total number of inter-chain H-bonds, $N_{H\text{-bond}}$, and the potential energy $E^*$ at room temperature (just below the melting temperature of the monomer) using $0.5 \times 10^{10}$ (black) and $10^{10}$ (red) MC steps.

Figure 3.6: Sampling convergence of the Aβ<sub>37−42</sub> trimer REMC simulation. Shown are the distributions of various quantities including the order parameter $P_2$, the total number of inter-chain H-bonds, $N_{H\text{-bond}}$, and the potential energy $E^*$ at room temperature (just below the melting temperature of the monomer) using $0.5 \times 10^{10}$ (black) and $10^{10}$ (red) MC steps.
3.4.2 The need for the force field parametrization

Abeln and coworkers have used their protein lattice model described above and force field (Table 3.2) to study the folding, aggregation and amyloid formation of various proteins [100, 102, 103]. This force-field, named as Abeln force field, is essentially the Thirumalai-Betancourt force field [101] for describing the interactions between amino acids, and the interaction terms between the solvent and amino acids were parametrized by Abeln et al. [102]. They have shown that protein aggregation without water could result in artificial clustering due to the unrealistically attraction between polar residues [102].

To obtain the first impression whether the Abeln force field is applicable for our peptides, we carry out a REMC simulation for the Aβ_{16−22} dimer. Since the interaction parameters of charged termini are not available in this force-field, we first make an approximation by simply replacing (Lys^+, Glu^-) by (Lys, Glu). As in the original model, we also set the state energy \( \varepsilon_{ss} = 0 \). The distribution of the end-to-end distance \( P(d_{ee}) \) of single chains near the melting temperatures indicates that the peptides tend to fold, with the average distance \( \langle d_{ee} \rangle = 0.6 \) nm, much shorter than the fully stretched distance \( d_{ee} = 0.3 \times (7 - 1) = 1.8 \) nm, with 0.3 nm being the lattice constant [Fig.3.8(A)]. However, as shown by the all-atom simulation presented below and also from the other work [170], the Aβ_{16−22} peptide tends
Figure 3.8: The distribution of the end-to-end distance $d_{ee}$ of single chains of the Aβ$_{16-22}$ dimer obtained using (A) the original Abeln force-field, (B) the Abeln force field with parameters $\varepsilon_{+-} = -1.4$, $\varepsilon_{\pm \pm} = 0.7$, $\varepsilon_{\pm a} = 0.2$ and $\varepsilon_{ss} = 0$, (C) the OPEP force-field with the four-body H-bond term, $\varepsilon_{+-} = -1.4$, $\varepsilon_{\pm \pm} = 0.7$, $\varepsilon_{\pm a} = 0.2$, $\varepsilon_{ss} = 0$, and (D) same as (C) but $\varepsilon_{ss} = -0.3$ and no four-body H-bond term.

Table 3.2: The Abeln force field [102] of pairwise interaction matrix between two residues of the Aβ$_{16-22}$ (above) and Aβ$_{37-42}$ (below) peptides. The values are normalized with respect to the maximal hydrophobic - hydrophobic interactions Phe-Phe for Aβ$_{16-22}$ and Val-Val for Aβ$_{37-42}$. The values are in the dimensionless $k_BT$ unit.
to be extended in the dimeric conformations. As the discrepancy could be due to the above approximation, we repeat the simulation but now consider explicitly the charged termini by introducing the parameters: $\varepsilon_{+-} = -1.4$, $\varepsilon_{\pm\pm} = 0.7$, $\varepsilon_{\pm a} = 0.2$, $\varepsilon_{ss} = 0$ (note that the optimal values of these parameters are presented below). However, the distribution $P(d_{ee})$ indicates again that the peptides are still in the compact conformations [Fig.3.8(B)]. We also vary these parameters but the peptides are always in the folded states. Nevertheless, the highly similarity between results shown in Figs.3.8(A), (B) suggest that the conformational distribution should not be affected much by the approximation where the same value $\varepsilon_{\pm a}$ is assigned for the interaction between charged termini and inner residues, regardless the amino acid types.

As the OPEP force-field has the four-body H-bond interaction term [99], we expected that it could bias the lattice $A\beta_{16-22}$ dimer to the more extended structure. This term accounts for the cooperative H-bond energy if tight conditions on sequence-separation, $\Delta$, between four residues are verified. That is, if $(i, j)$ and $(k, l)$ are the residues involved in the two H-bonds, then $\Delta(ijkl) = 1$ if $(k, l) = (i + 1, j + 1)$ (for $\alpha$-helix), or $\Delta(ijkl) = 1$ if $(k, l) = (i + 2, j - 2)$ or $(i + 2, j + 2)$ (for $\beta$-sheets), otherwise $\Delta$ is set to 0. These conditions stabilize secondary structures. Using OPEP and simply setting the interaction strength of the four-body H-bond interaction term equals to half of that of the two-body H-bond term ($\varepsilon_{hb}/2$), we repeat the same simulation and the distribution $P(d_{ee})$ shown in Fig.3.8(C) indicates that the peptides, indeed, favor the extended structures with up to $\approx 29\%$ of population having maximal $d_{ee} = 1.8$ nm. We note, alternatively, the extended structures could also be populated by favoring the strand state of individual residues. With this in mind, we repeat the same simulation using OPEP force-field without the four-body H-bond term, but with the energy state $\varepsilon_{ss} = -0.3$ while the other parameters are as same as above ($\varepsilon_{+-} = -1.4$, $\varepsilon_{\pm\pm} = 0.7$, $\varepsilon_{\pm a} = 0.2$). The distribution $P(d_{ee})$ shows that the peptides are indeed unfolded with $\approx 28.5\%$ of population having maximal $d_{ee} = 1.8$ nm, in agreement with the simulation using four-body H-bond term. However, since the calculation of the four-body H-bond term is rather expensive, in the following simulations we only consider the state energy term whose best value is the subject of our parameterization process.

### 3.4.3 Force field parametrization for $A\beta_{16-22}$ dimer

We begin the parameterization process with the $A\beta_{16-22}$ dimer by varying systematically four parameters ($\varepsilon_{+-}$, $\varepsilon_{\pm\pm}$, $\varepsilon_{\pm a}$, $\varepsilon_{ss}$) from their intervals listed in Eq.3.1. For each set of parameters, we carry out a REMC simulation, calculate the populations of the parallel ($P_{pa}$) and anti-parallel ($P_{ap}$) 2-stranded $\beta$-sheets (see chapter 2, section 2.5.2.6 for definition) and construct their phase diagrams as a function of the varied parameter and temperature.
Suggested by all-atom simulation results, we first start with large values \( \varepsilon_{+-} = -1.4 \), \( \varepsilon_{ss} = -0.3 \) and a small value \( \varepsilon_{\pm a} = 0.2 \) such that the extended anti-parallel \( \beta \)-sheet is favorable. Fig. 3.9(A) shows the dependence of \( P_{ap} \) on \( (\varepsilon_{\pm\pm}, T^*) \). As expected, the dimer is more disordered as the temperature increases. For example, at the lowest temperature \( T^* = 0.15 \), \( P_{ap} \approx 100\% \), and this value decreases to \( \approx 50\% \) around the melting temperature \( T^*_m = 0.25 \). Importantly, no significant variations in \( P_{ap} \) are observed over the interval \( \varepsilon_{\pm\pm} \in [0.0, \cdots, 0.7] \) at a given temperature. This suggests that the structure of the \( A\beta_{16-22} \) dimer is hardly affected by the contact energy between like-charged residues. Thus, in the subsequent parameterization process, we always fix \( \varepsilon_{\pm\pm} = 0.7 \).

With this value, we now vary \( \varepsilon_{\pm a} \) and fix \( \varepsilon_{+-} = -1.4 \), \( \varepsilon_{ss} = -0.3 \) as in the previous step. Fig. 3.9(B) shows the dependence of \( P_{ap} \) on \( (\varepsilon_{\pm a}, T^*) \). Again, the population of the anti-parallel \( \beta \)-sheet is essentially constant over the interval \( \varepsilon_{\pm a} \in [0.0, \cdots, 0.7] \) at a given temperature. This suggests further that it is quite reasonable to also fix this contact energy value at, say \( \varepsilon_{\pm a} = 0.2 \) in the parametrization process. To understand why the populations do not depend on \( (\varepsilon_{\pm\pm}, \varepsilon_{\pm a}) \), we analyze structures in more details and find that the fully parallel or anti-parallel \( \beta \)-sheets are always in register, and obviously, with such conformations there are no \( \varepsilon_{\pm\pm} \) and \( \varepsilon_{\pm a} \) interactions.

Next, the two remained parameters \( (\varepsilon_{+-}, \varepsilon_{ss}) \) are varied from their ranges [Eq.3.1], the first one by increments of 0.1 and the second one by increments of 0.02 in REMC simulations, and the best values are determined from the comparison of the conformational parallel/antiparallel populations obtained from lattice and all-atom simulations.
Figure 3.10: (A) Probability distribution (in %) of the anti-parallel 2-stranded $\beta$-sheets and 3-stranded $\beta$-sheets obtained from all-atom simulations of the $\text{A}\beta_{16-22}$ dimer (red) and trimer (blue), respectively. (B) Probability distribution of the anti-parallel 3-stranded $\beta$-sheets of the $\text{A}\beta_{16-22}$ trimer obtained from lattice simulations (blue) using $\epsilon_{ss} = -0.12$, $\epsilon_{+-} = -1.1$, $\epsilon_{\pm\pm} = 0.7$, $\epsilon_{\pm a} = 0.2$. The square boxes denote the best match between all-atom and lattice simulations.

To this end, we calculate the populations of the parallel and anti-parallel 2-stranded $\beta$-sheets of the $\text{A}\beta_{16-22}$ dimer from the all-atom REMD trajectories, and as seen from Fig.3.10(A) (red line), we obtain $P_{ap} \approx 19\%$ at 300 K. This population decreases quickly and reaches $\approx 2\%$ at 380 K. The parallel 2-stranded $\beta$-sheets are hardly observed with $P_{pa} \leq 0.1\%$ at all temperatures.

Now, the question is that at which temperature of the lattice simulations, results are comparable with that of all-atom simulations done at a physiological temperature, say $T = 300$ K? We note that the physiological temperature of a protein should be close to its folding temperature $T^*_f$, and this suggests us to compare the structure of lattice oligomers simulated at the ”room temperature” which is just below $T^*_f$ of monomer [156], with that obtained from all-atom simulations at 300 K. Therefore, the first task is to determine the folding temperature of the monomer.

With fixed parameters $\epsilon_{\pm\pm} = 0.7$, $\epsilon_{\pm a} = 0.2$, $T^*_f$ depends only on two parameters ($\epsilon_{+-}$, $\epsilon_{ss}$). However, it turns out that for the 7-residue $\text{A}\beta_{16-22}$ monomer, the lattice topology does not allow to have conformations having contact between two termini [Fig.3.1(A)], thus $T^*_f$ actually depends only on $\epsilon_{ss}$. Thus, setting $\epsilon_{+-} = -1.4$, we obtained: $T^*_f(\epsilon_{ss} = -0.3) = 0.1856$, $T^*_f(\epsilon_{ss} = -0.2) = 0.1779$ and $T^*_f(\epsilon_{ss} = -0.12) = 0.1724$ [Fig.3.11(A)]. First, we carried out a REMC simulation for the $\text{A}\beta_{16-22}$ monomer using the these parameters and
we calculated the heat capacity based on the equation (2.35) (see chapter 2, section 2.5.1). The folding temperature $T_f^*$ was determined by the peak of the heat capacity curve.

Having located $T_f^*$, we next calculate the populations of the parallel and anti-parallel $\beta$-sheets of the $\text{A}_{16-22}$ dimer at temperatures just below $T_f^*$, and Fig.3.12 shows $P_{\text{pa}}$ and $P_{\text{ap}}$ as a function of ($\varepsilon_{+-}$, $\varepsilon_{ss}$). To be consistent with the all-atom simulation, we only focus on $\varepsilon_{+-} \leq -1$ where the population of the parallel $\beta$-sheet is very low ($\leq 0.1\%$). With this, the population of the anti-parallel 2-stranded $\beta$-sheet is too high, $P_{\text{ap}} \approx 80 \%$ with $\varepsilon_{ss} = -0.3$, and $\approx 50 \%$ with $\varepsilon_{ss} = -0.2$. We find that at $\varepsilon_{ss} = -0.12$, and $\varepsilon_{+-} = -1.1$ the population $P_{\text{ap}} \approx 16 \%$, which is in very good agreement with the all-atom population (19%). This suggests the best parameters for the lattice model of the $\text{A}_{16-22}$ dimer are:

$$
\varepsilon_{+-} = -1.1, \quad \varepsilon_{\pm\pm} = 0.7,
\varepsilon_{ss} = -0.12, \quad \varepsilon_{\pm\alpha} = 0.2.
$$

(3.3)

3.4.4 Force field validation for $\text{A}_{16-22}$ trimer, $\text{A}_{37-42}$ dimer, and $\text{A}_{37-42}$ trimer

To verify whether the optimal parameters set of the $\text{A}_{16-22}$ dimer, [Eq.3.3], is applicable for the other oligomers, we carry out lattice simulations using these parameters for the
Figure 3.12: The population (in %) of the parallel and anti-parallel β-sheets of the Aβ₁₆–₂₂ dimer as a function of ε⁺⁻ given εₛₛ = -0.12 (magenta, blue), εₛₛ = -0.2 (black, red) and εₛₛ = -0.3 (cyan, green). Here, the parallel conformations are shown in magenta, black, cyan and the anti-parallel conformations are shown in blue, red and green. Shown are results of the lattice simulations at room temperatures $T^*(\epsilon_{ss} = -0.12) = 0.1705$ ($< T^*_f = 0.1724$), $T^*(\epsilon_{ss} = -0.2) = 0.1705$ and $T^*(\epsilon_{ss} = -0.3) = 0.1779$. In all cases, $\epsilon_{\pm\pm} = 0.7$, $\epsilon_{\pm a} = 0.2$. The square box denotes the best population of 16% corresponding to $\epsilon_{ss} = -0.12$, $\epsilon_{+\mp} = -1.1$ match with all-atom REMD simulations at 300K.
Aβ₁₆₋₂₂ trimer as well as for the Aβ₃₇₋₄₂ dimer and trimer, and compare the structures with that of all-atom simulations.

The all-atom REMD simulation of the Aβ₁₆₋₂₂ trimer shows that the parallel 3-stranded β-sheets are hardly populated ($P_{pa} \leq 0.1\%$). The temperature-dependence of the population of the anti-parallel 3-stranded β-sheets shown in Fig.3.10(A, blue line) reveals that at 300 K we obtain $P_{ap} \approx 51\%$, and decreases slowly as temperature increases, with $P_{ap} \approx 19\%$ at 380 K, reflecting that the trimer is more stable than the dimer.

For the lattice model of the Aβ₁₆₋₂₂ trimer, the simulation shows that the parallel 3-stranded β-sheet is zero populated, consistent with the all-atom simulation. As shown in Fig.3.10, the population of the anti-parallel 3-stranded β-sheets $P_{ap} \approx 57\%$ at room temperature $T^*_f = 0.1705$, which is surprisingly in very good agreement with that of the all-atom simulation (51%). As a test case, we carry out an additional simulation with $\epsilon_{ss} = -0.2$ while fixing the other parameters and find that $P_{ap} \approx 75\%$, too high as compared to the all-atom value. This confirms that $\epsilon_{ss} = -0.12$ is, indeed, also the best parameter for the Aβ₁₆₋₂₂ trimer, suggesting that the optimal parameters of the Aβ₁₆₋₂₂ dimer are transferable to the trimer counterpart.

Similarly with Aβ₁₆₋₂₂, to compare the results from lattice simulations with those from all-atom simulations, we need to estimate the folding temperature of a monomer of Aβ₃₇₋₄₂. Figure 3.13 shows the heat capacity of Aβ₃₇₋₄₂ using the best parameters given by [Eq.3.3]. From the monomer’s heat capacity curve, we obtained the folding temperature of the monomer Aβ₃₇₋₄₂ is $T^*_f = 0.1922$, so the nearest-below-folding- temperature of monomer Aβ₃₇₋₄₂ is $T^*_f = 0.1821$. This temperature will be used to compare results with room temperature from all-atom simulation.

Figs.3.14 (A), (B) show the distribution of the order parameters $P_2$ [Eq.2.38] of the Aβ₃₇₋₄₂ dimer and trimer obtained from the all-atom REMD simulations at 300 K. As seen, the average value $P_2 \leq 0.3$, implying that these systems are quite disordered. A closer inspection reveals that the populations of the parallel and anti-parallel β-sheets are almost zeros. In this context, these two systems could serve as good test cases to verify whether the $(\epsilon_{ss}, \epsilon_{s+-})$ interactions bias the oligomers to the β-sheets as seen with the Aβ₁₆₋₂₂ systems. Using the optimal parameters of the Aβ₁₆₋₂₂ dimer [Eq.3.3], the simulations show that both lattice Aβ₃₇₋₄₂ dimer and trimer are, indeed, disordered with $P_2 \approx 0.2$ [Fig.3.14 (C), (D)], chains are largely in the folded states, and the parallel, anti-parallel β-sheets are not populated. This is consistent with the all-atom simulation results, indicating that the energy state $\epsilon_{ss} = -0.12$ does not artificially bias the Aβ₃₇₋₄₂ oligomers to the extended states, and confirms further that the parametrized parameters of the Aβ₁₆₋₂₂ dimer are transferable to another peptide sequence.
Figure 3.13: The constant volume heat capacity of the $\text{A}\beta_{37-42}$ monomer (a), dimer (b), trimer (c) and decamer (d). Shown are results obtained from simulations using the parameters set listed in Eq. 3.3 of this thesis.

Figure 3.14: Probability distribution (in %) of the order parameters $P_2$ of the $\text{A}\beta_{37-42}$ dimer (A), (C) and and trimer (B), (D). Shown are results obtained from the all-atom REMD simulations at 300 K (A), (B) and from the lattice simulations at $T^* = 0.1821$ using $\epsilon_{\pm a} = 0.2$, $\epsilon_{\pm\pm} = 0.7$, $\epsilon_{+-} = -1.1$, $\epsilon_{ss} = -0.12$ (C), (D).
3.4.5 Applications: Free energy landscape of $A\beta_{16-22}$ and $A\beta_{37-42}$ decamer

Figure 3.15: The free energy landscape (in $k_B T^*$) of the $A\beta_{16-22}$ decamer as a function of the order parameter $P_2$ and the total number of inter-hydrogen bonds at the room temperature $T^* (\varepsilon_{ss} = -0.12) = 0.1711$. The representative structures at the center of each state are also shown.

Encouraged by the fact that the best OPEP parameters of the $A\beta_{16-22}$ dimer are transferable for the $A\beta_{16-22}$ trimer as well as the $A\beta_{37-42}$ oligomers, we now in the position to study larger oligomers. Adopting the $A\beta_{16-22}$ and $A\beta_{37-42}$ decamers as two representative examples, we carry out REMC simulations and characterize their equilibrium structures. We remind that for both systems the OPEP force-field [Tab. 3.1] is employed to describe the interactions between amino acids, and the remained parameters are listed in Eq.3.3. Figs.3.15 and 3.16 show the free energy landscapes $F(N_{H\text{-bond}}, P_2)$ as a function of the total number of inter-chain H-bonds, $N_{H\text{-bond}}$, and the order parameter $P_2$ at room temperatures $T^* = 0.1711$ and $T^* = 0.1905$ which are close to the melting temperatures of the corresponding monomers $A\beta_{16-22}$ and $A\beta_{37-42}$, respectively [Figs.3.11 - 3.13]. The representative structures pertained to the free energy minima denoted as $S_i$ are also shown.
Figure 3.16: The free energy landscape \((k_B T^*)\) of the \(A\beta_{37-42}\) decamer as a function of the order parameter \(P_2\) and the total number of inter-hydrogen bonds at the room temperature \(T^*(\varepsilon_{ss} = -0.12) = 0.1905\). The representative structures at the center of each state are also shown.
As seen from Fig. 3.15, the free energy landscape (FEL) of the Aβ_{16−22} decamer exhibits several free energy minima, representing ≈ 99% of the ensemble, with fibril-like features characterized by large P2 values (≥ 0.8) and a high number of interpeptide H-bonds (N_{H−bond} ≥ 30). Suggested by experimental fibril structures where β-sheets are highly ordered, we consider a lattice fibril is fully formed if chains are fully stretched and there are at least 5 H-bonds (out of maximal value of 7 H-bonds) between two consecutive chains, and we obtain 19% and 13% of populations forming one anti-parallel 10-stranded β-sheet (state S1) and two anti-parallel 5-stranded β-sheets (state S2), respectively. Note that S1 and S2 match the experimental microcrystal and solid-state NMR structures [125, 143]. The remained conformations (54%) in this deep minimum exhibit fibril topology like S1 and S2 but only 3-4 H-bonds are formed between two consecutive chains, and can be considered as prototype fibril structures. We also find in this minimum two anti-parallel β-sheets, each contains n-strands and (10 − n)-strands (n = 1, ⋅⋅⋅, 8) (S3, 12%). The intermediate state (S4) is very low populated (≤ 1%) with a two-layer-like architecture where the first layer is formed by several (typically 3-4) fully anti-parallel chains and the second layer consists of partially folded chains. Interestingly, the state S5 (≤ 1%) can be considered as a compact counterpart of the two anti-parallel 5-stranded β-sheets shown in S2. Indeed, imaging that if each fully stretched chain in S2 folds to the β-hairpin structures with the turn at the residues 4 - 5 then we get state S5. Overall, the very high probability of fibrillar topologies indicate that the nucleus size of the Aβ_{16−22} peptide is about 10 chains.

Similar to the results of the dimers and trimers, the FEL of the Aβ_{37−42} decamer (Fig. 3. 16) shows that this oligomer is also more disordered than Aβ_{16−22} decamer, with up to ≈ 91% of population having P2 ≤ 0.3 and N_{H−bond} ≤ 12. Here, chains are folded into β-hairpin-like conformations with turns at the residues 3-4 or 4-5, and packed in different architectures. In the first state S1 (29.7%), five hairpins stack on the top of each other, stabilized by H-bonds between strands as well as by oppositely charged termini interactions. The other four hairpins form two layers, each consists of two hairpins in the same plan and interact via strand-strand contacts. The state S2 (23.7%) consists of two perpendicular fibrils, each is formed by five stacked hairpins. The state S3 (38.1%) composed of three parallel layers, each is formed by two or four hairpins in the same plan. The intermediate states S4, S5 and S6 (3%) are basically formed by two layers, chains in each layer are more extended (typically four out of six residues form extended structures) as compared to the hairpin structures, but still not fully stretched. Also, the chains are parallel or anti-parallel within one layer or between the two layers.

A detailed analysis of the ordered state (P2 ≥ 0.8, N_{H−bond} ≥ 25) reveals structures with two fully anti-parallel 5-stranded β-sheets with five antiparallel β-strands (S7, ≈ 1%) or...
Figure 3.17: The population of fibril state as a function of Aβ\textsubscript{16−22} chains at T = 0.1711 (the folding temperature of monomer, below the melting temperature of the aggregate).

five parallel β -strands (S8, ≈ 2%) within individual sheets. The remained conformations (≈ 2.5%) are formed by two β-sheets with mixed antiparallel and parallel β -strands. We note that S7 is similar to the state S2 of the Aβ\textsubscript{16−22} decamer. Most importantly, we are able to capture the experimental microcrystal fibril structure (state S8) where the β-strands are parallel with the two end terminus identically charged residues in contact [125, 158]. Our results show that the Aβ\textsubscript{37−42} decamer is still largely disordered, implying that the nucleus size for the fibril formation should be larger than 10 chains.

3.4.6 Nucleus size for Aβ\textsubscript{16−22} and Aβ\textsubscript{37−42} peptides

According to classical nucleation picture, the free energy profile as a function of aggregate sizes increases from the monomer up to a maximum, after which it monotonically decreases under supersaturated conditions. The point of the maximum free energy is related to the size of the critical nucleus [18]. The interpretation of the experimental sigmoidal kinetic profiles of amyloid formation has shown, however, that it is necessary to integrate secondary nucleation processes [77]. Here, to determine the exact nucleus size for the Aβ\textsubscript{16−22} peptide, we have performed additional REMC simulations for all aggregate sizes between 4-mer and 12-mer starting from disordered states at the folding temperature of the monomer. The population of fibril states with $P2 \geq 0.8$ as a function of the number of chains is shown on Fig.3.17. Based on the argument that the beta-sheet oligomers are stable only if larger than a critical nucleus size [77], the convergence of the well-ordered cross-beta structure population to 95 %, 99% and 100% for the 9-mer, 10-mer and 11-mer systems suggests a nucleus size of 10 chains for Aβ\textsubscript{16−22} peptide. Standard Monte Carlo simulations remain to be performed to characterize the nucleus and its detailed structure.
3.5 Discussions

In prior studies, Li and coworkers employed the H-P model [171], where all amino acid types are classified as either hydrophobic (H) or polar (P), to identify the aggregate-prone monomeric conformations and the factors governing fibrillogenesis [156, 159]. Irback and coworkers developed further this model by introducing the orientation of the side chains [172]. These models, however, do not account for side-chain specific interactions which are important for studying specific protein sequences. Abeln et al. have developed a more sophisticated lattice model by including specific side-chain interactions and the formation of backbone hydrogen bonds [100, 102, 103]. Notably, they studied the fibril formation of a model amyloid peptide with alternating polar and hydrophobic amino acids [100] and the interplay between folding and assembly of a silk-collagen fibril model [173]. In this work, we first coupled the Abeln lattice model with the OPEP force field and parametrized the missing parameters on the amyloid \( A\beta_{16-22} \) dimer. Then we demonstrated the suitability of all parameters on three other systems differing in sizes and sequences with the final aim to determine the nucleus size for amyloid fibril formation of both peptides.

Our bottom-up approach starts with the determination of the best lattice force parameters for the \( A\beta_{16-22} \) dimer by fitting its equilibrium parallel and anti-parallel \( \beta \)-sheet populations to all-atom REMD simulation results and then checks the transferability of the parameters on the myloid \( A\beta_{16-22} \) trimer and the amyloid \( A\beta_{37-42} \) peptide in dimer and trimer states by comparing the degree of ordered and disordered conformations with all-atom AMBER-f99SB-ILDN/TIP3P simulations. We showed that the OPEP four-body term is important to capture properly the equilibrium structures of the dimer and its contribution can be mimicked by the energy for \( \beta \)-strand of individual residues. Based on the present simulations and off-lattice OPEP simulations, all amino acids except proline have the same energy for \( \beta \)-strand [151].

There are several bottom-up approaches to systematically parameterize the coarse-grained potentials matching the behavior of all-atom simulations or experimental data [104, 174, 175, 176]. These include the relative entropy method by Shell [174], the iterative Boltzmann inversion method by Mueller-Plathe [175] and the multiscale method by Voth [176]. The relative entropy method searches for potentials that minimize the relative entropy \( S = P_T ln(\frac{P_T}{P_M}) \), where \( P_T \) and \( P_M \) are the probabilities of the configuration space in the all-atom and coarse-grained ensembles, respectively. Our approach can be considered as a simple version of the relative entropy method, where \( P_T \approx P_M \). The iterative Boltzmann inversion method starts with an initial guess of the potential and a set of collective variables used to compare the reference and coarse-grained force fields. The probability
distribution of these variables in the coarse-grained model is solved iteratively to converge to its counterpart of the reference model. Again, our approach can also be considered as a simplified version of this approach.

In the present study, we used the AMBER-f99SB-ILDN force field [97] to model the peptides and TIP3P water model to describe the solvent in the atomistic peptide REMD simulations. To determine whether the oligomer structures and populations vary with the all-atom force field used, we carried out REMD simulations (150 ns/replicas) using the CHARMM22* [98] and AMBER-f99SB-ILDN force fields on the monomer and trimer of the $A\beta_{16-22}$ peptide. These two force fields have been shown recently to perform consistently well in reproducing the experimental data on a set of folded proteins with various secondary structure compositions [98, 177, 178]. We found that the conformational ensembles of the $A\beta_{16-22}$ monomer and trimer obtained from these two force fields are very similar (Figs.3.18, 3.19, 3.20) indicating that our lattice force field parameters are also consistent with CHARMM22* simulations results. For example, Fig.3.18 shows the conformational distribution at 300 K of various quantities such as the end-to-end distance $d_{ee}$, the radius of gyration, the solvent accessible surface area and the secondary structures. As seen, the two force fields yield similar results. For instance, the distribution of the end-to-end distance obtained from the two force fields reveals two main states with $d_{ee} \approx 0.8$ nm and $d_{ee} \approx 1.7$ nm corresponding to the folded and extended structures, respectively. The free energy landscapes of the monomer obtained from the two force fields are also quite similar as shown in Fig. 3.19. Here, the FEL is plotted as a function of the first two principal components (V1,V2) obtained from the dihedral angle principal component analysis (dPCA) [122]. Note that we used the same components V1, V2 in Fig. 3.19 by combining the two simulations. For the trimer, the distributions of various structural quantities including the order parameter $P_2$, the cosine of the angle between two end-to-end vectors, the inter-chain side-chain contacts, the solvent accessible surface area and the secondary structures, shown in Fig. 3.20, are very similar for the two force fields. Both force fields favor anti-parallel peptides [large negative cosine, Fig. 3.20(b)] with high beta-strand content (Fig. 3.20(e) and no $\alpha$-helix structure (Fig. 3.20(f)).

Using the same set of lattice protein parameters that work well on dimers and trimers of both peptides, our replica exchange Monte Carlo simulations shows that $A\beta_{16-22}$ decamer forms a stable fibril with two anti-parallel 5-stranded $\beta$-sheets that matches the experimental structure. Simulations for all aggregates varying between 2 and 12 chains suggest that the nucleus size of the $A\beta_{16-22}$ peptide is 10 chains. In contrast, the $A\beta_{37-42}$ decamer is still largely disordered, implying that the nucleus size for fibril formation is much larger.
Figure 3.18: The distribution of various quantities of the $\alpha$-helix monomer including the end-to-end distance (dee), the radius of gyration ($R_g$), the solvent accessible surface area (SASA), and the secondary structures along residues. Shown are results obtained from REMD simulations at 300 K using the AMBER-f99SB-ILDN (black) and CHARMM22* (red) force fields.
than 10 chains. Encouragingly for the latter peptide, we are able to obtain $\approx 2\%$ of conformational states consisting of two anti-parallel 5-stranded $\beta$-sheets with the peptides in each $\beta$-sheet parallel to each other (while the peptides in each $\beta$-sheet are antiparallel in $A\beta_{16-22}$, matching again the experimental structure [125, 158, 157].

It is useful to compare our $n^*$ estimates with previous computational or experimental data. For example, Nasica-Labouze et al. studied the GNNQQNY fragment as a function of the number of monomers from coarse-grained OPEP MD simulations [141] and showed that $n^* = 5 - 6$ chains at 300 K. Hills et al. obtained $n^* = 5$ chains for the peptide STVIYE from all-atom simulation [179]. Rohrig et al. followed the stability of preformed $\beta$-sheet oligomers of different sizes by all-atom MD simulations and suggested that $n^* = 8 - 16$ chains for the $A\beta_{16-22}$ peptide [180]. Bafitzadeh et al. determined the free energy per chain as a function of three reactions coordinates using bias-exchange metadynamics of the 18 chains of Poly-Valine, each 8 residues long, and proposed an effective nucleus size $n^*$ on the order of 14 chains [181]. Nussinov et al. proposed that $n^*$ is 6 for the GVIGIAQ peptide based on all-atom simulations [182]. Using on-lattice Monte Carlo simulations, Co et al. have determined, from the number of forming template peptides above which the time to add a new monomer is independent of the template size, that $n^*$ is 11 chains for a 8-bead sequence modeling the $A\beta_{1-40}$ peptide [82], and De Simone et al. found that $n^*$ is at least equal to 9 using linear peptides of 7 amino acids with an alternative hydrophobic
Figure 3.20: The distribution of various quantities of the \( A\beta_{16-22} \) trimer at 300 K, including the order parameter \( P_2 \) (a), the cosine of the angle between two end-to-end vectors (b), the inter-chain side-chain contacts \( (N_c) \) (c), the solvent accessible surface area (SASA)(d) and the secondary structures (e - h) along residues, averaged over three chains. Shown are results obtained from REMD simulations using the AMBER-f99SB-ILDN (black) and CHARMM22* (red) force fields. The secondary structures are calculated using STRIDE [183].
and hydrophilic pattern [184]. Overall, our \( n^* \) estimates of 10 chains for \( \text{A}\beta_{16-22} \) and > 10 chains for \( \text{A}\beta_{37-42} \) are consistent with all these previous simulations, and two independent theoretical approaches that establish that the most likely range of \( n^* \) is between 7 and 14 [185]. We must keep in mind however that the nucleus size or critical \( \beta \)-domain size with a probability of 50\% to form a fibril can vary from 4 to 35 chains depending on the population of the amyloid-competent monomeric state [186] and the population of the latter state is modulated by the balance between electrostatic and hydrophobic interactions [82, 159], so a generalization of this estimate to any amyloid sequence is not possible. In addition, variation in the supersaturation of the phase can cause \( n^* \) to increase from 15 to 40 chains of \( \text{A}\beta_{1-40} \) as determined by atomistic nucleation theory [85], in agreement with the experimental estimate (\( n^* > 29 \)) using fluorescence correlation spectroscopy at a supersaturation of 100 \( \mu \text{M} \) \( \text{A}\beta_{1-40} \) solution [187]. Finally, it has to be emphasized that the present lattice model has several limitations because, at variance with experimental studies and many off-lattice simulations, we cannot observe \( \alpha \)-helical and \( \beta \)-barrel intermediates [125, 188, 189]. We are currently employing our model and force field to estimate the exact nucleus size for the \( \text{A}\beta_{37-42} \) peptide and other amyloid sequences to help design more efficient drugs against Alzheimer’s disease [88, 89].

### 3.6 Conclusions

We have presented a comprehensive force-field parameterization strategy for the protein lattice model with the primary aim is to study structure and aggregation mechanism of large oligomers. The force field of our lattice model includes the interactions between (i) inner amino acids described by the OPEP force field [104], (ii) like-charged termini \( \epsilon_{\pm\pm} \), (iii) oppositely charged termini \( \epsilon_{\pm-} \), (iv) charged termini and inner residues \( \epsilon_{\pm a} \), and (v) the state energy of individual residue \( \epsilon_{ss} \). Adopting the \( \text{A}\beta_{16-22} \) dimer as a prototypical example, the parametrization is carried out by fitting the populations of the parallel and anti-parallel \( \beta \)-sheets obtained from the lattice simulations to that of all-atom simulations. With the parameters set

\[
\epsilon_{\pm a} = 0.2, \epsilon_{\pm\pm} = 0.7, \epsilon_{ss} = -0.12, \epsilon_{+-} = -1.1,
\]

we obtain the best agreement with the populations of the anti-parallel 2-stranded \( \beta \)-sheets are \( \approx 16\% \) and \( \approx 19\% \) for the lattice and all-atom models, respectively. Using these parameters, the lattice simulations show that the \( \text{A}\beta_{16-22} \) trimer does not form parallel 3-stranded \( \beta \)-sheets, while the population of the anti-parallel 3-stranded \( \beta \)-sheets is \( \approx 57\% \). The lattice simulations of the \( \text{A}\beta_{37-42} \) dimer and trimer show that these oligomers are essentially
disordered. All these results are consistent with the all-atom simulations, suggesting that our force-field parameters might be transferable to the other systems. With this in mind, we study the $\alpha \beta_{16-22}$ and $\alpha \beta_{37-42}$ decamers, and show that $\alpha \beta_{16-22}$ decamer forms a stable fibril with two anti-parallel 5-stranded $\beta$-sheets, consistent with experimental structure. This indicates that the nucleus size of the $\alpha \beta_{16-22}$ peptide is about 10 chains. Pushing the simulations for aggregates between 4-mer and 12-mer suggests a nucleus size for fibril formation of 10 chains. In contrast, the $\alpha \beta_{37-42}$ decamer is still largely disordered, implying that the nucleus size for the fibril formation should be larger than 10 chains. Encouragingly, we are able to obtain $\approx 2\%$ of population consisting of two anti-parallel 5-stranded $\beta$-sheets and chains in each $\beta$-sheet are parallel with each other, capturing the experimental structure [125, 158]. We are currently studying larger $\alpha \beta_{37-42}$ oligomers to identify the nucleus size.
Chapter 4

Results from all-atom model: Structures of $A\beta_{1-40}$ dimer Wild-Type

Sumary

We have studied the dimer of amyloid beta peptide $A\beta$ of 40 residues by means of all-atom replica exchange molecular dynamics. The $A\beta$-dimers have been found to be the smallest toxic species in Alzheimers disease, but its inherent flexibilities have precluded structural characterization by experimental methods. Though the 24-µs-scale simulation reveals a mean secondary structure of 18% $\beta$-strand and 10% $\alpha$ helix, we find transient configurations with an unstructured N-terminus and multiple $\beta$-hairpins spanning residues 17-21 and 30-36, but the antiparallel and perpendicular peptide orientations are preferred over the parallel organization. Short-lived conformational states also consist of all $\alpha$ topologies, and one compact peptide with $\beta$-sheet structure stabilized by a rather extended peptide with $\alpha$-helical content. Overall, this first all-atom study provides insights into the equilibrium structure of the $A\beta_{1-40}$ dimer in aqueous solution, opening a new avenue for a comprehensive understanding of the impact of pathogenic and protective mutations in early-stage Alzheimers disease on a molecular level. The contents of this chapter have been published in our recent paper: Bogdan Tarus, Thanh T. Tran, Jessica Nasica-Labouze, Fabio Sterpone, Phuong H. Nguyen, and Philippe Derreumaux, "Structures of the Alzheimer’s Wild-Type $A\beta_{1-40}$ Dimer from Atomistic Simulations", The Journal of Physical Chemistry B, 119, 10478-10487, 2015, DOI: 10.1021/acs.jpch.5b05593 (July 2015) [90].

Sommaire

Nous avons étudié le dimère du peptide $A\beta_{1-40}$ par simulations de dynamique moléculaire par échange de répliques. Le dimère $A\beta_{1-40}$ est la plus entité toxique dans la maladie d’Alzheimer, et sa flexibilité a empêché toute détermination structurale par les méthodes...
4.1 Introduction

Alzheimer's disease features extracellular plaques in the hippocampus and cortex of the human brain. The major component of senile fibrils is the amyloid beta protein (Aβ) of 39-43 amino acids, although post-translational modifications of Aβ peptides with the cyclizing of E3 or E11 to a pyroglumate are also observed [18]. Interfering with Aβ production, self-assembly, and clearance is a potential treatment for preventing or delaying the onset of the disease [18, 30].

The Aβ$_{1-40}$ peptide of sequence DAEFRHDSGYEVHHQKLVFFAEDVGSNKGAIGLMVGGVV has a hydrophilic patch E22-G29 (loop region) separating two hydrophobic patches at L17-A21 (central hydrophobic core, CHC) and A30-V40 (C-terminus). The N-terminus spanning residues D1-K16 is also very hydrophilic, with residues D1, H6, H13, and H14 found to play a dominant role in metal ion interactions [18]. It has been shown experimentally that toxicity results from low- and high-molecular-mass oligomers as well as fibril fragmentation [18, 190]. Since Aβ dimers are the smallest species to induce cognitive deficits [191], τ-hyperphosphorylation, and neuritic degeneration [33], their structures have been the focus of many studies.

Very little information has emerged from experimental studies because the Aβ$_{1-40}$ dimer exists in equilibrium with monomers and higher-molecular-mass oligomers. Using photo-induced cross-linking and circular dichroism (CD), it was shown the dimer of Aβ$_{1-40}$ has α-helix content of 10.5% and β-strand content of 38% [192]. The secondary structure is, however, sensitive to the experimental conditions [192, 193]. Also, two ion mobility mass spectrometry (IM-MS) studies provided cross collision sections of 1142 and 1245 Å$^2$ for the Aβ$_{1-40}$ dimer [32, 194] and detected two alternative structural forms in an ensemble of low-molecular-mass Aβ$_{1-40}$ oligomers [194]. Finally, Fourier transform infrared spectroscopy and solid-state NMR experiments indicate that small oligomers have some antiparallel β-sheet structure rather than the parallel β-structure as observed in amyloid fibrils [18].

Due to the experimental limitations, several computational studies have been performed on the Aβ$_{1-40}$ dimer using either coarse-grained (CG) protein models or all-atom representations in implicit solvent. The most recent simulations include 900 ns/replica Hamiltonian replica exchange molecular dynamics (REMD) with the six-bead CG OPEP force field [104, 71], all-atom REMD with a solvent-accessible surface area implicit solvent [69], discontinuous MD with a four-bead CG model [195], and all-atom MD simulations starting from CG DMD structures [72]. The results of these simulations, however, show many discrepancies in the total and per residue percentages of secondary structures and the 3D
intramolecular/intermolecular structures [18]. So we still lack a detailed understanding of the structures and dynamics of the Aβ1−40 dimer.

In this study we have characterized the intrinsic disorder of the Aβ1−40 dimer by means of extensive REMD simulations, 400 ns per replica, in a physiological buffer using the all-atom CHARMM22* force field. CHARMM22*, applied with success to the folding of structurally diverse proteins by MD [98] and simulated tempering [196], is considered to be one of the best force fields, at least for folded proteins. REMD simulations with CHARMM22* also led to results consistent with low-resolution data for the Aβ1−28 monomer [188] and helped clarify why the NQTrp molecule is not an ideal inhibitor of amyloid formation and toxicity [197, 198].

4.2 Simulation details

The dimer was built by using the third centroid with a population of 5% from the microsecond/replica simulation of the Aβ1−40 monomer [60]. This structure (Figure 4.1) has a disordered N-terminus and residues 17-23 and 31-36 forming a β-hairpin with a turn at positions 24-29. The second chain was randomly oriented with a distance of 5.6 Å between the peptide centers of mass. Next, the dimer was centered in a truncated octahedron box of 214 nm³ with 6684 TIP3P water molecules [199], resulting in a peptide concentration of 15.5 mM. To mimic a 20 mM phosphate buffer, often used in amyloid experimental studies, we added one H2PO4− ion and one H2PO2− ion. The peptide at pH 7 has NH3+ and CO2− termini, protonated Arg and Lys, deprotonated Glu and Asp, and neutral His with a protonated Ne atom. Finally, the system was neutralized by adding 9K+ ions, resulting in 21270 atoms.

The GROMACS program was used with periodic boundary conditions, and the bond lengths with hydrogen atoms were fixed with SHAKE, allowing a time step of 2 fs using the velocity Verlet integrator [200]. The electrostatic interactions were determined with the particle mesh Ewald method and a cutoff of 1.1 nm [201]. A cutoff of 1.2 nm was employed for the van der Waals interactions. The nonbonded pair lists were updated every 10 fs. The velocity-rescaling thermostat found to sample the canonical ensemble [106] was used to control the temperatures. REMD was carried out with 60 replicas from 300 to 448 K using the temperature-predictor method [166]. Exchanges between neighboring replicas were attempted every 2 ps, leading to a mean acceptance ratio of 25%. Each replica ran for 400 ns.
Figure 4.1: Structures of Aβ₁₋₄₀ dimer at 315 K. The Cα positions of D1 are represented by a ball. The structure at time 0 ns (a) consists of two identical peptide conformations, randomly oriented and separated by two solvation shells. At time 50 ns (b), the chains have different intramolecular structures and the radius of gyration of the dimer is 1.2 nm vs 1.6 nm at t = 0 ns.
4.3 Data analysis

4.3.1 Secondary structure

The secondary structure was calculated using the STRIDE program [112]. The percentage of $3_{10}$-helix amounting to a mean value of 2% is included in the $\alpha$-helix value, and that of $\pi$-helix is negligible in the present simulation.

4.3.2 $\beta$-hairpin

A $\beta$-hairpin was defined if there were at least two backbone H-bonds formed between consecutive $\beta$-strands and at least three consecutive residues belonging to the Ramachandran $\beta$-strand region in each strand.

4.3.3 Clustering

The dimer conformations were analyzed using the principal component analysis (PCA) of the inverse distances between $C_\alpha$ atoms [123]. In our system, the first three principal components account for 60% of the fluctuations. To identify the clusters in this subspace, we used the method described in ref [123]. Statistical errors were estimated by time interval averaging.

4.3.4 Solvent Accessible Surface Area

In all-atom conformation, we calculated the solvent-accessible surface area (SASA) per amino acid as implemented in GROMACS.

4.3.5 Collision Cross Section, CCS.

The CCS was determined by the trajectory method of the MOBCAL software which treats the molecule as a collection of atoms represented by a 12-6-4 potential [202]. This method used for monomeric proteins [202] has led to values consistent with experiments for the simulations of A$\beta_{1-40}$ dimer with D7N and H6R mutations [203, 204].

4.4 Results

The starting structure is based on a previous REMD simulation of the A$\beta_{1-40}$ monomer using the OPLS and TIP3P force fields [60]. This choice is justified because it would take more than several microseconds per replica to capture equilibrium states from a randomly
chosen intramolecular state of each peptide. Furthermore, though dimerization affects the internal structure of each peptide, there is experimental evidence that this $\beta$-hairpin spanning CHC and residues 31-36 persists upon association of the $A\beta_{1-40}$ monomer to the homodimeric $Z\alpha\beta_3$ protein of 58 residues [61]. Also the dimer at 50 ns and $T = 315$ K displays two different intramolecular conformations (Figure 4.1b). One chain has a $\beta$-hairpin formed by residues 18-19 and 36-37 whereas the other chain has residues 8-29 in turn and residues 31-36 in the $\alpha$-helix. In addition, at 50 ns the first 28 replicas (300-361 K) show 5-32% $\beta$-strand, 0-19% $\alpha$-helix, 27-62% turn, and 17-44% coil (Figure 4.2), indicating that our simulations are not biased toward a specific intramolecular conformation.

REMD convergence at 315 K, near the physiological temperature, was assessed by seven metrics calculated over time intervals of 50-300 and 50-400 ns. These include the radius of gyration ($R_g$) of each chain and the dimer, the end-to-end distance between the $C_\alpha$ atoms of the first and last residues, the total number of residues that are in contact via backbone-backbone (BB-BB) and side-chain-side-chain (SC-SC) interpeptide interactions, the solvent-accessible surface area of each amino acid, and the percentage of each amino

Figure 4.2: Secondary structure compositions of the first 28 replicas from 300 to 361 K. The values at 315 K are highlighted in red. For clarity, we present the population of each secondary structure on a different scale.
It can be seen in Figure 4.3 that the system has reached equilibrium after 400 ns, with all metrics remaining unchanged over the two time intervals. Overall, the mean $R_g$ value of each chain is $1.17 \pm 0.09$ nm and the mean $R_g$ value of the dimer is $1.28 \pm 0.06$ nm (Figure 4.3a). The end-to-end distance distribution is rather broad with a mean value of $2.22 \pm 0.53$ nm, and there are 99, 65, and 7% averaged conformations of two chains with end-to-end distances of $>1$, $>2$, and $>3$ nm, respectively (Figure 4.3b). The distribution of BB-BB and SC-SC interpeptide interactions is also broad (Figure 4.3c). All residues are exposed to solvent (SASA $>0.8$ nm$^2$), with the exception of V12, residues 17-20 of CHC, residues 24 and 26-27 of the loop region, and residues 30-39 (Figure 4.3d). In what follows, analysis was performed on the conformations at 315 K within the time interval of 50-400 ns. Statistical deviations were estimated by calculating block averages over different time intervals.

The percentage of secondary structure averaged over all residues is $18.7 \pm 3.3\%$ for the $\beta$-strand, $10 \pm 2.7\%$ for the $\alpha$-helix, $43 \pm 3.7\%$ for the turn, and $28 \pm 3.1\%$ for the random coil. Figure 4.4 shows the secondary structure along the sequence. There are four regions that populate the $\beta$-strand: residues 3-5 and 10-12 with populations of 22 and 30%, the CHC with a maximum of 57%, and residues 31-36 with a maximum of 53%. These four transient $\beta$-strands are separated by three turns (Figures 4.3e and 4.4c) at positions 7-9, 13-15, and 23-29, with probabilities of around 75%. We also find a fourth turn at position 37-38 with a population of 57%. The $\alpha$-helix profile shows populations of 5, 9, 16 and 17% for the N-terminus, CHC, and residues 22-29 and 30-38, respectively.

To obtain a first picture of the conformations of the peptides within the dimer, Figure 4.5 shows the BB-BB and SC-SC intrapeptide contact maps. Both contact maps are rather sparse, indicating a rather limited number of strong interactions. Looking at the contacts separated by at least four residues with probabilities of $>15\%$, the intrapeptide BB-BB contact map shows probabilities of 41% between V18 and M35, 39% between F20 and G33, 15% (17%) between F19 and L34 (G33), and 23% between K16 and G37. Using the same criteria, the intrapeptide SC-SC contact map reveals hydrophobic interactions with a probability of 18% between F4 and Y10, 19% between F4 and F19, 46% between L17 and L34, 37% between residues V18 and M35, 31% (42%) between F19 and I32 (L34), and 45% between F20 and M35. All intramolecular salt bridges separated at least by three residues have probabilities lower than 18%. The contact probability is 13% between R5 and E11, 17% between E11 and K16, 14% between E22 and K28, and 16% between D23 and K28, indicating a rather weak interaction between the charged residues in the loop region.
Figure 4.3: Convergence of REMD simulations at 315 K. (a) Monomer (continuous lines) and dimer (dotted line) radius of gyration ($R_g$), (b) the $C_\alpha$ end-to-end distance ($d_{ee}$) of each chain, (c) the total number of residues ($N_c$) that are in contact by interpeptide backbone-backbone (continuous lines) and side-chain-side-chain (dotted line) interactions ($P_c$), (d) the solvent-accessible surface area of each residue ($R_{SASA}$), and (e) the propensity of each residue to adopt a conformation. The results were calculated for two time intervals 50-300 ns (black curves) and 50-400 ns (red curves). The statistical deviations in (e) were estimated by calculating block averages over three equal time intervals. The values from panels b to e are averaged over the residues of both chains. In panel b, we computed the distribution of the averaged data of the two chains.
Figure 4.4: Secondary structure propensities of each amino acid of the Aβ₁−₄₀ dimer at 315 K. (a) β-strand, (b) α-helix, (c) turn, and (d) coil. Results were obtained over the time interval of 50-400 ns. The population of each secondary structure indicated in parentheses was averaged over all residues and both chains.

Figure 4.5: Intrapeptide backbone-backbone (BB-BB) and side-chain-side-chain (SC-SC) contact probabilities at 315 K. The populations are averaged over the two chains. The residues with medium and high β-strand signals are highlighted in green on the x and y axes. The main contacts are enclosed in pink boxes. For simplicity, BB-BB and SC-SC contacts, separated at most by two amino acids, are represented by red lines.
To characterize the interface formed by the dimer, Figure 4.6 shows the interpeptide BB-BB and SC-SC contact maps. There is not a well-defined pattern of BB-BB interactions. Using a 4.5% probability threshold, we find that the dominant interactions involve residues 32-35 of both chains (probability of 7.5%), A21 with residues 33-36 (7%), CHC with residues 32-34 (6%), and A21 with residues 29-30 (6%). The N-terminus cannot be ignored, as there are backbone contacts with residues A30, A33-A34, and V40. Looking at the interpeptide SC-SC map, we identify 13 spots with probabilities of between 10 and 20%, involving F4, Y10, residues 17-20, and residues 31-36 of one chain interacting with F4, and residues 10-12, 17-20 and 31-36 of the other chain. The interactions of decreasing probabilities are F19-L34 (19.6%), F19-F19 (18.2%), F20-F19 and I31-M35 (17%), M35-I31 and F20-I31 (16%), and L34-F19 (15%). Again, side-chain interactions among F4, L17, and F19 and among Y10, I32, and M35 cannot be ignored, all having probabilities of 10-15%. Weak interpeptide electrostatic interactions are detected between residues E22-D23 and residues K16 and K28, all four with populations of between 4 and 6%. The interaction free energy between the two peptides in Figure 4.7 shows the dominant contribution of the hydrophobic interactions to the dimer stability and emphasizes the unfavorable desolvation of polar residues at the interface.

Prior to the analysis of dimer conformations, we calculated the free-energy landscape (FEL) of a single molecule using backbone dihedral angle PCA analysis [204].
Figure 4.7: Aβ1−40−Aβ1−40 binding free energy at 315 K. The total, polar, and hydrophobic free-energy terms are in red, blue, and green, respectively.

single-molecule states are of interest because the dimer conformations result from their combinations [123]. The FEL projected on the first two principal components, which account for about 65% of the systems fluctuation, is shown in Figure 4.8. Analysis using the k-means clustering method [116] reveals eight free-energy minima, denoted as Ss1-8. Structures closest to the center of each minimum are shown in Figure 4.8. Using all conformations of each state, the eight Ss states are described in Table 4.1 by their Boltzmann populations, the β-strand and α-helix populations at residues 15-21 and 30-36, the population of turn at residues 23-29, and the end-to-end distance. We also give the population of side-chain contacts between the N-terminus and CHC and between the C-terminus and CHC.

The first two states account for 43% of the ensemble. Ss1 with a population of 23% is characterized by a β-hairpin spanning residues 19-22 and 31-34, with residues 3-7 and 10-13 residues having 15 and 5% to form a β-strand and α-helix, respectively (Figure 4.8). Ss2 with a population of 19% has a probability of 38% to form a β-hairpin at positions 15-21 and 30-36 vs 49% for Ss1, and its structure is compact with an end-to-end distance of 1.2 nm vs 3.1 nm for Ss1. The Ss6 and Ss3 states are essentially coil/turn, although they have probabilities of 46 and 38% to form two β-strands at 15-21 and 33-36. In contrast, the Ss4, Ss5, and Ss8 states with a total population of 27% display transient α-helices at
Figure 4.8: Free-energy landscape (in kcal/mol) of the single-molecule state of Aβ1−40 projected onto the two principal components V1 and V2. Shown are structures closest to the cluster centers with an all-atom representation of the D1 residue.
residues 20-26 and 28-32 (Ss4), residues 15-20 and 32-37 (Ss5), and residues 14-18 and 23-26 (Ss8). These states are extended with end-to-end distances varying between 2.3 and 2.9 nm. The Ss7 state with a population of 8% also displays transient α-helices at positions 24-30 and 31-37 and a transient β-hairpin at residues 5-6 and 10-11. The eight states have an average number of four side-chain contacts between the N-terminus and CHC and differ in the number of side-chain contacts between the C-terminus and CHC, with a minimal value of two contacts for Ss4 and a maximal value of six contacts for Ss6.

Finally we characterized the dominant clusters of the Aβ1–40 dimer. The centers of the first 20 most populated clusters (denoted as to S1-S20) with populations varying between 3.3 and 1.3% are displayed in Figure 4.9. The characteristics of each cluster are listed in Table 4.2. The descriptors include the cluster population, the surface of the interface, the percentages of α-helix and β-strand at the N-terminus, CHC, and C-terminus, the percentage of turn conformation in the loop region, and the collision cross section CCS. Also we list the total number of interpeptide side-chain contacts between the N-terminal and CHC regions (N1), the C-terminal and CHC regions (N2), the C-terminal and C-terminal regions (N3), and the two CHC-CHCs (N4). All values were obtained by using all conformations belonging to each cluster.

Table 4.1: Single-Molecule States of Aβ1–40: For the eight Ss states, shown are the population P in %, the end-to-end distance in nm, the population of (β-strand, α-helix) spanning residues 15-21 and 30-36, the population of turn-spanning residues 23-29, and the number of side-chain contacts between the N-terminus and CHC (N-CHC) and between the C-terminus (C-CHC) and CHC.

<table>
<thead>
<tr>
<th>Ss</th>
<th>P</th>
<th>End-To-End D.</th>
<th>β, α % residue 15-21</th>
<th>Turn % residues 23-29</th>
<th>β, α % residue 30-36</th>
<th>N-CHC</th>
<th>C-CHC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>23.1</td>
<td>3.1</td>
<td>59, 0</td>
<td>63</td>
<td>64, 0</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>19.6</td>
<td>1.2</td>
<td>49, 0</td>
<td>72</td>
<td>38, 3</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>14.6</td>
<td>2.2</td>
<td>39, 2</td>
<td>63</td>
<td>38, 2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>11.7</td>
<td>2.9</td>
<td>18, 6</td>
<td>57</td>
<td>17, 31</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>9.2</td>
<td>2.3</td>
<td>2, 50</td>
<td>51</td>
<td>0, 87</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>8.0</td>
<td>0.7</td>
<td>46, 1</td>
<td>68</td>
<td>48, 0</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>7</td>
<td>7.9</td>
<td>3.0</td>
<td>10, 20</td>
<td>44</td>
<td>2, 60</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>8</td>
<td>5.9</td>
<td>2.5</td>
<td>22, 17</td>
<td>66</td>
<td>5, 48</td>
<td>5</td>
<td>3</td>
</tr>
</tbody>
</table>

In the first dimer state, S1, with a population of 3.3% and consisting of the Ss1 state for both chains, the CHC and C-terminus form β-hairpins in both chains, the α-helix spans
Figure 4.9: Representative structures of the first 20 overall states of the Aβ1−40 dimer at 315 K. The population of each state is given in parentheses. The Cα atom of D1 is represented by a sphere.
Table 4.2: Characterization of the First 20 States of the Aβ<sub>1−40</sub> Dimer: Shown are the state population P in %, the (β, α) structure populations of the N-terminus (index a), CHC (index b), and C-terminus (index d), the turn population of the loop region (index c), the interface surface in nm<sup>2</sup>, the number of intermolecular side-chain –side-chain contacts among N-terminal CHC (N1), C-terminal CHC (N2), C-terminal –C-terminal (N3), and CHC –CHC regions (N4) and the collision cross section CCS in Å.

<table>
<thead>
<tr>
<th>S</th>
<th>P</th>
<th>(α,β)&lt;sub&gt;a&lt;/sub&gt;</th>
<th>(β, α)&lt;sub&gt;b&lt;/sub&gt;</th>
<th>(turn)&lt;sub&gt;c&lt;/sub&gt;</th>
<th>(β, α)&lt;sub&gt;d&lt;/sub&gt;</th>
<th>Sint</th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
<th>N4</th>
<th>CCS</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3.35</td>
<td>7, 3</td>
<td>56, 0</td>
<td>70</td>
<td>44, 0</td>
<td>16.5</td>
<td>6</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>1230</td>
</tr>
<tr>
<td>2</td>
<td>3.33</td>
<td>11, 10</td>
<td>8, 6</td>
<td>48</td>
<td>11, 9</td>
<td>16.6</td>
<td>7</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1243</td>
</tr>
<tr>
<td>3</td>
<td>3.04</td>
<td>3, 7</td>
<td>0, 47</td>
<td>59</td>
<td>0, 62</td>
<td>17.0</td>
<td>1</td>
<td>6</td>
<td>5</td>
<td>3</td>
<td>1225</td>
</tr>
<tr>
<td>4</td>
<td>4.28</td>
<td>17, 2</td>
<td>21, 12</td>
<td>65</td>
<td>13, 11</td>
<td>16.3</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1195</td>
</tr>
<tr>
<td>5</td>
<td>2.47</td>
<td>17, 1</td>
<td>76, 0</td>
<td>69</td>
<td>39, 1</td>
<td>15.0</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>3</td>
<td>1322</td>
</tr>
<tr>
<td>6</td>
<td>2.15</td>
<td>12, 3</td>
<td>57, 4</td>
<td>65</td>
<td>43, 2</td>
<td>16.5</td>
<td>6</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1244</td>
</tr>
<tr>
<td>7</td>
<td>2.09</td>
<td>12, 6</td>
<td>45, 1</td>
<td>58</td>
<td>30, 3</td>
<td>15.4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>1208</td>
</tr>
<tr>
<td>8</td>
<td>1.94</td>
<td>8, 7</td>
<td>44, 1</td>
<td>55</td>
<td>42, 3</td>
<td>15.5</td>
<td>4</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1195</td>
</tr>
<tr>
<td>9</td>
<td>1.93</td>
<td>15, 1</td>
<td>48, 2</td>
<td>74</td>
<td>33, 1</td>
<td>11.2</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1303</td>
</tr>
<tr>
<td>10</td>
<td>1.87</td>
<td>3, 1</td>
<td>59, 0</td>
<td>54</td>
<td>38, 1</td>
<td>15.7</td>
<td>4</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1272</td>
</tr>
<tr>
<td>11</td>
<td>1.84</td>
<td>15, 2</td>
<td>65, 1</td>
<td>75</td>
<td>41, 1</td>
<td>13.1</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>1222</td>
</tr>
<tr>
<td>12</td>
<td>1.81</td>
<td>12, 1</td>
<td>57, 0</td>
<td>62</td>
<td>50, 0</td>
<td>14.0</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>3</td>
<td>1292</td>
</tr>
<tr>
<td>13</td>
<td>1.72</td>
<td>6, 3</td>
<td>50, 0</td>
<td>66</td>
<td>43, 1</td>
<td>12.8</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>2</td>
<td>1258</td>
</tr>
<tr>
<td>14</td>
<td>1.47</td>
<td>8, 6</td>
<td>48, 0</td>
<td>74</td>
<td>36, 5</td>
<td>12.1</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>3</td>
<td>1323</td>
</tr>
<tr>
<td>15</td>
<td>1.42</td>
<td>22, 10</td>
<td>44, 1</td>
<td>76</td>
<td>21, 1</td>
<td>13.6</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1241</td>
</tr>
<tr>
<td>16</td>
<td>1.40</td>
<td>1, 1</td>
<td>64, 0</td>
<td>51</td>
<td>40, 0</td>
<td>16.0</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>2</td>
<td>1221</td>
</tr>
<tr>
<td>17</td>
<td>1.38</td>
<td>4, 0</td>
<td>0, 0</td>
<td>42</td>
<td>5, 48</td>
<td>16.3</td>
<td>1</td>
<td>5</td>
<td>7</td>
<td>4</td>
<td>1186</td>
</tr>
<tr>
<td>18</td>
<td>1.33</td>
<td>31, 11</td>
<td>45, 0</td>
<td>66</td>
<td>28, 0</td>
<td>15.0</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>1</td>
<td>1288</td>
</tr>
<tr>
<td>19</td>
<td>1.32</td>
<td>12, 6</td>
<td>39, 2</td>
<td>65</td>
<td>27, 0</td>
<td>15.1</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>1244</td>
</tr>
<tr>
<td>20</td>
<td>1.29</td>
<td>25, 13</td>
<td>21, 26</td>
<td>51</td>
<td>14, 28</td>
<td>13.2</td>
<td>5</td>
<td>3</td>
<td>5</td>
<td>2</td>
<td>1254</td>
</tr>
</tbody>
</table>
residues 9-12 in the first chain, and a β-strand covers residues 8-13 in both chains. The second chain is more compact ($R_g = 1.16\text{nm}$) than the first chain ($R_g = 1.13\text{nm}$), and the radius of gyration of the dimer is $1.26\text{ nm}$. The interface area is $16.5\text{ nm}^2$. This state is dominated by intermolecular interactions between the N-terminal and CHC and the C-terminal and CHC ($N1 = 6$ and $N2 = 4$) with two contacts between the two C-termini and only one contact between the two CHCs. The S2 state of population 3.3% has a completely different topology with one chain compact with transient β-strands at residues 15-21 and 33-36 (Ss6 state) and the second chain extended with α-helix character at residues 20-26 and 28-32 (Ss4 state). Among all 20 states, S2 has the largest number of intermolecular interactions between the N-terminal and CHC ($N1 = 7$).

The minimal interfacial area is observed in S9, $11\text{ nm}^2$, consisting of two Ss6 states with low β-strand content. The maximal interfacial area is observed in S3, $17\text{ nm}^2$. There is no β-strand content, the α-helix content is 32.5%, both chains are compact ($R_g = 1.15$ and 1.19 nm), and the long C-terminus α-helix of one chain is intercalated between the helical domains of the second chain located at the CHC and C-terminus. S17 and S19 states with a population of 2.7% are also all α, while states S4, S8, S14, S15, S18, S19, and S20, representing 11.3% of all configurations, have mixed αβ topologies with different intramolecular conformations and interfaces. S17 is characterized by seven intermolecular contacts between the two C-terminal regions, while S19 displays the same number of contacts between the two C-terminal regions and between the two CHCs ($N2 = 4$ and $N3 = 3$). Among all 20 states, S17 has the highest number of intermolecular contacts between the two CHCs ($N4 = 4$). S4 consists of Ss6 and Ss8 states, while S19 consists of two Ss5 states, i.e., α-helix at residues 15-20 and 32-37 (Ss5) and residues 14-18 and 23-26 (Ss8).

The S5, S6, S7, S8, S11, S12, and S13 states representing 14% of all configurations have β-topologies. However, they have different intramolecular conformations. S1 and S5 are characterized by two Ss1 states; S6, S7, and S11, by Ss2 and Ss1 states; S12, by Ss6 and Ss1 states; and S13, by two SS6 states. Also the states display various orientations of the chains. S1 is characterized by an interpeptide antiparallel β-sheet between residues 3-8 and 32-36. The S5, S11, and S12 states display an interpeptide parallel β-sheet involving residues 16-22 (for S5 and S12) or 15-21 (for S11) and residues 31-36. The S6 state has an interpeptide antiparallel β-sheet between the two CHCs, namely, residues 15-21 and 16-22. Finally, the two chains are almost perpendicular in S7 and 13, while they are rather parallel in S5. S5 is further dominated by intermolecular interactions between the C-terminus and CHC ($N2 = 5$), while S11, S12, and S13 have around four N1, N2, and N3 contacts.
4.5 Discussion

The size of the Aβ1–40 dimer was investigated by two IM-MS studies using distinct sample preparations and leading to mean cross-collision sections (CCSs) of 1142 and 1245 Å² [32, 194]. What is evident from experiment is that the dimer experimental arrival time distribution is broader than predicted for a unique species indicating the presence of multiple structures with different mobilities and cross sections. The CCS values of the first 3 clusters are 1230, 1243, and 1225 Å², the CCS values of the 20 clusters range between 1195 and 1322 Å², and the averaged CCS value using the Boltzmann population of each cluster is 1255 Å².

Using two sample preparations, CD analysis reported α-helix, β-strand, and random coil/turn contents of 10.5, 38.6, and 50.9% and 0, 12, and 78% [192, 193]. Our simulation gives 10, 18.7, and 71%, i.e., in between the two CD-derived values. For comparison, previous simulations gave α-helix and β-strand contents of 1.3 and 12.6% [71] and 0.1 and 13.6% [195], respectively, with coarse-grained models and 0.5 and 5.5% with all-atom MD simulations starting from coarse-grained DMD structures [72], i.e. negligible α-helix contents. By contrast, our simulations show that residues 30-38 and the loop region have the same probabilities (17%) for an α-helix followed by the CHC (9%) and the N-terminus (7%). Looking at the β-strand propensity, the CHC and residues 31-36 have an averaged value of 53%, followed by residues 3-5 and 10-12, with β-strand populations of 22 and 30%. Again, these propensities along the sequence are very different from previous simulations on the Aβ1–40 dimer. For instance, in ref [72], the largest β-strand propensity never exceeds 30%.

Interestingly, a very similar β-strand profile with four regions was obtained for the Aβ1–40 monomer from MD simulation in explicit SPC water with the GROMOS force field [205] and microsecond/relica REMD simulation in explicit TIP3P water with the OLPS-AA force field [60]. In the latter study, it was found that the maximum population of β-strand is 20% for residues 2-7 and 10-14 and around 60% for the CHC and C-terminus [60]. As in our study, these transient β-strands are connected by turns. However, the mean REMD-OPLS α-helix content is only 2 vs 10% here, with a maximal value of 6% in the 13-17 region vs 17% for residues 22-38 here. Because of the similarity in the populations and positions of β-strands between the monomer simulation and the present study, the dimerization of Aβ1–40 does not significantly enhance the β-strand content, though the impact of different force fields cannot be totally excluded [170].

Our simulations demonstrate the inherent plasticity of the dimer and the very large ensemble of conformations with high coil/turn content characteristics of an intrinsic dis-
ordered protein. It is, however, possible to identify transient configurations persisting from the monomer [60, 205] to the dimer simulations, namely, structures with multiple β-hairpins spanning the CHC and residues 30-36 and a flexible N-terminus. Though the network of main-chain hydrogen bonds and side-chain –side-chain interactions varies between these β-hairpins, it is interesting that the dominant interaction involves the F19-L34 hydrophobic contact with a probability of 34%. The role of this nonlocal contact in the early oligomers and even toxicity has been recently discussed experimentally [18, 206, 207].

We also find transient configurations with (i) antiparallel β-sheets between the two CHCs, consistent with many NMR-derived models of Aβ oligomers from 4 to 33 peptides [18] that may act as seeds for fibrils composed of antiparallel β-sheets as observed for the D23 Aβ1−40 peptide [52]; (ii) parallel β-sheets as observed in the ss-NMR and EM-derived model of the Aβ1−40 fibril [52]; their packing between the CHC and the C-terminus differ, however, from the fibrillar state where CHC/CHC and C-terminus/C-terminus interactions are observed; and (iii) perpendicular β-sheets as observed by coarse-grained and all-atom simulations of amyloid peptides [147, 208, 209, 210, 211]. Also our ensemble reveals the existence of off-pathway configurations with mixed-αβ or all-α contents, which have already been discussed in atomistic REMD-OPLS simulations of the Aβ1−42 dimer [89] and a NMR-guided metadynamics simulation of the Aβ1−40 monomer [212]. The finding of compact and extended dimer configurations with small and large end-to-end distances of the peptides is also consistent with the recent IM-MS analysis of Aβ1−40 oligomers [194, 213] and all-atom simulations of the Aβ1−28 and Aβ1−40 monomers in explicit solvent [188, 212].

Recent experiments and simulations have emphasized the role of the N-terminus in self-assembly. The D7N and H6R mutations accelerate the kinetics of transition from random coil states to β-sheet–rich configurations and fibrils. The substitution of A2 with V or T alters the kinetics and protects from AD in their heterozygous forms [18]. A toxic Aβ1−40 oligomer of high molecular mass with a β-sheet at the N-terminus was determined using solid-state NMR [214]. Also, the N-terminus was found to play a substantial role in dimer Aβ1−40/42 interactions using single-molecule atomic force spectroscopy [215]. Our interpeptide BB-BB and SC-SC contact maps indicate that the N-terminus affects the structures and dynamics of the dimer, and their hydrophobic interactions with the CHC and C-terminus cannot be neglected. The intramolecular interaction probabilities between (F4, Y10) and (L17, F19, F20) are between 11.3 and 18.3%. The probabilities of the intermolecular interactions between F4 and (L17, F19) and between Y10 and (I32, M35) are between 11 and 16%.
4.6 Conclusions

In this study we have determined the equilibrium ensemble of the Aβ₁−₄₀ dimer using extensive atomistic REMD simulations at pH 7. As Aβ₁−₄₀ peptide is much less prone to aggregation than the more toxic Aβ₁−₄₂ peptide, our predictions can be more readily verified by experimental means. Using the CHARMM22* force field that cannot fit normal mode frequencies with high accuracy [91, 216] but reproduces well the structural properties of many peptides [98, 196, 217], the Aβ₁−₄₀ dimer at 315 K is highly disordered with a very large number of structures differing in secondary structure composition and tertiary and quaternary contacts. However, it is possible to identify transient configurations with β-hairpin structures that persist from the monomer [60, 205, 218] to dimer simulations, are consistent with in vitro experiments of oligomers, and may act as seeds for polymerization into parallel or antiparallel β-sheets. For all generated structures, a large structural rearrangement is still necessary to fit the fibrillar-like structures. The simulations also report a detailed description of the N-terminus and provide a framework for a comprehensive understanding of the impact of pathogenic and protective mutations on a molecular level in early-stage Alzheimers disease.
Chapter 5

Coarse-grained and all-atom simulations towards the early and late steps of amyloid fibril formation

Summary

Experiments and computer simulations can complement one another to provide a full and in-depth understanding of many aspects in the amyloid field at an atomistic level. In this chapter, we review the results of our coarse-grained and all-atom simulations in aqueous solution aimed at determining: (i) early aggregation steps of short linear peptides, (ii) nucleation size number, (iii) solution structure of the Aβ_{1-40}/Aβ_{1-42} wild-type dimers, (iv) impact of FAD (Familial forms of Alzheimer’s disease) mutations on the structure of Aβ_{1-40}/Aβ_{1-42} peptides and (v) impact of protective mutations on the structure of Aβ_{1-40}/Aβ_{1-42} peptides. The content of the chapter is in press in the Israel Journal of Chemistry (June 2016) [219].

Sommaire chapitre 5: Simulations atomiques et gros grains pour comprendre les premières et dernières étapes de la formation des fibres amyloïdes.

Les données expérimentales et les simulations se complètent pour fournir une compréhension approfondie et complète atomique sur de nombreux aspects du domaine des amyloïdes. Dans ce chapitre, nous présentons une revue des résultats de nos simulations atomiques et gros grains en solution aqueuse visant à comprendre: (1) les premières étapes de l’agrégation de peptides linéaires, (2) la taille du noyau de nucléation, (3) les structures des dimères sauvages Aβ_{1-40}/Aβ_{1-42}, (4) l’impact des mutations FAD (Forme Familiales de la maladie d’Alzheimer) sur la structure des dimères Aβ_{1-40}/Aβ_{1-42} et (5) l’impact de mutations
protectrices sur les dimères des peptides $\text{A}_\beta_{1-40}/\text{A}_\beta_{1-42}$. Ce chapitre est sous presse dans Israel Journal of Chemistry (2016, Manuscript number: ijch.201600048R1) [219].

5.1 Paper
Israel Journal of Chemistry
Coarse-grained and all-atom simulations towards the early and late steps of amyloid fibril formation
---Manuscript Draft---

Manuscript Number: 
Article Type: Full Paper
Corresponding Author: Philippe Derreumaux
CNRS
Paris, FRANCE
Corresponding Author E-Mail: philippe.derreumaux@ibpc.fr
Other Authors: Mara chiricotto
Thanh Thuy Tran
Phuong Nguyen
Simone Melchionna
Fabio Sterpone
Keywords: Alzheimer disease; nucleation; oligomer structures; mutations; coarse-grained/all-atom simulations
Suggested Reviewers: Buyong Ma
mabuyong@mail.nih.gov
Birgit Strodel
b.strodel@fz-juelich.de
Joan Shea
shea@chem.ucsb.edu
Opposed Reviewers:

Abstract: Alzheimer's disease (AD) is the most common neurodegenerative disease. Experiments and computer simulations can complement one another to provide a full and in-depth understanding of many aspects in the amyloid field at atomistic level. Here we review results of our coarse-grained and all-atom simulations in aqueous solution aimed at determining: (i) early aggregation steps of short linear peptides, (ii) nucleation size number, (iii) solution structure of the Aβ1-40/Aβ1-42 wild-type dimers, (iv) impact of FAD (Familial forms of Alzheimer's disease) mutations on the structure of Aβ1-40/Aβ1-42 peptides and (v) impact of protective mutations on the structure of Aβ1-40/Aβ1-42 peptides.

Author Comments: Dear Yifat,

Please find my contribution to your volume dedicated to amyloids.

Best wishes,

Philippe

Section/Category: Amyloid Aggregation - Yifat Miller, John Straub

Additional Information:

<table>
<thead>
<tr>
<th>Question</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dedication</td>
<td></td>
</tr>
<tr>
<td>Submitted solely to this journal?</td>
<td>Yes</td>
</tr>
<tr>
<td>Has there been a previous version?</td>
<td>No</td>
</tr>
</tbody>
</table>
Coarse-grained and all-atom simulations towards the early and late steps of amyloid fibril formation

Mara Chiricotto,¹ Thanh Thuy Tran,¹ Phuong H. Nguyen,¹ Simone Melchionna,² Fabio Sterpone,¹ Philippe Derreumaux¹&

¹Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Université Paris Diderot, Sorbonne Paris Cité, IBPC, 13 Rue Pierre et Marie Curie, 75005 Paris, France

²Instituto Processi Chimico-Fisici, CNR-IPCF, Consiglio Nazionale delle Ricerche, Roma, Italy.

* Both authors contribute equally
& Corresponding Author: philippe.derreumaux@ibpc.fr, Tel: 33 1 58 41 51 72

Abstract

Alzheimer’s disease (AD) is the most common neurodegenerative disease. Experiments and computer simulations can complement one another to provide a full and in-depth understanding of many aspects in the amyloid field at atomistic level. Here we review results of our coarse-grained and all-atom simulations in aqueous solution aimed at determining: (i) early aggregation steps of short linear peptides, (ii) nucleation size number, (iii) solution structure of the Aβ1-40/Aβ1-42 wild-type dimers, (iv) impact of FAD (Familial forms of Alzheimer’s disease) mutations on the structure of Aβ1-40/Aβ1-42 peptides and (v) impact of protective mutations on the structure of Aβ1-40/Aβ1-42 peptides.

Key words: Alzheimer disease, amyloid simulations, nucleation, oligomer structures, mutations, coarse-grained/all-atom force fields
Introduction

Alzheimer’s disease (AD) is the most common neurodegenerative disease and is pathologically characterized by neurofibrillary tangles resulting from the accumulation of hyper-phosphorylated tau protein and by amyloid plaques made of the amyloid beta (Aβ) protein which results from cleavage of the transmembrane amyloid precursor protein (APP) by beta-secretase (BACE1) and gamma-secretase.\(^1\) Despite continuous debate, there is strong evidence that an imbalance between production and clearance of Aβ1-40/1-42 and related Aβ proteins (either truncated or post-translational modified) plays a key role in initiating AD.\(^2,3\)

Though researches aimed at targeting BACE1,\(^1\) modulating the response of the innate immune system,\(^4\) interfering on Apolipoprotein E4 and other components of cholesterol metabolism, and regulating endosomal vesicle recycling are pursued,\(^5\) controlling Aβ self-assembly with inhibitors is considered as one of the most promising solutions to delay the onset or stop the progression of AD.\(^6\) The challenge arises first from the intrinsic disorder structure of the human wild-type (WT) Aβ monomer in aqueous solution. Aβ1-42 WT sequence, DAEFRHDSGYEVHHQKLVFFAEDVGSNK GAIIGLMVGGVVIA, has two hydrophobic patches L17-A21 (CHC) and A30-A42 (C-terminus) and two hydrophilic patches E22-G29 (loop region) and D1-K16 (N-terminus).\(^7\) The challenge also comes from the lack of high-resolution structures and formation/dissociation rates of the low molecular weight Aβ1-40/1-42 oligomers, including dimers, which are believed to be the most critical players in the pathology;\(^8,9\) and for these oligomers, we have at hand low-resolution structural data.\(^1\) Finally, the experimental sigmoidal kinetics of amyloid formation is the result of a linear combination of microscopic reactions involving primary classical and secondary (fragmentation and surface-dependent lateral) nucleation processes and we know little on the topology and size of the primary nucleus.\(^10-13\) The kinetics is also sensitive to the experimental conditions and the sequences, with mutations enhancing or reducing fibrillogenesis and toxicity.\(^6,10\)

Here we review the contribution of my group and collaborators to understand the early and late aggregation steps of amyloids in aqueous solution based on coarse-grained (CG) OPEP and all-atom simulations. We focus on five aspects and compare with other simulation results and experimental data when possible. Simulations of Aβ and related peptides interacting with
inhibitor candidates such as EGCG, NQTrp, N-methylated peptides, and carbon nanotubes or aimed at dissociating amyloid fibrils are described elsewhere.\textsuperscript{14-22}

Prior to amyloid results, we recall that the OPEP (Optimized Potential for Efficient protein structure Prediction) CG model represents the amino acid by six centers of force. Each side-chain is represented by a unique bead and the backbone uses an atomic resolution with N, HN, C\textsubscript{\alpha}, C and O atoms. Proline is an exception, represented by all its heavy atoms (Figure 1A).\textsuperscript{23,24} The implicit solvent OPEP model retains chemical specificity and is free of any biases. This runs in contrast with the Martini CG model that imposes secondary structure constraints,\textsuperscript{(25)} and the CG Caflisch\textsuperscript{(26)} and Shea\textsuperscript{(27)} models that tune the probability of the monomer to form \(\beta\)-strand. The OPEP energy function is expressed as a sum of local, nonbonded, and hydrogen bonding (H-bond) terms and all analytical terms are given in Ref. 28,29. Notably, H-bonds between backbone atoms are modeled by two- and four-body potentials, rather than Coulomb interactions. In contrast to other CG models used for amyloid proteins\textsuperscript{(30,31)} OPEP has been successfully tested on many non-amyloid proteins, recovering experimental structures and thermodynamics properties,\textsuperscript{29,33-37} and protein/protein complexes using various advanced sampling methods.\textsuperscript{38}

**Early aggregation steps of short linear peptides**

Independently of the sampling method [molecular dynamics (MD), replica exchange molecular dynamics (REMD) and replica exchange Monte Carlo (REMC)], whether CG or all-atom force field with explicit/implicit solvent was used, the self-assembly of short linear peptides starts with the formation of partially ordered oligomers, which is modulated by the hydrophobic character of the system. Then, the formation of H-bonds drives oligomers to transient and marginally populated \(\beta\)-rich aggregates.\textsuperscript{(39-43)}

Using OPEP, we were the first to observe that these \(\beta\)-rich oligomers have various sheet-to-sheet pairing angles,\textsuperscript{(44)} a prediction that was confirmed by X-ray structures of macrocyclic \(\beta\)-sheet mimics\textsuperscript{(45)} and other force field calculations.\textsuperscript{(28,40)} We predicted that these \(\beta\)-rich oligomers can form transient \(\beta\)-barrels,\textsuperscript{(44,46)} and this was validated by the microcrystal structure of an 11-residue amyloid peptide\textsuperscript{(47)} and other simulations\textsuperscript{(27,48)} and are compatible with IM-MS data\textsuperscript{(49)}. We also identified rare events involving the reptation moves of the \(\beta\)-strands,\textsuperscript{(50,51)} allowing the change in the register of the H-bonds without full detachment of the peptides, prior to Fourier transform infrared spectroscopy (FTIR)\textsuperscript{(52)} and atomistic simulation results.\textsuperscript{(53,54)}

Of particular interest from all simulations of 7-mers to 20-mers is that the
β-rich oligomers are characterized by a predominance of mixed parallel–antiparallel β-strands, independently of the sequence.\(^{30,39-41,46,55-57}\) This runs in contrast with the final products, which display either antiparallel or parallel β-strands within the sheets.\(^{58}\) This β-strand orientation mismatch, also observed experimentally for Aβ1-40/1-42 peptides,\(^{1}\) raises the question when the transition towards fully parallel or antiparallel intermolecular β-sheets occurs during oligomerization.

Most computer studies of amyloid aggregation have focused on thermodynamics rather than dynamics because enhanced advanced sampling from REMD, ST and metadynamics does not provide direct information about kinetics. In addition, an accurate description of dynamics obtained from CG models with implicit solvent requires inclusion of computationally demanding hydrodynamic interactions that water exerts on the solute.\(^{59,60}\) Hydrodynamics interactions arise from the motion of atoms that generate a velocity field in surrounding aqueous environment and the resulting flow acts on other protein atoms. While hydrodynamic interactions do not impact the equilibrium distribution of states, they affect dynamics and escape from metastable states and have striking effects on the simulated diffusion and folding of proteins.\(^{61}\)

In this context, we recently presented a novel computational framework that integrates the OPEP CG model for proteins with the Lattice Boltzmann Molecular Dynamics (LBMD) methodology to account for the fluid as a continuum in a probabilistic sense and determine the explicit and on-the-fly solution of the fluid dynamics and kinetics.\(^{29}\) Protein particles are advanced in time by MD, fluid populations are used to represent the solvent and are advanced in time by the Lattice Boltzmann equation, and the coupling between the motion of a solute particle and fluid is based on the assumption that momenta exchange in a Stokes fashion, i.e. modeled by a drag force between each particle and the fluid with one term proportional to the friction coefficient, \(\gamma.\)^\(^{62}\) This parameter is empirically tuned so that the diffusion constant of the molecular system matches the experimental value.

Figure 2 shows the time evolution of the size of the largest cluster, and the number of free monomers using Langevin Dynamics and LBMD simulations of 100 Aβ16-22 peptides, starting from 100 randomly placed peptides in a cubic box of size \(L=150\ \text{Å}\). The aggregation process is characterized by two different time-scales. A first one, \(\approx 10\ \text{ns}\), controls the first encounter of the peptides and the formation of small oligomers, while a second slower one, \(\approx 10^2\ \text{ns}\), controls the fusion and the growth of larger aggregates. The first striking effect of the hydrodynamics interactions (HI) is to speed up the first aggregation.
phase as highlighted in the inset graphs. The second key effect is that HI favors the growth of the largest cluster and its size fluctuations. The aggregates behave as active particles and their change in shape and size alter the surrounding fluid with effect on the inclusions of free monomers in larger clusters as well as the fusion of separate aggregates. Note that after 300 ns of LBMD (data not show) all the monomers fuse in a unique elongated 100-mer with 17% of β-content.

In Figure 3 we present the results of a LBMD simulation of a system of unprecedented size composed of 1000 Aβ16-22 peptides, placed initially in random positions in a cubic box of size L=250 A. Transposed to an explicit solvent all-atom resolution, this system would count 2.4 millions of particles. A system of such a size allows, when compared to smaller systems, to explore a large number of intermediate states. At the time scale explored in the simulation, 200 ns, we observe a continuous growth of the larger clusters. The jumps observed along the growth curves mirror the sudden absorption of smaller in size entities. At the end of the simulation about 30% of the system is assembled in the two larger clusters. Two representative configurations of the largest structures formed at different time are also presented. Interestingly, during the growth we observe the presence of a branched structure with a β-content of 14% (see the snapshot at 200 ns) as predicted by lateral secondary nucleation and observed for some proteins experimentally.

**Nucleation size number**

A very important question is related to the size of the primary nucleus or critical nucleation number, N*. In addition to pH, concentration and temperature, many factors are known to modulate N* and the lag-phase time from experimental studies as measured by Thioflavin-T Fluorescence\(^\text{12,63,64}\) and theoretical studies as determined by kinetics models,\(^\text{65}\) atomistic nucleation theory,\(^\text{66}\) and simulations of mesoscopic and on-lattice models.\(^\text{11,26,27,67,68}\) These factors range from salt and metal concentration, energy landscape of the monomer and population of the monomeric aggregation-prone state, shear flows to the supersaturation of the protein solution. Using simplified models, it has been shown that increasing the total side-chain hydrophobicity switches the fibrilization mechanism from one- to two-step nucleation, where in the one-step nucleation, the β-sheet–enriched nucleus forms directly from the solution, and in the two-step nucleation, soluble monomers first assemble into disordered oligomers, which subsequently convert into a β-sheet nucleus.\(^\text{12,69}\)

All-atom simulations in explicit solvent were performed to determine N*. By following the stability of preformed β-sheet oligomers by MD or REMD, and
characterizing the free energy landscape from disordered aggregates by REMD or bias-exchange metadynamics simulations, it was suggested that $n^*$ varies between 7 to 16 depending on the peptide (e.g., Aβ16-22, STVIYE, GVIGIAQ, Vals).\textsuperscript{70-73} Also, the highest free energy barrier could be associated with the transition from mixed β-strand orientation to native β-strands and the formation of highly interdigitated side-chains, the so-called steric zipper.\textsuperscript{56,74} The limitation of standard atomistic simulations is that rare events and high-energy states are not properly explored and metadynamics simulation results are very sensitive to the choice of the collective variables.

On-lattice simulations were also conducted to determine $N^*$. Using a hydrophobic-polar energy model, Li et al. found that $N^*$ = 11 for Aβ1-42.\textsuperscript{75} By introducing the orientation of the side chains, Irback et al. showed that the highest free energy barriers for an oligomer to form a fibril is accompanied by a change in width.\textsuperscript{76} Using specific side-chain interactions, backbone H-bonds and solvent effects, Frenkel et al. found that $N^*$ = 10 for the amyloid peptide TFTFTFTFT with alternating polar and hydrophobic amino acids.\textsuperscript{77}

Recently, we went on step beyond by presenting an OPEP force field parametrization for the lattice model developed by Frenkel\textsuperscript{77} in order to determine the critical nucleus size of the experimentally well-characterized Aβ16-22 and Aβ37-42 peptides.\textsuperscript{78} A representative structure of Aβ37-42 in the lattice representation is shown in Figure 1B. There are various bottom-up approaches to develop coarse-grained potentials matching all-atom simulations or experimental data. Our bottom-up approach starts with the optimization of the lattice force parameters for the Aβ16-22 dimer by fitting its equilibrium parallel and antiparallel β-sheet populations to all-atom REMD simulation results using both CHARMM22*/TIP3P and AMBER-f99SB-ILDN/TIP3P force fields. We found that the OPEP four-body H-bond interaction plays a crucial role in correct description of secondary structures and end-to-end distributions, and this force field is transferable to the Aβ16-22 trimer and the 2-mers and 3-mers of Aβ37-42, again by comparing with all-atom REMD simulations.\textsuperscript{78} Using this set of parameters and extensive REMC simulations at the calculated folding temperature of the monomer, we characterized the free energy landscapes (FEL) of the 10-mers.

The most populated structures of the Aβ16-22 10-mers, representing more than 90% of the ensemble, display one 10-stranded β-sheet or two 5-stranded β-sheets (Figure 4A) matching the microcrystal and solid-state NMR structures of the amyloid fibril. Experimentally, the Aβ16-22 fibril forms anti-parallel β-sheets with anti-parallel β-strands within individual sheet.\textsuperscript{79,80}
Additional OPEP-REMC simulations for aggregates between 4- and 12-mers indicate a nucleus size of 10 chains.\(^{78}\)

The FEL changes completely for Aβ37-42 10-mer. At room temperature, this oligomer is very disordered and 91% of the conformational ensemble has a total of intermolecular H-bonds \((N_{\text{H-bond}}) \leq 12\) and an order parameter \(P_2 \leq 0.3\). This ensemble is characterized by Aβ37-42 peptides folded into β-hairpin conformations with turns at residues 3-4 or 4-5, and packed in different architectures. Analysis of the ordered states shows that the experimental fibril state is present with a Boltzmann probability of 2%, indicating that \(N^* > 10\) chains. \(^{78}\) Importantly, in contrast to Aβ16-22 fibril, the Aβ37-42 amyloid fibril features experimentally parallel β-strands (i.e. with the two end termini identically charged residues in contact) and antiparallel layers of β-sheets.\(^{79}\)

Because the determination of critical nucleus sizes is important in understanding fibril formation mechanisms, we present new OPEP-REMC simulation results of the Aβ37-42 peptide for 15-mer and 20-mer starting from disordered states at the folding temperature of the monomer. Figure 4B shows the population of one, two and three layers of β-sheets composed of n-stranded β-strands. The one-layer architecture is more populated than the two-layer one and the three-layer structure is hardly formed, and overall, both systems are mainly disordered as reported by the low population (< 15%) of the layers of β-sheets. These results indicate that \(N^*\) is > 20 chains for Aβ37-42.

**Solution structure of the Aβ1-40/Aβ1-42 dimers**

The Aβ1-40 and Aβ1-42 dimers (168) in aqueous solution were studied by Hamiltonian replica exchange molecular dynamics coupled to OPEP, all-atom REMD with a solvent-accessible surface area implicit solvent,\(^{81}\) discontinuous MD with a four-bead CG model,\(^{31}\) and all-atom MD simulations starting from CG DMD structures.\(^{82}\) The results of these simulations show many discrepancies in the 3D intra-/inter-molecular structures.\(^{1}\) To get a better understanding of the equilibrium structures, we studied the Aβ1-40 WT dimer by REMD simulations, 400 ns per replica,\(^{83}\) using the all-atom CHARMM22* force field, considered as one of the best force fields, at least for folded proteins.\(^{84,85}\) The representative structures of the first 10 overall states of the Aβ1-40 WT dimer at 315 K are reported in Ref. 83.

The cross collision sections (CCS) of the first 20 clusters for Aβ1-40 WT dimer were found to vary between 1195 and 1322 Å\(^2\), and using all conformations, the averaged CCS value is 1255 Å\(^2\).\(^{83}\) These values are compatible with two ion-mobility mass spectrometry (IM-MS) studies, based on
distinct sample preparations and leading to a mean cross collision section (CCS) of 1142 and 1245 Å².⁸⁶,⁸⁷ Note there is experimental evidence that there are multiple structures with different mobility and cross section.⁸⁷ Using all REMD-generated structures, the percentage of secondary structure averaged over all residues of Aβ1-40 dimer is found to be 18.7±3.3% for β-strand, 10±2.7% for α-helix, 43±3.7% for turn and 28±3.1% for random coil.⁸³ This calculated 2D structure is consistent with circular dichroism (CD) analysis, using two sample preparations, which reported α-helix, β-strand and random coil/turn contents of (10.5, 38.6 and 50.9%) and (0, 12 and 78%).⁸⁸,⁸⁹

This first all-atom extensive simulation demonstrates the inherent disorder structure of the Aβ1-40 WT dimer with high coil/turn content. We observe, however, multiple transient intramolecular β-hairpins spanning the CHC and residues 30-36 that persist from the monomer to the dimer simulations, and involve the F19-L34 contact with a lifetime of 34%. The role of this contact on the early oligomers and toxicity has been discussed experimentally.¹,⁹⁵ Consistent with many models of Aβ oligomers derived from nuclear magnetic resonance (NMR) spectroscopy,¹ we find transient antiparallel β-sheets between the two CHCs. Our equilibrium ensemble also reveals short-lived all-α topologies (Fig 4A), all-beta topologies (Fig 4B) with two perpendicular beta-sheets, mixed αβ topologies characterized by one compact peptide with β-sheet structure stabilized by a rather extended peptide with α-helical content, and parallel β-sheets between the CHC and the C-terminus rather than between the two CHC’s and between the two C-termini as observed in the fibrillar states.

Overall, a large structural rearrangement is necessary to fit the fibrillar-like states. We are exploring the Aβ1-40 WT dimer with other atomistic force fields by extending each replica to the microsecond time-scale. Indeed it has been shown that there is some dependence of Aβ monomer dynamics and thermodynamics on protein force fields.⁹⁴,⁹⁶,⁹⁷

**FAD mutations on the structure of Aβ1-40/Aβ1-42 peptides**

Familial forms of Alzheimer’s disease (FAD) represent only a small fraction of all AD cases. Here, we only discuss the mutations located within the residues 672-714 of APP from which Aβ42 is processed. Although the addition of the residues IA at positions 41 and 42 increases the pathogenic character of the Aβ peptide and a high percentage of Aβ with a Met-sulfoxide at position 35 is present in the AD brain, the C-terminus is devoid of FAD mutations.¹
The most common mutations located in the loop and near CHC, the Flemish (A21G), Dutch (E22Q), Italian (E22K), Arctic (E22G) and Iowa (D23N) mutations, and in the N-terminus, FAD H6R (English), D7H (Taiwanese), and D7N (Tottori), are known to increase Aβ propensity to aggregate in vitro and the toxicity mediated by Aβ. Using IM-MS experiment, Gessel et al. showed that the D7N, A21G, and E22G peptides display very different oligomer distributions with respect to the WT results, that also vary from Aβ1-40 to Aβ1-42. Two other FAD mutations exist near the CHC: the Osaka E22Δ mutation consisting of a deletion of residue 22, and K16N, which is itself not harmful, but becomes toxic when mixed upon equimolar ratio of WT.

Using Markov state models and potential of mean force calculations, Xu et al. showed that one FAD mutation changes the rugged free energy landscapes of Aβ1-42 monomer by altering the energy barriers around basins. The E22 (E22Δ, E22G, E22K, and E22Q) and D23N mutants generate more hub-like microstates than Aβ42 WT, offering therefore alternative pathways for transitions that could explain enhanced aggregation kinetics. Based on thousands of all-atom MD simulations of Aβ1-42 monomer in explicit solvent, a link between α-helix propensity and aggregation kinetics was proposed by Lin et al. Based on REMD simulations with various force fields, Garcia et al. showed that these mutations increases interactivity of the N-terminus in β pairing that could allow for the seeding of different oligomers and faster aggregation pathways.

Insights into the impact of FAD mutations on Aβ1-40 and Aβ1-42 dimers were investigated using different simulation conditions and extent of sampling. Atomistic MD simulations reported a decrease in β-strand propensity, an increase of the flexibility of CHC and a change in contact maps in Aβ1-40 and Aβ1-42 dimers upon A21G substitution. CG DMD simulations with implicit solvent showed the destabilizing effect of E22G mutation on the structure of residues 20-30 and its increased β-strand impact on the N-terminus of dimers. The other mutations at the position 22 were also found to impact the CHC and the global topologies. OPEP CG REMD simulations of Aβ1-40/1-42 dimers in implicit solvent showed that the Aβ1-40 D23N dimer exhibits structural motifs that differ from those observed in Aβ1-40 WT and Aβ1-42 WT. For instance, while its C-terminal has a higher β-strand propensity than in Aβ1-40, its CHC is almost free of secondary structure as opposed to Aβ1-42.

All-atom MD simulations of the Aβ1-40/1-42 dimers in explicit solvent proposed different mechanisms for the increased Aβ aggregation upon FAD D7N and H6R mutations. The D7N mutation could accelerate the kinetics by
reducing the bending free energy of the loop region;\textsuperscript{(103)} while, upon H6R mutation, the aggregation kinetics of Aβ1-42 could increase due to enhanced β-strand at the C-terminus and higher stability of the salt bridge D23-K28.\textsuperscript{(104)}

While these simulations help understand the increase in aggregation kinetics upon mutations, the structural characterization of FAD dimers remains to our opinion very elusive. What is clear from experiments and simulations is that the results obtained on Aβ1-40 cannot be transposed to Aβ1-42, but whether the variations observed between the mutants arise from differences in simulation details or analysis has to be explored.

**Protective mutations on Aβ1-40/Aβ1-42 dimer structures**

While the FAD and A2V mutations increase aggregation Aβ kinetics, the A2T mutation and the equimolar mixture of the WT and A2V peptides (WT/A2V) and of the WT and A2T peptides (WT/A2T) retard kinetics.\textsuperscript{(105-109)} Heterozygous carriers of A2V, and both homozygous and heterozygous carriers of A2T are protected against AD.\textsuperscript{(105,110)} A2T reduces the production of Aβ from APP by 20-40\%, in contrast to A2V, which enhances Aβ production.\textsuperscript{(105,110)} In vitro experiments have shown that A2V, WT/A2V and WT/A2T change the oligomer size distributions and the stability of the oligomers.\textsuperscript{(105-109)} For instance, by using IM-MS, A2V caused Aβ1-40 to aggregate similarly to Aβ1-42 WT with the formation of dimers, tetramers, hexamers and dodecamers, while the WT/A2V mixture inhibited formation of hexamers and dodecamers.\textsuperscript{(111)} Unique morphologies of the A2T aggregates were also observed using atomic force microscopy.\textsuperscript{(109)}

As a first step towards determining the impact of the single A2V mutation, Nguyen et al. found by atomistic REMD simulations that the Aβ1-28 A2V monomer is much less intrinsically disordered than the WT peptide, has a higher propensity to form β-hairpins, and displays a conformational ensemble totally different from that observed in WT.\textsuperscript{(112)} The monomer structures of Aβ1-40/1-42 A2T were also investigated by two REMD simulations differing in force field and simulation time, 175 ns/replica \textsuperscript{(113)} vs. 500-1000 ns/replica \textsuperscript{(94)}. It was shown that this mutation encourages the N-terminus to engage distant regions of the peptide, increases the N-terminus in β pairing, leading to the possibility of more diverse β topologies.

Next, we compared the equilibrium structures of the Aβ1-40 A2V and A2V/WT dimers with the structures of the WT dimer using extensive atomistic REMD simulations at pH 7 with 400 ns per replica. As Aβ1-40 peptide is much less prone to aggregation than the more Aβ1-42 toxic peptide, our predictions
can be more easily tested experimentally. Our simulations reveal that while the mean secondary structure composition is almost unchanged, there are drastic differences in the intramolecular conformations, and tertiary and quaternary structures upon single and double A2V mutation. \(^{(113)}\)

The intrinsic disorder and the intermolecular potential energies are reduced upon A2V mutation with respect to WT. In contrast, the A2V and WT have similar intrinsic disorder and the A2V dimer states have more favourable interpeptide energies than the WT states. The WT and A2V peptides display many all-\(\alpha\) topologies (Fig 4A), whereas WT/A2V is almost devoid of them. Very interestingly the population of the intramolecular 3-stranded \(\beta\)-sheet spanning Nter-CHC-Cter (Fig. 4C) ranks in the order: WT/A2V (23%) > WT (15%) and A2V (9%), correlating with the increase in the experimental lag phases. The presence of this transient N-terminal \(\beta\)-strand in A\(\beta\)1-40 dimers upon single A2V mutation is likely to increase the free energy barrier to convert one molecule to its aggregation-prone state. \(^{(113)}\)

Whether the protective effect of A2T in the heterozygous form can be rationalized similarly on the A\(\beta\)1-40 dimer was recently examined by atomistic REMD. \(^{(114)}\) We find that the calculated binding free energies correlate well with the observed kinetics of fibril formation, and the intramolecular 3-stranded beta-sheet is an appropriate variable to differentiate fast (A2V, WT) from slow (WT/A2V, WT/A2T) aggregation-prone sequences. The corresponding values are 9% in A2V and 15% in WT vs. 23% in WT/A2V and 25% in WT/A2T. \(^{(114)}\) It would be interesting to study the A\(\beta\)1-40 A2T dimer to determine whether we can propose a theoretical framework that unifies the experimental results on the assembly kinetics of the protective mutations in heterozygous and homozygous cases.

**Conclusion**

We have reviewed what our computer simulations based on off-lattice and on-lattice protein models can tell us about A\(\beta\) self-assembly and its link to Alzheimer’s disease. For each of the five aspects considered, we have looked at dynamics and thermodynamics properties in aqueous solution. It is clear that these simulations should be repeated including metal ions, main protein receptors and the membrane so as to be closer to in cell conditions.

**ACKNOWLEDGEMENT**
We acknowledge support from the GRAL ANR SIMI 12-BS07-0017-01, Pierre Gilles de Gennes Foundation, "DYNAMO" ANR-11-LABX-0011, IDRIS computer time (grant x2015077198), and the European Research Council (FP7/2007–2013, Grant Agreement no. 258748).

REFERENCES


(14) Y. Chebaro, P. Derreumaux, Proteins 2009, 75, 442–452.


(38) P. Kynast, P. Derreumaux, B. Strodel, BMC Biophys. 2016, 9, 4.


(44) (a) G. Wei, N. Mousseau, P. Derreumaux, Biophys. J. 2004, 87, 3648-3656; (b) G. Wei, N. Mousseau, P. Derreumaux, Prion. 2007, 1, 3-8.


(72) Y. Zou, Y. Sun, Y. Zhu, B. Ma, R. Nussinov, Q. Zhang, ACS Chem. Neurosci. DOI: 10.1021/ac-schemneuro.5b00242 (2016).


(114) P.H. Nguyen, R. Pouplana, F. Sterpone, P. Derreumaux, J. Campanera, submitted to *JACS.*
Figure 1. OPEP models. (A) The peptide Ala-Lys-Phe-Pro-Val in its zwitterion form shows the details of the backbone and the side-chains for the off-lattice model. (B) One on-lattice structure of the Aβ37-42 peptide Gly\(^+\)(red)-Gly (yellow)-Val (grey)-Val (grey)-Ile (green)-Ala\(^-\)(blue). For simplicity the aliphatic hydrogen of Gly side-chain is also shown.
Figure 2. Aggregation results of 100 Aβ(16–22) peptides. Time evolution of the size of the largest cluster (blue), the number of free monomers (green and red) in solution using Langevin dynamics and Lattice-Boltzmann molecular dynamics (LBMD). The inset shows the 1-mer population in the first 10 ns. The clusters are defined by considering the distance between the monomer center of mass and considering a cut-off of 12 Å.
Figure 3. **System with 1000 Aβ(16–22) peptides.** Time evolution of the size of the largest cluster (blue), the second largest cluster (green) and the number of free monomers in solution. The panels A) and B) show representative configurations of the largest aggregates explored between 110 and 115 ns and between 210 and 215 ns.
Figure 4. On-lattice OPEP-REMC simulations of Aβ16-22 and Aβ37-42 peptides. (A) The free energy landscape (in kBT\*) of Aβ16-22 10-mers as a function of the order parameter P2 and the total number of intermolecular H-bonds below the melting temperature of the aggregate. Representative structures at the center of all minima are depicted. (B) The population of one (black), two (red) and three (green) β-sheet layers as a function of the number of β-strands. Shown are results of the 15-mer (upper) and 20-mer (lower) Aβ37-42 below the melting temperature of the aggregates.
Figure 5. Representative structures of Aβ1-40 dimer at 315 K for the WT, WT/A2V, A2V and A2T peptides. The Cα atom of D1 is represented by a sphere. All-alpha topologies (A), all-beta topologies with two double beta-hairpins (B) and an intramolecular 3-stranded beta-sheet (C).
Conclusion

The main findings of my thesis can be summarized as follows.

Firstly, I determined the critical nucleus sizes of the two experimentally well-characterized peptide fragments Aβ₁₆–₂₂ and Aβ₃₇–₄₂ of the full length Aβ₁–₄₂. As fibril formation takes place on a day time-scale in vitro, all-atom simulations are too expensive to capture this time scale. In contrast, coarse-grained lattice models which approximate atomistic details and keep the essential interaction, are suitable for capturing generic properties of amyloid formation at low computational cost, but use approximate force fields.

To this end, I coupled the lattice model developed by Frenkel. et al. with the off-lattice coarse-grained OPEP force field, by implementing the four-body H-bond interactions of OPEP, and determining the best values for some missing parameters. There are various bottom-up approaches to develop coarse-grained potentials matching all-atom simulations or experimental data. Our bottom-up approach starts with the optimization of the lattice force parameters for the Aβ₁₆–₂₂ dimer by fitting its equilibrium parallel and antiparallel beta-sheet populations to all-atom replica exchange molecular dynamics (REMD) simulation results using both CHARMM22*/TIP3P and AMBER-f99SB-ILDN/TIP3P force fields. We found that the OPEP four-body H-bond interaction plays a crucial role in correct description of secondary structures and end-to-end distributions, and this force field is transferable to the Aβ₁₆–₂₂ trimer and the 2-mers and 3-mers of Aβ₃₇–₄₂, again by comparing with all-atom REMD simulations.

Using this set of parameters and extensive replica exchange Monte Carlo (REMC) simulations at the calculated folding temperature of the monomer, we characterized the free energy landscapes (FEL) of the 10-mers. The most populated structures of the Aβ₁₆–₂₂ 10-mers, representing more than 90% of the ensemble, display one 10-stranded beta-sheet or two 5-stranded beta-sheets matching the microcrystal and solid-state NMR structures of the amyloid fibril. Experimentally, the Aβ₁₆–₂₂ fibril forms anti-parallel beta-sheets with anti-parallel beta-strands within individual sheet. Additional OPEP-REMC simulations for aggregates between 4- and 12-mers indicate a nucleus size $N^*$ of 10 chains.
The FEL changes completely for Aβ37–42 10-mer. At room temperature, this oligomer is very disordered. This ensemble is characterized by peptides folded into beta-hairpin conformations with turns at residues 3-4 or 4-5, and packed in different architectures. Analysis of the ordered states shows that the experimental fibril state is present with a Boltzmann probability of 2%, indicating that N* > 10 chains. Importantly, in contrast to Aβ16–22 fibril, the Aβ37–42 amyloid fibril features experimentally parallel beta-strands (i.e. with the two end termini identically charged residues in contact) and antiparallel layers of beta-sheets. Because the determination of critical nucleus sizes is important in understanding fibril formation mechanisms, we performed additional OPEP-REMC simulations of the Aβ37–42 peptide for 15-mer and 20-mer starting from disordered states at the folding temperature of the monomer. Analysis of the population of one, two and three layers of beta-sheets composed of n-stranded beta-strands shows that N* is > 20 chains for Aβ37–42.

Extrapolating the results of these two hydrophobic fragments to the full length peptide is difficult, but they suggest that the nucleus size of the full length peptide is very likely much larger than 10, in contrast to what has been predicted by previous coarse-grained simulations.

Secondly, the structures of the Aβ1–40 WT dimer were investigated by means of extensive atomistic REMD simulations at pH 7 using CHARMM22* force field. Though our 24 microsecond-scale simulation reveal a mean secondary structure of 18% beta-strand and 10% alpha-helix, we find transient configurations with an unstructured N-terminus and multiple beta-hairpins spanning residues 17-21 and 30-36, but the antiparallel and perpendicular peptide orientation is preferred over the parallel organization. Short-lived conformational states also consist of all alpha topologies, and one compact peptide with beta-sheet structure stabilized by a rather extended peptide with alpha-helical content. Overall, this study provides, for the first time, insights on the equilibrium structure of the Aβ1–40 WT dimer in explicit aqueous solution, opening a new avenue for a comprehensive understanding of the impact of pathogenic and protective mutations in early-stage Alzheimer disease at a molecular level.

The results of these simulations have been used in order to clarify the fact that (1) the A2V mutation was reported to protect from Alzheimer disease in its heterozygous form and cause an early Alzheimer disease-type dementia in its homozygous form, and (2) experimental studies showed that the aggregation rate follows the order A2V > WT > A2V-WT. To understand the impact of this mutation, Phuong et al. carried out replica exchange molecular dynamics simulations of Aβ1–40 WT-A2V and A2V-A2V dimers and compared to the WT dimer. The atomistic simulations reveal that the mean secondary structure remains constant, but there are substantial differences in the intramolecular and
intermolecular conformations upon single and double A2V mutation. Upon single muta-
tion, the intrinsic disorder is reduced, the intermolecular potential energies are reduced, the 
population of intramolecular 3-stranded beta-sheets is increased and the number of all al-
pha dimer topologies is decreased. Taken together, these results offer an explanation for the 
reduced aggregation rate of the Aβ1−40 A2V-WT peptides and the protective effect of A2V 
in heterozygotes (Nguyen PH, Sterpone F, Campanera JM, Nasica-Labouze J, Derreumaux 

Finally, I present a review on the early and late steps of amyloid fibril formation from 
the coarse-grained and all-atom simulations in aqueous solution performed by the labora-
tory of biochemical chemistry over the years and aimed at determining: (i) early aggrega-
tions steps of short linear peptide, (ii) nucleation size number, (iii) solution structure of the 
Aβ1−40/Aβ1−42 wild-type dimer, (iv) impact of Familiar forms of Alzheimer’s disease mu-
tations on the structure of Aβ1−40/Aβ1−42 peptides and (v) impact of protective mutations 
on the structure of Aβ1−40/Aβ1−42 peptides.

On-going studies and perspectives

The parameters obtained from our on-lattice model are robust since they are transfer-
able from the Aβ16−22 dimer to the Aβ16−22 trimer as well as the dimer and trimer Aβ37−42. 
This set of parameters is currently used to determine (1) the nucleus size of the Aβ13−23 
peptide, also known to form amyloid fibrils in vitro, (2) the self-assembly mechanisms of 
the Aβ16−22 10-mers by using standard Monte Carlo simulations, and (3) the conforma-
tional ensemble of the Aβ1−40/Aβ1−42 peptides ranging from monomers to 12-mers by 
using REMC simulations.

Another perspective is to repeat the on-lattice amyloid peptide simulations in the vicin-
ity of an on-lattice membrane model so as to be closer to in cell conditions.
Bibliography


[32] Bernstein S. L., Dupuis N. F., Lazo N. D., Wytenbach T., Condron M. M., Bitan G.,
Oligomerization and the Importance of Tetramers and Dodecamers in the Aetiology

[33] Jin M., Shepardson N., Yang T., Chen G., Walsh D., Selkoe D. J., "Soluble amyloid
beta-protein dimers isolated from Alzheimer cortex directly induce Tau hyperphos-


[35] Sarroukh R., Goormaghtigh E., Ruysschaert J., Raussens V., "ATR-FTIR: A "rejuve-
nated" tool to investigate amyloid proteins", Biochimica et Biophysica Acta (BBA)-

[36] Miller Y., Ma B., Nussinov "Polymorphism in Alzheimer Aβ Amyloid Organization
Reflects Conformational Selection in a Rugged Energy Landscape", R. Chem. Rev.,

[37] Faendrich M., Schmidt M., Grigorieff N., "Recent progress in understanding

[38] Meinhardt J., Sachse C., Hortschansky P., Grigorieff N., Faendrich, "Abeta(1-40)
fibril polymorphism implies diverse interaction patterns in amyloid fibrils.", M. J.


[40] Paravastu A. K., Leapman R. D., Yau W. M., Tycko R., "Molecular structural basis

of amyloid fibril mass-per-length by tilted-beam transmission electron microscopy.",


[81] Paolo Arosio, Thomas C.T. Michaels, Sara Linse, Cecilia Mnsson, Cecilia Emanuelsen, Jenny Presto, Jan Johansson, Michele Vendruscolo, Christopher M. Dobson and Tuomas P.J. Knowles, "Kinetic analysis reveals the diversity of microscopic mechanisms through which molecular chaperones suppress amyloid formation", Nature Communications, 2016, DOI: 10.1038/ncomms10948


[98] Stefano Piana, Kresten Lindorff-Larsen and David E. Shaw, "How robust are protein-folding simulations with respect to force field parameterization?", Biophysical Journal, 100(9), L47-L49 (2011).


