Network Models and Structural Characteristics of Oligomer Formation of Amyloid-β Peptides and Huntingtin Proteins Based on Microsecond Timescale Molecular Dynamics Simulations

Inaugural Dissertation

presented by

Mohammed Khaled

from

Qalqilia, Palestine

for the attainment of the title of doctor
in the Faculty of Mathematics and Natural Sciences
at the Heinrich-Heine-Universität Düsseldorf

Jülich, May 2023

Supervisor:
Prof. Dr. Birgit Strodel

from the Institute of Theoretical and Computational Chemistry
at the Heinrich Heine University Düsseldorf

Published by permission of the
Faculty of Mathematics and Natural Sciences at
Heinrich-Heine-Universität Düsseldorf

Supervisor: Prof. Dr. Birgit Strodel
Co-supervisor: Prof. Dr. Gunnar Schröder

Date of the oral examination: 08.05.2023

Statement of Author

I, Mohammed Khaled, hereby certify that I wrote the thesis titled ”Network Models and Structural Characteristics of Oligomer Formation of Amyloid-β Peptides and Huntingtin Proteins Based on Microsecond Timescale Molecular Dynamics Simulations”, without any unauthorized external assistance and used only sources acknowledged in the work. All texts excerpted verbatim or paraphrased from published and unpublished texts as well as all information obtained from oral sources are duly indicated and listed in accordance with bibliographical rules. My thesis contains no material published elsewhere or extracted in whole or part from a thesis submitted for a degree at this or any other university. Where the results are produced in collaboration with others, my contributions are clearly stated.

Signed: _____ Date: _____

Acknowledgments

I am deeply grateful to my supervisor, Prof. Dr. Birgit Strodel, for providing me the opportunity to conduct my doctoral research under her guidance. Throughout my Ph.D. studies, she always welcomed scientific discourse and generously provided me with beneficial guidance and advice. I am thankful to her for providing me the opportunity to fully realize my potential and for supporting me to pursue a career in research. It was an honor and a privilege to work under her guidance, and I am grateful for the experience.

I express my gratitude to the Forschungszentrum Jülich for providing an exceptional research environment that continuously fuels my inspiration. I extend my gratefully acknowledge to the Jülich Supercomputing Center (JSC) for their valuable computing resource on Jureca Supercomputer. I acknowledge the financial support for this project from the German Federal Ministry of Education and Research (BMFB) via the Palestinian-German Science Bridge (PGSB). I would like to thank my institution, the Institute of Biological Information Processing (IBI-7), for providing such an inspiring and enjoyable working atmosphere. I would also like to express my appreciation to the staff members at IBI-7 for their support.

I would like to express my gratitude to Prof. Gunnar Schröder for his dedicated effort in providing me with detailed feedback on my thesis. I would also like to thank my examiners: Prof. Renu Batra-Safferling, Prof. Wolfgang Hoyer, and Prof. Claus Seidel. I am thankful to Prof. Abdallah Sayyed-Ahmad for his support as well as I would like to express my gratitude for fruitful scientific discussions with collaboration partners Dr. Nicklas Österlund and Prof. Astrid Gräslund.

I am thankful for the friendship and memories shared with my colleagues and former members of the Strodel lab, Dr. Hebah Fatafta, Dr. Maryam Olagunju, Dr. Suman Samantray, Moritz Schäffler, Jennifer Loschwitz, Wibke Schumann, Anna Jäckering and Lara Scharbert. Your support has been invaluable and I am truly grateful for the time spent together. I would like to extend my gratitude to my friends at Jülich for the enjoyable moments we shared together.

I am deeply thankful to my family, particularly my mother, father and siblings for

their constant emotional, mental and spiritual support throughout my journey. I am incredibly grateful for your unwavering encouragement and belief in me as I worked toward my goals. Thank you all for being there for me.

Above all, I am truly thankful to Allah for providing me with the strength and support to overcome all obstacles and for giving me the opportunity to continue my research.

List of Publications

- Paul A., Samantray, S., Anteghini, M., **Khaled, M.**, Strodel, B. (2021). Thermodynamics and kinetics of the amyloid-β peptide revealed by markov state models based on MD data in agreement with experiment. Chem. Sci., 12(19), 6652-6669.

- Fatafta H., **Khaled, M.**, Owen M. C., Sayyed-Ahmad, A., Strodel, B. (2021). Amyloid-β peptide dimers undergo a random coil to β-sheet transition in the aqueous phase but not at the neuronal membrane. PNAS, 118(39), e2106210118.

- *Schäffler, M., ***Khaled, M.**, Strodel, B. (2022). ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins (* contributed equally). Methods. Volume 206, Pages 18-26, ISSN 1046-2023.

- **Khaled, M.**, Strodel, B., Sayyed-Ahmad, A. (2023). Comparative molecular dynamics simulation studies of pathogenic and non-pathogenic Huntingtin protein monomer and dimer. Frontiers in Molecular Biosciences, Volume 10:1143353.

- **Khaled, M.**, Rönnbäck, I., L. Ilag, L., Gräslund, A., Strodel, B., Österlund, N. (2023). A hairpin motif in the Amyloid-β peptide is important for formation of disease-related oligomers. (Manuscript)

Author Contributions

Chapter 2.3

Article: ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins.

Authors: Moritz Schäffler, Mohammed Khaled & Birgit Strodel.

Contributions: Mohammed Khaled: software, validation, data curation, writing part of the original draft. **Moritz Schäffler:** methodology, software, validation, formal analysis, Investigation, data curation, writing part of the original draft, visualization. **Birgit Strodel:** methodology, resources, writing - reviewing & editing, supervision, project administration, funding acquisition.

Chapter 3.2

Article: A hairpin motif in the Amyloid-β peptide is important for formation of disease-related oligomers.

Authors: Mohammed Khaled, Isabel Rönnbäck, Leopold L Ilag, Astrid Gräslund, Birgit Strodel & Nicklas Österlund.

Contributions: Mohammed Khaled: performing and analysis of molecular dynamics simulations, producing the simulation-related figures, writing part of the first draft of the paper, analyzing data, reviewing. **Isabel Rönnbäck:** performing research. **Leopold L Ilag:** performing research. **Astrid Gräslund:** Conceptualizing, writing - reviewing & editing, supervising the work. **Birgit Strodel:** Conceptualizing, writing - reviewing & editing, supervising the work. **Nicklas Österlund:** Conceptualizing, writing the paper, performing research, analyzing data, reviewing.

Chapter 3.3

Article: Amyloid-β peptide dimers undergo a random coil to β-sheet transition in the aqueous phase but not at the neuronal membrane.

Authors: Hebah Fatafta, Mohammed Khaled, Michael C Owen, Abdallah Sayyed-Ahmad & Birgit Strodel.

Contributions: Mohammed Khaled: performing research, analyzing data. **Hebah Fatafta:** writing the first draft, designing research, performing research, analyzing data. **Michael C Owen:** performing research, analyzing data. **Abdallah Sayyed-Ahmad:** writing the paper, supervising the work. **Birgit Strodel:** designing research, writing the paper, supervising the work.

Chapter 4

Article: Comparative molecular dynamics simulation studies of pathogenic and non-pathogenic Huntingtin protein monomer and dimer.

Authors: Mohammed Khaled, Birgit Strodel & Abdallah Sayyed-Ahmad.

Contributions: Mohammed Khaled: writing the first draft of the manuscript, designing of the study, performed the simulations, analyzing the simulation data ,reviewing. **Birgit Strodel:** writing the paper, supervising the work, reviewing. **Abdallah Sayyed-Ahmad:** writing the paper, supervising the work, reviewing.

Abstract

Intrinsically disordered proteins (IDPs) are abundant in the human proteome and often associated with amyloid-related diseases, such as Alzheimer's and Huntington's disease. While IDPs can aggregate into insoluble fibrillar β-sheet aggregates, they can also form soluble, intermediate-size aggregates called oligomers that are highly toxic and are polymorphic. Studying oligomers in detail has been a significant challenge, and the availability of experimental methods for their thermodynamic and kinetic characterization is limited due to the low abundance and the transit nature of the oligomer. Molecular dynamics (MD) simulation provides a powerful tool to investigate their behavior on the atomistic scale. Additionally, advanced methods for analyzing MD simulations, such as Markov state models and transition networks, have enable a detailed exploration of the conformational landscapes of the oligomers. To this end, we utilized all-atom MD simulations to investigate the aggregation of amyloid-β (Aβ) peptides and Huntingtin (Htt) proteins into small oligomers. Specifically, our studies include the following: (i) In our first study, we conducted multiple MD simulations totaling 0.3 ms of Aβ_{42} oligomers, including dimers, tetramers, and hexamers, which were then analyzed using Markov state models and transition networks. Our analysis revealed that the oligomers can occupy multiple states, and transitions between these states occur within microseconds. These findings demonstrate the feasibility of characterizing the kinetics and thermodynamics of Aβ_{42} oligomers using network analysis. (ii) We conducted research on the structures and mechanisms of aggregation for different Aβ variants into oligomers. Our findings indicated that Aβ variants lacking the ability to fold into a β-hairpin structure, as present in the Aβ_{40} peptide, are unable to form oligomers. (iii) We also compered the formation of Aβ_{42} dimers in solution and at the neuronal membrane. In solution, the dimerization process involves a transition from a random coil to a β-sheet structure, which is a key step towards amyloid aggregation. However, our findings show that when Aβ_{42} dimers interact with the neuronal membrane, they are less likely to adopt a β-sheet structure. (iv) Finally, we investigated the conformational and dynamical effects of polyglutamine expansion and its flanking domains on the folding and dimerization of Htt proteins. Our findings revealed significant distinctions between the nonpathogenic and pathogenic Htt monomers, which directly affect their dimerization. In summary, by utilizing MD simulations and advanced analysis methods, our approach provides a detailed understanding of the aggregation of Aβ oligomers and

Htt proteins at the atomic level. The investigations conducted in this thesis are valuable in comprehending the initial stages of the amyloid aggregation process, which can facilitate the development of prospective treatments for Amyloid-related diseases.

List of Abbreviations

AD	Alzheimer's Disease
HD	Huntington's Disease
PD	Parkinson's Disease
Aβ	Amyloid-β
Htt	Huntingtin
IDP	Intrinsically Disordered Protein
IDPR	Intrinsically Disordered Protein Region
APP	Amyloid Precursor Protein
PolyQ	Polyglutamine
Q	Glutamine
PolyP	Polyproline
PRD	Proline-Rich Domain
MD	Molecular Dynamics
MM	Molecular Mechanics
FF	Force Field
LJ	Lennard-Jones
vdW	van der Waals
UB	Urey-Bradley
TIP	Transferable Intermolecular Potential Function
PBC	Periodic Boundary Condition
PME	Particle Mesh Ewald Method
REMD	Replica Exchange Molecular Dynamics
MSM	Markov State Model
HMM	Hidden Markov State Model
PCA	Principal Component Analysis
CV	Collective Variable
TICA	Time-Lagged Independent Component Analysis
MEMM	Multiensemble Markov Model
US	Umbrella Sampling
TRAM	Transition-Based Reweighted Analysis Method
LEQ	Local Equilibrium Likelihood
MBAR	Multistate Bennett Acceptance Ratio
TPT	Transition Path Theory

MFPT	Mean First Passage Time
TN	Transition Networks
DSSP	Dictionary of Protein Secondary Structures
cMD	Conventional Molecular Dynamics
LINCS	Linear Constraint Solver
PCCA	Robust Perron Cluster Analysis
IC	Independent Component
SASA	Solvent Accessible Surface Areas
hSASA	Hydrophobic Solvent Accessible Surface Areas
MS	Mass Spectrometry
IM	Ion mobility
CCS	Collision Cross Section
CD	Circular Dichroism
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
CHOL	Cholesterol
SM	Sphingomyelin
GM1	Monosialotetrahexosylganglioside

Contents

List of Figures

Chapter

1.

Introduction

Proteins are vital components of living systems and play a crucial role in carrying out essential biological functions such as catalyzing biochemical reactions, transporting ions and molecules across membranes, and facilitating cell replication. The majority of proteins fold into unique, three-dimensional native structures that enable them to perform their functions. However, some proteins, known as intrinsically disordered proteins (IDPs), feature flexible, three-dimensional structures that are critical to their function.

One aspect of protein behavior that has received much attention is protein aggregation, where proteins self-assemble into large structures. The aggregation of amyloid proteins into neurotoxic oligomers and fibrils is a common characteristic of several neurodegenerative disorders, including Alzheimer's disease (AD), Huntington's disease (HD) and Parkinson's disease (PD) [1, 2]. Despite the significant global impact of these diseases, there is still no cure available for many of them [3]. In the progression of several neurodegenerative diseases, amyloid aggregation affects the brain or neurons, where the misfolded amyloid proteins assemble into soluble intermediate aggregates known as oligomers and protofibrils, which eventually accumulate into rigid structures known as amyloid fibrils (Figure 1.1). These fibrils are characterized by their rich intermolecular β-sheet structures. However, it is believed that the most toxic species are the smaller, heterogeneous, and mobile oligomers, as they can diffuse across cells and interact with cell membranes, disrupting membrane-related functions [4]. Despite the critical importance of understanding the aggregation of

amyloid proteins, it remains a challenging task due to the difficulties in structurally and dynamically characterizing IDPs by experimental and computational methods [5]. As such, there is an urgent need to gain a deeper understanding of the molecular basis of oligomer aggregation mechanisms.

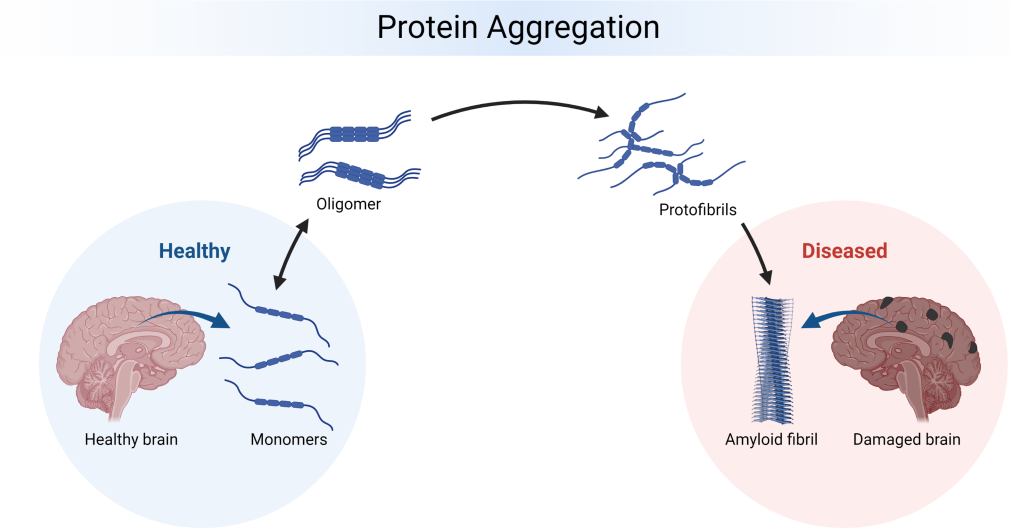

Figure 1.1.: Protein aggregation from monomers to fibrils that accumulate in the brain leads to neurodegenerative disease.

Computer simulations have become a valuable tool for modeling the aggregation of amyloid proteins at the atomic level, providing a better understanding of their mechanisms and structures on various time scales [6]. Despite the increasing computational costs associated with simulating large systems and longer simulation times, recent advancements in technology, software, and simulation techniques have made it possible to produce more accurate predictions. In this thesis, we employ computer simulations to study the aggregation mechanisms and structures of small oligomers of amyloid-β (Aβ) peptides and Huntingtin (Htt) proteins, which are challenging to study experimentally. Through molecular dynamics simulations, we aim to gain a deeper understanding of the conformations and aggregation kinetics of these amyloidogenic proteins. Additionally, we use network models such as transition networks and Markov state models to analyze the simulation data and shed light on their underlying mechanisms.

1.1. Intrinsically Disordered Proteins

Some proteins have well-defined three-dimensional structures that are inseparably associated with their biological functions. Indeed, knowing the structures of the protein is the way to define their function [7]. Such a protein is called a "folded" protein. Even though folded proteins are frequently described as having static three-dimensional structures, they can sample different conformations throughout their biological lifetime by responding to thermal fluctuations or other conditions [8]. In general, protein fluctuations allow them to carry out their prescribed functions. Given the fundamental role that protein motions play in defining their functions, the protein structures should be determined as a structural ensemble of accessible states that a protein can adopt [9]. Most of the experimental methods determine the protein structures by averaging over structural ensembles. In the case of a folded protein with well-defined free energy minima, such an approach is convenient, since long-timescale observations will show only slight variations in the measurements. Thus, the ensemble average captures a protein's essential structural features, which can help in understanding its function (Figure 1.2). In constant, IDPs do not have a single equilibrium structure and instead exist as dynamic, heterogeneous ensembles of conformations resulting from their relatively flat free-energy landscapes [10]. Furthermore, due to their disordered nature, structural characterization is difficult, as the average over long timescales or over the entire ensemble of structures obscure states and processes that occur on short timescales or that are less populated [11] (Figure 1.2). Thus, using the same approaches to determine the IDP structures as for folded protein is not convenient where the ensemble-averaged structure does not represent any structure in the ensemble. The flat energy landscape of IDPs consists of different local energy minima separated by small energy barriers, the transitions between those energy minima happen frequently and quickly, generating an ensemble of structurally distinct states with nearly equal energies. In fact, determination of the transition rates between the conformational states in the IDP ensemble is quite challenging to capture by experimental or computational methods. Consequently, advanced methods are required to understand the structures of IDPs and therefore get a better understanding of their function.

Disordered proteins are very common in nature and perform a variety of biological processes, such as cell cycle control, storage of small molecules, cellular signal

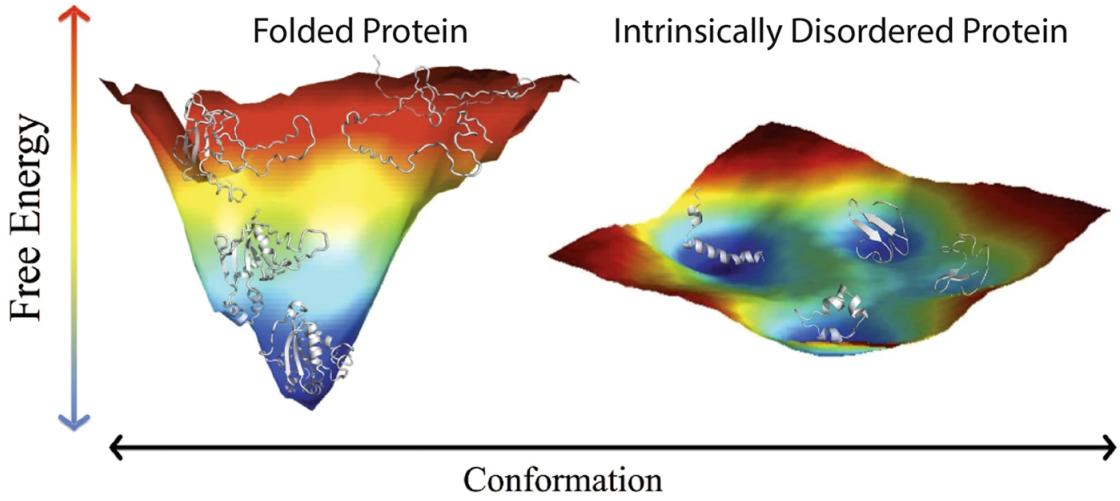

Figure 1.2.: Energy landscapes of a natively folded protein, human nucleoside diphosphate kinase (Left) and of an intrinsically disorder peptide, CcdA C-terminus (Right). The local free energy is shown as spectral, with the red highest and blue the lowest energies. Reproduced with Copyright ©2014 Polymers [12].

transduction, cell signaling, protein phosphorylation, protein-protein interaction, and self-assembly regulation. Around 25% of the human proteome is completely disordered and classified as IDPs and 40% contain one or more intrinsically disordered regions (IDPRs) that have lengths of at least 30 amino acids [13, 14]. The number of proteins identified as disordered protein is ever-increasing. A hallmark of IDPs (and IDPRs) is a low proportion of hydrophobic and aromatic amino acids, but a relatively high number of polar and charged amino acids, resulting in a high net charge. Many of these proteins play essential roles in many pathological pathways, particularly, the self-assembly of IDPs is linked to many neurodegenerative diseases known as amyloidoses. For example, aggregation of α-synuclein in the brains can lead to Parkinson's disease [15]. Huntington's disease is associated with the aggregation of Huntingtin proteins [2, 16]. Likewise, aggregation of the tau and Aβ proteins are the pathological features in the progression of Alzheimer's disease [17, 18]. Furthermore, malfunction of IDPs can as well lead to pathogenic faults in signaling pathways [19]. While the important role of IDPs in various biological functions and pathological pathways is prominent, their structural heterogeneity imposes notable challenges in studying those proteins and revealing their exact function.

Disordered proteins can be transformed into structured forms in multiple ways, driven

by either intramolecular or intermolecular interactions. A common way is known as folding-upon-binding in which the disordered protein converts to a structured form upon binding to a specific binding partner [20]. IDPs can also form ordered conformation driven mainly by intermolecular side-chain interactions *via* self-assembly into thermodynamically stable structures (amyloid fibrils) which will be explained in the next section.

1.2. Amyloid Aggregation

Protein aggregation, or the self-assembly of proteins, is a widespread phenomenon with significant impacts on biotechnology, protein biochemistry, and medicine [21]. Many folded proteins naturally form small, reversible oligomers through self-assembly *in vivo*. At high protein concentrations, these proteins aggregate to create larger oligomers that can eventually form mature fibrils. For instance, insulin proteins form reversible oligomers and fibrils [22], and interleukin-1 receptor antagonist proteins can create reversible oligomers at high protein concentrations [23]. However, certain proteins, particularly IDPs, tend to form irreversible aggregates associated with numerous neurodegenerative diseases [24, 25]. These diseases are characterized by disordered or misfolded proteins aggregating into intra-cellular inclusions or extracellular plaques (amyloids) in tissues. The progression of amyloidosis is thought to result from the accumulation of plaques, either through the "Gain of Toxicity" mechanism, which destroys the nearby tissues and cells, or through the "Loss of Normal Function" mechanism, which results in a dysfunctional cellular activity due to insulating other proteins in plaques and the subsequent development of disease [26, 27].

In protein aggregation, similar disordered proteins interact to form dimers, trimers, oligomers, protofibrils, and eventually amyloid fibrils. The structure of amyloid fibrils is highly ordered, characterized by cross β-sheet motifs, though it can exhibit polymorphism [28, 29]. The aggregation pathway of different proteins can vary and result in diverse fibrillar structures or amorphous aggregates. Various mechanisms have been proposed to explain the self-assembly of disorder proteins into amyloid fibrils, including nucleated or non-nucleated growth polymerization [30–32]. The nucleated growth polymerization pathway is divided into three stages: nucleation, elongation, and equilibrium (Figure 1.3) [30]. The nucleation phase involves the

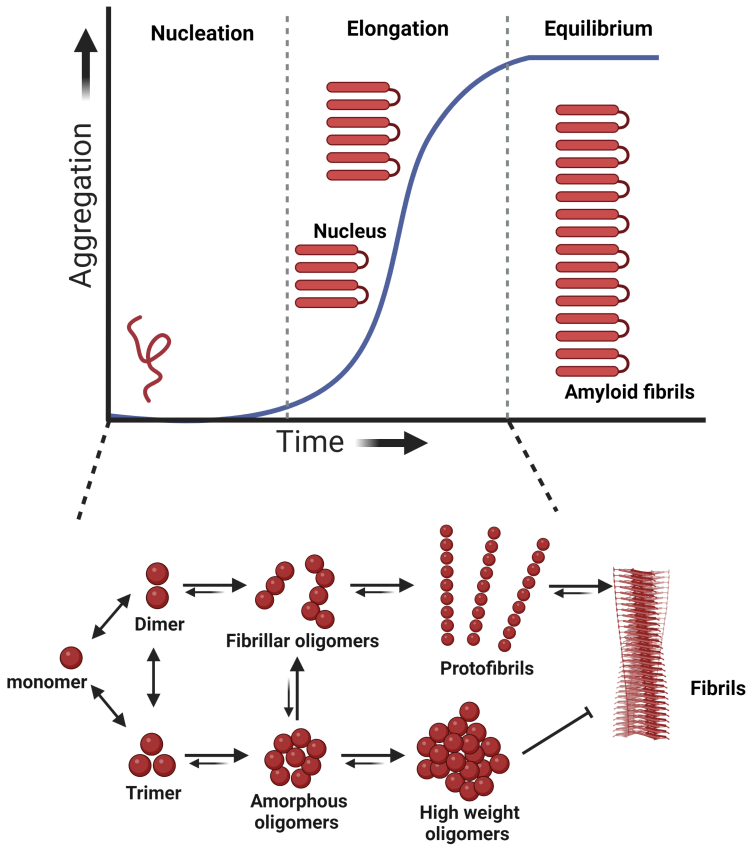

Figure 1.3.: Schematic representation of amyloid aggregation. The nucleated-polymerization model consists of three phases: nucleation, elongation, and equilibrium. Oligomers are intermediate structures in the fibril formation process and mainly exist in the nucleation phase and elongation phase. Some oligomers, labeled as fibrillar oligomers, can rapidly form protofibrils and fibrils *via* the fibril-formation pathways. Other oligomers, on the other hand, such as amorphous and high-weight oligomers are structurally distinguished from fibrillar oligomers, and they are off-pathway of the fibril formation, which could contribute to fibril formations after a structural conversion to fibrillar oligomers.

slow assembly of monomers into a thermodynamically unfavorable nucleus, followed by exponential growth of the fibril through the addition of more monomers in the elongation phase. Finally, the equilibrium phase is reached when no further elongation occurs and mature fibrils are formed through lateral association of soluble species. However, the aggregation mechanism can be influenced by environmental conditions, such as cell membranes, which play a role in the toxicity of the aggregates in neurodegenerative diseases such as Alzheimer's [33, 34]. More research is needed to fully understand the molecular mechanism of peptide-membrane interactions and

the resulting membrane damage.

So far, the majority of research on amyloids has been focused on mature fibrils due to the challenges of studying the early-stage oligomers that are transient and have a high tendency to aggregate. As a result, the current understanding of aggregation mechanisms is mostly based on studies of fibrils. Although molecular simulations have been helpful, their results are limited by the accuracy of the used force fields. Force fields are not capable of accurately predicting the aggregation potential of protein sequences, leading to fast aggregation kinetics in many cases. However, recent advancements have seen the development of force fields specifically designed for IDPs, making the predictions more precise and trustworthy [35–38]. Despite these improvements, the structures and formation mechanisms of oligomers remain elusive and poorly understood.

1.3. Amyloid-β Peptide

Aβ is a soluble disordered peptide that is naturally produced in neurons and other cells in the brain through the enzymatic cleavage of a transmembrane protein known as an amyloid precursor protein (APP). The cleavage process of APP is catalyzed by two enzymes, β-secretase and γ-secretase that cleave the N-terminal and C-terminal, respectively (Figure 1.4) [39, 40]. Normally, the clearance mechanism controls the peptide levels in the brain, however, defects in Aβ clearance mechanism cause abnormally high levels of the peptide leading to its accumulation and deposition as extracellular plaques in the brain; a characteristics feature of AD [41]. AD is a neurodegenerative disorder and the most prevalent form of dementia. It affects millions of people around the globe, particularly people aged 65 and older [1], and the numbers are going to escalate over the next decades. The disease was first diagnosed by Alois Alzheimer in 1907. Since then, the pathological cause of AD was ambiguous and several hypotheses suggested [42]. Nowadays, the most vastly accepted theory is the amyloid cascade hypothesis [43–45]. It proposes that Aβ peptide aggregates as the central pathological cause of the disease progression, suggesting that the Aβ peptide accumulates extracellularly in senile plaques, and the misfolded tau protein accumulates intracellularly in neurofibrillary tangles which leads to memory loss and a decline in cognitive function over time [46].

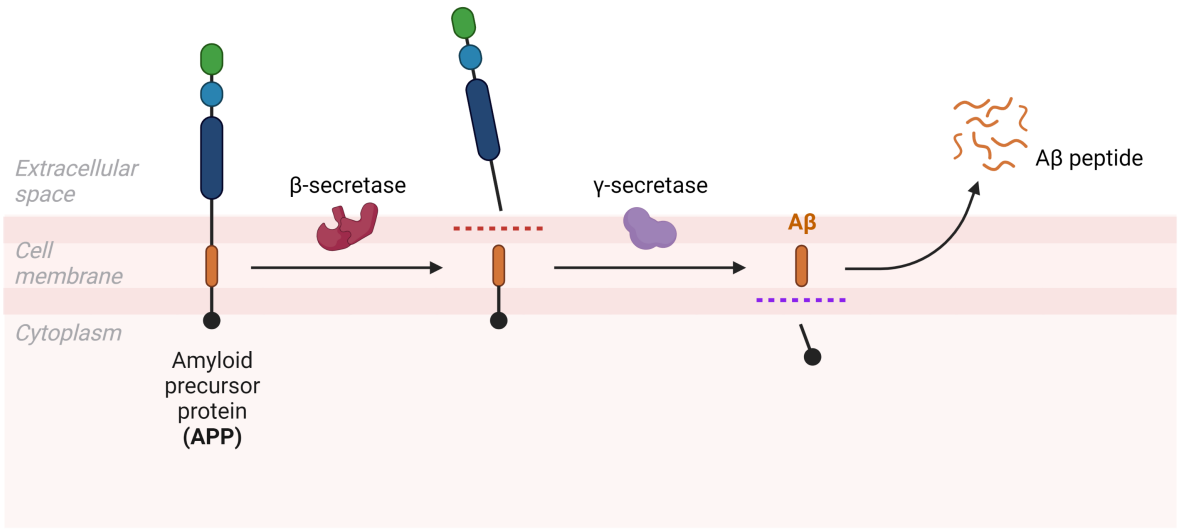

Figure 1.4.: The cleavage of the amyloid precursor protein by β-secretase and γ-secretase producing Aβ peptide.

The cleavage of APP at its C-terminal by γ-secretase generates Aβ peptides with 36 to 43 amino acid residues. The most prevalent species are Aβ_{40} and Aβ_{42} [47], where the latter has two additional hydrophobic residues at its C-terminal. The Aβ sequences are composed of four regions: the hydrophilic N-terminal region or the metal binding region and two hydrophobic segments in the center (residues 16-22) and in the C-terminal region of the sequence (residues 30-42), which lead to form a β-hairpin structure, separated by the central polar region (residues 23-29) (Figure 1.5). Aβ_{40} is the most abundant in the brain, while Aβ_{42}, which is more aggregation-prone and toxic, is more commonly found in senile plaques [48–50]. The higher aggregation tendency of Aβ_{42} is due to the additional two hydrophobic residues at the C-terminal that affect the flexibility of the peptide by enhancing the formation of a second β-hairpin by residues 31 to 34 and 38 to 41 [51].

In the last years, a considerable amount of literature has been published on the smaller soluble Aβ oligomers as the main toxic aggregate species instead of the mature fibrils [53–56]. Therefore, there is an increasing interest in an intensive investigation of Aβ oligomerization and its role in AD at the molecular scale to better understand the aggregation process.

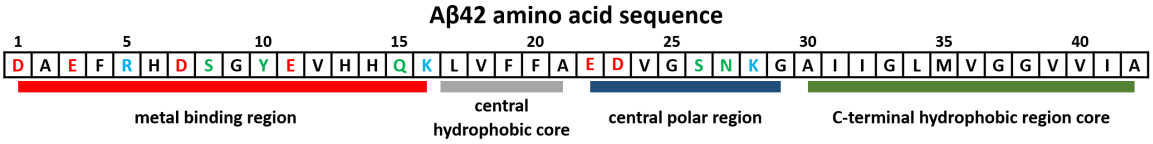

Figure 1.5.: The sequence of the Aβ_{42} can be divided into four regions: the metal-binding region (red), the central hydrophobic core region (gray), the central polar region (blue), and the C-terminal hydrophobic region (green). Residues marked in red, blue, green, and black are negatively charged, positively charged, polar and hydrophobic, respectively. Reproduced with Copyright ©2017 Elsevier [52].

1.4. Huntingtin Proteins

Huntingtin (Htt) protein is linked to the neurodegenerative disorder Huntington's disease (HD), which is classified as a polyglutamine (polyQ) disease along with a group of other inherited disorders which are all characterized by the abnormal expansion of a polyQ stretch due to a genetic mutation of trinucleotide CAG repeats encoding glutamine amino acids (Figure 1.6A) [57, 58]. In HD, the mutated Htt protein misfolds and aggregates within brain neurons, leading to their malfunction and death, resulting in cognitive decline and motor disturbance for HD patients [59, 60]. The Htt protein is a large protein consisting of 3,144 amino acids, with the polyQ stretch located at the very N-terminal region (Htt exon1). Normal (non-mutated) Htt has a polyQ stretch length ranging from 5 to 35, while mutated Htt in HD patients has a polyQ length expanded to over 35 repeats. The length of the polyQ is highly correlated with the age of onset and disease severity, with those having a polyQ length of 36 to 39 repeats having a later onset compared to those with 40 or more polyQ repeats [61, 62]. Numerous studies have demonstrated the cytotoxicity of mutated Htt in both its monomer and aggregate forms [63–66]. These studies suggest that mutated Htt exon1 causes neuronal toxicity through the formation of amyloid fibrils within neurons [67] or from monomers and oligomers interfering with cellular proteins and causing their malfunction [68].

Htt protein takes a role in many biological functions and widely interacts with various interaction partners, thus, the exact function of this protein is still unknown. Up to today, the crystallographic structure information at an atomistic resolution of the entire Htt is inaccessible. As IDP, the monomer state of Htt exon1 is a mostly random coil

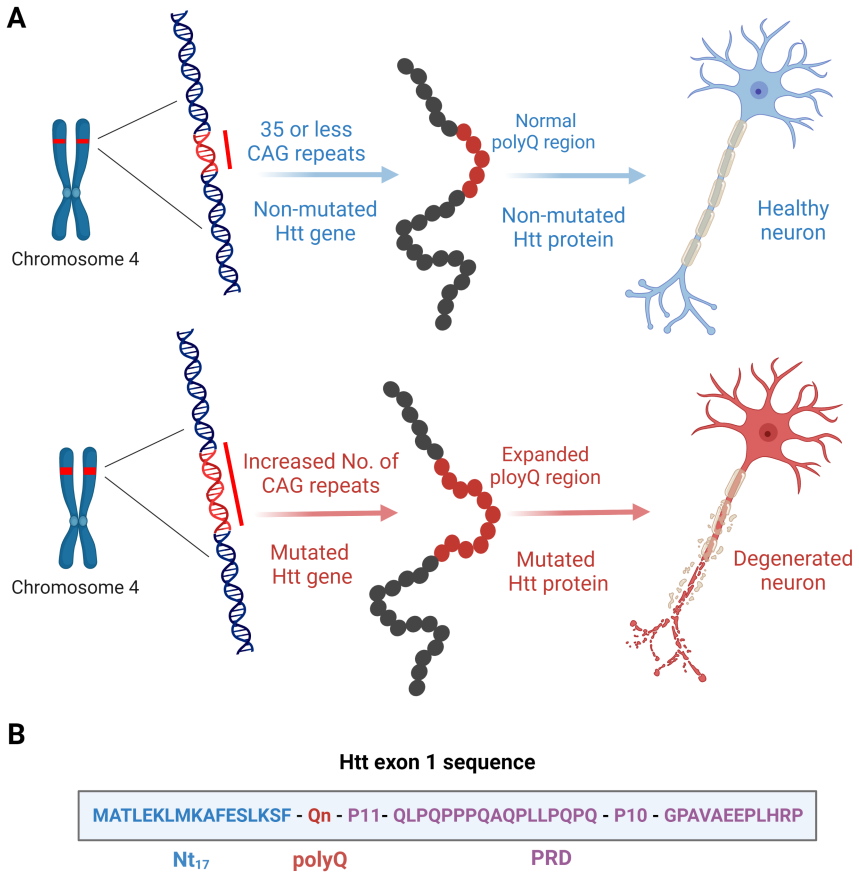

Figure 1.6.: **(A)** Htt protein is encoded by Htt gene located on chromosome 4. The Htt gene contains repeats of the CAG codon that encoded the glutamine (Q) amino acids. If the CAG repeats expand to 36 or more the Htt protein becomes pathogenic. **(B)** Htt exon 1 amino acid sequences. The Nt_{17} region is highlighted in blue, polyQ in red, and PRD in violet.

conformation [69–72]. Current experiments and simulation studies revealed that Htt exon1 has polymorphic structures and can be found in multiple conformations [73, 74]. Nevertheless, mutated Htt exon1 proteins transform into β-sheet-rich structures formed by the polyQ stretches, which assemble into amyloid fibril structures similar to Aβ fibrils found in AD [75–77]. Outside the polyQ stretch of Htt exon1, there are two flanking regions of interest: the N-terminal region consisting of 17 amino acids (Nt_{17}) that precedes the polyQ stretch and the proline-rich domain (PRD) that follow the polyQ stretch (Figure 1.6B). The polyQ flanking domains have different structures and play distinct roles in the aggregation mechanism [78–81]. Interestingly, Nt_{17} forms amphipathic α-helices [82–84] that modulate the aggregation mechanism through helix-helix interactions [85–89]. On the other hand, the PRD has a rela-

tively rigid structure promoted by PPII helix formation, which suppresses the β-sheet formation and thereby decreases the aggregation rate [90–94].

Unveiling the aggregation mechanism and discovering the structures of early-stage Htt-exon1 oligomers is crucial in comprehending the pathological effects of Htt-exon1. However, studying these proteins is challenging due to their high heterogeneity and limitations in computational resources. Currently, there are conflicting computational observations due to the usage of different force fields, water models, and protein models [72], leading to a need for extensive atomistic simulations of full-length Htt-exon1 using an appropriate force field designed for IDPs.

1.5. Aims

A vast number of researchers are conducting scientific investigations on amyloid proteins. These studies are crucial in filling our current knowledge gaps about how amyloid diseases develop and potentially finding new treatments. Oligomers, which are known to be the primary cause of neurotoxicity in the aggregation process, are highly dynamic and exhibit various sizes and heterogeneous structures of aggregates. Experimental techniques have limitations in fully resolving the structures, thermodynamics, and kinetics of these oligomers. Moreover, previous computational studies that used older force fields not suitable for IDPs have resulted in inaccurate conformational ensembles of IDPs. Recent advancements in force fields have led in particular to the all-atom Charmm36m force field [35] which has been shown to accurately predict the conformational ensemble of IDPs [36, 37] and amyloid aggregates [38], make it essential to use Charmm36m to obtain accurate conformational ensembles of Aβ aggregates. We aim to investigate the conformational ensembles of small Aβ oligomers and to extract thermodynamics and kinetics information using network models like transition networks and Markov state models. In this thesis, we focus on studying the aggregation of Aβ_{42} into small oligomers by conducting multiple MD simulations with an accumulated simulation time of 0.3 ms, enabling us to explore the oligomers' conformational ensembles. This understanding of Aβ_{42} aggregation at the atomic level can provides valuable insights into disease-related oligomers. Furthermore, our aim is to understand how the structures, dynamics and aggregation of the Htt-exon 1 proteins are influenced by the polyQ expansions and their adjacent

domains.

The thesis consists of three chapters: In chapter 2, we introduce the research methods and computational approaches used in this thesis. In chapter 3, we extensively discuss the Aβ peptide aggregation by simulating the aggregation of various Aβ species into small oligomers (dimers, tetramers, and hexamers) which is supported by experimental results done by our collaboration partners from Stockholm University. We further study the effects of the neuronal membrane on Aβ dimerization. In chapter 4, we investigate the folding and the dimerization of Htt-exon1 proteins by conducting MD simulation of Htt-exon1 monomer and dimer with normal polyQ region for the non-pathogenic protein and expanded polyQ region for the pathogenic one. The thesis finish by drawing overall conclusion and providing an outlook on future research that could emerge from this work.

Chapter

2.

Methods

The methods chapter introduces the computational approaches and the models used to describe the mechanism of the process under investigation. It is divided into three parts: First, a general overview of molecular dynamics simulation principles and some practical aspects of the method and algorithms are discussed. Afterward, we introduce the theoretical framework for constructing Markov state models for biomolecular systems, as well as some related techniques as dimensional reduction methods such as time-lagged independent component analysis, and K-means clustering. Lastly, the theoretical background of transition networks for structural characterization and aggregations of IDPs is given. In addition, the descriptor functions and the program for the construction of the transition networks are also presented.

2.1. Molecular Dynamics

Molecular dynamics (MD) simulations are essential tools aimed to characterize and predict the macroscopic properties of molecular systems at the atomic level [95, 96]. In recent years, with the availability of more powerful and faster computers, advanced algorithms, and software, MD simulations are increasingly used to explore complex biological processes. As well as providing information about conformations and dynamics of biological processes, MD simulation can also provide thermodynamic and kinetic data. This approach is widely applied to different applications in-

cluding: predict protein structures and dynamics [97], elucidating the protein folding problem [98], protein complexes interactions [99], ligand docking [100], membrane-embedded proteins [101], etc. Computer simulations can be considered powerful complementary tools to experimental methods due to their capability to provide high spatial (down to single atoms) and high temporal (femtoseconds) resolutions as well as the probe of the conformational energy landscape, hence, it may be used to obtain insights into mechanisms and processes that are difficult to capture by the experiments in the laboratory. Consequently, simulations can be used as a link between theoretical modeling and experimental results [102].

Molecular dynamics (MD) simulations use classical molecular mechanics (MM) to model molecular systems and predict their behavior [103, 104]. In a molecular model, atoms are represented as point charges with soft repulsive potentials, and their dynamics are determined by Newtonian mechanics [105]. The forces between particles in the model consist of bonded and nonbonded interactions, collectively referred to as the forcefield. Bonded interactions are between atoms that are chemically bonded, such as covalent bonds, while nonbonded interactions are between atoms that are not chemically bonded. The forcefield is typically described by simple mathematical formulas, including Hooke's law, the Lennard-Jones potential, and the Coulomb potential, which enable Newton's equations of motion to be numerically solved. Through these simulations, we can evaluate structural fluctuations over time, which provide insights into molecular behavior and the conformational energy landscape.

2.1.1. Molecular Dynamics Theory

Molecular dynamics simulations describe the motions of molecular systems by solving Newton's 2^{nd} law of motions for N-particle system. The net force $\mathbf{F}_i$ acting on the particle i with mass m_i and acceleration $\mathbf{a}_i$ within the system is given as:

$$\mathbf{F}_i = m_i \mathbf{a}_i = m_i \sum_{j}^{i \neq j} \frac{\mathbf{F}_{ij}}{m_j} \tag{2.1}$$

where F_{ij} is the force of particle j acting on particle i. Alternatively, the force $\mathbf{F}_i$ can

be derived from the gradient of the system's potential energy function $U(\mathbf{q})$ as:

$$\mathbf{F}_i(\mathbf{q}) = -\frac{\partial U(\mathbf{q}_i)}{\partial \mathbf{q}_i} \tag{2.2}$$

Here, $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2, \ldots, \mathbf{q}_N)$ is the complete set of $3N$ coordinates. Using equations 2.1 and 2.2, the change of positions and velocities as a function of time can be related to the gradient of the potential energy function:

$$-\frac{\partial U(\mathbf{q}_i)}{\partial \mathbf{q}_i} = m_i \frac{d^2\mathbf{q}_i}{d^2t} \tag{2.3}$$

Equation 2.3 represents the fundamental relation of MD simulations, the solution of this equation required defining the initial configuration of positions and velocities of the system particles at time t, then determining the new positions and velocities at time $t + \Delta t$. The repetition of this process generates a successive sequence over time of the position and velocity of the system particles called a trajectory (Figure 2.1). The initial structure of the biomolecular systems is commonly determined through experimental methods such as NMR spectroscopy, electron microscopy and X-ray crystallography [106]. Thus, the computational simulation can be applied to investigate the dynamics of the biomolecular system once it is structure is known.

2.1.2. Force Fields

In computational simulations of biomolecules, force fields (FFs) are used to estimate the potential energy surface. These FFs can be classified into two categories: atomistic or coarse-grained, depending on how the atoms in the system are modeled [107]. Atomistic force fields use classical mechanics at an atomic level and represent the molecular system as a ball-and-stick model. The potential energy function is made up of mathematical formulas that describe the bonded and nonbonded interactions and their associated constant parameters. These parameters are primarily modeled based on experimental data [108] or *ab initio* quantum calculations [109]. Bonded interactions describe the energetic contributions of covalent bond motions, bond angles, and dihedral torsion angles. Nonbonded interactions include Lennard-Jones and

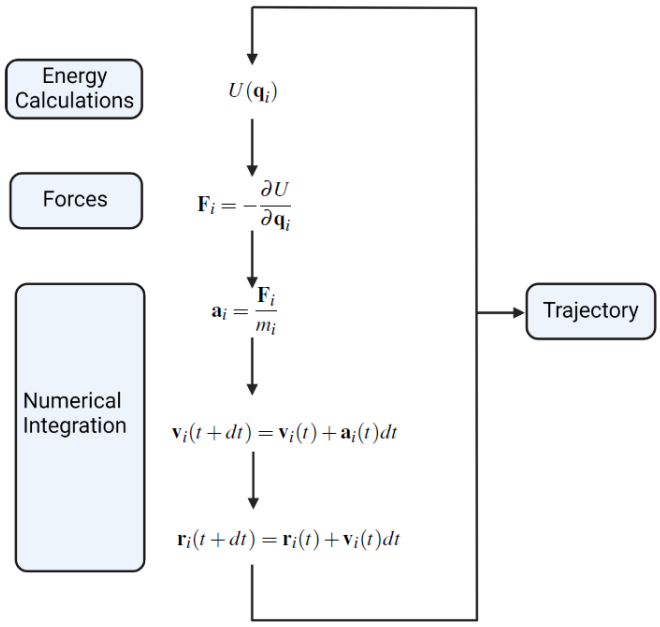

Figure 2.1.: An overview of the basic steps in MD simulation.

electrostatic interactions (Figure 2.2). Here, we provide the potential energy function $U(\mathbf{q})$ applied by several CHARMM FFs:

$$\begin{aligned} U(\mathbf{q}) = & \sum_{bonds} k_b(b-b_0)^2 + \sum_{angles} k_\theta(\theta-\theta_0)^2 \\ & + \sum_{dihedrals} V_\varphi(1+\cos(n\varphi-\delta)) + \sum_{impropers} k_\omega(\omega-\omega_0)^2 \\ & + \sum_{i,j(LJ)} \varepsilon_{ij}^{min} \left[\left(\frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] + \sum_{i,j(Coul.)} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \\ & + \sum_{Urey-Bradly} k_{UB}(s-s_0)^2 + \sum_{CAMP} U_{CAMP}(\phi,\psi) \end{aligned} \tag{2.4}$$

The potential energy function describes the energetic contributions of intramolecular motions, including bond oscillation around equilibrium bond length b_0 with spring constant k_b, angle oscillation about the equilibrium angle θ_0 with force constant k_θ, torsions around two sequentially covalent bonds, which can be defined using the dihedral angle φ, periodicity n, phase shift δ, and energy barrier V_φ, as well as improper torsion (out-of-plane bending) with the equilibrium angle at ω_0 and force constant

k_ω. The intermolecular interactions contribute energy through both Lennard-Jones (LJ) and Coulombic interactions between two non-bonded particles i and j. The LJ potential, a $12-6$ potential, consists of two terms: The repulsive term $1/r^{12}$ approximates the strong repulsion at short distances between interacting particles caused by overlapping electron orbitals, and the attractive term $1/r^6$ accounts for the attraction forces induced by London dispersion forces, which are also referred to as van der Waals (vdW) interactions. The distance separating a pair of interacting particles i and j is denoted by r_{ij}, the potential well-depth is represented by ε_{ij}, as well as, $R_{min,ij}$ corresponds to the distance where the LJ potential between two interacting particles reaches its minimum value. This distance is associated with the van der Waals radius of particles i and j. The Coulomb potential approximates the electrostatic interactions between atoms i and j, where each atom has a partial atomic charge q_i and q_j, respectively, and are separated by distance r_{ij}. The permittivity of vacuum is represented by ε_0. The bonded and nonbonded terms described above are common to all atomistic biomolecular force fields, such as CHARMM [110], AMBER [111], GROMOS [112], and OPLS-AA [113]. However, different parameterization and constraint methods may result in additional terms.

The potential energy of CHARMM force fields includes two correction terms to improve its accuracy: (i) The Urey-Bradley (UB) term was added to refine the accuracy of angle bending modeling due to the coupling between the bond length and the bond angle, it is estimated by harmonic function describing oscillation around equilibrium distance s_0 with force constant k_{UB}, where s represents the distance between the first and third atoms that determine the bond angle. This term was added to improve the consent of the vibrational spectra, thus the term didn't add to most force fields. (ii) The second correction term the CMAP potential was developed to refine the sampling of the protein backbone conformations. This correction term is a two-dimensional grid-based energy correction that considers the correlation effects between backbone dihedral angles ϕ and ψ. The CMAP potential is calculated using quantum mechanics calculations and further optimized using experimental data [114–116]. Although CMAP correction was initially implemented to CHARMM22 FF, it has since been applied in other force fields, such as AMBER12SB [117], AMBER19SB [118] and AMBER99IDPs [119] to improve the protein ensemble modeling, particularly for IDPs.

The potential energy in equation 2.4 depends on the positions of all particles ($3N$) in

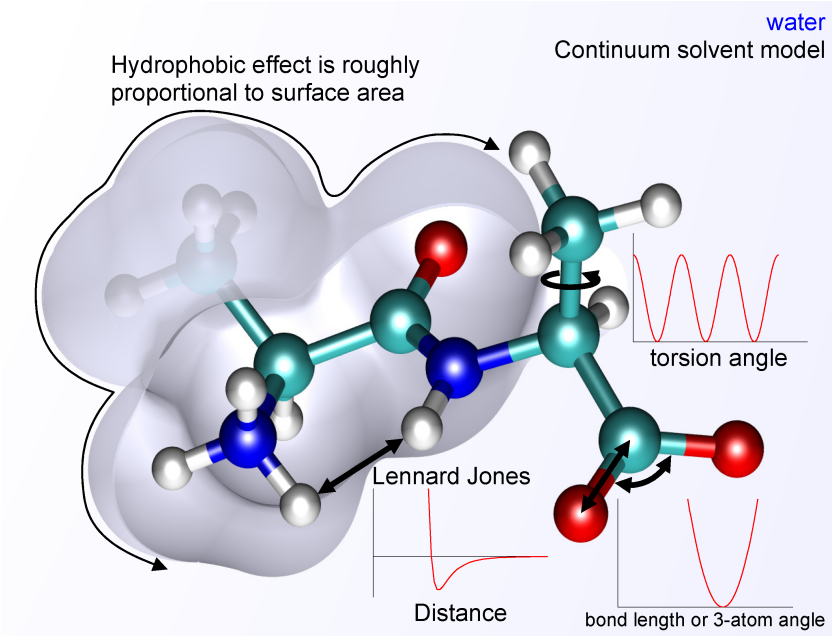

Figure 2.2.: A schematic representation of the components of all-atom force fields used in molecular dynamics simulations. The interactions among the atoms within the system can be categorized into two types: bonded and nonbonded interactions. The harmonic potentials are utilized to characterize the oscillation of bonds and bond-angle bending and the torsion around bonds is modeled by periodic functions. The non-bonded interactions, described by Coulomb and Lennard-Jones potentials, consider interactions between atoms that separated by a minimum of three bonds or between atoms belonging to different molecules. Reprinted with permission (https://en.wikipedia.org/wiki/Force field (chemistry)).

the system, leading to certain complexity for systems containing many particles. This complexity imposes solving the equations of motion by numerical methods, which is not possible by using the analytical solution. The integration algorithm used in this work will be introduced later in Section 2.1.4.

2.1.3. Water Models

Given the central role of water in biological systems, there has been of significant interest in developing high-quality interaction potentials on the molecular scale to simulate its properties. Therefore, many water models have emerged and improved over the past decades [120]. The water surrounding a protein can be modeled uti-

lizing explicit water models, such as TIP3P, TIP4P, TIP5P [121–123] and Berendsen's SPC/E models [124], or using an implicit solvent model [125]. In the case of IDPs, explicit modeling of the water molecules generally produces better outcomes.

The simplest and most commonly used atomistic water models are rigid, fixed-charge, and non-polarizable models [121, 124]. The TIP3P water model is one of the popular models due to it is computational efficiency as well as it is simplicity making it convenient for atomistic simulations in most cases [126]. TIP3P represents a rigid, 3-site water molecule whose three atoms are charged and assigned Lennard-Jones parameters. The model was reparameterized by Jorgensen with 3 site models of transferable intermolecular potential (TIP) functions [121, 127]. Lifson et al. describe the interactions between the sites of the water molecules by 12-6-1 functions which consist of the Lennard-Jones potentials of the short-range interactions between the oxygen atoms plus Coulomb potentials for long-range interactions between all intermolecular pairs of charges, respectively [128]. The overall TIP3P energy is the sum of dimerization energy for water molecules m and n:

$$E_{TIP3P} = \sum E_{mn} = \sum \left[\frac{A}{R_{oo}^{12}} - \frac{C}{R_{oo}^{6}} + \frac{1}{4\pi\varepsilon_0} \sum_{i}^{on\,m} \sum_{j}^{on\,n} \frac{q_i q_j}{r_{ij}} \right] \tag{2.5}$$

where i and j represent the 3 sites of water atoms, A and C are the interaction parameters, and r_{oo} is the distance between two oxygens of the molecules m and n.

2.1.4. Integration Algorithm: The Leap-Frog Algorithm

Several numerical algorithms used to integrate the equations of motion 2.2, include the leap-frog algorithm [129, 130], Verlet algorithm [131, 132], and velocity Verlet algorithm [133, 134]. These algorithms use a Taylor series expansion to estimate the positions and velocities of the particles at each time step, which is defined as a discrete interval of simulation time with a fixed duration Δt [135, 136]. To simulate the behavior of particles over long periods of time, numerical algorithms must be able to maintain energy and momentum conservation while accurately resolving the fastest motions of the system

The leap-frog algorithm is a popular choice for MD simulations due to its time-reversibility and computational efficiency, making it well-suited for simulating complex systems with large numbers of particles. This integrator used a second-order Taylor expansion in which the positions and velocities are not calculated at the same time points, but at alternating time points (Figure 2.3) where the positions are calculated at time t and velocities at time $t+\frac{1}{2}\Delta t$, as in the following equations:

$$\mathbf{v}_i(t+\frac{1}{2}\Delta t) = \mathbf{v}_i(t-\frac{1}{2}\Delta t) + \frac{1}{2}\mathbf{a}_i(t)\Delta t + \mathscr{O}(\Delta t)^2 \tag{2.6}$$

$$\mathbf{q}_i(t+\Delta t) = \mathbf{q}_i(t) + \mathbf{v}_i(t+\frac{1}{2})\Delta t + \mathscr{O}(\Delta t)^2 \tag{2.7}$$

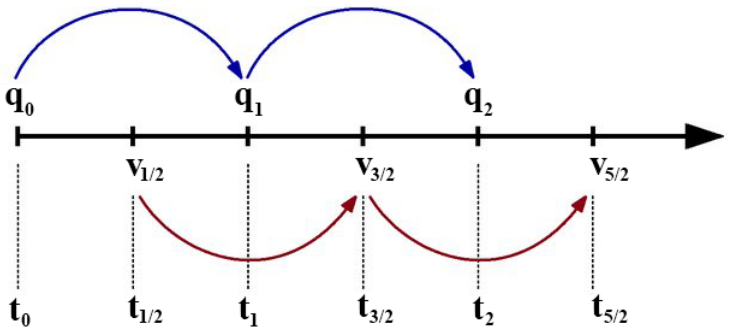

Figure 2.3.: Schematic representation illustrates the workflow of the leap-frog algorithm.

The leap-frog integrator has an advantage in that it explicitly incorporates the velocities in the integration scheme, enabling the system to be coupled to a thermal bath. The velocity at time t can be approximated as:

$$\mathbf{v}_i(t) = \frac{1}{2}\left[\mathbf{v}_i(t-\frac{1}{2}\Delta t) + \mathbf{v}_i(t+\frac{1}{2}\Delta t)\right] \tag{2.8}$$

Thus, the evolution of the positions and velocities can be started by assigning the initial positions and velocities of all the system particles. The initial velocities $\mathbf{v}(t=0) = \mathbf{v}_0$ of the system can be drawn from the Maxwell-Boltzmann distribution at the target temperature T:

$$\rho(v_i) = \left(\frac{m_i}{2\pi k_B T}\right)^{\frac{1}{2}} exp\left(-\frac{m_i v_i^2}{2k_B T}\right) \tag{2.9}$$

where $\rho(v_i)$ is the probability distribution of velocities for particle i, and k_B is the Boltzmann constant. In Eq. 2.9 the velocities are randomly assigned, therefore the initial thermodynamic state of the system could be far from equilibrium. As a result, the velocities required further equilibration, and to achieve this, the temperature is gradually increased until thermal equilibrium is reached at the desired temperature. Next, the rescaled velocities and positions are calculated, and so on until the desired simulation time has been achieved. Here, the estimated errors in the simulation are a result of the numerical approximation used in the integration scheme, which scales with the time step Δt. the estimated errors of the positions and velocities of the leap-frog integrator are on the order of Δt^2. A higher-order integrator, such as the velocity Verlet algorithm, provides a more accurate numerical approximation by using higher-order derivatives of the particle positions and velocities. However, this comes at the cost of increased computational complexity and may not be necessary for all simulations.

2.1.5. Simulation Ensembles

Early MD simulations were mostly conducted on thermally isolated systems, generating microcanonical ensembles (NVE) [137]. In such ensembles, the system's volume V, number of atoms N, and energy E remain constant in the equilibrium thermodynamic state. However, laboratory experiments are usually conducted under constant temperature or pressure, making it desirable to simulate the same experimental conditions. Therefore, simulations in canonical ensembles (NVT) at constant temperature and volume [138, 139] or isobaric-isothermal ensembles (NPT) at constant pressure and temperature [140, 141] ensembles are commonly used to sample configurations.

In the microcanonical ensemble, the total energy is conserved during the simulation. Therefore, fluctuations in temperature of the system result from the exchange of kinetic energy and potential energy. The equipartition theorem can couple thermal energy to kinetic energy, and the atomic velocities can be used to calculate the

instantaneous temperature.

$$\frac{1}{2}Nk_BT = \frac{1}{2}m_iv_i^2 \tag{2.10}$$

To maintain a constant temperature during simulations, atomic velocities must be rescaled. This is accomplished by using a thermostat. Similarly, to maintain constant pressure, the system density must be constant, which requires adjusting the simulation box volume by modifying its dimensions. To achieve this, a barostat is applied. There are several algorithms developed to control the temperature and pressure during MD simulations, such as the extended-ensemble Nosé-Hoover thermostat [142–144], Berendsen thermostat [142, 145, 146], and the velocity rescaling thermostat [147, 148] for simulations at constant temperature. In the case of the *NPT* ensemble, the thermostat is coupled with a barostat such as Berendsen algorithm [149], or the extended-ensemble Parrinello-Rahman barostat [150]. In most cases, the Parrinello-Rahman barostat is employed to preserve the pressure during the simulations and generate the *NPT* ensemble. In the following sections, the Nosé-Hoover thermostat and Parrinello-Rahman barostat, which are commonly applied to preserve constant temperature and pressure in MD simulations, will be introduced in more detail.

Nosé-Hoover Thermostat

The Nosé-Hoover thermostat was first introduced by Nosé [151] and improved by Hoover [152]. In this approach, an additional degree of freedom ζ is added as an external heat path, which allows for maintaining a constant temperature during simulations. The Hamiltonian of the system is modified by adding a potential term $3Nk_B\bar{T}\ln\zeta$ to produce the dynamic equations of ζ, thus ensuring that canonical ensemble averages are retrieved. The resulting Hamiltonian is:

$$H = K + U + \frac{p_\zeta^2}{2Q} + 3Nk_B\bar{T}\ln\zeta \tag{2.11}$$

where K is the kinetic energy, p_ζ is the momentum of the heat path parameter, $\bar{T}$ is the target temperature and Q is the mass parameter of the reservoir that defines the

coupling strength. The modified Newton's equations of motion are given as follows:

$$\frac{d^2\mathbf{q}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} - \frac{p_\zeta}{Q}\frac{d\mathbf{q}_i}{dt} \tag{2.12}$$

The equation of motion of the heat bath ζ given by:

$$\frac{dp_\zeta}{dt} = (T - \bar{T}) \tag{2.13}$$

Here, T is the current system temperature.

Parrinello-Rahman Barostat

In the Parrinello-Rahman barostat, an additional degree of freedom is introduced to modify the Hamiltonian of the system. To maintain constant pressure and density in a simulation, the volume of the simulation box must be adjusted by modifying its dimensions using a barostat. This approach generates the NPT ensemble and the modified Hamiltonian as follows:

$$H = K + U - \sum_i P_{ii}V + \frac{1}{2}\mathbf{W}\sum_{i,j}\left(\frac{db_{ij}}{dt}\right)^2 \tag{2.14}$$

Here, W is the piston mass that determines the strength of the coupling and the box deformation, P_{ii} represents the diagonal matrix of the pressure P, and b_{ij} is the box vectors. The equation of motion of the box vectors is

$$\frac{d^2\mathbf{b}}{dt^2} = \frac{V}{\mathbf{W}\mathbf{b}'}(\mathbf{P} - \bar{\mathbf{P}}) \tag{2.15}$$

where $\mathbf{b}$ represent the box matrix and $\bar{\mathbf{P}}$ are the target pressures.

2.1.6. Periodic Boundary Conditions

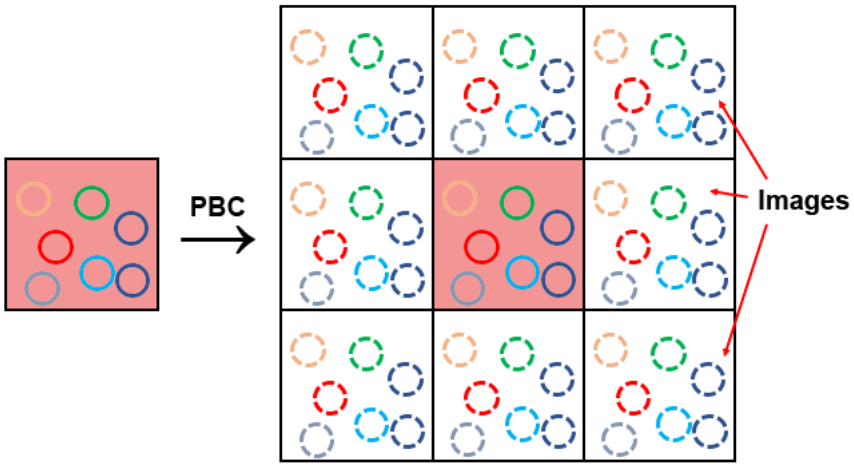

Figure 2.4.: A schematic of a two-dimensional periodic system. The original box is shown with light-red background whereas the image boxes have with a white background.

The periodic boundary conditions (PBCs) are used to enable us to simulate (artificial) large-volume systems and avoid producing artifacts at the simulation boundaries due to the long-range interactions. The PBCs in MD simulations are applied by replicating the system periodically in all dictations, creating an infinite lattice (Figure 2.4). The utilized approaches enables the simulation of a large volume system by means of a finite number of particles [153]. To ensure that the simulated system's particles interact with their closest neighboring particles, regardless of whether they are in the same box or periodic images, the minimum image convention is applied [154]. During the simulation, particles within the periodic images move in a manner identical to that of the particles within the original box. Thus, if a particle exits a box, its image enters the box from the opposite side through a corresponding face. These image boxes have the same volume and shape and contain the same number of particles as the original box. Thus, the coordinates are only saved for the particles in the original box. It is essential to note that imposing PBCs resolve artifacts originating from short-range non-bonded interactions such as the Lennard-Jones forces by truncating the intermolecular forces using a cut-off radius, in this case only images adjoining the original are considered. But, it fails for long-range electrostatic interactions which extend beyond the boundary of a box; truncating these interactions using a cut-off radius produces nonphysical distributions of the particles at the cutoff

boundary along accompanied by discontinuous forces and energies. Consequently, lattice sum methods such as Particle Mesh Ewald (PME) [155] are required for the estimation of electrostatic interactions under PBCs.

An advantage of applying PBC is that it enables us to avoid artificial distributions and at the same time we could reduce computational costs by reducing the number of solvent molecules in the simulation box by selecting a suitable box geometry. Nevertheless, the box should be sufficiently large to avoid self interaction of the images. The simulation box geometry might be chosen as a simple cubic box, as well as an octahedron or a rhombic dodecahedron since most of the protein systems, adopt an approximately spherical shape (Figure 2.5).

Figure 2.5.: Typical box types used in MD simulations of biomolecular systems.

2.1.7. Temperature Replica Exchange MD Simulations

While conventional MD simulations are useful for studying biomolecular systems, they are limited in their ability to explore the free-energy landscape within reasonable simulation timescales and it usually gets trapped in one of the local energy minima, particularly for complex biomolecular systems and protein aggregations. Therefore, many advanced sampling protocols have been improved in the last years to overcome this difficulty [156, 157], including metadynamics [158], Gaussian accelerated molecular dynamics [159], and replica exchange molecular dynamics (REMD) [160]. In this context, REMD is presented as a highly efficient advanced sampling technique utilized for investigating the structure and dynamics of biomolecular systems.

The REMD method was initially introduced by Okamoto et al. to study protein folding [160]. The method simulates several copies (replicas) of the original system in parallel, independently at different temperatures, or using different Hamiltonians at

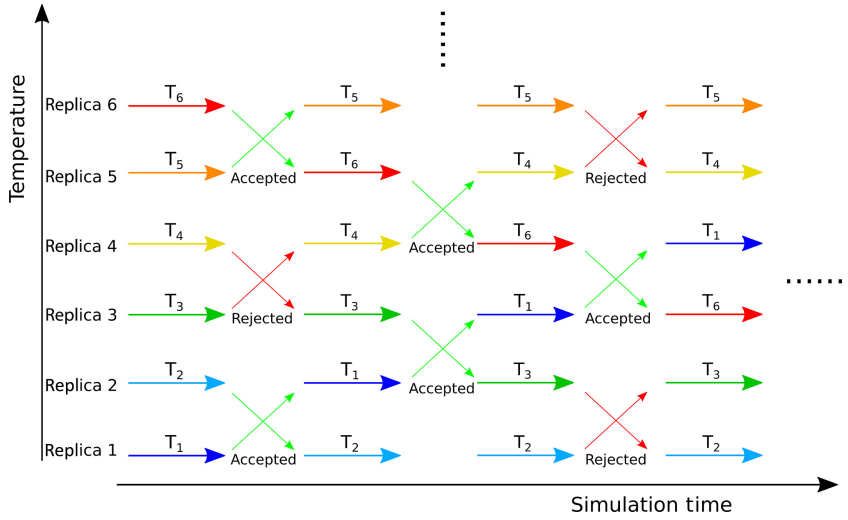

Figure 2.6.: Illustration of REMD protocol. A number of replicas (MD simulations) are carried out in parallel at different temperatures. The exchange attempts every few steps of neighboring replicas with transition probability identified by the Metropolis criterion. Reprinted with Copyright ©2020 Elsevier [162].

the same temperature, to generate a generalized ensemble of the system. In the REMD protocol as shown in Figure 2.6, an exchange is attempted between neighbored replicas every recurring interval with transition probability given by Metropolis criterion [161]. Through this, REMD is capable to enhance the sampling of the conformational landscape by facilitating the crossing of the high-energy barriers at higher temperatures. Another advantage of REMD is that it does not require a prior selection of collective variables, unlike some other advanced sampling techniques. However, REMD demands high computational resources due to the increased size of the system, which requires a high increase in the number of replicas and, therefore, the number of degrees of freedom

REMD Method. Assume a system of N particles of mass m_k $(k = 1,\ldots,N)$, coordinate vectors $q \equiv \{\mathbf{q}_1 \ldots \mathbf{q}_N\}$ and momentum vectors $p \equiv \{\mathbf{p}_1 \ldots \mathbf{p}_N\}$, the Hamiltonian of the system $H(q,p)$ [156, 163]:

$$H(q,p) = K(p) + U(q) \tag{2.16}$$

where $U(q)$ is the potential energy and $K(p)$ is the kinetic energy:

$$K(p) = \sum_{k=1}^{N} \frac{\mathbf{p}_k^2}{2m_k} \tag{2.17}$$

The probability of a system to be in a specific state $x \equiv (q,p)$ and temperature T can be determined using the canonical ensemble as:

$$\rho_B(x,T) = \exp[-\beta H(q,p)] \tag{2.18}$$

The inverse temperature, denoted as β, is defined as $\beta = 1/k_B T$, where k_B is the Boltzmann constant. The average ensemble value of the kinetic energy at temperature T can be determined as:

$$< K(p) >_T = \left\langle \sum_{k=1}^{N} \frac{\mathbf{p}_k^2}{2m_k} \right\rangle_T = \frac{3}{2} N k_B T \tag{2.19}$$

In REMD, the generalized ensemble comprising of M non-interacting replicas, each in the canonical ensemble but at different temperatures T_m where m ranges from 1 to M. This arrangement guarantees that each temperature has one corresponding replica. In other words, the mapping between replicas and temperatures is one-to-one, where the replica i $(i = 1,\ldots,M)$ is a permutation of temperature T_m $(m = 1,\ldots,M)$, and conversely stated:

$$\begin{cases} i = i(m) \\ m = m(i) \end{cases} \tag{2.20}$$

Let $X = (X_i^{[i(1)]},\ldots,X_i^{[i(M)]}) = (X_{m(1)}^{[i]},\ldots,X_{m(M)}^{[M]})$ represent the state of the generalized ensemble consisting of M sets of substates $x_m^i \equiv (q^{[i]},p^{[i]})_m$ that are characterized by specific coordinates $q^{[i]}$ and momentums $p^{[i]}$ in replica i at temperature T_m. Since the replicas are independent of each other, the probability of state X in the generalized ensembles is:

$$\begin{aligned}\rho_{REM}(X) &= \exp\left\{-\sum_{i=1}^{M}\beta_{m(i)}H(q^{[i]},p^{[i]})\right\} \\ &= \exp\left\{-\sum_{m=1}^{M}\beta_{m}H(q^{[i(m)]},p^{[i(m)]})\right\}\end{aligned} \tag{2.21}$$

The summations in the last equation are taken over each replica or temperature. Let's now consider swapping a pair of replicas in the generalized ensemble. Assuming exchange replicas i at temperatures T_n and j at temperatures T_m:

$$X = (\ldots, x_m^{[i]}, \ldots, x_n^{[j]}, \ldots) \rightarrow X' = (\ldots, x_m^{[j]'}, \ldots, x_n^{[i]'}, \ldots) \tag{2.22}$$

or in more particular as:

$$\begin{cases} x_m^{[i]} \equiv (q^{[i]}, p^{[i]})_m \rightarrow x_m^{[j]'} \equiv (q^{[j]}, p^{[j]'})_m \\ x_n^{[j]} \equiv (q^{[j]}, p^{[j]})_n \rightarrow x_n^{[i]'} \equiv (q^{[i]}, p^{[i]'})_n \end{cases} \tag{2.23}$$

It should be noted that exchanging temperatures T_m and T_n between the two replicas i and j is equivalent to exchanging the two configurations at the two temperatures. In the last equation, the momentums should be rescaled by the square root of the two temperatures ratio:

$$\begin{cases} p^{[i]'} \equiv \sqrt{\frac{T_n}{T_m}}p^{[i]} \\ p^{[j]'} \equiv \sqrt{\frac{T_m}{T_n}}p^{[j]} \end{cases} \tag{2.24}$$

Thus, the velocities of all the particles in the replica were rescaled uniformly to satisfy the desired temperature conditions in Eq. 2.19. The replica exchange transition probability $w(X \rightarrow X')$ and the backward transition $w(X' \rightarrow X)$ are required to satisfy the detailed balance conditions as the following:

$$\rho_{REM}(X)w(X \to X') = \rho_{REM}(X')w(X' \to X) \tag{2.25}$$

Using Eqs. 2.16, 2.17, 2.21, 2.24 and 2.25, we have:

$$\frac{w(X \to X')}{w(X' \to X)} = \frac{\rho_{REM}(X')}{\rho_{REM}(X)} = \exp(-\Delta) \tag{2.26}$$

Since all the kinetic energy terms in the last equation are canceled out, $\Delta \equiv [\beta_n - \beta_m](U(q^{[i]}) - U(q^{[j]}))$. Thus, the transition probability of replica exchange $w(X \to X')$ is determined *via* the Metropolis criterion [161]:

$$w(X' \to X) \equiv w(x_m^{[i]} \mid x_n^{[j]}) = \min\left(1, \frac{\rho_{REM}(X')}{\rho_{REM}(X)}\right) = \min(1, \exp(-\Delta)) \tag{2.27}$$

According to Eq.2.27, the transition acceptance probability depends on the overlap between the potential energy distributions of the two neighboring replicas. This criterion allows for efficient exploration of the phase space and ensures that the samples generated are representative of the desired distribution.

2.2. Markov State Models

Due to the increasing complexity of simulated biomolecular systems and the need for thermodynamic and kinetic information, common MD analysis is insufficient for interpreting MD simulation data. To address these challenges, several discrete kinetic models, including master equation models or Markov state models (MSMs), have been successfully employed. These approaches are based on the mathematical theory of conformational dynamics introduced by Schütte et al. [164]. Markov state models, in particular, can map free energy minima and transitions (memoryless jumps) between them in a network that underlies the essential structures and dynamics of biological processes, making it easily understandable to humans. Markovian models assume that states transition only based on their current state, not on their previous state [165]. MSMs have the distinct advantage of being able to predict long-timescale

processes by utilizing several shorter trajectories [165]. This is a significant benefit as it allows to model biological processes that occur over extended timescales.

The following sections outline a general workflow for estimating MSMs for molecular systems, time-lagged independent component analysis as a dimension reduction technique, the k-means clustering algorithm, and the theoretical background for estimating Markov state models and multi-ensemble Markov models is briefly discussed.

2.2.1. An Overview of Constructing MSMs

Constructing MSMs involves several steps, including defining the relevant features, dimensionality reduction, microstate identification, transition matrix construction, and model validation. Here, we present a general workflow for estimating MSMs from MD simulation data. The initial step in estimating MSMs is identifying a set of coordinates, known as features, that accurately reflect the process under investigation. For example, intermolecular distance is a useful feature for analyzing protein aggregation, while backbone torsion angles can be used to describe protein folding. However, the features selected can be very high dimensional, making discretization difficult and potentially leading to poor results [166]. To address this issue, linear transformation methods such as time-lagged independent component analysis can be utilized for dimension reduction. This method identifies a set of slow reaction coordinates with maximum autocorrelation at a specific lag time, which is followed by discretizing the conformational landscape into a defined number of microstates using clustering algorithms like k-means clustering. From these microstates, MSMs can be estimated by computing the transition probability matrix at a specific lag time τ. The optimal lag time can be determined by testing different models at various lag times to ensure convergence of the implied timescales. Finally, the Chapman-Kolmogorov test can be used to validate the Markovianity of the model. Additionally, coarse-grained Markov models or Hidden Markov state models (HMM) with fewer states can be generated for a simpler description of the thermodynamics and kinetics of the investigated system.

There are several software packages that can be used to generate MSMs from MD trajectories of molecular systems, such as PyEMMA [167], MSMBuilder [168], and Deeptime [169]. These packages offer a range of methods, such as dimension reduc-

tion techniques, clustering algorithms, Markov modeling, and model validation, to aid in the construction and analysis of MSMs.

2.2.2. Dimension Reduction

Generating extensive MD simulation data for complex biomolecular systems has become increasingly feasible. However, interpreting the high-dimensional time series output from these simulations can be difficult. While a larger number of dimensions provides better insight into biomolecular dynamics, not all dimensions are essential for the underlying slow dynamic events in the data. Therefore, it is possible to reduce the dimensionality drastically while keeping a high number of essential features.

For Markov models, it is desirable to use a linear transformation that maximizes the timescale separation of the components rather than maximizing the variance of the components, as in principal component analysis (PCA) [170–172]. The latter is not necessarily associated with slow kinetics. Therefore, the optimal method for identifying slow collective variables (CVs) is time-lagged independent component analysis (TICA). This method is widely used due to its ability to identify a set of slow reaction coordinates with maximum autocorrelation at a specific lag time [173, 174].

Time-Lagged Independent Component Analysis

TICA is a linear transformation method that was proposed by Schuster et al. [175] for identifying collective variables with maximum autocorrelation at a given lag time. The goal is to define a transformation matrix $\mathbf{U}$ consisting of a set of basis vectors $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_m]$, where $\mathbf{u}_i$ is a collective coordinate vector with m components. Let $\mathbf{x}(t) = [\mathbf{x}_1, \ldots, \mathbf{x}_d] \in \mathbb{R}^d$ be the future vectors with d components. The first step of TICA is to define the mean-free future vectors by subtracting the means $\bar{x}$ from every feature vector $\mathbf{x}_i(t)$, where the mean-free future vectors $\tilde{\mathbf{x}}_i(t)$ obtain by $\tilde{\mathbf{x}}_i(t) = \mathbf{x}_i(t) - \bar{x}_i$. Then, the instantaneous covariance matrix $\mathbf{C}(0)$ and the time-lagged covariance matrix $\mathbf{C}(\tau)$ are calculated:

$$c_{ij}(0) = \langle \tilde{x}_i(t)\tilde{x}_j(t)\rangle_t \tag{2.28}$$

$$c_{ij}(\tau) = \langle \tilde{x}_i(t)\tilde{x}_j(t+\tau)\rangle_t \tag{2.29}$$

The independent components $\mathbf{u}_i$ are obtained after solving the generalized eigenvalue problem:

$$\mathbf{C}(\tau)\mathbf{u}_i = \mathbf{C}(0)\lambda_i(\tau)\mathbf{u}_i \tag{2.30}$$

where $\lambda_i(\tau)$ are the eigenvalues which are related to relaxation timescales as follows:

$$t_i(\tau) = -\frac{\tau}{\ln|\lambda_i(\tau)|} \tag{2.31}$$

Next, the new set of coordinates $\mathbf{y}(t)$ are give by the projection of the future vector $\mathbf{x}(t)$ using the transformation matrix $\mathbf{U}$:

$$\mathbf{y}(t) = \mathbf{U}^T\mathbf{x}(t) \tag{2.32}$$

By selecting a subspace of the transformation matrix $\mathbf{U}$ consisting of m dominant independent components, the dimensions can be reduced from d to m. It has been demonstrated that the slowest TICA component provides optimal approximations to the first eigenfunction and their associated relaxation timescales underlying Markov models [173].

2.2.3. K-means Clustering Algorithm

MSMs require a discrete configuration space, so the input data must be discretized. Clustering algorithms are typically used to achieve this, which group data points into clusters. Clustering aims to maintain high similarity within clusters and maximum

difference between clusters. The K-means algorithm is a popular method for clustering, developed by Lloyd [176]. It partitions data points into predefined clusters iteratively, using centroids as cluster centers. The centroid is usually the mean value of data points within a cluster, minimizing the squared Euclidean distance between all points and the cluster center [177].

The K-means algorithm involves the following steps: (1) randomly select K data points as initial cluster centroids; (2) assign data points to the nearest cluster centroid; (3) calculate new centroids for each cluster; (4) keep iterating steps 2 and 3 until the optimal centroids are defined, where data points are no longer changing clusters. However, different initial centroids can lead to different final clusters since the K-means algorithm may not converge to a global minimum but is stuck in a local minimum. To reach optimal convergence, it is recommended to use the K-means++ algorithm, which initializes the algorithm from different initial centroids [178].

2.2.4. Estimation of Markov State Models

A Markov State Model (MSM) approximates the kinetics of a molecular system by a conditional transition probability matrix that describes transitions among discrete states. Let n be the number of discrete states, denoted as $s_1(t), \ldots, s_n(t)$. The transition probability between state s_i at time t and state s_j after a lag time τ is given by:

$$p_{ij}(\tau) = \mathbb{P}(s_j(t+\tau)|s_i(t)) \tag{2.33}$$

An essential property of an MSM is its ability to predict longer time-scale kinetics using the power of the transition matrix:

$$\mathbb{P}(s_j(t+k\tau)|s_i(t)) = \left[\mathbf{P}^k(\tau)\right]_{ij} \tag{2.34}$$

In case of an infinitely long trajectory, the transition matrix $\mathbf{P}(\tau)$ can be obtained in terms of the count matrix $\mathbf{C}(\tau)$. Each element c_{ij} represents the number of transitions from state s_i to state s_j:

$$\mathbf{P}(\tau) = \frac{c_{ij}}{\sum_j c_{ij}} \tag{2.35}$$

Here, $\sum_j c_{ij}$ is the normalization of the row of the count matrix, representing the number of times the trajectory visited state s_i. However, for limited-length trajectories, the transition matrix is not unique. Different transition matrices can be generated by different count matrices. To construct a certain transition matrix, the product of all these transition probabilities can be taken into account by defining the maximum likelihood transition matrix [179]:

$$\mathbb{P}(\mathbf{C}(\tau)|\mathbf{P}(\tau)) = \prod_{i,j=1}^{n} p_{ij}^{c_{ij}(\tau)} \tag{2.36}$$

and the transition matrix according to the posterior probability is given by:

$$\mathbb{P}(\mathbf{P}(\tau)|\mathbf{C}(\tau)) = \mathbb{P}(\mathbf{P}) \prod_{i,j=1}^{n} p_{ij}^{c_{ij}(\tau)} \tag{2.37}$$

For equilibrium molecular dynamics, an MSM can predict a unique stationary distribution π in terms of the transition matrix by:

$$\pi^T = \pi^T \mathbf{P}^k(\tau) \tag{2.38}$$

which also fulfills the detailed balance (reversibility condition):

$$\pi_i p_{ij} = \pi_j p_{ji} \tag{2.39}$$

Further useful information is obtained by determining the eigenvectors r_i of transition matrix $\mathbf{P}(\tau)$ and eigenvalues λ_i which can provide valuable insights into the dynamics of the system:

$$\mathbf{P}(\tau)\mathbf{r}_i = \mathbf{r}_i\lambda_i \tag{2.40}$$

Each eigenvector $\mathbf{r}_i$ of the transition matrix can be associated with a dynamic process of the system, while its corresponding eigenvalue λ_i provides a metric for the relaxation time of that process. The maximum eigenvalue, $\lambda_0 = 1$, corresponds to the stationary distribution π of the system, towards which the system tends to shift in the long-time limit. All other eigenvalues $\lambda_{i\neq 0}$ correspond to a finite timescale, and higher eigenvalues represent slower processes while lower values represent faster ones. These implied timescales t_i are related to eigenvalues λ_i as:

$$t_i(\tau) = -\frac{\tau}{\ln|\lambda_i(\tau)|} \tag{2.41}$$

As $\lambda_i \to 1$, $t_i \to \infty$, which provides the basis for predicting the stationary distribution π as in Eq. 2.38 from Eq. 2.40.

Despite the advantage of MSMs in estimating the kinetic relations between microstates using the transition matrix, the resulting model might have systematic errors that depend on the state-space discretization and the lag time [180, 181]. To select an appropriate lag time, it should be sufficiently long to guarantee that the system is memoryless but short enough to describe dynamic processes that occur on time scales exceeding the lag time. In other words, the implied timescale for the chosen lag time should be approximately constant within the statistical error. Therefore, the model should be able to predict results at longer timescales ($k\tau$) within statistical error estimated at lag time τ. To validate the kinetic model and determine whether it is Markovian, one can use the Chapman-Kolmogorov test [179]:

$$\mathbf{P}(k\tau) \approx \mathbf{P}^k(\tau) \tag{2.42}$$

The Chapman-Kolmogorov test examines whether the transition matrix $\mathbf{P}(\tau)$, which is calculated using a lag time τ, matches the transition matrix $\mathbf{P}(k\tau)$ calculated at longer lag times $k\tau$ within statistical errors.

2.2.5. Multiensembles Markov Models

Despite the potential of Markov models to reduce the MD sampling problem through adaptive sampling techniques, which involve running multiple short trajectories rather than one long trajectory, sampling rare events and satisfying the detailed balance requirements remains challenging due to the limitations of unbiased MD simulations. The complexity of larger protein complexes also exacerbates the problem, as computer resources are often insufficient to model these processes. To accelerate rare-event sampling, various enhanced sampling methods could be used, including metadynamics [158], parallel tempering [182, 183], umbrella sampling (US) [184, 185], and REMD [160]. These methods use bias potentials or higher temperatures at different ensembles to facilitate sampling. Multiensemble Markov models (MEMMs), developed by Wu et al., integrate the advantages of enhanced sampling methods and MSMs [186]. Wu et al. improved the transition-based reweighted analysis method (TRAM) [187–189] as estimator for MEMM. TRAM has the advantage of extracting thermodynamics and kinetics from high-dimensional MD data gathered from multiple ensembles. In the following, we briefly present the theoretical background of the TRAM.

TRAM Suppose simulations contain K different ensembles with a dimensionless bias potential has a unit of thermal energy $\beta^{-1} = k_B T$. The bias potential of the kth ensemble $b^k(x)$ is modeled depending on the sampling method being used as detailed in the following examples:

For umbrella sampling:

- The simulations are performed with different potential energy functions, where the potential energy of the kth simulation is the sum of potential energy function $U(x)$ and the kth umbrella potential $B^k(x)$.
- The bias potential is given by $b^k(x) = \beta B^k(x)$, where $\beta^{-1} = k_B T$ is the thermal energy unit.

For replica exchange:

- The simulations are carried out at different temperatures $T_1, \ldots, T_k$.
- The bias potential $b^k(x)$ of the kth temperature is related to the reference ensemble as $b^k(x) = U(x)(\beta^k - \beta)$.

The local equilibrium distribution $\mu_i^k(x)$ for the kth ensemble and the state s_i is given by:

$$\mu_i^k(x) = e^{f_i^k - f^k(x)} \mu^k(x) \quad \text{for } x \in s_i \tag{2.43}$$

Here, f_i^k is the local free energy of state s_i in the kth ensemble, f^k is the relative free energy needed to normalize the equilibrium distribution $\mu^k(x)$, and $\mu^k(x) = e^{f^k - b^k(x)} \mu(x)$ is the equilibrium distribution of the kth ensemble, where $b^k(x)$ is the bias potential for the kth ensemble. The simulations are assumed to sample from these local equilibrium distributions, but they may not be in global equilibrium distributions, which is a critical aspect of the MSM framework.

In the TRAM estimator, the local equilibrium likelihood (LEQ) and the MSMs likelihood are combined as follows:

$$\begin{aligned} L_{TRAM} &= \prod_k L_{MSM}^k . L_{LEQ}^k \\ &= \prod_{k,i,j} (p_{ij}^k)^{c_{ij}^k(\tau)} \prod_{x \in X_i^k} \mu(x) e^{f_i^k - b^k(x)} \end{aligned} \tag{2.44}$$

Here, X_i^k is the set of all configurations sampled in the kth ensemble and in the state s_i. The count matrix $c_{ij}^k(\tau)$ represents the number of times the transition from state i to state j is observed in the kth ensemble after a lag time τ. The TRAM likelihood represents the probability that a set of trajectories, sampled from different ensembles, visit a particular sequence of discrete states (L_{MSM}^k) and sample local configurations within these discrete states (L_{LEQ}^k). In the last equations, the transition probabilities p_{ij}^k, the point densities $\mu(x)$, and local free energies f_i^k are unknown. The optimal solution of the TRAM problem can be obtained by maximizing the TRAM likelihood

under the detailed balance and normalization constraints:

$$e^{-f_i^k} p_{ij}^k = e^{-f_j^k} p_{ji}^k \quad \text{for all } i,j,k \tag{2.45}$$

$$\sum_j p_{ij}^k = 1 \quad \text{for all } i,k \tag{2.46}$$

$$\sum_{x \in X} \mu(x) = 1 \tag{2.47}$$

where Eq. 2.45 is the detailed balance equation coupled with the transition probabilities and local equilibrium distributions. However, the detailed balanced constraints impose difficulties in finding the optimal solution. A fixed-point iteration method can be used to solve this problem (for more details see [186]).

TRAMMBAR Estimator The TRAM estimator requires simulations that are longer than the lag time. In REMD simulations, however, the exchange between ensembles is usually short, which makes it difficult to use TRAM to extract thermodynamics and kinetics information. Therefore, a new estimator called TRAMMBAR has been developed. This estimator combines the TRAM method with the multistate Bennett acceptance ratio (MBAR) to extract information from REMD simulations with unbiased MD simulation data [190–192]. The TRAMMBAR estimator analyzes simulation data by splitting it into two parts: (i) REMD simulations data is analyzed with the MBAR estimator, which assumes that the data achieve the equilibrium distributions of the relevant ensembles. (ii) Unbiased MD simulations data, which is not sufficient to achieve the equilibrium distributions of the relevant ensembles, is analyzed with the TRAM estimator. The TRAMMBAR likelihood is obtained as a product of the TRAM and MBAR likelihoods:

$$L_{TRAMMBAR} = L_{TRAM} . L_{MBAR} \tag{2.48}$$

The TRAM likelihood considers the probability of transitions between states and the equilibrium distribution of each state, while the MBAR likelihood considers the free energy differences between states [190, 193]:

$$L_{TRAM} = \prod_{k,i,j} (p_{ij}^k)^{c_{ij}^k(\tau)} \prod_{\mathbf{x} \in X_{TRAM}^k \bigcap s_i} \mu(\mathbf{x}) e^{f_i^k - b^k(\mathbf{x})} \tag{2.49}$$

$$L_{MBAR} = \prod_{k} \prod_{\mathbf{x} \in X_{MBAR}^k} \mu(\mathbf{x}) e^{f_i^k - b^k(\mathbf{x})} \tag{2.50}$$

The solutions of the TRAMMBAR problem can be obtained under the same constraints as the TRAM problem. The algorithm for finding the optimal solution of TRAMMBAR by maximizing the likelihood can be found in [194]. According to the assumption that the X_{MBAR} achieves the global equilibrium distributions, every state s_i visited in the X_{MBAR} must be sampled in a reversible manner. Hence, any irreversible transition in the unbiased MD simulations implicitly becomes reversible if its end states are visited in an X_{MBAR}.

2.3. Transition Networks

Declaration Parts of this section are published in Schäffler, M., Khaled, M., Strodel, B. ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins. Methods. Volume 206, Pages 18-26, ISSN 1046-2023 (2022).

2.3.1. Introduction

The transition networks (TN) approach utilizes geometric clustering, as opposed to the kinetic clustering involved in Markov state modeling, to identify the essential features of protein conformational transitions, including protein folding, IDP conformation switching, and protein aggregation [10, 195–199]. The advantage of utilizing network models in analyzing MD trajectory data lies in their ability to reveal the hidden mechanisms of molecular processes that are masked by the vast amount of MD simulation data. While MSMs can provide insights into biological pathways and their kinetics, the construction of MSMs from MD data is challenging for large biomolecular systems due to fulfilling the criteria, such as detailed balance, that require extensive

MD simulations [200]. In such cases, the TN approach may serve as an alternative to MSMs for the study of nonequilibrium simulations.

To construct a TN, a set of descriptor functions is selected to characterize the process being studied, such as the amount of specific secondary structure elements or the aggregate size and the type and number of intermolecular contacts in protein aggregation studies. The descriptors are applied to each frame of the MD trajectory, mapping it to a set of states and the transitions between these states. This information can then be used to form a transition matrix and create a visual representation.

2.3.2. Background

In order to construct a TN, one has to define a set of descriptor functions $\{f_i\}$ that discretize a given MD trajectory. Application of these descriptor functions to each time point t of the MD trajectory (also denoted as MD frame) generates a state S which contains the values of the descriptor functions for the molecular conformations $x(t)$,

$$S(t) = [f_1(x(t)), f_2(x(t)), \ldots, f_n(x(t))] \tag{2.51}$$

where n is the number of descriptor functions chosen. The key step in the TN analysis is to select a set of descriptor functions that optimally reflects the structures and dynamics of the system. Here, a balance between complexity and simplicity has to be found. While using more descriptors could provide a more precise picture of the process under study, it may also increase the complexity of the network to the point that humans no longer understand it intuitively. The balancing between these two aspects can be achieved by quantifying the quality of descriptor combinations. This requires the selection of descriptors with high sensitivity and no or only small correlation between them. The sensitivity of a descriptor is defined by how much of the theoretically possible value range of a descriptor is indeed sampled during the simulations. A descriptor is highly sensitive toward the system if the complete theoretical value range is present in the data set. The correlation between two descriptors f_i and f_j is defined by the correlation coefficient c_{corr}

$$c_{corr} = \frac{\Sigma_t (f_i(x(t)) - \overline{f}_i)(f_j(x(t)) - \overline{f}_j)}{\sqrt{\Sigma_t (f_i(x(t)) - \overline{f}_i)^2} \sqrt{\Sigma_t (f_j(x(t)) - \overline{f}_j)^2}} \tag{2.52}$$

where $\overline{f}_i$ and $\overline{f}_j$ are arithmetic mean values of the corresponding descriptor values along the trajectory. If the correlation between two descriptors is high, using both of them would not add substantial information to a TN yet would increase its complexity. Therefore, it is recommended to choose only one of the correlated descriptors to generate the TN.

Descriptor Functions

In the following, we describe a number of descriptor functions that are typically used to characterize protein structures and aggregates and may thus be used to generate a TN. All descriptor functions explained are available in ATRANET.

Secondary Structure. One typically wants to know about the formation of α-helices, β-sheets, or other secondary structure elements during conformational transitions of a protein. In terms of a TN, this can be quantified by the number of residues adopting a specific secondary structure, which is calculated by the dictionary of protein secondary structures (DSSP) method [201, 202].

Backbone Torsion Angles. For short peptides, such the Aβ_{16-22} fragment of Aβ it can be advantageous to use the Ramachandran angles φ ($\mathrm{C}-\mathrm{C}_\alpha-\mathrm{N}-\mathrm{C}$) and ψ ($\mathrm{N}-\mathrm{C}-\mathrm{C}_\alpha-\mathrm{N}$) to assign the secondary structure of the peptide instead of applying DSSP. The angles φ and ψ allow identifying extended peptide structures belonging to the β-sheet basin of the Ramachandran space, which by DSSP would be assigned to the random coil state. Here, the peptide- and time-averaged Ramachandran plot is generated first for the MD trajectory under analysis and a k-means clustering is applied to cluster the (φ, ψ) data to a specified number of cluster centers. If N_{res} is the number of peptide residues and F is the number of frames in the MD trajectory, the $N_{res}-1$ (φ, ψ)-tuples of the peptide are saved in the matrix $\mathbf{A} \in \mathbb{R}^{F \times N_{res}-1}$:

$$\mathbf{A} = \begin{pmatrix} (\varphi,\psi)_{11} & (\varphi,\psi)_{12} & \cdots & (\varphi,\psi)_{1(N_{res}-1)} \\ \vdots & \ddots & \ddots & \vdots \\ (\varphi,\psi)_{F1} & (\varphi,\psi)_{F1} & \cdots & (\varphi,\psi)_{F(N_{res}-1)} \end{pmatrix} \tag{2.53}$$

This matrix is then being discretized, producing $\mathbf{A}_c$, by assigning each element to its closest cluster center where whole numbers are used to denote the cluster centers, e.g., 0 for the β-sheet basin and 1 for the right-handed α-helical region. To assign an overall structural tendency to the peptide at the different time points t, the matrix entries along the rows are summed up, producing one number per time frame:

$$a_i = \sum_j (\mathbf{A}_c)_{ij} \equiv a(t) \tag{2.54}$$

It should be mentioned that the numbers chosen to denote the cluster centers need consideration. If one only wants to distinguish between β-strands and α-helices, one can use 0 and 1. Then, low values of a_i correspond to extended peptide structures, while high values indicate that a high amount of the dihedral angles are located in the α-helical area. If one wants to add random coil or right-handed helices to this analysis, the number for this new cluster needs to be larger than $N_{res}-1$ in order to avoid overlap with the case $a_i = N_{res}-1$ where all $N_{res}-1$ residues adopt an α-helical conformation.

Intra- or Interprotein Contacts. Protein structure formation or aggregation are driven by favorable residue–residue contacts, in particular by hydrophobic contacts. Such contacts are monitored by the descriptor that counts the number of intra- or interprotein contacts. For instance, in the case of hydrophobic contacts, amino acids A, F, I, L, M, V, W, and Y are included in the contact calculations. To this end, the minimum distances between all possible hydrophobic residue pairings are determined, and two residues are considered to be in contact with each other if their minimum distance is within the pre-defined cutoff distance. The readout value is the total number of the contacts under consideration for each MD frame.

Salt-bridge Contacts. A special case of contacts are salt bridges that form between oppositely charged amino acids or protein termini. The descriptor that reports on the number of salt bridges counts the number of intra- or interpeptide attractive Coulomb interactions using the distance between the oppositely charged groups, which has to fall below a defined cutoff distance to be counted as a salt bridge.

End-to-end Distance. This descriptor measures the spatial proximity between the termini of the protein. For each MD frame, the distance between the N atom of the first amino-acid residue and the C atom of the last residue of the protein is calculated and assigned to the bin it falls into. For the binning of the distances, a user-defined bin width is used. In order to provide a sensible bin width, one should first identify the minimum and maximum end-to-end distance.

Compactness. The compactness measures how compact or extended a protein or protein aggregate is. It is defined as the ratio between the lowest and largest moment of inertia, multiplied by 10 and rounded to the nearest integer. Thereby, the descriptor's theoretical value range is between 0 and 10, corresponding to a completely extended structure (a stick as the extreme case of a prolate spheroid) and a globular structure (a sphere), respectively.

Aggregate Size. The aggregate size is the number of proteins that assembled into a higher-order complex. To determine this quantity, it is assessed which of the proteins are in contact with each other and are thus neighbors. All proteins that are direct or indirect neighbors form one aggregate. Two proteins are considered to be in contact if any two atoms of them are within a certain cutoff. In the case of a monomer, the descriptor function returns the value 1, for a dimer it is 2, etc. During an aggregation process, monomers and oligomers of different sizes can be present in the system. By default, the descriptor function reports the largest aggregate size present at a particular MD time point.

Spatial Orientation of β-strands. If one studies the aggregation of proteins into amyloid structures, which are characterized by β-sheet layers that form fibrils, one is

usually interested to know whether these are parallel or antiparallel β-sheets. This information can be deduced from the polar and nematic order parameters P_1 and P_2, respectively, that characterize the spatial orientation of the β-strands [203, 204]. These two order parameters complement each other. P_2 describes the systems' orientational order and distinguishes between ordered and disordered conformations, while P_1 indicates the polarity of the system and discriminates between parallel, antiparallel, and mixed conformations. The descriptor function returns values of $+1$ for parallel β-strands, -1 for antiparallel alignments, and 0 or neither of them, including mixed parallel and antiparallel alignments as well as disordered β-strands:

$$f = \begin{cases} 1, & P_2 \geq 0.7 \text{ and } P_1 \geq 0.7 \\ -1, & P_2 \geq 0.7 \text{ and } P_1 \leq -0.7 \\ 0, & \text{else} \end{cases} \tag{2.55}$$

Transition-Network Calculation and Visualization

ATRANET is a Python script developed by our group that calculates transition networks using a combination of descriptor functions as described in the previous section [197]. This tool is user-friendly and can accommodate the addition of new descriptor functions to optimize the analysis for specific systems. The ATRANET script utilizes MDTraj [205] and MDAnalysis [206] to calculate the specified descriptors and generates a transition matrix file that contains the populations and number of transitions between states. The transition matrix is saved as a *.gexf* file and can be visualized using Gephi [207]. The transition matrix visualization presents the states as disks with varying sizes based on their population and edges connecting the nodes, with thickness indicating transition probability and arrows indicating transition directions.

Chapter

3.

Amyloid-β Aggregation

The primary objective of this thesis is to examine the structures and aggregation of Aβ peptides. In this chapter, a summary of our research on Aβ aggregation into oligomers is presented. To achieve this, all-atom MD simulations were performed to investigate the following topics: (i) the formation of small oligomers of Aβ_{42}, (ii) the aggregation of various Aβ variants (Manuscript I), (iii) Aβ_{42} dimerization in both the aqueous phase and at the neuronal membrane (Publication II). The original manuscript and publication are included in this thesis as sections A.2 and A.3.

3.1. Amyloid-β_{42} Oligomers

3.1.1. Introduction

Aβ_{42} peptide is known to be more prone to aggregation and consequently more toxic than Aβ_{40} peptide in both *in vitro* and *in vivo* [208–211]. Due to the disordered nature of Aβ peptides they can not maintain a stable structure in an aqueous environment [212, 213]. Instead, these peptides aggregate into oligomers, protofibrils, and fibrils, with the former two considered intermediate aggregates with lower mass compared to the fibril. Although oligomers and protofibrils lack a specific structure unlike the Aβ fibril, the β-content increases with the increase of the oligomer molecular weight

[214]. Additionally, the β-sheet structures within the oligomers can be both antiparallel and parallel, indicating that they undergo structural rearrangements to form fibrils where the β-sheet arrangement is parallel [215, 216]. Therefore, understanding the aggregation mechanism leading to low-molecular-weight Aβ oligomers and their structures is critical for identifying the oligomerization pathways and designing potential inhibitors of AD.

In our study, we aimed to gain insights into the aggregation mechanism of Aβ_{42}, focusing on dimers, tetramers, and hexamers. To achieve this, we conducted extensive all-atom MD simulations. To ensure accurate modeling of the Aβ_{42} oligomers, we employed the Charmm36m force field, which has been shown to outperform other force fields in the context of peptide aggregation [38]. Our investigation of Aβ_{42} dimerization involved analyzing the thermodynamics and kinetics ensembles obtained through MSM coupled with both conventional MD (cMD) and temperature REMD simulations. MSM is a powerful technique for analyzing the behavior of complex systems over long timescales, by constructing a model that describes the transitions between a set of discrete states. By combining MSM with both cMD and REMD simulations, we were able to explore both the thermodynamic and kinetic properties of Aβ_{42} dimers. Similarly, we explored the thermodynamic and structural properties of Aβ_{42} tetramers through both cMD and REMD simulations, which allowed us to sample a wide range of conformations. To further refine our analysis, we used dimensionality reduction techniques, such as TICA and k-means clustering, to reduce the dimensionality of the simulation data and identify clusters of similar conformations. Finally, we investigated the aggregation into Aβ_{42} hexamers through cMD simulations and TN analysis, which yielded insights into the hexamer conformations and transitions between them. Overall, our study provides a detailed understanding of the structures and aggregation mechanism of low-molecular-weight Aβ_{42} oligomers and may have important implications for the development of therapies for AD.

3.1.2. Results and Discussion

3.1.3. Amyloid-β_{42} Dimer

We utilized multi-ensemble Markov models (MEMM) to characterize the Aβ_{42} dimerization process. Specifically, we employed a transitions-based reweighing analysis method (TRAM) to construct an MSM of Aβ_{42} dimerization, which integrates 40 μs of unbiased MD simulations with 64 μs of REMD simulations (see Materials and Methods). This approach effectively enhances the sampling of the conformational landscape, which overcomes the limitation of cMD simulations in capturing rare events and their reversible transitions. As a result, the MSM provides a comprehensive understanding of the thermodynamic and kinetic properties of the Aβ_{42} dimerization process.

TICA suggests compact and extended Aβ_{42} dimer structures

The time-lagged independent component analysis (TICA) analysis was applied to reduce the dimensions as described in Materials and Methods. The projection of the interpeptide trajectories to the first two TICA reveals slow kinetics across the dimerization landscape. When projecting the sampled configuration space of the 40 μs cMD into these two modes it becomes obvious that IC_1 separates the Aβ_{42} dimer configurations into two states with $IC_1 < 0$ and $IC_1 > 0$ (Figure A.1). A single transition between the two dimer states ($IC_1 > 0$) and ($IC_1 < 0$) was observed through the cMD simulations, and therefore an MSM constructed from the cMD contained two disconnected states. According to TICA analysis, the transitions between the two states along the first independent component constitute the slowest dynamics. The motions along the second independent component are faster, and more transitions between the substates present for $IC_1 > 0$ were observed. These transitions were better sampled in the REMD simulation. Moreover, a larger variety of dimer structures was observed in the REMD simulation, leading to a larger basin for this dimer form.

To obtain a better understanding of how the conformations vary over the TICA space,

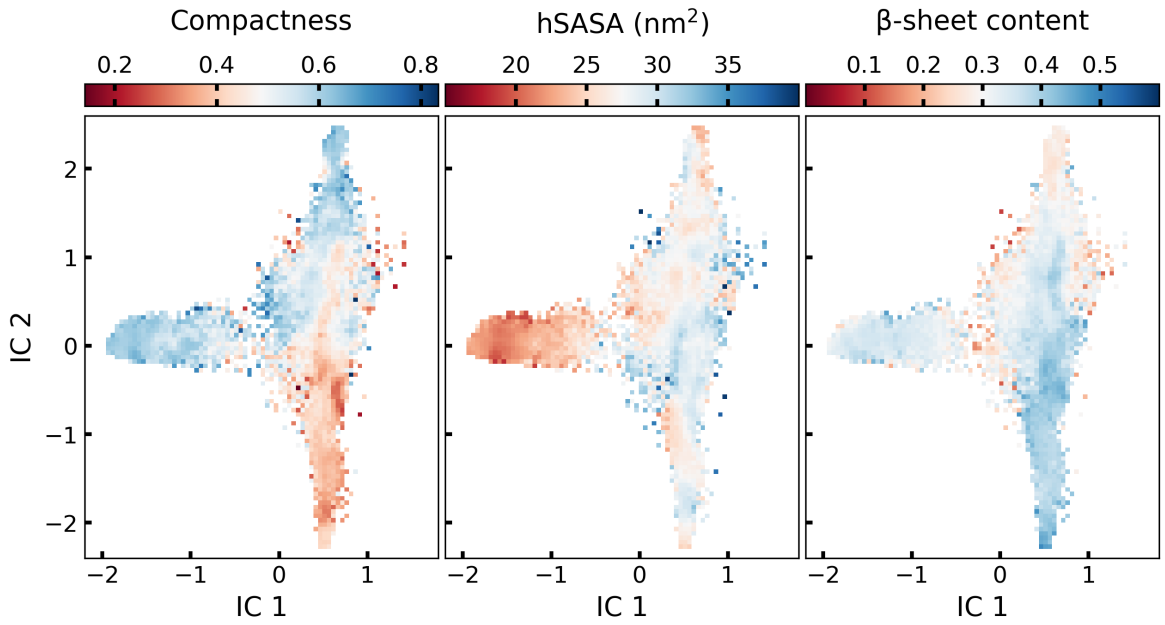

Figure 3.1.: The conformational landscape of Aβ_{42} dimers colored by structural property values, the compactness (left), hSASA (middle), and the β-sheet content propensity (right).

we computed the structure compactness, hydrophobic solvent accessible surface area (hSASA), and the β-sheet content probability (Figures 3.1 and A.2). The compactness decreases with increasing IC_1 values, while it shows a bimodal distribution along IC_2, centering around $IC_2 = 0$ and distinguish between compact structures ($IC_2 > 0$) and extended structures ($IC_2 < 0$). The formation of compact structures ($IC_1 > 0$) is associated with a decrease in hSASA. Moreover, the variation of the structures along IC_2 is also associated with changes in the β-sheet amount, which reduces by the increase of IC_2. The conformational variations along IC_2 suggest a correlation between compactness and the β-sheet content. This can be seen by increasing the β-sheet content, the dimer tends to adopt more extended structures.

Markov state model of Aβ_{42} dimer

To obtain an overview of the Aβ_{42} dimer conformations, an MSM was constructed as described in the Materials and Methods. The conformational landscape was discretized into eight metastable states and an MSM was built using these eight states (Figure 3.5A). An overview of the MSM is presented as a network plot in Figure

3.2 and a representative dimer structure of the metastable states are shown. Each metastable state is represented as a node whose radius is proportional to its stationary distribution. The stationary distribution of the states are shown in table A.1. Edges connecting the nodes correspond to state transformations, with line thickness proportional to transition rates between pairs of states. The MSM has two highly populated metastable states with probabilities of 24-25%, four states with probabilities of 9-13%, and two states with a 3-4% occupancy. In order to identify the different conformations across the metastable states the compactness, β-sheet content probability and hSASA for the metastable states were calculated (Figure 3.2B, C and D).

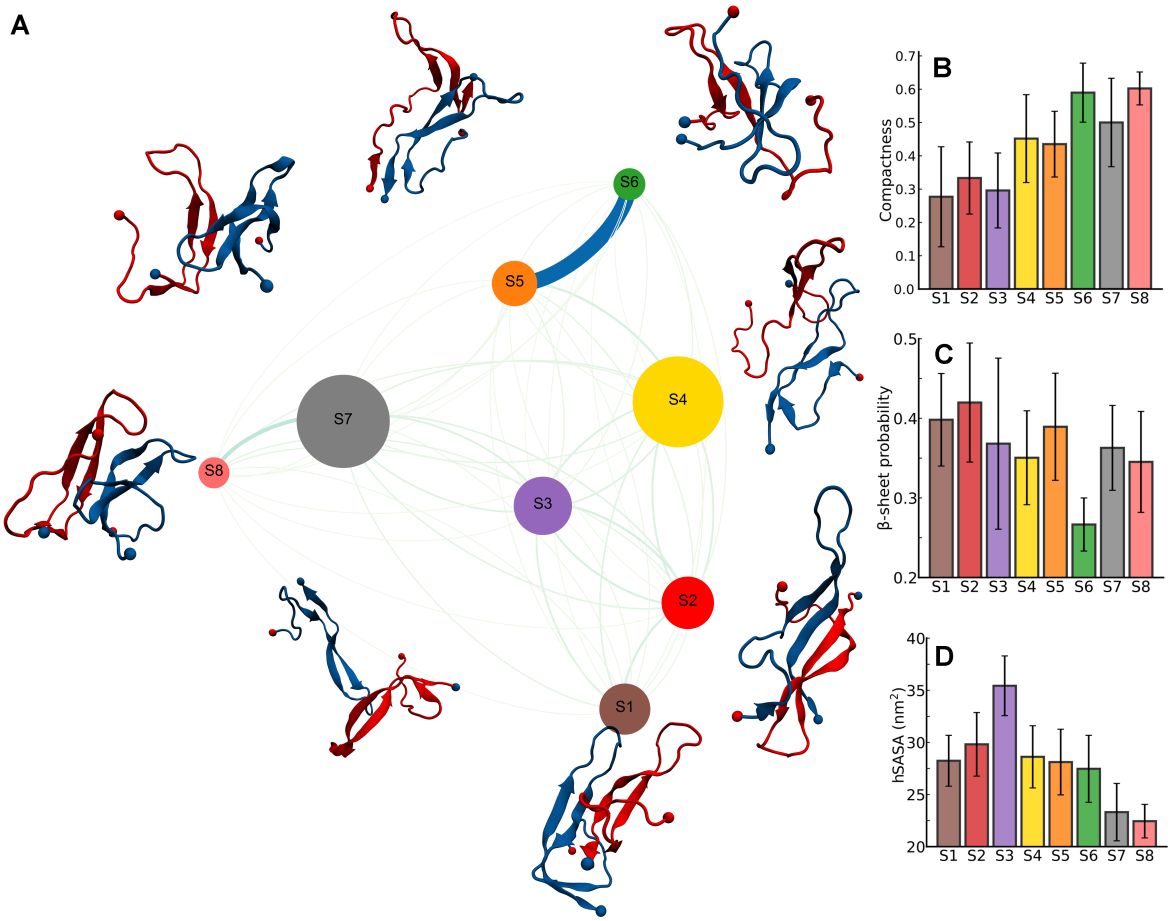

Figure 3.2.: **(A)** MSM of Aβ_{42} dimers with representative structures. The states are shown as nodes and the transitions as edges (the direction of transition is clockwise). The dimers are shown in secondary structure representation and the peptides are colored red and blue, and the N and C-terminus is indicated as red and blue ball, respectively. The average compactness **(B)**, β-sheet content probability **(C)** and hSASA **(D)** for the states are shown as bar plots. The error bars represent the standard deviations.

The analysis of compactness distributions among the eight metastable states of the dimer confirms our earlier observations that it can adopt both extended and compact conformations. States S6, S7, and S8, which together account for 33% of the dimers, mostly contain compact dimer structures, while states S1, S2, and S3 (occupancy of 35%) exhibit more extended dimers, and states S4 and S5 (occupancy of 32%) show intermediate extended structures. The dominant state S7, which has a population of 25%, contains a bimodal distribution of compact and extended conformations (Figure A.3), with a higher population of compact structures likely resulting from transitions from more extended conformations in states S2, S3, and S4.

All states display an average β-sheet content of around 38%, which increases to 42% in S2 and decreases to 27% in S6 (Figure 3.2B). Our previous study using the same force field demonstrated that the Aβ monomer mostly adopts disordered conformations with an average β-sheet content of around 15% during a 30 μs MD simulation [36]. Therefore, the dimerization enhanced the β-sheet formation. Nevertheless, the amount of β-sheet sampled through this study exceedes that of previous simulations, where the β-sheet amount ranged between 11.3% to 24.3%, depending on the force field used [217]. Even when using the same force field Charmm36m with different water model (an implicit Generalized Born (GB) water model) the amount is about 13% [218]. Moreover, we also observed a correlation between compactness and β-sheet content, with an increase in β-sheet content generally leading to more extended dimer conformations. For example, states S1 and S2, which exhibit the most extended conformations, also have the highest β-sheet content, while states S6 and S8, which have the lowest β-sheet content, display the most compact conformations. This relationship is also reflected in the decreased hSASA of these dimers (Figure 3.2D), indicating the formation of a hydrophobic core in the more compact structures.

Structures analysis

We performed secondary structure analysis to quantify the differences in conformation between the metastable states of the Aβ_{42} dimers. Specifically, we calculated the average probabilities of the secondary structure elements forming β-sheets, distinguishing between intra- and interpeptide β-sheets and β-turns/bends for each state

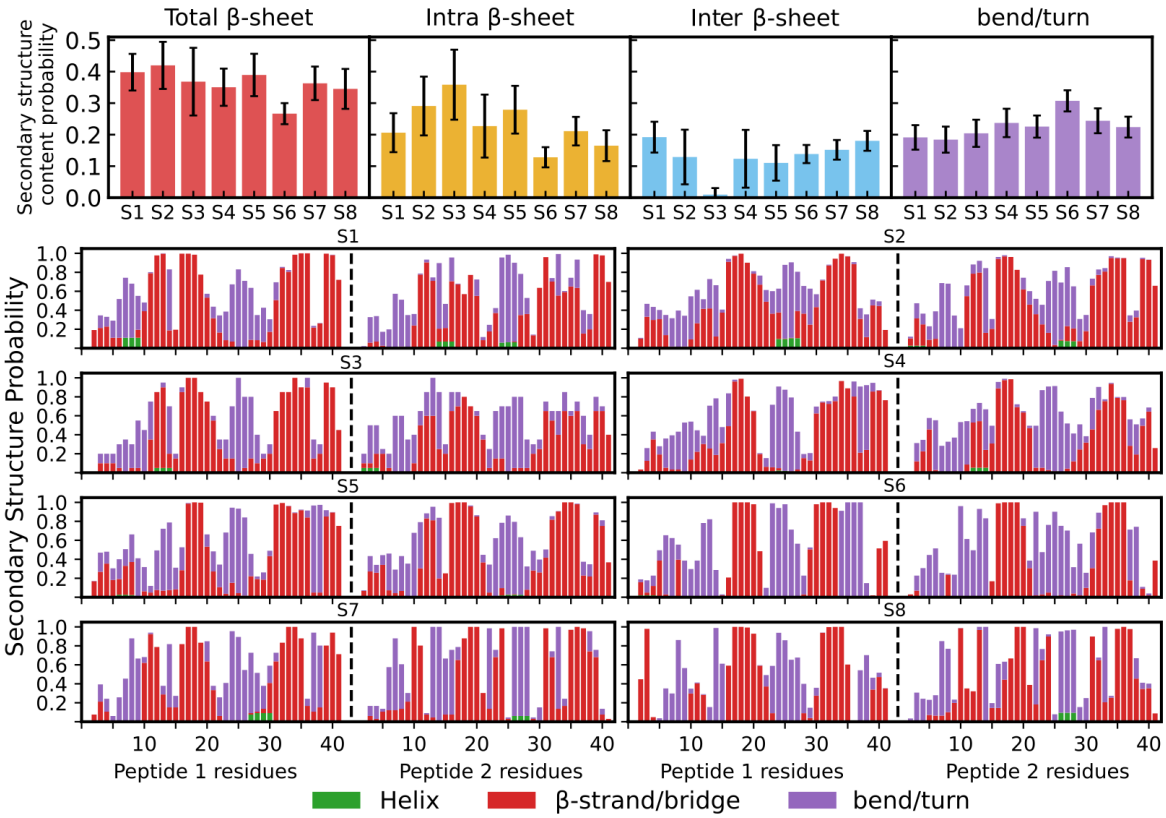

Figure 3.3.: **(A)** The average secondary structure contents of total β-sheet, intra- and interpeptide β-sheet and bend/turn for the MSM states. **(B)** The secondary structure probability of Aβ_{42} dimer peptide residues for the MSM states. The bars represent the additive secondary structure probabilities consisting of helix (green), β-strand/bridge (red), and turn or bend (violet). The difference to 1.0 presents the random coil.

(Figure 3.3A). Our analysis revealed that for most states (S1, S2, S3, S4, and S5), the dimer had more intraprotein β-sheet content. However, this was not the case for the compact states (S6, S7 and S8), where the intra- and interpeptide β-sheet contents were relatively equal. We found that the differences in the structure of the states primarily arose from the formation of turns/bends rather than β-sheet structures. In fact, the average turn/bend content of the dimer was 19% that ranged between 15% to 26%, while random coil structures made up 34% with a range of 32% to 36% of the conformations observed. Interestingly, we did not observe any significant α-helical structures in the Aβ_{42} dimer.

To gain a deeper understanding of the secondary structures present in the Aβ_{42} dimer, we analyzed the probabilities of different secondary structures per residue within

each MSM state (Figure 3.3B). Our results revealed that the central hydrophobic core region (residues Q15 to F20) and the C-terminal hydrophobic region (residues A30 to V40) exhibited high β-sheet probabilities across most states. Specifically, S1 and S2 displayed two long β-hairpins in each peptide, formed in the regions E11 to F20 and I31 to I41, with the exception of H14/Q15 and G37/G38, which tended to form turns. Notably, the residues E11-H14 had a significantly higher β-sheet content in S1 and S2 than in other states, up to 50%. In contrast, S6 generally had two β-strands for each peptide in the central hydrophobic core and G29 to V36 regions of the C-terminal hydrophobic region, occasionally with a third β-strand involving the C-terminal residues I40-A41. Interestingly, we observed β-strand formation in the N-terminal region (residues A2-H6) of S1, S2, and S5, accounting for around 10-20% of the β-sheet content. This is consistent with a disease-relevant Aβ_{42} fibril model [219, 220]. Overall, the Aβ dimer had heterogeneous structures, and high β-sheet content was mainly formed in the two hydrophobic regions but was also observed in other regions, including residues 11-14, C-terminal residues 37-42, and the N-terminal region (residues 2-6). Various experiments have suggested that Aβ adopts a long hairpin structure where both peptides exhibit β-hairpin structures in the regions 10-23 and 30-42 [215, 220].

Intra- and interpeptide interactions

By analyzing the interactions between residues, we observed conformational differences between the various Aβ_{42} dimer states. Intrapeptide distance matrices (see Figure 3.4A) showed that all states had similar contact likelihood in the Y10 to V24 and Q27 to V40 regions, which correspond to the formation of a β-hairpin stabilized by strong hydrophobic interactions in the V18-F20 and I32-L34 regions. Two intrapeptide salt bridge contacts were also present in these regions, specifically E22-K28 and D23-K28 in S1, and D23-K28 in S2 and S3 (see Figure A.4A). The different conformational states of the Aβ_{42} dimer exhibit varying tendencies to form further intrapeptide contacts, resulting in extended and compact conformations. States S6, S7, and S8 have a higher likelihood of forming these contacts, which occur in regions such as A2-Y10/I32-V40 and D1-Y10/Q15-E22 in one of the dimer peptides. The most compact state, S8, has a higher number of contacts within the two peptides, including regions D1-D7/L17-M35 of peptide 1 and A2-V12/Q15-I41 of peptide 2. This

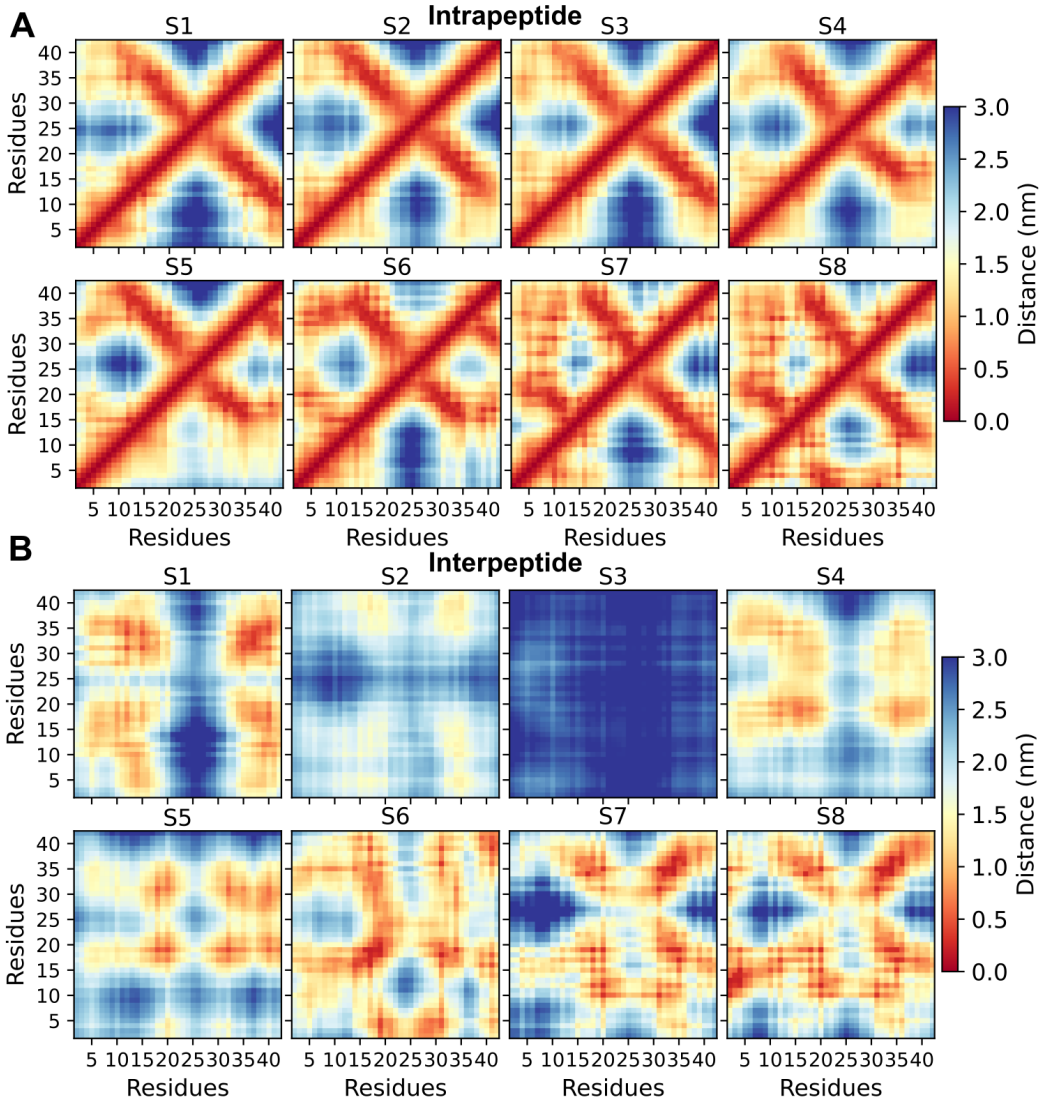

Figure 3.4.: The intra- and interpeptide contact maps for Aβ_{42} dimers. **(A)** The average intrapeptide contacts of the MSM states. The intrapeptide contacts within peptide 1 are shown below the main diagonal and within peptide 2 above the main diagonal. **(B)** The average interpeptide contacts of the MSM states. The color bar on the right shows the average distance between the residues (in nm).

state also exhibits a salt bridge between the N-terminal residue with D22 and E23 (Figure A.4A). The N-terminal salt bridges between neighboring residues E3-R5 and D7-R5 were also observed in this study and are consistent with experimental results [221].

To better understand the interaction interfaces between the two peptides in the

dimers, we calculated the residue-residue distance between the peptides and generated distance matrices for each state (Figure 3.4B). Our analysis revealed that the central hydrophobic core and C-terminal hydrophobic regions are the primary interpeptide interface in all states, except for S3, which represents the central hub state of MSM and harbors weakly connected dimers or two monomers. For S1, the highest contact densities were observed between the two hydrophobic C-terminal regions, A30 to V40, which have high β-propensity. In addition, the dimers exhibited preferred contacts between Y10 to F20 region interacted with the C-terminal region of the other peptide, as well as interactions between the central hydrophobic core regions of both peptides. The two peptides in this state tend to arrange themselves in a parallel β-sheet conformation between the two C-terminal regions (Figure A.5). S1 also showed preferred contact regions between regions 3-10 of one peptide with regions 11-16 and region 35-42 of the other peptide, which could include some interpeptide salt-bridge contacts like E3-K28, D7-K16, and C_{ter}-R5 (Figure A.4B). State S2 shares the same contact-density regions with S1 but with weaker contacts, and it does not show dominant antiparallel or parallel interpeptide β-sheet conformations (Figure A.5).

The compact Aβ_{42} states exhibit distinct interaction interfaces, including the interaction of N-terminal residues with the central hydrophobic core region of the other peptide. In S6, we observe contacts between the N-terminal (residues 1-5) and residues 26-30, with a high interaction tendency between the central hydrophobic core region that favors the formation of a parallel interpeptide β-sheet. On the other hand, S8 displays interactions between the N-terminal and residues 10-15, and frequent interactions between the C-terminal hydrophobic regions of the two peptides, which preferentially form a parallel interpeptide β-sheet. Notably, the interpeptide C-terminal/C-terminal interactions are crucial for stabilizing compact structures and are observed in S6, S7, and S8, with S7 being the most populated state (Figure A.6). This observation suggests that the stability of conformation S7 and its compactness are primarily attributed to the interpeptide C-terminal/C-terminal interactions and the formation of parallel interpeptide β-sheets.

The compact states of Aβ_{42} show a remarkable formation of interpeptide salt bridges involving the C-terminal, including C_{ter}-K16 in S6 and C_{ter}-K28 in S7 and S8 (Figure A.4B). These findings support previous experimental and computational evidence that highlight the crucial role of C-terminal interactions in forming highly stable

dimer structures and driving the aggregation process [222–224]. The importance of C-terminal interactions is further emphasized by the fact that the two hydrophobic residues at the C-terminal of Aβ_{42} enhance the aggregation propensity and neurotoxicity more than other alloforms, such as Aβ_{40}.

In summary, while the primary stabilization of the Aβ_{42} dimer is due to interpeptide hydrophobic interactions between central hydrophobic core regions and C-terminal hydrophobic regions, our study also reveals other potential interpeptide interaction interfaces. These include the N-terminal with central hydrophobic core and N-terminal with residues 10-15. Despite the unstructured nature of the N-terminal region in most amyloid fibrils, it plays a crucial role in forming oligomer species by influencing both intrapeptide and interpeptide interactions. Such interactions have a direct impact on aggregation kinetics by stabilizing the β-hairpin structures.

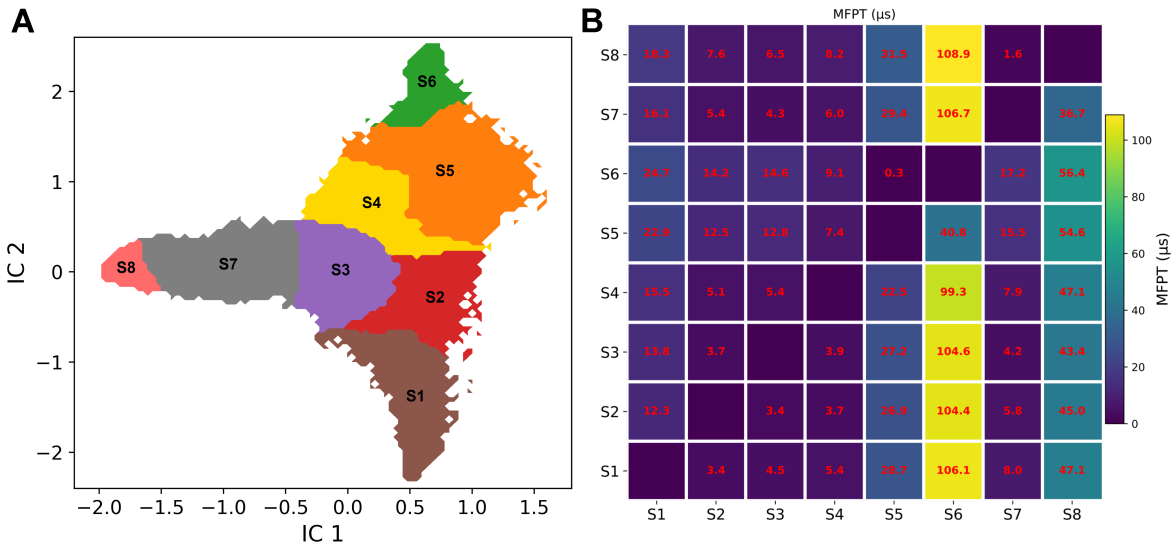

Figure 3.5.: **(A)** The coarse-grained MSM states projected on the first two TICA components. **(B)** The MFPT between the coarse-grained MSM states. The direction of the transitions is from row to column. The color bar on the right shows the MFPT range, while the exact MFPT is given as red number (in μs).

MSM reveal the slow dimerization kinetics

To gain a better understanding of how a protein transitions between different states, we used transition path theory (TPT) to calculate the mean first passage times (MFPT)

between the coarse-grained MSM states (Figure 3.5). Our analysis showed that the time it takes for the dimer to transition between states varies between a few nanoseconds and a hundred microseconds. Of particular interest were the transitions to and from the compact S6 and S8 states. These transitions had the slowest kinetics, with MFPT values of approximately 45 and 97 μs, respectively. In contrast, the transitions from these states to their neighboring states S5 and S7 were the fastest, taking only 1.6 and 0.3 μs, respectively. The transitions between states S2, S3, and S4 were relatively faster among the metastable states. These states are characterized by low-density profiles of intra-/interpeptide interactions. We also observed that the association and disassociation of the two peptides took approximately 0.45 and 5.83 μs, respectively. Overall, our findings provide insights into the kinetics of dimer conformational transitions, which have important implications for understanding Aβ aggregation.

Collisions cross-sections

State	S1	S2	S3	S4
CCS (Å^2)	1421 ± 54	1446 ± 68	1556 ± 90	1408 ± 52
State	S5	S6	S6	S8
CCS (Å^2)	1422 ± 54	1362 ± 61	1463 ± 88	1336 ± 46

Table 3.1.: Collision cross-sections of the MSM states for the Aβ_{42} dimer.

The collision cross-sections (CCS) values were obtained in our simulations by calculating the average of the top 30 cluster structures of the coarse-grained MSM states (Table 3.1). The average CCS over all the states is 1427 ± 91 Å^2 which ranges between 1336 Å^2 for the most compact dimer state to 1446 Å^2 for the most extended dimer state, excluding S3 which contains two monomer structures resulting in higher CCS values. These calculated CCS values are larger compared to the experimental results estimated by an ion-mobility mass-spectrometry (IM-MS) with a value of 1256 Å^2 [225]. In addition, a computational study using different force fields estimates a CCS value in the range 1255-1286 Å^2. However, β-sheet content probabilities in this study ranged between 11-20% [217]. The higher CCS values observed in our simulations can be explained by the fact that our Aβ_{42} dimers have more β-sheets and therefore adopt more extended structures compared to the earlier studies. The CCS

values obtained from the MD simulations are higher than the experimental values, likely because the structures become more compact when transitioning from solution to the gas phase [226]. Furthermore, it is important to mention that CCS values are only approximate and are also subject to uncertainty from the calculation tool.

3.1.4. Amyloid-β_{42} Tetramer

We conducted 22 μs of cMD and 80 μs of REMD simulations to explore the structures of Aβ_{42} tetramers. The resulting trajectories were projected using TICA onto two dimensions to provide a more comprehensive view of the conformational landscape (Figure A.7). Subsequently, we discretized the conformational landscape into 100 microstates, which were then grouped into 15 cross-grained states. Finally, we employed the MBAR algorithm to reweight the probabilities of these states (see Materials and Methods).

Conformational landscape of Aβ_{42} tetramer

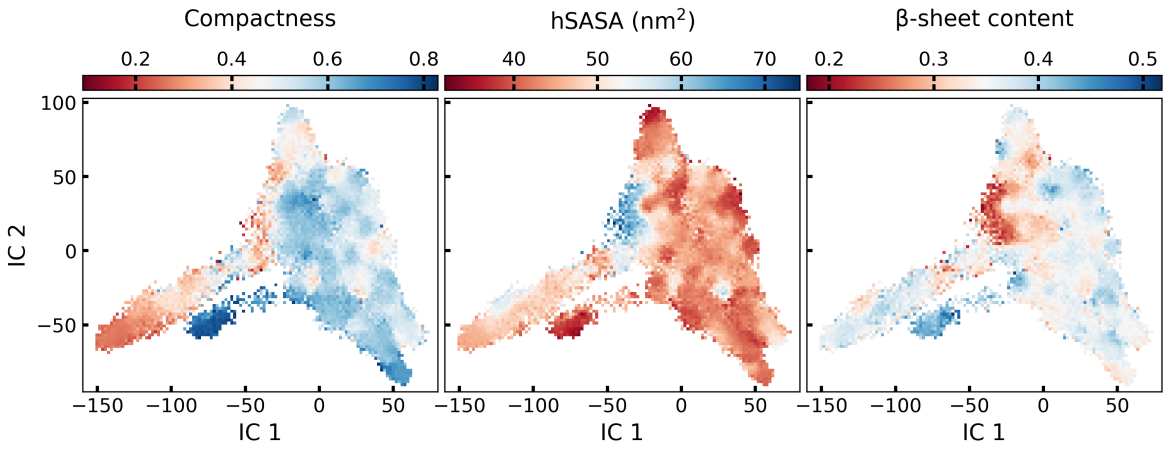

Figure 3.6.: The conformational landscape colored by structural property values, the compactness (left), hSASA (middle), and the β-sheet content propensity (right).

Figure 3.6 demonstrates the conformational landscape of the Aβ_{42} tetramers, which we characterized by three structural properties: compactness, hydrophobic hSASA and β-sheet content propensity. To analyze this landscape, we used TICA to project

the interpeptide Cα-distances trajectories onto the first two TICA components. This analysis suggests that the slow kinetics of the tetramer is associated with forming highly compact structures. The tetramer's conformational landscape is diverse, with multiple polymorphic structures characterized by different distributions of compactness, β-sheet content propensity, and hSASA. By inspecting these projections, we can gain insight into the correlation between these features. However, the non-linear correlation between these features is a result of the intrinsic disorder of the Aβ peptide. The distribution of compactness is partially positively correlated with the amount of β-sheet content, and the distributions of compactness and hSASA suggest that the tetramer's transition to compact structures is a slow process achieved by minimizing hSASA and forming a hydrophobic core.

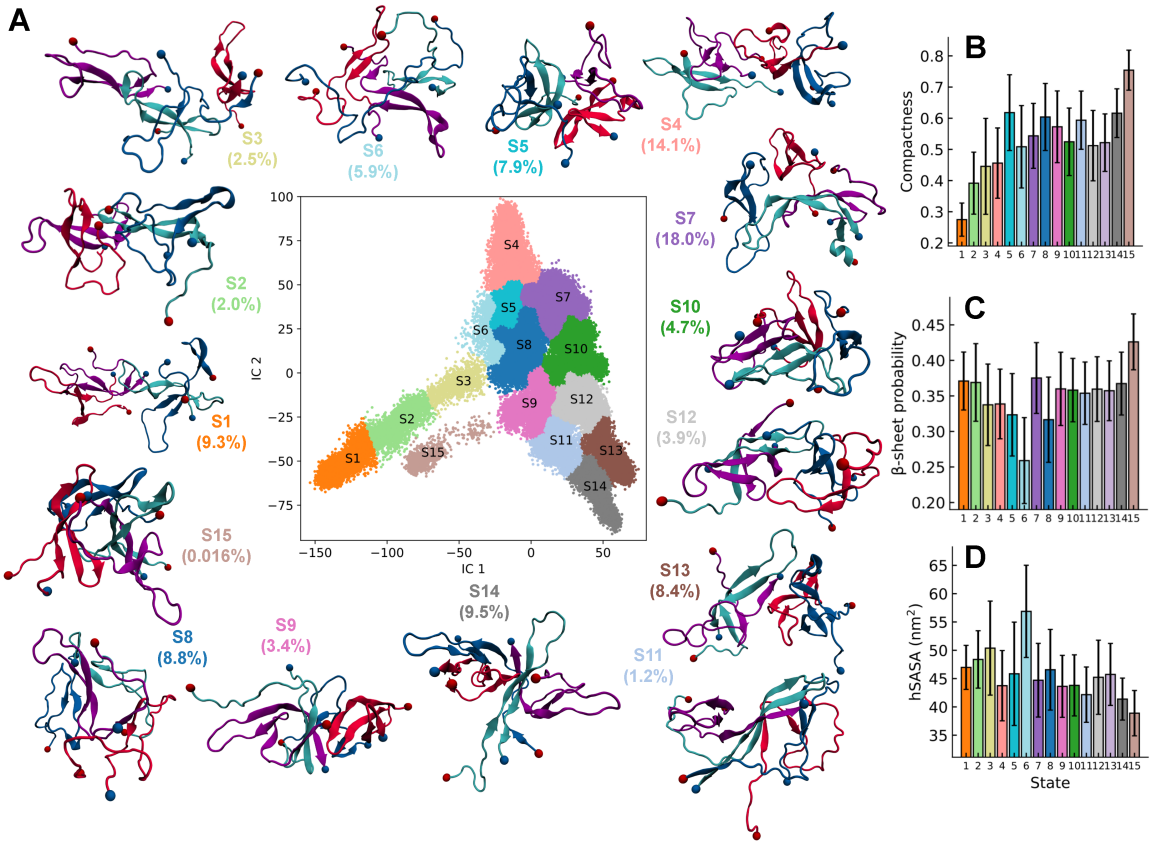

Figure 3.7.: **(A)** The discretized conformational landscape of Aβ_{42} tetramer with representative structures. The tetramers are shown in secondary structure representation and the peptides are colored differently, and the N- and C-terminis are shown as red and blue ball, respectively. The average compactness **(B)**, β-sheet content probability **(C)** and hSASA **(D)** for tetramer states. The error bars represent the standard deviations.

The conformational landscape of the Aβ_{42} tetramer was grouped into 15 states; and their probabilities were obtained by the MBAR reweighting method. Figure 3.7A displays the conformational landscape and representative structures of the tetramer states. The states exhibit diversity in their distributions of compactness, β-sheet content propensity, and hSASA (Figure 3.7B, C, and D). Based on their compactness values, the tetramer shapes can be classified as extended (compactness ≤ 0.5) with a probability of 27.9% (states S1, S2, S3, and S4), compact (compactness ≥ 0.6) with a probability of 27.4% (states S5, S8, S11, S14, and S15), and intermediate compact structures ($0.5 \leq$ compactness ≤ 0.6) in the remaining states. These results suggest that the tetramer's extended structures are less favorable. The formation of compact and extended tetramer structures is associated with lower and higher hSASA, respectively. It is important to note that states that are far apart in terms of their shape index are considered metastable since the transition between them is practically slow. In summary, the tetramer shapes can be classified into extended, compact, and intermediate compact structures based on their compactness values, and the formation of these structures is associated with different hSASA values and β-sheet content probability.

The disordered nature of Aβ implies that even if the structures are kinetically connected, there is still a high degree of variation and flexibility in the configurations within the same states. This can be observed from the large standard deviation error bars in Figure 3.7B, C, and D.

Although the Aβ_{42} tetramer structure is highly disordered, it has a tendency to form both short and long β-sheet structures, with 26% to 42% of peptide residues adopting a β-sheet conformation and an average β-sheet content of 35% (Figure 3.7C). In the majority of states, the peptides tend to form β-hairpin structures or hairpin-like conformations (Figure A.8). The most stable β-sheet structures observed in the residues are those involving the two hydrophobic regions and transit β-strands found in residues 2-5, 10-13, and 38-41. While most states exhibit a tendency for β-hairpin with 2 or 3 β-strands (Figure A.8), some states like S6 and S8 show a reduced probability of β-sheet formation due to completely disordered peptides. The β-sheet conformations, on the other hand, are predominantly composed of interpeptide β-sheet structures formed by different peptides, with an average probability of approximately 22%. Among these, around 13% of the β-sheet content is intrapeptide (Figures A.9 and A.10). State S6 stands out from the others due to the unfolding of one pep-

tide, which leads to a decrease in interpeptide β-sheet content and an increase in hSASA.

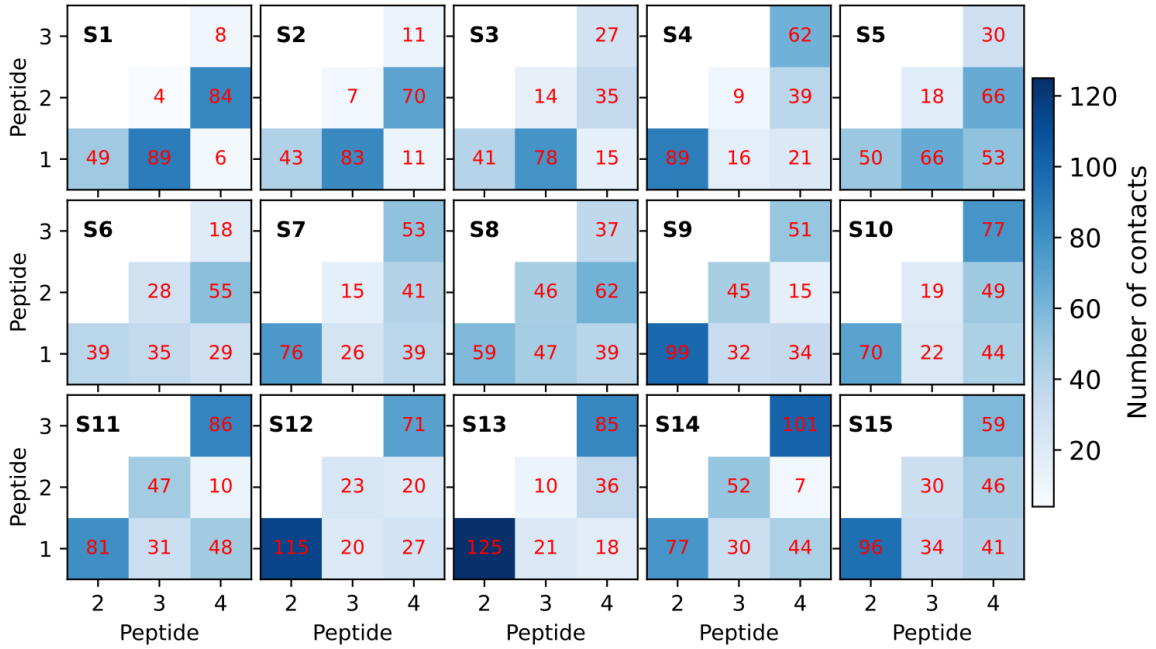

Figure 3.8.: The number of inter-chains contacts for the tetramer states.

Figure 3.8 shows the number of side chain contacts between tetramer peptides. We observed significant differences in the interaction among peptide pairs. In the case of extended states like S1 and S2, the peptide is arranged in a way such as three pairs show notable levels of contact between them and the other three pairs show a few numbers of contacts. In S1, the tetramer is formed by two peptide pairs, (1, 3) and (2, 4), that show high contacts propensity where the peptides 1 and 2 are arranged in the tetramer center and 3 and 4 are at the edges. The formation of compact tetramer structures, on the other hand, leads to more contact formation among the different peptides pairings. Based on this, it can be inferred that the peptides within the tetramer possess a distinct character in which the properties of chains are heterogeneous. This variation in characteristics may be attributed to differences in peptide exposure to the solvent and varying β-sheet content among the tetrameric structures. These findings suggest that the peptides within the tetramer possess a unique character that still differs from fibril structures, where the properties of peptides are homogeneous.

β-barrel structure

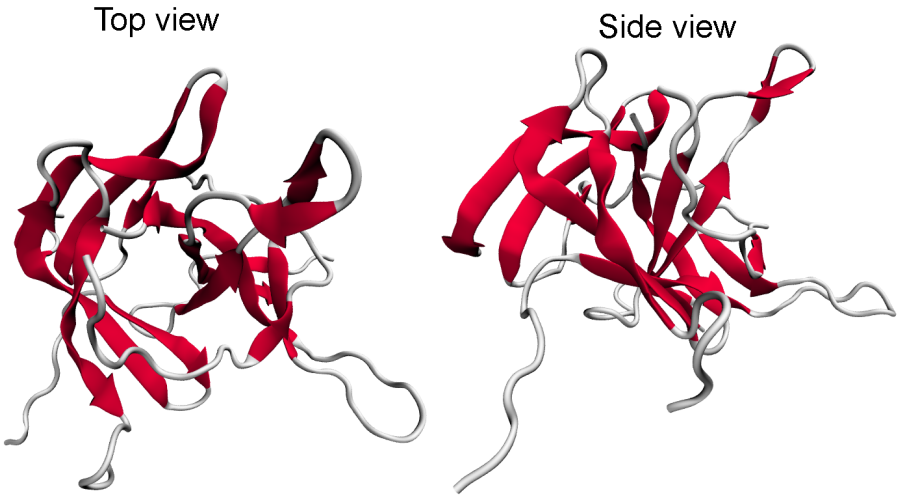

Figure 3.9.: A snapshot for the β-barrel-like structures observed in S15. The peptides are shown in the secondary structures representation and the β-strands are colored in red and other conformations in white.

The experimental structure of Aβ_{42} oligomer is not yet known, but it is suggested to adopt different conformations. One such conformation is the β-barrel structure, which has previously been modeled using truncated Aβ_{42} peptides [227–229]. In addition, a β-barrel conformation has been observed experimentally form by Aβ_{42} hexamer in peptide-micelle environments [230]. The β-barrel oligomer refers to a closed form of the β-sheet, where each β-strand is adjacent to at least two other β-strands. One example of such a β-barrel-like structure is S15 (Figure 3.9), where the tetramer is compact and the peptides have significant interchain contacts (Figure 3.8). However, our simulations show that the probability of this β-barrel-like structure formation is low and unstable (probability 0.016%). This observation may explain why peptide-micelle environments are required to stabilize the β-barrel conformation in experimental studies.

3.1.5. Amyloid-β_{42} Hexamer

Aβ_{42} hexamer structures were sampled by performing 80 μs MD simulations. The resulting trajectories were analyzed based on the transition networks using three descriptors compactness, hSASA, and the β-sheet content probability (see Materials

and Methods). The resulting TN is shown in Figure 3.10, where the clusters are colored differently. The TN for the hexamer has distinct regions, i.e. a region (S1, S2, S3 and S4) containing extended hexamer structures and a region with compact states (S5 and S6), indicating that the hexamer can adopt different shapes. The TN shows that the hexamer prefers to sample more extended conformations with 56% total occupation, while compact conformations are sampled with 44% probability. Moreover, a significant difference in the compactness index between states suggests the presence of a high free energy barrier separating them. Such states, which are distant in terms of their compactness index, are considered metastable due to the slow transition between them.

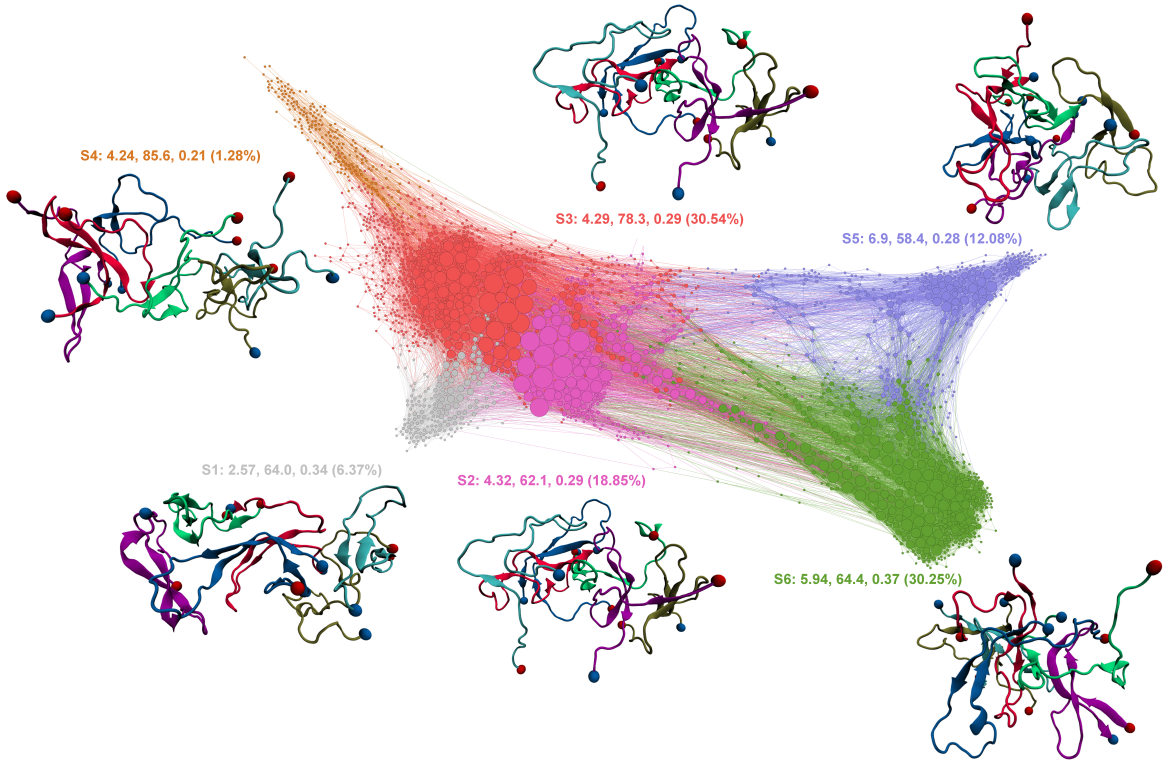

Figure 3.10.: Transition network of Aβ_{42} hexamers using three descriptors: i) the compactness, ii) the hSASA and iii) the β-sheet content. The nodes are colored according to their cluster membership. Representative structures for the most populated state of each cluster are shown. The average descriptor values for compactness, hSASA (nm^2), β-sheet content propensity and probability per cluster in parentheses are given.

The compact clusters can be classified into two groups, based on their characteristics. The larger, dominant cluster S6, consists of structures with 37% β-sheet content and a compactness index of 5.9. The smaller, but the most compact cluster S5, possesses

low hSASA and β-sheet content structures. The clusters S2 and S3 include the nodes with the largest size, which means they correspond to more stable configurations that are visited more often compared to the other states. The clusters have a similar compactness index (4.3) and β-sheet contact propensity (0.29) but differ in their hSASA. The other extended clusters S1 and S4 include nodes with small sizes indicating their low probability and thus the instability of the configurations. Cluster S1 contains the most extended structures with high β-sheet content compared to other clusters. On the contrary, the other cluster S4 is visited less frequently, and with average descriptor values of (4.24, 86, 0.21); the configurations exhibit a large hSASA and lower β-sheet content. The transition from clusters S1 and S4 to the compact clusters S5 and S6 includes S2 and S3 as intermediate states. In summary, the TN reveals that the hexamer samples compact and extended configuration with the latter being more favorable, and the transitions to the most compact states are expected to be a slow process.

3.1.6. Amyloid-β_{42} Oligomers

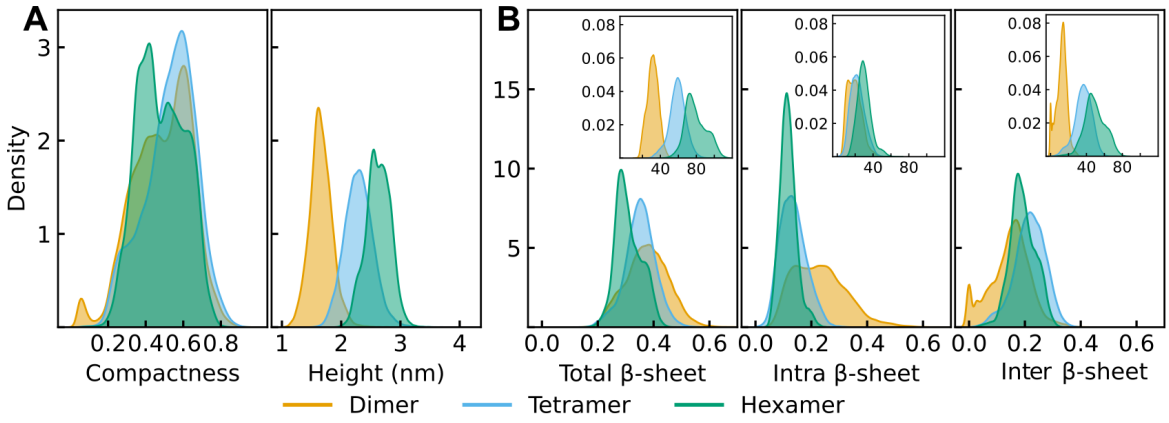

Figure 3.11.: The probability distributions of structural properties for Aβ_{42} oligomers. **(A)** The probability distributions of the oligomer shape index include compactness and height. **(B)** The probability distributions of total β-sheet content, intrapeptide β-sheet content and interpeptide β-sheet content. The inserts show the probability distributions of the number of residues of the respective β-sheet analysis.

To compare the Aβ_{42} dimers, tetramers and hexamers structures, we analyzed the probability distributions of oligomer shape, including compactness and height (Figure

3.11A). Our findings indicate that dimers tend to be more compact. The tetramers are similar to the dimer but have less extended configurations. Hexamers, on the other hand, tend to form more extended structures. Our observations are consistent with data obtained from mass spectrometry, which showed that the isotopic growth of the oligomer deviates from linear growth after reaching tetramer size (see Section 3.2). This indicates that the oligomers undergo a change in growth pattern after reaching a certain size, which is supported by our findings. We also measured the heights of the oligomers, which ranged from 1.2-2.2 nm, 1.6-3.0 nm, and 2.0-3.1 nm for dimers, tetramers, and hexamers, respectively. These results are consistent with early experimental studies that measured oligomer heights ranging from 2-5 nm [231, 232].

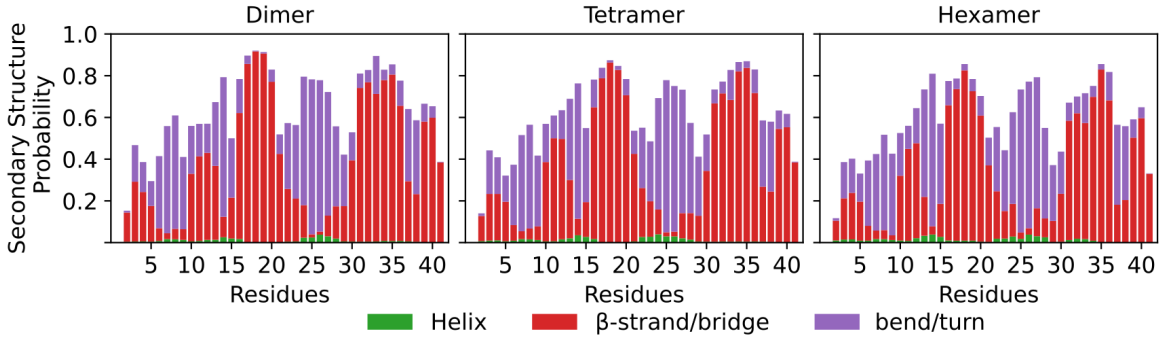

Figure 3.12.: The average secondary structure probability of Aβ_{42} peptide residues calculated from concatenated trajectories for dimer, tetramer and hexamer. The bars represent the additive secondary structure probabilities consisting of helix (green), β-strand/bridge (red), and turn or bend (violet).

We have analyzed the probability distributions of β-sheet content in Aβ_{42} oligomers, including total, intra- and interpeptide β-sheet structures (Figure 3.11B). Our results indicate that Aβ_{42} oligomers are primarily composed of disordered structures, but with a significant amount of β-sheet content. The β-sheet mainly forms by the two hydrophobic regions (Figure 3.12) and lees frequently by residues 2-5, 10-13, and 38-41. Experimental studies have identified turns in the Aβ_{42} oligomer at residues 13-15; 24-27, 25-28, 25-29; and 37-38, which connect β-strands at regions 13-23, 15-24, 17-21; 31-36; and 28-42, 29-42 [232–234]. The amount of the of the β-sheet content decreases with the oligomer size (Figure 3.11B). In particular, the dimer can adopt structures with high amount of β-sheet, which results from the intrapeptide β-hairpin formation. In contrast, these interactions are destabilized with an increase in the

oligomer size, leading to a decrease in intrapeptide β-sheet content. The interpeptide β-sheet distributions, on the other hand, are shifted towards higher probabilities for tetramers and hexamers.

3.1.7. Conclusion

In our study, we investigated how Aβ_{42} assembles into β-sheet rich oligomers by employing all-atom simulations using the force field Charmm36m and the TIP3P water model. A combination of computational approaches containing MD simulations and enhanced sampling MD simulations on microseconds timescales, MSM and TNs analysis was applied, enabling us to reveal the mechanism of Aβ_{42} assembly into dimers, tetramers and hexamers with different structures, containing compact and extended conformations. The compact and extended conformations of Aβ oligomers were previously observed in ion-mobility mass-spectrometry [235], atomic force microscopy [236] and MD simulations [237].

Our analysis reveals that the oligomer contains significant β-sheet conformations, primarily composed of two hydrophobic regions that form β-hairpin structures. However, we also observed transitions to β-sheet structures in other peptide regions, such as residues 2-5, 10-14, and 38-41. These findings are consistent with previous theoretical and experimental studies of Aβ_{42} oligomers [232–234, 238] and experimental data on Aβ_{42} fibrils [239, 240]. Recent studies by Im et al. have shown that mutating hydrophobic region residues with polar uncharged amino acids suppresses the aggregation of Aβ_{42} [241]. This replacement disturbs the hydrophobic interactions and affects the clustering of Aβ_{42} hydrophobic residues, interfering with the assembly dynamics to effectively inhibit amyloid aggregation. Oligomerization provides stability to the β-sheet structure in the oligomer, although the resulting β-content is lower than that found in fibrils. This indicates that the oligomer undergoes structural rearrangement to transition into the fibrils. Understanding the mechanisms behind this transition is crucial, as it may provide insights into the pathogenesis of AD.

The dimerization of Aβ_{42} exhibits a wide range of heterogeneous structures and kinetic features, as revealed by MSM analysis. The kinetic model shows that the transition timescales range from a few nanoseconds to a hundred microseconds. The formation of a stable compact dimer occurs primarily due to inter-/intrapeptide contacts

in the dimer. The extended and compact dimers exhibit rich β-sheet conformations but are aggregated through different mechanisms. The compact dimers are stabilized by interactions within the central hydrophobic core and the C-terminal hydrophobic region, as well as intrapeptide interactions of the N-terminal region. The strong hydrophobic interaction between peptides in the hydrophobic C-terminus makes it more stable than the N-terminus and central hydrophobic core. This highlights the vital role of the C-terminus in stabilizing the oligomer, particularly in the late stages of aggregation when fibril-like aggregates are formed. Targeting the C-terminal region by small compounds has been shown to increase protofibril and fibril formation in several studies [242, 243]. Similarly, targeting the N-terminal region alters the progress of aggregation [220, 231]. These findings suggest that interactions of the N-terminal and C-terminal regions could lead to the aggregation into different oligomer conformations on- or off-pathway toward amyloid fibrils.

The findings of this study have important implications for understanding the contributions of different segments of Aβ_{42} to the aggregation process. Our results show that intrapeptide interactions generally facilitate the formation of β-hairpin structures. However, these interactions can also hinder the unfolding of complexes, leading to a delay in the formation of fibril-like structures. On the other hand, interpeptide contacts are essential for stabilizing interpeptide β-sheets and clustering hydrophobic residues, which promotes oligomerization. These findings suggest that multiple oligomerization pathways exist, and the oligomers must rearrange their conformation to form fibrils or favor the oligomeric state due to strong hydrophobic interactions.

3.1.8. Materials and Methods

Simulations Setup

The MD simulations of the dimer, tetramer, and hexamer systems were conducted using GROMACS 2020.4 [244, 245], with the Charmm36m force field [35] and TIP3P water model. The systems were solvated in a dodecahedron box with water molecules and 150 mM NaCl, and neutralized by adding more Na^+ than Cl^-. Table 3.2 shows the MD sampling methods, atom numbers, box dimensions and temperature for each simulated system. After solvating the peptides, their initial structures were energy-

minimized using the steepest-descent algorithm [246]. The minimized systems were then equilibrated in the NVT ensemble for 200 ps at 298 K using the velocity-rescale thermostat method [247] and a pressure of 1.0 bar using the Parrinello-Rahman pressure coupling method [248, 249]. This was followed by a 200 ps equilibration in the NpT ensemble at 1.0 bar, also maintained using the Parrinello-Rahman pressure coupling method [248, 249]. The production MD simulations were run using the NpT ensemble, with periodic boundary conditions and a real-space cutoff of 1.2 nm. The Particle-mesh Ewald method [250] was used to calculate electrostatics. Bond lengths were constrained using the LINCS algorithm [251], and the equation of motions was integrated using the leapfrog integrator with a time step of 2 fs [252].

Aβ_{42} oligomer	**MD**	**Simulation lengths**	**Atom numbers**	**Box dimensions (nm^3)**	**Temperatures (K)**
Dimer	cMD	$20 \times 2\ \mu s$	43,000	$8.5 \times 8.5 \times 6.0$	298
	REMD	$64 \times 1\ \mu s$	45,000	$8.6 \times 8.6 \times 6.1$	298-400
Tetramer	cMD	$6 \times 3\ \mu s$ $2 \times 2\ \mu s$	100,000	$11.2 \times 11.2 \times 8.0$	298
	REMD	$80 \times 1\ \mu s$	82,000	$11.2 \times 11.2 \times 8.0$	298-400
Hexamer	cMD	$8 \times 10\ \mu s$	152,000	$13.0 \times 13.0 \times 9.2$	298

Table 3.2.: MD simulations setups

REMD Simulations

The temperature REMD simulations [160, 162] were conducted using the same simulation setup, with exchange attempts between neighboring replicas set every 2 ps. For the dimer systems, simulations were performed with temperatures ranging from 298 K to 400 K for 64 replicas, yielding 64 μs of data. The exchange probabilities for all replicas were above 0.30, with an average of approximately 0.34. Similarly, for the tetramer systems, simulations were performed with temperatures ranging from 298 K to 400 K for 80 replicas, yielding 80 μs of data. The exchange probabilities for all replicas were above 0.27, with an average of approximately 0.31.

The consistently high exchange probabilities between neighboring replicas indicate that configurations were successfully exchanged between different temperatures. This is further supported by the observed random walk pattern in the replica space (data

not shown), demonstrating that exchange events occurred between any neighboring replicas.

Starting Structures and Modeling

To explore the dimer conformation landscape, we initially performed unbiased MD simulations for a total of 40 μs (20 $\times$ 2 μs). We started four simulations from 3 μs cluster structures of Aβ_{42} monomers by ensuring that the two monomers were at least 1.0 nm apart. The peptides were added into the simulation boxes using PACKMOL [253]. The initial monomer structures were folded structures similar to β-hairpin conformations, or hairpin-like conformations. The rest of the dimer simulations were started from randomly selected dimer structures from previous simulations. Subsequently, we initiated REMD to accelerate sampling simulations using six initial dimer structures that had different conformational states from the unbiased MD simulations to accelerate sampling.

For the Aβ_{42} tetramer system, we performed unbiased MD simulations for a total of 22 μs by adding two monomers to the dimer structures obtained from the previous dimer simulations. The added monomer structures were folded structures with β-hairpin conformations or hairpin-like conformations. We placed the two monomers at least 1.0 nm away from the dimer structures. To initiate REMD simulations, we selected six tetramer structures from the unbiased tetramer MD simulations from different conformation states.

For the Aβ_{42} hexamer systems, we performed MD simulations for a total of 80 μs (8 $\times$ 10 μs) by adding two monomers to tetramer structures from our previous simulations. The added monomer structures were folded structures with β-hairpin conformations or hairpin-like conformations. We placed the two monomers at least 1.0 nm away from the tetramer structures. The tetramer structures were selected from different conformational states to improve the sampling.

Analysis

We employed various tools to analyze the MD simulation data, including the GROMACS, MDAnalysis [254], and MDTraj Python packages [205]. The analysis of the dimer and tetramer states was conducted by averaging and calculating the standard deviations over all structures in the states. The secondary structure analysis was conducted by DSSP [201, 202]. The secondary structure content probability was calculated as the number of residues in specific secondary structures divided by the total number of residues, and was used to characterize the secondary structure content of the oligomers. The β-sheet structures were classified as either intrapeptide or interpeptide β-sheet, depending on whether they had contact within 0.3 nm with a β-strand in another peptide. The interpeptide β-sheet structures were further categorized as parallel or antiparallel based on the cross product of the residues that make up the interpeptide β-sheet. GROMACS analysis tools were used for contact matrix calculations, number of contacts, and solvent-accessible surface area (SASA) analysis. To determine the residue contacts, the minimum distances were calculated for all pairs of residues, resulting in distance matrices that present the ensemble-averaged distances between the residues. The number of contacts was calculated using a cutoff of 0.3 nm. The hydrophobic SASA (hSASA) values were calculated for specific residue sets (ALA, VAL, ILE, LEU, MET, PHE, TYR, GLY) to characterize the hydrophobic surface exposure of the oligomers. Compactness, defined as the ratio between the lowest and largest moment of inertia of the oligomers, was used to measure how extended or compact the oligomers were. The theoretical range for the compactness value is between 0 and 1, with 0 representing a completely extended structure, like a prolate spheroid, and 1 representing a globular structure, like a sphere. To calculate the height of the oligomers, the semiaxes a, b, c were used, which were obtained from the moments of inertia I_1, I_2, I_3. The calculations were performed using the following formulas: $a^2 = \frac{5}{4m}(I_2 + I_3 - I_1)$, $b^2 = \frac{5}{4m}(I_1 + I_3 - I_2)$, and $c^2 = \frac{5}{4m}(I_1 + I_2 - I_3)$; where m represents the mass of the oligomer. The height of the oligomer was determined by multiplying the smallest semiaxis by 2. The collision cross-section (CCS) values were calculated using Collidoscope with the trajectory method [255].

Markov State Model of Aβ_{42} Dimers

To create a Markov state model (MSM) of the Aβ_{42} dimer, we used unbiased MD and REMD trajectories. First, we selected the Cα–Cα distances between the peptides as features, and used augmented time-lagged independent component analysis (TICAgg) [256, 257] to reduce the number of dimensions. TICAgg was applied to the 40 μs cMD trajectories with a lag time of 5 ns, resulting in a reduction of dimensions from 1764 to two dimensions that describe the slowest kinetics of the system (Figure A.1). Next, we applied k-means clustering [258] to discretize the trajectories into 100 clusters. We then projected the combined data from cMD and REMD simulations onto the first two eigenvectors obtained from TICA of unbiased MD data, and assigned them to the 100 k-means cluster centers. The resulting 100 microstates were used for multi-ensembles Markov model (MEMM) constructions at a lag time of 100 ns using the TRAMMBAR method [186, 194] as implemented in the PyEMMA package [167]. TRAMMBAR combines the Bennett acceptance ratio (MBAR) as a free energy estimator and transition-based reweighing analysis methods (TRAM) to estimate the kinetic. The lag time of 100 ns for MSM was selected based on the convergence of the implied time scales obtained from the cMD data (Figure A.11). We used Robust Perron cluster cluster analysis (PCCA++) [259] to group the microstates into eight macrostates (Figure 3.5). Finally, we used the transition path theory (TPT) [260, 261] to calculate the transitions rate and mean first-passage time (MFPT) between the macrostates.

Dimension Reduction and Clustering of Aβ_{42} Tetramers

To reduce the dimensionality of the Aβ_{42} tetramer trajectories, we applied the time-lagged independent component analysis (TICA) method [256, 257] to all intermolecular Cα-Cα distances, using a lag time of 50 ns. The resulting data was discretized into 100 clusters using k-means clustering [258]. To calculate the free energy of the states, we used the Bennett acceptance ratio (MBAR) approach [186, 192]. In order to refine the clustering results, we grouped the 100 clusters into 15 states using k-means clustering. This was achieved by reducing the variance between three cluster features, namely compactness, β-sheet content, and hSASA. The probabilities of the states are obtained by the summations of the reweighted probabilities of the clusters

in the state.

Transition Networks of Aβ_{42} Hexamer

We constructed transition networks for the Aβ_{42} hexamer using ATRANET [262, 263], using three descriptors to discretize the states and build the transition matrix. These descriptors were the compactness of the protein, hSASA, and the β-sheet content. To discretize these descriptors, we multiplied the compactness indexes by 10 and rounded them to the nearest integer, while hSASA was grouped into 5 nm^2 wide bins (e.g., 65-70 $\rightarrow$ 70 nm^2). We visualized the resulting transition network as a set of states represented by nodes in the network, and the transitions between states were represented as edges connecting the nodes. Using the Gephi software [207], we further clustered the nodes into modularity classes to identify groups of states that are highly interconnected and potentially related [264].

3.2. Manuscript I: Amyloid-β Variants

A hairpin motif in the Amyloid-β peptide is important for formation of disease-related oligomers

Khaled, M., Rönnbäck, I., L. Ilag, L., Gräslund, Strodel, B., A Österlund, N. (2023)

The full manuscript is reproduced in Section A.2.

3.2.1. Background

In this study, we utilized a combination of MD simulations and native mass spectrometry (MS) to investigate the structures and aggregation mechanisms of Aβ oligomers. MD simulations offer a powerful complementary approach to experimental techniques, providing insights into high-resolution structures and dynamics at the atomistic scale. The MS provided detailed information about the different assembly sizes present in a heterogeneous mixture, while the coupling of MS with ion mobility (IM) spectrometry offered additional insight into the oligomer shape through collision cross section (CCS) measurements [265]. We examined Aβ peptides of various lengths, including Aβ(1-40), Aβ(1-38), Aβ(1-28), and Aβ(1-16), as well as Aβ(1-40)$_{CC}$, which contains a double mutation of A21C and A30C in the hinge region of Aβ that forms an intramolecular disulfide bond. The intrapeptide disulfide bond was found to stabilize the oligomers, but it did not result in the formation of amyloid fibrils [266].

3.2.2. Results

MS revealed a different distribution of oligomer states among different Aβ variants (data not shown, see Manuscript I). Specifically, Aβ(1-40) exhibited a wide range of oligomers, from dimers to octamers, without a clear preference for any specific oligomeric state. In contrast, the Aβ(1-40) scrambled variant (Aβ(1-40)$_{Scr}$), which has the same amino acid sequence as Aβ(1-40) but in a random order, showed a limited tendency to aggregate, with the highest oligomeric state detected being a

dimer. Our observations of the wide range of oligomeric states present in Aβ(1-40) indicate a frustrated energy landscape for Aβ(1-40) oligomers, where no state has a clear energetic advantage over the others. Aβ(1-40)$_{CC}$ showed similar oligomeric states as Aβ(1-40), but with a significant increase in larger oligomers, particularly those larger than tetramers. However, when the C-terminus was truncated, the number of larger oligomers was reduced. Aβ(1-28) were mostly observed in dimeric and trimeric states, with the hexameric state being the largest oligomeric state detected.

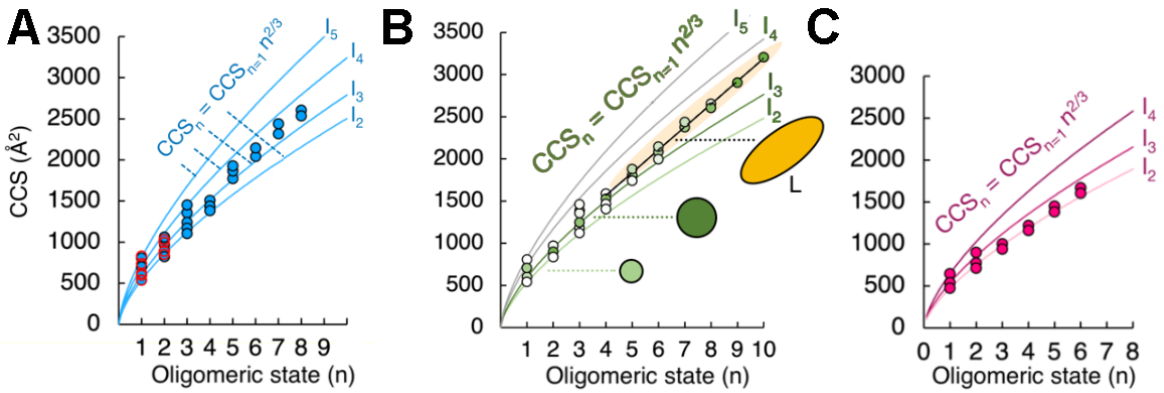

Figure 3.13.: **(A)** Measured collision cross sections (CCS) for Aβ(1-40) (blue circles) and Aβ(1-40)$_{Scr}$ (red circles) oligomeric states. **(B)** Measured CCS for Aβ(1-40)$_{CC}$ oligomeric states are shown as circles colored according to relative intensity within an oligomeric state (white to green). **(C)** Measured CCS for Aβ(1-28) oligomeric states (pink circles). The solid lines represent the theoretical growth behavior of isotropic growth (I) and linear growth (L).

The results from IM spectrometry show that small oligomers of Aβ(1-40) grow in an isotropic manner (Figure 3.13A), meaning that as the oligomeric state (n) increases, the oligomers expand in a spherical shape following the equation $CCS_n = CCS_{n=1} \times n^{2/3}$. The IM spectrometry also revealed different conformations for different oligomeric states, as indicated by their varying CCS values. However, it was observed that Aβ(1-40) oligomeric species continue to expand isotropically up to oligomeric states $n = 4$ or $n = 6$, after which the growth deviates from isotropic expansion and becomes linear. This deviation from isotropic growth suggests that aggregate expansion is occurring at a faster rate in one particular direction, leading to the formation of extended structures. It has been shown that deviation from isotropic growth observed in other amyloidogenic peptides is associated with an increase in β-sheet content [267]. Moreover, this may indicate that monomers, dimers, and

trimers are largely unstructured, while oligomers larger than tetramers begin to form more extended β-sheet structures. It is notable that the Aβ(1-40)$_{CC}$ variant showed a significant increase in larger and linearly growing oligomers, suggesting that the disulfide bridge specifically enhanced the peptides to arrange into oligomer structures (Figure 3.13B). In contrast, it was observed that the formation of Aβ(1-28) oligomers only occurred within an isotropic growth manner, providing additional evidence that the formation of the hairpin β-sheet structure is linked to the aggregation of extended states (Figure 3.13C).

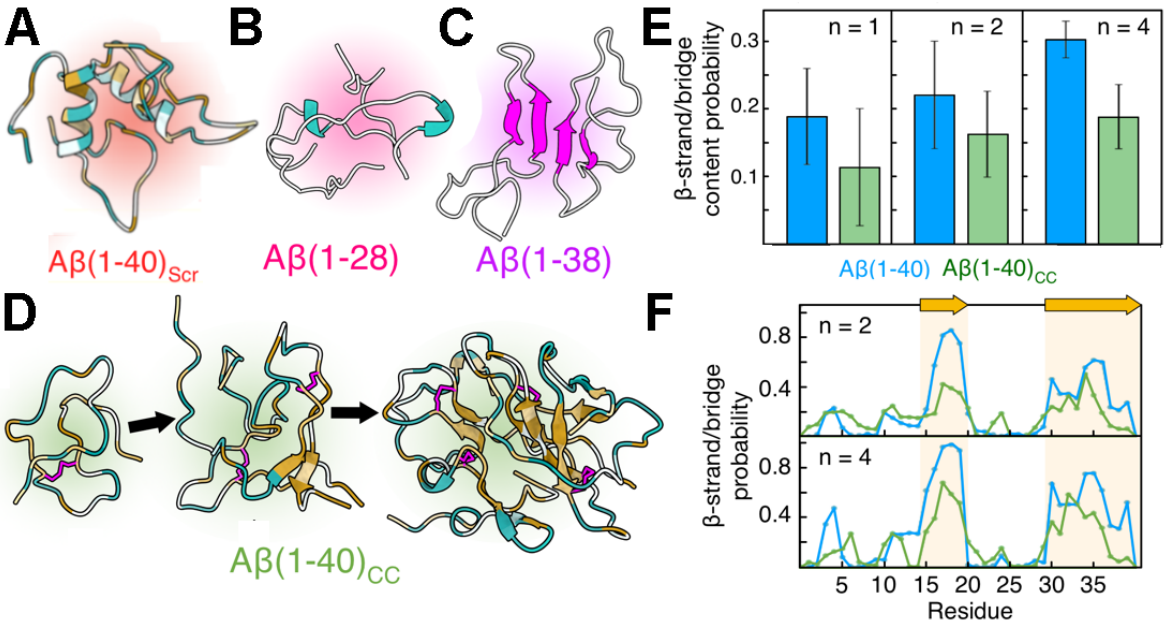

Figure 3.14.: **(A)** The top cluster structure of Aβ(1-40)$_{Scr}$ dimer, colored according to hydrophobicity (orange = hydrophobic, blue = hydrophilic). The top cluster of Aβ(1-28) dimer **(B)** and Aβ(1-38) dimer **(C)**, colored according to secondary structure (coil = gray, sheet = magenta, helix = cyan). **(D)** The top cluster structure of Aβ(1-40)$_{CC}$ monomer, dimer and tetramer. Colored according to hydrophobicity (orange = hydrophobic, blue = hydrophilic). **(E)** Average β-strand/bridge content probability for Aβ(1-40) (blue) and Aβ(1-40)$_{CC}$ (green) for monomers, dimers and tetramers. **(F)** Probability of β-strand/bridge of peptides residue for Aβ(1-40) (blue) and Aβ(1-40)$_{CC}$ (green) for dimers (top) and tetramers (bottom). The two regions with high β-strand/bridge probability are marked in yellow.

The MD simulations ranged from 2 μs to 10 μs per simulation, depending on the system. These simulations were used to determine the secondary structures and other structural properties of the Aβ variants. The Aβ(1-40)$_{Scr}$ variant was found to fold

into amphipathic α-helix through molecular dynamics simulations (Figure 3.14A), which was observed to form only dimers through the MS experiments. This amphipathic helix structure appeared to have the ability to form a relatively stable dimer. Furthermore, the truncation of the C-terminal hydrophobic region in Aβ(1-28) results in the loss of the propensity to form a β-hairpin structure at the monomeric state, which causes it assembly into unstructured aggregates instead of β-sheet oligomers (Figure 3.14B). This is consistent with the low aggregation tendency of this peptide observed by circular dichroism (CD) spectroscopy and the loss of large $(n > 6)$ oligomers in MS. This suggests that the isotropic growth may be associated with less structured aggregates that do not effectively nucleate into amyloid. In contrast, Aβ(1-40) and Aβ(1-40)$_{CC}$ oligomers were found to form β-hairpin structures which are critical for the aggregation growth mechanism (Figure 3.14D). The β-hairpin structures are not stable in the monomeric state which is highly disordered. The MD simulations showed an increase in β-sheet content as the oligomeric state increases, i.e., the β-hairpin structure becomes more stable (Figure 3.14E). However, the increase of the β-sheet with oligomer size of Aβ(1-40)$_{CC}$ is lower compared to the Aβ(1-40) variant. This reduction in the β-sheet content in the CC variant could also be seen by the β-sheet formation propensity by the two hydrophobic regions of Aβ peptide. Both variants formed an antiparallel β-sheet in the central hydrophobic core and the C-terminal hydrophobic regions. However, Aβ(1-40) showed a higher probability of forming β-sheets (Figure 3.14F). The formation of antiparallel structure in Aβ(1-40)$_{CC}$ is compatible with the disulfide bridge, which is also consistent with the oligomer increase observed by MS. The MD simulation of the monomeric state of Aβ(1-38) shows a β-hairpin structure as for Aβ(1-40), which also stays stable in the dimeric state (Figure 3.14C). According to these observations, the formation of the β-hairpin structure and oligomerization rely significantly on the two hydrophobic regions of the Aβ peptide.

3.2.3. Conclusion

Here, we have emphasized using computational and experimental methods that the folding into the β-hairpin motif in the monomeric state is essential for larger Aβ oligomer formation, especially for oligomeric species larger than tetramers $(n \geq 4)$. This β-hairpin consists of the two hydrophobic regions of the Aβ peptides that show

a high propensity to form β-strands. The folding into the β-hairpin enables the formation of oligomers with larger β-sheet structures, which is not possible in the case of Aβ(1-40)$_{Scr}$ that folded into a helical structure. Furthermore, our results showed that the intramolecular disulfide bond does not stabilize the β-hairpin structure in the Aβ(1-40)$_{CC}$ monomer state, which mostly adopts a random coil structure. In addition, the truncation of the C-terminal hydrophobic region of the peptide leads to more disordered structures by critically reducing it is tendency for intrapeptide interactions.

The elongated structures observed for oligomeric states are larger than tetramers, which could be explained as a result of the formation of extended β-sheet structures. These oligomers are likely to have distinct structures, which could later undergo a structural rearrangement to transform into fibrillar structures. The absence of significantly larger oligomers in MS can be explained by the nucleation process starting from elongated oligomeric states. The observation of increased populations of Aβ_{CC} oligomers can be traced back to the inability of Aβ_{CC} to adopt a fibrillar conformation and initiate amyloid nucleation.

3.3. Publication II: Amyloid-β and the Neuronal Membrane

Amyloid-β peptide dimers undergo a random coil to β-sheet transition in the aqueous phase but not at the neuronal membrane

Fatafta H., **Khaled, M.**, Owen M. C., Sayyed-Ahmad, A., Strodel, B. (2021). PNAS, 118(39), e2106210118.

The original publication is reproduced in Section A.3.

3.3.1. Background

Several studies suggested that the source of the toxicity of Aβ oligomers results from its interactions with the neuronal membrane, causing membrane damage or altering the lipid bilayer features. However, capturing these interactions *via* experimental methods is difficult due to their transient nature. Here, we conducted MD simulations of Aβ_{42} dimers in solution and at the neuronal membrane, detecting the effect of the neuronal membrane on the dimerization pathway.

3.3.2. Results

We compared the dimerization of Aβ_{42} in solution and in the existence of the neuronal membrane *via* accumulating 24 μs of all-atom MD simulations. The model lipid bilayer was built to mimic the structure of a neuronal cell membrane. We designed a model lipid bilayer consisting of six lipid types based on the composition of the membrane of a neuron (Figure 3.15) [268–270]: 38% 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 24% 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), 5% 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS), 20% cholesterol (CHOL), 9% sphingomyelin (SM), and 4% monosialotetrahexosylganglioside (GM1). Our simulations excel previous simulations in two terms: First, the simulation time scales of this work exceed previous studies investigating interactions of Aβ and membrane by an order of magnitude. Second, the lipid membrane modeled at the atomistic level with six lipid types is more complex.

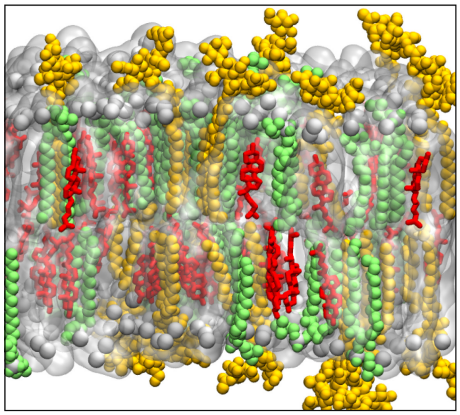

Figure 3.15.: A snapshot of the neuronal membrane containing 38% POPC, 24% POPE, 5% POPS (collectively shown as gray surface with their phosphorous atoms indicated by gray spheres), 20% CHOL (red sticks), 9% SM (green spheres), and 4% GM1 (yellow spheres).

The findings show that dimer formation occurs in both systems but with significant structural differences. The dimer in solution transforms from random coil to β-sheet structures, comparable to those detected in Aβ fibrils. We concluded that Aβ dimers formed in the solution are on the pathway to amyloid aggregation. On the other hand, the Aβ peptide's propensity of β-sheet formation is impaired by the interactions with the neuronal membrane (Figure 3.16A).

The simulations of the Aβ_{42} dimer on the neuronal membrane have revealed that the dimer adsorbs to the membrane, but no insertion of the peptide into the hydrophobic region of the membrane is observed. The membrane adsorption of Aβ is primarily induced by electrostatic and Lennard-Jone interactions between the charged N-terminal region and the central polar region (residues F20 to A30) of Aβ_{42} with the lipid headgroups of POPC, POPE, and POPS, as well as hydrogen bonds formed nearly between all peptide residues with the sugar groups of the GM1 lipids (Figure 3.16B). Since the peptides interact only with the headgroups of the lipids and are not embedded into the membrane, the membrane structure is only slightly affected by those interactions. Moreover, the simulations have show that the main interaction type between the peptides and the membrane are the hydrogen bonds formed *via* the headgroups of GM1 and the peptides, which appear preferable compared to the intrapeptide and interpeptide hydrogen bonds. As a result of that, the β-sheet formation is reduced at the neuronal membrane.

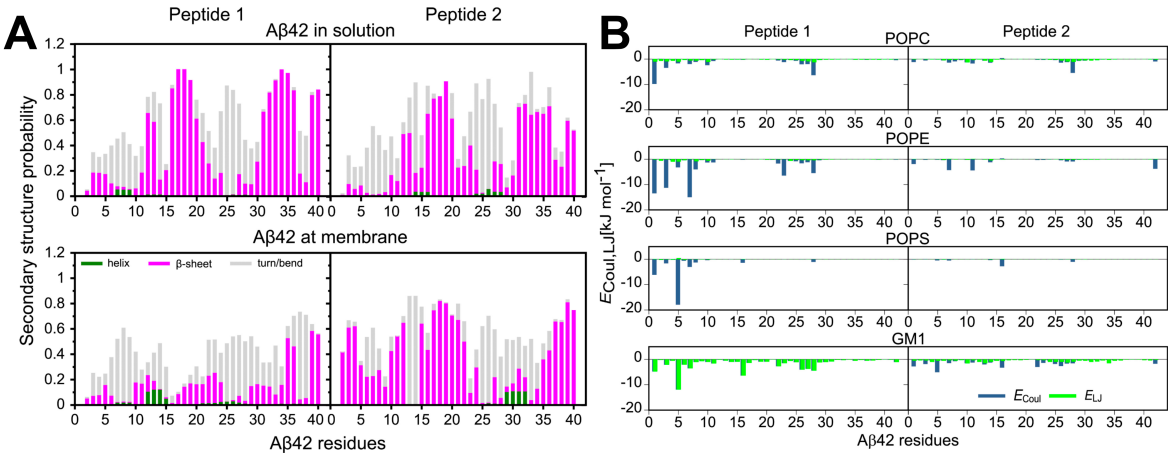

Figure 3.16.: **(A)** Probability of secondary structures to form in each residue of the peptides. The bars represent the cumulative secondary structure probabilities consisting of helix (green), β-strand/bridge (magenta), and turn or bend (gray). The difference from 1.0 presents the probability of the random coil state. **(B)** The average interaction energies of peptide 1 (Left) and peptide 2 (Right) with each lipid of the neuronal membrane. Electrostatic and Lennard-Jones energies are shown in blue and green, respectively. The more negative the energy is, the more attractive is the corresponding interaction.

Furthermore, we used transition networks to elucidate the dimerization pathways for Aβ_{42} both in solution and at neuronal membrane. The TNs were computed using three descriptors to define the conformational states: oligomer size, number of interpeptide hydrophobic contacts, and number of residues in β-strand conformation (Figure 3.17). According to the TNs, the dimerization pathways started with two peptides in monomeric random coil conformations represent by state [1, 0, 0] that lead to dimer assembly with few hydrophobic contacts and β-strand conformation as shown in states [2, h1, b1-b6], which subsequently forms more stable dimers as the interpeptide hydrophobic contacts increase which goes along with the increase of the β-strand content. The TNs exhibit that the dimer in solution has a higher conformational flexibility than the dimer at the membrane, wherein in the latter the dimer-membrane interactions limit the flexibility of the peptides.

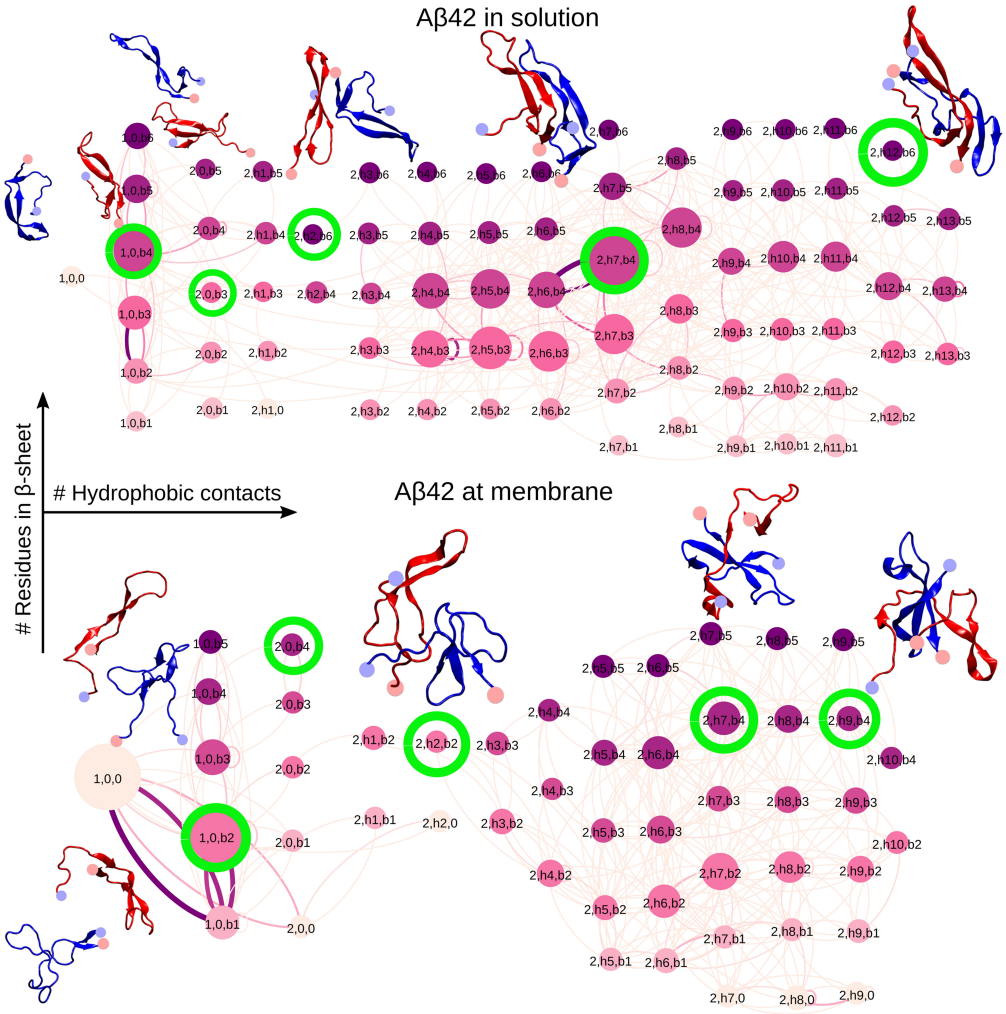

Figure 3.17.: The TN for Aβ_{42} dimerization in the aqueous phase (Top) and in the presence of the neuronal membrane (Bottom). Each node is defined by three descriptors: oligomer size, number of interpeptide hydrophobic contacts, and number of residues in β-strand conformation. The last two descriptors are grouped in blocks of five and are named h1 to h12 for hydrophobic contacts and b1 to b6 for the number of residues in β-strand conformation. The nodes are connected by edges that represent transitions between the connected peptide states. The size of the nodes and the thickness of the edges are proportional to the respective state or transition probability. For the nodes circled in green representative peptide conformations are shown (peptide 1 and peptide 2 are shown as cartoons in red and blue, respectively, with their termini indicated by spheres (N, light blue; C, light red)).

3.3.3. Conclusion

We observed dimerization of Aβ_{42} whether in solution or in the presence of a neuronal membrane, but with significant structural differences. The dimer in solution is distinguished by its high proportion of β-sheet structure, revealing a structural transition from a random coil configuration to a β-sheet structure. The dimer structures in solution exhibit resemblances to the U-shaped structure of Aβ_{42} fibrils. The study found that in the presence of a neuronal membrane, the dimer structure of Aβ_{42} is more disordered and contains less β-sheet content. Additionally, it was discovered that the GM1 lipids of the membrane are the preferred sites for Aβ_{42} to interact with, which is consistent with the findings of previous research. As a result of these observations, the conclusion was drawn that GM1 may provide neuroprotective effects against toxicity caused by Aβ.

Chapter

4.

Huntingtin Proteins

In this chapter, we present a brief overview of the research on the folding and dimerization of pathogenic and non-pathogenic Huntingtin proteins.

4.1. Publication III: Monomers and Dimers of Huntingtin Proteins

Comparative molecular dynamics simulation studies of pathogenic and non-pathogenic Huntingtin protein monomer and dimer

Khaled, M., Strodel, B., Sayyed-Ahmad, A. (2023). Frontiers in Molecular Biosciences, Volume 10:1143353.

The full publication is reprinted in section (A.4).

4.1.1. Background

The expansion of the polyQ tract of Htt-ex1 has significant effects on the proteins' structures and dynamics. In HD, the length of the polyQ tract in the Htt-ex1 proteins must exceed 35 glutamine amino acids in order for it to be considered pathogenic.

Several studies showed that pathogenic Htt-ex1 undergoes a misfolding to β-sheet structures that afterward aggregate into amyloid fibrils. Many pieces of evidence indicate that Htt-ex1 oligomers are the most toxic species to the brain's neuron cells. However, the disordered nature of Htt-ex1 proteins, particularly in the case of the pathogenic ones, imposes difficulties to capture the structures and the aggregation mechanism by experimental techniques. In this study, we performed MD simulations to investigate the structural properties of Htt-Q_n monomers and dimers with $n = 23$ for non-pathogenic Htt-ex1 and $n = 48$ for pathogenic Htt-ex1 (Figure 4.1). These simulations help us to understand the differences between non-pathogenic and pathogenic Htt-ex1 at the molecular level.

Htt-Q_{23} sequence

N_{17}	polyQ$_{23}$	polyP$_{11}$	PRD	polyP$_{10}$	
MATLEKLMKAFESLKSF	-Q23-	-P11-	QLPQPPPQAQPLLPQPQ	-P10-	GPAVAEEPLHRP
17	40	51	68	78	90

Htt-Q_{48} sequence

N_{17}	polyQ$_{48}$	polyP$_{11}$	PRD	polyP$_{10}$	
MATLEKLMKAFESLKSF	-Q48-	-P11-	QLPQPPPQAQPLLPQPQ	-P10-	GPAVAEEPLHRP
17	65	76	93	103	115

Figure 4.1.: Sequence of Htt-Q_{23} and Htt-Q_{48} studied in this work. The Nt_{17} region is highlighted in red, polyQ in blue, and the PRD is shown in yellow for polyP$_{11}$ and polyP$_{10}$ and rose for the other PRD parts.

4.1.2. Results

Htt-ex1 monomers

Our MD simulations, for 5 μs per protein, illustrate the significant structural differences between Htt-Q_{23} and Htt-Q_{48} monomers. Figure 4.2A and 4.2B elucidates the differences in the secondary structure profiles. Htt-Q_{48} has a higher propensity to form β-sheet, bend, and turn conformations, while Htt-Q_{23} has a higher tendency to

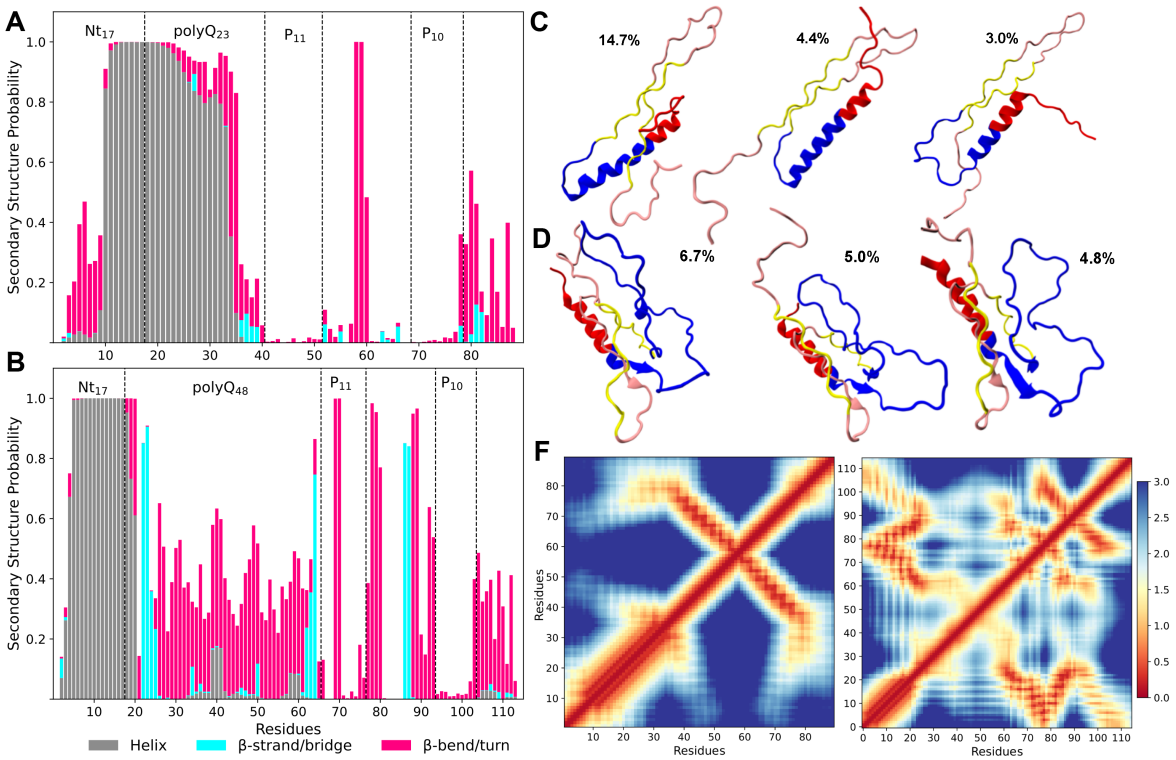

Figure 4.2.: Secondary structure preferences per residue of Htt-Q_{23} **(A)** and Htt-Q_{48} **(B)** monomers. The bars represent the additive secondary structure probabilities consisting of α-helix (gray), β-strand/bridge (cyan), and β-turn/bend (magenta). The difference to 1.0 presents the random coil state. Representative structures were determined from conformational clustering using an RMSD cutoff of 0.5 nm for Htt-Q_{23} **(C)** and Htt-Q_{48} **(D)**. The population of each cluster is given. The Nt_{17} region is shown in red, polyQ in blue, $polyP_{11}$ and $polyP_{10}$ in yellow, and the rest of the PRD in rose. **(F)** Intraprotein distance matrices of Htt-Q_{23} (Left) and Htt-Q_{48} (Right). The color bar on the right shows the average residue-residue distance (in nm).

form α-helix and random coil conformations. The higher β-sheet content of Htt-Q_{48} is consistent with previous experimental data and MD simulations [271–273]. Figure 4.2C and 4.2D show the three most populated clusters for the proteins. The Nt_{17} region of Htt-Q_{48} tends to form an α-helix, which is contrary to the Nt_{17} region Htt-Q_{23} that tends to form disordered structures in the first 10 residues. The following residues form an α-helix that extends up to 16 glutamines into the polyQ region. On the other hand, the polyQ region of Htt-Q_{48} mainly adopts a randomly collapsed coil or PPII helix conformations, while the terminal residues of the polyQ region show a considerable tendency to form β-sheet structures. The PRD in both proteins are

highly disordered but have a high tendency of forming PPII helices. While the PRD of the Htt-Q_{23} forms a PPII-turn-PPII motif, the PRD of Htt-Q_{48} contains more kinks and turns.

Further structural differences are defined by the distance maps (Figure 4.2F) to characterize the coupling between Htt-ex1 domains. The results confirm our observation that Htt-Q_{48} tends to adopt more compact structures, indicated by the significant contacts between Htt-Q_{48} residues. A notable difference between the two proteins is the strong contacts of the Nt_{17} region with the PRD region in Htt-Q_{48}, while this is not the case for Htt-Q_{23}. These interactions seem to stabilize the helical structures of the first ten residues of Nt_{17} by forming a hydrophobic core in Htt-Q_{48}. In general, the intraprotein interactions among Htt-ex1 are largely influenced by the increasing polyQ length and increased protein flexibility in the polyQ region, allowing a hydrophobic core to form in Htt-Q_{48}. Many experiments and simulations indicate that polyQ tracts prefer to adopt collapsed conformations because they are poorly soluble in water [273]. However, we find that our findings only partly agree with this, as the polyQ tract of Htt-Q_{48} shields the hydrophobic residues of Nt_{17} from the water, resulting in more collapsed conformations as compared to Htt-Q_{23} with shorter polyQ tract that does not provide equivalent shielding.

Htt-ex1 dimers

To gain insights into the first step of oligomerization, we conducted MD simulations for a total of 6 μs per protein to study the dimerization between Htt-ex1 monomers. The secondary structures and intra-peptide interactions are very similar to those of the monomers (Figure 4.2A and B). Therefore, our simulations did not show considerable structural transitions as a result of the aggregation. Figures 4.3A and B display the representative structures of the two most populated clusters. In general, the interaction interfaces between the Htt-Q_{23} proteins are primarily governed by the interactions of the partly amphipathic Nt_{17}/polyQ helices. These two helices are typically oriented parallel to each other, and their interactions contribute to helical stability. In contrast, the polyP regions have little involvement in the interaction interface and are oriented away from it (Figure 4.3C). The Htt-Q_{48} dimer clusters, on the other hand, exhibit a more diverse aggregation behavior, with various interaction

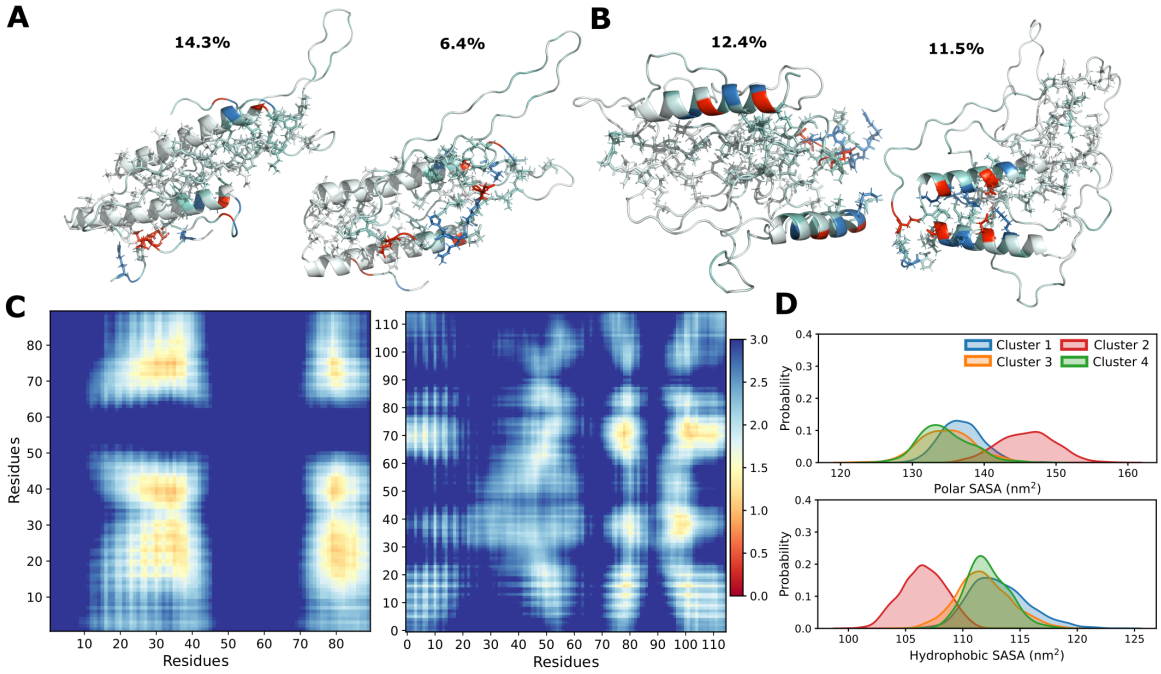

Figure 4.3.: The two most populated cluster structures for the Htt-Q_{23} **(A)** and Htt-Q_{48} **(B)** dimer. The proteins are shown as cartoons and colored red for the negatively charged residues, blue for the positively charged ones, and white otherwise. The side chains at the protein interfaces are shown as sticks. **(C)** The interprotein contacts between residues for Htt-Q_{23} (Left) and Htt-Q_{48} (Right) dimers. The color bar on the right shows the average distances (in nm). **(D)** The distribution of the polar (Top) and hydrophobic (Bottom) SASAs of the four most populated clusters Htt-Q_{48} dimers are shown.

interfaces involving both the polyP and the PRD. The primary interactions observed involve the polyP regions, as well as interactions between polyP and mixed regions in the PRD, and between polyP and polyQ. In addition, the polyQ interacts with the entire protein, while the Nt_{17} region interacts with $polyP_{11}$, polyQ, and the C-terminal of the PRD (Figure 4.3C).

Several studies proposed that the Htt-ex1 aggregation is initiated *via* the Nt_{17} headpiece, by forming oligomers with helical Nt_{17} in the core [89, 94]. In our simulations, we captured the interaction between Nt_{17} headpieces for Htt-Q_{48}. Dimerization *via* the Nt_{17} helices results in a decrease of the hydrophobic surface area and increases the polar surface area, compared to the other dimerization pattern (Figures 4.3D and B, cluster 2). This interaction pattern is expected to stabilize the helix-mediated

dimer, making it very likely aggregation mechanism.

4.1.3. Conclusion

Our MD simulations revealed significant structural differences between the Htt-ex1 monomers. Specifically, the non-pathogenic Htt-Q_{23} forms a long helix that extends from the second half of the Nt_{17} region to the majority of the polyQ region, while its PRD forms a PPII-turn-PPII motif. On the other hand, the Htt-Q_{48} monomer tends to be disordered, with the Nt_{17} region forming a helical structure while the polyQ region remains largely disordered with high PPII helix and random coil conformations. Consequently, the disordered polyQ leads to compact structures, with multiple intraprotein interactions and the formation of short β-sheet conformations. The helical structure of the Nt_{17} is stabilized by hydrophobic interactions between the Nt_{17} and the PRD. Both proteins exhibited mainly PPII structures in their polyP regions. However, in Htt-Q_{48}, these structures were disrupted by kinks, which is consistent with the compact conformations observed in this protein. In contrast, in Htt-Q_{23}, the $polyP_{11}$ and $polyP_{10}$ segments were straight and aligned antiparallel to each other, resulting in an elongated shape due to the 25-residue N-terminal helix and the PPII-turn-PPII motif. This overall fold was relatively stable, with structural fluctuations primarily arising from the first and last ten residues of Htt-Q_{23}.

The results for the dimers indicate that the N-terminal helix plays a crucial role in the dimerization of both Htt-Q_{23} and Htt-Q_{48}, with a more significant impact observed in the Htt-Q_{23}. In this protein, the helix-helix interaction stabilized the second part of the helix, which includes almost residues from the $polyQ_{23}$ region, explaining the resistance of this region against β-sheet formation. In contrast to Htt-Q_{23}, different modes of dimerization were observed in the Htt-Q_{48}, one of which involves the two N-terminal helices. However, we expect the dimerization mediated by the Nt_{17} helix to be more stable due to the larger burying of hydrophobic residues. A crucial aspect of the Htt-Q_{48} dimerization is the significant involvement of the disordered polyQ regions, which ultimately enable the formation of β-sheets over a longer timescale (not simulated in this study). Nevertheless, the observable interactions with the PRD obstruct this structural transition.

Overall, our results indicate that the extension of polyQ affects the conformational

ensemble of Htt-ex1 at the monomer level and has a significant impact on dimerization. Our findings shed light on how the flanking domains of polyQ modulate conformations and aggregation pathways.

Chapter

5.

Conclusions

In this thesis, we employed computational approaches to investigate the aggregation of amyloid proteins into small oligomers. Our focus was on two IDPs: the Aβ peptide, which is a peptide involved in pathology of AD, and the Htt protein, which is associated with HD. These diseases are characterized by the aggregation of amyloid proteins into amyloid fibrils, as well as pre-fibrillar aggregates such as protofibrils and amyloid oligomers. Experimental evidence suggests that the toxicity of the oligomers generated during the aggregation pathway is a crucial factor in the degeneration of neuronal cells in the brain. The oligomers are of low abundance, conformational heterogeneity, and high transit nature which makes them challenging to study through experiments. Therefore, computational approaches have become an intense area of investigation to understand the oligomer aggregation mechanism. To date, many computational studies on amyloid aggregation have used unsuitable force fields and water models, resulting in inaccurate ensembles that are not representative of IDPs. Accurately modeling the aggregation process requires careful consideration of the force fields and water models employed. By investigating the aggregation process of Aβ peptides and Htt proteins using appropriate force fields such as Charmm36m, we can enhance our understanding of amyloid proteins and their role in the related diseases.

In chapter 3, we focused on investigating the aggregation of the Aβ peptide into small oligomers through three topics: the aggregation into small oligomers of Aβ_{42}, the aggregation of various Aβ variants, and the effects of the neuronal membrane

on Aβ_{42} dimerization. In chapter 3.1, we have extensively simulated the aggregation of Aβ_{42} peptides into dimer, tetramer, and hexamer structures. We utilized long-timescale MD simulations of Aβ_{42} dimers to construct an MSM, which revealed conversion timescales between states ranging from a few nanoseconds to a hundred microseconds, with the slowest process being the transitions to the most compact structures. The formation of extended and compact dimer structures results from inter- and intrapeptide interactions. The compact dimers are stabilized by interactions within the central hydrophobic core, the C-terminal hydrophobic region, and intrapeptide interactions of the N-terminal region. The strong hydrophobic interaction among peptides in the C-terminus highlights its crucial role in stabilizing Aβ_{42} peptide oligomers. Additionally, the presence of two extra residues in the C-terminus of Aβ_{42} compared to Aβ_{40} enhances oligomerization. These findings suggest that interactions between the N-terminal and C-terminal regions can result in the formation of different oligomer conformations on or off pathways leading to amyloid fibrils. Our analysis demonstrates that the oligomer is rich in β-sheet conformations, mainly comprising two hydrophobic regions that form β-hairpin structures. However, we also observed transitions to β-sheet structures in other areas, such as residues 2-5, 10-14, and 38-41. Although the β-content of the oligomers is lower than that found in fibrils, it indicates that the oligomer undergoes structural changes to transform into the fibril structure. The oligomers can form compact structures by reducing their hydrophobic SASA. Although there is no dominant structure among the three oligomer sizes studied here, a slight preference to form compact structures in dimers and tetramers, and an extended structure in the hexamer, was observed. In summary, our findings suggest that intrapeptide interactions generally facilitate the formation of β-hairpin structures, but they can also hinder the unfolding of complexes, delaying the formation of fibril-like structures. Conversely, interpeptide interactions are crucial for stabilizing β-contacts across peptides and clustering hydrophobic residues, which promotes oligomerization. These findings suggest that multiple oligomerization pathways exist, and the oligomers must rearrange their conformation to form fibrils or prefer the oligomeric state due to strong hydrophobic interaction.

In chapter 3.2, we used computational and experimental approaches to investigation the aggregation of different Aβ variants. The finding demonstrated that a hairpin structure is crucial for the formation of oligomeric species of Aβ peptides. In particular, oligomers larger than tetramers require the folding of Aβ peptides into the

hairpin structure. This hairpin structure results from the high β-strand propensity of two hydrophobic regions in the Aβ peptide, which transiently fold onto each other. The formation of a hairpin structure in Aβ allows for the addition of monomers to the oligomer, resulting in the formation of larger β-sheet structures. As the size of the oligomer increases, the propensity for β-sheet formation also increases, suggesting that the folding process is cooperative. The elongated structures observed for oligomeric states larger than tetramers are believed to be the result of the formation of extended β-sheet structures. Nonetheless, these elongated oligomers have structures distinct from fibrils. Our MD simulations revealed that Aβ peptides adopt characteristic anti-parallel β-sheet structures in the oligomers. Therefore, nucleation requires a structural rearrangement that involves breaking intrapeptide hydrogen bonds and forming new interpeptide hydrogen bonds. Elongated oligomers are more likely to undergo conversion as their size increases, but it remains a rare event [274].

This study provides insights into the structural changes that occur during the initial stages of Aβ aggregation, which are thought to be particularly toxic and relevant to the development of AD. Our findings indicate that oligomers are inherently polydisperse and possess a complex energy landscape. This may contribute to their toxicity, as they have a strong tendency to interact with other cellular components and other biomolecules to minimize their energy levels [275–280]. Inhibiting the folding of Aβ into the hairpin structure could be a potential strategy for preventing Aβ toxicity. This could be accomplished through the development of therapeutic peptides or proteins that can effectively bind to regions of Aβ with a high β-sheet propensity, thus competing with Aβ-Aβ interactions. Our findings indicate that the interaction between two hydrophobic regions plays a critical role in oligomer formation, suggesting that targeting these regions could effectively reduce oligomerization. Alternatively, therapeutic molecules designed to specifically recognize and bind to Aβ in the hairpin structure could also be suggested as a potential approach.

In chapter 3.3, we simulated the dimerization of Aβ_{42} with a complex model lipid bilayer at the atomistic level. The bilayer consisted of six lipid components, which mimicked the composition of the neuronal membrane. We observed that Aβ_{42} dimerizes both in solution and in the presence of the neuronal membrane, but with notable structural differences compered to solution. We found that GM1 lipids of the neuronal membrane are the preferred lipid interaction sites of Aβ_{42}. Moreover, the interactions

of the dimer with monosialotetrahexosylganglioside (GM1) lipids in the membrane inhibit the formation of β-sheet and reduce the order level of the dimer in contrast to the Aβ_{42} dimer in solution. These results support earlier studies that suggest the neuroprotective role of GM1 lipid. Furthermore, we showed that the dimer in solution has a high proportion of β-sheet, indicating a transition from a random coil conformation to a β-sheet structure.

In chapter 4, we conducted multiple MD simulations to explore the structures of two Htt-ex1 proteins: the pathogenic Htt-Q_{48} and the nonpathogenic Htt-Q_{23} at both the monomer and dimer levels. Our results indicate distinct monomer structures for these two proteins. In the Htt-Q_{23}, the N-terminal helix includes both the Nt_{17} and most of the polyQ_{23} residues, whereas in the Htt-Q_{48}, the Nt_{17} tend to form helical structure and the polyQ tract is disordered, resulting in more bend/turn and PPII helix conformations across the polyQ and the PRD regions. This leads to compact protein structures with small intraprotein β-sheets. Although both Htt-Q_{23} and Htt-Q_{48} exhibit PPII structures in their PRD regions, they differ in terms of the PRD shape. The PRD of the Htt-Q_{48} is disrupted by kinks, consistent with its compact conformations, whereas the PRD in the Htt-Q_{23} has an elongated shape due to the presence of a PPII-turn-PPII motif and a 25-residue N-terminal helix. Our results on dimerization demonstrate that the N-terminal helix mediates the process for both Htt-Q_{23} and Htt-Q_{48} but to a greater extent in Htt-Q_{23}. The helix interaction with itself and the polyQ_{23} region stabilizes the helix and inhibits β-sheet formation. Moreover, the amphipathic α-helix Nt_{17} headpieces interact and reduce the hydrophobic surface, increasing the stability of this dimer conformation. In contrast, we observed different dimerization modes in the Htt-Q_{48}, with the one involving the two N-terminal helices. However, we predict that the Nt_{17} helix-mediated dimer is more stable because of greater burial of hydrophobic residues. Dimerization of the Htt-Q_{48} involves the disordered polyQ regions, allowing for β-sheet formation on a longer timescale, but interactions with the PRD slow down the formation. Our findings reveal that the length of polyQ has a significant impact on the structures of the Htt-ex1 monomers level, which subsequently affects its dimerization. Our results provide insight into how the adjacent domains of polyQ can influence the conformations and aggregation pathways of Htt-ex1, and offer an explanation for several experimental findings at the atomic scale.

It is important to note that there are various challenges and limitations associated

with all-atom MD simulations used to model protein aggregation. These simulations require a significant amount of computing resources, which restricts the time scales that can be simulated. Furthermore, simulating the complex process of aggregation involves addressing various issues. Firstly, a large number of intramolecular interactions with intricate behavior need to be considered. Secondly, the intermolecular behavior of aggregation-prone proteins presents similar difficulties, as the sheer number of possible interactions requires long sampling times to properly account for them. This will result in slowing down the conformational transitions and make it difficult to obtain the required data for constructing thermodynamic or kinetic ensembles. Even with applying enhanced sampling techniques, capturing sufficient transitions among different conformations to construct kinetic models can be difficult and require long timescale simulations. To overcome these challenges and fulfill the requirements of building a kinetics model, such as detailed balance, transition networks provide a valuable alternative to kinetic models as they are solely based on geometric clustering by extracting the essential features of the conformational transitions. When working with Markov models, it is important to consider statistical limitations that affect both temporal and spatial resolution. These limitations arise from the lag time and conformational space discretization, which can impact the accuracy of the resulting model. Nevertheless, the limited conformational space can also have some benefits, as it may aid in the interpretability of the results. In this work, we have employed conventional MD and REMD simulations to model the protein aggregation process. These simulations were able to sample a wide range of conformations, covering significant parts of the conformational space, which was essential in successfully modeling the aggregation process. Overall, through the use of computational approaches, we have gained a deeper understanding of the molecular mechanisms behind amyloid aggregation, supported by available experimental data. Our findings provide high resolution details of the systems under investigation that cannot be obtained through experimental means.

Chapter A.

Appendix

A.1. Appendix to Amyloid-β_{42} Oligomers

State	S1	S2	S3	S4
Stationary population	10.6%	11.1%	13.0%	23.9%
State	S5	S6	S7	S8
Stationary population	8.7%	4.2%	24.8%	3.9%

Table A.1.: The stationary distributions of the MSM states for the Aβ_{42} dimer.

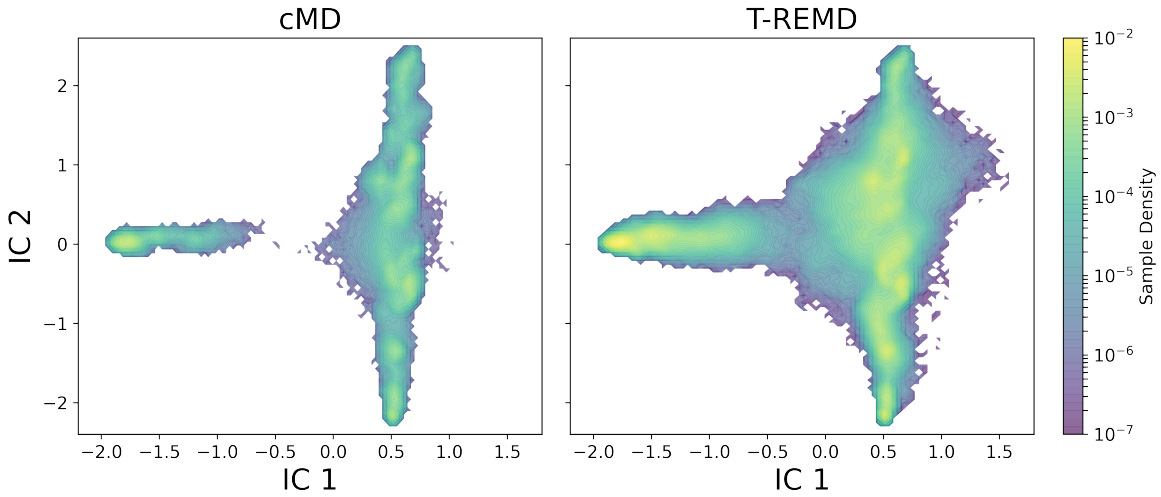

Figure A.1.: The sample density along the first two TICA components for unbiased MD and REMD dimer simulations.

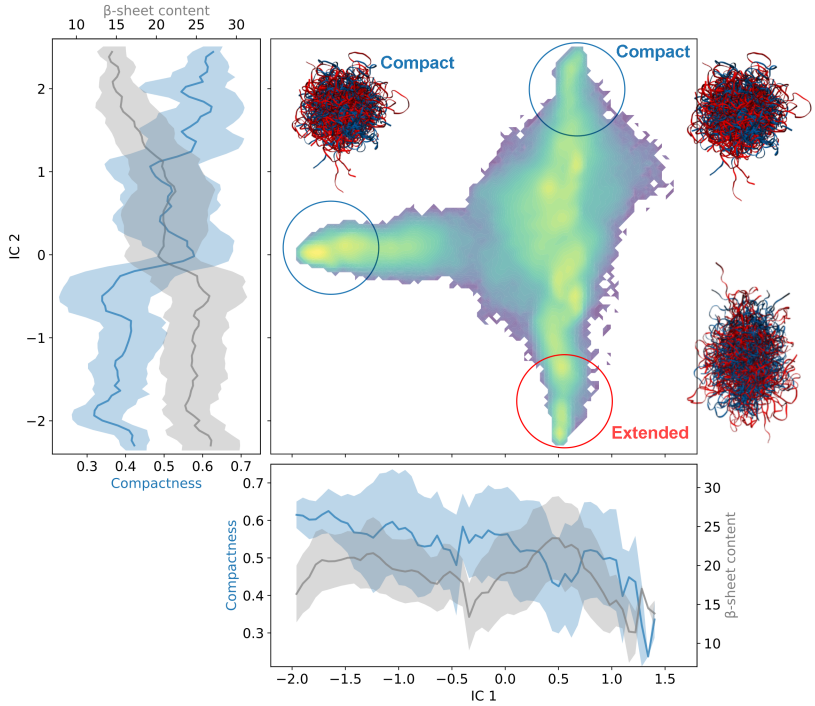

Figure A.2.: The sample density of the projections of the combined cMD and REMD dimer trajectories into the first two slowest collective coordinates. The average compactness and the average number of residues in β-sheet conformation with their standard deviations along the IC_1 and IC_2 are shown in blue and gray, respectively. The structures illustrating 100 representative conformations for extended and compact conformations.

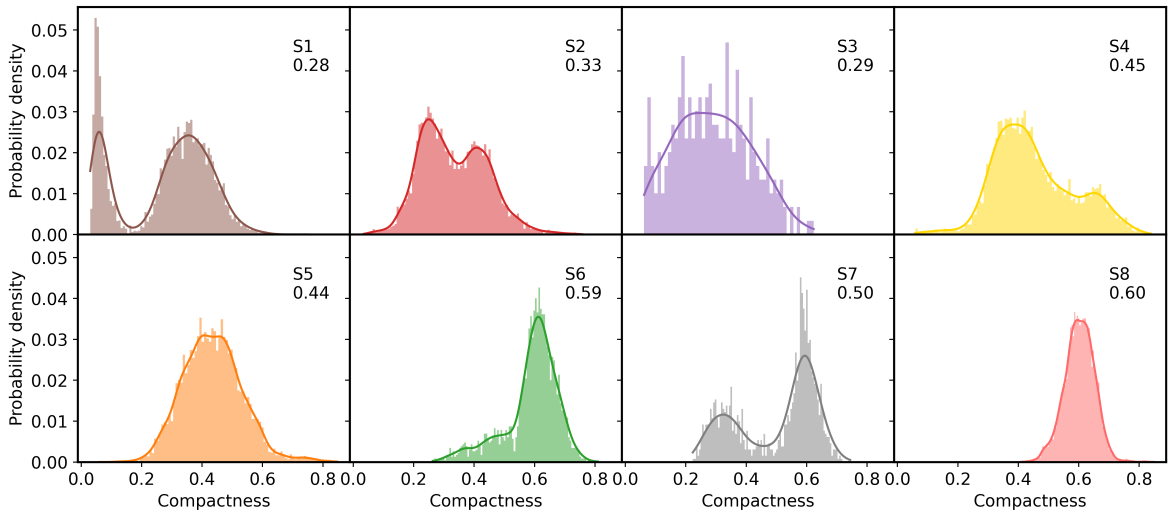

Figure A.3.: The compactness distributions of the MSM states for the Aβ_{42} dimer. The state number and compactness average value are printed in the top corner for each subplot.

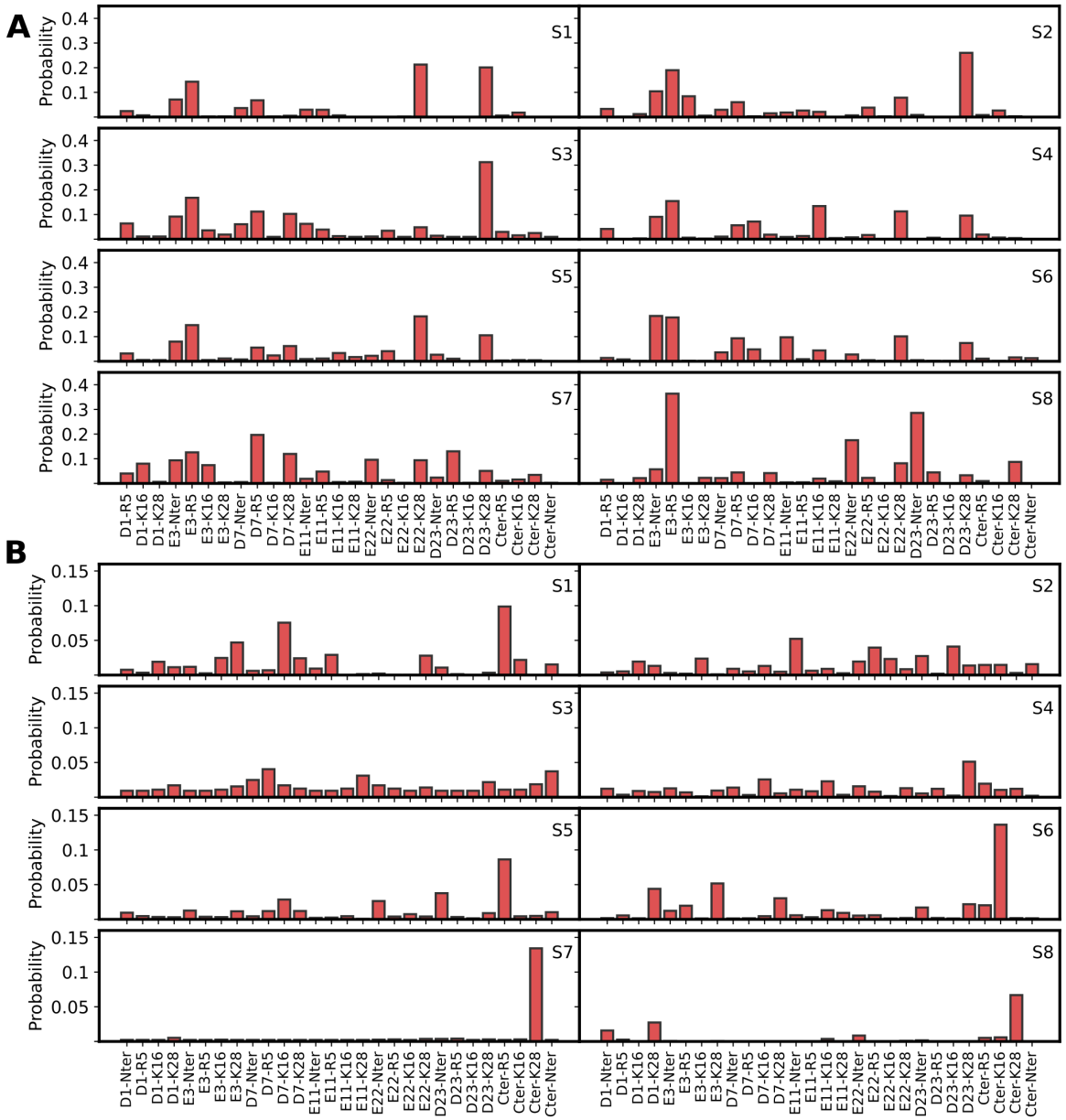

Figure A.4.: The probabilities of intrapeptide **(A)** and interpeptide **(B)** salt-bridges formation within of the MSM states for the Aβ_{42} dimer. The salt bridge is defined between the charged side chain atoms when the distance between two specified atoms is within 4.6 Å. The possible atoms are: N atom in the N-terminal, CG atom in Asp or CD atom in Glu; NH2 atom in Arg, NZ atom in Lys, or C atoms in the C-terminal.

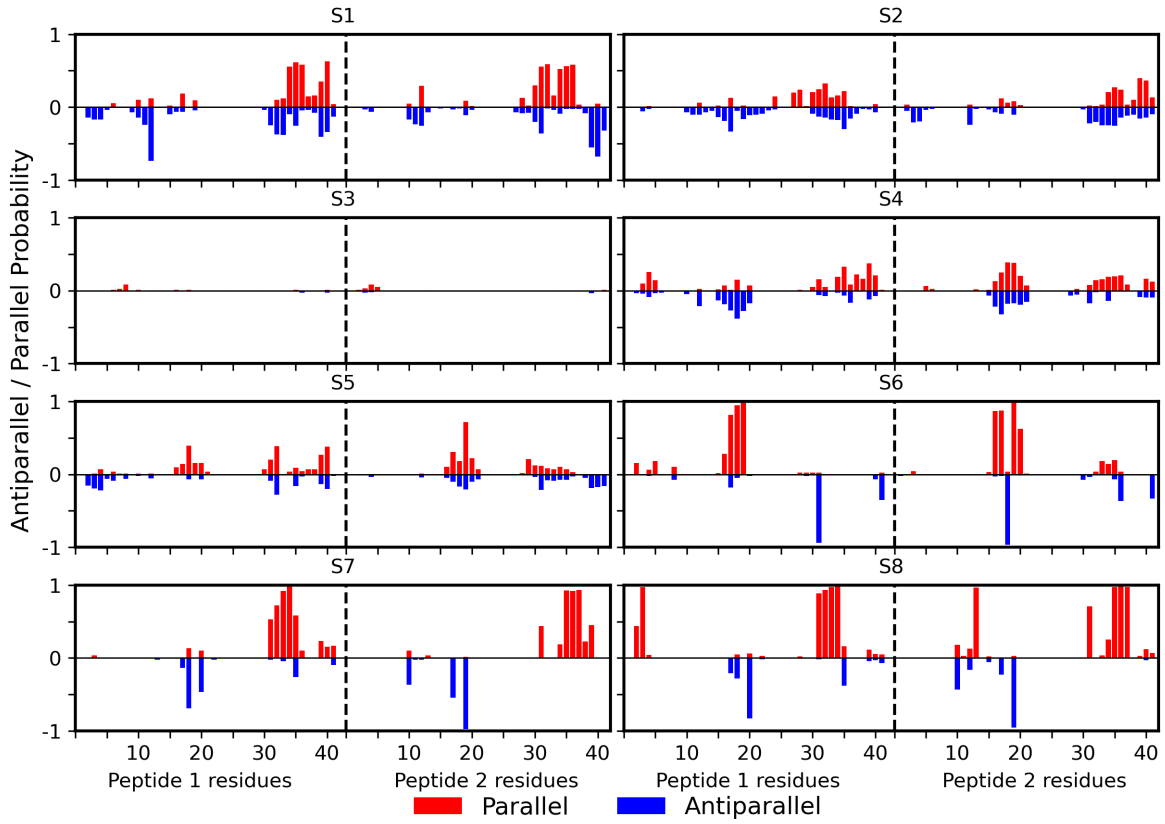

Figure A.5.: The probability of Aβ_{42} residues of the dimer to be in parallel (red) or antiparallel (blue) interpeptide β-sheet conformation. The antiparallel β-sheet probabilities are showed in negative y-axis.

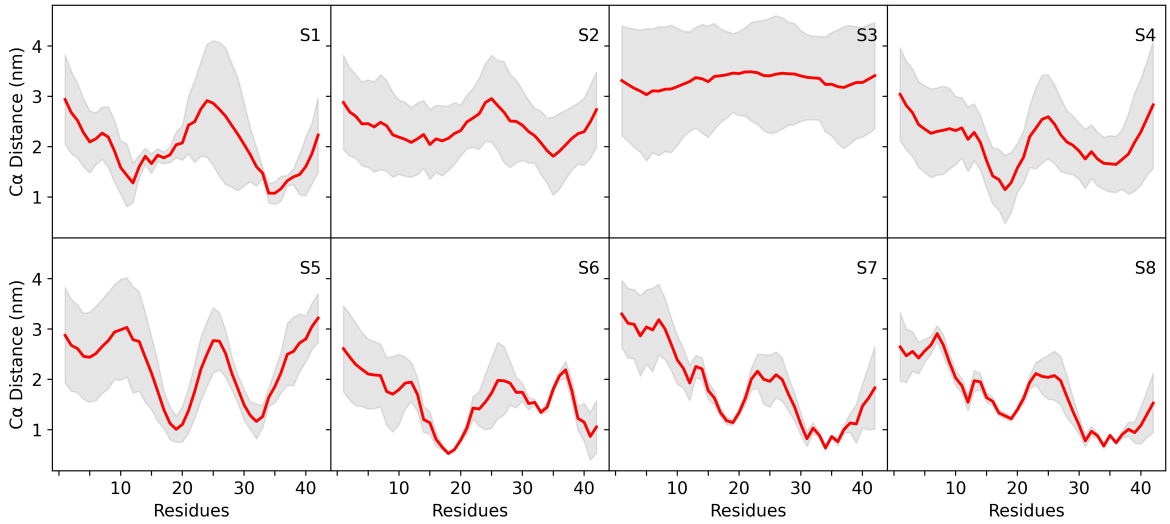

Figure A.6.: The distances between the Cα atoms of the residue pairs (i,i) in the Aβ_{42} dimer for the MSM states, where the residue index i ranges from 1 to 42. The gray shades represented the standard deviations.

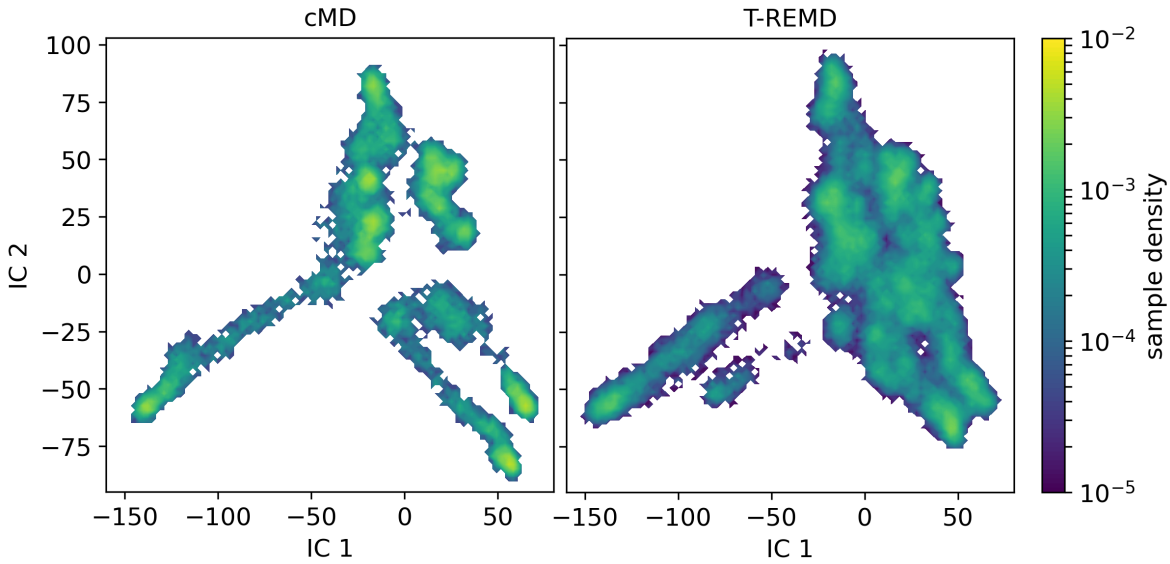

Figure A.7.: The sample density along the first two TICA components for unbiased MD and REMD tetramer simulations.

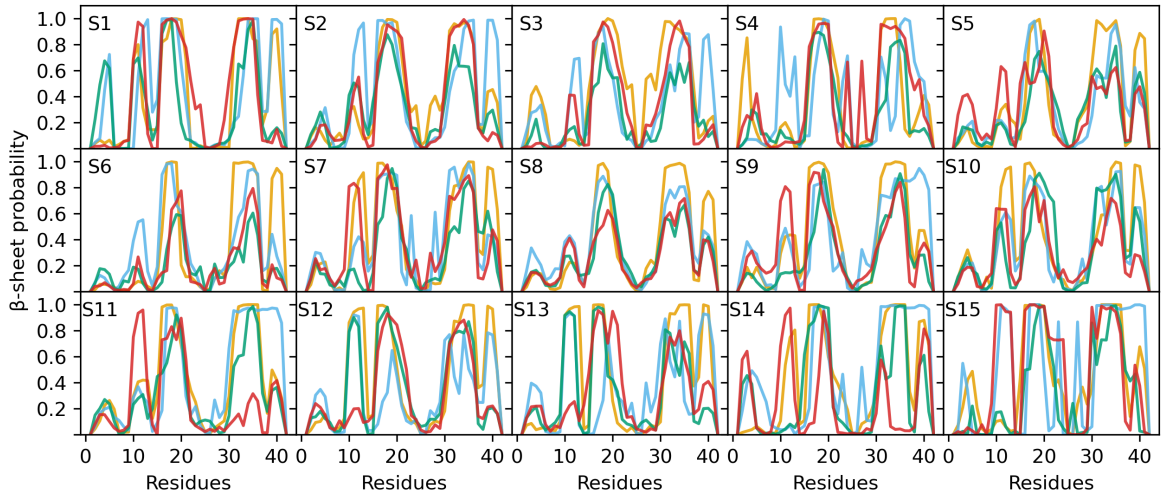

Figure A.8.: The β-sheet probability of Aβ_{42} residues of the four peptides in the different Aβ_{42} tetramers states. The peptides 1, 2, 3 and 4 are colored yellow, blue, green, and red, respectively.

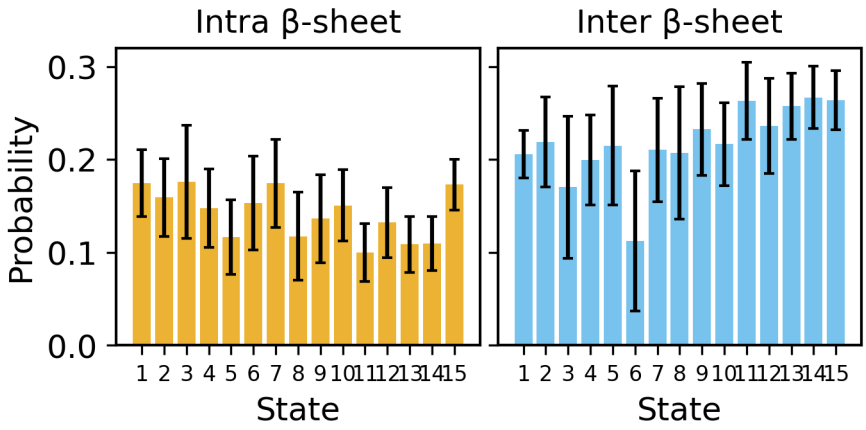

Figure A.9.: The average content probabilities of intra- and interpeptide β-sheet for the tetramer states.

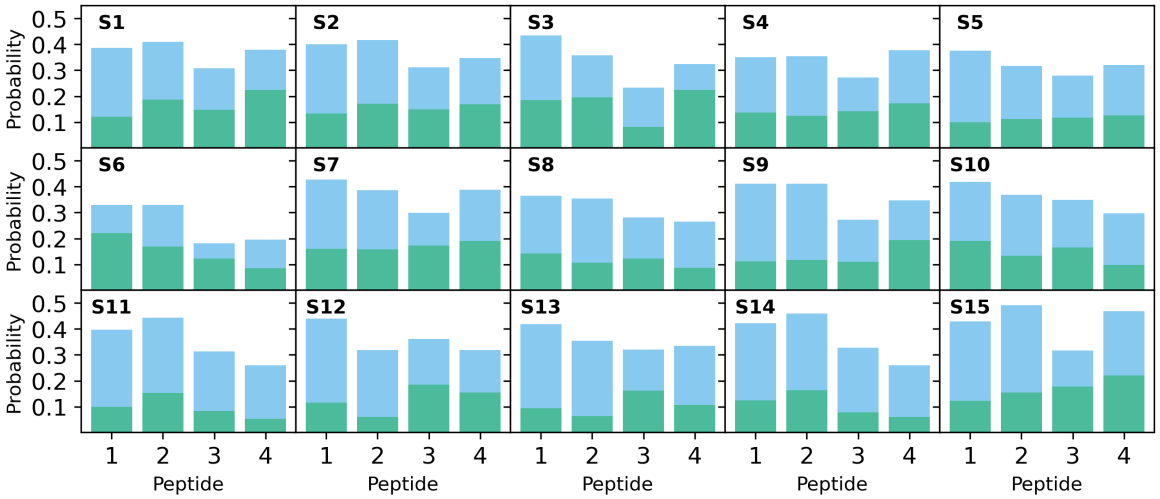

Figure A.10.: The β-sheet probability of Aβ_{42} tetramer per peptide for the tetramer states. The bars are colored in green and blue for intra- and interpeptide β-sheet probabilities, respectively.

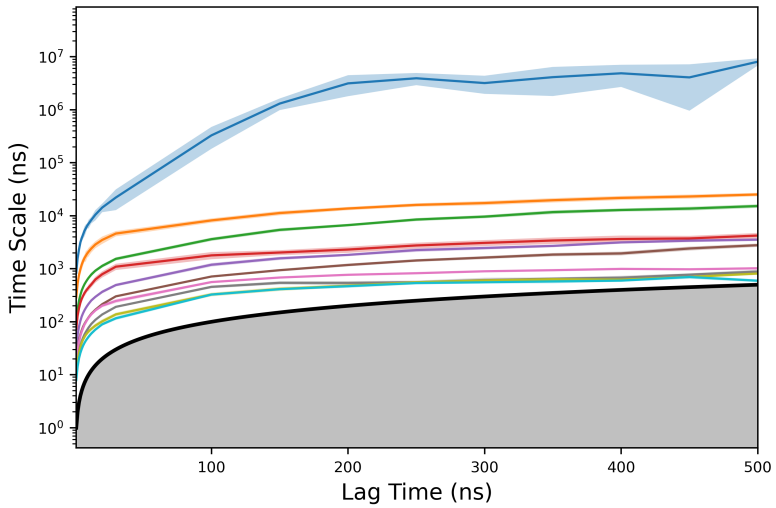

Figure A.11.: Implied time scales of the slowest MSMs process at different lag times obtained from the cMD simulations of Aβ_{42} dimer.

A.2. Manuscript I

A hairpin motif in the Amyloid-β peptide is important for formation of disease-related oligomers

Khaled, M., Rönnbäck, I., L. Ilag, L., Gräslund, A., Strodel, B., Österlund, N. (2023)

A hairpin motif in the Amyloid-β peptide is important for formation of disease-related oligomers

Mohammed Khaled[1], Isabel Rönnbäck[2], Leopold L. Ilag[3], Astrid Gräslund[2], Birgit Strodel[1,4], Nicklas Österlund[2]

1. Institute of Biological Information Processing: Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52428 Jülich, Germany.
2. Department of Biochemistry and Biophysics, Stockholm University, SE-114 18 Stockholm, Sweden
3. Department of Materials and Environmental Chemistry, Stockholm University, SE-114 18 Stockholm, Sweden
4. Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

ABSTRACT

The Amyloid-β (Aβ) peptide is an aggregation-prone peptide linked to neurodegeneration in Alzheimer's disease (AD). Aβ self-assembles spontaneously in aqueous solution to form aggregates of various sizes, with smaller pre-fibrillar oligomeric aggregates being especially neurotoxic. Such small oligomers are however difficult to study as they are transient, low abundant and heterogenous. Here we use a combination of native ion mobility-mass spectrometry and molecular dynamics simulations to systematically study the structure and assembly mechanisms of Aβ oligomers in vitro. It is found that oligomers cannot be formed by a peptide variant that does not have propensity to fold into a β-hairpin motif present in the wild type Aβ peptide. This specific structure motif seems to be a more important determinant for aggregation than the overall hydrophobicity of the peptide. Introduction of an intramolecular disulfide bond in the Aβ peptide increases oligomerization, even though the monomeric peptide is not stabilized in the hairpin conformation. This is probably achieved by pre-arranging the peptide in a conformation which is compatible with oligomeric, but not fibrillar structures. As oligomerization is driven by formation of the hairpin motif it was furthermore possible to decrease the oligomer population by truncating one of the β-strands, and thus decreasing the hairpin propensity of the peptide. These studies provide increased understanding of the earliest steps in Aβ aggregation where species related to AD toxicity might be formed. Prevention of Aβ folding into the hairpin conformation, or specific binding to the hairpin motif could be strategies to design AD therapies.

INTRODUCTION

The amyloid-β (Aβ) peptide is a neurotoxic peptide formed by enzymatic processing of the Amyloid Precursor Protein (APP), a type-I membrane protein found in the neuronal membrane. N-terminal cleavage by β-secretases in the soluble extracellular part of APP, in combination with C-terminal cleavage by γ-secretase in the transmembrane domain, releases Aβ peptides consisting of 38-43 amino acids (1). These amphiphilic peptides are highly aggregation prone and spontaneously assemble in aqueous solution at concentrations higher than their equilibrium solubility (2, 3). Aggregation starts with an unstructured Aβ monomer which evolves into β-sheet structured amyloid fibrils, typically deposited in neuritic plaques associated with Alzheimer's disease (AD). The exact mechanism for the peptide's toxicity, and its link to AD pathology is not known in detail. The so-called Amyloid cascade hypothesis suggests that it is the accumulation and aggregation of Aβ in the brain that is the main cause of AD (4, 5). Aβ populates several intermediate aggregation states on the way towards the fibrillar end-product, and a solution of very high polydispersity is consequently formed (6). The fact that Aβ peptides convert from an unstructured state into a β-sheet state means that the aggregation occurs by non-classical nucleation, in which both aggregate size and β-sheet content must increase on the path towards nucleation (7, 8). This structure conversion is associated with a high energy barrier, and primary nucleation is thus relatively slow (9). Structural templating using existing fibrils, so-called secondary nucleation, greatly reduces the activation barrier and is therefore the dominating nucleation mechanism for Aβ aggregation (10).

The pre-fibrillar Aβ aggregates have risen as especially interesting species in the aggregation process, as they have been observed to correlate better with neurodegeneration compared to the mature fibrillar aggregates (11, 12). Pre-fibrillar aggregates range from dimers all the way up to large megadalton particles (6, 13–15) with the smaller aggregates sometimes called oligomers ("oligo", a few), while larger and more fibrillar-like aggregates are often termed protofibrils (16). However, no universally accepted terminology for different types of aggregates exists within the amyloid field. Pre-fibrillar aggregates have been the target for novel strategies to treat AD (17, 18). Most recently monoclonal antibodies that specifically target such aggregates have been developed, which are promising drug candidates (19–21). Much is however still elusive about pre-fibrillar aggregates on the molecular level. The main reason for this is that they remain difficult to study due to their low abundance, low stability, and polydispersity. A recent study found that Aβ(1-42) oligomers are very dynamic and rapidly dissociate into monomers (7). It was also quantified that the pool of soluble Aβ(1-42) oligomers never reaches more than 1.5% of the total peptide concentration in a buffered aqueous solution (7).

The Aβ peptide contains two hydrophobic segments, in the middle (the central hydrophobic core, CHC, residues 16-22) and in the most C-terminal parts of the sequence (residues 30-42), separated by a relatively hydrophilic "hinge" segment (Figure S1). Proline scanning of the Aβ sequence has revealed that the E22P substitution in the hinge region increases the toxicity of the peptide, possibly due to stabilization of a turn motif that enables the two hydrophobic segments to interact more favorably (22). Phosphorylation of the S28 residue in the hinge segment has on the other hand been found to destabilize this turn-motif due to repulsive electrostatic effects,

which also leads to a loss in aggregation propensity (23). Transient back-folding of the C-terminal segment and formation of a transient β-hairpin structure **(Figure 1A)** is consequently believed to be important in Aβ aggregation and its associated disease-related toxicity (24). This is further supported by the fact that the hinge-region is a segment where many disease-related mutations are located. Such disease related variants include the Flemish (A21G), Dutch (E22Q), Italian (E22K), and Arctic (E22G) variants (25–28). The Arctic variant is particularly interesting as this variant is enriched in toxic pre-fibrillar aggregates, which have been used as targets to develop the promising anti-AD antibody drug Lecanemab (29). Stabilization of a partially folded hairpin state in Aβ has also been reported using an intermolecular disulfide-bridge in the hinge region (A21C A30C double-mutant, $A\beta_{CC}$). This restriction of conformational dynamics under oxidizing conditions similarly to the Arctic variant leads to formation of stable and highly toxic pre-fibrillar oligomers, but no amyloid fibrils (30). These $A\beta_{CC}$ structures have recently also been used to develop anti-AD antibodies to target especially neurotoxic Aβ species (31).

In this study the role of the seemingly very important β-hairpin motif in the earliest steps in Aβ peptide self-assembly is systematically examined biophysically. We use a combination of experimental in vitro techniques and molecular dynamics (MD) simulations to study the smallest Aβ aggregates, oligomers consisting of only a few monomeric units. Experimental studies of oligomers are generally challenging as only a small fraction of the total peptide ensemble is in the oligomeric state at each given moment (7). One well-suited technique for studying Aβ oligomerization is native mass spectrometry (MS), which provides highly resolved size information on the different co-existing assembly states in a heterogenous mixture, even for sparsely populated states (15, 32). Coupling of MS to ion mobility (IM) spectrometry furthermore provides low resolution information on oligomer shape. The combination of IM-MS and MD simulations is very complementary: MD provides high resolution information on oligomer structure and dynamics, which is not obtainable by IM-MS. MS on the other hand reports on the relative populations of assembly states, which is challenging to obtain by MD as the entire energy landscape can rarely be sufficiently sampled (9, 33). This combination of MD and native IM-MS to systematically examine different Aβ sequence variants here gives us valuable insights into the molecular mechanisms for Aβ aggregation and the nature of neurotoxic pre-fibrillar oligomers which are highly interesting species in AD biology and drug development.

RESULTS AND DISCUSSION

Aβ oligomerization is driven by specific structure motifs

Mass spectrometry analysis of Aβ(1-40) results in detection of various oligomeric states at low relative abundance (Figure 1A), as has been previously reported (14, 15, 34). We here annotate the oligomers by their oligomeric state/charge state (n/z) ratio, as electrospray ionization generates species with multiple charges. It should however be noted that oligomers can overlap in the m/z dimension of the mass spectrum, and for example the n/z = 1/2 peak could hence consist of monomeric (+2), dimeric (+4), trimeric (6+)… components. These components can however usually be deconvoluted using the ion mobility (IM) dimension where ions are separated by a shape factor termed the collision cross section (CCS). There have previously been various reports

regarding the oligomeric states populated by Aβ, with an early paper claiming that Aβ(1-40) only oligomerizes up to a tetrameric state (34). We here observe no such limitations with Aβ(1-40) oligomers detected at least up to an octameric state. There is also no specificity in which states are populated, with no preferred oligomeric state and a gradual decrease in intensity with increasing oligomeric state. This is in contrast to other more specific proteins detected by native MS (35), as well as to Aβ(1-42) oligomers formed in the presence of zwitterionic detergent, where a preferred assembly state is formed (36, 37). This indicates that Aβ oligomers in solution exist on a frustrated energy landscape where no single aggregate state is significantly more favorable than the other.

Observation of oligomers in ESI-MS could theoretically be due to artificial clustering during the ionization process, if two species are present in the same electrospray droplet upon desolvation. Low peptide concentrations and small droplet sizes (nanoESI) are here used to counteract this. Monte-Carlo simulations have previously suggested that such non-specific artifacts become relevant in nanoESI at protein concentrations greater than 50 μM (38), which is well above the concentration used here (20 μM).

To confirm that the detected oligomers are indeed specific aggregates we analyzed a scrambled Aβ variant, $Aβ(1\text{-}40)_{Scr}$ under identical conditions as the wild type Aβ(1-40). The $Aβ(1\text{-}40)_{Scr}$ peptide has the same amino acid composition as Aβ(1-40) but hydrophobic residues are more evenly distributed of over the sequence, in contrast to Aβ(1-40) where the hydrophobic residues are clustered in two segments **(Figure S1).** A scrambled peptide is a perfect control in a native MS experiment as it has all the same average physico-chemical properties as the original peptide, including the same mass. The mass spectrum for $Aβ(1\text{-}40)_{Scr}$ shows a very similar charge state envelope as Aβ(1-40), centered around the +3 state **(Figure 1B)**. Two differences are however noted: $Aβ(1\text{-}40)_{Scr}$ significantly populates also the +5 state, indicating a slightly larger solvent assessable surface area (SASA). $Aβ(1\text{-}40)_{Scr}$ does also importantly not form any oligomers other than a dimer. The slightly larger size and decrease in oligomers for the scrambled peptide are seen also in the size exclusion chromatograms of the two peptides (Figure S2).

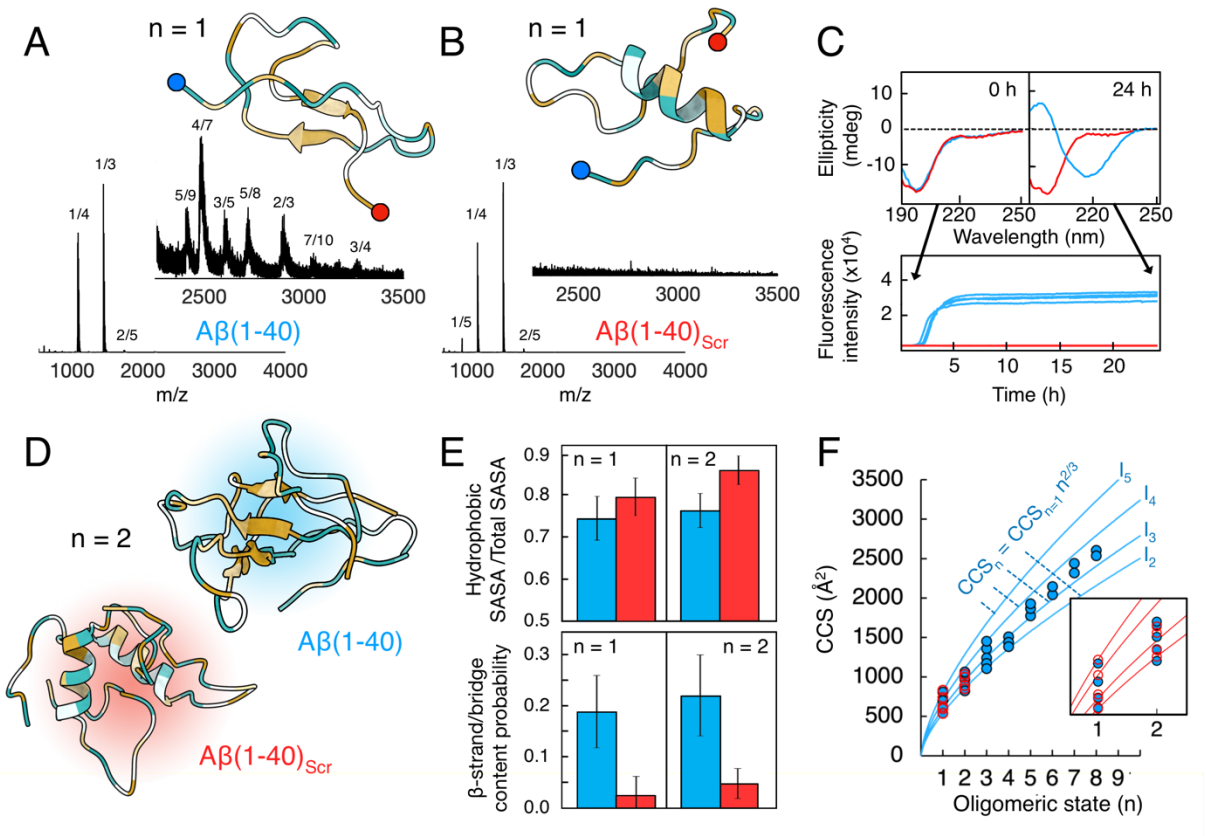

Figure 1. Self-assembly of Aβ(1-40) and Aβ(1-40)Scr. Mass spectra of 20 μM **(A)** Aβ(1-40) and **(B)** Aβ(1-40)Scr, in 200 mM ammonium acetate pH 6.8, with the oligomeric region magnified. Peaks are annotated by their oligomeric state/charge state ratio (n/z). The top cluster monomeric structures from 2 μs MD simulations are shown for each peptide, colored according to hydrophobicity (orange = hydrophobic, blue = hydrophilic). The N-terminus is highlighted in blue, while the C-terminus is highlighted in red. **C)** Time dependent aggregation assays on Aβ(1-40) (blue) and Aβ(1-40) Scr (red) aggregation, as monitored by CD spectroscopy (top) and ThT fluorescence (bottom). **D)** Top cluster structures from 10 μs MD simulations of Aβ(1-40) (blue) and Aβ(1-40)Scr (red) dimers, colored according to hydrophobicity. **E)** Calculated hydrophobic Solvent Accessible Surface Area (SASA)/total SASA ratio (top) and β-strand/bridge content probability (bottom) from MD simulations, for Aβ(1-40) (blue) and Aβ(1-40)Scr (red) monomers and dimers. **F)** Measured collision cross sections (CCS) for each oligomeric state (blue circles for Aβ(1-40), red circles for Aβ(1-40)Scr). Solid lines represent the theoretical isotropic growth originating from each monomer state. The insert shows a magnification of the n = 1 and n = 2 oligomeric states.

The reason for this observed difference in oligomerization propensity must be that the specific sequence in Aβ(1-40) is responsible for the clustering into oligomers, rather than a more general phenomenon such as clustering of hydrophobic peptides in aqueous solution or desolvation of multiple species from the same electrospray droplet. An explanation for the difference in oligomerization propensity could be found by examining the structure propensity of the two peptides. Structure prediction using AlphaFold2 predicts that Aβ(1-40) forms a β-hairpin structure where the two hydrophobic segments self-interact (Figure S3A). This is

similar to the NMR structure of Aβ(1-40) formed in the presence of an affibody (39). The more even distribution of hydrophobicity in $Aβ(1-40)_{Scr}$ is however predicted by AlphaFold2 to facilitate folding into an amphipathic helix (Figure S3B). To test if such structures could be stable in solution, we performed 2 μs MD simulations for each peptide monomer. While the helical motif in $Aβ(1-40)_{Scr}$ remained during the MD simulation, most of the β-hairpin structure was not present in the top cluster MD model of the Aβ(1-40) monomer **(Figure 1A, top)**. Circular dichroism (CD) spectroscopy confirms that fresh samples of both Aβ(1-40) and $Aβ(1-40)_{Scr}$ are mostly unstructured, as seen by a spectral minima around 196 nm, which is characteristic for a random coil **(Figure 1B, top)**. The Aβ(1-40) peptide, as expected, spontaneously evolves into large ThT-active aggregates upon incubation **(Figure 1C, bottom blue),** which have β-sheet structure as seen by CD spectroscopy **(Figure 1C, top blue)**. $Aβ(1-40)_{Scr}$ does however not form any such β-sheet amyloid aggregates **(Figure 1C, red)** upon incubation.

Further 10 μs MD simulations were performed to study the differences between dimers formed by the two peptides. $Aβ(1-40)_{Scr}$ was observed to interact within the dimer with the hydrophobic faces of the amphiphilic helices facing inwards **(Figure 1D,** Figure S4). This structure where hydrophobic patches have specific preferred interactions could explain why $Aβ(1-40)_{Scr}$ is less frustrated compared to Aβ(1-40) and does not go on to form higher oligomers. Aβ(1-40) instead forms a dimeric structure by interpeptide interactions involving the β-strands in the most hydrophobic parts of the sequence **(Figure 1D)**. Aβ(1-40) is found in the MD simulations to expose slightly fewer hydrophobic residues to the surrounding solvent compared to $Aβ(1-40)_{Scr}$ **(Figure 1E, top)**, highlighting that average hydrophobicity might not be the only driving force for oligomerization. Instead, the large difference in β-sheet propensity **(Figure 1E, bottom)** and distribution of hydrophobic residues over the sequence could be important. It is easy to imagine how a β-sheet structure would enable continued growth of larger aggregates by addition of monomeric units, without satisfying any particular aggregate number. This is demonstrated by modeling Aβ(1-40) tetramerization in MD simulations (also for 10 μs), where an even larger β-sheet core is formed (Figure S5A). Previous MD simulations of Aβ oligomers combined with transition network analysis have shown that several open and closed oligomeric structures of various aggregate sizes can be populated during further aggregation, in agreement with the mass spectrum in **Figure 1A** (40). The β-strand/bridge content of Aβ(1-40) increases in our MD simulations with increasing assembly state, from 19% in the monomer, to 22% in the dimer and 30% in the tetramer (Figure S5B). Proline substitutions that reduce aggregation propensity and neurotoxicity were also found to be in or close to the segments modeled as β-strands in our MD models of Aβ(1-40) (Figure S6A). Proline is a hydrophobic residue that acts as a β-sheet breaker. Loss of the β-hairpin motif in Aβ(1-40) is indeed suggested by AlphaFold2 when modeling the F19P variant (Figure S6B), which has previously been found to have decreased oligomerization propensity (41).

Small Aβ oligomers grow isotropically

We next turned to ion mobility spectrometry to study the shapes of the Aβ(1-40) and $Aβ(1-40)_{Scr}$ oligomers. All oligomeric states of the peptides display several conformations as seen by different collision cross sections (CCS) **(Figure 1F)**, which could represent either different solution state structures or gas phase structures. The

experimental CCS values for Aβ(1-40) and $Aβ(1-40)_{Scr}$ are similar for low z but are slightly increased for $Aβ(1-40)_{Scr}$ at higher z (**Figure 1F, insert)**. Calculated CCS values for the top clusters from MD simulations are also similar, with the Aβ(1-40) monomer having a CCS of 740 Å^2 and the $Aβ(1-40)_{Scr}$ monomer having a CCS of 780 Å^2. This is fairly close to the experimental values for monomers that range between 530 Å^2 (z = 2) and 830 Å^2 (z = 5).

A more careful analysis shows that the experimental CCS start to increase significantly above z = 3 for monomers, and above z = 4 for dimers (Figure S7). High charge density in the gas phase will lead to strong Coulomb repulsion due to the low permittivity of vacuum. We therefore conclude that the compact ion mobility conformations observed at z = +2 and +3 for the monomer and +3 and +4 for the dimer are solution state-like structures while the higher charge states which have significantly increased CCSs are structures that may have been significantly altered in the gas phase by repulsive intrachain coulomb-coulomb interactions. It can also be seen that most oligomers observed have CCS values that agree with isotropic growth starting from the compact z = +2 or +3 monomers, which we annotate as I_2 and I_3 respectively **(Figure 1F, blue solid lines).** Isotropic growth means that the oligomers grow as spheres according to: $CCS_n = CCS_{n=1} * n^{2/3}$ (42). Higher oligomers above n = 6 do however deviate from this isotropic growth, indicating a change in aggregate shape. The top cluster structures for n = 1, 2 and 4 generated from MD simulations also appear to grow isotopically (Figure S8). The absolute CCS values of the MD models are however larger than the experimental values, which is probably due to compaction of the structures upon transfer from solution into the gas phase (43).

An intramolecular crosslink in the hinge region increases oligomerization

As the sequence specific β-hairpin structure motif seems to be important for the oligomerization of Aβ(1-40) we next attempted to stabilize this structure to see if this leads to an enhancement of the peptide's oligomerization propensity. Stabilization of the β-hairpin state of the monomer has previously been explored by the introduction of an intramolecular disulfide bridge in the hinge-segment (A21C, A30C) (30). This double-mutant peptide called $Aβ_{CC}$ does not have a higher tendency to form amyloid fibrils, aggregation is instead halted at pre-fibrillar states. The $Aβ(1-40)_{CC}$ is co-expressed together with an affibody, which stabilizes the monomeric peptide in the β-hairpin conformation for which the structure has been solved using solution state NMR (30, 44). An MD simulation starting from an $Aβ(1-40)_{CC}$ monomer in this hairpin state performed here for 2 μs suggests that the hairpin motif is not stable without the affibody, even though the two hydrophobic segments remain close to each other **(Figure 2A)**. CD spectroscopy confirms this loss of β-sheet structure, as the $Aβ(1-40)_{CC}$ variant displays a spectrum typical for a random coil, very similar to wildtype Aβ(1-40) **(Figure 2A)**. Mass spectrometry analysis of $Aβ(1-40)_{CC}$ reveals a general increase in oligomeric species compared to the wildtype variant, with a most notable increase for larger oligomers (n>6) **(Figure 2B, orange fields)**. Notable shifts in the ion mobility dimension are also seen, with shifts towards larger oligomers in peaks that overlap in the m/z dimension (Figure S9). This increase in oligomers again points towards an important role of the hairpin structure in formation of oligomers.

Taking the relative intensity of charge states within an oligomeric state into account reveals yet more interesting features. The most populated charge states for each oligomeric state ≤ 4 follow isotropic growth that originates from the z = +3 monomeric state (I_3). The I_3 family extends from n=1 to n=6 but is not present for larger oligomers **(Figure 2C, dark green line)**. A less populated and more compact isotropic family I_2 also exists that originates from the z = +2 monomeric state, that is populated between n=1 and n=4 (**Figure 2C, light green line)**. Above n = 4 the most populated charge states have CCSs that instead belong to a linearly growing family, L **(Figure 2C, orange)** that extends for as long as oligomers are detectable (at least n = 10). The ratio between the experimental population weighted average CCS for an oligomeric state and its theoretical isotropic I_3 CCS for each oligomeric state clearly illustrates that the Aβ oligomers grow almost perfectly isotropic until n = 4, above which they start to deviate from isotropic growth **(Figure 2D)**. Such a deviations indicate that the growth rate of the aggregates is faster in one specific direction, leading to formation of elongated structures. Deviation from isotropic growth has also been shown to correlate with an increase in β-sheet content in studies of other amyloidogenic peptides (45). This could therefore indicate that monomers, dimers and trimers are predominantly unstructured while oligomers >4 start to form more extended β-sheets. If the relative oligomer signals in the mass spectra for wildtype and CC variants are compared it can be seen that indeed it is the oligomers with n ≥ 4 and especially with n ≥ 8 that are increased for the CC variant. Interestingly the CC variant has similar or slightly lower amounts of dimers and trimers compared to the WT **(Figure 2E)**.

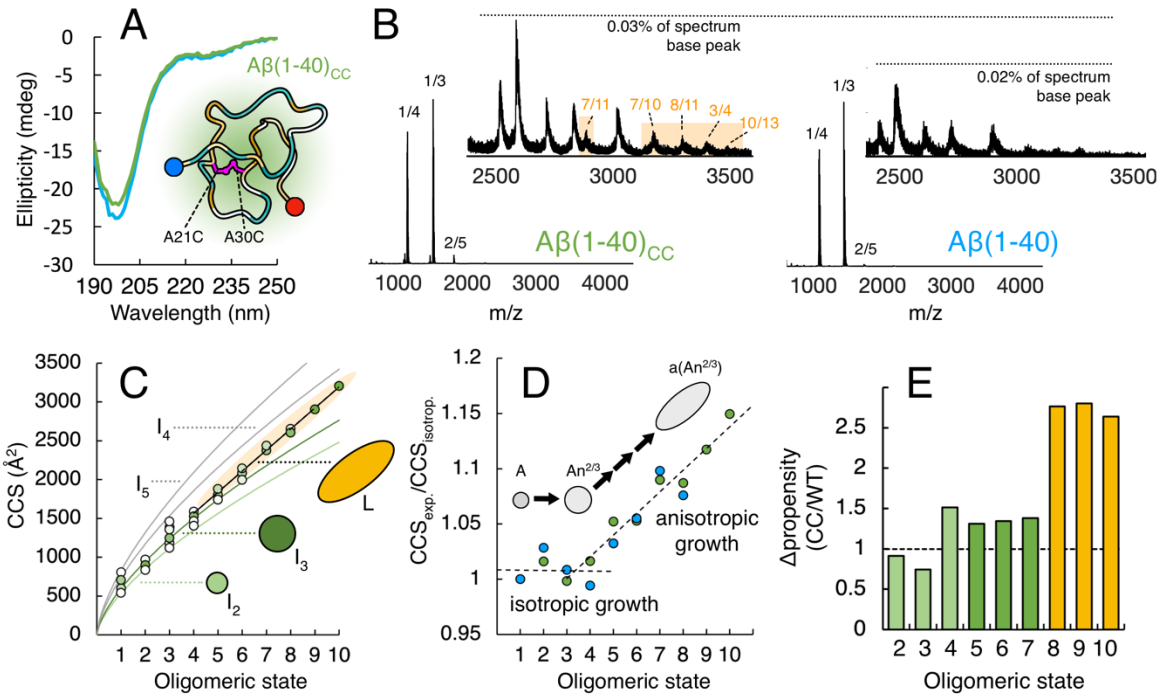

Figure 2. Self-assembly of Aβ(1-40) and Aβ(1-40)CC A) Far-UV CD spectra of 40 µM Aβ(1-40)CC (green) and Aβ(1-40) (blue) in 20 mM phosphate buffer pH 7.4, shown together with the top cluster structure after 2 µs MD simulations of the Aβ(1-40)CC monomer, colored according to hydrophobicity (orange = hydrophobic, blue = hydrophilic). The N-terminus is highlighted in blue, while the C-terminus is highlighted in red. The disulfide bond formed between C21 and C30 is shown in magenta. **B)** Mass spectra of 20 µM Aβ(1-40)CC (left) and Aβ(1-40)

(right) in 200 mM ammonium acetate pH 6.8, with the oligomeric region magnified. Peaks are annotated by their oligomeric state/charge state ratio (n/z). States that are especially enriched in the CC variant are marked in orange. **C)** CCS of oligomers plotted against oligomeric state. The experimental measurements are shown as circles colored according to relative intensity within an oligomeric state (white to green). Solid lines represent the theoretical growth behavior of isotropic growth (I) and linear growth (L). **D)** Ratio between detected experimental CCS (intensity-weighted average) and the theoretical isotropic growth according to I_3. The dashed lines represent fits to the data points, between n = 1 and n = 4, and between n = 4 and n = 10. **E)** Ratio between the relative intensity of oligomers in Aβ(1-40)$_{CC}$ and in Aβ(1-40). Values above 1 indicate that the oligomeric state is increased in Aβ(1-40)$_{CC}$ compared to in Aβ(1-40).

Oligomer growth is linked to folding into the hairpin motif

Native IM-MS suggest growth of oligomers with a shift in structure occurring upon formation of larger assemblies. The technique is however not capable of determining the structure of the oligomers. For this we instead analyze MD simulations of some of the species detected by mass spectrometry. We let the unstructured Aβ(1-40)$_{CC}$ assemble into dimers, which in turn assemble into tetramers **(Figure 3A)**. The β-sheet propensity of Aβ(1-40)$_{CC}$ increases when the oligomer size increases, but is overall lower compared to the wild-type variant **(Figure 3B)**. The increase in β-sheet propensity can also be seen in the top cluster structures **(Figure 3A)**, as unstructured Aβ(1-40)$_{CC}$ aggregates into a tetramer where a core of short antiparallel β-sheets is formed by the most hydrophobic segment. Three out of four monomer units in the top cluster tetramer have folded into a hairpin structure similar to the β-hairpin found in the affibody complex from which the starting monomeric structure was derived (Figure S10). It can be seen that the two hydrophobic segments (Figure S1) in the central hydrophobic core (CHC, residues 16 - 22) and the C-terminal hydrophobic regions (residues 30 - 40) are indeed the segments with β-sheet propensity both in the wild type and CC variant. These regions show β-sheet probabilities of around 60 to 90% in wild-type tetramers and around 40% to 70% in the CC variant tetramers **(Figure 3C)**. For both peptides, the N-terminal region (residues 1 to 15) and the hinge region (residues 21 to 28), which link the β-strands, are highly disordered and tend to form β-bend/turn or random coil conformations (Figure S11).

The flexibility of the N-terminal region is also seen by analyzing the intramolecular contacts where residues 5-10 make multiple short-range contacts with the entire peptide chain both in the dimer **(Figure 3D, top)** and in the tetramer (Figure S12). The other significant interaction pattern is the intrapeptide interaction between the CHC and the C-terminal region, which enables the formation of the hairpin motif. In Aβ(1-40)$_{CC}$ the interaction between the most C-terminal residues (35-40) and the CHC is reduced, which results in shorter β-sheets compared to the wild type oligomers. The interpeptide contacts in the oligomers are also mostly between the hydrophobic segments with high β-sheet propensity. These interactions are CHC with CHC, CHC with C-terminal hydrophobic region, and C-terminal hydrophobic region with C-terminal hydrophobic region **(Figure 3D, bottom)**. The interpeptide contacts are reduced in the Aβ(1-40)$_{CC}$ dimers compared to the wild type. For the wild-type dimer, the highest contact density is observed between the two CHC regions, the regions with the

highest β-sheet propensity **(Figure 3C)**, which also display stronger interaction energies than interactions between the C-terminal hydrophobic regions (Figure S13). The interaction energies reveal that the Lennard-Jones interactions within the wild-type dimer are stronger than those within the Aβ(1-40)$_{CC}$ dimer. Moreover, attractive interactions between the oppositely charged residues K16-D23 in Aβ(1-40) and E11-K28 in Aβ(1-40)$_{CC}$ are observed. Similar patterns for interpeptide interactions are observed in the wild-type and CC tetramers (Figure S12). Overall, Aβ(1-40)$_{CC}$ appears to be less ordered than the wild-type variant, especially for the monomeric and dimeric states. This can also be seen by higher root mean square fluctuations (RMSF) of the Cα atoms in the Aβ(1-40)$_{CC}$ dimer compared to the Aβ(1-40) dimer (Figure S14). It seems that the disulfide bond adds extra forces to the peptide that increase its flexibility, thereby suppressing intrapeptide interactions and intramolecular hydrogen bonds away from the enforced intramolecular A21C-A30C contact. Aβ(1-40)$_{CC}$ furthermore has a slightly higher amount of exposed hydrophobic surfaces compared to wild-type Aβ(1-40) in monomers and dimers **(Figure 3E)**.

The flexibility of the peptides is further seen by analyzing the compactness of the oligomeric species **(Figure 3F)**. Dimers of both peptide variants populate a wide distribution. The Aβ(1-40) dimer population is centered around a compactness value of 0.5, while the Aβ(1-40)$_{CC}$ dimer displays a bimodal distribution with two peaks around 0.4 and 0.6. Aβ(1-40)$_{CC}$ dimers display both a wider distribution of compactness and also populate slightly more extended (less compact) states **(Figure 3F, top)**. The tetramers of both peptide variants are less polydisperse, with sharper distributions **(Figure 3F, bottom)**, compared to the dimers. This is in agreement with folding into more well-defined β-sheet structures upon oligomer growth. Aβ(1-40)$_{CC}$ tetramers populate more compact conformations represented by values between 0-6-0.8, in addition to structures with the same compactness as Aβ(1-40), around 0.5.

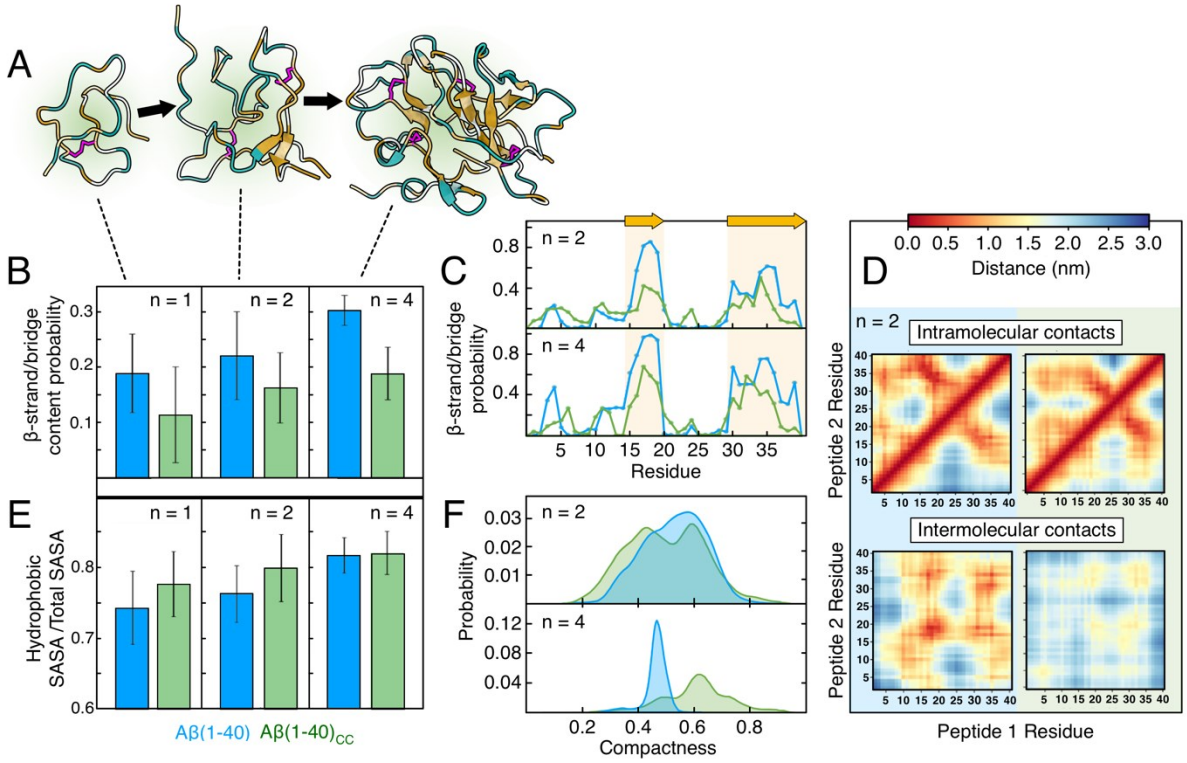

Figure 3. Structural properties obtained by MD simulations for Aβ(1-40) and Aβ(1-40)cc. A) The most populated cluster structure of Aβ(1-40)cc monomer, dimer and tetramer. Colored according to hydrophobicity (orange = hydrophobic, blue = hydrophilic). The clustering was performed by gromos algorithm and a cutoff distance of 0.4 nm. **B)** Average β-strand/bridge content probability for Aβ(1-40) (blue) and Aβ(1-40)cc (green) for monomers, dimers and tetramers. **C)** Probability of β-strand/bridge of peptides residue for Aβ(1-40) (blue) and Aβ(1-40)cc (green) for dimers (top) and tetramers (bottom). The two regions with high β-strand/bridge probability are marked in yellow. **D)** The intra- (top) and interpeptide (bottom) contacts between residues for Aβ(1-40) dimer (left, blue) and Aβ(1-40)cc dimer (right, green). The intrapeptide contacts within peptide 1 are shown below the main diagonal and within peptide 2 above the main diagonal. The color bar shows the average intra/inter-residue distance (in nm). **E)** The ratio of the average hydrophobic solvent accessible surface area (SASA) over the average total SASA ratio for Aβ(1-40) (blue) and Aβ(1-40)cc (green) for monomers, dimers and tetramers. F) The compactness distribution probabilities of Aβ(1-40) (blue) and Aβ(1-40)cc (green) for dimers (top) and tetramers (bottom). The compactness values range from 0 (extended) to 1 (compact).

C-terminal truncation leads to a decrease in oligomerization propensity

The results indicate that folding of monomeric units into the β-hairpin structure is essential for formation of larger Aβ oligomers, as they assemble by formation of extended antiparallel β-sheets. We next tested this by analyzing Aβ peptide variants which have been truncated from the C-terminal end, namely Aβ(1-38), Aβ(1-28) and Aβ(1-16) **(Figure 4A)**. As the hairpin is formed by interactions between the C-terminal and the middle segments of Aβ the truncation should lead to a lower propensity to fold into the hairpin structure, and therefore a lower propensity to form higher oligomers. In Aβ(1-38) two hydrophobic valine residues, V39 and V40, are removed which has previously been found to greatly reduce the amyloid aggregation rate (46). In Aβ(1-28), the

part of the peptide that is found extracellularly in APP, the entire C-terminal hydrophobic region ($_{29}$GAIIGLMVGGVV$_{40}$) is removed. The peptide thus only contains the N-terminal region, the CHC, and the hinge region. The N-terminal region in isolation, Aβ(1-16), was also analyzed.

An MD simulation of the Aβ(1-38) monomer shows that a compact hairpin-like state where the CHC and the C-terminal region are close to each other is still able to form **(Figure 4B, top)**. MD simulations do however unsurprisingly show a complete loss of hairpin structure in Aβ(1-28), as the entire second β-strand has been removed. The Aβ(1-28) monomer instead forms an entirely unstructured random coil **(Figure 4B, bottom)**. CD spectrometry indicates that the truncated variants have similar random coil-dominated spectral signatures as Aβ(1-40). Truncation does however lead to a decrease in the aggregation propensity when incubated without agitation at 37 °C **(Figure 4C)**.

Aβ(1-38) populates multiple oligomeric states **(Figure 4D)**, similar to those populated by Aβ(1-40). This is in agreement with MD simulations of the Aβ(1-38) dimer that shows formation of a similar antiparallel β-sheet structure as in Aβ(1-40). Truncation of the entire C-terminal region has a more drastic effect with the oligomeric population of Aβ(1-28) shifted towards mostly dimeric and trimeric states **(Figure 4E)**. As an example, the n/z = 2/3 is especially highlighted **(Figure 4D, Figure 4E, colored boxes)**. This signal contains overlapping dimeric, tetrameric and hexameric states in Aβ(1-38), with the tetramer being the main component. In Aβ(1-28) the distribution is heavily shifted towards the dimer, and the hexameric component is completely absent. Only minor pentameric (n/z = 5/7, 5/6) and hexameric (n/z = 3/4, 6/7) signals are overall seen for Aβ(1-28), with the hexamer being the highest oligomeric state detected in the spectrum. Ion mobility analysis show that Aβ(1-38) oligomers follow similar growth trends as Aβ(1-40) (Figure S16). Aβ(1-28) on the other hand seems to only grow isotropically, according to the most compact I_2 family **(Figure 4F)**. Aβ(1-28) forms a mostly unstructured dimer in MD simulations **(Figure 4E)**, rather than β-strand interactions between the CHC regions in the two monomeric units as in Aβ(1-38) and Aβ(1-40) (Figure S17). The CHC is probably however involved in the formation of the observed small Aβ(1-28) oligomers, as the N-terminal segment Aβ(1-16) only forms monomers and dimers (Figure S18). This can be seen in the intermolecular contact map for the Aβ(1-28) dimer, where the interactions are vague but clearly involves residues 15-20 (Figure S17).

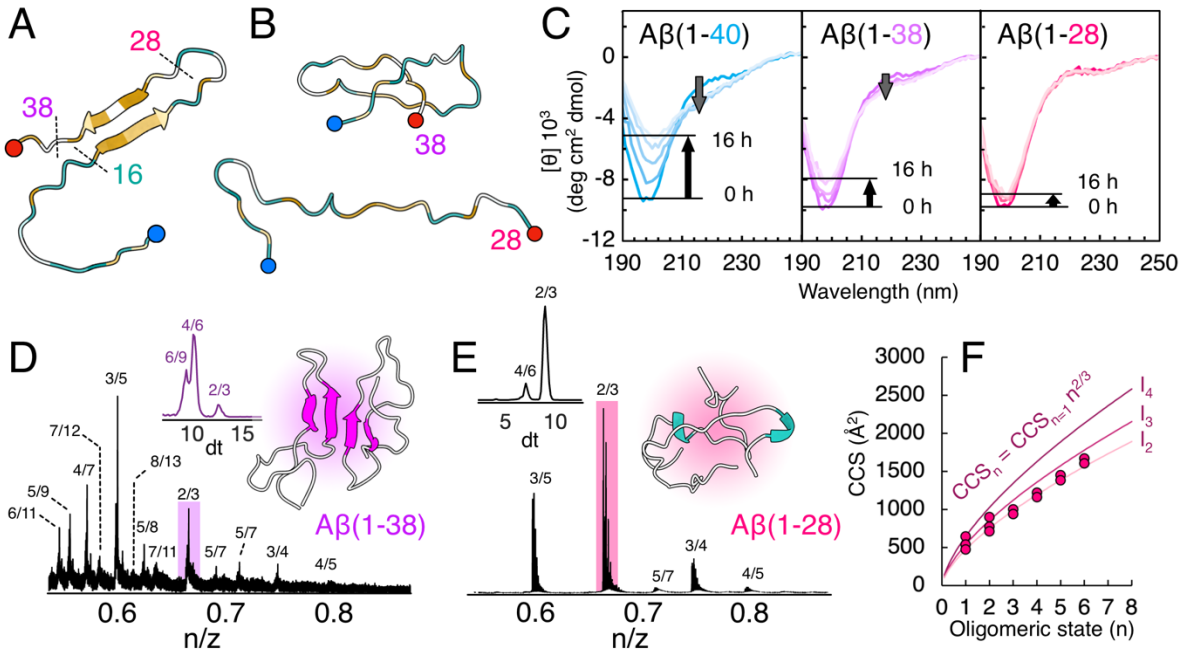

Figure 4. C-terminally truncated Aβ variants. A) Overview of the different truncation variants. **B)** Top cluster structures of Aβ(1-38) (top) and Aβ(1-28) (bottom) after 2 μs MD simulations. Colored according to hydrophobicity (orange = hydrophobic, blue = hydrophilic). The N-termina are highlighted in blue, while the C-termina are highlighted in red. **C)** Far-UV CD spectra of 40 μM Aβ(1-40) (blue), Aβ(1-38) (purple), and Aβ(1-28) (pink) in 20 mM phosphate buffer pH 7.4. CD spectra recorded every 2 h for 16 hour are shown. Arrows indicate changes at 196 nm (random coil signal) and 215 nm (β-sheet signal). **(D-E)** The oligomeric regions from the mass spectra of 20 μM Aβ(1-38) (D) and Aβ(1-28) (E) in 200 mM ammonium acetate pH 6.8. Peaks are annotated by their oligomeric state/charge state ratio (n/z). Note that the x-axes of the mass spectra are shown as oligomeric state (n) / charge state (z) (m/z divided by monomeric mass of the peptide variant) to enable easy comparison between the two variants. The ion mobility (drift time, dt) distribution for the n/z = 2/3 state (highlighted in a colored box) is shown as an insert for both variants. The top clusters of dimers of Aβ(1-38) (D) and Aβ(1-28) (E) are also shown as inserts, colored according to secondary structure (coil = gray, sheet = magenta, helix = cyan). **F)** Measured collision cross sections (CCS) for each oligomeric state in Aβ(1-28), solid lines represent the theoretical isotropic growth originating from each monomer state

Discussion

We have here exemplified, using both experimental and computational methods and by probing different Aβ variants, that a hairpin motif is important for formation of oligomeric species. Especially the formation of larger oligomers (n ≥ 4) seems to require folding into the hairpin motif **(Figure 2E)**. This hairpin motif arises due to the high β-strand propensity of two hydrophobic segments in the Aβ peptide, and the transient folding of these segments onto each other. Such a β-hairpin structure enables continued addition of monomeric units to form a larger β-sheet structure, which is not possible in the $Aβ(1-40)_{Scr}$ peptide which displays high helical propensity **(Figure 1)**. The β-sheet propensity of Aβ increases with aggregation number (n), indicating some cooperativity

in the folding process. The monomeric hairpin fold does not seem to be stable by itself in solution **(Figure 2A)**, which is also supported by solution state NMR spectroscopy data (47). Assembly of a larger β-sheet core does however seem to make the folding process more favorable and increases the stability of the structures. This could be compared to a "folding upon binding" event commonly observed for intrinsically disordered proteins, with the unstructured Aβ monomer folding upon binding to a more structured Aβ oligomer. A gradual increase in β-sheet structure has previously been observed by CD spectroscopy of isolated Aβ(1-40) oligomers stabilized by intermolecular photo-induced crosslinking of unmodified peptides (PICUP) (48) Those PICUP-CD spectroscopy results report a β-sheet content of 24% in the monomer, which increases to 45% in the crosslinked tetramer. This can be compared to 19% for monomer and 30% for tetramer in our here reported MD simulations.

An elongation in the structure occurs for oligomers larger than the tetramer. This could be interpreted as formation of an extended sheet structure **(Figure 5A)**, as deviation from isotropic growth has been previously reported to correlate with an increase in β-sheet content in other amyloidogenic peptides (45). It is also interesting to note that our previous study in membrane-mimicking micelles showed that all detected Aβ oligomers follow isotropic growth in such an environment, and that oligomerization stops at the hexamer (49). We here observe that isotropic growth according the I_3 family also occurs until at the hexamer in a simple aqueous solution **(Figure 1F, Figure 2C)**. The conformation distribution for the hexamer is however heavily shifted towards the linearly growing oligomer family, which goes on to form even larger oligomers. This would suggest that the enrichment of oligomers in the micellar environment arises from stabilization of the globular-like (isotropic) Aβ structures by interactions with the micelle. This inhibits continued growth into the extended and linearly growing oligomers observed in absence of micelles. The micelles also inhibit formation of amyloid fibrils (49, 50), indicating that the extended oligomers could be on-pathway for fibril formation. The same holds true for the Aβ(1-28) peptide which has very low aggregation propensity and does also not form the extended oligomers **(Figure 4F)**.

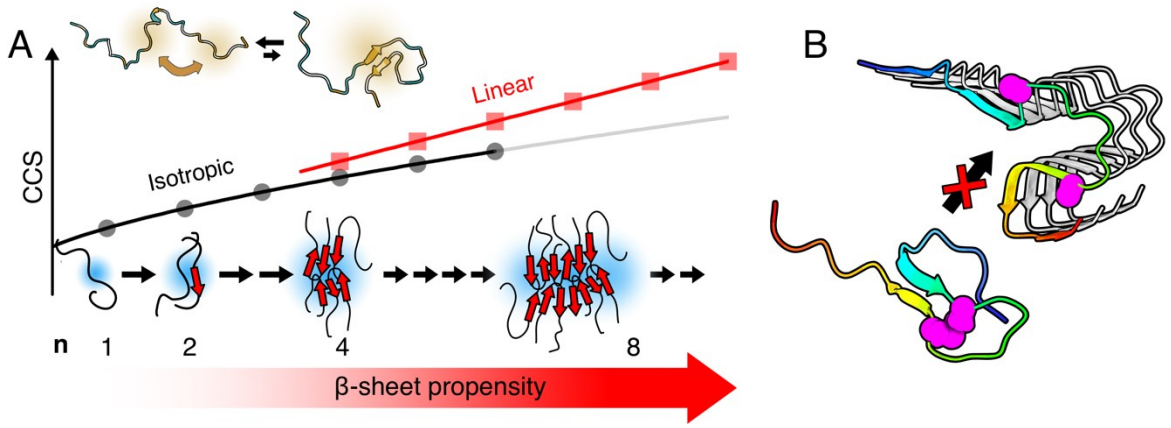

Figure 5. Overview of suggested steps in the early self-assembly process of Aβ peptides. A) Aβ peptides have a propensity to fold into a β-hairpin structure by interactions between the two hydrophobic segments in the peptide. Growth of Aβ oligomers follow isotropic growth until n = 4, at which point linear growth starts to dominate, indicating formation of more elongated structures. A gradual increase of β-sheet content is also seen

upon formation of larger oligomers in a "folding upon self-assembly" mechanism**. B)** The structure in oligomers (generated by MD simulations) is distinct from the structure of fibrils (cryo-EM pdb 7q4b (51)) and formation of amyloid fibril require a structural rearrangement. Residues 21 and 30, which are substituted to cysteines in $A\beta_{CC}$ are shown as pink spheres.

It is however important to note that the elongated oligomers are most likely not fibrillar-like structures. The different oligomer conformations detected by IM was recently suggested to correspond to different fibrillar morphologies (14). We consider this unlikely as oligomers are known to be structurally distinct from fibrils, as they for example are recognized by different antibodies (52, 53). The difference between fibrils and oligomers has also been recently exemplified by fitting the growth of oligomers over time to kinetic models (7). Such models show that the occurrence of detectable metastable oligomers is not in agreement with direct formation of fibrils by elongation of oligomers as was suggested in ref (14). This is because elongation is energetically a very favorable process and would result in five orders of magnitude lower amounts of oligomers than what is being experimentally observed (7). Our MD simulations also give rise to Aβ peptides in distinct anti-parallel β-sheet structures which have been classified as being characteristic for oligomeric Aβ aggregates (37, 54–56). It is on the contrary known from solid state NMR spectroscopy and cryo-electron microscopy that mature fibrils typically have parallel β-sheets (51, 57). A structural rearrangement therefore needs to occur upon nucleation where the β-strands twist 90°, which involves breaking of intramolecular hydrogen bonds and formation of new intermolecular hydrogen bonds **(Figure 5B)**. It is also from this obvious why the $A\beta_{CC}$ peptides cannot form mature fibrils in their oxidized state, as such a rearrangement is not possible if residues 21 and 30 are crosslinked together **(Figure 5B, pink spheres)**. If the extended oligomers observed by IM were fibrillar-like we would therefore expect them to be absent rather than increased in $A\beta_{CC}$.

Oligomers with n = 12 (52 kDa) are also not large complexes for native MS detection. It is therefore intriguing why we do not observe significantly larger oligomers in the mass spectra. One interesting and related question is at which point the Aβ oligomers convert into amyloid-like structures. Determining the critical size of the Aβ nucleus (n_c) is difficult experimentally as the nucleus represents a state with high free energy and is thus sparsely populated. Modeling however suggests that n_c occur at relatively low n, typically between n = 6 and 14 (58–60). The critical size has also been shown in such studies to depend on peptide concentration (7, 60), and is also most likely also highly dependent on other solution conditions. A recent experimental study using fluorescence correlation spectroscopy showed that the BRICHOS chaperone, which inhibits fibril-dependent secondary nucleation, binds to Aβ aggregates as small as n = 8 (61). This indicates that such small aggregates might already have converted to fibrillar-like aggregates which can induce secondary nucleation. It is thus possible that the elongated oligomers that we observe in MS are species which are able to convert into fibrillar states which quickly elongate. The propensity to convert should increase with increasing n according to modeling, but conversion is still a rare event (7). The inability for $A\beta_{CC}$ to convert **(Figure 5B)** into the fibril state agrees with the observed increase of larger $A\beta_{CC}$ oligomers. We also observe rather surprisingly that the introduction of the intramolecular disulfide bond in $A\beta(1\text{-}40)_{CC}$ does not stabilize the monomer in the β-hairpin state. We however

observe folding into this motif, similar to the wild-type peptide, in the tetrameric state, in a "folding upon self-assembly" mechanism. The experimental IM-MS results show similar charge state distributions, detected oligomeric states and collision cross sections for Aβ(1-40)$_{CC}$ compared to the wild type, indicating that the disulfide bond is compatible with the wild-type oligomer structures. The experimentally observed enrichment of large oligomers and decrease in smaller oligomers in Aβ(1-40)$_{CC}$ could therefore also (in addition to the inhibition of the rearrangement step into fibril) be due to structural effects caused by the introduced crosslink that affects the monomeric state more than the oligomeric state.

CONCLUDING REMARKS

This study increases our understanding of the structural transitions occurring at the very early stages of Aβ aggregation. Such early structures are believed to be especially toxic and could therefore be related to AD pathology. Our results show that oligomers are intrinsically polydisperse, indicating an underlying frustrated energy landscape. This is probably in part a reason for the toxicity of these species as they are very prone to interact with other cellular species to reduce their chemical potential. Such cellular species could for example be the cellular membrane, resulting in a loss of cell or organelle integrity and uncontrolled leakage (62–64). The oligomers could also co-aggregate with other biomolecules such as functionally important cellular proteins, resulting in breakdown of cellular function and loss of cellular proteostasis (65, 66). Such events would under normal cellular conditions be prevented by control systems such as chaperone proteins. Interestingly it has been suggested that some chaperone proteins with anti-amyloid activity such as BRICHOS and the Hsp40-type DNAJB6 chaperone might bind their clients by forming complementary β-strand/β-strand interactions (67). DNAJB6 is known to bind Aβ oligomers rather than monomers or fibrils, and therefore must be able to discriminate between these structures (68, 69). DNAJB6 also has a higher affinity for tetramers compared to dimers (69), which could indicate that it is indeed the β-hairpin motif that is recognized by the chaperone, as the β-sheet propensity is here shown to clearly increase upon going from dimer to tetramer. Such binding therefore inhibits the formation of the extended oligomers which are able to nucleate into amyloid.

The findings presented here combined with observations from nature's own anti-amyloid systems could provide useful insights into how to rationally design therapies against AD. Antibodies can be raised against epitopes of unknown structure, but other approaches would require more in-depth information about the molecular structures of the target species. It would be reasonable to imagine that inhibition of folding into the hairpin motif could be a possible strategy for preventing Aβ toxicity. This could be achieved be design of therapeutic peptides or proteins that bind with high affinity to the regions in Aβ with high β-sheet propensity and in that way out-compete the Aβ-Aβ interactions. The perhaps best example of this is anti-amyloid peptides that use the Aβ(16-20) sequence (KLVFF) from the CHC (70–74). Our results clearly show how CHC-CHC interactions are important interactions in oligomer formation, meaning that this segment is indeed very promising to target to decrease oligomer formation. Another approach would be to design therapeutic molecules that recognize and bind Aβ specifically in the hairpin motif, similar to the above-described chaperones. This could be achieved by designed small affinity proteins (so-called affibodies) (75, 76) or RNA aptamers (77). A combination of MD

simulations and IM-MS could furthermore prove to be important tools in drug-discovery studies. It would be anticipated that the β-sheet oligomers observed as linearly growing species in IM could be interesting to monitor upon introducing modulating molecules. IM-MS provides very quick detection of effects on the oligomer population and could be used to screen many molecules (78). MD simulations could then be employed to pinpoint the effects on oligomer structure by interesting molecules identified from the experiments.

METHODS

Peptide preparation

Lyophilized recombinant Aβ(1-40) was purchased from AlexoTech (Umeå, Sweden). Lyophilized recombinant Scrambled Aβ(1-40) (Figure S1) ($A\beta(1\text{-}40)_{Scr}$), Aβ(1-38) and Aβ(1-28), as well as lyophilized synthetic Aβ(1-16) was purchased from rPeptide (Georgia, USA). All peptides were dissolved in 6 M guanidine hydrochloride and purified and buffer exchanged into 200 mM ammonium acetate pH 6.8 (for IM-MS) or 20 mM phosphate buffer pH 7.4 (for optical spectroscopy) using either a Superdex Increase 75 10/300 (Cytiva, Sweden) or Superdex Icrease 30 10/300 (Cytiva, Sweden) size exclusion column.

Recombinant $A\beta(1\text{-}40)_{CC}$ peptides were kindly gifted by Professor Torleif Härd, SLU, and Professor Cecilia Emanuelsson, Lund University and were provided as a co-expressed Aβ -affibody complex as earlier described (30). This Aβ-affibody complex was immobilized on a 5 mL HisTrap (Cytiva, Sweden) column. $A\beta(1\text{-}40)_{CC}$ was separated from the immobilized affibody by washing with 20 mM phosphate buffer pH 7.7 with 150 mM NaCl 10 mM imidazole and 6 M guanidine hydrochloride. The $A\beta(1\text{-}40)_{CC}$ peptides were then further purified and buffer exchanged using size exclusion chromatography as described above.

The concentration of the obtained peptide fractions was determined using UV absorption of the Y10 (ε = 1490 $M^{-1}cm^{-1}$) residue by near-UV spectroscopy at 280 nm.

Native IM-MS

Native ESI-IM-IM was performed on a Synapt G2-S equipped with an ion mobility cell. The peptides were diluted to a final concentration of 20 μM in 200 mM ammonium acetate buffer, pH 6.8. Samples were ionized in a nanoESI source using commercial metal coated borosilicate spray emitters (ThermoScientific). All peptides were ionized in the positive ion mode using a capillary voltage of 1.5 kV. The remaining settings were as follow: cone voltage 50 V at an offset of 50 V, 25 °C source temperature, trap gas 10 mL/min. IM parameters were set at wave height 40 V and wave velocity at 1200 m/s. The collision energy in the ion mobility cell was set at 5 V.

Drift time information from each measurement was retained using the program DriftScope (Waters, USA). Drift times from IM-MS for peptides were calibrated to obtain CCS values as previously describes (79). Bovine ubiquitin (Sigma-Aldrich), Bovine β-Lactoglobulin (Sigma-Aldrich) and honeybee Melittin (Sigma-Aldrich) were used to create the collision cross-section calibration curve as these proteins span the size scale of Aβ oligomers.

The reference CCS values for the calibrant proteins were obtained from literature (35, 80) The obtained raw data from mass spectrometry were analyzed using the software massLynx V4.1.(Waters, USA).

Circular dichroism spectroscopy

Circular dichroism (CD) spectrometry was performed on a Chirascan spectrometer (Applied Photophysics, U.K.) using a 2 mm quartz cuvette. The ellipticity between 250-190 nm (1 nm step size, 4 s sampling time per point) was measured on samples of 40 µM peptide in 20 mM phosphate buffer, pH 7.4. The secondary structures of Aβ peptides were studied over time in aggregation kinetics experiments for a total of 18 hours at 37°C without agitation.

Molecular dynamics simulations

The simulations of the wild-type Aβ(1-40) monomer were initiated from a β-hairpin conformation (pdb 2otk (30)). The mutated $A\beta(1\text{-}40)_{CC}$ with the disulfide bond was built from the same structure by mutating the wild-type Aβ(1-40) amino acids A21 and A30 to C21 and C30 using Charmm-GUI (81), and, to mimic the experimental conditions, a methionine amino acid was added to the N-terminal using PyMOL (Schrödinger) (82) and the Modloop server (83) to relax the N-terminal residues. The initial structures of Aβ(1-40) scrambled ($A\beta(1\text{-}40)_{Scr}$), Aβ3(1-38) and Aβ(1-28) peptides were predicted by AlphaFold2 (84). The initial structures were placed in a cubic box, water molecules and 150 mM NaCl were added, and the systems were neutralized by inserting extra Na^+ ions. After that, the systems underwent energy minimization using the steepest descent algorithm to remove clashes between atoms (85), followed by two equilibration steps each for 1 ns under canonical (NVT) and isobaric-isothermic (NPT) ensemble conditions. During the equilibration, the pressure was kept at 1.0 bar using the Parrinello-Rahman pressure coupling method (86, 87) and the temperature was kept at 298 K using the velocity-rescale thermostat method (88). The production MD simulations were conducted under NPT ensemble conditions. Electrostatic interactions were calculated using the particle mesh Ewald method (PME) with a real-space cutoff distance of 1.2 nm; van der Waals interactions were also cut off at 1.2 nm (89). Each monomer was simulated for 2 µs. For the dimer systems, two monomers obtained from the highest populated clusters were added to the simulation box at 0.8 nm away from each other to initiate the dimer simulations. The dimers were simulated for 10 µs for Aβ(1-40) and $A\beta(1\text{-}40)_{CC}$ and 2 µs for $A\beta(1\text{-}40)_{Scr}$, Aβ(1-38), and Aβ(1-28). Moreover, in the cases of Aβ(1-40) and $A\beta(1\text{-}40)_{CC}$, tetramers were also simulated, each for 10 µs, starting from the highest populated dimer structures placed at 0.8 nm away from each other. The same simulation setups as used in the monomer simulations were used for the dimer and tetramer simulations. All MD simulations were performed using GROMACS 2020/2022 (90, 91) with the Charmm36m force field (92) and the TIP3P water model.

ACKNOWLEDGEMENTS

These studies were supported by grants from the Swedish Research Council (to LLI), the Swedish Brain Foundation, the Stockholm Regional Council (to AG), and the Palestinian-German Science Bridge financed by the German Federal Ministry of Education and Research (BMBF) (to MK). We gratefully acknowledge the computing time granted through JARA - High Performance Computing (Project AMYLOID-MSM) on the

supercomputer JURECA at Forschungszentrum Jülich. Professors Cecilia Emanuelsson, Torleif Härd, and Wolfgang Hoyer are thanked for providing peptide samples, Professors Mikael Oliveberg and Cecilia Emanuelsson are thanked for valuable discussions, and Dr. Claudia Möckel of the Stockholm University mass spectrometry facility is thanked for excellent technical assistance.

REFERENCES

1. Zheng, H., and Koo, E. H. (2011) Biology and pathophysiology of the amyloid precursor protein. Mol Neurodegener. 6, 1–16
2. Hellstrand, E., Boland, B., Walsh, D. M., and Linse, S. (2010) Amyloid β-protein aggregation produces highly reproducible kinetic data and occurs by a two-phase process. ACS Chem Neurosci. 1, 13–18
3. Pike, C. J., Walencewicz, A. J., Glabe, C. G., and Cotman, C. W. (1991) In vitro aging of ß-amyloid protein causes peptide aggregation and neurotoxicity. Brain Res. 563, 311–314
4. Hardy, J., and Higgins, G. (1992) Alzheimer's disease: the amyloid cascade hypothesis. Science (1979). 256, 184–185
5. Selkoe, D. J., and Hardy, J. (2016) The amyloid hypothesis of Alzheimer's disease at 25 years. EMBO Mol Med. 8, 595–608
6. Tiiman, A., Jarvet, J., Gräslund, A., and Vukojevic, V. (2015) Heterogeneity and intermediates turnover during amyloid-β (Aβ) peptide aggregation studied by Fluorescence Correlation Spectroscopy. Biochemistry. 10.1021/acs.biochem.5b00976
7. Michaels, T. C. T., Šarić, A., Curk, S., Bernfur, K., Arosio, P., Meisl, G., Dear, A. J., Cohen, S. I. A., Dobson, C. M., Vendruscolo, M., Linse, S., and Knowles, T. P. J. (2020) Dynamics of oligomer populations formed during the aggregation of Alzheimer's Aβ42 peptide. Nat Chem. 12, 445–451
8. Šarić, A., Michaels, T. C. T., Zaccone, A., Knowles, T. P. J., and Frenkel, D. (2016) Kinetics of spontaneous filament nucleation via oligomers: Insights from theory and simulation. Journal of Chemical Physics. 10.1063/1.4965040
9. Strodel, B. (2021) Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins: Energy landscapes of IDPs and protein aggregation. J Mol Biol. 10.1016/J.JMB.2021.167182
10. Cohen, S. I. A., Linse, S., Luheshi, L. M., Hellstrand, E., White, D. A., Rajah, L., Otzen, D. E., Vendruscolo, M., Dobson, C. M., and Knowles, T. P. J. (2013) Proliferation of amyloid- 42 aggregates occurs through a secondary nucleation mechanism. Proceedings of the National Academy of Sciences. 110, 9758–9763
11. Mclean, C. A., Cherny, R. A., Fraser, F. W., Hons, B., Fuller, S. J., Smith, M. J., Beyreuther, K., Bush, A. I., and Masters, C. L. (1999) Soluble Pool of A Amyloid as a Determinant of Severity of Neurodegeneration in Alzheimer's Disease. 10.1002/1531-8249
12. Lue, L. F., Kuo, Y. M., Roher, A. E., Brachova, L., Shen, Y., Sue, L., Beach, T., Kurth, J. H., Rydel, R. E., and Rogers, J. (1999) Soluble Amyloid β Peptide Concentration as a Predictor of Synaptic Change in Alzheimer's Disease. Am J Pathol. 155, 853–862
13. Yu, L., Edalji, R., Harlan, J. E., Holzman, T. F., Lopez, A. P., Labkovsky, B., Hillen, H., Barghorn, S., Ebert, U., Richardson, P. L., Miesbauer, L., Solomon, L., Bartley, D., Walter, K., Johnson, R. W., Hajduk, P. J., and Olejniczak, E. T. (2009) Structural characterization of a soluble amyloid β-peptide oligomer. Biochemistry. 48, 1870–1877
14. Lieblein, T., Zangl, R., Martin, J., Hoffmann, J., Hutchison, M., Stark, T., Stirnal, E., Schrader, T., Schwalbe, H., and Morgner, N. (2020) Structural rearrangement of amyloid-β upon inhibitor binding suppresses formation of Alzheimer's disease related oligomers. Elife. 10.7554/eLife.59306
15. Österlund, N., Lundqvist, M., Ilag, L. L., Gräslund, A., and Emanuelsson, C. (2020) Amyloid-β oligomers are captured by the DNAJB6 chaperone: Direct detection of interactions that can prevent primary nucleation. Journal of Biological Chemistry. 295, 8135–8144
16. Fändrich, M. (2012) Oligomeric Intermediates in Amyloid Formation: Structure Determination and Mechanisms of Toxicity. J Mol Biol. 421, 427–440
17. Hefti, F., Goure, W. F., Jerecic, J., Iverson, K. S., Walicke, P. A., and Krafft, G. A. (2013) The case for soluble Aβ oligomers as a drug target in Alzheimer's disease. Trends Pharmacol Sci. 34, 261–266

18. Cummings, J., Lee, G., Nahed, P., Zadeh, M. E., Kambar, N., Zhong, K., Fonseca, J., and Taghva, K. (2022) Alzheimer's disease drug development pipeline: 2022. Alzheimer's & Dementia: Translational Research & Clinical Interventions. 8, e12295
19. Lannfelt, L., Möller, C., Basun, H., Osswald, G., Sehlin, D., Satlin, A., Logovinsky, V., and Gellerfors, P. (2014) Perspectives on future Alzheimer therapies: Amyloid-β protofibrils-A new target for immunotherapy with BAN2401 in Alzheimer's disease. Alzheimers Res Ther. 6, 1–8
20. Söderberg, L., Johannesson, M., Nygren, P., Laudon, H., Eriksson, F., Osswald, G., Möller, C., and Lannfelt, L. (2022) Lecanemab, Aducanumab, and Gantenerumab — Binding Profiles to Different Forms of Amyloid-Beta Might Explain Efficacy and Side Effects in Clinical Trials for Alzheimer's Disease. Neurotherapeutics. 1, 1–12
21. Siemers, E., Hitchcock, J., Sundell, K., Dean, R., Jerecic, J., Cline, E., Iverson, K., Moore, J., Edgar, C., Manber, R., Fuin, N., Poppe, T., and Barton, R. (2022) ACU193, a Monoclonal Antibody that Selectively Binds Soluble Aβ Oligomers: Development Rationale, Phase 1 Trial Design, and Clinical Development Plan. Journal of Prevention of Alzheimer's Disease. 10.14283/JPAD.2022.93/TABLES/1
22. Morimoto, A., Irie, K., Murakami, K., Masuda, Y., Ohigashi, H., Nagao, M., Fukuda, H., Shimizu, T., and Shirasawa, T. (2004) Analysis of the secondary structure of β-amyloid (Aβ42) fibrils by systematic proline replacement. Journal of Biological Chemistry. 279, 52781–52788
23. Rezaei-Ghaleh, N., Amininasab, M., Giller, K., Kumar, S., Stündl, A., Schneider, A., Becker, S., Walter, J., and Zweckstetter, M. (2014) Turn plasticity distinguishes different modes of amyloid-β aggregation. J Am Chem Soc. 136, 4913–4919
24. Abelein, A., Abrahams, J. P., Danielsson, J., Gräslund, A., Jarvet, J., Luo, J., Tiiman, A., and Wärmländer, S. K. T. S. (2014) The hairpin conformation of the amyloid β peptide is an important structural motif along the aggregation pathway. J Biol Inorg Chem. 19, 623–34
25. Nilsberth, C., Westlind-Danielsson, A., Eckman, C. B., Condron, M. M., Axelman, K., Forsell, C., Stenh, C., Luthman, J., Teplow, D. B., Younkin, S. G., Näslund, J., and Lannfelt, L. (2001) The "Arctic" APP mutation (E693G) causes Alzheimer's disease by enhanced Aβ protofibril formation. Nature Neuroscience 2001 4:9. 4, 887–893
26. Rossi, G., Macchi, G., Porro, M., Giaccone, G., Bugiani, M., Scarpini, E., Scarlato, G., Molini, G. E., Sasanelli, F., Bugiani, O., and Tagliavini, F. (1998) Fatal familial insomnia. Neurology. 50, 688–692
27. Levy, E., Carman, M. D., Fernandez-Madrid, I. J., Power, M. D., Lieberburg, I., van Duinen, S. G., Bots, G. T. A. M., Luyendijk, W., and Frangione, B. (1990) Mutation of the Alzheimer's Disease Amyloid Gene in Hereditary Cerebral Hemorrhage, Dutch Type. Science (1979). 248, 1124–1126
28. Hendriks, L., van Duijn, C. M., Cras, P., Cruts, M., van Hul, W., van Harskamp, F., Warren, A., McInnis, M. G., Antonarakis, S. E., Martin, J. J., Hofman, A., and van Broeckhoven, C. (1992) Presenile dementia and cerebral haemorrhage linked to a mutation at codon 692 of the β–amyloid precursor protein gene. Nature Genetics 1992 1:3. 1, 218–221
29. Swanson, C. J., Zhang, Y., Dhadda, S., Wang, J., Kaplow, J., Lai, R. Y. K., Lannfelt, L., Bradley, H., Rabe, M., Koyama, A., Reyderman, L., Berry, D. A., Berry, S., Gordon, R., Kramer, L. D., and Cummings, J. L. (2021) A randomized, double-blind, phase 2b proof-of-concept clinical trial in early Alzheimer's disease with lecanemab, an anti-Aβ protofibril antibody. Alzheimers Res Ther. 13, 1–14
30. Sandberg, A., Luheshi, L. M., Sollvander, S., Pereira de Barros, T., Macao, B., Knowles, T. P. J., Biverstal, H., Lendel, C., Ekholm-Petterson, F., Dubnovitsky, A., Lannfelt, L., Dobson, C. M., and Hard, T. (2010) Stabilization of neurotoxic Alzheimer amyloid- oligomers by protein engineering. Proceedings of the National Academy of Sciences. 107, 15595–15600
31. Sandberg, A., Rodriguez, R. C., Kettunen, P., Rosengren, M., Correa, E. B., Morrema, T., Hoozemans, J. J. M., and Scheper, W. (2020) Specific targeting of a highly toxic subpopulation of Aβ42 oligomers for the treatment of Alzheimer's disease. Alzheimer's & Dementia. 16, e043003
32. Scarff, C. A., Ashcroft, A. E., and Radford, S. E. (2015) Characterization of amyloid oligomers by electrospray ionization-ion mobility spectrometry-mass spectrometry (ESI-IMS-MS). in Methods in Molecular Biology, pp. 115–132, 1345, 115–132
33. Marklund, E. G., and Benesch, J. L. (2019) Weighing-up protein dynamics: the combination of native mass spectrometry and molecular dynamics simulations. Curr Opin Struct Biol. 54, 50–58
34. Bernstein, S. L., Dupuis, N. F., Lazo, N. D., Wyttenbach, T., Condron, M. M., Bitan, G., Teplow, D. B., Shea, J. E., Ruotolo, B. T., Robinson, C. v., and Bowers, M. T. (2009) Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. Nat Chem. 1, 326–331

35. Salbo, R., Bush, M. F., Naver, H., Campuzano, I., Robinson, C. v., Pettersson, I., Jørgensen, T. J. D., and Haselmann, K. F. (2012) Traveling-wave ion mobility mass spectrometry of protein complexes: Accurate calibrated collision cross-sections of human insulin oligomers. Rapid Communications in Mass Spectrometry. 26, 1181–1193
36. Król, S., Österlund, N., Vosough, F., Jarvet, J., Wärmländer, S., Barth, A., Ilag, L. L., Magzoub, M., Gräslund, A., and Mörman, C. (2021) The amyloid-inhibiting NCAM-PrP peptide targets Aβ peptide aggregation in membrane-mimetic environments. iScience. 10.1016/j.isci.2021.102852
37. Ciudad, S., Puig, E., Botzanowski, T., Meigooni, M., Arango, A. S., Do, J., Mayzel, M., Bayoumi, M., Chaignepain, S., Maglia, G., Cianferani, S., Orekhov, V., Tajkhorshid, E., Bardiaux, B., and Carulla, N. (2020) Aβ(1-42) tetramer and octamer structures reveal edge conductivity pores as a mechanism for membrane damage. Nat Commun. 11, 3014
38. Benesch, J. L. P., Ruotolo, B. T., Simmons, D. A., and Robinsons, C. v. (2007) Protein complexes in the gas phase: Technology for structural genomics and proteomics. Chem Rev. 107, 3544–3567
39. Hoyer, W., Gronwall, C., Jonsson, A., Stahl, S., and Hard, T. (2008) Stabilization of a -hairpin in monomeric Alzheimer's amyloid- peptide inhibits amyloid formation. Proceedings of the National Academy of Sciences. 105, 5099–5104
40. Barz, B., Liao, Q., and Strodel, B. (2018) Pathways of Amyloid-β Aggregation Depend on Oligomer Shape. J Am Chem Soc. 140, 319–327
41. Bernstein, S. L., Wyttenbach, T., Baumketner, A., Shea, J. E., Bitan, G., Teplow, D. B., and Bowers, M. T. (2005) Amyloid β-protein: Monomer structure and early aggregation states of Aβ42 and its Pro19 alloform. J Am Chem Soc. 127, 2075–2084
42. Bleiholder, C., Dupuis, N. F., Wyttenbach, T., and Bowers, M. T. (2010) Ion mobility–mass spectrometry reveals a conformational conversion from random assembly to β-sheet in amyloid fibril formation. Nature Chemistry 2010 3:2. 3, 172–177
43. Breuker, K., and McLafferty, F. W. (2008) Stepwise evolution of protein native structure with electrospray into the gas phase, 10-12 to 102 s. Proceedings of the National Academy of Sciences. 105, 18145–18152
44. Hoyer, W., Gronwall, C., Jonsson, A., Stahl, S., and Hard, T. (2008) Stabilization of a beta-hairpin in monomeric Alzheimer's amyloid-beta peptide inhibits amyloid formation. Proc Natl Acad Sci U S A. 105, 5099–5104
45. Seo, J., Hoffmann, W., Warnke, S., Huang, X., Gewinner, S., Schöllkopf, W., Bowers, M. T., von Helden, G., and Pagel, K. (2017) An infrared spectroscopy approach to follow β-sheet formation in peptide amyloid assemblies. Nat Chem. 9, 39–44
46. Braun, G. A., Dear, A. J., Sanagavarapu, K., Zetterberg, H., and Linse, S. (2022) Amyloid-b peptide 37, 38 and 40 individually and cooperatively inhibit amyloid-b 42 aggregation †. 10.1039/d1sc02990h
47. Roche, J., Shen, Y., Lee, J. H., Ying, J., and Bax, A. (2016) Monomeric Aβ1-40 and Aβ1-42 Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil. Biochemistry. 55, 762–775
48. Ono, K., Condron, M. M., and Teplow, D. B. (2009) Structure-neurotoxicity relationships of amyloid β-protein oligomers. Proc Natl Acad Sci U S A. 106, 14745–14750
49. Österlund, N., Moons, R., Ilag, L. L., Sobott, F., and Graslund, A. (2019) Native ion mobility-mass spectrometry reveals the formation of β-barrel shaped amyloid-β hexamers in a membrane-mimicking environment. J Am Chem Soc. 141, 10440–10450
50. Österlund, N., Kulkarni, Y. S., Misiaszek, A. D., Wallin, C., Krüger, D. M., Liao, Q., Mashayekhy Rad, F., Jarvet, J., Strodel, B., Wärmländer, S. K. T. S., Ilag, L. L., Kamerlin, S. C. L., and Gräslund, A. (2018) Amyloid-β Peptide Interactions with Amphiphilic Surfactants: Electrostatic and Hydrophobic Effects. ACS Chem Neurosci. 9, 1680–1692
51. Yang, Y., Arseni, D., Zhang, W., Huang, M., Lövestam, S., Schweighauser, M., Kotecha, A., Murzin, A. G., Peak-Chew, S. Y., MacDonald, J., Lavenir, I., Garringer, H. J., Gelpi, E., Newell, K. L., Kovacs, G. G., Vidal, R., Ghetti, B., Ryskeldi-Falco, B., Scheres, S. H. W., and Goedert, M. (2022) Cryo-EM structures of amyloid-b 42 filaments from human brains. Science (1979). 375, 167–172
52. Kayed, R., and Glabe, C. G. (2006) Conformation-Dependent Anti-Amyloid Oligomer Antibodies. Methods Enzymol. 413, 326–344
53. Kayed, R., Head, E., Sarsoza, F., Saing, T., Cotman, C. W., Necula, M., Margol, L., Wu, J., Breydo, L., Thompson, J. L., Rasool, S., Gurlo, T., Butler, P., and Glabe, C. G. (2007) Fibril specific, conformation dependent antibodies recognize a generic epitope common to amyloid fibrils and fibrillar oligomers that is absent in prefibrillar oligomers. Mol Neurodegener. 2, 1–11

54. Zanjani, A. A. H., Reynolds, N. P., Zhang, A., Schilling, T., Mezzenga, R., and Berryman, J. T. (2020) Amyloid Evolution: Antiparallel Replaced by Parallel. Biophys J. 118, 2526–2536
55. Cerf, E., Sarroukh, R., Tamamizu-Kato, S., Breydo, L., Derclayes, S., Dufrênes, Y. F., Narayanaswami, V., Goormaghtigh, E., Ruysschaert, J. M., and Raussens, V. (2009) Antiparallel β-sheet: a signature structure of the oligomeric amyloid β-peptide. Biochemical Journal. 421, 415–423
56. Baronio, C. M., Baldassarre, M., and Barth, A. (2019) Insight into the internal structure of amyloid-b oligomers by isotope-edited Fourier transform infrared spectroscopy †. Phys. Chem. Chem. Phys. 21, 8587
57. Gallardo, R., Ranson, N. A., and Radford, S. E. (2020) Amyloid structures: much more than just a cross-β fold. Curr Opin Struct Biol. 60, 7–16
58. Ghosh, P., Vaidya, A., Kumar, A., and Rangachari, V. (2016) Determination of critical nucleation number for a single nucleation amyloid-β aggregation model. Math Biosci. 273, 70–79
59. Shoghi-Jadid, K., Barrio, J. R., Kepe, V., Wu, H. M., Small, G. W., Phelps, M. E., and Huang, S. C. (2005) Imaging β-amyloid fibrils in Alzheimer's disease: A critical analysis through simulation of amyloid fibril polymerization. Nucl Med Biol. 32, 337–351
60. Šarić, A., Chebaro, Y. C., Knowles, T. P. J., and Frenkel, D. (2014) Crucial role of nonspecific interactions in amyloid nucleation. Proc Natl Acad Sci U S A. 111, 17869–17874
61. Leppert, A., Tiiman, A., Kronqvist, N., Landreh, M., Abelein, A., Vukojević, V., and Johansson, J. (2021) Smallest Secondary Nucleation Competent Aβ Aggregates Probed by an ATP-Independent Molecular Chaperone Domain. Biochemistry. 60, 678–688
62. Glabe, C. G., and Kayed, R. (2006) Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis. Neurology. 66, S74–S78
63. Demuro, A., Mina, E., Kayed, R., Milton, S. C., Parker, I., and Glabe, C. G. (2005) Calcium Dysregulation and Membrane Disruption as a Ubiquitous Neurotoxic Mechanism of Soluble Amyloid Oligomers. Journal of Biological Chemistry. 280, 17294–17300
64. Sciacca, M. F. M., Kotler, S. A., Brender, J. R., Chen, J., Lee, D. K., and Ramamoorthy, A. (2012) Two-step mechanism of membrane disruption by Aβ through membrane fragmentation and pore formation. Biophys J. 103, 702–710
65. Martin Vabulas, R., and Ulrich Hartl, F. (2011) Aberrant protein interactions in amyloid disease. Cell Cycle. 10, 1512–1513
66. Radwan, M., Wood, R. J., Sui, X., and Hatters, D. M. (2017) When proteostasis goes bad: Protein aggregation in the cell. IUBMB Life. 69, 49–54
67. Österlund, N., Vosselman, T., Leppert, A., Gräslund, A., Jörnvall, H., Ilag, L. L., Marklund, E. G., Elofsson, A., Johansson, J., Sahin, C., and Landreh, M. (2022) Mass spectrometry and machine learning reveal determinants of client recognition by anti-amyloid chaperones. Molecular & Cellular Proteomics. 21, 100413
68. Månsson, C., Arosio, P., Hussein, R., Kampinga, H. H., Hashem, R. M., Boelens, W. C., Dobson, C. M., Knowles, T. P. J., Linse, S., and Emanuelsson, C. (2014) Interaction of the molecular chaperone DNAJB6 with growing amyloid-beta 42 (Aβ42) aggregates leads to sub-stoichiometric inhibition of amyloid formation. Journal of Biological Chemistry. 289, 31066–31076
69. Österlund, N., Lundqvist, M., Ilag, L. L., Gräslund, A., and Emanuelsson, C. (2020) Amyloid-β oligomers are captured by the DNAJB6 chaperone: Direct detection of interactions that can prevent primary nucleation. Journal of Biological Chemistry. 295, 8135–8144
70. Henning-Knechtel, A., Kumar, S., Wallin, C., Król, S., Wärmländer, S. K. T. S., Jarvet, J., Esposito, G., Kirmizialtin, S., Gräslund, A., Hamilton, A. D., and Magzoub, M. (2020) Designed Cell-Penetrating Peptide Inhibitors of Amyloid-beta Aggregation and Cytotoxicity. Cell Rep Phys Sci. 1, 100014
71. Stark, T., Lieblein, T., Pohland, M., Kalden, E., Freund, P., Zangl, R., Grewal, R., Heilemann, M., Eckert, G. P., Morgner, N., and Göbel, M. W. (2017) Peptidomimetics That Inhibit and Partially Reverse the Aggregation of Aβ1-42. Biochemistry. 56, 4840–4849
72. Tjernberg, L. O., Näslundt, J., Lindqvist, F., Johansson, J., Karlström, A. R., Thyberg, J., Tereniust, L., and Nordstedt, C. (1996) Arrest of -Amyloid Fibril Formation by a Pentapeptide Ligand (∗). Journal of Biological Chemistry. 271, 8545–8548
73. Arai, T., Sasaki, D., Araya, T., Sato, T., Sohma, Y., and Kanai, M. (2014) A Cyclic KLVFF-Derived Peptide Aggregation Inhibitor Induces the Formation of Less-Toxic Off-Pathway Amyloid-β Oligomers. ChemBioChem. 15, 2577–2583

74. Chafekar, S. M., Malda, H., Merkx, M., Meijer, E. W., Viertl, D., Lashuel, H. A., Baas, F., and Scheper, W. (2007) Branched KLVFF Tetramers Strongly Potentiate Inhibition of β-Amyloid Aggregation. ChemBioChem. 8, 1857–1864
75. Ståhl, S., Gräslund, T., Eriksson Karlström, A., Frejd, F. Y., Nygren, P. Å., and Löfblom, J. (2017) Affibody Molecules in Biotechnological and Medical Applications. Trends Biotechnol. 35, 691–712
76. Boutajangout, A., Lindberg, H., Awwad, A., Paul, A., Baitalmal, R., Almokyad, I., Höidén-Guthenberg, I., Gunneriusson, E., Frejd, F. Y., Härd, T., Löfblom, J., Ståhl, S., and Wisniewski, T. (2019) Affibody-Mediated Sequestration of Amyloid β Demonstrates Preventive Efficacy in a Transgenic Alzheimer's Disease Mouse Model. Front Aging Neurosci. 11, 64
77. Ylera, F., Lurz, R., Erdmann, V. A., and Fürste, J. P. (2002) Selection of RNA Aptamers to the Alzheimer's Disease Amyloid Peptide. Biochem Biophys Res Commun. 290, 1583–1588
78. Young, L. M., Saunders, J. C., Mahood, R. A., Revill, C. H., Foster, R. J., Tu, L. H., Raleigh, D. P., Radford, S. E., and Ashcroft, A. E. (2014) Screening and classifying small-molecule inhibitors of amyloid formation using ion mobility spectrometry–mass spectrometry. Nature Chemistry 2014 7:1. 7, 73–81
79. Ruotolo, B. T., Benesch, J. L. P., Sandercock, A. M., Hyung, S. J., and Robinson, C. v. (2008) Ion mobility-mass spectrometry analysis of large protein complexes. Nat Protoc. 3, 1139–1152
80. Bush, M. F., Hall, Z., Giles, K., Hoyes, J., Robinson, C. v., and Ruotolo, B. T. (2010) Collision cross sections of proteins and their complexes: A calibration framework and database for gas-phase structural biology. Anal Chem. 82, 9557–9565
81. Jo, S., Kim, T., Iyer, V. G., and Im, W. (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem. 29, 1859–1865
82. Schrödinger, L. L. C. (2015) The PyMOL Molecular Graphics System, Version 1.8
83. Fiser, A., and Sali, A. (2003) ModLoop: automated modeling of loops in protein structures. Bioinformatics. 19, 2500–2501
84. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. Nature. 596, 583–589
85. Zhang, J. (2015) The Hybrid Idea of (Energy Minimization) Optimization Methods Applied to Study PrionProtein Structures Focusing on the beta2-alpha2 Loop. Biochem Pharmacol (Los Angel). 10.4172/2167-0501.1000175
86. Parrinello, M., and Rahman, A. (1998) Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys. 52, 7182
87. Parrinello, M., and Rahman, A. (1998) Strain fluctuations and elastic constants. J Chem Phys. 76, 2662
88. Bussi, G., Donadio, D., and Parrinello, M. (2007) Canonical sampling through velocity rescaling. J Chem Phys. 126, 014101
89. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1998) A smooth particle mesh Ewald method. J Chem Phys. 103, 8577
90. van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005) GROMACS: Fast, flexible, and free. J Comput Chem. 26, 1701–1718
91. Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindah, E. (2015) GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX. 1–2, 19–25
92. Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., de Groot, B. L., Grubmüller, H., and MacKerell, A. D. (2016) CHARMM36m: an improved force field for folded and intrinsically disordered proteins. Nature Methods 2016 14:1. 14, 71–73
93. Cole, C., Barber, J. D., and Barton, G. J. (2008) The Jpred 3 secondary structure prediction server. Nucleic Acids Res. 36, W197–W201
94. Zirah, S., Kozin, S. A., Mazur, A. K., Blond, A., Cheminant, M., Ségalas-Milazzo, I., Debey, P., and Rebuffat, S. (2006) Structural changes of region 1-16 of the Alzheimer disease amyloid beta-peptide upon zinc binding and in vitro aging. J Biol Chem. 281, 2151–61

Supporting Information

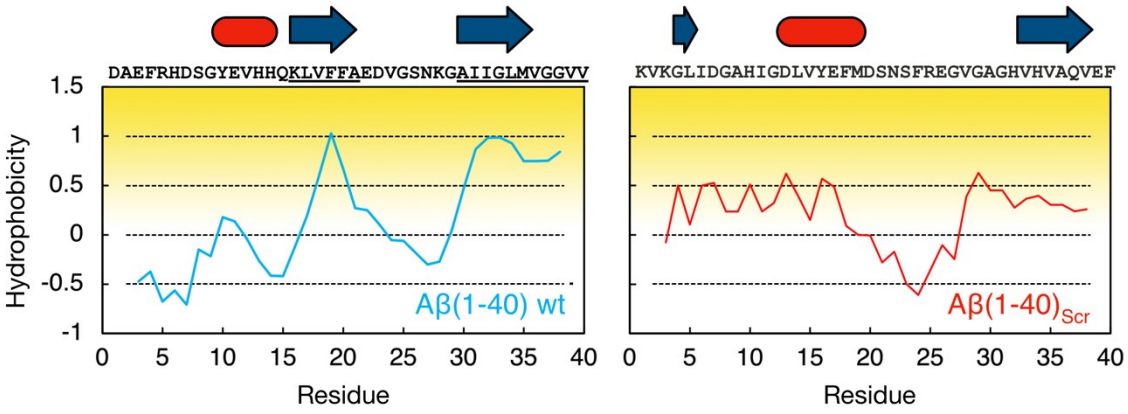

Figure S1. Overview of peptide sequences, Aβ(1-40) (left) and Aβ(1-40)$_{Scr}$ (right). Structure predictions by jpred (93) are shown on the top. Sheets are represented by blue arrows, helices by red ovals. Calculated hydrophobicity by the Eisenberg method is plotted against the sequence, using an averaging window of 5 residues.

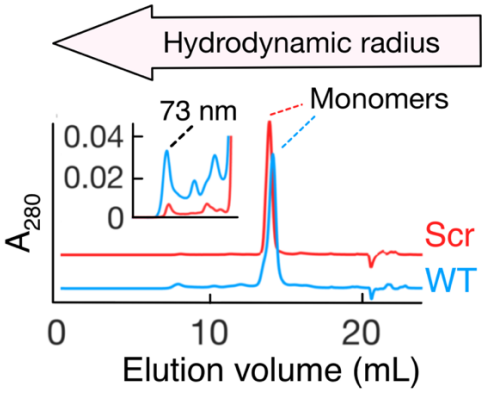

Figure S2. Size exclusion chromatography of Aβ(1-40) (blue) and Aβ(1-40)$_{Scr}$ (red) performed on a Superdex 75 10/300 Increase (Cytiva) column. The insert shows a magnification of aggregates in the chromatogram.

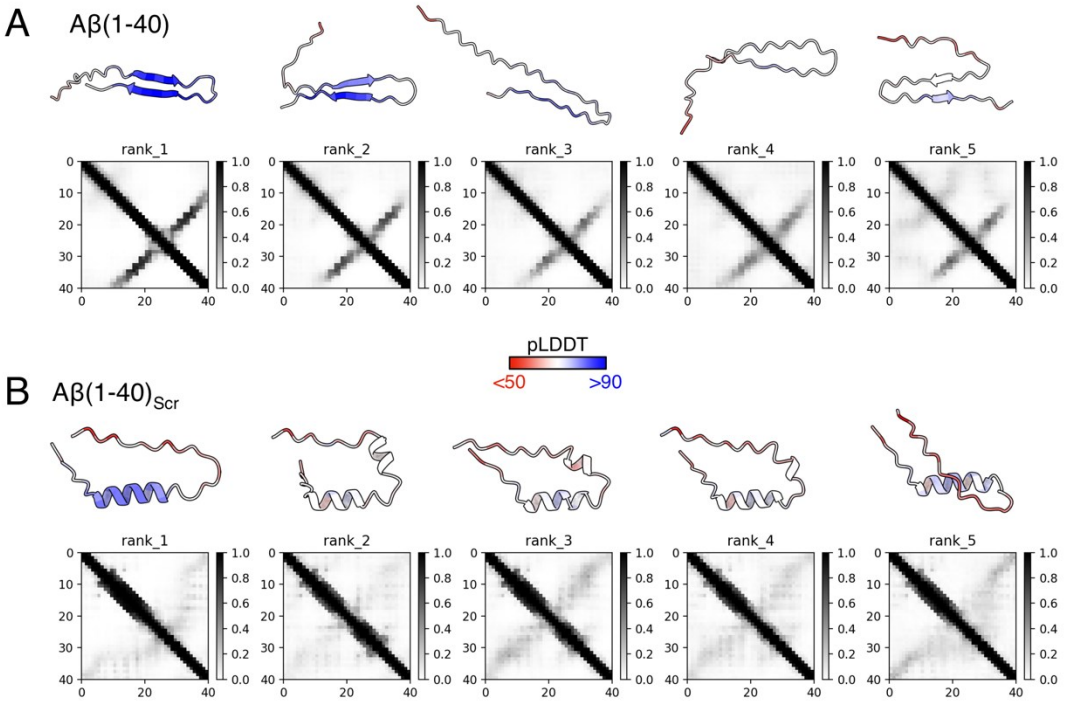

Figure S3. Predicted AlphaFold2 Models for A) Aβ(1-40) and B) Aβ(1-40)$_{Scr}$. The generated models are colored according to their pLDDT-score (red to blue). Scores >90 (blue) are expected to be modeled with high accuracy. Scores between 70-90 (white to blue) are expected to be modeled well. Intramolecular contact maps are shown below each structure model.

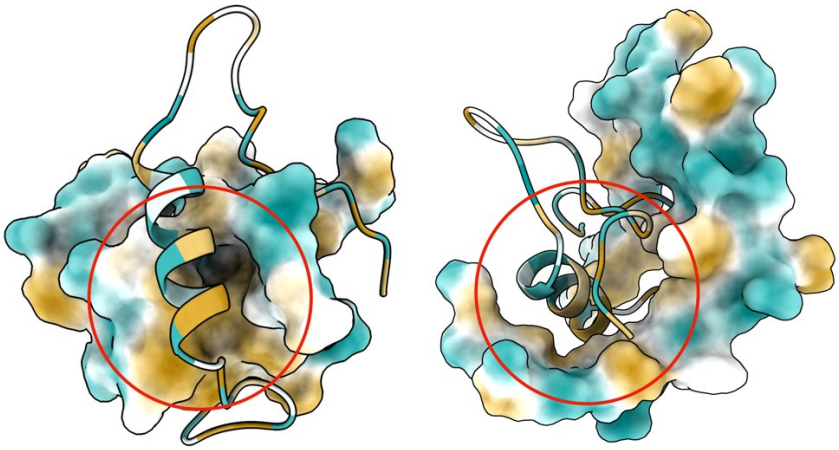

Figure S4. Top cluster structure of Aβ(1-40)$_{Scr}$ dimers obtained from MD simulation, colored according to hydrophobicity (orange = hydrophobic, blue = hydrophilic). One peptide in the dimer is shown as cartoon and the other one is shown as surface, illustrating the binding between the hydrophobic helix face and hydrophobic patches.

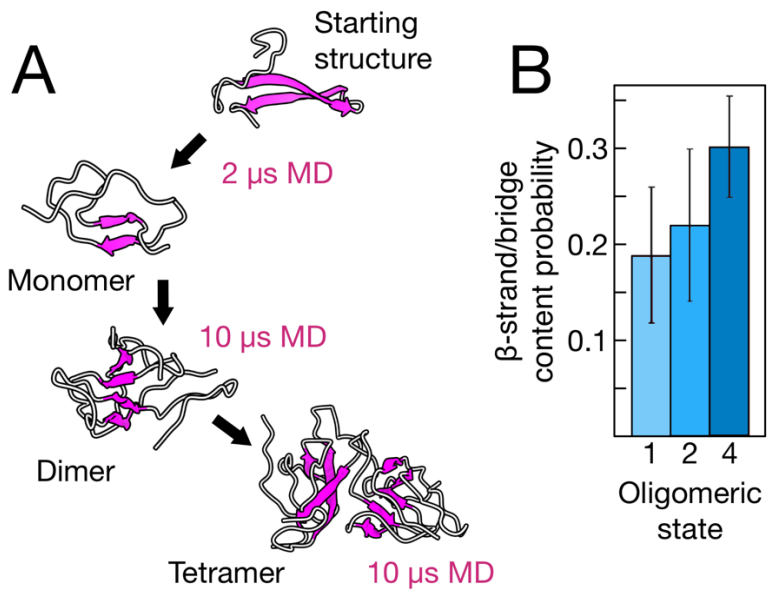

Figure S5. A) Structural development of Aβ(1-40) monomer and oligomers in MD simulations. The starting structure is taken from the NMR-derived structure of Aβ(1-40) in an affibody complex (pdb 2otk). Sheets are colored magenta. B) β-strand/bridge propensity (average plus standard deviation) for Aβ(1-40) species calculated from the MD simulations.

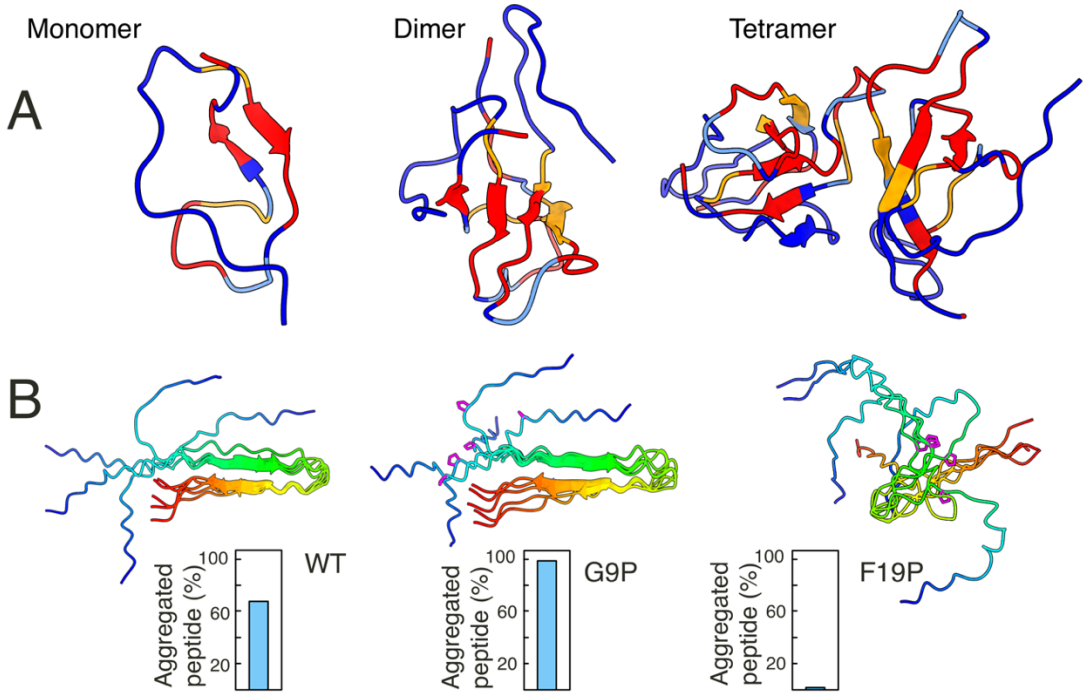

Figure S6. A) Top cluster structures from MD simulations of Aβ(1-40) colored according to the effect of proline substitutions in reference (22) (orange-red represents inhibition of aggregation, blue represents no effect on aggregation). B) AlpaFold2 models of wild-type Aβ(1-40) and two proline substitutions. The G9P substitution does not affect the hairpin fold and does not inhibit aggregation. The F19P substitution breaks the hairpin and inhibits aggregation. Data on aggregation propensity was collected from reference (22).

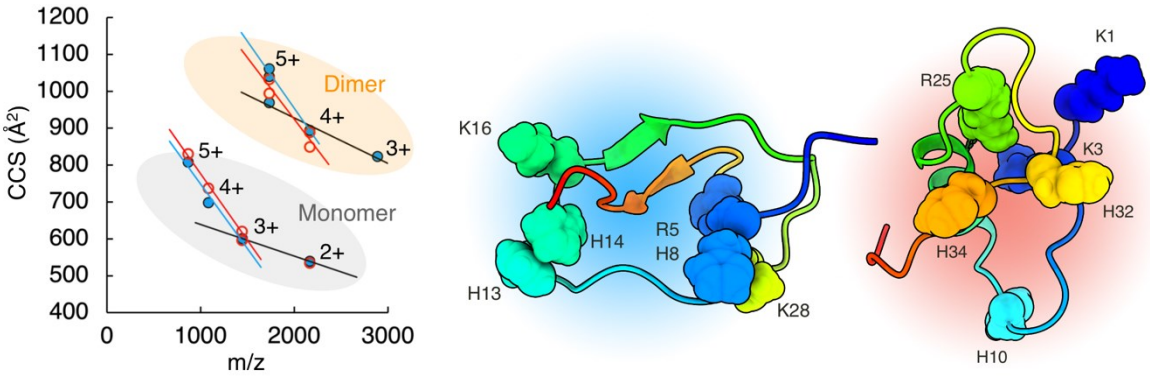

Figure S7. (Left) Experimental CCS values of Aβ(1-40) (blue) and Aβ(1-40)$_{Scr}$ (red) for different charge states. Solid lines represent linear fits to the experimental data. Higher charge states display a steep increase in CCS with charge, likely due to intrachain repulsion in the gas phase. (Right) Positively charged residues in the two peptide variants are shown as spheres in the MD derived structures.

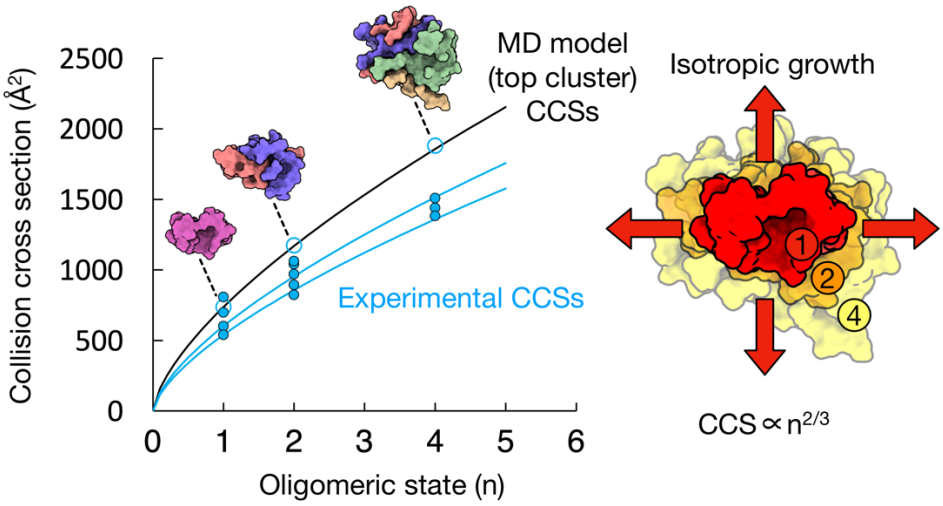

Figure S8. Calculated CCS values for the top clusters of Aβ(1-40) generated from MD simulations are shown as open blue circles. The theoretical isotropic growth originating from the MD monomer is shown as a black solid line, indicating that the MD models grow isotopically (spherically) with respect to oligomeric state. Experimental CCS values are shown with solid blue circles for comparison.

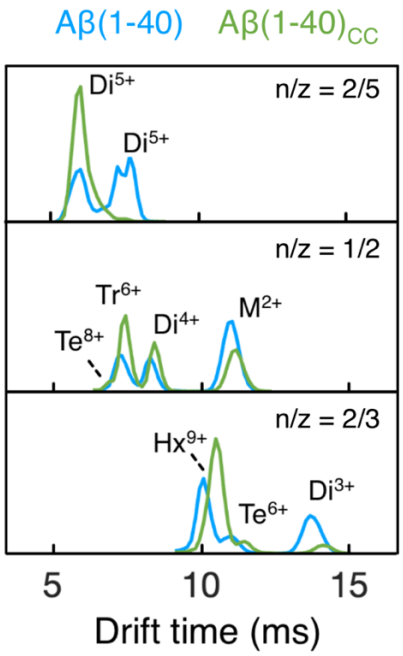

Figure S9. Example experimental drift time profiles for some n/z signals from Aβ(1-40) (blue) and Aβ(1-40)$_{CC}$ (green). The profiles are normalized with respect to their total signal intensity to compare the relative abundances within each n/z signal between the two variants.

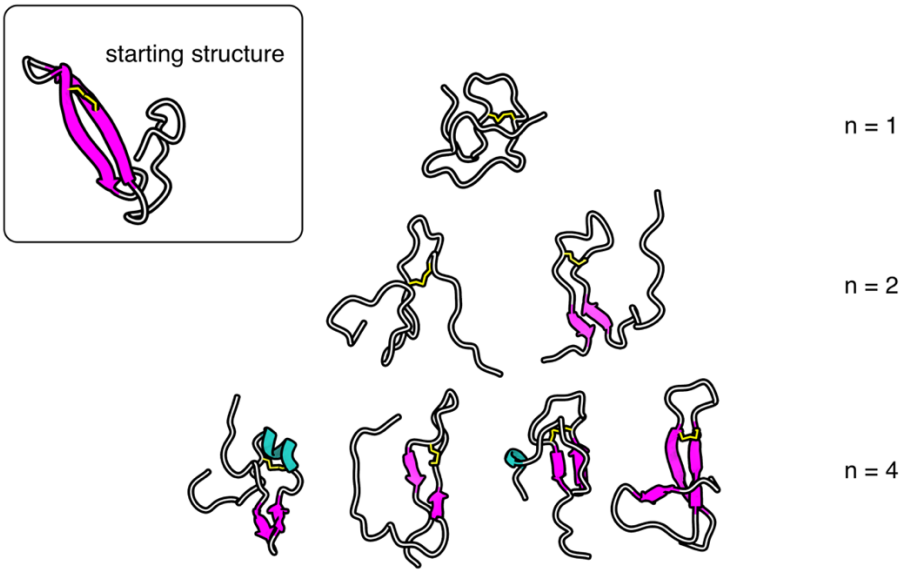

Figure S10. Structures of monomeric units of Aβ(1-40)$_{CC}$ in the top clusters of oligomers obtained from MD simulations, illustrating the increase in β-sheet propensity upon oligomer growth. The starting structure for the MD simulation (derived from pdb 2otk) of the monomer is also shown for comparison. The structures are colored according to secondary structure, gray for coil, magenta for sheets and cyan for helices. The disulfide bond is to shown in yellow.

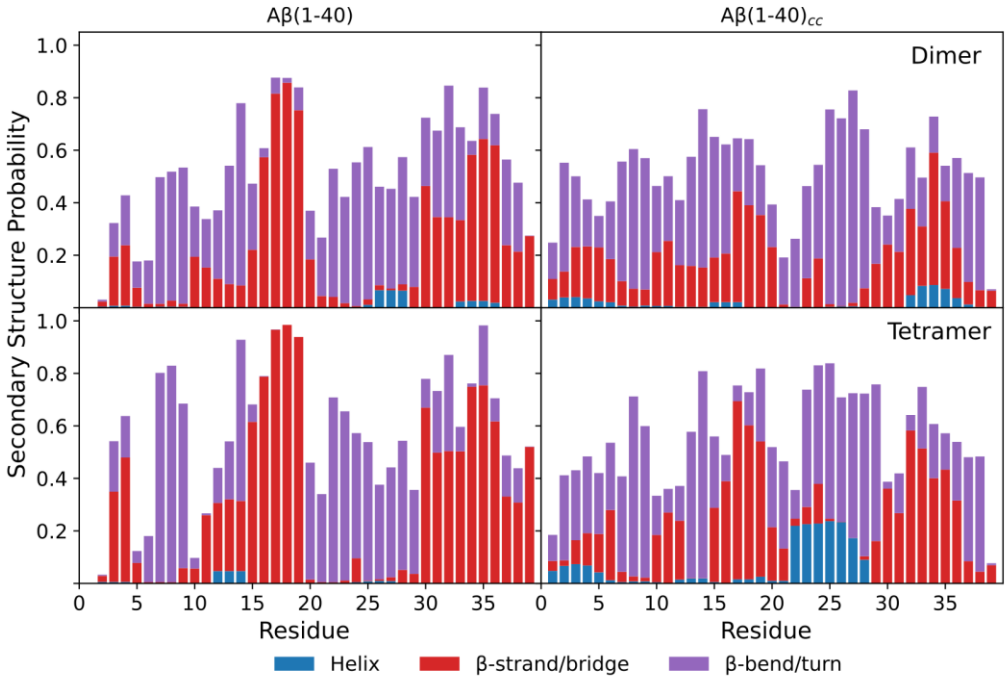

Figure S11. Probability of secondary structure of the residues of Aβ(1-40) (left) and Aβ(1-40)$_{CC}$ (right) for dimers (top) and tetramers (bottom). The bars represent the average additive secondary structure probabilities consisting of helix (blue), β-strand/bridge (red), and turn or bend (violet). The difference to 1.0 (shown in white) presents the random coil state.

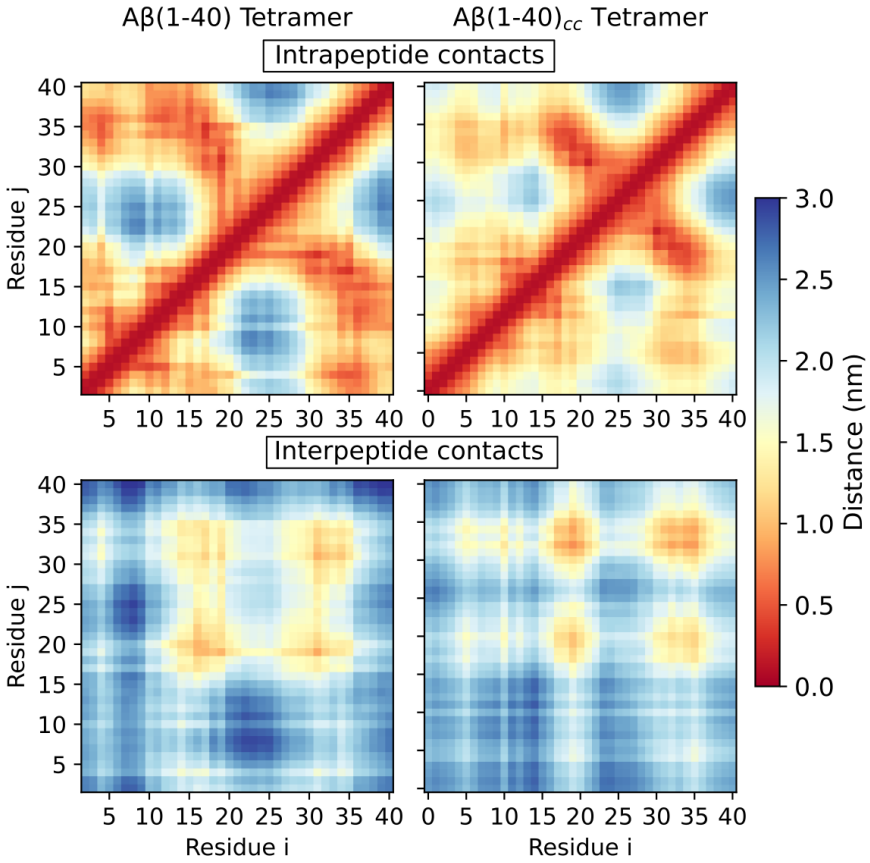

Figure S12. The average intra- (top) and interpeptide (bottom) contacts, shown as distances between residues for tetramers of Aβ(1-40) (left) and Aβ(1-40)$_{CC}$ (right), obtained from MD simulations. The intrapeptide contacts were averaged over the four peptides in the tetramers and the results are shown as symmetric matrices. For the interpeptide contacts, the distances were averaged over all possible peptide pairings. The resulting matrices are almost symmetric as in converged simulations one should obtain d(i,j') = d(i',j) where the prime indicates that the two residues i and j' belong to different peptides. The color bar on the right shows the average intra-/inter-residue distance (in nm).

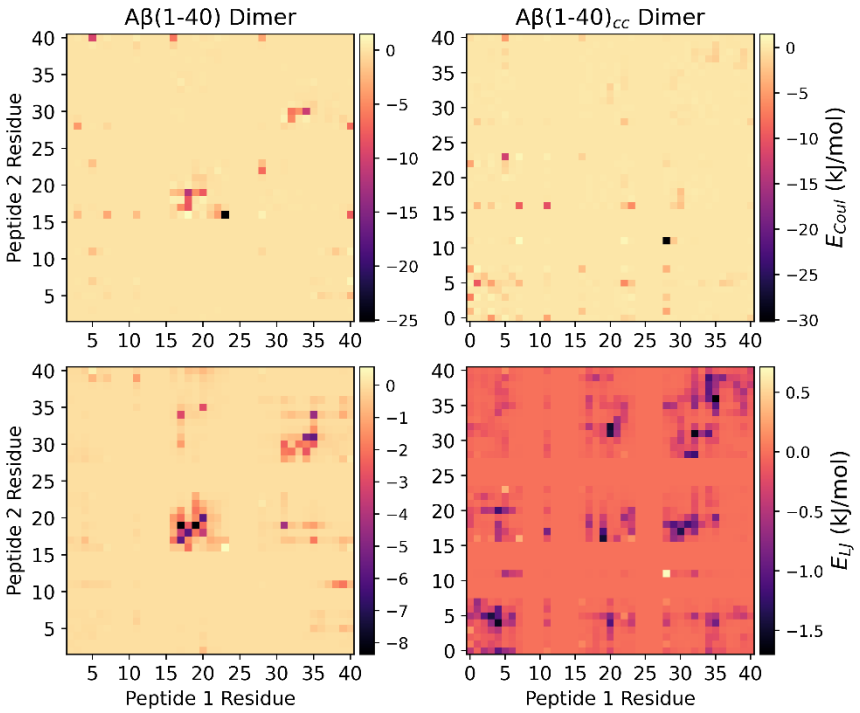

Figure S13. Average interpeptide interaction energies between residues, decomposed into Coulomb (top) and Lennard-Jones (bottom) interactions, for Aβ(1-40) (left) and Aβ(1-40)$_{CC}$ dimers (right). Please note that energies are only calculated for relevant residue numbers: 1-7, 11, 16-23 and 28-40. Moreover, the energy scales in the four plots are different from each other (see color bars) in order to provide a good resolution of the energies.

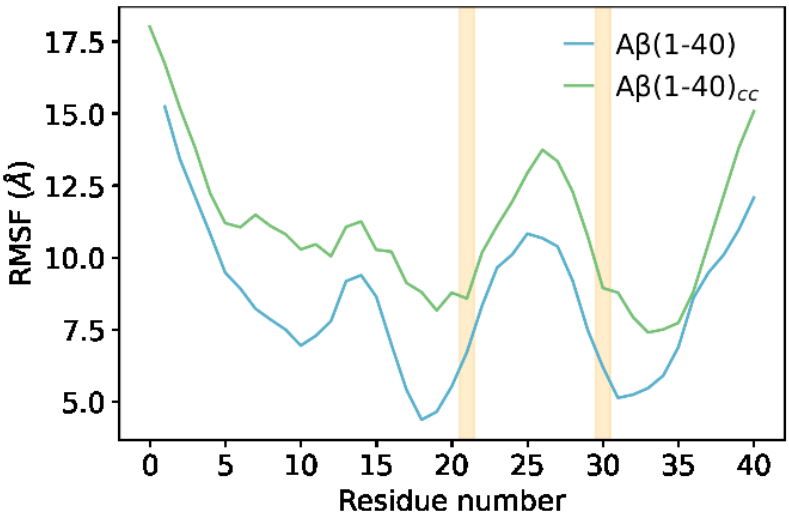

Figure S14. The root means square fluctuations (RMSF) of the C_α atoms, averaged over both peptides of the Aβ(1-40) dimer (blue) and the Aβ(1-40)$_{CC}$ dimer (green). The RMSF values were calculated after alignment of the trajectory in question and using the average structure as reference. The residues 21 and 30 are highlighted by orange bars.

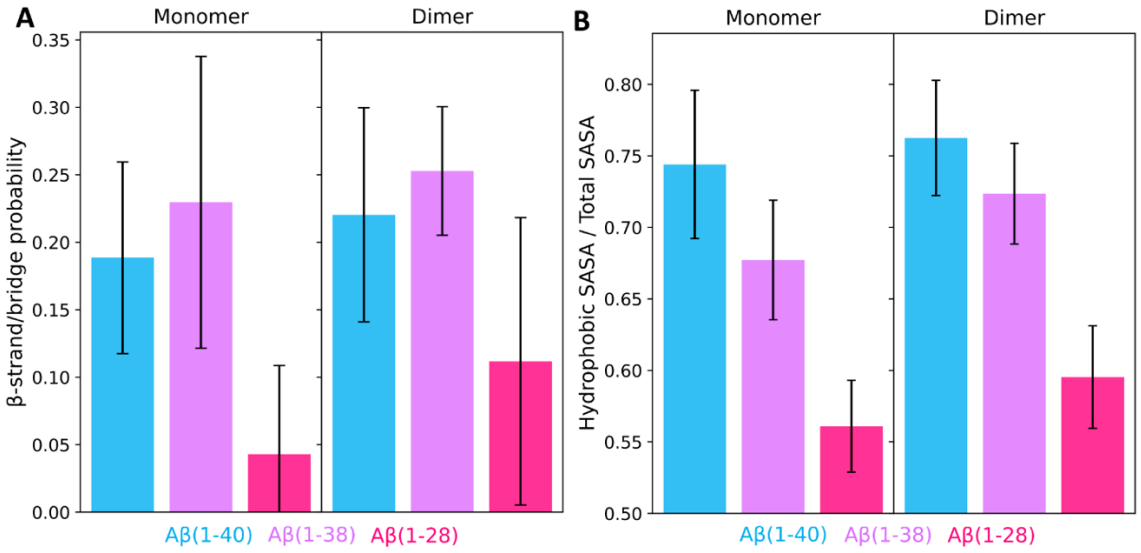

Figure S15. A) Average β-strand/bridge probability for Aβ(1-40), Aβ(1-38) and Aβ(1-28) monomers and dimers. B) The ratio between the average hydrophobic SASA and the average total SASA for Aβ(1-40), Aβ(1-38) and Aβ(1-28) monomers and dimers.

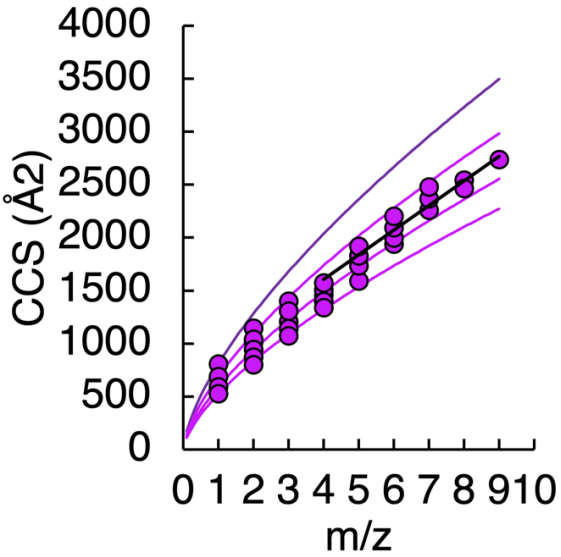

Figure S16. Measured collision cross sections (CCS) for each oligomeric state of Aβ(1-38). Solid purple lines represent the theoretical isotropic growth originating from each monomer state. A black solid line represents a linear fit to the experimental values for larger oligomers that deviate from isotropic growth.

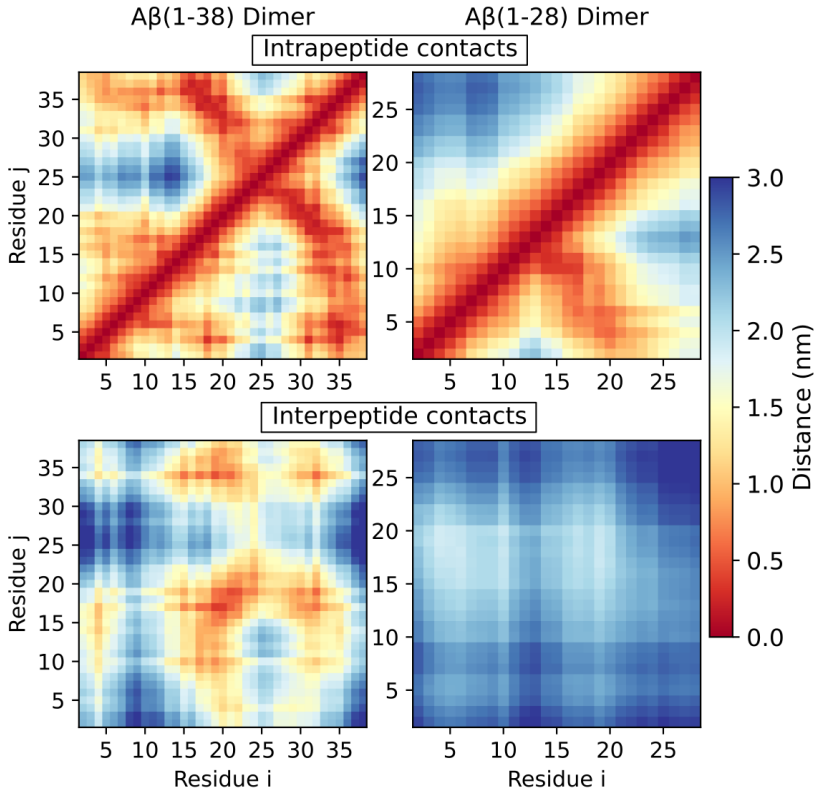

Figure S17. The average intra- (top) and interpeptide (bottom) contacts, shown as distances between residues for dimers of Aβ(1-38) (left) and Aβ(1-28) (right), as obtained from MD simulations. The intrapeptide contacts are shown per peptide composing the dimers in the upper left and lower right triangle of the maps. The interpeptide contacts are shown as distances d(i,j') where the prime indicates that the two residues i and j' belong to different peptides. The resulting matrices are almost symmetric as in converged MD simulations one should obtain d(i,j') = d(i',j). The color bar on the right shows the average intra-/inter-residue distance (in nm).

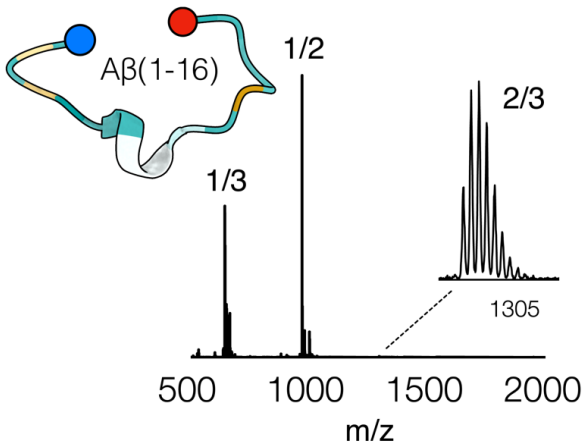

Figure S18. Mass spectrum of Aβ(1-16) with peaks annotated by their oligomeric state/charge state. The NMR structure of Aβ(1-16) (pdb 1ze7 (94)) is shown and is colored according to hydrophobicity (orange = hydrophobic, blue = hydrophilic). The N-terminus is highlighted in blue, while the C-terminus is highlighted in red.

A.3. Publication II

Amyloid-β peptide dimers undergo a random coil to β-sheet transition in the aqueous phase but not at the neuronal membrane

Fatafta H., **Khaled, M.**, Owen M. C., Sayyed-Ahmad, A., Strodel, B. (2021).

PNAS, 118(39), e2106210118.

Amyloid-β peptide dimers undergo a random coil to β-sheet transition in the aqueous phase but not at the neuronal membrane

Hebah Fatafta[a], Mohammed Khaled[a], Michael C. Owen[b,c], Abdallah Sayyed-Ahmad[d], and Birgit Strodel[a,e,1]

[a]Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich, 52425 Jülich, Germany; [b]Central European Institute of Technology, Masaryk University, Brno 625 00, Czech Republic; [c]Institute of Chemistry, University of Miskolc, 3515 Miskolc-Egyetemváros, Hungary; [d]Department of Physics, Birzeit University, 71939 Birzeit, Palestine; and [e]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

Edited by Gerhard Hummer, Max Planck Institute for Biophysics, Frankfurt am Main, Germany, and accepted by Editorial Board Member Angela M. Gronenborn August 16, 2021 (received for review March 31, 2021)

Mounting evidence suggests that the neuronal cell membrane is the main site of oligomer-mediated neuronal toxicity of amyloid-β peptides in Alzheimer's disease. To gain a detailed understanding of the mutual interference of amyloid-β oligomers and the neuronal membrane, we carried out microseconds of all-atom molecular dynamics (MD) simulations on the dimerization of amyloid-β (Aβ)42 in the aqueous phase and in the presence of a lipid bilayer mimicking the in vivo composition of neuronal membranes. The dimerization in solution is characterized by a random coil to β-sheet transition that seems on pathway to amyloid aggregation, while the interactions with the neuronal membrane decrease the order of the Aβ42 dimer by attenuating its propensity to form a β-sheet structure. The main lipid interaction partners of Aβ42 are the surface-exposed sugar groups of the gangliosides GM1. As the neurotoxic activity of amyloid oligomers increases with oligomer order, these results suggest that GM1 is neuroprotective against Aβ-mediated toxicity.

Alzheimer's disease | amyloid-β | neuronal membrane | molecular dynamics | transition network

In Alzheimer's disease (AD), amyloid-β peptide (Aβ) aggregates into fibrils and subsequently accumulates as plaques within the neural tissue (1). An increasing number of studies suggest that the smaller soluble oligomers formed in the earlier stages of the aggregation process are the main cytotoxic species affecting the severity and progression of AD (2–4). Aβ dimers have been reported to be the smallest toxic oligomer that affects synaptic plasticity and impairs memory (5, 6). Therefore, a detailed characterization of Aβ dimerization is an essential step toward developing a better understanding of the aggregation process. However, its transient nature (resulting from its high aggregation tendency), its plasticity, and its equilibrium with both the monomer and higher-order oligomers all make the Aβ dimer extremely challenging to study experimentally. In fact, a large amount of the experimental studies performed on Aβ dimers employ some kind of cross-linking to stabilize them (7–9). On the other hand, covalently cross-linked Aβ dimers are certainly of biological relevance, as such species have been retrieved from the brains of AD patients and their neurotoxicity has been demonstrated (6, 10). Apart from this, recent technological developments, such as advanced single-molecule fluorescence spectroscopy and imaging, opened the way to characterize amyloid oligomers without the need to stabilize them by cross-linking (11, 12). Molecular dynamics (MD) simulations are also able to provide atomic insight into the temporal evolution of the dimer structure without the need of cross-linking (13, 14). Previous simulations of Aβ dimers were modeled in the aqueous phase only, and thus they lacked essential details from the cellular context. Consideration of the latter is particularly important if one wishes to reveal the mechanism of toxicity that has been shown to rely on direct contact with the lipid membrane of neurons by Aβ oligomers (15, 16).

Many studies have been done to understand the consequences of Aβ–membrane interactions; however, it is extremely difficult to capture these transient interactions with experimental methods. This becomes possible with MD simulations and this problem is addressed in the current work. We use an aggregate of 24 μs of MD simulations to investigate the dimerization of the full-length Aβ42 peptide both in solution and in the presence of a model lipid bilayer including six lipid types to mimic the composition of a neuronal cell membrane (17–19): 38% 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), 24% 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE), 5% 1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (POPS), 20% cholesterol (CHOL), 9% sphingomyelin (SM), and 4% monosialotetrahexosylganglioside (GM1) (Fig. 1*A*). For modeling Aβ we employ Charmm36m, a force field adjusted for intrinsically disordered proteins (IDPs), to model their preference to adopt extended structures. When applied to monomeric Aβ, Charmm36m yields more than 80% of the structures in a random coil and extended

Significance

The aggregation of the amyloid-β peptide (Aβ) into neurotoxic oligomers is central to the development of Alzheimer's disease. One possible source of their toxicity results from interactions of the Aβ oligomers with the neuronal membrane, damaging membrane integrity and thus neurons. However, molecular details of these interactions are unclear. Here, we contrast the dimerization of Aβ in solution and at the neuronal membrane. Our results clearly indicate that the sugar moieties of GM1 sequester Aβ by forming key hydrogen bonds with the peptide, which diverts the configuration of the Aβ dimers away from damaging β-sheet–rich structures. These findings underline the importance of GM1 in Alzheimer's disease progression and provide a nanoscopic basis for its reported neuroprotective effect.

Author contributions: H.F., M.C.O., and B.S. designed research; H.F., M.K., and M.C.O. performed research; H.F., M.K., and M.C.O. analyzed data; H.F., A.S.-A., and B.S. wrote the paper; and A.S.-A. and B.S. supervised the work.

The authors declare no competing interest.

This article is a PNAS Direct Submission. G.H. is a guest editor invited by the Editorial Board.

[1]To whom correspondence may be addressed. Email: b.strodel@fz-juelich.de.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2106210118/-/DCSupplemental.

Published September 20, 2021.

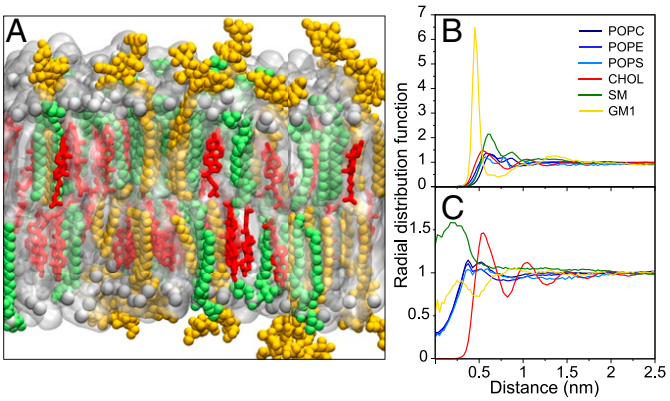

Fig. 1. (*A*) A snapshot of the neuronal membrane containing 38% POPC, 24% POPE, 5% POPS (collectively shown as gray surface with their phosphorous atoms indicated by gray spheres), 20% CHOL (red sticks), 9% SM (green spheres), and 4% GM1 (yellow spheres). In the following, PC, PE, and PS are synonymously used for POPC, POPE, and POPS, respectively. (*B* and *C*) Radial distribution functions for (*B*) lipid pairings of identical type and (*C*) lipid–CHOL pairings. The P atoms of PC, PE, PS, and SM and the O atoms of CHOL and GM1 were used as reference atoms for the RDF calculations. The RDFs are averaged over both membrane leaflets. The *x* axis shows the distances between the respective atom pairs. Since CHOL resides deeper inside the membrane, it is possible that the O atom of CHOL and the reference atoms of the other lipids are above each other, explaining why not all of the RDFs approach zero for $x = 0$. The colors of the functions refer to the lipids as indicated in the color key in *B*. Pairs with RDF > 1 are considered to form clusters.

state, and the remaining ones feature transient β-hairpins, which is in acceptable agreement with experimental data (20). Moreover, Charmm36m outperforms other force fields when it comes to modeling peptide aggregation (21, 22). To the best of our knowledge, this simulation study breaks ground on two fronts: 1) It exceeds the simulation time of previous studies modeling Aβ–membrane interactions by an order of magnitude, and 2) it studies the aggregation of Aβ on a bilayer containing more than three different lipid types. Lipid bilayers of a complexity comparable to the one modeled here have been thus far studied only at the coarse-grained level (23, 24). We also analyze the aggregation pathways by transition networks (25–27), which elucidate the similarities and differences between Aβ dimerization steps both in solution and at the neuronal membrane. We find that the neuronal membrane reduces the dynamics of membrane-bound Aβ42 while it also inhibits β-sheet formation. Here, the sugar groups of GM1 form hydrogen bonds with the peptide, thereby reducing the possibilities for other hydrogen bonds to otherwise form. In contrast, the dimerization in the aqueous phase is characterized by a random coil to β-sheet transition, leading to β-sheet structures similar to the ones found in Aβ fibrils.

Results

The Neuronal Membrane Is in a Liquid Ordered Phase. Before we analyze the interaction of Aβ with the neuronal membrane, we determine the characteristics of the latter. The mass density profile of each lipid and water along the membrane z axis (*SI Appendix*, Fig. S1) shows the distribution of the bilayer components, as well as the bilayer thickness. The positions of the headgroups are at similar locations for POPC (PC), POPE (PE), POPS (PS), and SM. CHOL, on the other hand, is shifted toward the hydrophobic core of the bilayer, while GM1 is farther away from the bilayer center, due to the protrusion of the sugar groups from the xy surface of the membrane (Fig. 1*A*). The headgroup-to-headgroup distance of PC, PE, and PS indicates a bilayer thickness of 4.65 ± 0.03 nm.

We calculated the acyl chain order parameter S_{CH} of the C–H bonds of all the lipid tails (*SI Appendix*, Fig. S2) to gain insight into their arrangement within the membrane. Values of 0.35 to 0.4 for the order parameters of carbon atoms 4 to 10 are reached, which is an increase compared to the order parameters found in other membranes (28, 29). This is due to the effects of cholesterol and sphingomyelin, which are known for their role in increasing lipid order. Notably, we find the acyl chains of GM1 and SM to be the most ordered. We can thus conclude that the neuronal membrane is in the liquid-ordered state, which is in agreement with previous observations (24, 30).

GM1 Forms Ganglioside Clusters. The radial distribution function (RDF) of all possible lipid pairings was calculated to monitor the effect of these pairwise interactions on lipid clustering (Fig. 1 *B* and *C* and *SI Appendix*, Fig. S3). A distinct RDF peak is seen at ≈0.45 nm for the self-clustering of GM1 and pronounced peaks are seen at 0.55 and 0.6 nm for the formation of CHOL and SM clusters, respectively, while all other lipids do not tend to self-associate. The self-clustering of GM1 is considerably stronger than that of the other lipids. Thus, taking the relatively low concentration of GM1 (4%) into account, one can conclude that GM1 has a strong tendency to self-associate that can result in its sorting. No strong clustering between mixed lipid pairs is observed. Notable coassociation is seen only for SM with POPE, CHOL, and GM1. Interestingly, the RDF of PE–PS has a higher peak compared to that of PE–PE and PS–PS, respectively. The dispersion of PS is understandable given that it is negatively charged. The negative charge of both GM1 and PS also explains why these two lipids avoid coclustering.

To elucidate the dominant lipid–lipid interactions underlying the RDFs, the average numbers of hydrogen bonds (H bonds) between the different lipid pairs were evaluated. *SI Appendix*, Fig. S4 shows that the sorting of GM1 results from its ability to build a network of H bonds via its sugar headgroups, despite its negative charge. The propensity of SM to form H bonds with itself also gives rise to its self-clustering, whereas the minor self-clustering seen for CHOL is a result of the cholesterol condensing effect. This effect does not result from attractive van der Waals interactions between CHOL molecules, but from a reduced membrane perturbation energy if small cholesterol domains are formed (31). However, such cholesterol clusters are not particularly stable, as evidenced by only a small peak in the RDF for CHOL–CHOL. The coclustering of CHOL and SM is facilitated by H bonds formed between the hydroxyl of CHOL and the amide group of SM, which agrees with previous findings (32). The RDF profile of SM–GM1 can also be explained by H-bond formation. We conclude that H bonds play an essential role in stabilizing lipid clusters within the neuronal membrane.

Aβ42 Dimerizes at the Neuronal Membrane and Interacts with GM1. To understand the effects of the neuronal membrane on the aggregation of Aβ42, we analyzed the 6×2 μs of MD data in the presence of the lipid bilayer and compared the aggregation to the 6×2 μs of MD simulations done in the aqueous phase. We first assess whether and how the two peptides bind to and interact with the membrane.

To follow the association between Aβ42 and the neuronal membrane, we calculated the minimum distance of both peptides from the lipid bilayer surface for each of the six simulations (*SI Appendix*, Fig. S5). It can be seen that peptide 1 usually interacts with the membrane at a closer distance than peptide 2 does, which can be explained by the fact that the initial structures of five of the six simulations were selected from the initial 2-μs simulation. This allows us to better elucidate the effects of the membrane on the preferentially membrane-bound

peptide as their mutual interaction time is larger than it would have been if both peptides had the same interaction probability. Nonetheless, both peptides tend to be associated with the membrane as an intact dimer, since if one peptide is >0.5 nm away from the membrane, implying that this peptide is in solution, then very often this is also the case for the other peptide. Fig. 2 shows representative snapshots for the membrane association of Aβ42, including one for loose binding in Fig. 2*A*. Fig. 2*B* represents the situation where peptide 1 is in close contact with the membrane, while peptide 2 is a bit farther away. The opposite, less prevalent situation with peptide 2 being closer is depicted in Fig. 2*C* and is less common, while Fig. 2*D* shows how both peptides can bind tightly to the membrane. Fig. 2 further suggests that Aβ42 tends to interact with GM1 instead of the other lipids and that β-sheets are the dominating secondary structure in peptide 2 but not in the more membrane-bound peptide 1. The analysis of the contacts between Aβ42 and the various lipids confirms that the peptide has a high tendency to associate with GM1, followed by PC, PE, and PS (*SI Appendix*, Fig. S6). Here, we emphasize that these contacts are not normalized but absolute values. Considering that only 4% of the lipids are GM1 while the phospholipids make up for more than two-thirds of the membrane, one can thus conclude that Aβ42 is highly attracted to GM1. Interestingly, almost no contacts are made with CHOL or SM.

To rationalize the driving force that controls Aβ42 interaction with the membrane surface, the interaction energy of each Aβ42 residue with each of the lipid components was calculated and partitioned into its electrostatic (E_{Coul}) and Lennard-Jones (E_{LJ}) contributions (Fig. 3). Notably, the lipid interactions of peptide 1 are more favorable than those of peptide 2, agreeing with the observation that peptide 1 interacted more strongly with the membrane. Our results suggest that the major driving force for the association of the peptides to the membrane is the electrostatic attraction to PC, PE, and PS, especially via the highly charged N-terminal region and residues F20 to A30. Residues at the N terminus had the strongest interactions with the membrane, such as D1, E3, and D7 with PE; D1 with PC; and R5 with PS. The latter interaction involves H-bond formation (*SI Appendix*, Fig. S7), which is enabled via the carboxylate group of PS, whereas the primary ammonium group of PE forms H bonds with D1, E3, and D7. The tertiary ammonium group of PC, on the other hand, does not support H-bond formation, leading to a relatively low H-bond propensity between PC (via its phosphate group) and Aβ42. The interactions between GM1 and Aβ42 are driven by both Coulomb and Lennard-Jones energies (Fig. 3) and are facilitated by the sugar headgroups of GM1, which protrude from the membrane and are therefore particularly accessible to Aβ42. Moreover, the interactions with GM1 derive from a considerable number of H bonds, which involve almost all residues of both peptides, but particularly those of peptide 1.

No direct interaction between Aβ42 and CHOL was observed (*SI Appendix*, Fig. S6), due to the deeper, unexposed position of CHOL within the membrane. Interestingly, even though SM has the same headgroup as PC, which is also located at a similar position along the bilayer normal, Aβ42 hardly interacted with SM. This can be understood by considering the preference of SM to form H bonds with other lipids including itself, which reduces its tendency to create H bonds with the peptide.

The Aβ42 Dimer Does Not Affect the Neuronal Membrane. To determine whether the peptides affect the structure of the lipid membrane, we calculated the lipid order parameter for the lipids that are within 0.5 nm of the peptide when adsorbed to the membrane (*SI Appendix*, Fig. S2). The results suggest no notable change in the lipid order parameter due to the interactions with Aβ42. Moreover, only a slight deviation of about $\pm$0.1 nm was seen in the bilayer thickness (*SI Appendix*, Fig. S8) at the site of peptide interaction. We thus conclude that the peptides interact only with the lipid headgroups without inserting into the membrane, thereby preventing larger changes in the membrane order and thickness.

Different Aggregation Pathways in Solution and at the Neuronal Membrane. To unravel differences within the aggregation pathways, we computed transition networks (TNs) for the Aβ42 dimerization both in the aqueous phase and in the presence of the neuronal membrane. To this end, we characterized the conformations by assigning the aggregate state (monomer or dimer), the number of hydrophobic contacts between the peptides in a dimer, and the number of residues in β-strand conformation as descriptors. To further simplify the TNs, we grouped both the number of hydrophobic contacts and the number of residues in β-strand conformation in blocks of five such that we end up with ranges h1 to h12 and b1 to b6. For example, h1 and b1 stand for hydrophobic contacts and the number of residues in β-strand conformation, respectively, ranging from 1 to 5. The maximum state h12 involves between 56 and 60 hydrophobic contacts and the b6 state means that between 26 and 30 residues per peptide adopted a β-strand conformation.

The resulting TNs (Fig. 4) are characterized by two regions: the monomeric region (on the left side of the TNs) and the dimeric region (in the middle and the right side of the TNs), where the former evolves into the latter. These regions are

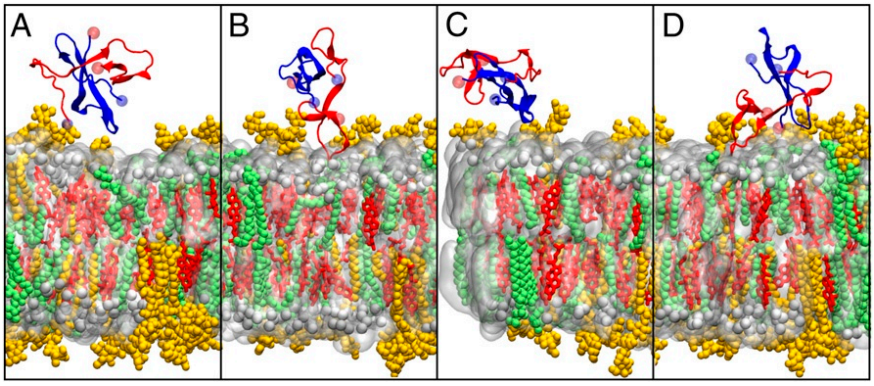

Fig. 2. Snapshots of Aβ42 interacting with the neuronal membrane. Peptide 1 and peptide 2 are shown as cartoons in red and blue, respectively, with their termini indicated by spheres (N, light blue; C, light red). The color coding for the membrane is the same as in Fig. 1*A*. Representative interaction patterns are provided: (*A*) both peptides being loosely attached to the bilayer surface, (*B*) peptide 1 being in close interaction with the membrane and peptide 2 being bound to peptide 1, (*C*) the opposite situation to that in *B*, and (*D*) both peptides being in close contact with the membrane.

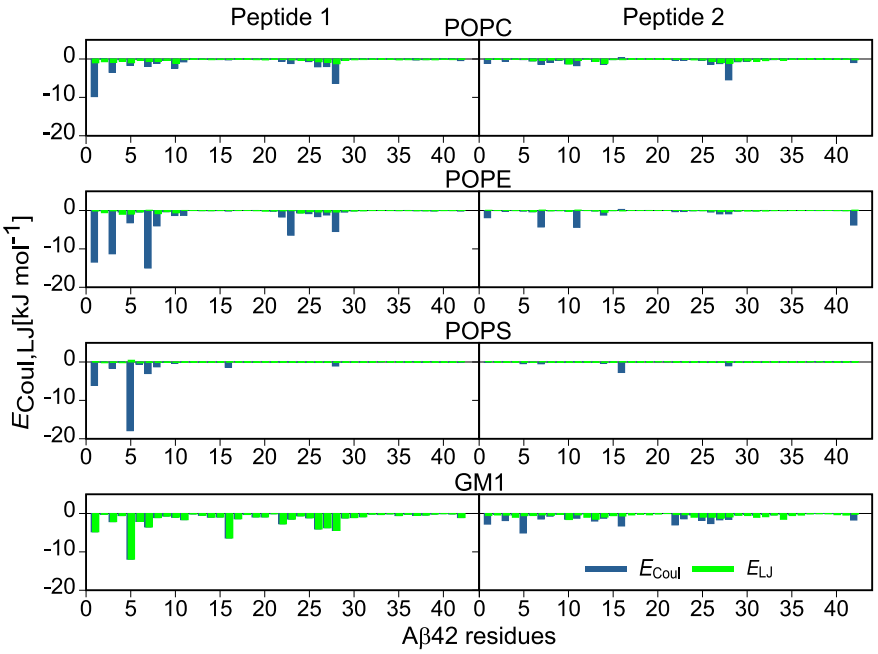

Fig. 3. The average interaction energies of peptide 1 (*Left*) and peptide 2 (*Right*) with each lipid of the neuronal membrane. Electrostatic and Lennard-Jones energies are shown in blue and green, respectively. The more negative an energy is, the more attractive is the corresponding interaction.

connected by several bridging nodes, which, on average, are characterized by a higher amount of β-sheet (i.e., larger n in the descriptor bn) in the case of the solution system. In both TNs, a representative bridging node is indicated by a green circle, [2, h2, b6] for the solution system and [2, h2, b2] for the membrane system, which are further augmented by a characteristic structure. In solution, there are more transitions between monomers and dimers, which indicates a higher number of association and dissociation events. In general, the TN for the solution system exhibits more nodes and transitions.

A closer inspection of both TNs reveals how the two peptides evolve from the monomeric random coil state, which is represented by node [1, 0, 0] with no interpeptide hydrophobic contacts and no residues in β-strand conformation, to dimers with only a few hydrophobic contacts, as present in states [2, h1, bn]. Here bn ranges from b1 to b6, indicating an increase in β-strand content as the structure changes along the path through nodes [1, 0, bn] and [2, 0, bn]. The dimers with no hydrophobic contacts are so-called encounter complexes, where the minimal distance between the two monomers fell below 4.5 nm, and subsequently form stable dimers by increasing their contact area as interpeptide contacts form. This process stabilizes the dimer and is accompanied by an increase in β-strand content. In solution, the dimers form more interpeptide hydrophobic contacts, reaching states [2, h12, b6] and [2, h13, b5] wherein 50 to 70% of all Aβ42 residues form a β-sheet. In the presence of the neuronal membrane, both the hydrophobic contact area and β-sheet content are reduced, with the maximal values being [2, h10, b4] and [2, h9, b5], explaining the smaller number of nodes in this TN. Some of the interpeptide contacts are replaced by peptide–lipid contacts, which in turn inhibits β-sheet formation. This conclusion is confirmed by the representative structures shown in Fig. 4 and those illustrating the membrane adsorption of the dimer (Fig. 2). The membrane-adsorbed dimer structures are more compact than the dimer structures in solution, which feature extended β-sheets.

Long β-Strands in Solution and Compact Aβ42 Structures at the Membrane. To quantify the effect of both aggregation and membrane adsorption on the peptide secondary structure, we determined the propensity of each residue to adopt a helical conformation, to be part of a β-sheet, or to be in a turn or bend conformation (Fig. 5*A*). For the dimer both in solution and on the membrane, β-sheet formation is observed. Using the same force field, mostly disordered conformations were sampled for the Aβ monomer during a 30-μs MD simulation, with an average β-sheet content of about 15% (20). This rises to 36% for the dimer in solution, which indicates that dimerization causes Aβ42 to undergo a disorder-to-order transition with β-sheet folding. The β-sheet content for the membrane-adsorbed dimer is 28% and thus smaller compared to that for the solvated dimer. This decrease is particularly pronounced for peptide 1, which interacts more strongly with the membrane than peptide 2 does. Instead, peptide 1 exhibits more turns, bends, and random coil structures, which suggests that the membrane inhibits β-sheet formation. Also, no pronounced helix formation is observed for the membrane-bound dimer, which one might expect based on NMR results (33) and previous simulation studies of Aβ that employed implicit membrane models (34, 35). However, a closer inspection of these studies reveals that for helices to be present, the affected Aβ residues need to be inserted into the hydrophobic membrane core, which did not occur here. It remains to be shown what comes first: helix formation or membrane insertion. In solution, both peptides feature a very similar secondary structure pattern along their primary structure. They display a particularly high propensity for a β-sheet in the regions Q15 to F20 of the central hydrophobic core (CHC) and A30 to V40 from the C-terminal hydrophobic region. This excludes the residue pair G37/G38, which has a tendency to form a turn as previously shown in simulations (36) and NMR spectroscopy (37).

The analysis of the intrapeptide contacts, derived from interresidue distances (Fig. 6), indicates that in solution long β-hairpins between two antiparallel strands involving residues Y10 to V24 and Q27 to V40 formed in both peptides. These β-hairpins are particularly stable since the β-sheet propensity of the strongly hydrophobic regions ^{18}VFF^{20}and ^{32}IGL^{34} even reaches values above 90%. In previous simulation studies, β-sheet formation upon Aβ dimerization has also been the prevailing finding (see table 2 of ref. 36 and references therein

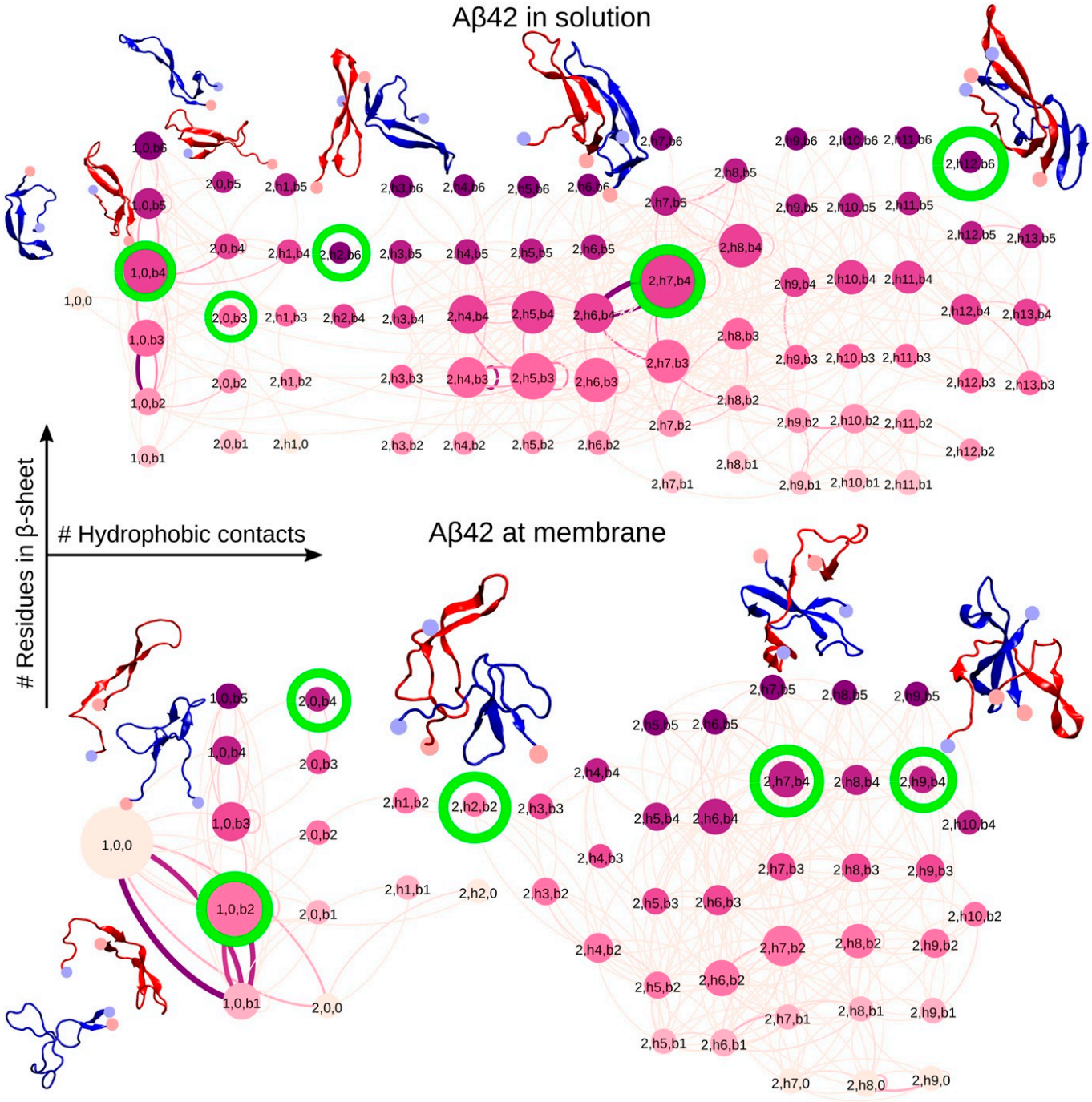

Fig. 4. The TN for Aβ42 dimerization in the aqueous phase (*Top*) and in the presence of the neuronal membrane (*Bottom*). Each node is defined by three descriptors: oligomer size, number of interpeptide hydrophobic contacts, and number of residues in β-strand conformation. The last two descriptors are grouped in blocks of five and are named h1 to h12 for hydrophobic contacts and b1 to b6 for the number of residues in β-strand conformation. The nodes are connected by edges that represent transitions between the connected peptide states. The size of the nodes and the thickness of the edges are proportional to the respective state or transition probability. They are colored based on the descriptor reflecting the number of residues in β-strand conformation (from light pink for no β-sheets to dark purple for the maximum amount of β-sheets in b6). For the nodes circled in green representative peptide conformations are shown (see color code in Fig. 2).

as well as refs. 38–41). As found here, the β-sheets are preferentially formed between the C-terminal hydrophobic regions, followed by the involvement of the CHC. However, in most of these previous studies, the β-sheets are shorter and the overall dimer appearance is more compact. This likely resulted from the usage of older force fields, which were not optimized for IDPs and are known to provide too compact IDP conformations (42). Exceptions are a coarse-grained discrete MD study (38) and a structure-prediction study for transmembrane Aβ oligomers (35) that yielded similarly extended β-sheets. This is confirmed by comparing the intrapeptide contacts that are present in the different Aβ42 dimers, as shown in *SI Appendix*, Fig. S9. *SI Appendix*, Fig. S9 further shows that the β-hairpin centered at G25/S26 coincides with the peptide regions that are involved in the cross–β-sheet structure found in U-shaped Aβ fibrils (43, 44).

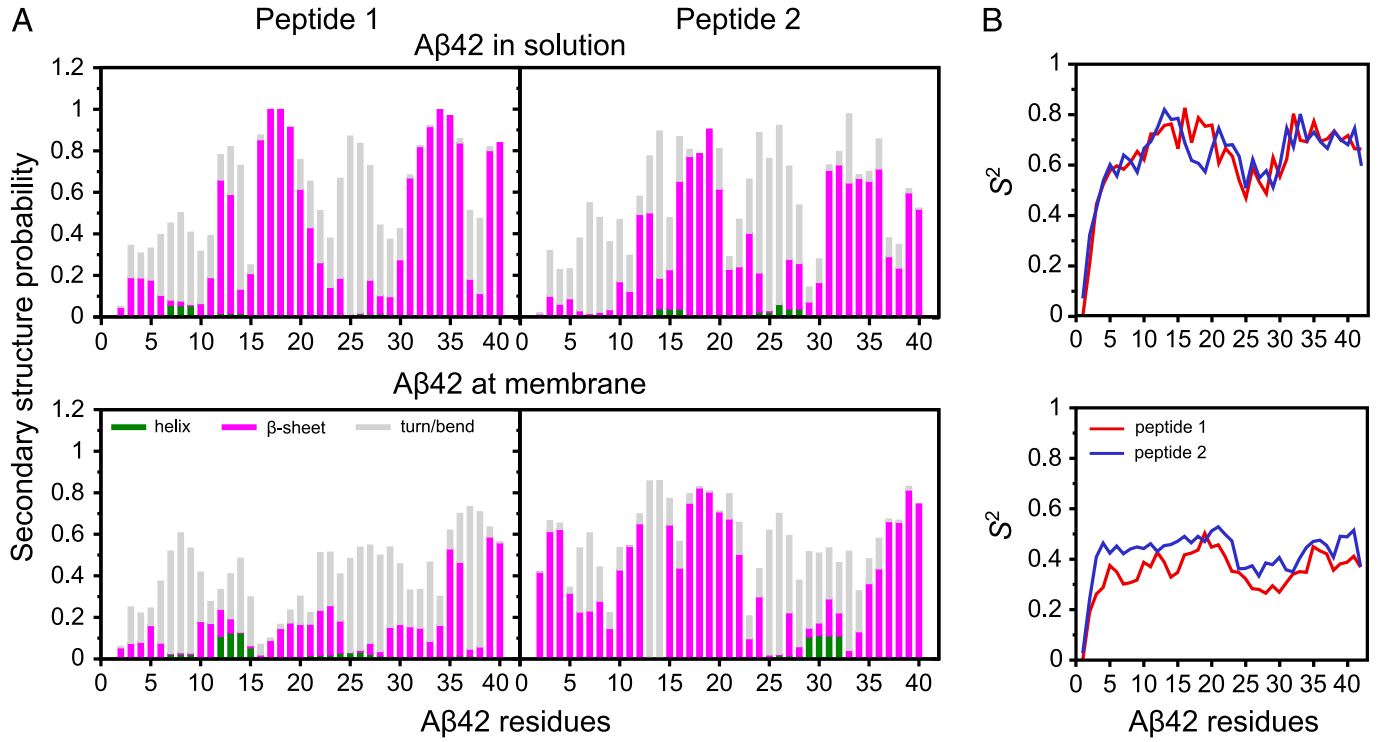

Fig. 5. Structural characteristics of the dimer in the aqueous phase (*Top*) and in the presence of the neuronal membrane (*Bottom*). (*A*) Probability of secondary structures to form in each residue of the peptides. The bars represent the cumulative secondary structure probabilities consisting of helix (green), β-strand/bridge (magenta), and turn or bend (gray). The difference from 1.0 presents the probability of the random coil state. (*B*) The average order parameter S^2 of each residue and peptide.

The intrapeptide contacts present in the membrane-adsorbed dimer are more diverse and different in the two peptides. For peptide 2 they reveal the prevalence of two shorter hairpins, one centered at H14 and the other one at G25, and several contacts between N- and C-terminal residues. The more membrane-adsorbed peptide 1, on the other hand, is devoid of noteworthy contacts involving its N-terminal residues. These are the amino acids that preferentially interact with the membrane and are therefore not available for interresidue interactions. In the C-terminal region of peptide 2 the formation of a very short β-hairpin is visible. Overall, the intrapeptide contacts corroborate the conclusion that at the membrane Aβ42 adopts more compact conformations with less β-sheet than the dimer in solution.

Dimerization in Solution Is Mainly Driven via the Hydrophobic C-Terminal Region. To obtain an overview of how the two peptides are arranged with respect to each other as dimers, we calculated the interpeptide distances on a per-residue basis. The resulting distance matrices for the two dimer systems (Fig. 6) are almost symmetric with respect to their diagonal and are characterized by areas of high contact density along the diagonal as well as in the upper left and lower right quadrants. Only the D23 to K28 region in both peptides and in both environments does not show a noteworthy contact propensity. This is the same region of the peptide that we assigned a turn or bend conformation (Fig. 5). It can thus be concluded that this bend/turn region does not form the interpeptide interface.

For the dimer in solution, the highest contact density is observed between the two C-terminal regions, A30 to A42, which are the same regions where a high β-propensity was identified. Therefore, these two C-terminal regions not only are involved in intrapeptide β-sheets, but also form an interpeptide β-sheet in solution. This is confirmed when analyzing the residue–residue interaction energies between the two peptides, which involve Coulomb interactions deriving from backbone H bonds and Lennard-Jones energies originating from interactions between hydrophobic residues (*SI Appendix*, Fig. S10). While the distance matrix does not show a clear preference for either an antiparallel or a parallel β-sheet between the two C-terminal regions, and both arrangements are indeed possible (see the representative conformations for nodes [2, h7, b4] [parallel] and [2, H12, b6] [antiparallel] of the corresponding TN in Fig. 4), the interaction energies indicate that the antiparallel arrangement is favored. This is different from Aβ fibrils where only parallel β-sheets are found. Other preferred contacts in the dimers form between the CHC of one peptide and the C-terminal region of the other peptide. The fourth area with a certain, yet smaller probability of interpeptide contact is between the CHC regions of both peptides. However, these contacts are weaker than those between the two C-terminal regions, as the corresponding interaction energies are smaller in magnitude (*SI Appendix*, Fig. S10). The dissection of the interaction energies further reveals that attraction between the oppositely charged residues E22/D23 and K28 is involved in the association process, which is in agreement with previous findings (45).

The distance matrix of the membrane-adsorbed dimer looks similar to the one of the dimer in solution. However, the contact areas are more pronounced, indicating less structural diversity in the internal arrangement of the dimers. Second, the area without interpeptide contacts around residues D23 to K28 is larger. This applies to peptide 1 in particular and can be explained by the contacts that this peptide forms with the membrane instead. Third, the order of areas with the highest contact probability is different from those of the solution system. The shortest distances in the membrane-adsorbed dimer are observed between the CHC of peptide 1 and the C-terminal region of peptide 2, followed by the contacts between both CHC regions. However, based on the secondary structure analysis, β-sheet formation between the two peptides is less likely and is largely limited to within peptide

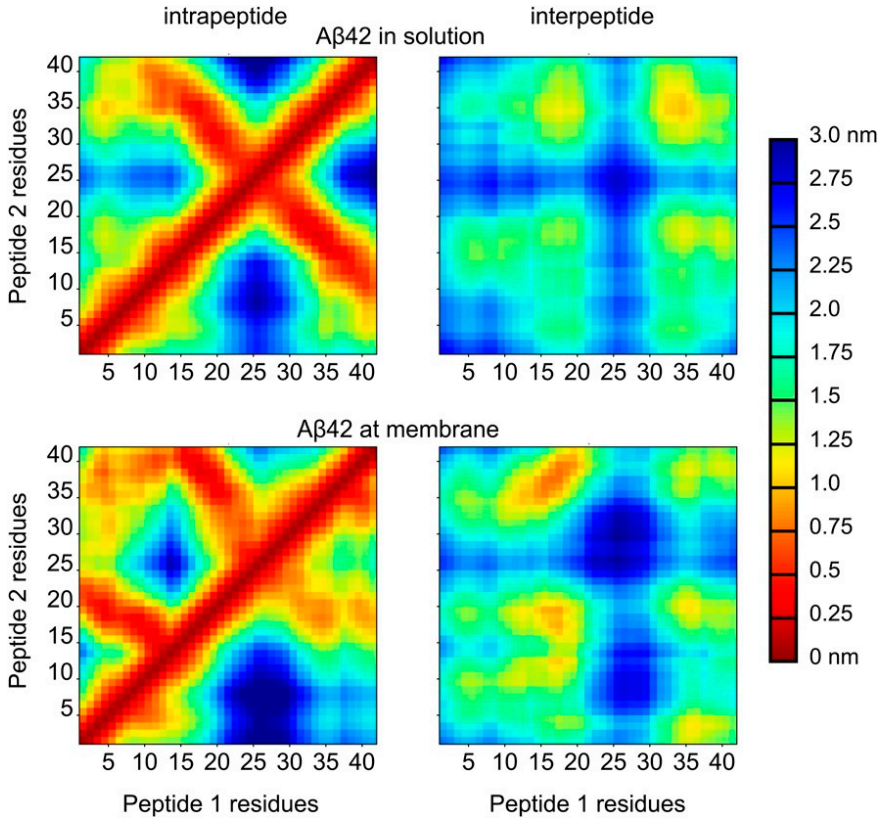

Fig. 6. The distance matrices illustrating intra- and interpeptide contacts between residues for the dimer in the aqueous phase (*Top*) and in the presence of the neuronal membrane (*Bottom*). The intrapeptide contacts within peptide 1 are shown below the main diagonal and those within peptide 2 above it. The interpeptide contacts are shown for peptide 1 and peptide 2 composing the dimer. The color bar on the right indicates the average intra- and interresidue distances (in nanometers).

2. Contacts between the C-terminal region of peptide 1 and the N-terminal region of peptide 2 are also observed. As a result, the latter region exhibits an increased β-sheet propensity, which extends up to residue Y10 (Fig. 5*A*). It initially was assumed that the N-terminal region of Aβ is always disordered. However, this was later refuted, first by simulations (*SI Appendix*, Fig. S9) and then by cryo-electron microscopy (35, 46). Contacts between the C-terminal regions are of less relevance for the membrane-adsorbed dimer due to the competition between peptide–peptide and peptide–membrane interactions. The ranking of the interpeptide contact preferences is confirmed by the analysis of the interaction energies (*SI Appendix*, Fig. S10). Unlike in solution, attraction between E22/D23 and K28 does not play a role during the dimerization of Aβ, which can be explained by the preference of K28 to associate with the membrane.

Reduced Global Motions but Increased Local Disorder in the Membrane-Adsorbed Dimer. To quantify the peptide dynamics, we calculated the S^2 order parameters to monitor the mobility of the N–H bond vectors of the peptide backbone along with the average global rotational correlation times, $\langle\tau\rangle$ (Fig. 5*B*). These quantities would be directly comparable to those determined by NMR spectroscopy, which, however, are not available yet. The global rotational dynamics of the Aβ42 dimer in solution occur on the low nanosecond time scale with $\langle\tau\rangle = 20 \pm 10$ ns. The S^2 values reflect the different secondary structure propensities of the various residues. They are above $\approx$0.7 for the residues in a β-conformation, while the more mobile turn region and neighboring residues ranging from E22 to A30 have S^2 values between 0.5 and 0.7, and the disordered N-terminal region has order parameters below 0.5. The comparison to the S^2 values of the Aβ40 monomer confirm that the dimer in solution is considerably more folded, since for the monomer all S^2 values are below 0.4 (47). The global rotational dynamics of the Aβ42 dimer in the presence of the neuronal membrane are by a factor of 5 slower than in solution: $\langle\tau\rangle = 108 \pm 30$ ns. Interestingly, the slower motion is accompanied by decreased order parameters compared to that seen in solution; the S^2 values range from 0.25 to 0.55 (and below 0.25 for the N-terminal residues, similar to the situation in the solution dimer). The overall reduction in S^2 for the membrane-adsorbed dimer implies that the peptides are generally less folded than they are in solution, which agrees with the observed reduction in β-sheet and increase in random coil. Thus, a picture emerges where on the one hand the overall peptide dynamics are reduced due to the adsorption on the membrane, while at the same time the interactions with the membrane reduce the local peptide order as reflected by the S^2 values.

Discussion

In the present study, all-atom MD simulations on the microsecond time scale have been performed to elucidate the mechanism of Aβ42 dimerization in pure water and in the presence of a neuronal membrane. The consideration of a neuronal membrane consisting of six components (PC, PE, PS, CHOL, SM, and GM1) is a major step forward compared to previous simulation studies on Aβ–membrane interactions, which included three lipid types or fewer. Dimerization was observed in the aqueous phase as well as at the neuronal membrane. However, the resulting dimer structures showed significant differences. Our simulations of Aβ42 dimerization in solution revealed a coil-to-β transition that is the first step along the amyloid

aggregation pathway. The dimer conformations sampled in solution bear certain similarities to the β-sheets found in the U-shaped Aβ42 amyloid fibrils. To our knowledge, a dimer structure with such a high β-sheet content and overall order has never been reported from all-atom MD simulations where the aggregation of Aβ progresses from disordered monomers into oligomers. We conclude that only the MD sampling of several microseconds and the use of a force field well suited to Aβ allow the random coil to β-sheet transition to be observed in a simulation (22). Thus, with these simulations we finally shed light on the structural transitions that might lead to nuclei enabling amyloid formation. Our future simulations will test whether the dimers that formed in solution here are indeed on pathway toward amyloid fibrils.

On the neuronal membrane, the dimer conformations are generally less ordered than in solution. The dimerization took place on the membrane, with one of the two peptides being preferentially adsorbed to the membrane and the other one associating with the already membrane-attached peptide without noteworthy interacting with the membrane itself. The directly adsorbed peptide in particular has a higher amount of random coil and less β-sheet. The membrane adsorption is mainly driven by electrostatic interactions between the charged N-terminal residues of Aβ and the headgroups of PC, PE, and PS, in addition to hydrogen bonding between the sugar moieties of the GM1 lipids and Aβ42 residues across its whole primary structure. GM1 is found to form clusters within the neuronal membrane, which are the preferable site for Aβ to bind to the membrane surface. This is in line with experimental results that revealed GM1 as part of a neuronal membrane to be the main interaction partner of Aβ, whereas less binding was seen for SM and also PC (48). No insertion of the peptides into the hydrophobic region of the membrane was observed in our simulations. Instead, the interactions with the membrane stiffened both peptides, restricting their conformational diversity compared to the Aβ42 dimer simulated in the aqueous phase. Not only did the transition networks reveal a reduction in the number of conformational states, but also the correlation times of the N–H bond vector motions indicated an impaired peptide motion. However, while adsorption was found to have profound effects on the Aβ42 dimer, the membrane was only marginally affected.

Our observations are in agreement with a large and diverse set of experimental results. Of special note is a study that analyzed the effects of glucose on Aβ42 aggregation (49). In this study, Kedia et al. (49) found that Aβ42 forms low-molecular-weight oligomers in the presence of sugars and that these oligomers do not adopt a β-sheet structure. This agrees with our observation that Aβ42 dimers that preferentially interact with the glycans of GM1 form fewer β-sheets than Aβ42 dimers that form in solution do. Moreover, another study revealed that Aβ oligomers that are present in the brain interstitial fluid are sequestered from that fluid by strongly binding to GM1, which also prevented the further aggregation of Aβ (48). We are aware of studies by Ikeda et al. (50) and Matsuzaki (51) that concluded that GM1 exhibits a strong Abeta fibril seeding potential following the formation of β-sheet–rich oligomers on GM1 clusters. However, these clusters are much larger than those formed in our simulations, as Ikeda et al. (50) and Matsuzaki (51) employed ganglioside-rich (>20 mol% vs. the 4 mol% used in our study) membranes, where GM1 forms an interconnected network of micrometer size yielding glycan platforms in liquid-ordered membranes. As elaborated by Hof and coworkers (52), the scenarios for membranes with high and low GM1 contents are not necessarily contradicting each other but rather complementary.

Another finding by the study of Kedia et al. (49) was that the unstructured Aβ42 oligomers that formed in the presence of glucose are able to interact with membrane bilayers. Their diffusion decreased by a factor of about 4 upon membrane adsorption, which agrees nicely with our observation that membrane interactions reduce the dynamics of the dimer. Moreover, no incorporation of the unstructured Aβ42 oligomers into the membrane was recorded (49), which also concurs with our findings. We conclude that, if a β-sheet structure should be required for membrane insertion of Aβ aggregates to occur, GM1 in the neuronal membrane has a neuroprotective effect as it could break the β-sheet structure in the Aβ dimer. This finding would be in agreement with the neuroprotective and neurogenerative effects reported for GM1 (53–55) and the conclusion that the neurotoxic activity of amyloid oligomers increases with their β-sheet content (8). On the other hand, Selkoe and coworkers (48) found that even though GM1 sequesters Aβ from the brain interstitial fluid, thereby inhibiting the aggregation of Aβ, the binding of the peptide to GM1 alone mediates neurotoxic effects. This once more highlights that the interplay between Aβ, its aggregation, and the neuronal membrane is far from trivial and despite the wealth of already published studies on that matter, further studies are needed to fully solve this puzzle.

Materials and Methods

Setup of the Simulated Systems. The systems modeled are composed of two Aβ42 peptides, which were simulated in the aqueous phase and in the presence of the neuronal lipid membrane. The initial Aβ42 structures were taken from the most populated clusters from a preceding 3-μs MD simulation of monomeric Aβ42 in solution. The neuronal membrane model composed of 152 PC, 96 PE, 20 PS, 80 CHOL, 36 SM, and 16 GM1 molecules was generated as symmetric lipid bilayer using the CHARMM-GUI interface (56).

The simulated membrane system also contained water layers above the upper and beneath the lower membrane leaflet, using the three-site transferable intermolecular potential (TIP3P) for modeling the water molecules, with sodium and chloride ions added at the physiological concentration of 150 mM. The two Aβ42 peptides were placed in the upper water layer at a distance of $\approx$2 nm from the equilibrated lipid bilayer surface and at a distance of $>$1 nm between the closest atoms from the two peptides. All distances from the peptides to any of the simulation box edges were at least 1.2 nm to avoid interactions between the peptides and their periodic images. The total number of atoms in the modeled membrane system was $\approx$160,000 atoms and the box size was about $9.6 \times 9.6 \times 13.6$ nm^3. The setup of the system in the aqueous phase was similar, but without a lipid bilayer, resulting in a system size of about $9.2 \times 9.2 \times 6.5$ nm^3, and contained $\approx$54,760 atoms. This amounts to peptide concentrations of 4 and 6 mM, respectively. This is two to three orders of magnitude higher than the concentrations used in corresponding in vitro experiments. However, it is beyond our computational capabilities to model μM peptide concentrations at the atomistic level. Moreover, simulations at such low concentrations would most of the time simulate only the diffusion of monomeric peptides (57). We therefore aim to model the oligomerization of Aβ in a stepwise fashion (57, 58), starting here with simulations of dimers.

MD Simulation Conditions. The all-atom MD simulations were performed using GROMACS/2018.2 (59) along with the CHARMM36m force field for Aβ42 (60) and Charmm36 for the lipids (61). Each system was first energy minimized using the steepest-descent algorithm to remove atomic clashes. This was followed by equilibration in the canonical ensemble where a temperature of 310 K was regulated with the velocity-rescale thermostat (62). Next, the system was equilibrated under isobaric–isothermic conditions to obtain a pressure of 1.0 bar, where the pressure was regulated using a semi-isotropic Parrinello–Rahman pressure coupling scheme (63). Periodic boundary conditions were set in all directions. Both the van der Waals and Coulomb force cutoffs were set to 1.2 nm in real space. The particle mesh Ewald (PME) method was applied for calculating the electrostatic interactions. Before we studied the interaction of Aβ42 with the neuronal membrane, we equilibrated the membrane without peptides being present for 1 μs. For Aβ42 dimer systems, an initial simulation was run for 2 μs, from which different snapshots were randomly selected and used as starting structures for the next 5×2-μs simulations. For the subsequent analysis, we combined the data from the six independent simulations and derived the results presented in this study.

Analysis of the Lipid Bilayer Properties. For the determination of the order parameter of the lipid acyl chains, S_{CH}, one uses the C–H bond vectors

present in the lipid tails and calculates the orientation of these vectors with respect to the bilayer normal (the z axis) using

$$S_{CH} = \frac{\langle 3\cos^2\theta - 1\rangle}{2}, \quad [1]$$

where θ is the angle between the C–H bond vector and the bilayer normal. The angular brackets indicate the ensemble average. This calculation was accomplished with a Python script available at https://github.com/NMRLipids/MATCH (64).

The mass density profiles along the bilayer normal were calculated using the "gmx density" tool. The distance between the peaks of the total density gives an estimate of the bilayer thickness. Furthermore, the bilayer thickness was calculated as the z-position difference between the P atoms of the lipid headgroups in the upper and lower leaflets using the "gmx distance" tool. The RDF provides information about the probability of finding a particle at a certain distance from another particle. We calculated the radial distribution functions of different lipid pairs in two dimensions (the xy plane) using the "gmx rdf" tool. The hydrogen bonds between different lipid pairs were determined using "gmx hbond." A hydrogen bond was recorded when the angle between the donor and acceptor bonded hydrogen was between 150 and 180° and the distance between the two atoms was within 0.35 nm.

Analysis of Aβ42 Properties. The secondary structure of each Aβ42 residue was determined using the "define secondary structure program" (DSSP) (65) invoked via the GROMACS tool "do dssp." To facilitate a clear representation, the data of similar secondary structures are grouped together: β-strand and β-bridge are combined as β-sheet and β-turn and bend as turn/bend; and the helix includes α, π, and 3_{10}-helices.

For the calculation of the S^2 order parameter we used the MOPS[2] (Molecular Order Parameter S[2]) software developed in ref. 66 to calculate S^2 from the N–H bond vector autocorrelation function. To facilitate the calculation, each trajectory was divided into subtrajectories of $t_{sub} = 100$ ns length. For each of the subtrajectories the S^2 values and the rotational correlation times, τ, were calculated and subsequently averaged over all subtrajectories. The rotational correlation times were further averaged over all residues and both peptides, denoted as $\langle\tau\rangle$, whereas S^2 is provided per residue and peptide. Since $\langle\tau\rangle$ for the membrane system is in the same range as t_{sub}, we checked on the convergence for the S^2 calculation in this case (*SI Appendix*, Fig. S11).

Transition Networks. For the generation of the TNs to characterize the assembly of peptides into dimers we used the ATRANET (Automated Transition Network) software (https://github.com/strodel-group/ATRANET) (27). It defines the oligomerization state by a number of descriptors, depending on the properties of interest. In our case, three descriptors are used: The first one is the oligomer size, which can be 1 in the case of monomer or 2 in the case of a dimer. To define a dimer, the minimum distance between any atom of peptide 1 and any atom of peptide 2 along with the requirement of this distance to be within 0.45 nm was used. The second descriptor, the number of hydrophobic contacts between both peptides, counts the possible interpeptide atom pairs formed between the hydrophobic amino acids of Aβ42 that are within a certain cutoff (also 0.45 nm). The third descriptor is the number of residues in β-strand conformation, which is evaluated using DSSP and averaged over both peptides. Feeding these descriptors to ATRANET leads to a transition matrix that can be visualized using Gephi (67). Snapshots of the representative structures from the transition network were rendered using the visual molecular dynamics (VMD) program (68).

Calculation of Aβ42–Bilayer Interactions. The peptide–lipid interactions were analyzed by calculating the interaction energy between each Aβ42 residue and the headgroup of each lipid component using "gmx energy." The "gmx mindist" program was employed to determine the number of contacts between each Aβ42 residue and each lipid component in the neuronal membrane. A contact was recorded when the distance between any two nonhydrogen atoms from a residue and a lipid was within 0.5 nm. The H-bond propensity was determined by the number of times an H bond was formed between hydrogen bond donating and accepting atoms in lipid pairs.

Data Availability. The MD trajectories and the analysis scripts are available at Mendeley Data, https://data.mendeley.com/datasets/92mkp4pk86. All data resulting from the analysis of this raw data is shown in the main text or *SI Appendix*.

ACKNOWLEDGMENTS. H.F., M.K., A.S.-A., and B.S. acknowledge funding for this project from the Palestinian-German Science Bridge financed by the German Federal Ministry of Education and Research. B.S. received funding for this project from the Deutsche Forschungsgemeinschaft (German Research Foundation, https://www.dfg.de/) through Grant 267205415 (CRC 1208, Project A07). M.C.O. received funding for this project from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant and it is cofinanced by the South Moravian Region under Grant 665860. This article reflects only the authors' view and the European Union is not responsible for any use that may be made of the information it contains. This research was supported by the European Union and the Hungarian State, cofinanced by the European Regional Development Fund in the framework of the GINOP-2.3.4-15-2016-00004 project, aimed to promote the cooperation between the higher education and the industry. M.C.O. and B.S. gratefully acknowledge the Gauss Centre for Supercomputing (GCS) e.V. (www.gauss-centre.eu) for funding this project by providing computing time on the GCS Supercomputer SuperMUC-NG at Leibniz Supercomputing Centre (https://www.lrz.de/). H.F., M.K., and B.S. gratefully acknowledge the computing time granted through Jülich Aachen Research Alliance - High Performance Computing (Projects JICS6C and AMYLOID-MSM) on the supercomputer Jureca at Forschungszentrum Jülich.

1. S. H. Barage, K. D. Sonawane, Amyloid cascade hypothesis: Pathogenesis and therapeutic strategies in Alzheimer's disease. *Neuropeptides* **52**, 1–18 (2015).
2. C. A. McLean *et al*., Soluble pool of Abeta amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. *Ann. Neurol.* **46**, 860–866 (1999).
3. M. D. Kirkitadze, G. Bitan, D. B. Teplow, Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: The emerging role of oligomeric assemblies. *J. Neurosci. Res.* **69**, 567–577 (2002).
4. K. Broersen, F. Rousseau, J. Schymkowitz, The culprit behind amyloid beta peptide related neurotoxicity in Alzheimer's disease: Oligomer size or conformation? *Alzheimers Res. Ther.* **2**, 12 (2010).
5. A. Müller-Schiffmann *et al*., Amyloid-β dimers in the absence of plaque pathology impair learning and synaptic plasticity. *Brain* **139**, 509–525 (2016).
6. G. Brinkmalm *et al*., Identification of neurotoxic cross-linked amyloid-β dimers in the Alzheimer's brain. *Brain* **142**, 1441–1457 (2019).
7. G. Bitan, D. B. Teplow, Rapid photochemical cross-linking—A new tool for studies of metastable, amyloidogenic protein assemblies. *Acc. Chem. Res.* **37**, 357–364 (2004).
8. K. Ono, M. M. Condron, D. B. Teplow, Structure-neurotoxicity relationships of amyloid beta-protein oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14745–14750 (2009).
9. B. O'Nuallain *et al*., Amyloid beta-protein dimers rapidly form stable synaptotoxic protofibrils. *J. Neurosci.* **30**, 14411–14419 (2010).
10. A. Vázquez de la Torre *et al*., Direct evidence of the presence of cross-linked a β dimers in the brains of Alzheimer's disease patients. *Anal. Chem.* **90**, 4552–4560 (2018).
11. F. Castello *et al*., Two-step amyloid aggregation: Sequential lag phase intermediates. *Sci. Rep.* **7**, 40065 (2017).
12. J. Yang *et al*., Direct observation of oligomerization by single molecule fluorescence reveals a multistep aggregation mechanism for the yeast prion protein ure2. *J. Am. Chem. Soc.* **140**, 2493–2503 (2018).
13. J. Nasica-Labouze *et al*., Amyloid β protein and Alzheimer's disease: When computer simulations complement experimental studies. *Chem. Rev.* **115**, 3518–3563 (2015).
14. P. H. Nguyen *et al*., Amyloid oligomers: A joint experimental/computational perspective on Alzheimer's disease, Parkinson's disease, type II diabetes, and amyotrophic lateral sclerosis. *Chem. Rev.* **121**, 2545–2647 (2021).
15. S. J. C. Lee, E. Nam, H. J. Lee, M. G. Savelieff, M. H. Lim, Towards an understanding of amyloid-β oligomers: Characterization, toxicity mechanisms, and inhibitors. *Chem. Soc. Rev.* **46**, 310–323 (2017).
16. M. C. Owen *et al*., Effects of in vivo conditions on amyloid aggregation. *Chem. Soc. Rev.* **48**, 3946–3996 (2019).
17. Z. Korade, A. K. Kenworthy, Lipid rafts, cholesterol, and the brain. *Neuropharmacology* **55**, 1265–1273 (2008).
18. E. Posse de Chaves, S. Sipione, Sphingolipids and gangliosides of the nervous system in membrane function and dysfunction. *FEBS Lett.* **584**, 1748–1759 (2010).
19. Y. C. Kao, P. C. Ho, Y. K. Tu, I. M. Jou, K. J. Tsai, Lipids and Alzheimer's disease. *Int. J. Mol. Sci.* **21**, 1505 (2020).
20. A. Paul, S. Samantray, M. Anteghini, M. Khaled, B. Strodel, Thermodynamics and kinetics of the amyloid-β peptide revealed by Markov state models based on MD data in agreement with experiment. *Chem. Sci. (Camb.)* **12**, 6652–6669 (2021).
21. S. Samantray, F. Yin, B. Kav, B. Strodel, Different force fields give rise to different amyloid aggregation pathways in molecular dynamics simulations. *J. Chem. Inf. Model.* **60**, 6462–6475 (2020).
22. B. Strodel, Amyloid aggregation simulations: Challenges, advances and perspectives. *Curr. Opin. Struct. Biol.* **67**, 145–152 (2021).
23. H. Koldsø, D. Shorthouse, J. Hélie, M. S. Sansom, Lipid clustering correlates with membrane curvature as revealed by molecular simulations of complex lipid bilayers. *PLOS Comput. Biol.* **10**, e1003911 (2014).
24. H. I. Ingólfsson *et al*., Computational lipidomics of the neuronal plasma membrane. *Biophys. J.* **113**, 2271–2280 (2017).

25. B. Barz, D. J. Wales, B. Strodel, A kinetic approach to the sequence-aggregation relationship in disease-related protein assembly. *J. Phys. Chem. B* **118**, 1003–1011 (2014).
26. B. Barz, Q. Liao, B. Strodel, Pathways of amyloid-β aggregation depend on oligomer shape. *J. Am. Chem. Soc.* **140**, 319–327 (2018).
27. A. M. Illig, B. Strodel, Performance of Markov state models and transition networks on characterizing amyloid aggregation pathways from MD data. *J. Chem. Theory Comput.* **16**, 7825–7839 (2020).
28. J. P. Jämbeck, A. P. Lyubartsev, An extension and further validation of an all-atomistic force field for biological membranes. *J. Chem. Theory Comput.* **8**, 2938–2948 (2012).
29. I. Ermilova, A. P. Lyubartsev, Extension of the slipids force field to polyunsaturated lipids. *J. Phys. Chem. B* **120**, 12826–12842 (2016).
30. J. Egawa, M. L. Pearn, B. P. Lemkuil, P. M. Patel, B. P. Head, Membrane lipid rafts and neurobiology: Age-related changes in membrane lipids and loss of neuronal function. *J. Physiol.* **594**, 4565–4579 (2016).
31. V. Pata, N. Dan, Effect of membrane characteristics on phase separation and domain formation in cholesterol-lipid mixtures. *Biophys. J.* **88**, 916–924 (2005).
32. H. Ohvo-Rekilä, B. Ramstedt, P. Leppimäki, J. P. Slotte, Cholesterol interactions with phospholipids in membranes. *Prog. Lipid Res.* **41**, 66–97 (2002).
33. M. Coles, W. Bicknell, A. A. Watson, D. P. Fairlie, D. J. Craik, Solution structure of amyloid beta-peptide(1-40) in a water-micelle environment. Is the membrane-spanning domain where we think it is? *Biochemistry* **37**, 11064–11077 (1998).
34. N. Miyashita, J. E. Straub, D. Thirumalai, Structures of β−amyloid peptide 1–40, 1–42, and 1–55 – the 672–726 fragment of app – in a membrane environment with implications for interactions with γ-secretase. *J. Am. Chem. Soc.* **131**, 17843–17852 (2009).
35. B. Strodel, J. W. L. Lee, C. S. Whittleston, D. J. Wales, Transmembrane structures for Alzheimer's Aβ(1-42) oligomers. *J. Am. Chem. Soc.* **132**, 13300–13312 (2010).
36. L. Nagel-Steger, M. C. Owen, B. Strodel, An account of amyloid oligomers: Facts and figures obtained from experiments and simulations. *ChemBioChem* **17**, 657–676 (2016).
37. T. Kakeshpour *et al*., A lowly populated, transient β-sheet structure in monomeric aβ1-42 identified by multinuclear NMR of chemical denaturation. *Biophys. Chem.* **270**, 106531 (2020).
38. B. Urbanc *et al*., Molecular dynamics simulation of amyloid beta dimer formation. *Biophys. J.* **87**, 2310–2321 (2004).
39. S. Côté, R. Laghaei, P. Derreumaux, N. Mousseau, Distinct dimerization for various alloforms of the amyloid-beta protein: Aβ1–40, β1–42, and β1–40(d23n). *J. Phys. Chem. B* **116**, 4043–4055 (2012).
40. V. H. Man, P. H. Nguyen, P. Derreumaux, High-resolution structures of the amyloid-β–42 dimers from the comparison of four atomistic force fields. *J. Phys. Chem. B* **121**, 5977–5987 (2017).
41. B. Mehrazma, A. Rauk, Exploring amyloid-β dimer structure using molecular dynamics simulations. *J. Phys. Chem. A* **123**, 4658–4670 (2019).
42. S. Rauscher *et al*., Structural ensembles of intrinsically disordered proteins depend strongly on force field: A comparison to experiment. *J. Chem. Theory Comput.* **11**, 5513–5524 (2015).
43. T. Lührs *et al*., 3D structure of Alzheimer's amyloid-β (1-42) fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 17342–17347 (2005).
44. A. K. Paravastu, R. D. Leapman, W. M. Yau, R. Tycko, Molecular structural basis for polymorphism in Alzheimer's beta-amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 18349–18354 (2008).
45. B. Tarus, J. E. Straub, D. Thirumalai, Structures and free-energy landscapes of the wild type and mutants of the Aβ (21-30) peptide are determined by an interplay between intrapeptide electrostatic and hydrophobic interactions. *J. Mol. Biol.* **379**, 815–829 (2008).
46. L. Gremer *et al*., Fibril structure of amyloid-β(1-42) by cryo-electron microscopy. *Science* **358**, 116–119 (2017).
47. N. Rezaei-Ghaleh, G. Parigi, M. Zweckstetter, Reorientational dynamics of amyloid-β from NMR spin relaxation and molecular simulation. *J. Phys. Chem. Lett.* **10**, 3369–3375 (2019).
48. S. Hong *et al*., Soluble Aβ oligomers are rapidly sequestered from brain ISF in vivo and bind GM1 ganglioside on cellular membranes. *Neuron* **82**, 308–319 (2014).
49. N. Kedia, M. Almisry, J. Bieschke, Glucose directs amyloid-β into membrane-active oligomers. *Phys. Chem. Chem. Phys.* **19**, 18036–18046 (2017).
50. K. Ikeda, T. Yamaguchi, S. Fukunaga, M. Hoshino, K. Matsuzaki, Mechanism of amyloid β-protein aggregation mediated by GM1 ganglioside clusters. *Biochemistry* **50**, 6433–6440 (2011).
51. K Matsuzaki, Formation of toxic amyloid fibrils by amyloidβ protein on ganglioside clusters. *Int. J. Alzheimers Dis.* **2011**, 956104 (2011).
52. M. Cebecauer, M. Hof, M. Amaro, Impact of GM1 on membrane-mediated aggregation/oligomerization of β-amyloid: Unifying view. *Biophys. J.* **113**, 1194–1199 (2017).
53. T. V. Sokolova, I. O. Zakharova, V. V. Furaev, M. P. Rychkova, N. F. Avrova, Neuroprotective effect of ganglioside GM1 on the cytotoxic action of hydrogen peroxide and amyloid beta-peptide in PC12 cells. *Neurochem. Res.* **32**, 1302–1313 (2007).
54. F. Kreutz *et al*., Amyloid-β induced toxicity involves ganglioside expression and is sensitive to GM1 neuroprotective action. *Neurochem. Int.* **59**, 648–655 (2011).
55. R. Yang *et al*., Monosialoanglioside improves memory deficits and relieves oxidative stress in the hippocampus of rat model of Alzheimer's disease. *Neurol. Sci.* **34**, 1447–1451 (2013).
56. J. Lee *et al*., CHARMM-GUI input generator for NAMD, GROMACS, AMBER, openMM, and CHARMM/openMM simulations using the CHARMM36 additive force field. *J. Chem. Theory Comput.* **12**, 405–413 (2016).
57. M. Carballo-Pacheco, B. Strodel, Advances in the simulation of protein aggregation at the atomistic scale. *J. Phys. Chem. B* **120**, 2991–2999 (2016).
58. U. Sengupta, M. Carballo-Pacheco, B. Strodel, Automated Markov state models for molecular dynamics simulations of aggregation and self-assembly. *J. Chem. Phys.* **150**, 115101 (2019).
59. M. J. Abraham *et al*., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015).
60. J. Huang *et al*., CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **14**, 71–73 (2017).
61. J. B. Klauda *et al*., Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *J. Phys. Chem. B* **114**, 7830–7843 (2010).
62. G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101 (2007).
63. H. J. Berendsen, J. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).
64. H. Antila *et al*., Headgroup structure and cation binding in phosphatidylserine lipid bilayers. *J. Phys. Chem. B* **123**, 9066–9079 (2019).
65. Y. Zhang, C. Sagui, Secondary structure assignment for conformationally irregular peptides: Comparison between DSSP, STRIDE and KAKSI. *J. Mol. Graph. Model.* **55**, 72–84 (2015).
66. C. Möckel *et al*., Integrated NMR, fluorescence, and molecular dynamics benchmark study of protein mechanics and hydrodynamics. *J. Phys. Chem. B* **123**, 1453–1480 (2019).
67. M. Bastian *et al*., Gephi: An open source software for exploring and manipulating networks. *Proc. Third Int. ICWSM Conf.* **8**, 361–362 (2009).
68. W. Humphrey, A. Dalke, K. Schulten, VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38 (1996).

PNAS
www.pnas.org

Supplementary Information for

Amyloid-β peptide dimers undergo a random coil to β-sheet transition in the aqueous phase but not at the neuronal membran

Hebah Fatafta, Mohammed Khaled, Michael C. Owen, Abdallah Sayyed-Ahmad, Birgit Strodel

Corresponding Author name: Birgit Strodel.
E-mail: b.strodel@fz-juelich.de

This PDF file includes:

Figs. S1 to S11 (not allowed for Brief Reports)

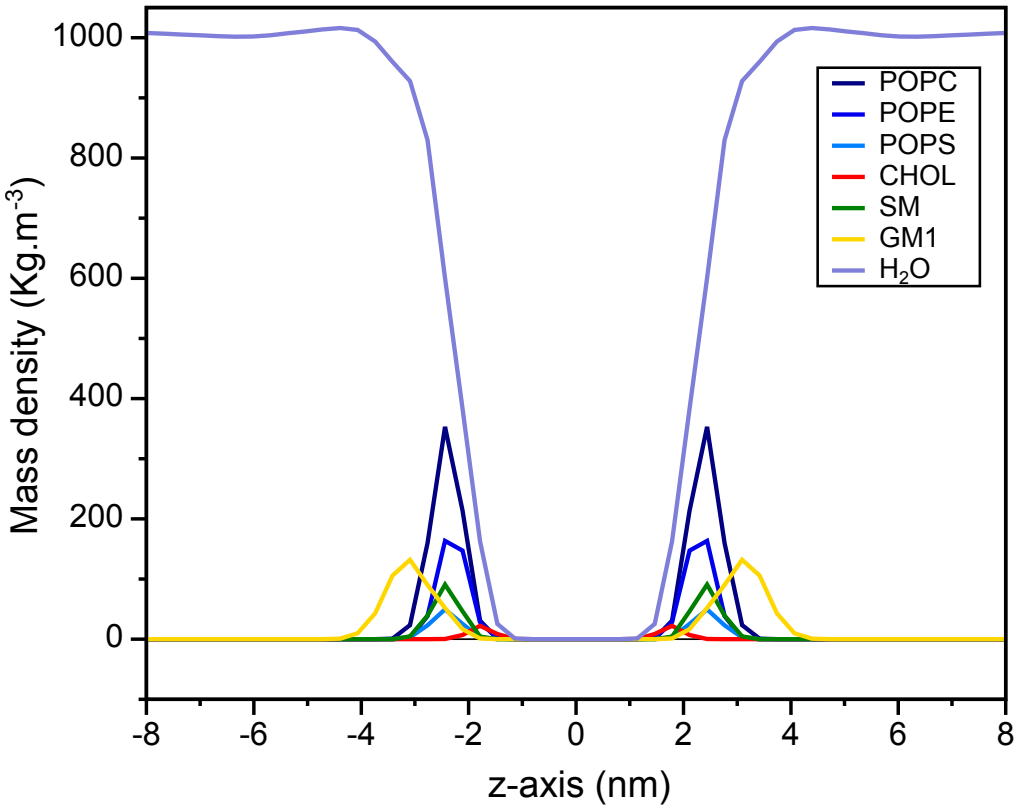

Fig. S1. The average mass density profile of lipids and water along the z-direction corresponding to the bilayer normal. Colors are chosen according to the legend given on the right.

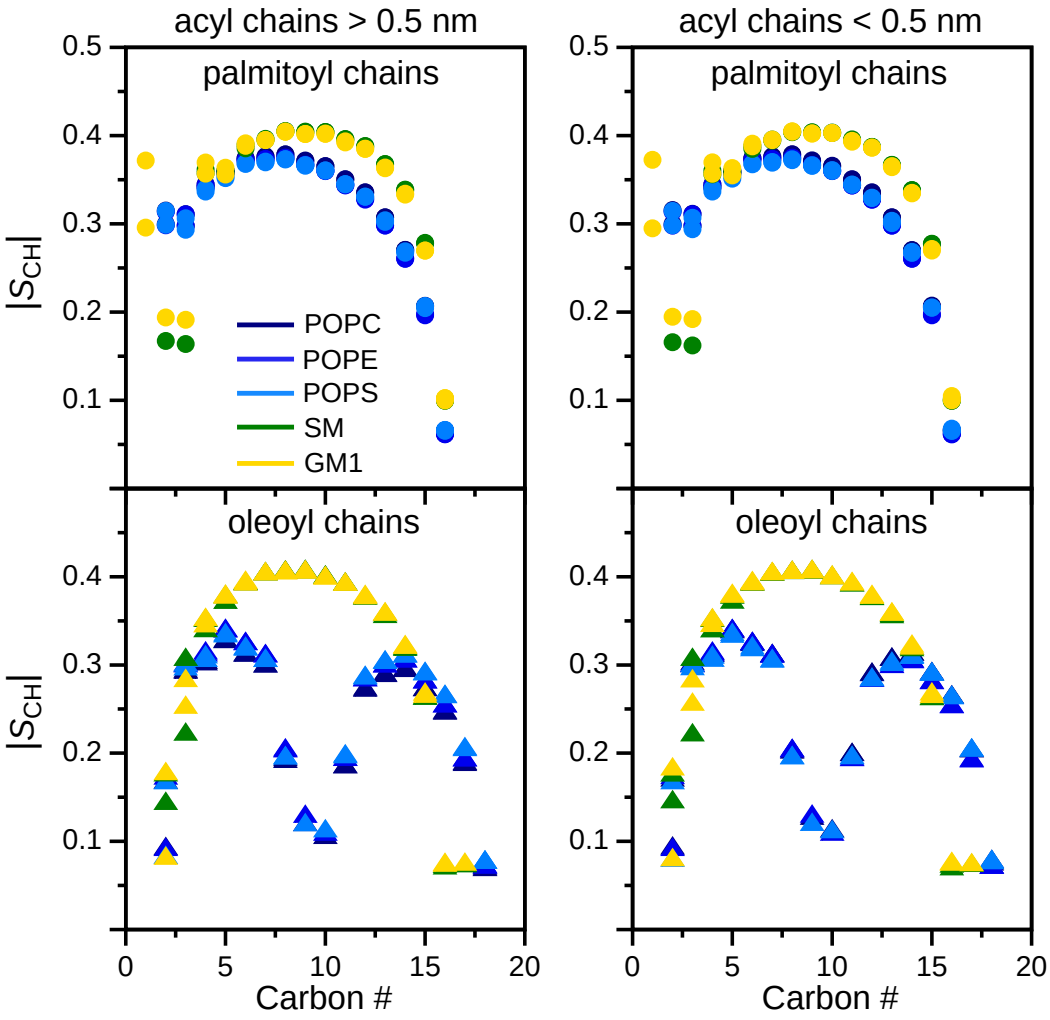

Fig. S2. Average order parameters of the acyl chains (top: palmitoyl chains; bottom: oleoyl chains) of each lipid component of the neuronal membrane, distinguishing between lipids that are more than 0.5 nm away from Aβ42 (left) and those within 0.5 nm of the peptide (right).

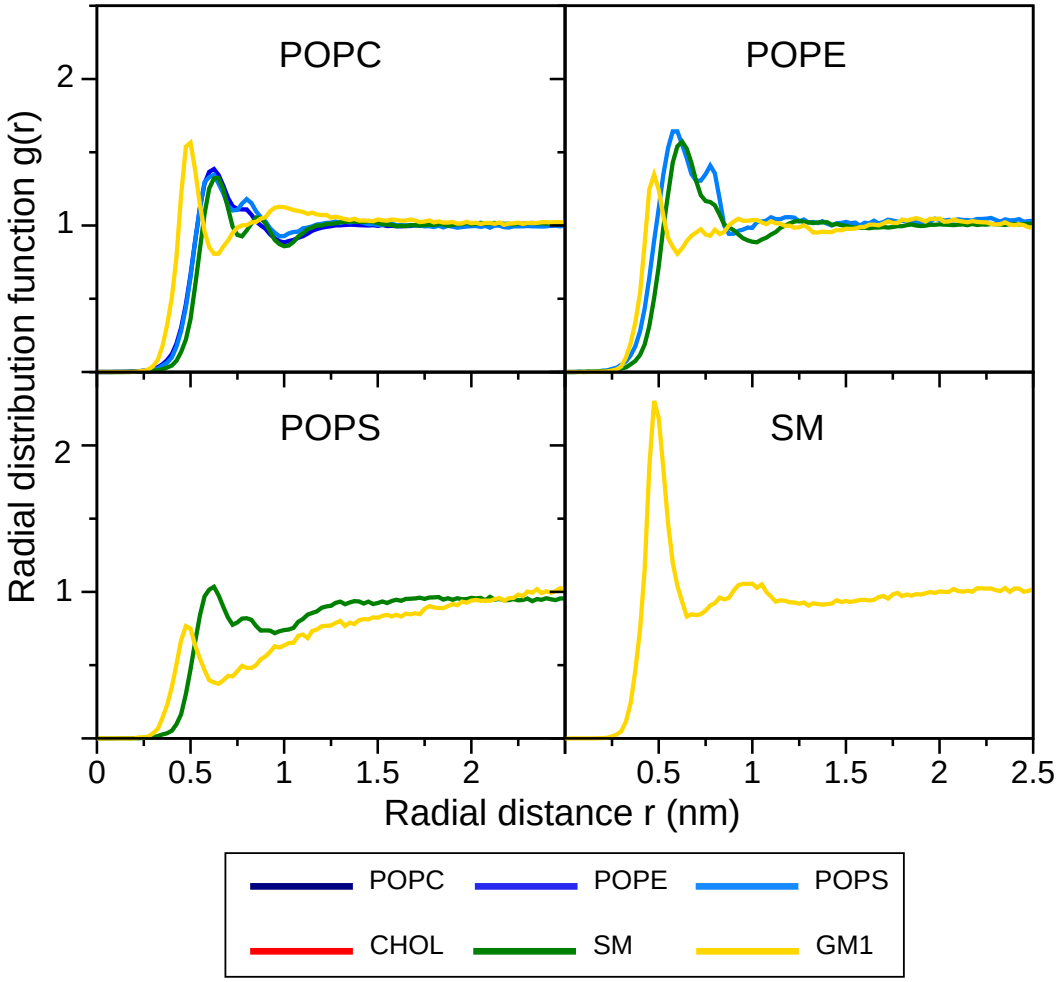

Fig. S3. Radial distribution functions for mixed lipid pairings: (top left) POPC–lipid pairs; (to pright) POPE–lipid pairs; (bottom left) POPS–lipid pairs; (bottom right) SM–lipid pairings). Pairings between identical lipid types are shown in the main text in Figure 2A and pairing involving CHOL are presented in Figure 2B. The P atoms of POPC, POPE, POPS, and SM and the O atoms of CHOL and GM1 were used for the RDF calculations. The x-axes show the distances between the respective atom pairs. The colors of the graphs refer to the lipids as indicated in the color key below the plots. Pairs with RDF > 1 are considered to cluster.

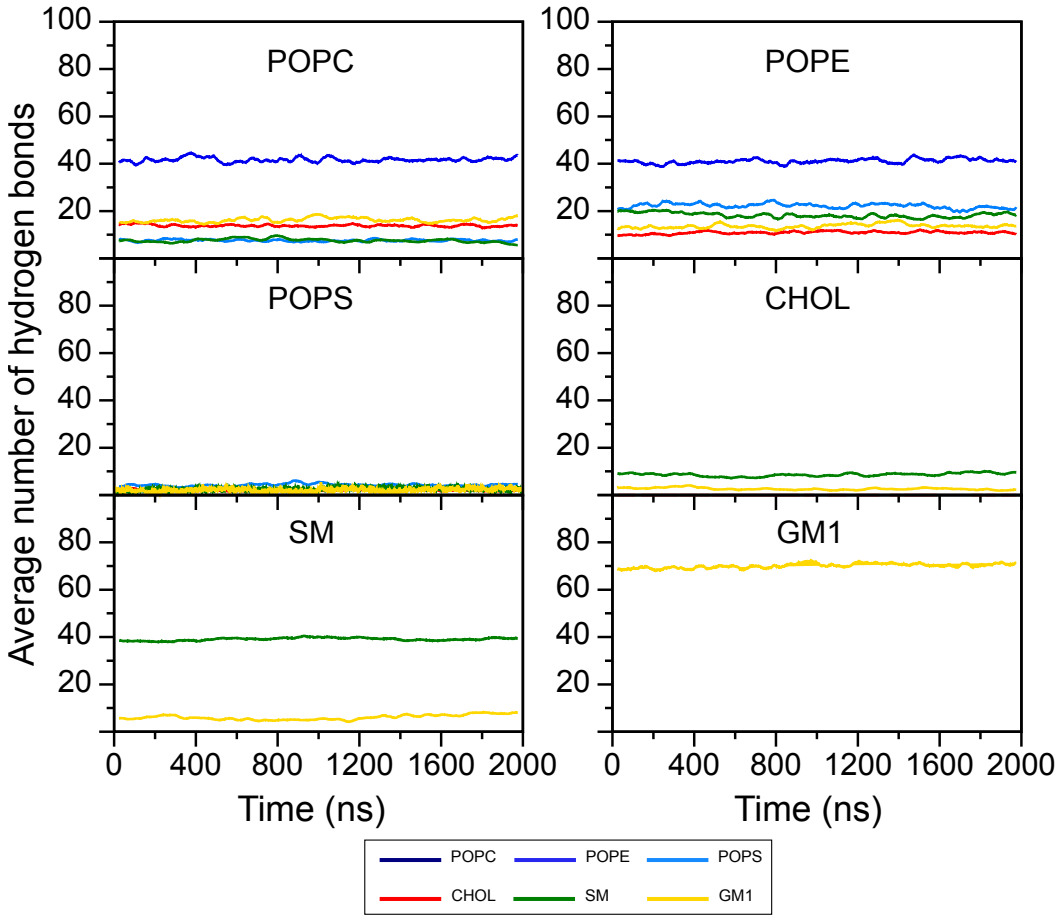

Fig. S4. The average number of hydrogen bonds between different lipid pairs. The colors of the graphs refer to the lipids as indicated in the color key below the plots.

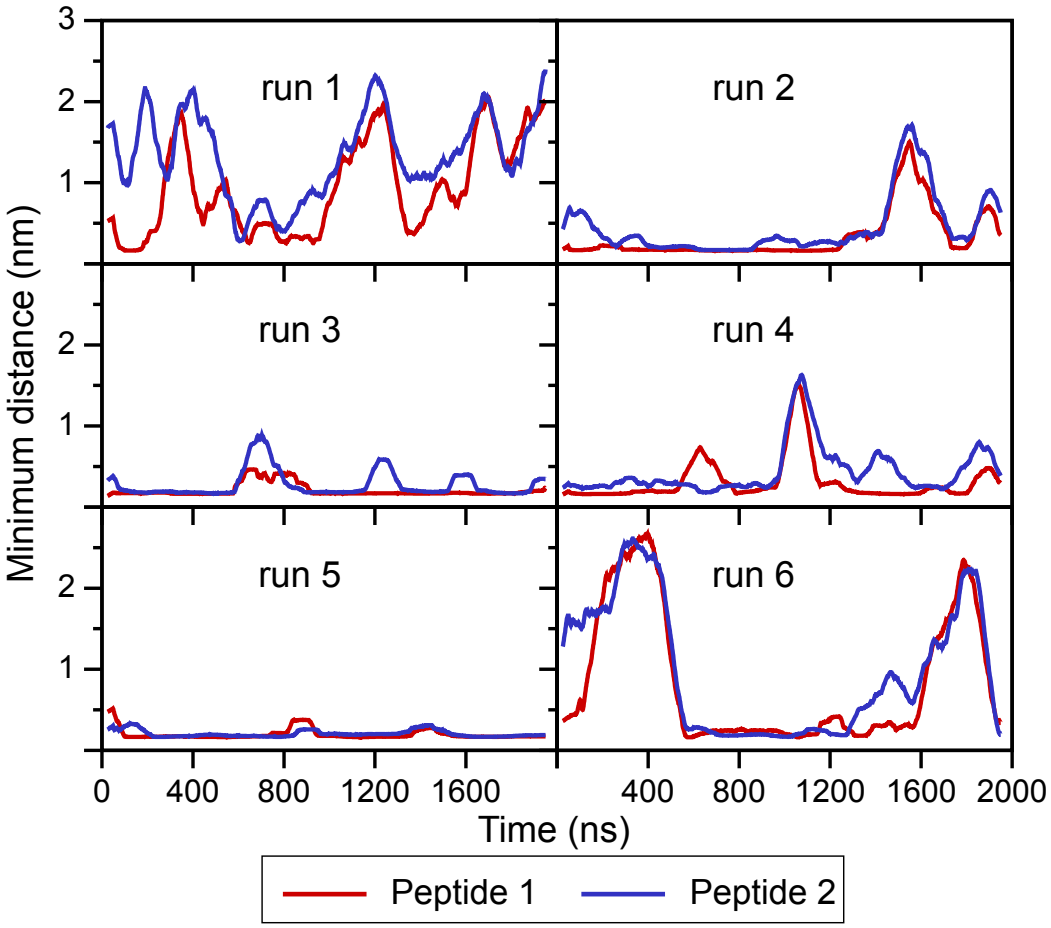

Fig. S5. The average number of Aβ42–lipid contacts (and standard error of the mean) calculated for peptide 1 (left) and peptide 2 (right) with each of the components of the neuronal membrane (lipid names shown above the panels).

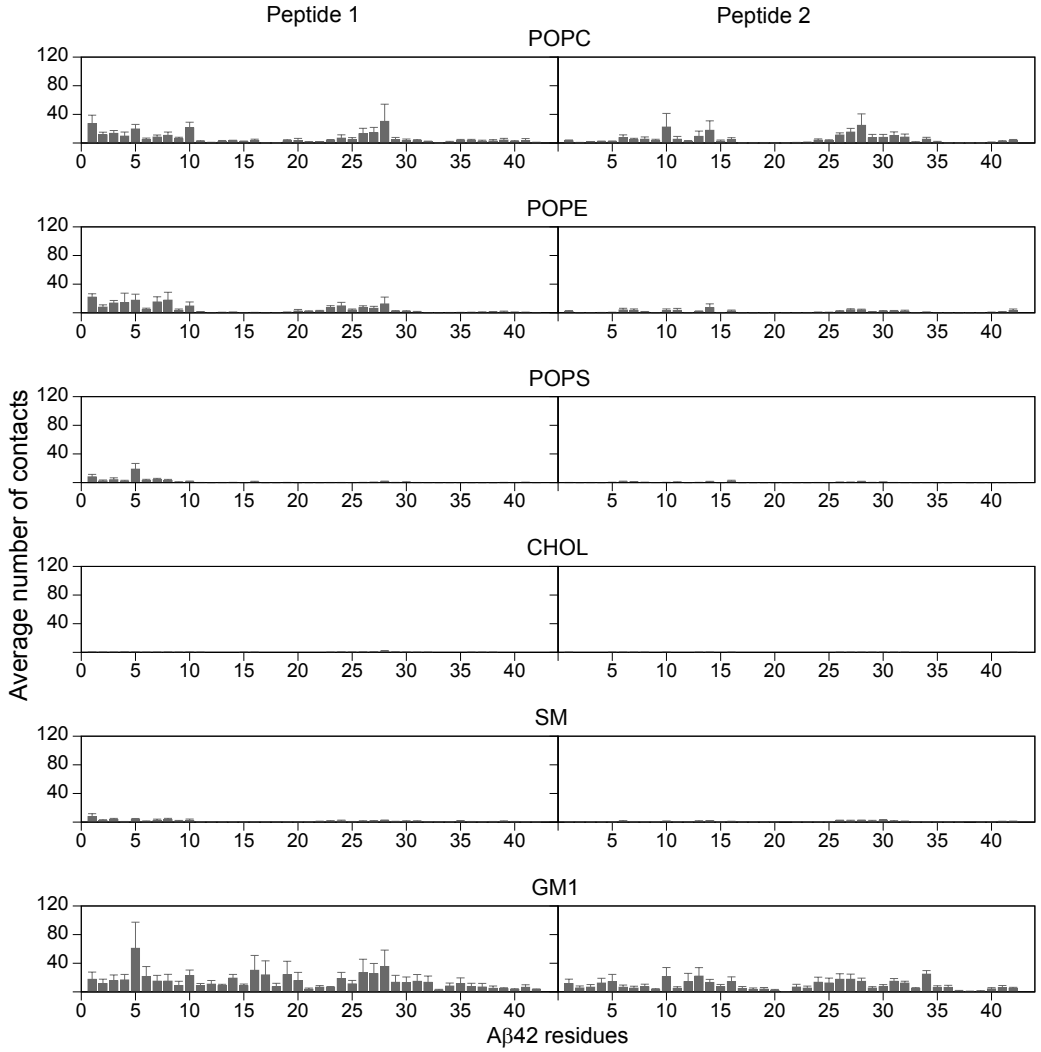

Fig. S6. The average number of Aβ42–lipid contacts (and standard error of the mean) calculated for peptide 1 (left) and peptide 2 (right) with each of the components of the neuronal membrane (lipid names shown above the panels).

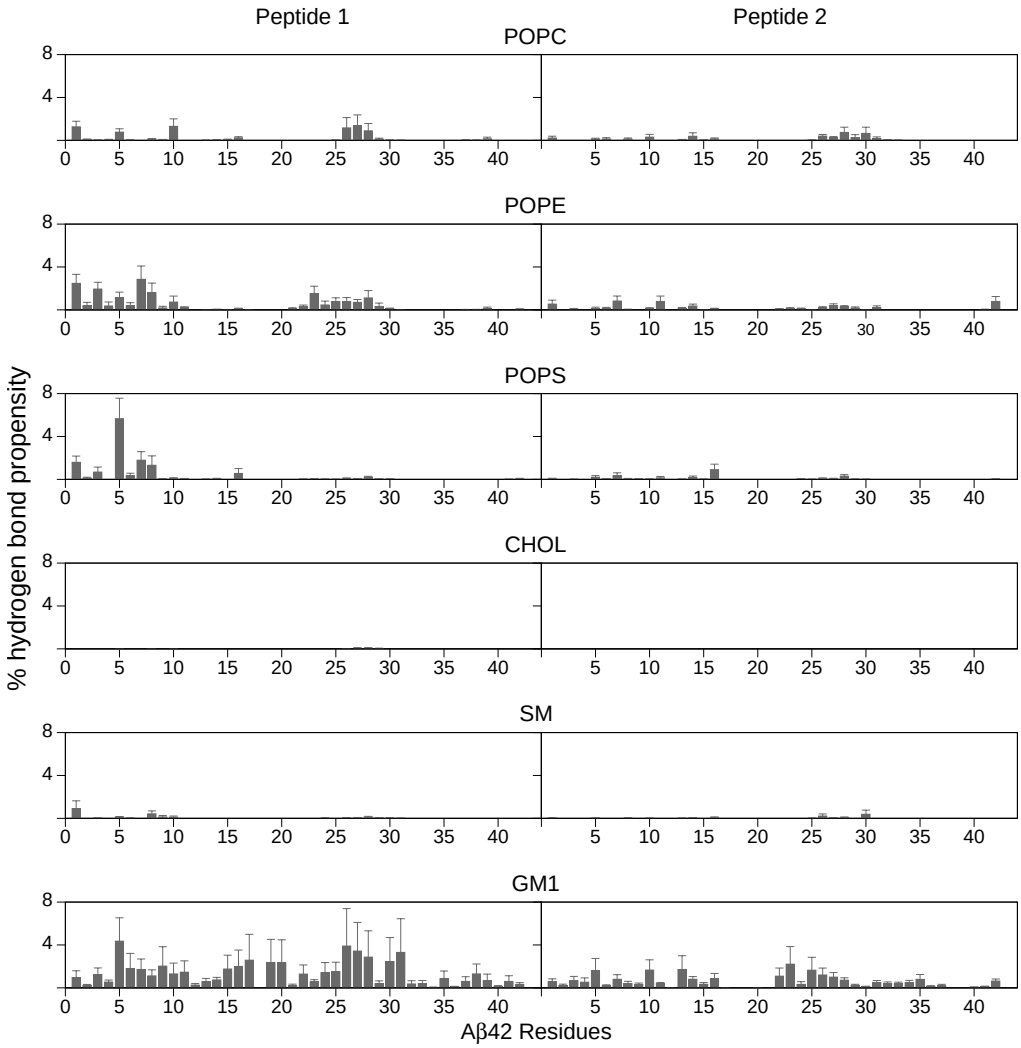

Fig. S7. The average hydrogen bond propensity between Aβ42 and lipids (and standard error of the mean) calculated for peptide 1 (left) and peptide 2 (right) with each of the components of the neuronal membrane (lipid names shown above the panels).

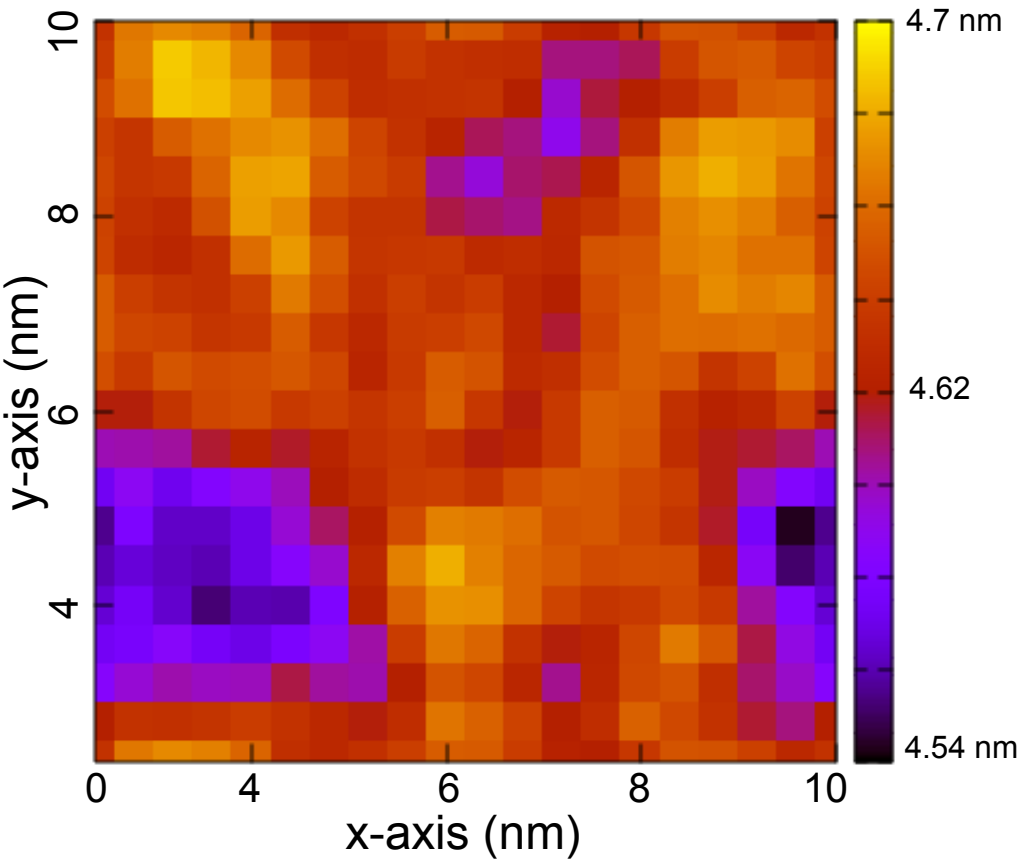

Fig. S8. Average bilayer thickness calculated when the protein is within 0.5 nm of the membrane. The x and the y-axes represent the unit cell dimension in nm. The color bar shows the thickness range in nm.

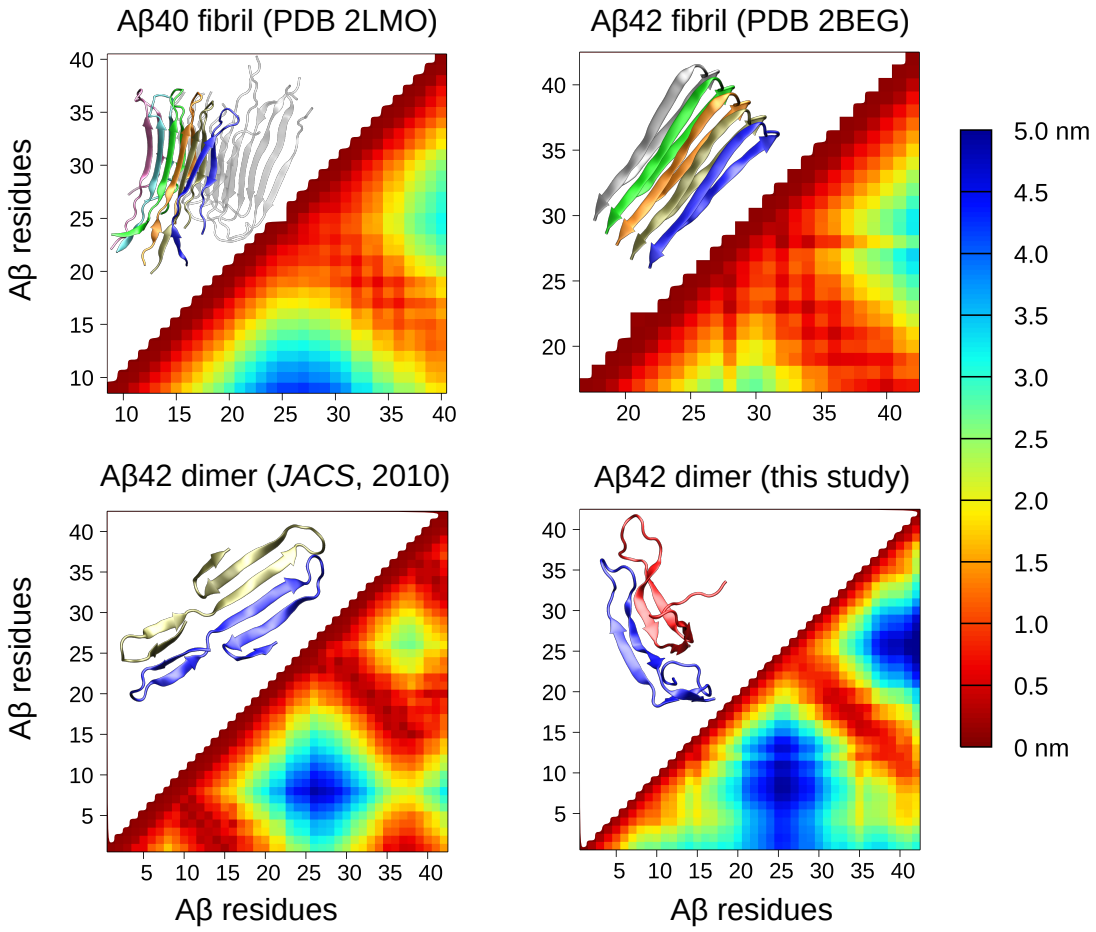

Fig. S9. Comparison of the intrapeptide contacts present in Aβ fibrils (top) and Aβ dimers (bottom). The contacts are provided as minimal distances between the residues of Aβ. For the fibrils, the structures from PDB entry 2LMO for Aβ40 (ref. 44 in the main text) and PDB entry 2BEG for Aβ42 (ref. 43 in the main text) were chosen. Since these fibril models only start with residue 9 for Aβ40 and residue 17 for Aβ42, the distance presentations are limited to residues G9–V40 and L17–A42, respectively. The corresponding fibril structure is shown as cartoon in the upper left triangle of either plot. The average intrapeptide distance matrix of the Aβ42 dimer structures obtained from the MD simulations in solution performed in the current study are shown in the lower right panel. As comparison, the distance matrix for one of the Aβ42 dimer structures obtained from a structure prediction approach for transmembrane Aβ42 (ref. 46 in the main text) is shown in the lower left panel. Representative dimer structures are shown in both panels. The contacts present in the fibrils and those in the dimers obtained here are very similar for residue numbers $\gtrsim 20$ and are dominated by a β-hairpin centered at G25/S26. Only in the Aβ42 fibril (upper right panel), residue D23 forms more contacts with other residues in its vicinity, as it is pointing towards the interior of the fibril in order to enable salt bridge formation with K28. The β-hairpin centered at G25/S26 is also present in the Aβ42 dimer model determined by structure prediction for transmembrane Aβ42 and is accompanied by two further β-hairpins: an N-terminal one centered at G9/Y10 and a C-terminal one at G37/G38 (lower left panel). These two hairpins are only weakly visible in the distance map for the dimers sampled in the current study (lower right panel).

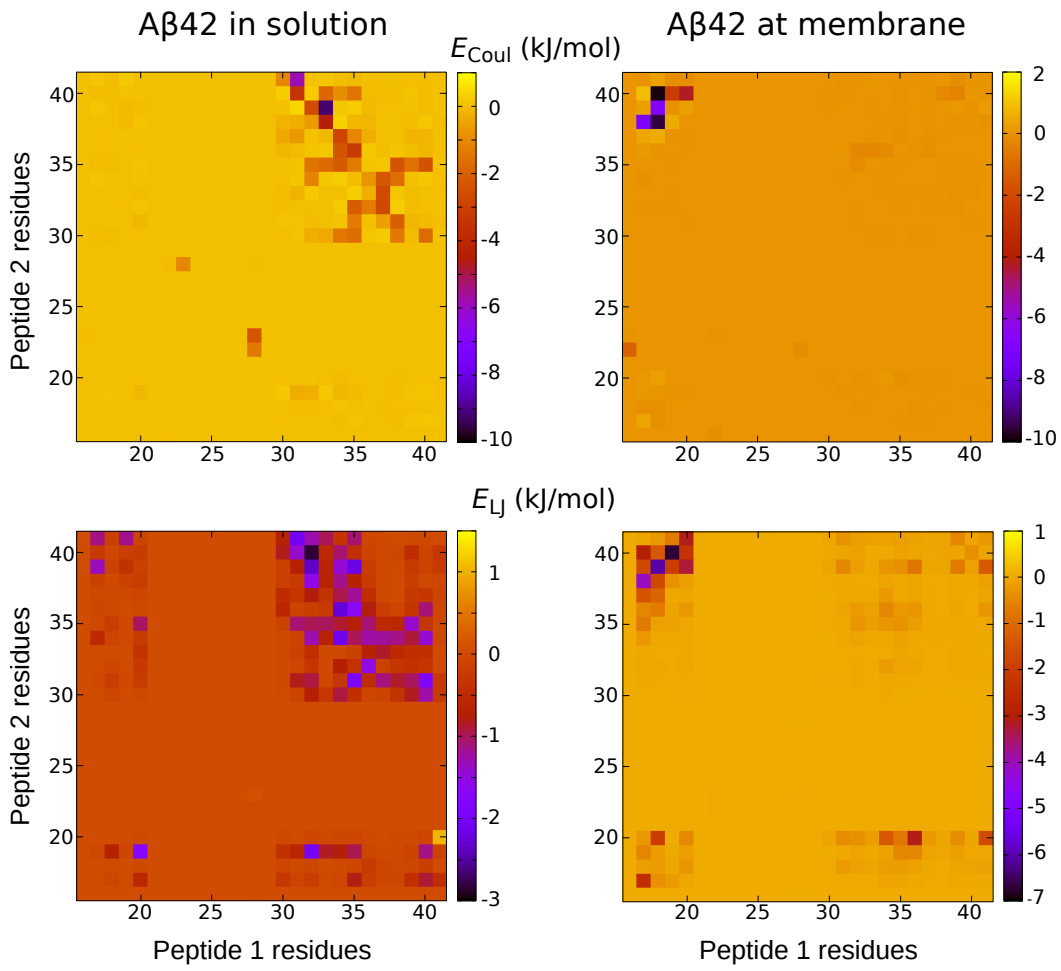

Fig. S10. Average interpeptide interaction energies between residues, decomposed into Coulomb (top) and Lennard-Jones (bottom) interactions, for Aβ42 dimers in solution (left) and at the membrane (right). The peptide association in solution is dominated by interactions between the C-terminal regions of both peptides, while the peptides of the membrane-bound dimers mainly interact with each other via the central hydrophobic core of peptide 1 and the C-terminal residues of peptide 2. Please note that results are only shown for residue numbers > 15, as for the N-terminal residues no noteworthy interaction energies were recorded. Moreover, the energy scales in the four plots are slightly different from each other (see color bars) in order to provide a good resolution of the energies.

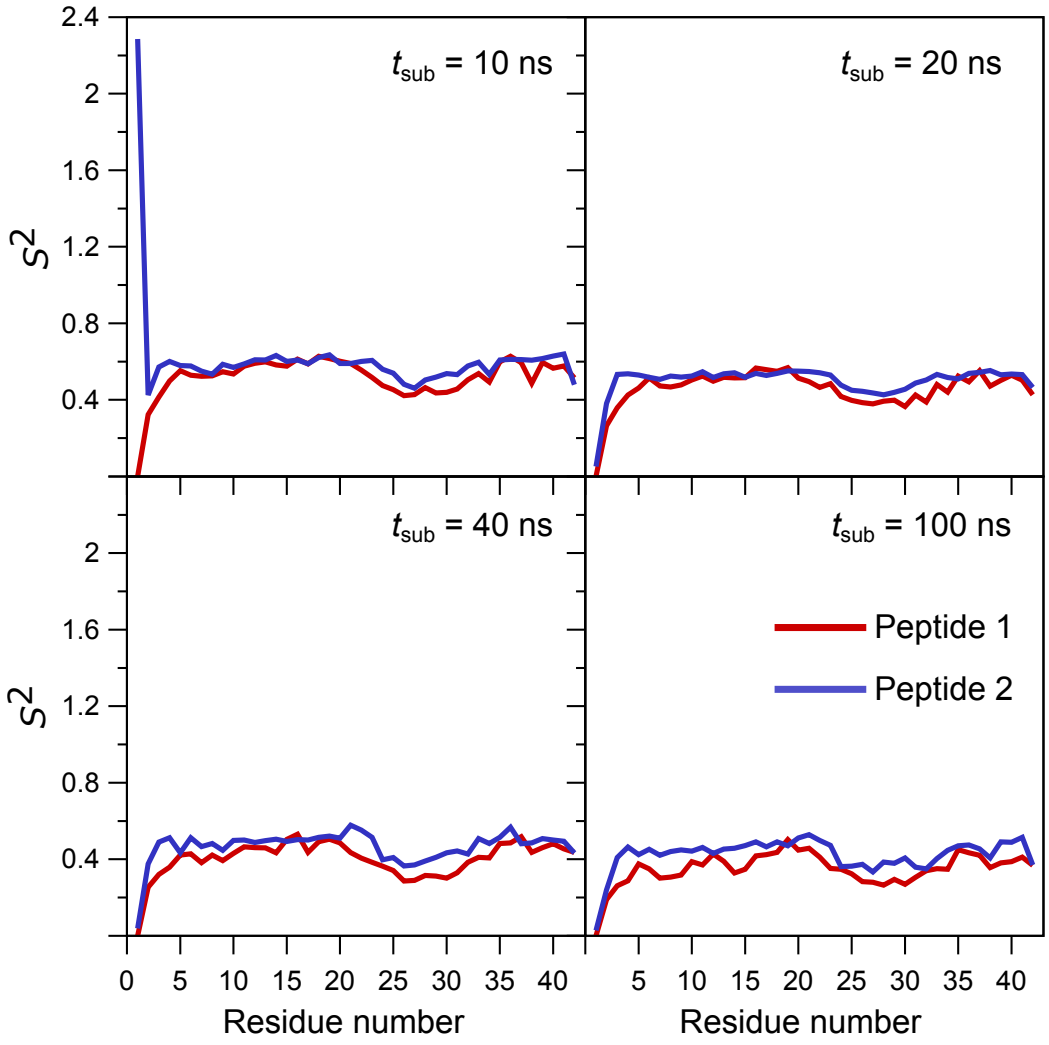

Fig. S11. The average order parameter S^2 of each residue and peptide in the Aβ42 dimer at the neuronal membrane. Results are shown for increasing lengths of the subtrajectories, t_{sub}, used for calculating and fitting the N–H bond vector autocorrelation functions. A sufficient level of convergence is reached for $t_{\mathrm{sub}} \geq 40$ ns.

A.4. Publication III

Comparative molecular dynamics simulation studies of pathogenic and non-pathogenic Huntingtin protein monomer and dimer

Khaled, M., Strodel, B., Sayyed-Ahmad, A.(2023).

Frontiers in Molecular Biosciences, Volume 10:1143353.

frontiers | Frontiers in Molecular Biosciences

TYPE Original Research
PUBLISHED 10 April 2023
DOI 10.3389/fmolb.2023.1143353

OPEN ACCESS

EDITED BY
Alfonso De Simone,
University of Naples Federico II, Italy

REVIEWED BY
Phuong Nguyen,
UPR9080 Laboratoire de Biochimie Théorique (LBT), France
Davide Mercadante,
The University of Auckland, New Zealand

*CORRESPONDENCE
Birgit Strodel,
b.strodel@fz-juelich.de
Abdallah Sayyed-Ahmad,
asayyeda@birzeit.edu

SPECIALTY SECTION
This article was submitted to Protein Folding, Misfolding and Degradation, a section of the journal Frontiers in Molecular Biosciences

RECEIVED 12 January 2023
ACCEPTED 29 March 2023
PUBLISHED 10 April 2023

CITATION
Khaled M, Strodel B and Sayyed-Ahmad A (2023), Comparative molecular dynamics simulations of pathogenic and non-pathogenic huntingtin protein monomers and dimers.
Front. Mol. Biosci. 10:1143353.
doi: 10.3389/fmolb.2023.1143353

Comparative molecular dynamics simulations of pathogenic and non-pathogenic huntingtin protein monomers and dimers

Mohammed Khaled[1], Birgit Strodel[1,2]* and Abdallah Sayyed-Ahmad[3]*

[1]Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich, Jülich, Germany, [2]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf, Germany, [3]Department of Physics, Birzeit University, Birzeit, Palestine

Polyglutamine expansion at the N-terminus of the huntingtin protein exon 1 (Htt-ex1) is closely associated with a number of neurodegenerative diseases, which result from the aggregation of the increased polyQ repeat. However, the underlying structures and aggregation mechanism are still poorly understood. We performed microsecond-long all-atom molecular dynamics simulations to study the folding and dimerization of Htt-ex1 (about 100 residues) with non-pathogenic and pathogenic polyQ lengths, and uncovered substantial differences. The non-pathogenic monomer adopts a long α-helix that includes most of the polyQ residues, which forms the interaction interface for dimerization, and a PPII-turn-PPII motif in the proline-rich region. In the pathogenic monomer, the polyQ region is disordered, leading to compact structures with many intra-protein interactions and the formation of short β-sheets. Dimerization can proceed *via* different modes, where those involving the N-terminal headpiece bury more hydrophobic residues and are thus more stable. Moreover, in the pathogenic Htt-ex1 dimers the proline-rich region interacts with the polyQ region, which slows the formation of β-sheets.

KEYWORDS

polyglutamine, huntingtin, molecular dynamics, oligomer, aggregation

1 Introduction

Huntington's disease (HD) is an inherited neurodegenerative disease caused by an abnormal expansion in the polyglutamine (polyQ) tract of the N-terminal Huntingtin (Htt-ex1) protein. The elongated polyQ tract mutation is caused by the expansions of nucleotide CAG repeats in exon-1 of the HD gene that encodes the elongated polyQ tract within the Htt-ex1 protein. Furthermore, the expansion of polyQ tracts and their toxicity are correlated to the age of the onset of HD (MacDonald et al., 1993). In HD, Htt-ex1 becomes pathogenic beyond a threshold of 36 glutamine repeats (Takeuchi and Nagai, 2017). The full-length Htt-ex1 is over 3,100 amino acids expressed in all mammalian cells including nerve cells in the brain where it has higher concentrations. Htt-ex1 interacts with a wide range of proteins involved in many cellular processes (Gusella and MacDonald, 2000; Li and Li, 2004; Saudou and Humbert, 2016), but the exact structure and function of Htt-ex1 are still poorly understood (Mangiarini et al., 1996). Nevertheless, HD onset and progression are associated with the misfolding of Htt-ex1 which eventually forms amyloid aggregates

Htt-Q_{23} sequence

Nt_{17}	$polyQ_{23}$	$polyP_{11}$	PRD	$polyP_{10}$	
MATLEKLMKAFESLKSF	-Q23-	-P11-	QLPQPPPQAQPLLPQPQ	-P10-	GPAVAEEPLHRP
17	40	51	68	78	90

Htt-Q_{48} sequence

Nt_{17}	$polyQ_{48}$	$polyP_{11}$	PRD	$polyP_{10}$	
MATLEKLMKAFESLKSF	-Q48-	-P11-	QLPQPPPQAQPLLPQPQ	-P10-	GPAVAEEPLHRP
17	65	76	93	103	115

FIGURE 1
Sequence of Htt-Q_{23} and Htt-Q_{48} studied in this work. The Nt_{17} region is highlighted in red, polyQ in blue, and the PRD is shown in yellow for $polyP_{11}$ and $polyP_{10}$ and rose for the other PRD parts.

that are connected to neuronal cell death (Perutz et al., 1994). Toxic aggregation of Htt-ex1 into amyloid cannot only be observed *in vivo*, but can also be reproduced *in vitro* (Nagai et al., 2007). Several hypotheses have been suggested to explain the aggregation behavior and thereby the toxicity of Htt-ex1. One of these hypotheses proposes that pathogenic Htt-ex1 accumulates into insoluble aggregates in neurons as amyloid fibrillar structures (Miller et al., 2011). Other hypotheses suggest that Htt-ex1 monomers or oligomers with extended polyQ tracts interact with other cellular proteins and alter their functions, which leads to neuronal cell death (Williams and Paulson, 2008; Gkekas et al., 2021).

Importantly, the expansion of the polyQ tract is also associated with many other inherited neurodegenerative diseases (Williams and Paulson, 2008). Thus, several studies have been conducted to investigate the structures and aggregation mechanisms of isolated polyQ peptides. Experiments under different solutions conditions found that isolated polyQ tracts could sample various conformations as collapsed structures (Crick et al., 2006) with random coils (Chen et al., 2001; Klein et al., 2007), α-helix (Bhattacharyya et al., 2006), β-sheets (Nagai et al., 2007; Darnell et al., 2009) and PPII helix (Chellgren et al., 2006; Darnell et al., 2007; Darnell et al., 2009). The conformational flexibility of polyQ tracts has also been confirmed by computational modeling using molecular dynamics (MD) simulations, as was comprehensively reviewed by Moldovean and Chis (Moldovean and Chiș, 2019). Of particular note are the studies (Długosz and Trylska, 2011; Vöpel et al., 2017; Priya and Gromiha, 2019) as they did not only simulate isolated polyQ stretches as done in the majority of the other simulation studies, but included the first 17 N-terminal residues directly preceding the polyQ sequence and some of the following C-terminal residues. Despite the various experimental and simulation efforts, there is no consensus yet on the preferred polyQ structure in solution. Almost all secondary structures have been suggested, ranging from α-helical structures (Elena-Real et al., 2022) to coil and β-sheets (Moldovean and Chiș, 2019). Nonetheless, there is ample evidence that indicates that longer polyQ increases the β-sheet propensity and aggregation rates (Thakur and Wetzel, 2002; Klein et al., 2007; Sivanandam et al., 2011).

The Htt-ex1 consists of three regions (Figure 1): the N-terminal 17 amino acid region (Nt_{17}), the polyQ tract region, followed by a proline-rich region (PRD). The N-terminal and proline-rich regions surrounding the polyQ region have critical roles in modulating the aggregation mechanism as has been illustrated in computational and experimental studies (Chow et al., 2012; Qin et al., 2004; Duennwald et al., 2006a; Duennwald et al., 2006b). The Nt_{17} region is known to accelerate the aggregation by forming prefibrillar spherical oligomers with Nt_{17} at their core (Tam et al., 2009; Thakur et al., 2009). Experimentally determined structures of Nt_{17} suggested that it folds into an amphipathic α-helix (Kim et al., 2009; Kim, 2013; Michalek et al., 2013). In addition, it has been suggested that the α-helix Nt_{17} may initialize the Htt-ex1 aggregation by forming α-helix rich oligomers (Hoop et al., 2014; Sahoo et al., 2014; Pandey et al., 2018). In general, it appears that Nt_{17} is mostly helical in Htt-ex1 aggregates (Sivanandam et al., 2011; Jayaraman et al., 2012; Hoop et al., 2014). However, nuclear magnetic resonance (NMR) spectroscopy showed that structures of Nt_{17} are intrinsically disordered and adopt different conformations (Thakur et al., 2009). The PRD, on the other hand, is found to decrease the stability and rate of amyloid-like aggregation without changing the fundamental mechanism of aggregation (Bhattacharyya et al., 2006). NMR and electron paramagnetic resonance (EPR) spectroscopy showed that the PRD tends to adopt similar structures with rich PPII helical structures in both monomers and fibril structures (Bugg et al., 2012; Isas et al., 2015).

Proteins with glutamine repeats are structurally unstable. This makes it difficult to resolve the structure of regions surrounding the polyQ tracts (Takeuchi and Nagai, 2017). As of today, only structures of the Htt-ex1 containing 17 glutamine repeats were resolved by X-ray crystallography (Kim et al., 2009). Thus, various computational studies of model Htt-ex1 have been conducted to illustrate the effect of polyQ tract length on protein structure (Moldovean and Chiș, 2019). Other studies provided insights into the effects of Htt-ex1 flanking domains on its structure (Długosz and Trylska, 2011; Vöpel et al., 2017; Priya and Gromiha, 2019). By adding or removing these domains, the PRD was found to destabilize the protein and inhibited β-sheet formation (Lakhani et al., 2010; Williamson et al., 2010). Kang and coworkers carried out MD simulations of full-length Htt-ex1 monomers with different polyQ lengths. Their results suggest a positive correlation between polyQ tract length and the β-sheet

content (Kang et al., 2017). Finally, the polyQ region was found to be the main driver for Htt-ex1 aggregation (Williamson et al., 2010).

In this work, we performed all-atom MD simulations to investigate the structural properties of Htt-Q_n monomers and dimers with n = 23 for non-pathogenic Htt-ex1 and n = 48 for disease-causing Htt-ex1. In addition to including the Nt_{17} region, we also added 50 amino acids to model the impact of the PRD on the polyQ region (Figure 1). With this, our simulated systems are longer than those usually simulated. In addition, we used the Charmm36m force field, which has been adapted to proteins with low-complexity sequences (Huang et al., 2017) and has been shown to be useful in predicting the conformational ensemble of intrinsically disordered proteins (Robustelli et al., 2018; Paul et al., 2021) and amyloid aggregates (Samantray et al., 2020), but has not yet been applied to Htt-ex1 proteins. Furthermore, we performed simulations on the microsecond time scale, which is longer than in most preceding simulations. Our study therefore provides updated insights into the combined effects of Nt_{17}, PRD, and polyQ tract length on the Htt-ex1 structure and aggregation.

2 Materials and methods

2.1 Initial structures and system preparations

The two simulated sequences are MATLEKLMKAFESLKSF-Q_n-P_{11}-QLPQPPPQAQPLLPQPQ-P_{10}-GPAVAEEPLHRP with n = 23 and n = 48 glutamine residues (Figure 1). The X-ray crystal structure of Htt-ex1 (PDB ID: 3IOT) (Kim et al., 2009) consists of an α-helix for Nt_{17} and the α-helix extends up to 15 glutamines into the $polyQ_{17}$ tract. For the $polyP_{11}$ region, extended loops, random coil conformations and also PPII-helix formation were observed. However, since the structure of the full-length Htt-ex1 protein is not completely resolved, the initial structures were built using the Avogadro software (Hanwell et al., 2012) and PyMOL (Schrödinger, 2022) by adding the peptide fragment Q_{11}-PRD and Q_{36}-PRD to the N-terminal part Nt_{17}-Q_{12} taken from the PDB structure 3IOT for modeling the Htt-Q_{23} and Htt-Q_{48}, respectively. This resulted in extended structures for both proteins, which were solvated with water and then collapsed through initial 20 ns MD simulations to reduce the system size in the production MD simulations.

2.2 MD simulations

All MD simulations were carried out using GROMACS 2020 (Van Der Spoel et al., 2005; Abraham et al., 2015) as MD program, employing the atomistic force field Charmm36m along with the Charmm-modified TIP3P water model (Huang et al., 2017). The initial (already collapsed) monomer structures were solvated in a ~2,000 nm^3 cubic box, including 150 mM Na^+ and Cl^- ions and overall system neutralization, leading to about 200,000 atoms. The energy of each system was initially minimized using the steepest descent algorithm (Müller and Brown, 1979; Zhang, 2015), followed by 1.0 ns equilibration and then the production runs under NPT ensemble conditions. The pressure and temperature were maintained at 1.0 bar and 298 K using the Parrinello-Rahman pressure coupling method (Parrinello and Rahman, 1981; Parrinello and Rahman, 1982) and a velocity-rescale thermostat method (Bussi et al., 2007), respectively. All simulations implied periodic boundary conditions applied in all directions and using a cutoff distance of 1.2 nm for the calculation of the non-bonded interactions in real space. The electrostatic interactions were computed using the particle mesh Ewald (PME) method (Essmann et al., 1995). The LINCS algorithm was applied (Hess et al., 1997) to constrain bond lengths, and a leapfrog integrator (Van Gunsteren and Berendsen, 1988) was used to integrate the equations of motions with a 2 fs time step. For all simulations it was checked that there were no self-interactions of the proteins with their periodic images during the course of the simulations as a result of conformational reorientations.

For each monomer system, 5 × 1.0 μs simulations were carried out. Simulation 2 was initiated from the final snapshot of the first simulation, while simulations 3 through 5 were started from the central structure of the most populated conformational cluster of the respective preceding simulation. For the analysis, the first simulation per sequence was discarded due to the bias of the extended initial structures, despite the initial 20 ns MD simulation to collapse these structures. This bias is visible in the time traces of various observables (Supplementary Figures S1–S3). To assesses the convergence of the monomer simulations, the autocorrelation functions of various structural properties were calculated for the individual 1.0 μs runs (Supplementary Figure S4). They all tend to zero within 500 ns, indicating loss of memory from initial structures. Nonetheless, while general convergence of the individual trajectories is found, certain fluctuations of the various structural properties persist, which is an expected behavior for intrinsically disordered proteins (Paul et al., 2021). We further conducted the Augmented Dickey Fuller (ADF) test as a unit-root test to assess the stationarity of the monomer data. The ADF test was applied to the concatenated time series of various observables (Supplementary Table S1). The obtained p-values for all the tested observables are below 0.05, and the test statistics are also smaller than the critical values, thereby rejecting the null hypothesis and indicating that the data are stationary.

For both proteins, 3 × 2.0 μs dimer simulations were carried out in a cubic simulation box with a volume of about 1,925 nm^3 and a total number of 190,000 atoms. The six most dominant conformational clusters determined from the previous monomer simulations were used as starting structures for the monomers in the dimer simulations. The simulations were initiated by randomly orienting two monomers in the simulation box, ensuring a minimum distance of 5 nm between them.

2.3 Analysis

The MD simulations were analyzed using different tools, which were invoked from GROMACS or from the MDAnalysis Python package (Michaud-Agrawal et al., 2011; Gowers et al., 2016). For the analysis of the monomers, the 4 × 1.0 μs simulations per protein (ignoring the initial simulation) were concatenated. In the case of the dimers, the 3 × 2.0 μs simulations per protein were joined for analysis. The determination of the secondary structure was done with the DSSP (Define Secondary Structure of Proteins)

program (Kabsch and Sander, 1983) as available *via* GROMACS. The PPII-helix structures were determined based on the dihedral angles of the protein backbone, ϕ and ψ, which fall into the range of $-104 \leq \phi \leq -46$ and $116 \leq \psi \leq 174$ for a PPII helix (Mansiaux et al., 2011; Yu et al., 2021). Previous studies have demonstrated that the assignment of the PPII helix using either DSSP-PPII or the backbone dihedral-angles approach provide highly similar results (Jephthah et al., 2021; McIvor et al., 2022). Conformational clustering of the trajectories was accomplished using the algorithm by Daura et al. (Daura et al., 1999), which is a nearest neighbor algorithm, using the root mean square deviation (RMSD) between all trajectory snapshots together with an RMSD cutoff of 0.5 nm to assign the neighbors. For the RMSD calculations, only the Cα atoms were used for both the alignment and actual calculation. To determine the contacts between residues, we calculated minimum distances for all residue pairs, within the proteins and also between the proteins in the case of the dimers. The time-averaged distances between the residues are presented as distance matrices. The contacts were further analyzed based on their interaction type, i.e., hydrophobic, H-bond, or salt-bridge interactions, using the CONAN software (Mercadante et al., 2018). This tool was also employed to identify correlations in the motions of the proteins in their monomeric form. Further analysis involved the calculation of the radius of gyration (R_g) of the proteins, the distance between the Cα atoms of the protein termini (denoted as end-to-end distance or d_{ee}), and the root mean square fluctuations (RMSFs) of the Cα atoms after alignment to the time-averaged structure. The free energy as a function of R_g and d_{ee} was calculated as $\Delta G(R_g, d_{ee}) = -k_B T[\ln P(R_g, d_{ee}) - \ln P_{max}(R_g, d_{ee})]$, where k_B is the Boltzmann constant, T is room temperature, $P(R_g, d_{ee})$ is the probability of the protein to have given values (R_g, d_{ee}), and $P_{max}(R_g, d_{ee})$ is the maximum of that probability distribution. The free energy was also determined as a function of the first two principal components, PC1 and PC2, that were determined from a principal component analysis (PCA) using the Cartesian coordinates of the Cα atoms.

The solvent-accessible surface area (SASA) was calculated for both the whole proteins and selected residues. The SASA of the proteins was further distinguished into the hydrophobic and polar solvent-accessible surface areas using the residue sets (Met, Ala, Phe, Leu, Pro, Gly, Val) and (Asp, Glu, His, Lys, Arg, Ser, Thr, Gln), respectively. The standard errors of the SASA values were determined *via* block averaging with 4 blocks, each containing 1,000 data points.

3 Results and discussion

3.1 The pathogenic Htt-ex1 monomer is more compact and has a higher β-sheet propensity than its non-pathogenic counterpart

In this section, we illustrate the structural differences between Htt-Q_{23} and Htt-Q_{48} monomers by the analysis of various structural properties, including the RMSD, R_g, d_{ee}, secondary structure, and the SASA. The evolution of these structural properties are presented in Supplementary Figures S1–S3.

3.1.1 Secondary structure and structural flexibility

Figures 2A, B indicate significant differences in the secondary structure profiles of the two proteins. Htt-Q_{48} has a greater tendency to adopt β-sheet, bend and turn conformations, while Htt-Q_{23} seems to favor α-helix and random coil conformations. The first ten residues of the Nt_{17} region of Htt-Q_{48} form a stable α-helix, which is contrary to Htt-Q_{23} where these N-terminal residues prefer disordered structures with some bends and turns. However, the following residues of Htt-Q_{23} form an α-helix that extends up to 16 glutamines into the polyQ region. The last five residues of $polyQ_{23}$ are mostly disordered with a negligible β-sheet content between residues 36–39. The polyP regions of Htt-Q_{23} adopt straight PPII helical structures, which is revealed by an analysis of the Ramachandran angles (Supplementary Figure S5). To further evaluate the propensities of the individual residues to adopt a PPII helix, we calculated the backbone dihedral angles ϕ and ψ and used them to calulate the PPII-helix probabilities (Supplementary Figure S6). These probabilities confirm a high propensity of PPII-helix formation in $polyP_{11}$ and $polyP_{10}$ of both proteins. In particular, the average ϕ and ψ angles are very similar for the residues of these regions. A rather high probability of PPII-helix formation is also found for most of the other residues of the PRD of both Htt-Q_{23} and Htt-Q_{48}, yet with a higher number of kinks and turns present in the PRD of Htt-Q_{48}. The α-helical regions encompassing the second half of Nt_{17} and most of the polyQ region of Htt-Q_{23} and most of the Nt_{17} region of Htt-Q_{48} have zero propensity to adopt PPII-helix structures. This suggests a more disordered polyQ region in Htt-Q_{48} than in Htt-Q_{23}, which in the former can also adopt ϕ and ψ angles that fall within the PPII-helix region with an average probability of 44% (Supplementary Figure S6).

These observations are consistent with previous NMR solution results which showed that the wild-type Huntingtin protein Nt_{17} has a propensity to adopt helical structures that extended to the polyQ domains (Baias et al., 2017; Newcombe et al., 2018). The N-terminal helix in Htt-Q_{48} is mainly limited to the Nt_{17} region, while the majority of its polyQ region forms a random coil or a bend/turn conformation. However, the terminal residues of the extended polyQ region shows a high propensity to adopt β-sheet structures. This higher β-sheet content compared to Htt-Q_{23} is consistent with previous experimental and MD simulation results (Nagai et al., 2007; Heck et al., 2014; Kang et al., 2017). These studies suggested that the β-sheet content within the polyQ region is associated with longer polyQ lengths and may play a role in amyloid fibril formation (Nagai et al., 2007; Kang et al., 2017). Experiments with synthesized β-sheet in the polyQ domains showed an increased rate of aggregation (Kar et al., 2013). This suggests that the β-sheet conformations within the polyQ region represent the aggregation-prone structures giving rise to amyloid fibrils (Kar et al., 2013).

We continued with a structural cluster analysis to see how the different secondary structures are arranged in the proteins. The six most populated clusters per protein are shown in Figures 2C, D. Five of the shown cluster structures of Htt-Q_{48} have β-sheet conformations that involve polyQ regions (residues 22–25 and 62–64) and parts of the PRD (residues 86–87), with two or three β-strands per sheet. However, in the fourth cluster structure coil conformations outside the helical Nt_{17} region dominate, indicating that the β-sheet formation in Htt-Q_{48} is transient. Htt-Q_{23} samples a

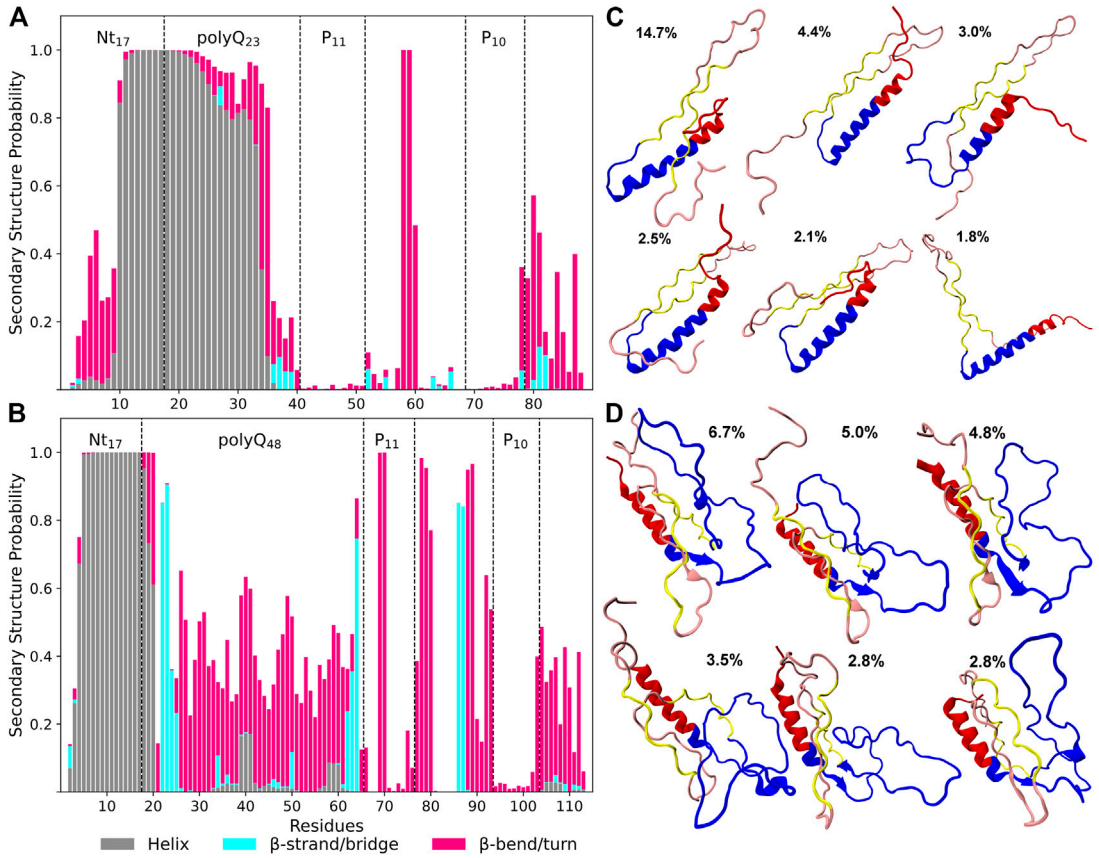

FIGURE 2
Secondary structure preferences per residue of Htt-Q_{23} **(A)** and Htt-Q_{48} **(B)** monomers. The bars represent the additive secondary structure probabilities consisting of α-helix (gray), β-strand/bridge (cyan), and β-turn/bend (magenta). The difference to 1.0 presents the other structures, in particular PPII and random coil. Representative structures were determined from conformational clustering using an RMSD cutoff of 0.5 nm for Htt-Q_{23} **(C)** and Htt-Q_{48} **(D)**. The population of each cluster is given. The Nt_{17} region is shown in red, polyQ in blue, $polyP_{11}$ and $polyP_{10}$ in yellow, and the rest of the PRD in rose.

long helix in the Nt_{17} and most of the polyQ region. The whole protein is less collapsed than Htt-Q_{48}, but rather extended, which results from the long N-terminal α-helix plus the two polyP stretches in PPII conformation that prefer to align in an antiparallel arrangement in Htt-Q_{23}. This PPII-turn-PPII arrangement is also made possible because residues Pro58/Glu59 between $polyP_{11}$ and $polyP_{10}$ allow for a turn to be formed with 100% probability. Apart from that bend, only some further bends and turns are formed in the last ten residues of the PRD of Htt-Q_{23}, while in Htt-Q_{48} several residues along the PRD sequence have a high bend/turn propensity. Together with the disordered polyQ region, this gives the longer protein a more collapsed and also more disordered shape. The Htt-Q_{23} results suggest that the region reaching from P_{11} to P_{10} strongly affects the folding of this protein by forming PPII helices that also have an ordering effect on the polyQ region, leading to helix formation in most of the 23 glutamine residues.

The different structural flexibilities of Htt-Q_{23} and Htt-Q_{48} are also visible in the RMSFs of their Cα atoms (Figure 3A). The RMSFs reveal for Htt-Q_{23} a higher flexibility in the Nt_{17}, where the first ten residues did not form a helix as in Htt-Q_{48}, while the C-terminal of the PRD region is flexible in both proteins. However, the rest of the Htt-Q_{23} is more stable than Htt-Q_{48}. In the latter, the polyQ region is particularly flexible, especially in its middle. These results suggest that longer polyQ tracts stabilize the N-terminal half of Nt_{17}, while inducing disorder and flexibility in the rest of the protein.

3.1.2 Intra-protein contacts and solvent accessibility

To further define the structural differences, we calculated the average distances between the protein residues to characterize the coupling between different domains (Figure 3B). The distance maps confirm the tendency of Htt-Q_{48} to adopt more compact structures as indicated by the significant contacts among Htt-Q_{48} residues. These contacts enable β-sheet formation in Htt-Q_{48} to take place. The contact map of Htt-Q_{23} is defined by the long α-helix in the N-terminal protein part and the alignment between $polyP_{11}$ and $polyP_{10}$ in the C-terminal half. The alignment of the polyP regions in Htt-Q_{23} is stabilized by hydrophobic interactions (Supplementary Figure S7). In addition, the polyQ region sporadically interacts with $polyP_{10}$ and the C-terminal residues (79–90). Conversely, the Nt_{17} of Htt-Q_{48} tends to interact with $polyP_{11}$, the middle of PRD (77–87)—both mediated by hydrophobic interactions –, and the C-terminal (98–110). Moreover, the polyQ region of Htt-Q_{48} can also interact with PRD domains excluding the PRD C-terminal, which involves

10.3389/fmolb.2023.1143353

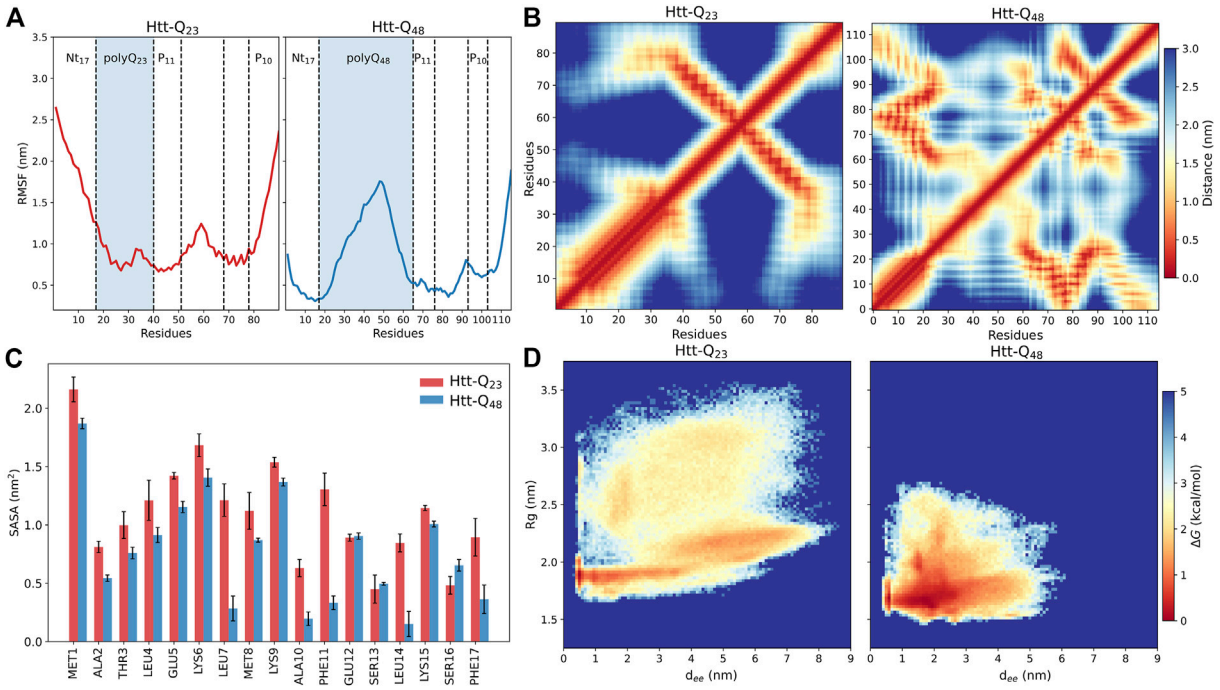

FIGURE 3
Structural properties of Htt-Q_{23} and Htt-Q_{48} monomers. **(A)** The RMSFs of the Cα atoms of Htt-Q_{23} (left) and Htt-Q_{48} (right). The polyQ regions are highlighted by a blue shade. **(B)** Intra-protein distance matrices of Htt-Q_{23} (left) and Htt-Q_{48} (right). The color bar on the right shows the average residue-residue distance (in nm). **(C)** The average SASA of the Nt_{17} residues (with standard errors) for Htt-Q_{23} (red) and Htt-Q_{48} (blue). **(D)** The free energy surface of Htt-Q_{23} (left) and Htt-Q_{48} (right) was calculated using the radius of gyration and the end-to-end distance. The color bar shows the value of the free energy (ΔG) in kcal/mol.

H-bond formation. Strikingly, the polyP regions interact with different domains of Htt-Q_{48}. A remarkable difference between the two proteins is that the Nt_{17} region of Htt-Q_{48} shows strong contacts with PRD regions, which is hardly the case in Htt-Q_{23}. The stability of the intra-protein interaction was further assessed by calculating their probability expressed as percent lifetime (Supplementary Figure S8). This analysis reveals that the hydrophobic interactions between the polyP regions in Htt-Q_{23} and those involving Nt_{17} in Htt-Q_{48} are the most stable interactions. The other very stable interactions are the H-bonds in α-helices that are present in either protein. The terminal interactions in Htt-Q_{48} seem to stabilize the helix in the first ten Nt_{17} residues, and also induce bends/turns with some helical propensity in the last ten C-terminal residues. It has been shown that the helical structure of Nt_{17} is characterized by amphipathic properties (Kelley et al., 2009; Długosz and Trylska, 2011). Thus, the hydrophobic residues of the Nt_{17} minimize their contact with water by interacting with the hydrophobic $polyP_{11}$ and the following PRD mixed region. The different intra-protein interactions in the two proteins also manifest themselves in different coupling between their domains, as revealed by dynamic cross-correlation maps that uncover correlations in the motions (Supplementary Figure S9). In Htt-Q_{23}, all residues spanning from the start of $polyP_{11}$ to the end of $polyP_{10}$ move in a correlated fashion, as they are coupled *via* their strong intra-protein interactions, while this region moves anticorrelated to residues 1–40 containing Nt_{17} and $polyQ_{23}$. In Htt-Q_{48} the correlated and anticorrelated motions are more fragmented along the sequence as a result of the more distributed intra-protein contacts. Most notably, the motions of the $polyQ_{48}$ are mainly anticorrelated to the motions of all other residues on either side of that region.

An analysis of the SASA per residue of Nt_{17} confirms that in Htt-Q_{48} these residues are better shielded from water; this difference is particularly noticeable for the hydrophobic residues Leu4, Leu7, Ala10, Phe11, Leu14, and Phe17 (Figure 3C). This is consistent with previous observations that the Nt_{17}, $polyP_{11}$ and mixed region of Htt-ex1 form a hydrophobic core (Williamson et al., 2010; Długosz and Trylska, 2011) and is confirmed by the hydrophobic and polar SASA of the whole proteins. Htt-Q_{48} exposes more polar residues and prefers to bury the hydrophobic ones compared to Htt-Q_{23} (Supplementary Figure S10A). However, the polar SASA results are not directly comparable, as Htt-Q_{48} has 25 more polar residues than Htt-Q_{23}. Supplementary Figure S10B shows that the polyQ stretch of Htt-Q_{48} is solvent-exposed, shielding the hydrophobic core from the solvent from one side, while on the other side the shielding is done by the hydrophilic face of the N-terminal helix. In Htt-Q_{23}, on the other hand, the hydrophobic residues cluster around the polyP PPII helices and interact with the hydrophobic residues of the Nt_{17} helix (Supplementary Figure S10B), leaving them more solvent-exposed as the 23 glutamine residues are not enough to cover them in that geometry. Nonetheless, this protein structure is stable as the pronounced intra-protein contacts attest. Previous studies suggested that the increase of the interaction surface of the polyQ tract could enhance the protein-protein interactions in the

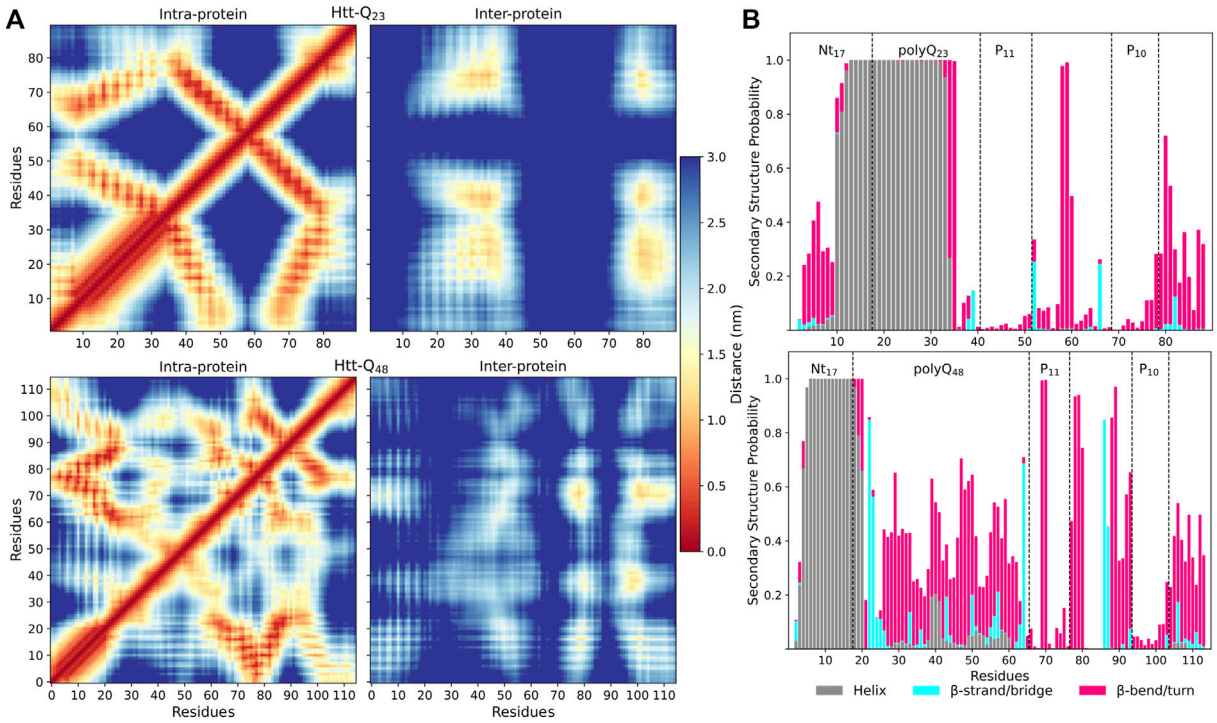

FIGURE 4
Residue-residue contacts and secondary structures in Htt-Q_{23} and Htt-Q_{48} dimers. **(A)** The intra- and inter-protein contacts between residues for Htt-Q_{23} (top) and Htt-Q_{48} (bottom) dimers. The intra-protein contacts within protein 1 are shown below the main diagonal and within protein 2 above the main diagonal. The color bar on the right shows the average distances (in nm). **(B)** The time- and protein-averaged probability of secondary structure per residue for Htt-Q_{23} (top) and Htt-Q_{48} (bottom) dimers. The bars represent the additive secondary structure probabilities consisting of α-helix (gray), β-strand/bridge (cyan), and β-turn/bend (magenta). The difference to 1.0 presents the other structures, in particular PPII and random coil.

cell by enhancing the binding (Warner et al., 2017; Newcombe et al., 2018), which according to the solvent exposure of the polyQ tract seen here for Htt-Q_{48} seems plausible. Overall, the intra-protein interactions of Htt-ex1 are highly associated with the increase of polyQ length, and are mainly governed by the increased protein flexibility in that region, allowing the formation of a hydrophobic core in Htt-Q_{48}.

3.1.3 Conformational variability

To highlight the conformational differences between the two proteins and get better insights into how the elongated polyQ tract affects the conformational energy landscape, we calculated the free energy surface (FES) using the radius of gyration and the end-to-end distance as order parameters. Figure 3D illustrates the structural differences between the two proteins. The FES of Htt-Q_{23} shows that it covers a wider range of the conformational (R_g, d_{ee}) space than Htt-Q_{48}, which populates regions with more compact structures, confirming our observations made thus far. Htt-Q_{23} prefers conformations with larger R_g values and end-to-end distances due to its elongated shape. Its greater flexibility results mainly from the larger conformational freedom of the first ten residues, that together with the C-terminal motions lead to a large distribution of the R_g and d_{ee} values in Htt-Q_{23}. As a result, there are structures where the two termini are very close to each other (e.g., clusters 1 and 5 in Figure 2C), structures where they are directed away from each other (cluster 2), and the intermediate cases (clusters 3, 4 and 6). Nonetheless, it should be reiterated that the overall fold of Htt-Q_{23} defined by the N-terminal helix and the PPII-turn-PPII motif reaching from polyP$_{11}$ to polyP$_{10}$ is rather stable despite the FES suggesting otherwise. Since d_{ee} and R_g are partially correlated, although in Htt-Q_{23} small d_{ee} values are possible for large R_g and the converse is also true, i.e., small R_g with large d_{ee} values, we also created FESs as a function of the two main pincipal (orthogonal) components, PC1 and PC2 (Supplementary Figure S11A). These suggest that the two proteins cover an equally large conformational space. However, the two FESs cannot be directly compared because the principal components of Htt-Q_{23} and Htt-Q_{48} are not identical, as the two proteins have different sizes. Projection of the individual trajectories onto PC1-PC2 space (Supplementary Figure S11B) shows that Htt-Q_{48} tends toward more compact structures with increasing simulation time, whereas the shorter protein continues to explore the entire conformational space. In summary, the main conclusion from the FES analysis is that Htt-Q_{48} prefers to form more collapsed structures, which is in agreement with its numerous intra-protein contacts. Several experiments and simulations studies suggested that the polyQ tracts prefer to adopt collapsed conformations due to their poor water solubility (Kang et al., 2017). However, this only partly agrees to our findings as we find the long polyQ tract of Htt-Q_{48} to shield the more hydrophobic residues from the water, giving rise to more collapsed conformations than on average seen for Htt-Q_{23} where the polyQ tract is not long enough to provide sufficient shielding and forms a helix instead.

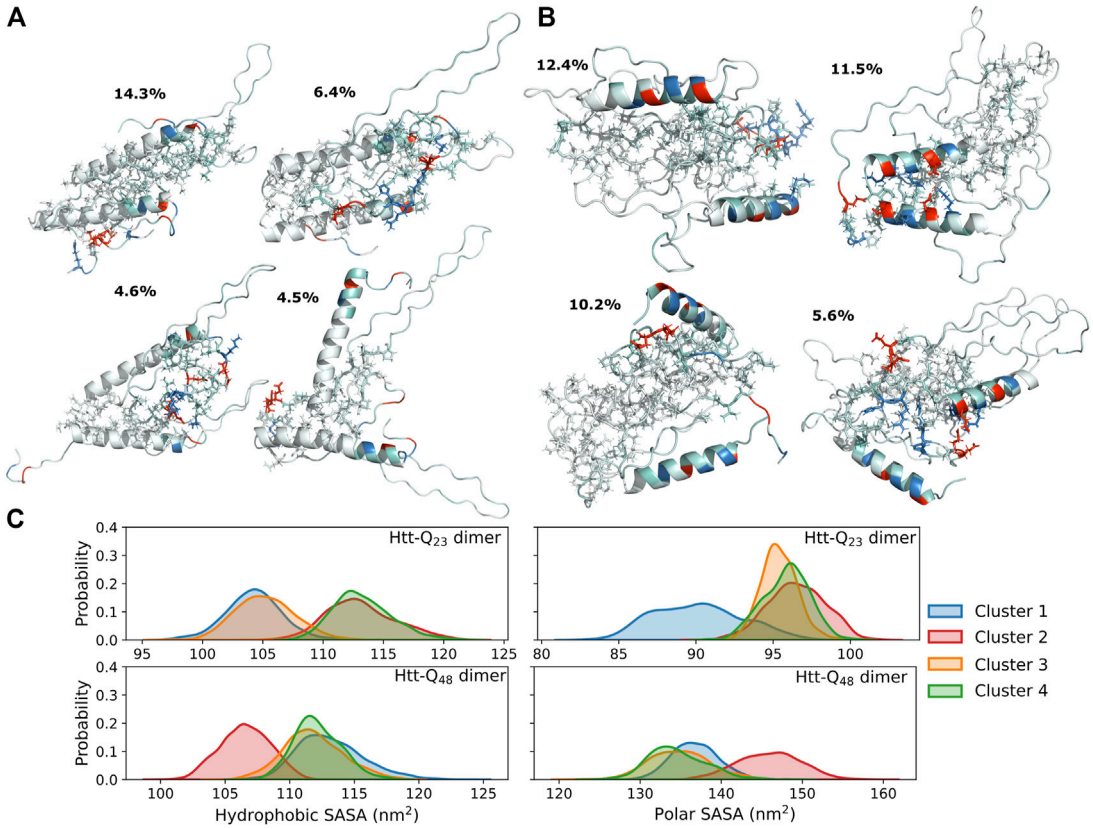

FIGURE 5
Structures and SASAs of dimers. The four most populated cluster structures for the Htt-Q_{23} **(A)** and Htt-Q_{48} **(B)** dimer. The proteins are shown as cartoons and colored red for the negatively charged residues, blue for the positively charged ones, and white otherwise. The side chains at the protein interfaces are shown as sticks. **(C)** The distribution of the hydrophobic (right) and polar (left) SASAs of the four most populated clusters for Htt-Q_{23} (top) and Htt-Q_{48} (bottom) dimers are shown. The results for the four clusters are shown in different colors according to the color mapping on the right.

3.2 Non-pathogenic and pathogenic Htt-ex1 dimerize in distinct patterns

We investigated the dimerization between the Htt-ex1 monomers to obtain information on the first oligomerization step. For either protein, we observed an association between the monomers within ~100 ns of the simulation time (Supplementary Figure S12). The dimerization modes are analyzed in more detail in the following.

3.2.1 Interaction interfaces and dimer structures

To understand how the proteins interact with each other, we calculated intra- and inter-protein residue-residue distance matrices (Figure 4A). They reveal different interaction patterns within the Htt-Q_{23} and Htt-Q_{48} dimers. In Htt-Q_{23}, there are four pronounced interaction regions between the two proteins: i) polyQ/polyP$_{11}$ with polyQ/polyP$_{11}$, ii) polyQ/polyP$_{11}$ with polyP$_{10}$/PRD (79–90), iii) polyP$_{10}$/PRD(79–90) with polyQ/ploylyP$_{11}$, iv)polyP$_{10}$/PRD(79–90) with ployP$_{10}$/PRD(79–90). Nt$_{17}$ and PRD(52–68) are not involved at all in the dimerization of Htt-Q_{23}. Htt-Q_{48}, on the other hand, features intermolecular interactions across the whole sequence, which are, however, less clear-cut. This indicates that different dimer configurations are possible for Htt-Q_{48}. The most prominent interactions involve the polyP regions, polyP with mixed regions in PRD, and polyP with polyQ. Furthermore, the polyQ is seen to interact with the whole protein, while the Nt$_{17}$ shows interactions with polyP$_{11}$, polyQ, and the C-terminal of the PRD.

The intra-protein interactions in the two proteins of the Htt-Q_{48} dimers are very similar to the ones found for the corresponding monomer (Figure 3B). This implies that the aggregation did not cause major structural transitions, which is confirmed by the secondary structure analysis (Figure 4B vs. Figure 2A). A small difference in Htt-Q_{23} is that dimerization stabilized the Nt$_{17}$/polyQ helix, while in Htt-Q_{48} a small increase in the tendency for β-sheet formation is observed across the polyQ region. For more β-sheet formation as a result of Htt-Q_{48} dimerization to occur, we expect that more simulation time than the 6×2 μs sampled here is required. In Htt-Q_{23}, in addition to the intra-protein contacts that were already present in the monomer, two more contact areas within the proteins emerged: the Nt$_{17}$ and first half of the polyQ region interacts with both polyP$_{11}$ and polyP$_{10}$. This speaks for small reorientations of the secondary-structure elements with respect to each other. In order to further understand the interaction interfaces between the proteins, we performed structural clustering and analyzed the hydrophobic and the polar SASAs of the dominant

dimer clusters. Figures 5A, B show the representative structures of the four most populated clusters. In general, the interaction interfaces between the Htt-Q_{23} proteins are mainly governed by the interactions of the partly amphipathic Nt_{17}/polyQ helices. The two helices tend to be in a parallel orientation to each other, and their interactions add helical stability. The polyP regions, on the other hand, are hardly involved in the interaction interface and instead point away from it. On the other hand, the Htt-Q_{48} dimer clusters reveal a more diverse aggregation behavior with different interaction interfaces where the polyP and the PRD take part in them.

3.2.2 Comparison with experimental observations

Our observations are in agreement with atomic force microscopy (AFM) results that identified oligomeric aggregates of Htt-Q_{20} to be structurally different from those of Htt-ex1 with disease-causing polyQ lengths (Legleiter et al., 2010). The Htt-Q_{20} oligomers did also not transit to form fibril structures. In another study it was shown that the addition of isolated polyP to Htt-ex1 could block the initial aggregation by interacting with Nt_{17} and reduce the ability of headpiece dimerization (Arndt et al., 2019). This speaks for the general involvement of Nt_{17} in the aggregation of Htt-ex1, which is also seen here for both Htt-Q_{23} and Htt-Q_{48}. Another possible aggregation-modifying effect of added polyP is that it can also interact with the polyQ region and stabilize it by promoting it to form PPII helix (Darnell et al., 2007; Darnell et al., 2009), which could prevent the nucleation in the oligomers or promote aggregation pathways that are not mediated by the Nt_{17} headpieces (Arndt et al., 2019). Pandey et al. proposed that the Htt-ex1 aggregation is initiated *via* the Nt_{17} headpiece, suggesting that the aggregation pathway of synthetic Htt-ex1 protein into oligomers starts with the self-assembly into an α-helix rich oligomer with helical Nt_{17} in the core, followed by β-sheet formation within the polyQ tracts. The final and slowest step is the structural maturation of the PRD (Pandey et al., 2018). Through the simulations performed here, we were able to capture the event of Nt_{17} headpiece interactions for Htt-Q_{48} (Figure 5B, cluster 2). Dimerization *via* the Nt_{17} helices results in a reduction of the hydrophobic surface area, while it increases the polar surface area, compared to the other dimerization modes (Figure 5C). This difference is expected to make the helix-mediated dimer more stable. Mutational studies have shown that altering the hydrophobic residues of Nt_{17} to negatively charged residues or by serine phosphorylation, strongly affects the aggregation kinetics by slowing down the aggregation (Gu et al., 2009; Thakur et al., 2009; Mishra et al., 2012). However, Htt-ex1 aggregation was not prevented by these amino-acid changes. In addition to that, according to the aggregation mechanism investigated by Williamson et al., metastable aggregates start to form *via* Nt_{17} packed cores and the polyQ regions being excluded from these cores. These initial aggregates develop then to amyloid nuclei composed of Nt_{17} headpieces and polyQ tracts (Williamson et al., 2010). Based on our dimer simulation results, this aggregation mechanism appears fully plausible. We further find that longer polyQ tracts are needed for amyloid formation as for shorter polyQ sequences, the Nt_{17} helix extends far into the polyQ region and is resistant against β-sheet formation. This observation is in agreement with recent NMR spectroscopy results (Elena-Real et al., 2022).

4 Conclusion

In this work, we carried out multiple MD simulations to investigate the monomer and dimer structures of a pathogenic Htt-ex1 protein (Htt-Q_{48}) and a non-pathogenic (Htt-Q_{23}) counterpart. We found rather different monomer structures for the two proteins. One of the most relevant distinctions is that in Htt-Q_{23} the N-terminal helix involves not only Nt_{17} but also the majority of the $polyQ_{23}$ residues. This can be explained by the rather high helix propensity of glutamine, which comes seventh on the helix propensity scale of the twenty proteinogenic amino acids (Pace and Scholtz, 1998). However, for longer polyQ sequences as in Htt-Q_{48}, this propensity is not sufficient and the polyQ tract becomes disordered. Here, it should also be considered that the length of an α-helix always results from a competition between α-helix folding, unfolding into a random coil and the formation of higher-order tertiary structures, and that this competition leads to naturally favored α-helix lengths of 9–17 amino acids (Qin et al., 2013). It is thus fully understandable that with increasing polyQ length, this region becomes disordered. As a result, more intra-protein interactions develop, which in turn can lead to β-sheet formation, even though glutamine has only a medium propensity to be in a β-sheet (Shingate and Sowdhamini, 2012). Here, we observed an increase in bend/turn conformations across the polyQ and PRD regions of Htt-Q_{48}, which yielded compact protein structures involving small intra-protein β-sheets. In both proteins, the polyP regions mainly adopted PPII structures. However, in Htt-Q_{48} they were disrupted by kinks, which is in line with the compact Htt-Q_{48} conformations, whereas in Htt-Q_{23} the $polyP_{11}$ and $polyP_{10}$ are straight segments that are aligned antiparallel to each other. As a result, Htt-Q_{23} has an elongated shape resulting from the 25 residue-long N-terminal helix and the PPII-turn-PPII motif. This overall fold appeared rather stable and structural fluctuations mainly derived from the first and last ten residues of Htt-Q_{23}.

With regard to dimerization, we found that for both Htt-Q_{23} and Htt-Q_{48} this process involves the N-terminal helix, yet to a larger degree in Htt-Q_{23}. Here, the helix-helix interaction further stabilized the helix, especially in its second part that involves residues from the $polyQ_{23}$ region, which explains why this region is resistant against β-sheet formation. Moreover, interaction through the amphipathic α-helix Nt_{17} headpieces minimizes the hydrophobic surface, making this dimer conformation particularly stable. In Htt-Q_{48}, on the other hand, different dimerization modes were observed, of which the one involving the two N-terminal helices is one possibility. However, we expect the Nt_{17} helix mediated dimer to be more stable than the others due to the larger burial of hydrophobic residues. An important aspect of Htt-Q_{48} dimerization is that the disordered polyQ regions are considerably involved in the process, allowing β-sheets to form on the long time scale (not simulated here). However, this structural transition is slowed down by visible interactions with the PRD. For both Htt-Q_{23} and Htt-Q_{48} the polyP regions are not involved in the dimerization, though their avoidance of the dimer interaction interface is more pronounced in Htt-Q_{23} due to its particular elongated protein shape, allowing the polyP regions to keep away from it.

Overall, we observed that polyQ extension affects the conformational ensemble of Htt-ex1 already at the monomer level and has a direct impact on dimerization. Our results provide insights into the roles of the flanking domains of polyQ in modulating the conformations and the aggregation pathways, thereby providing explanations for several experimental findings at the atomistic level.

Data availability statement

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Author contributions

MK, AS-A, and BS contributed to design of the study. MK performed and analyzed the simulations. MK wrote the first draft of the manuscript. All authors contributed to the manuscript revision, read, and approved.

Funding

The project was funded *via* the Palestinian-German Science Bridge financed by the German Federal Ministry of Education and Research (BMBF).

Acknowledgments

The authors gratefully acknowledge the computing time granted through JARA on the supercomputer Jureca at Forschungszentrum Jülich (project AMYLOID-MSM).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2023.1143353/full#supplementary-material

References

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., et al. (2015). Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 1, 19–25. doi:10.1016/j.softx.2015.06.001

Arndt, J. R., Chaibva, M., Beasley, M., Kiani Karanji, A., Ghassabi Kondalaji, S., Khakinejad, M., et al. (2019). Nucleation inhibition of huntingtin protein (htt) by polyproline ppii helices: A potential interaction with the n-terminal α-helical region of htt. *Biochemistry* 59, 436–449. doi:10.1021/acs.biochem.9b00689

Baias, M., Smith, P. E., Shen, K., Joachimiak, L. A., Zerko, S., Kozminski, W., et al. (2017). Structure and dynamics of the huntingtin exon-1 n-terminus: A solution nmr perspective. *J. Am. Chem. Soc.* 139, 1168–1176. doi:10.1021/jacs.6b10893

Bhattacharyya, A., Thakur, A. K., Chellgren, V. M., Thiagarajan, G., Williams, A. D., Chellgren, B. W., et al. (2006). Oligoproline effects on polyglutamine conformation and aggregation. *J. Mol. Biol.* 355, 524–535. doi:10.1016/j.jmb.2005.10.053

Bugg, C. W., Isas, J. M., Fischer, T., Patterson, P. H., and Langen, R. (2012). Structural features and domain organization of huntingtin fibrils. *J. Biol. Chem.* 287, 31739–31746. doi:10.1074/jbc.M112.353839

Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *Chem. Phys.* 126, 014101. doi:10.1063/1.2408420

Chellgren, B. W., Miller, A.-F., and Creamer, T. P. (2006). Evidence for polyproline ii helical structure in short polyglutamine tracts. *J. Mol. Biol.* 361, 362–371. doi:10.1016/j.jmb.2006.06.044

Chen, S., Berthelier, V., Yang, W., and Wetzel, R. (2001). Polyglutamine aggregation behavior *in vitro* supports a recruitment mechanism of cytotoxicity. *J. Mol. Biol.* 311, 173–182. doi:10.1006/jmbi.2001.4850

Chow, W. N. V., Luk, H. W., Chan, H. Y. E., and Lau, K.-F. (2012). Degradation of mutant huntingtin via the ubiquitin/proteasome system is modulated by fe65. *Biochem. J.* 443, 681–689. doi:10.1042/BJ20112175

Crick, S. L., Jayaraman, M., Frieden, C., Wetzel, R., and Pappu, R. V. (2006). Fluorescence correlation spectroscopy shows that monomeric polyglutamine molecules form collapsed structures in aqueous solutions. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16764–16769. doi:10.1073/pnas.0608175103

Darnell, G., Orgel, J. P., Pahl, R., and Meredith, S. C. (2007). Flanking polyproline sequences inhibit β-sheet structure in polyglutamine segments by inducing ppii-like helix structure. *J. Mol. Biol.* 374, 688–704. doi:10.1016/j.jmb.2007.09.023

Darnell, G. D., Derryberry, J., Kurutz, J. W., and Meredith, S. C. (2009). Mechanism of cis-inhibition of polyq fibrillation by polyp: Ppii oligomers and the hydrophobic effect. *Biophys. J.* 97, 2295–2305. doi:10.1016/j.bpj.2009.07.062

Daura, X., Gademann, K., Jaun, B., Seebach, D., Gunsteren, W., and Mark, A. (1999). Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.* 38, 236–240. doi:10.1002/(SICI)1521-3773(19990115)38:1/2⟨236::AID-ANIE236⟩3.0.CO;2-M

Długosz, M., and Trylska, J. (2011). Secondary structures of native and pathogenic huntingtin n-terminal fragments. *J. Phys. Chem. B* 115, 11597–11608. doi:10.1021/jp206373g

Duennwald, M. L., Jagadish, S., Giorgini, F., Muchowski, P. J., and Lindquist, S. (2006a). A network of protein interactions determines polyglutamine toxicity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11051–11056. doi:10.1073/pnas.0604548103

Duennwald, M. L., Jagadish, S., Muchowski, P. J., and Lindquist, S. (2006b). Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11045–11050. doi:10.1073/pnas.0604547103

Elena-Real, C. A., Sagar, A., Urbanek, A., Popovic, M., Morató, A., Estaña, A., et al. (2022). The structure of pathogenic huntingtin exon-1 defines the bases of its aggregation propensity. *bioRxiv*. doi:10.1101/2022.10.25.513661

Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H., and Pedersen, L. G. (1995). A smooth particle mesh ewald method. *Chem. Phys.* 103, 8577–8593. doi:10.1063/1.470117

Gkekas, I., Gioran, A., Boziki, M. K., Grigoriadis, N., Chondrogianni, N., and Petrakis, S. (2021). Oxidative stress and neurodegeneration: Interconnected processes in polyq diseases. *Antioxidants* 10, 1450. doi:10.3390/antiox10091450

Gowers, R. J., Linke, M., Barnoud, J., Reddy, T. J. E., Melo, M. N., Seyler, S. L., et al. (2016). "MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations," in Proceedings of the 15th Python in Science Conference. Editors S. Benthall and S. Rostrup, 98–105. doi:10.25080/Majora-629e541a-00e

Gu, X., Greiner, E. R., Mishra, R., Kodali, R., Osmand, A., Finkbeiner, S., et al. (2009). Serines 13 and 16 are critical determinants of full-length human mutant huntingtin induced disease pathogenesis in hd mice. *Neuron* 64, 828–840. doi:10.1016/j.neuron.2009.11.020

Gusella, J. F., and MacDonald, M. E. (2000). Molecular genetics: Unmasking polyglutamine triggers in neurodegenerative disease. *Nat. Rev. Neurosci.* 1, 109–115. doi:10.1038/35039051

Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., and Hutchison, G. R. (2012). Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *J. Cheminformatics.* 4, 17. doi:10.1186/1758-2946-4-17

Heck, B. S., Doll, F., and Hauser, K. (2014). Length-dependent conformational transitions of polyglutamine repeats as molecular origin of fibril initiation. *Biophys. Chem.* 185, 47–57. doi:10.1016/j.bpc.2013.11.008

Hess, B., Bekker, H., Berendsen, H. J. C., and Fraaije, J. G. E. M. (1997). Lincs: A linear constraint solver for molecular simulations. *J. Comput. Chem.* 18, 1463–1472. doi:10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.0.co;2-h

Hoop, C. L., Lin, H.-K., Kar, K., Hou, Z., Poirier, M. A., Wetzel, R., et al. (2014). Polyglutamine amyloid core boundaries and flanking domain dynamics in huntingtin fragment fibrils determined by solid-state nuclear magnetic resonance. *Biochemistry* 53, 6653–6666. doi:10.1021/bi501010q

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B. L., et al. (2017). Charmm36m: An improved force field for folded and intrinsically disordered proteins. *Nat. Methods.* 14, 71–73. doi:10.1038/nmeth.4067

Isas, J. M., Langen, R., and Siemer, A. B. (2015). Solid-state nuclear magnetic resonance on the static and dynamic domains of huntingtin exon-1 fibrils. *Biochemistry* 54, 3942–3949. doi:10.1021/acs.biochem.5b00281

Jayaraman, M., Kodali, R., Sahoo, B., Thakur, A. K., Mayasundari, A., Mishra, R., et al. (2012). Slow amyloid nucleation via α-helix-rich oligomeric intermediates in short polyglutamine-containing huntingtin fragments. *J. Mol. Biol.* 415, 881–899. doi:10.1016/j.jmb.2011.12.010

Jephthah, S., Pesce, F., Lindorff-Larsen, K., and Skepö, M. (2021). Force field effects in simulations of flexible peptides with varying polyproline ii propensity. *J. Chem. Theory Comput.* 17, 6634–6646. doi:10.1021/acs.jctc.1c00408

Kabsch, W., and Sander, C. (1983). Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* 22, 2577–2637. doi:10.1002/bip.360221211

Kang, H., Vázquez, F. X., Zhang, L., Das, P., Toledo-Sherman, L., Luan, B., et al. (2017). Emerging β-sheet rich conformations in supercompact huntingtin exon-1 mutant structures. *J. Am. Chem. Soc.* 139, 8820–8827. doi:10.1021/jacs.7b00838

Kar, K., Hoop, C. L., Drombosky, K. W., Baker, M. A., Kodali, R., Arduini, I., et al. (2013). β-hairpin-mediated nucleation of polyglutamine amyloid formation. *J. Mol. Biol.* 425, 1183–1197. doi:10.1016/j.jmb.2013.01.016

Kelley, N. W., Huang, X., Tam, S., Spiess, C., Frydman, J., and Pande, V. S. (2009). The predicted structure of the headpiece of the huntingtin protein and its implications on huntingtin aggregation. *J. Mol. Biol.* 388, 919–927. doi:10.1016/j.jmb.2009.01.032

Kim, M. W., Chelliah, Y., Kim, S. W., Otwinowski, Z., and Bezprozvanny, I. (2009). Secondary structure of huntingtin amino-terminal region. *Struc* 17, 1205–1212. doi:10.1016/j.str.2009.08.002

Kim, M. (2013). Beta conformation of polyglutamine track revealed by a crystal structure of huntingtin n-terminal region with insertion of three histidine residues. *Prion* 7, 221–228. doi:10.4161/pri.23807

Klein, F. A., Pastore, A., Masino, L., Zeder-Lutz, G., Nierengarten, H., Oulad-Abdelghani, M., et al. (2007). Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: Towards a length-dependent toxicity gradient. *J. Mol. Biol.* 371, 235–244. doi:10.1016/j.jmb.2007.05.028

Lakhani, V. V., Ding, F., and Dokholyan, N. V. (2010). Polyglutamine induced misfolding of huntingtin exon1 is modulated by the flanking sequences. *PLoS Comput. Biol.* 6, e1000772. doi:10.1371/journal.pcbi.1000772

Legleiter, J., Mitchell, E., Lotz, G. P., Sapp, E., Ng, C., DiFiglia, M., et al. (2010). Mutant huntingtin fragments form oligomers in a polyglutamine length-dependent manner *in vitro* and *in vivo*. *J. Biol. Chem.* 285, 14777–14790. doi:10.1074/jbc.M109.093708

Li, S.-H., and Li, X.-J. (2004). Huntingtin–protein interactions and the pathogenesis of huntington's disease. *Trends Genet.* 20, 146–154. doi:10.1016/j.tig.2004.01.008

MacDonald, M. E., Ambrose, C. M., Duyao, M. P., Myers, R. H., Lin, C., Srinidhi, L., et al. (1993). A novel gene containing a trinucleotide repeat that is expanded and unstable on huntington's disease chromosomes. *Cell* 72, 971–983. doi:10.1016/0092-8674(93)90585-E

Mangiarini, L., Sathasivam, K., Seller, M., Cozens, B., Harper, A., Hetherington, C., et al. (1996). Exon 1 of the hd gene with an expanded cag repeat is sufficient to cause a progressive neurological phenotype in transgenic mice. *Cell* 87, 493–506. doi:10.1016/S0092-8674(00)81369-0

Mansiaux, Y., Joseph, A. P., Gelly, J.-C., and de Brevern, A. G. (2011). Assignment of polyproline ii conformation and analysis of sequence – structure relationship. *PLOS ONE* 6, e18401–e18415. doi:10.1371/journal.pone.0018401

McIvor, J. A. P., Larsen, D. S., and Mercadante, D. (2022). Simulating polyproline ii-helix-rich peptides with the latest kirkwood–buff force field: A direct comparison with amber and charmm. *J. Am. Phys. Chem. B* 126, 7833–7846. doi:10.1021/acs.jpcb.2c03837

Mercadante, D., Gräter, F., and Daday, C. (2018). Conan: A tool to decode dynamical information from molecular interaction maps. *Biophys. J.* 114, 1267–1273. doi:10.1016/j.bpj.2018.01.033

Michalek, M., Salnikov, E. S., and Bechinger, B. (2013). Structure and topology of the huntingtin 1–17 membrane anchor by a combined solution and solid-state nmr approach. *Biophys. J.* 105, 699–710. doi:10.1016/j.bpj.2013.06.030

Michaud-Agrawal, N., Denning, E. J., Woolf, T. B., and Beckstein, O. (2011). Mdanalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* 32, 2319–2327. doi:10.1002/jcc.21787

Miller, J., Arrasate, M., Brooks, E., Libeu, C. P., Legleiter, J., Hatters, D., et al. (2011). Identifying polyglutamine protein species *in situ* that best predict neurodegeneration. *Nat. Chem. Biol.* 7, 925–934. doi:10.1038/nchembio.694

Mishra, R., Hoop, C. L., Kodali, R., Sahoo, B., van der Wel, P. C., and Wetzel, R. (2012). Serine phosphorylation suppresses huntingtin amyloid accumulation by altering protein aggregation properties. *J. Mol. Biol.* 424, 1–14. doi:10.1016/j.jmb.2012.09.011

Moldovean, S. N., and Chiş, V. (2019). Molecular dynamics simulations applied to structural and dynamical transitions of the huntingtin protein: A review. *ACS Chem. Neurosci.* 11, 105–120. doi:10.1021/acschemneuro.9b00561

Müller, K., and Brown, L. D. (1979). Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta* 53, 75–93. doi:10.1007/BF00547608

Nagai, Y., Inui, T., Popiel, H. A., Fujikake, N., Hasegawa, K., Urade, Y., et al. (2007). A toxic monomeric conformer of the polyglutamine protein. *Nat. Struct. Mol. Biol.* 14, 332–340. doi:10.1038/nsmb1215

Newcombe, E. A., Ruff, K. M., Sethi, A., Ormsby, A. R., Ramdzan, Y. M., Fox, A., et al. (2018). Tadpole-like conformations of huntingtin exon 1 are characterized by conformational heterogeneity that persists regardless of polyglutamine length. *J. Mol. Biol.* 430, 1442–1458. doi:10.1016/j.jmb.2018.03.031

Pace, C. N., and Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophys. J.* 75, 422–427. doi:10.1016/S0006-3495(98)77529-0

Pandey, N. K., Isas, J. M., Rawat, A., Lee, R. V., Langen, J., Pandey, P., et al. (2018). The 17-residue-long n terminus in huntingtin controls stepwise aggregation in solution and on membranes via different mechanisms. *J. Biol. Chem.* 293, 2597–2605. doi:10.1074/jbc.M117.813667

Parrinello, M., and Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* 52, 7182–7190. doi:10.1063/1.328693

Parrinello, M., and Rahman, A. (1982). Strain fluctuations and elastic constants. *Chem. Phys.* 76, 2662–2666. doi:10.1063/1.443248

Paul, A., Samantray, S., Anteghini, M., Khaled, M., and Strodel, B. (2021). Thermodynamics and kinetics of the amyloid-β peptide revealed by markov state models based on md data in agreement with experiment. *Chem. Sci.* 12, 6652–6669. doi:10.1039/D0SC04657D

Perutz, M. F., Johnson, T., Suzuki, M., and Finch, J. T. (1994). Glutamine repeats as polar zippers: Their possible role in inherited neurodegenerative diseases. *Proc. Natl. Acad. Sci. U.S.A.* 91, 5355–5358. doi:10.1073/pnas.91.12.5355

Priya, S. B., and Gromiha, M. M. (2019). Structural insights into the aggregation mechanism of huntingtin exon 1 protein fragment with different polyq-lengths. *J. Cell. Biochem.* 120, 10519–10529. doi:10.1002/jcb.28338

Qin, Z.-H., Wang, Y., Sapp, E., Cuiffo, B., Wanker, E., Hayden, M. R., et al. (2004). Huntingtin bodies sequester vesicle-associated proteins by a polyproline-dependent interaction. *J. Neurosci.* 24, 269–281. doi:10.1523/JNEUROSCI.1409-03.2004

Qin, Z., Fabre, A., and Buehler, M. (2013). Structure and mechanism of maximum stability of isolated alpha-helical protein domains at a critical length scale. *Eur. Phys. J. E* 36, 53. doi:10.1140/epje/i2013-13053-8

Robustelli, P., Piana, S., and Shaw, D. (2018). Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* 115, E4758–E4766. doi:10.1073/pnas.1800690115

Sahoo, B., Singer, D., Kodali, R., Zuchner, T., and Wetzel, R. (2014). Aggregation behavior of chemically synthesized, full-length huntingtin exon1. *Biochem. J.* 53, 3897–3907. doi:10.1021/bi500300c

Samantray, S., Yin, F., Kav, B., and Strodel, B. (2020). Different force fields give rise to different amyloid aggregation pathways in molecular dynamics simulations. *J. Chem. Inf. Model.* 60, 6462–6475. PMID: 33174726. doi:10.1021/acs.jcim.0c01063

Saudou, F., and Humbert, S. (2016). The biology of huntingtin. *Neuron* 89, 910–926. doi:10.1016/j.neuron.2016.02.003

Schrödinger, L. L. C. (2022). *The pymol molecular graphics system*. Schrödinger. version 2.5.0.

Shingate, P., and Sowdhamini, R. (2012). Analysis of domain-swapped oligomers reveals local sequence preferences and structural imprints at the linker regions and swapped interfaces. *PloS one* 7, e39305. doi:10.1371/journal.pone.0039305

Sivanandam, V., Jayaraman, M., Hoop, C. L., Kodali, R., Wetzel, R., and van der Wel, P. C. (2011). The aggregation-enhancing huntingtin n-terminus is helical in amyloid fibrils. *J. Am. Chem. Soc.* 133, 4558–4566. doi:10.1021/ja110715f

Takeuchi, T., and Nagai, Y. (2017). Protein misfolding and aggregation as a therapeutic target for polyglutamine diseases. *Brain Sci.* 7, 128. doi:10.3390/brainsci7100128

Tam, S., Spiess, C., Auyeung, W., Joachimiak, L., Chen, B., Poirier, M. A., et al. (2009). The chaperonin tric blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nat. Struct. Mol. Biol.* 16, 1279–1285. doi:10.1038/nsmb.1700

Thakur, A. K., and Wetzel, R. (2002). Mutational analysis of the structural organization of polyglutamine aggregates. *Proc. Natl. Acad. Sci. U.S.A.* 99, 17014–17019. doi:10.1073/pnas.252523899

Thakur, A. K., Jayaraman, M., Mishra, R., Thakur, M., Chellgren, V. M., Byeon, I.-J. L., et al. (2009). Polyglutamine disruption of the huntingtin exon 1 n terminus triggers a complex aggregation mechanism. *Nat. Struct. Mol. Biol.* 16, 380–389. doi:10.1038/nsmb.1570

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. (2005). Gromacs: Fast, flexible, and free. *J. Comput. Chem.* 26, 1701–1718. doi:10.1002/jcc.20291

Van Gunsteren, W. F., and Berendsen, H. J. (1988). A leap-frog algorithm for stochastic dynamics. *Mol. Simul.* 1, 173–185. doi:10.1080/08927028808080941

Vöpel, T., Bravo-Rodriguez, K., Mittal, S., Vachharajani, S., Gnutt, D., Sharma, A., et al. (2017). Inhibition of huntingtin exon-1 aggregation by the molecular tweezer clr01. *J. Am. Chem. Soc.* 139, 5640–5643. doi:10.1021/jacs.6b11039

Warner, J. B., IV, Ruff, K. M., Tan, P. S., Lemke, E. A., Pappu, R. V., and Lashuel, H. A. (2017). Monomeric huntingtin exon 1 has similar overall structural features for wild-type and pathological polyglutamine lengths. *J. Am. Chem. Soc.* 139, 14456–14469. doi:10.1021/jacs.7b06659

Williams, A. J., and Paulson, H. L. (2008). Polyglutamine neurodegeneration: Protein misfolding revisited. *Trends Neurosci.* 31, 521–528. doi:10.1016/j.tins.2008.07.004

Williamson, T. E., Vitalis, A., Crick, S. L., and Pappu, R. V. (2010). Modulation of polyglutamine conformations and dimer formation by the n-terminus of huntingtin. *J. Mol. Biol.* 396, 1295–1309. doi:10.1016/j.jmb.2009.12.017

Yu, L., Li, D.-W., and Brüschweiler, R. (2021). Systematic differences between current molecular dynamics force fields to represent local properties of intrinsically disordered proteins. *J. Am. Phys. Chem. B* 125, 798–804. PMID: 33444020. doi:10.1021/acs.jpcb.0c10078

Zhang, J. (2015). The hybrid idea of (energy minimization) optimization methods applied to study prion protein structures focusing on the beta2-alpha2 loop. *Biochem. Pharmacol.* 4, 2167–0501. doi:10.4172/2167-0501.1000175

frontiers

Supplementary Material

Comparative Molecular Dynamics Simulations of Pathogenic and Nonpathogenic Huntingtin Protein Monomers and Dimers

Mohammed Khaled [1], **Birgit Strodel** [1,2,*] **and Abdallah Sayyed-Ahmad** [3,*]

[1]Institute of Biological Information Processing (IBI-7: Structural Biochemistry), Forschungszentrum Jülich, 52425 Jülich, Germany
[2]Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany
[3]Department of Physics, Birzeit University, 71939 Birzeit, Palestine
[*]b.strodel@fz-juelich.de, asayyeda@birzeit.edu

ADF test	**Htt-Q$_{23}$**				
	R$_g$	**d$_{ee}$**	**RMSD**	**helix**	**SASA**
Test Statistic	-4.41	-6.51	-3.7	-3.29	-4.26
p-Value	2.81×10^{-4}	1.14×10^{-8}	4.17×10^{-3}	5.16×10^{-4}	5.16×10^{-4}
Critical Value 1%	-3.43	-3.43	-3.43	3.43	-3.43
Critical Value 5%	-2.86	-2.86	-2.86	-2.86	-2.86
Critical Value 10%	-2.57	-2.57	-2.57	-2.57	-2.57
	Htt-Q$_{48}$				
	R$_g$	**d$_{ee}$**	**RMSD**	**helix**	**SASA**
Test Statistic	-6.04	-4.47	-8.61	-6.17	-5.02
p-Value	1.34×10^{-7}	2.22×10^{-4}	6.51×10^{-14}	6.95×10^{-8}	2.04×10^{-5}
Critical Value 1%	-3.43	-3.43	-3.43	-3.43	-3.43
Critical Value 5%	-2.86	-2.86	-2.86	-2.86	-2.86
Critical Value 10%	-2.57	-2.57	-2.57	-2.57	-2.57

Table S1. **The Augmented Dickey-Fuller (ADF) test for the Htt-ex1 monomer simulation data.** The ADF test was conducted as a unit-root test to assess the stationary of the Htt-ex1 monomer simulations. The test was applied to the concatenated time series of various observables: the radius of gyration (R_g), end-to-end distance (d_{ee}), root mean square deviation (RMSD), α-helix content, and solvent accessible surface area (SASA).

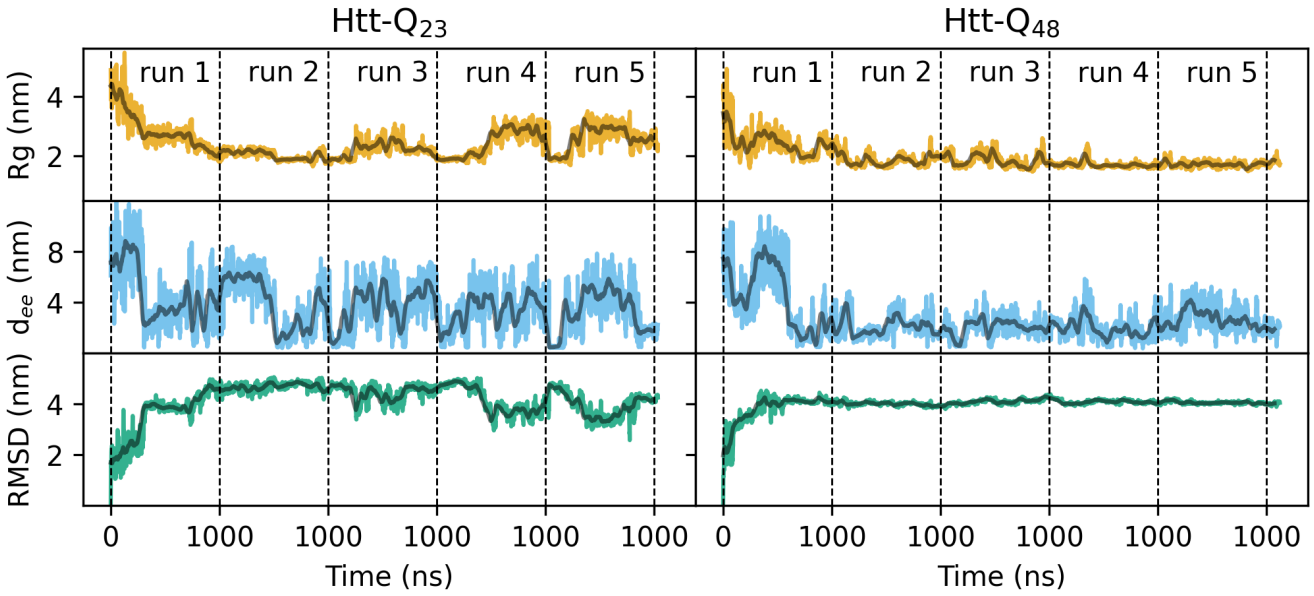

Figure S1: **Results for the 5×1 μs MD simulations of the Htt-ex1 monomers.** The evolution of R_g, RMSD, and d_{ee} for Htt-Q_{23} (left) and Htt-Q_{48} (right) are shown. The thick lines represent 50 ns running averages.

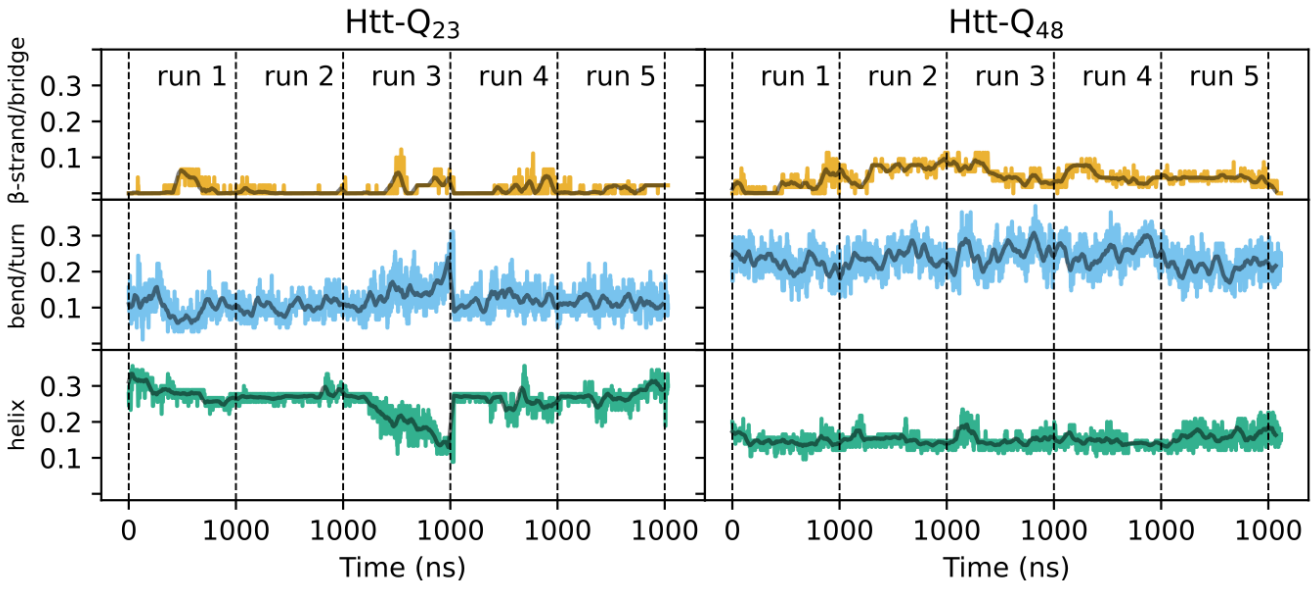

Figure S2: **Secondary structures during the** 5×1 μ**s MD simulations of the Htt-ex1 monomers.** The evolution of the β-strand/bridge, bend/turn, and α-helix for Htt-Q_{23} (left) and Htt-Q_{48} (right) are shown. The thick lines represent 50 ns running averages.

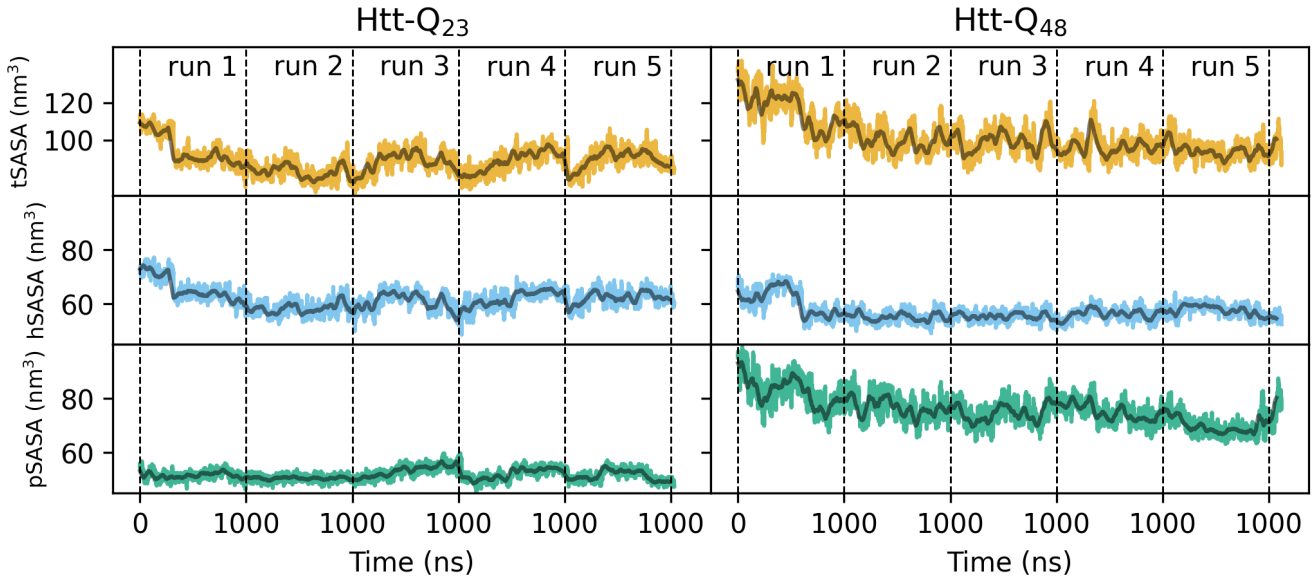

Figure S3: **Solvent accessible surface areas of the Htt-ex1 monomers during the** 5×1 μ**s MD simulations.** The evolution of the total SASA (tSASA), the hydrophobic SASA (hSASA), and the polar SASA (pSASA) for Htt-Q_{23} (left) and Htt-Q_{48} (right) are shown. The thick lines represent 50 ns running averages.

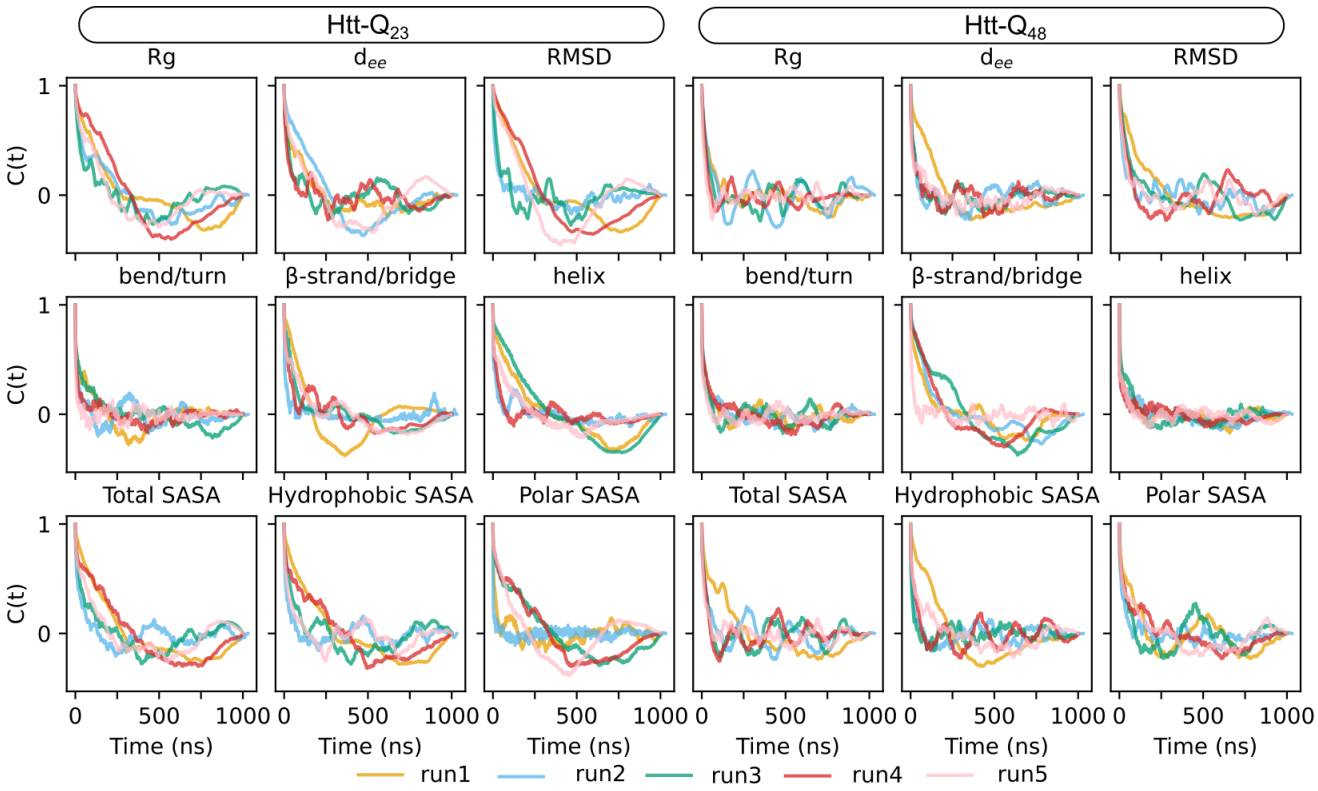

Figure S4: **The autocorrelation functions for a set of structural properties of Htt-ex1 monomers.** The top panels show the autocorrelation functions ($C(t)$) of R_g, d_{ee}, and RMSD; the middle panels include the $C(t)$ of the content of secondary structures, i.e., bend/turn, β-strand/bridge, and helix; and the bottom panels display the $C(t)$ of the total, hydrophobic, and polar SASA. The results for Htt-Q_{23} are shown on the left and those for Htt-Q_{48} on the right. All autocorrelations vanish within 100-500 ns, indicating loss of memory from initial structures.

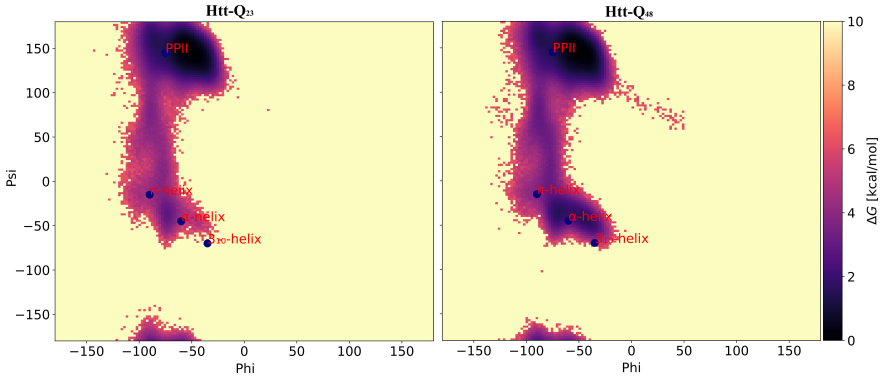

Figure S5: **Ramachandran maps for the polyP regions of Htt-Q_{23} (left) and Htt-Q_{48} (right) monomers.** The color bar shows the value of the free energy (ΔG) in kcal/mol. The blue dots indicate the ideal location of the corresponding secondary structure conformations.

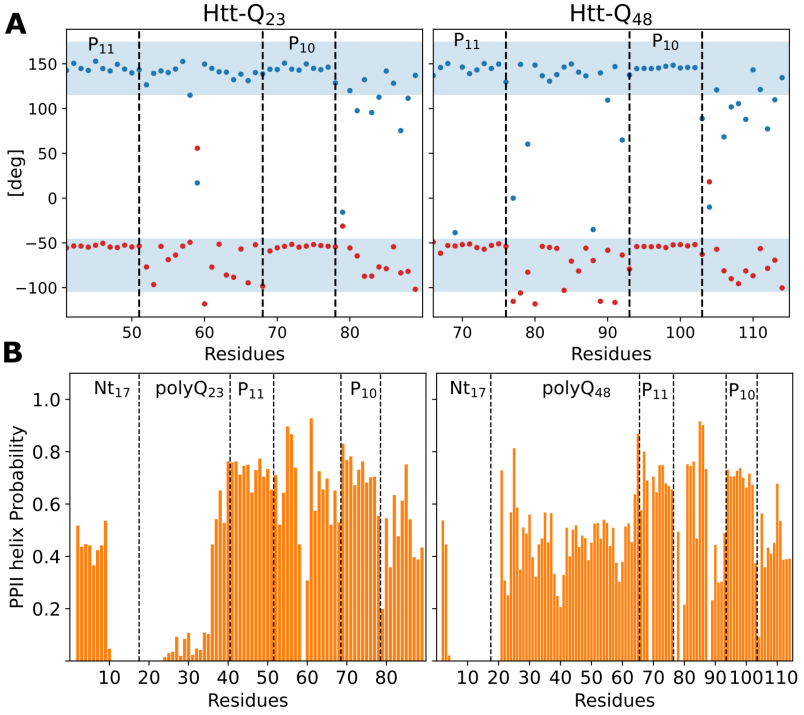

Figure S6: **PPII-helix formation within the Htt-ex1 monomers. (A)** The residue-resolved average ϕ (red) and ψ (blue) dihedral angles of the PRD region of Htt-Q_{23} (left) and Htt-Q_{48} (right) monomers. The shaded areas indicate the region that defines the PPII helix as in the DSSP-PPII algorithm. **(B)** The PPII-helix propensity of the residues of Htt-Q_{23} (left) and Htt-Q_{48} (right) monomers calculated from the ϕ and ψ dihedral angle distributions during the simulations.

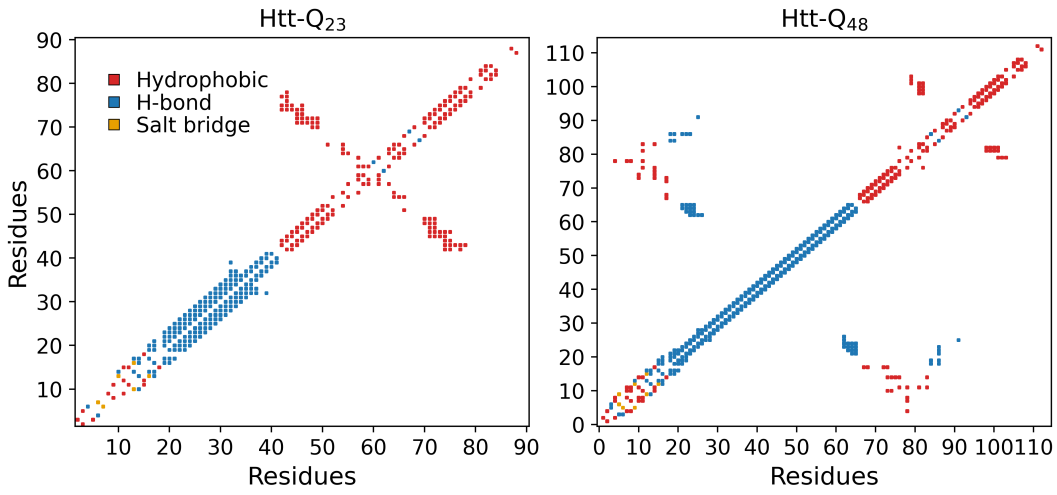

Figure S7: **Types of intra-protein interactions in the Htt-ex1 monomers.** The intra-protein contact maps are colored based on the type of interaction for Htt-Q_{23} (left) and Htt-Q_{48} (right), using red for hydrophobic interactions, blue for H-bonds, and orange for salt bridges.

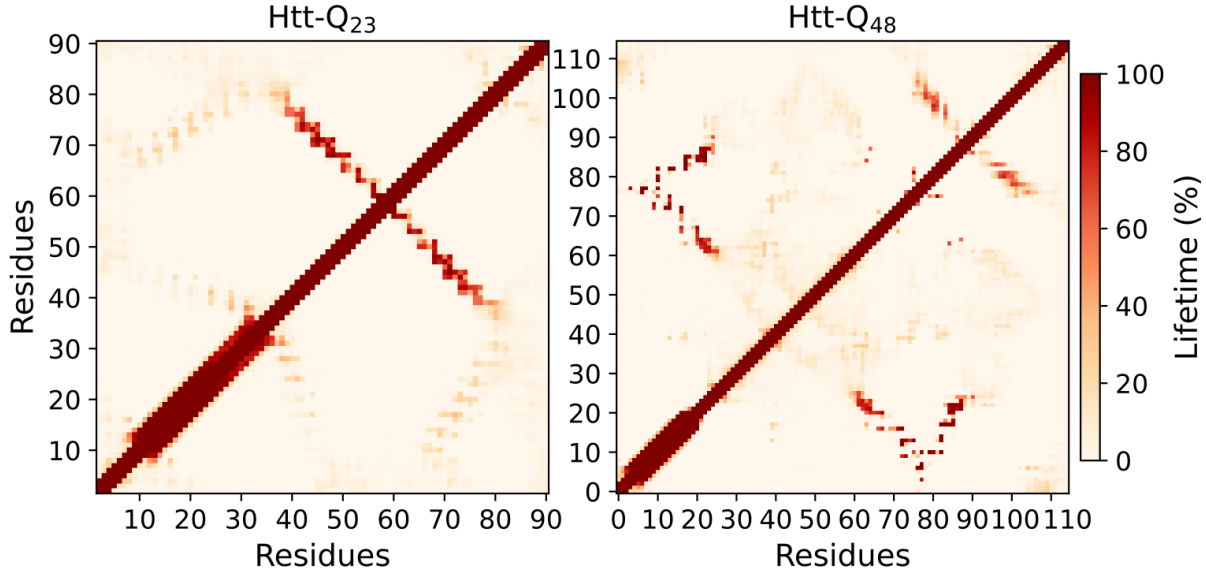

Figure S8: **The lifetime probability of the intra-protein interactions in the Htt-ex1 monomers.** The percent lifetimes for Htt-Q_{23} (left) and Htt-Q_{48} (right) are colored based on the scale on the right.

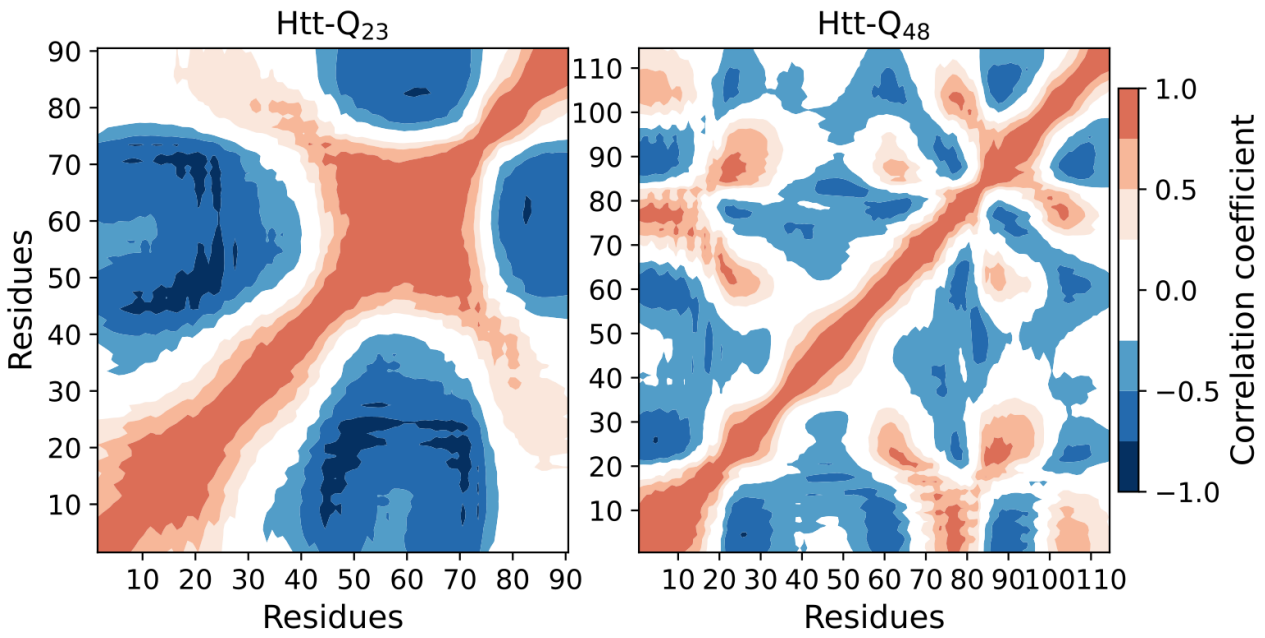

Figure S9: **Dynamic cross-correlation maps of the Htt-ex1 monomers.** Positively and negatively correlated motions are represented in red and blue, respectively, for Htt-Q_{23} (left) and Htt-Q_{48} (right).

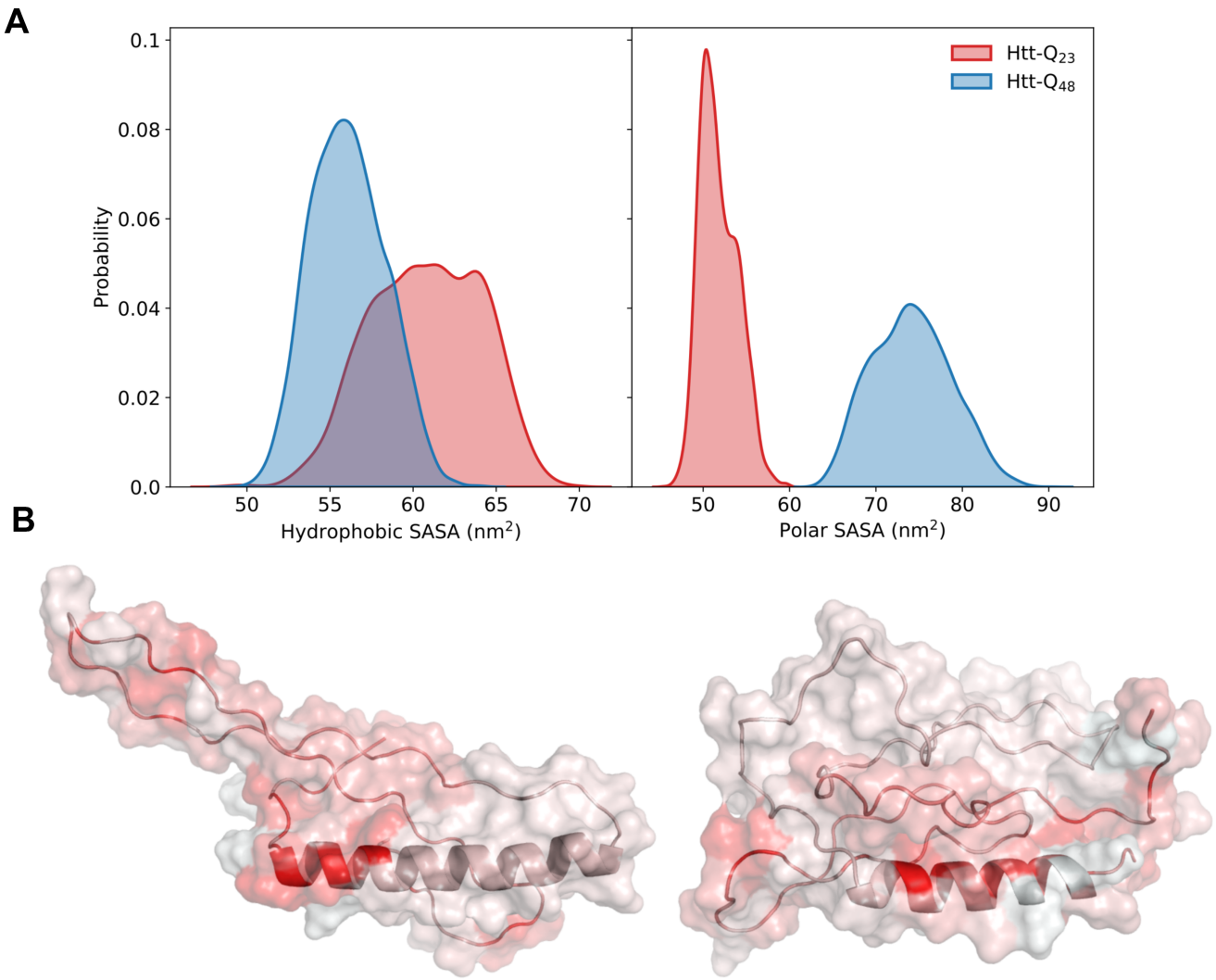

Figure S10: **Solvent accessible surface areas and hydrophobicity of the Htt-ex1 monomers. (A)** The hydrophobic (left) and polar (right) SASA distribution for Htt-Q_{23} (red) and Htt-Q_{48} (blue). **(B)** The most populated cluster structures for Htt-Q_{23} (left) and Htt-Q_{48} (right) colored according their hydrophobicity. The proteins are shown as cartoon plus their van der Waals surface (in transparent). The color represents the hydrophobicity with colors changing from white (not hydrophobic) to red for the amino acids of highest hydrophobicity.

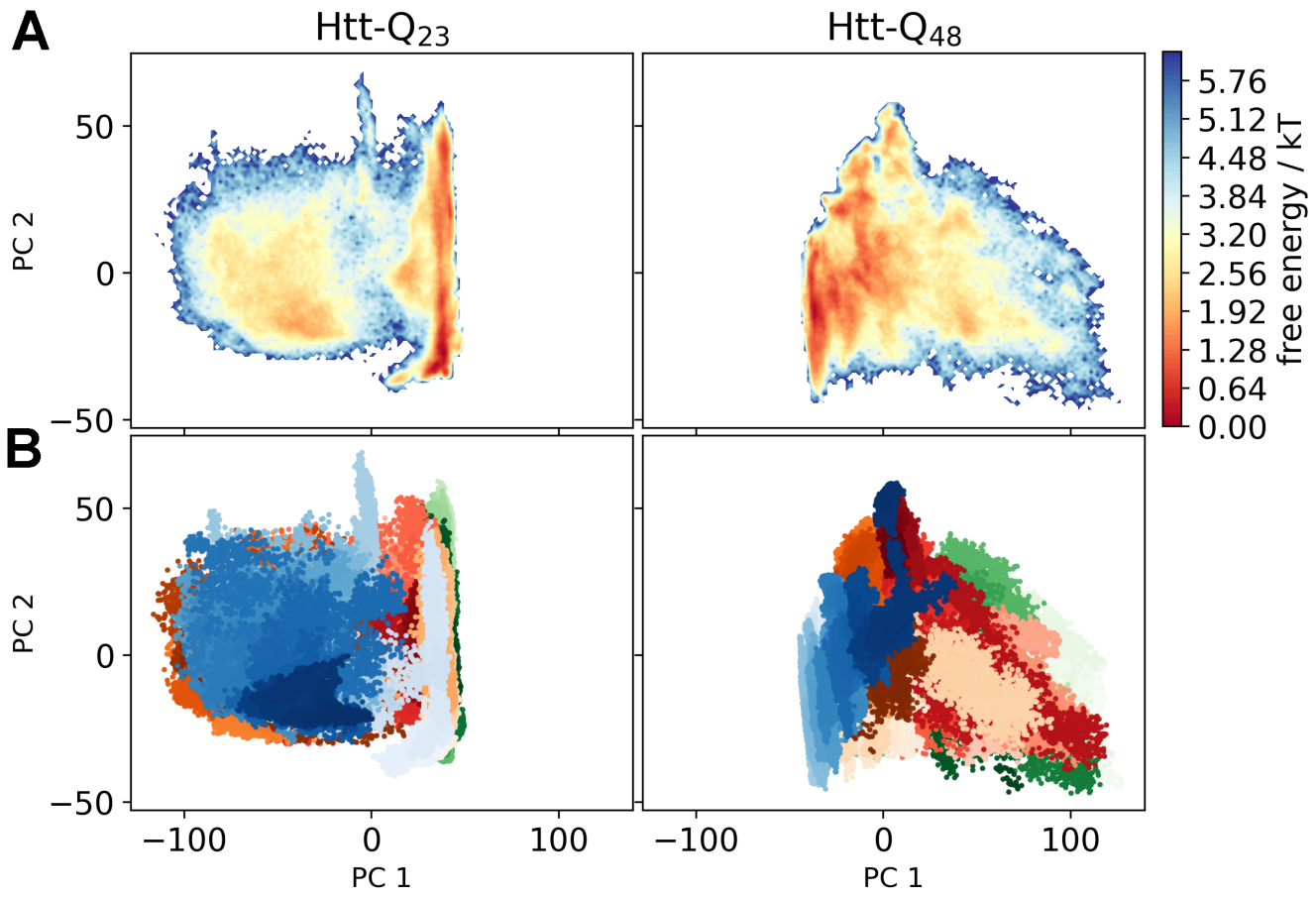

Figure S11: **Principal component analysis for the monomers of Htt-Q_{23} (left) and Htt-Q_{48} (right). (A)** The free energy surface as a function of the first two PCs. **(B)** Projection of individual trajectories onto their joint PC1-PC2 space. The color scale represents the space coverage of the individual trajectories (run 2, green; run 3, red; run 4, orange; run 5, blue), showing i) the conformational overlap between the individual trajectories and ii) that also trajectory-specific conformations were sampled.

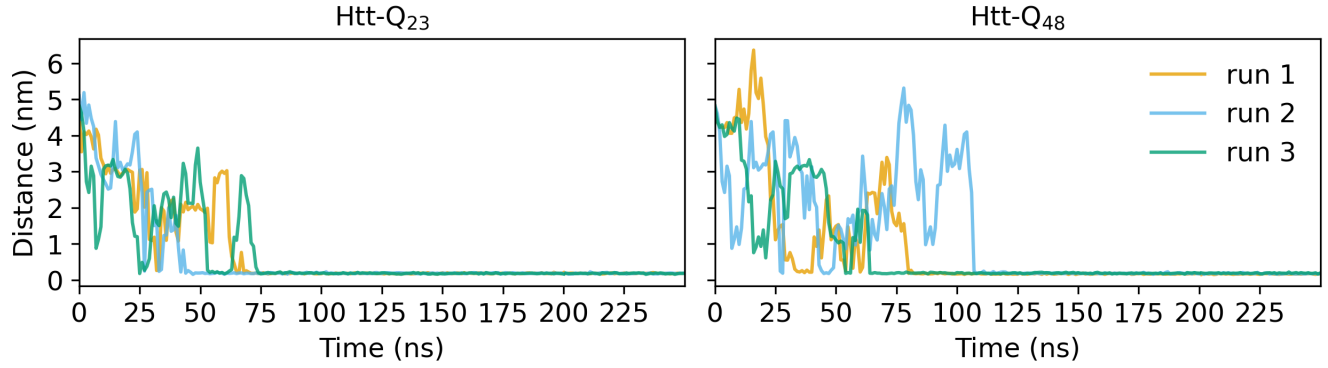

Figure S12: **The minimum distance between the two proteins in the dimer simulations.** The evolution of the inter-protein minimum distance during the three dimer simulations of Htt-Q_{23} (left) and Htt-Q_{48} (right).

Bibliography

1. Burns, A. & Iliffe, S. Alzheimer's disease. *BMJ* **338.** ISSN: 0959-8138 (2009).
2. Williams, A. J. & Paulson, H. L. Polyglutamine neurodegeneration: protein misfolding revisited. *Trends in neurosciences* **31,** 521–528 (2008).
3. Ankarcrona, M. *et al.* Current and future treatment of amyloid diseases. *Journal of Internal Medicine* **280,** 177–202 (2016).
4. Nagel-Steger, L., Owen, M. C. & Strodel, B. An Account of Amyloid Oligomers: Facts and Figures Obtained from Experiments and Simulations. *ChemBioChem* **17,** 657–676 (2016).
5. Pentony, M. M. & Jones, D. T. Modularity of intrinsic disorder in the human proteome. *Proteins: Structure, Function, and Bioinformatics* **78,** 212–221 (2010).
6. Carballo-Pacheco, M. & Strodel, B. Advances in the Simulation of Protein Aggregation at the Atomistic Scale. *The Journal of Physical Chemistry B* **120.** PMID: 26965454, 2991–2999 (2016).
7. Richardson, J. S. The anatomy and taxonomy of protein structure. *Advances in protein chemistry* **34,** 167–339 (1981).
8. Vihinen, M. Relationship of protein flexibility to thermostability. *Protein Engineering, Design and Selection* **1,** 477–480. ISSN: 1741-0126 (June 1987).
9. Fisher, C. K. & Stultz, C. M. Protein Structure along the Order–Disorder Continuum. *Journal of the American Chemical Society* **133.** PMID: 21650183, 10022–10025 (2011).
10. Strodel, B. Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins. *Journal of Molecular Biology* **433.** From Protein Sequence to Structure at Warp Speed: How Alphafold Impacts Biology, 167182. ISSN: 0022-2836 (2021).

11. Heller, G. T., Aprile, F. A. & Vendruscolo, M. Methods of probing the interactions between small molecules and disordered proteins. *Cellular and Molecular Life Sciences* **74,** 3225–3243 (2017).

12. Burger, V. M., Gurry, T. & Stultz, C. M. Intrinsically disordered proteins: where computation meets experiment. *Polymers* **6,** 2684–2719 (2014).

13. Oldfield, C. J. *et al.* Comparing and Combining Predictors of Mostly Disordered Proteins. *Biochemistry* **44.** PMID: 15697224, 1989–2000 (2005).

14. Uversky, V. N. & Dunker, A. K. Understanding protein non-folding. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1804,** 1231–1264. ISSN: 1570-9639 (2010).

15. Conway, K. A. *et al.* Acceleration of oligomerization, not fibrillization, is a shared property of both α-synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy. *Proceedings of the National Academy of Sciences* **97,** 571–576 (2000).

16. Perutz, M. F. Glutamine repeats and neurodegenerative diseases: molecular aspects. *Trends in Biochemical Sciences* **24,** 58–63. ISSN: 0968-0004 (1999).

17. Hardy, J. & Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **297,** 353–356 (2002).

18. Pitschke, M, Prior, R, Haupt, M & Riesner, D. Detection of single amyloid β-protein aggregates in the cerebrospinal fluid of Alzheimer's patients by fluorescence correlation spectroscopy. *Nature medicine* **4,** 832–834 (1998).

19. Iakoucheva, L. M., Brown, C. J., Lawson, J., Obradović, Z. & Dunker, A. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *Journal of Molecular Biology* **323,** 573–584. ISSN: 0022-2836 (2002).

20. Robustelli, P., Piana, S. & Shaw, D. E. Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. *Journal of the American Chemical Society* **142.** PMID: 32323533, 11092–11101 (2020).

21. McManus, J. J., Charbonneau, P., Zaccarelli, E. & Asherie, N. The physics of protein self-assembly. *Current opinion in colloid & interface science* **22,** 73–79 (2016).

22. Nielsen, L. *et al.* Effect of Environmental Factors on the Kinetics of Insulin Fibril Formation:. Elucidation of the Molecular Mechanism. *Biochemistry* **40.** PMID: 11352739, 6036–6046 (2001).

23. Alford, J. R., Kendrick, B. S., Carpenter, J. F. & Randolph, T. W. High concentration formulations of recombinant human interleukin-1 receptor antagonist: II. Aggregation kinetics. *Journal of pharmaceutical sciences* **97,** 3005–3021 (2008).

24. Li, D. & Liu, C. Structural Diversity of Amyloid Fibrils and Advances in Their Structure Determination. *Biochemistry* **59.** PMID: 31967790, 639–646 (2020).

25. Fitzpatrick, A. W. & Saibil, H. R. Cryo-EM of amyloid fibrils and cellular aggregates. *Current opinion in structural biology* **58,** 34–42 (2019).

26. Stefani, M. in *Protein Misfolding and Cellular Stress in Disease and Aging: Concepts and Protocols* (eds Bross, P. & Gregersen, N.) 25–41 (Humana Press, Totowa, NJ, 2010). ISBN: 978-1-60761-756-3.

27. Stefani, M. & Dobson, C. M. Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *Journal of molecular medicine* **81,** 678–699 (2003).

28. Knowles, T. P., Vendruscolo, M. & Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nature reviews Molecular cell biology* **15,** 384–396 (2014).

29. Selkoe, D. J. Folding proteins in fatal ways. *nature* **426,** 900–904 (2003).

30. Ferrone, F. in *Amyloid, Prions, and Other Protein Aggregates* 256–274 (Academic Press, 1999).

31. Wetzel, R. Kinetics and Thermodynamics of Amyloid Fibril Assembly. *Accounts of Chemical Research* **39.** PMID: 16981684, 671–679 (2006).

32. Kodali, R. & Wetzel, R. Polymorphism in the intermediates and products of amyloid assembly. *Current Opinion in Structural Biology* **17.** Folding and binding / Protein-nucleic interactions, 48–57. ISSN: 0959-440X (2007).

33. Sciacca, M. F., Tempra, C., Scollo, F., Milardi, D. & La Rosa, C. Amyloid growth and membrane damage: Current themes and emerging perspectives from theory and experiments on Aβ and hIAPP. *Biochimica et Biophysica Acta (BBA) - Biomembranes* **1860.** Protein Aggregation and Misfolding at the Cell Membrane Interface, 1625–1638. ISSN: 0005-2736 (2018).

34. J Fernandez-Perez, E., Peters, C. & G Aguayo, L. Membrane damage induced by amyloid beta and a potential link with neuroinflammation. *Current Pharmaceutical Design* **22,** 1295–1304 (2016).

35. Huang, J. *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature methods* **14,** 71–73 (2017).

36. Paul, A., Samantray, S., Anteghini, M., Khaled, M. & Strodel, B. Thermodynamics and kinetics of the amyloid-β peptide revealed by Markov state models based on MD data in agreement with experiment. *Chem. Sci.* **12,** 6652–6669 (19 2021).

37. Robustelli, P., Piana, S. & Shaw, D. Developing a molecular dynamics force field for both folded and disordered protein states. *Proc. Natl. Acad. Sci. U.S.A.* **115,** 201800690 (May 2018).

38. Samantray, S., Yin, F., Kav, B. & Strodel, B. Different Force Fields Give Rise to Different Amyloid Aggregation Pathways in Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **60.** PMID: 33174726, 6462–6475 (2020).

39. Chen, G.-f. *et al.* Amyloid beta: structure, biology and structure-based therapeutic development. *Acta Pharmacologica Sinica* **38,** 1205–1235 (2017).

40. Kang, J. *et al.* The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* **325,** 733–736 (1987).

41. Yoon, S.-S. & Jo, S. A. Mechanisms of amyloid-β peptide clearance: potential therapeutic targets for Alzheimer's disease. *Biomolecules & therapeutics* **20,** 245 (2012).

42. Liu, P.-P., Xie, Y., Meng, X.-Y. & Kang, J.-S. History and progress of hypotheses and clinical trials for Alzheimer's disease. *Signal transduction and targeted therapy* **4,** 1–22 (2019).

43. Haass, C. & Selkoe, D. J. Cellular processing of beta-amyloid precursor protein and the genesis of amyloid beta-peptide. *Cell* **75,** 1039–1042 (1993).

44. Glenner, G. G. & Wong, C. W. Alzheimer's disease: initial report of the purification and characterization of a novel cerebrovascular amyloid protein. *Biochemical and biophysical research communications* **120,** 885–890 (1984).

45. Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO molecular medicine* **8,** 595–608 (2016).

46. Bloom, G. S. Amyloid-β and tau: the trigger and bullet in Alzheimer disease pathogenesis. *JAMA neurology* **71,** 505–508 (2014).

47. Nunan, J. & Small, D. H. Regulation of APP cleavage by α-, β- and γ-secretases. *FEBS Letters* **483,** 6–10 (2000).

48. Marshall, K. E. *et al.* A critical role for the self-assembly of Amyloid-β1-42 in neurodegeneration. *Scientific reports* **6,** 1–13 (2016).

49. Ono, K., Condron, M. M. & Teplow, D. B. Structure–neurotoxicity relationships of amyloid β-protein oligomers. *Proceedings of the National Academy of Sciences* **106,** 14745–14750 (2009).

50. El-Agnaf, O. M., Mahil, D. S., Patel, B. P. & Austen, B. M. Oligomerization and Toxicity of β-Amyloid-42 Implicated in Alzheimer's Disease. *Biochemical and Biophysical Research Communications* **273,** 1003–1007. ISSN: 0006-291X (2000).

51. Findeis, M. A. The role of amyloid β peptide 42 in Alzheimer's disease. *Pharmacology & therapeutics* **116,** 266–286 (2007).

52. Kepp, K. P. Alzheimer's disease: How metal ions define β-amyloid function. *Coordination Chemistry Reviews* **351,** 127–159 (2017).

53. Broersen, K., Rousseau, F. & Schymkowitz, J. The culprit behind amyloid beta peptide related neurotoxicity in Alzheimer's disease: oligomer size or conformation? *Alzheimer's research & therapy* **2,** 1–14 (2010).

54. Müller-Schiffmann, A. *et al.* Amyloid-β dimers in the absence of plaque pathology impair learning and synaptic plasticity. *Brain* **139,** 509–525. ISSN: 0006-8950 (Dec. 2015).

55. Kirkitadze, M. D., Bitan, G. & Teplow, D. B. Paradigm shifts in Alzheimer's disease and other neurodegenerative disorders: The emerging role of oligomeric assemblies. *Journal of Neuroscience Research* **69,** 567–577 (2002).

56. Brinkmalm, G. *et al.* Identification of neurotoxic cross-linked amyloid-β dimers in the Alzheimer's brain. *Brain* **142,** 1441–1457 (2019).

57. Takeuchi, T. & Nagai, Y. Protein misfolding and aggregation as a therapeutic target for polyglutamine diseases. *Brain sciences* **7,** 128 (2017).

58. Orr, H. T. Polyglutamine neurodegeneration: expanded glutamines enhance native functions. *Current opinion in genetics & development* **22,** 251–255 (2012).

59. Ross, C. A. & Tabrizi, S. J. Huntington's disease: from molecular pathogenesis to clinical treatment. *The Lancet Neurology* **10,** 83–98. ISSN: 1474-4422 (2011).

60. Walker, F. O. Huntington's disease. *The Lancet* **369,** 218–228. ISSN: 0140-6736 (2007).

61. Kremer, B. *et al.* A worldwide study of the Huntington's disease mutation: the sensitivity and specificity of measuring CAG repeats. *New England Journal of Medicine* **330,** 1401–1406 (1994).

62. Lee, J.-M. *et al.* CAG repeat expansion in Huntington disease determines age at onset in a fully dominant fashion. *Neurology* **78,** 690–695 (2012).

63. Bäuerlein, F. J. *et al.* In situ architecture and cellular interactions of PolyQ inclusions. *Cell* **171,** 179–187 (2017).

64. Arrasate, M. & Finkbeiner, S. Protein aggregates in Huntington's disease. *Experimental Neurology* **238.** Special issue: Mechanisms of protein misfolding in stressors, 1–11. ISSN: 0014-4886 (2012).

65. Hoffner, G. & Djian, P. Monomeric, Oligomeric and Polymeric Proteins in Huntington Disease and Other Diseases of Polyglutamine Expansion. *Brain Sciences* **4,** 91–122. ISSN: 2076-3425 (2014).

66. Lieberman, A. P., Shakkottai, V. G. & Albin, R. L. Polyglutamine Repeats in Neurodegenerative Diseases. *Annual Review of Pathology: Mechanisms of Disease* **14.** PMID: 30089230, 1–27 (2019).

67. Miller, J. *et al.* Identifying polyglutamine protein species in situ that best predict neurodegeneration. *Nature chemical biology* **7,** 925–934 (2011).

68. Gkekas, I. *et al.* Oxidative Stress and Neurodegeneration: Interconnected Processes in PolyQ Diseases. *Antioxidants* **10,** 1450 (2021).

69. Bennett, M. J. *et al.* A linear lattice model for polyglutamine in CAG-expansion diseases. *Proceedings of the National Academy of Sciences* **99,** 11634–11639 (2002).

70. Guo, Q. *et al.* The cryo-electron microscopy structure of huntingtin. *Nature* **555,** 117–120 (2018).

71. Warner, J. B. I. *et al.* Monomeric Huntingtin Exon 1 Has Similar Overall Structural Features for Wild-Type and Pathological Polyglutamine Lengths. *Journal of the American Chemical Society* **139.** PMID: 28937758, 14456–14469 (2017).

72. Moldovean, S. N. & Chiş, V. Molecular Dynamics Simulations Applied to Structural and Dynamical Transitions of the Huntingtin Protein: A Review. *ACS Chemical Neuroscience* **11.** PMID: 31841621, 105–120 (2020).

73. Feng, X., Luo, S. & Lu, B. Conformation polymorphism of polyglutamine proteins. *Trends in biochemical sciences* **43,** 424–435 (2018).

74. Birol, M. & Melo, A. M. Untangling the Conformational Polymorphism of Disordered Proteins Associated With Neurodegeneration at the Single-Molecule Level. *Frontiers in Molecular Neuroscience* **12.** ISSN: 1662-5099 (2020).

75. Klein, F. A. *et al.* Pathogenic and non-pathogenic polyglutamine tracts have similar structural properties: towards a length-dependent toxicity gradient. *Journal of molecular biology* **371,** 235–244 (2007).

76. Sivanandam, V. *et al.* The aggregation-enhancing huntingtin N-terminus is helical in amyloid fibrils. *Journal of the American Chemical Society* **133,** 4558–4566 (2011).

77. Thakur, A. K. & Wetzel, R. Mutational analysis of the structural organization of polyglutamine aggregates. *Proceedings of the National Academy of Sciences* **99,** 17014–17019 (2002).

78. Chow, W. N. V., Luk, H. W., Chan, H. Y. E. & Lau, K.-F. Degradation of mutant huntingtin via the ubiquitin/proteasome system is modulated by FE65. *Biochemical Journal* **443,** 681–689 (2012).

79. Qin, Z.-H. *et al.* Huntingtin bodies sequester vesicle-associated proteins by a polyproline-dependent interaction. *Journal of Neuroscience* **24,** 269–281 (2004).

80. Duennwald, M. L., Jagadish, S., Muchowski, P. J. & Lindquist, S. Flanking sequences profoundly alter polyglutamine toxicity in yeast. *Proceedings of the National Academy of Sciences* **103,** 11045–11050 (2006).

81. Duennwald, M. L., Jagadish, S., Giorgini, F., Muchowski, P. J. & Lindquist, S. A network of protein interactions determines polyglutamine toxicity. *Proceedings of the National Academy of Sciences* **103,** 11051–11056 (2006).

82. Kim, M. W., Chelliah, Y., Kim, S. W., Otwinowski, Z. & Bezprozvanny, I. Secondary structure of Huntingtin amino-terminal region. *Structure* **17,** 1205–1212 (2009).

83. Kim, M. Beta conformation of polyglutamine track revealed by a crystal structure of Huntingtin N-terminal region with insertion of three histidine residues. *Prion* **7,** 221–228 (2013).

84. Michalek, M., Salnikov, E. S. & Bechinger, B. Structure and topology of the huntingtin 1–17 membrane anchor by a combined solution and solid-state NMR approach. *Biophysical journal* **105,** 699–710 (2013).

85. Tam, S. *et al.* The chaperonin TRiC blocks a huntingtin sequence element that promotes the conformational switch to aggregation. *Nature structural & molecular biology* **16,** 1279–1285 (2009).

86. Thakur, A. K. *et al.* Polyglutamine disruption of the huntingtin exon 1 N terminus triggers a complex aggregation mechanism. *Nature structural & molecular biology* **16,** 380–389 (2009).

87. Sahoo, B., Singer, D., Kodali, R., Zuchner, T. & Wetzel, R. Aggregation behavior of chemically synthesized, full-length huntingtin exon1. *Biochemistry* **53,** 3897–3907 (2014).

88. Hoop, C. L. *et al.* Polyglutamine amyloid core boundaries and flanking domain dynamics in huntingtin fragment fibrils determined by solid-state nuclear magnetic resonance. *Biochemistry* **53,** 6653–6666 (2014).

89. Pandey, N. K. *et al.* The 17-residue-long N terminus in huntingtin controls stepwise aggregation in solution and on membranes via different mechanisms. *Journal of Biological Chemistry* **293,** 2597–2605 (2018).

90. Isas, J. M., Langen, R. & Siemer, A. B. Solid-state nuclear magnetic resonance on the static and dynamic domains of huntingtin exon-1 fibrils. *Biochemistry* **54,** 3942–3949 (2015).

91. Bugg, C. W., Isas, J. M., Fischer, T., Patterson, P. H. & Langen, R. Structural features and domain organization of huntingtin fibrils. *Journal of Biological Chemistry* **287,** 31739–31746 (2012).

92. Bhattacharyya, A. *et al.* Oligoproline effects on polyglutamine conformation and aggregation. *Journal of molecular biology* **355,** 524–535 (2006).

93. Lakhani, V. V., Ding, F. & Dokholyan, N. V. Polyglutamine induced misfolding of huntingtin exon1 is modulated by the flanking sequences. *PLoS computational biology* **6,** e1000772 (2010).

94. Williamson, T. E., Vitalis, A., Crick, S. L. & Pappu, R. V. Modulation of polyglutamine conformations and dimer formation by the N-terminus of huntingtin. *Journal of molecular biology* **396,** 1295–1309 (2010).

95. Hollingsworth, S. A. & Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **99,** 1129–1143. ISSN: 0896-6273 (6 Sept. 2018).

96. Allen, M. P. *et al.* Introduction to molecular dynamics simulation. *Computational soft matter: from synthetic polymers to proteins* **23,** 1–28 (2004).

97. Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature 2004 433:7022* **433,** 128–132. ISSN: 1476-4687 (7022 Jan. 2005).

98. Lane, T. J., Shukla, D., Beauchamp, K. A. & Pande, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Current Opinion in Structural Biology* **23.** Folding and binding / Protein-nucleic acid interactions, 58–65. ISSN: 0959-440X (2013).

99. Różycki, B. & Boura, E. Large, dynamic, multi-protein complexes: a challenge for structural biology. *Journal of Physics: Condensed Matter* **26,** 463103 (2014).

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

101. Goossens, K. & De Winter, H. Molecular Dynamics Simulations of Membrane Proteins: An Overview. *Journal of Chemical Information and Modeling* **58.** PMID: 30336018, 2193–2202 (2018).

102. Meller, J. *et al.* Molecular dynamics. *Encyclopedia of life sciences* **18** (2001).

103. Boeyens*, J. C. & Comba*, P. Molecular mechanics: theoretical basis, rules, scope and limits. *Coordination Chemistry Reviews* **212,** 3–10. ISSN: 0010-8545 (2001).

104. Allinger, N. L. *Molecular structure: understanding steric and electronic effects from molecular mechanics* (John Wiley & Sons, 2010).

105. ANDO, T. & YAMATO, I. in *Quantum Bio-Informatics II* 228–241 ().

106. Becker, O. M., MacKerell, A. D., Roux, B. & Watanabe, M. *Computational biochemistry and biophysics* (Marcel Dekker New York, 2001).

107. González, M. Force fields and molecular dynamics simulations. *École thématique de la Société Française de la Neutronique* **12,** 169–200 (2011).

108. Jorgensen, W. L., Maxwell, D. S. & Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society* **118,** 11225–11236 (1996).

109. Hornak, V. *et al.* Comparison of multiple Amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function, and Bioinformatics* **65,** 712–725 (2006).

110. Brooks, B. R. *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4,** 187–217 (1983).

111. Cornell, W. D. *et al.* A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society* **117,** 5179–5197 (1995).

112. Scott, W. R. P. *et al.* The GROMOS Biomolecular Simulation Program Package. *The Journal of Physical Chemistry A* **103,** 3596–3607 (1999).

113. Jorgensen, W. L. & Tirado-Rives, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society* **110,** 1657–1666 (1988).

114. Mackerell Jr., A. D. Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* **25,** 1584–1604 (2004).

115. Mackerell Jr., A. D., Feig, M. & Brooks III, C. L. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *Journal of Computational Chemistry* **25,** 1400–1415 (2004).

116. MacKerell, A. D. J., Feig, M. & Brooks, C. L. Improved Treatment of the Protein Backbone in Empirical Force Fields. *Journal of the American Chemical Society* **126.** PMID: 14733527, 698–699 (2004).

117. Perez, A., MacCallum, J. L., Brini, E., Simmerling, C. & Dill, K. A. Grid-Based Backbone Correction to the ff12SB Protein Force Field for Implicit-Solvent Simulations. *Journal of Chemical Theory and Computation* **11.** PMID: 26574266, 4770–4779 (2015).

118. Tian, C. *et al.* ff19SB: Amino-Acid-Specific Protein Backbone Parameters Trained against Quantum Mechanics Energy Surfaces in Solution. *Journal of Chemical Theory and Computation* **16.** PMID: 31714766, 528–552 (2020).

119. Ye, W., Ji, D., Wang, W., Luo, R. & Chen, H.-F. Test and Evaluation of ff99IDPs Force Field for Intrinsically Disordered Proteins. *Journal of Chemical Information and Modeling* **55.** PMID: 25919886, 1021–1029 (2015).

120. Vega, C. & Abascal, J. L. F. Simulating water with rigid non-polarizable models: a general perspective. *Phys. Chem. Chem. Phys.* **13,** 19663–19688 (44 2011).

121. Jorgensen, W. L. Quantum and statistical mechanical studies of liquids. 10. Transferable intermolecular potential functions for water, alcohols, and ethers. Application to liquid water. *Journal of the American Chemical Society* **103,** 335–340 (1981).

122. Mahoney, M. W. & Jorgensen, W. L. Diffusion constant of the TIP5P model of liquid water. *The Journal of Chemical Physics* **114,** 363–366 (2001).

123. 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. *The Journal of Chemical Physics* **112,** 8910–8922 (2000).

124. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The missing term in effective pair potentials. *The Journal of Physical Chemistry* **91,** 6269–6271 (1987).

125. Cramer, C. J. & Truhlar, D. G. Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics. *Chemical Reviews* **99.** PMID: 11849023, 2161–2200 (1999).

126. Izadi, S. & Onufriev, A. V. Accuracy limit of rigid 3-point water models. *The Journal of Chemical Physics* **145,** 74501. ISSN: 00219606 (7 Aug. 2016).

127. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* **79,** 926–935 (1983).

128. Lifson, S., Hagler, A. T. & Dauber, P. Consistent force field studies of intermolecular forces in hydrogen-bonded crystals. 1. Carboxylic acids, amides, and the C=O...H– hydrogen bonds. *Journal of the American Chemical Society* **101,** 5111–5121 (1979).

129. Gunsteren, W. F. V. & Berendsen, H. J. C. A Leap-frog Algorithm for Stochastic Dynamics. *Molecular Simulation* **1,** 173–185 (1988).

130. Toxvaerd, S. Algorithms for canonical molecular dynamics simulations. *Molecular Physics* **72,** 159–168 (1991).

131. Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **159,** 98–103 (1 1967).

132. Grubmüller, H., Heller, H., Windemuth, A. & Schulten, K. Generalized Verlet Algorithm for Efficient Molecular Dynamics Simulations with Long-range Interactions. *Molecular Simulation* **6,** 121–142 (1991).

133. Martys, N. S. & Mountain, R. D. Velocity Verlet algorithm for dissipative-particle-dynamics-based models of suspensions. *Phys. Rev. E* **59,** 3733–3736 (3 1999).

134. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics* **76,** 637–649 (1982).

135. Streett, W., Tildesley, D. & Saville, G. Multiple time-step methods in molecular dynamics. *Molecular Physics* **35,** 639–648 (1978).

136. Akimov, A. V. & Kolomeisky, A. B. Recursive Taylor Series Expansion Method for Rigid-Body Molecular Dynamics. *Journal of Chemical Theory and Computation* **7.** PMID: 26598150, 3062–3071 (2011).

137. Andersen, H. C. Molecular dynamics simulations at constant pressure and/or temperature. *The Journal of Chemical Physics* **72,** 2384–2393 (1980).

138. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **52,** 255–268 (1984).

139. Nose, S. Constant-temperature molecular dynamics. *Journal of Physics: Condensed Matter* **2,** SA115 (1990).

140. Corti, D. S. Isothermal-isobaric ensemble for small systems. *Phys. Rev. E* **64,** 016128 (1 2001).

141. Yamauchi, M., Mori, Y. & Okumura, H. Molecular simulations by generalized-ensemble algorithms in isothermal–isobaric ensemble. *Biophysical Reviews* **11,** 457–469 (2019).

142. Rühle, V. Berendsen and nose-hoover thermostats. *Am. J. Phys,* 1–4 (2007).

143. Braga, C. & Travis, K. P. A configurational temperature Nosé-Hoover thermostat. *The Journal of Chemical Physics* **123,** 134101 (2005).

144. Brańka, A. C., Kowalik, M. & Wojciechowski, K. W. Generalization of the Nosé-Hoover approach. *The Journal of Chemical Physics* **119,** 1929–1936 (2003).

145. Lemak, A. S. & Balabaev, N. K. On The Berendsen Thermostat. *Molecular Simulation* **13,** 177–187 (1994).

146. Morishita, T. Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath. *The Journal of Chemical Physics* **113,** 2976–2982 (2000).

147. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics* **126,** 014101 (2007).

148. Bussi, G., Zykova-Timan, T. & Parrinello, M. Isothermal-isobaric molecular dynamics using stochastic velocity rescaling. *The Journal of Chemical Physics* **130,** 074101 (2009).

149. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81,** 3684–3690 (1984).

150. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* **52,** 7182–7190 (1981).

151. Nosé, S. A unified formulation of the constant temperature molecular dynamics methods. *The Journal of Chemical Physics* **81,** 511–519 (1984).

152. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31,** 1695–1697 (3 1985).

153. Weber, W, ‡, Hünenberger, P. H. & Mccammon, J. A. Molecular Dynamics Simulations of a Polyalanine Octapeptide under Ewald Boundary Conditions: Influence of Artificial Periodicity on Peptide Conformation. *Journal of Physical Chemistry B* **104,** 3668–3675 (2000).

154. Smith, W. G. *The Minimum Image Convention in Non-cubic Md Cells* in (1989).

155. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103,** 8577–8593 (1995).

156. Mitsutake, A., Mori, Y. & Okamoto, Y. in *Biomolecular Simulations: Methods and Protocols* (eds Monticelli, L. & Salonen, E.) 153–195 (Humana Press, Totowa, NJ, 2013). ISBN: 978-1-62703-017-5.

157. Barducci, A., Pfaendtner, J. & Bonomi, M. in *Molecular Modeling of Proteins* (ed Kukol, A.) 151–171 (Springer New York, New York, NY, 2015). ISBN: 978-1-4939-1465-4.

158. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences* **99,** 12562–12566 (2002).

159. Miao, Y., Feher, V. A. & McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **11.** PMID: 26300708, 3584–3595 (2015).

160. Sugita, Y. & Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **314,** 141–151. ISSN: 0009-2614 (1999).

161. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. Equation of State Calculations by Fast Computing Machines. *The Journal of Chemical Physics* **21,** 1087–1092 (1953).

162. Liao, Q. in *Computational Approaches for Understanding Dynamical Systems: Protein Folding and Assembly* (eds Strodel, B. & Barz, B.) 177–213 (Academic Press, 2020).

163. Qi, R., Wei, G., Ma, B. & Nussinov, R. in *Peptide Self-Assembly: Methods and Protocols* (eds Nilsson, B. L. & Doran, T. M.) 101–119 (Springer New York, New York, NY, 2018). ISBN: 978-1-4939-7811-3.

164. Schütte, C., Fischer, A, Huisinga, W & Deuflhard, P. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *Journal of Computational Physics* **151,** 146–168. ISSN: 0021-9991 (1999).

165. Noé, F. & Rosta, E. Markov Models of Molecular Kinetics. *The Journal of Chemical Physics* **151,** 190401 (2019).

166. Noé, F. & Clementi, C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *Journal of Chemical Theory and Computation* **11.** PMID: 26574285, 5002–5011 (2015).

167. Scherer, M. K. *et al.* PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *Journal of Chemical Theory and Computation* **11.** PMID: 26574340, 5525–5542 (2015).

168. Harrigan, M. P. *et al.* MSMBuilder: Statistical Models for Biomolecular Dynamics. *Biophysical Journal* **112,** 10–15. ISSN: 15420086 (1 Jan. 2017).

169. Hoffmann, M. *et al.* Deeptime: a Python library for machine learning dynamical models from time series data. *Machine Learning: Science and Technology* **3,** 015009 (2021).

170. Li, T., Ogihara, M. & Tzanetakis, G. *Music data mining* (CRC Press Boca Raton, FL, 2012).

171. Van Der Maaten, L., Postma, E., Van den Herik, J., *et al.* Dimensionality reduction: a comparative. *J Mach Learn Res* **10,** 13 (2009).

172. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science* **2,** 559–572 (1901).

173. Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noé, F. Identification of slow molecular order parameters for Markov model construction. *The Journal of Chemical Physics* **139,** 015102 (2013).

174. Schwantes, C. R. & Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *Journal of Chemical Theory and Computation* **9.** PMID: 23750122, 2000–2009 (2013).

175. Molgedey, L. & Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. *Phys. Rev. Lett.* **72,** 3634–3637 (23 1994).

176. Lloyd, S. Least squares quantization in PCM. *IEEE Transactions on Information Theory* **28,** 129–137 (1982).

177. Steinley, D. K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology* **59,** 1–34. ISSN: 2044-8317 (2006).

178. Vassilvitskii, S. & Arthur, D. *k-means++: The advantages of careful seeding* in *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (2006), 1027–1035.

179. Prinz, J.-H. *et al.* Markov models of molecular kinetics: Generation and validation. *The Journal of Chemical Physics* **134,** 174105 (2011).

180. Djurdjevac, N., Sarich, M. & Schütte, C. Estimating the Eigenvalue Error of Markov State Models. *Multiscale Modeling & Simulation* **10,** 61–81 (2012).

181. Sarich, M., Noé, F. & Schütte, C. On the Approximation Quality of Markov State Models. *Multiscale Modeling & Simulation* **8,** 1154–1177 (2010).

182. Marinari, E & Parisi, G. Simulated Tempering: A New Monte Carlo Scheme. *Europhysics Letters (EPL)* **19,** 451–458 (1992).

183. Hansmann, U. H. Parallel tempering algorithm for conformational studies of biological molecules. *Chemical Physics Letters* **281,** 140–150. ISSN: 0009-2614 (1997).

184. Souaille, M. & Roux, B. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Computer Physics Communications* **135,** 40–57. ISSN: 0010-4655 (2001).

185. Torrie, G. & Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **23,** 187–199. ISSN: 0021-9991 (1977).

186. Wu, H., Paul, F., Wehmeyer, C. & Noé, F. Multiensemble Markov models of molecular thermodynamics and kinetics. *Proceedings of the National Academy of Sciences* **113,** E3221–E3230 (2016).

187. Wu, H. & Noé, F. Optimal Estimation of Free Energies and Stationary Densities from Multiple Biased Simulations. *Multiscale Modeling & Simulation* **12,** 25–54 (2014).

188. Mey, A. S. J. S., Wu, H. & Noé, F. xTRAM: Estimating Equilibrium Expectations from Time-Correlated Simulation Data at Multiple Thermodynamic States. *Phys. Rev. X* **4,** 041018 (4 2014).

189. Wu, H., Mey, A. S. J. S., Rosta, E. & Noé, F. Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states. *The Journal of Chemical Physics* **141,** 214106 (2014).

190. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *The Journal of Chemical Physics* **129,** 124105 (2008).

191. Kong, A., McCullagh, P., Meng, X.-L., Nicolae, D. & Tan, Z. A theory of statistical models for Monte Carlo integration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **65,** 585–604 (2003).

192. Bartels, C. Analyzing biased Monte Carlo and molecular dynamics simulations. *Chemical Physics Letters* **331,** 446–454. ISSN: 0009-2614 (2000).

193. Vardi, Y. Empirical Distributions in Selection Bias Models. *The Annals of Statistics* **13,** 178–203. ISSN: 00905364. (2022) (1985).

194. Paul, F. *et al.* Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nature Communications 2018 8:1* **8,** 1–10. ISSN: 2041-1723 (1 Oct. 2017).

195. Rao, F. & Caflisch, A. The Protein Folding Network. *J. Mol. Biol.* **342,** 299–306. ISSN: 0022-2836 (2004).

196. Barz, B., Wales, D. J. & Strodel, B. A Kinetic Approach to the Sequence–Aggregation Relationship in Disease-Related Protein Assembly. *J. Phys. Chem. B* **118,** 1003–1011 (2014).

197. Illig, A.-M. & Strodel, B. Performance of Markov State Models and Transition Networks on Characterizing Amyloid Aggregation Pathways from MD Data. *J. Chem. Theory Comput.* **16,** 7825–7839 (2020).

198. Fatafta, H., Khaled, M., Owen, M. C., Sayyed-Ahmad, A. & Strodel, B. Amyloid-β peptide dimers undergo a random coil to β-sheet transition in the aqueous phase but not at the neuronal membrane. *Proceedings of the National Academy of Sciences* **118,** e2106210118 (2021).

199. Schäffler, M., Khaled, M. & Strodel, B. ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins. *Methods.* ISSN: 1046-2023 (2022).

200. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52.** Protein Folding, 99–105. ISSN: 1046-2023 (2010).

201. Kabsch, W. & Sander, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 (1983).

202. Touw, W. G. *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Research* **43,** D364–D368. ISSN: 0305-1048 (Oct. 2014).

203. Cecchini, M., Rao, F., Seeber, M. & Caflisch, A. Replica exchange molecular dynamics simulations of amyloid peptide aggregation. *J. Chem. Phys.* **121,** 10748–10756 (2004).

204. Osborne, K. L., Bachmann, M. & Strodel, B. Thermodynamic analysis of structural transitions during GNNQQNY aggregation. *Proteins: Struct. Funct. Bioinf.* **81,** 1141–1155 (2013).

205. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109,** 1528 –1532 (2015).

206. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *J. Comput. Chem.* **32,** 2319–2327 (2011).

207. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. *Proceedings of the International AAAI Conference on Web and Social Media* **3,** 361–362 (2009).

208. Dahlgren, K. N. *et al.* Oligomeric and Fibrillar Species of Amyloid-β Peptides Differentially Affect Neuronal Viability*. *Journal of Biological Chemistry* **277,** 32046–32053. ISSN: 0021-9258 (2002).

209. Iijima, K. *et al.* Dissecting the pathological effects of human Aβ40 and Aβ42 in *Drosophila*: A potential model for Alzheimer's disease. *Proceedings of the National Academy of Sciences* **101,** 6623–6628 (2004).

210. Jarrett, J. T., Berger, E. P. & Lansbury, P. T. The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry* **32,** 4693–4697. ISSN: 0006-2960 (18 May 1993).

211. JARRETT, J. T., BERGER, E. P. & LANSBURY JR., P. T. The C-Terminus of the β Protein is Critical in Amyloidogenesisa. *Annals of the New York Academy of Sciences* **695,** 144–148 (1993).

212. Hou, L. *et al.* 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. *Journal of the American Chemical Society* **126,** 1992–2005 (2004).

213. Riek, R., Güntert, P., Döbeli, H., Wipf, B. & Wüthrich, K. NMR studies in aqueous solution fail to identify significant conformational differences between the monomeric forms of two Alzheimer peptides with widely different plaque-competence, Aβ (1–40) ox and Aβ (1–42) ox. *European journal of biochemistry* **268,** 5930–5936 (2001).

214. Chiti, F. & Dobson, C. M. Protein Misfolding, Amyloid Formation, and Human Disease: A Summary of Progress Over the Last Decade. *Annual Review of Biochemistry* **86.** PMID: 28498720, 27–68 (2017).

215. Yu, L. *et al.* Structural Characterization of a Soluble Amyloid β-Peptide Oligomer. *Biochemistry* **48.** PMID: 19216516, 1870–1877 (2009).

216. Chandra, B. *et al.* Major Reaction Coordinates Linking Transient Amyloid-β Oligomers to Fibrils Measured at Atomic Level. *Biophysical Journal* **113,** 805–816. ISSN: 0006-3495 (2017).

217. Man, V. H., Nguyen, P. H. & Derreumaux, P. High-Resolution Structures of the Amyloid-β 1–42 Dimers from the Comparison of Four Atomistic Force Fields. *The Journal of Physical Chemistry B* **121.** PMID: 28538095, 5977–5987 (2017).

218. Dehabadi, M. H. & Firouzi, R. Constructing conformational library for amyloid-β42 dimers as the smallest toxic oligomers using two CHARMM force fields. *Journal of Molecular Graphics and Modelling* **115,** 108207. ISSN: 1093-3263 (2022).

219. Wälti, M. A. *et al.* Atomic-resolution structure of a disease-relevant Aβ(1-42) amyloid fibril. *Proceedings of the National Academy of Sciences* **113,** E4976–E4984 (2016).

220. Haupt, C. *et al.* Structural Basis of β-Amyloid-Dependent Synaptic Dysfunctions. *Angewandte Chemie International Edition* **51,** 1576–1579 (2012).

221. Vemulapalli, S. P. B., Becker, S., Griesinger, C. & Rezaei-Ghaleh, N. Combined High-Pressure and Multiquantum NMR and Molecular Simulation Propose a Role for N-Terminal Salt Bridges in Amyloid-Beta. *The Journal of Physical Chemistry Letters* **12.** PMID: 34617758, 9933–9939 (2021).

222. Huang, D. *et al.* Antiparallel β-Sheet Structure within the C-Terminal Region of 42-Residue Alzheimer's Amyloid-β Peptides When They Form 150-kDa Oligomers. *Journal of Molecular Biology* **427,** 2319–2328. ISSN: 0022-2836 (2015).

223. Schmidt, M. *et al.* Peptide dimer structure in an Aβ(1-42) fibril visualized with cryo-EM. *Proceedings of the National Academy of Sciences* **112,** 11858–11863 (2015).

224. Cerf, E. *et al.* Antiparallel β-sheet: a signature structure of the oligomeric amyloid β-peptide. *Biochemical Journal* **421,** 415–423. ISSN: 0264-6021 (July 2009).

225. Amyloid-β 2 protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimerś disease. *Nature Chemistry* **1,** 326–331. ISSN: 17554330 (4 July 2009).

226. Breuker, K. & McLafferty, F. W. Stepwise evolution of protein native structure with electrospray into the gas phase, 10E-12 to 10E2 s. *Proceedings of the National Academy of Sciences* **105,** 18145–18152 (2008).

227. Pham, J. D., Chim, N., Goulding, C. W. & Nowick, J. S. Structures of Oligomers of a Peptide from β-Amyloid. *Journal of the American Chemical Society* **135.** PMID: 23927812, 12460–12467 (2013).

228. Laganowsky, A. *et al.* Atomic View of a Toxic Amyloid Small Oligomer. *Science* **335,** 1228–1231 (2012).

229. Berhanu, W. M. & Hansmann, U. H. E. The stability of cylindrin β-barrel amyloid oligomer models—A molecular dynamics study. *Proteins: Structure, Function, and Bioinformatics* **81,** 1542–1555 (2013).

230. Österlund, N., Moons, R., Ilag, L. L., Sobott, F. & Gräslund, A. Native Ion Mobility-Mass Spectrometry Reveals the Formation of β-Barrel Shaped Amyloid-β Hexamers in a Membrane-Mimicking Environment. *Journal of the American Chemical Society* **141.** PMID: 31141355, 10440–10450 (2019).

231. Fu, Z. *et al.* Capping of Aβ42 Oligomers by Small Molecule Inhibitors. *Biochemistry* **53.** PMID: 25422864, 7893–7903 (2014).

232. Ahmed, M. *et al.* Structural conversion of neurotoxic amyloid-β1–42 oligomers to fibrils. *Nature structural & molecular biology* **17,** 561–567 (2010).

233. Gu, L. *et al.* Antiparallel triple-strand architecture for prefibrillar Aβ42 oligomers. *Journal of Biological Chemistry* **289,** 27300–27313 (2014).

234. Pan, J., Han, J., Borchers, C. H. & Konermann, L. Conformer-Specific Hydrogen Exchange Analysis of Aβ(1–42) Oligomers by Top-Down Electron Capture Dissociation Mass Spectrometry. *Analytical Chemistry* **83.** PMID: 21635007, 5386–5393 (2011).

235. Kłoniecki, M. *et al.* Ion Mobility Separation Coupled with MS Detects Two Structural States of Alzheimer's Disease Aβ1–40 Peptide Oligomers. *Journal of Molecular Biology* **407,** 110–124. ISSN: 0022-2836 (2011).

236. Banerjee, S., Sun, Z., Hayden, E. Y., Teplow, D. B. & Lyubchenko, Y. L. Nanoscale Dynamics of Amyloid β-42 Oligomers As Revealed by High-Speed Atomic Force Microscopy. *ACS Nano* **11.** PMID: 29165985, 12202–12209 (2017).

237. Barz, B., Liao, Q. & Strodel, B. Pathways of Amyloid-β Aggregation Depend on Oligomer Shape. *Journal of the American Chemical Society* **140.** PMID: 29235346, 319–327 (2018).

238. Nguyen, H. L., Krupa, P., Hai, N. M., Linh, H. Q. & Li, M. S. Structure and Physicochemical Properties of the Aβ42 Tetramer: Multiscale Molecular Dynamics Simulations. *The Journal of Physical Chemistry B* **123.** PMID: 31365254, 7253–7269 (2019).

239. Lührs, T. *et al.* 3D structure of Alzheimer's amyloid-β(1-42) fibrils. *Proceedings of the National Academy of Sciences* **102,** 17342–17347 (2005).

240. Xiao, Y. *et al.* Aβ(1–42) fibril structure illuminates self-recognition and replication of amyloid in Alzheimer's disease. *Nature Structural & Molecular Biology 2015 22:6* **22,** 499–505. ISSN: 1545-9985 (6 May 2015).

241. Im, D. *et al.* Kinetic Modulation of Amyloid-β (1–42) Aggregation and Toxicity by Structure-Based Rational Design. *Journal of the American Chemical Society* **144.** PMID: 35073692, 1603–1611 (2022).

242. Richter, L. *et al.* Amyloid beta 42 peptide (Aβ-42)-lowering compounds directly bind to Aβ and interfere with amyloid precursor protein (APP) transmembrane dimerization. *Proceedings of the National Academy of Sciences* **107,** 14597–14602 (2010).

243. Bieschke, J. *et al.* Small-molecule conversion of toxic oligomers to nontoxic β-sheet–rich amyloid fibrils. *Nature chemical biology* **8,** 93–101 (2012).

244. Van Der Spoel, D. *et al.* GROMACS: fast, flexible, and free. *Journal of computational chemistry* **26,** 1701–1718 (2005).

245. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1,** 19–25 (2015).

246. Zhang, J. The Hybrid Idea of (Energy Minimization) Optimization Methods Applied to Study Prion Protein Structures Focusing on the beta2-alpha2 Loop. *Biochem Pharmacol (Los Angel)* **4,** 2167–0501 (2015).

247. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *The Journal of chemical physics* **126,** 014101 (2007).

248. Parrinello, M & Rahman, A. Strain fluctuations and elastic constants. *The Journal of Chemical Physics* **76,** 2662–2666 (1982).

249. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied physics* **52,** 7182–7190 (1981).

250. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of chemical physics* **103,** 8577–8593 (1995).

251. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18,** 1463–1472 (1997).

252. Van Gunsteren, W. F. & Berendsen, H. J. A leap-frog algorithm for stochastic dynamics. *Molecular Simulation* **1,** 173–185 (1988).

253. Martínez, L., Andrade, R., Birgin, E. G. & Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *Journal of Computational Chemistry* **30,** 2157–2164 (2009).

254. Michaud-Agrawal, N., Denning, E. J., Woolf, T. B. & Beckstein, O. MDAnalysis: A toolkit for the analysis of molecular dynamics simulations. *Journal of Computational Chemistry* **32,** 2319–2327 (2011).

255. Ewing, S. A., Donor, M. T., Wilson, J. W. & Prell, J. S. Collidoscope: an improved tool for computing collisional cross-sections with the trajectory method. *Journal of The American Society for Mass Spectrometry* **28,** 587–596 (2017).

256. Pérez-Hernández, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noé, F. Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139,** 015102 (2013).

257. Sengupta, U., Carballo-Pacheco, M. & Strodel, B. Automated Markov state models for molecular dynamics simulations of aggregation and self-assembly. *The Journal of Chemical Physics* **150,** 115101 (2019).

258. Prinz, J.-H. *et al.* Markov models of molecular kinetics: Generation and validation. *J. Chem. Phys.* **134,** 174105 (2011).

259. Röblitz, S. & Weber, M. Fuzzy spectral clustering by PCCA+: Application to Markov state models and data classification. *Adv. Data Anal. Classi.* **7,** 147–179 (June 2013).

260. Metzner, P., Schütte, C. & Vanden-Eijnden, E. Transition Path Theory for Markov Jump Processes. *Multiscale Modeling & Simulation* **7,** 1192–1219 (2009).

261. E, W. & Vanden-Eijnden, E. Towards a Theory of Transition Paths. *Journal of Statistical Physics 2006 123:3* **123,** 503–523. ISSN: 1572-9613 (3 May 2006).

262. Schäffler, M., Khaled, M. & Strodel, B. ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins. *Methods* **206,** 18–26. ISSN: 1046-2023 (2022).

263. Illig, A.-M. & Strodel, B. Performance of Markov State Models and Transition Networks on Characterizing Amyloid Aggregation Pathways from MD Data. *Journal of Chemical Theory and Computation* **16.** PMID: 33233894, 7825–7839 (2020).

264. Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* **2008,** P10008 (2008).

265. Scarff, C. A., Ashcroft, A. E. & Radford, S. E. in *Protein Amyloid Aggregation: Methods and Protocols* (ed Eliezer, D.) 115–132 (Springer New York, New York, NY, 2016). ISBN: 978-1-4939-2978-8.

266. Sandberg, A. *et al.* Stabilization of neurotoxic Alzheimer amyloid-β oligomers by protein engineering. *Proceedings of the National Academy of Sciences* **107,** 15595–15600 (2010).

267. Seo, J. *et al.* An infrared spectroscopy approach to follow β-sheet formation in peptide amyloid assemblies. *Nature chemistry* **9,** 39–44 (2017).

268. Posse de Chaves, E. & Sipione, S. Sphingolipids and gangliosides of the nervous system in membrane function and dysfunction. *FEBS Letters* **584.** Frontiers in Membrane Biochemistry, 1748–1759. ISSN: 0014-5793 (2010).

269. Korade, Z. & Kenworthy, A. K. Lipid rafts, cholesterol, and the brain. *Neuropharmacology* **55,** 1265–1273. ISSN: 0028-3908 (2008).

270. Kao, Y.-C., Ho, P.-C., Tu, Y.-K., Jou, I.-M. & Tsai, K.-J. Lipids and Alzheimer's Disease. *International Journal of Molecular Sciences* **21.** ISSN: 1422-0067 (2020).

271. Nagai, Y. *et al.* A toxic monomeric conformer of the polyglutamine protein. *Nature structural & molecular biology* **14,** 332–340 (2007).

272. Heck, B. S., Doll, F. & Hauser, K. Length-dependent conformational transitions of polyglutamine repeats as molecular origin of fibril initiation. *Biophys. Chem.* **185,** 47–57 (2014).

273. Kang, H. *et al.* Emerging β-sheet rich conformations in supercompact huntingtin exon-1 mutant structures. *J. Am. Chem. Soc.* **139,** 8820–8827 (2017).

274. Michaels, T. C. *et al.* Dynamics of oligomer populations formed during the aggregation of Alzheimer's Aβ42 peptide. *Nature chemistry* **12,** 445–451 (2020).

275. Glabe, C. G. & Kayed, R. Common structure and toxic function of amyloid oligomers implies a common mechanism of pathogenesis. *Neurology* **66,** S74–S78. ISSN: 0028-3878 (2006).

276. Nicastro, L. K. *et al.* Cytotoxic Curli Intermediates Form during Salmonella Biofilm Development. *Journal of Bacteriology* **201,** e00095–19 (2019).

277. Demuro, A. *et al.* Calcium Dysregulation and Membrane Disruption as a Ubiquitous Neurotoxic Mechanism of Soluble Amyloid Oligomers. *Journal of Biological Chemistry* **280,** 17294–17300 (2005).

278. Sciacca, M. F. *et al.* Two-step mechanism of membrane disruption by Aβ through membrane fragmentation and pore formation. *Biophysical journal* **103,** 702–710 (2012).

279. Vabulas, R. M. & Hartl, F. U. *Aberrant protein interactions in amyloid disease* 2011.

280. Radwan, M., Wood, R. J., Sui, X. & Hatters, D. M. When proteostasis goes bad: Protein aggregation in the cell. *IUBMB Life* **69,** 49–54 (2017).