Atomistic Simulations of Amyloidogenic Peptides and Their Aggregation

Von der Fakultät für Mathematik, Informatik und Naturwissenschaften der RWTH Aachen University zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften genehmigte Dissertation

vorgelegt von

M.Sc. Martín Carballo Pacheco

aus Comodoro Rivadavia, Argentinien

Berichter: Universitätsprofessor Dr.rer.nat. Arne Lüchow
Juniorprofessorin Dr.rer.nat. Birgit Strodel
Professor Ahmed E. Ismail, Ph.D.

Tag der mündlichen Prüfung: 06.03.2017

Diese Dissertation ist auf den Internetseiten der Universitätsbibliothek online verfügbar.

I hereby declare that I composed this work independently and used no other than the specified sources and tools and that I marked all quotes as such.

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet sowie Zitate kenntlich gemacht habe.

Aachen, den 09.11.2016

Martín Carballo Pacheco

"Procurad también que, leyendo vuestra historia, el melancólico se mueva a risa, el risueño la acreciente, el simple no se enfade, el discreto se admire de la invención, el grave no la desprecie, ni el prudente deje de alabarla. En efecto, llevad la mira puesta a derribar la máquina mal fundada de estos caballerescos libros, aborrecidos de tantos y alabados de muchos más; que, si esto alcanzásedes, no habríades alcanzado poco."

Miguel de Cervantes Saavedra in *El ingenioso hidalgo don Quijote de la Mancha*.

Bibliographic information

Publications related to this thesis

- Carballo-Pacheco M., Ismail A. E. and Strodel B., Oligomer Formation of Toxic and Functional Amyloid Peptides Studied with Atomistic Simulations, *J. Phys. Chem. B* 119, 9696-9705 (2015).
- Carballo-Pacheco M., Barz B., Strodel B., Functional Amyloids, Proceedings of the 46^{th} IFF Spring School: Functional Soft Matter, Jülich, Germany (2015).
- Carballo-Pacheco M. and Strodel B., Advances in the Simulation of Protein Aggregation at the Atomistic Scale, *J. Phys. Chem. B* 120, 2991-2999 (2016).
- Carballo-Pacheco M. and Strodel B., Comparison of Force Fields for Alzheimer's Aβ_{42}: a Case Study for Intrinsically Disordered Proteins, *Protein Sci.* 26, 174-185 (2017).

Other publications

- Carballo-Pacheco M., Vancea I. and Strodel B., Extension of the FACTS Implicit Solvation Model to Membranes, *J. Chem. Theory Comput.* 10, 3163-3176 (2014).
- Hora M., Carballo-Pacheco M., Weber B., Morris V.K., Wittkopf A., Buchner J., Strodel B. and Reif B., Epigallocatechin-3-gallate Preferentially Induces Aggregation of Amyloidogenic Immunoglobulin Light Chains, *Sci. Rep.* 7, 41515 (2017).

Abstract

Protein aggregation into highly structured amyloid fibrils is associated both with devastating diseases, including Alzheimer's disease and type 2 diabetes, and functional roles, such as the storage of neuropeptides. Experimental evidence shows that the toxic species in amyloid diseases are small oligomers. These oligomers are transient and, hence, are hard to characterize experimentally. In this thesis, I study the aggregation of amyloidogenic peptides into oligomers using classical molecular dynamics simulations. Most of the peptides simulated are variants of the amyloid-β peptide (Aβ), which is involved in the development of Alzheimer's disease. First, I investigate the aggregation of Aβ_{25-35} and two functional amyloidogenic tachykinin peptides: kassinin and neuromedin K. The three peptides have similar primary sequences, yet, while Aβ_{25-35} is toxic, tachykinin peptides are not. In my simulations, tachykinin peptides aggregate faster than Aβ_{25-35}, which suggests that functional amyloids may avoid toxicity by rapidly aggregating into the non-toxic fibril phase. Furthermore, I observe that peptides that exist in extended conformations as monomers aggregate faster than those in hairpin-like conformations. Next, I compare the ability of different force fields in modeling intrinsically disordered proteins (IDPs) and protein aggregation. In recent years, new force fields have been developed to balance different secondary structures in protein folding simulations. These new force fields should perform better than older ones for IDPs or protein aggregation. In my simulations of Aβ_{42}, which is an IDP, the new force fields, particularly CHARMM22*, reproduce experimental nuclear magnetic resonance data better than the older force fields under study. In the simulations of protein aggregation, none of the force fields is able to distinguish between slowly, fast and non-aggregating peptides. However, the force fields predict similar inter-peptide contacts for aggregating peptides, indicating that protein aggregation is driven by the same interactions with all force fields. Finally, I study the monomer dynamics of multiple mutants of Aβ_{16-22} with different aggregation propensity. No correlation is observed between ensemble averaged properties and aggregation propensity. However, the implied time scale of the slowest process of the monomer dynamics correlates with aggregation propensity, which shows that amyloidogenic peptide aggregation is

encoded in the dynamical properties of the monomer. This thesis presents an advance in the simulations of protein aggregation, providing new insight into the formation of amyloid oligomers. Only if the physico-chemical principles of this process are understood, one can rationally design therapeutic agents against amyloid diseases and create novel amyloid-based nanomaterials.

Zusammenfassung

Die Ausbildung hoch-strukturierter Amyloid-Fibrillen ist zum einen mit schweren Krankheiten, einschließlich der Alzheimerschen Demenz und Typ II Diabetes, als auch mit funktionalen Rollen wie zum Beispiel der Speicherung von Neuropeptiden assoziiert. Experimentelle Daten belegen, dass die toxische Spezies in amyloidogenen Erkankungen niedermolekulare Oligomere sind. Diese Oligomere sind allerdings kurzlebig und daher experimentell schwer zu charakterisieren. In dieser Arbeit untersuche ich die Aggregation von amyloidogenen Peptiden zu Oligomeren mithilfe von klassichen Molekülardynamik-Simulationen. Die Mehrzahl der simulierten Peptide sind Varianten des Amyloid-β-Peptids (Aβ), welches in der Entstehung der Alzheimer-Krankheit involviert ist. Zunächst untersuche ich die Aggregation von Aβ_{25-35} und zwei funktionellen amyloidogenen Tachykininen (Kassinin und Neuromedin K). Diese drei Peptide weisen eine hohe Sequenzidentität auf, jedoch ist Aβ_{25-35} toxisch, wohingegen die Tachykinin-Peptide es nicht sind. In meinen Simulationen aggregieren die Tachykinin-Peptide schneller als Aβ_{25-35}, was nahelegt, dass funktionelle Amyloide möglicherweise schneller in die nicht-toxischen Fibrillen aggregieren und dadurch die toxische Oligomer-Phase meiden. Darüber hinaus beobachte ich, dass die Peptide, welche als Monomer in einer ausgedehnten Konformation vorliegen, schneller aggregieren als jene, die in einer Haarnadel-Stuktur existieren. Anschließend vergleiche ich unterschiedliche Kraftfelder im Bezug auf ihre Fähigkeit, intrinsisch ungeordnete Proteine (intrinsically disordered proteins, IDPs) und Proteinaggregation zu modellieren. In den letzten Jahren wurden neue Kraftfelder entwickelt, um Sekundärstrukturen während der Proteinfaltung ausgeglichen simulieren zu können. Die Vermutung ist, dass diese neuen Kraftfelder auch besser geeignet sind für die Simulation von IDPs und Proteinaggregation als die älteren Kraftfelder. In meinen Simulationen des IDPs Aβ_{42} kann das CHARMM22*-Kraftfeld experimentelle Kernspinresonanzspektroskopie-Daten deutlich besser reproduzieren als ältere von mir getestete Kraftfelder. In den Proteinaggregation-Simulationen kann keins der Kraftfelder überzeugen, da kein Kraftfeld fähig ist, zwischen schnell, langsam und nicht-aggregierenden Peptiden zu unterscheiden. Allerdings sagen die Kraftfelder ähnliche Kontakte zwischen den Peptiden voraus, was da-

rauf hindeutet, dass Proteinaggregation bei allen untersuchten Kraftfeldern durch die gleichen Interaktionen ausgelöst wird. Schließlich untersuche ich die Monomerdynamik mehrerer Aβ_{16-22}-Varianten mit unterschiedlichen Aggregationseigenschaften. Zwischen den Ensemble-gemittelten Eigenschaften und den experimentell bestimmten Aggregationskinetiken kann ich keinen Zusammenhang beobachten. Allerdings korreliert die Zeitskala des langsamsten Prozesses der Monomerdynamik mit der Aggregationsneigung, was zeigt, dass die Aggregation von amyloidogenen Peptiden bereits in den dynamischen Eigenschaften des Monomers kodiert ist. Diese Arbeit stellt einen Fortschritt in der Simulation von Protein-Aggregation dar und präsentiert neue Erkenntnisse über die Bildung amyloidogener Oligomere. Nur wenn die physikochemischen Prinzipien dieses Prozesses verstanden sind, können therapeutische Mittel gegen amyloidogene Erkankungen durch rationelles Design sowie neue amyloid-basierte Nanomaterialien entwickelt werden.

Acknowledgements

There are many people that I would like to thank for playing an important role in my Ph.D. First and foremost, I would like to thank my supervisor Jun-Prof. Dr. Birgit Strodel, for giving me the opportunity to work on this topic. I would also like to thank my other supervisors Jun-Prof. Ahmed E. Ismail, Ph.D. and Prof. Dr. Arne Lüchow. I also thank Prof. Dr. Bettina Keller from the Freie Universität Berlin for her help in applying the variational approach to study the dynamics of the monomers of amyloidogenic peptides. I also thank Prof. Miguel Ladero from the Complutense University of Madrid for introducing me to the scientific world.

I would like to thank all my fellow group members in Forschungszentrum Jülich: Oliver Schillinger, Dr. Michael Owen, Dr. Philipp Kynast, Dusan Petrovic, Dr. Falk Hoffmann, Marianne Schulte, Dr. Qinghua Liao and Dr. Bogdan Barz, for many scientific and non-scientific discussions. I also thank my fellow group members in Aachen: Dr. Brooks D. Rabideau, Dr. Pan Chen, Igor Traskunov, Daniel Tameling and Dr. Marcus Schmidt. I particularly want to thank Brooks and Pan for many conversations and lunches together. I also thank the students that worked with me during my Ph.D.: Matthias Henzgen, Ekaterina Ponomareva and Katee Driscoll.

I thank all of my friends in Germany. In particular, I would like to thank César Ojeda for many scientific discussions. Moreover, I thank all researchers I have met in various scientific workshops and conferences, for sharing their passion for science and ideas with me.

No science would be possible without resources and funding. Hence, I would like to thank the DFG for funding my Ph.D. and JARA-HPC for granting access to computer resources. I thank the teams that maintain both the RWTH and AICES cluster, because none of the simulations in this thesis would have been possible without them. I also acknowledge the entire AICES service team for help in administrative matters.

I would like to thank my family and, in particular, my parents for being next to me during this entire experience. I finally thank my aunt Dr. Miriam Pacheco for being the first person who ever introduced me to science.

Contents

List of Figures

List of Tables

1. Introduction

1.1. Protein Aggregation

Protein aggregation is the process by which individual proteins come together to form stable non-covalent aggregates. Individual proteins in these aggregates may stay in conformations which are similar to the non-aggregated monomers, or, more commonly, change drastically from the conformations of the monomers in solution. Among the latter, the most studied aggregates are amyloids, in which proteins and peptides which are often intrinsically disordered proteins (IDPs) and do not adopt stable 3-dimensional structures, aggregate into highly-structured fibrils mostly formed by β-sheets. Amyloids have been studied at length because of their association with devastating diseases, such as Alzheimer's disease, Parkinson's disease or type 2 diabetes.[6,7] However, amyloids can also have functional roles in nature[8,9] and could be the basis for the design of novel biocompatible nanomaterials.[10] It has been hypothesised that most proteins can form amyloids under the right environmental conditions[11] even if most proteins do not form them under normal conditions.

The most common disease associated with amyloid formation is Alzheimer's disease, which is also the most common type of dementia (about 50-70 % of the patients with dementia in Europe have Alzheimer's disease[12]). Alzheimer's disease features a slow progressive decline in mental ability in which memory, thinking and judgement deteriorate and the personality of the patients may change. An estimated 7.3 million people suffer dementia in Europe[12] (slightly over 1 % of the population). Because dementia is age-related (only 0.09 % of women and 0.16 % of men between 30-59 suffer dementia but 36 % of women and 31.58 % of men between 95 and 99 suffer dementia[12]) and the population of Europe is ageing,[13] the number of affected people will most probably increase in the future. The cost of Alzheimer's disease in 2008 in Europe was estimated to be 160.3 billion €, with 71.7 billion € for direct costs and 88.6 billion € for informal care,[14] which demonstrates the high economic burden of the disease. The hallmark of Alzheimer's disease is the existence of intracellular amyloid plaques in the patient's brain, formed by amyloid fibrils of the amyloid-β (Aβ) protein, a peptide which is an IDP in its monomer state.

There is no known cure for Alzheimer's disease, and current treatment only masks the symptoms.[15]

Parkinson's disease is the second most common neurodegenerative condition after Alzheimer's disease, affecting an estimated 1.2 million Europeans.[16] Parkinson's disease impairs the neurons that control movement, and typical symptoms include tremor, muscle rigidity and slowness of movement. Non-motor symptoms include pain, anxiety and depression. There is also no cure for Parkinson's disease and current treatments only control the symptoms. One of the main physiological changes in Parkinson's disease is the aggregation of α-synuclein, a presynaptic neuronal protein, into amyloids.[17] These amyloids form Lewy bodies which can be found in the brains of patients with Parkinson's disease.

Another major disease associated with amyloids is type 2 diabetes. Diabetes mellitus is a chronic disease, associated with a higher than normal blood glucose concentration resulting from defects on insulin action. Type 2 diabetes is the most common type of diabetes with between 85-95 % of diabetics suffering from it.[18] Unlike type 1 diabetes, patients only suffer from relative insulin deficiency caused by insulin resistance or secretory defects. Untreated type 2 diabetes can lead to blindness or visual disability, kidney failure, heart disease and neuropathy. In 2010, there were approximately 33 million diabetics in Europe (around 9 % of the adult population).[18] The treatment of diabetes is a major economical burden, with countries such as Germany or Spain spending 8.6 % of its total health expenditures for the treatment of diabetes.[18] One of the most common pathological features of type 2 diabetes is the existence of amyloid deposits in pancreatic islets, made of the islet amyloid polypeptide (IAPP), also known as amylin, a short peptide which is secreted when insulin is produced from β-cells.[19]

Alzheimer's disease, Parkinson's disease and type 2 diabetes are probably the most commonly known diseases associated with protein aggregation into amyloids. However, there are at least 40 other diseases associated with amyloids including Huntington's disease, spongiform encephalopaties (including mad cow disease), amyotrophic lateral sclerosis, amyloid light-chain (AL) amyloidosis and fibrinogen amyloidosis.[6] It has even been hypothesised that p53, a protein intrinsically related to the development of cancer, may also aggregate into amyloids.[20] Understanding the molecular mechanisms by which normally soluble proteins aggregate into amyloid fibrils and their relationship to the toxicity in these diseases is essential to find cures for them.

Even though amyloids are mostly known because of their association with

diseases, they can also have normal physiological roles, in which case they are referred to as *functional amyloids*. Functional amyloids are found in a wide variety of organisms: all the way from bacteria to humans.[8,9] For example, in bacteria amyloids formed by the protein curli play an essential role in biofilm formation and host invasion.[21] In the fungus *Podospora anserina*, amyloids formed by the protein Het-S participate in the process of heterokaryon incompatibility, in which two nuclei check their compatibility.[22] There are also several examples of functional amyloids in animals: the protein Pmel17 forms an amyloid in *Homo sapiens* which serves as a scaffold for melanin formation;[23] various protein and peptide hormones are stored as amyloids in the secretory granules of the endocrine system;[24] and the kinases RIP1 and RIP3 form amyloids which work as a signaling complex in programmed necrosis.[25] Amyloids also play a role in the storage of long-term memory in *Drosophila melanogaster*.[26] The existence of functional amyloids calls into question the fact that amyloids are toxic by their own nature, because functional amyloids must have had evolutionary pressure to lower their toxicity.

Finally, amyloids have also been proposed lately as the possible basis for novel nanomaterials.[10,27] Amyloids offer many advantages as a material: their structure is not dependant on complex side-chain interactions but on universal physico-chemical properties; they are biocompatible, stable, self-aggregating and can be designed to have multiple functions; they can be as rigid as natural materials such as silk and bones; and they assemble from simple building blocks such as soluble precursors in a bottom-up fashion. As a proof of concept, amyloids were shown to self-assemble into two-dimensional films.[28] Moreover, amyloids can also be employed in enzyme design, in which an existing enzymatic pocket can be copied into an amyloid-forming peptide and aggregate into a catalytic amyloid.[29] Amyloids can also be used as storage for long-lasting drugs in the pharmaceutical industry.[30] In this case, the drug is injected into the patient as an amyloid, which later slowly disassembles, and the released monomers act as the active drug. It should be noted though, that aggregation is often deleterious in drug action because they can cause an immune response in patients and drugs can stop being efficient.[31] Another possible use of amyloids could be obtained by fusing amyloid-forming peptides with other proteins with particular properties and self-assemble them in an ordered manner. For example, the amyloid-forming SH3 domain was fused to a cytochrome protein, which contains heme molecules with Fe^{2+} ions, and could work as biologically compatible nanowires.[32] Finally, amyloids have also been used as a material for carbon capture, which can be easily regenerated through heating.[33]

Because of the involvement of amyloids in multiple devastating diseases, their growing number of identified functional roles, and the possible construction of new amyloid-based nanobiomaterials, it is essential to understand the physicochemical principles that drive their formation. In such a way, we can help in finding cures for the diseases caused by aberrant amyloids and rationalise the design of functional amyloids.

1.2. Amyloid structure

From a structural biology point of view, the most important question regarding amyloids is to determine its molecular structure. William Astbury was the first one to study the diffraction pattern that amyloids form under X-rays in 1935[34] and obtain some structural details about them. He found amyloids form a cross-β diffraction pattern, characterised by a reflection at $\sim 4.8\,\text{Å}$ which represents the distance between β-strands, and a second reflection at $\sim 10\,\text{Å}$ which represents the distance between β-sheets. The amyloid structure is then a long fibril formed by a number of β-sheets which lie perpendicular to the fibril axis. An example of an amyloid structure is shown in figure 1.1, including the hydrogen bond pattern between the protein backbones that stabilizes the individual β-sheets. Eventually, more detailed structures of amyloid were determined. For example, Nelson et al. in 2005[35] showed using X-ray crystallography that various short amyloidogenic peptides form so-called dry steric zippers, where no water molecules are found between sheets. Structures of amyloids can also be determined using solid-state nuclear magnetic resonance[36] and electron microscopy.[37] Moreover, amyloid formation can also be followed using dies such as thioflavin T and Congo red.

The structure of amyloid fibrils is now well understood for small amyloidogenic peptides. However, there are still many open questions about the structure of other relevant peptides such as Alzheimer's Aβ,[38] which is complicated to study because of its polymorphicity. Nonetheless, it is now clear that in many amyloid-related diseases it is not fibrils which are the toxic agents but small, usually more disordered, oligomers.[39–41] Hence, there is now a growing interest in understanding the structure of oligomers both on- and off-pathway toward fibril formation.[42] However, because of the transient nature of many of these oligomers, it is often hard to obtain information about them experimentally without stabilizing them artificially, and multiple structures have been determined for the same peptide.[38,42] Even with these obstacles, it is now known that the oligomers of Alzheimer's Aβ,[43–47] type 2 diabetes' amylin[48,49]

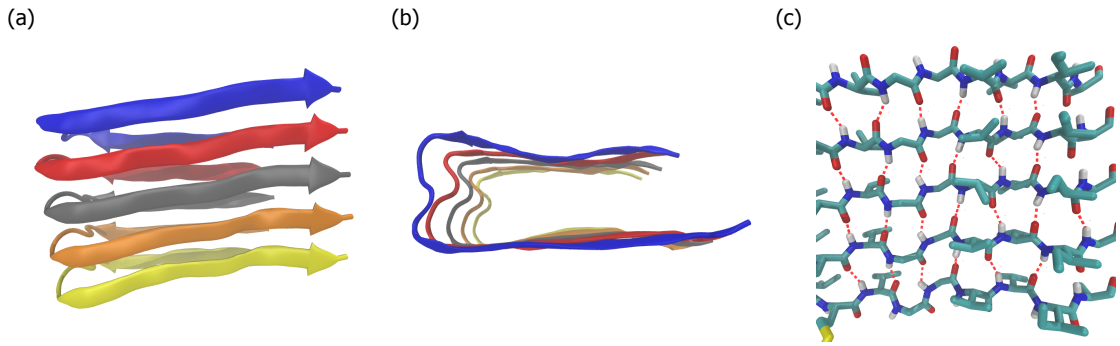

Figure 1.1.: Structure of residues 18-42 of Alzheimer's Aβ_{42} (PDB code 2BEG),[36] an example of an amyloid fibril. (a) Amyloid structure from the front. Each peptide is plotted in a different colour. In this particular fibril, only 5 peptides are plotted, however an actual amyloid could be made of thousands or millions of proteins. (b) The same structure rotated 90° and seen through its main axis. (c) Close-up on the amyloid structure. The hydrogen bond network between backbone atoms of different peptides that stabilize the structure are plotted in red.

and the prion protein[50] have a certain amount of β-sheet content. However, there are contradictory results for Aβ, for which oligomers with little β-sheet content have also been found.[51] Even with these important results, the ensemble of paths from the monomers, which are usually disordered, up to the highly-structured fibrils, passing through a multitude of ordered and disordered oligomers, some of which are toxic, is still unknown.

1.3. Molecular simulations

Understanding the behaviour of proteins at the atomic detail is often hard to do experimentally. In these cases, molecular simulations, and in particular molecular dynamics simulations, can play an essential role by simulating the dynamics of proteins at spatial and temporal resolutions which are impossible to reach experimentally, where either only static information or ensemble average properties can be obtained.[52] These simulations have become an important tool for studying both protein aggregation and already formed amyloid fibrils.[38,53–59] Recently, we have published a review paper that summarises the newest results of atomistic simulation studies of protein aggregation.[2]

One of the most important questions to solve in the process of amyloid formation is how monomers add to fibrils. By means of atomistic simulations, it was possible to understand its molecular mechanism: it follows the so-called

dock-lock mechanism, in which the monomer first docks to the fibrils followed by a drastic increase of the β-sheet content of the monomer, which then slowly locks to the fibril by reorganising itself to optimize the hydrogen bonds with the fibril.[60–62] Regarding the issue of fibril stability, it has been found that toxic mutations of Alzheimer's Aβ are more stable than the wild type fibril,[63] and that they are stable in more conformations than the wild type.[64] Another interest aspect that can also be studied by means of molecular dynamics simulations is the effect of secondary nucleation (the fact that the fibril surface acts as a nucleation surface) on the conformation sampling of monomers.[65]

A particularly relevant and hard question that can be studied using molecular simulations is the formation of the initial oligomers, which are relevant because of their toxicity, and are hard to study experimentally because of their transient nature. Atomistic simulations of protein aggregation are extremely expensive, particularly if performed with explicitly represented water molecules. Hence, coarse-grained models with implicitly represented water, in which multiple atoms are represented as one bead[66–70] or even peptides represented as tubes, are normally used to study protein aggregation.[71–74] These simulations revealed that non-specific interactions are essential for amyloid formation[75] and that crowding agents play an important role in the initial oligomer formation.[76] It has even been possible to build a phase diagram dependant on concentration and temperature for peptide aggregation.[77]

However, to understand the details of protein aggregation, such as the effect of mutations on the aggregation of aberrant amyloids, it is essential to perform atomistic simulations of the oligomer formation of amyloidogenic peptides. The easiest way of performing simulations of protein aggregation is by introducing peptides in a simulation box and allowing them to freely aggregate.[78–82] The first study in which this approach was used was in 2003 by Klimov and Thirumalai,[78] where they investigated the aggregation of a section of Alzheimer's Aβ, Aβ_{16-22} and harmonically restrained the center of the water box and the oligomer center of mass to speed up aggregation. They observed, that aggregation passes through an α-helical intermediate before the antiparallel β-sheet oligomers were formed. Unrestrained protein aggregation simulations were used by Matthes et al. to determine that aggregating peptides first collapse into partially ordered aggregates driven by lowering the solvation free energy and then reorganize to optimize intermolecular contacts.[81] Also, Barz et al. observed that different peptides aggregate following different pathways dependant on their sequence.[82] Even with the advent of special-purpose molecular dynamics computers[83] and highly optimized parallelized[84,85] and

GPU-accelerated codes,[86,87] protein aggregation simulations are still extremely expensive, particularly because of the large amount of solvent needed. Hence, it is customary to study protein aggregation using enhanced sampling algorithms,[88] which provide a more efficient use of the computational resources by speeding up conformational sampling. These methods can provide a converged picture of the free energy landscape but provide limited information on the kinetics of the simulated process. One of the most commonly used enhanced sampling algorithms is replica exchange molecular dynamics (REMD), in which multiple replicas of the system are simulated at different temperatures to enhance the sampling of the system.[89] There are several studies in which REMD was used to investigate the aggregation of amyloidogenic peptides.[90–96] For example, Nguyen and Derreumaux[94] studied the aggregation of 16 Aβ_{37-42} peptides and observed that the free energy minimum is characterized by 2- and 3-stranded β-sheets. Another enhanced sampling algorithm commonly used is metadynamics, in which a history dependence bias is applied in a number of collective variables to sample new sections of configuration space.[97,98] When studying protein aggregation the combination of REMD and metadynamics, the so-called bias-exchange metadynamics[99] has been used.[100–102] Using this methodology, it was found that the aggregation of polyvaline proceeds via an intermediate characterized by antiparallel β-sheets, before reaching the free energy minimum mostly composed of parallel β-sheets. Hamiltonian REMD, in which the Hamiltonian or energy function is changed between different replicas has also been used to study protein aggregation.[103]

Even with the limitations caused by the high unrealistic concentration at which most simulations of protein aggregation are performed,[2] atomistic simulations have provided important information to understand the initial and final events that drive the formation of amyloids, including a detailed understanding of the intermediates followed before reaching the final state.

1.4. Motivation and outline

There are still many open questions regarding amyloid formation, and particularly, how the initial oligomers are formed. Molecular dynamics simulations are specially well-suited for the study of the initial oligomers. This thesis concentrates on understanding the aggregation of biologically relevant amyloidogenic peptides using molecular dynamics simulations and on understanding the limitations of such simulations in the study of protein aggregation.

In Chapter 2, the basics of molecular dynamics simulations, which will be

used during the rest of the thesis, are introduced.

One of the most important questions still to answer is what are the differences in the oligomer formation between aberrant and functional amyloids. Considering that oligomers are the toxic agents in amyloidoses, there should be differences in the structure and/or dynamics of the oligomers formed between toxic and non-toxic amyloids. In Chapter 3, the aggregation of the functional tachykinin peptides and Alzheimer's Aβ_{25-35} is investigated. These peptides are particularly interesting because, even though they share a large amount of sequence identity, they have different toxicity.

Because of the fact that the application of all-atom explicit solvent simulations is recent, we still do not know if the force fields used to model proteins are accurate enough for the study of intrinsically disordered peptides and protein aggregation. In Chapter 4, multiple force fields are tested in how good they are for the modeling of both an intrinsically disordered protein, Aβ_{42}, which is usually the state in which monomers associated with amyloids are found; and in the aggregation of Aβ_{16-22} and two mutants of this peptide with different aggregation propensities.

Finally, one of the most important hypotheses regarding what drives protein aggregation is that the aggregation propensity is encoded in the dynamics of the monomer. In Chapter 5, a number of mutations of Aβ_{16-22} with different aggregation propensities are simulated for the first time and both their structural and dynamical properties are calculated to correlate them with their aggregation propensity.

The conclusions from this thesis and an outlook for future studies are presented in Chapter 6.

2. Methodology

2.1. Molecular dynamics simulations

Molecular dynamics (MD) is a simulation method to calculate equilibrium and dynamical properties of a system of interacting atoms.[104,105] Even though quantum mechanical effects can be explicitly included, for many systems, including proteins whenever enzymatic reactions are not studied, it is often enough to model the system using classical mechanics. The first application of MD simulations to proteins were performed in the 1970,[106,107] and since then the technique has become a standard tool to study the dynamics of proteins.[52,108,109] In 2013, the Nobel Prize in Chemistry was awarded to Martin Karplus, Michael Levitt and Arieh Warshel who have worked at large in the field of MD simulations of proteins. Today, there exist many high quality programs to perform MD simulations including GROMACS,[110] NAMD,[85] ACEMD,[87] OpenMM,[86] Desmond[111] and LAMMPS.[112]

The basic theory behind MD simulations is reasonably simple. First, the initial coordinates for all atoms in the simulated system must be specified. In the case of folded proteins, it is customary to start from a structure determined either using X-ray crystallography or NMR. The potential energy of the interactions between atoms are then calculated with a so-called force field, which will be explained in detail in the next section. Then, forces are calculated as the gradient of the potential energy:

$$\mathbf{F}_i = -\nabla_i U, \tag{2.1}$$

where $\mathbf{F}_i$ is the force acting on particle i, U is the potential energy, and ∇_i is the gradient with respect to the coordinates of particle i. Then, Newton's second law[113] is used to calculate the acceleration of the particles:

$$\mathbf{F}_i = m_i \mathbf{a}_i, \tag{2.2}$$

where m_i is the mass and $\mathbf{a}_i$ is the acceleration of particle i. The acceleration is defined as the derivative of the velocity ($\mathbf{v}_i$), which is defined as the derivative

of the particle's coordinates ($\mathbf{r}_i$):

$$\mathbf{a}_i = \frac{d\mathbf{v}_i}{dt} = \frac{d^2\mathbf{r}_i}{dt^2}. \tag{2.3}$$

In this way, the new coordinates $\mathbf{r}_i$ are calculated for all the atoms in the system. The algorithm can then be used iteratively, until the system has converged to its free energy minimum.

The simulation of large systems of thousands of atoms cannot be solved analytically and, hence, the equations of motion must be discretized to solve the problem numerically. In GROMACS,[110,114] the software used for all of the simulations in this thesis, the standard integrator is the leap-frog integrator,[115] which is of third order in the coordinates $\mathbf{r}$ and time-reversible. The leap-frog algorithm receives its name from the fact that coordinates and velocities are calculated at different times. This algorithm uses coordinates $\mathbf{r}$ at time t and velocities $\mathbf{v}$ at time $t - \frac{1}{2}\Delta t$, where Δt is the timestep. Δt should be small enough to correctly sample the fastest motion simulated, and is usually in the order of 1 fs in MD simulations of proteins. Modern MD softwares often use newly developed techniques to increase Δt, which will be explained in further sections. The new coordinates and velocities are then calculated with the following equations:

$$\mathbf{v}_i(t + \frac{1}{2}\Delta t) = \mathbf{v}_i(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m_i}\mathbf{F}_i(t), \tag{2.4}$$

$$\mathbf{r}_i(t + \Delta t) = \mathbf{r}_i(t) + \Delta t\mathbf{v}_i(t + \frac{1}{2}\Delta t). \tag{2.5}$$

These equations must be modified in order to include constraints, temperature and pressure coupling.

In the following sections, I will explain the details of how to perform realistic and fast MD simulations.

2.2. Force fields

The potential energy is calculated with a so-called force field, which is a complex system of equations including multiple terms and parameters that estimate the interaction between the different atoms in a protein and the solvent, in the case that the solvent is explicitly represented. There are a number of families of force fields, which have been developed over time and have had multiple updates of their functional forms and parameters.[116] Some of the most famous force fields families are CHARMM (Chemistry at Harvard

Macromolecular Mechanics),[117] OPLS (Optimized Potentials for Liquid Simulations),[118] AMBER (Assisted Model Building with Energy Refinement)[119] and GROMOS (GROningen MOlecular Simulation).[120] Most force fields use the following standard functional form:

$$\begin{aligned} U(\mathbf{r}) = & \sum_{\text{bonds}} K_b(b-b_0)^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_0)^2 \\ & + \sum_{\text{dihedrals}} \sum_{n} K_\chi(1+\cos(n\chi-\delta)) + \sum_{\text{impropers}} K_\phi(\phi-\phi_0)^2 \\ & + \sum_{\text{nonbonded}} \left\{ \epsilon_{ij} \left[\left(\frac{r_{ij,0}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{ij,0}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \right\}. \end{aligned} \tag{2.6}$$

The first four terms include the bonded interactions while the last term represent the non-bonded interactions.

Bonded interactions include bonds, angles, dihedral angles and improper angles. The first term in equation 2.6 represents bonds, which are interactions between two covalently bonded atoms. The interactions are modeled with a harmonic potential, in which K_b is the force constant which establishes the strength of the interaction, b represents the bond distance, and b_0 is the equilibrium bond distance, which is equivalent to the minimum in the potential energy well. Figure 2.1 shows an example of this functional form. K_b and b_0 are parameters that need to be defined for each bond in the protein and its parametrization is needed for the development of a force field. The second term in equation 2.6 represents angles, which model the angle θ which is formed between three covalently bonded atoms. Again, there are two parameters for each angle: the force constant K_θ and the equilibrium angle θ_0. These parameters need to be defined for each angle formed between every three bonded atoms in the protein. The third term represents dihedral angles χ, which are the angles defined between the two planes formed by the first and last three atoms of four atoms that are consecutively covalently bound. In this case, the term is represented with a sinusoidal function, which allows for the definition of multiple minima. The parameters that need to be defined are the force constant K_χ, the multiplicity n and the phase angle δ. An example of the functional form used to model dihedral angles is plotted in figure 2.2. Finally, the fourth term represent improper dihedral angles ϕ, which are also defined by four covalently bonded atoms and are used to keep certain chemical groups, such as aromatic rings, planar. There are two parameters that need to be parametrized for each improper dihedral angle: the force constant K_ϕ and the equilibrium improper dihedral angle ϕ_0. Bonded interactions may be aug-

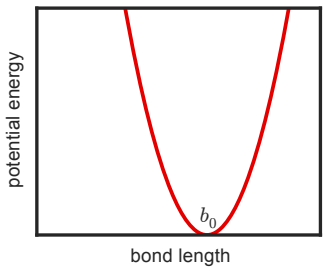

Figure 2.1.: Harmonic potential for bonded interactions. b_0 is the equilibrium bond length and it represents the minimum in the potential energy. Because of the harmonic nature of the potential, bonds cannot be broken.

mented by coupling or correction terms, such as the Urey-Bradley potential,[121] which is a distance potential between atoms connected through a third atom, and the CMAP correction,[122] a grid-based correction to accurately model the preferences of the backbone dihedral angles in proteins. These two terms are used in the CHARMM force field.

The last term in equation 2.6 represents non-bonded interactions, which includes two terms: Lennard-Jones interactions and electrostatic interactions. Lennard-Jones interactions (see figure 2.3 for its functional form) are use to model interactions between all atoms, even if they have no partial charges. It includes two terms: the first one is a short-range repulsive interaction and models the Pauli exclusion principle, which impedes the overlap of electron orbitals; and the second term is attractive and represents long-range dispersion interactions, which are interactions that develop between instantaneously induced dipoles caused by polarization. In equation 2.6, r_{ij} represents the distance between atoms i and j for which the interactions are calculated. The parameters needed are $r_{ij,0}$ which is the distance of the potential energy minimum and ϵ_{ij}, the Lennard-Jones potential energy depth between atoms i and j. Non-bonded interactions also include electrostatic interactions, which are essential for protein function. q_i and q_j are the static partial charges of atoms i and j respectively, which need to be parametrized for each atom, and ϵ_0 is the permittivity of the free space. Non-bonded interactions for covalently bonded atoms and for atoms separated by two covalent bonds are usually neglected and their interactions are dictated by bonded terms. However, for atoms separated by three covalent bonds the non-bonded interactions are included in the force field. In the CHARMM force field, these interactions are not scaled, but in OPLS a scaling factor of 0.5 is used. In AMBER, scaling factors of 0.5 and 0.83 are used for Lennard-Jones and electrostatic interactions, respectively.

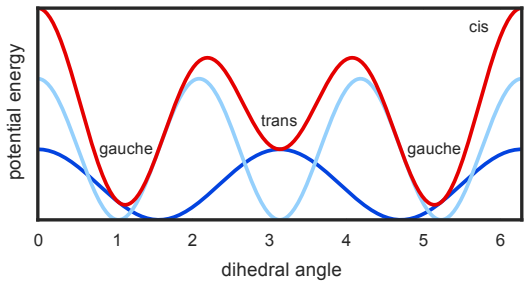

Figure 2.2.: Example of a dihedral angle potential. In this case, it represents a potential that follows the equation $U(\chi) = K_2(1 + \cos(2\chi)) + K_3(1 + \cos(3\chi))$. The first term is plotted in blue, the second term in light blue and the sum of them in red. It represents a standard dihedral angle potential energy function for a case where the two terminal groups sterically clash. In this case, the gauche conformations are the lowest potential energy minima, the trans conformation is a metastable energy minimum, and the cis conformation is the least stable conformation. In this example, the dihedral potential function only has two terms, but force fields include interactions with up to four terms. By summing them up, potentials with multiple minima at different energy levels can be created.

This limits the transferability between force fields.

The Lennard-Jones parameters ϵ_{ij} and $r_{ij,0}$ must be estimated for each possible pair of atoms that exist in the force field, which means that thousands of parameters need to be defined. Hence, it is standard to develop atomic ϵ_i and $r_{i,0}$ parameters which are later combined to obtain the final parameters. CHARMM and AMBER use the Lorentz-Berthelodt rules: i.e., ϵ_{ij} are obtained using the geometric mean $\epsilon_{ij} = \sqrt{\epsilon_i \epsilon_j}$ and $r_{ij,0}$ are estimated using the arithmetic mean $r_{ij,0} = (r_{i,0} + r_{ij0})/2$. OPLS, however, uses the geometric mean for both parameters. This further limits the transferability between force fields.

Equation 2.6 is the standard equation for the so-called class I additive force fields. There are also force fields referred to as class II force field, which include higher order terms for bonds and angle, and also cross terms.[123] Further improvements of classical force fields include the use of the Morse function for bonds, which allows for the simulation of bond breaking,[124] and polarizable force fields.[125] These terms, however, increase the computational costs of the simulation and are usually not used for proteins.

Even if many force fields have similar functional forms, they can perform drastically differently when used in the simulation of a particular protein.[126,127] One of the main reasons for this, is that the parameters are different between

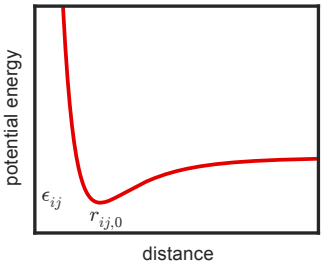

Figure 2.3.: Lennard-Jones potential for non-bonded interactions. ϵ_{ij} is equivalent to the depth of the energy well and models the strength of the interaction. $r_{ij,0}$ is the distance at which the minimum is located. The left repulsive interaction for short distances comes from the r_{ij}^{-12} term and arouses from the Pauli exclusion principle. The right attractive interaction for long distances comes from the r_{ij}^{-6} and comes from dispersion interactions.

different force fields, and even between different versions of the same force field family. In particular, for both OPLS and CHARMM the partial atomic charges were determined using HF/6-31*G quantum mechanical calculations, while AMBER partial atomic charges were calculated based on RESP charges fit to the same level of theory. The Lennard-Jones parameters of the three force fields were estimated based on condensed phase simulations of small organic compounds. However, the water model (see section 2.3) used for the parametrization of CHARMM and AMBER was TIP3P, while OPLS was parametrized to work with TIP3P, TIP4P and SPC water models. Bonded parameters for CHARMM, AMBER and OPLS were estimated to reproduce quantum mechanical and experimental data for model compounds. The GROMOS force field is the only of the most commonly used force field which still uses united atoms, where methyl groups are considered as one atom, and it has been parametrized with a focus on the reproduction of condensed phase properties. All force fields are regularly updated as computers have become faster and inaccuracies of the force fields become evident from longer simulations. The complexity of developing accurate force fields means that it is hard to know which force field is the best for a particular system, and it is often needed to test and update force fields when new systems are studied.

2.3. Water models

Proteins cannot properly function without water surrounding them and, hence, the correct treatment of solvation is essential for modeling the dynamics of pro-

teins. In the past, implicit solvents,[128] in which water models are not explicitly represented but implicitly included in the potential energy function, were normally used and are still needed when simulating extremely large systems such as protein aggregation at very low concentrations.[65,129] However, there is now clear evidence that implicit solvents do not represent water accurately[130–133] and, therefore, it is now customary to represent water explicitly, even if it is computationally more expensive.

The most common water models are the TIP3P[134] and the SPC[135] models, which have had multiple corrections and extensions.[136–140] Both TIP3P and SPC are three-site models, which means that they have three interaction sites which correspond to the the three atoms in the water molecule, which carry a partial charge, while the oxygen atom also has Lennard-Jones interactions. The TIP3P model is probably the most used water model and most force fields have been developed to be used in combination with it. The three atoms form an angle of 104.5°, which is the value observed experimentally. Simulations with TIP3P accurately reproduce water energetics and the first two hydration shells. However, they underestimate the radial distribution function for long-range water structures and the diffusion constant is larger than the experimental value by a factor of 2. The SPC model is similar to TIP3P, but it has an ideal tetrahedral angle of 109.47° and, when simulated, results in more defined peaks in the radial distribution function.

Many of the extensions of the TIP3P model are part of the family of the TIP4P water model. TIP4P is a four-site model, with an extra dummy atom with a negative charge near the oxygen along the bisector of the angle between the three real atoms. The dummy atom improves the electrostatic distribution around the water model. One of the most current extensions of this model is the TIP4P-Ew[138] water model, which has been developed to be used in combination with long-range Ewald summation methods.

2.4. Constraints

When discretizing the equations of motion, a timestep Δt, must be used as shown in equations 2.4 and 2.5. This timestep should be small enough to correctly sample the fastest motion in the system, which are usually bond vibrations. A timestep of 1 fs is usually needed to avoid the system from becoming unstable and the simulation from crashing. This means that to model a system for 1 µs, a billion steps are necessary. Any method that allows to increase the timestep, enables considerable larger simulations to be performed.

One of the most common methods to increase the timestep is to constrain the bonds. In classical MD, reactions are not being simulated and, hence, the changes in bond lengths are usually not of interest. In classical protein simulations, the most relevant effects are produced by non-bonded interactions and dihedral angles. Therefore, constraining the bonds usually has no major effects on the simulation and, because it allows for the timestep to be doubled to 2 fs, simulations twice as large can be performed with the same amount of computational resources. The constraint algorithm used for the simulations in this thesis is LINCS (LINear Constraint Solver).[141,142] LINCS uses Lagrange multipliers to constrain the bonds. The constraints create new forces that must be added to the equations of motion. The most important parameter to select when using LINCS is the order of the expansion for the numerical estimation of the Jacobian, which represents the constraining coupling matrix. For normal MD simulations, an order of 4 is enough, but for simulations with virtual sites (see next section) an order of 6 is necessary. In the case of parallel simulations, atoms which are constrained may be found in different processors and the parallel LINCS algorithm is needed.[142] In this case, when an atom crosses the boundary between processors, atoms up to `lincs_order + 1` are communicated between processors and the normal LINCS algorithm is then used.

For water molecules, the SETTLE algorithm, which is based on the SHAKE algorithm,[143] is used.[144] The SHAKE algorithm also solves the constrained equations of motion using a Lagrange multiplier for each constraint. The Lagrange multipliers are calculated with the Gauss-Seidel method. The algorithm is iteratively used until a relative tolerance for the constraint is reached. SETTLE is an analytical solution of SHAKE which has been designed specifically for the simulation of water. Instead of calculating constraint forces to correct the unconstrained move, SETTLE uses quasi-Euler angles to directly correct the update, and is faster than the original SHAKE algorithm.

2.5. Virtual sites

The next fastest motions after bond vibrations are angle and out-of-plane motions. However, constraint algorithms such as LINCS have limitations when the number of constraints are too high, therefore both bonds and angles cannot be constrained using LINCS. Hence, another strategy must be implemented to further increase the timestep. One possibility is the use of virtual sites for hydrogens, which allows to increase the timesteps from 2 fs up to 7 fs.[145]

Under this strategy, the positions of the hydrogen atoms are not updated based on bonded interactions with other atoms, but they are calculated at every step based on the positions of the three closest heavy atoms. The forces on the hydrogen atoms are then redistributed on the heavy atoms, and because virtual sites are massless, the mass of the hydrogens must also be redistributed on the heavy atoms. In this way, the faster degrees of freedom of the hydrogen atoms are removed from the system. However, certain degrees of freedom such as the rotational freedom of the hydrogen atoms in amine groups must still be simulated to accurately model proteins.

2.6. Parallelization

Many of the advances in theory and implementation of MD simulations allow scientists to perform longer and bigger simulations, to obtain converged simulations and to study new systems, such as protein aggregation. Probably, the most important advances in the implementation of MD codes is their parallelization and the design of MD codes for newer computer architectures. In particular, GROMACS,[110] LAMMPS[112] and NAMD[85] are highly optimized parallel MD codes. On the other side, ACEMD[87] and OpenMM[86] have been optimized for their use with graphical processing units (GPUs), while Desmond[111] has been developed to be used with the specialized hardware Anton.[83]

In parallel codes such as GROMACS,[114] the computation is divided between a number of processors to increase the speed at which the simulation is performed. Because in MD simulations most interactions are of local character, it is reasonable to use a domain decomposition scheme, in which each processor is in charge of simulating the atoms in a certain region of space. Atoms, however, may interact with other atoms located in different processors and, hence, information from the neighbouring processors needs to be communicated at every step. GROMACS also uses dynamic load balance to optimize the space assigned to each processor and maximize the division of work. Even though most interactions are local, electrostatic interactions are long-range and therefore, global communication between all processors is needed for an accurate calculation of them. In GROMACS, a number of processors are used only for the calculation of electrostatic interactions, usually using the particle mesh Ewald (PME) summation method (see next section). In such a way, only a few processors must communicate globally and this limits the computing time lost because of the communication between processors.

2.7. Long range interactions

The most computationally expensive part of an MD simulation is the calculation of non-bonded interactions, and particularly electrostatic interactions. Because these interactions depend on the distance between atoms, the forces between atoms that are located far apart are negligible. Therefore, truncating them after a certain cutoff was considered a reasonable approach to decrease the amount of computing time needed. Originally, cutoffs of 0.8 or 0.9 nm were used in MD simulations and force fields were parametrized considering such cutoffs.[116] However, important effects are observed when truncating long-range electrostatic interactions. For example, simulations of the TIP3P water model reproduce the experimental diffusion constant better without cutoffs[146] and protein simulations are better at reproducing experimental observables when considering long-range electrostatics.[147,148] Hence, long-range electrostatics are essential to model proteins correctly.

The easiest way to account for long-range electrostatics would be to have no cutoffs and to take into account all atoms in the force calculation. However, this method is extremely computationally expensive and alternatives are usually needed. The most commonly used methods to account for long-range electrostatics are based on the Ewald summation method,[149] which divides the space into two based on the distance from each atom. Short-range interactions are calculated using a direct calculation, while long-range interactions are calculated using an efficient reciprocal space treatment. To be able to use Ewald summation methods, the system under study must be periodic, hence it is customary to simulate proteins with periodic boundary conditions. Periodic boundary conditions were originally introduced because of other advantages such as avoiding the unwanted effects caused by finite systems like an increased surface tension. When using periodic boundary conditions, the system is surrounded by copies of itself. Hence, one should always be careful to leave enough space around the protein so that it does not interact with itself via its copies. In general, in a system with periodic boundary conditions, the total electrostatic energy is calculated with:

$$V = \frac{f}{2} \sum_{n_x} \sum_{n_y} \sum_{n_z^*} \sum_{i}^{N} \sum_{j}^{N} \frac{q_i q_j}{\mathbf{r}_{ij,\mathbf{n}}}, \tag{2.7}$$

where N is the number of particles, $f = \frac{1}{4\pi\epsilon_0}$ is the electric conversion factor, $\mathbf{n} = (n_x, n_y, n_z)$ is the box index vector, and the star is to indicate that when $i = j$ and $(n_x, n_y, n_z) = (0, 0, 0)$ the terms should be omitted, and $\mathbf{r}_{ij,\mathbf{n}}$ is

the real distance between particles. In the Ewald summation method, the conditionally convergent equation 2.7 is divided into three terms to increase convergence:

$$V = V_{\text{dir}} + V_{\text{rec}} + V_0, \tag{2.8}$$

where

$$V_{\text{dir}} = \frac{f}{2} \sum_{i,j}^{N} \sum_{n_x} \sum_{n_y} \sum_{n_z^*} q_i q_j \frac{\text{erfc}(\beta r_{ij,\mathbf{n}})}{r_{ij,\mathbf{n}}}, \tag{2.9}$$

$$V_{\text{rec}} = \frac{f}{2\pi V} \sum_{i,j}^{N} q_i q_j \sum_{m_x} \sum_{m_y} \sum_{m_z^*} \frac{\exp(-(\pi \mathbf{m}/\beta)^2 + 2\pi i \mathbf{m} \cdot (\mathbf{r}_i - \mathbf{r}_j))}{\mathbf{m}^2} \tag{2.10}$$

and

$$V_0 = -\frac{f\beta}{\sqrt{\pi}} \sum_{i}^{N} q_i^2, \tag{2.11}$$

where V_{dir} is the calculated in direct space, V_{rec} is calculated in reciprocal space and V_0 is constant. β is a parameter that balances the accuracy of the direct and reciprocal space, V is the volume of the box and $\mathbf{m} = (m_x, m_y, m_z)$ is the wave vector.

In the case of GROMACS, a variation of the Ewald summation method, called particle mesh Ewald (PME),[150] is used. In PME a grid is used to interpolate particles instead of directly summing wave vectors. The grid is Fourier transformed with the efficient Fast Fourier Transform algorithm, which scales as $N \log(N)$ instead of N^2, and is then more efficient than the original Ewald summation method.

Even if taking into account long-range electrostatics is now customary for protein simulations, it is still common to use cutoffs for Lennard-Jones interactions because of their dependence on r^{-6} instead of r^{-1} like electrostatic interactions. If the space outside of the cutoff is considered as homogeneous, corrections can be applied to take into account the truncation.[151] It should be noted that in the case of surface simulations, it is now becoming normal to also consider long-range interactions for dispersion interactions.[152–156]

2.8. Thermostats

If the equations 2.4 and 2.5 are used to perform an MD simulation, it would be performed in the microcanonical or NVE ensemble, i.e., the number of particles, volume and energy would be conserved. However, it is often more realistic to sample the system in the canonical or NVT ensemble (i.e., constant

number of particles, volume and temperature) or in the isothermal-isobaric or NPT ensemble (i.e., constant number of particles, pressure and temperature). In an MD simulation, the temperature can be calculated as a function of the kinetic energy of the system, which depends on the velocity of the particles:

$$T = \frac{1}{N_{\mathrm{df}} k_{\mathrm{B}}} \sum_{i=1}^{N} m_i v_i^2, \tag{2.12}$$

where $N_{\mathrm{df}} = 3N - N_{\mathrm{c}} - N_{\mathrm{com}}$ is the number of degrees of freedom, N is the number of particles, N_{c} is the number of constraints, $N_{\mathrm{com}} = 3$ is the number of degrees of freedom of the center of mass velocities, k_{B} is the Boltzmann constant, m_i is the mass and v_i is the velocity of particle i.

The easiest way to keep the temperature constant is to rescale the velocities of the particles after each timestep to maintain the kinetic energy at the correct value. However, with such a simple method, the system would not sample a proper canonical ensemble. During production runs, I performed all simulations using the Nosé-Hoover chain method.[157,158] In the original Nosé-Hoover method, a thermal reservoir and friction term are added to the equations of motions 2.2. The equations of motion are then:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \frac{p_\epsilon}{Q}\frac{d\mathbf{r}_i}{dt}, \tag{2.13}$$

where p_ϵ is the momentum of the friction parameter ϵ, which keeps the temperature close to the reference temperature T_0, and Q is the coupling strength. The equation of motion of ϵ is:

$$\frac{dp_\epsilon}{dt} = (T - T_0), \tag{2.14}$$

where T is the current temperature. The period of the oscillation τ_T, a parameter that must be preselected in GROMACS[114] to define the strength of the coupling, can be defined as:

$$\tau_T = \sqrt{\frac{4\pi^2 Q}{T_0}}. \tag{2.15}$$

However, the Nosé-Hoover thermostat can be non-ergodic, which is essential for a correct sampling of the free energy landscape. Hence, I used the Nosé-Hoover chain method, in which each Nosé-Hoover thermostat has its own Nosé-Hoover thermostat controlling its temperature. When infinite chains are used, the method is known to be ergodic. I used ten chains in all of simulations

presented in this thesis. The equations of motion for N chains are:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \frac{p_{\epsilon_1}}{Q_1}\frac{d\mathbf{r}_i}{dt}, \tag{2.16}$$

$$\frac{dp_{\epsilon_1}}{dt} = (T - T_0) - p_{\epsilon_1}\frac{p_{\epsilon_2}}{Q_2}, \tag{2.17}$$

$$\frac{dp_{\epsilon_{i=2\ldots N}}}{dt} = \left(\frac{p^2_{\epsilon_{i-1}}}{Q_{i-1}} - kT\right) - p_{\epsilon_i}\frac{p_{\epsilon_{i+1}}}{Q_{i-1}}, \tag{2.18}$$

$$\frac{dp_{\epsilon_N}}{dt} = \left(\frac{p^2_{\epsilon_{N-1}}}{Q_{N-1}} - kT\right). \tag{2.19}$$

In the simulations in this thesis, the solvent and protein are coupled to different thermostats to stop the protein from freezing.[159]

2.9. Barostats

Experimental observables in condensed matter physics and protein physics are normally obtained at constant temperature and pressure. Therefore, it is common to perform production MD simulations in the NPT ensemble. To achieve this goal, a barostat that keeps the pressure at a constant value is needed. In such a simulation, the volume of the system will be allowed to fluctuate, and care must be taken so that the system does not blow up and voids of vacuum are created in the simulation box.

Pressure in an MD simulation can be calculated with the following equation:

$$\mathbf{P} = \frac{2}{V}(\mathbf{E}_{kin} - \Xi), \tag{2.20}$$

where V is the volume of the system and $\Xi = -\frac{1}{2}\sum_{i<j}\mathbf{r}_{ij} \otimes \mathbf{F}_{ij}$ is the virial tensor and $\mathbf{E}_{\text{kin}} = \frac{1}{2}\sum_i^N m_i\mathbf{v}_i \otimes \mathbf{v}_i$. The scalar pressure for isotropic systems is then calculated as:

$$P = \text{trace}(\mathbf{P}/3). \tag{2.21}$$

In all production runs, I used the Parrinello-Rahman barostat to mantain the pressure constant.[160,161] This method, which is similar to the Nosé-Hoover thermostat, is stable and samples a true NPT ensemble. The box vectors are represented by the matrix $\mathbf{b}$ and follow the following equation of motion:

$$\frac{d\mathbf{b}^2}{dt^2} = V\mathbf{W}^{-1}\mathbf{b}'^{-1}(\mathbf{P} - \mathbf{P}_0), \tag{2.22}$$

where $\mathbf{W}$ determines the coupling strength, $\mathbf{b}'$ is the corrected matrix to consider an anistropic system in a non-cubic box, $\mathbf{P}$ is the current pressure and $\mathbf{P}_0$ is the reference pressure. The equations of motion are then:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \mathbf{M}\frac{d\mathbf{r}_i}{dt}, \tag{2.23}$$

$$\mathbf{M} = \mathbf{b}^{-1}\left(\mathbf{b}\frac{d\mathbf{b}'}{dt} + \frac{d\mathbf{b}}{dt}\mathbf{b}'\right)\mathbf{b}'^{-1}, \tag{2.24}$$

The coupling strength is calculated as:

$$(\mathbf{W}^{-1})_{ij} = \frac{4\pi\beta_{ij}}{3\tau_p^2 L}, \tag{2.25}$$

where β_{ij} is the isothermal compressibility, τ_p is the pressure time constant and L is the largest box matrix element.

2.10. Analysis

The goal of performing MD simulations is to understand equilibrium or dynamical properties of the system under study. In the case of proteins, this may include contact between residues, major motions of protein sections, changes in secondary structure or the estimation of experimental observables. In general, for MD simulations, properties are estimated as time-averages:

$$\langle A(\mathbf{r},\mathbf{v})\rangle_t = \lim_{T\to\infty}\frac{1}{T}\int_{t=0}^{T} A(\mathbf{r}(t),\mathbf{v}(t))dt, \tag{2.26}$$

where t is time, T is the simulation time, and A is the property of interest. However, it is commonly more interesting to know the ensemble-averaged properties. In the case of the microcanonical ensemble the following equation can be used:

$$\langle A(\mathbf{r},\mathbf{v})\rangle_{\text{ensemble}} = \frac{\int A(\mathbf{r},\mathbf{v})\delta(H(\mathbf{r},\mathbf{v})-E)drdv}{\int \delta(H(\mathbf{r},\mathbf{v})-E)drdv}, \tag{2.27}$$

where H is the Hamiltonian of the system and E is the energy. According to the ergodic hypothesis $\langle A(\mathbf{r},\mathbf{v})\rangle_t = \langle A(\mathbf{r},\mathbf{v})\rangle_{\text{ensemble}}$. In practice, a simulation can only be performed for a limited amount of time. Selected observables are then calculated to estimate the convergence of the simulation. The first part of the simulation is then considered as equilibration and discarded. Ensemble properties are estimated from the converged simulation. The validity of this assumption depends on how much phase space has actually been sampled.

3. Simulations of Functional Amyloid Aggregation

The discovery of functional amyloids has disproved the notion that amyloids are inherently toxic, and opened multiple questions, such as what is the difference between functional and aberrant amyloids, and how do functional amyloids avoid toxicity.[8,162] Most evidence shows that small, soluble and more disordered oligomers and not amyloid fibrils are the main toxic agents in amyloid-related diseases.[42] Hence, in this part of the thesis I study the difference in oligomer formation between functional and aberrant amyloidogenic peptides. In particular, I study the oligomer formation of kassinin and neuromedin K, which are part of the tachykinin peptides family and form functional amyloids and the section 25-35 of Aβ (Aβ_{25-35}), the peptide associated with Alzheimer's disease.

Tachykinin peptides are a family of neuropeptides that can be found in animals ranging from amphibians to mammals and have different functions like contracting muscles, exciting neurons and dilating blood vessels.[163] It was recently found that many tachykinin peptides, other neuropeptides and hormone proteins are stored in secretory granules as amyloids.[24,164] Tachykinin peptides share a common C-terminus: -Phe-X-Gly-Leu-Met-NH_2, where X is an aromatic or aliphatic residue.[163] This C-terminus makes tachykinin peptides particularly interesting because of its similarity with Aβ_{25-35} (see table 3.1), a section of Aβ which forms amyloids[165,166] and is toxic[167,168] on its own. Because Aβ_{25-35} is shorter than Aβ and has similar properties as Aβ, it has been used as a proxy for the full-length peptide.[169] Moreover, tachykinin peptides can coaggregate with Aβ[170] and decrease its toxicity.[171] Tachykinin peptides and Aβ_{25-35} are a perfect example to study the difference in oligomer formation between toxic and functional amyloids, because they have similar primary structure, but only Aβ_{25-35} is toxic. Moreover, the peptides are short enough to obtain converged results when studying them with MD simulations.

In this chapter, I study the aggregation of two tachykinin peptides: kassinin and neuromedin K, and Aβ_{25-35} by means of atomistic explicit-solvent MD simulations. The results of this chapter have been published in *The Journal*

Table 3.1.: Primary structure of Aβ_{25-35}, kassinin and neuromedin K.

Peptide	Sequence
Aβ_{25-35}	H-GSNKGAIIGLM-OH
Kassinin	H-DVPKSDQFVGLM-NH_2
Neuromedin K	H-DMHDFFVGLM-NH_2

of Physical Chemistry B.[1]

3.1. Methodology

The aggregation of Aβ_{25-35}, kassinin and neuromedin K was studied by allowing them to freely aggregate in MD simulations. For each peptide, six monomers were randomly placed in a box with a side length of 10 nm (equivalent to a concentration of $\approx$ 10 mM) and simulated for 300 ns. Each simulation box contained around 130,000 atoms. Each system was replicated 5 times to take into account the effects of the stochasticity on the aggregation process. One simulation of each system was extended to 1 µs to study the changes in protein conformation in longer simulations. Neuromedin K was simulated with a charged His^3 residue. The simulation was repeated with an unprotonated His^3 (protonated only at N_ϵ) and the effect was found to be negligible. The systems were modeled with the OPLS force field[118,172] and the TIP4P water model.[134] The simulations were performed both at 0 mM NaCl concentration (only with enough counterions to neutralize the system) and at 150 mM NaCl. The simulations of Aβ_{25-35} and kassinin were performed with both free and amidated C-termini. Aβ_{25-35} was also studied in a restrained hairpin-like conformation.

The starting conformations for the aggregation simulations were obtained from short simulations of the monomer. The monomer simulations for Aβ_{25-35}, kassinin and neuromedin K were started from the structures with Protein Database code 1QYT (model 1),[173] 1MYU[174] and 1P9F[175] respectively. The monomer simulations were performed before the aggregation simulations, because the experimental monomer structures were obtained in micelles, so it is expected that the experimental conformations have a larger propensity towards α-helical conformations than the one expected in water. Each monomer was placed in a cubic box, then solvated, and enough counterions were added to neutralize the system. After an initial minimization using the steepest descent algorithm, the systems were equilibrated with successive 100 ps NVT and 100 ps NPT simulations. During the equilibration, the peptide coordi-

nates were constrained. 50 ns production runs were then performed for each peptide. Random monomer conformations were picked from the last 25 ns of the simulations to start the aggregation studies.

For the aggregation simulations, six peptides were introduced in a cubic box, the system was then solvated and enough counterions were added to neutralize the system. The systems were equilibrated with 1 ns NVT and 1 ns NPT simulations with the peptide coordinates being restrained during the equilibration. The systems were then allowed to evolve without restraints during the 300 ns production runs. In all simulations, a 1 nm cutoff was used for van der Waals interactions and short-range electrostatic interactions. Electrostatic interactions were calculated with the particle mesh Ewald algorithm[150] with a grid of 0.12 nm and periodic boundary conditions. The bond length in the peptides were constrained with the P-LINCS algorithm[142] with an expansion order of 6 and the water molecules were constrained with the SETTLE algorithm.[144] A 5 fs time step was used, permitted by the use of virtual sites for all peptide hydrogen atoms.[145] During both monomer and aggregation production runs, the temperature was kept constant at 310 K using the Nosé -Hoover thermostat[157,158] and pressure was kept constant at 1 bar using the Parrinello-Rahman barostat.[160,161]

All systems were simulated with Gromacs 4.6[84] and figures of the peptides were produced with VMD.[176] The simulations were analyzed using Gromacs tools[84] and the MDAnalysis package.[177] Monomers were clustered using the algorithm from Daura et al.[178] For the monomer contact map, two residues i and j are considered in contact if any atom of residue i is closer than 0.4 nm of any atom in residue j. Two peptides i and j are considered as aggregated if the distance between any atom of peptide i is closer than 0.5 nm from any atom of peptide j. For oligomer contact maps, residues i and j were considered in contact if any atom of residue i is closer than 0.5 nm of any atom in residue j. For these plots, all oligomers (i.e., from dimers to hexamers) of the aggregation simulations were considered. The solvent accessible surface area (SASA) for hydrophobic residues were calculated with the algorithm from Eisenhaber et al.,[179] with atoms with charges less than 0.2e considered as hydrophobic. The secondary structure was calculated using the Define Secondary Structure of Proteins (DSSP) algorithm.[180] To estimate how spherical the oligomers are, the asphericity was calculated:[181]

$$\Delta = \frac{3}{2}\frac{\sum_{i=1}^{3}(\lambda_i - \bar{\lambda})^2}{(\mathrm{tr}\,\mathbf{T})^2}, \tag{3.1}$$

with

$$\bar{\lambda} = \frac{\mathrm{tr}\,\mathbf{T}}{3}, \tag{3.2}$$

where $\mathbf{T}$ is the inertia tensor and λ_i are the eigenvalues of $\mathbf{T}$. The asphericity is a number between 0 and 1: when 0, the oligomer is perfectly spherical, and the closer the asphericity is to 1, the more aspherical the oligomer is. To estimate hexamer properties, snapshots were taken every 500 ps.

3.2. Results and Discussion

3.2.1. Aggregation of tachykinin peptides and Aβ_{25-35}

To study the difference in aggregation between functional and aberrant amyloidogenic peptides, I simulated the aggregation of Aβ_{25-35} and two amyloid-forming functional tachykinin peptides: kassinin and neuromedin K. In figure 3.1a, the average maximum oligomerization state for the five simulations for each peptide is plotted against time. The tachykinin peptides aggregate faster than Aβ_{25-35}. Both tachykinin peptides achieve the maximum oligomerization state in around 100 ns. However, Aβ_{25-35} does not reach the hexamer state even after 300 ns. To better quantify the aggregation process, average aggregation rates were calculated for each system and summarized in table 3.2. These results suggest that functional amyloids may avoid toxicity by aggregating faster than aberrant amyloids and surpassing the toxic oligomer phase. This has not been observed for tachykinin peptides but it has been observed for the functional amyloid Pmel17, which aggregates faster than Alzheimer's Aβ and Parkinson's α-synuclein.[23] To test if the results depend on a particular oligomerization state, I added more monomers at the end of each simulation so that the concentration of free monomers was again $\approx$ 10 mM and extended the simulation for 100 ns. The difference in oligomerization state relative to the beginning of this simulation is plotted in figure 3.2. Both kassinin and neuromedin K reach the maximum difference of six in around 100 ns, while Aβ_{25-35} reaches an average difference of of 3.2 by the end of the 100 ns simulation.

To understand the difference between tachykinin peptides and Aβ_{25-35} aggregation, I calculated different properties for the hexamers (see table 3.3). First, I studied the secondary structure content of hexamers. The β-sheet content is relatively low for all peptides, with a minimum of $3.2 \pm 0.2\%$ for neuromedin K and a maximum of $4.2 \pm 0.2\%$ for kassinin. Considering that β-sheet is a hallmark of amyloid aggregates,[9] I suspected that longer simulations are needed to observe a larger content of β-sheet as found in other

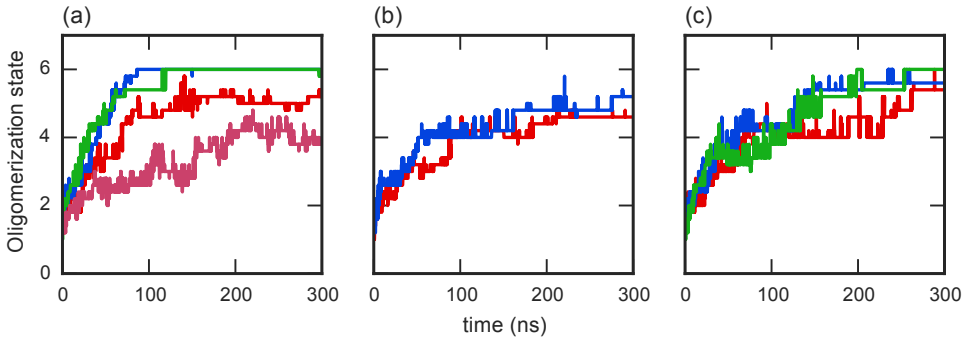

Figure 3.1.: Average oligomerization state as a function of time for Aβ_{25-35} (red), kassinin (blue) and neuromedin K (green) with (a) standard and (b) non-standard termini. In (a) the results for Aβ_{25-35} with uncapped termini and for kassinin and neuromedin K with an uncapped N-terminus and a capped C-terminus, are plotted. In (b) the results for Aβ_{25-35} with an uncapped N-terminus and a capped C-terminus, and for kassinin with uncapped termini, are plotted. In (c) the simulations with 150 mM of NaCl are plotted for the peptides with standard termini. In addition, in (a) the results for Aβ_{25-35} in a constrained hairpin-like conformation are plotted in pink (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

Table 3.2.: Average aggregation rate for each system studied.

System	NaCl (mM)	Aggregation rate (monomer/ns)
Aβ_{25-35}	0	0.0138
Kassinin	0	0.0519
Neuromedin K	0	0.0475
Aβ_{25-35} with capped C-terminus	0	0.0125
Kassinin with uncapped C-terminus	0	0.0123
Restrained Aβ_{25-35}	0	0.0085
Kassinin in a β-hairpin	0	0.0173
Aβ_{25-35}	150	0.0108
Kassinin	150	0.0157
Neuromedin K	150	0.0147

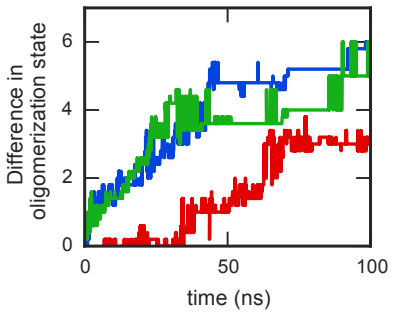

Figure 3.2.: Average difference in oligomerization state for the extension of the aggregation simulations of Aβ_{25-35} (red), kassinin (blue) and neuromedin K (green). The extension of these simulations were obtained by adding enough monomers in the system to attain a monomer concentration of ≈ 10 mM in each simulation (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

simulations of peptide aggregation.[80–82] Therefore, I extended one simulation for each system to 1 µs. At the end of these simulations, Aβ_{25-35}, kassinin and neuromedin K had β-sheet contents of 12 %, 14 % and 13 % respectively. This shows that longer simulations are needed to obtain large β-sheet content. The α-helix content is small for all systems. However, I observe a somewhat larger α-helical content for neuromedin K, which correlates which circular dichroism experiments.[30] There is some experimental evidence that oligomers with larger hydrophobic surface are more toxic,[182–185] hence, the average hydrophobic SASA was calculated for the simulations. No correlation between toxicity and hydrophobic SASA was observed. However, when the hydrophobic SASA was normalized by the total SASA, a slightly higher relative hydrophobic SASA was observed for Aβ_{25-35} (0.63 ± 0.01) than for the tachykinin peptides (0.52 ± 0.02 and 0.60 ± 0.02). When visualizing the oligomers with VMD, I observed that certain oligomers were more extended than others, so the asphericity was estimated for all peptides. The most aspherical oligomers are observed for kassinin, but the standard deviations of the measurements are too high to properly distinguish between the different peptides.

3.2.2. Influence of the C-termini

Aiming to gain a deeper understanding of what drives the faster aggregation of tachykinin peptides compared to Aβ_{25-35}, I plotted contact maps for the oligomers (see figure 3.3) to see which are the most probable contacts between

Table 3.3.: Summary of properties of aggregated hexamers.

Peptide	Hydrophobic SASA (nm^2)	Relative hydrophobic SASA (%)	Asphericity	β-sheet (%)	α-helix (%)
$A\beta_{25-35}$	31.2 ± 1.6	0.63 ± 0.01	0.18 ± 0.12	3.6 ± 0.2	0.19 ± 0.01
Kassinin	34.9 ± 3.1	0.60 ± 0.02	0.26 ± 0.16	4.2 ± 0.2	0.14 ± 0.01
Neuromedin K	26.3 ± 2.1	0.52 ± 0.02	0.14 ± 0.08	3.2 ± 0.2	1.72 ± 0.12

different peptides in the simulations. In both kassinin and neuromedin K, the C-termini interact strongly with each other. This interaction is permitted by the neutral C-termini and the C-terminal hydrophobic residues. However, in $A\beta_{25-35}$ the most relevant interactions are between the N-terminus and the C-terminus. This is expected because both termini have opposite charges and the interaction between C-termini is less likely because of their negative charge. This is relevant because it has been observed that an amidated $A\beta_{25-35}$ has lower toxicity than an $A\beta_{25-35}$ with a free C-terminus.[168] It is also common to observe that mutations of neutral to charged amino acids can abolish aggregation. For example, the mutation of the peptide STVIIE to STVIKT prevents amyloid formation.[186] It should also be mentioned that amidation is a standard post-transcriptional modification for peptides[187] and, in particular, it is essential for the agonist activity of tachykinin peptides.[188] The fact that C-terminal amidation is essential for tachykinin peptide aggregation suggests that tachykinin peptides have an amidated C-terminal to permit its correct aggregation and that tachykinin peptide receptors have evolved to adapt to this change. The contact maps further show that there is a also a strong interaction between the hydrophobic residues in the intermediate section of the peptides. For example, there is a strong interaction between the isoleucine residues at positions 7 and 8 from different $A\beta_{25-35}$ peptides in the oligomers and between residues 5–7 in neuromedin K.

To test if the C-terminal charges are essential for correct aggregation, I studied the aggregation of $A\beta_{25-35}$ with a non-standard amidated C-terminus and kassinin with a non-standard free carboxylate C-terminus. The oligomerization state as a function of time for both systems are plotted in figure 3.1b. As expected, the aggregation of kassinin with a free carboxylate C-terminus decreases considerably. This confirms that C-terminal amidation is important for a correct and fast aggregation of kassinin. However, surprisingly, the aggregation of $A\beta_{25-35}$ also decreases when the C-terminus being amidated. It should be considered that the aggregation simulation of $A\beta_{25-35}$ is performed at a concentration of ≈ 10 mM, which is much higher than the experimental

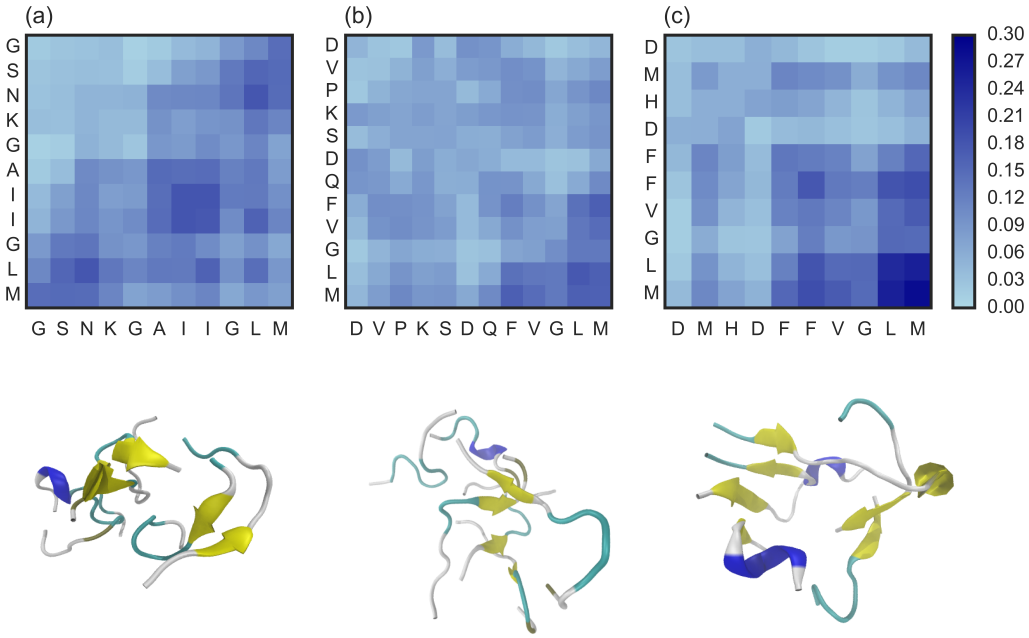

Figure 3.3.: Oligomer contact maps and example snapshots for the oligomers of the aggregation simulations of (a) Aβ_{25-35}, (b) kassinin and (c) neuromedin K (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

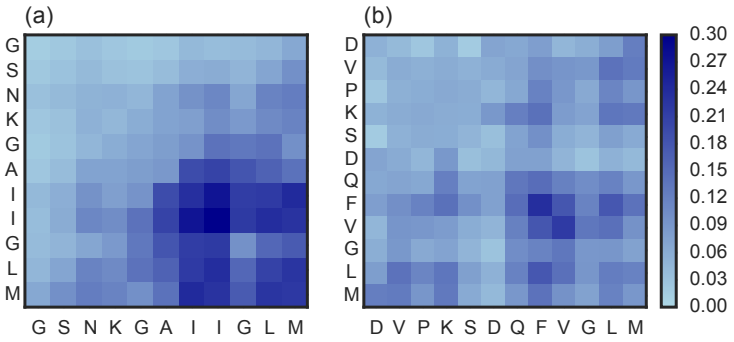

Figure 3.4.: Oligomer contact maps for (a) Aβ_{25-35} with an amidated C-terminus and (b) kassinin with a free C-terminus (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

concentration of amyloid forming peptides *in vitro*. Hence, some aggregation is expected in my simulations even if it would not happen *in vitro*. Hence, the decrease in aggregation of Aβ_{25-35} when amidated may represent a lack of aggregation at an experimental concentration. In figure 3.4, I plotted the oligomer contact maps for Aβ_{25-35} with an amidated C-terminus and kassinin with a free carboxylate C-terminus. As expected, there is a strong interaction between C-termini in the simulations of Aβ_{25-35}, while in the case of kassinin the interaction between C-termini diminishes. However, the interaction between N- and C-termini in kassinin is not as strong as in the simulation of Aβ_{25-35} with uncapped termini. These results show the importance of the peptide-specific charges of the termini in peptide aggregation.

3.2.3. Importance of monomer conformations

To understand what drives aggregation, I studied the conformations of the monomers before aggregation. Figure 3.5 shows the contact maps and most relevant conformations, according to the cluster analysis using the method of Daura et al.,[178] for the 50 ns monomer simulations of the three peptides. In the case of Aβ_{25-35}, the most stable conformation is a hairpin-like structure in which the positively charged Lys^4 interacts with the negatively charged terminal Met^{11}. There is also a strong interaction between the positively charged N-terminus and the negatively charged C-terminus. Similar results were obtained by Larini and Shea[93] with the OPLS force field and TIP3P water model, even though they observed a stronger interaction between the termini. Kassinin and neuromedin K have more extended conformations with little contact between the termini. Some helicity is observed in the N-terminus of neuromedin K. According to the DSSP algorithm, the central region of Aβ_{25-35} has a turn

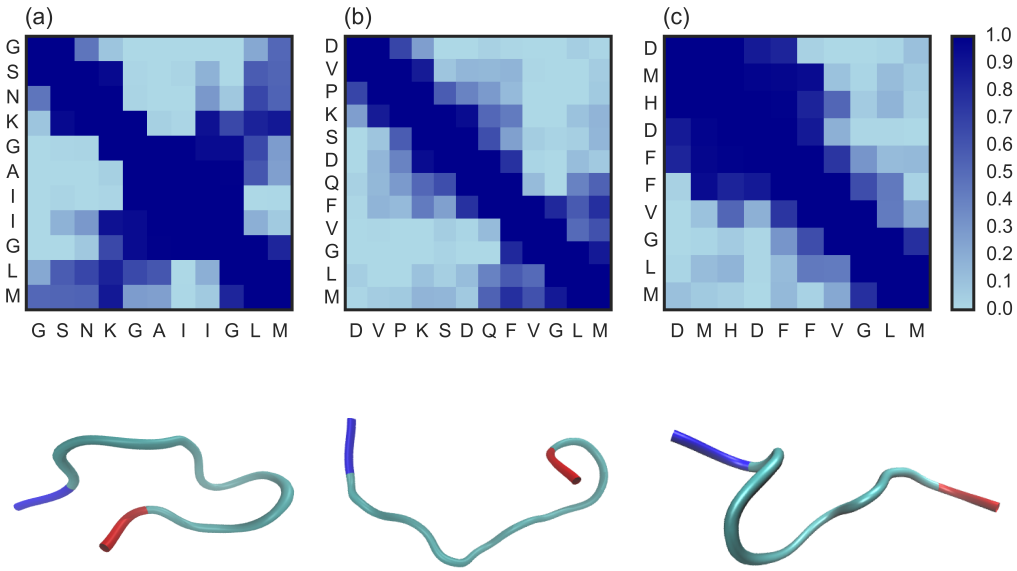

Figure 3.5.: Monomer contact maps and the most representative monomer conformations according to the cluster analysis for (a) Aβ_{25-35}, (b) kassinin and (c) neuromedin K for the 50 ns monomer simulations. The N-terminus is plotted in blue and the C-terminus in red (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

content between 70% and 80 % and these high values are not observed for any of the two tachykinin peptides (see figure 3.6).

To further understand the influence of the monomer conformations on aggregation, I plotted the monomer contact maps and most representative conformations for the simulations of Aβ_{25-35} with an amidated C-terminus and kassinin with a free C-terminus (see figure 3.7). I expected to observe hairpin-like structures for both peptides because their aggregation is slower than that for the peptides with standard termini. Hairpin-like structures are observed for both peptides. On the one hand, Aβ_{25-35} forms a hairpin-like structure because of a nonspecific interaction between Met^{11} and the central section of the peptide (between residues Ser^{2} and Ile^{7}). On the other hand, kassinin forms a hairpin-like structure because of a strong electrostatic interactions between the charged C-terminus and Lys^{4}, similar to the results observed for Aβ_{25-35} with charged termini. Haspel et al.[189] also observed a hairpin-like structure in the simulation of neuromedin K with charged termini using the AMBER03 force field.

One of the main advantages of simulations over experiments is that we can

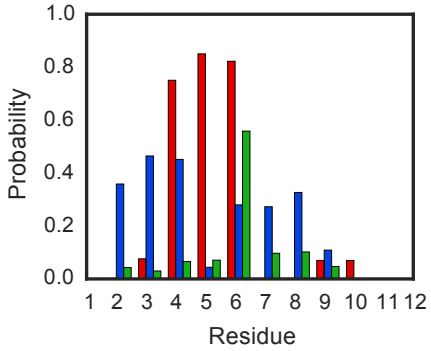

Figure 3.6.: Turn content per residue for the monomer simulations of Aβ_{25-35} (red), kassinin (blue) and neuromedin K (green) for the 50 ns monomer simulations (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

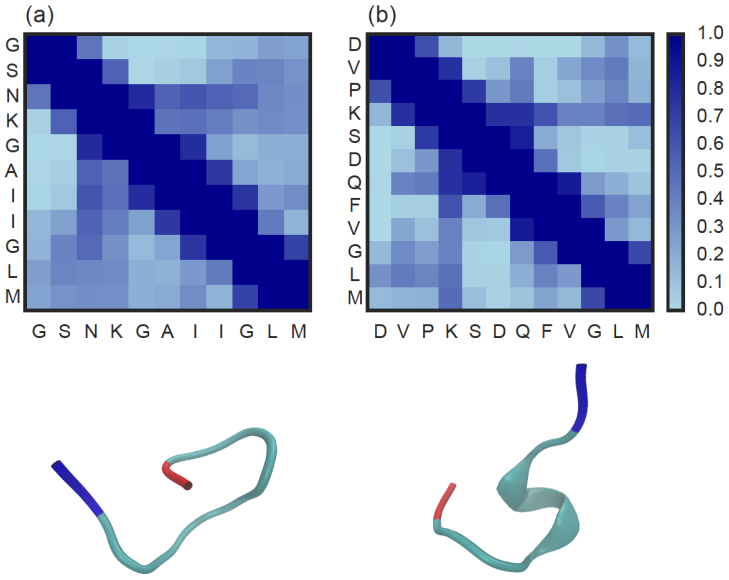

Figure 3.7.: Monomer contact maps and the most representative conformations according to the cluster analysis for the simulations of (a) Aβ_{25-35} with an amidated C-terminus and (b) kassinin with a free C-terminus for the 50 ns monomer simulations. The N-terminus is plotted in blue and the C-terminus is plotted in red (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

perform unrealistic simulations to study particular questions that cannot be performed experimentally. For example, it would be possible to delete electrostatic interactions to observe if they play an important influence on a certain dynamical feature of proteins. In our case, it would be of interest to force the peptides into a hairpin-like conformation to observe if this limits aggregation even further. I decided to repeat the simulation of Aβ_{25-35} aggregation with a restraint to force a close interaction between the charged C-terminus and Lys^4. The restraint was applied according to the following equation:

$$V(r_{ij}) = \begin{cases} \frac{1}{2}k(r_{ij} - r_0)^2, & r_{ij} < r_0, \\ 0, & r_0 \leq r_{ij} < r_1, \\ \frac{1}{2}k(r_{ij} - r_1)^2, & r_1 \leq r_{ij} < r_2, \\ \frac{1}{2}k(r_2 - r_1)(2r_{ij} - r_2 - r_1), & r_2 \leq r_{ij}, \end{cases} \tag{3.3}$$

where r_{ij} is the distance between the carboxylic acid carbon of the C-terminus and the side-chain nitrogen of Lys^4, $r_0 = 0.2\,\text{nm}$, $r_1 = 0.5\,\text{nm}$, $r_2 = 0.6\,\text{nm}$ and $k = 10^5\,\text{kJ mol}^{-1}\,\text{nm}^{-2}$. The oligomerization state as a function of time is plotted in figure 3.1a, and it can be observed that the aggregation diminishes considerably compared to that of the unconstrained Aβ_{25-35}. This confirms that forcing a peptide into a hairpin-like conformation can decrease its aggregation propensity.

Because of the importance of the monomer conformations on protein aggregation, I decided to extend the 50 ns monomer simulations up to 1 µs. The contact map and most representative structures for the three peptides are plotted in figure 3.8. The conformation of Aβ_{25-35} and neuromedin K did not change considerably, while the conformation of kassinin changed into a β-hairpin. The aggregation simulation of kassinin was then repeated to account for the hairpin-like conformations. To this end, monomer conformations were randomly picked from the last 800 ns of the 1 µs simulation and introduced in a cubic box and the same simulation protocol as before was applied. The oligomerization state as a function of time and the intermolecular contact map are plotted in figure 3.9. The aggregation kinetics decreases compared to the initial simulation, confirming that the conformation of the monomer is essential to determine the aggregation propensity. However, even with this decrease in aggregation, the aggregation of kassinin is still faster than the one of Aβ_{25-35}. From the intermolecular contact map, it can be observed that the contacts between the C-termini have diminished, which suggests that when the termini are free they drive aggregation, while when they form stable intramolecular

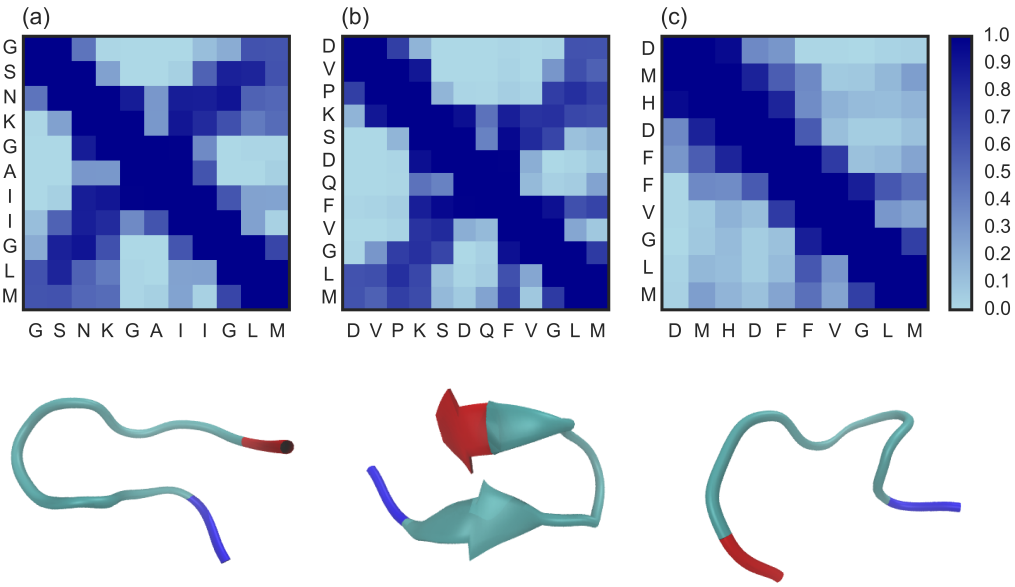

Figure 3.8.: Monomer contact maps and most representative conformations according to the cluster analysis for the 1 µs monomer simulations of (a) Aβ_{25-35}, (b) kassinin and (c) neuromedin K. The N-terminus is plotted in blue and the C-terminus is plotted in red (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

contacts, they cannot drive aggregation.

The results observed here regarding the importance of monomer conformations is probably extensible for small peptides. However, it is hard to know if they would be applicable for larger peptides because they have more complicated internal dynamics. However, it has been observed that when the residues 17–36 of Alzheimer's Aβ_{40} are restrained in a β-hairpin conformation, its aggregation is inhibited.[190] It is also known that small sections of proteins usually drive the aggregation of the entire proteins.[186] Hence, our conclusions could be of general validity.

3.2.4. Influence of salt concentration

Another interesting question regarding the formation of amyloid is the influence of external factors on aggregation. It has been observed that amyloid formation depends on pH,[191] temperature,[192] salt ions and ionic strength,[193–195] hydrostatic pressure[196] and the presence of crowding agents.[197] Most of the information on the influence of external factors is on fibril formation, but little is known about the effects on the kinetics and conformations of oligomers. Here,

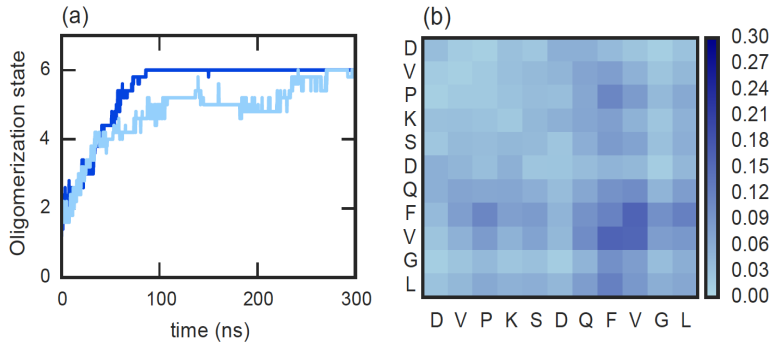

Figure 3.9.: (a) Oligomerization state for kassinin in an extended monomer conformation (dark blue) and a hairpin-like monomer conformation (light blue) and (b) oligomer contacts maps for the aggregation simulation of kassinin in a hairpin-like monomer conformation. (Adapted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

I repeated the monomer and aggregation simulations for the peptides with standard termini at 150 mM of NaCl to study the influence of ionic strength on oligomer formation. The aggregation of kassinin and neuromedin K are slower at high salt concentration compared to those at low salt concentration (i.e., only neutralizing ions), while the aggregation of Aβ_{25-35} is similar at both concentrations. Because of the change in aggregation kinetics, the three peptides now have comparable aggregation kinetics (see figure 3.1c). However, tachykinin peptides still aggregate faster. In figures 3.10 and 3.11, I plotted the intermolecular contact maps for the oligomers and the intramolecular contact maps for the monomers of the three peptides. The most important contacts between monomers are similar at low and high salt concentrations. Monomers are mostly found in hairpin-like structures for Aβ_{25-35} and kassinin, which explains the lower aggregation kinetics of kassinin. Our results are supported by the experimental results of Naldi et al.,[166] who found that salt concentration does not change the aggregation kinetics of Aβ_{25-35}. The fact that the results are so dependant on salt concentration demonstrates the importance of simulating protein aggregation at the correct salt concentration. Moreover, it displays the difficulty on making general conclusions about the intrinsic factors that drive protein aggregation, because small differences in external factors may change the results drastically. This fits the picture that functional amyloids need to be tightly regulated for their correct aggregation.[8]

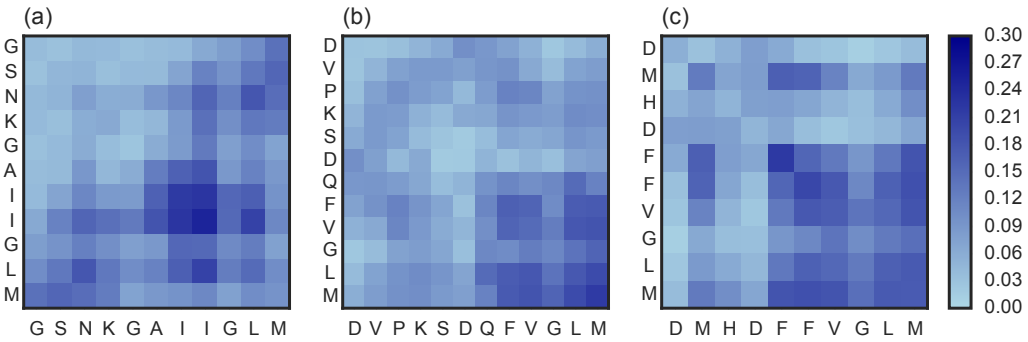

Figure 3.10.: Oligomer intermolecular contact maps of the aggregation simulations of (a) Aβ_{25-35}, (b) kassinin and (c) neuromedin K at 150 mM NaCl (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

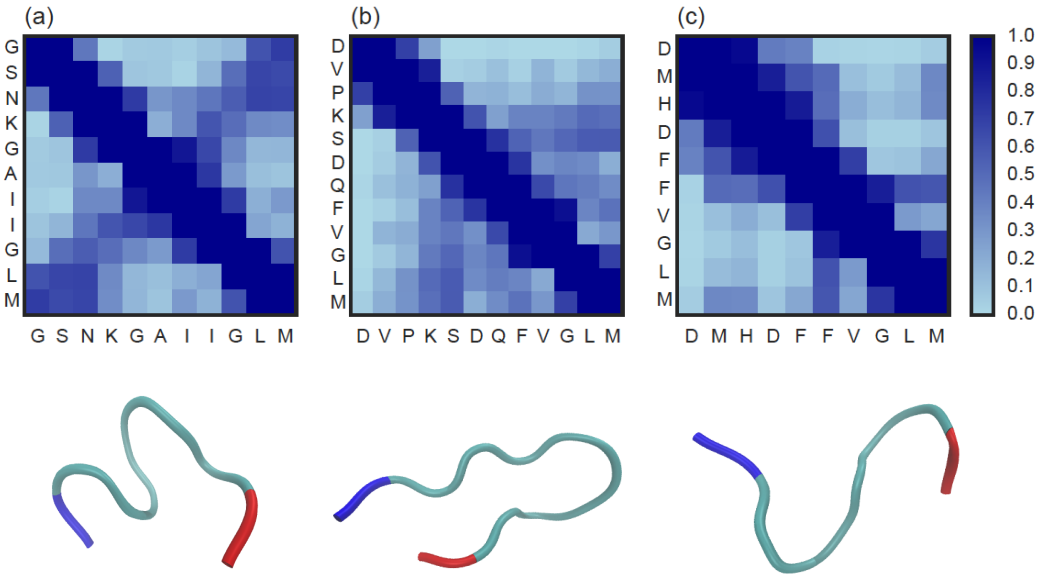

Figure 3.11.: Monomer contact maps for the monomer simulations at 150 mM NaCl of (a) Aβ_{25-35}, (b) kassinin and (c) neuromedin K. The N-terminus is plotted in blue and the C-terminus is plotted in red (Reprinted with permission from Carballo-Pacheco et al.[1] Copyright 2015 American Chemical Society).

3.3. Summary

In this chapter, I studied the aggregation of both functional and aberrant amyloids using atomistic explicit-solvent simulations. In particular, I studied the aggregation of two tachykinin peptides, kassinin and neuromedin K, and Alzheimer's Aβ_{25-35}. These peptides have a similar primary structure but, while the first two are functional and non-toxic, the latter is disease-related and toxic. Moreover, tachykinin peptides are known to be able to influence the aggregation of Aβ. Few simulations of functional amyloids have been performed so far,[198–200] and none of them have concentrated on the difference between functional and aberrant amyloids. Considering that it is now known that oligomers are the toxic species in amyloid diseases,[42] it is important to understand the difference in oligomer formation between functional and aberrant aggregation.

One of the key observation of this chapter is that functional amyloids seem to aggregate faster than aberrant amyloid peptides. This suggests that functional amyloids may avoid toxicity by aggregating faster and surpassing the toxic phase. This has not been observed for tachykinin peptides, but it has been experimentally verified for the functional amyloid Pmel17.[23] Functional amyloids may also avoid toxicity by being tightly regulated and only aggregating under certain conditions.[8] Another important observation is the fact that the charged termini are essential in modulating peptide aggregation. In particular, when peptides have a charged C-terminus, they seem to aggregate more slowly than with an amidated C-terminus. This is relevant because tachykinin peptides cannot perform their agonist activity when they are not amidated,[188] and if Aβ_{25-35} is amidated its toxicity is reduced.[168] Another key observation is that the conformation of the monomers are essential to understand the aggregation kinetics: when the peptides are extended their aggregation is fast, while when the peptides are in hairpin-like conformations their aggregation is limited. The reason for this could be that when peptides are in hairpin-like conformations the charged interactions sites in the termini are occupied and are not free to drive aggregation. However, when the peptides are in extended conformations, the charged interactions sites close to the termini are free to drive aggregation. There is evidence that stabilizing a β-hairpin in the central section of Aβ inhibits aggregation,[190] which means that our conclusion could be of general applicability. Finally, I have also studied the effect of changing the salt concentration from 0 mM to 150 mM on peptide aggregation. The salt concentration has a strong influence on protein aggregation for the tachykinin peptides. However, the effect on Aβ_{25-35} is minimal, which correlates to ex-

perimentally observed results.[166] The effect of the salt concentration on the results shows the importance of external factors on aggregation.

In this chapter, I performed all simulations using only one force field, namely OPLS. However, it is still an open question if the aggregation results depend strongly on the force field used, particularly considering that the observed difference in aggregation between different peptides were not too large. Moreover, when visualising the simulations with VMD, it appears that some oligomers are unphysically stable. Hence, I ask myself if force fields are overstabilising protein-protein aggregation. In the next chapter, I will study the effect of force fields both on intrinsically disordered peptides and protein aggregation.

4. Comparison of Force Fields for Intrinsically Disordered Peptides and Protein Aggregation

The two main limitations of MD simulations in accurately modeling protein dynamics are the lack of sampling of the free energy surface and the accuracy of the force fields. However, with the creation of highly optimized parallel[85,110] and GPU-specific[86,87] MD codes, the creation of special-purpose hardware[83] and the ability to join multiple short MD simulations for the construction of Markov state models,[201–203] sampling of small and middle-sized proteins is no longer a problem. These long simulations have made it clear that the second problem, the accuracy of the force fields, is still far from being solved. In this chapter, I will focus on the accuracy of force fields for simulating intrinsically disordered proteins (IDPs) and protein aggregation.

Force fields have been developed by fitting to quantum mechanical and experimental thermodynamical data of small molecules. Many have shown that protein simulations depend strongly on the used force field,[126,127,148,204–206] and multiple corrections have been proposed based on extensive simulations performed in the last few years. For example, AMBER99SB[207] has been recently updated to AMBER99SB*ILDN[208–210] by modifying backbone and side-chain torsional angles parameters. Also, CHARMM22[211] was recently updated to CHARMM22*[210] by modifying backbone and some side-chain parameters. Newer force fields such as AMBER99SB*ILDN and CHARMM22* were proven to fold both α-helical and β-sheet peptides.[126] Hence, I expect these new force fields to perform better than older force fields for IDPs and protein aggregation.

IDPs are proteins that, unlike folded proteins, do not have a stable 3-dimensional structure. They have important biological roles such as in cell signalling,[212] but are also related to amyloid-related diseases because many amyloidogenic peptides are IDPs in the monomer state. Because of their lack of stable structure it is hard to study IDPs with experimental methods such as X-ray crystallography, while nuclear magnetic resonance (NMR) can usually only provide ensemble averaged properties. Hence, computer simulations

are used, often in combination with experimental data, to obtain a detailed understanding of the structure and dynamics of IDPs. It has now become customary to study IDPs using atomistic explicit-solvent MD simulations.[213–220] Many studies have been performed to understand the influence of force fields on IDPs,[140,221–223] and the conclusion is that newer force fields are in general better than older force fields. Moreover, in most simulations the proteins have been found to be too collapsed compared to what is observed experimentally.

Simulations of protein aggregation are particularly difficult because of the low concentration at which most aggregation experiments are performed. However, it is now becoming common to study protein aggregation with atomistic MD simulations.[2,78–82,224,225] A few studies have concentrated on the effect of force fields on protein aggregation and the general conclusion is that most force fields overstabilize protein-protein interactions.[226–228] However, all of these studies were performed with proteins that do not suffer major conformational changes when interacting, unlike the proteins that form amyloids. Nguyen et al.,[229] on the other hand, studied the effect of force fields on amyloid formation and concluded that the main conformations sampled for the monomer, dimer and trimer varied between force fields. However, they only studied older force fields, and the effect on the final conformation and not on the aggregation kinetics. Finally, Bernahu and Hansmann[230] studied the influence of force fields on the stability of Aβ_{16-22} fibrils and found that AMBER99SB-ILDN performed the best.

In this chapter, I study two different systems to understand how well modern force fields model IDPs and protein aggregation. As a model system for IDPs, I study Aβ_{42}, a peptide that is associated with Alzheimer's disease. This protein was chosen because there is ample nuclear magnetic resonance (NMR) data[213,231–235] that can be used to validate the conformational ensemble sampled by the force fields, and because it has already been studied with computational methods that I can compare to.[213,233–249] However, no one has so far studied the influence of the latest protein force fields in long simulations of Aβ_{42}. As a model system for protein aggregation, I study the aggregation propensity of Aβ_{16-22} and two mutants: Aβ_{16-22}(F19V,F20V) and Aβ_{16-22}(F19L). Aβ_{16-22} is a section of Alzheimer's Aβ, which has shown to aggregate and form structured anti-parallel fibrils.[250,251] The first mutant was experimentally shown not to aggregate, while the latter one aggregates faster than the wild type.[252] I studied the aggregation of these three peptides with different force fields to analyse if they can discriminate the aggregation propensity of the different peptides. Parts of the results of this chapter have been

published in *Protein Science*.[3]

4.1. Methodology

4.1.1. Simulation of Aβ_{42}

To study if different force fields are good at modeling IDPs, I simulated Aβ_{42} using five different force fields. In particular, I used CHARMM22*,[210] AMBER99SB*ILDN[209,210,253] and AMBER99SBILDN-NMR,[254] each in combination with the TIP4P-Ew water model.[138] This water model was chosen because it has been shown to perform better than three-site water models.[255,256] These new force fields have been developed recently and have been shown to accurately model folded proteins and protein folding. Moreover, AMBER99SB[207] with the TIP4P-Ew water model[138] and OPLS[118,172] with the TIP3P water model[134] were also studied as these were considered to be the best force field combinations so far.[213,233–235,237] All simulations were performed using Gromacs 4.6.4.[84]

To obtain a converged ensemble of the conformational space sampled by each force field, I performed replica exchange MD (REMD) simulations,[89,257] a standard enhanced sampling algorithm.[88] In REMD, multiple replicas are run in parallel at different temperatures for each simulation. After a defined number of time steps, exchanges between replicas are attempted, which are accepted based on the Metropolis criteria:[258,259]

$$p = \min\left(1, e^{(E_i - E_j)\left(\frac{1}{k_B T_i} - \frac{1}{k_B T_j}\right)}\right) \tag{4.1}$$

where i and j are two different replicas, E is the energy of the system, T is the absolute temperature and k_B is the Boltzmann constant. In this way, detailed balance is preserved and the data obtained from the simulations are thermodynamically consistent. Sampling is enhanced because replicas at higher temperature are used to jump over free energy barriers, and the lower temperature replicas are used to sample the system at the relevant thermodynamic ensemble. A large amount of replicas may be needed because the energy of the system grows with the number of atoms, and if the difference in energy between replicas close in temperature is extremely high, the probability of an exchange being accepted is low. A recent improvement of the algorithm was proposed by Chodera and Shirts,[260] in which random exchanges are tested between all replicas (normally, exchanges are only tested between replicas that are located at

consecutive temperatures because they are the most likely to be accepted). In such a way, the sampling can be further enhanced and convergence is achieved faster. In my simulations, exchanges were attempted between adjacent replicas every 250 time steps. In addition, 32,768 (i.e., N^3, where $N = 32$ is the number of replicas, as suggested by Chodera and Shirts[260]) exchanges between random replicas were also attempted at each exchange window.

Simulations were started from the Aβ_{42} structure with PDB code 1IYT.[261] This structure was obtained in a solvent with a 80/20 ratio of hexafluoroisopropanol (HFIP) to water. The structure is mostly α-helical and is expected to be different in water. Therefore, I first performed a 500 K high temperature simulation to obtain a disordered structure. For this simulation, the structure was first minimized with the steepest descent algorithm. Then, a 10 ns MD simulation was performed with the AMBER99SB force field. This structure was used as the initial structure for all simulations.

All simulations were performed with standard protonation states at a neutral pH, with neutral histidines that are only protonated at the N_ϵ. For each REMD simulation, 32 replicas, which were exponentially distributed between 270 and 443 K, were used. First, the protein was introduced in a rhombic dodecahedron box with an edge length of 5.5 nm. The box was then solvated with between 3621 and 3636 water molecules. Each system was neutralized with 3 sodium atoms, and was then minimized with the steepest descent algorithm. Afterwards, the systems were equilibrated to the right temperature of each replica and 1 bar of pressure with a 1-ns NVT simulation and a 1-ns NPT simulation. During the equilibration, protein heavy atoms coordinates were constrained. A 200 ns production REMD was performed for each system. The temperature in each replica was kept constant with the Nosé-Hoover thermostat[157,158] and a time constant of 0.5 ps, and the pressure was kept constant with the Parrinello-Rahman barostat[160,161] with a time constant of 2 ps and a compressibility of $4.5 \times 10^{-5}\,\mathrm{bar}^{-1}$. The P-LINCS[142] and SETTLE[144] algorithms were used to constrain protein and water bonds, respectively. A time step of 4 fs was used, permitted by the use of virtual sites for protein hydrogens.[145] Van der Waals and short range electrostatic were calculated with a cutoff of 1.2 nm. Electrostatic interactions were calculated using the particle mesh Ewald method[150] with a Fourier spacing of 0.12 nm and periodic boundary conditions.

Considering that Aβ_{42} is an IDP, it is expected that it will sample an extended ensemble. Therefore, the distances between the peptide and its image were measured. The snapshots for peptides whose image was closer than 1 nm

were not considered for the analysis.

To establish which of the force fields is better, the ensembles obtained with the REMD simulations were compared with different experimental NMR observables. For calculating the observables, snapshots taken every 0.1 ns of the replicas between 290 and 310 K were used. First, C_α, C_β, H_α and H_N chemical shifts were calculated with SPARTA+.[262] The secondary chemical shifts were then determined by subtracting random coil chemical shifts[263] from the simulated value, and the resulting values were then compared with the experimental results by Hou et al.[231] Moreover, ${}^3J_{H_NH_\alpha}$ coupling constants were calculated using the Karplus equation:[264]

$$J(\phi) = A\cos^2(\phi) + B\cos(\phi) + C \tag{4.2}$$

where A, B and C are three empiric parameters estimated by Vuister and Bax: $A = 6.51$, $B = -1.76$ and $C = 1.60$.[265] The simulated values were compared to the experimental values obtained by Rosenman et al. with a J-resolved SOFAST HMQC.[233] Finally, residual dipolar couplings (RDCs) were calculated for the simulations with the PALES program.[266,267] To take into account the dependence of the alignment tensor magnitude with experimental conditions, the results were uniformly scaled for each force field to minimize the RMSD between the simulated and experimental values,. The simulated values were compared to the experimental values obtained by Yan et al.[232] Block averages of 25-ns were used to estimate the errors of the predictions (see Appendix A for a more detailed explanation on the estimation of errors for correlated data).

The structures sampled by the REMD simulations were analysed using Gromacs analysis tools[84] and the MDAnalysis toolkit.[177] Secondary structure was calculated using the DSSP algorithm. To simplify the analysis, β-sheets and β-bridges were plotted together as extended structures, and α-helices, 3_{10} helices and π-helices were plotted together as helices. Two residues are considered to be in contact when the distance between any two atoms of the residues is below 0.5 nm. The algorithm of Daura et al.[178] was used to cluster protein structures with a cutoff of 0.8 nm. Peptide figures were generated with VMD[176] and plots were created with Python with the help of Matplotlib.[268]

4.1.2. Simulations of protein aggregation

To understand the effect of force fields on protein aggregation, I study the aggregation of three different peptides: Aβ_{16-22} (KLVFFAE) and its mutants

Aβ_{16-22} (F19V,F20V) (KLVVVAE) and Aβ_{16-22}(F19L) (KLVLFAE). Experimentally, it is known that Aβ_{16-22}(F19L) aggregates faster than Aβ_{16-22} and Aβ_{16-22}(F19V,F20V) does not aggregate at all.[252] The peptides were studied with capped termini to reproduce experimental conditions. Five different force fields were used: Gromos54a7,[269] OPLS-AA,[118,172] CHARMM22*,[122,210,211,270] AMBER99SB*ILDN[207,209,253] and AMBER03WS.[227,271] Moreover, the simulation with AMBER99SB*ILDN was also repeated using long-range dispersion interactions (LRDI) to test their influence on protein aggregation. For the simulations with AMBER99SB*ILDN and CHARMM22*, the TIP4P-Ew water model[138] was used, the simulation with OPLS was performed with the TIP4P water model,[134] the simulation with AMBER03WS was performed with TIP4P/2005[139] and the simulation with Gromos54a7 was performed with the SPC water model.[135] The peptides were simulated at a concentration of $\approx$ 10 mM. It should be noted that at concentrations below 0.5 mM, Aβ_{16-22} does not aggregate.[272]

First, the monomer of each peptide was studied. Each monomer simulation was started with the peptide in an extended conformation, which was introduced in a cubic box with its boundaries at least 1.2 nm away from the peptide. The peptide was then solvated and NaCl ions were added up to a concentration of 150 mM. The system was then minimized with the steepest descent algorithm with peptide heavy atoms coordinates being constrained. Then, a 0.1 ns NVT and 0.1 ns NPT simulations were performed to equilibrate the system. Considering the importance of the monomer conformation observed in chapter 3, a long 1 µs production run was performed for each system.

Aggregation simulations were then performed for each peptide and force field. Six peptides were randomly introduced in a cubic box of side length 10 nm. The initial conformations were selected randomly from the last 800 ns of the 1 µs monomer simulation. NaCl ions were added up to a concentration of 150 mM. The systems were then minimized using the steepest descent algorithm and equilibrated, with peptides heavy atoms coordinates constrained, with a 1 ns NVT and a 1 ns NPT simulation. Production runs were performed for 300 ns or 500 ns, depending on the force field. Each system was studied five times to account for the stochasticity of protein aggregation.

For the production runs both for monomers and protein aggregation simulations, the following parameters were used. The temperature and pressure were kept at 310 K with the Nosé-Hoover thermostat[157,158] and at 1 bar with the Parrinello-Rahman barostat,[160,161] respectively. The P-LINCS algorithm[142] with an expansion order of 6 was used to constrain the bonds in the peptides,

and the SETTLE algorithm[144] was used to constrain water molecules. Virtual sites[145] were used for protein hydrogens, which allowed for a 4 fs timestep. For the simulations with Gromos54a7, a 0.9 nm cutoff was used for short-range electrostatic interactions and a 1.4 nm cut off was used for van der Waals interactions. For the simulation with CHARMM22*, a 1.2 nm cutoff was used for short-range electrostatic interactions and a switching function starting at 1.0 nm and finishing at 1.2 nm was used for van der Waals interactions. In all other simulations, a 1.0 nm cutoff was used for short-range electrostic interactions and van der Waals interactions. In all cases, the particle mesh Ewald method (PME)[150] was used for the calculation of electrostatic interactions. For the simulations using AMBER99SB*ILDN with LRDI, the PME algorithm was also used for dispersion interactions.[273] The MD simulations with LRDI were performed with Gromacs 5.0.4,[110] while all other simulations were performed with Gromacs 4.6.4.[84]

The simulations were analysed with MDAnalysis.[177] To calculate aggregation rates, peptides were considered to associate when the distance between any two atoms from different peptides is under 0.4 nm, and to disassociate when the distance between all atoms of the two peptides is at least 1 nm. To calculate contact maps, two residues were considered to be in contact whenever the distance between any two atoms from the the residues is under 0.4 nm. A peptide is considered to be in β-strand conformation when the dihedral angles ϕ and ψ are inside the polygon ($-180°$,$180°$), ($-180°$,$126°$), ($-162°$,$126°$), ($-162°$,$108°$), ($-144°$,$108°$), ($-144°$,$90°$), ($-50°$,$90°$) and ($-50°$,$180°$).[78,82] To calculate dissociation constants the unbinding rate k_{off} and binding rate k_{on} were estimated. The unbinding rate is $k_{\text{off}} = 1/t_{\text{off}}$ where t_{off} is the residence time before unbinding, i.e., the total time bound divided by the number of unbinding events. The binding rate is $k_{\text{on}} = 1/(t_{\text{on}}c_0)$, where t_{on} is the average residence time before binding, i.e., the total time unbound divided by the number of binding events, and c_0 is the protein concentration. In this way, the dissociation constant can be calculated as $K_d = k_{\text{off}}/k_{\text{on}}$.[227] The acceptance probability was calculated as the ratio between the difference between the number of binding and unbinding events, and the number of binding events. Kinetic networks were calculated for all aggregation simulations. The nodes in the networks are defined based on the oligomerization state (from 1 to 6) and β-strand content (from 1 to 5, where state 1 has 0 to 20 % β-strand content, state 2 has 20 to 40 %, state 3 has 40 to 60 %, state 4 has 60 to 80 % and state 5 has 80 to 100 % β-strand content). Simulation snapshots every 5 ps were considered for the construction of the kinetic networks. The size of the

nodes corresponds to the population and the size of the edges corresponds to the monomer flux between the nodes.

To understand the thermodynamic forces that drive protein aggregation, the formation of the dimer was studied in greater detail. Umbrella sampling simulations of the aggregation of the dimer were then performed. Umbrella sampling[274,275] is a method to improve sampling of the free energy along a predefined coordinate. In the case of the dimer, a simulation starting from the aggregated dimer would unlikely separate and, hence, it would be impossible to accurately measure the free energy difference between aggregated and unaggregated peptides. To perform an umbrella sampling simulation, one (or many) reaction coordinate, which must represent progress along the reaction pathway, must be chosen. In my simulation, the reaction coordinate is the distance between the two peptides. Furthermore, a number of overlapping windows between two points in the reaction coordinate must also be chosen. In this case, I chose to start from the peptides together up to 5 nm apart, where the peptides are considered as separated. To force each window to sample a particular section of the free energy landscape, a biasing potential is introduced to keep the peptide in a position close to its starting position. The distance between windows must be small enough so that the sampled sections of the the free energy profile overlap. In this way, the potential of mean force (PMF), which is the change in energy as a function of the reaction coordinate, can be calculated. The PMF is similar to the free energy difference, but the biasing potential alters the free energy landscape slightly and the relationship between PMFs and free energy is still under debate.[276,277] All simulations can then be joined using the weighted histogram analysis method (WHAM):[278]

$$P(z) = \frac{\sum_{i=1}^{N_{\text{sim}}} n_i(z)}{\sum_{i=1}^{N_{\text{sim}}} N_i \exp((F_i - U_{\text{bias},i}(z))/k_B T)} \tag{4.3}$$

and

$$F_i = -k_B T \ln \left(\sum_z P(z) \exp(-U_{\text{bias},i}(z)/k_B T) \right) \tag{4.4}$$

where N_{sim} is the number of simulations (i.e., the number of windows), n_i is the counts in the histogram associated with the reaction coordinate z, $U_{\text{bias},i}$ is the biasing potential and F_i is the the energy shift of simulation i, P is the unbiased probability distribution. Both F_i and P are unknown and must be

solved iteratively. The PMF is then calculated as:

$$\mathrm{PMF}(z) = -k_B T \ln(P(z)) \tag{4.5}$$

Umbrella sampling simulations have already been used to study how monomers bind to amyloid fibrils.[279,280]

To perform the umbrella sampling simulations, a dimer was picked from the aggregation simulations at random for each force field and peptide. The dimer was then introduced in a cubic box whose boundaries are at least 1.5 nm away from the peptides. The system was then solvated, minimized and equilibrated with 1 ns NVT and 1 ns NPT simulations. The system was then simulated for 300 ns to allow the dimer to relax to a local minimum. The final structure of this simulation was used for a pulling simulation to separate the dimers into monomers. For this simulation, the dimer was introduced in a box with an extra 5 nm in the z-dimension to account for the pulling. The system was solvated, minimized and equilibrated again. Then, one of the peptides was pulled away from the other during 500 ps with a rate of 0.01 nm ps^{-1} and a force constant of 1000 kJ mol^{-1} nm^{-2}. The distance between the center of mass of the peptides at the end of the simulations was around 5 nm. From this pulling simulation, snapshots were picked as starting configurations for the umbrella sampling simulation. The window spacing was 0.1 nm for $z \leq 2$ nm, and 0.2 nm between 2 and 5 nm. New windows were added if the overlap between windows was found to be insufficient. Production runs were then performed for each window for 50 ns with a force constant of 1000 kJ mol^{-1} nm^{-2}. The Gromacs implementation of WHAM[278,281] was used to calculate the PMF. All other simulation parameters were selected as before. The simulations of Aβ_{16-22} were also performed at 290, 300, 310, 320, 330, 340, 350 and 360 K, to estimate the enthalpic and entropic contributions. To this end, the following equation was fitted:[280,282,283]

$$\mathrm{PMF}(z,T) = \Delta H(z,T_0) - T\Delta S(z,T_0) + \Delta C_p(z,T_0)\left[(T-T_0) - T\log\left(\frac{T}{T_0}\right)\right] \tag{4.6}$$

where T is the temperature, z is the distance between the peptides, T_0 is the temperature for which the contributions are estimated (i.e., 310 K), ΔH is the difference in enthalpy, ΔS is the difference in entropy and ΔC_p is the difference in heat capacity.

4.2. Results and Discussion

4.2.1. Simulation of Aβ_{42}

Comparison to experimental data

In this section, the ensembles sampled by the REMD simulations of Aβ_{42} are compared to experimental NMR data. Because of the fact that Aβ_{42} is an IDPs, it is often hard to know if simulations of this peptide have converged. To check for convergence, the distribution of time spent by each replica at each temperature is plotted in figure 4.1. In all simulations, each replica visits each temperature and the acceptance ratio for exchanges between replicas was around 10 %. Considering that each simulation included 32 replicas, the distribution observed in figure 4.1 is considered as thorough. Then, the root mean square deviation (RMSD) and Pearson correlation coefficient (PCC) as a function of time comparing the calculated and experimental values for the six NMR observables were calculated and plotted in figures 4.2 and 4.3. In the case of the C_α, C_β and H_N chemical shifts, there is a clear decrease in the RMSD and an increase in the PCC during the first 100 ns for all force fields. In the case of J-couplings, the convergence seems to be faster, i.e., around 50 ns. However, there are no clear trends for H_α chemical shifts and RDCs. Hence, the first 100 ns will be considered as equilibration, and the last 100 ns will be used for all further analysis.

The converged ensemble is now used to understand the difference between force fields in reproducing the NMR observables. In figures 4.4 and 4.5, and tables 4.1 and 4.2, the results are summarized. For the C_α chemical shifts, the highest PCC and lowest RMSD is obtained for CHARMM22*. For all other force fields, the results are similar apart from AMBER99SBILDN-NMR, which has a much higher RMSD, because the simulated secondary shift is higher than the experimental results for the central region of Aβ_{42}, which corresponds to a high α-helical content as will be seen in the following section. For the C_β chemical shifts, the highest PCC and lowest RMSD is obtained for AMBER99SB, but the results for AMBER99SB*ILDN, AMBER99SBILDN-NMR and CHARMM22* are comparable. The results for H_α chemical shifts are similar for all force fields. However, most force fields other than CHARMM22* tend to overestimate the secondary shift of the C-terminus. In the case of the H_N chemical shifts, all force fields other than CHARMM22* underestimate the secondary shifts for the entire peptide. This error is particularly clear for OPLS. The better results using CHARMM22* are correlated to a high PCC and low RMSD. This underestimation of H_N chemical shifts has also been ob-

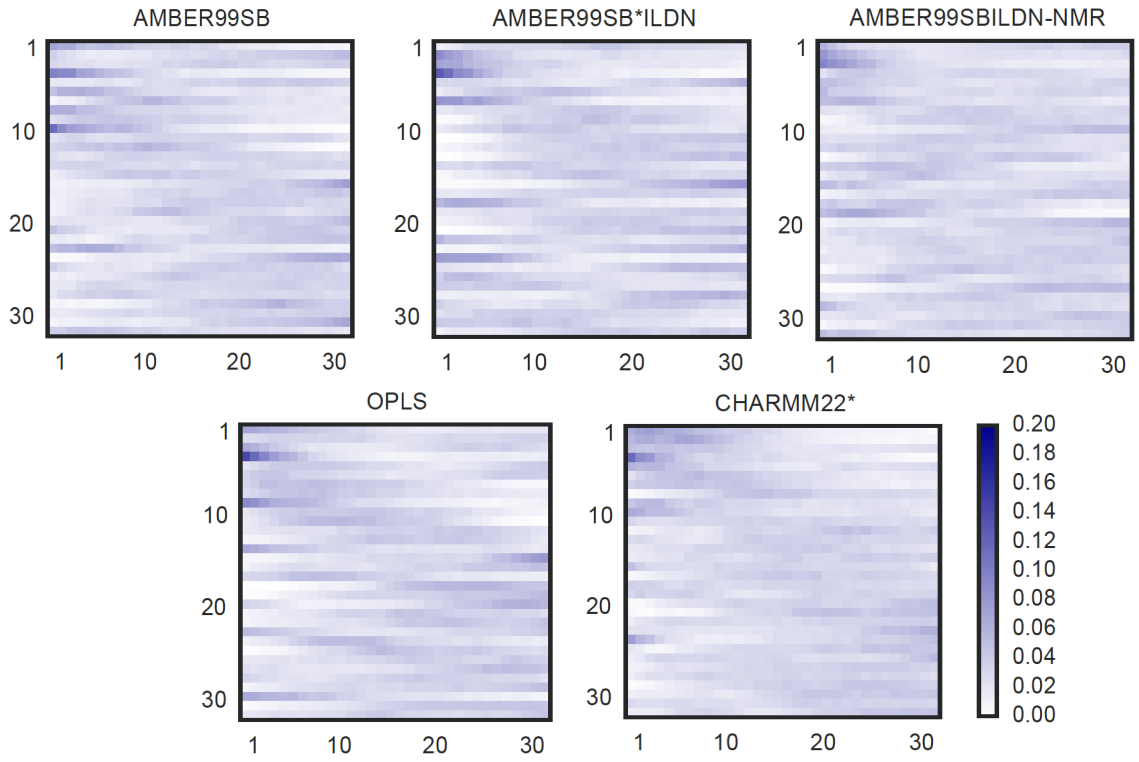

Figure 4.1.: Distribution of time spent by each replica (in the y-axis) at different thermodynamic states (x-axis, where 1 is the lowest temperature and 32 is the highest). The maximum time spent by any replica at one thermodynamic state is 12 % for AMBER99SB, 13 % for AMBER99SB*ILDN, 10 % for AMBER99SBILDN-NMR, 14 % for OPLS and 12 % for CHARMM22* (Reprinted with permission from Carballo-Pacheco and Strodel.[3] Copyright 2016 The Protein Society).

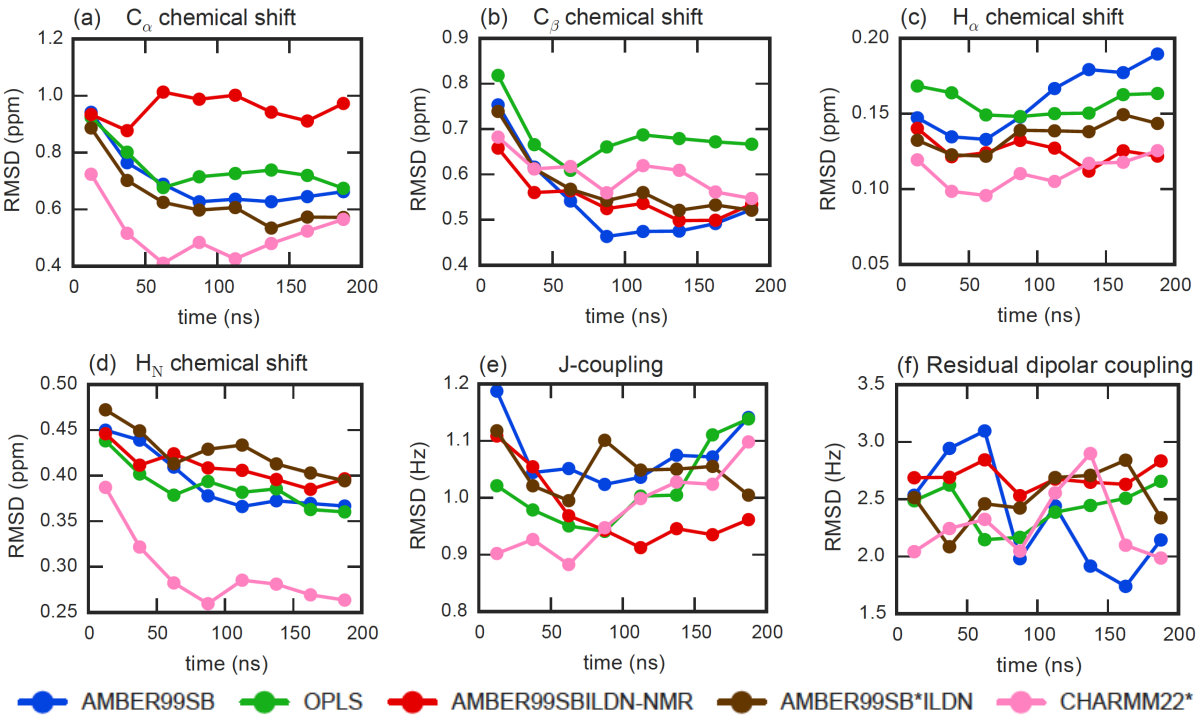

Figure 4.2.: RMSD between simulated and experimental NMR observables as a function of time for (a) C_α chemical shifts, (b) C_β chemical shifts, (c) H_α chemical shifts, (a) H_N chemical shifts, (e) J-couplings and (f) residual dipolar couplings for AMBER99SB (red), OPLS (green), AMBER99SBILDN-NMR (red), AMBER99SB*ILDN (brown) and CHARMM22*(pink) (Reprinted with permission from Carballo-Pacheco and Strodel.[3] Copyright 2016 The Protein Society).

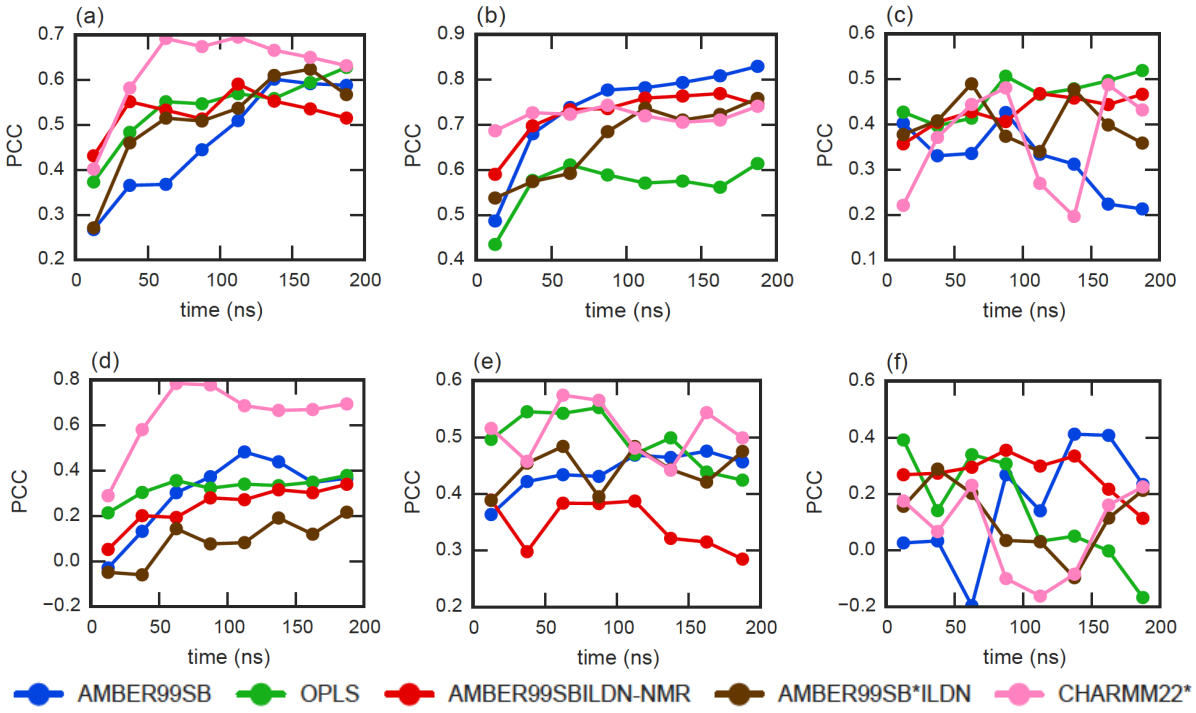

Figure 4.3.: PCC between simulated and experimental NMR observables as a function of time for (a) C_α chemical shifts, (b) C_β chemical shifts, (c) H_α chemical shifts, (a) H_N chemical shifts, (e) J-couplings and (f) residual dipolar couplings for AMBER99SB (red), OPLS (green), AMBER99SBILDN-NMR (red), AMBER99SB*ILDN (brown) and CHARMM22*(pink) (Reprinted with permission from Carballo-Pacheco and Strodel.[3] Copyright 2016 The Protein Society).

Table 4.1.: PCC and RMSD between simulated and experimental chemical shifts for each force field.

Force field	C_α chemical shift		C_β chemical shift		H_α chemical shift		H_N chemical shift	
	PCC	RMSD (ppm)	PCC	RMSD (ppm)	PCC	RMSD (ppm)	PCC	RMSD (ppm)
AMBER99SB	0.62	0.59	0.82	0.46	0.28	0.17	0.43	0.36
AMBER99SB*ILDN	0.60	0.55	0.75	0.50	0.41	0.14	0.16	0.41
AMBER99SBILDN-NMR	0.56	0.94	0.77	0.50	0.47	0.12	0.32	0.39
OPLS	0.59	0.70	0.59	0.67	0.50	0.15	0.36	0.37
CHARMM22*	0.68	0.48	0.73	0.56	0.38	0.11	0.71	0.27

Table 4.2.: PCC and RMSD between simulated and experimental J-couplings and RDCs for each force field.

Force field	J-coupling		RDC	
	PCC	RMSD (Hz)	PCC	RMSD (Hz)
AMBER99SB	0.49	1.06	0.37	1.75
AMBER99SB*ILDN	0.49	1.02	0.10	2.27
AMBER99SBILDN-NMR	0.35	0.91	0.26	2.64
OPLS	0.48	1.05	-0.03	2.40
CHARMM22*	0.51	1.03	0.02	2.19

served in the simulations of Ball et al.,[234] who used AMBER99SB with the TIP4P-Ew water model. The average RMSD between estimated and chemical shifts for folded proteins for SPARTA+ for C_α, C_β, H_α and H_N chemical shifts are 0.94, 1.14, 0.25 and 0.49 ppm, respectively. These are in the order of the RMSD for all simulations of Aβ_{42} for all force fields, which shows that even force fields which are not accurate for Aβ_{42} are in the order of the average error. It should be noted that force fields which represent completely different ensembles such as AMBER99SBILDN-NMR and CHARMM22* have average errors which are both comparable to the average RMSD for folded proteins for SPARTA+, which shows that it is hard to discriminate between ensembles using only chemical shifts. With respect to J-couplings, similar results are obtained for all force fields. It should be noted that the PCC and RMSD for the simulation with OPLS , which are 0.48 and 1.05 Hz respectively, are similar to the 0.48 and 1.25 Hz obtained by Rosenman et al.,[233] who performed an REMD study with 52 replicas and 1000 ns/replica for the same system. This further demonstrates that our simulations are converged. All the force fields I used here perform better that the older force fields studied by Sgourakis et al.[213] (i.e., Gromos, AMBER94 and AMBER96). In the case of RDCs, the best results are obtained for AMBER99SB. All other force fields perform similarly, apart from AMBER99SBILDN-NMR which overestimates the RDCs for the second part of the peptide.

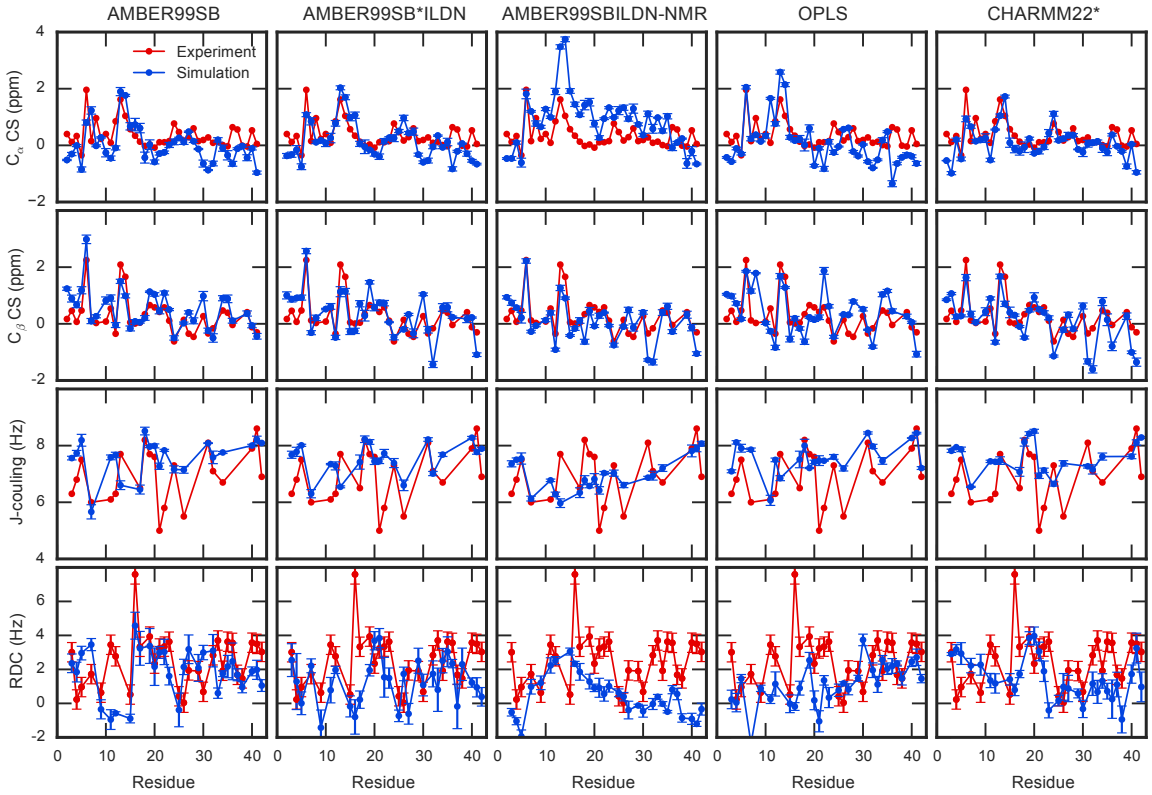

Figure 4.4.: Experimental (red) and simulation (blue) C_α and C_β chemical shifts, J-couplings and residual dipolar couplings per residue for all studied force fields (Reprinted with permission from Carballo-Pacheco and Strodel.[3] Copyright 2016 The Protein Society).

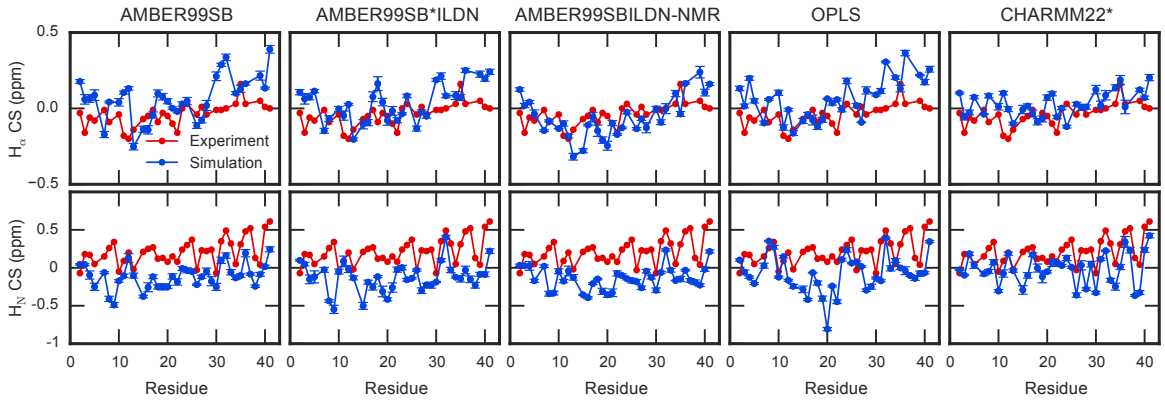

Figure 4.5.: Experimental (red) and simulation (blue) H_α and H_N chemical shifts per residue for all studied force fields (Reprinted with permission from Carballo-Pacheco and Strodel.[3] Copyright 2016 The Protein Society).

Table 4.3.: Radius of gyration of Aβ_{42} for each force field.

Force field	Radius of gyration (nm)
AMBER99SB	1.09 ± 0.12
AMBER99SB*ILDN	1.05 ± 0.08
AMBER99SBILDN-NMR	1.07 ± 0.07
OPLS	1.01 ± 0.08
CHARMM22*	1.11 ± 0.12

Finally, I estimated the radius of gyration for Aβ_{42}. The results can be found in table 4.3. The simulated radii of gyration are similar between all force fields and compare favourably with the experimentally estimated hydrodynamic radius of 0.9 nm determined with single-molecule fluorescence.[284] It should be noted that all snapshots in which the distance between the protein and its image were below 1.0 nm, were deleted. In the case of CHARMM22*, 4.7 % of snapshots were deleted, while in all other force fields less than 1.0 % of snapshots were deleted. Hence, in a larger simulation box, CHARMM22* could have had a larger radius of gyration. There is now ample evidence that force fields produce IDP ensembles that are too compact compared to the experimental ensembles,[140,204,221,227] which has prompted to rescale protein-water interactions.[140,227] It has also been shown that CHARMM22* is the force field which generates more extended conformations before recalibration,[244] which is similar to what I observe. In conclusion, in our simulations of Aβ_{42}, none of the force fields produced ensembles which are too collapsed when compared to experiment.

Our results correlate to other studies in which the ability of force fields to reproduce experimental observables of IDPs was investigated. For example, Rauscher et al.[222] simulated the disordered arginine/serine (RS) peptide and compared the results to NMR data and X-ray scattering data and observed that the best correlation was obtained with CHARMM22*. Somavarapu and Kepp[249] compared different force field for the simulation of Aβ_{40} and found that AMBER99SB-ILDN and CHARMM22* are the best when compared to J-couplings and chemical shifts. They also found that TIP3P produced more compact, helical and structured ensembles when compared to other water models. Hoffmann et al.[223] simulated amylin, the IDP associated with type 2 diabetes, and found that AMBER99SB*ILDN, AMBER03W and CHARMM22* were better than OPLS, Gromos and CHARMM22. There is a general consensus that newer force fields, particularly CHARMM22*, are an improvement over older force fields.

Conformational ensemble

To understand the difference on the ensembles sampled by different force fields and their correlation to experimental observables, a number of structural properties were calculated. In figure 4.6, the secondary structure is plotted for each force field. AMBER99SBILDN-NMR shows a high helical and small β content compared to the other force fields, which correlates with the high secondary C_α chemical shift. AMBER99SB, OPLS and AMBER99SB*ILDN show a helical content of about 20 % in the central section of the peptide (residues 11–17) similar to what was observed by Ball et al.[234] and Lin et al.[240,285] However, CHARMM22* shows less helical content with its peak for residues 33–35. Most force fields show a similar pattern in β content with peaks around residues 4–5, 11–12, 18–20, 30–31 and particularly in the C-terminus. The highest β content is obtained with CHARMM22* with a content of over 60 % for residues 40 and 41. My results are intermediate to the results obtained by other researchers. For example, Lin et al.[240,285] and Ball et al.[234] have observed mostly disordered structures with some helical content and little β-sheet content. On the other hand, Rosenman et al.[233] observed very high β-sheet content and little helicity. I observe both helical and β structures.

To further understand the ensemble sampled by each force field, intramolecular contact maps and the most representative clusters were calculated for each force field and plotted in figures 4.7 and 4.8, respectively. For AMBER99SBILDN-NMR, contacts between residues close in sequence are observed in the contact map as expected for structures with high α-helical content. This can also be observed in the most probable cluster which represents 50.4 % of the snapshots and is mostly α-helical. For all other force fields, the structures are mostly disordered with some local structure. The ensembles produced by CHARMM22*, which is the force field that reproduces the experimental NMR observables the best, OPLS and AMBER99SB show a high propensity for a β-hairpin of the C-terminus, which has been suggested to be the initial seed for the aggregation process for the β-sheet rich amyloids.[286–288] Some force fields such as AMBER99SBILDN-NMR or OPLS have clusters that represent a large number of snapshots, while others such as AMBER99SB or AMBER99SB*ILDN have clusters that represent smaller number of snaphots, which shows that they sample a more heterogeneous conformational ensemble. Rosenman et al.[244] have observed for $A\beta_{40}$ that different force fields produce different secondary structures but similar contact maps. In my simulations, there are differences between force fields in both contact maps and secondary structure.

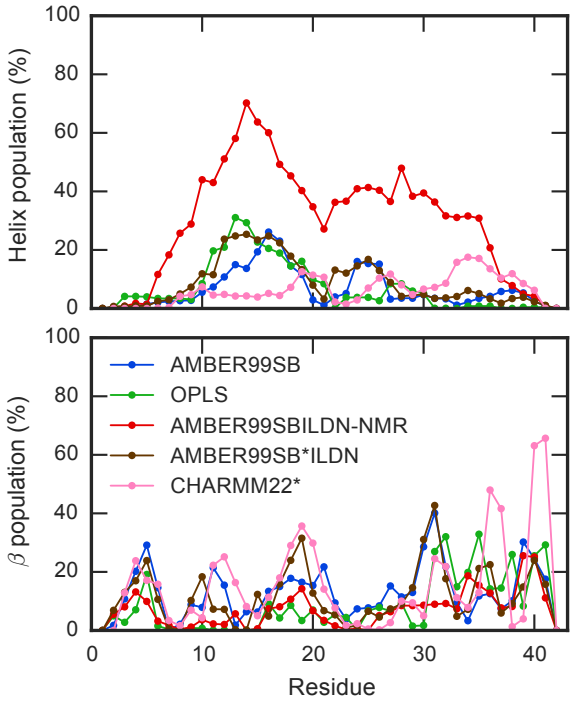

Figure 4.6.: Helix population (top) and β population (bottom) for AMBER99SB (blue), OPLS (green), AMBER99SBILDN-NMR (red), AMBER99SB*ILDN (brown) and CHARMM22* (pink) (Reprinted with permission from Carballo-Pacheco and Strodel.[3] Copyright 2016 The Protein Society).

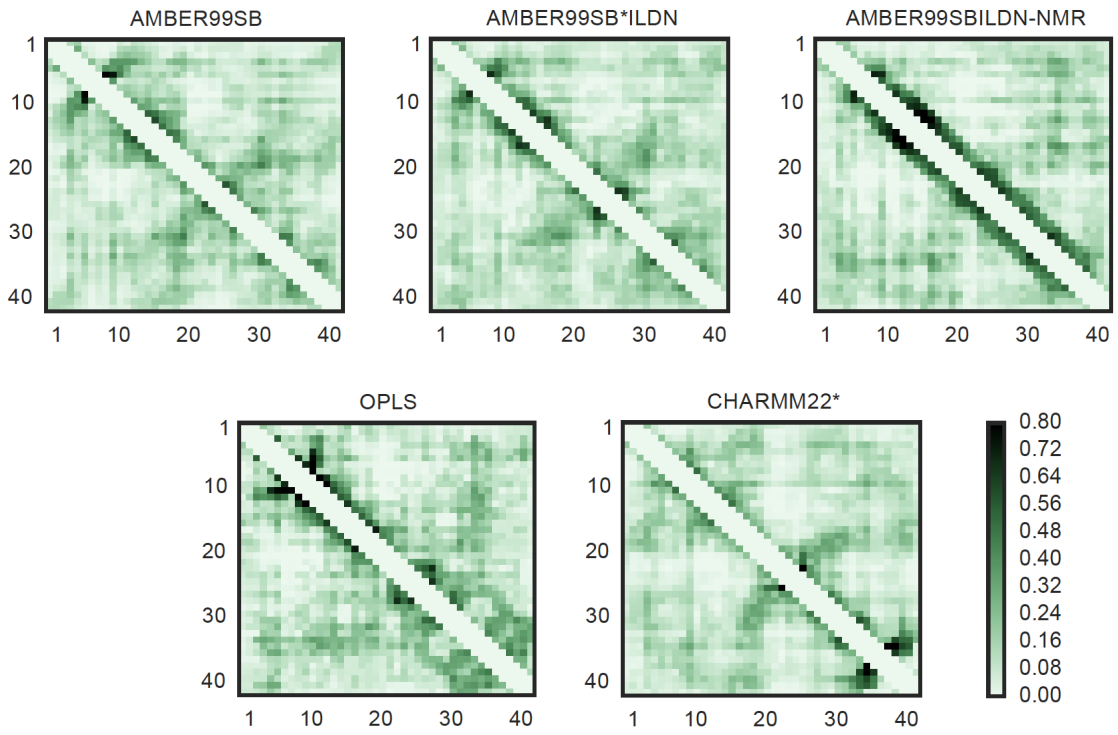

Figure 4.7.: Contact maps for the monomer of Aβ using the AMBER99SB, AMBER99SB*ILDN, AMBER99SBILDN-NMR, OPLS and CHARMM22* force fields (Reprinted with permission from Carballo-Pacheco and Strodel.[3] Copyright 2016 The Protein Society).

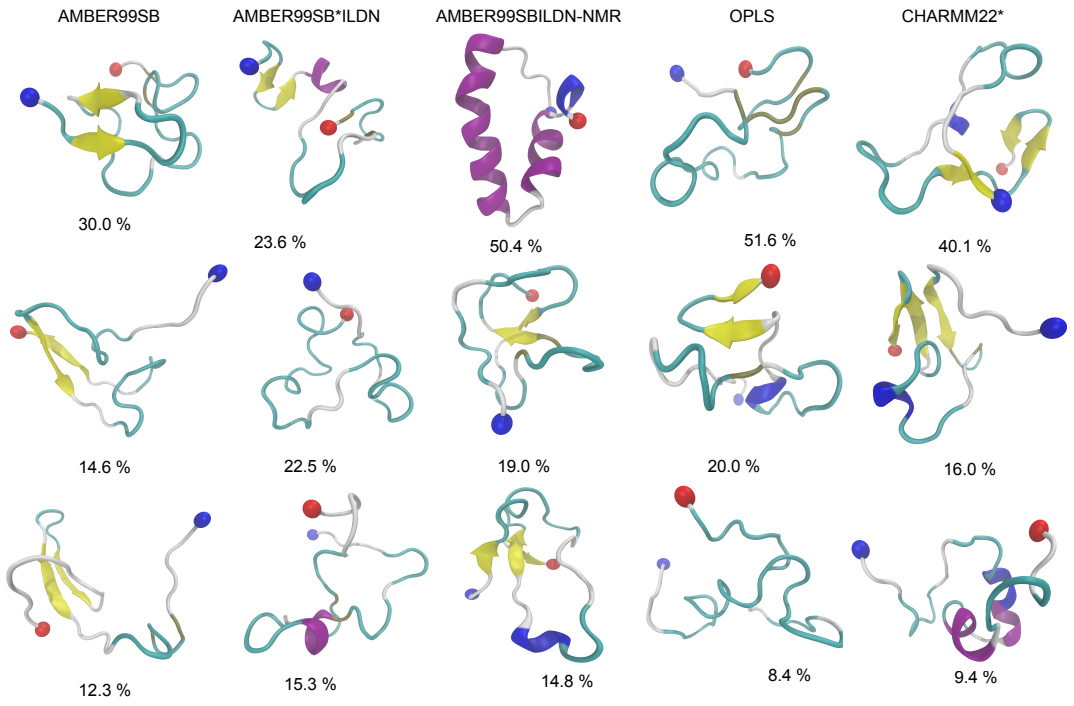

Figure 4.8.: Three most representative clusters for the AMBER99SB, AMBER99SB*ILDN, AMBER99SBILDN-NMR, OPLS and CHARMM22* force fields and the percentage of snapshots each cluster represents (Reprinted with permission from Carballo-Pacheco and Strodel.[3] Copyright 2016 The Protein Society).

4.2.2. Simulations of protein aggregation

Aggregation simulations

In this section, the ability of protein force fields to model protein aggregation is assessed. In particular, the aggregation of Aβ_{16-22}, Aβ_{16-22}(F19V,F20V) and Aβ_{16-22}(F19L) is studied. Experimentally, it is known that Aβ_{16-22}(F19L) aggregates faster than Aβ_{16-22} and that Aβ_{16-22}(F19V,F20V) does not aggregate.[289] In figure 4.9, the oligomerization state as a function of time is plotted for all the tested force fields. The three peptides aggregate to the maximum oligomerization state of six in the simulations with Gromos54a7. Specifically, the five replicas of each system have reached the maximum oligomerization state within 200 ns and no disassembly is observed after that. For the simulations with OPLS, all replicas also reach the maximum oligomerization state but it takes almost the entire 300 ns of simulation time. The results for CHARMM22* and AMBER99SB*ILDN are comparable: peptides aggregate fast, however oligomers break continuously. Moreover, it appears that Aβ_{16-22}(F19V,F20V) aggregates slower than the two other peptides. No major differences are observed in the simulations using AMBER99SB*ILDN with and without LRDI. Finally, the behaviour with AMBER03WS is strikingly different from the other force fields. Little aggregation is observed and the average oligomerization state is never larger than three. It should be noted that this force field was developed to correct for the overstabilization of protein-protein interactions by most force fields and has stronger water-protein interactions.[227]

To quantify the difference between force fields, dissociation constants K_D and their corresponding free energy differences were estimated as explained in section 4.1.2. (see table 4.4). The dissociation constant represents the equilibrium between the addition of a monomer to an oligomer and the dissociation of a monomer from an oligomer. The higher the value, the less aggregation prone the system is, as a higher concentration of monomers is needed for one to stably bind to an oligomer. Senguen et al.[289] calculated experimentally critical concentrations, which are similar to dissociation constants, for the same peptides. The values for Aβ_{16-22} and Aβ_{16-22}(F19L) are 33 ± 3 and 12 ± 1 µM, respectively. Because Aβ_{16-22}(F19V,F20V) does not aggregate, it is impossible to calculate the critical concentration for this peptide. The free energy difference associated with these critical concentrations are -26.8 ± 0.4 and -29.3 ± 0.4 kJ mol^{-1} for Aβ_{16-22} and Aβ_{16-22}(F19L), respectively. The critical concentration is the monomer concentration at which there is an equilibrium between adding a monomer to the fibril and a monomer dissociating

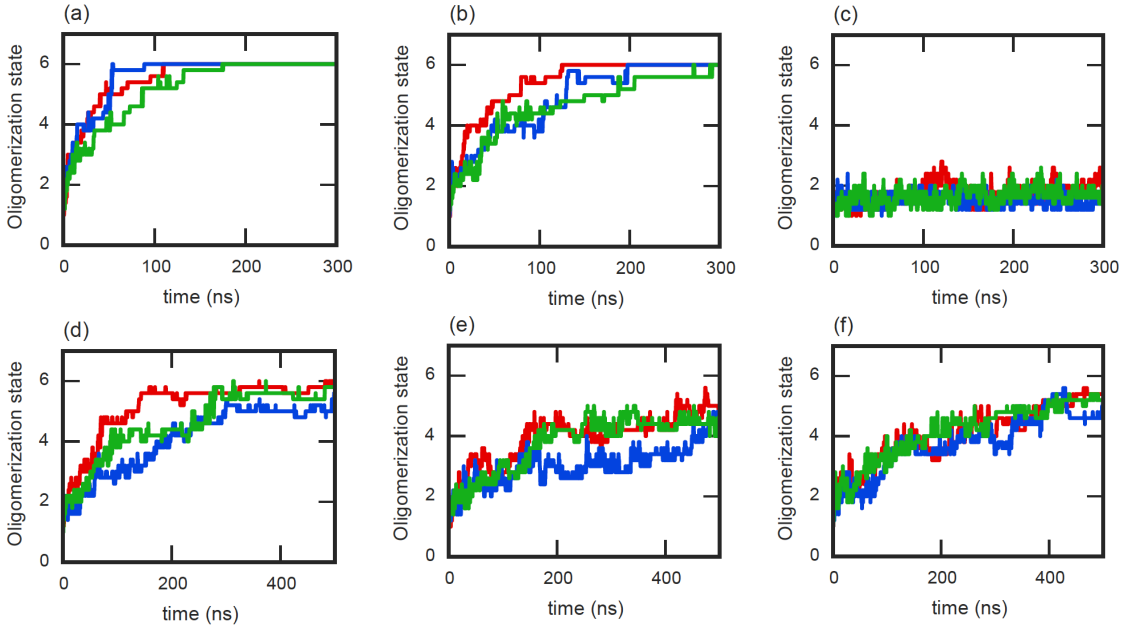

Figure 4.9.: Oligomerization state as a function of time for Aβ_{16-22} (red), Aβ_{16-22}(F19V,F20V) (blue) and Aβ_{16-22}(F19L) (green) for (a) Gromos54a7, (b) OPLS-AA, (c) AMBER03WS, (d) CHARMM22*, (e) AMBER99SB*ILDN and (f) AMBER99SB*ILDN with LRDI.

from the fibril. Even though it is similar to the dissociation constant, experimentally only the monomer addition to the fibril can be measured, whereas in my simulations, the monomer addition to oligomers is studied. Since oligomer formation is considered as the rate limiting step in amyloid formation,[290] the simulated dissociation constants should be much higher than the experimental critical concentrations. Indeed, for all force fields, the simulated dissociation constants are higher than the experimental value. However, it should be noted that for the simulations with OPLS, and particularly Gromos54a7, little disassembly of monomers is observed. This means that if the simulations were longer, the dissociation constant would considerably decrease, which would represent that the formation of oligomers would be more favourable than the monomer addition to fibrils. This suggests that these force fields probably overstabilize protein-protein interactions.

To further compare the results of the simulations with experiments, I calculated the collision acceptance probability. Because this calculation does not depend on reaching equilibrium, it should converge faster than the dissociation constant. In table 4.5, the acceptance probabilities for all force fields are summarized. The estimation of the collision acceptance probability is hard to obtain experimentally, and, as far as I know, it has not been estimated for

Table 4.4.: Equilibrium properties for the aggregation simulations.

Force field	Aβ_{16-22}		Aβ_{16-22}(F19V,F20V)		Aβ_{16-22}(F19L)	
	K_D (mM)	ΔG (kJ mol^{-1})	K_D (mM)	ΔG (kJ mol^{-1})	K_D (mM)	ΔG (kJ mol^{-1})
Gromos54a7	0.19 ± 0.09	-20.9 ± 0.7	0.25 ± 0.04	-21.8 ± 0.4	0.65 ± 0.18	-19.0 ± 0.9
OPLS-AA	0.6 ± 0.2	-19.8 ± 1.4	0.9 ± 0.2	-18.3 ± 1.2	1.1 ± 0.4	-16.7 ± 0.8
AMBER03WS	36 ± 8	-9.8 ± 0.4	61 ± 15	-8.6 ± 0.9	40 ± 8	-8.0 ± 0.7
CHARMM22*	1.3 ± 0.3	-18.0 ± 0.8	4.2 ± 1.4	-13.6 ± 1.0	1.4 ± 0.3	-17.8 ± 0.4
AMBER99SB*ILDN	3.8 ± 1.3	-14.6 ± 1.2	7.1 ± 2.5	-13.5 ± 1.5	3.7 ± 1.0	-15.0 ± 0.8
AMBER99SB*ILDN (LRDI)	3.4 ± 0.9	-15.3 ± 0.5	3.8 ± 0.4	-14.9 ± 0.2	2.5 ± 0.3	-15.6 ± 0.6

Aβ_{16-22}. However, it has been estimated for other proteins. In particular, Knowles et al.[291] measured an acceptance probability of 1 in 10,000 for insulin aggregation using a quartz crystal oscillator. Furthermore, Xue and Radford[292] also measured an acceptance rate of 1 in 10,000 for β_2 microglobulin using tapping-mode atomic force microscopy. However, these estimations were performed for monomer addition to fibrils and even lower acceptance probabilities are expected for oligomer formation. I must note that different results are expected for different peptides, so the picture could be completely different for Aβ_{16-22}. However, it is interesting that the simulated acceptance probabilities are orders of magnitude higher than the experimental values, which could be used as order of magnitude estimates. For example, for some of the simulations with Gromos54a7, no rejection is ever observed. The only force field which comes close to the experimental values is AMBER03WS, with an acceptance rate of 0.013 ± 0.005 for Aβ_{16-22}.

Table 4.5.: Collision acceptance probability in the aggregation simulations.

Force field	Aβ_{16-22}	Aβ_{16-22}(F19V,F20V)	Aβ_{16-22}(F19L)
Gromos54a7	0.65 ± 0.26	0.47 ± 0.19	0.31 ± 0.12
OPLS-AA	0.27 ± 0.11	0.42 ± 0.17	0.20 ± 0.08
AMBER03WS	0.013 ± 0.005	0.005 ± 0.002	0.006 ± 0.002
CHARMM22*	0.19 ± 0.07	0.07 ± 0.03	0.11 ± 0.04
AMBER99SB*ILDN	0.07 ± 0.03	0.04 ± 0.02	0.06 ± 0.02
AMBER99SB*ILDN (LRDI)	0.09 ± 0.03	0.07 ± 0.03	0.07 ± 0.03

Considering the strong differences in the aggregation kinetics between force fields, intermolecular contacts were calculated to understand which peptide-peptide interactions drive aggregation. In figures 4.10, 4.11 and 4.12, the intermolecular contact maps between peptides for each force field for Aβ_{16-22}, Aβ_{16-22}(F19V,F20V) and Aβ_{16-22}(F19L), respectively, are plotted. In the contact map of Aβ_{16-22}, the phenylalanines in the center of the peptides interact strongly between peptides, which is expected as phenylalanines have always been considered essential for the aggregation of many amyloidogenic

peptides,[293–296] and mutations of Phe^{19} and Phe^{20} strongly change the aggregation propensity of Aβ_{16-22}.[252,289] Moreover, the positively charged N-terminal Lys forms a salt bridge with the negatively charged C-terminal Glu. The interactions predicted by all force fields are similar, which demonstrates that force fields with different aggregation kinetics may have the same predominant contacts. On the contrary, the main interactions between peptides in the simulations of Aβ_{16-22}(F19V,F20V) differ between force fields. In most force fields, the interaction between the N-terminal Lys and the C-terminal Glu is still present, but there is no interaction between the central section of the peptides. In the case of CHARMM22* and AMBER99SB*ILDN though, the most relevant interaction is between residues Leu^{17}-Val^{19} of the different peptides. Similar results are obtained for Gromos54a7 but the interactions are weaker. In the case of OPLS, there are unspecific interactions between all residues. The simulation with AMBER03WS is particularly interesting because the contact map shows a perfect anti-parallel β-sheet, which means that whenever the peptides are in contact with each other, they form an ordered structured. I can hypothesise that the aggregation at low concentration will then not proceed, considering aggregates are only stable when a very specific interaction is formed and unstable otherwise, which is different for the results for other peptides or force fields where unspecific interactions are found. Finally, in the case of Aβ_{16-22}(F19L) the contacts are similar to the ones observed for Aβ_{16-22}.

Considering the importance of the monomer dynamics in the simulations of protein aggregation,[1] monomer intramolecular contact maps before aggregation were also calculated for all peptides and force fields. The contact maps are plotted in figures 4.13, 4.14 and 4.15 for Aβ_{16-22}, Aβ_{16-22}(F19V,F20V) and Aβ_{16-22}(F19L), respectively. All force fields sample similar, mostly extended, conformations for Aβ_{16-22} and Aβ_{16-22}(F19L). For Aβ_{16-22}(F19V,F20V), all force fields except for OPLS, also preferentially sample an extended conformation. However, OPLS predicts a bended conformation where the C- and N-termini are in close contact with each other. Considering the similarity of the conformations predicted by the different force fields, in this case the difference in aggregation propensity cannot be explained by the conformations of the monomer and must therefore arise from the interactions between the peptides. Nonetheless, these results show that, even at the monomer stage, there are some important differences between force fields.

To understand the intermediate conformations in the aggregation process, kinetic transition networks (KTNs) were calculated and plotted in figures 4.16,

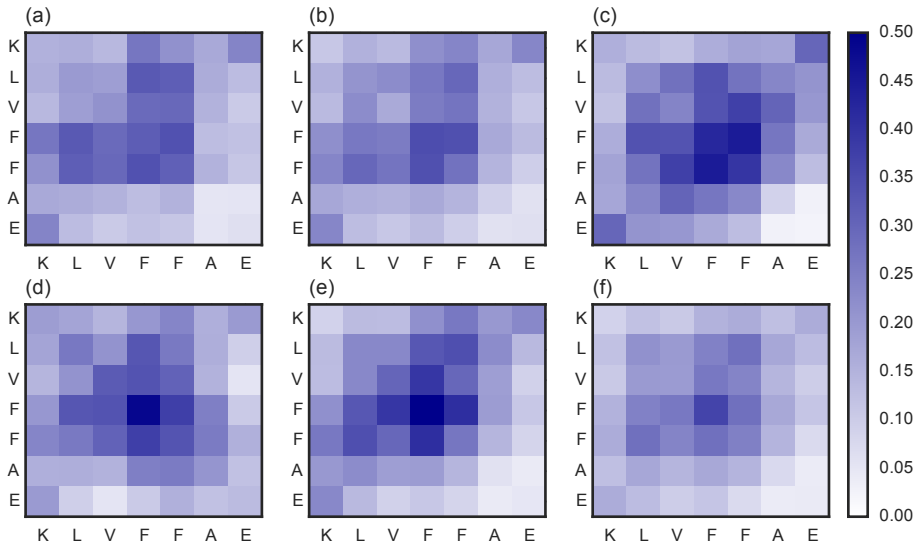

Figure 4.10.: Intermolecular contact maps for Aβ_{16-22} oligomers in the aggregation simulations using the following force fields: (a) Gromos54a7, (b) OPLS-AA, (c) AMBER03WS, (d) CHARMM22*, (e) AMBER99SB*ILDN and (f) AMBER99SB*ILDN with LRDI.

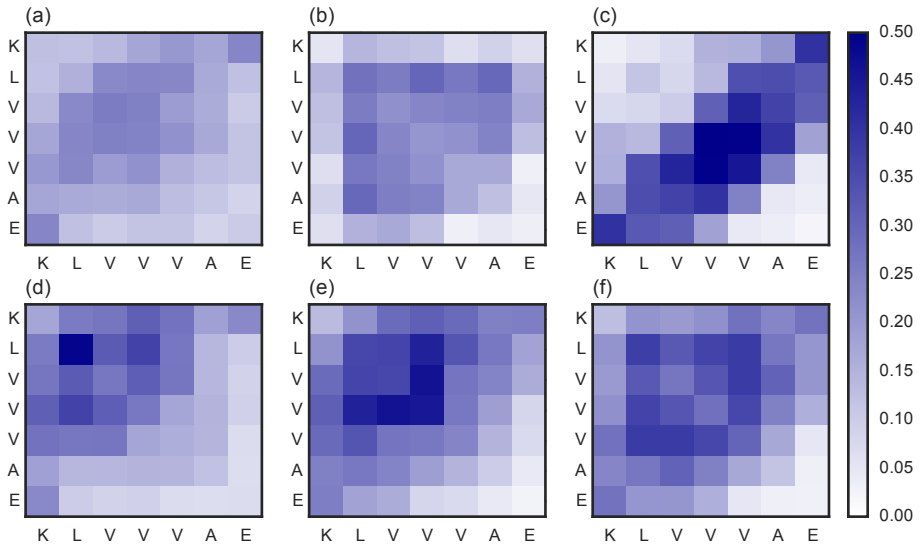

Figure 4.11.: Intermolecular contact maps for Aβ_{16-22}(F19V,F20V) oligomers in the aggregation simulations using the following force fields: (a) Gromos54a7, (b) OPLS-AA, (c) AMBER03WS, (d) CHARMM22*, (e) AMBER99SB*ILDN and (f) AMBER99SB*ILDN with LRDI.

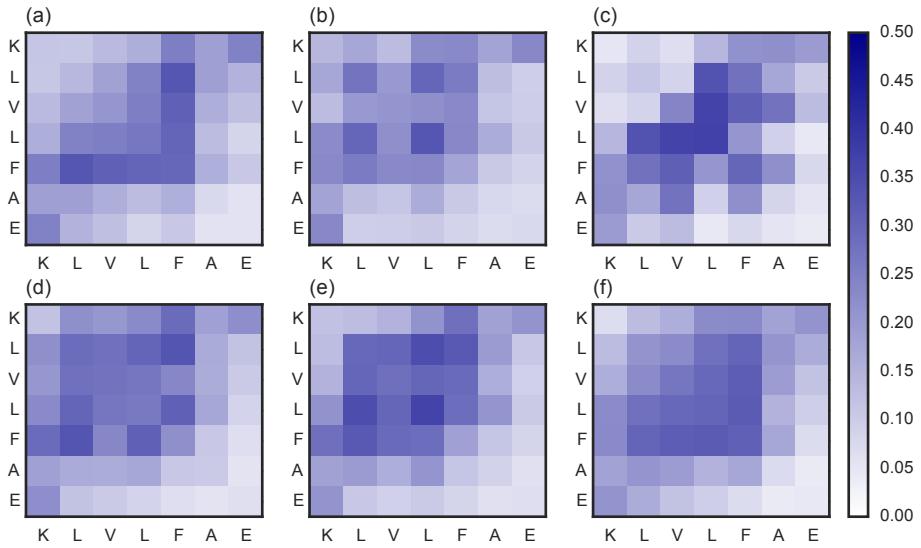

Figure 4.12.: Intermolecular contact maps for Aβ_{16-22}(F19L) oligomers in the aggregation simulations using the following force fields: (a) Gromos54a7, (b) OPLS-AA, (c) AMBER03WS, (d) CHARMM22*, (e) AMBER99SB*ILDN and (f) AMBER99SB*ILDN with LRDI.

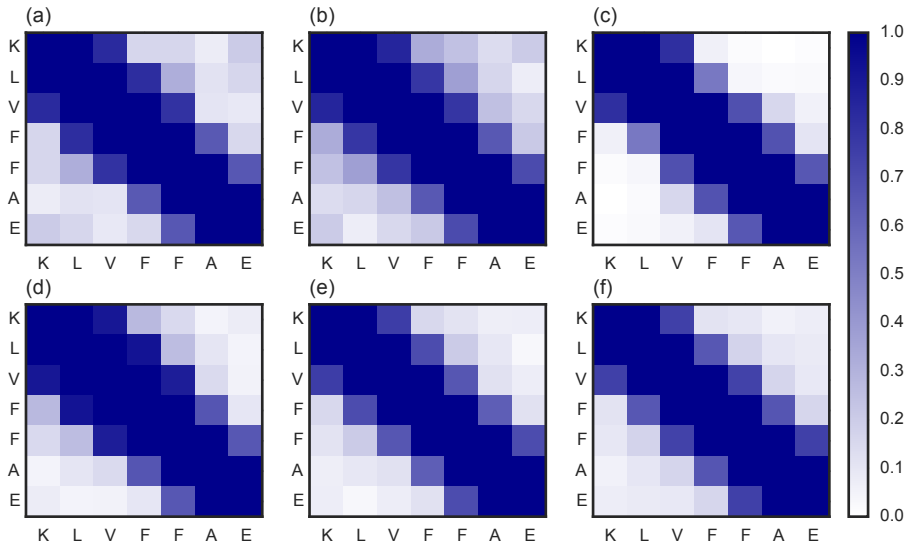

Figure 4.13.: Intramolecular contact maps for Aβ_{16-22} monomers in the simulations using the following force fields: (a) Gromos54a7, (b) OPLS-AA, (c) AMBER03WS, (d) CHARMM22*, (e) AMBER99SB*ILDN and (f) AMBER99SB*ILDN with LRDI.

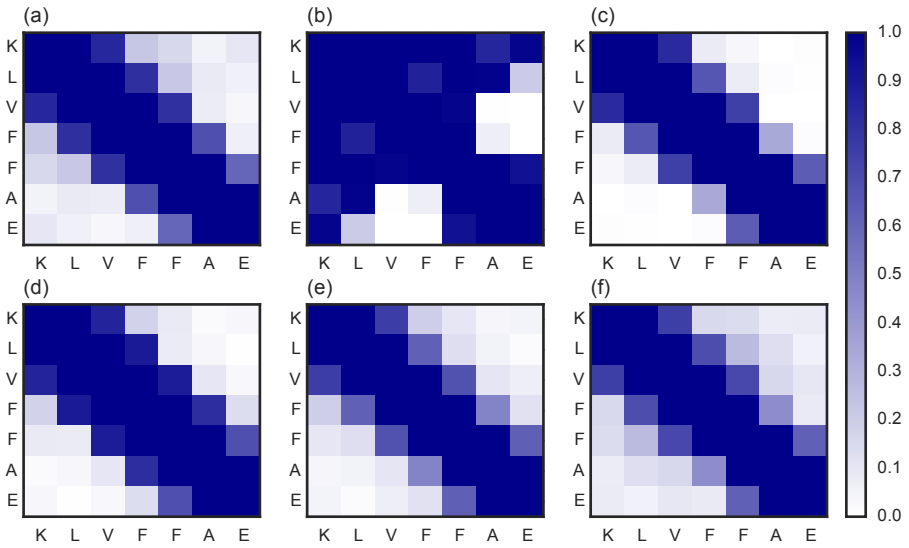

Figure 4.14.: Intramolecular contact maps for Aβ_{16-22}(F19V,F20V) monomers in the simulations using the following force fields: (a) Gromos54a7, (b) OPLS-AA, (c) AMBER03WS, (d) CHARMM22*, (e) AMBER99SB*ILDN and (f) AMBER99SB*ILDN with LRDI.

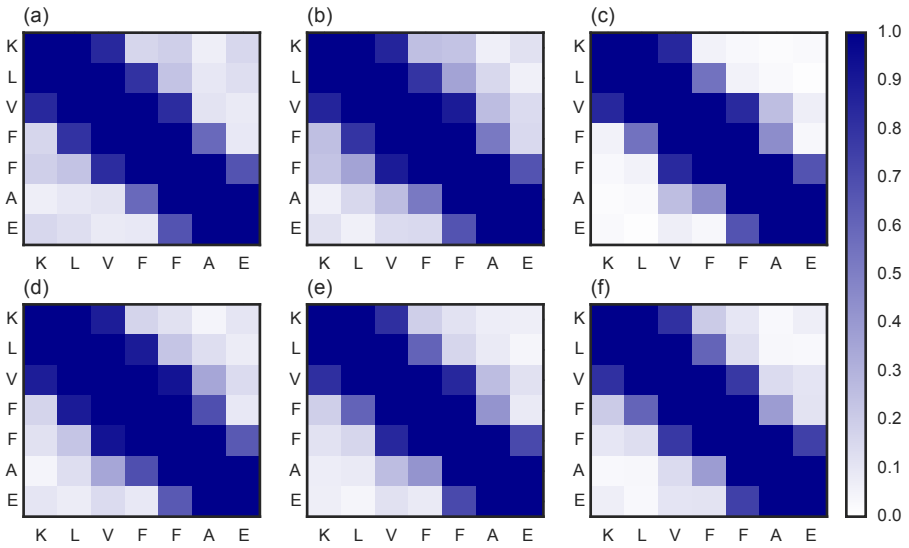

Figure 4.15.: Intramolecular contact maps for Aβ_{16-22}(F19L) monomers in the simulations using the following force fields: (a) Gromos54a7, (b) OPLS-AA, (c) AMBER03WS, (d) CHARMM22*, (e) AMBER99SB*ILDN and (f) AMBER99SB*ILDN with LRDI.

4.17 and 4.18 for Aβ_{16-22}, Aβ_{16-22}(F19V,F20V) and Aβ_{16-22}(F19L), respectively. KTNs are a method to visualize the most relevant states and connections of a free energy landscape[297–302] and have recently been applied to study protein aggregation.[79,82] Here, I used two properties to define the states: the oligomerization state in the vertical direction and the β-strand content in the horizontal direction. It should be noted that the definition of the β-strand is based on the dihedral angles, which means that even if the β-strand content of an oligomer is high, this does not mean that a β-sheet has been formed between the peptides. This definition was used because other standard definitions such as DSSP are based on the formation of hydrogen bonds between backbone atoms which means that, for these short peptides, the β-strand content of the monomer would always be close to zero. The KTNs of Aβ_{16-22} are similar for all force fields, except for AMBER03WS. In general, the most populated nodes are the high oligomerization states with high β-strand content. The nodes with low β-strand content and high oligomerization state are rarely populated, which shows that β-strand content rapidly increases during the aggregation process. Monomers can be found with both high and low β-strand content, though the ones with high β-strand content are more common, particularly for CHARMM22* and AMBER99SB*ILDN. The simulation with OPLS shows lower β-strand content than those with other force fields. In the case of AMBER03WS, the free energy landscape is tilted toward monomers and low-mass oligomers with high β-strand content. For the simulations of Aβ_{16-22}(F19V,F20V) and Aβ_{16-22}(F19L), similar results are observed. The biggest difference is that for OPLS and Aβ_{16-22}(F19V,F20V), the system has even lower β-strand content, which is expected from the monomer contact map described before. Also, the simulation of Aβ_{16-22}(F19V,F20V) with AMBER03WS tends to favour smaller oligomers than the other peptides with the same force field. All force fields predict high β-strand content, which has been observed experimentally for Aβ_{16-22}, in which peptides form β-sheets with interchangeable β-strands.[251]

It should be noted that the influence of long-range dispersion interactions (LRDI) is very limited on the results of the simulations with AMBER99SB*ILDN. Recently, it was found that systems that are inhomogeneous and anisotropic such as surfaces, LRDI can play an essential role.[152,155,156] Of particular interest for biomolecular systems is the result that the surface tension can change drastically in simulations of lipid membranes when using LRDI.[273] However, in the system studied here, the influence of LRDI is negligible.

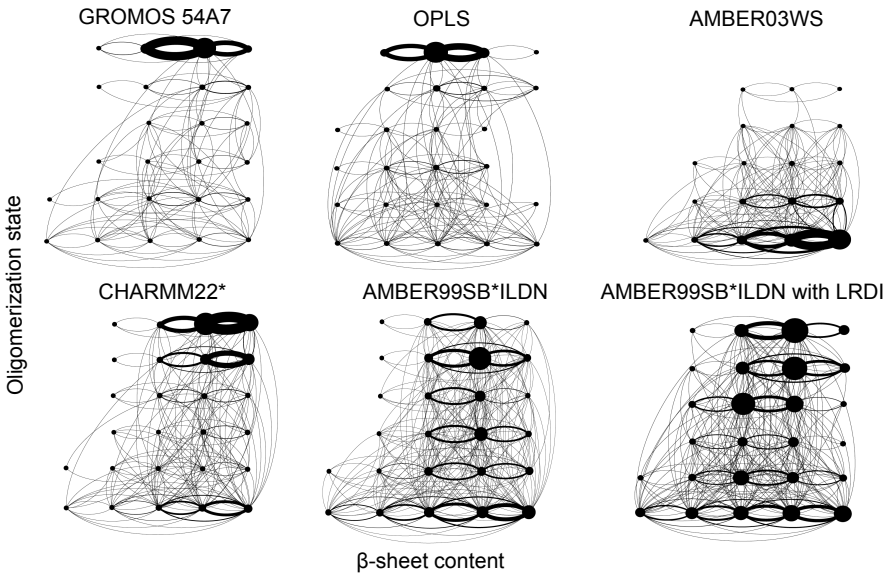

Figure 4.16.: Kinetic transition networks for Aβ_{16-22} aggregation using the oligomerization state (from monomer to hexamer, in the vertical axis) and the β-strand content (state 1 being from 0 to 20 % β-strand content up to state 5 with 80 to 100 % β-strand content, in the horizontal axis) in the simulations using the force fields: Gromos54a7, OPLS-AA, AMBER03WS, CHARMM22*, AMBER99SB*ILDN and AMBER99SB*ILDN with LRDI.

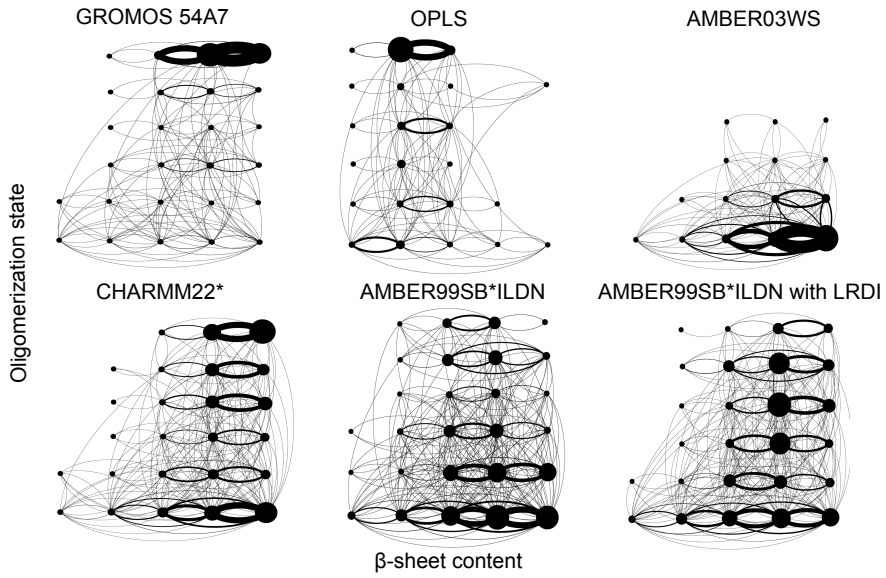

Figure 4.17.: Kinetic transition networks for Aβ_{16-22}(F19V,F20V) aggregation using the oligomerization state (from monomer to hexamer, in the vertical axis) and the β-strand content (state 1 being from 0 to 20 % β-strand content up to state 5 with 80 to 100 % β-strand content, in the horizontal axis) in the simulations using the force fields: Gromos54a7, OPLS-AA, AMBER03WS, CHARMM22*, AMBER99SB*ILDN and AMBER99SB*ILDN with LRDI.

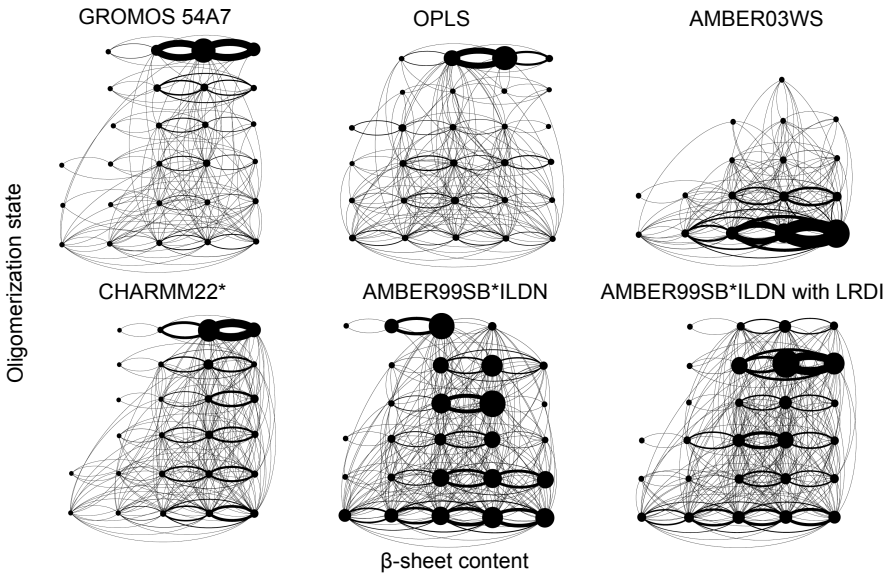

Figure 4.18.: Kinetic transition networks for Aβ_{16-22}(F19L) aggregation using the oligomerization state (from monomer to hexamer, in the vertical axis) and the β-strand content (state 1 being from 0 to 20 % β-strand content up to state 5 with 80 to 100 % β-strand content, in the horizontal axis) in the simulations using the force fields: Gromos54a7, OPLS-AA, AMBER03WS, CHARMM22*, AMBER99SB*ILDN and AMBER99SB*ILDN with LRDI.

Thermodynamics of dimer formation

To better understand the thermodynamic forces that drive protein aggregation, the simplest oligomer, i.e., the dimer, was studied in greater detail. In particular, one snapshot per force field and peptide was selected and separated into monomers for the generation of starting configurations to initiate umbrella sampling simulations and calculate PMFs between the aggregated and unaggregated peptides. The difference between the minimum and maximum in the PMFs for all peptides and force fields are summarized in table 4.6. The differences in free energy for Gromos54a7 are the highest for all the systems, which is expected from the fastest aggregation kinetics. The difference in free energy for all other force fields are smaller. However, the difference in free energy for AMBER03WS is not smaller than for the other force fields, which is surprising because of the slower aggregation kinetics obtained with this force field. It should be noted that the umbrella sampling simulations were started from dimers from the aggregation simulations, that are different between force fields (see figure 4.19), which may influence the PMF. The simulations could have been performed starting from the same initial structure, however, the reason for these simulations was to measure the driving forces behind the aggregation in our simulations and not to see the difference in stability predicted by different force fields.

Table 4.6.: ΔG based on the PMFs for dimer formation.

Force field	Aβ_{16-22}	Aβ_{16-22}(F19V,F20V)	Aβ_{16-22}(F19L)
	ΔG (kJ mol^{-1})	ΔG (kJ mol^{-1})	ΔG (kJ mol^{-1})
Gromos54a7	−30	−27	−37
OPLS-AA	−36	−13	−12
AMBER03WS	−15	−11	−11
CHARMM22*	−11	−7	−18
AMBER99SB*ILDN	−10	−11	−4
AMBER99SB*ILDN (LRDI)	−13	−9	−19

An important question that needs to be addressed regarding the aggregation of amyloidogenic peptides is to understand if the process is driven by entropy or enthalpy. Rao Jampani et al.[280] performed multiple umbrella sampling simulations at different temperatures to calculate the enthalpic and entropic contributions for monomer addition to a short fibril and a similar protocol is followed here to calculate the contributions for the dimer formation of Aβ_{16-22}. In figure 4.20, both contributions are plotted as a function of the distance between the two peptides. In most force fields, the aggregation is enthalpy driven and an enthalpy-entropy compensation is observed,[303–312] in which a decrease

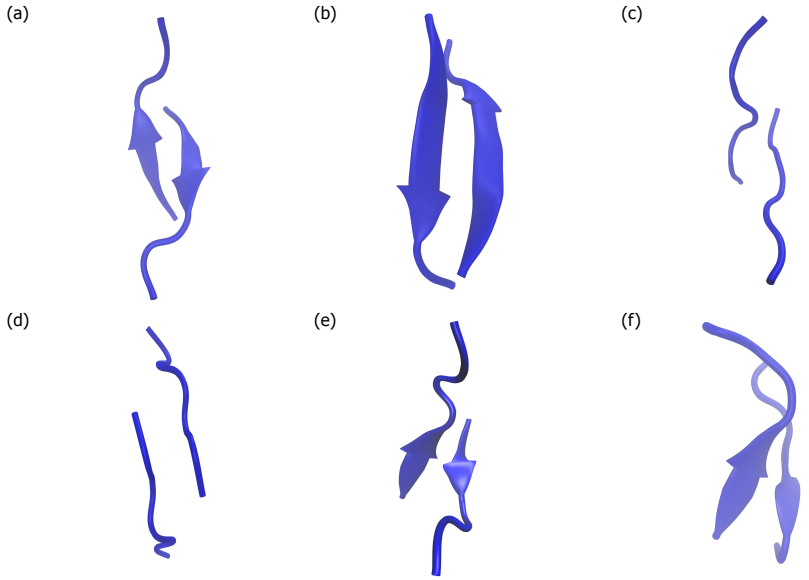

Figure 4.19.: Initial conformations for Aβ_{16-22} the dimer simulations using the following force fields: (a) Gromos54a7, (b) OPLS-AA, (c) AMBER03WS, (d) CHARMM22*, (e) AMBER99SB*ILDN and (f) AMBER99SB*ILDN with LRDI.

of enthalpy is partially compensated by a decrease in entropy. It has been suggested that this effect is a consequence of the formation of hydrogen bonds,[303] which are important in protein-protein interactions. The decrease in entropy is expected when free peptides lock into stable structures, and the decrease of enthalpy probably arouses from the strong contacts between peptides. Similar results are observed by Rao Jampani et al.[280] Interestingly, I observe a different mechanism for the stabilization of the dimer in the simulation with CHARMM22*. In this case, both the entropy and the enthalpy increase. This difference probably arises from the different initial conformation of the simulations: the initial conformation of the simulation with CHARMM22* did not have such an ordered structure as the ones with most of the other force fields (see figure 4.19). In this case, the increase in entropy probably arouses from the solvent that surrounds the peptide, and the increase in enthalpy from the unfavourable contacts between peptides. This shows that the aggregation of monomers to oligomers may follow a similar scheme to the dock-lock mechanism.[60] There are dock interactions which are non-specific and increase entropy and enthalpy but are still stable and lock interactions which lower both enthalpy and entropy by rearranging hydrogen bonds leading to ordered structures. Both of these mechanisms are stabilizing protein interactions in

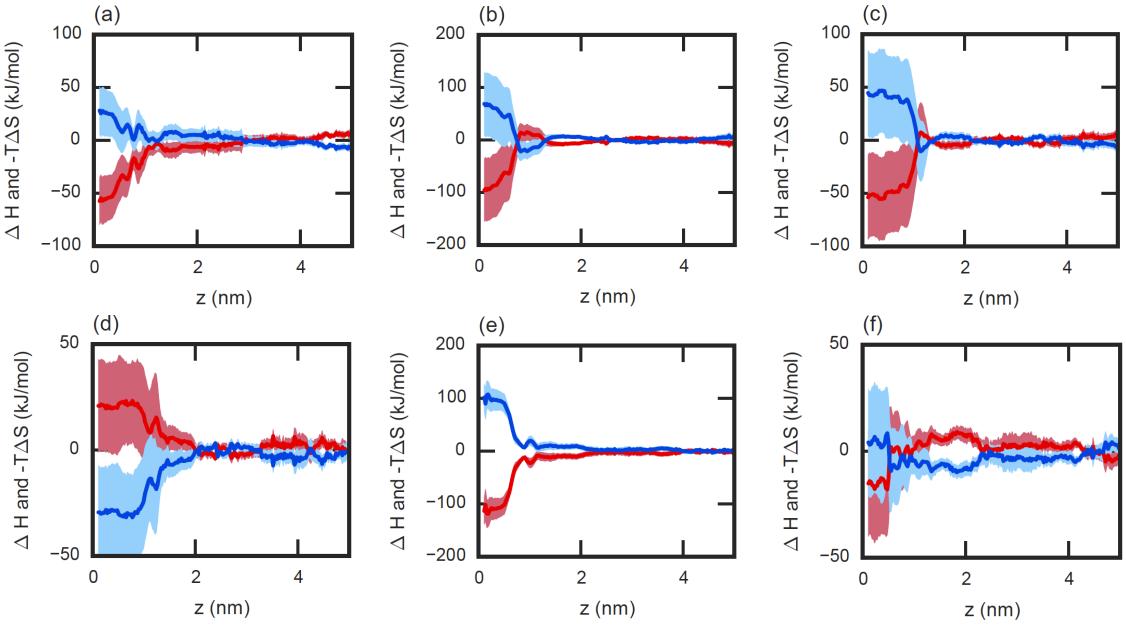

Figure 4.20.: Enthalpy (red) and entropy (blue) calculated from the umbrella sampling simulations for the formation of a Aβ_{16-22} dimer using the following force fields: (a) Gromos54a7, (b) OPLS-AA, (c) AMBER03WS, (d) CHARMM22*, (e) AMBER99SB*ILDN and (f) AMBER99SB*ILDN with LRDI. The error is colored as a shaded area.

my simulations.

4.3. Summary

To be able to use molecular simulations to properly simulate amyloid formation, the force fields used must accurately model the free energy landscape of both IDPs and protein-protein interactions. In this chapter, I studied how various force fields reproduce the free energy landscape of Aβ_{42}, an example for IDPs, and the aggregation of Aβ_{16-22} and two mutations: Aβ_{16-22}(F19V,F20V) and Aβ_{16-22}(F19L) to estimate if force fields accurately reproduce protein aggregation.

The comparison of the simulations of Aβ_{42} with experimental observables is straightforward because of the existence of multiple experimental NMR data, in particular J-coupling, chemical shifts and residual dipolar couplings, and the ease of calculating them from the simulated ensemble. Most of the tested force fields (CHARMM22*, AMBER99SB, OPLS and AMBER99SB*ILDN) are accurate enough in reproducing the experimental observables, with CHARMM22* being slightly better than others. However, in general the errors of the esti-

mation methods are high which means that it is often hard to differentiate which of the force field is better. It is interesting that even if the correlation between the experimental observables and simulated ensembles are comparable for the different force fields, they nonetheless sample distinct secondary structures and tertiary contacts. This means that based on the current experimental data that is available for Aβ_{42}, it is difficult to decide which of the sampled structures are the correct ones.

In the case of the simulations of peptide aggregation, the comparison with experimental observables is complicated. In this chapter, I simulated the aggregation of Aβ_{16-22} and Aβ_{16-22}(F19L) which are known to aggregate and Aβ_{16-22}(F19V,F20V) which is known not to aggregate. However, in most of the simulations all peptides behave rather similarly. While in the simulations with most force fields all peptides aggregate, in the simulation with AMBER03WS little aggregation is observed. Even if the aggregation kinetics and propensity between different force fields are diverse, the most relevant contacts between peptides are the same for all force fields for the peptides that are known to aggregate experimentally. However, different contacts are observed for Aβ_{16-22}(F19V,F20V). In particular, it is interesting that only when peptides found each other in a perfect antiparallel β-sheet, oligomers are stable in the simulation with AMBER03WS. One of the main limitation of these and many other studies is that the systems are studied at concentrations much higher than the experimental values.[2] Nonetheless, this study shows that care must be taken in the selection of force fields for the simulation of protein aggregation.

5. Aggregation propensity encoded in the monomer dynamics

There is no absolute theory regarding what makes certain protein and peptides aggregate into amyloid formation and others not. There are multiple algorithms that can predict amyloid propensity based on the sequence of a protein.[313–317] These algorithms are usually based on the hydrophobicity, β-sheet propensity, net charge and solubility of the proteins. However, these algorithms cannot be used to understand the physico-chemical aggregation process. The interaction between proteins is, as expected, essential for protein aggregation, however, it is still an open question if the conformations of the monomers can be used to predict protein aggregation.

It has been suggested that protein aggregation is encoded in the monomer conformations.[53,318–321] In particular, one of the strongest hypotheses is that proteins can be found in two different conformations, a conformation M which is not aggregation prone, and an aggregation prone conformation M*. The M* conformation usually has hydrophobic residues at the surface that can stabilize protein-protein interactions. Further, it has been suggested that the reconfiguration rate between these states is essential for understanding the propensity to protein aggregation[322] and there is some experimental evidence to support this theory.[323–325] So far, no one has performed long atomistic explicit-solvent molecular dynamics simulations of multiple peptides with different aggregation propensities and try to correlate their experimentally known aggregation propensity with the conformational ensembles of the monomers.

Here, I study the monomer dynamics by means of molecular simulations of Aβ_{16-22} and eight different mutants: Aβ_{16-22}(F19A), Aβ_{16-22}(F19L,F20L), Aβ_{16-22}(F19L), Aβ_{16-22}(F20L), Aβ_{16-22}(F20V), Aβ_{16-22}(F19Y), Aβ_{16-22}(F19V) and Aβ_{16-22}(F19V,F20V). The aggregation propensities of these peptides have been experimentally characterized.[252,289] The conformational ensembles of the peptides are described using standard analysis and the kinetic properties of the peptides are estimated using the recently developed variational approach to molecular kinetics.[326–328]

5.1. Theory

The description of the properties of the conformational ensembles sampled by the simulated peptides is performed with methods which are standard for analysing molecular dynamics simulations, and have been used in previous chapters. Hence, they are explained briefly in section 5.2. On the other hand, to estimate the kinetic properties of the peptides, a new method that uses a variational approach to molecular kinetics was used.[326–328] The theory behind the method is introduced in this section.

If we consider X to be the conformational space that a protein in water can access to be ergodic, reversible and Markovian, the probability density of a particular state $x \in X$ is $p_t(x)$ at time t. $p_t(x)$ will relax to the equilibrium probability distribution $\pi(x)$ for $t \to \infty$. The evolution of the probability density is:

$$p_{t+\tau} = \mathcal{P}(\tau)p_t(x), \tag{5.1}$$

where $\mathcal{P}(\tau)$ is the propagator operator, which calculates the new probability distribution after a time interval τ, called the lag time. The propagator operator has the following eigenvalue spectrum

$$\lambda_1 = 1 > |\lambda_2(\tau)| \geq |\lambda_3(\tau)| \ldots, \tag{5.2}$$

where λ_1 is the largest eigenvalue. Its associated eigenvector is the equilibrium probability distribution $\pi(x)$:

$$\mathcal{P}(\tau)\pi(x) = \lambda_1\pi(x), \tag{5.3}$$

which means that when the systems reaches $\pi(x)$, it will remain there. Because of ergodicity and reversibility, the eigenvalues $\lambda_i(\tau)$ and eigenvectors $l_i(x)$ have real values. Also, because of self-adjointness, the eigenfunctions form a complete basis and the probability density can be decomposed as:

$$p_t(x) = \sum_{i=1}^{\infty} c_i\lambda_i(\tau)l_i(x) = \pi(x) + \sum_{i=2}^{\infty} c_i \exp(-t/t_i)l_i(x), \tag{5.4}$$

where $t_i = -\frac{\tau}{\ln|\lambda_i(\tau)|}$ and c_i are the coefficients determined at time $t = 0$. $p_t(x)$ can be seen as a superposition of modes l_i, and considering that $|\lambda_i| < 1$ for $i > 2$, all modes decay with time as the process approaches the equilibrium probability distribution. The system cannot be solved analytically and, therefore, a variational approach is used to estimate the dominant eigenfunctions and eigenvalues, which hold important information of the most relevant

conformational states and the rates of exchanges between these states.

Because of the self-adjointness and the bounded eigenvalue spectrum of the propagator, a variational principle can be stated:

$$\langle f|\mathcal{P}(\tau)|f\rangle_{\pi^{-1}} \leq 1, \tag{5.5}$$

where the scalar product must be weighted by π^{-1}. Using the method of linear variation, the eigenfunctions can be approximated as linear combinations of basis functions $\phi_i(x)$:

$$l_i(x) \approx \hat{l}_i(x) = \sum_{j=1}^{M} a_{ji}\phi_j(x), \tag{5.6}$$

where M is the number of basis functions. If the expansion on the basis is introduced in the variational principle and the coefficients a_{ji} are used to maximize the variational problem, a generalized eigenvalue problem is obtained:

$$\mathbf{C}(\tau)\mathbf{a}_i = \mathbf{S}\lambda_i(\tau)\mathbf{a}_i, \tag{5.7}$$

where $\mathbf{C}(\tau)$ is the correlation matrix with elements $C_{ij}(\tau) = \langle\phi_i|\mathcal{P}(\tau)\phi_j\rangle_{\pi^{-1}}$ and $\mathbf{S}$ is the overlap matrix with elements $S_{ij}(\tau) = \langle\phi_i|\phi_j\rangle_{\pi^{-1}}$. For a large space such as X, the correlation matrix can be calculated as time-lagged correlation function, which for finite times is equal to:

$$\hat{C}_{ij}(\tau) = \frac{1}{N_T - n_\tau}\sum_{t=1}^{N_T-n_\tau} \chi_j(x_t)\chi_i(x_{t+n_\tau}), \tag{5.8}$$

where x_t is a time-discretized realization of the dynamical process, N_T is the length of the realization, and $n_\tau = \tau/\Delta t$. It should be noted, that the correlation functions are calculated with respect to the cofunctions χ_i and not the basis functions ϕ_i, where $\chi_i(x) = \pi^{-1}(x)\phi_i(x)$. The overlap matrix can be calculated as:

$$\hat{S}_{ij} = \frac{1}{N_T}\sum_{t=1}^{N_T} \chi_j(x_t)\chi_i(x_t). \tag{5.9}$$

In conclusion, to use the variational approach to MD simulations, first a realization x_t needs to be performed (i.e., a trajectory of an MD simulation). Then, after the basis functions are defined, x_t needs to be projected onto these basis functions. Next, with a chosen lag time, the correlation and overlap matrix are estimated. Finally, using equation 5.7 the eigenvalues and expansion coefficient for the first M eigenfunctions are calculated.

An essential step in using the variational approach is the definition of the basis sets that will be used for the projection of the trajectory. A correct basis set would be able to distinguish all major conformational changes, should be transferable (i.e., when applying the variational approach to different peptides, it should be easy to compare between them) and should be as small as possible. Vitalini et al.[328] designed a basis set based on the backbone torsion angles ϕ and ψ of the capped amino acids that can be used for the simulation of small peptides. In this thesis, the same basis set will be used. In short, the basis function χ for each peptide is formed by the tensor product of the residue basis function:

$$\chi(\phi_1, \psi_1, \phi_2, \psi_2, \ldots, \phi_N, \psi_N) = R(\phi_1, \psi_1) \otimes R(\phi_2, \psi_2) \otimes \ldots \otimes R(\phi_N, \psi_N) \quad (5.10)$$

where $R(\phi, \psi)$ are the dynamic modes for that particular residue. In this way, the basis sets are constructed by combining the dynamic modes of the individual residues. Physically, this means that the motions of each residue are considered independently. In this thesis, only the first three dynamic modes for each residue are used, with the first mode being the ground state of the residue, the second mode being the slowest transition in the residue and the third mode being the second slowest transition of the residue. In this way, the basis function $\chi_{1111111}$ represents the entire peptide (in this case a heptamer, which is why 7 residue basis functions are needed) being in the ground state, and the basis function $\chi_{1112111}$ represents all residues being in equilibrium but residue 4 having a transition in its slowest dynamical mode. The residue basis functions depend on the force field used and are constructed here for AMBER99SB*ILDN.

5.2. Computational Methodology

All peptides were simulated with capped termini, to reproduce experimental conditions.[252,289] To sample the conformational ensemble of the peptide thoroughly, multiple simulations were performed. To start these simulations from different relevant conformations, a short Hamiltonian replica exchange MD (HREMD) simulation was performed.[329] HREMD is similar to REMD, however, instead of the temperature, the Hamiltonian is changed between different replicas. One of the problems with REMD is that the number of replicas depend on the number of atoms in the system. If explicit solvent simulations are performed, the number of replicas can become very large and, thus, the simulation can become computationally too expensive. In HREMD, the sys-

tem is divided into a hot section, usually the protein, whose temperature is implicitly increased in the higher temperature replicas, and a cold section, usually the solvent and ions, whose temperature remains the same. In this way, the number of replicas depends only on the number of atoms in the hot region and not on the entire system. Considering that the goal of enhanced sampling algorithms is to enhance the sampling of the protein and not the solvent, HREMD is an efficient way of saving computing time. To separate the system into hot and cold sections, the Hamiltonian of the system must be modified. The algorithm used in this chapter is the REST2 scheme,[330–332] in which the Hamiltonian at each replica is scaled based on a parameter λ. In particular, charges are scaled by a factor of $\sqrt{\lambda}$, the Lennard-Jones parameters ϵ are scaled by a factor λ and the dihedral potentials are scaled by a factor λ. The parameter λ is 1 for the reference replica and between 0 and 1 for all other replicas. The HREMD simulations were performed with the PLUMED plug-in[333,334] in Gromacs 4.6.[84] HREMD has already been used to increase the sampling in simulations of peptides[335–337] and even in the study of oligomer formation by amyloidogenic peptides.[103]

All simulations were performed with the AMBER99SB*ILDN force field[208–210] and the TIP4-Ew water model.[138] For the HREMD simulation, each peptide was introduced in a dodecahedral box in an extended conformation, with a 1.2 nm distance between the peptide and the box walls. The system was solvated and ions were added up to a concentration of 150 mM. It was then minimized using the steepest descent algorithm, and equilibrated with 0.1 ns NVT and NPT simulations. A 100 ns HREMD simulation was then launched using 8 replicas. The scaling factor λ was evenly distributed between 1.0 and 0.5. The temperature was kept constant at 310 K using the Nosé-Hoover algorithm[157,158] and the pressure was kept contant at 1 bar with the Parrinello-Rahman algorithm.[160,161] Protein bonds and water molecules were constrained with the LINCS algorithm[142] and the Settle algorithm,[144] respectively. A 2 fs timestep was used. A 1 nm cutoff was used for both Lennard-Jones interactions and short range electrostatic interactions. Electrostatic interactions were calculated using the particle mesh Ewald method.[150]

From the last 80 ns of the HREMD simulations with $\lambda = 1$, five conformations were picked at random and standard MD simulations were started from them. The system was again equilibrated with 0.1 ns NVT and NPT simulations. Then, 2 µs production runs were performed for each replica with the same parameters as for the HREMD simulations, leading to a total of 10 µs simulation time per peptide. These simulations were performed with Gromacs

5.[110]

The analysis of the conformational ensemble sampled by the peptides was done with the help of MDAnalysis.[177] Similar to the previous chapter, peptides were considered to be in a β-strand conformation when the dihedral angles ϕ and ψ were inside the polygon (−180°,180°), (−180°,126°), (−162°,126°), (−162°,108°), (−144°,108°), (−144°,90°), (−50°,90°) and (−50°,180°).[78,82] For calculating intramolecular contact maps, two residues were considered to be in contact whenever the distance between any two atoms of the residues were under 0.4 nm.

To use the variational approach for understanding the kinetic properties of the peptides, first a basis set was created from the dynamics of the individual amino acids. For these simulations, the residues were capped (Ac-X-NHMe, where X is the residue under study) and introduced in dodecahedral box with a distance between the peptide and the walls of 1.2 nm. The system was solvated and NaCl ions were added to simulate a concentration of 150 mM. It was then minimized with the steepest descent algorithm, and then equilibrated with a 0.1 NVT and NPT simulations. Then, a 1 µs production run was performed with the same parameters as for the HREMD simulations.

For the construction of the residue-based functions, Markov state models (MSM) were constructed from the residue simulations with the help of PyEMMA.[338] For each trajectory, the backbone torsion angles were first extracted and projected onto a grid of 36 × 36 states. Then, an MSM was constructed using standard input parameters and used to calculate the main dynamical modes. The basis functions were constructed for the following amino acids: alanine, glutamate, leucine, lysine, phenylalanine, tyrosine and valine.

For the application of the variational approach to the peptides, only singly excited states (i.e., only one residue can be excited at a time) were used for the construction of the peptide basis set.

5.3. Results and Discussion

5.3.1. Conformational ensemble

In this section, the conformational ensemble sampled by the nine different peptides were studied. Among the nine peptides, six have different aggregation propensities, while three (Aβ_{16-22}(F19A), Aβ_{16-22}(F19V,F20V) and Aβ_{16-22}(F19Y)) do not aggregate.[252,289] In figure 5.1, the β-strand content is plotted. All peptides have similar β-strand content and there is no clear difference between the aggregating and non-aggregating peptides. All pep-

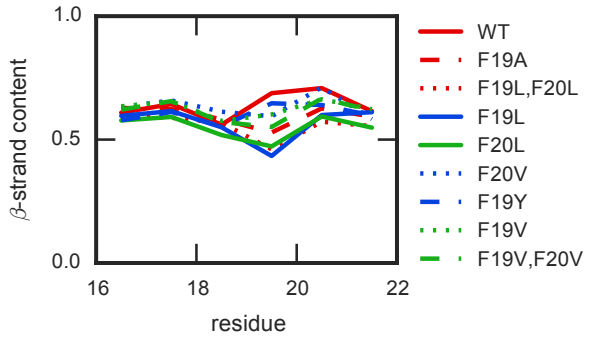

Figure 5.1.: Average β-strand content per residue for Aβ_{16-22} (red solid line), Aβ_{16-22}(F19A) (red dashed line), Aβ_{16-22}(F19L,F20L) (red dotted line), Aβ_{16-22}(F19L) (blue solid line), Aβ_{16-22}(F20L) (green solid line), Aβ_{16-22}(F20V) (blue dotted line), Aβ_{16-22}(F19Y) (blue dashed line), Aβ_{16-22}(F19V) (green dotted line) and Aβ_{16-22}(F19V,F20V) (green dashed line). The three peptides which do not aggregate are plotted as dashed lines.

tides have high β-strand content with a probability between 0.5 and 0.7 for all residues. There are only small differences in residues 19-21 and the non-aggregating peptides are some of the ones with higher β-strand content. To better understand the conformational ensemble of the peptides, and considering how important the monomer dynamics was to understand the difference in aggregation between functional and aberrant amyloids (see chapter 3), intramolecular monomer contact maps were plotted for each peptide in figure 5.2. There are no clear differences between aggregating and non-aggregating peptides. All peptides are mostly extended and very few contacts are observed between the termini. These results show that there is no clear correlation between the ensembles sampled by the peptides and their aggregation propensity.

5.3.2. Kinetic properties

In this section, the kinetic properties of the nine peptides are studied using the variational approach to molecular kinetics. First, MSMs were built for the residues to construct the residue-based basis functions. In figure 5.3, the implied timescales as a function of lag time for the MSM of phenylalanine is plotted. It is clear that the MSM converges fast, because the implied timescales do not change as a function of the lag time. Therefore, a lag time of 50 ps was used to construct the peptide basis functions. Similar plots are observed for all the residues. The first three dynamical modes for phenylalanine are plotted in figure 5.4. The first mode is the ground state which represents the

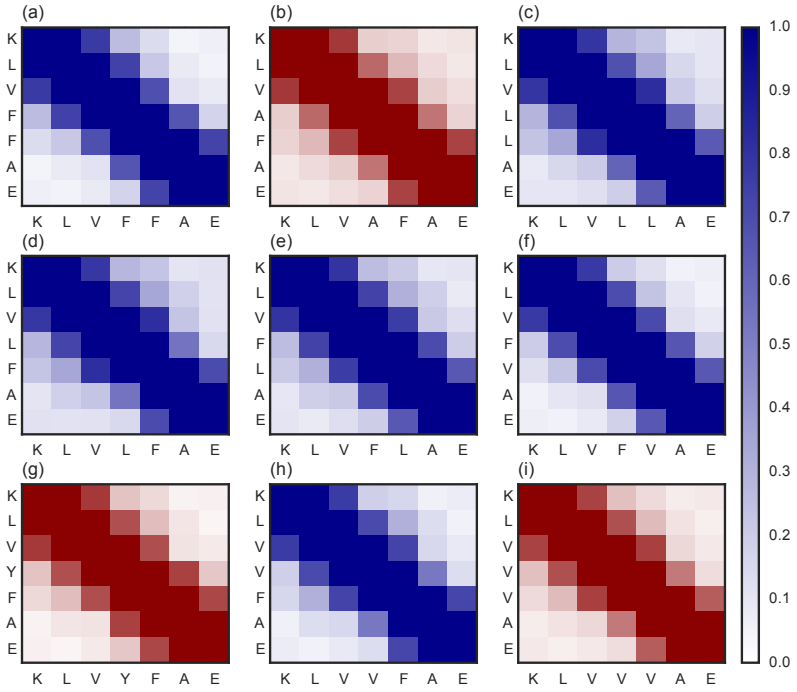

Figure 5.2.: Intramolecular contact maps for the monomers of (a) Aβ_{16-22}, (b) Aβ_{16-22}(F19A), (c) Aβ_{16-22}(F19L,F20L), (d) Aβ_{16-22}(F19L), (e) Aβ_{16-22}(F20L), (f) Aβ_{16-22}(F20V), (g) Aβ_{16-22}(F19Y), (h) Aβ_{16-22}(F19V) and (i) Aβ_{16-22}(F19V,F20V). The peptides that do not aggregate are plotted in red.

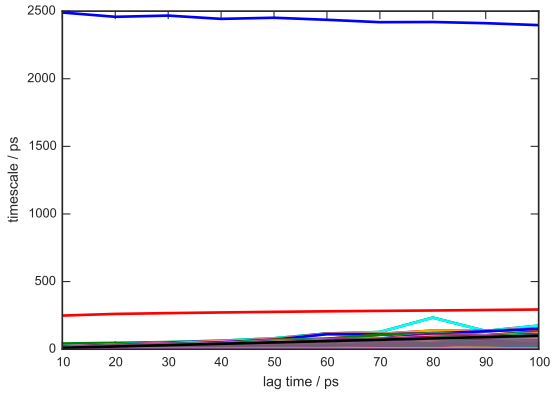

Figure 5.3.: Implied timescales as a function of the lag time for the Markov state model for the capped phenylalanine residue.

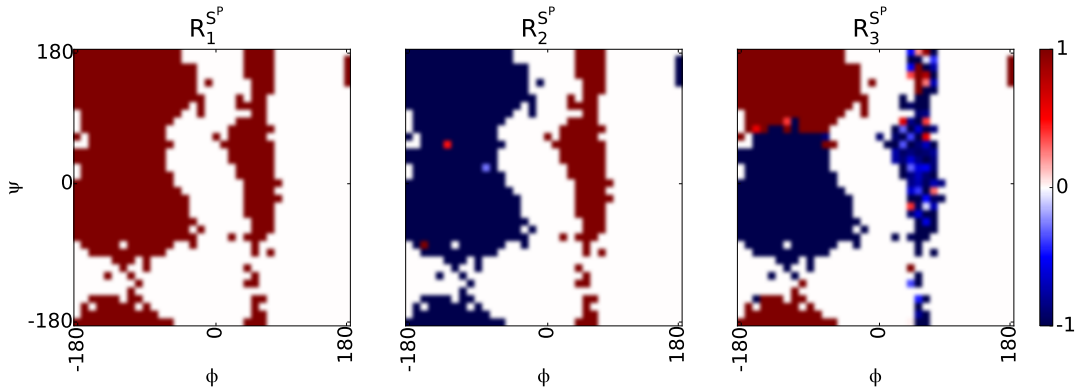

Figure 5.4.: Most relevant dynamical modes for the Markov state model for the capped phenylalanine residue. The first dynamical mode is the equilibrium distribution. The second dynamical mode is a change in the ϕ angle, while a transition along ψ occurs in the third dynamical mode.

equilibrium probability density. The second mode represents the transition between right-handed α-helices and β-strands, to left-handed α-helices. The third mode represents the transition between α-helices and β-strands. Similar modes are observed for all residues. Using these residue-based basis functions, peptide basis functions were constructed. In particular, only the singly excited (i.e., only one of the residues can be either in the second or third dynamical mode) were used. All of the used basis functions are listed in table 5.1.

With the use of the peptide basis functions, the variational approach was applied to the simulations of the 9 peptides. In figure 5.5, the implied timescale as a function of lag time is plotted for Aβ_{16-22}. In this case, the convergence is much slower than for the simulations of the capped residues. This likely occurs because of the more complex dynamics that the peptides have compared to the

Table 5.1.: Transitions represented by each basis functions.

Basis function	Transition
1	1111111
2	2111111
3	3111111
4	1211111
5	1311111
6	1121111
7	1131111
8	1112111
9	1113111
10	1111211
11	1111311
12	1111121
13	1111131
14	1111112
15	1111113

capped residues. Nevertheless, the timescale do stay reasonably constant after a lag time of around 5 ns. Similar results are observed for all other peptides. Therefore, this lag time is chosen for the analysis of the peptide dynamics. The main results from the variational approach are the timescales for the most relevant processes and the basis functions involved in such processes. In figure 5.6, the coefficients for the basis functions involved in the six slowest processes for Aβ_{16-22} are plotted. As expected, the first eigenvalue $\lambda = 1$ represents the equilibrium probability density (i.e., the first basis function). The second eigenvalue represents mostly basis function 8, with contributions of basis functions 4, 6 and 10. Basis function 8 represents the transition of residue 4 (i.e., Phe^{19}) from a right-handed α-helix and β-strand, to left-handed α-helix, which considering that the peptides are mostly extended can be understood as a closing of the peptides through its central residue. The third eigenvalue represents mostly the basis function 6 with contributions from 4, 8 and 10. Basis function 6 is similar to basis function 8, but for residue 3 (i.e., Val^{18}). In this way, the slowest transitions and their identities can be extracted from the variational method. In table 5.2, the two slowest transitions and the basis functions they represent are listed for each of the peptides.

Considering that no correlation was observed between static properties and the aggregation propensity of the different peptides, and that the dynamics of the monomer may be essential for aggregation, I decided to study if there is a correlation between the slowest timescales and the aggregation propensity. With this in mind, the slowest timescales were plotted as a function of the criti-

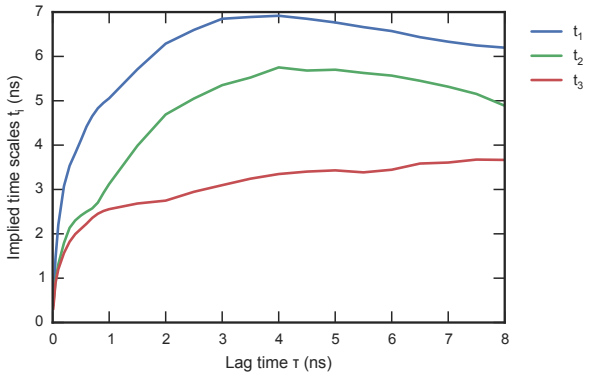

Figure 5.5.: Three slowest implied timescales as a function of lag time for the simulation of Aβ_{16-22}. After an initial increase in the timescale, they plateau around 5 ns, which is chosen as the lag time for estimating the timescales of the major processes.

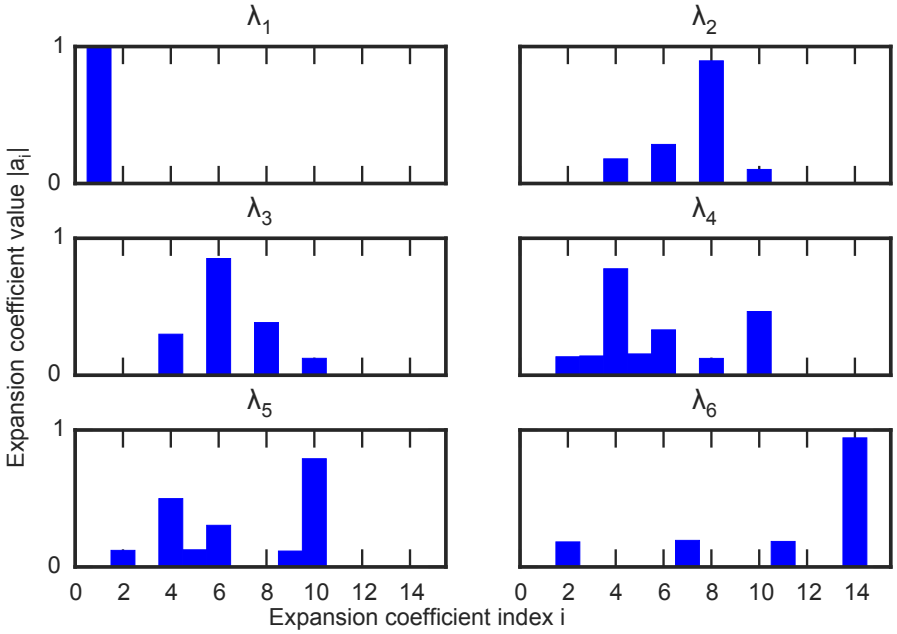

Figure 5.6.: Expansions coefficients for the six largest eigenvalues for the simulation of Aβ_{16-22}.

Table 5.2.: The two slowest timescales and basis functions they represent for each of the peptides under study. The non-aggregating peptides are highlighted in red.

Peptide	t_1 (ns)	Process	t_2 (ns)	Process
Aβ_{16-22}	6.8	1112111	5.7	1121111
Aβ_{16-22}(F19A)	5.5	1112111 and 1121111	4.3	1211111
Aβ_{16-22}(F19L,F20L)	12.8	1112111	6.2	1211111
Aβ_{16-22}(F19L)	13.2	1121111	4.8	1112111
Aβ_{16-22}(F20L)	10.0	1112111	4.0	1211111
Aβ_{16-22}(F20V)	11.1	1112111	4.2	1121111
Aβ_{16-22}(F19Y)	7.5	1112111	5.5	1121111
Aβ_{16-22}(F19V)	5.7	1112111	4.2	1111211
Aβ_{16-22}(F19V,F20V)	10.2	1112111	6.8	1121111

cal monomer concentration needed for the peptide to aggregate calculated from experiments in figure 5.7. As mentioned in the previous chapter, the critical concentration represents the concentration in which there is an equilibrium between adding a monomer to a fibril and a monomer separating from an existing fibril. Hence, the smaller the critical concentration, the higher the aggregation propensity of the peptide. In figure 5.7, the three peptides which do not aggregate *in vitro* (Aβ_{16-22}(F19A), Aβ_{16-22}(F19Y) and Aβ_{16-22}(F19V,F20V)) are plotted with critical concentrations of 100 µM, which is the maximum concentration experimentally studied. These peptides may aggregate at a higher concentration. Considering all peptides, there is no clear correlation between the data points in figure 5.7. However, there seems to be a correlation for the aggregating peptides. Hence, they were plotted separately in figure 5.8. Here, a very clear correlation is observed between the critical concentration of the peptides and its slowest implied timescale: the peptides with slower timescales are the ones that aggregate faster. Considering the theory from Lapidus,[322] it appears that for these small peptides, most of them lie in a regime in which the aggregation is slowed by the rapid reconfiguration rate of the peptides which does not allow them to stay in their aggregation prone position for enough time to rapidly aggregate. Hence, when the reconfiguration rate decreases, they stay longer in the aggregation prone position and their aggregation is faster. The fact that the non-aggregating peptides do not follow this trend probably arises from the fact that this theory is too simplistic. Other factors, such as the strength of the interactions between the peptides may also be important for protein aggregation.

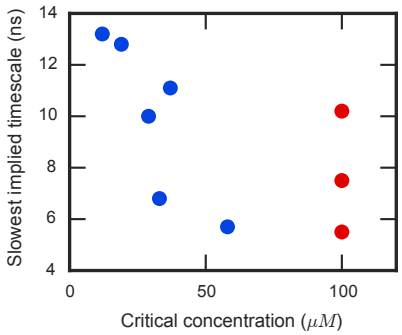

Figure 5.7.: Correlation between experimental critical concentrations measured by Senguen et al.[252,289] and the slowest implied timescale for Aβ_{16-22} and its eight simulated mutants. The three data points at the left represent peptides that do not aggregate (Aβ_{16-22}(F19A), Aβ_{16-22}(F19Y) and Aβ_{16-22}(F19V,F20V)) whose critical concentration is larger than 100 µM.

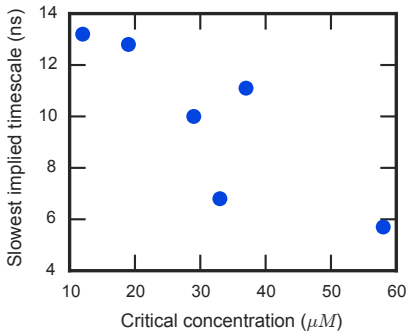

Figure 5.8.: Correlation between experimental critical concentrations measured by Senguen et al.[252,289] and the slowest implied timescale for Aβ_{16-22} and the 5 simulated mutants which aggregate.

5.4. Summary

There are still many open questions about what drives protein aggregation. In particular, it is still hard to predict if a peptide aggregates and if it will aggregate faster than other peptides. Even though there exist multiple bioinformatical programs to predict the aggregation propensity based on properties of the residues, they are not able to make predictions about the aggregation pathways. One of the most interesting hypotheses, is that the aggregation propensity is based on the conformational ensemble of the monomer. In this chapter, this hypothesis was tested by studying the monomer conformational ensemble of Aβ_{16-22} and eight different mutants: Aβ_{16-22}(F19A), Aβ_{16-22}(F19L,F20L), Aβ_{16-22}(F19L), Aβ_{16-22}(F20L), Aβ_{16-22}(F20V), Aβ_{16-22}(F19Y), Aβ_{16-22}(F19V) and Aβ_{16-22}(F19V,F20V). These peptides have different aggregation propensities which have been characterized experimentally, with three of them not aggregating at all.[252,289]

No correlation is observed between average ensemble properties of the peptides and their aggregation propensity. In particular, all peptides have a large percentage of β-strand content, which is essential for amyloid formation, and form mostly extended structures. This is true even for peptides which do not aggregate. However, when the kinetic properties of the peptides are studied, a clear correlation is observed between experimental and simulated properties. Here, a variational approach to molecular kinetics was used to measure the slowest transitions. A negative correlation is observed between the slowest implied timescales and the experimentally known critical concentration. This shows that these peptides are in the regime in which aggregation can be accelerated by decreasing its reconfiguration rate, which means that slower aggregating peptides reconfigure too fast to form stable contacts. It should be noted that non-aggregating peptides do not follow this correlation. This shows that the process of amyloid formation is more complicated and not solely determined by the structural preference and kinetics of the monomer.

6. Conclusions and Outlook

The aggregation of peptides into highly structured amyloid fibrils is associated with multiple diseases including Alzheimer's disease, Parkinson's disease and type 2 diabetes. However, amyloids are also associated with biologically functional roles such as storing of signalling peptides. Moreover, because of the intrinsic properties of amyloids (biocompatible, stable, self-aggregating, rigid and can be assembled from simple building blocks), they could be used as the basis for novel nanomaterials. If we could understand the driving forces behind the aggregation of peptides into amyloid fibrils, we could find drugs to treat amyloid-related diseases and design novel amyloid materials.

To design drugs for amyloid-related diseases, we should first understand which are their toxic agents. Most experimental evidence shows that they are not highly structured amyloid fibrils but smaller and usually more disordered oligomers which are the main toxic species. Hence, to understand the difference in toxicity between functional and aberrant amyloids, and to understand why certain mutations lead to higher toxicity, the conformational ensembles of oligomers need to be determined. However, most oligomer species are transient and can only be studied experimentally by artificially stabilizing them. Therefore, molecular simulations are an essential tool for the understanding of how amyloid oligomers behave. In particular, atomistic explicit solvent molecular dynamics (MD) simulations can provide a detailed understanding of how the aggregation process proceeds. In this thesis, I studied the formation of amyloid oligomers and their preceding monomers using extensive MD simulations.

In chapter 3, the difference in oligomer formation between functional and aberrant amyloidogenic peptides was studied. In particular, the aggregation of Alzheimer's Aβ_{25-35} and two functional tachykinin peptides: kassinin and neuromedin K. Aβ_{25-35} is a section of the Alzheimer's peptide which aggregates and is toxic on its own. Tachykinin peptides are neuropeptides which are stored in secretory granules as amyloids. Moreover, Aβ_{25-35} and tachykinin peptides have similar primary sequences, which is particularly interesting because of their dissimilar toxicity. In the simulations, tachykinin peptides aggregate faster than Aβ_{25-35}, which suggests that functional amyloids avoid being toxic by aggregating faster than toxic amyloids thereby surpassing the

toxic oligomer phase. Furthermore, I found that the charges of the C-terminus are essential in modulating aggregation. For example, when capping the C-terminus the aggregation kinetics of kassinin decreases considerably. This is important because tachykinin peptides need to be capped for their agonist activity. Moreover, I observed that peptides which exist in the monomer phase in extended conformations aggregate faster than those that exist in hairpin-like conformations.

The study of functional amyloids and the application of molecular simulations has only just started. Major experimental studies should be performed in the future to understand the difference in behaviour in the oligomer formation of functional and aberrant amyloids. Different properties that can be measured experimentally include aggregation kinetic constants and hydrophobic surface areas. This study is a first step towards understanding the difference in oligomer formation between functional and aberrant amyloids.

In chapter 4, the effect of multiple force fields on intrinsically disordered proteins (IDPs), which is relevant as most amyloid-forming peptides are IDPs in their monomer phase, and on protein aggregation was studied. Testing the accuracy of current force field is essential for correctly modeling biological processes. To study the effect of force fields on IDPs, various force fields were used to model Aβ_{42} and the results were compared to experimental nuclear magnetic resonance (NMR) data, such as J-couplings, chemical shifts and residual dipolar couplings. Most current force fields excluding AMBER99SBILDN-NMR are accurate enough at reproducing NMR observables, with CHARMM22* being slightly better than others. However, the error in the prediction of NMR observables from simulated ensembles is so high that it is hard to differentiate between force fields. In my simulations, I did not observe that the simulated Aβ_{42} conformational ensembles are too compact when compared to experiments, which has been observed in other simulations of IDPs. To study the effect of force fields on protein aggregation, various force fields were used to model the aggregation of Aβ_{16-22} and two mutants: Aβ_{16-22}(F19L), which aggregates faster, and Aβ_{16-22}(F19V,F20V), which does not aggregate *in vitro*. However, the three peptides behave similarly during the aggregation simulations with all force fields. The difference in behaviour between force fields is striking too: fast aggregation is observed with most force fields (particularly with Gromos and OPLS), but little aggregation is observed in the simulation with AMBER03WS. Even if the aggregation kinetics is different between force fields, the most relevant intermolecular contacts between peptides are similar. The most remarkable difference is seen for Aβ_{16-22}(F19V,F20V)

with AMBER03WS with which the peptide only aggregate when in a perfect antiparallel β-sheet, which suggests that this peptide only aggregates in this specific conformation, and its aggregation may not happen at low concentrations because of the unlikelihood of this conformation being formed. For all other force fields and peptide conformations, a large percentage of unspecific interactions are found to be stable.

Finally, in chapter 5, the monomers of different mutants of Aβ_{16-22} were simulated to understand if the aggregation propensity is encoded in the dynamics of the monomer. It is still hard to predict which peptides form amyloids and at which speed. Particularly, it is still unknown if this prediction can be based on the conformational ensemble of the monomer. In my simulations, no correlation was observed between ensemble average properties, such as β-strand content or intramolecular contact map, and aggregation propensity. All peptides sample similar β-strand content and intramolecular contact map. However, a strong correlation between aggregation propensity and the kinetic properties of the peptides was identified. In particular, there is a negative correlation between the experimental critical concentration and the implied timescales of the first dynamical process in the peptide. This suggests that more dynamic peptides aggregate slower because they stay in the aggregation prone conformation for too short a time to form stable intermolecular contacts. When the peptide dynamics decreases, the aggregation propensity of the peptides increases. However, it should be noted that this correlation does not include the mutants that do not aggregate (Aβ_{16-22}(F19A), Aβ_{16-22}(F19Y) and Aβ_{16-22}(F19V,F20V)), which suggests that aggregation is a more complex process where intermolecular contacts are also essential for predicting aggregation propensities and dynamics.

Atomistic simulations of protein aggregation is still in its infancy. Only a few years ago has it become possible to even perform such simulations because of the advent of faster computers and better MD codes. Hence, there are many open questions about how to perform simulations of protein aggregation. One of the main problems is the fact that simulations are often performed at much higher concentrations compared to the experimental set up. The reason for this is that if the experimental concentrations were used, simulation systems would be too large and thus the simulations too expensive to be performed. Moreover, most of the simulation time would be spent on water molecules and in the simulation of protein diffusion, which could be estimated with a simulation of the monomer. Furthermore, the conformational sampling of proteins is slow, which means that if simulations are performed at high concentration,

peptides do not have enough time to rearrange before other peptides collide with them. Hence, new multiscale methods should be created to simulate this process at low concentrations. For example, simulations of oligomer formation could be performed by means of monomer addition. The aggregation kinetics could then be corrected using the experimental or simulated diffusion constant. Moreover, care must be taken with the force fields used for simulating protein aggregation as most of the current force fields seem to overstabilize protein-protein aggregation. However, it should also be mentioned that better experimental methods are needed to be able to perform an accurate comparison between simulations and experiments, so as to have a clear picture of which is the correct force field to use. The advance in the simulations of protein aggregation will perform an important role in the understanding of protein aggregation and amyloid formation. Such an understanding will help us to find a cure for amyloid associated diseases and to precisely design amyloid materials.

Appendices

A. Errors in correlated data

In MD simulations, calculating the error of the simulated properties is not trivial because the snapshots used for the calculation are not independent from each other. Therefore, errors cannot be calculated with the normal equation for the standard error of the mean:

$$SE = \sqrt{\frac{\sum_{i=1}^{N} x_i - \bar{x}}{(N-1)N}} \tag{A.1}$$

where N is the number of snapshots, x_i is the value of the observable for each snapshot and $\bar{x}$ is the average value of the observable. Even if multiple simulations were started from the same starting configuration, these trajectories would still not be entirely independent.

Here, I use the blocking method, which is based on the renormalization group theory, summarized by Flyvbjerg and Petersen.[339] This method was probably invented by Wilson[340] and is also described by Whitmer[341] and Gottlieb et al.[342] The blocking method is based on repeatedly blocking the data and calculating the errors for this blocked data. In particular, if the data set is x_1, ..., x_n, the data can be transformed into a data set half as large x'_1,...,x'_n where:

$$x'_1 = \frac{1}{2}(x_{2i-1} + x_{2i}), \tag{A.2}$$

$$n' = \frac{1}{2}n. \tag{A.3}$$

The correlation function can then be defined as:

$$\gamma_{i,j} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle , \tag{A.4}$$

and because of invariance under time translation $\gamma_t \equiv \gamma_{ij}$ where $t = |i - j|$. For the prime variables:

$$\gamma'_t = \begin{cases} \frac{1}{2}\gamma_0 + \frac{1}{2}\gamma_1 & \text{for } t = 0, \\ \frac{1}{4}\gamma_{2t-1} + \frac{1}{2}\gamma_{2t} + \frac{1}{4}\gamma_{2t+1} & \text{for } t > 0. \end{cases} \tag{A.5}$$

The estimator of the variance is then:

$$\sigma^2(\bar{x}') = \frac{1}{n'^2} \sum_{i,j=1}^{n'} \gamma'_{i,j}. \tag{A.6}$$

The error can then be calculated as:

$$\sigma^2(\bar{x}) \geq \frac{\gamma_0}{n}. \tag{A.7}$$

Because the data becomes more uncorrelated with a larger number of block transformations, γ_0/n increases every time the data is blocked, up until it has converged to the real value. A possible estimator with a small bias for γ_t is:

$$c_t \equiv \frac{1}{n-t} \sum_{k=1}^{n-t} (x_k - \bar{x})(x_{k+t} - \bar{x}), \tag{A.8}$$

$$\sigma^2(\bar{x}) \geq \left\langle \frac{c_0}{n-1} \right\rangle. \tag{A.9}$$

Hence, the blocking method works in the following way. First $\frac{c_0}{n-1}$ is calculated. Then, the data is blocked and $\frac{c'_0}{n'-1}$ is calculated. The method is repeated iteratively. The value for $\frac{c_0}{n-1}$ should converge to a defined value. If we consider the variables as independent and following a Gaussian distribution, the variance of the error can be estimated as :

$$\sigma^2(\bar{x}) \approx \frac{c'_0}{n'-1} \pm \sqrt{\frac{2}{n'-1}} \frac{c'_0}{n'-1}. \tag{A.10}$$

In the case of the REMD simulation of Aβ_{42} in chapter 4, the last 100 ns of the C_α chemical shifts were used to estimate the convergence of the data. An example of the estimation of the error is plotted in figure A.1. It can be observed that the error grows as a function of the number of transformation until it converges. However, after too many transformation, it becomes impossible to measure convergence because the estimation of the standard deviation of the error is too high. A conservative 25-ns block, which is equivalent to 8 transformations, was used for the estimation of the errors for all observables.

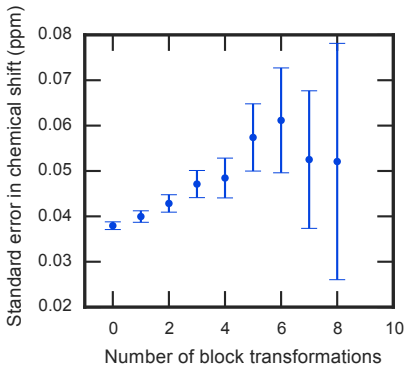

Figure A.1.: Standard error and estimation of the deviation of the standard error (in error bars) as a function of the number of block transformations for the Ile^{32} C_α chemical shift calculated with the blocking method. Similar results are observed for other residues (Reprinted with permission from Carballo-Pacheco and Strodel.[3] Copyright 2016 The Protein Society).

Bibliography

[1] Carballo-Pacheco, M.; Ismail, A. E.; Strodel, B. Oligomer Formation of Toxic and Functional Amyloid Peptides Studied with Atomistic Simulations. *J. Phys. Chem. B* **2015**, *119*, 9696–9705.

[2] Carballo-Pacheco, M.; Strodel, B. Advances in the Simulation of Protein Aggregation at the Atomistic Scale. *J. Phys. Chem. B* **2016**, *120*, 2991–2999.

[3] Carballo-Pacheco, M.; Strodel, B. Comparison of Force Fields for Alzheimer's Aβ_{42}: A Case Study for Intrinsically Disordered Proteins. *Protein Sci.* **2017**, *26*, 174–185.

[4] Carballo-Pacheco, M.; Vancea, I.; Strodel, B. Extension of the FACTS Implicit Solvation Model to Membranes. *J. Chem. Theory Comput.* **2014**, *10*, 3163–3176.

[5] Hora, M.; Carballo-Pacheco, M.; Weber, B.; Morris, V. K.; Wittkopf, A.; Buchner, J.; Strodel, B.; Reif, B. Epigallocatechin-3-gallate Preferentially Induces Aggregation of Amyloidogenic Immunoglobulin Light Chains. *Sci. Rep.* **2017**, *7*, 41515.

[6] Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006**, *75*, 333–366.

[7] Eisenberg, D.; Jucker, M. The Amyloid State of Proteins in Human Diseases. *Cell* **2012**, *148*, 1188 – 1203.

[8] Fowler, D. M.; Koulov, A. V.; Balch, W. E.; Kelly, J. W. Functional Amyloid - from Bacteria to Humans. *Trends Biochem. Sci.* **2007**, *32*, 217–224.

[9] Greenwald, J.; Riek, R. Biology of Amyloid: Structure, Function, and Regulation. *Structure* **2010**, *18*, 1244–1260.

[10] Knowles, T. P. J.; Buehler, M. J. Nanomechanics of Functional and Pathological Amyloid Materials. *Nat. Nanotechnol.* **2011**, *6*, 469–479.

[11] Goldschmidt, L.; Teng, P. K.; Riek, R.; Eisenberg, D. Identifying the Amylome, Proteins Capable of Forming Amyloid-like Fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 3487–3492.

[12] European Commission, Dementia. 2016; `http://ec.europa.eu/health/major_chronic_diseases/diseases/dementia`, [Online; accessed 1-May-2016].

[13] European Commission, Population structure and ageing. 2016; `http://ec.europa.eu/eurostat/statistics-explained/index.php/Population_structure_and_ageing`, [Online; accessed 2-May-2016].

[14] Alzheimer Europe, Cost of illness and burden of dementia - The base option. 2016; `http://www.alzheimer-europe.org/Research/European-Collaboration-on-Dementia/Cost-of-dementia/Cost-of-illness-and-burden-of-dementia`, [Online; accessed 2-May-2016].

[15] Alzheimer's Association, Current Alzheimer's Treatments. 2016; `http://www.alz.org/research/science/alzheimers_disease_treatments.asp`, [Online; accessed 2-May-2016].

[16] European Parkinson's Disease Association, 2016; `http://www.epda.eu.com/en/pd-info/`, [Online; accessed 2-May-2016].

[17] Stefanis, L. α-Synuclein in Parkinson's Disease. *Cold Spring Harbor Perspect. Med.* **2012**, *2*, a009399.

[18] European Commission, Diabetes. 2016; `http://ec.europa.eu/health/major_chronic_diseases/diseases/diabetes/`, [Online; accessed 2-May-2016].

[19] Jaikaran, E. T.; Clark, A. Islet Amyloid and Type 2 Diabetes: From Molecular Misfolding to Islet Pathophysiology. *Biochim. Biophys. Acta, Mol. Basis Dis.* **2001**, *1537*, 179–203.

[20] Ano Bom, A. P. D.; Rangel, L. P.; Costa, D. C. F.; de Oliveira, G. A. P.; Sanches, D.; Braga, C. A.; Gava, L. M.; Ramos, C. H. I.; Cepeda, A. O. T.; Stumbo, A. C.; De Moura Gallo, C. V.; Cordeiro, Y.; Silva, J. L. Mutant p53 Aggregates into Prion-like Amyloid Oligomers and Fibrils: Implications for Cancer. *J. Biol. Chem.* **2012**, *287*, 28152–28162.

[21] Chapman, M. R.; Robinson, L. S.; Pinkner, J. S.; Roth, R.; Heuser, J.; Hammar, M.; Normark, S.; Hultgren, S. J. Role of Escherichia Coli Curli Operons in Directing Amyloid Fiber Formation. *Science* **2002**, *295*, 851–855.

[22] Dos Reis, S.; Coulary-Salin, B.; Forge, V.; Lascu, I.; Bégueret, J.; Saupe, S. J. The HET-s Prion Protein of the Filamentous Fungus Podospora anserina Aggregates in Vitro into Amyloid-like Fibrils. *J. Biol. Chem.* **2002**, *277*, 5703–5706.

[23] Fowler, D. M.; Koulov, A. V.; Alory-Jost, C.; Marks, M. S.; Balch, W. E.; Kelly, J. W. Functional Amyloid Formation Within Mammalian Tissue. *PLoS Biol.* **2005**, *4*, e6.

[24] Maji, S. K.; Perrin, M. H.; Sawaya, M. R.; Jessberger, S.; Vadodaria, K.; Rissman, R. A.; Singru, P. S.; Nilsson, K. P. R.; Simon, R.; Schubert, D.; Eisenberg, D.; Rivier, J.; Sawchenko, P.; Vale, W.; Riek, R. Functional Amyloids As Natural Storage of Peptide Hormones in Pituitary Secretory Granules. *Science* **2009**, *325*, 328–332.

[25] Li, J.; McQuade, T.; Siemer, A. B.; Napetschnig, J.; Moriwaki, K.; Hsiao, Y.-S.; Damko, E.; Moquin, D.; Walz, T.; McDermott, A.; Chan, F. K.-M.; Wu, H. The RIP1/RIP3 Necrosome Forms a Functional Amyloid Signaling Complex Required for Programmed Necrosis. *Cell* **2012**, *150*, 339–350.

[26] Majumdar, A.; Cesario, W. C.; White-Grindley, E.; Jiang, H.; Ren, F.; Khan, M. R.; Li, L.; Choi, E. M.-L.; Kannan, K.; Guo, F.; Unruh, J.; Slaughter, B.; Si, K. Critical Role of Amyloid-like Oligomers of *Drosophila* Orb2 in the Persistence of Memory. *Cell* **2012**, *148*, 515–529.

[27] Dai, B.; Li, D.; Xi, W.; Luo, F.; Zhang, X.; Zou, M.; Cao, M.; Hu, J.; Wang, W.; Wei, G.; Zhang, Y.; Liu, C. Tunable Assembly of Amyloid-Forming Peptides into Nanosheets as a Retrovirus Carrier. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 2996–3001.

[28] Knowles, T. P. J.; Oppenheim, T. W.; Buell, A. K.; Chirgadze, D. Y.; Welland, M. E. Nanostructured Films from Hierarchical Self-Assembly of Amyloidogenic Proteins. *Nat. Nanotechnol.* **2010**, *5*, 204–207.

[29] Rufo, C. M.; Moroz, Y. S.; Moroz, O. V.; Stöhr, J.; Smith, T. A.; Hu, X.; DeGrado, W. F.; Korendovych, I. V. Short Peptides Self-Assemble to Produce Catalytic Amyloids. *Nat. Chem.* **2014**, *6*, 303–309.

[30] Maji, S. K.; Schubert, D.; Rivier, C.; Lee, S.; Rivier, J. E.; Riek, R. Amyloid As a Depot for the Formulation of Long-Acting Drugs. *PLoS Biol.* **2008**, *6*, e17.

[31] Roberts, C. J. Therapeutic Protein Aggregation: Mechanisms, Design, and Control. *Trends Biotechnol.* **2014**, *32*, 372–380.

[32] Baldwin, A. J.; Bader, R.; Christodoulou, J.; MacPhee, C. E.; Dobson, C. M.; Barker, P. D. Cytochrome Display on Amyloid Fibrils. *J. Am. Chem. Soc.* **2006**, *128*, 2162–2163.

[33] Li, D.; Furukawa, H.; Deng, H.; Liu, C.; Yaghi, O. M.; Eisenberg, D. S. Designed Amyloid Fibers As Materials for Selective Carbon Dioxide Capture. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 191–196.

[34] Astbury, W. T.; Dickinson, S.; Bailey, K. The X-ray Interpretation of Denaturation and the Structure of the Seed Globulins. *Biochem J.* **1935**, *29*, 2351–2360.

[35] Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the Cross-β Spine of Amyloid-Like Fibrils. *Nature* **2005**, *435*, 773–778.

[36] Lührs, T.; Ritter, C.; Adrian, M.; Riek-Loher, D.; Bohrmann, B.; Döbeli, H.; Schubert, D.; Riek, R. 3D Structure of Alzheimer's Amyloid-β(1–42) Fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17342–17347.

[37] Antzutkin, O. N.; Leapman, R. D.; Balbach, J. J.; ; Tycko, R. Supramolecular Structural Constraints on Alzheimer's β-Amyloid Fibrils from Electron Microscopy and Solid-State Nuclear Magnetic Resonance. *Biochemistry* **2002**, *41*, 15436–15450.

[38] Nagel-Steger, L.; Owen, M. C.; Strodel, B. An Account of Amyloid Oligomers: Facts and Figures Obtained from Experiments and Simulations. *ChemBioChem* **2016**, *17*, 657–676.

[39] Schmitz, C.; Rutten, B. P. F.; Pielen, A.; Schäfer, S.; Wirths, O.; Tremp, G.; Czech, C.; Blanchard, V.; Multhaup, G.; Rezaie, P.; Korr, H.; Steinbusch, H. W. M.; Pradier, L.; Bayer, T. A. Hippocampal Neuron

Loss Exceeds Amyloid Plaque Load in a Transgenic Mouse Model of Alzheimer's Disease. *Am. J. Pathol.* **2004**, *164*, 1495–1502.

[40] McLean, C. A.; Cherny, R. A.; Fraser, F. W.; Fuller, S. J.; Smith, M. J.; Vbeyreuther, K.; Bush, A. I.; Masters, C. L. Soluble Pool of Aβ Amyloid as a Determinant of Severity of Neurodegeneration in Alzheimer's Disease. *Ann. Neurol.* **1999**, *46*, 860–866.

[41] McDonald, J. M.; Savva, G. M.; Brayne, C.; Welzel, A. T.; Forster, G.; Shankar, G. M.; Selkoe, D. J.; Ince, P. G.; Walsh, D. M. The Presence of Sodium Dodecyl Sulphate-Stable Aβ Dimers is Strongly Associated with Alzheimer-Type Dementia. *Brain* **2010**, *133*, 1328–1341.

[42] Hayden, E.; Teplow, D. Amyloid β-Protein Oligomers and Alzheimer's Disease. *Alzheimers Res. Ther.* **2013**, *5*, 60.

[43] Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B. Amyloid β-Protein (Aβ) Assembly: Aβ40 and Aβ42 Oligomerize Through Distinct Pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 330–335.

[44] Ono, K.; Condron, M. M.; Teplow, D. B. Structure-Neurotoxicity Relationships of Amyloid β-Protein Oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 14745–14750.

[45] Yu, L. et al. Structural Characterization of a Soluble Amyloid β-Peptide Oligomer. *Biochemistry* **2009**, *48*, 1870–1877.

[46] Tay, W. M.; Huang, D.; Rosenberry, T. L.; Paravastu, A. K. The Alzheimer's Amyloid-β(1–42) Peptide Forms Off-Pathway Oligomers and Fibrils That are Distinguished Structurally by Intermolecular Organization. *J. Mol. Biol.* **2013**, *425*, 2494–2508.

[47] Chimon, S.; Shaibat, M. A.; Jones, C. R.; Calero, D. C.; Aizezi, B.; Ishii, Y. Evidence of Fibril-like β-Sheet Structures in a Neurotoxic Amyloid Intermediate of Alzheimer's β-Amyloid. *Nat. Struct. Mol. Biol.* **2007**, *14*, 1157–1164.

[48] Kayed, R.; Bernhagen, J.; Greenfield, N.; Sweimeh, K.; Brunner, H.; Voelter, W.; Kapurniotu, A. Conformational Transitions of Islet Amyloid Polypeptide (IAPP) in Amyloid Formation in Vitro. *J. Mol. Biol.* **1999**, *287*, 781–796.

[49] Buchanan, L. E.; Dunkelberger, E. B.; Tran, H. Q.; Cheng, P.-N.; Chiu, C.-C.; Cao, P.; Raleigh, D. P.; de Pablo, J. J.; Nowick, J. S.; Zanni, M. T. Mechanism of IAPP Amyloid Fibril Formation Involves an Intermediate with a Transient β-Sheet. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 19285–19290.

[50] Grabenauer, M.; Wu, C.; Soto, P.; Shea, J.-E.; Bowers, M. T. Oligomers of the Prion Protein Fragment 106–126 are Likely Assembled from β-Hairpins in Solution, and Methionine Oxidation Inhibits Assembly without Altering the Peptide's Monomeric Conformation. *J. Am. Chem. Soc.* **2010**, *132*, 532–539.

[51] Ahmed, M.; Davis, J.; Aucoin, D.; Sato, T.; Ahuja, S.; Aimoto, S.; Elliott, J. I.; Van Nostrand, W. E.; Smith, S. O. Structural Conversion of Neurotoxic Amyloid-β_{1-42} Oligomers to Fibrils. *Nat. Struct. Mol. Biol.* **2010**, *17*, 561–567.

[52] Dror, R. O.; Dirks, R. M.; Grossman, J.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annu. Rev. Biophys.* **2012**, *41*, 429–452.

[53] Thirumalai, D.; Klimov, D.; Dima, R. Emerging Ideas on the Molecular Basis of Protein and Peptide Aggregation. *Curr. Opin. Struct. Biol.* **2003**, *13*, 146–159.

[54] Straub, J. E.; Thirumalai, D. Principles Governing Oligomer Formation in Amyloidogenic Peptides. *Curr. Opin. Struct. Biol.* **2010**, *20*, 187–195.

[55] Wu, C.; Shea, J.-E. Coarse-Grained Models for Protein Aggregation. *Curr. Opin. Struct. Biol.* **2011**, *21*, 209–220.

[56] Straub, J. E.; Thirumalai, D. Toward a Molecular Theory of Early and Late Events in Monomer to Amyloid Fibril Formation. *Annu. Rev. Phys. Chem.* **2011**, *62*, 437–463.

[57] Morriss-Andrews, A.; Shea, J.-E. Simulations of Protein Aggregation: Insights from Atomistic and Coarse-Grained Models. *J. Phys. Chem. Lett.* **2014**, *5*, 1899–1908.

[58] Morriss-Andrews, A.; Shea, J.-E. Computational Studies of Protein Aggregation: Methods and Applications. *Annu. Rev. Phys. Chem.* **2015**, *66*, 643–666.

[59] Berhanu, W. M.; Alred, E. J.; Bernhardt, N. A.; Hansmann, U. H. All-atom Simulation of Amyloid Aggregates. *Phys. Procedia* **2015**, *68*, 61–68.

[60] Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. Monomer Adds to Preformed Structured Oligomers of Aβ-Peptides by a Two-Stage Dock-Lock Mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 111–116.

[61] Reddy, G.; Straub, J. E.; Thirumalai, D. Dynamics of Locking of Peptides onto Growing Amyloid Fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 11948–11953.

[62] Schor, M.; Mey, A. S. J. S.; Noé, F.; MacPhee, C. E. Shedding Light on the Dock-Lock Mechanism in Amyloid Fibril Growth Using Markov State Models. *J. Phys. Chem. Lett.* **2015**, *6*, 1076–1081.

[63] Berhanu, W. M.; Alred, E. J.; Hansmann, U. H. E. Stability of Osaka Mutant and Wild-Type Fibril Models. *J. Phys. Chem. B* **2015**, *119*, 13063–13070.

[64] Alred, E. J.; Scheele, E. G.; Berhanu, W. M.; Hansmann, U. H. E. Stability of Iowa Mutant and Wild Type Aβ-Peptide Aggregates. *J. Chem. Phys.* **2014**, *141*, 175101.

[65] Barz, B.; Strodel, B. Understanding Amyloid-β Oligomerization at the Molecular Level: The Role of the Fibril Surface. *Chem. – Eur. J.* **2016**, *22*, 8768–8772.

[66] Mousseau, N.; Derreumaux, P. Exploring the Early Steps of Amyloid Peptide Aggregation by Computers. *Acc. Chem. Res.* **2005**, *38*, 885–891.

[67] Chebaro, Y.; Pasquali, S.; Derreumaux, P. The Coarse-Grained OPEP Force Field for Non-Amyloid and Amyloid Proteins. *J. Phys. Chem. B* **2012**, *116*, 8741–8752.

[68] Bellesia, G.; Shea, J.-E. Self-Assembly of β-Sheet Forming Peptides into Chiral Fibrillar Aggregates. *J. Chem. Phys.* **2007**, *126*, 245104.

[69] Osborne, K. L.; Bachmann, M.; Strodel, B. Thermodynamic Analysis of Structural Transitions During GNNQQNY Aggregation. *Proteins: Struct., Funct., Bioinf.* **2013**, *81*, 1141–1155.

[70] Chiricotto, M.; Melchionna, S.; Derreumaux, P.; Sterpone, F. Hydrodynamic Effects on β-Amyloid (16-22) Peptide Aggregation. *J. Chem. Phys.* **2016**, *145*, 035102.

[71] Auer, S.; Dobson, C. M.; Vendruscolo, M. Characterization of the Nucleation Barriers for Protein Aggregation and Amyloid Formation. *HFSP J.* **2007**, *1*, 137–146.

[72] Bieler, N. S.; Knowles, T. P. J.; Frenkel, D.; Vácha, R. Connecting Macroscopic Observables and Microscopic Assembly Events in Amyloid Formation Using Coarse Grained Simulations. *PLoS. Comput. Biol.* **2012**, *8*, e1002692.

[73] Irbäck, A.; Jónsson, S. A.; Linnemann, N.; Linse, B.; Wallin, S. Aggregate Geometry in Amyloid Fibril Nucleation. *Phys. Rev. Lett.* **2013**, *110*, 058101.

[74] Linse, B.; Linse, S. Monte Carlo Simulations of Protein Amyloid Formation Reveal Origin of Sigmoidal Aggregation Kinetics. *Mol. BioSyst.* **2011**, *7*, 2296–2303.

[75] Šarić, A.; Chebaro, Y. C.; Knowles, T. P. J.; Frenkel, D. Crucial Role of Nonspecific Interactions in Amyloid Nucleation. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 17869–17874.

[76] Latshaw, D. C.; Cheon, M.; Hall, C. K. Effects of Macromolecular Crowding on Amyloid Beta (16–22) Aggregation Using Coarse-Grained Simulations. *J. Phys. Chem. B* **2014**, *118*, 13513–13526.

[77] Auer, S.; Kashchiev, D. Phase Diagram of α-Helical and β-Sheet Forming Peptides. *Phys. Rev. Lett.* **2010**, *104*, 168105.

[78] Klimov, D. K.; Thirumalai, D. Dissecting the Assembly of Aβ_{16-22} Amyloid Peptides into Antiparallel β Sheets. *Structure* **2003**, *11*, 295–307.

[79] Riccardi, L.; Nguyen, P. H.; Stock, G. Construction of the Free Energy Landscape of Peptide Aggregation from Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1471–1479.

[80] Matthes, D.; Gapsys, V.; Daebel, V.; de Groot, B. L. Mapping the Conformational Dynamics and Pathways of Spontaneous Steric Zipper Peptide Oligomerization. *PLoS One* **2011**, *6*, e19129.

[81] Matthes, D.; Gapsys, V.; de Groot, B. L. Driving Forces and Structural Determinants of Steric Zipper Peptide Oligomer Formation Elucidated by Atomistic Simulations. *J. Mol. Biol.* **2012**, *421*, 390–416.

[82] Barz, B.; Wales, D. J.; Strodel, B. A Kinetic Approach to the Sequence-Aggregation Relationship in Disease-Related Protein Assembly. *J. Phys. Chem. B* **2014**, *118*, 1003–1011.

[83] Shaw, D. E. et al. Anton, a Special-Purpose Machine for Molecular Dynamics Simulation. *Commun. ACM* **2008**, *51*, 91–97.

[84] Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

[85] Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.

[86] Friedrichs, M. S.; Eastman, P.; Vaidyanathan, V.; Houston, M.; Legrand, S.; Beberg, A. L.; Ensign, D. L.; Bruns, C. M.; Pande, V. S. Accelerating Molecular Dynamic Simulation on Graphics Processing Units. *J. Comput. Chem.* **2009**, *30*, 864–872.

[87] Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. *J. Chem. Theory Comput.* **2009**, *5*, 1632–1639.

[88] Abrams, C.; Bussi, G. Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. *Entropy* **2014**, *16*, 163–199.

[89] Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

[90] Bellesia, G.; Shea, J.-E. What Determines the Structure and Stability of KFFE Monomers, Dimers, and Protofibrils? *Biophys. J.* **2009**, *96*, 875–886.

[91] Rivera, E.; Straub, J.; Thirumalai, D. Sequence and Crowding Effects in the Aggregation of a 10-Residue Fragment Derived from Islet Amyloid Polypeptide. *Biophys. J.* **2009**, *96*, 4552–4560.

[92] Wei, G.; Jewett, A. I.; Shea, J.-E. Structural Diversity of Dimers of the Alzheimer Amyloid-β(25-35) Peptide and Polymorphism of the Resulting Fibrils. *Phys. Chem. Chem. Phys.* **2010**, *12*, 3622–3629.

[93] Larini, L.; Shea, J.-E. Role of β-Hairpin Formation in Aggregation: The Self-Assembly of the Amyloid-β(25-35) Peptide. *Biophys. J.* **2012**, *103*, 576–586.

[94] Nguyen, P. H.; Derreumaux, P. Conformational Ensemble and Polymorphism of the All-Atom Alzheimer's Aβ_{37-42} Amyloid Peptide Oligomers. *J. Phys. Chem. B* **2013**, *117*, 5831–5840.

[95] Zhang, T.; Zhang, J.; Derreumaux, P.; Mu, Y. Molecular Mechanism of the Inhibition of EGCG on the Alzheimer Aβ_{1-42} Dimer. *J. Phys. Chem. B* **2013**, *117*, 3993–4002.

[96] Zhang, T.; Xu, W.; Mu, Y.; Derreumaux, P. Atomic and Dynamic Insights into the Beneficial Effect of the 1,4-Naphthoquinon-2-yl-l-tryptophan Inhibitor on Alzheimer's Aβ1-42 Dimer in Terms of Aggregation and Toxicity. *ACS Chem. Neurosc.* **2014**, *5*, 148–159.

[97] Laio, A.; Parrinello, M. Escaping Free-Energy Minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.

[98] Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826–843.

[99] Piana, S.; Laio, A. A Bias-Exchange Approach to Protein Folding. *J. Phys. Chem. B* **2007**, *111*, 4553–4559.

[100] Baftizadeh, F.; Biarnés, X.; Pietrucci, F.; Affinito, F.; Laio, A. Multidimensional View of Amyloid Fibril Nucleation in Atomistic Detail. *J. Am. Chem. Soc.* **2012**, *134*, 3886–3894.

[101] Baftizadeh, F.; Pietrucci, F.; Biarnés, X.; Laio, A. Nucleation Process of a Fibril Precursor in the C-Terminal Segment of Amyloid-β. *Phys. Rev. Lett.* **2013**, *110*, 168103.

[102] Barducci, A.; Bonomi, M.; Prakash, M. K.; Parrinello, M. Free-Energy Landscape of Protein Oligomerization from Atomistic Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, E4708–E4713.

[103] Itoh, S. G.; Okumura, H. Oligomer Formation of Amyloid-β(29-42) from Its Monomers Using the Hamiltonian Replica-Permutation Molecular Dynamics Simulation. *J. Phys. Chem. B* **2016**, *120*, 6555–6561.

[104] Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*; Academic Press: San Diego, USA, 2002.

[105] Tuckerman, M. E. *Statistical mechanics: Theory and Molecular Simulation*; Oxford Graduate Texts: Oxford, UK, 2010.

[106] Levitt, M.; Warshel, A. Computer Simulation of Protein Folding. *Nature* **1975**, *253*, 694–698.

[107] Warshel, A. Bicycle-Pedal Model for the First Step in the Vision Process. *Nature* **1976**, *260*, 679–683.

[108] Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Mol. Biol.* **2002**, *9*, 646–652.

[109] Adcock, S. A.; ; McCammon, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chem. Rev.* **2006**, *106*, 1589–1615.

[110] Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

[111] Bowers, K. J.; Chow, D. E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. SC 2006 Conference, Proceedings of the ACM/IEEE. 2006; p 43.

[112] Plimpton, S. Fast Parallel Algorithms for Short-Range Molecular Dynamics. *J. Comput. Phys.* **1995**, *117*, 1–19.

[113] Newton, I. *Philosophiae Naturalis Principia Mathematica*; Royal Society: London, UK, 1687.

[114] van der Spoel, D.; Lindahl, E.; Hess, B.; and the GROMACS development team, GROMACS User Manual version 4.6.7, www.gromacs.org. 2014.

[115] Hockney, R.; Goel, S.; Eastwood, J. Quiet High-Resolution Computer Models of a Plasma. *J. Comput. Phys.* **1974**, *14*, 148–158.

[116] Mackerell, A. D. Empirical Force Fields for Biological Macromolecules: Overview and Issues. *J. Comput. Chem.* **2004**, *25*, 1584–1604.

[117] Brooks, B. R. et al. CHARMM: The Biomolecular Simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1614.

[118] Jorgensen, W. L.; Tirado-Rives, J. The OPLS Potential Functions for Proteins. Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

[119] Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.

[120] Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: The GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

[121] Scherer, J. R.; Overend, J. The application of the Urey-Bradley force field to the in-plane vibrations of benzene. *Spectrochim. Acta* **1961**, *17*, 719–730.

[122] MacKerell, A. D.; Feig, M.; Brooks III, C. L. Improved Treatment of the Protein Backbone in Empirical Force Fields. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.

[123] Allinger, N. L.; Yuh, Y. H.; Lii, J. H. Molecular Mechanics. The MM3 Force Field for Hydrocarbons. 1. *J. Am. Chem. Soc.* **1989**, *111*, 8551–8566.

[124] van Duin, A. C. T.; Dasgupta, S.; Lorant, F.; ; Goddard III, W. A. ReaxFF: A Reactive Force Field for Hydrocarbons. *J. Phys. Chem. A* **2001**, *105*, 9396–9409.

[125] Baker, C. M. Polarizable Force Fields for Molecular Dynamics Simulations of Biomolecules. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2015**, *5*, 241–254.

[126] Lindorff-Larsen, K.; Maragakis, P.; Piana, S.; Eastwood, M. P.; Dror, R. O.; Shaw, D. E. Systematic Validation of Protein Force Fields Against Experimental Data. *PLoS One* **2012**, *7*, e32131.

[127] Beauchamp, K. A.; Lin, Y.-S.; Das, R.; Pande, V. S. Are Protein Force Fields Getting Better? A Systematic Benchmark on 524 Diverse NMR Measurements. *J. Chem. Theory Comput.* **2012**, *8*, 1409–1414.

[128] Roux, B.; Simonson, T. Implicit Solvent Models. *Biophys. Chem.* **1999**, *78*, 1–20.

[129] Barz, B.; Olubiyi, O. O.; Strodel, B. Early Amyloid β-Protein Aggregation Precedes Conformational Change. *Chem. Commun.* **2014**, 5373–5375.

[130] Zhou, R.; Berne, B. J. Can a Continuum Solvent Model Reproduce the Free Energy Landscape of a β-Hairpin Folding in Water? *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777–12782.

[131] Zhou, R. Free Energy Landscape of Protein Folding in Water: Explicit Vs. Implicit Solvent. *Proteins: Struct., Funct., Bioinf.* **2003**, *53*, 148–161.

[132] Nymeyer, H.; Garca, A. E. Simulation of the Folding Equilibrium of α-Helical Peptides: A Comparison of the Generalized Born Approximation with Explicit Solvent. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13934–13939.

[133] Tan, C.; Yang, L.; ; Luo, R. How Well Does Poisson-Boltzmann Implicit Solvent Agree with Explicit Solvent? A Quantitative Analysis. *J. Phys. Chem. B* **2006**, *110*, 18680–18687.

[134] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.

[135] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, Netherlands, 1981; pp 331–342.

[136] Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

[137] Mahoney, M. W.; Jorgensen, W. L. A Five-Site Model for Liquid Water and the Reproduction of the Density Anomaly by Rigid, Nonpolarizable Potential Functions. *J. Chem. Phys.* **2000**, *112*, 8910–8922.

[138] Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an Improved Four-Site Water Model for Biomolecular Simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.

[139] Abascal, J. L. F.; Vega, C. A General Purpose Model for the Condensed Phases of Water: TIP4P/2005. *J. Chem. Phys.* **2005**, *123*, 234505.

[140] Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

[141] Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

[142] Hess, B. P-LINCS: A Parallel Linear Constraint Solver for Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.

[143] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

[144] Miyamoto, S.; Kollman, P. A. Settle: An Analytical Version of the SHAKE and RATTLE Algorithm for Rigid Water Models. *J. Comput. Chem.* **1992**, *13*, 952–962.

[145] Feenstra, K. A.; Hess, B.; Berendsen, H. J. C. Improving Efficiency of Large Time-Scale Molecular Dynamics Simulations of Hydrogen-Rich Systems. *J. Comput. Chem.* **1999**, *20*, 786–798.

[146] Feller, S. E.; ; Pastor, R. W.; Rojnuckarin, A.; Bogusz, S.; Brooks, B. R. Effect of Electrostatic Force Truncation on Interfacial and Transport Properties of Water. *J. Phys. Chem.* **1996**, *100*, 17011–17020.

[147] Piana, S.; Lindorff-Larsen, K.; Dirks, R. M.; Salmon, J. K.; Dror, R. O.; Shaw, D. E. Evaluating the Effects of Cutoffs and Treatment of Long-Range Electrostatics in Protein Folding Simulations. *PLoS One* **2012**, *7*, e39918.

[148] Lange, O. F.; van der Spoel, D.; de Groot, B. L. Scrutinizing Molecular Mechanics Force Fields on the Submicrosecond Timescale with NMR Data. *Biophys. J.* **2010**, *99*, 647–655.

[149] Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921**, *64*, 253–287.

[150] Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An $N \cdot \log(N)$ Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.

[151] Shirts, M. R.; Mobley, D. L.; Chodera, J. D.; ; Pande, V. S. Accurate and Efficient Corrections for Missing Dispersion Interactions in Molecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 13052–13063.

[152] in't Veld, P. J.; Ismail, A. E.; Grest, G. S. Application of Ewald Summations to Long-Range Dispersion Forces. *J. Chem. Phys.* **2007**, *127*, 144711.

[153] Isele-Holder, R. E.; Mitchell, W.; Ismail, A. E. Development and Application of a Particle-Particle Particle-Mesh Ewald Method for Dispersion Interactions. *J. Chem. Phys.* **2012**, *137*, 174107.

[154] Isele-Holder, R. E.; Mitchell, W.; Hammond, J. R.; Kohlmeyer, A.; Ismail, A. E. Reconsidering Dispersion Potentials: Reduced Cutoffs in Mesh-Based Ewald Solvers Can Be Faster Than Truncation. *J. Chem. Theory Comput.* **2013**, *9*, 5412–5420.

[155] Isele-Holder, R. E.; Berkels, B.; Ismail, A. E. Smoothing of Contact Lines in Spreading Droplets by Trisiloxane Surfactants and its Relevance for Superspreading. *Soft Matter* **2015**, *11*, 4527–4539.

[156] Isele-Holder, R. E.; Ismail, A. E. Requirements for the Formation and Shape of Microscopic Precursors in Droplet Spreading. *Langmuir* **2016**, *32*, 4472–4478.

[157] Nosé, S. A Unified Formulation of the Constant Temperature Molecular Dynamics Methods. *J. Chem. Phys.* **1984**, *81*, 511–519.

[158] Hoover, W. G. Canonical Dynamics: Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697.

[159] Lingenheil, M.; Denschlag, R.; Reichold, R.; Tavan, P. The "Hot-Solvent/Cold-Solute" Problem Revisited. *J. Chem. Theory Comput.* **2008**, *4*, 1293–1306.

[160] Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: A New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

[161] Nosé, S.; Klein, M. Constant Pressure Molecular Dynamics for Molecular Systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

[162] Pham, C. L.; Kwan, A. H.; Sunde, M. Functional Amyloid: Widespread in Nature, Diverse in Purpose. *Essays Biochem.* **2014**, *56*, 207–219.

[163] Maggio, J. E. Tachykinins. *Annu. Rev. Neurosci.* **1988**, *11*, 13–28.

[164] Singh, P. K.; Maji, S. K. Amyloid-like Fibril Formation by Tachykinin Neuropeptides and Its Relevance to Amyloid β-Protein Aggregation and Toxicity. *Cell Biochem. Biophys.* **2012**, *64*, 29–44.

[165] Pike, C. J.; Walencewicz-Wasserman, A. J.; Kosmoski, J.; Cribbs, D. H.; Glabe, C. G.; Cotman, C. W. Structure-Activity Analyses of β-Amyloid Peptides: Contributions of the β25-35 Region to Aggregation and Neurotoxicity. *J. Neurochem.* **1995**, *64*, 253–265.

[166] Naldi, M.; Fiori, J.; Pistolozzi, M.; Drake, A. F.; Bertucci, C.; Wu, R.; Mlynarczyk, K.; Filipek, S.; De Simone, A.; Andrisano, V. Amyloid β-Peptide 25-35 Self-Assembly and Its Inhibition: A Model Undecapeptide System to Gain Atomistic and Secondary Structure Details of the Alzheimer's Disease Process and Treatment. *ACS Chem. Neurosci.* **2012**, *3*, 952–962.

[167] Yankner, B.; Duffy, L.; Kirschner, D. Neurotrophic and Neurotoxic Effects of Amyloid β Protein: Reversal by Tachykinin Neuropeptides. *Science* **1990**, *250*, 279–282.

[168] Shearman, M. Cellular MTT Reduction Distinguishes the Mechanism of Action of β-Amyloid from That of Tachykinin Receptor Peptides. *Neuropeptides* **1996**, *30*, 125–132.

[169] Bleiholder, C.; Do, T. D.; Wu, C.; Economou, N. J.; Bernstein, S. S.; Buratto, S. K.; Shea, J.-E.; Bowers, M. T. Ion Mobility Spectrometry Reveals the Mechanism of Amyloid Formation of Aβ(25-35) and Its Modulation by Inhibitors at the Molecular Level: Epigallocatechin Gallate and Scyllo-Inositol. *J. Am. Chem. Soc.* **2013**, *135*, 16926–16937.

[170] Flashner, E.; Raviv, U.; Friedler, A. The Effect of Tachykinin Neuropeptides on Amyloid β Aggregation. *Biochem. Biophys. Res. Commun.* **2011**, *407*, 13–17.

[171] Pieri, M.; Amadoro, G.; Carunchio, I.; Ciotti, M.; Quaresima, S.; Florenzano, F.; Calissano, P.; Possenti, R.; Zona, C.; Severini, C. SP Protects Cerebellar Granule Cells Against β-Amyloid-Induced Apoptosis by Down-Regulation and Reduced Activity of Kv4 Potassium Channels. *Neuropharmacol.* **2010**, *58*, 268–276.

[172] Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

[173] D'Ursi, A. M.; Armenante, M. R.; Guerrini, R.; Salvadori, S.; Sorrentino, G.; Picone, D. Solution Structure of Amyloid β-Peptide (25-35) in Different Media. *J. Med. Chem.* **2004**, *47*, 4231–4238.

[174] Grace, R. C. R.; Lynn, A. M.; Cowsik, S. M. Lipid Induced Conformation of the Tachykinin Peptide Kassinin. *J. Biomol. Struct. Dyn.* **2001**, *18*, 611–625.

[175] Mantha, A. K.; Chandrashekar, I. R.; Baquer, N. Z.; Cowsik, S. M. Three Dimensional Structure of Mammalian Tachykinin Peptide Neurokinin B Bound to Lipid Micelles. *J. Biomol. Struct. Dyn.* **2004**, *22*, 137–147.

[176] Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.

[177] Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O. MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319–2327.

[178] Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed.* **1999**, *38*, 236–240.

[179] Eisenhaber, F.; Lijnzaad, P.; Argos, P.; Sander, C.; Scharf, M. The Double Cubic Lattice Method: Efficient Approaches to Numerical Integration of Surface Area and Volume and to Dot Surface Contouring of Molecular Assemblies. *J. Comput. Chem.* **1995**, *16*, 273–284.

[180] Kabsch, W.; Sander, C. Dictionary of Protein Secondary Structure: Pattern Recognition of Hydrogen-Bonded and Geometrical Features. *Biopolymers* **1983**, *22*, 2577–2637.

[181] Dima, R. I.; Thirumalai, D. Asymmetry in the Shapes of Folded and Denatured States of Proteins. *J. Phys. Chem. B* **2004**, *108*, 6564–6570.

[182] Campioni, S.; Mannini, B.; Zampagni, M.; Pensalfini, A.; Parrini, C.; Evangelisti, E.; Relini, A.; Stefani, M.; Dobson, C. M.; Cecchi, C.; Chiti, F. A Causative Link Between the Structure of Aberrant Protein Oligomers and Their Toxicity. *Nat. Chem. Biol.* **2010**, *6*, 140–147.

[183] Bolognesi, B.; Kumita, J. R.; Barros, T. P.; Esbjorner, E. K.; Luheshi, L. M.; Crowther, D. C.; Wilson, M. R.; Dobson, C. M.; Favrin, G.; Yerbury, J. J. ANS Binding Reveals Common Features of Cytotoxic Amyloid Species. *ACS Chem. Biol.* **2010**, *5*, 735–740.

[184] Krishnan, R.; Goodman, J. L.; Mukhopadhyay, S.; Pacheco, C. D.; Lemke, E. A.; Deniz, A. A.; Lindquist, S. Conserved Features of Intermediates in Amyloid Assembly Determine Their Benign or Toxic States. *Proc. Natl. Acad. Sci. U.S.A.* **2012**,

[185] Ladiwala, A. R. A.; Litt, J.; Kane, R. S.; Aucoin, D. S.; Smith, S. O.; Ranjan, S.; Davis, J.; Van Nostrand, W. E.; Tessier, P. M. Conformational Differences Between Two Amyloid β Oligomers of Similar Size and Dissimilar Toxicity. *J. Biol. Chem.* **2012**, *287*, 24765–24773.

[186] Esteras-Chopo, A.; Serrano, L.; de la Paz, M. L. The Amyloid Stretch Hypothesis: Recruiting Proteins Toward the Dark Side. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 16672–16677.

[187] Merkler, D. J. C-Terminal Amidated Peptides: Production by the in vitro Enzymatic Amidation of Glycine-Extended Peptides and the Importance of the Amide to Bioactivity. *Enzyme Microb. Technol.* **1994**, *16*, 450–456.

[188] Patacchini, R.; Quartara, L.; Rovero, P.; Goso, C.; Maggi, C. A. Role of C-Terminal Amidation on the Biological Activity of Neurokinin A Derivatives with Agonist and Antagonist Properties. *J. Pharm. Exp. Ther.* **1993**, *264*, 17–21.

[189] Haspel, N.; Laurent, A. D.; Zanuy, D.; Nussinov, R.; Alemán, C.; Puiggalí, J.; Revilla-López, G. Conformational Exploration of Two Peptides and Their Hybrid Polymer Conjugates: Potentialities As Self-Aggregating Materials. *J. Phys. Chem. B* **2012**, *116*, 13941–13952.

[190] Hoyer, W.; Grönwall, C.; Jonsson, A.; Ståhl, S.; Härd, T. Stabilization of a β-Hairpin in Monomeric Alzheimer's Amyloid-β Peptide Inhibits Amyloid Formation. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5099–5104.

[191] Xu, W.; Zhang, C.; Morozova-Roche, L.; Zhang, J. Z. H.; Mu, Y. pH-Dependent Conformational Ensemble and Polymorphism of Amyloid-β Core Fragment. *J. Phys. Chem. B* **2013**, *117*, 8392–8399.

[192] Kusumoto, Y.; Lomakin, A.; Teplow, D. B.; Benedek, G. B. Temperature Dependence of Amyloid β-Protein Fibrillization. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 12277–12282.

[193] Klement, K.; Wieligmann, K.; Meinhardt, J.; Hortschansky, P.; Richter, W.; Fändrich, M. Effect of Different Salt Ions on the Propensity of Aggregation and on the Structure of Alzheimer's Aβ(1-40) Amyloid Fibrils. *J. Mol. Biol.* **2007**, *373*, 1321–1333.

[194] Campioni, S.; Mannini, B.; López-Alonso, J. P.; Shalova, I. N.; Penco, A.; Mulvihill, E.; Laurents, D. V.; Relini, A.; Chiti, F. Salt Anions Promote the Conversion of HypF-N into Amyloid-like Oligomers and Modulate the Structure of the Oligomers and the Monomeric Precursor State. *J. Mol. Biol.* **2012**, *424*, 132–149.

[195] Buell, A. K.; Hung, P.; Salvatella, X.; Welland, M. E.; Dobson, C. M.; Knowles, T. P. Electrostatic Effects in Filamentous Protein Aggregation. *Biophys. J.* **2013**, *104*, 1116–1126.

[196] Chatani, E.; Naiki, H.; Goto, Y. Seeding-Dependent Propagation and Maturation of β_2-Microglobulin Amyloid Fibrils Under High Pressure. *J. Mol. Biol.* **2006**, *359*, 1086–1096.

[197] Breydo, L.; Reddy, K. D.; Piai, A.; Felli, I. C.; Pierattelli, R.; Uversky, V. N. The Crowd You're in With: Effects of Different Types of Crowding Agents on Protein Aggregation. *Biochim. Biophys. Acta, Proteins Proteomics* **2014**, *1844*, 346–357.

[198] Friedman, R.; Caflisch, A. Wild Type and Mutants of the HET-s(218-289) Prion Show Different Flexibility at Fibrillar Ends: A Simulation Study. *Proteins: Struct., Funct., Bioinf.* **2014**, *82*, 399–404.

[199] Tian, P.; Boomsma, W.; Wang, Y.; Otzen, D. E.; Jensen, M. H.; Lindorff-Larsen, K. Structure of a Functional Amyloid Protein Subunit Computed Using Sequence Variation. *J. Am. Chem. Soc.* **2015**, *137*, 22–25.

[200] Tian, P.; Lindorff-Larsen, K.; Boomsma, W.; Jensen, M. H.; Otzen, D. E. A Monte Carlo Study of the Early Steps of Functional Amyloid Formation. *PLoS One* **2016**, *11*, e146096.

[201] Bowman, G.; Pande, V. S.; Noé (eds.), F. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*; Advances in Experimental Medicine and Biology; Springer, 2014.

[202] Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* **2011**, *134*, 174105.

[203] Chodera, J. D.; Noé, F. Markov State Models of Biomolecular Conformational Dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135–144.

[204] Henriques, J.; Cragnell, C.; Skepö, M. Molecular Dynamics Simulations of Intrinsically Disordered Proteins: Force Field Evaluation and Comparison with Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 3420–3431.

[205] Cino, E. A.; Choy, W.-Y.; Karttunen, M. Comparison of Secondary Structure Formation Using 10 Different Force Fields in Microsecond Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 2725–2740.

[206] Martín-García, F.; Papaleo, E.; Gomez-Puertas, P.; Boomsma, W.; Lindorff-Larsen, K. Comparing Molecular Dynamics Force Fields in the Essential Subspace. *PLoS One* **2015**, *10*, e121114.

[207] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Bioinf.* **2006**, *65*, 712–725.

[208] Best, R. B.; Buchete, N.-V.; Hummer, G. Are Current Molecular Dynamics Force Fields Too Helical? *Biophys. J.* **2008**, *95*, L07–L09.

[209] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. Improved Side-Chain Torsion Potentials for the Amber ff99SB Protein Force Field. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 1950–1958.

[210] Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. How Robust are Protein Folding Simulations with Respect to Force Field Parameterization? *Biophys. J.* **2011**, *100*, L47–L49.

[211] MacKerell, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

[212] Madan Babu, M.; van der Lee, R.; Sanchez de Groot, N.; Gsponer, J. Intrinsically disordered proteins: regulation and disease. *Curr. Opin. Struct. Biol.* **2011**, *21*, 432–440.

[213] Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. The Alzheimer's Peptides Aβ40 and 42 Adopt Distinct Conformations in Water: A Combined MD / NMR Study. *J. Mol. Biol.* **2007**, *368*, 1448–1457.

[214] Cino, E. A.; Wong-ekkabut, J.; Karttunen, M.; Choy, W.-Y. Microsecond Molecular Dynamics Simulations of Intrinsically Disordered Proteins Involved in the Oxidative Stress Response. *PLoS One* **2011**, *6*, e27371.

[215] Venken, T.; Voet, A.; Maeyer, M. D.; Fabritiis, G. D.; Sadiq, S. K. Rapid Conformational Fluctuations of Disordered HIV-1 Fusion Peptide in Solution. *J. Chem. Theory Comput.* **2013**, *9*, 2870–2874.

[216] Mittal, J.; Yoo, T. H.; Georgiou, G.; Truskett, T. M. Structural Ensemble of an Intrinsically Disordered Polypeptide. *J. Phys. Chem. B* **2013**, *117*, 118–124.

[217] Qiao, Q.; Bowman, G. R.; Huang, X. Dynamics of an Intrinsically Disordered Protein Reveal Metastable Conformations That Potentially Seed Aggregation. *J. Am. Chem. Soc.* **2013**, *135*, 16092–16101.

[218] Stanley, N.; Esteban-Martín, S.; De Fabritiis, G. Kinetic modulation of a disordered protein domain by phosphorylation. *Nat. Commun.* **2014**, *5*, 5272.

[219] Stanley, N.; Esteban-Martín, S.; Fabritiis, G. D. Progress in studying intrinsically disordered proteins with atomistic simulations. *Prog. Biophys. Mol. Biol.* **2015**, *119*, 47–52.

[220] Mercadante, D.; Milles, S.; Fuertes, G.; Svergun, D. I.; Lemke, E. A.; Gräter, F. Kirkwood-Buff Approach Rescues Overcollapse of a Disordered Protein in Canonical Protein Force Fields. *J. Phys. Chem. B* **2015**, *119*, 7975–7984.

[221] Palazzesi, F.; Prakash, M. K.; Bonomi, M.; Barducci, A. Accuracy of Current All-Atom Force-Fields in Modeling Protein Disordered States. *J. Chem. Theory Comput* **2015**, *11*, 2–7.

[222] Rauscher, S.; Gapsys, V.; Gajda, M. J.; Zweckstetter, M.; de Groot, B. L.; Grubmüller, H. Structural Ensembles of Intrinsically Disordered Proteins Depend Strongly on Force Field: A Comparison to Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 5513–5524.

[223] Hoffmann, K. Q.; McGovern, M.; Chiu, C.-c.; de Pablo, J. J. Secondary Structure of Rat and Human Amylin across Force Fields. *PLoS One* **2015**, *10*, e134091.

[224] Srivastava, A.; Balaji, P. V. Molecular Events During the Early Stages of Aggregation of GNNQQNY: An all Atom MD Simulation Study of Randomly Dispersed Peptides. *J. Struct. Biol.* **2015**, *192*, 376–391.

[225] Matthes, D.; Gapsys, V.; Brennecke, J. T.; de Groot, B. L. An Atomistic View of Amyloidogenic Self-assembly: Structure and Dynamics of Heterogeneous Conformational States in the Pre-nucleation Phase. *Sci. Rep.* **2016**, *6*, 33156.

[226] Petrov, D.; Zagrovic, B. Are Current Atomistic Force Fields Accurate Enough to Study Proteins in Crowded Environments? *PLoS Comput. Biol.* **2014**, *10*, e1003638.

[227] Best, R. B.; Zheng, W.; Mittal, J. Balanced Protein-Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association. *J. Chem. Theory Comput.* **2014**, *10*, 5113–5124.

[228] Abriata, L. A.; Dal Peraro, M. Assessing the Potential of Atomistic Molecular Dynamics Simulations to Probe Reversible Protein-Protein Recognition and Binding. *Sci. Rep.* **2015**, *5*, 10549.

[229] Nguyen, P. H.; Li, M. S.; Derreumaux, P. Effects of All-Atom Force Fields on Amyloid Oligomerization: Replica Exchange Molecular Dynamics Simulations of the Aβ_{16-22} Dimer and Trimer. *Phys. Chem. Chem. Phys.* **2011**, *13*, 9778–9788.

[230] Berhanu, W. M.; Hansmann, U. H. E. Side-Chain Hydrophobicity and the Stability of Aβ_{16-22} Aggregates. *Protein Sci.* **2012**, *21*, 1837–1848.

[231] Hou, L.; Shao, H.; Zhang, Y.; Li, H.; Menon, N. K.; Neuhaus, E. B.; Brewer, J. M.; Byeon, I.-J. L.; Ray, D. G.; Vitek, M. P.; Iwashita, T.; Makula, R. A.; Przybyla, A. B.; ; Zagorski, M. G. Solution NMR Studies of the Aβ(1-40) and Aβ(1-42) Peptides Establish that the Met35 Oxidation State Affects the Mechanism of Amyloid Formation. *J. Am. Chem. Soc.* **2004**, *126*, 1992–2005.

[232] Yan, Y.; McCallum, S. A.; Wang, C. M35 Oxidation Induces Aβ40-like Structural and Dynamical Changes in Aβ42. *J. Am. Chem. Soc.* **2008**, *130*, 5394–5395.

[233] Rosenman, D. J.; Connors, C. R.; Chen, W.; Wang, C.; García, A. E. Aβ Monomers Transiently Sample Oligomer and Fibril-Like Configurations: Ensemble Characterization Using a Combined MD/NMR Approach. *J. Mol. Biol.* **2013**, *425*, 3338–3359.

[234] Ball, K. A.; Phillips, A. H.; Nerenberg, P. S.; Fawzi, N. L.; Wemmer, D. E.; Head-Gordon, T. Homogeneous and Heterogeneous Tertiary Structure Ensembles of Amyloid-β Peptides. *Biochemistry* **2011**, *50*, 7612–7628.

[235] Ball, K. A.; Phillips, A. H.; Wemmer, D. E.; Head-Gordon, T. Differences in β-Strand Populations of Monomeric Aβ40 and Aβ42. *Biophys. J.* **2013**, *104*, 2714–2724.

[236] Anand, P.; Nandel, F. S.; Hansmann, U. H. E. The Alzheimer's β Amyloid (Aβ_{1-39}) Monomer in an Implicit Solvent. *J. Chem. Phys.* **2008**, *128*, 165102.

[237] Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z.; Wang, C.; Garcia, A. E. Atomic-Level Characterization of the Ensemble of the Aβ(1-42) Monomer in Water Using Unbiased Molecular Dynamics Simulations and Spectral Algorithms. *J. Mol. Biol.* **2011**, *405*, 570–583.

[238] Côté, S.; Derreumaux, P.; Mousseau, N. Distinct Morphologies for Amyloid Beta Protein Monomer: Aβ_{1-40}, Aβ_{1-42}, and Aβ_{1-40}(D23N). *J. Chem. Theory Comput.* **2011**, *7*, 2584–2592.

[239] Olubiyi, O. O.; Strodel, B. Structures of the Amyloid β-Peptides Aβ_{1-40} and Aβ_{1-42} as Influenced by pH and a d-Peptide. *J. Phys. Chem. B* **2012**, *116*, 3280–3291.

[240] Lin, Y.-S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S. Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophys. J.* **2012**, *102*, 315–324.

[241] Chebaro, Y.; Mousseau, N.; Derreumaux, P. Structures and Thermodynamics of Alzheimer's Amyloid-β Aβ(16-35) Monomer and Dimer by Replica Exchange Molecular Dynamics Simulations: Implication for Full-Length Aβ Fibrillation. *J. Phys. Chem. B* **2009**, *113*, 7668–7675.

[242] Das, P.; Murray, B.; Belfort, G. Alzheimer's Protective A2T Mutation Changes the Conformational Landscape of the Aβ_{1-42} Monomer Differently Than Does the A2V Mutation. *Biophys. J.* **2015**, *108*, 738–747.

[243] Nasica-Labouze, J. et al. Amyloid β Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies. *Chem. Rev.* **2015**, *115*, 3518–.3563.

[244] Rosenman, D. J.; Wang, C.; García, A. E. Characterization of Aβ Monomers through the Convergence of Ensemble Properties among Simulations with Multiple Force Fields. *J. Phys. Chem. B* **2016**, *120*, 259–277.

[245] Fisher, C. K.; Ullman, O.; Stultz, C. M. Comparative Studies of Disordered Proteins with Similar Sequences: Application to Aβ40 and Aβ42. *Biophys. J.* **2013**, *104*, 1546–1555.

[246] Granata, D.; Baftizadeh, F.; Habchi, J.; Galvagnion, C.; De Simone, A.; Camilloni, C.; Laio, A.; Vendruscolo, M. The Inverted Free Energy Landscape of an Intrinsically Disordered Peptide by Simulations and Experiments. *Sci. Rep.* **2015**, *5*, 15449.

[247] Mitternacht, S.; Staneva, I.; Härd, T.; Irbäck, A. Comparing the folding free-energy landscapes of Aβ42 variants with different aggregation properties. *Proteins: Struct., Funct., Bioinf.* **2010**, *78*, 2600–2608.

[248] Song, W.; Wang, Y.; Colletier, J.-P.; Yang, H.; Xu, Y. Varied Probability of Staying Collapsed/Extended at the Conformational Equilibrium of Monomeric Aβ_{40} and Aβ_{42}. *Sci. Rep.* **2015**, *5*, 11024.

[249] Somavarapu, A. K.; Kepp, K. P. The Dependence of Amyloid-β Dynamics on Protein Force Fields and Water Models. *ChemPhysChem* **2015**, *16*, 3278–3289.

[250] Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. Amyloid Fibril Formation by Aβ16-22, a Seven-Residue Fragment of the Alzheimer's β-Amyloid Peptide, and Structural Characterization by Solid State NMR. *Biochemistry* **2000**, *39*, 13748–13759.

[251] Petty, S. A.; Decatur, S. M. Experimental Evidence for the Reorganization of β-Strands within Aggregates of the Aβ(16-22) Peptide. *J. Am. Chem. Soc.* **2005**, *127*, 13488–13489.

[252] Senguen, F. T.; Doran, T. M.; Anderson, E. A.; Nilsson, B. L. Clarifying the Influence of Core Amino Acid Hydrophobicity, Secondary Structure

Propensity, and Molecular Volume on Amyloid-β 16-22 Self-Assembly. *Mol. BioSyst.* **2011**, *7*, 497–510.

[253] Best, R. B.; Hummer, G. Optimized Molecular Dynamics Force Fields Applied to the Helix-Coil Transition of Polypeptides. *J. Phys. Chem. B* **2009**, *113*, 9004–9015.

[254] Li, D.-W.; Brüschweiler, R. NMR-Based Protein Potentials. *Angew. Chem., Int. Ed.* **2010**, *49*, 6778–6780.

[255] Fawzi, N. L.; Phillips, A. H.; Ruscio, J. Z.; Doucleff, M.; Wemmer, D. E.; Head-Gordon, T. Structure and Dynamics of the Aβ_{21-30} Peptide from the Interplay of NMR Experiments and Molecular Simulations. *J. Am. Chem. Soc.* **2008**, *130*, 6145–6158.

[256] Nath, A.; Sammalkorpi, M.; DeWitt, D. C.; Trexler, A. J.; Elbaum-Garfinkle, S.; O'Hern, C. S.; Rhoades, E. The Conformational Ensembles of α-Synuclein and Tau: Combining Single-Molecule FRET and Simulations. *Biophys. J.* **2012**, *103*, 1940–1949.

[257] Hansmann, U. H. Parallel Tempering Algorithm for Conformational Studies of Biological Molecules. *Chem. Phys. Lett.* **1997**, *281*, 140–150.

[258] Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. Equation of State Calculations by Fast Computing Machines. *J. Chem. Phys.* **1953**, *21*, 1087–1092.

[259] Hastings, W. K. Monte Carlo Sampling Methods Using Markov Chains and Their Applications. *Biometrika* **1970**, *57*, 97–109.

[260] Chodera, J. D.; Shirts, M. R. Replica Exchange and Expanded Ensemble Simulations as Gibbs Sampling: Simple Improvements for Enhanced Mixing. *J. Chem. Phys.* **2011**, *135*.

[261] Crescenzi, O.; Tomaselli, S.; Guerrini, R.; Salvadori, S.; D'Ursi, A. M.; Temussi, P. A.; Picone, D. Solution Structure of the Alzheimer Amyloid β-Peptide (1-42) in an Apolar Microenvironment. *Eur. J. Biochem.* **2002**, *269*, 5642–5648.

[262] Shen, Y.; Bax, A. SPARTA+: A Modest Improvement in Empirical NMR Chemical Shift Prediction by Means of an Artificial Neural Network. *J. Biomol. NMR* **2010**, *48*, 13–22.

[263] Wishart, D. S.; Bigam, C. G.; Holm, A.; Hodges, R. S.; Sykes, B. D. 1H, 13C and 15N Random Coil NMR Chemical Shifts of the Common Amino Acids. I. Investigations of Nearest-Neighbor Effects. *J. Biomol. NMR* **1995**, *5*, 67–81.

[264] Karplus, M. Contact Electron-Spin Coupling of Nuclear Magnetic Moments. *J. Chem. Phys.* **1959**, *30*, 11–15.

[265] Vuister, G. W.; Bax, A. Quantitative *J* Correlation: a New Approach for Measuring Homonuclear Three-Bond J($H^N H^\alpha$) Coupling Constants in ^{15}N-Enriched Proteins. *J. Am. Chem. Soc.* **1993**, *115*, 7772–7777.

[266] Zweckstetter, M.; Bax, A. Prediction of Sterically Induced Alignment in a Dilute Liquid Crystalline Phase: Aid to Protein Structure Determination by NMR. *J. Am. Chem. Soc.* **2000**, *122*, 3791–3792.

[267] Zweckstetter, M. NMR: Prediction of Molecular Alignment from Structure Using the PALES Software. *Nat. Protocols* **2008**, *3*, 679–690.

[268] Hunter, J. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.

[269] Schmid, N.; Eichenberger, A.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A.; van Gunsteren, W. Definition and Testing of the GROMOS Force-Field Versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.

[270] Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects from Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6*, 459–466.

[271] Best, R. B.; Mittal, J. Protein Simulations with an Optimized Water Model: Cooperative Helix Formation and Temperature-Induced Unfolded State Collapse. *J. Phys. Chem. B* **2010**, *114*, 14916–14923.

[272] Liang, Y.; Lynn, D. G.; Berland, K. M. Direct Observation of Nucleation and Growth in Amyloid Self-Assembly. *J. Am. Chem. Soc.* **2010**, *132*, 6306–6308.

[273] Wennberg, C. L.; Murtola, T.; Hess, B.; Lindahl, E. Lennard-Jones Lattice Summation in Bilayer Simulations Has Critical Effects on Surface Tension and Lipid Properties. *J. Chem. Theory Comput.* **2013**, *9*, 3527–3537.

[274] Torrie, G.; Valleau, J. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.

[275] Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 932–942.

[276] Wong, K.-Y.; York, D. M. Exact Relation between Potential of Mean Force and Free-Energy Profile. *J. Chem. Theory Comput.* **2012**, *8*, 3998–4003.

[277] den Otter, W. K. Revisiting the Exact Relation between Potential of Mean Force and Free-Energy Profile. *J. Chem. Theory Comput.* **2013**, *9*, 3861–3865.

[278] Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. the Method. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

[279] Lemkul, J. A.; Bevan, D. R. Assessing the Stability of Alzheimer's Amyloid Protofibrils Using Molecular Dynamics. *J. Phys. Chem. B* **2010**, *114*, 1652–1660.

[280] Rao Jampani, S.; Mahmoudinobar, F.; Su, Z.; Dias, C. L. Thermodynamics of Aβ_{16-21} Dissociation from a Fibril: Enthalpy, Entropy, and Volumetric Properties. *Proteins: Struct., Funct., Bioinf.* **2015**, *83*, 1963–1972.

[281] Hub, J. S.; de Groot, B. L.; van der Spoel, D. g_wham – A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J. Chem. Theory Comput.* **2010**, *6*, 3713–3720.

[282] Shimizu, S.; Chan, H. S. Temperature Dependence of Hydrophobic Interactions: A Mean Force Perspective, Effects of Water Density, and Nonadditivity of Thermodynamic Signatures. *J. Chem. Phys.* **2000**, *113*, 4683–4700.

[283] Dias, C. L.; Chan, H. S. Pressure-Dependent Properties of Elementary Hydrophobic Interactions: Ramifications for Activation Properties of Protein Folding. *J. Phys. Chem. B* **2014**, *118*, 7488–7509.

[284] Nag, S.; Sarkar, B.; Bandyopadhyay, A.; Sahoo, B.; Sreenivasan, V. K. A.; Kombrabail, M.; Muralidharan, C.; Maiti, S. Nature of the

Amyloid-β Monomer and the Monomer-Oligomer Equilibrium. *J. Biol. Chem.* **2011**, *286*, 13827–13833.

[285] Lin, Y.-S.; Pande, V. S. Effects of Familial Mutations on the Monomer Structure of Aβ_{42}. *Biophys. J.* **2012**, *103*, L47–L49.

[286] Urbanc, B.; Cruz, L.; Yun, S.; Buldyrev, S. V.; Bitan, G.; Teplow, D. B.; Stanley, H. E. In silico study of amyloid β-protein folding and oligomerization. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 17345–17350.

[287] Yan, Y.; Wang, C. Aβ42 is More Rigid than Aβ40 at the C Terminus: Implications for Aβ Aggregation and Toxicity. *J. Mol. Biol.* **2006**, *364*, 853–862.

[288] Wu, C.; Murray, M. M.; Bernstein, S. L.; Condron, M. M.; Bitan, G.; Shea, J.-E.; Bowers, M. T. The Structure of Aβ42 C-Terminal Fragments Probed by a Combined Experimental and Theoretical Study. *J. Mol. Biol.* **2009**, *387*, 492–501.

[289] Senguen, F. T.; Lee, N. R.; Gu, X.; Ryan, D. M.; Doran, T. M.; Anderson, E. A.; Nilsson, B. L. Probing Aromatic, Hydrophobic, and Steric Effects on the Self-Assembly of an Amyloid-β Fragment Peptide. *Mol. BioSyst.* **2011**, *7*, 486–496.

[290] Broersen, K.; Rousseau, F.; Schymkowitz, J. The Culprit Behind Amyloid Beta Peptide Related Neurotoxicity in Alzheimer's Disease: Oligomer Size or Conformation? *Alzheimers Res. Ther.* **2010**, *2*, 12.

[291] Knowles, T. P. J.; Shu, W.; Devlin, G. L.; Meehan, S.; Auer, S.; Dobson, C. M.; Welland, M. E. Kinetics and Thermodynamics of Amyloid Formation from Direct Measurements of Fluctuations in Fibril Mass. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 10016–10021.

[292] Xue, W.-F.; Radford, S. E. An Imaging and Systems Modeling Approach to Fibril Breakage Enables Prediction of Amyloid Behavior. *Biophys. J.* **2013**, *105*, 2811–2819.

[293] Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C. Molecular Basis for Amyloid Fibril Formation and Stability. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 315–320.

[294] Bemporad, F.; Taddei, N.; Stefani, M.; Chiti, F. Assessing the Role of Aromatic Residues in the Amyloid Aggregation of Human Muscle Acylphosphatase. *Protein Sci.* **2006**, *15*, 862–870.

[295] Boyd-Kimball, D.; Abdul, H. M.; Reed, T.; Sultana, R.; Butterfield, D. A. Role of Phenylalanine 20 in Alzheimer's Amyloid β-Peptide (1-42)-Induced Oxidative Stress and Neurotoxicity. *Chem. Res. Toxicol.* **2004**, *17*, 1743–1749.

[296] Gazit, E. Mechanisms of Amyloid Fibril Self-Assembly and Inhibition. *FEBS J.* **2005**, *272*, 5971–5978.

[297] Rao, F.; Caflisch, A. The Protein Folding Network. *J. Mol. Biol.* **2004**, *342*, 299–306.

[298] Noé, F.; Krachtus, D.; Smith, J. C.; ; Fischer, S. Transition Networks for the Comprehensive Characterization of Complex Conformational Change in Proteins. *J. Chem. Theory Comput.* **2006**, *2*, 840–857.

[299] Buchete, N.-V.; Hummer, G. Coarse Master Equations for Peptide Folding Dynamics. *J. Phys. Chem. B* **2008**, *112*, 6057–6069.

[300] Noé, F.; Fischer, S. Transition Networks for Modeling the Kinetics of Conformational Change in Macromolecules. *Curr. Opin. Struct. Biol.* **2008**, *18*, 154–162.

[301] Wales, D. J. Energy Landscapes: Some New Horizons. *Curr. Opin. Struct. Biol.* **2010**, *20*, 3–10.

[302] Gong, L.; Zhou, X. Kinetic Transition Network Based on Trajectory Mapping. *J. Phys. Chem. B* **2010**, *114*, 10266–10276.

[303] Gilli, P.; Ferretti, V.; Gilli, G.; Borea, P. A. Enthalpy-Entropy Compensation in Drug-Receptor Binding. *J. Phys. Chem.* **1994**, *98*, 1515–1518.

[304] Sharp, K. Entropy-Enthalpy Compensation: Fact or Artifact? *Protein Sci.* **2001**, *10*, 661–667.

[305] Starikov, E. B.; ; Nordén, B. Enthalpy-Entropy Compensation: A Phantom or Something Useful? *J. Phys. Chem. B* **2007**, *111*, 14431–14435.

[306] Freed, K. F. Entropy-Enthalpy Compensation in Chemical Reactions and Adsorption: An Exactly Solvable Model. *J. Phys. Chem. B* **2011**, *115*, 1689–1692.

[307] Starikov, E.; Nordén, B. Entropy-Enthalpy Compensation as a Fundamental Concept and Analysis Tool for Systematical Experimental Data. *Chem. Phys. Lett.* **2012**, *538*, 118–120.

[308] Starikov, E. Valid Entropy-Enthalpy Compensation: Its True Physical-Chemical Meaning. *J. Appl. Solution Chem. Model.* **2013**, *2*, 240–245.

[309] Breiten, B.; Lockett, M. R.; Sherman, W.; Fujita, S.; Al-Sayah, M.; Lange, H.; Bowers, C. M.; Heroux, A.; Krilov, G.; Whitesides, G. M. Water Networks Contribute to Enthalpy/Entropy Compensation in Protein-Ligand Binding. *J. Am. Chem. Soc.* **2013**, *135*, 15579–15584.

[310] Ryde, U. A Fundamental View of Enthalpy-Entropy Compensation. *Med. Chem. Commun.* **2014**, *5*, 1324–1336.

[311] Pan, A.; Biswas, T.; Rakshit, A. K.; Moulik, S. P. Enthalpy-Entropy Compensation (EEC) Effect: A Revisit. *J. Phys. Chem. B* **2015**, *119*, 15876–15884.

[312] Ahmad, M.; Helms, V.; Lengauer, T.; Kalinina, O. V. Enthalpy-Entropy Compensation upon Molecular Conformational Changes. *J. Chem. Theory Comput.* **2015**, *11*, 1410–1418.

[313] Fernandez-Escamilla, A.-M.; Rousseau, F.; Schymkowitz, J.; Serrano, L. Prediction of Sequence-Dependent and Mutational Effects on the Aggregation of Peptides and Proteins. *Nat. Biotech.* **2004**, *22*, 1302–1306.

[314] Pawar, A. P.; DuBay, K. F.; Zurdo, J.; Chiti, F.; Vendruscolo, M.; Dobson, C. M. Prediction of "Aggregation-Prone" and "Aggregation-Susceptible" Regions in Proteins Associated with Neurodegenerative Diseases. *J. Mol. Biol.* **2005**, *350*, 379–392.

[315] Tartaglia, G. G.; Pawar, A. P.; Campioni, S.; Dobson, C. M.; Chiti, F.; Vendruscolo, M. Prediction of Aggregation-Prone Regions in Structured Proteins. *J. Mol. Biol.* **2008**, *380*, 425–436.

[316] Maurer-Stroh, S.; Debulpaep, M.; Kuemmerer, N.; de la Paz, M. L.; Martins, I. C.; Reumers, J.; Morris, K. L.; Copland, A.; Serpell, L.; Serrano, L.; Schymkowitz, J. W. H.; Rousseau, F. Exploring the Sequence Determinants of Amyloid Structure Using Position-Specific Scoring Matrices. *Nat. Meth.* **2010**, *7*, 237–242.

[317] Belli, M.; Ramazzotti, M.; Chiti, F. Prediction of Amyloid Aggregation in vivo. *EMBO rep.* **2011**, *12*, 657–663.

[318] Massi, F.; Straub, J. E. Energy Landscape Theory for Alzheimer's Amyloid β-Peptide Fibril Elongation. *Proteins: Struct., Funct., Bioinf.* **2001**, *42*, 217–229.

[319] De Simone, A.; Kitchen, C.; Kwan, A. H.; Sunde, M.; Dobson, C. M.; Frenkel, D. Intrinsic Disorder Modulates Protein Self-Assembly and Aggregation. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 6951–6956.

[320] Zhuravlev, P. I.; Reddy, G.; Straub, J. E.; Thirumalai, D. Propensity to Form Amyloid Fibrils Is Encoded as Excitations in the Free Energy Landscape of Monomeric Proteins. *J. Mol. Biol.* **2014**, *426*, 2653–2666.

[321] Levine, Z. A.; Larini, L.; LaPointe, N. E.; Feinstein, S. C.; Shea, J.-E. Regulation and Aggregation of Intrinsically Disordered Peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 2758–2763.

[322] Lapidus, L. J. Understanding Protein Aggregation from the view of Monomer Dynamics. *Mol. BioSyst.* **2013**, *9*, 29–35.

[323] Ahmad, B.; Chen, Y.; Lapidus, L. J. Aggregation of α-Synuclein is Kinetically Controlled by Intramolecular Diffusion. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 2336–2341.

[324] Acharya, S.; Safaie, B. M.; Wongkongkathep, P.; Ivanova, M. I.; Attar, A.; Klrner, F.-G.; Schrader, T.; Loo, J. A.; Bitan, G.; Lapidus, L. J. Molecular Basis for Preventing α-Synuclein Aggregation by a Molecular Tweezer. *J. Biol. Chem.* **2014**, *289*, 10727–10737.

[325] Acharya, S.; Saha, S.; Ahmad, B.; Lapidus, L. J. Effects of Mutations on the Reconfiguration Rate of α-Synuclein. *J. Phys. Chem. B* **2015**, *119*, 15443–15450.

[326] Noé, F.; Nüske, F. A Variational Approach to Modeling Slow Processes in Stochastic Dynamical Systems. *Multiscale Model. Simul.* **2013**, *11*, 635–655.

[327] Nüske, F.; Keller, B. G.; Pérez-Hernández, G.; Mey, A. S. J. S.; Noé, F. Variational Approach to Molecular Kinetics. *J. Chem. Theory Comput.* **2014**, *10*, 1739–1752.

[328] Vitalini, F.; Noé, F.; Keller, B. G. A Basis Set for Peptides for the Variational Approach to Conformational Kinetics. *J. Chem. Theory Comput.* **2015**, *11*, 3992–4004.

[329] Fukunishi, H.; Watanabe, O.; Takada, S. On the Hamiltonian Replica Exchange Method for Efficient Sampling of Biomolecular Systems: Application to Protein Structure Prediction. *J. Chem. Phys.* **2002**, *116*, 9058–9067.

[330] Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Tempering: A Method for Sampling Biological Systems in Explicit Water. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.

[331] Moors, S. L. C.; Michielssens, S.; Ceulemans, A. Improved Replica Exchange Method for Native-State Protein Sampling. *J. Chem. Theory Comput.* **2011**, *7*, 231–237.

[332] Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A More Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431–9438.

[333] Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New Feathers for an Old Bird. *Comput. Phys. Commun.* **2014**, *185*, 604–613.

[334] Bussi, G. Hamiltonian Replica Exchange in GROMACS: a Flexible Implementation. *Mol. Phys.* **2014**, *112*, 379–384.

[335] Miller, C. M.; Brown, A. C.; Mittal, J. Disorder in Cholesterol-Binding Functionality of CRAC Peptides: A Molecular Dynamics Study. . *Phys. Chem. B* **2014**, *118*, 13169–13174.

[336] Stirnemann, G.; Sterpone, F. Recovering Protein Thermal Stability Using All-Atom Hamiltonian Replica-Exchange Simulations in Explicit Solvent. *J. Chem. Theory Comput.* **2015**, *11*, 5573–5577.

[337] Pang, X.; Zhou, H.-X. Disorder-to-Order Transition of an Active-Site Loop Mediates the Allosteric Activation of Sortase A. *Biophys. J.* **2015**, *109*, 1706–1715.

[338] Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory and Comput.* **2015**, *11*, 5525–5542.

[339] Flyvbjerg, H.; Petersen, H. G. Error Estimates on Averages of Correlated Data. *J. Chem. Phys.* **1989**, *91*, 461–466.

[340] Wilson, K. In *Recent Developments in Gauge Theories*; 't Hooft, G., Ed.; Plenum: New York, U.S.A., 1980; pp 363–402.

[341] Whitmer, C. Over-Relaxation Methods for Monte Carlo Simulations of Quadratic and Multiquadratic Actions. *Phys. Rev. D* **1984**, *29*, 306–311.

[342] Gottlieb, S.; Mackenzie, P. B.; Thacker, H.; Weingarten, D. Hadronic Coupling Constants in Lattice Gauge Rheory. *Nucl. Phys. B* **1986**, *263*, 704–730.