

# **Exploring the Free Energy Landscape and Conformational Dynamics of Intrinsically Disordered Proteins and Amyloid Aggregation**

Inaugural Dissertation

by

**Moritz Schäffler**

for the attainment of the title of doctor  
in the Faculty of Mathematics and Natural Sciences

at

Heinrich-Heine-Universität Düsseldorf

Cologne, November 2024

Supervisor:

Prof. Dr. Birgit Strodel

From the Institute for Biological Information Processing: Structural Biochemistry  
(IBI-7) at the Forschungszentrum Jülich

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

1. Reviewer: Prof. Dr. Birgit Strodel
2. Reviewer: Prof. Dr. Holger Gohlke

Date of the oral examination:  
13.06.2025



# Selbstständigkeitserklärung / Declaration of Authorship

Hiermit erkläre ich, Moritz Schäffler, dass ich die vorliegende Arbeit mit dem Titel „*Exploring the Free Energy Landscape and Conformational Dynamics of Intrinsically Disordered Proteins and Amyloid Aggregation*“ selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Die Stellen der Arbeit sowie eventuell beigefügte Zeichnungen, Skizzen, graphische Darstellungen oder Daten, die anderen Werken entnommen oder an diese angelehnt sind, habe ich unter Angabe der Quelle als Entlehnung kenntlich gemacht.

I, Moritz Schäffler, hereby declare that I have authored the thesis entitled “*Exploring the Free Energy Landscape and Conformational Dynamics of Intrinsically Disordered Proteins and Amyloid Aggregation*”. No other person’s work has been used without due acknowledgement in this thesis. Where I have quoted from or adapted the work of others, the source is always given. This includes the sources of graphs and data.

Düsseldorf, \_\_\_\_\_

\_\_\_\_\_  
Moritz Schäffler



# Contents

<b>Acknowledgements</b>	<b>ix</b>
<b>Abstract</b>	<b>xi</b>
<b>Publications</b>	<b>xiii</b>
<b>List of Abbreviations</b>	<b>xvi</b>
<b>List of Figures</b>	<b>xviii</b>
<b>List of Tables</b>	<b>xviii</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Aim of the Thesis . . . . .	3
1.2 Intrinsically Disordered Proteins . . . . .	3
1.3 Amyloid Aggregation . . . . .	5
1.4 Amyloid- $\beta$ peptide . . . . .	7
1.5 Parathyroid Hormone . . . . .	9
1.6 SH3-domain . . . . .	10
<b>2 Methods</b>	<b>13</b>
2.1 Molecular Dynamics Simulation . . . . .	13
2.1.1 Force Field . . . . .	13
2.1.2 The CHARMM36m Force Field . . . . .	16
2.1.3 TIP3P Water Model . . . . .	17
2.1.4 Periodic Boundary Conditions and Cutoff Methods . . . . .	17
2.1.5 Numerical Integration . . . . .	18
2.1.6 Thermodynamical Ensembles in MD Simulations . . . . .	20
2.2 Enhanced Sampling . . . . .	22
2.2.1 Multistate Bennett Acceptance Ratio method . . . . .	24
2.2.2 Principle Component Analysis . . . . .	25
2.3 Transition Networks . . . . .	25
2.4 Distribution of Reciprocal Interatomic Distances . . . . .	26
2.5 Free Energy Calculation . . . . .	27
2.6 Disconnectivity Graphs . . . . .	28
2.7 First Passage Times . . . . .	29

<b>3</b>	<b>Automated Generation of Transition Networks for the Structural Characterization of Intrinsically Disordered Proteins</b>	<b>31</b>
3.1	Methods . . . . .	33
3.2	Results . . . . .	34
3.3	Conclusion . . . . .	36
<b>4</b>	<b>Transition Networks Reveal the Conformational Switching of A<math>\beta</math><sub>42</sub></b>	<b>37</b>
4.1	Methods . . . . .	39
4.2	Results . . . . .	40
4.2.1	Transition Network of A $\beta$ in solution . . . . .	40
4.2.2	Transition Network of A $\beta$ -GAG . . . . .	42
4.2.3	Transition Network of A $\beta$ -POPC . . . . .	42
4.2.4	A $\beta$ -GAG interaction . . . . .	44
4.3	Discussion . . . . .	44
<b>5</b>	<b>Free Energy Surface and Timescale Analysis of the A<math>\beta</math><sub>42</sub> peptide</b>	<b>47</b>
5.1	Methods . . . . .	49
5.2	Results . . . . .	50
5.2.1	A $\beta$ <sub>42</sub> Monomer . . . . .	50
5.2.2	A $\beta$ <sub>42</sub> Dimer . . . . .	50
5.2.3	A $\beta$ <sub>42</sub> with a GAG . . . . .	53
5.2.4	A $\beta$ <sub>42</sub> with POPC lipids . . . . .	56
5.2.5	Timescale Analysis . . . . .	58
5.3	Discussion . . . . .	59
<b>6</b>	<b>Photocontrolled Reversible Amyloid Fibril Formation of PTH</b>	<b>63</b>
6.1	Structural Analysis . . . . .	66
6.2	Peptide Simulations . . . . .	66
6.2.1	Simulation Details . . . . .	67
6.2.2	Results . . . . .	67
6.3	Fibril Model . . . . .	70
6.3.1	Simulation Details . . . . .	70
6.3.2	Results . . . . .	70
6.4	Conclusion . . . . .	72
<b>7</b>	<b>Influence of pH on the Aggregation Behaviour of the SH3 Domain</b>	<b>73</b>
7.1	Motivation . . . . .	73
7.2	Structure Preparation . . . . .	74
7.3	Conventional MD - Monomer . . . . .	76
7.3.1	Simulation Methods . . . . .	77
7.3.2	Results . . . . .	77
7.3.3	Discussion . . . . .	80
7.4	Conventional MD - Dimer . . . . .	83

7.4.1	Simulation Methods . . . . .	83
7.4.2	Results . . . . .	83
7.4.3	Discussion . . . . .	83
7.5	Replica Exchange MD . . . . .	85
7.5.1	Simulation Methods . . . . .	85
7.5.2	Convergence . . . . .	85
7.5.3	Results . . . . .	87
7.5.4	Discussion . . . . .	91
7.6	Fibril Unfolding . . . . .	92
7.6.1	Simulation Methods . . . . .	92
7.6.2	Results . . . . .	92
7.6.3	Discussion . . . . .	94
7.7	Early Folding Events . . . . .	96
7.7.1	Simulation Methods . . . . .	96
7.7.2	Results . . . . .	97
7.7.3	Discussion . . . . .	97
7.8	Conclusion . . . . .	100
<b>8</b>	<b>Conclusion and Final Remarks</b>	<b>103</b>
	<b>Bibliography</b>	<b>107</b>
<b>A</b>	<b>Appendix</b>	<b>121</b>
A.1	Supplementary Figures . . . . .	121
A.2	Publication I: ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins . . . . .	122
A.3	Publication II: Transition Networks Unveil Disorder-to-Order Transformations in A $\beta$ Caused by Glycosaminoglycans or Lipids . . . . .	131
A.4	Publication III: The energy landscape of A $\beta$ 42 - a funnel to disorder for the monomer becomes a folding funnel for self-assembly . . . . .	155
A.5	Publication IV: Photocontrolled Reversible Amyloid Fibril Formation of Parathyroid Hormone-Derived Peptides . . . . .	168



# Acknowledgements

First and foremost, I would like to express my deepest gratitude to Prof. Dr. Birgit Strodel for her unwavering supervision and guidance throughout my PhD journey. Her support, insightful feedback, and encouragement have been invaluable, and I am truly grateful for the opportunity she gave me to pursue research in her group. I would also like to extend my sincere thanks to Prof. Dr. David Wales, with whom I had the privilege of studying in Cambridge. His expertise and insights elevated my research to new heights, and I deeply appreciate his significant contributions to my academic development.

Furthermore, I would like to extend my appreciation to all of my fellow group members for creating such a positive and collaborative working environment, where everyone is always willing to help one another. In particular, I want to thank Anna Jäckering and Lara Schabert, not only for making the occasional trip to Jülich worthwhile but also for their invaluable help with proofreading my thesis. Your friendship and support have been greatly appreciated.

I am deeply thankful to my friends and family for their endless support, patience, and belief in me during the challenging times of this journey. Their encouragement has kept me grounded and motivated throughout.

This thesis would not have been possible without the contributions of all these wonderful people, and I am truly grateful for their roles in my academic and personal growth.

Lastly, I gratefully acknowledge the computing time granted through JARA-HPC (project *idp*) on the supercomputer JURECA at Forschungszentrum Jülich, without which my research would not have been possible.





# Abstract

Intrinsically disordered proteins (IDPs) play a pivotal role in amyloid aggregation, a process linked to neurodegenerative diseases such as Alzheimer's and Parkinson's. Due to their flexible structures, IDPs can transition from disordered to ordered states upon interaction with molecular partners, often leading to the formation of amyloid fibrils. This thesis examines the mechanisms of IDP conformation switching and aggregation using molecular dynamics (MD) simulations and free energy surface (FES) analyses to elucidate structural transitions, fibrillization pathways and early aggregation events, i.e. oligomer formation. In this work, three IDPs involved in amyloid formation are studied: the amyloid- $\beta$  protein ( $A\beta_{42}$ ), a segment of the parathyroid hormone ( $PTH_{1-34}$ ), and the SH3 domain of the bovine phosphatidylinositol-3-kinase (PI3K-SH3). For  $A\beta_{42}$ , we analyze its configurational space in various molecular environments, highlighting its transition from a disordered state to a toxic  $\beta$ -hairpin structure. By employing transition networks and FES calculations, we reveal that  $A\beta_{42}$  as a monomer in solution exhibits a 'structurally inverted folding funnel', where disordered states dominate. Upon dimerization, the FES resembles that of a folded protein, with a singular folding funnel and a  $\beta$ -hairpin structure at the global minimum. A similar change in the FES is observed if  $A\beta_{42}$  interacts with either a glycoaminoglycan or lipids, which both promote the formation of intrapeptide  $\beta$ -sheets. For  $PTH_{25-37}$ , we study photoinduced reversible amyloid fibrillization by incorporating a photoswitch, 3-[(4-aminomethyl)phenyl]diazanylbenzoic acid, at various positions in the peptide. Through computational design, synthesis, and experimental validation, we demonstrate how the mutation position impacts fibril formation and develop a comprehensive fibril model for  $PTH_{25-37}$  and one of its mutants. In this study, we demonstrate how light-induced switching between the *cis* and *trans* conformations of the photoswitch enables controlled, reversible transitions between an amorphous state (*cis*) and fibril formation (*trans*). These structural propensities are further explained through comprehensive MD simulations. Lastly, we investigate pH-dependent amyloid aggregation in the context of the PI3K-SH3 domain. Using extensive MD simulations, we elucidate the structural transition from a folded state at neutral pH to a disordered state at acidic pH, which drives amyloid formation. Our analysis identifies key features of pH-driven unfolding and aggregation, providing a framework for future experimental validation. In summary, this thesis offers insights into IDP behavior during amyloid aggregation, focusing on disorder-to-order transitions and FES changes across different environments. Additionally, we examine the role of pH and photo-controlled aggregation in modulating fibril formation, shedding light on key features driving these processes.



# Publications

## Publications that are part of the thesis

- [I] M. Schäffler, M. Khaled and B. Strodel: ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins. *Methods* **206** (2022)
- [II] M. Schäffler, S. Samantray and B. Strodel: Transition Networks Unveil Disorder-to-Order Transformations in A $\beta$  Caused by Glycosaminoglycans or Lipids. *Int. J. Mol. Sci.* **24** (2023)
- [III] A. Paschold<sup>†</sup>, M. Schäffler<sup>†</sup>, X. Miao, L. Gardon, S. Krüger, H. Heise, M. I. S. Röhr, M. Ott, B. Strodel and W. H. Binder: Photocontrolled Reversible Amyloid Fibril Formation of Parathyroid Hormone-Derived Peptides. *Bioconjug. Chem.* **35** (2024)
- [IV] M. Schäffler, D. Wales and B. Strodel: The energy landscape of A $\beta_{42}$ : a funnel to disorder for the monomer becomes a folding funnel for self-assembly *Chem. Commun.* **60** (2024)

## Publications beyond the scope of this thesis

- [I] J. Loschwitz, N. Steffens, X. Wang, M. Schäffler, K. Pfeffer, D. Degrandi and B. Strodel: Domain motions, dimerization, and membrane interactions of the murine guanylate binding protein 2. *Sci Rep* **13** 679 (2023).
- [II] A. Jäckering, F. Göttisch, M. Schäffler, M. Doerr, U.T. Bornscheuer, R. Wei and B. Strodel: From bulk to binding: Decoding the entry of PET into hydrolase binding pockets. *JACS Au* (2024).

---

<sup>†</sup>The authors have contributed equally to the work.



# List of Abbreviations

**AA** All-Atom. 13

**AA MD** All-Atom Molecular Dynamics. 13

**AD** Alzheimer's disease. 1, 8, 73

**AFM** Atomic Force Microscopy. 2

**AMBER** Assisted Model Building and Energy Refinement. 14

**AMPB** 3-[[[4-aminomethyl]phenyl]diazenyl]benzoic acid. xvii, xviii, 9, 10, 63, 64, 66–72, 104, 105

**APP** Amyloid Precursor Protein. 7, 8

**A $\beta$**  Amyloid- $\beta$ . vi, xvii, 3, 4, 6–9, 33, 34, 36, 37, 39–61, 69, 73, 77, 103–105

**CG** Coarse Grained. 13

**CG MD** Coarse-Grained Molecular Dynamics. 13

**CGenFF** CHARMM General Force Field. 16

**CHARMM** Chemistry at HARvard Macromolecular Mechanics. 14, 16, 17

**CMAP** Correction Map. 16

**cryo-EM** single-particle cryo-Electron Microscopy. 1, 2

**CV** Collective Variable. 24, 25, 28

**DRID** Distribution of Reciprocal Interatomic Distance. 26–28, 49–57

**FES** Free Energy Surface. xvii, xviii, 3, 25, 27–29, 36, 40, 47, 49–61, 87–91, 99–101, 103–105

**FF** Force Field. 13, 14

**FPT** First Passage Time. 13, 29, 30, 49, 58, 59, 103, 104

**GAG** Glycosaminoglycan. xvii, 37, 40, 42–47, 53, 54, 56, 58–61, 103, 104

**IDP** Intrinsically disordered protein. 1–5, 7, 16, 23–26, 29, 31, 36, 44, 47, 60, 67, 73, 77, 103, 105

**MBAR** Multistate Bennett Acceptance Ratio. 24, 25

**MD** Molecular Dynamics. 2, 3, 7, 10, 13–22, 25–28, 31, 33, 36, 37, 56, 59, 60, 63, 67, 70–72, 74, 76, 77, 80, 83, 85, 87, 89, 91, 92, 96, 100, 103–105

**MFPT** Mean First Passage Time. 29

**NMR** Nuclear Magnetic Resonance. 1, 2, 17, 47, 53, 64, 70, 71, 73, 74, 77, 80, 81, 101, 105

**PBC** Periodic Boundary Conditions. 16–18, 77, 83

**PC** Principal Component. 25, 87–89

**PCA** Principal Component Analysis. 13, 25, 87

**PD** Parkinson’s disease. 1

**POPC** 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine. vi, xvii, 31, 33–37, 40, 42, 44, 46, 47, 50, 53, 56–61, 103, 104

**PTH** Parathyroid Hormone. xvii, xviii, 3, 9, 63, 64, 66–72, 104, 105

**QM** Quantum Mechanics. 3, 13, 16

**REMD** Replica Exchange Molecular Dynamics. 23–25, 74, 85, 91, 100, 103, 105

**SAXS** Small Angle X-ray Scattering. 2, 17

**SH3** Src-homology 3. vi, xviii, 3, 10, 73–96, 98–101

**TN** Transition Network. 13, 25, 26, 31, 33, 34, 36, 37, 39–44, 46, 47, 54, 103, 104

**UB** Urey-Bradley. 16

**WAXS** Wide Angle X-ray Scattering. 70, 71

# List of Figures

1.1	Schematic overview over time- and length scales in biological methods.	2
1.2	Schematic representation of the free energy landscape of a folded and disordered protein. . . . .	4
1.3	Schematic representation of amyloid aggregation phases. . . . .	5
1.4	Human APP proteolysis in the amyloidogenic pathway. . . . .	7
1.5	Sequence of the A $\beta_{42}$ peptide and prominent structures. . . . .	8
1.6	Structural overview of the PTH <sub>25–37</sub> peptide and AMPB photo switch. .	9
1.7	Structural overview of the bovine PI3K-SH3 domain. . . . .	11
2.1	Replica exchange molecular dynamics simulation scheme. . . . .	23
2.2	Schematic representation of a disconnectivity graph. . . . .	29
3.1	Transition network of A $\beta_{42}$ forming a complex with three POPC lipids.	35
4.1	Chemical structures of the glycosaminoglycan and lipid studied in this work. . . . .	40
4.2	Transition network of the A $\beta$ monomer in solution. . . . .	41
4.3	Transition network of the A $\beta$ peptide in the presence interaction partners.	43
4.4	Contact analysis of A $\beta$ with a GAG molecule. . . . .	45
5.1	Free energy disconnectivity graph for the FES of the A $\beta_{42}$ monomer. . .	51
5.2	Free energy disconnectivity graph for the FES of the A $\beta_{42}$ dimer. . . .	52
5.3	Fastest pathway from the A $\beta_{42}$ dimer disordered state to the global minimum . . . . .	53
5.4	Free energy disconnectivity graph for the FES of A $\beta_{42}$ with a GAG. . . .	55
5.5	Free energy disconnectivity graph for the FES of A $\beta_{42}$ with three POPC lipids. . . . .	57
5.6	First passage time probability distributions for interconversions between disordered and $\beta$ -hairpin states. . . . .	58
6.1	Schematic representation of light-induced (de-)aggregation of PTH <sub>25–37</sub> with embedded AMPB photoswitch. . . . .	63
6.2	Electrostatic potential surface of PTH <sub>25–37</sub> and <i>trans</i> -P4. . . . .	66
6.3	Comparison between PTH <sub>25–37</sub> and <i>trans/cis</i> -P4. . . . .	68
6.4	Average simulated secondary structure population of PTH <sub>25–37</sub> and simulated mutants in <i>trans/cis</i> conformation. . . . .	70

6.5	Fibril model of PTH <sub>25–37</sub> and <i>trans</i> -P4. . . . .	71
7.1	Overlay of bovine and human SH3-domains structure. . . . .	74
7.2	Charged sidechains of the SH3-domain at different pH. . . . .	75
7.3	Contact map of charged sidechains of the SH3-domain at pH 7. . . . .	76
7.4	Secondary structure of the SH3 monomer at different pH . . . . .	78
7.5	Contact rates of charged sidechains of the SH3 monomer from conventional MD. . . . .	79
7.6	Map of the native contacts of the SH3 monomer. . . . .	81
7.7	Contact map of the SH3 monomer at different pH. . . . .	82
7.8	Contact maps of the SH3 dimer at pH 2. . . . .	84
7.9	Average state per replica during an REMD simulation of the SH3 monomer. 86	
7.10	Free energy surface of the SH3 domain at different temperatures obtained from REMD simulations at pH 2 and pH 7. . . . .	88
7.11	Free energy surface of the SH3 domain at $T=300$ K and pH 2 with reference structures. . . . .	89
7.12	Contact map for the minima in the FES of the SH3 domain obtained from REMD simulations. . . . .	90
7.13	Fibril structure of the SH3 domain over time. . . . .	93
7.14	Contact maps of SH3 fibril unfolding. . . . .	95
7.15	Initial structural collapse of the disordered SH3 domain. . . . .	96
7.16	Contact map of SH3 early aggregation events. . . . .	98
7.17	Structural motives and sequence of the bovine PI3K-SH3 domain. . . .	101
A1	Free energy surface of the SH3 domain obtained from REMD with projected trajectory from conventional MD at pH 2 and pH 7. . . . .	121

## List of Tables

3.1	Contribution by author to Schäffler et al. 2022 . . . . .	32
4.1	Contribution by author to Schäffler et al. 2022 . . . . .	38
5.1	Contribution by author to Schäffler et al. 2022 . . . . .	48
6.1	Contribution by author to Pascholt et al. 2024 . . . . .	65
6.2	Designed PTH <sub>25–37</sub> peptides containing the AMPB photoswitch . . . . .	67



# Introduction

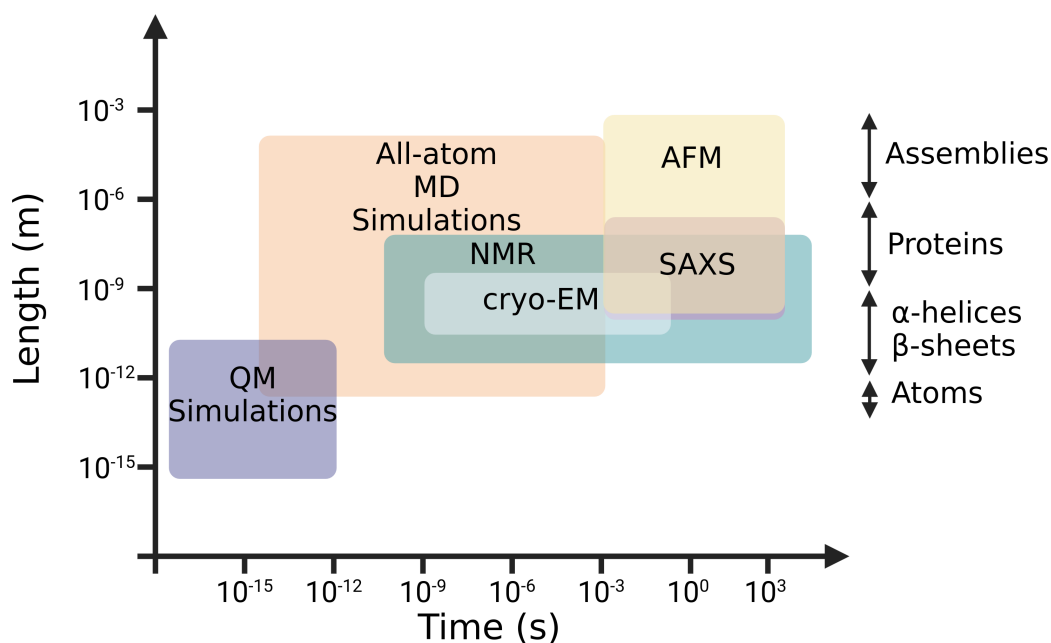
Proteins are fundamental molecules integral to the functioning of living systems, playing a critical role in virtually every biological activity. From catalyzing metabolic reactions to providing structural support, from mediating cell signaling to regulating gene expression, the versatility and indispensability of proteins are unparalleled. Understanding the structure, function, and dynamics of proteins is pivotal in fields ranging from biochemistry and molecular biology to medicine and biotechnology.

Proteins are polymers composed of amino acids linked by peptide bonds, resulting in complex three-dimensional structures that determine their function. Traditionally, it is believed that the amino acid sequence of a protein dictates its unique folding and conformation, which are crucial for its specific function. Specificity is often explained by the lock-and-key mechanism, where the protein's structure allows it to interact precisely with its target molecules [1]. This intricate relationship between sequence, structure, and function is a fundamental principle of molecular biology. Over the past decade, efforts to predict protein structure from amino acid sequences have made significant strides, culminating in the development of AlphaFold by DeepMind [2–4], which received the Nobel price in Chemistry in 2024. By leveraging deep-learning techniques, AlphaFold has achieved unprecedented accuracy in predicting protein structures, revolutionizing the field and rivaling traditional methods such as cryo-electron microscopy (cryo-EM), X-ray crystallography, and nuclear magnetic resonance spectroscopy (NMR) [5].

However, not all proteins conform to this paradigm. Intrinsically disordered proteins (IDPs) lack a unique three-dimensional structure under physiological conditions, instead existing in a dynamic ensemble of conformations. This inherent flexibility allows IDPs to engage in diverse and transient interactions with multiple binding partners, thereby playing crucial roles in various cellular processes such as signaling and regulation [6, 7]. However, their functional advantages sometimes come with the tendency for amyloid aggregation. This phenomenon, where proteins self-assemble into larger, often pathogenic structures, is a significant area of protein research. Amyloid aggregation is implicated to play a pivotal role in numerous neurodegenerative disorders, including Alzheimer's disease (AD) [8, 9], Parkinson's disease (PD) [10], and type II diabetes [11, 12]. These disorders are characterized by the accumulation of amyloid proteins into neurotoxic oligomers and fibrils. The process begins with the formation of soluble intermediate aggregates, such as oligomers and protofibrils, which eventually mature into rigid amyloid fibrils with rich inter-molecular  $\beta$ -sheet structures. It is widely believed that the smaller, heterogeneous, and mobile oligomers are the most toxic species, capable of diffusing across cells and interacting with cellular membranes, thereby disrupting cellular functions [13].

The conformational heterogeneity of IDPs poses major challenges for their structural characterization using experimental techniques. Standard techniques for studying molecular structures include NMR spectroscopy and cryo-EM. Small-angle X-ray scattering (SAXS) is frequently employed to investigate the secondary structure propensity and compactness of molecules. Additionally, atomic force microscopy (AFM) is utilized as an imaging technique to examine larger aggregates and assemblies. However, most experimental techniques already struggle to resolve dynamics that occur on short timescales (see Fig. 1.1) and in the case of IDPs, the diversity of structures further complicates techniques that average over an ensemble of probe molecules, making simulations all the more important [14, 15].

Molecular dynamics (MD) simulations offer a complementary approach for gaining insights into the structural properties of IDPs. MD simulations enable the study of IDPs in a dynamic and spatiotemporal manner by simulating the motion of individual atoms over time. This provides a molecular-level understanding of their conformational changes, allowing researchers to explore the diverse conformational landscapes of IDPs and their aggregation behaviors [5, 16–18]. Recent advancements in technology, software, and simulation methodologies have made large-scale



**Fig. 1.1.: Schematic overview over time- and length scales in biological methods.** The temporal and spatial resolution of the different simulation and experimental methods is indicated by colored boxes. The ranges are only approximate and meant to highlight the broad applicability for each method. Methods include: Nuclear magnetic resonance (NMR), single-particle electron cryo-microscopy (cryo-EM), small-angle X-ray scattering (SAXS), atomic force microscopy (AFM), all-atom molecular dynamics simulations and quantum-mechanical (QM) Simulations. [14, 15]

computer simulations feasible and established them as valuable tools for modeling amyloid protein aggregation at the atomic level [19]. These advancements have significantly enhanced our understanding of their mechanisms and structures on picosecond to millisecond time scales. While simulations at the quantum mechanical (QM) level would offer even higher accuracy, they remain computationally infeasible for simulating entire molecules or oligomers. MD simulations, on the other hand, provide a good balance between accuracy and computational feasibility, capturing the kinetic behavior of molecules, while being executable on contemporary hardware.

## 1.1 Aim of the Thesis

In this thesis, MD simulations are employed to investigate various IDPs and amyloid aggregation prone peptides, notably the amyloid- $\beta$  ( $A\beta$ ) peptide, a sub-sequence of the parathyroid hormone  $PTH_{25-37}$  and the Src-homology 3 (SH3) domain. By utilizing transition networks, the configurational space of the  $A\beta_{42}$  peptide is analyzed to study its disorder-to-order transitions upon binding and self-assembly, leading to qualitative visualizations of the free energy landscapes that provide valuable insights into aggregation mechanisms. Extending this analysis, the free energy surface (FES) is thoroughly examined, alongside timescale analyses that shed light on the dynamical properties of these systems. Additionally, the work explores photoinduced reversible amyloid aggregation in the context of  $PTH_{25-37}$  and investigates pH-dependent amyloid aggregation using the SH3 domain as a model system. Collectively, these approaches aim to deepen our understanding of the conformational dynamics and aggregation kinetics of IDPs and amyloid-forming peptides, contributing to the broader knowledge of amyloid aggregation and structural transitions.

## 1.2 Intrinsically Disordered Proteins

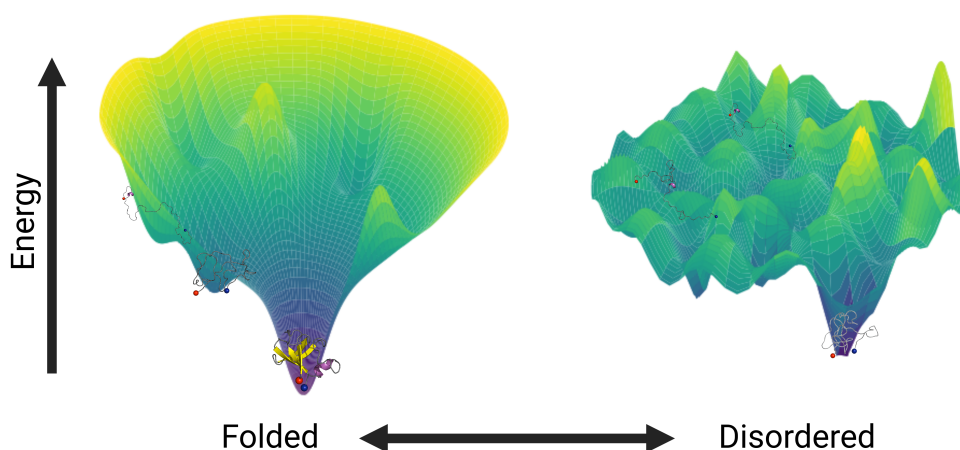
In the past, proteins have been understood as molecules with well-defined three-dimensional structures that are inseparably associated with their biological functions. These "folded" proteins, which turn out to only make up about 50% of all proteins in the human proteome [20], achieve their function through specific structural conformations. Another 20% of proteins exhibit intrinsically disordered regions longer than 30–40 consecutive residues that lack a well-defined structure, while a total of 30% are considered completely IDPs. A protein is classified as an IDP if less than 70% of its sequence can be identified with a unique folded structure.

In terms of their FES, IDPs differ significantly from folded proteins in their topologies. Folded proteins typically exhibit a singular folding funnel which drives the protein towards the global minimum, corresponding to its unique three-dimensional structure (see Figure 1.2). In contrast, IDPs do not have a single equilibrium structure. Instead, they exist as flexible, heterogeneous ensembles of conformations

resulting from relatively flat free energy landscapes with multiple competing funnels. This flat multifunnel landscape consists of multiple local energy minima of similar free energy separated by relatively small energy barriers, allowing for frequent and rapid transitions between these minima [21]. These multifunnel landscapes enable IDPs to interact with multiple partners and play versatile roles in cellular processes such as signal transduction, gene regulation, and protein-protein interactions.

To facilitate these processes, IDPs exhibit a unique property known as "folding upon binding," wherein they transition from a disordered state to a more ordered structure upon interacting with specific binding partners [22, 23]. This process enables IDPs to achieve high specificity in binding interactions without the necessity for strong affinity, facilitating rapid association and dissociation rates crucial for regulatory functions. Studies, such as those involving the measles virus nucleoprotein and its interaction with the phosphoprotein X domain, revealed that folding often occurs through induced-folding pathways [24]. This mechanism involves the formation of inter-molecular contacts prior to or concurrent with the secondary structure development of the IDP, suggesting a kinetic advantage of flexible, disordered states in navigating binding pathways and enhancing interaction efficiency.

A common structural motive across IDPs is the relatively high abundance of glutamic acids and serines in their amino acid sequence, resulting in low overall hydrophobicity and large net charge [25, 26]. This unique composition distinguishes them from folded proteins, which typically have hydrophobic cores that stabilize their three-dimensional structures. The reduced hydrophobicity in IDPs prevents the formation of such stable cores, contributing to their lack of a fixed structure. Despite these properties, which generally reduce the propensity for beta-sheet formation, several IDPs are notably involved in amyloid-related diseases. Proteins such as A $\beta$ , tau, and alpha-synuclein defy their structural predispositions and form amyloid



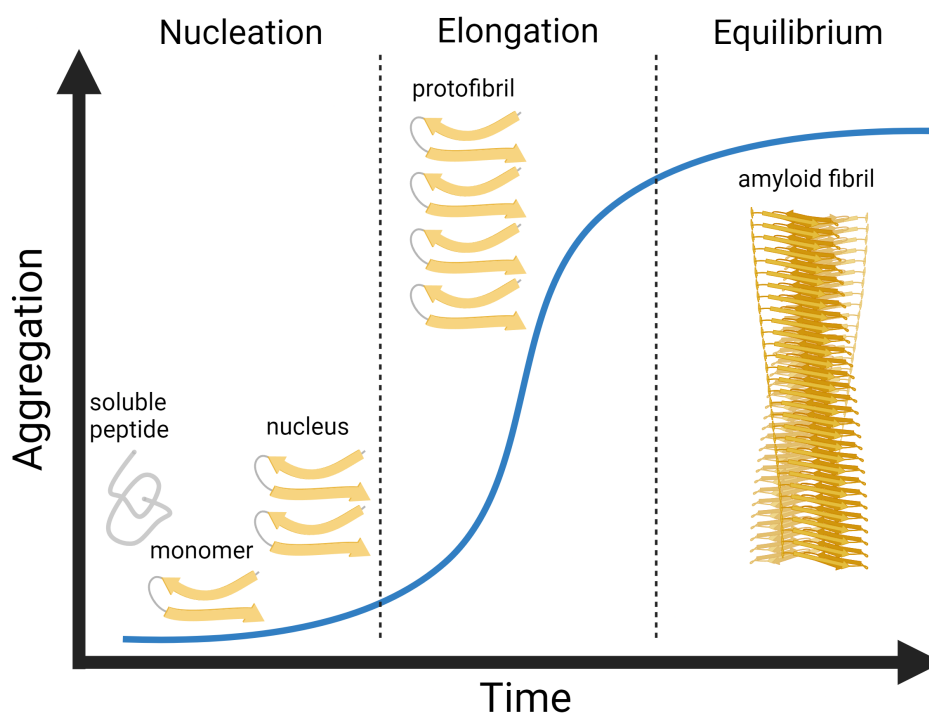
**Fig. 1.2.:** Schematic representation of the free energy landscape of a folded and disordered protein.

fibrils, playing critical roles in conditions like Alzheimer's [27, 28] and Parkinson's diseases [10]. These proteins contain specific segments that are capable of forming amyloid structures, highlighting the complex and dual nature of IDPs.

### 1.3 Amyloid Aggregation

The transition of soluble proteins to ordered, highly stable, and insoluble amyloid aggregates has gained significant research interest in recent decades. While some functional amyloids, such as the parathyroid hormone, can reversibly disintegrate after aggregation [29], numerous common disorders are associated with irreversible amyloid formation. Diseases linked to amyloid aggregation include Parkinson's disease, Type 2 Diabetes Mellitus, Huntington's disease, Amyotrophic Lateral Sclerosis, Prion diseases, and Alzheimer's disease. Alzheimer's disease, in particular, is of increasing importance in our aging society. It is estimated that by 2050, 150 million people worldwide will be affected by Alzheimer's disease [30].

The fibrillization pathway of amyloid fibrils is widely understood as a nucleated growth polymerization process, which can be divided in three phases: nucleation, elongation, and equilibrium (see Figure 1.3) [31–33]. The process begins when individual monomeric polypeptides assemble into oligomeric aggregates, which



**Fig. 1.3.: Schematic representation of amyloid aggregation phases.** In the nucleated-polymerization model, the aggregation pathway is divided into nucleation, elongation, and equilibrium. During nucleation phase a critical nucleus is established, which is a fibrillar/on-pathway oligomer. The nucleus grows to protofibrils during the elongation phase and finally forms mature amyloid fibrils in the equilibrium phase.

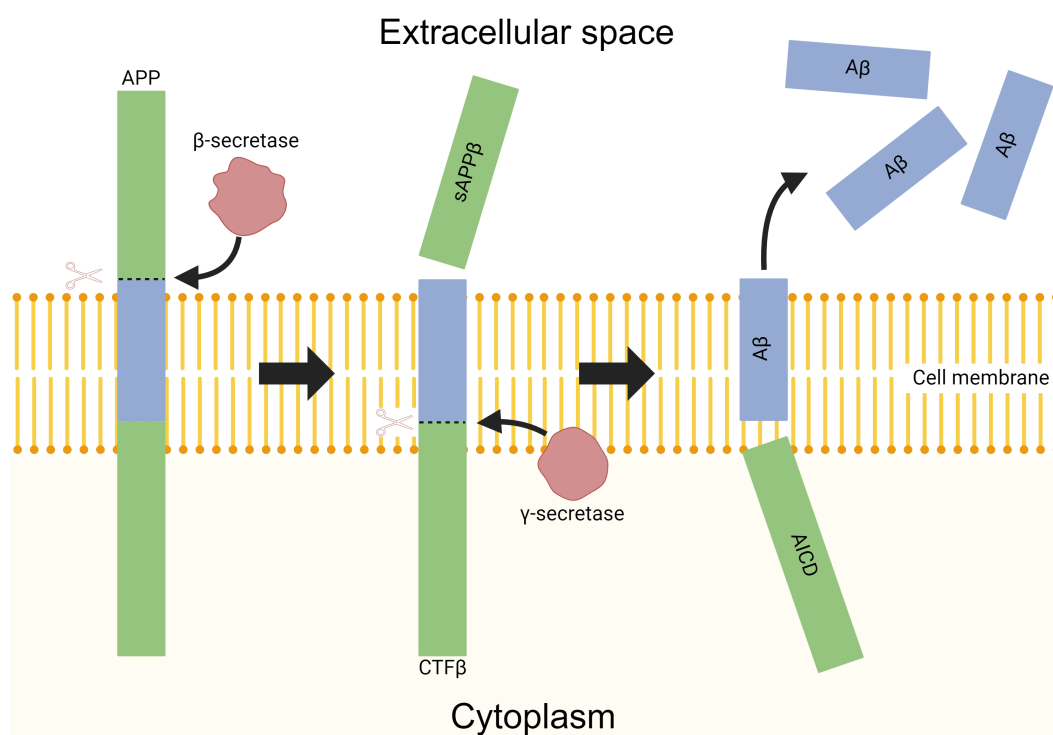
transform into protofibrils and eventually form fibrils. Mature fibrils typically exhibit highly ordered and rigid structures compared to their monomeric and oligomeric predecessors, due to the formation of highly stable inter-protein  $\beta$ -sheet structures. Compared to elongation, primary nucleation is a slower process due to the significant free energy barrier associated with the required structural rearrangements [5]. Once this barrier is surpassed, the critical nucleus forms, representing the smallest and least stable structure capable of promoting further aggregation [34, 35]. This nucleus is characterized by a high free energy state [36]. Following the formation of the critical nucleus, the addition of monomers or oligomers to the protofibril becomes thermodynamically favorable, leading to rapid addition of monomers to the established structure during the elongation phase. However, during elongation, larger assemblies can still dissociate into smaller ones, which can then accelerate the fibrilization by stimulating new fibril growth through the branched-chain mechanism [37]. Studies in the lab of Tuomas Knowles and Christopher Dobson demonstrated that once a small but critical concentration of amyloid fibrils has accumulated, toxic oligomeric species are predominantly formed from monomeric peptide molecules through a fibril-catalyzed secondary nucleation reaction. In secondary nucleation, the surface of existing fibrils serves as a template, catalyzing the formation of new toxic oligomers, accelerating the aggregation process and increasing the overall toxicity [38, 39]. These processes continue until equilibrium is reached and mature fibrils are formed. Environmental factors can influence each phase of the aggregation process, significantly impacting the morphology of the fibrils and the kinetics of their aggregation. Factors known to affect fibril formation include increased polypeptide concentration, post-translational modifications, pH, and interactions with lipid membranes [40–45]. Recently it has been suggested that oligomers are in fact the most toxic species, which exhibit toxicity through receptor-mediated, membrane-damaging, and intracellular mechanisms [46–48]. For example, it has been suggested that oligomeric A $\beta$  can cause cell death by interacting with the cell membrane disrupting Ca<sup>2+</sup> homeostasis [49] or lead to inflammation upon binding to microglia [50].

The primary challenges in studying amyloid aggregation lie in capturing the rapid timescales associated with oligomerization [51–53]. Furthermore, the low relative abundance of oligomers complicates their detection and characterization. For example, soluble A $\beta$  oligomers in buffered aqueous solution typically constitute less than 1.5% of the total peptide concentration [54]. Consequently, the small size (typically fewer than 10 proteins), low abundance and fast timescales present significant difficulties for experimental studies, resulting in most research focusing on the more stable, mature fibrils. Additionally, the polymorphism of amyloid aggregates complicates both experimental and computational studies. Amyloid polymorphism occurs when the same polypeptide can form different structural arrangements, which can manifest at multiple levels, including various packing arrangements of protofilaments, variations in the folds of specific protein regions, and different

interfaces of interaction. Experimental techniques commonly employed to investigate amyloid polymorphisms include cryo-EM [55], AFM, and NMR studies [56]. In this context, MD simulations offer a valuable tool for addressing the presented challenges by enabling the study of oligomerization and early aggregation events over short timescales. Additionally, structural clustering of the configurational space explored during MD simulations allows for the identification of oligomer polymorphisms and their relative formation propensities.

## 1.4 Amyloid- $\beta$ peptide

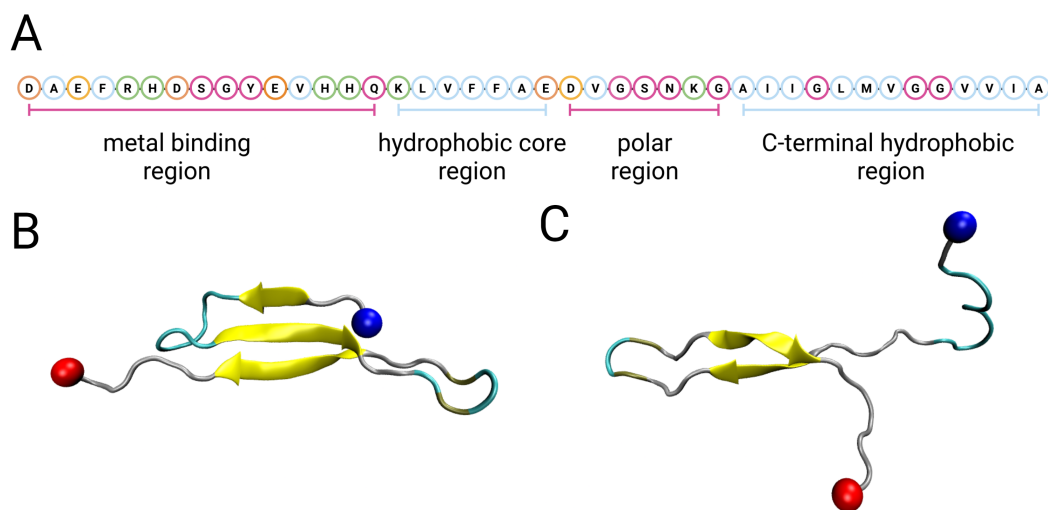
Amyloid- $\beta$  ( $A\beta$ ) is a soluble IDP, which is generated through the proteolytic cleavage of the transmembrane amyloid precursor protein (APP), by  $\beta$ -secretase and  $\gamma$ -secretase enzymes in the human brain. APP can undergo proteolysis via two distinct pathways involving either  $\alpha$ - or  $\beta$ -secretase, resulting in non-amyloidogenic or amyloidogenic pathways, respectively. In the amyloidogenic pathway, as depicted in Figure 1.4,  $\beta$ -secretase cleaves APP into the membrane-bound C-terminal fragment  $CTF\beta$  and N-terminal  $sAPP\beta$ . Subsequent cleavage of  $CTF\beta$  by  $\gamma$ -secretases produces extracellular  $A\beta$  and the intracellular APP domain AICD. The  $\gamma$ -secretases cleavage can yield  $A\beta$  peptides of varying length ranging from 38–51 amino acids, including the two most prevalent products of the amyloidogenic pathway:  $A\beta_{40}$  and  $A\beta_{42}$ , containing 40 and 42 amino acids respectively [57]. Conversely, in the



**Fig. 1.4.: Human APP proteolysis in the amyloidogenic pathway.** In the amyloidogenic pathway  $\beta$ -secretase cleaves APP into the membrane-bound C-terminal fragment  $CTF\beta$  and N-terminal  $sAPP\beta$ . Subsequent cleavage of  $CTF\beta$  by  $\gamma$ -secretases produces extracellular  $A\beta$  and the intracellular APP domain AICD.

non-amyloidogenic pathway, APP is cleaved by  $\alpha$ -secretase, which cuts within the  $A\beta$  sequence, thereby preventing the formation of  $A\beta$  peptides [58, 59]. While the  $A\beta$  fragments produced by the proteolysis of APP have been identified to be involved in the development of AD, its role in the healthy human brain is less well understood [60]. APP knockout studies in mice have shown reduced locomotor activity and eventual reactive gliosis in their brains [61]. Furthermore,  $A\beta_{40}$  has been suggested to act as an cellular antioxidant [62] and both  $A\beta_{40}$  and  $A\beta_{42}$  seem to modulate potassium channels in neurons, counteracting the effects of secretase inhibitors [63].

Both  $A\beta_{40}$  and  $A\beta_{42}$  are capable of forming amyloid plaques; however, while  $A\beta_{40}$  is the most abundant species in the brain,  $A\beta_{42}$  is identified as the more toxic variant and is widely considered the primary agent in AD pathogenesis [64]. The  $A\beta$  sequences can be divided into four distinct regions: (1) a metal-binding region at the N-terminus, primarily composed of hydrophilic amino acids and three histidine residues H6, H13 and H14 (residues 1–15); (2) a hydrophobic core region in the center of the peptide (residues 16–22); (3) a central polar region (residues 23–29); and (4) a hydrophobic C-terminal region (residues 30–40). Notably, the C-terminal region of  $A\beta_{42}$  is extended by two additional hydrophobic residues compared to  $A\beta_{40}$  (see Figure 1.5 A). The longer hydrophobic tail of  $A\beta_{42}$  facilitates the formation of a double hairpin, or S-shaped structure (see Figure 1.5 B), in addition to the prominent single  $\beta$ -hairpin structure (see Figure 1.5 C). A hairpin structure is a protein structural motif, where adjacent regions of the sequence form an anti-parallel  $\beta$ -sheet, through a loop involving 2–5 amino acids between the two strands of the  $\beta$ -sheet. Studies have shown that oligomeric  $\beta$ -hairpin structures may represent the more toxic species,



**Fig. 1.5.: Sequence of the  $A\beta_{42}$  peptide and prominent structures.** (A) The amino acids are colored according to their primary attribute: orange: acidic, green: basic, pink: polar, blue: non-polar/hydrophobic. Furthermore an example of (B) a S-shape structure and (C) a  $\beta$ -hairpin structure are shown. Here, the N-/C-termini are shown as blue and red spheres, respectively.

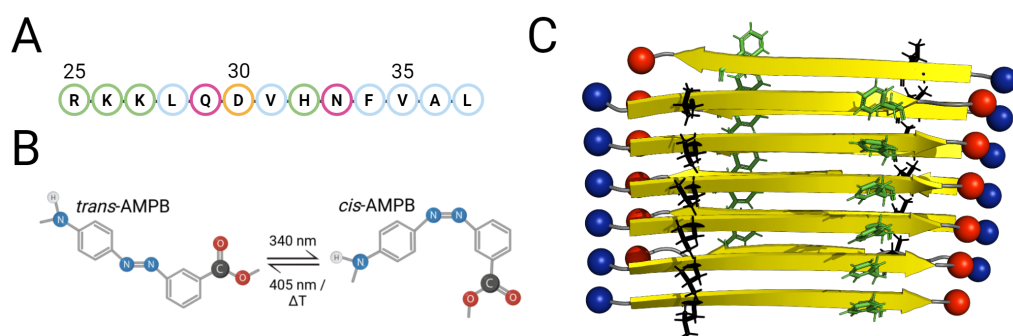


highlighting their potential role in the pathogenesis of Alzheimer's disease [36, 47, 65–67].

In chapter 3 and 4 of this thesis we will discuss the conformational space of  $A\beta_{42}$  and the change of structural propensity, specifically the  $\beta$ -sheet and  $\beta$ -hairpin propensity, in different environments. Furthermore, in chapter 5 we analyse the free energy surface of the  $A\beta_{42}$  monomer/dimer and identify key features of the pathway from a disordered to a  $\beta$ -hairpin state, using first passage time analysis.

## 1.5 Parathyroid Hormone

The parathyroid hormone (PTH) is a 84-amino acid hormone, which is capable of reversible amyloid fibril formation. It plays a crucial role in regulating calcium and phosphate balance in the human body [68]. The mature form,  $PTH_{1-84}$ , is stored as functional amyloids prior to secretion, with its N-terminal pro-sequence possibly preventing premature aggregation [69]. The N-terminal fragment with 34 amino acids of  $PTH_{1-84}$  is well-studied, particularly its importance in activating G-protein coupled receptors in bone and kidney cells and in maintaining calcium and phosphate homeostasis in the nervous system [70]. The  $PTH_{1-34}$  sequence became the first food and drug administration (FDA) approved drug for the treatment of severe osteoporosis in 2002 [71]. However, there is still a gap in understanding the fibrillization process and the characteristics of the resulting fibrils. Current research suggests that under physiological conditions, the thermodynamic stability of  $PTH_{1-84}$  fibrils is low enough to permit dissociation when diluted, with the fibril-forming sequence spanning amino acids R25 to L37 (see Figure 1.6 A). Previous studies have examined the influence of the pro-sequence and environmental factors on the fibrillization of PTH, yet detailed structural parameters governing fibril assembly and disassembly remain largely unexplored [69].



**Fig. 1.6.: Structural overview of the  $PTH_{25-37}$  peptide and AMPB photo switch.** (A) Sequence of the  $PTH_{25-37}$  peptide. The amino acids are colored according to their primary attribute: orange: acidic, green: basic, pink: polar, blue: non-polar/hydrophobic. (B) Chemical structure of AMPB photo switch in *trans* and *cis* conformation (C) fibril model of  $PTH_{25-37}$ . The N- and C-terminus are depicted as blue and red spheres, respectively. The  $\beta$ -sheets are colored yellow and  $\alpha$ -helices are colored purple. In (C) the side chains of L28 and F34 are highlighted as black and green sticks, respectively.

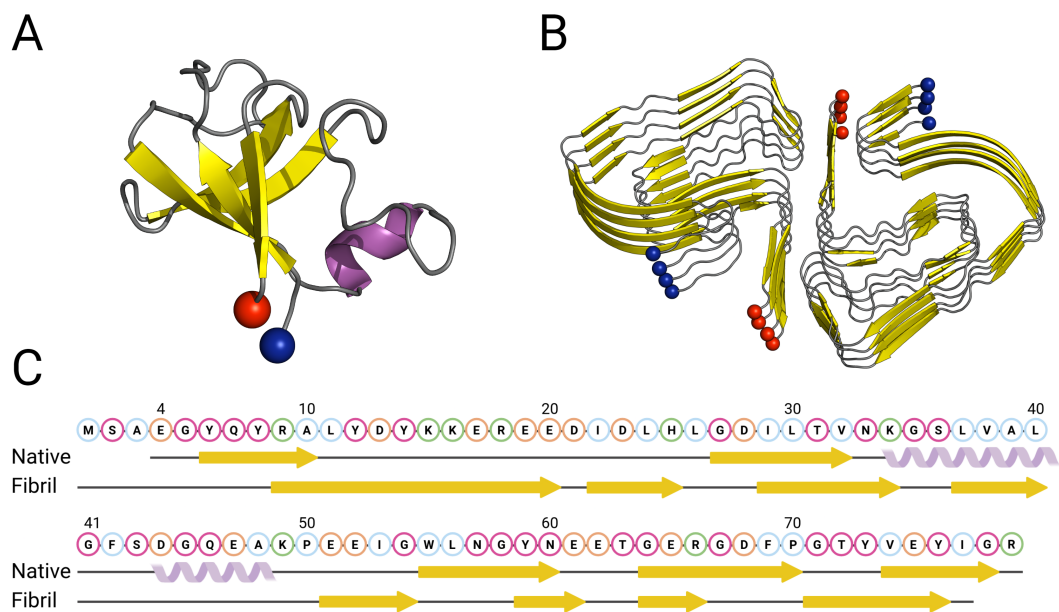
In chapter 6 we study the aggregation behaviour of the PTH<sub>25–37</sub> sequence using MD simulations, coupled with the findings by our experimental collaboration partners. In that study we investigate the light-induced transitions between aggregated and non-aggregated states of the PTH<sub>25–37</sub> sequence mutated with an artificial 3-[[4-aminomethyl]phenyl]diazenyl}benzoic acid (AMPB) photo switch, which shows to be a reversible process. The AMPB photoswitch can transition between two conformations: the open *trans* and closed *cis* states (see Figure 1.6 B), triggered by light activation at specific wavelengths. This light-activated mechanism presents a promising method for targeted drug delivery by enabling controlled release of reversibly fibrillating peptides. Finally, we develop a fibril model for the PTH<sub>25–37</sub> sequence and one mutated sequence, by combining both computational and experimental efforts (see Figure 1.6 C).

## 1.6 SH3-domain

The bovine SH3 domain of the p85 $\alpha$  subunit of PI3K (PI3K-SH3) is a small, well-characterized protein module composed of 84 amino acid residues. SH3 domains are known to play a key role in several cellular signaling-pathways due to their ability to mediate protein-protein interactions and recognizing proline-rich motifs [72, 73]. In the family of protein tyrosine kinases, for instance, SH3 domains play a crucial role in catalyzing the phosphorylation of tyrosine residues on protein substrates, a vital step in cellular signaling [74].

The structure of PI3K-SH3, shown in Figure 1.7 A, is a  $\beta$ -sheet core consisting of two perpendicular, antiparallel  $\beta$ -sheets, with three and two strands, respectively. This compact  $\beta$ -sheet structure lacks disulfide bridges, contributing to its flexibility and functionality. Notably, the folding and unfolding transitions of PI3K-SH3 at neutral pH are reversible and cooperative, with no significant accumulation of intermediates, ensuring a smooth conformational transition [75]. At acidic pH values, specifically below pH 4, the PI3K-SH3 domain undergoes significant conformational changes leading to denaturation and subsequent fibrillization. In this denatured state, the protein gradually aggregates into long fibrils that share many characteristics with amyloid fibrils (see Figure 1.7 B), despite the SH3 domain not being linked to any known disease [76]. Figure 1.7 C shows the sequence of the PI3K-SH3 domain along with its secondary structure in the native fold and the amyloid fibril. To transition from the folded structure to the fibril, the protein must undergo significant structural changes, necessitating the complete unfolding of the native fold.

The ability of the PI3K-SH3 domain to switch between a folded state at pH $\sim$ 7 and an amyloid forming disordered state at low pH, as shown in Figure 1.2, makes it a very well suited system to study amyloid aggregation. In chapter 7, we employ several simulation techniques to study the structural and kinetic differences at pH=2 and pH=7 and investigate early aggregation events.



**Fig. 1.7.: Structural overview of the bovine PI3K-SH3 domain.** (A) Folded structure of the bovine PI3K-SH3 domain (PDB:1PKS) and (B) fibril structure of the PI3K-SH3 domain (PDB:6R4R). The N- and C-terminus are depicted as blue and red spheres, respectively. The  $\beta$ -sheets are colored yellow and  $\alpha$ -helices are colored purple. (C) Sequence of the SH3-domain. The amino acids are colored according to their primary attribute: orange: acidic, green: basic, pink: polar, blue: non-polar/hydrophobic. Below the sequence the secondary structure in the native fold and amyloid fibril is shown.



The Methods chapter outlines the models and algorithms utilized in the presented work. First, the primary sampling tool, molecular dynamics (MD) simulations, is introduced, followed by a discussion of enhanced sampling methods. Additionally, various analysis methods to investigate the conformational space of molecules are presented, including principle component analysis (PCA) and the transition network (TN) methodology. The latter is further developed to produce a free energy surface from an MD simulation. Finally, the first passage time (FPT) analysis method is introduced, providing kinetic insights derived from the free energy surface.

## 2.1 Molecular Dynamics Simulation

MD simulations provide a powerful tool to study time and length scales of molecular systems, that are challenging to access experimentally (see Figure 1.1). Starting from an initial structure, the forces acting on all atoms are calculated, allowing the system to be integrated over small discrete time steps in a Newtonian fashion. Repeating the force calculation and integration ultimately yields a trajectory of the molecule's structural evolution over time. The accuracy, with which the simulation resembles a molecule's behavior *in vivo*, depends highly on the quality of the underlying potential energy function, also called force field, which is used for the force calculations.

### 2.1.1 Force Field

The force field (FF) is the fundamental potential energy function employed to describe all interaction energies and consequently forces in MD simulations. Force fields are parameterized to approximate the true potential energy function, which might vary on the application, resulting in multiple FFs optimized for different scenarios. In MD simulations, a primary distinction is made between all-atom (AA) and coarse-grained (CG) force fields, excluding QM level approaches. In coarse-grained MD (CG MD), the number of the system's degrees of freedom is significantly reduced compared to all-atom MD (AA MD) simulations by mapping clusters of atoms onto coarse-grained beads [77]. This coarse-graining approach extends the accessible temporal and spatial scales of simulations by orders of magnitude compared to AA MD, potentially covering biological scales, but at the cost of losing detailed structural information. The degrees of freedom averaged out in the CG FF may include critical structural information, as well as making fast transitions on the picosecond timescale inaccessible altogether [78].

In this work, we employed AA MD simulations (henceforth referred to as MD simulations) to study early aggregation events and protein interactions at an atomistic level. Although simulations at a quantum mechanical level would offer even higher accuracy, the computational cost of simulating an entire protein or even oligomers is currently not feasible. Future advancements in machine learning-based QM force

fields may enable such simulations in upcoming studies [79–81]. As mentioned, the choice of FF for an MD simulation depends on the application. The most popular FFs to date include those from the assisted model building and energy refinement (AMBER) and chemistry at Harvard macromolecular mechanics (CHARMM) families. While AMBER force fields, specifically AMBER99SB-ILDN and AMBER14SB, are well-suited for globular folded proteins and provide reasonable results for amyloid aggregation, they favor secondary-structure formation. Instead, we used the CHARMM36m force field, which was demonstrated to provide good models for IDPs and early peptide aggregation events [82, 83]. This FF includes a refined backbone correction map (CMAP) potential and improved specific salt bridge interactions, optimized for modeling IDPs.

The potential energy function  $U_{ff}(\mathbf{R})$  is mostly the same for all FFs, only differing in the parameters of the interaction terms. The types of interaction can be divided into bonded and non-bonded interactions

$$U_{ff}(\mathbf{R}) = U_{bonded}(\mathbf{R}) + U_{non-bonded}(\mathbf{R}), \quad (2.1)$$

which only depend on the position of all  $N$  atoms in the system with coordinates  $\mathbf{R} = (r_x^{(1)}, r_y^{(1)}, r_z^{(1)}, \dots, r_x^{(N)}, r_y^{(N)}, r_z^{(N)})$ .

The bonded interactions describe intra-molecular two-, three- and four-body terms, which describe bonds, angles and diheadral angles respectively:

$$U_{bonded}(\mathbf{R}) = \sum_{bonds} U_{ij}^{Bonds} + \sum_{angles} U_{ijk}^{Angles} + \sum_{dihedrals} U_{ijkl}^{Dihedrals} + \sum_{improper} U_{ijkl}^{Improper}. \quad (2.2)$$

The bond vibration of two bonded atoms  $i$  and  $j$  is described by a harmonic potential

$$U_{bond}(r_{ij}) = \frac{k_{ij}^b (r_{ij} - b_{ij}^0)^2}{2}, \quad (2.3)$$

where  $r_{ij}$  is their inter-atomic distance. The stretching force constant  $k_{ij}^b$  and the equilibrium bond length  $b_{ij}^0$  are specific to the atom type of  $i$  and  $j$ . It should be noted, that the parametrization of the bond term by a harmonic potential precludes the bond breaking in MD simulations.

The angle bending of three consecutively bound atoms  $i$ ,  $j$ , and  $k$ , is characterized by the harmonic potential

$$U_{angle}(\theta_{ijk}) = \frac{k_{ijk}^\theta (\theta_{ijk} - \theta_{ijk}^0)^2}{2}, \quad (2.4)$$

where  $k_{ijk}^\theta$  is the bending force constant and  $\theta_{ijk}^0$  the equilibrium angle.

The dihedral angle  $\Phi_{ijkl}$  is the angle between two planes, each defined by three of the four atoms  $i$ ,  $j$ ,  $k$  and  $l$ . To describe the rotation around the bond-axis between

atoms  $j$  and  $k$ . Most commonly the dihedral angle potential is modeled using periodic cosine functions

$$U_{dihedral}(\Phi_{ijkl}) = k_{ijkl}^{\Phi} [1 + \cos(n\Phi_{ijkl} - \Phi_{ijkl}^0)], \quad (2.5)$$

where  $k_{ijkl}^{\Phi}$  is the torsion force constant,  $\Phi_{ijkl}^0$  is the equilibrium angle, and  $n$  denotes the periodicity factor (sometimes called multiplicity). Multiple cosine functions may be combined to accurately describe cis/trans or gauche conformations, yielding a sum of cosine functions for a single dihedral angle. Furthermore, a distinction is made between proper and improper dihedral angles. Proper dihedral angles refer to torsion angles between consecutive bonds in a molecule, while improper dihedral angles involve three atoms bonded to a central atom, which are needed to model molecular planarity or chirality.

Besides the intra-molecular bonded interactions, non-bonded interactions govern the forces between all atoms whose interactions are not accounted for by the bonded terms. The non bonded interactions are divided into van-der-Waals forces and electrostatic forces

$$U_{non-bonded}(\mathbf{R}) = + \sum_{atom\ pairs} U_{ij}^{LJ} + \sum_{atom\ pairs} U_{ij}^{Coul}. \quad (2.6)$$

The van-der-Waals interaction is modeled by the Lennard-Jones (LJ) potential

$$U_{ij}^{LJ}(r_{ij}) = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right]. \quad (2.7)$$

The potential includes two components: a repulsive  $r^{-12}$  term that reflects the Pauli exclusion principle and nuclear repulsion in close proximity, and an attractive  $r^{-6}$  term accounting for dispersion forces or London interactions over longer distances. The parameters in the Lennard-Jones 12-6 potential,  $\sigma_{ij}$  and  $\epsilon_{ij}$ , define the characteristic length scale and depth of the potential well, respectively. At the minimum distance  $r_{min} = 2^{1/6}\sigma_{ij}$ , the potential energy  $U^{LJ}(r_{ij})$  reaches its lowest value  $-\epsilon_{ij}$ .

The electrostatic interactions between atoms  $i$  and  $j$  is modeled by Coulomb's (Coul) law

$$U_{ij}^{Coul}(r_{ij}) = \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}, \quad (2.8)$$

based on their partial charges  $q_i$  and  $q_j$  and the distance  $r_{ij}$  between them. Here,  $\epsilon_0$  denotes the vacuum permittivity, and  $\epsilon_r$  is the relative permittivity (usually set to 1 in MD simulations). The Coulomb interactions decay slowly with  $1/r_{ij}$ , classifying them as long-range interactions within the length scales of MD simulations.

As non-bonded interactions need to be computed between each pair of atoms, the computational cost scales as  $\sim N^2$ . Compared to bonded interactions, which only scale with  $\sim N$ , non-bonded interactions dominate the computational expense of

MD simulations. Furthermore, MD simulations typically employ periodic boundary conditions (PBC) to simulate an infinite environment and mitigate finite-size effects, necessitating careful consideration of long-range electrostatic interactions.

The full potential energy function employed in MD simulations including bonded and non-bonded interactions reads as:

$$U_{ff}(\mathbf{R}) = \sum_{bonds} U_{ij}^{Bonds} + \sum_{angles} U_{ijk}^{Angles} + \sum_{dihedrals} U_{ijkl}^{Dihedrals} + \sum_{improper} U_{ijkl}^{Improper} + \sum_{atom\ pairs} U_{ij}^{LJ} + \sum_{atom\ pairs} U_{ij}^{Coul}. \quad (2.9)$$

### 2.1.2 The CHARMM36m Force Field

The CHARMM all-atom force field, originally developed by Martin Karplus' lab at Harvard University [84, 85], is extensively used for simulating biomolecules, including lipids [86, 87], carbohydrates [88, 89], nucleic acids [90], IDPs [82], in addition to folded proteins as the most common application. CHARMM's parameterization strategy involves fitting experimental and *ab initio* QM data at the HF/6-31(g) level for small model compounds, which are transferable to larger molecules with further optimization. Furthermore, the CHARMM General Force Field (CGenFF) [91] facilitates the parameterization of compounds not originally included in the force field, such as drugs or other ligands [92].

Two correction terms are included in the CHARMM force field to enhance accuracy: the Urey-Bradley (UB) term and the CMAP potential [93, 94]. The UB term, estimated by a harmonic function, improves the description of angle bending by accounting for the coupling between bond length and bond angle:

$$U_{UB}(r) = k_{ub,i(i+2)} \left( r_{i(i+2)} - r_{i(i+2)}^0 \right)^2 \quad (2.10)$$

Here, the UB force constant  $k_{ub,i(i+2)}$  applies to atoms  $i$  and  $i + 2$  connected by two bonds. The CMAP potential, a grid-based energy correction factor, refines protein backbone conformations by considering the correlation between backbone dihedral angles  $\phi$  and  $\psi$ . CHARMM also recommends using the modified TIP3P water model for simulations, as its parameters for biomolecule-water interactions are optimized for this model. Additionally, CHARMM employs force-based cutoffs to smooth the force to zero at a distance of 1–1.2 nm, thus this cutoff has to be used in simulations involving a CHARMM force field. Non-bonded interactions between atoms separated by four bonds (1–4 interactions) are not scaled, necessitating an iterative optimization process for dihedral potential parameters based on electrostatic and Lennard-Jones parameters.

Especially the CHARMM36m force field represents a significant improvement for simulating IDPs [95, 96]. It includes refinements such as a modified backbone CMAP potential to better capture the conformational ensembles of IDPs, and an enhanced description of specific salt bridge interactions. Validation using a comprehensive



set of peptides and proteins demonstrated that CHARMM36m significantly reduces the overrepresentation of left-handed  $\alpha$ -helices seen in previous versions, aligning more closely with experimental data. The force field also improves the accuracy of predicted experimental observables, such as NMR scalar couplings and SAXS profiles. Overall, CHARMM36m provides more reliable modeling of protein dynamics and conformations, especially for systems with significant disorder [82].

### 2.1.3 TIP3P Water Model

Water is essential to all biological systems, acting as a solvent and reactant. Its unique properties facilitate numerous biochemical processes. In MD simulations, accurately modeling water is crucial for replicating these biological phenomena and understanding biomolecular interactions. However, in explicit water MD simulations, the majority of atoms are water molecules, driving up computational costs. Thus, a balance between accuracy and efficiency is often necessary.

The development of accurate interaction potentials for water has been a focus for decades. Various models, including TIP3P, TIP4P [97], TIP5P [98], and Berendsen's SPC/E [99], have been created. Although TIP4P offers a more accurate representation by including an additional pseudo-atom for the two lone electron pairs of the water oxygen atom, TIP3P remains widely used due to its lower computational cost and extensive optimization of parameters of various force fields, such as CHARMM. The TIP3P model, reparameterized by Jorgensen et al., is a rigid, non-polarizable three-site water model based on Transferable Inter-molecular Potential functions [100, 101]. It uses 12-6-1 functions, combining Lennard-Jones potentials for short-range interactions of oxygen atoms and Coulomb potentials for long-range interactions. The dimerization energy of a pair of TIP3P water molecules  $m$  and  $n$  is computed as:

$$E_{m,n}^{\text{TIP3P}} = \frac{A}{r_{OO}^{12}} - \frac{C}{r_{OO}^6} + \frac{1}{4\pi\epsilon_0} \sum_i^{on\ m} \sum_j^{on\ n} \frac{q_i q_j}{r_{ij}} \quad (2.11)$$

where  $i$  and  $j$  represent the atoms of the water molecule with  $q_i$  and  $q_j$  being their respective charges.  $A$  and  $C$  are Lennard-Jones interaction parameters, and  $r_{OO}$  is the distance between two oxygen atoms. In GROMACS the TIP3P model employs holonomic constraints, restraining the bond lengths within water molecules, and the SETTLE algorithm, adjusting the molecules velocities accordingly, to efficiently manage solvent simulations.

### 2.1.4 Periodic Boundary Conditions and Cutoff Methods

Periodic boundary conditions (PBC) are utilized in MD simulations to model large-volume systems and avoid artifacts at the simulation boundaries. PBCs work by periodically replicating the system in all directions, forming an infinite lattice. When a particle crosses the simulation box boundary, its position is updated according to the minimum image convention: During the simulation, particles in periodic images move identically to those in the original box; as a particle exits, its image enters

from the opposite side. The minimum image convention ensures that each particle interacts with its nearest neighbors, whether within the same box or in periodic images. These image boxes are identical in shape and particle configuration to the original box. While PBCs effectively handle short-range non-bonded interactions like Lennard-Jones forces, which are truncated after a hard cutoff of 8–12 Å, they are inadequate for long-range electrostatic interactions. Truncating these interactions results in nonphysical distributions and discontinuous forces/energies at the cutoff boundary. To address this, the particle mesh Ewald (PME) method is employed in MD simulations. PME improves the contribution of Coulomb interactions to the potential energy function by splitting the interaction into real space and reciprocal space terms, both described by fast converging series.

The Coulomb interactions in the PME method can be expressed by three terms:

$$U_{Coul} = U_{real} + U_{reciprocal} - U_{self}. \quad (2.12)$$

The potential  $U_{real}$  calculates the short-range interaction in real space by treating the charges  $q_i$  as a sum of point charges and Gaussian charge densities of opposing sign. The potential  $U_{reciprocal}$  computes the charge interaction as Gaussian densities in reciprocal space using a Fourier transformation. The term  $U_{self}$  cancels spurious self-interactions due to the summation of  $U_{real}$  and  $U_{reciprocal}$ . The Ewald sum (Eq. 2.12) is computed by mapping all charges onto a lattice, which reduces computational cost and allows for solving the Poisson equation using fast Fourier transforms. This approach results in a lattice sum of discrete charges, which, by construction, converges relatively quickly.

## 2.1.5 Numerical Integration

In MD simulations the evolution of a system with time is described by solving the equation of motion given by Newton's second law of motion. This fundamental equation of classical mechanics relates the net forces  $F_i$  acting on particle  $i$  with its mass  $m_i$  and acceleration  $a_i$ :

$$\mathbf{F}_i(\mathbf{R}(t)) = m_i \cdot \mathbf{a}_i(t) = m_i \frac{\partial^2 \mathbf{r}_i(t)}{\partial t^2}. \quad (2.13)$$

Here,  $\mathbf{R}(t) = (r_x^{(1)}, r_y^{(1)}, r_z^{(1)}, \dots, r_x^{(N)}, r_y^{(N)}, r_z^{(N)})$  is the position of all particles at time  $t$ . Furthermore, the force  $F_i$  can be calculated as the gradient of the potential energy function  $U_{ff}(\mathbf{R})$ :

$$\mathbf{F}_i(\mathbf{R}(t)) = -\frac{\partial U_{ff}(\mathbf{R}(t))}{\partial \mathbf{r}_i(t)}. \quad (2.14)$$

Using equation 2.13 and 2.14, the acceleration, which drives the evolution of positions and velocities over time, can be linked to the gradient of the potential energy function:

$$m_i \mathbf{a}_i(t) = -\frac{\partial U_{ff}(\mathbf{R}(t))}{\partial \mathbf{r}_i(t)}. \quad (2.15)$$

To solve the equation of motion (Eq. 2.15) and generate a trajectory of frames illustrating the system's movement over time in MD simulations, a numerical integration approach is employed. This method computes new positions  $\mathbf{r}_i$  and velocities  $\mathbf{v}_i$  for each atom  $i$  after a specific time step  $\Delta t$ .

**The Verlet Integrator** One of the most common integration schemes in MD simulations is the Verlet algorithm, named after the French physicist Loup Verlet [102]. The derivation of the new positions  $\mathbf{r}(t + \Delta t)$  relies on calculating the third-order Taylor expansions for the positions, forward as well as backward in time:

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 + \frac{1}{6}\frac{\partial^3 \mathbf{r}(t)}{\partial t^3}\Delta t^3 \quad (2.16)$$

$$\mathbf{r}(t - \Delta t) = \mathbf{r}(t) - \mathbf{v}(t)\Delta t + \frac{1}{2}\mathbf{a}(t)\Delta t^2 - \frac{1}{6}\frac{\partial^3 \mathbf{r}(t)}{\partial t^3}\Delta t^3 \quad (2.17)$$

Summation of both expansions yields an expression for the new positions:

$$\mathbf{r}(t + \Delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \Delta t) + \mathbf{a}(t)\Delta t^2 + \mathcal{O}(\Delta t^4), \quad (2.18)$$

which depends solely on the acceleration  $\mathbf{a}(t)$  calculated from Eq. 2.15 and the previous positions. Due to cancelling of the third term of the Taylor expansion, the new coordinates are computed with an error of order  $\mathcal{O}(\Delta t^4)$ . However, a limitation of the Verlet algorithm is the indirect calculation of velocities, which must be derived from the new positions  $\mathbf{r}(t + \Delta t)$ :

$$\mathbf{v}(t) = \frac{\mathbf{r}(t + \Delta t) - \mathbf{r}(t - \Delta t)}{2\Delta t} + \mathcal{O}(\Delta t^2), \quad (2.19)$$

resulting in an error of order  $\mathcal{O}(\Delta t^2)$ . This necessitates very short time steps (e.g.,  $\sim 1$  fs), and the delayed computation of velocities complicates the implementation of thermostats reliant on kinetic energy calculations at time  $t$ .

**The Leapfrog Algorithm** An algorithm which addresses the main shortcomings of the Verlet algorithm is the Leap-Frog algorithm [103], which was used for simulations in the frame of this work. In this method, velocities are first computed at half-integer time steps  $t + \frac{\Delta t}{2}$ , which are subsequently used to calculate positions at  $t + \Delta t$ .

Following a similar derivation as before employing a Taylor expansion, the algorithm yields:

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \mathbf{a}(t)\Delta t + \mathcal{O}(\Delta t^3) \quad (2.20)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right)\Delta t + \mathcal{O}(\Delta t^4). \quad (2.21)$$

This approach explicitly computes velocities, but they are not synchronized with positions at the same time step  $t$ . Velocities at time  $t$  can be approximated by:

$$\mathbf{v}(t) = \frac{1}{2} \left[ \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right) \right]. \quad (2.22)$$

Although the velocities at time  $t + \frac{\Delta t}{2}$  must first be computed to obtain  $\mathbf{v}(t)$ , which prevents the simultaneous calculation of kinetic and potential energy, the leapfrog algorithm allows for the integration of temperature and pressure coupling schemes in the MD code.

### 2.1.6 Thermodynamical Ensembles in MD Simulations

In MD simulations, the choice of thermodynamical ensemble is critical for accurately modeling the physical properties and behaviors of systems. The ensemble dictates the set of conditions under which the simulation is conducted, thereby influencing the results' comparability to processes in vivo and in vitro. The primary ensembles employed in MD simulations include the microcanonical (NVE), canonical (NVT), and isothermal-isobaric (NPT) ensembles.

#### Thermodynamical Ensembles

The microcanonical ensemble (NVE) maintains a constant number of particles ( $N$ ), volume ( $V$ ), and energy ( $E$ ). It is used to simulate isolated systems with no energy exchange with the surroundings. The system's total energy is conserved, and time averages of properties are equivalent to ensemble averages if the system explores all accessible phase space regions with the given energy. In these simulations, temperature fluctuations arise exclusively from exchanges between kinetic and potential energy.

In the canonical ensemble (NVT), the number of particles ( $N$ ), volume ( $V$ ), and temperature ( $T$ ) are held constant, making it suitable for studying systems in thermal equilibrium with a heat bath. The instantaneous temperature of a system can be derived from the atomic velocities  $v_i$  according to the equipartition theorem:

$$\frac{1}{2}Nk_{\text{B}}T = \frac{1}{2} \sum_i m_i v_i^2, \quad (2.23)$$

where  $k_{\text{B}}$  is the Boltzmann constant. Maintaining the system's temperature at a desired target is achieved by rescaling particle velocities. Common thermostats include the velocity rescaling thermostat [104], the Nosé-Hoover thermostat [105, 106]

and the Berendsen thermostat [107]. While the Berendsen and velocity rescaling thermostats efficiently reach the target temperature, they do not produce a true canonical ensemble. Therefore, the Nosé-Hoover thermostat is commonly used to generate the canonical ensemble.

The isothermal-isobaric ensemble (NPT) keeps the number of particles ( $N$ ), pressure ( $P$ ), and temperature ( $T$ ) constant, making it ideal for simulating systems under realistic conditions in experiments and in vivo where both temperature and pressure are regulated. Barostats adjust the system's volume to maintain the desired pressure. Common barostats include the Berendsen barostat [107] and the Parrinello-Rahman barostat [108]. In the MD simulations performed for this work, the Parrinello-Rahman barostat is usually employed with the Nosé-Hoover thermostat to accurately generate the NPT ensemble.

### Nosé-Hoover Thermostat

The Nosé-Hoover thermostat was first introduced by Nosé [109] in 1984 and later improved by Hoover [106] in 1985. This thermostat adds an additional degree of freedom  $\zeta$ , which couples the system to an external heat bath and effectively acts as a friction parameter either accelerating or decelerating particles to meet the target temperature  $\bar{T}$ . The Nosé-Hoover equations of motions read as:

$$m_i \frac{d^2 \mathbf{r}_i}{dt^2} = \mathbf{f}_i - \zeta m_i \mathbf{v}_i \quad (2.24)$$

$$\frac{d\zeta(t)}{dt} = \frac{1}{Q} \left[ \frac{1}{2} \sum_{i=1}^N m_i \mathbf{v}_i^2 - \frac{3N+1}{2} k_B \bar{T} \right], \quad (2.25)$$

where  $Q$  regulates the strength of the interaction with the external heat bath. Furthermore, if  $\frac{d\zeta(t)}{dt} = 0$  the equipartition theorem is recovered, only now with an additional degree of freedom leading to  $\nu = 3N + 1$  degrees of freedom in total.

### Parrinello-Rahman Barostat

Simulating a system at constant pressure is essential for characterizing the conditions of the isothermal-isobaric ensemble in MD simulations. The pressure in a simulation box can be related to the kinetic energy of all particles  $E_{kin}$  and the dispersion forces  $\mathbf{f}_{ij}$  the particles exert onto each other. Using Clausius theorem [110, 111] the pressure  $p$  can be expressed as:

$$p = \frac{2}{3V} (E_{kin} - \Xi), \quad (2.26)$$

where  $V$  is the volume of the system and  $\Xi$  is the inner virial tensor defined as:

$$\Xi = -\frac{1}{2} \sum_{i < j}^N \mathbf{r}_{ij} \cdot \mathbf{f}_{ij} \quad (2.27)$$

Keeping the pressure constant during an MD simulation can be achieved using a barostat. The Parrinello-Rahman barostat [112, 113], in particular, theoretically

reproduces the true NPT ensemble. Similar to the Nosé-Hoover thermostat, the Parrinello-Rahman barostat modifies the system's Hamiltonian by introducing an additional degree of freedom, the volume  $V$ , enabling the simulation volume to fluctuate during the simulation. The modified Hamiltonian is:

$$H = E_{kin} + E_{pot} - p_{ext}V + \frac{1}{2}W \sum_{i,j} \left( \frac{dh_{i,j}}{dt} \right)^2 \quad (2.28)$$

where  $W$  is a parameter determining the mass of a virtual piston (de-)compressing the system,  $p_{ext}$  the external pressure and  $\mathbf{h} = (b_1, b_2, b_3)$  is the matrix of box vectors. From this, the equation of motion for the particles in the box becomes:

$$\frac{d^2 \mathbf{r}_i}{dt^2} = \frac{\mathbf{f}_i}{m_i} \mathbf{h}^{-1} - \mathbf{G}^{-1} \dot{\mathbf{G}} \frac{d\mathbf{r}_i}{dt}, \quad (2.29)$$

where  $\mathbf{h}^{-1}$  its reciprocal inverse of  $\mathbf{h}$ , which is related to the volume as  $V = \det(\mathbf{h})$ .  $\mathbf{G} = \mathbf{h}^t \mathbf{h}$  is a metric tensor. The equation of motion of for the box vectors follows as:

$$\frac{d^2 \mathbf{h}_{i,j}}{dt^2} = \frac{1}{W} \sum_k (\Pi_{i,k} - p \delta_{i,k}) \sigma_{k,j}, \quad (2.30)$$

where  $\sigma_{i,j} = V(\mathbf{h}^{-1})_{j,i}$  and  $\Pi$  is the internal stress tensor [114].

Selecting the appropriate thermodynamical ensemble and implementing thermostats and barostats correctly are critical for accurately simulating physical systems in MD simulations. By choosing the suitable ensemble and control mechanisms, researchers can ensure that their simulations reflect realistic conditions, enabling the study of various phenomena from molecular interactions to phase transitions.

## 2.2 Enhanced Sampling

Enhanced sampling techniques are essential in MD simulations to overcome the limitations of conventional MD, which often fail to efficiently sample the complete free-energy landscape within reasonable simulation timescales. Conventional MD simulations model the system in thermal equilibrium, most commonly in the isothermal-isobaric (NPT) ensemble, exploring the free energy landscape and configurational states freely. For ergodic sampling, meaning the ensemble average matches the time average, the free energy  $G$  of a macrostate  $s$  can be determined using Boltzmann's constant  $k_B$ , temperature  $T$ , and the probability  $p$  of finding the system in state  $s$ :

$$G(s) = -k_B T \ln p(s). \quad (2.31)$$

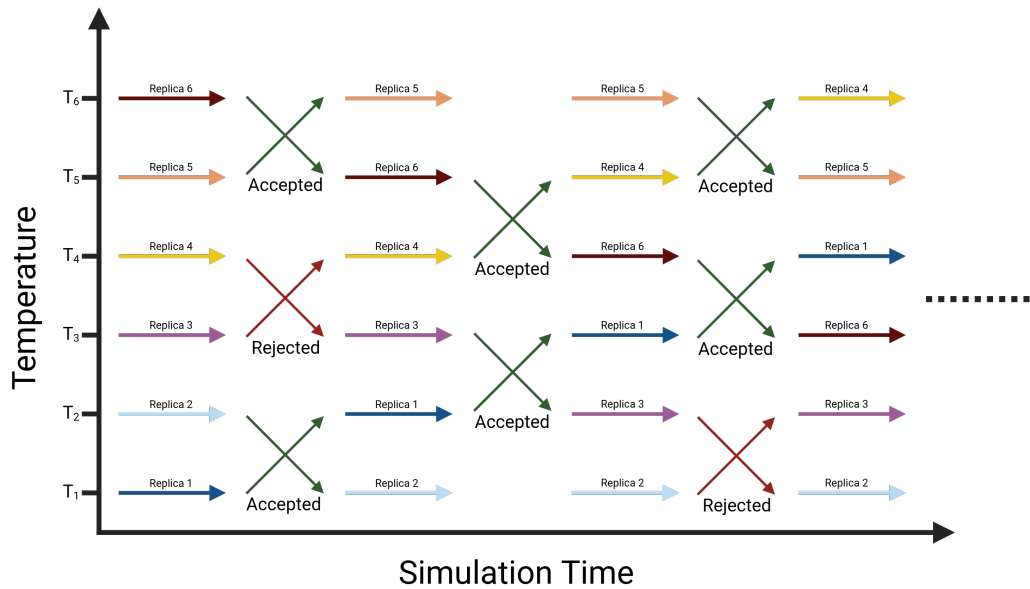
While theoretically, a sufficiently long runtime would sample the entire phase space and achieve ergodicity, in reality, this is rarely feasible. Therefore, the calculation of free energy difference  $\Delta G$  between two states must be considered cautiously, as conventional MD simulations can become trapped in local energy minima, ex-

haustively exploring that minima and skewing the free energy calculation. This issue is particularly challenging for simulations involving "rare events" or complex biomolecular systems such as intrinsically disordered proteins. To address these issues, various advanced sampling techniques, such as metadynamics [115], umbrella sampling [116], Gaussian accelerated molecular dynamics [117], and replica exchange molecular dynamics (REMD) [118, 119], have been developed.

Here, we focus on REMD as an efficient algorithm for studying the structure and dynamics of IDPs. In REMD, multiple copies (replicas) of the original system are simulated in parallel, each at different temperatures or using different Hamiltonians. Exchanges between neighbouring replicas are attempted at regular intervals, as shown in Figure 2.1, based on the Metropolis-Hastings criterion [120]:

$$P(i \leftrightarrow j) = \min \left( 1, \exp \left[ - \left( \frac{1}{k_B T_i} - \frac{1}{k_B T_j} \right) (U_j - U_i) \right] \right), \quad (2.32)$$

where  $T_i$  and  $T_j$  are the temperatures of replicas  $i$  and  $j$ , respectively, and  $U_i$  and  $U_j$  are their corresponding energies. For an exchange to occur, the potential energy distributions of both states must overlap. The overlap between adjacent replicas is typically adjusted to achieve an exchange probability in the range of  $P(i \leftrightarrow j) \sim 0.2 - 0.4$ . This ratio is influenced by the desired temperature range and the system size, as the number of required replicas to achieve the exchange probability scales with  $\sim \sqrt{N}$ . REMD enhances sampling over the conformational



**Fig. 2.1.: Replica exchange molecular dynamics simulation scheme.** Replicas (shown in different colors) start at various temperatures and periodically attempt exchanges with adjacent replicas. Exchanges are either rejected (red) or accepted (green). If accepted, the temperatures of the corresponding replicas are rescaled, and the simulations continue at the new temperatures.

landscape by facilitating the crossing of high-energy barriers at higher temperatures, thereby enabling the accurate computation of thermodynamic properties without prior selection of collective variables (CVs). Unlike CV-based methods (such as umbrella sampling), which require prior knowledge of the system under study, REMD alleviates this problem, making it particularly useful for studying IDPs and their conformational ensembles.

### 2.2.1 Multistate Bennett Acceptance Ratio method

Enhanced sampling techniques like REMD efficiently sample equilibrium ensembles at multiple states. However, combining these simulations could enhance the precision of thermodynamic observables. The multistate Bennett acceptance ratio (MBAR) [121, 122] method is a statistical estimator that facilitates accurate calculations of free energy differences and thermodynamic expectations by utilizing samples derived from multiple equilibrium states. Applicable in both simulation and experimental contexts, MBAR provides a flexible approach well suited to combine the data of multiple equilibrium samples from enhanced sampling methods, such as trajectories at different temperatures in REMD simulations. Unlike multiple histogram reweighting techniques [123], MBAR does not rely on energy binning, eliminating a significant bias of such approaches. Furthermore, MBAR yields precise estimates of free energy differences while also delivering statistical uncertainty estimates for all calculated quantities.

Suppose we have  $K$  thermodynamic states, each sampled with  $N$  uncorrelated equilibrium samples. In the isothermal–isobaric (NPT) ensemble, the reduced potential energy function for state  $i$  is defined by:

$$u_i(\mathbf{R}) = \beta_i [U_i(\mathbf{R}) + p_i V(\mathbf{R})], \quad (2.33)$$

where  $\mathbf{R}$  is the system's configuration in the given configurational space, with  $V(\mathbf{R})$  as its volume, and  $U_i(\mathbf{R})$  representing its potential energy. Here,  $\beta_i$  denotes the inverse temperature of state  $i$ , and  $p_i$  the external pressure. MBAR is derived from a set of  $K^2$  weighting functions, which is chosen to minimize variance within the reweighting calculations across all states involved. Under Boltzmann statistics, this leads to an equation for the dimensionless free energies  $\hat{f}_i$

$$\hat{f}_i = -\ln \sum_{j=1}^K \sum_{n=1}^N \frac{\exp[-u_i(\mathbf{R}_{j,n})]}{\sum_{k=1}^K N \exp[\hat{f}_k - u_k(\mathbf{R}_{j,n})]}, \quad (2.34)$$

which is solved self-consistently. Here, free energies are defined up to an additive constant, thus only the free energy differences  $\Delta\hat{f}_{ij} = \hat{f}_j - \hat{f}_i$  hold physical significance.



## 2.2.2 Principle Component Analysis

The high dimensionality of conformational data generated by MD simulations poses significant challenges for analysis, often requiring dimensionality reduction techniques to highlight key features within the dataset. Principal component analysis (PCA) is a widely used method for systematically determining CVs of maximum variance, thereby capturing the system's most significant motions [124–126].

Suppose we have an ensemble of structures  $\mathbf{r}_i \in \mathbf{R}$  with  $\dim(\mathbf{R}) = N$ , then their covariance matrix can be written as:

$$\mathbf{C}_{ij} = \langle (\mathbf{r}_i - \langle \mathbf{r}_i \rangle) \cdot (\mathbf{r}_j - \langle \mathbf{r}_j \rangle) \rangle. \quad (2.35)$$

Here,  $\mathbf{C}$  is a symmetric  $N \times N$  matrix, which can be diagonalized by a matrix  $\mathbf{D}$ ,

$$\mathbf{D}^T \mathbf{C} \mathbf{D} = \text{diag}(\lambda_1, \dots, \lambda_N), \quad (2.36)$$

where  $\lambda_i$  represents the eigenvalues of  $\mathbf{C}$ , and the rows  $\mathbf{d}_i$  of  $\mathbf{D}$  are the eigenvectors, known as principal components (PCs). The eigenvectors  $\mathbf{d}_i$  build an orthonormal basis set, spanning the principle component space. Although this space has the same dimensionality as the input space, only a few PCs are often required to describe most of the systems motion, particularly those with the largest eigenvalues, which capture the modes of highest amplitude.

In Chapter 7, we utilize PCA to visualize the FES of the SH3 domain, highlighting the dominant motions observed in an REMD simulation. For increased accuracy, the MBAR method is applied to combine the FES data across temperatures sampled during the REMD simulation.

## 2.3 Transition Networks

Combining the conventional MD approach with network-based models, such as Markov state models (MSMs) [127, 128] or conformational transition networks (TNs) [129, 130], provides a comprehensive understanding of the structural behavior of IDPs. These networks reveal the underlying mechanisms of molecular processes hidden within the vast amounts of MD simulation data by generating human-interpretable networks that help illustrate the molecular processes under investigation. The TN approach pursued by our group offers a solution for those seeking a network model of protein motions captured by MD simulations, including explicit modeling of the protein's environment, without needing a method that sets up a master equation for the dynamics. Instead, TNs rely purely on geometric clustering to extract crucial features of protein conformational transitions.

To construct a TN, one must define a set of  $n$  features that describe the process under study and discretize the frames of a given MD trajectory into macrostates. These

features are evaluated by descriptor functions  $\{f_i\}$  that act on a given conformation  $\mathbf{r}(t)$  and project the  $3N$ -dimensional phase space onto an  $n$ -dimensional state  $\mathbf{S}(t)$ :

$$\mathbf{r}(t) \mapsto \mathbf{S}(t) = [f_1(\mathbf{r}(t)), f_2(\mathbf{r}(t)), \dots, f_n(\mathbf{r}(t))]. \quad (2.37)$$

Here,  $\mathbf{r}(t)$  is the conformation of an MD simulation at time  $t$  and  $N$  is the number of particles within the system. The key step in TN analysis is selecting a set of descriptor functions that optimally reflect the structures and dynamics of the system. A balance between complexity and simplicity is crucial. While more descriptors might provide a precise picture of the process, they may also make the network overly complex and difficult to interpret intuitively. This balance is achieved by quantifying the quality of descriptor combinations, choosing those with high sensitivity and low correlation. Sensitivity is defined by how much of the theoretical value range of a descriptor is sampled during the simulations. Correlation between two descriptors  $f_i$  and  $f_j$  is defined by the correlation coefficient  $c_{corr}$ :

$$c_{corr} = \frac{\Sigma_t(f_i(\mathbf{r}(t)) - \bar{f}_i)(f_j(\mathbf{r}(t)) - \bar{f}_j)}{\sqrt{\Sigma_t(f_i(\mathbf{r}(t)) - \bar{f}_i)^2} \sqrt{\Sigma_t(f_j(\mathbf{r}(t)) - \bar{f}_j)^2}} \quad (2.38)$$

where  $\bar{f}_i$  and  $\bar{f}_j$  are the arithmetic mean values of the corresponding descriptor values along the trajectory. High correlation between two descriptors suggests that using both would not add significant information to the TN but would increase its complexity. Thus, it is recommended to use only one of the correlated descriptors in generating the TN.

To describe the process of conformational switching of IDPs and the aggregation propensity of amyloid formers, we chose the descriptors: i) the number of residues exhibiting  $\alpha$ -helical secondary structure ( $N_\alpha$ ), ii) the number of residues adopting a  $\beta$ -sheet structure ( $N_\beta$ ), and iii) the end-to-end distance ( $d_{NC}$ , in Å) of the protein as a measure of compactness in the studies presented in chapter 3 and 4<sup>1</sup>.

## 2.4 Distribution of Reciprocal Interatomic Distances

A method to define microstates with reduced bias involves structural clustering, rather than predefined descriptors. However, given the size of most molecular systems, structure-based clustering in Cartesian coordinates becomes a challenging task. One therefore usually relies on some form of dimensionality reduction, which should preserve as much of the kinetics and structural features as possible. It has been shown that the distribution of reciprocal interatomic distance (DRID) metric is a good candidate to meet both of these criteria [132, 133]. A key feature of the DRID metric is the use of the multiplicative inverse (reciprocal) distances, which highlights the difference in short distances, while not neglecting changes in large distances completely.

<sup>1</sup>Section 2.3 is largely taken from Schaeffler et al., 2022 [131].

To apply the DRID metric, two essential atom sets are defined: a set of  $m$  centroids  $\mathcal{C}$  representing key structural elements, and a set of  $N$  reference atoms  $\mathcal{A}$  (excluding atoms that are covalently bound to the centroid). For a given structure, the distribution of reciprocal interatomic distances for each centroid  $i \in \mathcal{C}$  and the first three moments of that distribution  $(\mu_i, \nu_i, \xi_i)$  are calculated, resulting in a  $3m$  dimensional vector for each structure (i.e., each frame of the MD trajectory). The moments are defined as follows,

$$\mu_i = \frac{1}{N-1-nb_i} \sum_i^N \frac{1}{d_{ij}}, \quad (2.39)$$

$$\nu_i = \left[ \frac{1}{N-1-nb_i} \sum_i^N \frac{1}{(d_{ij} - \mu_i)^2} \right]^{1/2}, \quad (2.40)$$

$$\xi_i = \left[ \frac{1}{N-1-nb_i} \sum_i^N \frac{1}{(d_{ij} - \mu_i)^3} \right]^{1/3}, \quad (2.41)$$

where  $d_{ij}$  denotes the distance of atom  $a_j \in \mathcal{A}$  to centroid  $c_i \in \mathcal{C}$  and  $nb_i$  is the number of covalent bonds of a centroid. The distance metric  $s_{jk}$  between a pair of conformations  $j$  and  $k$  in DRID space is defined as

$$s_{jk} = \frac{1}{3m} \sum_i^m \left[ (\mu_i^j - \mu_i^k)^2 + (\nu_i^j - \nu_i^k)^2 + (\xi_i^j - \xi_i^k)^2 \right]^{1/2}. \quad (2.42)$$

To group the structures into states, we performed regular space clustering in DRID space, as implemented in the PyEMMA python package [134], using the  $s_{jk}$  distance metric<sup>2</sup>.

## 2.5 Free Energy Calculation

The free energy surface (FES) of a protein dictates its structural and dynamical properties, making it crucial for understanding protein function. We calculate the free energies associated with states identified by structural clustering in DRID space, treating each state as a minimum in the FES. This method is applicable to any given transition/rate matrix. The free energy of the minima  $F_i$  is calculated using their occupation probability  $p_i$ ,

$$F_i = -k_B T \log(p_i), \quad (2.43)$$

where  $k_B$  is the Boltzmann constant and  $T$  is the temperature of the system.

<sup>2</sup>Section 2.4 is largely taken from the Supporting Information of Schaeffler et al., 2024 [135].

The rate matrix  $\mathbf{R}$ , which represents the transition rates  $r_{jk}$  between minima  $j$  and  $k$  observed in MD simulations, is used to estimate the transition state free energies  $F_{jk}$  from the interconversion rates  $k_{jk}$  between states,

$$F_{jk} = F_k - k_B T \log(k_{jk}) + k_B T \log\left(\frac{k_B T}{h}\right), \quad (2.44)$$

where  $h$  is the Planck constant. In a perfectly converged system, the transition state free energies for both interconversion rates between minima  $j$  and  $k$  should be equal, i.e.  $F_{jk} = F_{kj}$ . However, this equality is rarely achieved in finite MD trajectories. To minimize the error, we average both rates,

$$F_{jk}^{\text{ts}} = \frac{F_{jk} + F_{kj}}{2}, \quad (2.45)$$

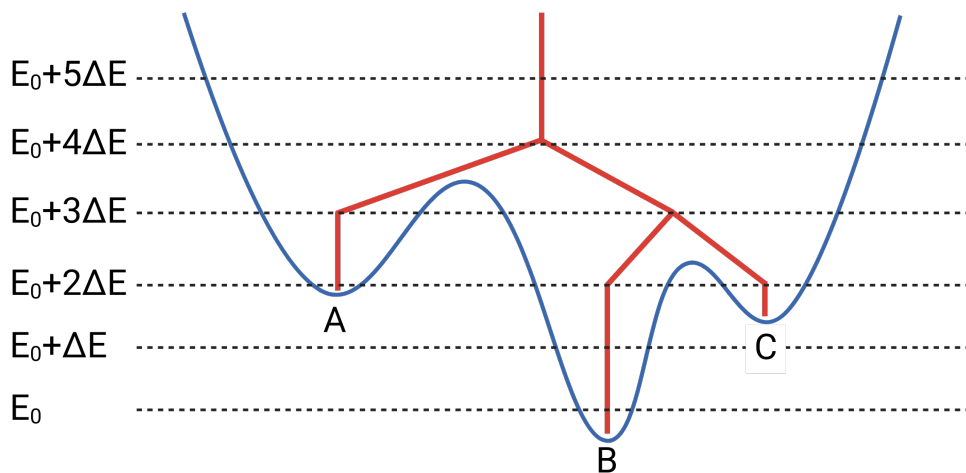
providing an estimate for the transition state free energy  $F_{jk}^{\text{ts}}$  between minima  $j$  and  $k$ <sup>3</sup>.

## 2.6 Disconnectivity Graphs

Disconnectivity graphs provide a powerful visualization of the connectivity within an energy landscape by grouping minima, or basins, at different energy thresholds, connected by transition states with energies below these specified thresholds (see Figure 2.2). The energy difference between thresholds ( $\Delta E$ ) should be chosen to balance the resolution of structurally distinct energy basins and the grouping of similar minima. In this study,  $\Delta E$  was set to 0.5 kT for all presented FESs, where  $k$  is the Boltzmann constant  $k_B$  and  $T$  the temperature of the system ( $\sim 310$  K). At each threshold, minima are grouped into superbasins, creating a tree-like structure where branches terminate at the energy of local minima, with the vertical scale representing free energy. Going upwards in energy, branches converge at thresholds where the free energy surpasses the highest transition state on the lowest path between them, indicating sets of minima that can interconvert at a given energy level. This method enables identification of folding funnels and a dimensionality-free visualization of the hierarchical connectivity in the FES, independent of CVs [136, 137]. In this study, each minimum represents an ensemble of structures, grouped either by clustering in DRID space or through descriptor functions<sup>4</sup>.

<sup>3</sup>Section 2.5 is largely taken from the Supporting Information of Schaeffler et al., 2024 [135].

<sup>4</sup>Section 2.6 is largely taken from Schaeffler et al., 2024 [135].



**Fig. 2.2.: Schematic representation of a disconnectivity graph.** Example of the construction of a disconnectivity graph from a database of minima (A-C) of a free energy landscape (blue). In the disconnectivity graph (red), each local minimum is depicted by a vertical line starting at its energy. At a specified energy threshold  $E + n\Delta E$ , where two or more minima are no longer separated by an energy barrier, i.e. their transition state lies below that threshold, they are grouped into disjoint sets and become part of the same superbasin within the FES.

## 2.7 First Passage Times

While the FES dictates a molecule's structural and dynamical properties, experiments often measure the relaxation times associated with specific processes. Studying the timescales for transitions between minima on the FES can bridge the gap between simulation and experiment, providing insights into these processes. Interconversion rates between minima, often quantified by the mean first passage time (MFPT), reveal the average time for a system to transition between reactant and product states. In the context of amyloid aggregation, the reactant state corresponds to the IDP state, and the product state corresponds to oligomeric or fibril structures. Analyzing the first passage time (FPT) distribution offers additional insights into the organizational structure of the underlying energy landscape, identifying distinct signatures linked to relaxation to different funnels in the FES [138–140].

For a given transition  $A \leftarrow B$  from reactant state B to product state A, the first passage time probability distribution  $p(t)$  is obtained by treating the product state as absorbing. By treating A as absorbing state all trajectories end in the product state, but the dynamics are not changed until the absorbing event. The master equation for the occupation probabilities  $P_\alpha(t)$  of the joined set of intervening states I and reactant states  $I \cup B$  reads as:

$$\begin{bmatrix} \dot{\mathbf{P}}_I(t) \\ \dot{\mathbf{P}}_B(t) \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{II} - \mathbf{D}_I & \mathbf{K}_{IB} \\ \mathbf{K}_{BI} & \mathbf{K}_{BB} - \mathbf{D}_B \end{bmatrix} \begin{bmatrix} \mathbf{P}_I(t) \\ \mathbf{P}_B(t) \end{bmatrix} = \mathbf{M}\mathbf{P}_{I \cup B}(t), \quad (2.46)$$

where  $\mathbf{K}_{XY}$  is the rate matrix for transitions between connected states, and  $\mathbf{D}_X$  is a diagonal matrix containing the escape rates of each state in  $X$ , i.e.,  $[\mathbf{D}_X]_{ii} = \sum_j \mathbf{K}_{ji}$ . For an initial distribution  $P_A(0) = \mathbf{0}_A$ , the probability  $p(t)dt$  that the first passage time lies between  $t$  and  $t + dt$ , which is the probability flux out of  $I \cup B$ , can be written as:

$$p(t)dt = -(\mathbf{1}_I \dot{\mathbf{P}}_I + \mathbf{1}_B \dot{\mathbf{P}}_B)dt \equiv -\mathbf{1}_{I \cup B} \dot{\mathbf{P}}_{I \cup B} dt, \quad (2.47)$$

where  $\mathbf{1}_X$  is a row vector of ones with dimension  $N_X$ . Rewriting Equation 2.46 as  $\dot{\mathbf{P}}_{I \cup B}(t) = \mathbf{M} \mathbf{P}_{I \cup B}(t)$ , we get the formal solution:

$$\mathbf{P}_{I \cup B}(t) = \exp(\mathbf{M}t) \mathbf{P}_{I \cup B}(0), \quad (2.48)$$

and thus,

$$p(t) = -\mathbf{1}_{I \cup B} \mathbf{M} \exp(\mathbf{M}t) \mathbf{P}_{I \cup B}(0). \quad (2.49)$$

Applying eigenvector decomposition to  $\mathbf{M}$ , we can derive an analytic solution for the first passage time distribution:

$$\mathbf{M} = -\sum_l \nu_l \mathbf{w}_l^R \otimes \mathbf{w}_l^L, \quad (2.50)$$

where  $\mathbf{w}_l^L$  and  $\mathbf{w}_l^R$  are the left and right eigenvectors, and  $\otimes$  denotes the dyadic (outer) product. Furthermore, the eigenvalues  $-\nu_l < 0$  are strictly negative for a connected network. Using this expansion, we obtain:

$$\exp(\mathbf{M}t) = \sum_l \exp(-\nu_l t) \mathbf{w}_l^R \otimes \mathbf{w}_l^L. \quad (2.51)$$

Together with Equation 2.49 we obtain the analytic equation for the first passage time distribution:

$$p(t) = \sum_l \nu_l e^{-\nu_l t} \mathbf{1}_{I \cup B} [\mathbf{w}_l^R \otimes \mathbf{w}_l^L] \mathbf{P}_{I \cup B}(0). \quad (2.52)$$

which can be further simplified to:

$$p(t) = \sum_l \nu_l e^{-\nu_l t} A_l, \quad (2.53)$$

where  $A_l$  are amplitudes depending on the eigenvectors of  $\mathbf{M}$ . The FPT can be represented by the probability distribution  $\mathcal{P}(y)$  for  $y = \log(t)$ :

$$\mathcal{P}(y) = \sum_l \nu_l e^{y - \nu_l \exp(y)} A_l, \quad (2.54)$$

which reveals distinct peaks corresponding to different features and relaxation time scales in the FPT distribution <sup>5</sup>.

---

<sup>5</sup>Section 2.7 is largely taken from the Supporting Information of Schaeffler et al., 2024 [135].

# Automated Generation of Transition Networks for the Structural Characterization of Intrinsically Disordered Proteins

Intrinsically disordered proteins (IDPs) lack a stable three-dimensional structure and instead sample a diverse array of configurations influenced by their environment. This structural heterogeneity presents challenges for characterization using experimental techniques alone. Molecular dynamics (MD) simulations offer a valuable complement, capturing the dynamic behavior of IDPs across multiple timescales. However, the vast data generated by MD simulations necessitates sophisticated post-processing to extract meaningful insights. In this chapter, we overview how transition networks (TN), derived from MD trajectories, effectively reveal the configurational ensemble and structural transitions of IDPs, using amyloid- $\beta$  in complex with 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids as a case study. We also discuss the impact of molecular descriptor selection on the resulting transition networks that were calculated with our own open-source Python script ATRANET. This chapter provides an overview of the results published in:

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

*Moritz Schäffler, Mohammed Khaled and Birgit Strodel*  
Methods, **206** (2022)

The full publication is provided in the appendix A.1. The contributions of the authors are listed in Table 3.1.

Author	Contribution	Tasks
<u>Moritz Schäffler</u>	~60%	Establishing the general methodology, Writing the ATRANET software package, Validation, Formal-analysis, Visualization, Writing-original-draft
Mohammed Khaled	~20%	Contribution to the ATRANET software package, Validation, Data-curation, Writing-original-draft
Birgit Strodel	~20%	Conceptualization, Methodology, Resources, Writing-review-editing, Supervision, Project-administration, Funding-acquisition

**Tab. 3.1.:** Contribution by author to Schäffler et al. 2022 [131].



## 3.1 Methods

The methods discussed in this paper are presented in chapter 2.3. For convenience they are summarized in condensed form in the following.

Constructing a TN requires defining a set of descriptor functions  $\{f_i\}$  that discretize the MD trajectory. By applying these descriptor functions to each time point  $t$  within the trajectory, a state trajectory  $\mathbf{S}(t)$  is generated, capturing the descriptor values for the molecular conformations  $\mathbf{r}(t)$ :

$$\mathbf{S}(t) = [f_1(\mathbf{r}(t)), f_2(\mathbf{r}(t)), \dots, f_n(\mathbf{r}(t))] \quad (3.1)$$

A critical step in TN analysis is the selection of descriptor functions that accurately capture the system's structural and dynamic properties. Choosing the type and number of descriptor functions, it is essential to find a balance between complexity and simplicity. In the present work we investigated how additional descriptors can offer a more detailed representation of the dynamics, but they may also complicate the network, rendering it less intuitive to interpret.

To address this, we generated TNs using two sets of descriptor functions. Initially, we employed three descriptors: (i) the number of residues exhibiting  $\alpha$ -helical secondary structure ( $N_\alpha$ ), (ii) the number of residues adopting a  $\beta$ -sheet structure ( $N_\beta$ ), and (iii) the end-to-end distance ( $d_{NC}$ , in Å). This approach led to a state space discretization defined by the parameters ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}$ ). Subsequently, we introduced a fourth descriptor,  $N_{p-l}$ , which refers to the number of contacts formed between the A $\beta_{42}$  peptide and the POPC lipids, and examined the TN resulting from the combined set of descriptors ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}$ ,  $N_{p-l}$ ).

The primary insights provided by TNs are encoded in their visualization, making this a critical component of the analysis. The transition matrices generated by ATRANET were visualized using Gephi [141, 142]. The node sizes were scaled proportionally to the population of the corresponding states, as computed by the ATRANET script. Here, the population of a state is defined by its diagonal entry in the transition matrix. This scaling enables the easy identification of predominant and thus more stable states by comparing node sizes. The layout of a TN is crucial, as an effective representation allows for the quick identification of node connectivity and distinct clusters within the system. We utilized the *Force Atlas 2* algorithm, a force-driven approach that considers the connectivity of node pairs and their relative degrees. Consequently, nodes with more frequent transitions between them are positioned closer together, meaning strong spatial separation in the TN layout indicates a high distance between the respective states in the high-dimensional phase space. Given the potentially overwhelming number of states resulting from multiple descriptors, we identified local communities within the network to highlight clusters of states with high connectivity.

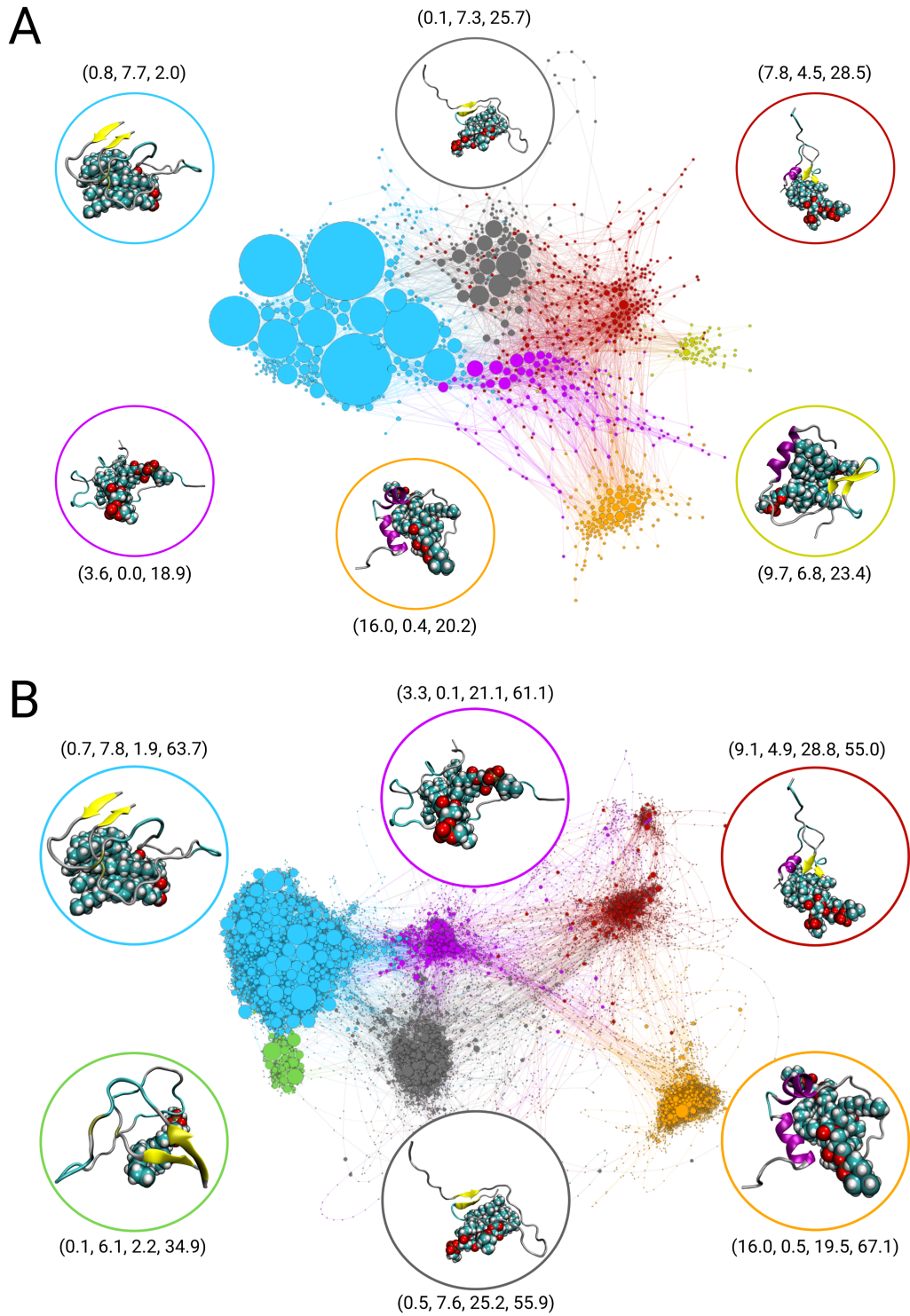
The simulations were conducted using the GROMACS software package [143]. Since the CHARMM36mW force field only contains parameters for polypeptides, the A $\beta$ <sub>42</sub> peptide was modeled with the CHARMM36mW force field [82], while the POPC lipids were modeled with the CHARMM36 force field [144]. Initially, the peptide and three lipid molecules were positioned apart in the simulation box, which was filled with TIP3P water [97] and NaCl at a physiological concentration of 150 mM. After equilibration, we performed four 1  $\mu$ s simulations under NPT conditions, totaling 4  $\mu$ s of simulation time. Pressure was regulated with the Parrinello-Rahman coupling scheme [108] ( $p = 1$  bar), and temperature was controlled using a Nosé–Hoover thermostat [105, 106] ( $T = 310$  K) for canonical sampling. Periodic boundary conditions were applied in all directions, with the particle-mesh Ewald method [145] used for electrostatic interactions, and cutoffs of 12 Å were set for both Coulomb and van der Waals interactions.

## 3.2 Results

First, we generated a TN using three descriptors ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}$ ) to characterize the conformational preferences of A $\beta$ <sub>42</sub>. The TN, shown in Fig. 3.1 A, includes representative structures for each of the six communities, corresponding to the most populated state within each cluster.

The TN is primarily dominated by a blue-colored community with average descriptor values of ( $N_\alpha=0.8$ ,  $N_\beta=7.7$ ,  $d_{NC}=2.0$  Å), indicating more stable configurations with a small end-to-end distance and a predominance of  $\beta$ -sheet structures. In contrast, the orange-colored cluster, visited less frequently, represents configurations with a larger end-to-end distance and mostly  $\alpha$ -helical structures ( $N_\alpha=16.0$ ,  $N_\beta=0.4$ ,  $d_{NC}=20.2$  Å). Intermediate clusters, such as the gray cluster ( $N_\alpha=0.1$ ,  $N_\beta=7.3$ ,  $d_{NC}=25.7$  Å), display extended structures with larger end-to-end distances and partial  $\beta$ -sheet formation. The transitions between these communities illustrate the structural interconversions of A $\beta$ <sub>42</sub>, with shifts from extended and mixed  $\alpha/\beta$  states to more compact  $\beta$ -sheet structures. However, the three descriptors used in this TN do not capture the interactions between the peptide and lipids, which are crucial for understanding structure formation in this context.

To address this, we extended our analysis by incorporating a fourth descriptor ( $N_{p-l}$ ), representing the number of contacts between the A $\beta$ <sub>42</sub> peptide and POPC lipids, in addition to the original three descriptors: ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}, N_{p-l}$ ). This modification resulted in a more distinct separation of clusters and clarified transitions between them, as shown in Fig. 3.1 B. The TN largely retained the original cluster structure, with the most significant change being the subdivision of the dominant blue-colored cluster into two clusters, differentiated by their peptide–lipid contact count. The newly defined TN highlights that structural transitions within the peptide occur without detachment from the lipid cluster, providing insights that were not captured in earlier analyses. This refined understanding emphasizes the role of



**Fig. 3.1.: Transition network of  $A\beta_{42}$  forming a complex with three POPC lipids.** The TNs are constructed using (A) three descriptors and (B) four descriptors. The descriptors used are: (i) the number of residues with  $\alpha$ -helical structure  $N_\alpha$ , (ii) the number of residues with  $\beta$ -sheet structure  $N_\beta$ , (iii) the end-to-end distance  $d_{NC}$  and (iv) the number of peptide-lipid contacts  $N_{p-l}$ . For (A) the descriptors (i–iii) were used and for (B) the descriptors (i–iv). The nodes are colored based on their community membership, with representative structures for the most populated state of each community displayed within circles matching the cluster's color. Additionally, the average values for the descriptors are provided for each community.

peptide–lipid interactions in the conformational dynamics of A $\beta$ <sub>42</sub>, particularly in the context of amyloid aggregation.

### 3.3 Conclusion

Studying the structural characteristics of IDPs poses significant challenges due to their lack of a unique three-dimensional structure and the complexity of their conformational transitions. MD simulations offer a powerful complement to experimental techniques for investigating these properties. In this study, we demonstrated the utility of TNs derived from MD simulations to elucidate the structural changes of an IDP, using A $\beta$ <sub>42</sub> in complex with POPC lipids as a model system.

We identified three key features of TNs: (i) the size of nodes indicates the most stable polypeptide configurations, (ii) the spatial layout of nodes reflects the phase-space distances between conformations, and (iii) clustering within the TN highlights significant conformational transitions. These features were demonstrated using two sets of descriptors: one with three descriptors ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}$ ) and another adding a fourth descriptor ( $N_{p-l}$ ) to capture peptide–lipid interactions. The inclusion of the fourth descriptor provided a more detailed separation of clusters, making the conformational transition pathways more apparent. It revealed that structural transitions within the peptide occur without detachment from the lipid cluster.

While TNs provide valuable insights, they are inherently influenced by the *a priori* selection of descriptors, which can impact both the network’s structure and the conclusions derived from it. Nonetheless, the choice of descriptors is often guided by the specific research question being addressed.

An intriguing aspect of TN is their similarity to FESs, as both represent the configurational space explored by a system. For IDPs, the presence of distinct clusters in TNs with similar occupation probabilities, as observed here, indicates an underlying multifunnel structure in the FES. While FESs offer quantitative insights into energy differences between states, they typically rely on projections onto one or more collective variables, which can lead to inaccuracies in capturing energy barriers. As an alternative, disconnectivity graphs, constructed from TN data, provide a more comprehensive view of the multidimensional energy landscape without the need for such projections. We will explore the insights gained from such an analysis in chapter 5.

## Transition Networks Reveal the Conformational Switching of A $\beta$ <sub>42</sub>

The aggregation of A $\beta$ -peptides, particularly A $\beta$ <sub>42</sub>, is closely associated with the pathogenesis of Alzheimer's disease. In this study, we investigate the conformational changes of A $\beta$ <sub>42</sub> in the presence of the glycosaminoglycan (GAG) chondroitin-4-sulfate with 8 sub units and three 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids using MD simulations. Chapter 3 introduces an initial analysis of A $\beta$ <sub>42</sub> with POPC lipids, which is extended here to illustrate changes in the sampled configurational space relative to A $\beta$ <sub>42</sub> in solution. GAGs, polysaccharides found in various tissues, are frequently associated with A $\beta$  plaques and have been shown to be associated with amyloid aggregation, although the precise molecular mechanisms remain unclear [146]. By analyzing the key structural features of A $\beta$ , we construct transition networks (TNs) that highlight the shifts from intrinsically disordered states, common in A $\beta$  in solution, to more compact states characterized by stable  $\beta$ -hairpin formation near GAG molecules, and even more compact structures exhibiting  $\alpha$ -helix or  $\beta$ -sheet formations upon interaction with POPC lipid clusters. Our findings reveal distinct molecular mechanisms driving these transitions: the hydrophobic environment provided by POPC lipids facilitates A $\beta$  folding, whereas GAGs, through the sequestration of sodium ions, enhance electrostatic interactions within A $\beta$ , stabilizing a  $\beta$ -hairpin structure. These insights deepen our understanding of how GAGs and lipids influence the conformational preferences of A $\beta$  peptides, impacting their aggregation into toxic oligomers and amyloid fibrils.

This chapter provides an overview of the results published in:

### **Transition Networks Unveil Disorder-to-Order Transformations in A $\beta$ Caused by Glycosaminoglycans or Lipids.**

*Moritz Schäffler, Suman Samantray and Birgit Strodel*

Int. J. Mol. Sci., **24** (2023)

The full publication is provided in the appendix A.1. The contributions of the authors are listed in Table 4.1.

Author	Contribution	Tasks
<u>Moritz Schäffler</u>	~80%	Methodology, Software, Investigation Data-curation, Validation, Formal-analysis, Visualization, Writing-original-draft, Writing-review-editing
Suman Samantray	~5%	Data-curation, Writing-review-editing
Birgit Strodel	~15%	Conceptualization, Methodology, Resources, Writing-review-editing, Supervision, Project-administration, Funding-acquisition

**Tab. 4.1.:** Contribution by author to Schäffler et al. 2023 [147].

## 4.1 Methods

The methods utilized in this study are outlined in Chapter 2.3 and 3.1. Here, we briefly summarize the key concepts; for more detailed information, the reader is referred to the Appendix A.1.

A critical aspect in constructing a TN is the selection of descriptor functions. As highlighted in our previous work [131] and summarized in Chapter 3, both the type and number of descriptors significantly influence the resulting TN. The type of descriptors should be related to the specific process under investigation, whereas determining the ideal number of descriptors often requires trial and error. Using fewer descriptors simplifies the TN but may result in the loss of crucial information, while using too many descriptors can make the TN overly complex and difficult to interpret. Based on these considerations and the results provided in chapter 3 we decided to use the same three descriptor functions across all systems for ease of comparison: i) the number of residues adopting an  $\alpha$ -helical structure ( $N_\alpha$ ), ii) the number of residues adopting a  $\beta$ -sheet structure ( $N_\beta$ ), and iii) the end-to-end distance ( $d_{NC}$ , in Å) as a measure of peptide compactness. The TNs were constructed using ATRANET, a Python package developed by our group [131] and visualized using Gephi [141, 142].

In all simulations, A $\beta_{42}$  was modeled with neutral histidine residues and no terminal capping, resulting in an overall peptide charge of  $q = -3$ . All three systems variants, A $\beta_{42}$  in solution, A $\beta_{42}$  with GAG, and A $\beta_{42}$  with three POPC lipids were simulated using the GROMACS simulation package [143]. As two of the three systems were adapted from previous studies (A $\beta_{42}$  in solution and A $\beta_{42}$  with three POPC lipids), some of the MD simulation settings differ slightly. While this is not ideal, these differences appear to be negligible compared to the remarkable distinctions in the results presented below. In all simulations the CHARMM36m force field [82] was used for modeling the A $\beta_{42}$  peptide, with POPC lipids parameterized by CHARMM36 [144] and GAG parameters provided by the Glycan Reader & Modeler module [148–150] of the CHARMM-GUI web server [151]. The structure of the simulated GAG molecule and POPC lipid is depicted in Figure 4.1.

All systems were prepared using a consistent protocol: solute(s) were placed in the simulation box, which was filled with TIP3P water molecules [97] and NaCl at a physiological concentration of 150 mM. After equilibration, simulations were conducted under NPT conditions at 1 bar, using a Parrinello–Rahman pressure coupling scheme [108]. The A $\beta$ -GAG and A $\beta$ -POPC simulations were run at 310 K with a Nosé–Hoover thermostat [105, 106], while the A $\beta$ -only system was run at 300 K with a velocity rescaling thermostat [104]. Simulation durations were as follows: A $\beta$ -GAG, 4  $\mu$ s; A $\beta$ -POPC,  $2 \times 2 \mu$ s; and A $\beta$ -only,  $1 \times 6 \mu$ s. Simulations were performed under periodic boundary conditions, with electrostatics computed via the particle-mesh Ewald method [145], and cutoffs for van der Waals and Coulomb interactions set to 12 Å.

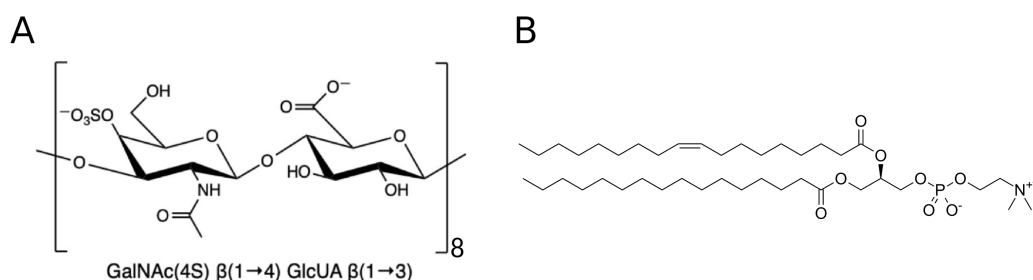
## 4.2 Results

In the following, we present the TNs of A $\beta$  in solution, A $\beta$  in complex with three POPC lipids, and A $\beta$  in the presence of a GAG. Additionally, we discuss the mode of interaction between GAGs and the A $\beta$  peptide.

### 4.2.1 Transition Network of A $\beta$ in solution

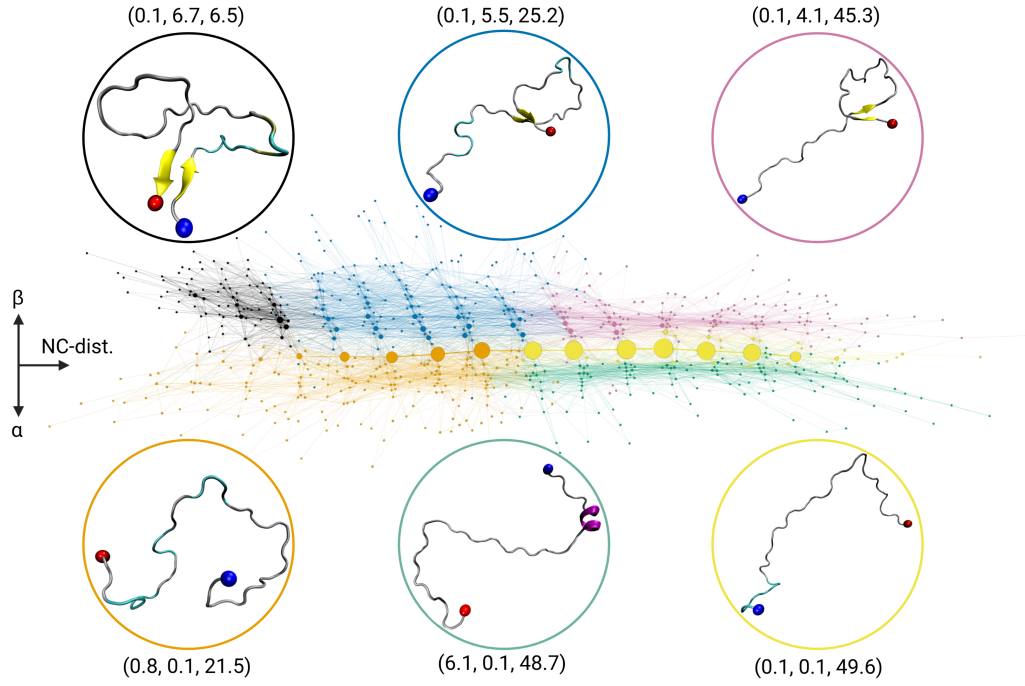
Figure 4.2 shows the TN of the A $\beta$  monomer in solution. The most populated states are located in the orange and yellow communities along the horizontal axis ( $x$ -axis), corresponding to primarily disordered conformations with varying degrees of expansion and no significant secondary structure formation, as indicated by descriptor values (0.8, 0.1, 21.5) and (0.1, 0.1, 49.6). States with increasing amounts of  $\beta$ -sheets are represented by the black, blue, and pink communities toward the upper part of the TN, while the green community at the bottom contains states with  $\alpha$ -helical structures. However, these more structured states are sparsely populated compared to the random-coil states in orange and yellow.

The TN layout effectively illustrates the conformational landscape of the A $\beta$  monomer in solution, where the  $x$ -axis correlates with the NC-distance and the  $y$ -axis reflects the degree of secondary structure. Unlike structured proteins, A $\beta$  shows a low propensity for forming  $\alpha$ -helices or  $\beta$ -sheets. This analysis confirms the classification of the A $\beta$  monomer as an intrinsically disordered protein, characterized by a flat free energy surface with many local minima separated by low energy barriers. As discussed in chapter 5, the underlying FES can be described as an "inverted folding funnel," where the global minimum corresponds to a disordered state, while more structured, partly folded states occupy higher energy and can thus be considered excited states.



**Fig. 4.1.: Chemical structures of the glycosaminoglycan and lipid studied in this work.** (A) Chondroitin-4-sulfate (C4S) glycosaminoglycan, modeled as an 8-subunit chain. The picture is reproduced from Samantray et al. (2021) [83]. (B) 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipid, simulated as a complex of three POPC molecules. The picture is taken from <https://en.wikipedia.org/wiki/POPC>.





**Fig. 4.2.:** Transition network of the A $\beta$  monomer in solution. For state assignment, three descriptors are used: i) the number of residues forming  $\alpha$ -helical structure ( $N_\alpha$ ), ii) the number of residues forming  $\beta$ -sheet structure ( $N_\beta$ ), iii) the NC-distance ( $d_{NC}$ ). The layout of the TN is such that  $d_{NC}$  increases from left to right along the  $x$ -axis,  $N_\beta$  increases in positive  $y$ -direction, and  $N_\alpha$  increases with negative  $y$ -direction. The nodes are colored according to their community membership, and the average descriptor values ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}$ ) of the communities are given. For the highest-populated node per community a representative structure is shown as cartoon ( $\beta$ -sheets in yellow,  $\alpha$ -helices in purple) with the N- and C-termini being indicated by blue and red spheres, respectively.

### 4.2.2 Transition Network of A $\beta$ -GAG

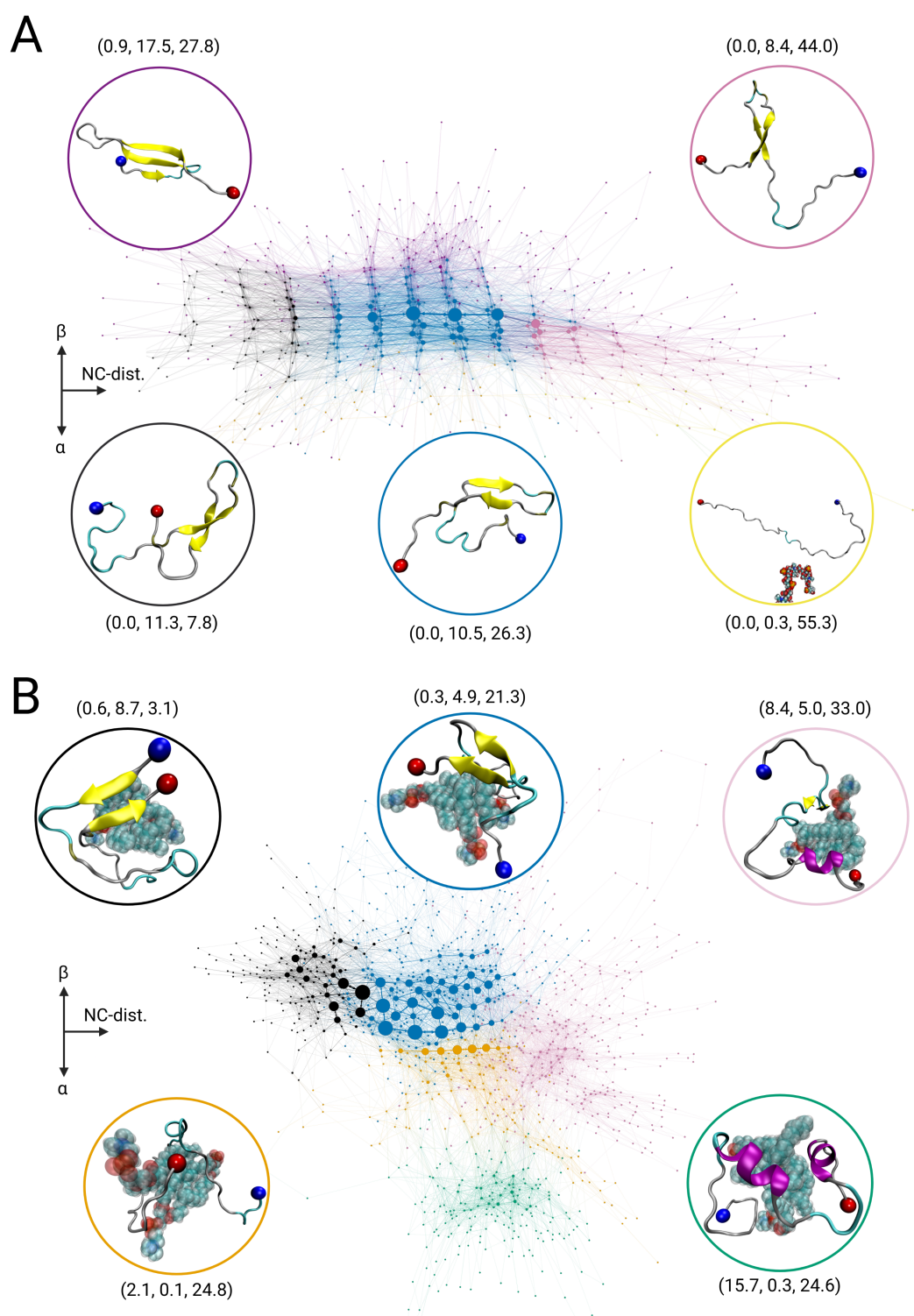
To investigate the conformational changes of A $\beta$  in the presence of a sulfated GAG compared to A $\beta$  in solution, we calculated a TN using the same three descriptors as before (see Figure 4.2). The color scheme for the communities was retained from the A $\beta$  in solution system, emphasizing the shift in population distribution. The layout remains consistent, with increasing NC-distance along the horizontal axis, while states with higher  $\beta$ -sheet content appear in the positive  $y$ -direction.

The TN for the A $\beta$ -GAG system, shown in Figure 4.3 A, exhibits significant differences compared to A $\beta$  in solution. Disordered states, now represented by a single yellow community (0.0, 0.3, 55.3), are scarcely populated. Instead, the most populated states belong to the blue community (0.0, 10.5, 26.3), which contains structures of intermediate compactness and notable  $\beta$ -sheet formation. Additional  $\beta$ -sheet rich states are present in the black (0.0, 11.3, 7.8) and pink (0.0, 8.4, 44.0) communities, indicating varying levels of compactness and NC-distance. At the top of the TN, the purple community (0.9, 17.5, 27.6) contains states with up to 42% of the residues involved in  $\beta$ -sheet formation, forming a three-stranded  $\beta$ -sheet structure.

In conclusion, the TN reveals a significant shift of the global minimum from disordered to  $\beta$ -sheet rich states, driven by the formation of a stable  $\beta$ -hairpin structures. This structural change is notable despite minimal direct contact between A $\beta$  and the GAG, due to electrostatic repulsion between them, which will be discussed below, along with the reason for the conformational shift.

### 4.2.3 Transition Network of A $\beta$ -POPC

To provide context for the conformational changes observed in the A $\beta$ -GAG system, this study also includes the TN for the interaction of A $\beta$  with three POPC lipids (see Figure 4.3 B), as discussed in Chapter 3. As demonstrated in a previous study [42], A $\beta$  undergoes a disorder-to-order transition in the presence of lipids, driven by hydrophobic interactions between A $\beta$  and lipid tails. In the A $\beta$ -POPC system, many conformational states that are not present in the A $\beta$ -only system are sampled, indicating significant differences in peptide dynamics. The TN for the A $\beta$ -POPC system reveals a shift towards more compact,  $\beta$ -sheet rich conformations, unlike the extended disordered states dominant in the A $\beta$ -only system. Furthermore, this system also exhibits a distinct community of  $\alpha$ -helical states, representing a local minimum in the free energy landscape. The comparison highlights the shift of A $\beta$  towards more stable, folded states when interacting with lipids, suggesting a change from a flat free energy surface (as seen in the A $\beta$ -only and A $\beta$ -GAG systems) to one with distinct energy basins, corresponding to semi-folded A $\beta$  conformations.



**Fig. 4.3.: Transition network of the A $\beta$  peptide in the presence interaction partners.** TNs of the A $\beta$  peptide in the presence of (A) a GAG molecule and (B) three POPC lipids are shown. See the caption of Figure 4.2 for explanations of the graphical representation of the TN.

#### 4.2.4 A $\beta$ -GAG interaction

To investigate the conformational changes of A $\beta$  in the presence of a GAG molecule, we examined their molecular interactions and the effect of the GAG on water dynamics and ion distribution around A $\beta$ . Our findings show that while there is minimal direct interaction between A $\beta$  and the GAG (see Figure 4.4 C), the latter induces a significant conformational shift in A $\beta$ . This mode of interaction is distinct from that observed in the A $\beta$ -POPC system, where hydrophobic interactions with lipid tails drive the conformational changes in A $\beta$  [42].

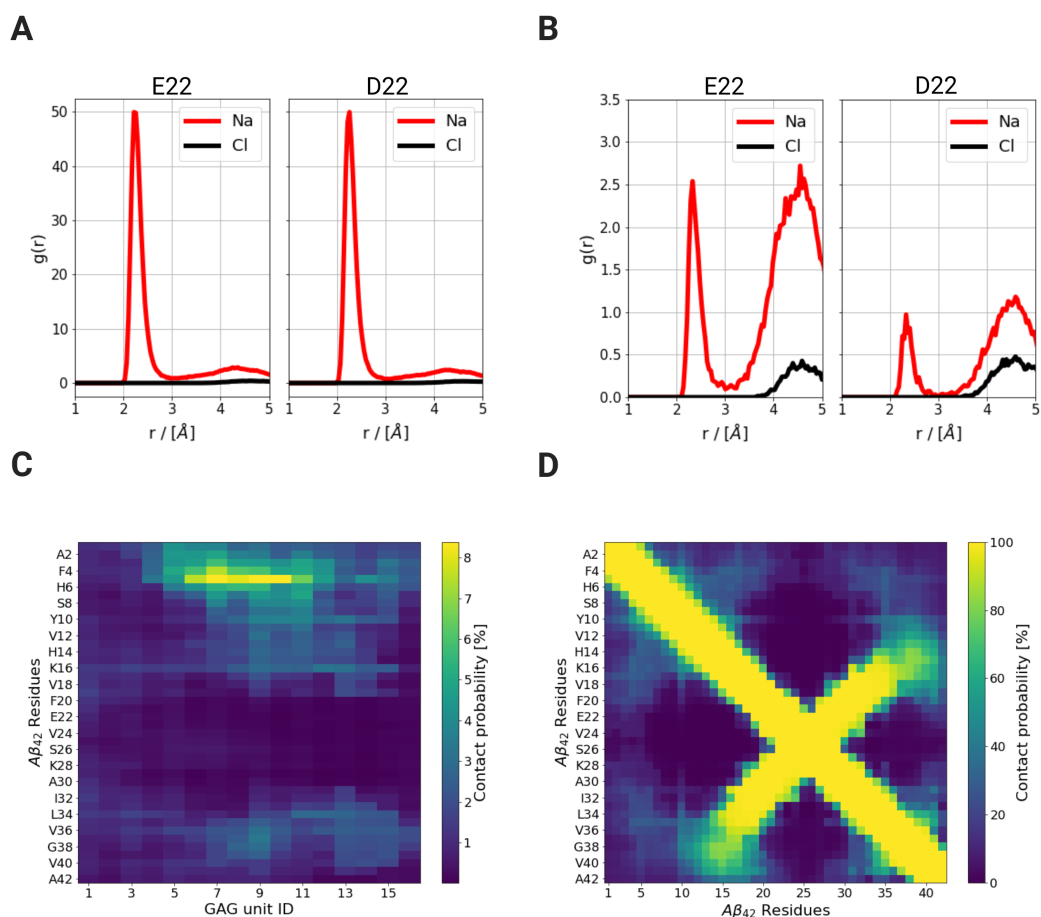
Further analysis of the water structure and dynamics around A $\beta$  indicated no significant differences between the A $\beta$ -only and A $\beta$ -GAG systems, suggesting that the GAG does not affect water-mediated interactions. Instead, the GAG's influence appears to stem from its effect on the ion distribution, particularly the dramatic reduction in Na<sup>+</sup> ions around A $\beta$ . Panels A and B of Figure 4.4 display the radial distribution of Na<sup>+</sup> ions around the carboxyl groups of residues E22 and D23 in A $\beta$  for the A $\beta$ -only and A $\beta$ -GAG systems respectively, which show a decrease in ion concentration of a factor 10 in presence of the GAG, since the Na<sup>+</sup> ions are strongly attracted by the highly negatively charged GAG. This ion redistribution promotes the formation of salt bridges between E22/D23 and K28, leading to  $\beta$ -hairpin formation. Intra-peptide contact maps confirmed the presence of these stable intra-peptide interactions in the A $\beta$ -GAG system (see Figure 4.4 D), in contrast to the transient contacts observed in the A $\beta$ -only system.

In summary, the GAG indirectly affects A $\beta$  by attracting Na<sup>+</sup> ions, which enhances electrostatic interactions within the peptide and drives the formation of stable  $\beta$ -hairpins. No significant direct GAG-A $\beta$  interactions or effects on water dynamics were observed to explain the conformational changes in A $\beta$ .

### 4.3 Discussion

We constructed TNs to explore the conformational space and identify differences of the amyloid  $\beta$ -peptide A $\beta_{42}$  (referred to as A $\beta$ ) under three different conditions: A $\beta$  as a single peptide in solution, A $\beta$  in the presence of the GAG chondroitin-4-sulfate with 16 subunits, and A $\beta$  in complex with three POPC lipids. To define the states for each TN, we used the same three descriptors: i) the number of residues with an  $\alpha$ -helical structure, ii) the number of residues with a  $\beta$ -sheet structure, and iii) the NC-distance of the peptide. This consistent descriptor selection allows for direct comparison of the TNs and highlights changes in the underlying free energy surfaces across the different systems.

The TN of A $\beta$  in solution displayed characteristics typical of an IDP, with many states, most of which were lowly populated and connected by numerous transitions. The most populated states corresponded primarily to random-coil conformations with varying degrees of spatial expansion. On the contrary, the TN for A $\beta$  in the presence of the GAG revealed a significant shift towards more compact and  $\beta$ -sheet-



**Fig. 4.4.:** Contact analysis of Aβ with a GAG molecule. Radial distributions  $g(r)$  of  $\text{Na}^+$  (red) and  $\text{Cl}^-$  (black) relative to the carboxyl groups of E22 and D23 of Aβ in (A) the Aβ-only system and (B) the Aβ-GAG system. (C) Intermolecular contacts between Aβ<sub>42</sub> and the GAG molecule, and (D) intramolecular contacts between Aβ<sub>42</sub> residues in the Aβ-GAG system. Interaction partners were considered in contact if they were closer than 10 Å. The number of contacts was normalized by the total number of time frames, yielding contact probabilities.

rich structures, highlighting the formation of a stable  $\beta$ -hairpin driven by changes in sodium ion distribution rather than direct peptide-GAG interactions. Finally, the TN of A $\beta$  with POPC lipids demonstrated a transition towards both  $\alpha$ -helical and  $\beta$ -sheet structures, with compact configurations facilitated by hydrophobic interactions between the peptide and the lipids. This interaction led to the formation of a free energy landscape with multiple competing minima. Here, the TN can be interpreted as a qualitative depiction of the underlying free energy landscape, as both represent the accessible conformational space. The detailed free energy landscapes for all systems will be discussed in chapter 5.

In conclusion, these analyses demonstrate how A $\beta$  undergoes a disorder-to-order transition when interacting with either GAGs or lipids, albeit through distinct modes of interaction. Our findings contribute to the growing body of knowledge on the role of GAGs and lipids in A $\beta$  aggregation and the development of Alzheimer's disease. In humans, two A $\beta$  fibril polymorphs, 'type I' and 'type II' filaments, are associated with sporadic and familial Alzheimer's disease, respectively [152]. Recent cryo-EM studies in transgenic mouse models reveal a wide variety of fibril polymorphs across different models, with additional unidentified electron densities on the fibril surfaces [40]. Our analyses suggest that these polymorphic variations may be driven by specific interaction partners influencing A $\beta$  fibril formation, leading to distinct structural outcomes.

# Free Energy Surface and Timescale Analysis of the A $\beta$ <sub>42</sub> peptide

In the previous chapter, we discussed how A $\beta$ <sub>42</sub> undergoes a disorder-to-order transition upon interacting with suitable interaction partners. Utilizing TNs, this analysis primarily focused on the shift in state populations, from disordered states in solution to more secondary structure-forming states upon interaction with GAGs or lipids. In this study, we aim to further explore this conformational change by constructing the free energy surface (FES) of A $\beta$ <sub>42</sub>.

The FES of intrinsically disordered proteins (IDPs) is typically flatter than that of folded proteins, characterized by complex topographies with multiple competing minima [5, 153, 154]. For amyloid aggregation, the FES is even more heterogeneous, reflecting diverse on- and off-pathway states [155, 156]. In a previous study using NMR-based metadynamics simulations of A $\beta$ <sub>40</sub>, an ‘inverted landscape’ was observed, where disordered structures occupy the global minimum, and higher-energy regions contain transiently structured conformations, with increased structuring at higher temperatures [157]. This stands in contrast to the traditional view of a rugged energy landscape for IDPs, where numerous local minima are separated by energy barriers that trap the system in metastable states. Here, we observe a similar FES for A $\beta$ <sub>42</sub>, where disordered states are the lowest in energy and conformations with partial secondary structure are excited states (higher free energy minima). We propose the term ‘structurally inverted funnel,’ as it is the structural aspect that is inverted, not the funnel shape. In contrast, the lowest minima for the dimer feature more ordered structures, predominantly  $\beta$ -hairpins, which form as A $\beta$ <sub>42</sub> folds upon interacting with the hydrophobic regions of another A $\beta$ <sub>42</sub> peptide.

While the original work only included a comparison of the FES of A $\beta$ <sub>42</sub> in solution and in a dimeric state, we now extend this analysis by incorporating the FES of A $\beta$ <sub>42</sub> interacting with both GAGs and three POPC lipids. This approach extends our analysis from a qualitative investigation of the configurational space using TNs in chapter 4 to a quantitative assessment of the underlying FES.

This chapter provides an overview of the results published in:

**The energy landscape of A $\beta$ <sub>42</sub>: a funnel to disorder for the monomer becomes a folding funnel for self-assembly.**

*Moritz Schäffler, David Wales and Birgit Strodel*

Chem. Commun., **60** (2024)

The full publication is provided in the appendix A.1. The contributions of the authors are listed in Table 5.1.

Author	Contribution	Tasks
<u>Moritz Schäffler</u>	~80%	Methodology, Software, Investigation Data-curation, Validation, Formal-analysis, Visualization, Writing-original-draft, Writing-review-editing, Funding-acquisition
David Wales	~10%	Methodology, Software, Investigation, Writing-review-editing, Supervision
Birgit Strodel	~10%	Conceptualization, Methodology, Resources, Writing-review-editing, Supervision, Project-administration

**Tab. 5.1.:** Contribution by author to Schäffler et al. 2024 [135].



## 5.1 Methods

The methods utilized in this study are outlined in Chapter 2.4 to 2.7. Here, we briefly summarize the key concepts; for more detailed information, please refer to the full publication in the Appendix A.1.

In contrast to previous studies, where states were determined using predefined descriptor functions, we employed structural clustering for state determination. Due to the high number of degrees of freedom in Cartesian coordinates ( $f = 3N \sim 10^3$ ) and the large number of frames in the trajectory ( $\sim 10^5$ ), direct structural clustering is not feasible. Therefore, dimensionality reduction was applied. We utilized the distribution of reciprocal interatomic distance (DRID) metric, which effectively preserves system kinetics during dimensionality reduction. A set of reference atoms was selected, from which the reciprocal interatomic distances to all other atoms in the molecule were calculated. The first three moments ( $\mu_i, \nu_i, \xi_i$ ) of these distributions were then used for state descriptions. In this study, six key residues (D1, F19, D23, K28, L34, A42) were chosen as reference atoms, resulting in an DRID space with reduced dimensionality of 18. The state vectors can be written as:

$$S(t) = (\mu_1, \nu_1, \xi_1, \dots, \mu_6, \nu_6, \xi_6)$$

Clustering these states enabled calculation of the transition rate matrix, from which free energy minima and transition states were derived (see chapter 2.5). The resulting FES was visualized using disconnectivity graphs, revealing the structural landscape and folding funnels (see chapter 2.6). First-passage time (FPT) analysis provided insights into the kinetics of conformational transitions between disordered and  $\beta$ -sheet-rich states (see chapter 2.7).

In all four simulations of A $\beta_{42}$ , in solution, as a dimer, with a GAG molecule and with three POPC lipids, the peptide was modeled with neutral histidine residues and no terminal capping, resulting in an overall peptide charge of  $q = -3$ . The simulations were performed using the GROMACS simulation package [143] with the CHARMM36m force field [158] to model the A $\beta_{42}$  peptide. The simulation details of A $\beta_{42}$  in solution, A $\beta_{42}$ -GAG and A $\beta_{42}$ -POPC are presented in chapter 4.1, however all simulations were extended to an accumulated simulation time of 6  $\mu$ s for each system. For the dimer simulations, three runs were initiated, each starting from two extended A $\beta_{42}$  monomers with an initial separation of at least 2 nm. The peptides were positioned within a simulation box with a minimum distance of 1.2 nm from any box face or edge, then solvated with TIP3P water molecules [97] and NaCl at a physiological concentration of 150 mM. After equilibration, simulations were conducted under NPT conditions at 1 bar using the Parrinello–Rahman pressure coupling scheme [108]. The dimer simulations were carried out at 298 K using a Nosé–Hoover thermostat [105, 106]. Periodic boundary conditions were applied in all dimensions, and electrostatic interactions were calculated with the particle-mesh

Ewald method [145]. Real-space cutoffs for van der Waals and Coulomb interactions were set to 1.2 nm. To minimize computational demands, the dimer simulation box size was reduced after 2  $\mu$ s, once dimer formation had been observed. Following brief NPT equilibration, each dimer system was further simulated for an additional 6  $\mu$ s, resulting in a total of  $3 \times 8 \mu$ s for the A $\beta_{42}$  dimer

## 5.2 Results

In the following we shortly present the FESs of the A $\beta_{42}$  monomer and dimer as discussed in the publication [135]. Furthermore, we include the FESs of A $\beta_{42}$  with POPC lipids and a GAG molecule, which has not been published to the present date.

### 5.2.1 A $\beta_{42}$ Monomer

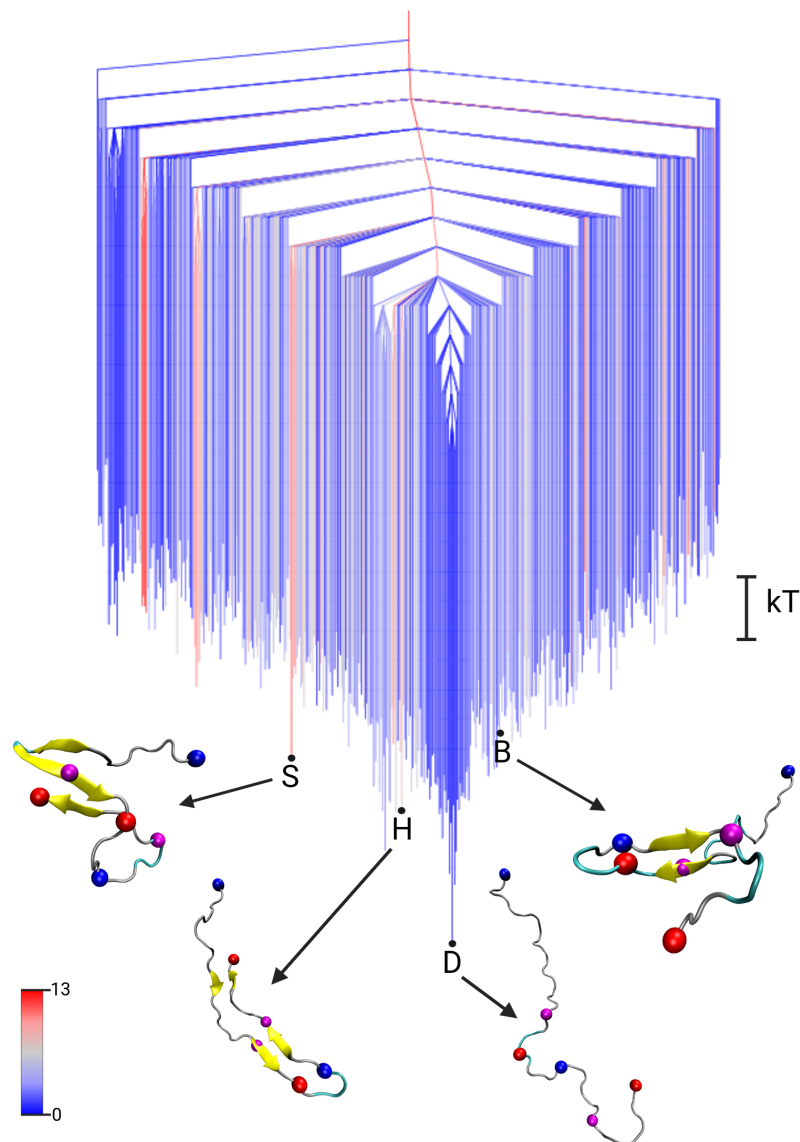
Figure 5.1 shows the FES of the A $\beta_{42}$  monomer, visualized as a disconnectivity graph. The FES displays a primary funnel leading to the global minimum. Unlike the FES of folded proteins, the global minimum here corresponds to disordered states of the A $\beta_{42}$  monomer, denoted as state D (disordered). Conformations with partial secondary structure, such as the  $\beta$ -hairpin, characteristic of A $\beta$  oligomers [36], or the S-shape conformation, a common building block of fibrils [159], appear as excited states in the FES with free energy differences of  $\Delta F_H^{\text{mon}} = 2.3$  kT and  $\Delta F_S^{\text{mon}} = 3.2$  kT, respectively.

This arrangement of the FES, where (partially) folded states occupy higher positions in the funnel, with disordered states at the bottom, has been previously referred to as an "inverted free energy landscape" [157]. We proposed the term "structurally inverted funnel" or simply "disordered funnel," as it is the structure, not the funnel, that is inverted.

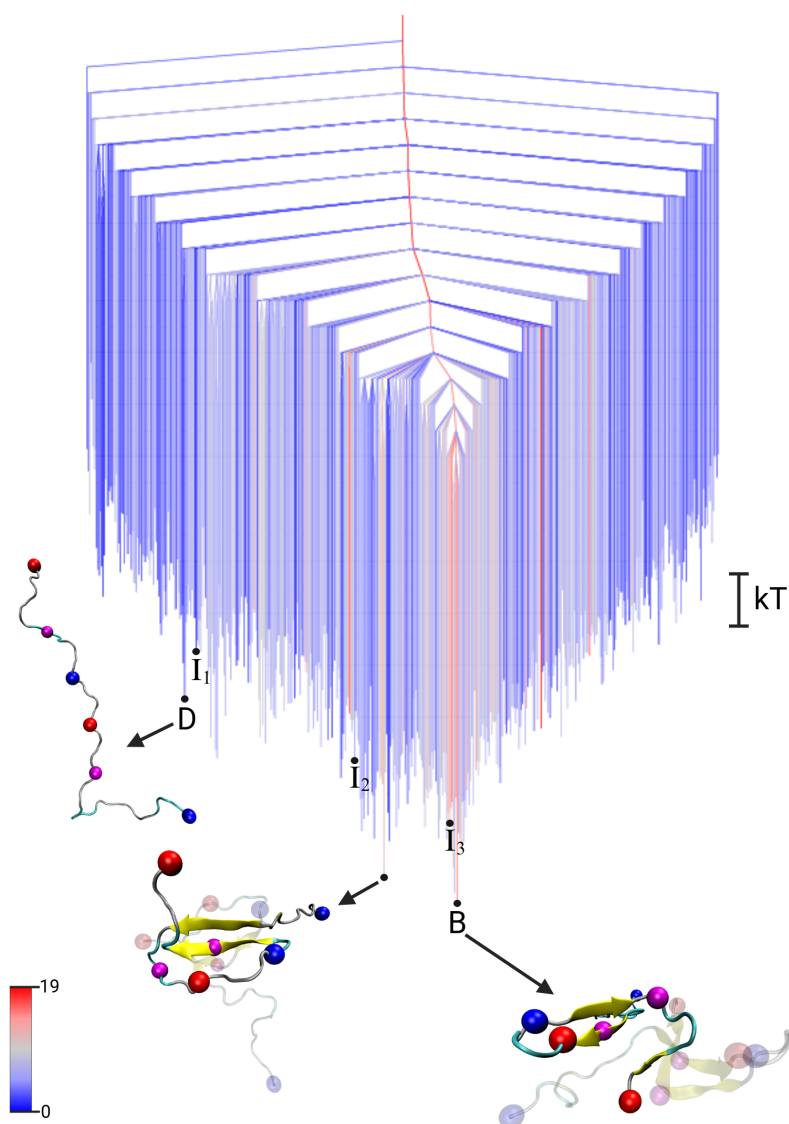
### 5.2.2 A $\beta_{42}$ Dimer

To compare the FES of the A $\beta_{42}$  dimer with that of the monomer, we employed the same DRID metric to calculate the states, treating the two chains in the dimer system as individual peptides. Thus, the dimer disconnectivity graph represents the FES of an A $\beta_{42}$  peptide in the presence of another A $\beta_{42}$  peptide. The resulting FES for the A $\beta_{42}$  dimer, shown in Figure 5.2, features the topology of a folding funnel leading to a  $\beta$ -hairpin state with substantial  $\beta$ -sheet content (labeled state B).

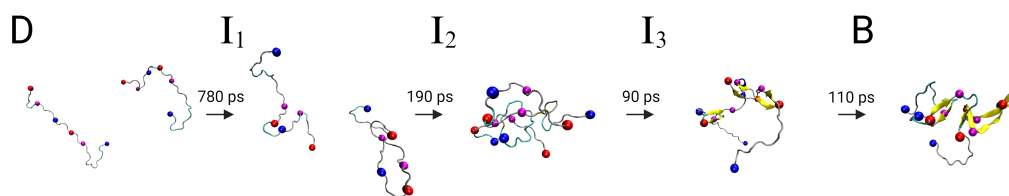
Projecting the global minimum of the monomer FES (state D) onto the dimer FES reveals that the disordered state appears as an excited state in the dimer FES, with a free energy difference of  $\Delta F_D^{\text{dim}} = 3.9$  kT. Similarly, the global minimum of the dimer FES (state B) projected onto the monomer FES (Figure 5.1) appears as an excited state there, with a free energy difference of  $\Delta F_B^{\text{mon}} = 3.6$  kT. It is important to note that state D in the dimer FES does not correspond to a dimer but rather to two disordered monomers, whereas the  $\beta$ -hairpin in state B of the dimer arises from interpeptide interactions resulting from dimerization. This finding demonstrates that the A $\beta_{42}$  peptide can stabilize into a  $\beta$ -hairpin by interacting with the hydrophobic



**Fig. 5.1.:** Free energy disconnectivity graph for the FES of the  $A\beta_{42}$  monomer. The energies are given in units of  $kT$  (see scale bar on the right), with  $k$  the Boltzmann constant and  $T$  the absolute temperature. The branches are colored according to the average number of residues in  $\beta$ -sheet conformation in the ensemble of structures belonging to the corresponding minimum, ranging from blue (no  $\beta$ -sheets) to red (13 residues involved in  $\beta$ -sheets). Representative structures of some minima are shown, where D (for ‘disordered’) is the global minimum of the monomer FES and B (for ‘ $\beta$ -sheet’) is the global minimum of the dimer FES projected onto the monomer FES. The structures are shown in the cartoon representation, with  $\beta$ -sheets highlighted in yellow and the centroids used in the DRID metric shown as spheres (blue for positive charge at the N-terminus and K28 side chain, red for negative charge at the C-terminus and D23, magenta for the hydrophobic F19 and L34).



**Fig. 5.2.: Free energy disconnectivity graph for the FES of the  $A\beta_{42}$  dimer.** The energies are given in units of  $kT$  (see scale bar on the right), with  $k$  the Boltzmann constant, and  $T$  the absolute temperature. The branches are colored according to the average number of residues in  $\beta$ -sheet conformation in the ensemble of structures belonging to the respective minimum, ranging from blue (no  $\beta$ -sheets) to red (19 residues involved in  $\beta$ -sheets). Representative structures of some minima are shown, where B (for ‘ $\beta$ -sheet’) is the global minimum of the dimer FES and D (for ‘disordered’) is the global minimum of the monomer FES projected onto the dimer FES. The structures are shown in the cartoon representation, with  $\beta$ -sheets highlighted in yellow and the centroids used in the DRID metric shown as spheres (blue for positive charge at the N-terminus and K28 side chain, red for negative charge at the C-terminus and D23, magenta for the hydrophobic F19 and L34). The second peptide of the dimer is only shown in faint color, and state D is in fact a monomeric state and thus shown as such.



**Fig. 5.3.: Fastest pathway from the  $A\beta_{42}$  dimer disordered state to the global minimum.** The states are labeled as presented in the disconnectivity graph of the  $A\beta_{42}$  dimer in Figure 5.2. The spheres represent the centroids used in the DRID metric. The first passage times associated with the fastest transition between the states are shown above the arrows, which were calculated separately for each transition and therefore do not add up to the total transition time associated with the full pathway.

region of another  $A\beta_{42}$  peptide, emphasizing the role of hydrophobic partners in facilitating folding during self-assembly.

To elucidate the structural transitions from the disordered state D to the folded  $\beta$ -hairpin state B in the  $A\beta_{42}$  dimer, we extracted the fastest pathways from the free energy surface using the PATHSAMPLE program [160]. The transition pathway, shown in Figure 5.3, includes three intermediate states ( $I_1$ ,  $I_2$ ,  $I_3$ ) marked on the dimer FES in Figure 5.2.

The  $D \rightarrow I_1$  transition is characterized by a scissor motion, driven by the formation of a D23–K28 salt bridge, followed by hydrophobic interactions between  $^{17}\text{LVFFA}^{21}$  and  $^{30}\text{AIIGLMV}^{36}$ . This process occurs cooperatively with the formation of inter-peptide contacts between the hydrophobic regions of both peptides. In the  $I_1 \rightarrow I_2$  transition, the key hydrophobic contact shifts from F19–L34 to F19–I32, facilitating the formation of a short antiparallel  $\beta$ -sheet in  $I_3$ . The final  $I_3 \rightarrow B$  transition further stabilizes the  $\beta$ -sheet, as  $I_3$  is already located within the folding funnel towards the global minimum.

This study confirms the significance of the D23–K28 salt bridge in initiating the turn and subsequent  $\beta$ -hairpin formation, which is essential for  $A\beta$  oligomerization and toxicity [36, 161]. Our findings align with recent NMR studies, which similarly identified this salt bridge [162, 163] and hydrophobic interactions [164–166] as crucial in driving amyloid formation and toxicity in both  $A\beta_{40}$  and  $A\beta_{42}$ . Specifically, Tycko and coworkers observed  $\beta$ -hairpin conformations with intra-peptide contacts between the main hydrophobic segments in  $A\beta_{40}$  forming within 1 ms. The primary hydrophobic interaction was found between the aromatic side chain of F19 and aliphatic residues 30–35, consistent with structures in the dimer FES folding funnel determined here [167].

### 5.2.3 $A\beta_{42}$ with a GAG

In the previous chapter 4, we compared the configurational space of the  $A\beta_{42}$  monomer with that of  $A\beta_{42}$  in complex with GAG and POPC lipids using TNs. Here, we extend this discussion by providing the FESs of the  $A\beta_{42}$ -POPC and  $A\beta_{42}$ -GAG

systems in terms of disconnectivity graphs and compare them to the  $A\beta_{42}$  monomer and dimer FESs.

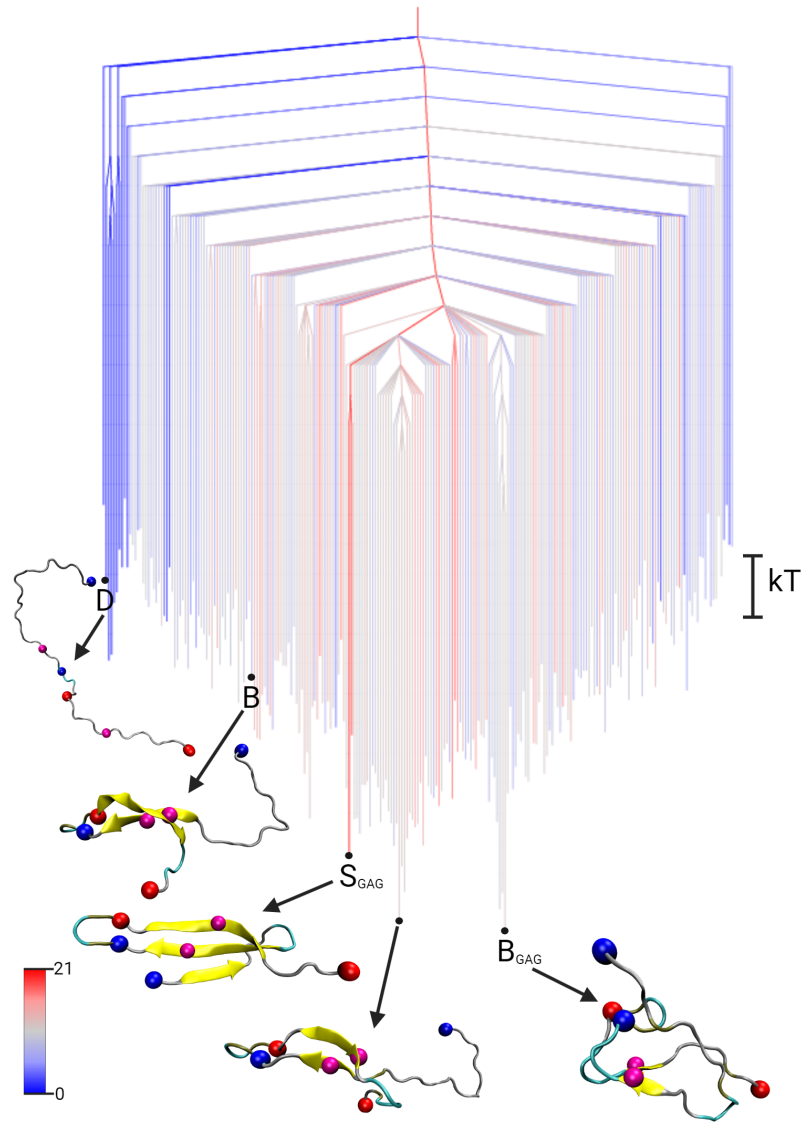
Figure 5.4 shows the FES of the  $A\beta_{42}$  monomer in the presence of a GAG. The same DRID metric as for the monomer and dimer system was employed to calculate the states, treating interactions between  $A\beta_{42}$  and the GAG only implicitly. Moreover, our previous study has revealed that due to minimal contacts between  $A\beta_{42}$  and the GAG molecule, there is no direct cooperative folding mechanism [147]. The FES features a single main funnel with two competing  $\beta$ -hairpin structures at the bottom. The global minimum  $B_{GAG}$  consists of more compact  $\beta$ -hairpin structures, where the D23–K28 salt bridge folds inward, interacting with the N-terminal and hydrophobic regions. The competing  $\beta$ -hairpin structure is only slightly higher in energy  $\Delta F^{GAG} = 0.2$  kT and resembles a more extended structure, similar to the global minimum of the dimer FES.

Projecting the global minimum of the dimer FES (state B) onto the  $A\beta_{42}$ -GAG FES reveals that, despite the structural similarity, it appears as an excited state with a free energy difference of  $\Delta F_B^{GAG} = 4.2$  kT. Similarly, the disordered state from the monomer FES (state D) appears as an excited state in the  $A\beta_{42}$ -GAG FES, with  $\Delta F_D^{GAG} = 5.9$  kT. Notably, this disordered state belongs to a side funnel of intrinsically disordered states, where the local minimum has  $\Delta F_{idp}^{GAG} = 4.5$  kT. As observed in the dimer FES, the system must overcome an initial energy barrier to escape the basin of disordered states and enter the main folding funnel.

The state with the highest  $\beta$ -sheet content can be identified as an S-shaped structure (state  $S_{GAG}$ ), with  $\Delta F_S^{GAG} = 1.3$  kT. Given that the S-shaped conformation has been identified as a common building block of fibrils [159], and the relatively low free energy in the  $A\beta_{42}$ -GAG FES, this suggests that GAGs may accelerate amyloid fibril formation.

Analysis of the  $A\beta_{42}$ -GAG FES reveals significant similarity between the global minimum of the dimer FES and the low-energy states in the presence of a GAG. However, the state with the highest overlap with the dimer FES global minimum appears more twisted, as a result of the interactions with the other  $A\beta_{42}$  molecule composing the dimer (see Figure 5.2). The presence of GAGs may promote a more ordered interaction pattern, warranting further investigation in future studies.

Compared to the TN analysis, the current FES yields similar conclusions. However, the identification of structures associated with the global minimum is significantly more precise, allowing us to distinguish two distinct  $\beta$ -hairpin structures that may have been grouped into the same state in the corresponding TN. Additionally, the dynamics from the initial disordered state to the folded  $\beta$ -hairpin state are more clearly elucidated, and the comparison of free energies provides insights into the relative propensities of these states, which could be tested experimentally. Biophysical studies have shown that GAGs promote the formation of amyloid fibrils [146], which in the light of our analysis identifies the  $\beta$ -hairpin structures as key structural motive in the process.



**Fig. 5.4.:** Free energy disconnectivity graph for the FES of  $A\beta_{42}$  with a GAG. The energies are given in units of  $kT$  (see scale bar on the right), with  $k$  the Boltzmann constant, and  $T$  the absolute temperature. The branches are colored according to the average number of residues in  $\beta$ -sheet conformation in the ensemble of structures belonging to the respective minimum, ranging from blue (no  $\beta$ -sheets) to red (21 residues involved in  $\beta$ -sheets). Representative structures of some minima are shown, where B (for ‘ $\beta$ -sheet’) is the global minimum of the dimer FES and D (for ‘disordered’) is the global minimum of the monomer FES projected onto the  $A\beta_{42}$ -GAG FES. Furthermore, the global minimum  $B_{GAG}$  and state with the highest  $\beta$ -sheet content  $S_{GAG}$  are highlighted. The structures are shown in the cartoon representation, with  $\beta$ -sheets highlighted in yellow and the centroids used in the DRID metric shown as spheres (blue for positive charge at the N-terminus and K28 side chain, red for negative charge at the C-terminus and D23, magenta for the hydrophobic F19 and L34). The GAG molecule is not depicted as there are only rarely contacts with the  $A\beta_{42}$  peptide.

## 5.2.4 $A\beta_{42}$ with POPC lipids

The conformational space of  $A\beta_{42}$  in complex with POPC lipids, investigated through MD simulations, has been extensively analyzed in our previous studies [42, 131, 147]. Here, we extend this analysis by constructing the FES similarly to previous systems. For consistency, we employed the same DRID metric as with other  $A\beta_{42}$  systems to calculate the states, treating  $A\beta_{42}$ -POPC interactions implicitly.

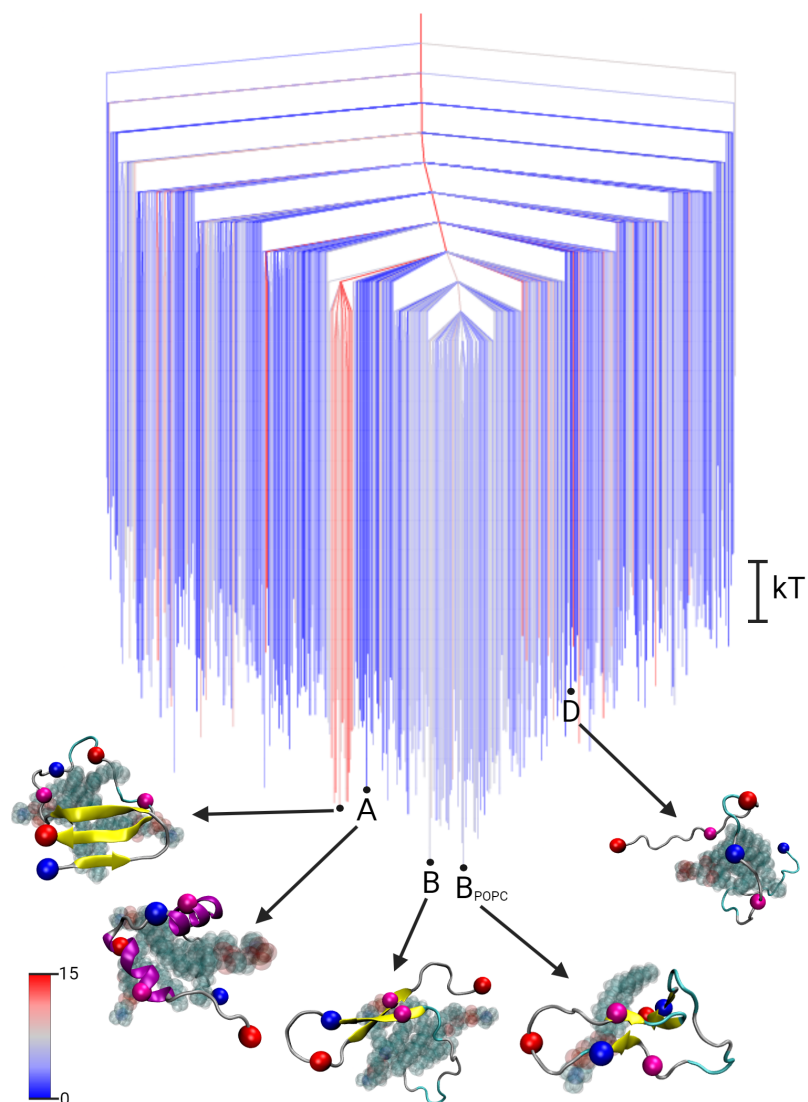
As discussed in chapter 4, the mode of interaction of  $A\beta_{42}$  with POPC lipids differs significantly from its interaction with GAGs. While almost no direct contacts are observed between  $A\beta_{42}$  and GAGs,  $A\beta_{42}$  forms a compact complex with the POPC cluster. In a previous study [42], Fatafta et al. demonstrated that  $A\beta_{42}$  undergoes a disorder-to-order transition upon forming a complex with POPC lipids. At a 1:3 ratio,  $A\beta_{42}$  exhibited either a random coil-to-helix or random coil-to- $\beta$ -sheet transition. The simulations revealed the formation of a stable helix-kink-helix structure in one run, while  $\beta$ -sheet formation dominated in the other runs. The increase in  $\alpha$ -helix content was correlated with specific peptide-lipid interactions, particularly with residues L17, A21, I32, and V36. Contacts between these residues and the lipid tails stabilized the helix, while  $\beta$ -sheet structures emerged in regions with different peptide-lipid interactions, indicating that structure formation in  $A\beta_{42}$  is driven by its interaction with the surrounding lipids.

Figure 5.5 shows the FES of the  $A\beta_{42}$  monomer in the presence of three POPC lipids, visualized as a disconnectivity graph. The FES reveals a singular funnel with a  $\beta$ -sheet forming state at the global minimum ( $B_{\text{POPC}}$ ), distinct from the typical  $\beta$ -hairpin structure, discussed above. In addition, the FES appears shallower compared to the previously discussed FESs, as indicated by the relatively small free energy differences between key states.

Projecting the disordered state from the monomer FES (state D) onto the  $A\beta_{42}$ -POPC FES, shows that it appears as an excited state in the  $A\beta_{42}$ -POPC FES, with  $\Delta F_D^{\text{POPC}} = 3.0$  kT, which is  $\sim 1$  kT lower compared to the dimer and GAG systems, indicating a flatter FES of  $A\beta_{42}$  in the presence of POPC. Note that the disordered state D in the POPC FES shows the highest overlap with the ensemble of structures belonging to state D in the monomer FES, however here  $A\beta_{42}$  forms a complex with the POPC lipids. Projecting the global minimum of the dimer FES (state B) onto the  $A\beta_{42}$ -POPC FES reveals that the  $\beta$ -hairpin state is positioned very close to the global minimum in the folding funnel, with a free energy difference of only  $\Delta F_B^{\text{POPC}} = 0.1$  kT. While previous studies indicated a high  $\beta$ -sheet propensity for  $A\beta_{42}$  in the presence of POPC lipids, the  $\beta$ -hairpin state had not been unambiguously identified as the predominant structure. The low free energy further suggests that this state is both likely to form and highly stable.

The state with the highest  $\alpha$ -helical content can be identified as the helix-kink-helix structure (state  $S_A$ ), which has been identified by Fatafta et al. [42], with  $\Delta F_A^{\text{POPC}} = 1.0$  kT. Given that  $A\beta_{42}$  has shown very low propensity to form  $\alpha$ -helices





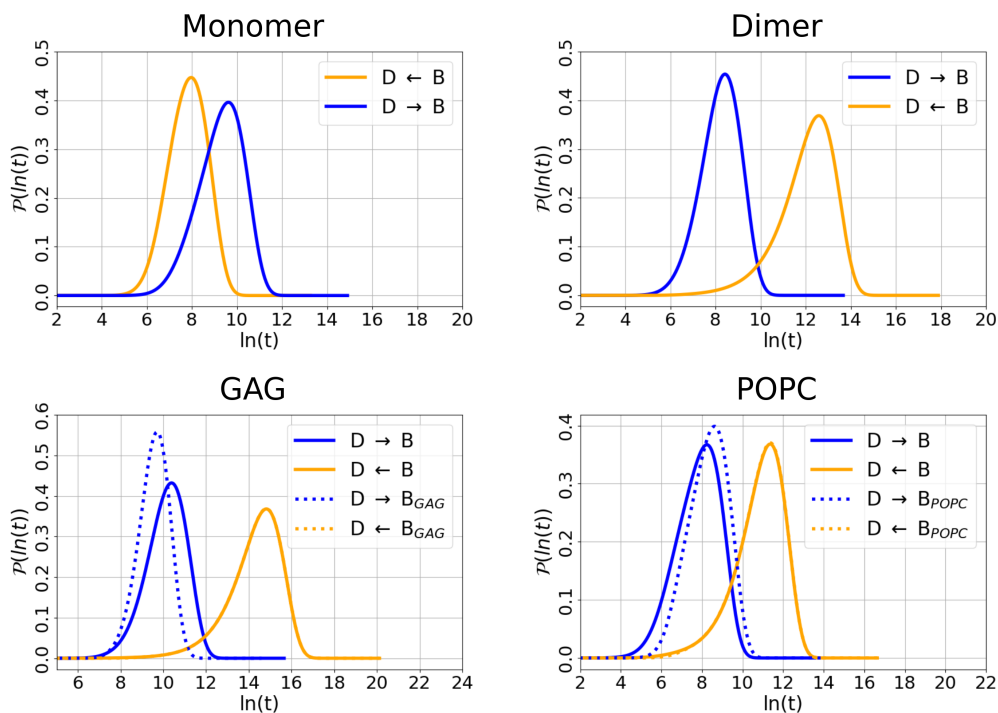
**Fig. 5.5.:** Free energy disconnectivity graph for the FES of  $A\beta_{42}$  with three POPC lipids. The energies are given in units of  $kT$  (see scale bar on the right), with  $k$  the Boltzmann constant, and  $T$  the absolute temperature. The branches are colored according to the average number of residues in  $\beta$ -sheet conformation in the ensemble of structures belonging to the respective minimum, ranging from blue (no  $\beta$ -sheets) to red (15 residues involved in  $\beta$ -sheets). Representative structures of some minima are shown, where B (for ' $\beta$ -sheet') is the global minimum of the dimer FES and D (for 'disordered') is the global minimum of the monomer FES projected onto the  $A\beta_{42}$ -POPC FES. The structures are shown in the cartoon representation, with  $\beta$ -sheets highlighted in yellow and the centroids used in the DRID metric shown as spheres (blue for positive charge at the N-terminus and K28 side chain, red for negative charge at the C-terminus and D23, magenta for the hydrophobic F19 and L34). The POPC lipids are shown as translucent spheres.

in the other systems, the low free energy of state A shows the distinct change on the conformational space of  $A\beta_{42}$  by POPC lipids, as well as the flatness of the FES.

### 5.2.5 Timescale Analysis

To analyze the timescales associated with the interconversions between the disordered state D and the  $\beta$ -hairpin state B, as previously determined from the monomer and dimer FES respectively [135], we calculated the FPT distribution for these transitions in the presence of GAG and POPC. The resulting FPTs for all four systems, are shown in Figure 5.6. For the  $A\beta_{42}$ -GAG systems and  $A\beta_{42}$ -POPC, the interconversion timescales between state D and B, as well as between state D and the respective global minimum are presented.

Focusing on the comparison between monomer and dimer, the transition times for the monomer are  $\tau_{D \rightarrow B}^{\text{mon}} \sim 15$  ns and  $\tau_{B \rightarrow D}^{\text{mon}} \sim 3$  ns. While the transition from the disordered state to the excited  $\beta$ -hairpin state is five times longer than the reverse transition, it remains relatively fast. In contrast, for the dimer, the transition times are  $\tau_{D \rightarrow B}^{\text{dim}} \sim 4$  ns and  $\tau_{B \rightarrow D}^{\text{dim}} \sim 300$  ns, with a ratio of about 75 for the transition from the global minimum to the excited disordered state, relative to the reverse transition. This reveals a clear separation of timescales in the dimer, which is absent in the



**Fig. 5.6.: First passage time probability distributions for interconversions between disordered and  $\beta$ -hairpin states.** The probability distribution  $\mathcal{P}(\ln t)$  of the first passage time  $t$  for transitions between the disordered state D and the  $\beta$ -hairpin state B is shown on a logarithmic scale. The states D and B were determined from the monomer and dimer FES, respectively, and projected onto the FES of the other systems to identify the state with the highest overlap. For the  $A\beta_{42}$ -GAG and  $A\beta_{42}$ -POPC systems, the FPTs between state D and the global minimum are also presented.

monomer, indicating a much flatter FES for the monomer compared to the dimer, despite both FESs exhibiting a singular funnel structure.

Similar to the dimer, in the case of  $A\beta_{42}$ -GAG, an even more pronounced separation of timescales is observed between the interconversion times of states D and B. The transition times are  $\tau_{D \rightarrow B}^{\text{GAG}} \sim 32$  ns and  $\tau_{B \rightarrow D}^{\text{GAG}} \sim 2800$  ns. The timescale for the reverse transition,  $\tau_{B \rightarrow D}^{\text{GAG}}$ , is so high that practically no transition from B to D is observed within common MD timescales, highlighting the strong disorder-to-order transition of  $A\beta_{42}$  in the presence of a GAG. Furthermore, the timescale associated with the transition to the global minimum, which is an alternative hairpin state, is  $\tau_{D \rightarrow B_{\text{GAG}}}^{\text{GAG}} \sim 17$  ns, making this transition twice as fast as the transition to state B. The back transition from the global minimum to state D ( $\tau_{B_{\text{GAG}} \rightarrow D}^{\text{GAG}} \sim 2800$  ns) is as slow as the  $B \rightarrow D$  transition, indicating that the disordered state D is positioned very high in the  $A\beta_{42}$ -GAG FES that returning to the disordered basin is highly improbable under the simulated conditions.

The transition times between states D and B for the  $A\beta_{42}$ -POPC system are  $\tau_{D \rightarrow B}^{\text{POPC}} \sim 4$  ns and  $\tau_{B \rightarrow D}^{\text{POPC}} \sim 86$  ns. This indicates that the transition from the disordered to the hairpin state is as fast as for the dimer, while the reverse transition is approximately four times faster than  $\tau_{B \rightarrow D}^{\text{dim}}$ , suggesting an overall flatter FES with a less steep slope toward the hairpin state. The transition times from the disordered state to the global minimum  $B_{\text{POPC}}$  are  $\tau_{D \rightarrow B_{\text{POPC}}}^{\text{POPC}} \sim 6$  ns and  $\tau_{B_{\text{POPC}} \rightarrow D}^{\text{POPC}} \sim 86$  ns. While the transition to the global minimum is slightly slower compared to  $\tau_{D \rightarrow B}^{\text{POPC}}$ , the reverse transition times are identical. This suggests that state B may serve as an intermediate on the pathway  $D \rightarrow B_{\text{POPC}}$ , with both states residing in the same basin at the bottom of the FES.

### 5.3 Discussion

Through the analysis of free energy landscapes (FES) and first passage time (FPT) probability distributions, we have identified key mechanisms underlying  $A\beta_{42}$ 's transition from a disordered state to a  $\beta$ -sheet-forming, hairpin-like structure. Our results reveal distinctive features in the free energy landscapes of  $A\beta_{42}$  monomers and dimers, highlighting the impact of the molecular environment on structural stability and propensity for aggregation. Specifically, we observe that the FES of the monomer resembles a structurally inverted folding funnel, where disordered states are lowest in energy. In contrast, the lowest minima for the dimer are characterised by more ordered structures, predominantly consisting of  $\beta$ -hairpins, which form as  $A\beta_{42}$  folds upon binding to the hydrophobic region of another  $A\beta_{42}$  peptide.

Our analysis of first-passage times indicates that the FES of the monomer is relatively flat, with excited states more readily accessible compared to the dimer. For the disordered funnel of the monomer, local minima with higher energy are readily accessible and the timescales of interconversions between them are of similar magnitude, while transitions to excited states from the ordered global minimum of

the dimer are relatively slow. Thus, while we characterize the monomer FES as a funnel to disorder, the dimer FES is closer to the expected form for a folded protein. Our FES and timescale analysis of  $A\beta_{42}$  further confirms the predictions made for the energy landscapes of folded and disordered proteins based on the slope of the free energy funnels. Chong and Ham determined the FESs of two folded proteins, which exhibited steep folding funnels, and of an IDP, which has a gentler slope and led to similar transition times to and from the global minimum [154].

The pathway analysis of the dimer FES identifies key intermediates and structural rearrangements involved in the transition from disordered to folded states, providing mechanistic insights into the initial aggregation steps for  $A\beta_{42}$ . Starting from an extended conformation, the peptide chain undergoes a scissor-like motion, establishing a salt bridge between residues D23 and K28, followed by the formation of hydrophobic contacts. Subsequently, these contacts reorganise, leading to  $\beta$ -hairpin structures. Our results should contribute to a deeper understanding of the molecular mechanisms underlying Alzheimer's disease pathology, and may inform the development of targeted therapeutic interventions.

In addition to the published results [135], we included an analysis of the FES of  $A\beta_{42}$  in the presence of a GAG molecule and a lipid cluster formed by three POPC lipids. For the  $A\beta_{42}$ -GAG FES, we observe a significant similarity to the dimer FES. As with the dimer, the FES reveals a folding funnel akin to that of a folded protein, with two competing  $\beta$ -hairpin structures at the bottom of the FES. Similar to the global minimum of the dimer, projecting state D onto the  $A\beta_{42}$ -GAG FES identifies it as an excited state. A closer inspection of the structures reveals that state B, characteristic for the dimer, exhibits a more twisted conformation compared to  $B_{GAG}$ , likely due to the cooperative folding resulting from binding to an interaction partner, whereas the GAG interacts indirectly with the  $A\beta_{42}$  peptide. Timescale analysis reveals that the transition from the disordered state to the global minimum is slower compared to the dimer FES, indicating a steeper folding funnel in the dimer. Additionally, the back transition from the hairpin state to the disordered basin in the GAG system is exceedingly slow, essentially inaccessible within MD timescales, which underscores the stability of the hairpin structure in the presence of a GAG.

Our findings, which indicate a strong correlation between the presence of a GAG molecule and the  $\beta$ -hairpin propensity of Abeta, elucidate the role of GAGs in the development of Alzheimer's disease. GAGs are major components of the extracellular matrix, providing structural support to brain tissues [168]. They help maintain the integrity and organization of neural tissue by forming a gel-like matrix that surrounds neurons and glial cells. It has been observed a significant presence of GAGs in  $A\beta$  aggregates sampled from Alzheimer's disease affected patients [146] and several biophysical studies have shown that GAGs promote aggregation, nucleation, and formation of amyloid fibrils [169–171]. However, their role in Alzheimer's disease pathology and the molecular details are not yet fully understood.

The FES in the presence of three POPC lipids exhibits significant differences compared to the previously discussed FES. The overall landscape is flatter, with the disordered state D still appearing as an excited state but about  $\sim 1$  kT lower in energy relative to the GAG FES. Moreover, the structure associated with the global minimum, while displaying a sheet-turn-sheet motif, lacks the D23-K28 salt bridge that is a key feature of the  $\beta$ -hairpin structure in the dimer and in the presence of GAG. Nonetheless, this structure (state B) resides in the same energy basin at the bottom of the FES and is only slightly higher in energy than the global minimum. Additionally, the helix-kink-helix structure, identified in previous studies, appears as an excited state in the FES. Timescale analysis of the  $A\beta_{42}$ -POPC FES reveals its relative flatness, with transition timescales comparable to those observed in the monomer FES, despite the high interaction propensity between  $A\beta_{42}$  and the POPC lipids.

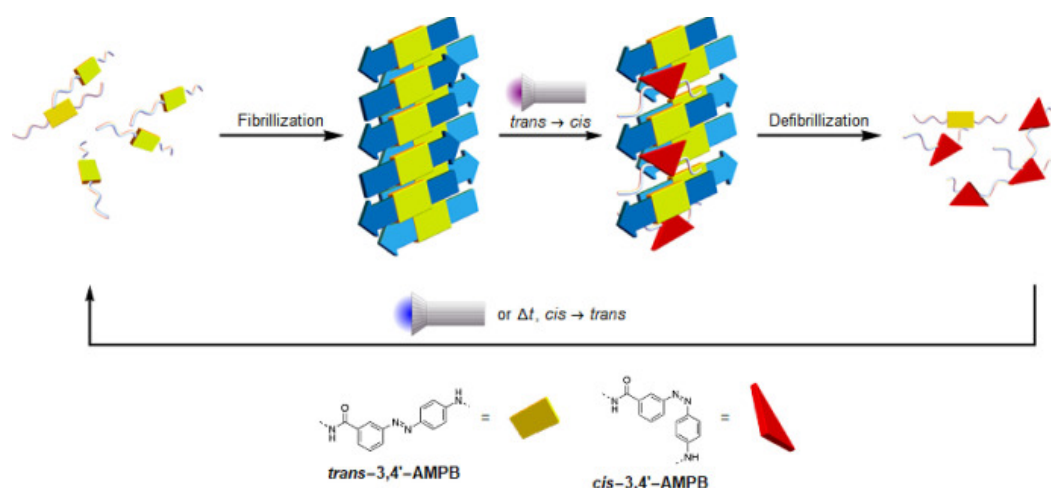
These findings highlight the disorder-to-order transition of  $A\beta_{42}$  upon lipid binding, reported in our previous studies [42, 147]. Lipids play a significant role in Alzheimer's disease pathology, particularly in the interaction of  $A\beta$  with lipid membranes, which has been linked to neuronal membrane disruption [172]. While many studies have examined  $A\beta$  solely in a lipid-membrane environment, recent work suggests that complex formation with free lipids in the extracellular space may drive membrane insertion, leading to pore formation and membrane thinning [173, 174]. Our results support this hypothesis, showing a conformational change in  $A\beta_{42}$  when complexed with POPC, similar to that in the dimer or  $A\beta_{42}$ -GAG systems, suggesting a shared mechanism underlying toxicity. Furthermore, our findings underscore the ability of  $A\beta$  to readily transition between  $\alpha$ - and  $\beta$ -structures, supporting the hypothesis that lipid-associated amyloid aggregation may proceed via a helical intermediate state [175].



# Photocontrolled Reversible Amyloid Fibril Formation of PTH

6

Peptide fibrillization plays a critical role in biological processes such as amyloid-related diseases and hormone storage, characterized by transitions between folded, unfolded, and aggregated states. The parathyroid hormone (PTH) is of particular interest due to its ability to undergo reversible amyloid fibril formation, while regulating calcium and phosphate balance in the human body [68]. The PTH<sub>1–34</sub> sequence was also the first FDA-approved drug for the treatment of severe osteoporosis [71] and its therapeutic efficacy could be enhanced through controlled transitions between its stored fibril state and its active monomeric form. In this study, we leverage light to induce reversible transitions between aggregated and non-aggregated states of the PTH<sub>25–37</sub> peptide segment, establishing a framework for targeted drug delivery using the full PTH<sub>1–34</sub> sequence. Specifically, we investigate the PTH<sub>25–37</sub> peptide, a segment central to the fibril of PTH and independently capable of fibril formation [69]. The light-responsive switch 3-{[(4-aminomethyl)phenyl]diazenyl}benzoic acid (AMPB) was incorporated into PTH<sub>25–37</sub> to control aggregation, revealing position-dependent effects (see Figure 6.1). Through a combined effort of *in silico* design, synthesis, and experimental validation, we predicted and confirmed the amyloid-forming potential of 11 novel AMPB-containing PTH<sub>25–37</sub> mutants. Molecular dynamics (MD) simulations revealed structural and interaction pattern changes based on AMPB positioning, emphasizing the importance of  $\pi$ - $\pi$  interactions be-



**Fig. 6.1.: Schematic representation of light-induced (de-)aggregation of PTH<sub>25–37</sub> with embedded AMPB photoswitch.** The artificial light-switch, 3-{[(4-aminomethyl)phenyl]diazenyl}benzoic acid (*cis/trans* - AMPB), is incorporated at various positions within the PTH peptide. *Cis/trans* isomerization regulates the reversible fibril formation of the peptide. The figure was taken from Pascholt et al. (2024) [176].

tween AMPB segments. Together with solid-state NMR studies, fibril models were developed for both PTH<sub>25–37</sub> and an AMPB-containing variant. This multifaceted approach enabled the identification of a peptide capable of phototriggered transitions between fibrillated and defibrillated states, confirmed by various spectroscopic techniques. Our strategy offers the first example of a photo-stimulus-responsive hormone, unlocking the potential to manipulate PTH's reversible switch between active and inactive aggregated states.

This chapter is meant as an overview of the published results in:

### **Photocontrolled Reversible Amyloid Fibril Formation of Parathyroid Hormone-Derived Peptides.**

*Andr  Paschold<sup>†</sup>, Moritz Sch ffler<sup>†</sup>, Xincheng Miao, Luis Gardon, Stephanie Kr ger, Henrike Heise, Merle I. S. R hr, Maria Ott, Birgit Strodel and Wolfgang H. Binder*  
Bioconjugate Chemistry, 35 7 (2024)

The full publication is provided in the appendix A.1. The contributions of the authors are listed in Table 6.1.

---

<sup>†</sup>AP and MS contributed equally.



Author	Contribution	Tasks
<u>Moritz Schäffler</u>	~25%	Methodology, Investigation (MD, fibril modelling), MD-simulations (Peptides, Fibrils, pathway for QM simulations), Forcefield Parametrization, Data-curation, Formal-analysis, Validation, Visualization, Writing-original-draft, Writing-review-editing
André Pascholt	~25%	Methodology, Investigation (ThT-Essays, Toxicity-Essays), Data-curation, Validation, Formal-analysis, Peptide Synthesis, Visualization, Writing-original-draft, Writing-review-editing
Xincheng Miao	~4%	Investigation (QM-calculations) Writing-review-editing
Luis Gardon	~4%	Investigation (NMR-measurements) Writing-review-editing
Stephanie Krüger	~4%	Investigation (Transmission-EM-imaging)
Henrike Heise	~4%	Methodology (NMR-measurements) Writing-review-editing
Merle I. S. Röhr	~4%	Methodology (QM-calculations) Writing-review-editing
Maria Ott	~4%	Investigation (WAXS-measurements)
Birgit Strodel	~13%	Conceptualization, Methodology, Resources, Writing-review-editing, Supervision, Project-administration, Funding-acquisition
Wolfgang H. Binder	~13%	Conceptualization, Methodology, Resources, Writing-review-editing, Supervision, Project-administration, Funding-acquisition

**Tab. 6.1.:** Contribution by author to Pascholt et al. 2024 [176].

## 6.1 Structural Analysis

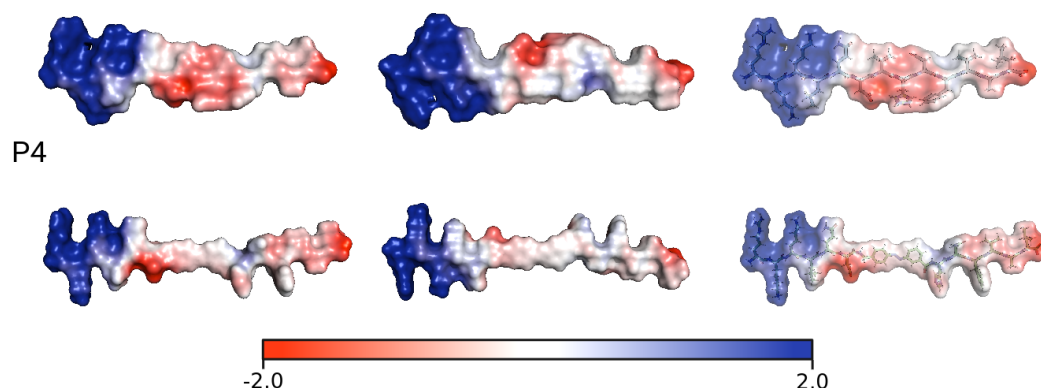
First we studied the structural characteristics of the PTH<sub>25–37</sub> peptide and one mutant with an incorporated AMBP photoswitch (called P4). The PTH<sub>25–37</sub> sequence contains three positively charged residues (RKK) at the N-terminus, followed by a combination of hydrophobic, polar, and one negatively charged residue (D30) in the central region. The C-terminal residues are predominantly hydrophobic (see Figure 1.5 and Table 6.2). This uneven distribution of physicochemical properties is reflected in the electrostatic potential surface, which shows a highly positively charged N-terminus and a more hydrophobic C-terminal half, with minor negative charge accumulation beyond the first three residues (see Figure 6.2). Therefore we hypothesized that amyloid aggregation of PTH<sub>25–37</sub> is primarily driven by residues following the initial RKK sequence.

The *trans*-P4 mutant, where V31 is substituted by AMPB, carries the photoswitch in the central region of the peptide, resulting in accelerated amyloid fibril formation within 10 hours, as opposed to 15 hours for the unmodified peptide. In contrast, the *cis*-P4 mutant initially forms amorphous aggregates that convert into amyloid fibrils after approximately 50 hours. The replacement of V31 with AMPB in the P4 mutant increases the overall hydrophobicity of the peptide, while the AMPB group introduces a slight positive charge, partially offsetting the primarily negative charge in the center (see Figure 6.2). These structural changes account for the enhanced aggregation kinetics of *trans*-P4 relative to PTH<sub>25–37</sub>.

## 6.2 Peptide Simulations

To study the effects of the insertion of an AMBP photoswitch at different positions of the PTH<sub>25–37</sub> sequence on the peptides' conformational space and aggregation

PTH<sub>25–37</sub>



**Fig. 6.2.: Electrostatic potential surface of PTH<sub>25–37</sub> and *trans*-P4.** Front and back views of the structure with a transparent representation, showing the peptides backbone structure. The color scale represents the electrostatic potential in units of  $kTe^{-1}$ , as indicated at the bottom.

propensity, we performed MD simulations of various mutants and the original sequence.

### 6.2.1 Simulation Details

MD simulations were run for the PTH<sub>25–37</sub> peptide and five selected mutants from a total of twelve designed variants, namely: P1, P3, P4, P8, and P12. These mutants incorporate the AMPB modification at various positions within the peptide chain, with variations including both full-chain insertions and deletions of existing residues. An overview of the designed mutants is provided in Table 6.2. All MD simulations were conducted using the GROMACS package. Given that PTH<sub>25–37</sub> is classified as an IDP, we employed the CHARMM36m force field. For the AMPB photoswitch, parameters were derived using the cgenff protocol [91] and refined with QM/MM data to reproduce the correct *cis*- and *trans*-state geometries [177].

AMPB position	peptide	primary sequence	modification
	PTH <sub>25–37</sub>	<sup>25</sup> RKKLQ <sup>30</sup> DVHNF <sup>35</sup> VAL	-
Central	P1	<sup>25</sup> RKKLQ <sup>30</sup> D-AMPB-VHNF <sup>35</sup> VAL	D30-AMPB-V31
	P3	<sup>25</sup> RKKLQ-AMPB-VHNF <sup>35</sup> VAL	D30→AMPB
	P4	<sup>25</sup> RKKLQ <sup>30</sup> D-AMPB-HNF <sup>35</sup> VAL	V31→AMPB
N-terminal	P8	<sup>25</sup> RK-AMPB-KLQ <sup>30</sup> DVHNF <sup>35</sup> VAL	K26-AMPB-K27
C-terminal	P12	<sup>25</sup> RKKLQ <sup>30</sup> DVHNF-AMPB-VAL	V35→AMPB

**Tab. 6.2.:** Designed PTH<sub>25–37</sub> peptides containing the AMPB photoswitch

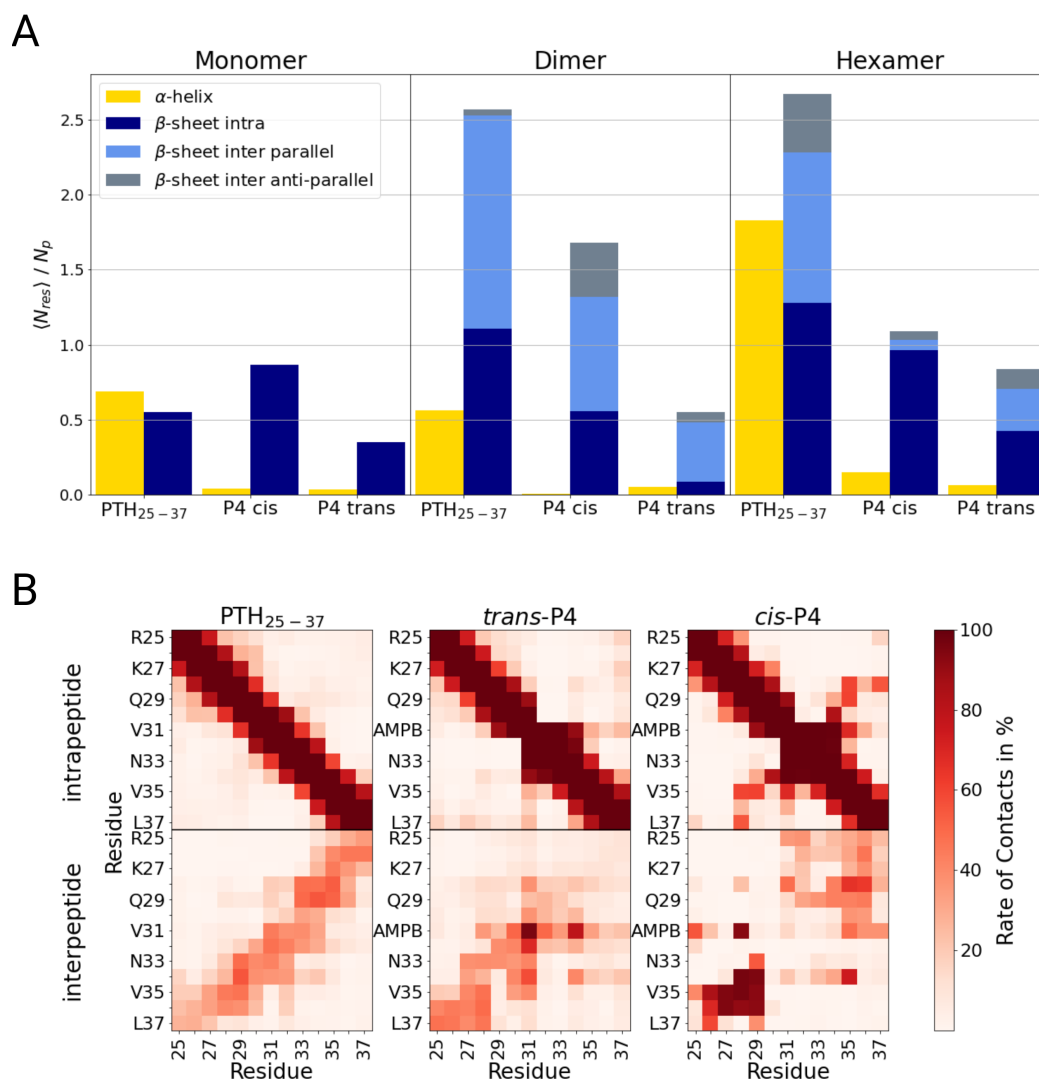
In the simulations, the  $\angle$ CNNC dihedral was fixed to either the *cis*- or *trans*-state, and transitions between the two states were not modeled. Peptides were placed in a simulation box containing TIP3P water, Na<sup>+</sup>, and Cl<sup>−</sup> ions at a physiological concentration of 150 mM. Simulations were run under NPT conditions at 300 K and 1 bar, using the Parrinello-Rahman barostat and Nosé-Hoover thermostat. A production run of 10  $\mu$ s per system was performed, with periodic boundary conditions applied in all directions and electrostatic interactions calculated using the particle-mesh Ewald method. In total, 285  $\mu$ s of simulation time were generated across all systems.

### 6.2.2 Results

To understand the aggregation mechanisms and elucidate the structures that form during oligomerization, we first focus on the MD simulations performed of the monomer, dimer and hexamer for both PTH<sub>25–37</sub> and P4, the latter simulated in the *trans*- and *cis*-states. Simulating the P4 mutant in its *cis*-state, we aim to elucidate the structural basis for its slower and reduced fibrillization. Figure 6.3 A shows the average secondary structure elements formed during the simulations.

All monomers exhibit mainly a random coil structure, with on average less than one residue forming secondary structure elements. Nevertheless, there are still subtle differences between the peptides indicating their overall structural propensities.

Notably, PTH<sub>25–37</sub> shows a small tendency to form an  $\alpha$ -helix. This helix is formed at the N-terminal by residues K27 to D30, a feature that is reduced in the P4 mutants when V31 is replaced with AMPB. Furthermore, the *cis*-isomer promotes intrapeptide interactions, resulting in  $\beta$ -hairpin formation, while the *trans*-P4 favours a fully extended conformation, thus the  $\beta$ -sheet content is highest for the *cis*-P4 structure. Dimer simulations revealed an increase in  $\beta$ -sheet content for all three peptides, likely due to aggregation. The highest  $\beta$ -sheet propensity is found in PTH<sub>25–37</sub>, predominantly forming antiparallel  $\beta$ -sheet segments. The contact maps as observed in the dimer simulations (see Figure 6.3 B) reveal interaction between the positively



**Fig. 6.3.: Comparison between PTH<sub>25–37</sub> and *trans/cis*-P4.** (A) Average simulated secondary structure population of PTH<sub>25–37</sub> and *trans/cis*-P4. The secondary structure propensity is categorized into  $\alpha$ -helical, intrapeptide  $\beta$ -sheets, and interpeptide parallel and antiparallel  $\beta$ -sheets, as indicated by the color code. (B) Contact map from dimer simulations for PTH<sub>25–37</sub>, *trans*-P4, and *cis*-P4. The rate of contacts between pairs of inter-/intramolecular residues is shown, averaged over the full simulation. Two residues are considered to be in contact if any pair of atoms is closer than 6 Å.

charged RKK N-terminal residues and the negatively charged C-terminus of the adjacent peptides, promoting an anti-parallel alignment of the peptides.

Unexpectedly, despite the faster aggregation rates, *trans*-P4 displayed fewer  $\beta$ -sheets at the oligomer level, where interactions were dominated by AMPB-AMPB contacts, favoring an anti-parallel alignment. In contrast, *cis*-P4, forming intrapeptide  $\beta$ -hairpins, mainly aggregated randomly, consistent with the amorphous structures observed experimentally for larger aggregates. These differences between *cis*- and *trans*-AMPB demonstrate molecular-level control over fibril formation through isomerization, paving the way for photoregulated amyloid formation.

The hexamer simulations for PTH<sub>25–37</sub>, *trans*-P4, and *cis*-P4 did not yield fibrillar structures, suggesting that the system size or simulation time was insufficient. Interestingly, the hexamers were less ordered than dimers, supporting the experimental findings that PTH<sub>25–37</sub> has lower amyloid-forming potential compared to other peptides, such as A $\beta$ <sub>16–22</sub> [178]. Unlike A $\beta$ <sub>16–22</sub>, which forms ordered hexamers in MD simulations with identical conditions as applied here, both PTH<sub>25–37</sub> and P4 showed a decrease in interpeptide  $\beta$ -sheet content with increased system size, likely due to the expansion of the conformational space. PTH<sub>25–37</sub> hexamers also exhibited greater helicity compared to their monomeric and dimeric forms, reflecting the peptide's inherent helical propensity, which is often stabilized by interpeptide interactions in intrinsically disordered peptides [179].

To assess the effect of AMPB at different sequence positions on peptide aggregation, we additionally performed MD simulations of the P1, P3, P8, and P12 variants (see Table 6.2), examining both *cis*- and *trans*-configurations as monomers and dimers. Figure 6.4 shows the average secondary structure elements formed during all simulations. Simulations of P1 revealed a tendency for  $\beta$ -sheet formation comparable to that of P4, due to the similar AMPB positioning in the peptide sequence. In P1, however, AMPB is inserted between D30 and V31, extending the hydrophobic C-terminal region (see Figure 1.6), potentially explaining the more rapid aggregation kinetics compared to P4. In P3, AMPB shifts one position toward the N-terminus compared to P4, replacing D30, which enhances the peptides hydrophobicity, supporting aggregation. On the contrary, the removal of D30's negative charge increases the net positive charge, potentially counteracting aggregation. The simulations indicated that the hydrophobic effect outweighs the electrostatic repulsion, as we observe random dimerization with primarily intra-peptide rather than inter-peptide  $\beta$ -sheet formation in *trans*-P3. To probe the effect of a mutation by AMPB near the N-terminus, we simulated the P8 mutant with AMPB inserted between K26 and K27. Although, the peptide displays increased hydrophobicity due to AMPB insertion into the positively charged N-terminal region, P8 favors  $\alpha$ -helix over  $\beta$ -sheet structures, in both *cis*- and *trans*-configurations, which may counteract amyloid fibrillization. In P1, P3, and P4  $\alpha$ -helix formation seems to be largely suppressed, due to the central positioning of the AMPB photoswitch in the peptide sequence. To examine the effects of the AMPB mutation on the C-terminal region, we simulated the P12 mutant. Here,

AMPB replaces V35 in the C-terminal aggregation-prone region, which substantially reduces  $\beta$ -sheet formation relative to PTH<sub>25–37</sub> and other variants. Similar to P8, AMPB positioning in P12 favors  $\alpha$ -helix formation, which may further inhibit amyloid aggregation.

## 6.3 Fibril Model

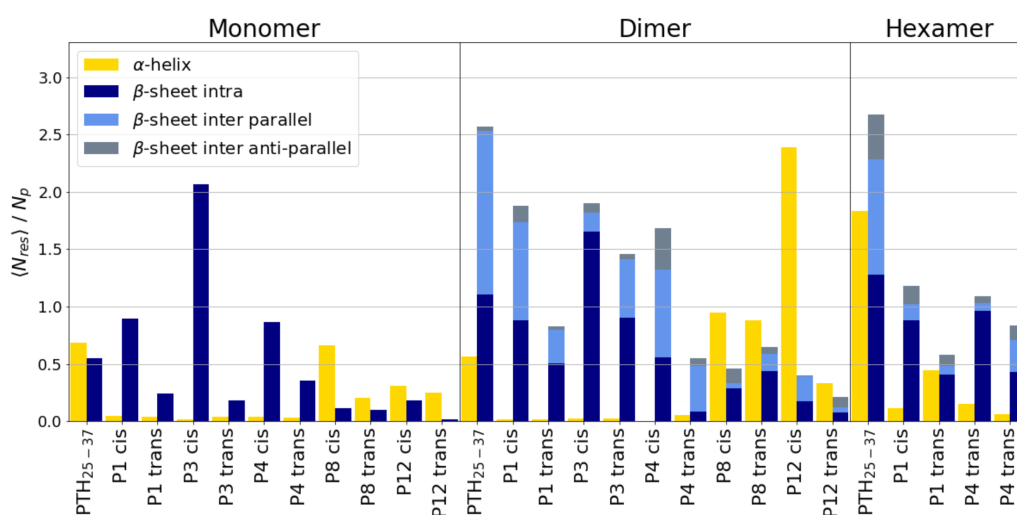
To develop a comprehensive fibril model for PTH<sub>25–37</sub> and *trans*-P4, we integrated experimental data of our collaboration partners from wide-angle X-ray scattering (WAXS) and solid-state NMR measurements with structural insights from our oligomer simulations. Based on the combined data we proposed multiple fibril structures and assessed their stability using MD simulations.

### 6.3.1 Simulation Details

The system was prepared according to the standard protocol for MD simulations, using the same parameters as described for the peptide simulations in section 6.2.1. All fibril models containing 12 peptides were simulated for 1  $\mu$ s.

### 6.3.2 Results

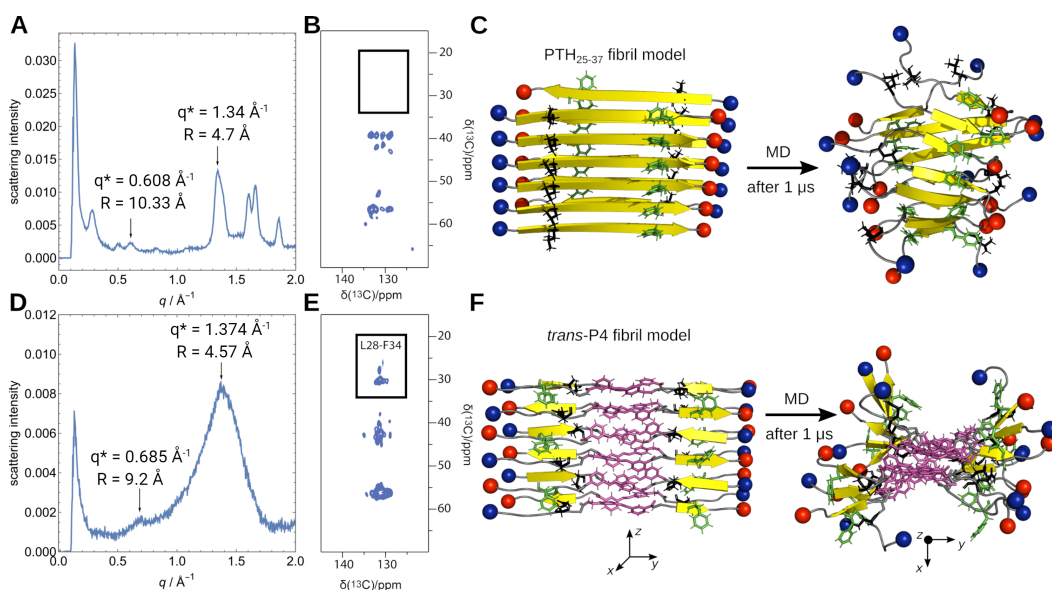
The amyloid fibrils of PTH<sub>25–37</sub> and *trans*-P4 were characterized using WAXS and solid-state NMR, with MD simulations employed to assess the stability of the resulting structural models. WAXS results for PTH<sub>25–37</sub> (see Figure 6.5 A/D) displayed a typical  $\beta$ -sheet fibril pattern with reflections at 4.7 Å (inter-strand distance) and 10.3 Å (inter-sheet distance), indicating parallel  $\beta$ -sheets. In contrast, the *trans*-P4 fibrils exhibited broader peaks, indicating a less structured fibril arrangement, yet



**Fig. 6.4.:** Average simulated secondary structure population of PTH<sub>25–37</sub> and simulated mutants in *trans/cis* conformation. The secondary structure propensity is categorized into  $\alpha$ -helical, intrapeptide  $\beta$ -sheets, and interpeptide parallel and antiparallel  $\beta$ -sheets, as indicated by the color code.

retained dominant  $\beta$ -sheet spacing's at 4.6 Å and 9.4 Å, consistent with antiparallel alignment. Solid-state NMR cross-peaks between L28 and F34 further supported the antiparallel arrangement of *trans*-P4 peptides within the fibrils, while the absence of that cross-peak for PTH<sub>25–37</sub> indicates a parallel alignment (see Figure 6.5 B/E).

Combining the experimental and computational efforts to characterize the structural properties PTH<sub>25–37</sub>/*trans*-P4 peptide interactions, we constructed several potential fibril models for both systems. MD simulations of these models revealed that the stable structure for PTH<sub>25–37</sub> consisted of parallel, in-register  $\beta$ -sheets, oriented antiparallel between sheets, stabilized by electrostatic interactions between the N- and C-terminal charges. For *trans*-P4, the fibrils are stabilized by  $\pi$ - $\pi$  stacking interactions between the central AMPB groups, leading to shorter  $\beta$ -sheets and consequently less structured fibrils. The MD-averaged structure displayed larger deviations from idealized fibrils, which explains the broadness of the corresponding WAXS signal. Both models are shown in Figure 6.5.



**Fig. 6.5.: Fibril model of PTH<sub>25–37</sub> and *trans*-P4.** (A, D) WAXS measurements of PTH<sub>25–37</sub> and *trans*-P4. (B, E) Solid-state NMR 2D <sup>13</sup>C-<sup>13</sup>C proton driven spin diffusion spectra close to rotational resonance (mixing time 1 s) of PTH<sub>25–37</sub> and *trans*-P4 fibrils with uniformly <sup>13</sup>C-labeled L28 and F34. (C, F) Fibril models of PTH<sub>25–37</sub> and *trans*-P4, initially (left) and at the end of MD simulations (right). Here,  $\beta$ -sheets are shown in yellow, random coil in gray, the N- and C-termini as blue and red spheres, respectively, AMPB in purple, and the side chains of L28 and F34 are highlighted as black and green sticks, respectively. Note that in panel F, the structure on the right-hand side is rotated by 90° to better represent the fibrillar arrangement. The figure was taken from Pascholt et al. (2024) [176].

## 6.4 Conclusion

In the presented study, we achieved precise control over reversible peptide fibrillization by embedding the photoswitch AMPB into the PTH<sub>25–37</sub> peptide. Through a combination of MD simulations, peptide synthesis, aggregation assays, and structural analysis, we demonstrated that the positioning of AMPB significantly affects fibrillization behavior. Peptides with AMPB embedded in the fibril-forming region, such as P4, exhibited enhanced aggregation in the *trans* form, while fibrillization was reduced in the *cis* form. MD simulations highlighted that AMPB facilitates  $\pi$ - $\pi$  interactions, promoting aggregation. Structural studies further confirmed that  $\beta$ -strands in P4 fibrils are arranged antiparallel, unlike the unmodified peptide. The experimental and structural insights were used to construct fibril models of the PTH<sub>25–37</sub> and P4 peptide, which proved stable in MD simulations. Notably, the strategic positioning of the AMPB photoswitch was demonstrated to enable reversible control of amyloid aggregation. Experimental results showed that phototriggered degradation of fibrils in the *trans*-state is a reversible process, permitting repeated and controlled peptide aggregation. This light-activated approach presents a promising method for controlled drug delivery via the release of reversibly fibrillating peptides, particularly given the nontoxic nature of the studied fibrils [176].



# Influence of pH on the Aggregation Behaviour of the SH3 Domain

## 7.1 Motivation

Understanding the factors influencing amyloid formation is crucial for advancing our knowledge of neurodegenerative diseases such as Alzheimer's disease (AD). One of the critical factors in this context is the pH, which plays a pivotal role in the aggregation behavior of amyloid-fibrils.

Previous research indicates that the ionization states of amino acid residues, which are pH-dependent, significantly influence amyloidogenesis by either facilitating or inhibiting protein aggregation through electrostatic interactions [180]. Specifically, for amyloid-beta ( $A\beta$ ), studies have shown that changes in net charge and the reduction of electrostatic repulsion at low pH levels greatly affect the self-association of peptides during the initial nucleation phase, but they have low impact on fibril elongation and nucleation on fibril surfaces [181, 182]. Additionally, in Alzheimer's pathology, the oligomeric form of  $A\beta$  has been identified as the primary neurotoxic species. These metastable oligomers, exceeding 50 kDa, form a unique assembly type, distinct from the intermediates in amyloid fibril formation. It has been shown that  $A\beta$  oligomer formation is highly pH-dependent, with a massive acceleration when transitioning from neutral to more acidic pH levels [183]. Consequently, understanding the early stages of amyloid aggregation and oligomerization, particularly the role of pH, is vital for elucidating the mechanisms underlying related disease pathologies.

The SH3 domain from the p85 $\alpha$  subunit of bovine phosphatidylinositol-3-kinase (PI3K-SH3) has been studied excessively and served as a model system for investigating protein folding, aggregation, and fibril formation. Recent investigations have examined the aggregation kinetics of the SH3 domain under low pH conditions using a combination of solution NMR, high-resolution magic angle spinning (HR-MAS) NMR, and solid-state NMR techniques [76]. These studies utilized recombinantly expressed, low pH-denatured [ $^{13}\text{C}$ ,  $^{15}\text{N}$ ] labeled bovine PI3K-SH3, employing a fibrillization protocol similar to that for  $A\beta$  [184]. Similar to the aggregation of  $A\beta$  peptides, the acidic conditions often accelerate the aggregation, leading to faster fibril formation. For the SH3 domain, which is a folded protein at neutral pH, experiments suggest that acidic conditions (pH  $\sim$  2), induce a structural change leading to denaturation of the secondary structure. Thus, at pH 2 it behaves effectively like an intrinsically disordered protein (IDP) and subsequently aggregates



**Fig. 7.1.: Overlay of bovine and human SH3-domains structure.** The structure of the human SH3 domain is shown in red, and the structure of the bovine SH3 domain is shown in blue.

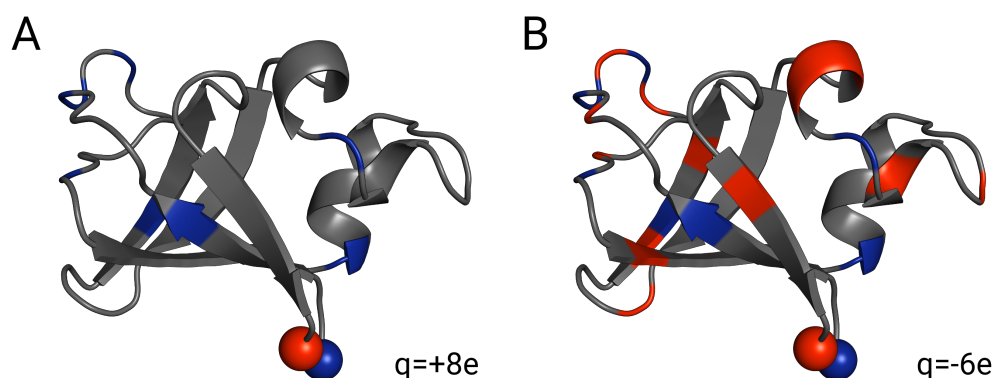
into amyloid fibrils. Recent studies have further shown that this process is highly temperature-dependent and reversible with changes in temperature and pH.

In this study, we aim to elucidate the structural changes of the SH3 domain following a pH change by manipulating the protonation states of titratable sidechains and performing conventional MD simulations along with enhanced sampling REMD simulations. Highlighting the difference between physiological and acidic conditions, we compare simulations at pH  $\sim 2$  and pH  $\sim 7$ . The focus is on investigating the differences in kinetics and structure on short timescales, which are not observable in experiments, thus providing insights into the early (un-)folding processes driving amyloid aggregation.

## 7.2 Structure Preparation

Starting from the published structure of the human SH3 domain (PDB:1PKS), we mutated the amino acid sequence to match that of the bovine SH3 domain used in experiments by Luis Gardon, Melinda Jaspert, Magdalena Kuom, Nils Lakomek, Lothar Grämer, and Henrike Heise, which also matches the sequence reported by Röder et al. [185] (see Figure 1.7). The secondary structure of both the human and bovine SH3 domains exhibits the same characteristics as described in Chapter 1.6. Figure 7.1 presents an overlay of the mutated bovine 1PKS structure (red) with that of the 1PNJ bovine structure (blue), both determined from solution-state NMR [186, 187]. Both structures display the characteristic  $\beta$ -sheet core structure with a flexible region from residues A10 to D27, which is mostly in a random coil, and a  $\alpha$ -helical region from residue K34 to I53.

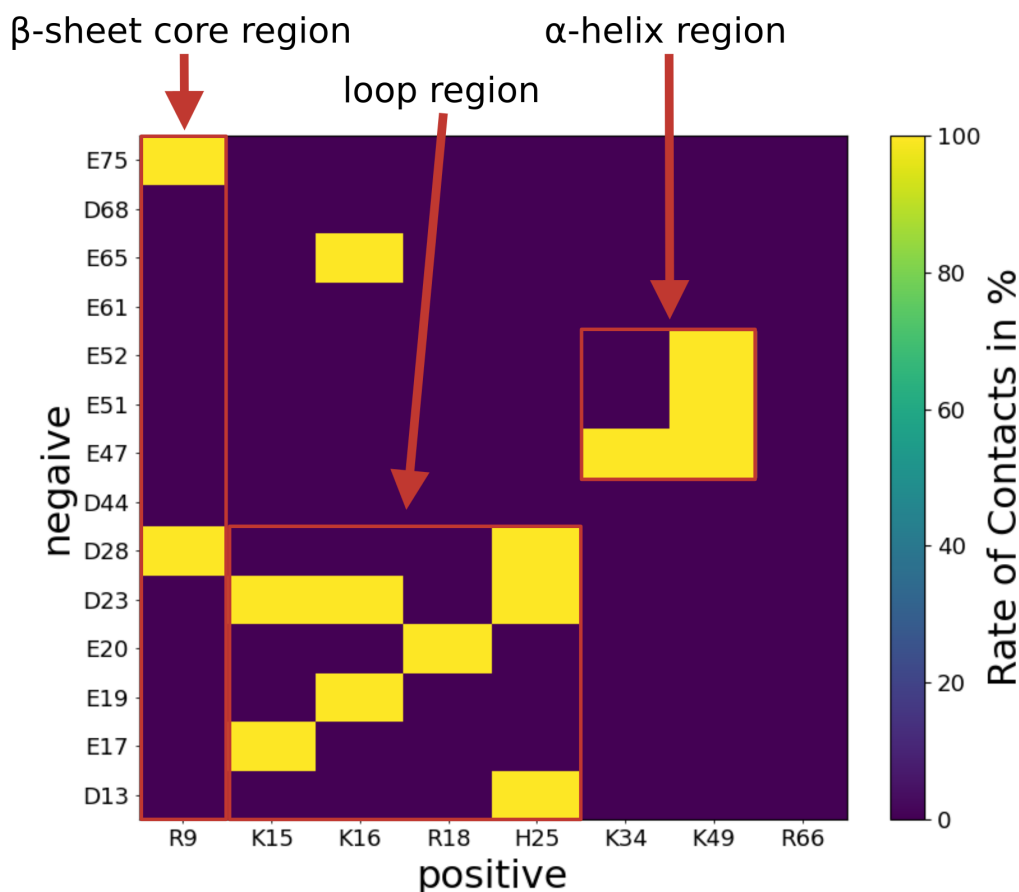
In conventional MD simulations, the pH is typically accounted for indirectly by assigning the protonation states of titratable residues to their presumed most probable state of a given pH, at the start of the simulation. Since electrostatic interactions



**Fig. 7.2.: Charged sidechains of the SH3-domain at different pH.** (A) shows the charged sidechains at pH=2 and (B) at pH=7. Positively charged sidechains are colored in blue and negatively charged ones are colored in red. The total charge of each structure is shown in the lower right corner.

are strongly affected by the protonation state of residues, the pH can influence the sampled conformational ensemble strongly. Therefore, simulations that only sample the most probable protonation states at the start may not explore the full conformational space. Furthermore, conformation can affect the proton affinity (pKa) of residues, establishing a direct correlation between pH and conformational dynamics that cannot be accurately captured if the protonation state remains fixed during the simulation [188]. In recent years, several methods have been proposed to model the influence of pH on the structural ensemble dynamically, by allowing the protonation state of the sidechains to change during the course of the simulation [189]. This approach has been improved to work for large systems and incorporated into the GROMACS simulation package [188]. This method is particularly useful if a protein has multiple sidechains with pKa values close to the desired pH of the simulation, as these are similarly likely to be protonated or deprotonated.

In this study, we aim to investigate the differences in conformational propensity in extreme conditions of very acidic solution (pH=2) and physiological conditions (pH=7). Therefore, we fix the protonation states during the simulation, which conserves resources and speeds up the process. The protonation state of the titratable side chains was chosen according to the pKa values calculated with the H++ webserver [190]. The calculations of the individual pKa values are performed using the standard continuum solvent methodology within the Poisson-Boltzmann (PB) model framework [191, 192]. Figure 7.2 shows the structures of the SH3-domain with their sidechains colored according to their net charge at pH 2 and pH 7. At pH 2, all sidechains that were previously negative at pH 7 become protonated and lose their net charge, namely: D13, E17, E19, E20, D23, D28, D44, E47, E51, E52, E61, E65, D68 and E75. The salt bridges formed at pH 7 are shown in Figure 7.3 as a contact map. The figure provides an overview of where these contacts are formed. The contacts between R9 with D28 and E75 occur between neighboring  $\beta$ -sheets in the  $\beta$ -sheet core structure. The positively charged residues K15, K16, R18, and



**Fig. 7.3.: Contact map of charged sidechains of the SH3-domain at pH 7.** A contact is assumed to be formed if two oppositely charged residues are closer than 4 Å, excluding next neighbours. As the contact map of the native structure is displayed, contacts occur only with a rate of 0% or 100%. Furthermore, the region involved in the formation of specific contacts is highlighted.

H25 are part of the flexible loop region and primarily form contacts with negatively charged residues in the same region, such as E17, E19, E20, D23, and D28. However, the contact between K16 and E65 attaches the loop region to the backside of the  $\beta$ -sheet core structure, resulting in an overall globular protein structure. Lastly, the contacts between K34 and K49 with E47, E51, and E52 stabilize the  $\alpha$ -helical region. Structurally, the loss of these salt bridges should weaken both the  $\beta$ -sheet core structure and the  $\alpha$ -helical region.

## 7.3 Conventional MD - Monomer

Initially, we conducted a 10  $\mu$ s MD simulation of the SH3 domain monomer at pH 2 and pH 7 to observe possible pH-induced conformational changes in the protein's structure. The experimental observation of unfolding at acidic pH, coupled with the loss of stabilizing salt bridges, suggests significant structural changes at pH 2 to take place.

### 7.3.1 Simulation Methods

The simulations of the SH3-domain were performed using the GROMACS simulation package [193]. In each simulation the protein was modeled using the CHARMM36m force field [82] together with the TIP3P water model [97]. It has been found in previous studies, that the CHARMM36m force field is best suited for modeling both monomeric A $\beta$ <sub>42</sub> and other IDPs [96], but is also applicable to folded proteins [158, 194].

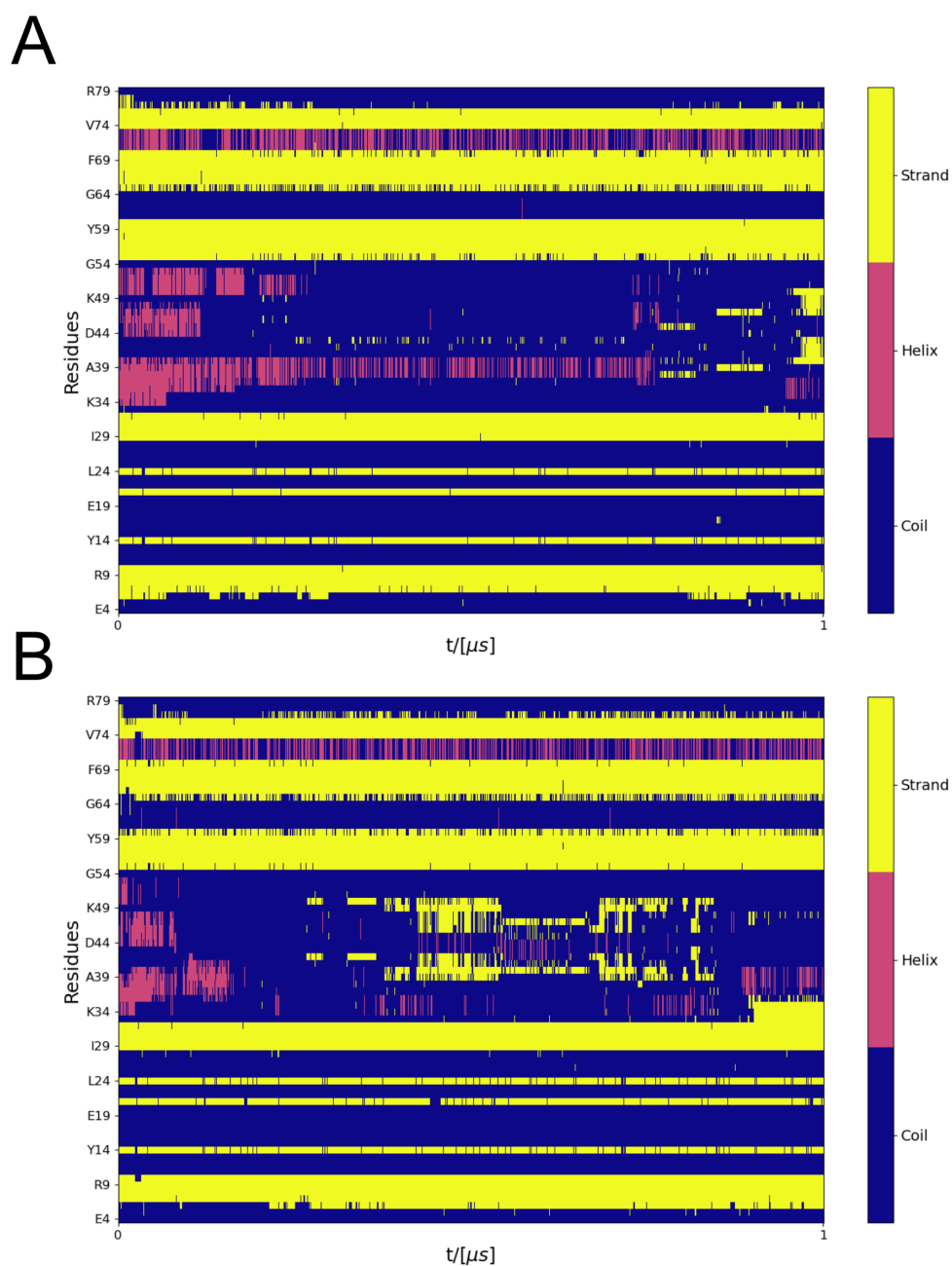
The systems were prepared according to the standard protocol for MD simulations. Initially, proteins were placed in a simulation box of sufficient size to prevent self-interaction across the PBCs, even in the event of partial unfolding. The simulation box was then filled with TIP3P water molecules. Sodium (Na) and chloride (Cl) ions were added to neutralize the system electrostatically. Following system equilibration, we conducted simulations for 10  $\mu$ s under NPT conditions at 1 bar, using the Parrinello-Rahman pressure coupling scheme [195]. The temperature was maintained at 308 K using the Nosé-Hoover thermostat [106]. For integration, the leap-frog algorithm was used with an integration time step of 2 fs [103]. All simulations were performed under PBCs in all directions, and the particle-mesh Ewald method [145] was employed to calculate electrostatic interactions. The cutoffs for van der Waals and Coulomb interactions calculated in real space were both set to 12 Å.

### 7.3.2 Results

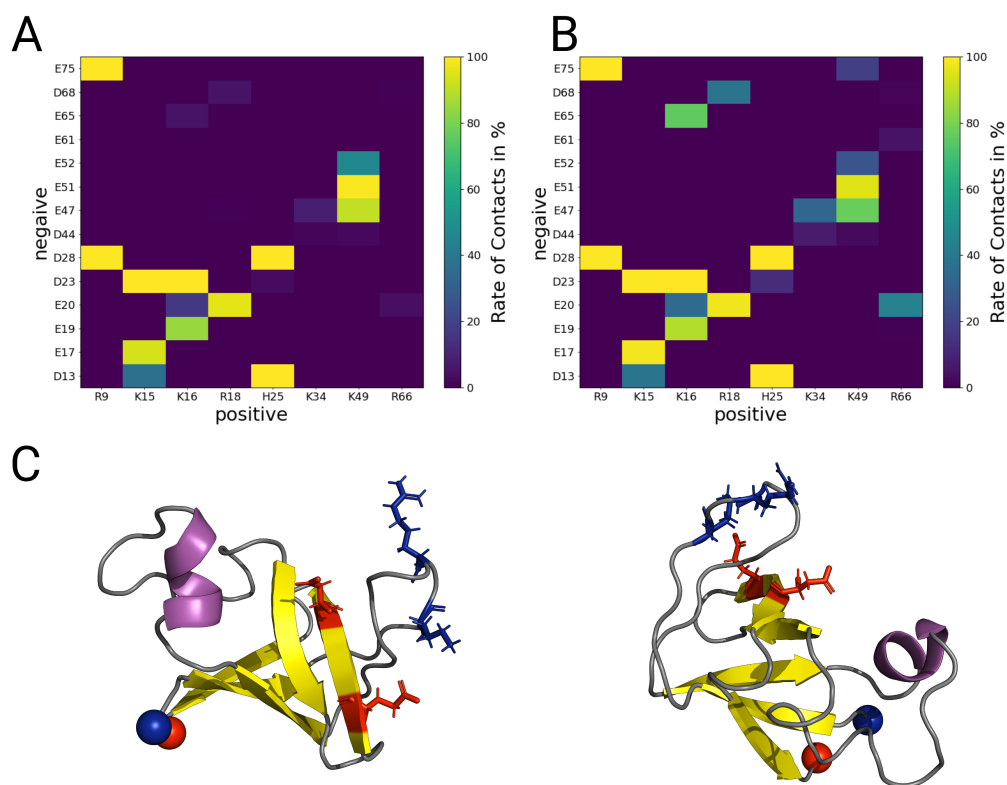
Figure 7.4 shows the evolution of the secondary structure over time for the SH3-domain at pH 2 and pH 7. During the simulation, structural changes occur primarily in the  $\alpha$ -helical region of the protein between residues K34 and G54. At pH 2, the initial  $\alpha$ -helical structure mostly unfolds, whereas at pH 7, it begins to form  $\beta$ -sheets towards the end of the simulation. Outside the  $\alpha$ -helical region, the rest of the protein secondary-structure remains stable at both pH levels.

This is further supported by inspecting the intra-molecular contact rates of charged residues and comparing them to the native contacts in Figure 7.3. Figure 7.5 shows the rate of contact between residues which are charged at pH 7, but become neutral at pH 2. The contacts formed in the  $\beta$ -sheet core between R9 and D28/E75 remain intact, which is consistent with the secondary structure plot in Figure 7.4. Furthermore, the contacts in the  $\alpha$ -helical region remain mostly intact at both pH 2 and pH 7, despite the secondary structure plot showing temporary reorganization into  $\beta$ -sheet structures. Only the contacts between the loop region and the  $\beta$ -sheet core structure, specifically between K16/R18 and E65/D68 as well as between R66 and E20, are broken at pH 2, indicating a dissociation of these two regions or at least weaker binding.

Examining the full contact map of the native fold and comparing it to the contact rates during the simulation can yield further insights into these structural changes. Figure 7.6 displays the contact map of the initial NMR structure (PDB: 1pks), used for



**Fig. 7.4.: Secondary structure of the SH3 monomer at different pH.** Secondary Structure over time of the SH3 domain at (A) pH=2 and (B) pH=7. The y-axis shows all residues in the SH3 protein, which are classified as either in random coil (blue), forming an  $\alpha$ -helix (pink) or forming a  $\beta$ -sheets (yellow).



**Fig. 7.5.: Contact rates of charged sidechains of the SH3 monomer from conventional MD.** The rate of contact for (A) pH 2 and (B) pH 7 is averaged over the full 10  $\mu$ s simulation. A contact is assumed to form if two oppositely charged residues are closer than 4 Å. In (C) the crystal structure of the SH3 domain is shown from two different angles. The two residues 65-GLU/68-ASP, which are negatively charged at pH 7, are shown in red and the two positively charged residues 16-LYS/18-ARG are shown in blue.

subsequent simulations. Examining the full contact map reveals the contacts formed during a simulation and provides insights into the secondary structure present. In a contact map,  $\alpha$ -helices are identified by a trace of contacts parallel and close to the diagonal, as these contacts are formed between residue  $i$  and its nearest neighbors  $i + 2, i + 3$  and  $i + 4$ . This is observed in the natively folded structure within the central region from residue K34 to G54. Similarly, connected traces orthogonal to the diagonal indicate interactions between neighboring antiparallel  $\beta$ -sheets. In Figure 7.6, all  $\beta$ -sheets comprising the  $\beta$ -sheet core structure can be identified by such traces. Additionally, the contacts in the central region of the triangular matrix represent interactions between the two stacked  $\beta$ -sheet regions, as well as interactions of the flexible loop region with the  $\beta$ -sheet core region, maintaining the structure's globular form.

Figure 7.7 depicts the rate of contacts during the simulation at pH 2 and pH 7. Here, the lower triangular matrix shows the contacts observed during the simulation, while the upper triangular matrix highlights the deviations from the initial folded structure. At both pH 2 and pH 7, most of the changes to the initial fold occur in

the central  $\alpha$ -helical region. The difference matrix clearly shows the loss of helix-stabilizing contacts. However, at pH 7, these reorganize into a pattern resembling the formation of an antiparallel  $\beta$ -sheet, indicative of a  $\beta$ -hairpin structure. In contrast, at pH 2, the new contacts appear random, suggesting a compact disordered region. The unfolding of the  $\alpha$ -helical region is associated with the loss of contacts between the regions around K49 and V74, disrupting its connection with the  $\beta$ -sheet core. While at pH 2 these contacts reestablish differently, at pH 7 this region remains detached, indicating the protrusion of the newly formed  $\beta$ -hairpin.

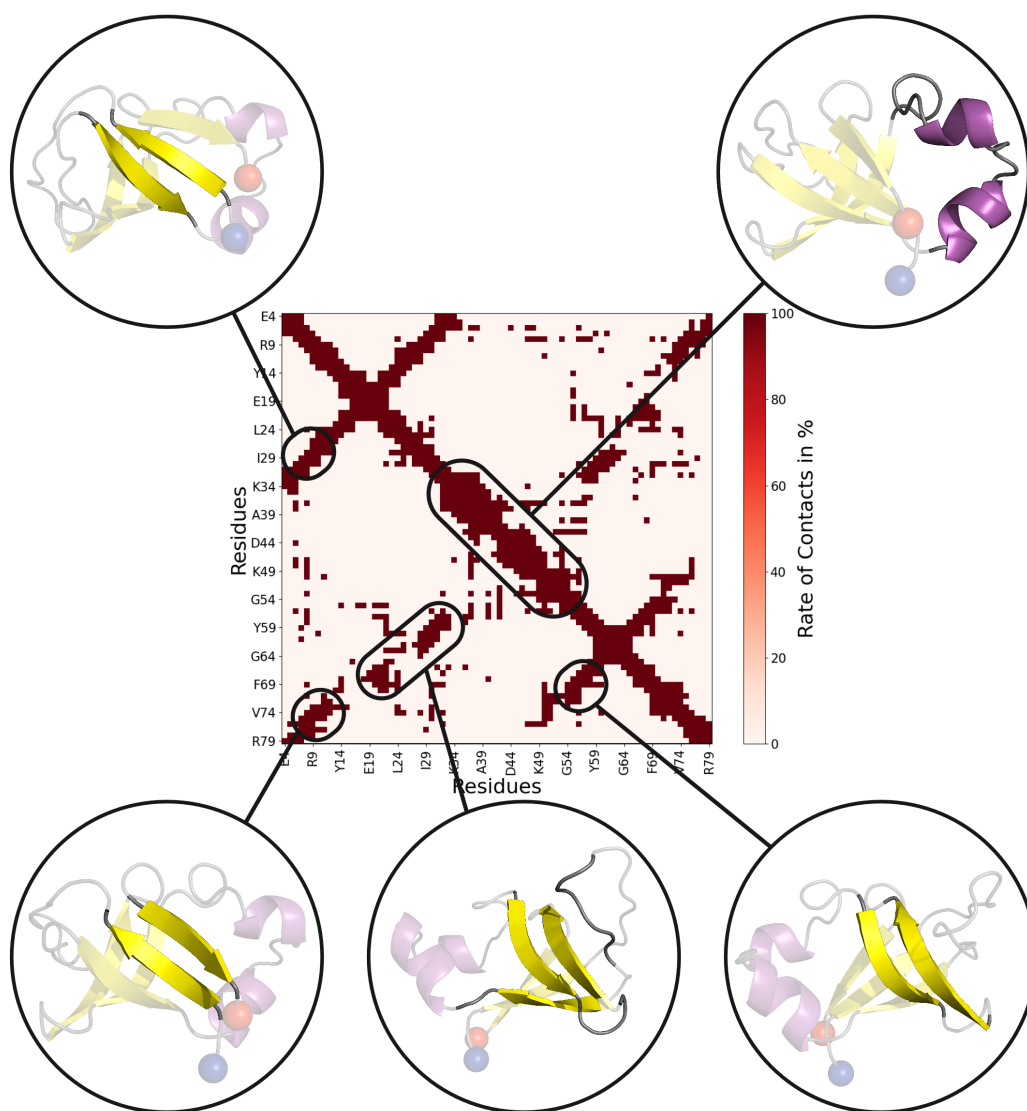
Another similarity observed between the two simulations is the reorientation of the N-terminal region, which disrupts its initial contacts with the C-terminal and the beginning of the loop region. At pH 7, the N-terminal does not appear to bind with another region. In contrast, at pH 2, it "folds" inward, contacting the more central part of the  $\alpha$ -helical region. This folding possibly hinders the dissociation of the formerly  $\alpha$ -helical now flexible region and the subsequent formation of a  $\beta$ -sheet.

Lastly, at pH 7, we observe a significant loss of contacts between the flexible loop region and the  $\beta$ -sheet core region, specifically the regions between residues Y59 to F69, as well as between E19 to I29. Although these regions do not fully dissociate, the loss of contacts that maintain the protein's globular form is unexpected at this pH, where it is generally considered most stable. This observation may be explained by the relaxation of the structure into a more favorable orientation compared to the NMR structure, where due to a sparsity of distance restraints for flexible regions, it is difficult to determine the exact average representation of dynamic regions during a standard NMR structure calculation.

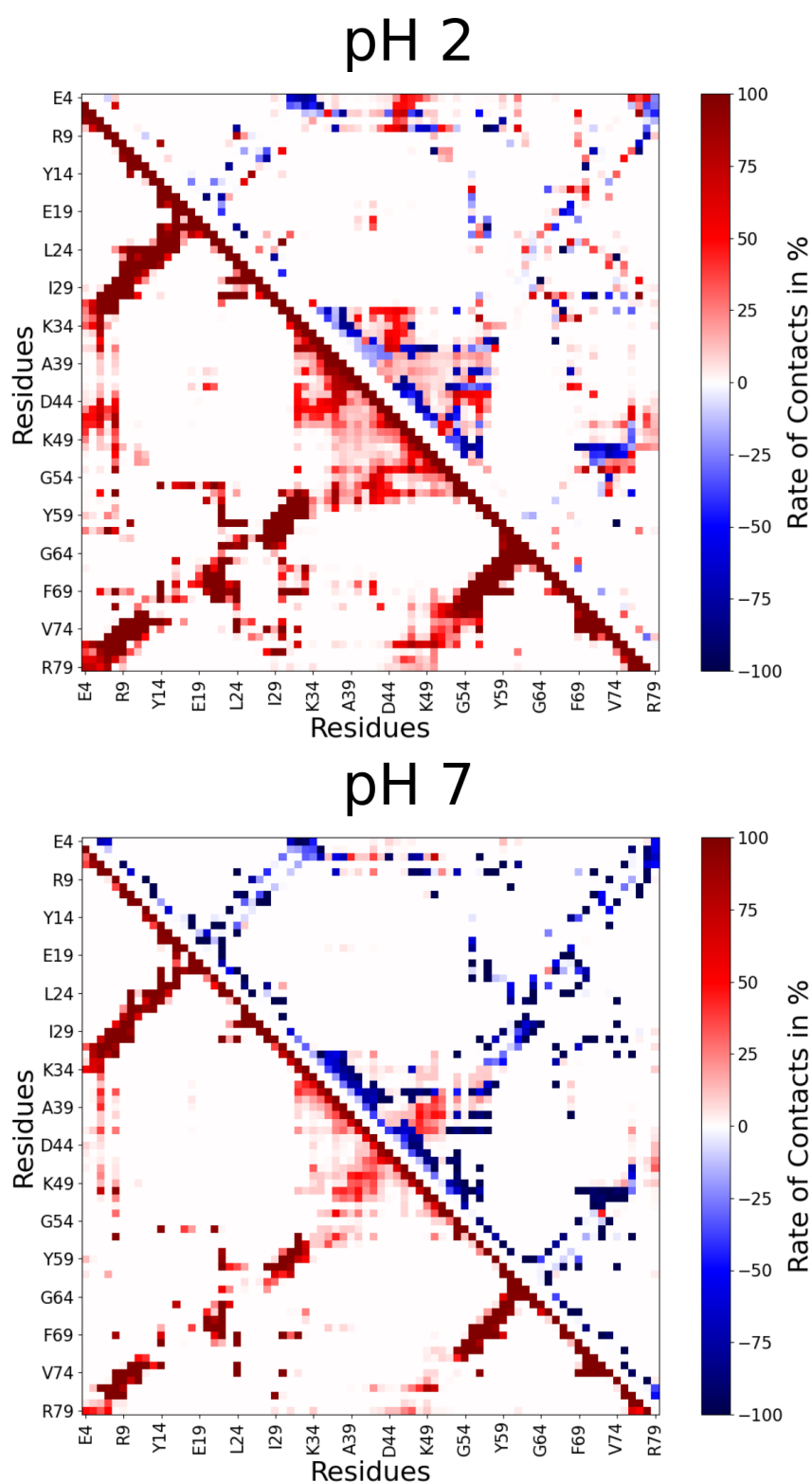
### 7.3.3 Discussion

Overall, conventional MD simulation of the monomer did not replicate the unfolding of the robust  $\beta$ -sheet core structure, even not at pH 2. Only the initial  $\alpha$ -helical region undergoes significant structural changes, mostly unfolding at pH 2 and forming a  $\beta$ -hairpin at pH 7. Protonation of the titratable side chains alone does not appear sufficient to destabilize the  $\beta$ -sheet structures. Thus, the unfolding pathway likely involves additional environmental factors, such as interaction partners or excess ions present in the experimental buffer conditions.





**Fig. 7.6.:** Map of the native contacts of the SH3 monomer. Contacts in the NMR structure (1pks), together with structures highlighting the respective contact region. Two residues are considered to be in contact if any pair of atoms from the separate residues are within a distance of 6 Å.



**Fig. 7.7.:** Contact map of the SH3 monomer at different pH. The lower triangular matrix shows the contacts present during the simulation, while the upper triangular matrix indicates the difference from the initial fold. In the latter, positive rates (red) denote the formation of new contacts, while negative rates (blue) indicate the loss of contacts present in the initial structure. Two residues are considered to be in contact if any pair of atoms from the separate residues are within a distance of 6 Å.

## 7.4 Conventional MD - Dimer

Since the monomer simulations did not reproduce the experimental results of the SH3 domain unfolding at pH 2, we conducted a dimer simulation at pH 2. This approach was used to investigate whether the unfolding pathway could be facilitated through interactions with other SH3 proteins.

### 7.4.1 Simulation Methods

The system was prepared according to the standard protocol for MD simulations, using the same parameters as described for the monomer simulations in section 7.3.1. The monomers were placed in a simulation box with an initial minimum distance of 2 nm between them and with a box volume of  $V=1300 \text{ nm}^3$ , resulting in a protein concentration of 2.6 mM. This relatively low protein concentration is necessary to avoid artifacts due to interactions across the PBCs in the event of unfolding. We conducted three simulations, each for 4  $\mu\text{s}$ , resulting in a total accumulated simulation time of 12  $\mu\text{s}$ .

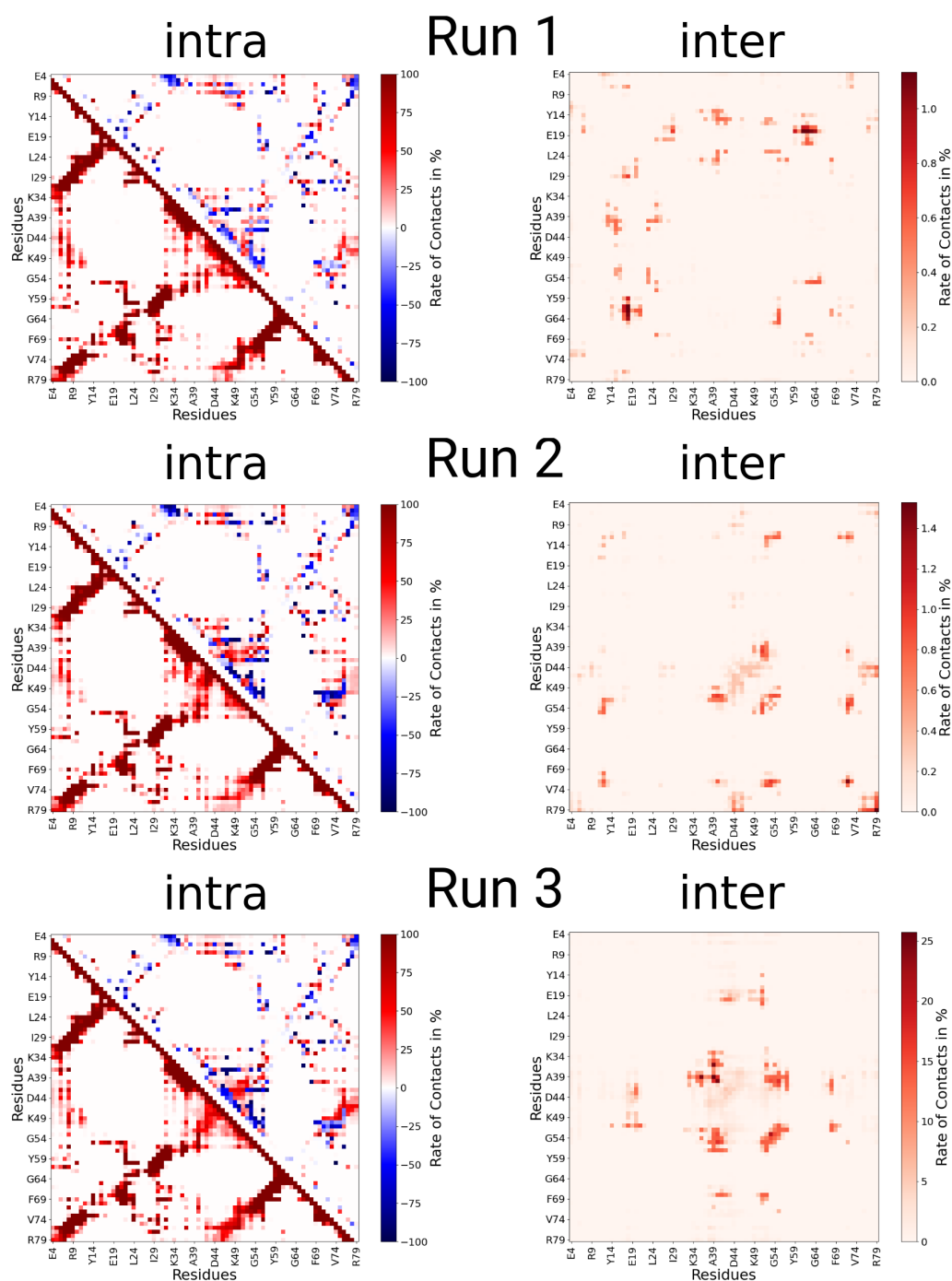
### 7.4.2 Results

Figure 7.8 displays the inter- and intra-molecular contact maps from the dimer simulations. Interestingly, the two SH3 monomers exhibit very low interaction propensity. Inter-molecular interactions were observed only in one of the three simulations, specifically run 3. In this case, the interactions are primarily facilitated by the  $\alpha$ -helical region of the SH3 domain, which forms contacts mainly with the  $\alpha$ -helical region of the other monomer, as well as with the  $\beta$ -sheet region between residues G64 to P70 and parts of the flexible loop region between residues Y14 to E20.

Examining the intra-molecular contacts reveals a similar structural pattern as observed in the monomer simulation. However, the contacts stabilizing the central helix between residues K34 and A39 remain intact in all dimer simulations. Consequently, the  $\alpha$ -helix remains stable throughout the majority of all three simulations, in contrast to the monomer simulation.

### 7.4.3 Discussion

Our simulations reveal a very low interaction propensity between two SH3 domain proteins. Significant inter-molecular interactions were observed in only one of the three simulations, primarily facilitated by the  $\alpha$ -helical region of the SH3 domain. Additionally, this region remained stable in all dimer simulations, even in the absence of direct interactions. Although these simulations did not replicate the unfolding of the SH3 domain at acidic pH, they emphasize the structural importance of the  $\alpha$ -helical region, which may play a crucial role in the oligomerization process.



**Fig. 7.8.: Contact maps of the SH3 dimer at pH 2.** The intra- and inter-molecular contact maps for all three runs are shown. For the intra-molecular contact map, the lower triangular matrix shows the contacts present during the simulation, while the upper triangular matrix indicates the difference from the initial fold. In the latter, positive rates (red) denote the formation of new contacts, while negative rates (blue) indicate the loss of contacts present in the initial structure. Two residues are considered to be in contact if any pair of atoms from the separate residues are within a distance of 6 Å. Note that the color scale for the inter-molecular interactions is adjusted relative to the maximum rate of contacts in the respective simulation.

## 7.5 Replica Exchange MD

Conventional MD simulations did not demonstrate the unfolding of the SH3 domain at acidic pH, as observed experimentally. Therefore, it is assumed that even at pH 2, the folded state remains in a local free-energy minimum, which is sufficiently stable to prevent major unfolding events in conventional MD simulations. To sample the unfolding of the SH3 domain and observe the unfolding pathway, we performed replica exchange MD (REMD) simulations. In REMD, multiple replicas of the system are run in parallel at different temperatures, i.e. allowing to access conformational states with higher energies (see Chapter 2.2). Thus, the system should acquire enough energy in higher temperature replicas to escape the local minimum and sample unfolding events.

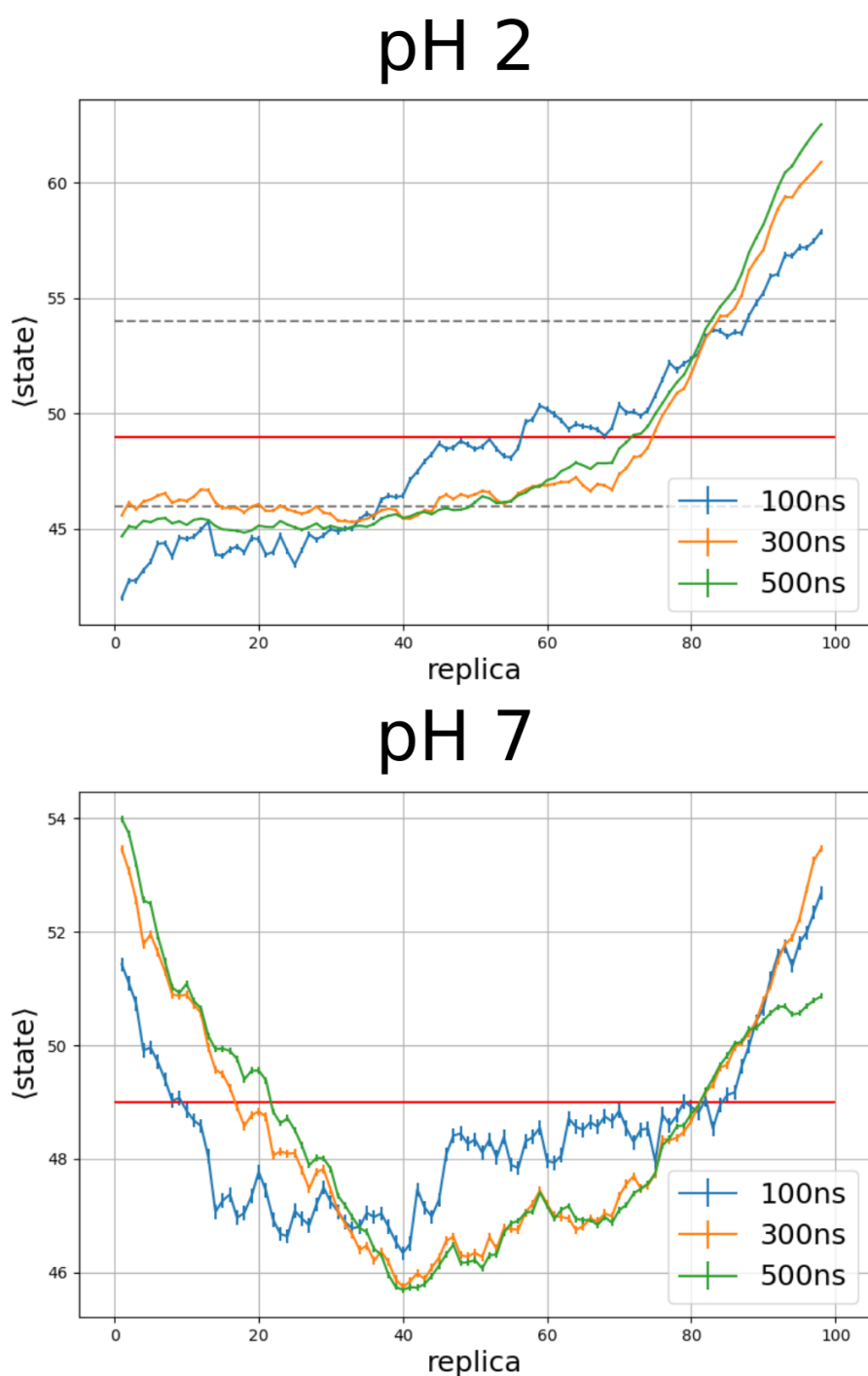
### 7.5.1 Simulation Methods

The system was prepared following the standard protocol for MD simulations, using the same structure and parameters as described for the monomer simulations in section 7.3.1. The starting structure was identical for all replicas. A total of 98 replicas were simulated over a temperature range from 300 K to 450 K, achieving an average acceptance rate for replica exchange of approximately 0.3. The simulations were run for 500 ns and 450 ns at pH 2 and pH 7, respectively. The duration of the REMD simulation was determined by the convergence of the sampling, as discussed below.

### 7.5.2 Convergence

In an REMD simulation, each replica of the system explores both the configurational space and the temperature space by exchanging temperatures with neighboring replicas. This process should be ergodic, i.e. for a sufficiently long simulation, each replica should assume each temperature state with equal likelihood, resulting in the average state of  $N_{\text{replica}}/2 = 54$  for all replicas. Here, replicas are numbered 1, 2, ...,  $N_{\text{replica}}$  with increasing temperatures from 300 K to 450 K.

Figure 7.9 shows the average state per replica during the REMD simulation at pH 2 and pH 7. Here, the average states are plotted for the full 500 ns simulation, as well as for the first 100 ns and 300 ns. It can be observed that the states converge reasonably well towards the average state of 49. However, comparing replicas 80-98 at 300 ns and 500 ns reveals that these states actually start diverging at pH 2. Consequently, replicas 1-79 tend to populate lower energy states. The divergence of the simulation can be explained by the unfolding of the SH3 domain at high temperatures. When the structure unfolds, it becomes significantly different from its folded structure, resulting in a large difference in the protein's potential energy. This leads to a low acceptance probability for exchanges between replicas across these two regimes. Since folding is a much slower process compared to unfolding [196], we do not expect the structure to return to its folded state during our simulation, thus a longer simulation would not lead to better convergence of the REMD in this case.



**Fig. 7.9.: Average state per replica during an REMD simulation of the SH3 monomer.** The average state for each replica over the first 100 ns and 300 ns as well as the full 500 ns is shown for the REMD simulation at pH 2 and pH 7. Absolute convergence would be reached if all replicas acquire an average state of  $N_{\text{replica}}/2$  (red line). For comparison, in case of pH 2 the limits of the pH 7 figure are plotted as gray lines.

At pH7, however, better convergence is observed, reflecting fewer unfolding events, consistent with experimental results. As the simulation is reasonably converged at 500 ns and shows first unfolding events, we decided to stop the simulation and use it for consequent analysis.

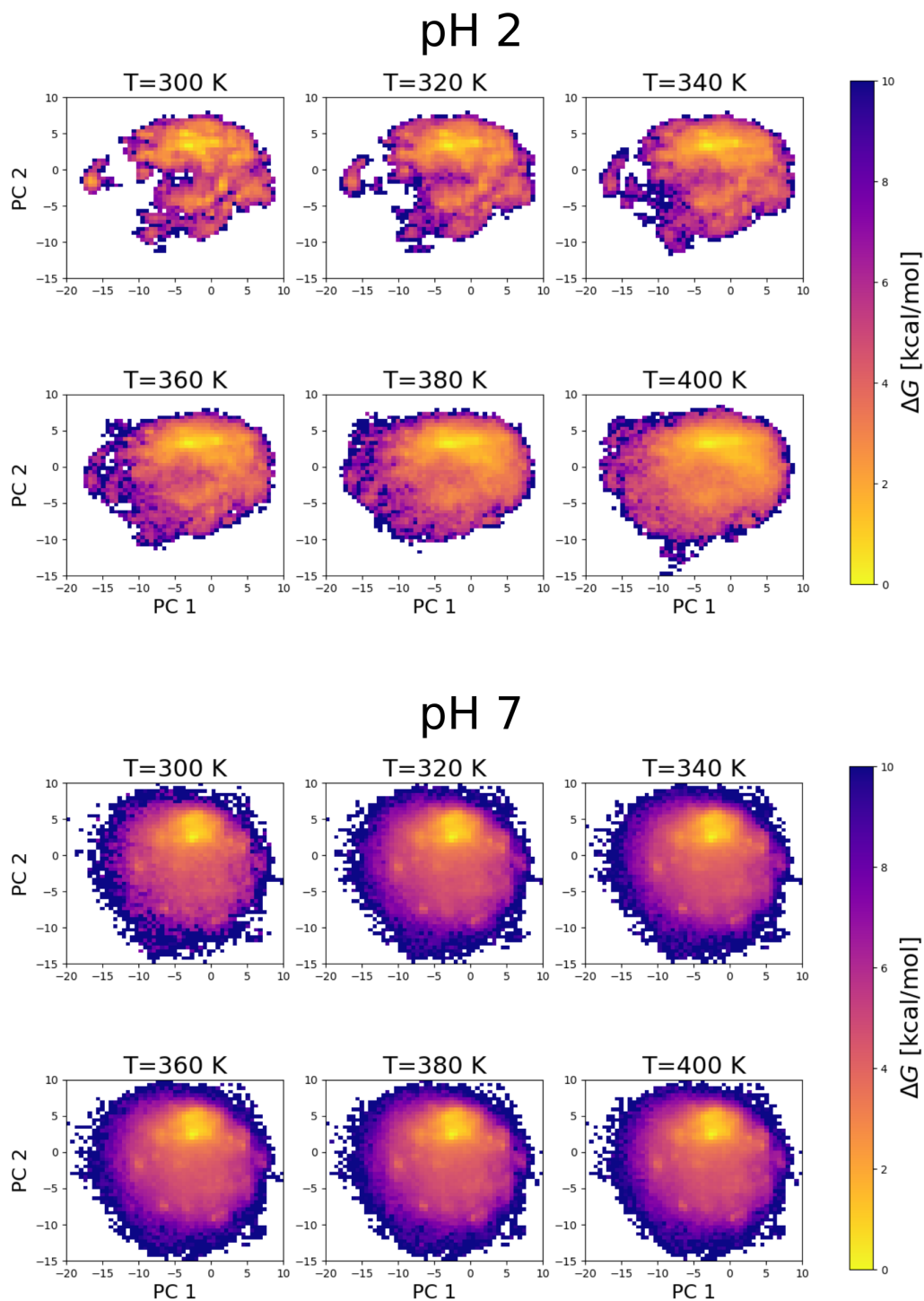
### 7.5.3 Results

To visualize the high-dimensional configurational space and capture the most prominent structural changes during the simulation, we performed a principal component analysis (PCA) and calculated the FES in the space spanned by the first two principal components. The PCA was conducted on the full dataset of trajectories from all 98 replicas at pH 2. We then projected the individual trajectories into this space to determine which parts of the FES are explored at different temperatures. Additionally, we employed the multistate Bennett acceptance ratio (MBAR) method [121, 122] to include the information from all replicas at each temperature. The MBAR method, similar to the weighted histogram analysis method (WHAM) [123], reweighs the data from different temperatures based on their relative Boltzmann weights, thus utilizing the full dataset to obtain the FES at a specific temperature.

Analysis of the first two principal components (PCs) reveals that the predominant motion involves the turning and twisting of the region between residues K34 to G54. This region corresponds to the initial  $\alpha$ -helical segment, which unfolds in the conventional monomer simulations and facilitates binding in the dimer simulations. The resulting FES projected onto the first two PCs is shown in Figure 7.10. The figure illustrates the changes in the FES with increasing temperature, ranging from 300 K to 400 K, for both pH 2 and pH 7.

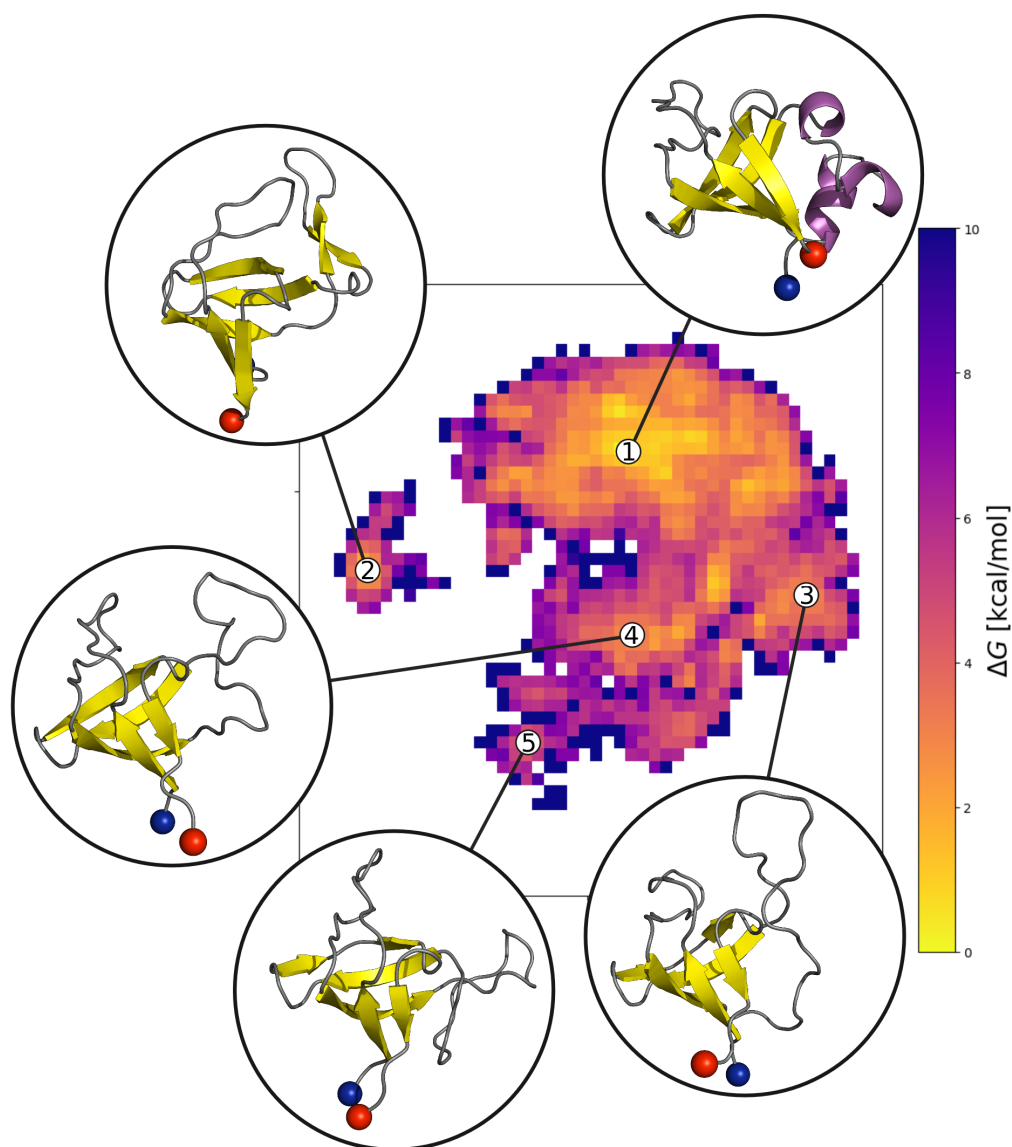
At pH 2, distinct minima are observed in the FES, which broaden with increasing temperature and eventually merge into a large basin at 400 K. In contrast, at pH 7, only a single basin is observed at all temperatures, corresponding to the native fold of the SH3 domain. The FES at pH 2 reveals more accessible states besides the folded structure even at 300 K, indicating a less stable folded structure compared to pH 7. Additionally, at 400 K, the protein explores a broader conformational space away from the initial fold at pH 2 compared to pH 7. A closer inspection of the FES at  $T=300$  K for pH 2, as depicted in Figure 7.11, reveals the predominant motions during the initial stages of the unfolding process. The PCs accounting for the highest variance in the system primarily describe the movements of the  $\alpha$ -helical region of the SH3 domain, as evidenced by the structures representing the local minima. In the following, we discuss the structural features of the minima in the FES, using the contact maps of the ensemble of structures corresponding to these minima, as shown in Figure 7.12.

Minima (1) is associated with the native fold of the SH3 domain. The contact map shows, that the contacts stabilizing central  $\alpha$ -helix between residues K34 to A39 remains largely intact. Other than that, the contact map resembles that from the conventional MD at pH 2, further supporting that during the former simulations



**Fig. 7.10.:** Free energy surface of the SH3 domain at different temperatures obtained from REMD simulations at pH 2 and pH 7. The PCs on which the FES are projected were calculated from all replicas at pH2. Both the FES at pH 2 and pH 7 are shown in the same PC space. For better sampling at each temperature, the corresponding FES was reconstructed using the pymbar python package, which employs the multistate Bennett acceptance ratio (MBAR) method [121].

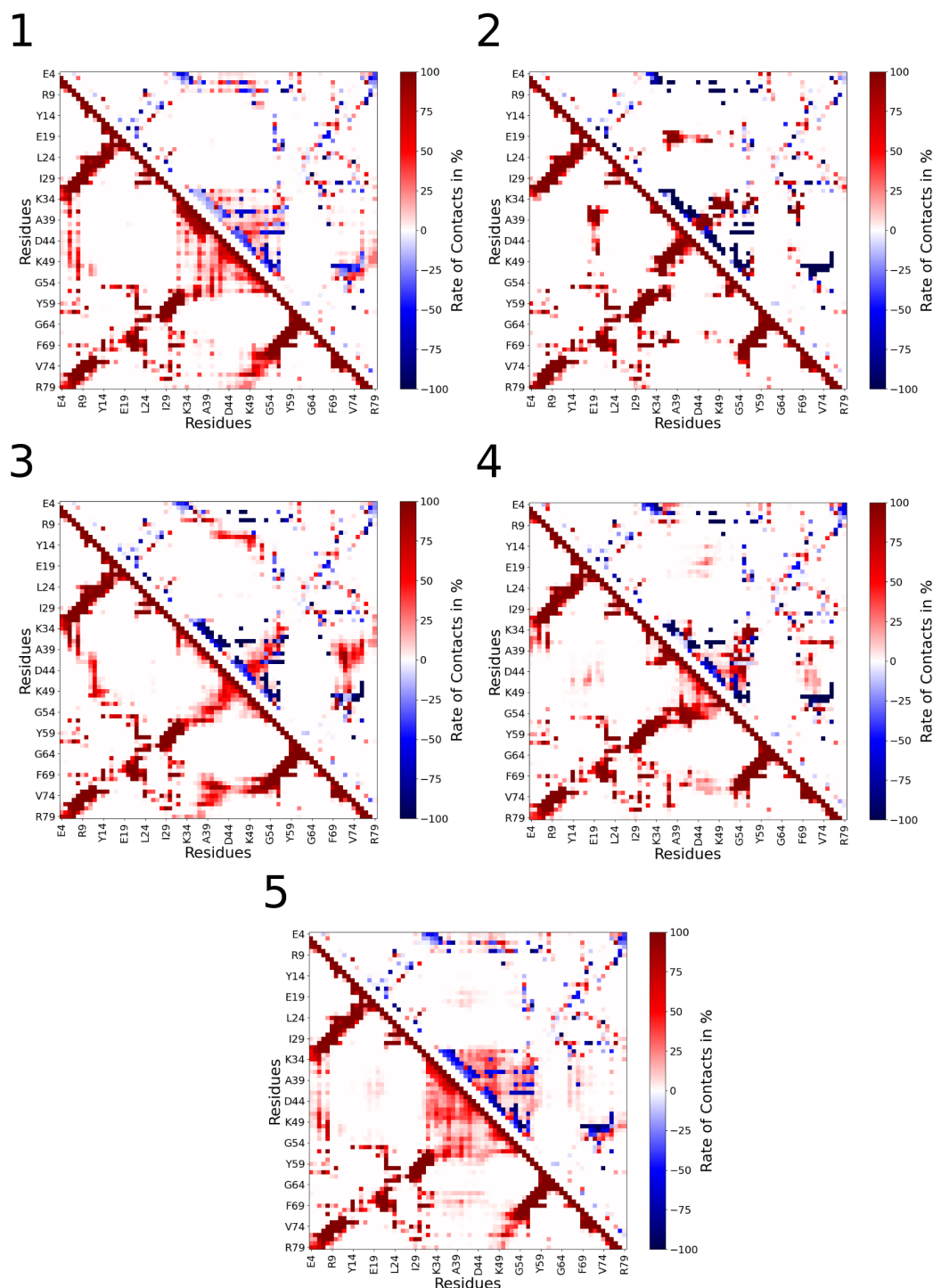




**Fig. 7.11.:** Free energy surface of the SH3 domain at  $T=300$  K and pH 2 with reference structures. The PCs on which the FES are projected were calculated from all replicas at pH 2. For better sampling at each temperature the FES was reconstructed using the pymbar method.

mostly the local minimum belonging to the native fold was explored. This can also be seen from the projection of the conventional MD trajectory onto the FES obtained from REMD (see Figure A1).

To the left of the central basin, minimum (2) is separated from the main basin by an infinitely high energy barrier at  $T=300$  K. In this state, the central  $\alpha$ -helix refolds into a  $\beta$ -hairpin structure, as indicated by the representative structure and contact map. The contact map reveals this transition by showing a loss of contacts stabilizing the helical fold and the appearance of a contact trace perpendicular to the diagonal, between residues K34 to G54. Additionally, new contacts are formed between the former  $\alpha$ -helical region and the flexible loop region, primarily involving



**Fig. 7.12.:** Contact map for the minima in the FES of the SH3 domain obtained from REMD simulations. The lower triangular matrix shows the contacts present during the simulation, while the upper triangular matrix indicates the difference from the initial fold. In the latter, positive rates (red) denote the formation of new contacts, while negative rates (blue) indicate the loss of contacts present in the initial structure. Two residues are considered to be in contact if any pair of atoms from the separate residues are within a distance of 6 Å.

residues V38 and E19. This structural refolding resembles the changes observed in the conventional MD simulation at pH 7, as further supported by the projection of the conventional MD trajectory onto the FES shown in Figure A1. Together with the high-energy barrier at  $T=300$  K of that minimum, we conclude this minimum must be off pathway for the unfolding and thus the subsequent amyloid formation.

Minima (3) and (4) are both directly connected to the global minimum (1) and represent the initial steps in the unfolding of the  $\alpha$ -helical region. Unlike minimum (2), this region remains disordered. In minimum (3), new contacts resemble a perpendicular trace to the diagonal but are not stable enough to facilitate the formation of  $\beta$ -sheets. Additionally, the disordered region around A39 forms a contact with the globular core region between F69 and V74, preventing the protrusion and refolding of the former  $\alpha$ -helical region. Minimum (4) exhibits fewer contacts between the helical and core regions, although it maintains a high structural similarity to minimum (3).

Finally, in minimum (5), the initially  $\alpha$ -helical region becomes completely disordered, as no clear pattern is visible in the corresponding region of the contact map. Additionally, contacts with the core region are broken, and new contacts are only formed sporadically. Here, the initially  $\alpha$ -helical region protrudes from the globular fold while being unfolded, probably marking the first step towards further unfolding.

#### 7.5.4 Discussion

Even in the REMD simulation of the SH3 domain at acidic pH, we did not observe complete unfolding at  $T=300$  K. The REMD simulation primarily captures early unfolding events, as it was terminated before significant divergence could occur, due to the lack of exchanges between high-temperature unfolded replicas and low-temperature mostly folded replicas.

At pH 2, the SH3 domain exhibits a flat FES with several competing minima at  $T=300$  K, which merge into one large basin at higher temperatures. At pH 7, the FES shows a single minimum across the entire temperature range, corresponding to the natively folded structure. Thus, the simulations indicate a similar loss of structural stability of the SH3 domain at acidic pH 2, also observed experimentally. As previously noted, the  $\alpha$ -helical region of the initially folded structure seems crucial in the early stages of the unfolding process, as here the most structural changes occur. Consequently, the first two principal components predominantly describe the motion of this region, while the rest of the protein remains relatively stable.

Examining the distinct minima in the FES at pH 2 reveals structures towards further unfolding. The minimum separated by an infinitely high energy barrier from the main basin, thus normally inaccessible at this temperature, probably hinders further unfolding due to its refolding into a  $\beta$ -hairpin structure, which stabilizes the  $\beta$ -sheet core structure. Therefore, we identify the states towards the bottom of the FES as pathway to complete unfolding, where the  $\alpha$ -helical region unfolds into a protruding disordered region.

## 7.6 Fibril Unfolding

The formation of mature amyloid fibrils typically takes multiple hours or even days, which are timescales not accessible by MD simulations. However, studying the reverse process of unfolding a fibril structure is feasible. The unfolding process can identify which contacts in the fibril are the weakest, as they are broken first, providing insights into the aggregation pathway itself.

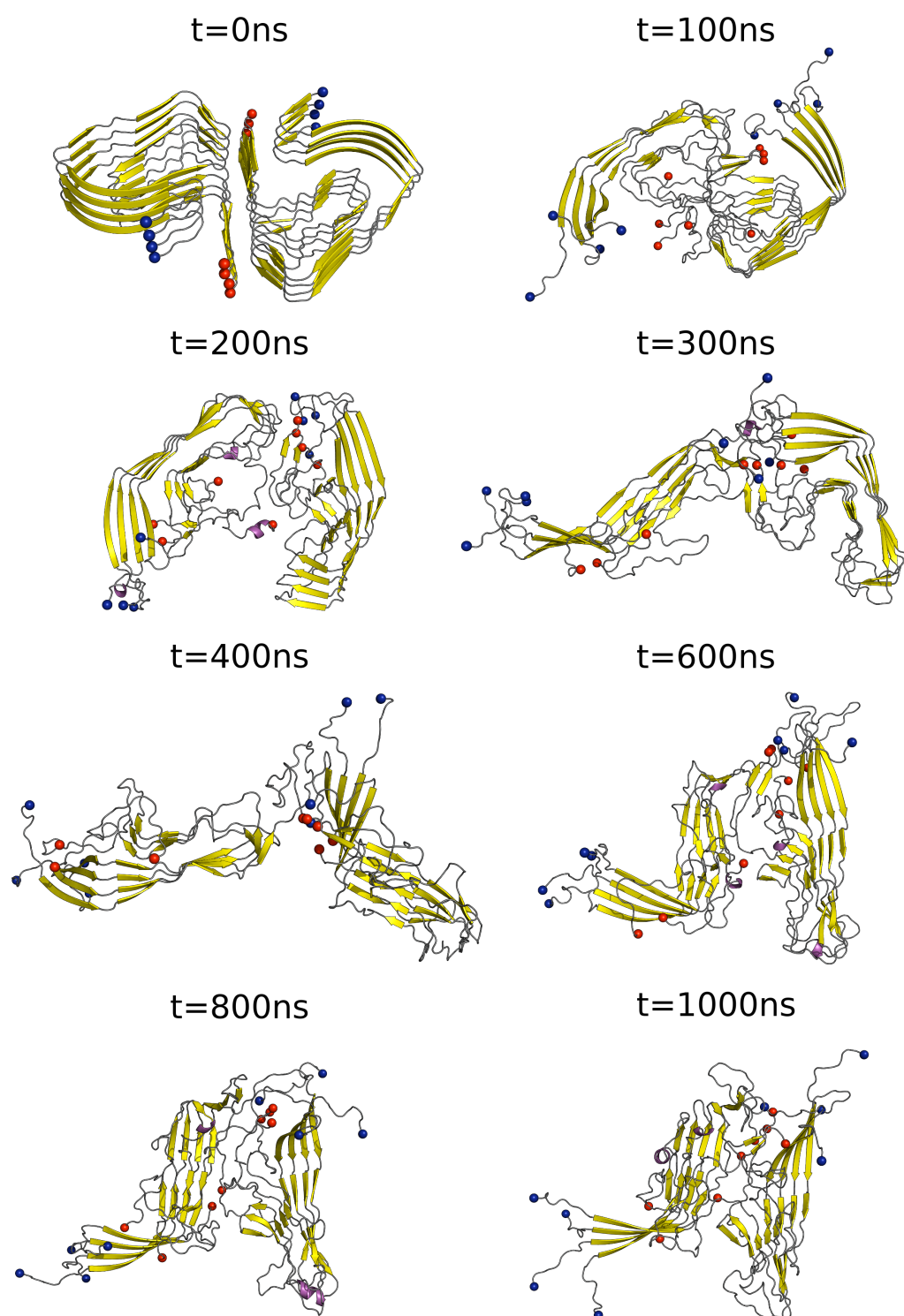
### 7.6.1 Simulation Methods

The system was prepared following the standard protocol for MD simulations, using mostly the same parameters as described for the monomer simulations in section 7.3.1. Only the temperature was set to  $T=400$  K to facilitate unfolding of the fibril. As starting structure we used the fibril structure obtained by Röder et al. (PDB:6R4R) [185]. The complex is made up of 8 monomers, which are ordered into two intertwined protofilaments sharing an interface and are stacked in four layers (see Figure 7.13 at  $t = 0$ ). Note that the sequence of the natively folded structure discussed in sections 7.3, 7.4, and 7.5 differs slightly from the fibril structure analyzed here. Specifically, the fibril structure includes three additional residues at the N-terminus (M1, S2, and A3) but is two residues shorter at the C-terminus, lacking G78 and R79 (see Figure 1.7 in chapter 1.6).

### 7.6.2 Results

Figure 7.13 shows the fibril structure at several point during the simulation. The initial fibril structure is composed of two protofibrils, aligned in a point-symmetrical manner relative to their interface. Within each protofibril (referred to as "sheets"), the protein chains are stacked in register in a parallel alignment, forming a rigid  $\beta$ -sheet structure. This parallel alignment is depicted as a trace along the diagonal in the initial intra-sheet contact map in Figure 7.14. Within each protofibril, the SH3 domain adopts a hairpin structure, observable as a trace perpendicular to the diagonal in both intra-protein and intra-sheet contacts. These contacts, highlighting the antiparallel alignment, occur within the protein chain and between neighboring proteins in the sheet. The slight twist of the fibril shifts these antiparallel intra-protein interactions to be visible in the intra-sheet contact map. Thus, the N-terminal residues of one protein form primarily contacts with the C-terminal of their next neighbour that further stabilize the protofibril. The interface between the two protofibrils involves only a few contacts, specifically between the two C-terminal regions spanning from G67 to T72, as well as hydrophobic interactions between the region from L37 to F42 and the C-terminal region from F69 to I76. The latter contact involves the central protein region, which is associated with the  $\alpha$ -helical region in the native folded structure.

After 100 ns, the fibril structure looks still relatively stable, however the N-terminal region within the sheets dissociate and the hairpin structure opens slightly. However, already in the initial structure the stacked N-terminal regions were only partly



**Fig. 7.13.:** Fibril structure of the SH3 domain over time. Snapshots from a  $1\ \mu\text{s}$  simulation at  $T=400\ \text{K}$ . The N-/C-termini are shown as blue/red spheres respectively.

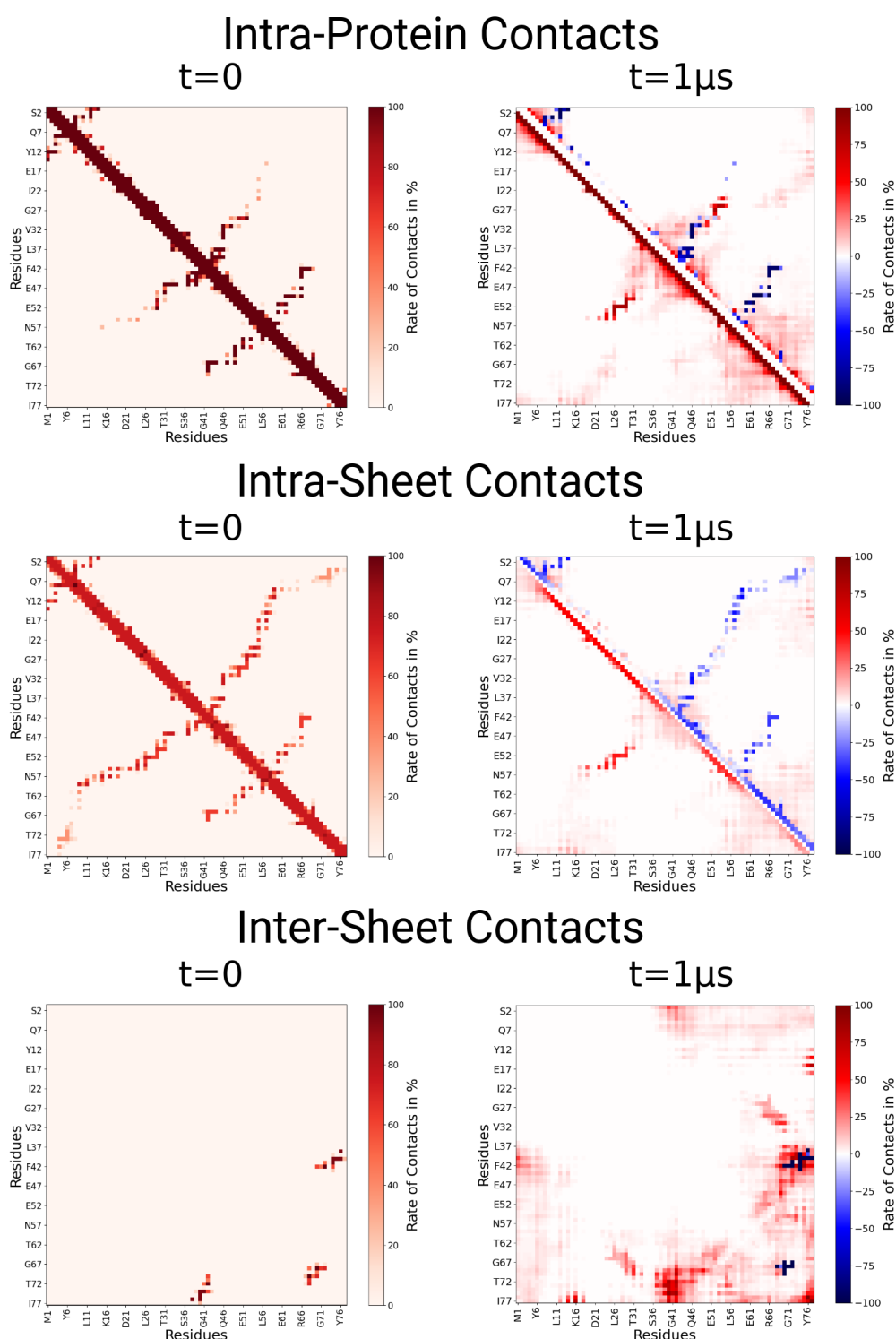
identified to form  $\beta$ -sheets, thus a flexible structure seems likely even in the fibril structure. Subsequently, the two protofibrils dissociate and are almost fully separated at  $T=400$  ns, while the hairpin structure and most of the  $\beta$ -sheets remain intact. Towards the end of the simulation, the two protofibrils reassemble into a complex with a new interface.

Throughout the simulation, the hairpin structure of the protofibrils, stabilized by inter-protein  $\beta$ -sheets, remained largely intact. However, an analysis of the intra-protein/sheet contacts, as shown in Figure 7.14, reveals key structural changes. Both contact maps show a significant loss of antiparallel contacts that support the hairpin structure. Some of these contacts shifted from neighboring proteins in the intra-sheet contacts to intra-protein contacts, particularly the contacts between the regions from I22 to T31 and from P50 to W55. This dissociation resolves the twist in the fibril structure, resulting in a newly stacked sheet of hairpin structures primarily stabilized by intra-protein contacts in the central region of the protein chain. At the termini, the antiparallel contacts between neighboring proteins do not reform, leading to an open hairpin structure with flexible termini.

A loss of intra-sheet contacts among the stacked proteins is also observed, shown as a reduction in contacts along the diagonal in the contact map. This primarily affects the region from L40 to P50, as well as the terminal regions from M1 to R9 and G58 to I77. The most stable region, which remains intact throughout the simulation, spans from L11 to D33. Notably, this region is associated with the disordered loop in the native fold, containing a high number of aspartic and glutamic acids that are negatively charged at pH 7 and become neutral at pH 2.

### 7.6.3 Discussion

The simulation of the SH3-domain fibril at  $T=400$  K and pH 2 revealed partial unfolding of the fibrillar structure, highlighting key structural features. The initial contacts to break were those forming the interface between the two protofibrils, supporting the widely accepted view that these interactions are part of a fibril-catalyzed secondary nucleation reaction [38, 39]. Additionally, the involvement of the  $\alpha$ -helical region from the native fold in the interface supports previous observations in section 7.4 that this region plays a role in facilitating complex formation of the SH3 domain. While the overall hairpin structure of the protofibrils stayed intact, by unbinding from the interface, the twist within the structure was reversed, shifting the primary interaction partner in the U-bend from the neighbouring protein in the sheet, to intra-protein contacts. Furthermore, the termini become more flexible with increasing simulation time and unfold the  $\beta$ -sheet structure. Finally, we were able to identify the region spanning from residue L11 to N33 as the most stable region of the  $\beta$ -sheet structure within the protofibril, as this is the only region exhibiting almost no breakage of contacts after 1  $\mu$ s simulations. Notably, this region is associated with the disordered loop region in the natively folded structure.



**Fig. 7.14.: Contact maps of SH3 fibril unfolding.** The contact matrices are categorized into intra-protein contacts (within single protein chains), intra-sheet contacts (formed between proteins within the same protofibril forming  $\beta$ -sheets), and inter-sheet contacts (formed at the interface between two protofibrils that constitute the amyloid fibril). The contact map of the initial fibril and after  $1\mu s$  are shown. In the latter, the lower triangular matrix displays the contacts present during the simulation, while the upper triangular matrix shows the differences from the initial fibril. Positive rates (red) indicate the formation of new contacts, while negative rates (blue) represent the loss of contacts present in the initial structure. Two residues are considered in contact if any pair of atoms from separate residues are within a distance of  $6\text{ \AA}$ .

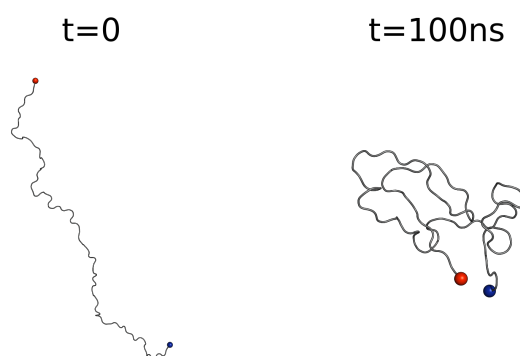
## 7.7 Early Folding Events

Thus far, we have examined the unfolding of both the native SH3 domain structure and its fibrillar form, identifying key structural features that destabilize the native structure and stabilize the amyloid fibril. In this section, we shift our focus to the aggregation pathway starting from the initially unfolded state rather than the reverse process studied before. Experimental studies of the SH3 domain indicate that upon reducing the pH to acidic conditions (approximately pH 2), the SH3 domain first unfolds and subsequently aggregates into amyloid fibrils. This observation is further supported from a structural perspective, as the secondary structure of the SH3 domain in its native state differs significantly from that in the amyloid fibril (Figure 1.7). To investigate the early stages of the aggregation pathway, we performed simulations of the SH3 monomer, starting from a completely disordered extended state, and monitored the propensity for intra-protein contact formation.

### 7.7.1 Simulation Methods

The system was prepared following the standard protocol for MD simulations, using the same parameters as described for the monomer simulations in section 7.3.1. The initial structure was modeled as a fully extended protein chain of the same sequence as presented in the natively folded structure (PDB:1PKS), leading to a large simulation box of 20,000 nm<sup>3</sup> with 1,900,000 atoms.

To reduce computational costs, we conducted an initial short simulation of 100 ns, allowing the extended protein chain to collapse into a globular, yet unfolded, structure as shown in Figure 7.15. After this initial simulation, the protein was resolvated in a smaller simulation box, counter ions were added, and the system was equilibrated with position restraints. This process significantly reduced the system size to approximately 500,000 atoms, thereby decreasing the computational cost drastically. The simulation was then extended for an additional 1  $\mu$ s. This protocol was applied to three simulations at pH 2 and pH7, resulting in a total simulation time of 3.3  $\mu$ s for each condition.



**Fig. 7.15.: Initial structural collapse of the disordered SH3 domain.** The structure is shown in the beginning ( $t=0$ ) and end ( $t=100$  ns) of the simulation. The N-/C-termini are shown as blue/red spheres respectively.



## 7.7.2 Results

Figure 7.16 shows the combined contact maps of the three folding trajectories for pH 2 and pH 7, respectively. To facilitate comparison of contacts formed during early stages of the folding process with those in the native fold and fibril structure, we show the contact map of these structures together with difference contact maps at both pH values. In the difference contact maps, the lower triangular matrix displays the contacts observed during the simulation, while the upper triangular matrix shows the differences to the reference structure.

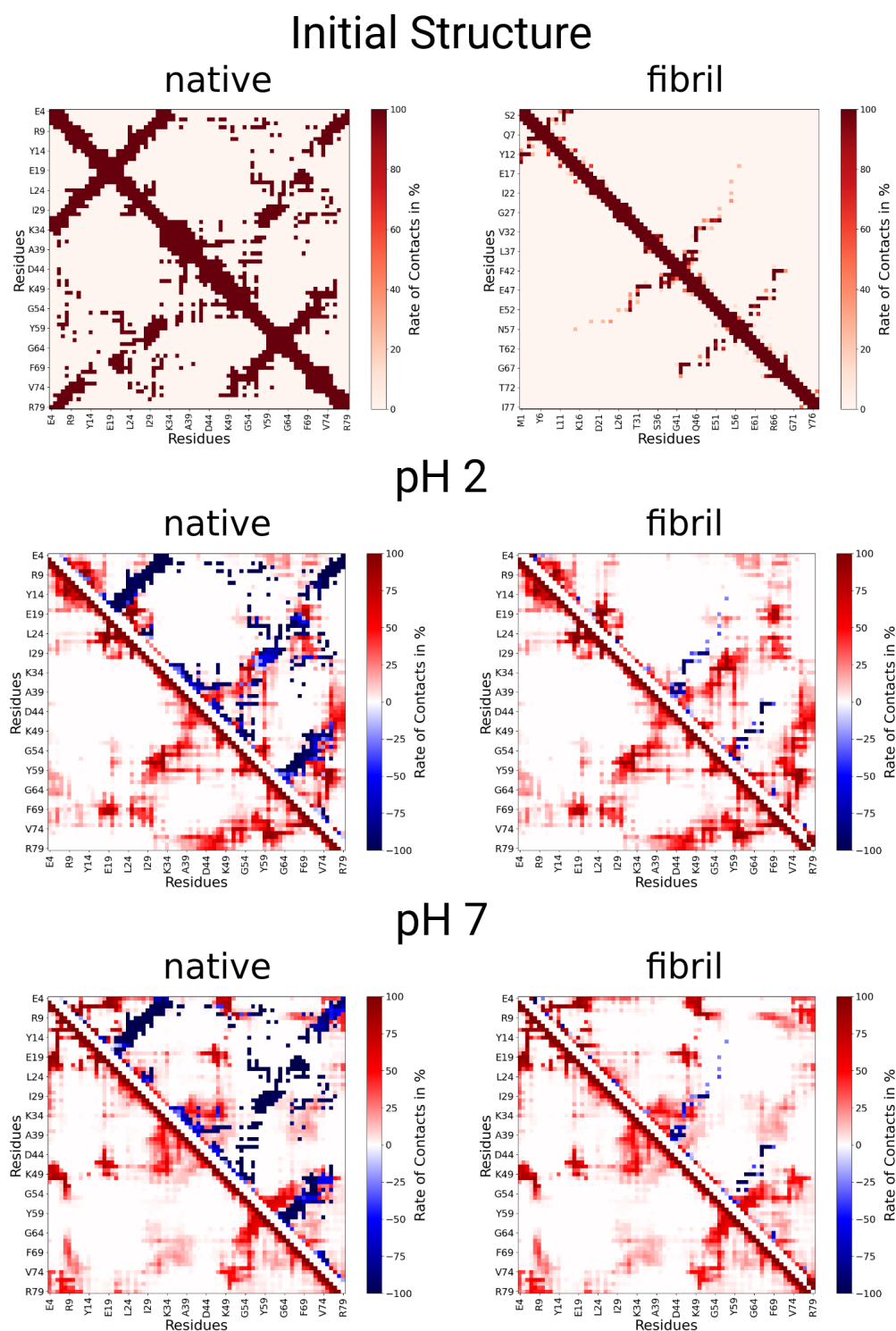
Comparing the contacts formed at pH 2 with those at pH 7 (focusing on the lower triangular matrix of the contact maps) reveals significant differences in their contact propensity. At pH 2 contacts are observed between the regions around E19 and P70 as well as T31 and N57, which do not form at pH 7. These contacts roughly correspond to the contacts formed between the  $\beta$ -sheet core regions and the flexible loop-region (see Figure 7.6). However, the hairpin structure typically associated with this  $\beta$ -sheet does not appear to fully form at pH 2.

Additionally, several contacts observed at pH 2 do not align with those in the reference structures. Notably, there is a high interaction propensity between the C-terminal region and the helical region in the native fold, as well as with the region around N60. Furthermore, the contact between regions around K15 and G25 does not correspond to any reference structure and appears to involve interactions within the flexible loop region in the native fold. Lastly, the N-terminal region shows early signs of forming a hairpin motif, akin to that in the fibrillar structure, although it occurs in a different region. Here, the formation of a hairpin in a similar region to the fibril structure may be hindered by the slight truncation of the N-terminus.

In the simulation at pH 7, the N-terminus does not form a hairpin structure. Instead, it interacts with three distinct regions of the protein chain: the region around E20, the region around K49, and the C-terminal region. None of these interactions align with those observed in the reference structures. The interaction of the N-terminal region with the region around E20 and the C-terminal may be promoted by electrostatic interactions (compare with Figure 1.7). Another notable contact is observed between the regions around R18 and E47, driven by electrostatic interactions. This includes the interaction between E19 and K49, which is absent in both reference structures. Finally, the hairpin structure of the  $\beta$ -sheet core in the native fold begins to emerge in the region spanning from P50 to E61. Although the contacts appear slightly shifted, resulting in a trace parallel to that of the original fold, this may represent an early stage in the development of this specific structural motif.

## 7.7.3 Discussion

Studying the early folding at pH 2 and pH 7 by observing the intra-molecular interactions during the conformational transition of the protein chain from an extended to a globular structure, we identified key differences between the two conditions. In both cases, the contact maps reveal minimal overlap with the reference struc-



**Fig. 7.16.: Contact map of SH3 early aggregation events.** The contact maps of the folding trajectories is compared to the contacts map in the SH3 domain fibrillar and natively folded state. Here, the lower triangular matrix displays the contacts present during the simulation, while the upper triangular matrix shows the differences from the initial fibril. Positive rates (red) indicate the formation of new contacts, while negative rates (blue) represent the loss of contacts present in the reference structure. Two residues are considered in contact if any pair of atoms from separate residues are within a distance of 6 Å.

tures of the native fold or the fibril structure. The simulation at pH 7 underscores the predominance of electrostatic interactions during the early stages of folding. However, at short timescales as simulated here, these interactions may also lead to misfolding. Consequently, at pH7, the SH3 domain exhibits a rugged FES characterized by numerous competing low-energy local minima, which act as kinetic traps. Therefore, the folding process to achieve the natively folded structure may be slow or could require additional interaction partners to facilitate the folding pathway. At pH 2, where all negatively charged residues are protonated, these electrostatic interactions cease to exist. Instead contacts which stabilize the globular form of the native fold start appearing, which may also be an early stage of a hairpin motif found in the fibril structure. Both simulations show a high propensity of forming a short hairpin structure in the N-terminal region, which showed a high degree of flexibility in previous regions as well.

## 7.8 Conclusion

This chapter presents a summary of the SH3 domain's structural dynamics under various conditions, highlighting key differences between monomeric, dimeric, and fibrillar forms, as well as the effects of pH on their behavior. Figure 7.17 offers an overview of the structural motifs central to each simulation study.

Conventional MD simulations revealed that the SH3 domain's robust  $\beta$ -sheet core structure remains largely intact, even at pH 2, with only the initial  $\alpha$ -helical region showing significant structural changes. The  $\alpha$ -helical region primarily unfolds at pH 2 and forms a  $\beta$ -hairpin at pH 7. These results suggest that either the simulation time might be too short or protonation alone is insufficient to destabilize the  $\beta$ -sheet structures, indicating that additional factors, such as interaction partners or excess ions, may play a crucial role in the unfolding process. For example, the presence of  $H_3O^+$  in experiments is typically ignored in MD simulations.

The dimer simulations demonstrated a low interaction propensity between folded SH3 monomers, with significant inter-molecular interactions observed in only one of the three simulations. These interactions were primarily mediated by the  $\alpha$ -helical region, which remained stable across all dimer simulations, emphasizing its potential role in oligomerization. Despite not replicating the unfolding observed at acidic pH in experiments, the simulations underscore the structural importance of the  $\alpha$ -helical region in the SH3 domain. The low interaction propensity between the two SH3 monomers suggests that their aggregation leading to amyloid structures must proceed through prior unfolding.

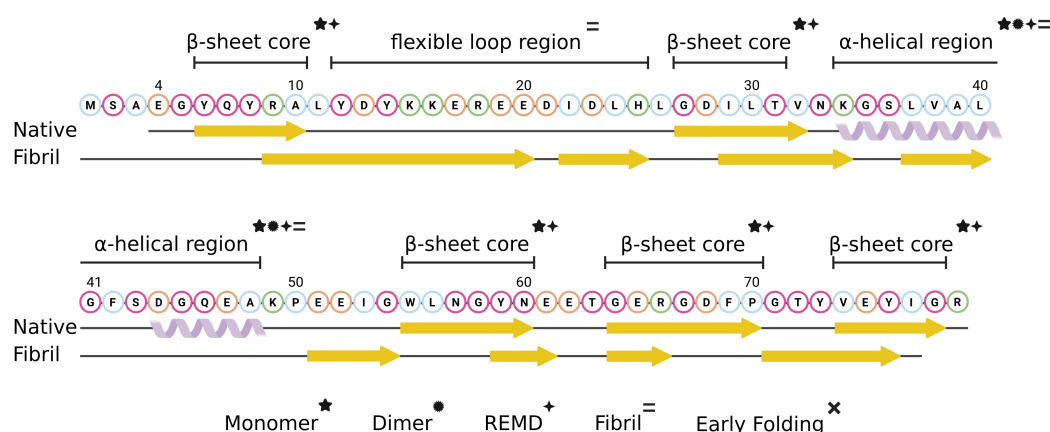
The REMD simulations predominantly captured early unfolding events, constrained by limited exchanges between high-temperature unfolded replicas and low-temperature folded replicas. At pH 2, the SH3 domain exhibited a flat FES with multiple competing minima at  $T = 300$  K, merging into a single basin at higher temperatures. In contrast, at pH 7, the FES displayed a single minimum corresponding to the natively folded structure. Analyzing the different minima of the FES at pH 2 suggests that the  $\alpha$ -helical region plays a critical role in the early stages of unfolding, with the rest of the protein remaining relatively stable at  $T = 300$  K.

The simulation of the SH3-domain fibril at  $T = 400$  K and pH 2 provided insights into the partial unfolding of the fibrillar structure. The initial breakage of contacts at the protofibril interface supports the view that these interactions are part of a secondary aggregation process. Additionally, the involvement of the  $\alpha$ -helical region in the interface aligns with previous observations of its role in SH3 domain complex formation. Despite the overall retention of the hairpin structure within the protofibrils, the unbinding from the interface reversed the twist within the fibril structure, shifting the primary interactions to intra-protein contacts. The termini became more flexible over time, leading to partial unfolding of the  $\beta$ -sheet structure. The region spanning residues L11 to N33 emerged as the most stable within the

$\beta$ -sheet structure, showing minimal loss of contacts, and is notably associated with the disordered loop region in the native folded structure.

Early folding effects were studied by observing intra-molecular interactions during the conformational transition from an extended to a globular structure at pH 2 and pH 7. Significant differences were identified between the two conditions. At pH 7, electrostatic interactions dominate the early stages of folding, but lead to misfolding on short timescales as a result of a rugged FES with numerous competing local minima that act as kinetic traps. At pH 2, where electrostatic interactions are considerably reduced due to protonation of negatively charged residues, contacts begin to form, which potentially mark the early stages of a hairpin motif observed in the fibril structure. Both simulations revealed a strong tendency to form a short hairpin structure in the N-terminal region, which can also be observed in the fibril structure but was also previously identified as highly flexible.

In summary, we have identified the  $\alpha$ -helical region as a key player in the early stages of the SH3 domain unfolding pathway, with its functional role likely related to mediating inter-protein interactions. Additionally, the disordered loop region was found to be the most stable sequence within the  $\beta$ -sheet structure of the fibril, suggesting it forms early in the aggregation process, potentially facilitated by the formation of a hairpin structure in the disordered state of the SH3 domain. Finally, our comprehensive analysis of contact propensities at acidic and neutral pH provides a foundation for future studies, potentially in concert with complementary NMR measurements, to verify our findings and further elucidate the full aggregation pathway of the SH3 domain.



**Fig. 7.17.: Structural motives and sequence of the bovine PI3K-SH3 domain.** Each structural motif is marked according to the simulation studies providing insights into its structure or function. Amino acids are color-coded by primary properties: orange for acidic, green for basic, pink for polar, and blue for non-polar/hydrophobic residues. The native fold and amyloid fibril secondary structures are displayed below the sequence.



In this thesis, we investigated the free energy surface (FES) and conformational switching of intrinsically disordered proteins (IDPs), as well as their aggregation into amyloid fibrils. These different aspects of IDPs were studied considering three different proteins: A $\beta_{42}$ , a segment of the parathyroid hormone (PTH<sub>25–37</sub>), and bovine phosphatidylinositol-3-kinase (PI3K-SH3). We utilized extensive all-atom molecular dynamics (MD) simulations along with enhanced sampling techniques, such as replica-exchange molecular dynamics (REMD), to thoroughly explore the conformational space of all systems. For the analysis, we combined established methodologies with novel approaches, including our own software package ATRANET, to investigate the configurational space of IDPs. Additionally, we calculated FESs and analyzed transition timescales using first passage time (FPT) analysis. This comprehensive approach allowed us to explore various stages of the amyloid pathway, from early folding events to oligomerization and fibrillization, as well as elucidate the conformational transitions that are closely coupled to the aggregation process.

Following the introduction of the thesis topic in Chapter 1 and the theoretical methods in Chapter 2, in Chapter 3 we utilized transition networks (TNs) to explore the configurational space of IDPs from MD simulations, as demonstrated in the case of A $\beta_{42}$  interacting with POPC lipids. The ATRANET software package, developed for this purpose and available on GitHub<sup>1</sup>, facilitates the construction of TNs using predefined descriptor functions from MD data. By comparing two TNs constructed with different descriptor sets, one using three descriptors ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}$ ) and the other incorporating a fourth descriptor ( $N_{p-l}$ ) to account for peptide-lipid interactions, we identified key aspects of the TN analysis. These include: (i) the node size represents the stability of conformations, (ii) the spatial layout of the nodes reflects the phase-space distances between conformations, and (iii) clustering within the TN highlights significant conformational transitions. While TNs offer valuable structural insights, they are heavily influenced by the choice of descriptors, emphasizing the importance of aligning descriptor selection with the research objective.

Subsequently, we applied the TN methodology to A $\beta_{42}$  to study its disorder-to-order transition upon binding. In Chapter 4, we compared the TNs of A $\beta_{42}$  as a single peptide in solution, A $\beta_{42}$  in the presence of a glycosaminoglycan (GAG), and A $\beta_{42}$  in complex with three POPC lipids. The TN of A $\beta_{42}$  in solution exhibited characteristics typical of IDPs, with many low-population states connected by frequent transitions. The most populated states corresponded to random-coil conformations of varying spatial expansion. In contrast, the TN of A $\beta_{42}$  in the presence of a GAG demonstrated a shift towards more compact,  $\beta$ -sheet-rich structures. We uncovered that the

<sup>1</sup><https://github.com/strodel-group/ATRANET>

formation of this stable  $\beta$ -hairpin structure is driven by changes in local sodium ion concentration, rather than originating from direct interactions with the GAG. The TN of A $\beta$  in complex with POPC lipids revealed transitions towards both  $\alpha$ -helical and  $\beta$ -sheet structures, driven by hydrophobic peptide-lipid interactions. These interactions resulted in a rougher free energy landscape with multiple distinct minima.

We employed structural clustering in DRID space for state assignment and visualized the FESs as disconnectivity graphs. These graphs are a projection-free representation of the FES based on the connectivity between minima.

We applied structural clustering in DRID space for state assignment, visualizing the resulting FESs as disconnectivity graphs. These graphs represent the connectivity within an energy landscape by grouping minima at various energy thresholds if they are connected by transition states below these thresholds. Minima are split at thresholds where their transition state surpasses the specified energy level, creating a tree-like structure.

To further investigate the oligomerization pathway of A $\beta_{42}$ , in Chapter 5 we compared the FES of the A $\beta_{42}$  monomer and dimer, focusing on early aggregation events. Here, we employed structural clustering in DRID space for state assignment and visualized the FESs as disconnectivity graphs. These graphs are a projection-free representation of the FES based on the connectivity between minima. Significant differences were observed between the two FESs. The FES of A $\beta_{42}$  in solution exhibits a structurally inverted folding funnel, with disordered states at the bottom of the FES. In contrast, the dimer FES resembles that of a folded protein, where the global minimum corresponds to ordered structures forming a  $\beta$ -hairpin. FPT analysis revealed the fastest transition pathway from the disordered state to the ordered  $\beta$ -hairpin in the dimer FES, emphasizing the critical role of salt-bridge formation between D23 and K28 and cooperative interactions between hydrophobic regions of the A $\beta_{42}$  peptides. Furthermore, we included the FESs of A $\beta_{42}$  with a cluster of three POPC lipids and A $\beta_{42}$  in the presence of a GAG. Both of which exhibit significant similarities to the dimer FES, with a similar  $\beta$ -hairpin state at their global minimum.

Our findings demonstrate the significant conformational heterogeneity of A $\beta_{42}$ , which explores extensive regions of the FES even on short timescales in MD simulations due to relatively low energy barriers between states. Upon binding to an interaction partner, the FES shifts, promoting a transition to more structured conformations dependent on the partner involved. The interaction-specific conformational changes may elucidate the diverse fibril polymorphisms observed in humans [152] and in various mouse models [40], potentially influenced by the agent initiating the aggregation process. These findings contribute to a deeper understanding of Alzheimer's disease pathology and may guide the development of targeted therapeutic strategies.

In Chapter 6, we investigated amyloid fibrillization using the PTH<sub>25–37</sub> peptide. By incorporating the photoswitch 3-[(4-aminomethyl)phenyl]diazonyl}benzoic acid (AMPB) into the sequence, we achieved precise control over reversible peptide fibril-



lization. Through a combined effort of *in silico* design, synthesis, and experimental validation, we demonstrated that fibrillization depends on the position of the mutation within the peptide sequence. Extensive MD simulations, focusing on PTH<sub>25–37</sub> and the P4 mutant (where the mutation is located in the central peptide region), revealed that AMPB promotes  $\pi$ - $\pi$  interactions, enhancing aggregation. Additionally, we observed that  $\beta$ -strands in P4 fibrils adopt an antiparallel arrangement, in contrast to the unmodified peptide. From these results, we constructed a fibril model for both PTH<sub>25–37</sub> and the P4 mutant, which agree with solid-state NMR data and proved stable in MD simulations. This study presents a potential method for controlling reversible amyloid aggregation and highlights applications in drug delivery using light-triggered peptide systems.

Chapter 7 presents a comprehensive simulation study of the PI3K-SH3 domain, focusing on pH-dependent unfolding and subsequent aggregation at acidic pH. In addition to conventional MD simulations of the folded monomer and dimer, we conducted REMD simulations and simulated both the full fibril and unfolded dimer. Our analysis identified the  $\alpha$ -helical region as a critical agent in the early stages of SH3 domain unfolding, potentially mediating inter-peptide interactions. Furthermore, we identified the disordered loop region in the native fold as the most stable sequence within the  $\beta$ -sheet fibril structure, indicating it forms early during aggregation, potentially facilitated by the formation of a hairpin structure in the disordered state. Finally, a detailed analysis of contact propensities at both acidic and neutral pH provides a foundation for future studies, potentially alongside NMR experiments, to confirm our findings and further elucidate the full aggregation pathway of the SH3 domain.

In summary, this study provides valuable insights into the roles of IDPs, with a focus on amyloid aggregation. We employed a diverse range of simulation and analysis methods to investigate IDPs of varying sizes, from the 12-residue PTH<sub>25–37</sub> to the 75-residue SH3 domain. Specifically, we explored the configurational space of IDPs using A $\beta$ <sub>42</sub> as a model system, focusing on the disorder-to-order transition and the associated changes in the FES in response to different interaction partners. Additionally, we investigated the pH-dependent aggregation of the SH3 domain, which becomes disordered at acidic pH. Finally, we studied the full aggregation pathway of PTH<sub>25–37</sub>, culminating in the successful development of fibril models.

In future work, expanding the analysis of FESs to various IDPs would provide insights into the unique and shared features of these landscapes that may influence amyloid disease progression. A promising direction would involve testing the hypothesis that the topography and steepness of energy landscapes may be predictive of the speed at which the diseases develop. For instance, in Alzheimer's disease, a prolonged latency precedes symptom onset by approximately 10–20 years, followed by a gradual decline, suggesting a relatively shallow and less deterministic energy landscape. In contrast, prion diseases like Creutzfeldt-Jakob disease exhibit rapid progression [197, 198], potentially indicative of a steeper, more restrictive energy

landscape that favors irreversible aggregation. Characterizing the energy landscape of prion proteins, although challenging due to their structural stability and greater sequence length, would enable comparative analyses. This could ultimately illuminate factors governing the progression rates of different amyloid diseases and identify landscape features critical to aggregation pathways.

# Bibliography

1. Fischer, E. Einfluss der Configuration auf die Wirkung der Enzyme. *Berichte der deutschen chemischen Gesellschaft* **27**, 2985–2993 (Oct. 1894) (cit. on p. 1).
2. Senior, A. W., Evans, R., Jumper, J., *et al.* Improved protein structure prediction using potentials from deep learning. en. *Nature* **577**, 706–710 (Jan. 2020) (cit. on p. 1).
3. Callaway, E. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. en. *Nature* **588**, 203–204 (Dec. 2020) (cit. on p. 1).
4. Jumper, J., Evans, R., Pritzel, A., *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (July 2021) (cit. on p. 1).
5. Strodel, B. Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins. *Journal of Molecular Biology* **433**, 167182 (Oct. 2021) (cit. on pp. 1, 2, 6, 47).
6. Wright, P. E. & Dyson, H. Intrinsically disordered proteins in cellular signalling and regulation. *Nature Reviews Molecular Cell Biology* **16**, 18–29 (2014) (cit. on p. 1).
7. Dunker, A. K., Brown, C. J., Lawson, J. D., Iakoucheva, L. M. & Obradović, Z. Intrinsic Disorder and Protein Function. *Biochemistry* **41**, 6573–6582 (May 2002) (cit. on p. 1).
8. Yankner, B., Duffy, L. & Kirschner, D. A. Neurotrophic and neurotoxic effects of amyloid beta protein: reversal by tachykinin neuropeptides. *Science* **250** **4978**, 279–82 (1990) (cit. on p. 1).
9. Hardy, J. A. & Higgins, G. A. Alzheimer's Disease: The Amyloid Cascade Hypothesis. *Science* **256**, 184–185 (Apr. 1992) (cit. on p. 1).
10. Conway, K. A., Lee, S.-J., Rochet, J.-C., *et al.* Acceleration of oligomerization, not fibrillization, is a shared property of both  $\alpha$ -synuclein mutations linked to early-onset Parkinson's disease: Implications for pathogenesis and therapy. *Proceedings of the National Academy of Sciences* **97**, 571–576 (Jan. 2000) (cit. on pp. 1, 5).
11. Cooper, G., Willis, A., Clark, A., *et al.* Purification and characterization of a peptide from amyloid-rich pancreases of type 2 diabetic patients. *Proceedings of the National Academy of Sciences of the United States of America* **84** **23**, 8628–32 (1987) (cit. on p. 1).
12. Otzen, D. Functional amyloid. en. *Prion* **4**, 256–264 (Oct. 2010) (cit. on p. 1).
13. Bucciantini, M., Giannoni, E., Chiti, F., *et al.* Inherent toxicity of aggregates implies a common mechanism for protein misfolding diseases. *Nature* **416**, 507–511 (Apr. 2002) (cit. on p. 1).

14. Dror, R. O., Jensen, M. Ø., Borhani, D. W. & Shaw, D. E. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *Journal of General Physiology* **135**, 555–562 (May 2010) (cit. on p. 2).
15. Hsu, C. C., Buehler, M. J. & Tarakanova, A. The Order-Disorder Continuum: Linking Predictions of Protein Structure and Disorder through Molecular Simulation. *Scientific Reports* **10** (Feb. 2020) (cit. on p. 2).
16. Fatafta, H., Samantray, S., Sayyed-Ahmad, A., Coskuner-Weber, O. & Strodel, B. in *Dancing Protein Clouds: Intrinsically Disordered Proteins in the Norm and Pathology, Part C* 135–185 (Elsevier, 2021) (cit. on p. 2).
17. Shrestha, U. R., Smith, J. C. & Petridis, L. Full structural ensembles of intrinsically disordered proteins from unbiased molecular dynamics simulations. *Communications Biology* **4** (Feb. 2021) (cit. on p. 2).
18. Chu, X., Nagpal, S. & Muñoz, V. in *Protein Folding* 343–362 (Springer US, Nov. 2021) (cit. on p. 2).
19. Fatafta, H., Khaled, M., Kav, B., Olubiyi, O. O. & Strodel, B. A brief history of amyloid aggregation simulations. *WIREs Comput. Mol. Sci.* **14**, e1703 (2024) (cit. on p. 3).
20. Deiana, A., Forcelloni, S., Porrello, A. & Giansanti, A. Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell. *PLOS ONE* **14** (ed Permyakov, E. A.) e0217889 (Aug. 2019) (cit. on p. 3).
21. Chebaro, Y., Ballard, A. J., Chakraborty, D. & Wales, D. J. Intrinsically Disordered Energy Landscapes. *Scientific Reports* **5** (May 2015) (cit. on p. 4).
22. Receveur-Bréchet, V., Bourhis, J.-M., Uversky, V. N., Canard, B. & Longhi, S. Assessing protein disorder and induced folding. en. *Proteins* **62**, 24–45 (Jan. 2006) (cit. on p. 4).
23. Bourhis, J.-M., Johansson, K., Receveur-Bréchet, V., *et al.* The C-terminal domain of measles virus nucleoprotein belongs to the class of intrinsically disordered proteins that fold upon binding to their physiological partner. *Virus Research* **99**, 157–167 (Feb. 2004) (cit. on p. 4).
24. Robustelli, P., Piana, S. & Shaw, D. E. Mechanism of coupled folding-upon-binding of an intrinsically disordered protein. en. *J. Am. Chem. Soc.* **142**, 11092–11101 (June 2020) (cit. on p. 4).
25. Uversky, V. N., Gillespie, J. R. & Fink, A. L. Why are natively unfolded proteins unstructured under physiologic conditions? *Proteins: Structure, Function, and Genetics* **41**, 415–427 (2000) (cit. on p. 4).
26. Uversky, V. N. The alphabet of intrinsic disorder: II. Various roles of glutamic acid in ordered and intrinsically disordered proteins. en. *Intrinsically Disord. Proteins* **1**, e24684 (Jan. 2013) (cit. on p. 4).
27. Selkoe, D. J. & Hardy, J. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol. Med.* **8**, 595–608 (2016) (cit. on p. 5).
28. Pitschke, M., Prior, R., Haupt, M. & Riesner, D. Detection of single amyloid  $\beta$ -protein aggregates in the cerebrospinal fluid of Alzheimer's patients by fluorescence correlation spectroscopy. en. *Nat. Med.* **4**, 832–834 (July 1998) (cit. on p. 5).

29. Maji, S. K., Perrin, M. H., Sawaya, M. R., *et al.* Functional Amyloids As Natural Storage of Peptide Hormones in Pituitary Secretory Granules. *Science* **325**, 328–332 (July 2009) (cit. on p. 5).
30. C, P. *The State of the Art of Dementia Research: New Frontiers* London (UK), 2018 (cit. on p. 5).
31. Ferrone, F. in *Amyloid, Prions, and Other Protein Aggregates* 256–274 (Elsevier, 1999) (cit. on p. 5).
32. Wetzel, R. Kinetics and Thermodynamics of Amyloid Fibril Assembly. *Accounts of Chemical Research* **39**, 671–679 (July 2006) (cit. on p. 5).
33. Jarrett, J. T. & Lansbury, P. T. Seeding “one-dimensional crystallization” of amyloid: A pathogenic mechanism in Alzheimer’s disease and scrapie? *Cell* **73**, 1055–1058 (June 1993) (cit. on p. 5).
34. Schmit, J. D., Ghosh, K. & Dill, K. What Drives Amyloid Molecules To Assemble into Oligomers and Fibrils? *Biophysical Journal* **100**, 450–458 (Jan. 2011) (cit. on p. 6).
35. Šarić, A., Chebaro, Y. C., Knowles, T. P. J. & Frenkel, D. Crucial role of nonspecific interactions in amyloid nucleation. *Proceedings of the National Academy of Sciences* **111**, 17869–17874 (Dec. 2014) (cit. on p. 6).
36. Khaled, M., Rönnbäck, I., Ilag, L. L., *et al.* A Hairpin Motif in the Amyloid- $\beta$  Peptide Is Important for Formation of Disease-Related Oligomers. *J. Am. Chem. Soc.* **145**, 18340–18354 (2023) (cit. on pp. 6, 9, 50, 53).
37. Baskakov, I. V. Branched chain mechanism of polymerization and ultrastructure of prion protein amyloid fibrils. *The FEBS Journal* **274**, 3756–3765 (July 2007) (cit. on p. 6).
38. Törnquist, M., Michaels, T. C. T., Sanagavarapu, K., *et al.* Secondary nucleation in amyloid formation. *Chemical Communications* **54**, 8667–8684 (2018) (cit. on pp. 6, 94).
39. Cohen, S. I. A., Vendruscolo, M., Dobson, C. M. & Knowles, T. P. J. Nucleated polymerization with secondary pathways. II. Determination of self-consistent solutions to growth processes described by non-linear master equations. *The Journal of Chemical Physics* **135** (Aug. 2011) (cit. on pp. 6, 94).
40. Zielinski, M., Peralta Reyes, F. S., Gremer, L., *et al.* Cryo-EM of A $\beta$  fibrils from mouse models find tg-APP<sup>ArcSwe</sup> fibrils resemble those found in patients with sporadic Alzheimer’s disease. *Nature Neuroscience* **26**, 2073–2080 (Nov. 2023) (cit. on pp. 6, 46, 104).
41. Owen, M. C., Gnutt, D., Gao, M., *et al.* Effects of in vivo conditions on amyloid aggregation. *Chemical Society Reviews* **48**, 3946–3996 (2019) (cit. on p. 6).
42. Fatafta, H., Kav, B., Bundschuh, B. F., Loschwitz, J. & Strodel, B. Disorder-to-order transition of the amyloid- $\beta$  peptide upon lipid binding. *en. Biophys. Chem.* **280**, 106700 (Jan. 2022) (cit. on pp. 6, 42, 44, 56, 61).
43. Schlenzig, D., Röncke, R., Cynis, H., *et al.* N-Terminal pyroglutamate formation of A $\beta$ 38 and A $\beta$ 40 enforces oligomer formation and potency to disrupt hippocampal long-term potentiation. *en. J. Neurochem.* **121**, 774–784 (June 2012) (cit. on p. 6).

44. Schieb, H., Kratzin, H., Jahn, O., *et al.* Beta-amyloid peptide variants in brains and cerebrospinal fluid from amyloid precursor protein (APP) transgenic mice: comparison with human Alzheimer amyloid. *en. J. Biol. Chem.* **286**, 33747–33758 (Sept. 2011) (cit. on p. 6).
45. Faller, P., Hureau, C. & Berthoumieu, O. Role of metal ions in the self-assembly of the Alzheimer's amyloid- $\beta$  peptide. *en. Inorg. Chem.* **52**, 12193–12206 (Nov. 2013) (cit. on p. 6).
46. Kaye, R., Head, E., Thompson, J. L., *et al.* Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis. *Science* **300**, 486–489 (Apr. 2003) (cit. on p. 6).
47. Walsh, D. M., Klyubin, I., Fadeeva, J. V., *et al.* Naturally secreted oligomers of amyloid  $\beta$  protein potently inhibit hippocampal long-term potentiation in vivo. *Nature* **416**, 535–539 (Apr. 2002) (cit. on pp. 6, 9).
48. Shankar, G. M., Li, S., Mehta, T. H., *et al.* Amyloid- $\beta$  protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nature Medicine* **14**, 837–842 (June 2008) (cit. on p. 6).
49. Kawahara, M. Neurotoxicity of  $\beta$ -Amyloid Protein: Oligomerization, Channel Formation and Calcium Dyshomeostasis. *Current Pharmaceutical Design* **16**, 2779–2789 (Aug. 2010) (cit. on p. 6).
50. Wilkinson, K. & El Khoury, J. Microglial Scavenger Receptors and Their Roles in the Pathogenesis of Alzheimer's Disease. *International Journal of Alzheimer's Disease* **2012**, 1–10 (2012) (cit. on p. 6).
51. Nguyen, P. H., Ramamoorthy, A., Sahoo, B. R., *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) (cit. on p. 6).
52. Gallardo, R., Ranson, N. A. & Radford, S. E. Amyloid structures: much more than just a cross- $\beta$  fold. *Current Opinion in Structural Biology* **60**, 7–16 (Feb. 2020) (cit. on p. 6).
53. Cawood, E. E., Karamanos, T. K., Wilson, A. J. & Radford, S. E. Visualizing and trapping transient oligomers in amyloid assembly pathways. *Biophysical Chemistry* **268**, 106505 (Jan. 2021) (cit. on p. 6).
54. Michaels, T. C. T., Šarić, A., Curk, S., *et al.* Dynamics of oligomer populations formed during the aggregation of Alzheimer's A $\beta$ 42 peptide. *Nature Chemistry* **12**, 445–451 (Apr. 2020) (cit. on p. 6).
55. Kollmer, M., Close, W., Funk, L., *et al.* Cryo-EM structure and polymorphism of A $\beta$  amyloid fibrils purified from Alzheimer's brain tissue. *Nature Communications* **10** (Oct. 2019) (cit. on p. 7).
56. Fändrich, M., Nyström, S., Nilsson, K. P. R., *et al.* Amyloid fibril polymorphism: a challenge for molecular imaging and therapy. *Journal of Internal Medicine* **283**, 218–237 (Feb. 2018) (cit. on p. 7).
57. Matsumura, N., Takami, M., Okochi, M., *et al.*  $\gamma$ -Secretase Associated with Lipid Rafts. *Journal of Biological Chemistry* **289**, 5109–5121 (Feb. 2014) (cit. on p. 7).

58. Kang, J., Lemaire, H.-G., Unterbeck, A., *et al.* The precursor of Alzheimer's disease amyloid A4 protein resembles a cell-surface receptor. *Nature* **325**, 733–736 (Feb. 1987) (cit. on p. 8).
59. Chen, G.-f., Xu, T.-h., Yan, Y., *et al.* Amyloid beta: structure, biology and structure-based therapeutic development. *Acta Pharmacologica Sinica* **38**, 1205–1235 (July 2017) (cit. on p. 8).
60. Findeis, M. A. The role of amyloid  $\beta$  peptide 42 in Alzheimer's disease. *Pharmacology & Therapeutics* **116**, 266–286 (Nov. 2007) (cit. on p. 8).
61. Zheng, H., Jiang, M., Trumbauer, M. E., *et al.*  $\beta$ -amyloid precursor protein-deficient mice show reactive gliosis and decreased locomotor activity. *Cell* **81**, 525–531 (May 1995) (cit. on p. 8).
62. Teng, F. Y. H. & Tang, B. L. Widespread gamma-secretase activity in the cell, but do we need it at the mitochondria? en. *Biochem. Biophys. Res. Commun.* **328**, 1–5 (Mar. 2005) (cit. on p. 8).
63. Plant, L. D., Boyle, J. P., Smith, I. F., Peers, C. & Pearson, H. A. The Production of Amyloid  $\beta$  Peptide Is a Critical Requirement for the Viability of Central Neurons. *The Journal of Neuroscience* **23**, 5531–5535 (July 2003) (cit. on p. 8).
64. Walsh, D. M. & Selkoe, D. J. A beta oligomers - a decade of discovery. en. *J. Neurochem.* **101**, 1172–1184 (June 2007) (cit. on p. 8).
65. Abelein, A., Abrahams, J. P., Danielsson, J., *et al.* The hairpin conformation of the amyloid  $\beta$  peptide is an important structural motif along the aggregation pathway. *JBIC Journal of Biological Inorganic Chemistry* **19**, 623–634 (Apr. 2014) (cit. on p. 9).
66. Lue, L.-F., Kuo, Y.-M., Roher, A. E., *et al.* Soluble Amyloid  $\beta$  Peptide Concentration as a Predictor of Synaptic Change in Alzheimer's Disease. *The American Journal of Pathology* **155**, 853–862 (Sept. 1999) (cit. on p. 9).
67. McLean, C. A., Cherny, R. A., Fraser, F. W., *et al.* Soluble pool of A $\beta$  amyloid as a determinant of severity of neurodegeneration in Alzheimer's disease. *Annals of Neurology* **46**, 860–866 (Dec. 1999) (cit. on p. 9).
68. Martin, T. J., Sims, N. A. & Seeman, E. Physiological and Pharmacological Roles of PTH and PTHrP in Bone Using Their Shared Receptor, PTH1R. *Endocrine Reviews* **42**, 383–406 (Feb. 2021) (cit. on pp. 9, 63).
69. Gopalswamy, M., Kumar, A., Adler, J., *et al.* Structural characterization of amyloid fibrils from the human parathyroid hormone. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* **1854**, 249–257 (Apr. 2015) (cit. on pp. 9, 63).
70. Rendina-Ruedy, E. & Rosen, C. J. Parathyroid hormone (PTH) regulation of metabolic homeostasis: An old dog teaches us new tricks. *Molecular Metabolism* **60**, 101480 (June 2022) (cit. on p. 9).
71. Ishtiaq, S., Fogelman, I. & Hampson, G. Treatment of post-menopausal osteoporosis: beyond bisphosphonates. *Journal of Endocrinological Investigation* **38**, 13–29 (Sept. 2014) (cit. on pp. 9, 63).
72. Musacchio, A., Wilmanns, M. & Saraster, M. Structure and function of the SH3 domain. *Progress in Biophysics and Molecular Biology* **61**, 283–297 (1994) (cit. on p. 10).

73. Dionne, U., Percival, L. J., Chartier, F. J., Landry, C. R. & Bisson, N. SRC homology 3 domains: multifaceted binding modules. *Trends in Biochemical Sciences* **47**, 772–784 (Sept. 2022) (cit. on p. 10).
74. Stout, T., Foster, P. & Matthews, D. High-Throughput Structural Biology in Drug Discovery: Protein Kinases. *Current Pharmaceutical Design* **10**, 1069–1082 (Apr. 2004) (cit. on p. 10).
75. Guijarro, J. I., Sunde, M., Jones, J. A., Campbell, I. D. & Dobson, C. M. Amyloid fibril formation by an SH3 domain. en. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 4224–4228 (Apr. 1998) (cit. on p. 10).
76. Gardon, L., Becker, N., Rähse, N., *et al.* Amyloid fibril formation kinetics of low-pH denatured bovine PI3K-SH3 monitored by three different NMR techniques. en. *Front. Mol. Biosci.* **10**, 1254721 (Nov. 2023) (cit. on pp. 10, 73).
77. Zwanzig, R. *Nonequilibrium statistical mechanics* (Oxford university press, 2001) (cit. on p. 13).
78. Kinjo, T. & Hyodo, S.-a. Equation of motion for coarse-grained simulation based on microscopic description. *Physical Review E* **75** (May 2007) (cit. on p. 13).
79. Noé, F., Tkatchenko, A., Müller, K.-R. & Clementi, C. Machine Learning for Molecular Simulation. *Annual Review of Physical Chemistry* **71**, 361–390 (Apr. 2020) (cit. on p. 14).
80. Wang, J., Olsson, S., Wehmeyer, C., *et al.* Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Central Science* **5**, 755–767 (Apr. 2019) (cit. on p. 14).
81. Unke, O. T., Chmiela, S., Sauceda, H. E., *et al.* Machine Learning Force Fields. *Chemical Reviews* **121**, 10142–10186 (Mar. 2021) (cit. on p. 14).
82. Huang, J., Rauscher, S., Nawrocki, G., *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nature Methods* **14**, 71–73 (Nov. 2016) (cit. on pp. 14, 16, 17, 34, 39, 77).
83. Samantray, S., Yin, F., Kav, B. & Strodel, B. Different force fields give rise to different amyloid aggregation pathways in molecular dynamics simulations. *J. Chem. Inf. Model.* **60**, 6462–6475 (2021) (cit. on pp. 14, 40).
84. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., *et al.* CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4**, 187–217 (June 1983) (cit. on p. 16).
85. Zhu, X., Lopes, P. E. M. & MacKerell, A. D. Recent developments and applications of the CHARMM force fields. *WIREs Computational Molecular Science* **2**, 167–185 (June 2011) (cit. on p. 16).
86. Pastor, R. W. & Mackerell Jr, A. D. Development of the CHARMM force field for lipids. en. *J. Phys. Chem. Lett.* **2**, 1526–1532 (2011) (cit. on p. 16).
87. Klauda, J. B., Monje, V., Kim, T. & Im, W. Improving the CHARMM force field for polyunsaturated fatty acid chains. en. *J. Phys. Chem. B* **116**, 9424–9431 (Aug. 2012) (cit. on p. 16).



88. Raman, E. P., Guvench, O. & MacKerell, A. D. CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses. *The Journal of Physical Chemistry B* **114**, 12981–12994 (Sept. 2010) (cit. on p. 16).
89. Mallajosyula, S. S., Guvench, O., Hatcher, E. & MacKerell, A. D. CHARMM Additive All-Atom Force Field for Phosphate and Sulfate Linked to Carbohydrates. *Journal of Chemical Theory and Computation* **8**, 759–776 (Jan. 2012) (cit. on p. 16).
90. MacKerell, A. D., Banavali, N. & Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **56**, 257–265 (2000) (cit. on p. 16).
91. Vanommeslaeghe, K., Hatcher, E., Acharya, C., *et al.* CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **31**, 671–690 (July 2009) (cit. on pp. 16, 67).
92. Loschwitz, J., Olubiyi, O. O., Hub, J. S., Strodel, B. & Poojari, C. S. in *Progress in Molecular Biology and Translational Science* 273–403 (Elsevier, 2020) (cit. on p. 16).
93. Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *Journal of Computational Chemistry* **25**, 1584–1604 (July 2004) (cit. on p. 16).
94. Buck, M., Bouguet-Bonnet, S., Pastor, R. W. & MacKerell, A. D. Importance of the CMAP Correction to the CHARMM22 Protein Force Field: Dynamics of Hen Lysozyme. *Biophysical Journal* **90**, L36–L38 (Feb. 2006) (cit. on p. 16).
95. 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**, 6462–6475 (Nov. 2020) (cit. on p. 16).
96. Paul, A., Samantray, S., Anteghini, M., Khaled, M. & Strodel, B. Thermodynamics and kinetics of the amyloid- $\beta$  peptide revealed by Markov state models based on MD data in agreement with experiment. *Chem. Sci.* **12**, 6652–6669 (19 2021) (cit. on pp. 16, 77).
97. Jorgensen, W., Chandrasekhar, J., Madura, J., Impey, R. & Klein, M. Comparison of simple potential functions for simulating liquid water. *J Chem Theory Comput* **79**, 926–935 (1983) (cit. on pp. 17, 34, 39, 49, 77).
98. 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 (May 2000) (cit. on p. 17).
99. 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 (Nov. 1987) (cit. on p. 17).
100. 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 (Jan. 1981) (cit. on p. 17).
101. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983) (cit. on p. 17).

102. Verlet, L. Computer "experiments" on classical fluids. II. Equilibrium correlation functions. *Phys Rev* **165**, 201–214 (1 1968) (cit. on p. 19).
103. Gunsteren, W. F. V. & Berendsen, H. J. C. A leap-frog algorithm for stochastic dynamics. *Mol Simulat* **1**, 173–185 (1988) (cit. on pp. 19, 77).
104. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126** (2007) (cit. on pp. 20, 39).
105. Nosé, S. Molecular-dynamics method for simulations in the canonical ensemble. *Mol Phys* **52**, 255–268 (1984) (cit. on pp. 20, 34, 39, 49).
106. Hoover, W. Canonical dynamics - equilibrium phase-space distributions. *Phys Rev A* **31**, 1695–1697 (1985) (cit. on pp. 20, 21, 34, 39, 49, 77).
107. 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. *J Chem Phys* **81**, 3684–3690 (1984) (cit. on p. 21).
108. Parrinello, M. & Rahman, A. Polymorphic transitions in single-crystals - a new molecular-dynamics method. *Mol Phys* **52**, 7182–7190 (1981) (cit. on pp. 21, 34, 39, 49).
109. Nosé, S. Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **52**, 255–268 (1984) (cit. on p. 21).
110. Clausius, R. XVI. On a mechanical theorem applicable to heat. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* **40**, 122–127 (Aug. 1870) (cit. on p. 21).
111. Tsai, D. H. The virial theorem and stress calculation in molecular dynamics. *The Journal of Chemical Physics* **70**, 1375–1382 (Feb. 1979) (cit. on p. 21).
112. Parrinello, M. & Rahman, A. Crystal structure and pair potentials: A molecular-dynamics study. *Phys Rev Lett* **45**, 1196–1199 (14 1980) (cit. on p. 21).
113. Parrinello, M. & Rahman, A. Strain fluctuations and elastic constants. *J Chem Phys* **76**, 2662–2666 (1982) (cit. on p. 21).
114. Nosé, S. & Klein, M. Constant pressure molecular dynamics for molecular systems. *Molecular Physics* **50**, 1055–1076 (Dec. 1983) (cit. on p. 22).
115. Bussi, G. & Laio, A. Using metadynamics to explore complex free-energy landscapes. *Nat Rev Phys* **2**, 200–212 (2020) (cit. on p. 23).
116. Torrie, G. & Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **23**, 187–199 (Feb. 1977) (cit. on p. 23).
117. 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**, 3584–3595 (July 2015) (cit. on p. 23).
118. Swendsen, R. H. & Wang, J.-S. Replica Monte Carlo Simulation of Spin-Glasses. *Physical Review Letters* **57**, 2607–2609 (Nov. 1986) (cit. on p. 23).
119. Okabe, T., Kawata, M., Okamoto, Y. & Mikami, M. Replica-exchange Monte Carlo method for the isobaric–isothermal ensemble. *Chemical Physics Letters* **335**, 435–439 (Mar. 2001) (cit. on p. 23).

120. Hastings, W. K. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97–109 (Apr. 1970) (cit. on p. 23).
121. Shirts, M. R. & Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys* **129**, 124105 (2008) (cit. on pp. 24, 87, 88).
122. Shirts, M. R. & Ferguson, A. L. Statistically optimal continuous free energy surfaces from biased simulations and multistate reweighting. *J Chem Theory Comput* **16**, 4107–4125 (2020) (cit. on pp. 24, 87).
123. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H. & Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J Comput Chem* **13**, 1011–1021 (1992) (cit. on pp. 24, 87).
124. Pearson, K. LIII. On lines and planes of closest fit to systems of points in space. *London Edinburgh Dublin Philos Mag J Sci* **2**, 559–572 (1901) (cit. on p. 25).
125. Jolliffe, I. T. *Principal component analysis* (Springer New York, NY, 2006) (cit. on p. 25).
126. Jolliffe, I. T. & Cadima, J. Principal component analysis: A review and recent developments. *Philos T R Soc A* **374**, 20150202 (2016) (cit. on p. 25).
127. Chodera, J. D. & Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**. Theory and simulation / Macromolecular machines, 135–144 (2014) (cit. on p. 25).
128. Husic, B. E. & Pande, V. S. Markov State Models: From an Art to a Science. *J. Am. Chem. Soc.* **140**, 2386–2396 (2018) (cit. on p. 25).
129. Rao, F. & Caflisch, A. The Protein Folding Network. *J. Mol. Biol.* **342**, 299–306 (2004) (cit. on p. 25).
130. Noé, F. & Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Current Opinion in Structural Biology* **18**. Theory and simulation / Macromolecular assemblages, 154–162 (2008) (cit. on p. 25).
131. Schäffler, M., Khaled, M. & Strodel, B. ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins. *Methods* **206**, 18–26 (2022) (cit. on pp. 26, 32, 39, 56).
132. Zhou, T. & Caflisch, A. Distribution of Reciprocal of Interatomic Distances: A Fast Structural Metric. *Journal of Chemical Theory and Computation* **8**, 2930–2937 (July 2012) (cit. on p. 26).
133. Chakraborty, D., Straub, J. E. & Thirumalai, D. Energy landscapes of A $\beta$  monomers are sculpted in accordance with Ostwald’s rule of stages. *Science Advances* **9**, eadd6921 (2023) (cit. on p. 26).
134. Scherer, M. K., Trendelkamp-Schroer, B., Paul, F., *et al.* PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015) (cit. on p. 27).
135. Schaeffler, M., Wales, D. J. & Strodel, B. The energy landscape of A $\beta$ 1-42: a funnel to disorder for the monomer becomes a folding funnel for self-assembly. *Chemical Communications* (2024) (cit. on pp. 27, 28, 30, 48, 50, 58, 60).

136. Becker, O. M. & Karplus, M. The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics. *J. Chem. Phys.* **106**, 1495–1517 (1997) (cit. on p. 28).
137. Wales, D. J., Miller, M. A. & Walsh, T. R. Archetypal energy landscapes. *Nature* **394**, 758–760 (1998) (cit. on p. 28).
138. Swinburne, T. D., Kannan, D., Sharpe, D. J. & Wales, D. J. Rare events and first passage time statistics from the energy landscape. *The Journal of Chemical Physics* **153** (Oct. 2020) (cit. on p. 29).
139. Wales, D. J. Dynamical Signatures of Multifunnel Energy Landscapes. *J. Phys. Chem. Lett.* **13**, 6349–6358 (2022) (cit. on p. 29).
140. Woods, E. J. & Wales, D. J. Analysis and interpretation of first passage time distributions featuring rare events. *Physical Chemistry Chemical Physics* **26**, 1640–1657 (2024) (cit. on p. 29).
141. Bastian, M., Heymann, S. & Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks (2009) (cit. on pp. 33, 39).
142. 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 (Oct. 2008) (cit. on pp. 33, 39).
143. Abraham, M., Murtola, T., Schulz, R., *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015) (cit. on pp. 34, 39, 49).
144. Bowman, G. R., Ensign, D. L. & Pande, V. S. Enhanced Modeling via Network Theory: Adaptive Sampling of Markov State Models. *Journal of Chemical Theory and Computation* **6**, 787–794 (Feb. 2010) (cit. on pp. 34, 39).
145. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* (1993) (cit. on pp. 34, 39, 50, 77).
146. Iannuzzi, C., Irace, G. & Sirangelo, I. The Effect of Glycosaminoglycans (GAGs) on Amyloid Aggregation and Toxicity. *Molecules* **20**, 2510–2528 (2015) (cit. on pp. 37, 54, 60).
147. Schäffler, M., Samantray, S. & Strodel, B. Transition Networks Unveil Disorder-to-Order Transformations in A $\beta$  Caused by Glycosaminoglycans or Lipids. *International Journal of Molecular Sciences* **24**, 11238 (July 2023) (cit. on pp. 38, 54, 56, 61).
148. Jo, S., Song, K. C., Desaire, H., MacKerell, A. D. & Im, W. Glycan reader: Automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *Journal of Computational Chemistry* **32**, 3135–3141 (Aug. 2011) (cit. on p. 39).
149. Park, S.-J., Lee, J., Patel, D. S., *et al.* Glycan Reader is improved to recognize most sugar types and chemical modifications in the Protein Data Bank. *Bioinformatics* **33** (ed Valencia, A.) 3051–3057 (June 2017) (cit. on p. 39).
150. Park, S.-J., Lee, J., Qi, Y., *et al.* CHARMM-GUIGlycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **29**, 320–331 (Feb. 2019) (cit. on p. 39).

151. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *Journal of Computational Chemistry* **29**, 1859–1865 (June 2008) (cit. on p. 39).
152. Yang, Y., Arseni, D., Zhang, W., *et al.* Cryo-EM structures of amyloid- $\beta$  42 filaments from human brains. *Science* **375**, 167–172 (Jan. 2022) (cit. on pp. 46, 104).
153. Jensen, M. R., Zweckstetter, M., Huang, J.-r. & Blackledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chemical Reviews* **114**, 6632–6660 (Apr. 2014) (cit. on p. 47).
154. Chong, S.-H. & Ham, S. Folding Free Energy Landscape of Ordered and Intrinsically Disordered Proteins. *Sci. Rep.* **9**, 14927 (Oct. 2019) (cit. on pp. 47, 60).
155. Muschol, M. & Hoyer, W. Amyloid oligomers as on-pathway precursors or off-pathway competitors of fibrils. *Frontiers in Molecular Biosciences* **10** (Feb. 2023) (cit. on p. 47).
156. Powers, E. T. & Powers, D. L. Mechanisms of Protein Fibril Formation: Nucleated Polymerization with Competing Off-Pathway Aggregation. *Biophysical Journal* **94**, 379–391 (Jan. 2008) (cit. on p. 47).
157. Granata, D., Baftizadeh, F., Habchi, J., *et al.* The inverted free energy landscape of an intrinsically disordered peptide by simulations and experiments. *Sci. Rep.* **5**, 15449 (Oct. 2015) (cit. on pp. 47, 50).
158. Huang, J., Rauscher, S., Nawrocki, G., *et al.* CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nature Meth.* **14**, 71–73 (2017) (cit. on pp. 49, 77).
159. Willbold, D., Strodel, B., Schröder, G. F., Hoyer, W. & Heise, H. Amyloid-type Protein Aggregation and Prion-like Properties of Amyloids. *Chem. Rev.* **121**, 8285–8307 (2021) (cit. on pp. 50, 54).
160. PATHSAMPLE: A program for generating connected stationary point databases and extracting global kinetics <http://www-wales.ch.cam.ac.uk/software.html>. <http://www-wales.ch.cam.ac.uk/software.html> (cit. on p. 53).
161. Sandberg, A., Luheshi, L., Söllvander, S., *et al.* Stabilization of neurotoxic Alzheimer amyloid- oligomers by protein engineering. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15595–600 (2010) (cit. on p. 53).
162. Sciarretta, K. L., Gordon, D. J., Petkova, A. T., Tycko, R. & Meredith, S. C. A $\beta$ 40-Lactam(D23/K28) Models a Conformation Highly Favorable for Nucleation of Amyloid. *Biochemistry* **44**, 6003–6014 (Mar. 2005) (cit. on p. 53).
163. Mithu, V. S., Sarkar, B., Bhowmik, D., *et al.* Zn<sup>++</sup> Binding Disrupts the Asp23-Lys28 Salt Bridge without Altering the Hairpin-Shaped Cross- $\beta$  Structure of A $\beta$ 42 Amyloid Aggregates. *Biophysical Journal* **101**, 2825–2832 (Dec. 2011) (cit. on p. 53).
164. Adler, J., Scheidt, H. A., Krüger, M., Thomas, L. & Huster, D. Local interactions influence the fibrillation kinetics, structure and dynamics of A $\beta$ (1–40) but leave the general fibril structure unchanged. *Phys. Chem. Chem. Phys.* **16**, 7461–7471 (2014) (cit. on p. 53).
165. Das, A. K., Rawat, A., Bhowmik, D., *et al.* An Early Folding Contact between Phe19 and Leu34 is Critical for Amyloid- $\beta$  Oligomer Toxicity. *ACS Chemical Neuroscience* **6**, 1290–1295 (May 2015) (cit. on p. 53).

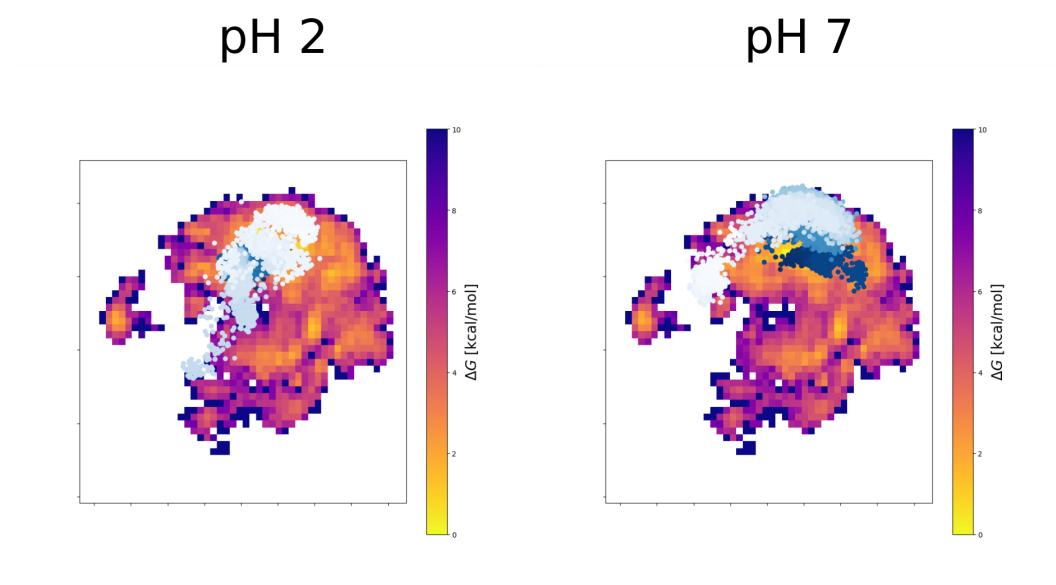
166. Korn, A., McLennan, S., Adler, J., *et al.* Amyloid  $\beta$  (1–40) Toxicity Depends on the Molecular Contact between Phenylalanine 19 and Leucine 34. *ACS Chemical Neuroscience* **9**, 790–799 (Dec. 2017) (cit. on p. 53).
167. Jeon, J., Yau, W.-M. & Tycko, R. Early events in amyloid- $\beta$  self-assembly probed by time-resolved solid state NMR and light scattering. *Nature Commun.* **14**, 2964 (May 2023) (cit. on p. 53).
168. Linhardt, R. J. & Toida, T. Role of Glycosaminoglycans in Cellular Communication. *Accounts of Chemical Research* **37**, 431–438 (May 2004) (cit. on p. 60).
169. Van Horssen, J., Wesseling, P., van den Heuvel, L. P., de Waal, R. M. & Verbeek, M. M. Heparan sulphate proteoglycans in Alzheimer's disease and amyloid-related disorders. *The Lancet Neurology* **2**, 482–492 (Aug. 2003) (cit. on p. 60).
170. Díaz-Nido, J., Wandosell, F. & Avila, J. Glycosaminoglycans and  $\beta$ -amyloid, prion and tau peptides in neurodegenerative diseases. *Peptides* **23**, 1323–1332 (July 2002) (cit. on p. 60).
171. Stewart, K. L., Hughes, E., Yates, E. A., *et al.* Atomic Details of the Interactions of Glycosaminoglycans with Amyloid- $\beta$  Fibrils. *Journal of the American Chemical Society* **138**, 8328–8331 (July 2016) (cit. on p. 60).
172. Cline, E. N., Bicca, M. A., Viola, K. L. & Klein, W. L. The Amyloid- $\beta$  Oligomer Hypothesis: Beginning of the Third Decade. *Journal of Alzheimer's Disease* **64** (eds Perry, G., Avila, J., Moreira, P., Sorensen, A. & Tabaton, M.) S567–S610 (June 2018) (cit. on p. 61).
173. La Rosa, C., Scalisi, S., Lolicato, F., Pannuzzo, M. & Raudino, A. Lipid-assisted protein transport: A diffusion-reaction model supported by kinetic experiments and molecular dynamics simulations. *The Journal of Chemical Physics* **144** (May 2016) (cit. on p. 61).
174. Sciacca, M. F., Lolicato, F., Tempra, C., *et al.* Lipid-Chaperone Hypothesis: A Common Molecular Mechanism of Membrane Disruption by Intrinsically Disordered Proteins. *ACS Chemical Neuroscience* **11**, 4336–4350 (Dec. 2020) (cit. on p. 61).
175. Abedini, A. & Raleigh, D. P. A critical assessment of the role of helical intermediates in amyloid formation by natively unfolded proteins and polypeptides. *Protein Engineering Design and Selection* **22**, 453–459 (July 2009) (cit. on p. 61).
176. Paschold, A., Schäffler, M., Miao, X., *et al.* Photocontrolled Reversible Amyloid Fibril Formation of Parathyroid Hormone-Derived Peptides. *Bioconjugate Chemistry* **35**, 981–995 (June 2024) (cit. on pp. 63, 65, 71, 72).
177. Böckmann, M., Peter, C., Site, L. D., *et al.* Atomistic Force Field for Azobenzene Compounds Adapted for QM/MM Simulations with Applications to Liquids and Liquid Crystals. *Journal of Chemical Theory and Computation* **3**, 1789–1802 (Aug. 2007) (cit. on p. 67).
178. 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**, 7825–7839 (Nov. 2020) (cit. on p. 69).
179. Dyson, H. & Wright, P. E. Coupling of folding and binding for unstructured proteins. *Current Opinion in Structural Biology* **12**, 54–60 (Feb. 2002) (cit. on p. 69).

180. Guo, M., Gorman, P. M., Rico, M., Chakrabartty, A. & Laurents, D. V. Charge substitution shows that repulsive electrostatic interactions impede the oligomerization of Alzheimer amyloid peptides. *FEBS Letters* **579**, 3574–3578 (June 2005) (cit. on p. 73).
181. Tian, Y. & Viles, J. H. pH Dependence of Amyloid- $\beta$  Fibril Assembly Kinetics: Unraveling the Microscopic Molecular Processes. *Angewandte Chemie International Edition* **61** (Oct. 2022) (cit. on p. 73).
182. Hortschansky, P., Schroeckh, V., Christopeit, T., Zandomenighi, G. & Fändrich, M. The aggregation kinetics of Alzheimer's  $\beta$ -amyloid peptide is controlled by stochastic nucleation. *Protein Science* **14**, 1753–1759 (July 2005) (cit. on p. 73).
183. Schützmann, M. P., Hasecke, F., Bachmann, S., *et al.* Endo-lysosomal A $\beta$  concentration and pH trigger formation of A $\beta$  oligomers that potently induce Tau missorting. *Nature Communications* **12** (July 2021) (cit. on p. 73).
184. Fezoui, Y., Hartley, D. M., Harper, J. D., *et al.* An improved method of preparing the amyloid  $\beta$ -protein for fibrillogenesis and neurotoxicity experiments. *Amyloid* **7**, 166–178 (Jan. 2000) (cit. on p. 73).
185. Röder, C., Schröder, G. F. & *et al.* Atomic structure of PI3-kinase SH3 amyloid fibrils by cryo-electron microscopy. *Nature Communications* **10**, 3754 (2019) (cit. on pp. 74, 92).
186. KOYAMA, S. Structure of the PI3K SH3 domain and analysis of the SH3 family. *Cell* **72**, 945–952 (Mar. 1993) (cit. on p. 74).
187. Booker, G. W., Gout, I., Kristina<sup>D</sup>owning, A., *et al.* Solution structure and ligand-binding site of the SH3 domain of the p85 $\alpha$  subunit of phosphatidylinositol 3-kinase. *Cell* **73**, 813–822 (May 1993) (cit. on p. 74).
188. Warshel, A., Sussman, F. & King, G. Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry* **25**, 8368–8372 (Dec. 1986) (cit. on p. 75).
189. Knight, J. L. & Brooks, C. L.  $\lambda$ -Dynamics free energy simulation methods. *Journal of Computational Chemistry* **30**, 1692–1700 (May 2009) (cit. on p. 75).
190. Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. H++ 3.0: Automating PK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research* **40** (2012) (cit. on p. 75).
191. Nielsen, J. E. & Vriend, G. Optimizing the hydrogen-bond network in Poisson–Boltzmann equation-based pKa calculations. *Proteins: Structure, Function, and Bioinformatics* **43**, 403–412 (Apr. 2001) (cit. on p. 75).
192. Baker, N. & Bashford, D. Case DA In New Algorithms for Macromolecular Simulation. *Leimkuhler B., Chipot C., Elber R., Laaksonen A., Mark A., Schlick T., Schütte C., Skeel R., Eds*, 263–295 (cit. on p. 75).
193. Abraham, M. J., Murtola, T., Schulz, R., *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1**, 19–25 (2015) (cit. on p. 77).
194. Robustelli, P., Piana, S. & Shaw, D. E. Developing a molecular dynamics force field for both folded and disordered protein states. *Proceedings of the National Academy of Sciences* **115** (May 2018) (cit. on p. 77).

195. Berendsen, H. J., Postma, J. v., 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) (cit. on p. 77).
196. Glyakina, A. & Galzitskaya, O. How Quickly Do Proteins Fold and Unfold, and What Structural Parameters Correlate with These Values? *Biomolecules* **10**, 197 (Jan. 2020) (cit. on p. 85).
197. Uttley, L., Carroll, C., Wong, R., Hilton, D. A. & Stevenson, M. Creutzfeldt-Jakob disease: a systematic review of global incidence, prevalence, infectivity, and incubation. *The Lancet Infectious Diseases* **20**, e2–e10 (Jan. 2020) (cit. on p. 105).
198. Rudge, P., Jaunmuktane, Z., Hyare, H., *et al.* Early neurophysiological biomarkers and spinal cord pathology in inherited prion disease. en. *Brain* **142**, 760–770 (Mar. 2019) (cit. on p. 105).



## A.1 Supplementary Figures



**Fig. A1.:** Free energy surface of the SH3 domain obtained from REMD with projected trajectory from conventional MD at pH 2 and pH 7. The principal components on which the FES is projected were calculated from all replicas at pH 2. For improved sampling, the FES was reconstructed using the pymbar method. The trajectories from conventional MD are projected into the same PC space and are plotted on top of the FES. The projected frames are colored from dark to light blue according to their appearance in the trajectory, from start to end, respectively.



# ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins

Moritz Schäffler<sup>a,1</sup>, Mohammed Khaled<sup>a,1</sup>, Birgit Strodel<sup>a,b,\*</sup>

<sup>a</sup> Institute of Biological Information Processing (IBI-7:Structural Biochemistry), Forschungszentrum Jülich, 52425 Jülich, Germany

<sup>b</sup> Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

## ARTICLE INFO

### Keywords:

Intrinsically disordered proteins  
Molecular dynamics simulations  
Transition networks  
Amyloid- $\beta$   
Disorder-to-order transition

## ABSTRACT

Intrinsically disordered proteins (IDPs) do not fold into a unique three-dimensional structure but sample different configurations of different probabilities that further change with the surrounding of the IDPs. The structural heterogeneity and dynamics of IDPs pose a challenge for the characterization of their structures by experimental techniques only. Molecular dynamics (MD) simulations provide a powerful complement to experimental approaches for that purpose. However, MD simulations on the micro- to millisecond timescale generate a lot of data of protein motions, necessitating advanced post-processing techniques to extract the relevant information. Here, we demonstrate how transition networks created from MD trajectories allow revealing the configurational ensemble and structural interconversions of IDPs, using the amyloid- $\beta$  peptide as example. The construction of transition networks relies on molecular descriptors as input, and we show how the choice of descriptors influences the resulting transition network. The transition networks are generated with the open-source Python script ATRANET, and we explain the usage of ATRANET by providing a detailed workflow and exemplary analysis for amyloid- $\beta$ , which can be easily generalized to other IDPs and even protein aggregation.

## 1. Introduction

A wide range of the human proteins belongs to the class of intrinsically disordered proteins (IDPs) or contain intrinsically disordered regions [1,2]. IDPs are commonly found in all living organisms and are frequently involved in various biological processes, such as cell cycle control, protein–protein interaction, cell signaling, and disease pathways, including neurodegenerative diseases [3]. These proteins are structurally and functionally different from folded proteins and usually lack a unique three-dimensional structure under physiological conditions [4]. Instead, IDPs adopt an ensemble of different conformations and can even undergo disordered-to-ordered transitions upon forming complexes with specific binding partners, thereby comprising structural plasticity needed for various functions of the same IDP [5,6]. The wide conformational heterogeneity of IDPs imposes difficulties in studying their structural characteristics in experiments [7]. Common experimental techniques for the investigation of protein structures include nuclear magnetic resonance (NMR) spectroscopy, small angle X-ray scattering (SAXS), cryo-electron microscopy, and single-molecule Förster resonance energy transfer (smFRET) spectroscopy [8]. Since

these techniques usually average over a wide range of conformations, the obtained structural information is often limited in the case of IDPs. Consequently, special experimental and computational tools are required to identify and analyze the conformational ensembles of IDPs.

In terms of computational approaches, the combination of molecular dynamics (MD) simulations and network models, such as Markov state models (MSMs) or conformational transition networks (TNs) provide advanced capabilities for elucidating structural preferences of IDPs and the kinetics of conformational transitions [9–11]. The advantage of applying network models to MD trajectories is that they can extract the mechanisms of the molecular processes hidden within the large amounts of MD simulation data, producing human-readable networks of the processes under study [12–16]. It should be noted that MSMs fall in the broader category of methods that aim to set up a master equation for the dynamics [17–19]. Although MSMs can help in revealing pathways of biological processes and the associated kinetics [12,14,15], fulfilling the requirements to calculate MSMs from MD data, like the detailed balance requirement, makes the MSM approach challenging for large biomolecular systems that for this reason requires extensive MD simulations [20]. Alternatives to MD simulation-based techniques would be

\* Corresponding author at: Institute of Biological Information Processing (IBI-7:Structural Biochemistry), Forschungszentrum Jülich, 52425 Jülich, Germany.

E-mail address: [b.strodel@fz-juelich.de](mailto:b.strodel@fz-juelich.de) (B. Strodel).

<sup>1</sup> Contributed equally.

approaches involving Monte Carlo methods [21] or geometry optimization techniques such as discrete path sampling [22,23]. The latter method characterizes the energy landscape in terms of local minima and the transition states that connect them, thereby providing a direct way of coarse-graining the energy landscape, which can be compared to the identification of stable states from the MD simulations. Moreover, discrete path sampling also provides the framework to calculate rate constants from the database of local minima and transition states and master equation dynamics [22,23]. Both, Monte Carlo approaches as well as discrete path sampling have been successfully applied to IDPs [24,25]. They have the advantage of being able to sample processes beyond the MD time scale. However, they are typically only employed together with implicit solvent models, where the crucial and specific interactions between proteins and water molecules or between proteins and lipids when modeling protein–membrane interactions are being ignored [26,27].

If one does not request a method that aims at setting up a master equation for the dynamics but one nonetheless wants to have a network model of the protein motions sampled by MD simulations (with explicit modeling of the protein's surroundings), TNs provide a solution. The TN approach is purely based on geometric clustering – and not on kinetic clustering as involved in Markov state modeling – to extract the essential features of protein conformational transitions, ranging from protein folding to IDP conformation switching and protein aggregation [16,28,9,29]. In order to construct a TN, one chooses a set of descriptor functions that characterize the process under study, such as the amount of specific secondary structure elements or the aggregate size and the number and type of intermolecular contacts if one studies protein aggregation. The descriptors are applied to each frame of the MD trajectory, thereby mapping the trajectory to a set of states with the information of the transitions between these states. This allows to construct a transition matrix, which can then be visualized. In our group, we have improved and successfully applied the TN approach for protein aggregation using our self-developed Python script ATRANET [29,30].

Here, we apply the TN approach to study conformation switching of the amyloid- $\beta$  ( $A\beta$ ) peptide.  $A\beta$  is an IDP that is being studied intensively, as it is strongly linked to the development of Alzheimer's disease [31–33]. In solution, it mostly adopts expanded random-coil conformations with a very low tendency to form transient  $\beta$ -sheets or  $\alpha$ -helices [11,34,35]. However, we recently showed that it undergoes a disordered-to-ordered transition upon forming a complex with three 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids [36]. Here, we demonstrate how TNs can be used to identify the predominant states of the  $A\beta$ -POPC complex and illustrate how the complexity of the resulting network can be regulated by the choice of the molecular descriptors. However, before showcasing TNs in action, we provide the theory behind TNs and how to generate them with ATRANET (Automated TRAnSition NETWORKs).

## 2. Background

### 2.1. Transition networks

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  $x(t)$ ,

$$S(t) = [f_1(x(t)), f_2(x(t)), \dots, f_n(x(t))] \quad (1)$$

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  $c_{\text{corr}}$

$$c_{\text{corr}} = \frac{\sum_t (f_i(x(t)) - \bar{f}_i)(f_j(x(t)) - \bar{f}_j)}{\sqrt{\sum_t (f_i(x(t)) - \bar{f}_i)^2} \sqrt{\sum_t (f_j(x(t)) - \bar{f}_j)^2}} \quad (2)$$

where  $\bar{f}_i$  and  $\bar{f}_j$  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.

### 2.2. 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  $\alpha$ -helices,  $\beta$ -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 [37,38].

**Backbone torsion angles.** For short peptides, such as the  $A\beta_{16-22}$  fragment of  $A\beta$  it can be advantageous to use the Ramachandran angles  $\varphi$  ( $C - C_\alpha - N - C$ ) and  $\psi$  ( $N - C - C_\alpha - N$ ) to assign the secondary structure of the peptide instead of applying DSSP. The angles  $\varphi$  and  $\psi$  allow to identify extended peptide structures belonging to the  $\beta$ -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  $(\varphi, \psi)$  data to a specified number of cluster centers. If  $N_{\text{res}}$  is the number of peptide residues and  $F$  is the number of frames in the MD trajectory, the  $N_{\text{res}} - 1$   $(\varphi, \psi)$ -tuples of the peptide are saved in the matrix  $\mathbf{A} \in \mathbb{R}^{F \times N_{\text{res}} - 1}$ :

$$\mathbf{A} = \begin{pmatrix} (\varphi, \psi)_{11} & (\varphi, \psi)_{12} & \dots & (\varphi, \psi)_{1(N_{\text{res}}-1)} \\ \vdots & \vdots & \ddots & \vdots \\ (\varphi, \psi)_{F1} & (\varphi, \psi)_{F1} & \dots & (\varphi, \psi)_{F(N_{\text{res}}-1)} \end{pmatrix} \quad (3)$$

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  $\beta$ -sheet basin and 1 for the right-handed  $\alpha$ -helical region. To assign an overall structural tendency to the peptide at the different time points  $t$ , the matrix entries along the rows are being summed up, producing one number per time frame:

$$a_i = \sum_j (\mathbf{A}_c)_{ij} \equiv a \left( \begin{pmatrix} t \end{pmatrix} \right) \quad (4)$$

It should be mentioned that the numbers chosen to denote the cluster centers need consideration. If one only wants to distinguish between  $\beta$ -strands and  $\alpha$ -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  $\alpha$ -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  $N_{\text{res}} - 1$  in order to avoid overlap with the case  $a_i = N_{\text{res}} - 1$  where all  $N_{\text{res}} - 1$  residues adopt an  $\alpha$ -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  $\beta$ -strands.** If one studies the aggregation of proteins into amyloid structures, which are characterized by  $\beta$ -sheet layers that form fibrils, one is usually interested to know whether these are parallel or antiparallel  $\beta$ -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  $\beta$ -strands [39,40]. 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  $\beta$ -strands, −1 for antiparallel alignments, and 0 or neither of them, including mixed parallel and antiparallel alignments as well as disordered  $\beta$ -strands:

$$f = \begin{cases} 1, & P_2 \geq 0.7 \text{ and } P_1 \geq 0.7 \\ -1, & P_2 \geq 0.7 \text{ and } P_1 \leq -0.7 \\ 0, & \text{else} \end{cases} \quad (5)$$

### 2.3. ATRANET software for transition-network calculation

Transition networks based on a combination of the descriptor functions explained in the previous section can be calculated with ATRANET, which is a Python script developed for this purpose by our group [29]. The script is designed in a way that it allows the user to easily add further descriptor functions if needed, that is if they seem to be beneficial for the system under investigation. To perform a TN analysis using ATRANET, the topology and trajectory files of all-atom MD simulations containing any number of polypeptide chains have to be provided. In its current version, these input files have to be in Gromacs format (.gro and .xtc) [41], but in the future this will be extended so that trajectory files originating from other MD software can be directly processed too. The ATRANET script invokes the MD trajectory analysis software MDTraj [42] and MDAnalysis [43] to calculate the descriptors specified by the user. It then produces a file that contains the transition matrix, including the states' populations and the number of transitions between the states. The transition matrix is saved as a .gexf file, which can be visualized by the network visualization software Gephi [44]. The ATRANET script is available at <https://github.com/strodel-group/ATRANET>.

### 2.4. Gephi software for transition-network visualization

The visualization of the transition matrix creates a plot that shows the states as circles of different sizes according to their population and edges between the nodes. The edges can be shown with different thickness reflecting the underlying transition probability and arrows to indicate the direction of the transition from one node to another node. In addition to visualizing the transition matrix, the open-source software Gephi [45] also provides a number of powerful tools to further analyze some basic properties of TNs. One often used analysis function is the calculation of the in-/out-degree of the nodes. The degree of a node in a network is the number of edges the node shares with other nodes. If a network is directed, the in-degree corresponds to the number of incoming edges, and the out-degree is the number of outgoing edges. Gephi further allows to visualize the network based on the calculated properties, e.g., to color the nodes according to their modularity class and choose the node size with respect to their degree or population. Lastly, one can choose different algorithms to get a layout of the network based on internal parameters, such as the Yifan Hu or ForceAtlas 2 algorithm.

## 3. Materials and methods

In order to demonstrate the use of TNs in studying the conformational transitions of IDPs, we apply the transition-network procedure explained above to  $4 \times 1 \mu\text{s}$  MD trajectories of  $A\beta_{42}$  in complex with three POPC lipid molecules.

### 3.1. Molecular dynamics simulations

The MD simulations were originally performed in the framework of another study that unraveled a disorder-to-order transition of  $A\beta_{42}$  following its binding to individual lipid molecules [36]. The simulations were performed using the Gromacs 2018 software [41]. The  $A\beta_{42}$  peptide was modeled using the CHARMM36mW force field [46], as it was found to be suitable for modeling both monomeric  $A\beta_{42}$  [11] as well as amyloid aggregation [47]. As CHARMM36mW only contains parameters for polypeptides, we used the original force field CHARMM36 [48] for modeling the POPC lipids. The peptide and three lipid molecules were initially placed away from each other in the simulation box, which was filled with TIP3P water molecules [49] and NaCl at a physiological salt concentration of 150 mM. After appropriate equilibration of the system it was simulated four times under  $NpT$  conditions (i.e., a constant number of atoms  $N$ , a constant pressure  $p$  of 1 bar, and a constant temperature of 310 K) for 1  $\mu\text{s}$  each, amounting to a total of 4  $\mu\text{s}$  of

simulation time. Data was written out every 100 ps, resulting in 40,000 MD frames for analysis. For pressure regulation the Parrinello-Rahman pressure coupling scheme [50] was used, while the temperature was controlled via a velocity-rescaling thermostat allowing for canonical sampling [51]. Simulations were carried out under periodic boundary conditions in all directions using the particle-mesh Ewald method [52] for calculating the electrostatic interactions. Cutoffs for the Coulomb interactions calculated in real space and the van der Waals interactions were both set to 12 Å.

### 3.2. Transition networks

The TNs of  $A\beta_{42}$  in the presence of three POPC lipids were calculated with the ATRANET Python script. To demonstrate the influence of the choice of descriptor functions on the resulting networks, two descriptor sets were used. First, we chose three descriptors to characterize the conformations of  $A\beta_{42}$ : i) the number of residues exhibiting  $\alpha$ -helical secondary structure, ii) the number of residues adopting a  $\beta$ -sheet structure, and iii) the end-to-end distance. For the latter we used a bin size of 5 Å to transform the continuous distance into a discrete variable. For the second choice of descriptors, we used the same three descriptors as before and added a fourth one, which is the number of  $A\beta_{42}$  residues being in contact with any of the three POPC lipids. We defined a residue to be in contact if any of its atoms is within a cutoff of 6 Å of a lipid.

To better understand how these descriptors lead to the assignment of a particular  $A\beta_{42}$  conformation to a node, we provide an example. Let us assume we sampled an  $A\beta_{42}$  structure with a short  $\alpha$ -helix involving only five residues in the N-terminal part of the peptide and a  $\beta$ -hairpin in the C-terminal peptide half where in total eight of the residues are in a  $\beta$ -sheet conformation. The two peptide ends are rather close to each other, having a distance of 7.8 Å from each other. The three descriptors thus are (5, 8, 7.8). Since the end-to-end distances are assigned to 5 Å-wide bins, which can be understood as coarse-graining, this  $A\beta_{42}$  conformation will belong to the node with descriptors (5, 8, [5, 10]), since  $5 \leq 7.8 < 10$  Å, and all structures with these three descriptors will belong to that node. If we take the fourth descriptor, the number of peptide-lipid contacts into account and let us suppose that  $A\beta_{42}$  formed 43 atomic contacts with the lipid, the node description changes to (5, 8, [5, 10], 43).

### 3.3. Visualization using Gephi

The transition matrices calculated by ATRANET were then visualized using Gephi. Here, we briefly present our choice of TN visualization to give the reader a rough guideline for useful Gephi functionality selections; however, the optimal TN representation depends on the system under study and research question being addressed.

For the size of the nodes we chose it to be proportional to the population of the corresponding state, which is calculated by the ATRANET script. This allows to easily identify predominant and therefore more stable states by their relative size compared to those of the other nodes. This Gephi setting is being accomplished via the *Size* → *Ranking* tab and selecting the attribute *Size*.

The layout of a TN is important as a good representation allows to quickly grasp the connectedness of the nodes and identify the different clusters of the system. When the transition-matrix data is loaded into Gephi, the nodes are initially randomly placed. In order to give the relative placement of each node meaning, one can choose between different algorithms that determine the relative placement of the nodes based on their interconnectedness. The algorithms can be executed by selecting them in the *Layout* window and pressing *Run*. Here, we first ran the *Yifan Hu Proportional* algorithm to get a rough outline of the network, followed by the *Force Atlas 2* algorithm together with the *Prevent Overlap* option, which expands the network and makes sure that there are no nodes hidden behind one another.

Depending on the number of descriptors used, the resulting number

of nodes can be quite overwhelming. For this reason, we used the *Modularity* feature in the *Statistics* tab of Gephi to identify clusters of nodes. This clustering method, which is known as the Louvain method, orders the nodes into communities (called clusters herein) based on their connectedness and a user-defined resolution. Details of this method can be found in [53]. In a nutshell, the algorithm is divided into two phases that are repeated iteratively. It starts with the weighted network of  $N$  nodes and initially, each node is assigned to a different community. Then, for each node  $i$  it is checked if the modularity increases if it is considered to form a community with another node  $j$ . The modularity is a measure of the density of links inside communities as compared to links between communities. It ranges between  $-1$  and  $1$  (optimal) and can thus be used as an objective function to be maximized. Node  $i$  is placed in the community for which the gain in modularity is maximal. The process of the assignment of nodes to communities is applied sequentially to all nodes and repeated until no further improvement can be achieved, finishing the first phase. The second phase of the algorithm consists in building a new network whose nodes are now the communities found during the first phase. To this end, the weights of the links between the new nodes have to be determined, which are the sums of the weights of the links between nodes of two communities. Once this second phase is completed, one can reapply the first phase of the algorithm to the resulting weighted network and iterate. With each such iteration the number of communities decreases, resulting in a coarser clustering. In Gephi, the number of iterations is controlled by (but is not identical to) the user-defined resolution, which we chose as  $\approx 1$  for the TN based on three descriptors and  $\approx 5$  when four descriptors were used, as this choice led to six clusters per network, allowing for a direct comparison between them. The nodes were then colored according to their cluster (or community) membership. This can be achieved via the *Color* → *Partition* tab, selecting there the attribute *Modularity Class*.

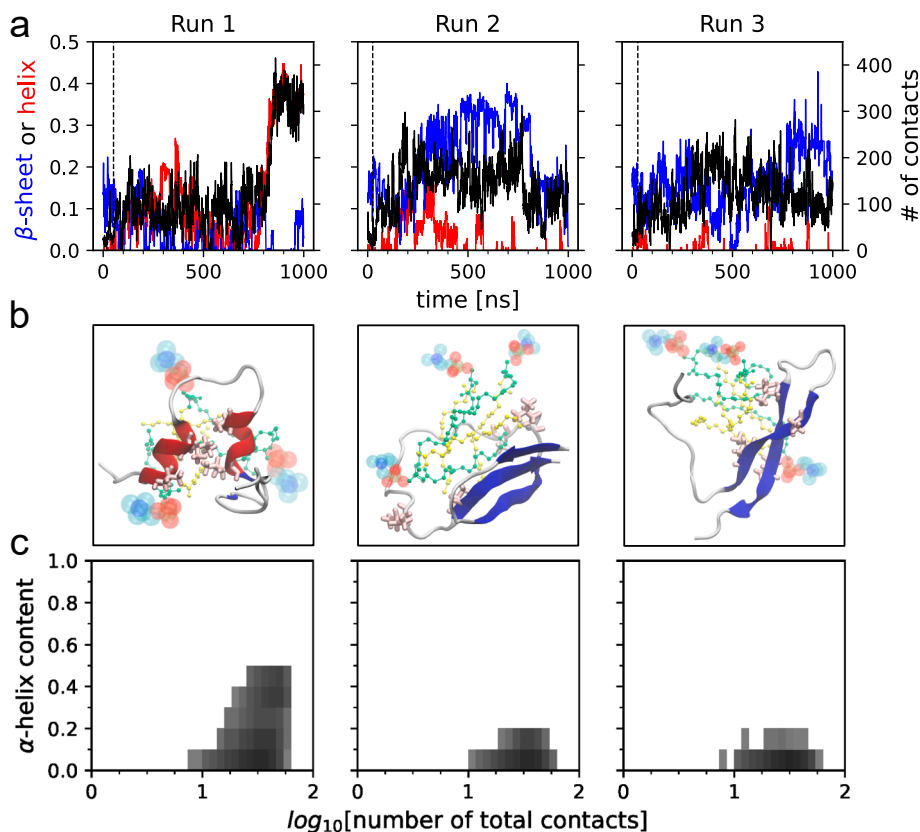
## 4. Results and discussion

### 4.1. Disorder-to-order transition of $A\beta_{42}$ upon lipid binding

In our previous study [36] we demonstrated that  $A\beta_{42}$  becomes folded when it forms a complex with more than one lipid. At a 1:1 ration the peptide remained mainly disordered, yet at peptide:lipid ratio of 1:3,  $A\beta_{42}$  underwent a random coil-to-helix transition or a random coil-to- $\beta$ -sheet transition (Fig. 1a). That study was based on  $3 \times 1 \mu\text{s}$  MD simulations of 1:3 complexes. In all three simulations (denoted as run 1 to run 3) 1:3 complex formation readily occurred within the first 50 ns, and the complexes did not disintegrate in the remainder of the simulations. In run 1 we observed a high  $\alpha$ -helix content, where about 45% of the  $A\beta_{42}$  residues adopted a helical conformation after  $\approx 800$  ns of simulation time. Inspection of the helical conformation that formed reveals that it is a helix-kink-helix structure (Fig. 1b), which was stable in the last 200 ns of the simulation. In the other two runs, transient helix formation was observed too, especially in the first 500 ns of run 2, yet in the end the dominating secondary structure were  $\beta$ -sheets (Fig. 1b). We thus concluded that, when  $A\beta_{42}$  is surrounded by a sufficient number of lipids, it undergoes a disorder-to-order transition, which is a characteristic of IDPs, which commonly adopt more ordered states upon binding with their targets.

In order to understand the driving force behind the folding of  $A\beta_{42}$ , we calculated the number of contacts formed between the peptide and the POPC lipids. The number of these contacts is correlated with the structure formation in  $A\beta_{42}$  (Fig. 1c). For run 1, one can clearly see that the more residues adopt a helical structure, the more contacts between  $A\beta_{42}$  and the lipids exist. However, a large number of  $A\beta_{42}$ -POPC contacts is not a guarantee for an  $\alpha$ -helix to be formed, as for the other two runs, where the helical content did not increase beyond 20%, a similar amount of peptide-lipid contacts were formed. We thus assumed that the  $\alpha$ -helix content is likely to be correlated with specific residue-POPC contacts, rather than the total number of contacts. This hypothesis was





**Fig. 1.** Disorder-to-order transition of  $A\beta_{42}$  following its binding to three POPC lipids. (a) Evolution of the  $\beta$ -sheet (blue) and helix content (red) as well as the number of atom–atom contacts formed between the POPC lipids and  $A\beta_{42}$  residues Leu17, Ala21, Ile32, and Val36 obtained from the three simulations of 1:3  $A\beta_{42}$ –POPC complexes. The dashed vertical lines mark time when all three POPC lipids had bound to  $A\beta_{42}$ . (b) Representative snapshots showing the helix-kink-helix structure that formed at the end of run 1 and  $\beta$ -sheet structures sampled in runs 2 and 3. The peptide is shown as cartoon and the sidechains of Leu17, Ala21, Ile32, and Val36 are explicitly shown. The lipid headgroups are indicated by blue and red spheres, and the oleoyl and palmitoyl chains are represented as ball-and-stick model in green and yellow color, respectively. (c) Two-dimensional histograms of the number of  $A\beta_{42}$ –POPC contacts plotted against the  $\alpha$ -helix content. The darker the color of a bin, the more likely this contact-helix combination is. This figure was reproduced from Ref. [36] with permission from Elsevier.

tested by further analysis, and indeed we found that the stability of the helix-kink-helix structure formed in run 1 results from contacts between the POPC lipids and residues Leu17, Ala21, Ile32, and Val36. This can be seen from the plot of the evolution of the number of atom–atom contacts formed between POPC and Leu17, Ala21, Ile32, and Val36 along with the evolution of the  $\alpha$ -helix content (Fig. 1a). For run 1 we notice that the helical structure is only stable when the mentioned residues are simultaneously in contact with the lipids. This is also highlighted in the representative helix-kink-helix structure that is shown in Fig. 1b. It shows the interactions between the lipid tails and Leu17 and Ala21 of the first helix as well as Ile32 and Val36 of the second helix. In run 2, the amount of these contacts initially rises, leading to a transient increase in the helical content, yet at  $\approx 500$  ns both the number of these contacts and the helical content drop. Instead,  $\beta$ -sheet structures developed in runs 2 and 3, for which representative structures are shown in Fig. 1b. One can see that  $\beta$ -sheets can develop in different parts of the peptide. Visualization of the trajectories and analysis of the time evolution of the secondary structure on a per-residue basis further revealed that the formation of  $\beta$ -sheets at different time points of the same trajectory seems to be independent of each other, but it is highly correlated with intrapeptide and peptide–lipid interactions.

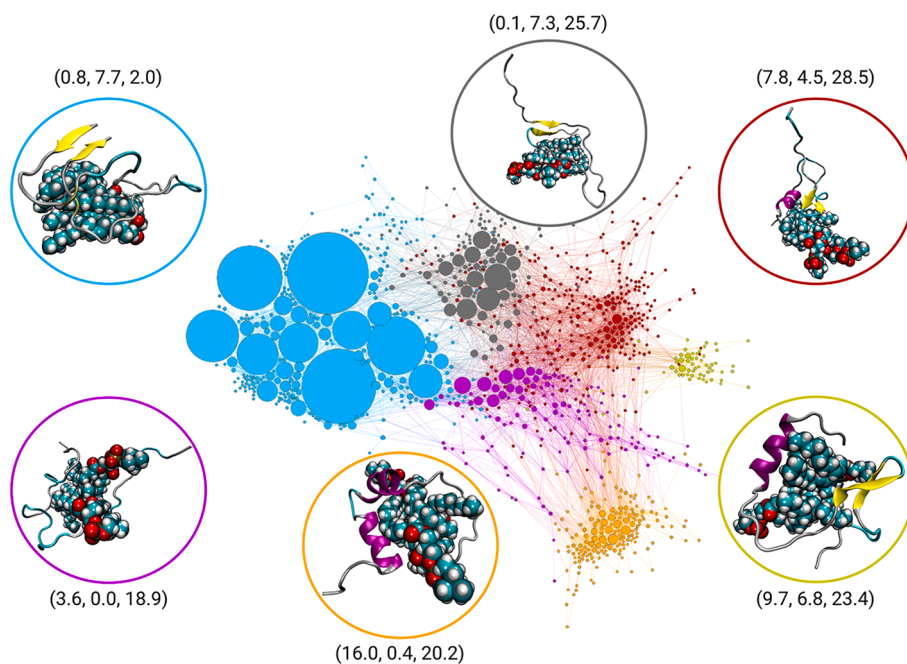
Here, we revisited the complex formed by one  $A\beta_{42}$  peptide and three POPC lipids. First, we ran an additional 1  $\mu$ s simulation and then analyzed the concatenated trajectories based on TNs using either three or four descriptors for characterizing  $A\beta_{42}$ . By concatenating the trajectories, we obtain three spurious transitions, where the previous trajectory ends and the next trajectory begins. One can manually remove these extra transitions from the transition matrix calculated by AT-RNET. However, as these are only three extra transitions, they will either show up as faint edges in the TN visualized by Gephi when not sampled otherwise, or, if the same kind of transitions also occurred naturally, then the corresponding transition counts are only increased by one transition. Thus, even if one decides not to remove these spurious

transitions from the transition matrix, the appearance of the resulting TN will not be affected.

#### 4.2. Transition network using three descriptors for $A\beta_{42}$

First, we produced a TN using three descriptors to describe the conformational preferences of  $A\beta_{42}$ : i) the number of residues exhibiting  $\alpha$ -helical secondary structure ( $N_\alpha$ ), ii) the number of residues adopting a  $\beta$ -sheet structure ( $N_\beta$ ), and iii) the end-to-end distance ( $d_{NC}$ , in Å). The resulting TN is shown in Fig. 2 together with a representative structure for each of the six clusters, for which the most populated state per cluster was used. The TN is dominated by the cluster colored in blue. Other noteworthy clusters are the ones shown in orange and lime, while the three other clusters seem to be intermediate states that partly overlap and are not well separated in this TN representation.

The cluster shown in blue includes the nodes with the largest size, which means they correspond to more stable configurations that are visited more often compared to the other states. With average descriptor values of ( $N_\alpha = 0.8, N_\beta = 7.7, d_{NC} = 2.0$  Å) the cluster encompasses states with a small end-to-end distances that exhibit predominantly  $\beta$ -sheet structures. On the contrary, the other distinct cluster shown in orange, is visited less frequently and with average descriptor values of ( $N_\alpha = 16.0, N_\beta = 0.4, d_{NC} = 20.2$  Å) the configurations exhibit a rather large end-to-end distance with mostly  $\alpha$ -helical structures. The other clusters seem to be intermediate states between the ones colored in blue and orange, like the one shown in gray with average values of ( $N_\alpha = 0.1, N_\beta = 7.3, d_{NC} = 25.7$  Å). While it is similar to the blue-colored cluster in terms of secondary structure, the configurations within the gray cluster are more extended with larger end-to-end distances. Thus, a transition from the blue- to the gray-colored cluster happens when the termini move away from each other while there is still a  $\beta$ -sheet present in the peptide. A transition from the gray- to the orange-colored cluster involves the intermediate clusters colored in red



**Fig. 2.** Transition network of  $A\beta_{42}$  forming a complex with three POPC lipids using three descriptors: i) the number of residues having an  $\alpha$ -helical structure, ii) the number of residues having a  $\beta$ -sheet structure, iii) the end-to-end distance. The nodes are colored according to their cluster membership. Representative structures for the most populated state of each cluster are shown within circles of the same color as used for the corresponding cluster. The average descriptor values ( $N_\alpha, N_\beta, d_{NC}$ ) per cluster are provided too.

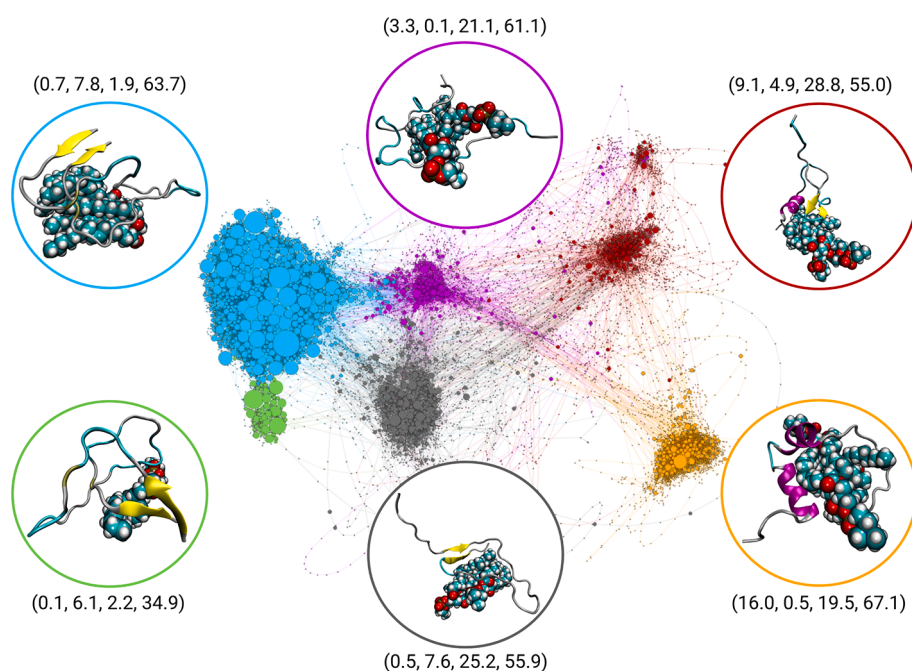
and magenta. The red-colored states exhibit a mixture of  $\alpha$ -helical and  $\beta$ -sheet structures and can lead to the lime-colored cluster with average descriptor values of ( $N_\alpha = 9.7, N_\beta = 6.8, d_{NC} = 23.4 \text{ \AA}$ ) where on the one end of the peptide a helix is present and a  $\beta$ -hairpin dominates the structure at the other peptide end. The alternative route implies the complete dissolution of the  $\beta$ -sheet, giving rise to an intermediate cluster colored in magenta where  $A\beta_{42}$  is in a random-coil state with ( $N_\alpha = 3.6, N_\beta = 0.0, d_{NC} = 18.9 \text{ \AA}$ ), which eventually leads to alpha-helical structure represented by the orange-colored cluster.

In summary, the TN based on three descriptors nicely reveals the different structures that  $A\beta_{42}$  adopts when being bound to a small lipid cluster formed by three POPC lipids, how these structures can interconvert, and that the  $\beta$ -sheet structure is the favored one. With the three

descriptors chosen, it is however not possible to unveil the interplay between structure formation and peptide–lipid interactions. This becomes only possible when including a descriptor for the peptide–lipid interactions, which is therefore added to the three existing descriptors. The resulting TN based on these four descriptors is discussed next.

#### 4.3. Transition network using four descriptors of $A\beta_{42}$

We performed the same analysis for the four descriptors ( $N_\alpha, N_\beta, d_{NC}, N_{p-l}$ ) where  $N_{p-l}$  refers to the number of contacts formed between the  $A\beta_{42}$  peptide and the POPC lipids. As before, the clusters were calculated using the modularity class feature of Gephi and the coloring of the cluster was chosen to be directly comparable to that of the clusters in



**Fig. 3.** Transition network of  $A\beta_{42}$  forming a complex with three POPC lipids using four descriptors: i) the number of residues having an  $\alpha$ -helical structure, ii) the number of residues having a  $\beta$ -sheet structure, iii) the end-to-end distance, iv) the number of peptide–lipid contacts. The nodes are colored according to their cluster membership. Representative structures for the most populated state of each cluster are shown within circles of the same color as used for the corresponding cluster. The average descriptor values ( $N_\alpha, N_\beta, d_{NC}, N_{p-l}$ ) per cluster are provided too.

Fig. 2; however, the cluster membership of a specific configuration might have changed due to the new definition of states. Fig. 3 shows the resulting TN together with representative structures taken from the most populated state of each cluster.

A first look at this TN reveals that, thanks to the additional fourth descriptor, the clusters have become better separated from each other and the transitions between them are clearer. This implies that this extra descriptor  $N_{p-1}$  contains additional information not included yet in the first three descriptors ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}$ ). Nonetheless, the clusters remain largely the same when using the four instead of only three descriptors. Only the former lime-colored cluster has become part of the red-colored cluster, while the most populated blue-colored cluster has split into two clusters: one dominant cluster (still shown in blue) as well as a smaller cluster (shown in green). The main difference between them is the number of peptide–lipid contacts, which is almost half the size in the green-colored cluster compared to the blue-colored cluster:  $N_{p-1} = 34.9$  and  $N_{p-1} = 63.7$ , respectively. The other three descriptors have very similar values in these two clusters and correspond to a compact  $\beta$ -sheet structure where the N- and C-terminus are close to each other. The connectedness of the green-colored cluster reveals that dissociation of the peptide from the lipid cluster does not lead to states with more  $\alpha$ -helical or random coil structures, as the green-colored cluster is almost exclusively connected to the blue-colored cluster. Transitions to these other states start from the blue-colored cluster, either leading to the magenta-colored cluster harboring random coil conformations of  $A\beta_{42}$  or to the gray-colored cluster where the  $\beta$ -sheet is still present and attached to the lipid cluster, while the N-terminus has detached from the rest of the peptide and the lipids. Mixed  $\alpha$ -helical/ $\beta$ -sheet structures, which are represented by the red-colored cluster, can be formed from both the magenta- and gray-colored cluster, while the helix-only structure contained within the orange-colored cluster evolves mainly from the random-coil conformation represented by the magenta-colored cluster or, to a lesser degree, via a transient mixture of both  $\alpha$ -helix and  $\beta$ -sheet (red-colored cluster).

With the extra descriptor  $N_{p-1}$ , the different pathways between the clusters have become much better visible. Another important finding is that the helix-only state (orange-colored cluster) has the highest average number of contacts with the lipids. The representative structure shows that in this case the peptide wraps around the lipids, whereas the  $\beta$ -sheet is rather formed on the surface of the lipid cluster. This observation confirms the observations from our previous study [36] and might be of relevance for the process of amyloid aggregation in the presence of lipid membranes. Another important finding is that the structural transitions within the peptide can take place without the peptide needing to detach from the lipid cluster. This finding was not uncovered previously from our more standard analysis of the behavior of  $A\beta_{42}$  in complex with POPC lipids [36].

## 5. Conclusions

Studying the structural characteristics of IDPs remains challenging to the present date. Compared to folded proteins, IPDs do not exhibit a unique three-dimensional structure. The absence of a well-defined structure in combination with the conformational transitions spanning several time scales makes it difficult to study the corresponding characteristics of IDPs in experiments. A complement to experiments is provided by MD simulations. Here, we demonstrated how the structural changes of an IDP sampled during all-atom MD simulations can be elucidated by calculating transition networks (TNs) from the MD data.

We identified three main features of TNs by analyzing 4  $\mu$ s of MD trajectories of  $A\beta_{42}$  in complex with three POPC lipids. Firstly, from the size of the nodes one can directly identify the most stable polypeptide configurations, which are represented by the largest nodes. Secondly, one can directly comprehend the connectedness between adjacent nodes. We chose the layout of the TNs such that, when there are fewer

transitions between them, they were placed further away in the network. This also extends to the cases where the shortest pathway between two nodes requires the system to visit many intermediate nodes, which are then also located further away from each other. Thus, the spatial distance between nodes in the TN is directly related to the phase-space distance between the corresponding configuration of the polypeptide. Lastly, one can generate a coarser view of a TN by identifying clusters within the TN, making the TN and the underlying dynamics more comprehensible. The transition from one cluster to another one entails significant conformational changes in the polypeptide, whereas the conformations represented by nodes of the same cluster are relatively similar.

The choice of the descriptors to characterize the polypeptide influences the features of the resulting TN. We demonstrated this by creating TNs of  $A\beta_{42}$  for two different sets of descriptors. First we used three descriptors, namely the numbers of residues exhibiting  $\alpha$ -helical secondary structure, the number of residues forming  $\beta$ -sheets, and the end-to-end distance. Then, we generated the TN for a second set of descriptors where we used the same descriptors as before and added a fourth one, which is the number of contacts between  $A\beta_{42}$  and the lipids. The additional descriptor yielded a more granular network, implying a generally lower population per node. On the other hand, this additional descriptor turned out to be advantageous, as it helped to separate the clusters from each other and thus made the conformational transition pathways better visible.

Limitations of the analysis using TNs are founded in the assumptions one imposes on the system by the choice of the descriptors. As demonstrated, that choice might influence the appearance of the TN and therefore the conclusions drawn from it. On the other hand, considering the research questions surrounding structural changes in IPDs, it is necessary to use descriptors that characterize the polypeptide configurations, such as the secondary structure or some measure of the polypeptide compactness. Moreover, since the calculation of a TN with ATRANET is rather straightforward and computationally not expensive, the user can try various descriptor combinations until they are satisfied with the resulting TN. ATRANET further helps in assessing the quality of the chosen descriptor combinations by quantifying the sensitivity of the descriptors and the correlation between them [29]. Another constraint of TNs is that, when they are based on insufficient sampling of the phase space, the resulting TN has limited meaning. However, one can also consider the possibility of creating TNs from incomplete phase-space sampling as an advantage, as this is not possible for Markov state modeling and allows early insight into molecular processes tested by MD simulations.

An interesting feature of TNs are their parallels to free energy surfaces (FESs). The reader might have noticed the similarities between analyzing a TN and how one would interpret the FES of a polypeptide, for the reason that they are both representations of the visited configurational space. In the case of IDPs, the presence of multiple distinct clusters in the TN with similar occupation probabilities, as found here for  $A\beta_{42}$ , corresponds to an underlying multifunnel structure for the energy landscape [24]. FESs have the advantage of containing information about the actual energy difference between states, which is a quantitative measure of the underlying dynamics. They usually rely on a projection of the free energy onto one or more (collective) variables, which are similar to the descriptors used in our TN analysis. However, the projection on one or two dimensions has the danger that minima or the transitions between them are not properly resolved, leading to an under- or overestimation of the free energy barrier for configurational transitions. An alternative is to use disconnectivity graphs to visualize the multidimensional free energy landscape of polypeptides [54–56]. They have the advantage that they do not rely on projecting the free energy on (collective) variables. In one of our earlier works we also showed how disconnectivity graphs can be created from TNs, using the node and edge data from the TNs [28]. Moreover, disconnectivity graphs created for IDPs are able to nicely reveal the multifunnel structure of



their energy landscapes [24]. Further analysis that can be extracted from TNs is information regarding the kinetics through first passage time distributions for reaching the most populated states. In our future work, we will implement the possibility of such additional free-energy and kinetics analysis into our ATRANET software.

## CRediT authorship contribution statement

**Moritz Schäffler:** Methodology, Software, Validation, Formal-analysis, Investigation, Data-curation, Writing-original-draft, Visualization. **Mohammed Khaled:** Software, Validation, Data-curation, Writing-original-draft. **Birgit Strodel:** Conceptualization, Methodology, Resources, Writing-review-editing, Supervision, Project-administration, Funding-acquisition.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The authors thank Dr. Hebah Fatafta for providing some of the MD simulation data and Dr. Batuhan Kav for fruitful discussions. M.K. and B. S. acknowledge funding for this project from the Palestinian-German Science Bridge financed by the German Federal Ministry of Education and Research (BMBF).

## References

- [1] V.N. Uversky, P. Kulkarni, Intrinsically disordered proteins: Chronology of a discovery, *Biophys. Chem.* 279 (2021), 106694, <https://doi.org/10.1016/j.bpc.2021.106694>.
- [2] N. Perdigão, J. Heinrich, C. Stolte, K. Sabir, M. Buckley, B. Tabor, B. Signal, B. Gloss, C. Hammang, B. Rost, A. Schafferhans, S. O' Donoghue, Unexpected features of the dark proteome, *Proc. Natl. Acad. Sci. U.S.A.* 112 (2015) 15898–15903, <https://doi.org/10.1073/pnas.1508380112>.
- [3] S. DeForte, V.N. Uversky, Order, disorder, and everything in between, *Molecules* 21 (8), doi:10.3390/molecules21081090.
- [4] A. Deiana, S. Forcelloni, A. Porrello, A. Giansanti, Intrinsically disordered proteins and structured proteins with intrinsically disordered regions have different functional roles in the cell, *PLOS ONE* 14 (2019), e0217889, <https://doi.org/10.1371/journal.pone.0217889>.
- [5] V. Uversky, Functional roles of transiently and intrinsically disordered regions within proteins, *FEBS J.* 282 (2015) 1182–1189, <https://doi.org/10.1111/febs.13202>.
- [6] V. Uversky, Intrinsically disordered proteins and their "mysterious" (meta)physics, *Front. Phys.* 7 (2019) 10, <https://doi.org/10.3389/fphy.2019.00010>.
- [7] D. Eliezer, Biophysical characterization of intrinsically disordered proteins, *Curr. Opin. Struct. Biol.* 19 (1) (2009) 23–30.
- [8] B.B. Kragelund, K. Skriver, *Intrinsically Disordered Proteins: Methods and Protocols*, Springer, 2020.
- [9] B. Strodel, Energy landscapes of protein aggregation and conformation switching in intrinsically disordered proteins, *J. Mol. Biol.* 433 (20) (2021), 167182, <https://doi.org/10.1016/j.jmb.2021.167182>.
- [10] H. Fatafta, S. Samantray, A. Sayyed-Ahmad, O. Coskuner-Weber, B. Strodel, Chapter five - molecular simulations of idps: From ensemble generation to idp interactions leading to disorder-to-order transitions, in: V.N. Uversky (Ed.), *Dancing Protein Clouds: Intrinsically Disordered Proteins in the Norm and Pathology*, Part C, Vol. 183 of Progress in Molecular Biology and Translational Science, Academic Press, 2021, pp. 135–185, doi:10.1016/bs.pmbts.2021.06.003.
- [11] A. Paul, S. Samantray, M. Anteghini, M. Khaled, B. Strodel, Thermodynamics and kinetics of the amyloid- $\beta$  peptide revealed by markov state models based on md data in agreement with experiment, *Chem. Sci.* 12 (2021) 6652–6669, <https://doi.org/10.1039/D0SC04657D>.
- [12] J.D. Chodera, F. Noé, Markov state models of biomolecular conformational dynamics, *Curr. Opin. Struct. Biol.* 25 (2014) 135–144, theory and simulation/Macromolecular machines, doi:10.1016/j.sbi.2014.04.002.
- [13] B.E. Husic, V.S. Pande, Markov state models: From an art to a science, *J. Am. Chem. Soc.* 140 (7) (2018) 2386–2396, <https://doi.org/10.1021/jacs.7b12191>.
- [14] F. Noé, E. Rosta, Markov models of molecular kinetics, *J. Chem. Phys.* 151 (19) (2019), 190401, <https://doi.org/10.1063/1.5134029>.
- [15] G.R. Bowman, D.L. Ensign, V.S. Pande, Enhanced modeling via network theory: Adaptive sampling of Markov state models, *J. Chem. Theory Comput.* 6 (3) (2010) 787–794, <https://doi.org/10.1021/ct900620b>.
- [16] F. Rao, A. Caflisch, The protein folding network, *J. Mol. Biol.* 342 (1) (2004) 299–306, <https://doi.org/10.1016/j.jmb.2004.06.063>.
- [17] F. Noé, S. Fischer, Transition networks for modeling the kinetics of conformational change in macromolecules, *Curr. Opin. Struct. Biol.* 18 (2) (2008) 154–162, <https://doi.org/10.1016/j.sbi.2008.01.008>.
- [18] D. Prada-Gracia, J. Gómez-Gardeñes, P. Echenique, F. Falo, Exploring the free energy landscape: From dynamics to networks and back, *PLOS Comput. Biol.* 5 (2009) 1–9, <https://doi.org/10.1371/journal.pcbi.1000415>.
- [19] D.J. Wales, Energy landscapes: some new horizons, *Curr. Opin. Struct. Biol.* 20 (1) (2010) 3–10, <https://doi.org/10.1016/j.sbi.2009.12.011>.
- [20] V.S. Pande, K. Beauchamp, G.R. Bowman, Everything you wanted to know about Markov state models but were afraid to ask, *Methods* 52 (1) (2010) 99–105, <https://doi.org/10.1016/j.ymeth.2010.06.002>.
- [21] N. Heilmann, M. Wolf, M. Kozłowska, E. Sedghamiz, J. Setzler, M. Brieg, W. Wenzel, Sampling of the conformational landscape of small proteins with monte carlo methods, *Sci. Rep.* 10, doi:10.1038/s41598-020-75239-7.
- [22] D.J. Wales, Discrete path sampling, *Mol. Phys.* 100 (2002) 3285–3305, <https://doi.org/10.1080/00268970210162691>.
- [23] D.J. Wales, Some further applications of discrete path sampling to cluster isomerization, *Mol. Phys.* 102 (9–10) (2004) 891–908, <https://doi.org/10.1080/00268970410001703363>.
- [24] Y. Chebaro, A.J. Ballard, D. Chakraborty, D.J. Wales, Intrinsically disordered energy landscapes, *Sci. Rep.* 5, doi:10.1038/srep10386, URL: doi:10.1038/srep10386.
- [25] M. Kurcinski, A. Kolinski, S. Kmiecik, Mechanism of folding and binding of an intrinsically disordered protein as revealed by ab initio simulations, *J. Chem. Theory Comput.* 10 (2014) 2224–2231, <https://doi.org/10.1021/ct500287c>.
- [26] B. Strodel, D.J. Wales, Implicit solvent models and the energy landscape for aggregation of the amyloidogenic kffe peptide, *J. Chem. Theory Comput.* 4 (2008) 657–672, <https://doi.org/10.1021/ct700305w>.
- [27] M. Carballo-Pacheco, I. Vancea, B. Strodel, Extension of the facts implicit solvation model to membranes, *J. Chem. Theory Comput.* 10 (8) (2014) 3163–3176, <https://doi.org/10.1021/ct500084y>.
- [28] 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 (4) (2014) 1003–1011, <https://doi.org/10.1021/jp412648u>.
- [29] 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 (12) (2020) 7825–7839, <https://doi.org/10.1021/acs.jctc.0c00727>.
- [30] H. Fatafta, M. Khaled, M.C. Owen, A. Sayyed-Ahmad, B. Strodel, Amyloid- $\beta$  peptide dimers undergo a random coil to  $\beta$ -sheet transition in the aqueous phase but not at the neuronal membrane, *Proc. Natl. Acad. Sci. U.S.A.* 118 (39) (2021), e2106210118, <https://doi.org/10.1073/pnas.2106210118>.
- [31] J.A. Hardy, G.A. Higgins, Alzheimer's disease: The amyloid cascade hypothesis, *Science* 256 (5054) (1992) 184–185, <https://doi.org/10.1126/science.1566067>.
- [32] G.F. Chen, T.H. Xu, Y. Yan, Y.R. Zhou, Y. Jiang, K. Melcher, H.E. Xu, Amyloid beta: structure, biology and structure-based therapeutic development, *Acta Pharmacologica Sinica* 2017 38:9 38 (2017) 1205–1235, doi:10.1038/aps.2017.28.
- [33] P.H. Nguyen, A. Ramamoorthy, B.R. Sahoo, J. Zheng, P. Faller, J.E. Straub, L. Dominguez, J.-E. Shea, N.V. Dokholyan, A. De Simone, B. Ma, R. Nussinov, S. Najafi, S.T. Ngo, A. Loquet, M. Chiricotto, P. Ganguly, J. McCarty, M.S. Li, C. Hall, Y. Wang, Y. Miller, S. Melchionna, B. Habenstein, S. Timr, J. Chen, B. Hnath, B. Strodel, R. Kaye, S. Lesné, G. Wei, F. Sterpone, A.J. Doig, P. Derreumaux, Amyloid oligomers: A joint experimental/computational perspective on alzheimer's disease, parkinson's disease, type II diabetes, and amyotrophic lateral sclerosis, *Chem. Rev.* 121 (4) (2021) 2545–2647, <https://doi.org/10.1021/acs.chemrev.0c01122>.
- [34] J. Roche, Y. Shen, J.H. Lee, J. Ying, A. Bax, Monomeric  $\alpha$ 1–40 and  $\beta$ 1–42 peptides in solution adopt very similar ramachandran map distributions that closely resemble random coil, *Biochemistry* 55 (5) (2016) 762–775, <https://doi.org/10.1021/acs.biochem.5b01259>.
- [35] T. Kakeshpor, V. Ramanujam, C.A. Barnes, Y. Shen, J. Ying, A. Bax, A lowly populated, transient  $\beta$ -sheet structure in monomeric  $\alpha$ 1–42 identified by multinuclear NMR of chemical denaturation, *Biophys. Chem.* 270 (2021), 106531, <https://doi.org/10.1016/j.bpc.2020.106531>.
- [36] H. Fatafta, B. Kav, B.F. Bundschuh, J. Loschwitz, B. Strodel, Disorder-to-order transition of the amyloid- $\beta$  peptide upon lipid binding, *Biophys. Chem.* 280 (2022), 106700, <https://doi.org/10.1016/j.bpc.2021.106700>.
- [37] W. Kabsch, C. Sander, Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features, *Biopolymers* 22 (12) (1983) 2577–2637, <https://doi.org/10.1002/bip.360221211>.
- [38] W.G. Touw, C. Baakman, J. Black, T.A. te Beek, E. Krieger, R.P. Joosten, G. Vriend, A series of PDB-related databanks for everyday needs, *Nucleic Acids Res.* 43 (D1) (2014) D364–D368, <https://doi.org/10.1093/nar/gku1028>.
- [39] M. Cecchini, F. Rao, M. Seeber, A. Caflisch, Replica exchange molecular dynamics simulations of amyloid peptide aggregation, *J. Chem. Phys.* 121 (21) (2004) 10748–10756, <https://doi.org/10.1063/1.1809588>.
- [40] K.L. Osborne, M. Bachmann, B. Strodel, Thermodynamic analysis of structural transitions during GNNQQNY aggregation, *Proteins: Struct. Funct. Bioinf.* 81 (7) (2013) 1141–1155, <https://doi.org/10.1002/prot.24263>.
- [41] M.J. Abraham, T. Murtola, R. Schulz, S. Páll, J.C. Smith, B. Hess, E. Lindahl, Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, *SoftwareX* 1 (2015) 19–25.
- [42] R.T. McGibbon, K.A. Beauchamp, M.P. Harrigan, C. Klein, J.M. Swails, C. X. Hernández, C.R. Schwantes, L.-P. Wang, T.J. Lane, V.S. Pande, MDTraj: A

- modern open library for the analysis of molecular dynamics trajectories, *Biophys. J.* 109 (8) (2015) 1528–1532, <https://doi.org/10.1016/j.bpj.2015.08.015>.
- [43] N. Michaud-Agrawal, E.J. Denning, T.B. Woolf, O. Beckstein, MDAnalysis: A toolkit for the analysis of molecular dynamics simulations, *J. Comput. Chem.* 32 (10) (2011) 2319–2327, <https://doi.org/10.1002/jcc.21787>.
- [44] M. Bastian, S. Heymann, M. Jacomy, Gephi: An open source software for exploring and manipulating networks, *Proceedings of the International AAAI Conference on Web and Social Media* 3 (1) (2009) 361–362.
- [45] M. Bastian, S. Heymann, M. Jacomy, Gephi: An open source software for exploring and manipulating networks (2009).
- [46] J. Huang, S. Rauscher, G. Nawrocki, R. Ting, M. Feig, B. de Groot, H. Grubmüller, A. MacKerell, Charmm36m: An improved force field for folded and intrinsically disordered proteins, *Nature Meth.* 14 (2017) 71–73, <https://doi.org/10.1038/nmeth.4067>.
- [47] 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 (12) (2021) 6462–6475.
- [48] R.W. Pastor, A.D. MacKerell, Development of the charmm force field for lipids, *J. Phys. Chem. Lett.* 2 (13) (2011) 1526–1532, <https://doi.org/10.1021/jz200167q>.
- [49] W.L. Jorgensen, J. Chandrasekhar, J.D. Madura, R.W. Impey, M.L. Klein, Comparison of simple potential functions for simulating liquid water, *J. Chem. Phys.* 79 (2) (1983) 926–935, <https://doi.org/10.1063/1.445869>.
- [50] H.J. Berendsen, J.v. Postma, W.F. van Gunsteren, A. DiNola, J.R. Haak, Molecular dynamics with coupling to an external bath, *J. Chem. Phys.* 81 (8) (1984) 3684–3690.
- [51] G. Bussi, D. Donadio, M. Parrinello, Canonical sampling through velocity rescaling, *J. Chem. Phys.* 126 (1) (2007), 014101.
- [52] T. Darden, D. York, L. Pedersen, Particle mesh Ewald: An Nlog(N) method for Ewald sums in large systems, *J. Chem. Phys.* doi:10.1063/1.464397.
- [53] V. Blondel, J.-L. Guillaume, R. Lambiotte, E. Lefebvre, Fast unfolding of communities in large networks, *J. Stat. Mech.* 2008 (2008) P10008, <https://doi.org/10.1088/1742-5468/2008/10/P10008>.
- [54] O.M. Becker, M. Karplus, The topology of multidimensional potential energy surfaces: Theory and application to peptide structure and kinetics, *J. Chem. Phys.* 106 (4) (1997) 1495–1517, <https://doi.org/10.1063/1.473299>.
- [55] D. Wales, M. Miller, T. Walsh, Archetypal energy landscapes, *Nature* 394 (1998) 758–760, <https://doi.org/10.1038/29487>.
- [56] J.A. Joseph, K. Röoder, D. Chakraborty, R.G. Mantell, D.J. Wales, Exploring biomolecular energy landscapes, *Chem. Commun.* 53 (2017) 6974–6988, <https://doi.org/10.1039/C7CC02413D>.



Article

# Transition Networks Unveil Disorder-to-Order Transformations in A $\beta$ Caused by Glycosaminoglycans or Lipids

Moritz Schäffler <sup>1,2</sup>, Suman Samantray <sup>1</sup>  and Birgit Strodel <sup>1,2,\*</sup> 

<sup>1</sup> Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52428 Jülich, Germany; mo.schaeffler@fz-juelich.de (M.S.); sumansamantray06@gmail.com (S.S.)

<sup>2</sup> Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

\* Correspondence: b.strodel@fz-juelich.de

**Abstract:** The aggregation of amyloid- $\beta$  (A $\beta$ ) peptides, particularly of A $\beta$ <sub>1–42</sub>, has been linked to the pathogenesis of Alzheimer’s disease. In this study, we focus on the conformational change of A $\beta$ <sub>1–42</sub> in the presence of glycosaminoglycans (GAGs) and 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids using molecular dynamics simulations. We analyze the conformational changes that occur in A $\beta$  by extracting the key structural features that are then used to generate transition networks. Using the same three features per network highlights the transitions from intrinsically disordered states ubiquitous in A $\beta$ <sub>1–42</sub> in solution to more compact states arising from stable  $\beta$ -hairpin formation when A $\beta$ <sub>1–42</sub> is in the vicinity of a GAG molecule, and even more compact states characterized by a  $\alpha$ -helix or  $\beta$ -sheet structures when A $\beta$ <sub>1–42</sub> interacts with a POPC lipid cluster. We show that the molecular mechanisms underlying these transitions from disorder to order are different for the A $\beta$ <sub>1–42</sub>/GAG and A $\beta$ <sub>1–42</sub>/POPC systems. While in the latter the hydrophobicity provided by the lipid tails facilitates the folding of A $\beta$ <sub>1–42</sub>, in the case of GAG there are hardly any intermolecular A $\beta$ <sub>1–42</sub>–GAG interactions. Instead, GAG removes sodium ions from the peptide, allowing stronger electrostatic interactions within the peptide that stabilize a  $\beta$ -hairpin. Our results contribute to the growing knowledge of the role of GAGs and lipids in the conformational preferences of the A $\beta$  peptide, which in turn influences its aggregation into toxic oligomers and amyloid fibrils.

**Keywords:** intrinsically disordered proteins; molecular dynamics simulations; transition networks; amyloid- $\beta$ ; disorder-to-order transition



**Citation:** Schäffler, M.; Samantray, S.; Strodel, B. Transition Networks Unveil Disorder-to-Order Transformations in A $\beta$  Caused by Glycosaminoglycans or Lipids. *Int. J. Mol. Sci.* **2023**, *24*, 11238. <https://doi.org/10.3390/ijms241411238>

Academic Editor: David J. Wales

Received: 23 May 2023

Revised: 2 July 2023

Accepted: 4 July 2023

Published: 8 July 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Intrinsically disordered proteins (IDPs) are a class of proteins that do not exhibit a well-defined three-dimensional structure in their native state. Instead, IDPs adopt an ensemble of different conformations, which allows them to perform a variety of functions, such as cell signaling, cell cycle control, and protein–protein interaction, but are also associated with a variety of disease pathways. IDPs are able to bind to a wide range of interaction partners and often undergo a disorder-to-order transition upon binding, which can lead to the formation of new structures and the initiation of interaction-specific functions [1,2]. One heavily studied example of an IDP is the amyloid- $\beta$  peptide (A $\beta$ ), which is involved in the development of Alzheimer’s disease [3,4]. A $\beta$  is a peptide that can form aggregates called amyloid plaques in the brain, which are thought to contribute to the cognitive decline associated with Alzheimer’s disease. As typical for an IDP, A $\beta$  is able to bind to various interaction partners, resulting in conformational changes in A $\beta$ . For example, in a previous study we showed that A $\beta$  undergoes a disorder-to-order transition when in complex with three 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) lipids [5].

Molecular examination of A $\beta$  aggregate samples from Alzheimer’s disease affected patients have revealed a significant presence of charged polyelectrolytes, especially polysaccharides, belonging to the class of glycosaminoglycans (GAGs) [6]. GAGs are long chains

of repeating disaccharide units and are found in various tissues, including cartilage and the extracellular matrix. They play a critical role in maintaining the structural integrity of these tissues and also act as lubricants and shock absorbers [7–9]. Their elevated presence in A $\beta$  amyloid deposits suggests that GAGs may be involved in the formation and stability of A $\beta$  plaques. Biophysical studies have shown that GAGs promote aggregation, nucleation, and formation of amyloid fibrils; however, the molecular details are not yet known [6].

Due to the wide conformational heterogeneity of IDPs such as A $\beta$ , the experimental characterization of their structures is a challenging task [10,11]. Experimental techniques that can be applied to this goal usually average over a wide range of conformations, e.g., nuclear magnetic resonance (NMR) spectroscopy, small angle X-ray scattering (SAXS), cryo-electron microscopy, or single-molecule Förster resonance energy transfer (smFRET) spectroscopy. Consequently, the structural information obtained by these techniques is limited in the case of IDPs. Molecular dynamics (MD) simulations offer a complementary approach for gaining insights into the structural properties of IDPs, as they allow for the study of IDPs in a dynamic and spatiotemporal manner by simulating the motion of individual atoms over time, thus providing a molecular-level understanding of their conformational changes. Moreover, combining the MD approach with network-based models, such as Markov state models (MSMs) or conformational transition networks (TNs), yields a comprehensive understanding of the structural behavior of IDPs [12–16], as the networks reveal the underlying mechanisms of molecular processes that are hidden within the vast amounts of MD simulation data by generating human-interpretable networks that help to illustrate the molecular processes under investigation [17–19]. The TN approach pursued by our group is a solution for those who want a network model of protein motions captured by MD simulations, including explicit modeling of the protein's environment, but do not need a method that sets up a master equation for the dynamics, since TNs rely purely on geometric clustering to extract the crucial features of protein conformational transitions (and not on kinetic clustering as in Markov state modeling) [20].

In this study we compare the conformational ensembles of A $\beta$  under different external conditions: A $\beta$  alone in solution [15], A $\beta$  in contact with a small lipid cluster consisting of three POPC lipids [5], and A $\beta$  in interaction with a GAG. For the GAG, we chose a polymer involving sixteen chondroitin-4-sulfate subunits, which are the alternating monosaccharides D-galactosamine (GalNAc) sulfated at position 4 and D-glucuronic acid (GlcUA), resulting in -GalNAc(4S)- $\beta$ (1 $\rightarrow$ 4)-GlcUA- $\beta$ (1 $\rightarrow$ 3)-. In all simulations, A $\beta$  was modeled as the alloform having 42 amino acid residues (known as A $\beta$ <sub>1–42</sub>). To compare the conformational ensembles of A $\beta$  under the different conditions, we created TNs based on 4 or 6  $\mu$ s MD data per system. In order to obtain TNs that are comparable with each other, we used the same molecular features (or descriptors) to define the states of the underlying transition matrix. To capture the conformational changes and identify possible disorder-to-order transitions, we used the number of residues forming  $\alpha$ -helical or  $\beta$ -sheet structures as well as the N-to-C distance of the peptide. The resulting TNs confirmed that A $\beta$  in solution is an IDP that undergoes an unstructured-to-structured transition upon interaction with either the POPC cluster or the GAG. However, the causes of the emergence of structural order in A $\beta$  are completely different, as shown here.

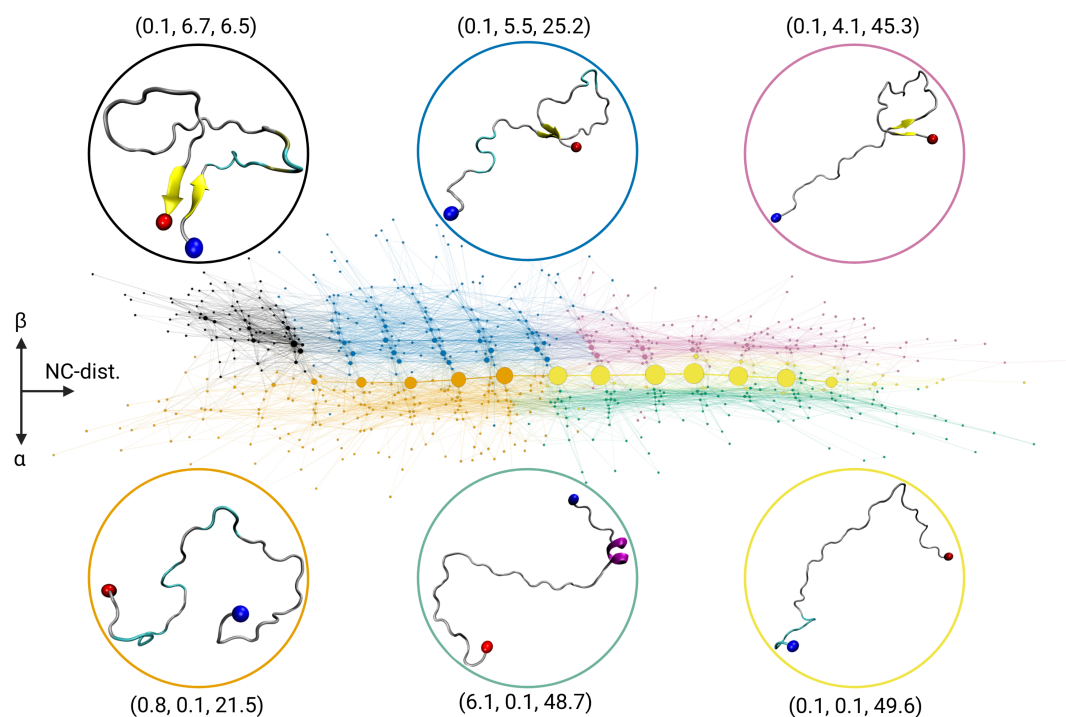
## 2. Results and Discussion

### 2.1. Transition Network of A $\beta$ in Solution

The A $\beta$  monomer in solution classifies as an IDP [21], thus its kinetics can be described by a flat free-energy surface that consists of many local minima that are separated by low-energy barriers [13,15]. In other words, the peptide does not exhibit a unique three-dimensional folded structure that would be connected to a surface-dominating minimum of low free energy, but has a low propensity towards forming  $\alpha$ -helices or  $\beta$ -sheets. This behavior is very well reflected by the TN in Figure 1. The most populated states belong to the community of states colored in orange and yellow, which are aligned along the horizontal axis (called *x*-axis in the following). These two communities represent states that



are primarily disordered with average descriptor values of (0.8, 0.1, 21.5) and (0.1, 0.1, 49.6), respectively, which means that there are neither  $\alpha$ -helices nor  $\beta$ -sheets present in these  $A\beta$  conformations (first two descriptors) and the N-to-C distance varies, on average, between 21 and 50 Å. Thus, the most populated states only vary in their expansion, which increases from left to right in the TN, while they adopt random coil structures, reflecting the intrinsically disordered nature of the  $A\beta$  monomer in solution.



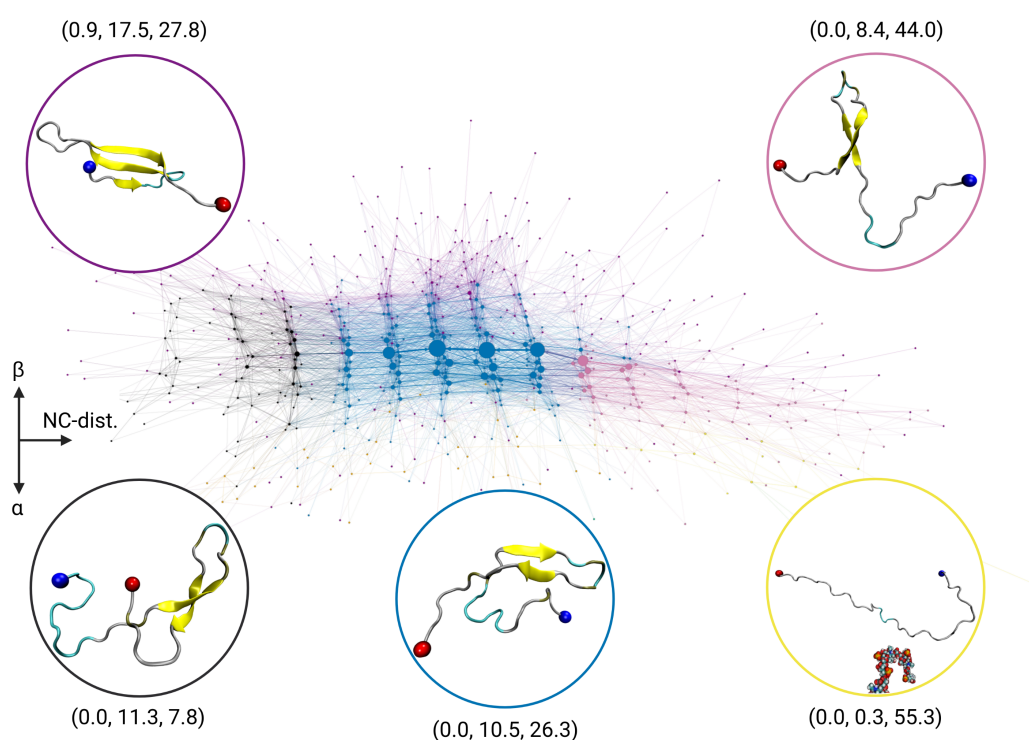
**Figure 1.** Transition network of the  $A\beta$  monomer in solution. For state assignment, three descriptors are used: (i) the number of residues forming  $\alpha$ -helical structure ( $N_\alpha$ ), (ii) the number of residues forming  $\beta$ -sheet structure ( $N_\beta$ ), (iii) the N-to-C distance ( $d_{NC}$ ). The layout of the TN is such that  $d_{NC}$  increases from left to right along the  $x$ -axis,  $N_\beta$  increases in positive  $y$ -direction, and  $N_\alpha$  increases with negative  $y$ -direction. The nodes are colored according to their community membership, and the average descriptor values ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}$ ) of the communities are given. For the highest-populated node per community, a representative structures is shown as cartoon ( $\beta$ -sheets in yellow,  $\alpha$ -helices in purple) with the N- and C-termini being indicated by blue and red spheres, respectively.

Communities shown in black with descriptor values (0.1, 6.7, 6.5), blue with (0.1, 5.5, 25.2), and pink with (0.1, 4.1, 45.3) represent the states with increasing amounts of  $\beta$ -sheets (second descriptor value) as they move toward the top of the network. In contrast, the states exhibiting  $\alpha$ -helical structures are located towards the bottom of the network, and are contained within the community shown in green with descriptor values (6.1, 0.1, 48.7). However, the states containing  $\beta$ -sheets or  $\alpha$ -helices are only sparsely populated compared to the completely random-coil communities shown in orange and yellow. This is even better seen from the distribution of the descriptor values, which are shown in Figure S1 along with the averages and variances in Table S1 in the Supplementary Materials. The population of the different communities is provided in Table S2.

In summary, the arrangement of the states of the TN allows to impose a coordinate system onto its layout, where the N-to-C distance is on the  $x$ -axis with increasing values from left to right, and the secondary structure is resolved along the  $y$ -axis, with disordered states at  $y \approx 0$ , increasing amounts of  $\beta$ -sheet structures for  $y > 0$ , and increasing  $\alpha$ -helical structures for  $y < 0$ . It has to be stressed that this layout mainly arose from the *Force Atlas 2* optimization and was only slightly adjusted for visualization.

## 2.2. Transition Network of A $\beta$ in the Presence of a GAG Molecule

In order to study the conformational change of A $\beta$  in the presence of a sulfated GAG compared to A $\beta$  alone, we calculated the TN using the same three descriptors as before. The resulting TN is shown in Figure 2. Also the coloring of the communities has been chosen as for the A $\beta$  monomer in solution to highlight the shift in community population. Thus, nodes that were displayed in a certain color in Figure 2 are displayed in that color again, with the descriptor values serving as the basis for the color assignment. However, depending on the existence and population of the various states, the relevant communities may appear in somewhat different parts of the TN than in the TN of the A $\beta$ -only system. Nonetheless, the coordinate system is still the same, i.e., increasing N-to-C distances correspond to an increase along the horizontal axis, while states with increasing amounts of  $\beta$ -sheets are found in the positive  $y$ -direction, and in the opposite direction, there are very few sparsely populated states exhibiting  $\alpha$ -helical structures.

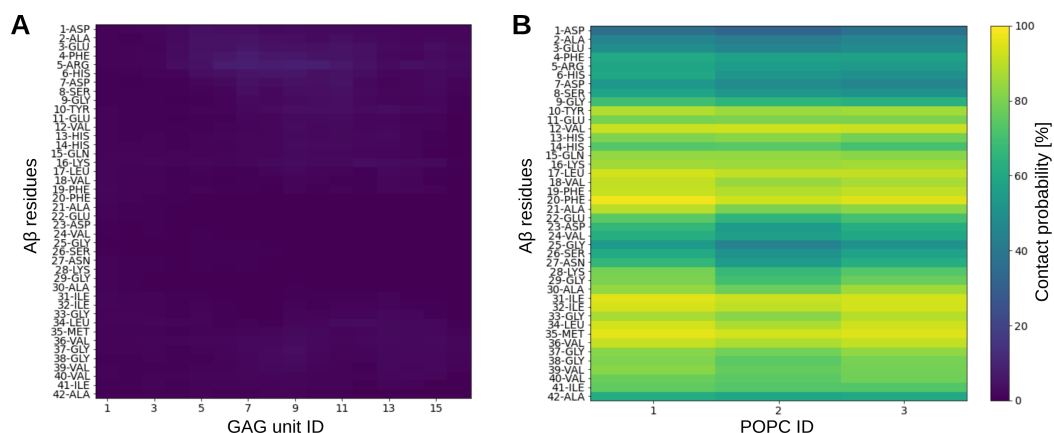


**Figure 2.** Transition network of the A $\beta$  in the presence of a GAG molecule. The nodes are colored according to their community membership, and the average descriptor values ( $N_\alpha$ ,  $N_\beta$ ,  $d_{NC}$ ) of the communities and a representative snapshot are provided. The color of the communities was chosen as in Figure 1, so that states with the same or similar descriptor values are represented with the same color as in the A $\beta$ -only system. See the caption of Figure 1 for further explanations of the graphical representation of the TN.

The TN for the A $\beta$ -GAG system looks quite different from that for the A $\beta$ -only system. First, there is a drastic shift in the state population away from the disordered states, which are now presented by only one community, shown in yellow and with average descriptor values of (0.0, 0.3, 55.3). These disordered states are hardly populated with A $\beta$ -GAG. Instead, the most populated states belong to the community shown in blue, which has average descriptor values of (0.0, 10.5, 26.3) and thus contains structures of intermediate compactness and a substantial amount of  $\beta$ -sheets. The community to the left shown in black with average descriptor values of (0.0, 11.3, 7.8) represents states with comparable  $\beta$ -sheet propensity, yet higher compactness. However, these states are only sparsely populated (Table S2). To the right of the blue-colored community, there is the community shown in pink with average descriptor values of (0.0, 8.4, 44.0), which

also harbors  $\beta$ -sheet rich states, yet with large N-to-C distances. The population of this community is between those of the highest populated community shown in blue and the black-colored community. All three communities involve A $\beta$  structures where about one fourth of the residues is part of a  $\beta$ -hairpin structure, as the representative structures in Figure 2 show, which mainly differ in the orientation of the N- and C-terminals with respect to each other. At the very top of the TN, mainly above the blue community, there are states combined into the community shown in purple, which are not present in the A $\beta$ -only system. With average descriptor values of (0.9, 17.5, 27.6), these are states with even higher amounts of  $\beta$ -sheet structures. Here, about 42% of the A $\beta$  residues are involved in  $\beta$ -sheet formation, which is accomplished by a  $\beta$ -sheet with three strands, where the third strand aligns with the previously mentioned  $\beta$ -hairpin.

In summary, compared to the TN of the A $\beta$ -only system, there is a significant population shift away from disordered states, towards states with considerable amounts of  $\beta$ -sheets (Figure S1 and Table S1), which results from the formation of a stable  $\beta$ -hairpin. This drastic change in secondary structure is astonishing, as A $\beta$  was rarely in direct contact with the GAG, as the time-averaged distance matrix in Figure 3A shows. This molecular avoidance is understandable, given that both molecules are negatively charged (3<sup>−</sup> and 16<sup>−</sup>, respectively). Nevertheless, the contact map of A $\beta$ -GAG interactions with populations shown only up to the maximum value of  $\approx 8.5\%$  (Figure S2) identify the positively charged A $\beta$  residues Arg5 (in particular) and Lys16 as preferred binding sites for GAG. All other interactions that appear in the contact matrix simply result from their proximity to Arg5 and Lys16. The preference of GAG binding to arginine compared to lysine has been reported previously for other proteins [22]. However, these sparsely populated interactions are not sufficient to explain the GAG-induced structural changes in A $\beta$ , and further reasons are given below.

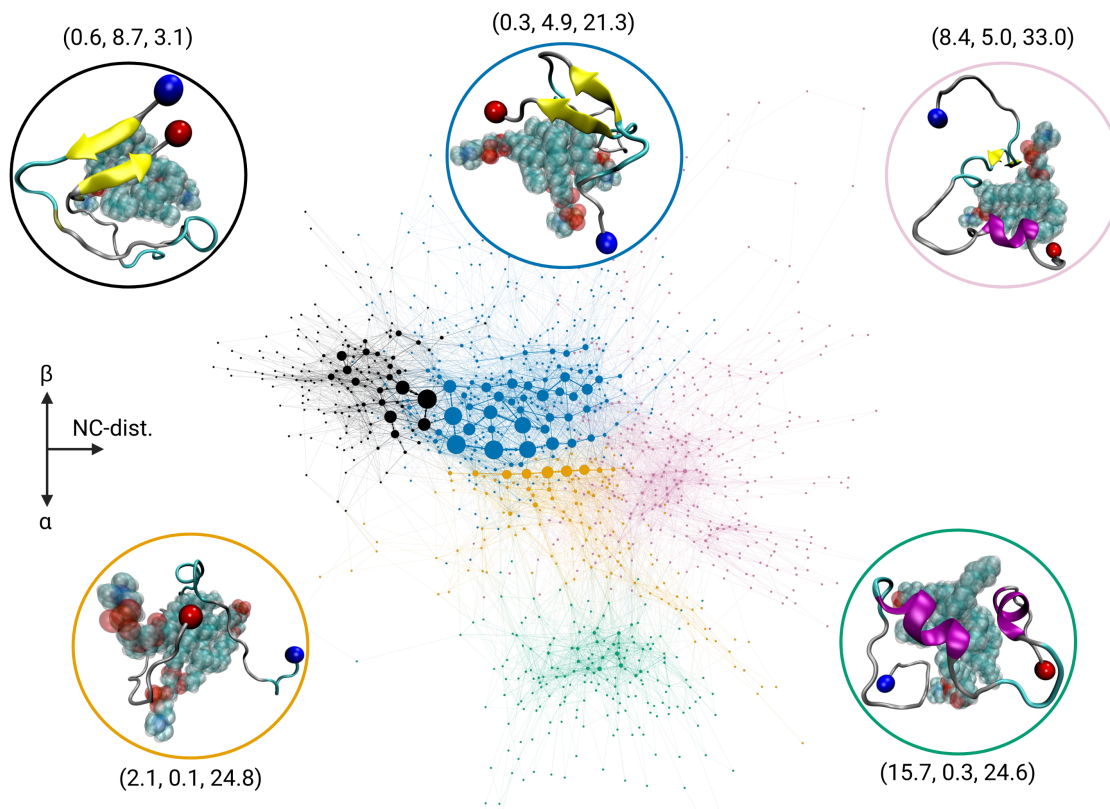


**Figure 3.** Intermolecular contact maps for A $\beta$  interacting with (A) a GAG molecule and (B) three POPC lipids. The interactions are separated into residue–monosaccharide interactions for the A $\beta$ -GAG system and residue–lipid interactions for the A $\beta$ -POPC system. Two interaction partners were considered to be in contact if in a given frame of the trajectory they are closer than 10 Å. The resulting number of contacts were normalized with respect to the total number of time frames per trajectory, yielding a contact probability (see color scale on the right).

### 2.3. Transition Network of A $\beta$ Interacting with a POPC Cluster

To put the TN of the A $\beta$ -GAG system into perspective, we also generated the TN of A $\beta$  in interaction with three POPC lipids. In our previous study, we showed that under these conditions the A $\beta$  monomer undergoes a disorder-to-order transition, which is facilitated by interactions between hydrophobic residues of A $\beta$  and the lipid tails (Figure 3B) [5]. Here we use the simulations of that study to calculate the TN using the same three descriptors as before, resulting in the TN shown in Figure 4. The structures of the A $\beta$  peptide in contact with POPC differ significantly from those of the A $\beta$ -only system, resulting in many

new states. For this reason, direct mapping of the color code of the communities was not possible. Nevertheless, the coloration was chosen to be as close as possible to the previous representations. Again, the communities were distributed along the three axes  $d_{NC}$  (positive  $x$ -axis),  $N_{\beta}$  (positive  $y$ -axis), and  $N_{\alpha}$  (negative  $y$ -axis).



**Figure 4.** Transition network of A $\beta$  in complex with three POPC lipids. The color of the communities was chosen as in Figure 1, so that states with the same or similar descriptor values are represented with the same color as in the A $\beta$ -only system. However, the communities shown here in pink and green are not found in either the A $\beta$ -only or A $\beta$ -GAG systems. See the caption of Figure 1 for further explanations of the graphical representation of the TN.

As for the A $\beta$ -only system, there is a community colored orange that harbors disordered states with average descriptor values of (2.1, 0.1, 24.8). The underlying states have slight deviations in their N-to-C distance, with increasing values from left to right. However, they are less expanded than the states of the yellow-colored community found for the A $\beta$ -only and A $\beta$ -GAG systems, which is not present here. Moreover, some of the A $\beta$ -POPC states belonging to the orange-colored community feature some  $\alpha$ -helical structure, which is also different from the corresponding A $\beta$ -only community. These states are located towards the bottom of the community, since, as before, the  $y$ -axis distinguishes between  $\alpha$ -helical and  $\beta$ -sheet structures. Another difference is that the disordered states are only sparsely populated in the A $\beta$ -POPC system. Here, the two most populated communities are the ones colored black and blue, which have average descriptor values of (0.6, 8.7, 3.1) and (0.3, 4.9, 21.3), respectively, and thus mainly contain compact states with considerable  $\beta$ -sheet content. On the opposite side along the  $y$ -axis, at the bottom of the TN, there is a distinct cluster of states shown in green with average descriptor values of (15.7, 0.3, 24.6). This community exhibits high amounts of  $\alpha$ -helical structures. These states are infrequently visited, but their spatial separation, determined using the *Force Atlas 2* algorithm based on their low connectivity with other communities, suggests that this community corresponds to a local minimum in the free-energy landscape, separated from the other communities by a high-energy barrier. The rightmost community along the  $x$ -axis, colored pink, exhibits the states with the largest N-to-C distance of the peptide in the A $\beta$ -POPC system, as indicated



by the average descriptor values of (8.4, 5.0, 33.0). The first two descriptors indicate that the states of this community feature both an  $\alpha$ -helix and a  $\beta$ -sheet. This structural finding, as also the location of this community in the TN, suggests that these states are intermediates between the  $\beta$ -sheet-rich states of the blue community and the  $\alpha$ -helical-rich states of the green community. The fact that the pink community is also next to the orange community results from the projection of the TN onto a two-dimensional plane. In a third dimension, the orange community would appear closer to the viewer.

Comparison of this TN to the TN of the A $\beta$ -only system highlights the conformational change of A $\beta$  towards folded states when in complex with lipids, as also revealed by the descriptor distributions (Figure S1 and Table S1) and community populations (Table S2). One can see a notable shift in the state population towards more compact states; the current TN has its maximum extension in the  $y$ -direction, while for the A $\beta$ -only system the maximum TN extension is in the  $x$ -direction. Such a change in TN geometry did not occur in the A $\beta$ -GAG system where A $\beta$  remained very expanded and only adopted  $\beta$ -sheet structures but not  $\alpha$ -helices. The change of TN geometry for the A $\beta$ -POPC system also reveals a change in the underlying free-energy surface, from being rather flat with many local minima (A $\beta$ -only and A $\beta$ -GAG) towards multiple definite energy basins of (semi-)folded A $\beta$  conformations.

#### 2.4. Interactions in the A $\beta$ -GAG System

In order to understand the conformation switching of A $\beta$  in presence of the GAG, we investigated their molecular interactions as well as the impact of the GAG on the water dynamics and Na<sup>+</sup> distribution. As already mentioned, the GAG induces the structural change in A $\beta$  with hardly any direct interactions between the peptide and the GAG. Thus, the mode of interaction differs notably to that of the A $\beta$ -POPC interactions, as the comparison between the two contact maps in Figure 3 shows. In the case of the A $\beta$ -GAG system, only some of the positively charged side chains of A $\beta$ , in particular Arg5 and the neighbored residue His6, are in direct contact with the GAG for about 9% of the time. In contrast, the contact map of the A $\beta$ -POPC system shows that some of the hydrophobic residues interact with POPC during the whole course of the simulation, which causes the conformation switching in A $\beta$  [5].

Since A $\beta$  and the GAG do not interact directly with each other, the effect of the GAG on the peptide must therefore be indirect. We considered two possibilities for this, which both could arise due to the high negative charge of the GAG: (i) a change in the water ordering and dynamics, and (ii) a change in the Na<sup>+</sup> ion distribution. To address (i), we determined the water structure in close proximity to A $\beta$  using the translational and orientational order parameters  $T$  and  $Q$  (see Equations (2) and (3)). Table 1 shows the ensemble and time average for both quantities considering all water molecules in the vicinity (i.e., within 10 Å) of the A $\beta$  peptide. As one can see, the order parameters for the water surrounding A $\beta$  show no significant difference between the A $\beta$ -only and A $\beta$ -GAG simulations. This also applies to the water around the GAG molecule. Furthermore, the results did also not change considerably when reducing the radius of the water molecules to be considered in the calculation to 5 Å within the solutes. To probe the dynamics of the water molecules, we used the lifetimes of the H-bonds formed between water and either A $\beta$  or the GAG (see Equations (4) and (5)). The results, also listed in Table 1, show that in either system and for either molecule, the values for the stretching exponent  $\beta$  are about 0.5, reflecting that the relaxation of the H-bonds deviates from exponential behavior, which is caused by interactions between water and A $\beta$  or the GAG. The deviation is strongest for the GAG and smallest for A $\beta$  in the A $\beta$ -GAG system, which suggests that the presence of the GAG molecule weakens the A $\beta$ -water interactions. However, the differences in the water dynamics between the molecules are only minor, as confirmed by the lifetimes of the H-bonds. Interestingly, the lifetime of the H-bonds formed between the GAG and water is an intermediate of that of A $\beta$  in different environments, i.e., the high negative charge of the GAG does not slow down the water dynamics. Overall, we did not find noteworthy effects

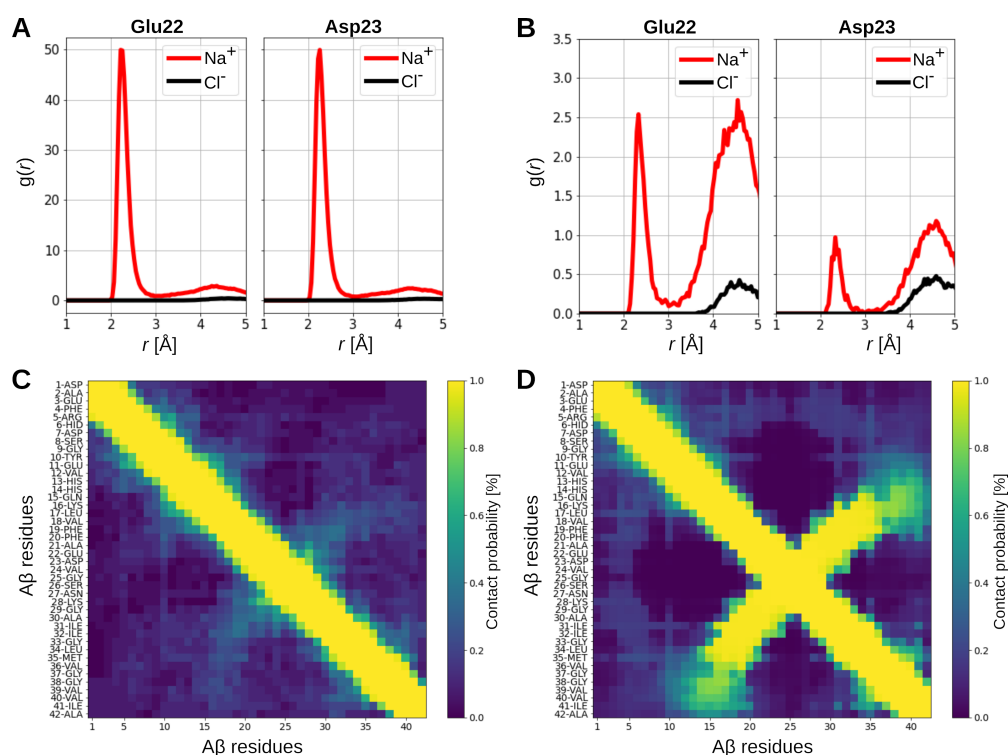
of the GAG on the water structure and dynamics around A $\beta$  that would explain its drastic change in conformation.

**Table 1.** Translational and rotational order parameters  $T$  and  $Q$  of water in the vicinity ( $<10$  Å) of A $\beta$  and the GAG molecule, as well as the average stretching exponents  $\beta$  and mean lifetimes  $\langle\tau\rangle$  of H-bonds formed between water and A $\beta$  or the GAG.

Molecule	$T$	$Q$	$\beta$	$\langle\tau\rangle$ /[ps]
A $\beta$ -only	0.37	0.44	0.555	16.45
A $\beta$ in A $\beta$ -GAG	0.34	0.44	0.565	14.54
GAG in A $\beta$ -GAG	0.34	0.38	0.499	15.39

To assess whether the negative charge has considerable effects on the distribution of the ions in the system, we calculated the radial distributions,  $g(r)$ , of Na $^+$  and Cl $^-$  with respect to the negatively and positively charged A $\beta$  residues in the A $\beta$ -only and A $\beta$ -GAG system. For the latter, we also determined the ion distribution around the GAG molecule. While the Cl $^-$  distributions are not noteworthy affected by the GAG, the Na $^+$  distribution changed dramatically. Panels A and B of Figure 5 show the results for the carboxyl groups of residues Glu22 and Asp23 for A $\beta$  in the two systems, while all other distributions are shown in Figures S3–S6 in the Supplementary Materials. Comparison of these distributions reveals that in the A $\beta$ -GAG system, the interaction between these two residues and the surrounding sodium ions is an order of magnitude smaller compared to the simulation of A $\beta$  alone. This is a result of the strong attraction between the negatively charged COO $^-$  and OSO $_3^-$  groups of the GAG and Na $^+$  (Figure S7, discussed in detail in our previous work on GAGs [23]), making the peptide a less favorable interaction partner. The withdrawal of Na $^+$  from A $\beta$  descreens the electrostatic interactions between the peptide residues, which enables a strong attraction between Glu22/Asp23 and Lys28. This is visible from intrapeptide residue–residue contact maps shown in Figure 5C,D for the A $\beta$ -only and A $\beta$ -GAG system, respectively. The salt bridge between Glu22/Asp23 and Lys28 in the A $\beta$ -GAG system results in a  $\beta$ -hairpin, which was already mentioned when discussing the TN of that system, which shows as strong contacts perpendicular to the diagonal that reflect the neighboring contacts along the sequence. In the A $\beta$ -only system, such a perpendicular trace of contacts is only slightly visible, resulting from short-lived interactions. Otherwise the peptide is devoid of interactions beyond  $i, i + 3$  along the sequence ( $i$  referring to the residue number), which is in line with the observation from the TN of the A $\beta$ -only system.

In summary, the effect of the GAG on A $\beta$  results from notably changing the distribution of the sodium ions in the vicinity of the peptide, as these ions are strongly attracted by the GAG. As a consequence, the formation of intrapeptide salt bridges is facilitated, in particular between residues Glu22/Asp23 and Lys28, which is further stabilized by subsequent  $\beta$ -hairpin formation. There are no notable direct interactions between the GAG and A $\beta$  nor relevant effects of the GAG on the water order and dynamics, which could serve as alternative explanations for the conformational changes in A $\beta$  in the presence of the GAG.



**Figure 5.** The radial distributions  $g(r)$  of  $\text{Na}^+$  (red) and  $\text{Cl}^-$  (black) relative to the carboxyl groups of the side chains of Glu22 and Asp23 of  $\text{A}\beta$  in (A) the  $\text{A}\beta$ -only system and (B) the  $\text{A}\beta$ -GAG system. The intra-peptide contacts between the  $\text{A}\beta$  residues in (C) the  $\text{A}\beta$ -only system and (D) the  $\text{A}\beta$ -GAG system. Two interaction partners were considered to be in contact if in a given frame of the trajectory they are closer than 10 Å. The resulting number of contacts were normalized with respect to the total number of time frames per trajectory, yielding a contact probability.

### 2.5. Discussion

First, we discuss the TN of the  $\text{A}\beta$ -only system, which is a prime example of what one would expect the TN of an IDP to look like. This TN is characterized by the presence of many states, most of which have small populations and many connections to other states, corresponding to a broad but flat free-energy landscape. Furthermore, the vast majority of states belongs to  $\text{A}\beta$  conformations that are purely random coils and vary only in their spatial expansion. Upon running the *Force Atlas 2* optimization algorithm of Gephi, the states differing mainly in their N-to-C distance aligned along one axis, which we assigned as the  $x$ -axis. The second axis in our TN representation, the  $y$ -axis, turned out to represent the change in secondary structure, with positive  $y$ -values being associated with more  $\beta$ -sheet structure and negative  $y$ -values with more  $\alpha$ -helical structures. Translating the TN to a free-energy surface, one can conclude that the peptide populates a shallow energy basin corresponding to random-coil structures and large N-to-C distance variations. Free-energy states with a defined secondary structure are rarely visited and quickly return to the disordered states.

The TN of  $\text{A}\beta$  together with a GAG differs significantly from that of  $\text{A}\beta$  alone. The TN was calculated using the same three descriptors as before, and the automatically assigned layout of the nodes along the  $x$ - and  $y$ -axes did not change either. The TN clearly revealed that  $\text{A}\beta$  undergoes a structural change in the presence of the GAG. While  $\text{A}\beta$  can still adopt fully extended structures in the presence of the GAG, the most populated states are found for more compact and more  $\beta$ -sheet rich structures. These structures contain a very stable  $\beta$ -hairpin maintained by the presence of a salt bridge between Glu22/Arg23 and Lys28, which is made possible by a descreening of the electrostatic interaction between these residues due to a shift in the local sodium ion concentration away from the peptide. These ions are instead attracted to the strongly negatively charged GAG. This distinct

structural switching of A $\beta$  in the presence of a GAG agrees with the experimental finding that in the presence of GAGs, the random-coil to amyloidogenic  $\beta$ -sheet transition of A $\beta$  is accelerated, leading to a more rapid fibril formation [6]. This observation is of biological relevance because GAGs are important components of the extracellular space, where they can be located on the cell surface or within the extracellular matrix. There, they exist in two forms: covalently attached to the protein core of proteoglycans or as independent macromolecules. Studies have revealed a close connection between GAGs and amyloid fibrils extracted from humans. Evidence indicates that GAGs actively participate in the formation and stabilization of amyloid fibrils [9,24,25]. Here, we show for the first time how GAGs can cause a structural change in A $\beta$  from a random coil to a  $\beta$ -sheet structure. This structural change is even more remarkable considering that A $\beta$  and the GAG show almost no direct interaction with each other and only GAG-induced changes in the local peptide environment trigger a conformational change in A $\beta$ .

Some of these observations should be investigated in further studies. We chose the protonation state of A $\beta$  to correspond to the physiological conditions of the extracellular space, i.e., about pH 7.4. At this pH, the three histidine residues of A $\beta$  can be assumed to be neutral (the pK<sub>a</sub> value of the free His is 6.0). On the other hand, the local environment of His6, His13, and His14 can change their pK<sub>a</sub> value. Moreover, it is known that aging and Alzheimer's disease can decrease the pH of the extracellular brain space to below 7 [26]. Therefore, it would be justified to simulate other protonation states of A $\beta$  as we have already carried out for A $\beta$  in solution [27,28], which resulted in random coil to  $\beta$ -sheet formation at the isoelectric point of A $\beta$  (pI of 5.3) due to altered intrapeptide electrostatics. We expect that modeling the histidine residues as His<sup>+</sup>, which would yield neutral A $\beta$ , would increase the interaction between the GAG and A $\beta$ . This in turn would increase the local Na<sup>+</sup> concentration around A $\beta$  since it would now be in close proximity to the GAG molecule, while neutral A $\beta$  would have a higher preference to form  $\beta$ -sheets. It will therefore be interesting to observe how these different forces affect the behavior of the peptide. Moreover, the observation that the GAG-induced decrease in sodium ion concentration near A $\beta$  leads to structural changes in the peptide should also be further investigated by a titration simulation in which the concentration (and also the type of salt) is gradually increased, starting at zero, to determine the dependence of the A $\beta$  structure on the salt. It is known from experiments that both the type and concentration of salts, particularly the type of cation, have significant effects on the rate of aggregation and the morphology of the resulting A $\beta$  fibrils [29,30], but the effects of the salts on the peptide monomer are still unknown.

The TN of A $\beta$  interacting with three POPC lipids also revealed a disordered-to-ordered transition in A $\beta$ . However, certain differences exist compared to the A $\beta$ -GAG system. First, A $\beta$  forms a complex with the lipids over the whole time of the simulation. This leads to more compact peptide structures; fully extended structures did not occur anymore. Second, A $\beta$  can also adopt  $\alpha$ -helical structures and not only  $\beta$ -sheet-rich states. Either folded state is facilitated by hydrophobic interactions between the peptide and the lipids. The TN revealed that the  $\alpha$ -helical structures populate a community that is somewhat separated from the other communities, which allowed us to conclude that the  $\alpha$ -helical states correspond to a separate local minimum in a multi-funnel free-energy landscape. Thus the interaction with the POPC lipids not only shifts the main energy basin towards more compact configurations with structure formation, but also leads to a rougher free-energy landscape. The results of these simulations are consistent with a large body of experimental studies that attest to a central role for lipids in amyloid aggregation and disease development. Lipids are an integral component of many amyloid deposits in vivo, where their presence can influence fibril nucleation, morphology, and mechanical properties [31]. It was demonstrated that the toxicity of A $\beta$  aggregates correlates with the amount of their  $\beta$ -sheet content, which, in turn, is increased by lipids present during A $\beta$  aggregation [32]. With respect to helical A $\beta$  structures, this appears to be the prerequisite for membrane incorporation of A $\beta$  [33–37].

### 3. Materials and Methods

#### 3.1. Molecular Dynamics Simulations

In all simulations, A $\beta$  was simulated as A $\beta_{1-42}$  with the histidine residues modeled as neutral and no terminal capping groups, leading to an overall peptide charge of 3 $^-$ . The three systems of only A $\beta$ , A $\beta$  plus GAG, and A $\beta$  with three POPC lipids (henceforth called A $\beta$ -only, A $\beta$ -GAG, and A $\beta$ -POPC, respectively) were simulated using the GRO-MACS simulation package [38]. Since two of the three systems were originally part of different studies [5,15,19], some of the MD simulation settings (but not the force field parameters and ion concentration) differ slightly between the simulations. While this is not ideal, these differences appear to be negligible given the remarkable results presented below. In each simulation, the A $\beta$  peptide was modeled using the CHARMM36m force field [39]. It has been found in previous studies that the CHARMM36m force field is best suited for modeling both monomeric A $\beta$  [15] as well as amyloid aggregation [40]. CHARMM36m is a polypeptide force field that can be combined with the original force field CHARMM36 [41] for modeling the POPC lipids. For the GAGs, we used the parameters as available through the Glycan Reader & Modeler module [42–44] of the CHARMM-GUI web server [45], as in our previous studies of GAGs alone and in interaction with multiple A $\beta_{16-22}$  peptides [23,46]. The preparation of the systems followed the same standard protocol: First the solute(s) were placed in the simulation box, which was then filled with TIP3P water molecules [47] together with Na $^+$  and Cl $^-$  ions to neutralize the systems and achieve a physiological salt concentration of 150 mM. After equilibration of the systems, each of them was simulated under  $NpT$  conditions at 1 bar, which was accomplished using a Parrinello–Rahman pressure coupling scheme [48]. The simulations of A $\beta$ -GAG and A $\beta$ -POPC were carried out at 310 K using a Nosé–Hoover thermostat [49,50], while the A $\beta$  system was simulated at 300 K using a velocity rescaling thermostat [51]. In the case of the A $\beta$ -GAG and A $\beta$ -POPC, the simulations were carried out for 4  $\mu$ s involving 1  $\times$  4  $\mu$ s and 2  $\times$  2  $\mu$ s, respectively, while the A $\beta$  system was simulated for 1  $\times$  6  $\mu$ s. All simulations were achieved under periodic boundary conditions in all directions and the particle-mesh Ewald method [52] was used for calculating the electrostatic interactions. The cutoffs for van der Waals and Coulomb interaction calculations in real space were both set to 12 Å. The minimum distance between any solute atom and any face or edge of the simulation box was set to 12 Å. All MD simulations were run on the supercomputer JURECA [53].

#### 3.2. Transition Networks

In order to construct a TN, one chooses a set of  $n$  features that describe the process under study. These features are evaluated by descriptor functions  $\{f_i\}$  that act on a given conformation  $x(t)$  and project the  $3N$ -dimensional phase space onto an  $n$ -dimensional state  $S(t)$

$$x(t) \mapsto S(t) = [f_1(x(t)), f_2(x(t)), \dots, f_n(x(t))]. \quad (1)$$

Here,  $x(t)$  is the conformation of an MD simulation at time  $t$  and  $N$  is the number of particles within the conformation. A crucial point in constructing a TN is the choice of descriptor functions. As discussed in our previous work [19,20], the number and type of descriptors have a huge impact on the resulting TN. While the choice of the type of descriptor functions is closely related to the process under study, the ideal number of descriptors is often up to trial and error. Choosing fewer descriptors yields a simpler TN, though more information is lost due to the projection onto a low-dimensional space. On the contrary, choosing more descriptors might yield a TN that is too complex to be intuitively interpreted. Here we decided to use the same three descriptor functions for the studied systems, which allows easy comparison between them. To describe the process of conformation switching of A $\beta$ , we chose (i) the number of residues adopting an  $\alpha$ -helical structure ( $N_\alpha$ ), (ii) the number of residues adopting a  $\beta$ -sheet structure ( $N_\beta$ ), and (iii) the distance between N- and C-terminus, called the N-to-C distance ( $d_{NC}$ , in Å) as a measure of compactness of the peptide. The TNs were calculated with ATRANET, which is a Python



package developed by our group. The software has been optimized to handle large amounts of MD data with many different descriptor functions to choose from, while still providing a dynamic framework to easily add custom descriptor functions [19,20]. ATRANET is available at <https://github.com/strodel-group/ATRANET> (accessed on 1 May 2023).

The transition matrix created by ATRANET can be visualized with the network analysis and visualization software Gephi 0.10 [54,55]. For the layout of the networks we chose the *Force Atlas 2* algorithm, which is a force-driven algorithm that takes into account the connectivity of pairs of nodes and their relative degree. As a result, nodes that have more transitions between them are displayed closer to each other. Thus, a large spacial separation in the layout of the TN corresponds to a large distance between the respective states in the high-dimensional phase space. Additionally, we chose the size of the nodes to be proportional to their diagonal entries in the transition matrix as a state with more self-transitions is more stable (i.e., has a lower free energy), while nodes with fewer self-transitions reflect higher-energy states. For visualization purposes, the depicted sizes are adjusted based on the minimum and maximum values for each network on a linear scale from 1 to 10. In the following, the size of the nodes will also be referred to as the population of the states. In addition, we used Gephi's modularity class feature to divide the network into local communities, which makes it easier to identify groups of nodes that are strongly connected and have high similarity between states. In terms of a free-energy perspective, the states of a community belong to the same energy basin of an underlying multi-funnel free-energy landscape [56].

### 3.3. Analysis of Water around the Solutes

To elucidate possible effects of the water solvating the peptide or the GAG, we analyzed the water structure and dynamics around the solutes. For determining the water structure, we used the translational order parameter  $T$  and orientational order parameter  $Q$  [57]. The translational order parameter is given by

$$T = \frac{1}{\zeta_c} \int_0^{\zeta_c} |g(\zeta) - 1| d\zeta \quad (2)$$

where  $g$  refers to the oxygen–oxygen radial distribution function (RDF) and  $\zeta = r \cdot \rho^{1/3}$  is a dimensional variable dependent on the radial distance  $r$  and the density of the water-oxygen atoms ( $\rho = N_O/V$ ). The parameter  $\zeta_c = 2.8$  is chosen such that  $g(\zeta_c) \sim 0$ . The order parameter  $T$  can be used to measure whether or not long-range interactions are present in a medium. For an ideal gas, the RDF is equal to 1 and hence  $T = 0$ . In the case of a crystal, the RDF is different from 1 even for large distances, so  $T$  is large in a system with long-range order. The orientational parameter  $Q$  measures the ability of neighboring water molecules to produce tetrahedral arrangements. It is given by

$$Q = 1 - \frac{3}{8} \sum_{j=1}^3 \sum_{k=j+1}^4 \left( \cos \psi_{jk} + \frac{1}{3} \right)^2 \quad (3)$$

where  $\psi_{jk}$  is the angle between neighboring O atoms  $j$  and  $k$  with central atom  $i$ . The value of  $Q$  can range from 0 to 1, where 0 corresponds to a random distribution of water molecules and 1 to a perfect tetrahedral arrangement.

To probe the dynamics of the water in the first solvation shell, we made use of the lifetimes of the hydrogen bonds (H-bonds) between water and the solutes (A $\beta$  or GAG). To this end, we performed short simulations of 100 ns for the A $\beta$  and A $\beta$ -GAG system, using the same parameters as described above but writing out data every 0.5 ps to resolve the H-bond lifetimes. These simulations were analyzed in terms of the H-bond existing function  $h(t_0 + t)$ , which is either 1 or 0 at a given time, depending on whether or not a specific H-bond is present. To improve statistics, multiple time origins  $t_0$  are used in the calculation and the average is taken over all time origins and possible H-bonds. The mean

H-bond lifetime  $\langle \tau \rangle$  can then be determined by calculating the autocorrelation function of the averaged  $h(t_0 + t)$  and fitting a stretched exponential function to it:

$$c(t) = \exp \left\{ \left( -\frac{t}{\tau} \right)^\beta \right\} \quad (4)$$

where  $c(t)$  refers to the autocorrelation function,  $\tau$  is the lifetime, and  $\beta$  is the stretching factor. From this, the mean lifetime  $\langle \tau \rangle$  can be calculated via integration, which is analytically solved by the gamma function  $\Gamma$ :

$$\langle \tau \rangle = \frac{\tau}{\beta} \Gamma \left( \frac{1}{\beta} \right) \quad (5)$$

#### 4. Conclusions

We constructed the transition networks (TNs) revealing the conformational preferences and conversions of the amyloid  $\beta$ -peptide  $A\beta_{1-42}$  (here simply called  $A\beta$ ) under different conditions: as a single peptide in solution,  $A\beta$  in the presence of the GAG chondroitin-4-sulfate with sixteen subunits, and  $A\beta$  in complex with three POPC lipids. For defining the states of each TN, we chose the same three descriptors: (i) the number of residues with an  $\alpha$ -helical structure, (ii) the number of residues with a  $\beta$ -sheet structure, (iii) the peptide distance from end-to-end. Using the same descriptors allows direct comparison of the resulting TNs and identification of changes in the underlying free-energy surfaces between the different systems. Moreover, the choice of a low-dimensional projection of the phase space due to using only three descriptors allows a very intuitive interpretation of the resulting TNs and directly visualizes the conformation switching of  $A\beta$ . In particular, we have shown how the interaction of  $A\beta$  with a GAG or POPC lipids leads to a transition from disorder to order of the intrinsically disordered monomer. Taking advantage of the similarities of the transition network layout, we can infer a shift of the main basin of the underlying free-energy surface from disordered conformations with large end-to-end separations to more compact conformations with high amounts of  $\beta$ -sheet. The overall increase in  $\beta$ -sheet-rich structures could, in turn, serve as a nucleus for amyloid aggregation and the formation of toxic oligomers. Our findings contribute to the growing body of knowledge on the role of GAGs and lipids in  $A\beta$  aggregation and the development of Alzheimer's disease.

**Supplementary Materials:** The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms241411238/s1>.

**Author Contributions:** Conceptualization, B.S.; methodology, M.S. and B.S.; software, M.S.; formal analysis, M.S.; validation, M.S. and B.S.; investigation, M.S.; data curation, M.S. and S.S.; visualization, M.S.; writing—original draft preparation, M.S.; writing—review and editing, M.S., S.S., and B.S.; supervision, B.S.; project administration, B.S.; funding acquisition, B.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** The authors gratefully acknowledge computing time on the supercomputer JURECA at Forschungszentrum Jülich under grant name IDP.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The code for the analysis of the MD simulations is available at <https://github.com/strodel-group/ATRANET> (accessed on 1 May 2023). The trajectory files are available from the authors upon request.

**Acknowledgments:** The authors thank Anna Jäckering for fruitful discussions.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References


- Uversky, V.N.; Kulkarni, P. Intrinsically disordered proteins: Chronology of a discovery. *Biophys. Chem.* **2021**, *279*, 106694. [\[CrossRef\]](#)
- Perdigão, N.; Heinrich, J.; Stolte, C.; Sabir, K.; Buckley, M.; Tabor, B.; Signal, B.; Gloss, B.; Hammang, C.; Rost, B.; et al. Unexpected features of the dark proteome. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 15898–15903. [\[CrossRef\]](#) [\[PubMed\]](#)
- Hardy, J.A.; Higgins, G.A. Alzheimer's Disease: The Amyloid Cascade Hypothesis. *Science* **1992**, *256*, 184–185. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nguyen, P.H.; Ramamoorthy, A.; Sahoo, B.R.; Zheng, J.; Faller, P.; Straub, J.E.; Dominguez, L.; Shea, J.E.; Dokholyan, N.V.; De Simone, A.; 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.* **2021**, *121*, 2545–2647. [\[CrossRef\]](#)
- Fatafta, H.; Kav, B.; Bundschuh, B.F.; Loschwitz, J.; Strodel, B. Disorder-to-order transition of the amyloid- $\beta$  peptide upon lipid binding. *Biophys. Chem.* **2022**, *280*, 106700. [\[CrossRef\]](#)
- Iannuzzi, C.; Irace, G.; Sirangelo, I. The Effect of Glycosaminoglycans (GAGs) on Amyloid Aggregation and Toxicity. *Molecules* **2015**, *20*, 2510–2528. [\[CrossRef\]](#)
- Snow, A.D.; Wight, T.N. Proteoglycans in the pathogenesis of Alzheimer's disease and other amyloidoses. *Neurobiol. Aging* **1989**, *10*, 481–497. [\[CrossRef\]](#)
- Ancsin, J.B. Amyloidogenesis: Historical and modern observations point to heparan sulfate proteoglycans as a major culprit. *Amyloid* **2003**, *10*, 67–79. [\[CrossRef\]](#)
- Van Horsen, J.; Wesseling, P.; Van Den Heuvel, L.P.; De Waal, R.M.; Verbeek, M.M. Heparan sulphate proteoglycans in Alzheimer's disease and amyloid-related disorders. *Lancet Neurol.* **2003**, *2*, 482–492. [\[CrossRef\]](#) [\[PubMed\]](#)
- Eliezer, D. Biophysical characterization of intrinsically disordered proteins. *Curr. Opin. Struct. Biol.* **2009**, *19*, 23–30. [\[CrossRef\]](#)
- Kragelund, B.B.; Skriver, K. *Intrinsically Disordered Proteins: Methods and Protocols*; Springer: New York, NY, USA, 2020.
- Löhr, T.; Kohlhoff, K.; Heller, G.; Camilloni, C.; Vendruscolo, M. A kinetic ensemble of the Alzheimer's A $\beta$  peptide. *Nat. Comput. Sci.* **2021**, *1*, 71–78. [\[CrossRef\]](#)
- Strodel, B. Energy Landscapes of Protein Aggregation and Conformation Switching in Intrinsically Disordered Proteins. *J. Mol. Biol.* **2021**, *433*, 167182. [\[CrossRef\]](#)
- Fatafta, H.; Samantray, S.; Sayyed-Ahmad, A.; Coskuner-Weber, O.; Strodel, B. Chapter Five—Molecular simulations of IDPs: From ensemble generation to IDP interactions leading to disorder-to-order transitions. In *Dancing Protein Clouds: Intrinsically Disordered Proteins in the Norm and Pathology, Part C*; Uversky, V.N., Ed.; Academic Press: Cambridge, MA, USA, 2021. [\[CrossRef\]](#)
- Paul, A.; Samantray, S.; Anteghini, M.; Khaled, M.; Strodel, B. Thermodynamics and kinetics of the amyloid- $\beta$  peptide revealed by Markov state models based on MD data in agreement with experiment. *Chem. Sci.* **2021**, *12*, 6652–6669. [\[CrossRef\]](#)
- Chakraborty, D.; Straub, J.; Thirumalai, D. Energy landscapes of A $\beta$  monomers are sculpted in accordance with Ostwald's rule of stages. *Sci. Adv.* **2023**, *9*, eadd6921. [\[CrossRef\]](#)
- Rao, F.; Caflisch, A. The Protein Folding Network. *J. Mol. Biol.* **2004**, *342*, 299–306. [\[CrossRef\]](#)
- Barz, B.; Wales, D.J.; Strodel, B. A Kinetic Approach to the Sequence–Aggregation Relationship in Disease-Related Protein Assembly. *J. Phys. Chem. B* **2014**, *118*, 1003–1011. [\[CrossRef\]](#) [\[PubMed\]](#)
- Schäffler, M.; Khaled, M.; Strodel, B. ATRANET – Automated generation of transition networks for the structural characterization of intrinsically disordered proteins. *Methods* **2022**, *206*, 18–26. [\[CrossRef\]](#)
- 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.* **2020**, *16*, 7825–7839. [\[CrossRef\]](#)
- Roche, J.; Shen, Y.; Lee, J.H.; Ying, J.; Bax, A. Monomeric A $\beta$ 1–40 and  $\beta$ 1–42 Peptides in Solution Adopt Very Similar Ramachandran Map Distributions That Closely Resemble Random Coil. *Biochemistry* **2016**, *55*, 762–775. [\[CrossRef\]](#) [\[PubMed\]](#)
- Joseph, P.; Sawant, K.; Iwahara, J.; Garofalo, R.; Desai, U.; Rajarathnam, K. Lysines and Arginines play non-redundant roles in mediating chemokine-glycosaminoglycan interactions. *Sci. Rep.* **2018**, *8*, 12289. [\[CrossRef\]](#) [\[PubMed\]](#)
- Samantray, S.; Olubiyi, O.O.; Strodel, B. The Influences of Sulphation, Salt Type, and Salt Concentration on the Structural Heterogeneity of Glycosaminoglycans. *Int. J. Mol. Sci.* **2021**, *22*. [\[CrossRef\]](#) [\[PubMed\]](#)
- Diaz-Nido, J.; Wandosell, F.; Avila, J. Glycosaminoglycans and  $\beta$ -amyloid, prion and tau peptides in neurodegenerative diseases. *Peptides* **2002**, *23*, 1323–1332. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gruys, E.; Ultee, A.; Upragarin, N. Glycosaminoglycans are part of amyloid fibrils: Ultrastructural evidence in avian AA amyloid stained with cuproinic blue and labeled with immunogold. *Amyloid* **2006**, *13*, 13–19. [\[CrossRef\]](#) [\[PubMed\]](#)
- Decker, Y.; Németh, E.; Schomburg, R.; Chemla, A.; Fülöp, L.; Menger, M.D.; Liu, Y.; Fassbender, K. Decreased pH in the aging brain and Alzheimer's disease. *Neurobiol. Aging* **2021**, *101*, 40–49. [\[CrossRef\]](#)
- Olubiyi, O.O.; Strodel, B. Structures of the Amyloid  $\beta$ -Peptides A $\beta$ 1–40 and A $\beta$ 1–42 as Influenced by pH and a D-Peptide. *J. Phys. Chem. B* **2012**, *116*, 3280–3291. [\[CrossRef\]](#)
- Liao, Q.; Owen, M.C.; Bali, S.; Barz, B.; Strodel, B. A $\beta$  under stress: The effects of acidosis, Cu<sup>2+</sup>-binding, and oxidation on amyloid  $\beta$ -peptide dimers. *Chem. Commun.* **2018**, *54*, 7766–7769. [\[CrossRef\]](#)
- Klement, K.; Wieligmann, K.; Meinhardt, J.; Hortschansky, P.; Richter, W.; Fändrich, M. Effect of Different Salt Ions on the Propensity of Aggregation and on the Structure of Alzheimer's A $\beta$ (1–40) Amyloid Fibrils. *J. Mol. Biol.* **2007**, *373*, 1321–1333. [\[CrossRef\]](#)



30. Wang, H.; Wu, J.; Sternke, R.; Zheng, W.; Mörmann, C.; Luo, J. Multivariate effects of pH, salt, and Zn ions on A $\beta$ 40 fibrillation. *Commun. Chem.* **2022**, *5*, 171. [\[CrossRef\]](#)
31. Sanderson, J. The Association of Lipids with Amyloid Fibrils. *J. Biol. Chem.* **2022**, *298*, 102108. [\[CrossRef\]](#)
32. Zhaliyazka, K.; Matveyenka, M.; Kurouski, D. Lipids uniquely alter the secondary structure and toxicity of amyloid beta 1–42 aggregates. *FEBS J.* **2023**, *290*, 3203–3220. [\[CrossRef\]](#)
33. Coles, M.; Bicknell, W.; Watson, A.A.; Fairlie, D.P.; Craik, D.J. Solution Structure of Amyloid  $\beta$ -Peptide(1–40) in a Water–Micelle Environment. Is the Membrane-Spanning Domain Where We Think It Is? *Biochemistry* **1998**, *37*, 11064–11077. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Shao, H.; Jao, S.; Ma, K.; Zagorski, M.G. Solution structures of micelle-bound amyloid beta-(1–40) and beta-(1–42) peptides of Alzheimer’s disease. *J. Mol. Biol.* **1999**, *285*, 755–773. [\[CrossRef\]](#)
35. Jarvet, J.; Danielsson, J.; Damberg, P.; Oleszczuk, M.; Gräslund, A. Positioning of the Alzheimer A $\beta$ (1–40) peptide in SDS micelles using NMR and paramagnetic probes. *J. Biomol. NMR* **2007**, *39*, 63–72. [\[CrossRef\]](#) [\[PubMed\]](#)
36. La Rosa, C.; Scalisi, S.; Lolicato, F.; Pannuzzo, M.; Raudino, A. Lipid-assisted protein transport: A diffusion-reaction model supported by kinetic experiments and molecular dynamics simulations. *J. Chem. Phys.* **2016**, *144*, 184901. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Sciacca, M.F.; Lolicato, F.; Tempa, C.; Scollo, F.; Sahoo, B.R.; Watson, M.D.; García-Viñuales, S.; Milardi, D.; Raudino, A.; Lee, J.C.; et al. Lipid-Chaperone Hypothesis: A Common Molecular Mechanism of Membrane Disruption by Intrinsically Disordered Proteins. *ACS Chem. Neurosci.* **2020**, *11*, 4336–4350. [\[CrossRef\]](#)
38. Abraham, M.J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J.C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1*, 19–25. [\[CrossRef\]](#)
39. Huang, J.; Rauscher, S.; Nawrocki, G.; Ting, R.; Feig, M.; de Groot, B.; Grubmüller, H.; MacKerell, A. CHARMM36m: An Improved Force Field for Folded and Intrinsically Disordered Proteins. *Nat. Methods* **2017**, *14*, 71–73. [\[CrossRef\]](#)
40. Samantray, S.; Yin, F.; Kav, B.; Strodel, B. Different force fields give rise to different amyloid aggregation pathways in molecular dynamics simulations. *J. Chem. Inf. Model.* **2021**, *60*, 6462–6475. [\[CrossRef\]](#)
41. Pastor, R.W.; MacKerell, A.D. Development of the CHARMM Force Field for Lipids. *J. Phys. Chem. Lett.* **2011**, *2*, 1526–1532. [\[CrossRef\]](#)
42. Jo, S.; Song, K.C.; Desaire, H.; MacKerell, A.D.; Im, W. Glycan reader: Automated sugar identification and simulation preparation for carbohydrates and glycoproteins. *J. Comput. Chem.* **2011**, *32*, 3135–3141. [\[CrossRef\]](#)
43. Park, S.J.; Lee, J.; Patel, D.S.; Ma, H.; Lee, H.S.; Jo, S.; Im, W. Glycan Reader is improved to recognize most sugar types and chemical modifications in the Protein Data Bank. *Bioinformatics* **2017**, *33*, 3051–3057. [\[CrossRef\]](#)
44. Park, S.J.; Lee, J.; Qi, Y.; Kern, N.R.; Lee, H.S.; Jo, S.; Joong, I.; Joo, K.; Lee, J.; Im, W. CHARMM-GUI Glycan Modeler for modeling and simulation of carbohydrates and glycoconjugates. *Glycobiology* **2019**, *29*, 320–331. [\[CrossRef\]](#) [\[PubMed\]](#)
45. Jo, S.; Kim, T.; Iyer, V.G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865. [\[CrossRef\]](#)
46. Samantray, S.; Strodel, B. The Effects of Different Glycosaminoglycans on the Structure and Aggregation of the Amyloid- $\beta$  (16–22) Peptide. *J. Phys. Chem. B* **2021**, *125*, 5511–5525. [\[CrossRef\]](#)
47. Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.D.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935. [\[CrossRef\]](#)
48. Parrinello, M.; Rahman, A. Polymorphic Transitions in Single-Crystals—A New Molecular-Dynamics Method. *Mol. Phys.* **1981**, *52*, 7182–7190. [\[CrossRef\]](#)
49. Nosé, S. Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.* **1984**, *52*, 255–268. [\[CrossRef\]](#)
50. Hoover, W.G. Canonical Dynamics—Equilibrium Phase-Space Distributions. *Phys. Rev. A* **1985**, *31*, 1695–1697. [\[CrossRef\]](#) [\[PubMed\]](#)
51. Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101. [\[CrossRef\]](#)
52. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N-log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [\[CrossRef\]](#)
53. Krause, D.; Thörnig, P. JURECA: Modular supercomputer at Jülich Supercomputing Centre. *J. Large-Scale Res. Facil. JLSRF* **2018**, *4*, A132. [\[CrossRef\]](#)
54. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In Proceedings of the Third International AAAI Conference on Weblogs and Social Media, San Jose, CA, USA, 17–20 May 2009.
55. Blondel, V.D.; Guillaume, J.L.; Lambiotte, R.; Lefebvre, E. Fast unfolding of communities in large networks. *J. Stat. Mech. Theory Exp.* **2008**, *2008*, P10008. [\[CrossRef\]](#)
56. Chebaro, Y.; Ballard, A.J.; Chakraborty, D.; Wales, D.J. Intrinsically Disordered Energy Landscapes. *Sci. Rep.* **2015**, *5*, 10386. [\[CrossRef\]](#)
57. Giovambattista, N.; Debenedetti, P.G.; Sciortino, F.; Stanley, H.E. Structural order in glassy water. *Phys. Rev. E* **2005**, *71*, 061505. [\[CrossRef\]](#) [\[PubMed\]](#)

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

# Supplementary Materials: Transition Networks Unveil Disorder-to-Order Transformations in A $\beta$ Caused by Glycosaminoglycans or Lipids

Moritz Schäffler<sup>1,2</sup>, Suman Samantray<sup>1</sup>  and Birgit Strodel<sup>1,2,\*</sup> 

<sup>1</sup> Institute of Biological Information Processing: Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52428 Jülich, Germany

<sup>2</sup> Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

\* Correspondence: b.strodel@fz-juelich.de

---

## Supplementary Tables

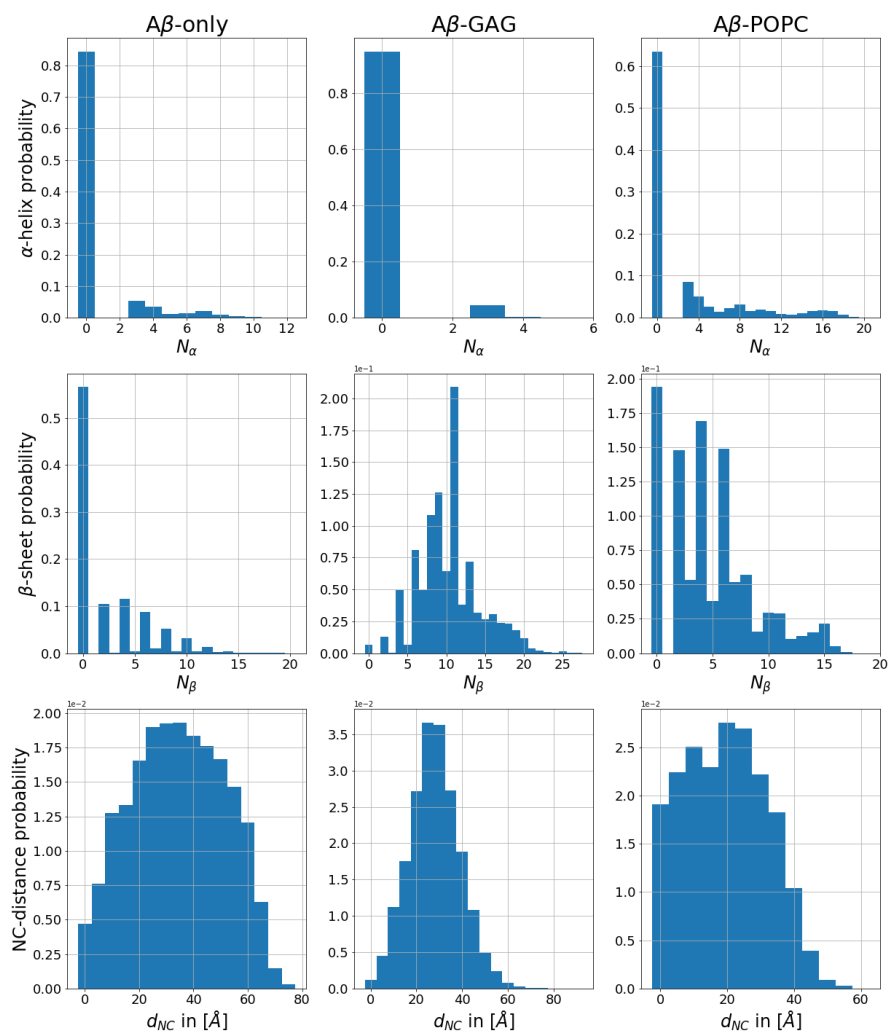
	$N_\alpha$	$N_\beta$	$d_{\text{NC}}/[\text{\AA}]$
A $\beta$ only	$0.8 \pm 1.9$	$2 \pm 3$	$34 \pm 17$
A $\beta$ in A $\beta$ -GAG	$0.2 \pm 0.7$	$10 \pm 4$	$28 \pm 11$
A $\beta$ in A $\beta$ -POPC	$3 \pm 5$	$5 \pm 4$	$20 \pm 12$

**Table S1.** Mean and variance of the three descriptors used for the transitions networks of the A $\beta$ -only, A $\beta$ -GAG, and A $\beta$ -POPC systems. The descriptors are: i) the number of residues forming  $\alpha$ -helical structure ( $N_\alpha$ ), ii) the number of residues forming  $\beta$ -sheet structure ( $N_\beta$ ), iii) the N-to-C distance ( $d_{\text{NC}}$ ).

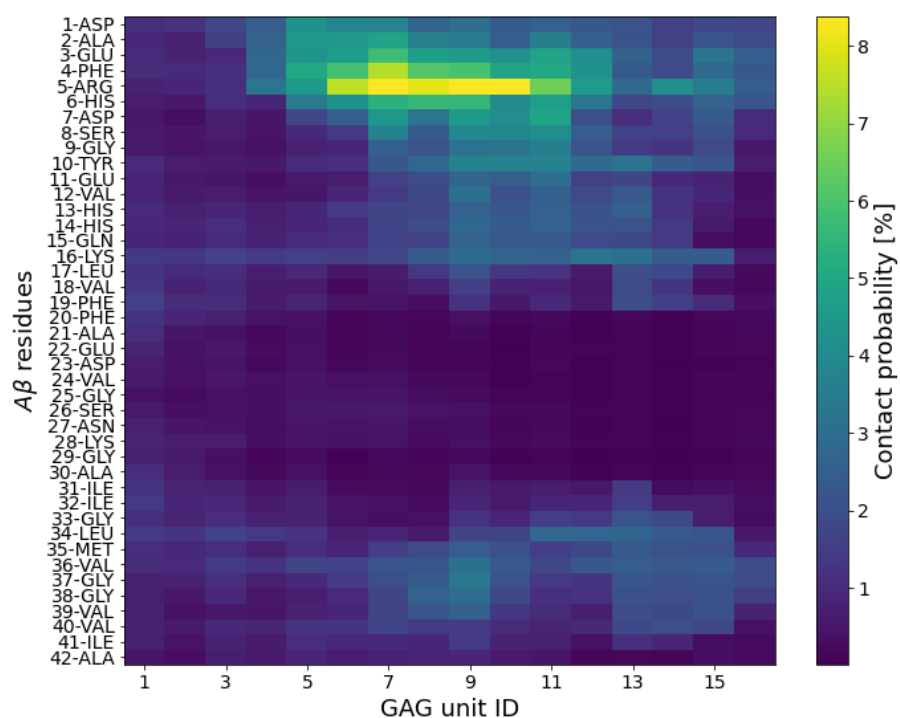
Community	Orange	Yellow	Black	Blue	Pink	Other
A $\beta$ -only	20.0%	40.2%	8.7%	19.7%	7.1%	4.3%
A $\beta$ -GAG	0.1%	1.0%	8.3%	70.3%	16.7%	3.6%
A $\beta$ -POPC	13.1%	-	22.3%	54.9%	4.5%	5.2%

**Table S2.** Population of the transition network communities (identified by color) of the three A $\beta$  systems.

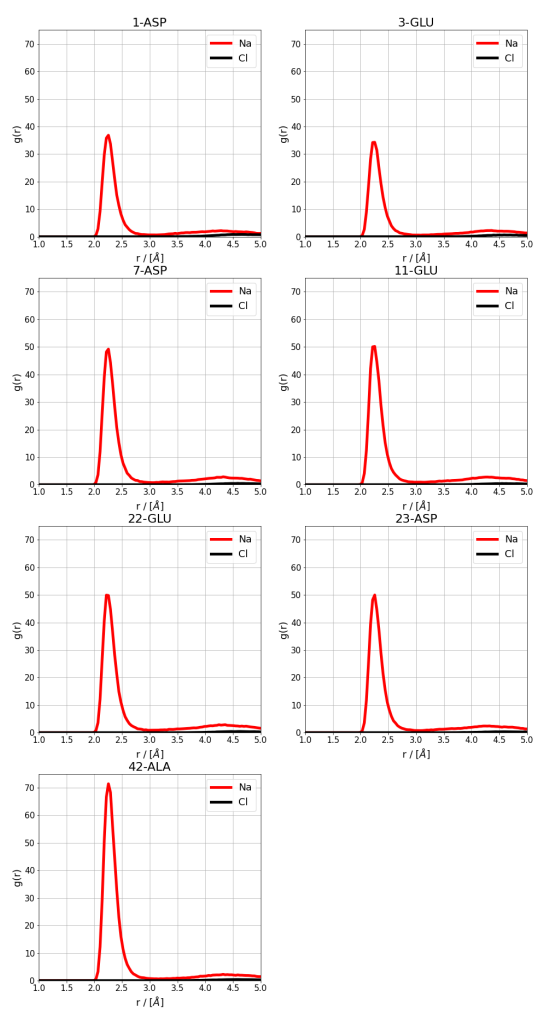
## Supplementary Figures



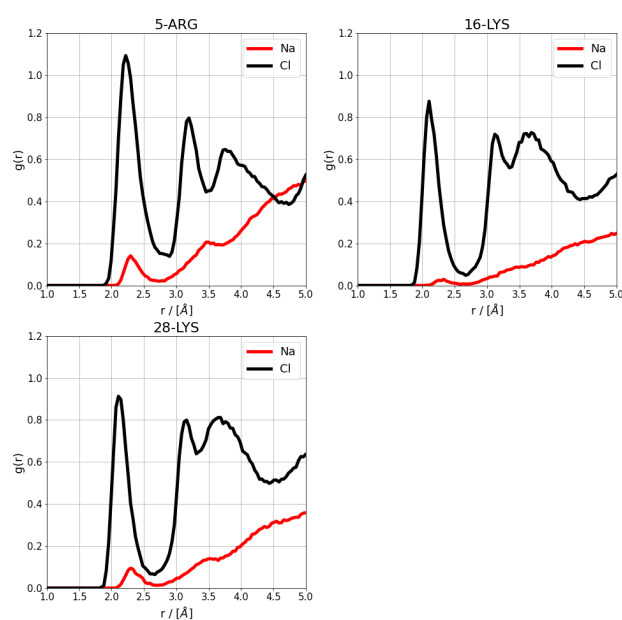
**Figure S1.** Probability distributions for the three descriptors used in the transitions network of the A $\beta$ -only system (left), A $\beta$ -GAG system (center), and A $\beta$ -POPC system (right). The descriptors are: i) the number of residues forming  $\alpha$ -helical structure ( $N_\alpha$ , top), ii) the number of residues forming  $\beta$ -sheet structure ( $N_\beta$ , middle), iii) the N-to-C distance ( $d_{NC}$ , bottom).



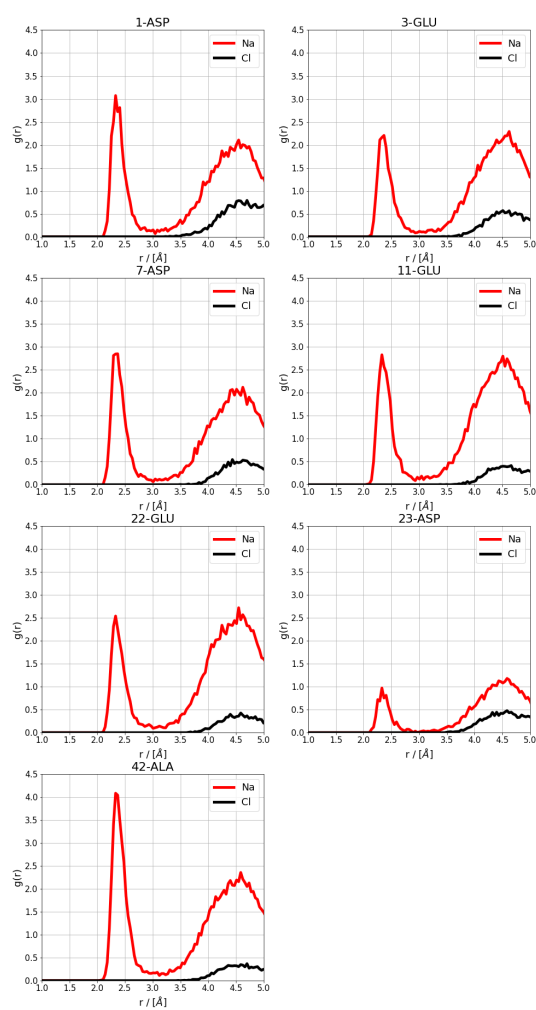
**Figure S2.** Intermolecular contact maps for Aβ interacting with a GAG molecule. The interactions are separated into residue–monosaccharide interactions. Two interaction partners were considered to be in contact if in a given frame of the trajectory they are closer than 10 Å. The resulting number of contacts were normalized with respect to the total number of time frames per trajectory, yielding a contact probability with a maximum value of about 8.5% (see color scale on the right).



**Figure S3.** The radial distribution  $g(r)$  of Na<sup>+</sup> (red) and Cl<sup>-</sup> (black) relative to the carboxyl groups of negatively charged Aβ residues in the Aβ-only system.

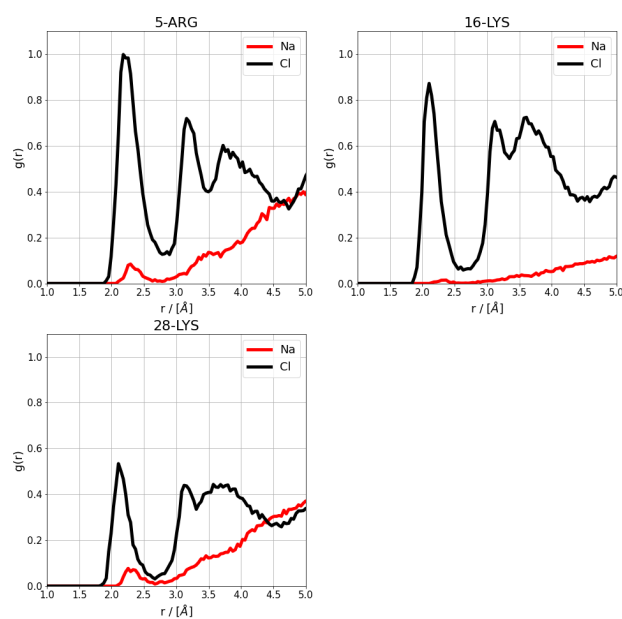


**Figure S4.** The radial distribution  $g(r)$  of Na<sup>+</sup> (red) and Cl<sup>-</sup> (black) relative to the charged groups of positively charged Aβ residues in the Aβ-only system.

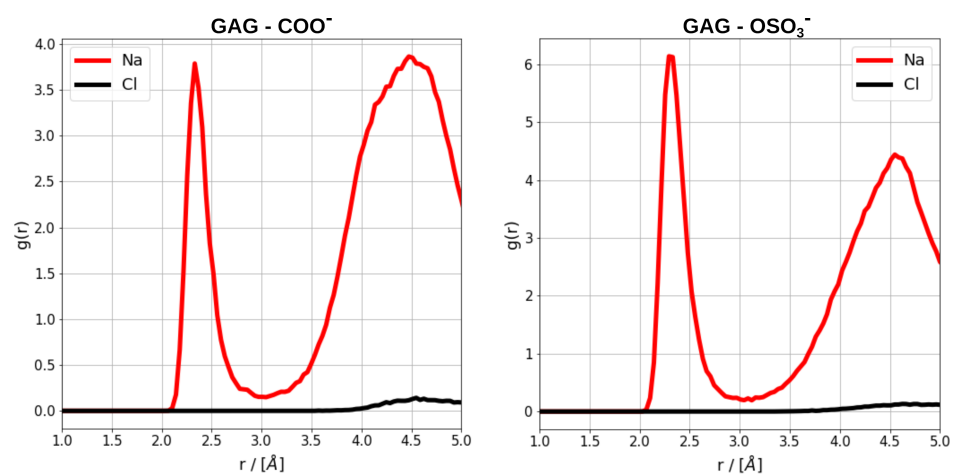


**Figure S5.** The radial distribution  $g(r)$  of  $\text{Na}^+$  (red) and  $\text{Cl}^-$  (black) relative to the carboxyl groups of negatively charged A $\beta$  residues in the A $\beta$ -GAG system.





**Figure S6.** The radial distribution  $g(r)$  of Na<sup>+</sup> (red) and Cl<sup>-</sup> (black) relative to the charged groups of positively charged Aβ residues in the Aβ-GAG system.



**Figure S7.** The radial distribution  $g(r)$  of  $\text{Na}^+$  (red) and  $\text{Cl}^-$  (black) relative to the  $\text{COO}^-$  (left) and  $\text{OSO}_3^-$  (right) groups of the GAG in the  $\text{A}\beta$ -GAG system.



Cite this: DOI: 10.1039/d4cc02856b

Received 12th June 2024,  
Accepted 24th October 2024

DOI: 10.1039/d4cc02856b

rsc.li/chemcomm

# The energy landscape of A $\beta$ <sub>42</sub>: a funnel to disorder for the monomer becomes a folding funnel for self-assembly†

Moritz Schäffler,<sup>ab</sup> David J. Wales<sup>id</sup><sup>c</sup> and Birgit Strodel<sup>id</sup><sup>\*ab</sup>

The aggregation of amyloid- $\beta$  (A $\beta$ ) peptides, particularly A $\beta$ <sub>1–42</sub>, plays a key role in Alzheimer's disease pathogenesis. In this study, we investigate how dimerisation transforms the free energy surface (FES) of the A $\beta$ <sub>1–42</sub> monomer when it interacts with another A $\beta$ <sub>1–42</sub> peptide. We find that the monomer FES is a structurally inverted funnel with a disordered state at the global minimum. However, in the presence of a second A $\beta$ <sub>1–42</sub> peptide, the landscape becomes a folding funnel, leading to a  $\beta$ -hairpin state. Using first passage time analysis, we analyse the pathway for the transition from disordered to the  $\beta$ -hairpin state, which highlights the initial formation of a D23–K28 salt bridge as the driving force, together with hydrophobic contacts.

Alzheimer's disease, a neurodegenerative disorder, is linked to the aggregation and misfolding of A $\beta$  peptides, with A $\beta$ <sub>42</sub> being especially prone to forming insoluble fibrillar structures in the brain.<sup>1</sup> As a monomer, A $\beta$  is an intrinsically disordered peptide (IDP), lacking a stable secondary or tertiary structure.<sup>2,3</sup> However, during the aggregation process, A $\beta$  undergoes conformational changes, leading to more ordered structures.<sup>4</sup> Understanding these structural transitions is crucial for elucidating the molecular mechanisms underlying Alzheimer's pathology, as they are pivotal in the formation of toxic oligomers and fibrils.<sup>5</sup> Studying amyloid aggregation at the structural level in experiments is challenging due to the heterogeneity and transient nature of amyloid oligomers, as their diverse and short-lived states complicate detailed characterisation. Techniques such as nuclear magnetic resonance (NMR) spectroscopy, cryogenic electron microscopy (cryo-EM), and X-ray crystallography usually struggle to resolve these small, dynamic oligomeric intermediates.<sup>6</sup> Molecular simulations can

therefore be helpful, providing insights into the conformational dynamics and aggregation pathways of A $\beta$  peptides.<sup>7</sup>

The energy landscapes of IDPs feature complex topographies with multiple minima representing various conformational states, which may correspond to an evolved intrinsically multifunctional capability.<sup>8</sup> The energy landscapes for amyloid aggregation are even more heterogeneous, reflecting the diverse conformational and aggregation states amyloid proteins can adopt.<sup>9</sup> In a previous study that applied NMR-based metadynamics simulations to A $\beta$ <sub>40</sub>, the free energy landscape revealed what the authors called an 'inverted landscape', in which the global minimum consists of disordered structures, while higher energy regions contain transiently structured conformations, suggesting increased structuring at higher temperatures.<sup>10</sup> This result contrasts with the multifunnel energy landscapes characterised for IDPs,<sup>8,11</sup> which feature low-lying minima of similar energy separated by substantial energy barriers that trap the system in metastable states. In this study, we determine the free energy surface (FES) of A $\beta$ <sub>42</sub> as a monomer and for an initial aggregation step into  $\beta$ -sheet-rich dimers, offering fresh insights into the structural dynamics during A $\beta$ <sub>42</sub> aggregation. We suggest that the monomer landscape might be described as a 'disordered funnel', as explained below.

To obtain the FES of the monomer and the dimer, we performed molecular dynamics (MD) simulations using the all-atom force field CHARMM36m<sup>12</sup> and sampled 6 and 24  $\mu$ s for the A $\beta$ <sub>42</sub> monomer and dimer, respectively. The A $\beta$ <sub>42</sub> peptides in each frame of the MD simulations of both monomer and dimer are characterised by the distribution of reciprocal interatomic distance (DRID) metric, a structure-specific dimensionality reduction method known for preserving system kinetics effectively.<sup>13,14</sup> We used the C $\alpha$  atoms of structurally important residues of A $\beta$ <sub>42</sub>, namely D1, F19, D23, K28, L34, and A42, as reference atoms to define the DRID metric. For each of the reference atoms, the first three moments of the DRID are used to define the state vector, resulting in a  $3N \rightarrow 3N_c$  dimensionality reduction, where  $N$  is the number of atoms in the molecule and  $N_c$  the number of reference atoms. After clustering, the trajectory of states is used to calculate the rate matrix of transitions between them, which is translated into free energies of the states (minima) and

<sup>a</sup> Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany. E-mail: b.strodel@fz-juelich.de

<sup>b</sup> Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52428, Jülich, Germany

<sup>c</sup> Yusuf Hamied Department of Chemistry, University of Cambridge, CB2 1EW Cambridge, UK

† Electronic supplementary information (ESI) available. See DOI: <https://doi.org/10.1039/d4cc02856b>



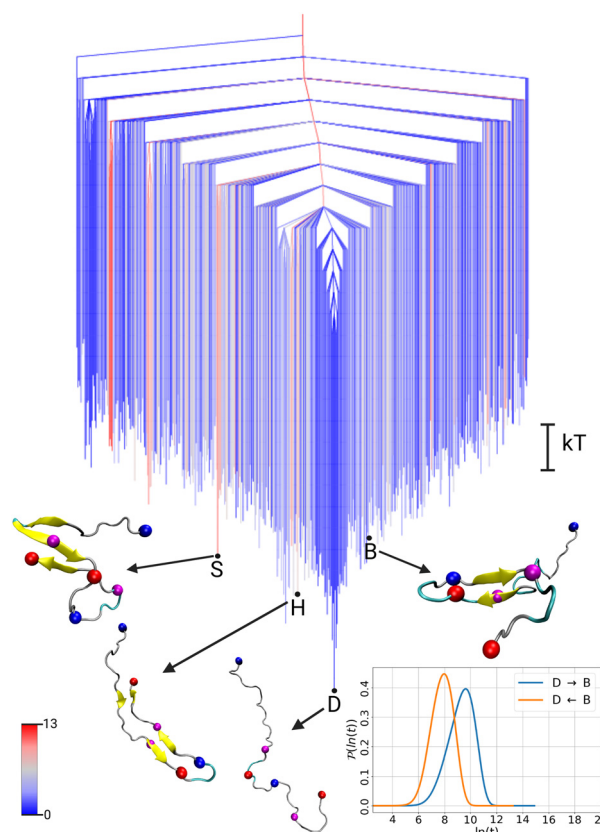
## Communication

transition states that connect them.<sup>15,16</sup> The resulting FESs are visualised using disconnectivity graphs,<sup>17,18</sup> where the branches terminate at the energies of local minima, and the vertical scale is the free energy. These branches are joined at the threshold where the free energy exceeds the value for the highest transition state on the lowest path between them. Hence the branches correspond to a segregation of the minima into sets that can interconvert amongst themselves at any given level on the vertical scale. Similar states leading to the same local minimum are grouped in the same basin, facilitating intuitive analysis of the underlying structure of the landscape and identification of folding funnels. Central to our investigation is the application of first-passage time (FPT) distribution analysis to extract the kinetics of transitions between different states.<sup>16,19</sup> This approach allows us to address the temporal aspects of the conformational changes, shedding light on the pathways and rates governing the transitions between disordered and  $\beta$ -sheet-rich conformations. Detailed information on the methods can be found in the ESI†

First, we discuss the free energy surface of the  $A\beta_{42}$  monomer, for which the disconnectivity graph is shown in Fig. 1. The FES exhibits one main funnel leading to the global minimum. In contrast to the usual landscapes of folded proteins, the global minimum corresponds to a disordered state of the  $A\beta_{42}$  monomer (labeled as state D for disorder). Conformations featuring partial secondary structure, such as a  $\beta$ -hairpin characteristic for  $A\beta$  oligomers<sup>20</sup> or the S-shape conformation that is the usual building block of fibrils,<sup>4</sup> are found as excited states in the FES of the monomer with energies of  $\Delta F_H^{\text{mon}} = 2.3$  kT and  $\Delta F_S^{\text{mon}} = 3.2$  kT respectively. Here, an excited state is defined as a higher free energy minimum, not to be confused with an electronic excited state. This organisation of the FES with (partially) folded states at the top of the funnel, and the disordered states at the bottom, has previously been called an 'inverted free energy landscape'.<sup>10</sup> We suggest the nomenclature 'structurally inverted funnel' or simply 'disordered funnel', since it is the structure that is inverted, not the funnel.

To illustrate the FES of the  $A\beta_{42}$  dimer and allow direct comparison to the monomer, we treat the two chains in the dimer systems as individual peptides and present the disconnectivity graph of the dimer for an  $A\beta_{42}$  peptide in the environment of the second  $A\beta_{42}$  peptide. This representation further allows us to use the same DRID metric to calculate the states, highlighting the structural changes in the presence of another  $A\beta_{42}$  peptide. The resulting FES for the  $A\beta_{42}$  dimer, shown in Fig. 2, exhibits the topology of a folding funnel leading to a  $\beta$ -hairpin state with significant  $\beta$ -sheet content (hence labeled state B). Projecting the global minimum of the monomer FES (state D) onto the dimer FES, reveals that the disordered state appears as an excited state in the dimer FES, with a free energy difference of  $\Delta F_D^{\text{dim}} = 3.9$  kT. Similarly, the global minimum of the dimer FES (state B) can be projected onto the monomer FES (see Fig. 1), where it is likewise an excited state, with a free energy difference of  $\Delta F_B^{\text{mon}} = 3.6$  kT. We note that the D state in the dimer FES is not yet a dimer, but two disordered monomers, while the  $\beta$ -hairpin in the B state of the dimer is a result of interpeptide interactions due to dimerisation (Fig. S1, ESI†).

To analyse the timescales associated with the interconversions between the disordered state D and the  $\beta$ -hairpin state B, we calculated the first passage time distribution for these transitions



**Fig. 1** Free energy disconnectivity graph for the FES of the  $A\beta_{42}$  monomer. The energies are given in units of kT (see scale bar on the right), with  $k$  the Boltzmann constant, and  $T$  the absolute temperature. The branches are colored according to the average number of residues in  $\beta$ -sheet conformation in the ensemble of structures belonging to the corresponding minimum, ranging from blue (no  $\beta$ -sheets) to red (13 residues involved in  $\beta$ -sheets). Representative structures of some minima are shown, where D (for 'disordered') is the global minimum of the monomer FES and B (for ' $\beta$ -sheet') is the global minimum of the dimer FES projected onto the monomer FES. The structures are shown in the cartoon representation, with  $\beta$ -sheets highlighted in yellow and the centroids used in the DRID metric shown as spheres (blue for positive charge at the N-terminus and K28 side chain, red for negative charge at the C-terminus and D23, magenta for the hydrophobic F19 and L34). In the bottom right corner, the probability distribution ( $\ln(t)$ ) of the first passage time  $t$  for transitions between states D and B is shown on a logarithmic scale.

(see ESI† for details). The resulting FPTs for the monomer and dimer are shown in the lower right corner of Fig. 1 and 2, respectively. For the monomer, the transition times are  $\tau_{D \rightarrow B}^{\text{mon}} \sim 15$  ns and  $\tau_{B \rightarrow D}^{\text{mon}} \sim 3$  ns. While the transition time from the disordered state to the excited  $\beta$ -hairpin state is five times longer than the reverse transition, the transition to the excited state is still relatively fast. In contrast, for the dimer the transition times are  $\tau_{D \rightarrow B}^{\text{dim}} \sim 4$  ns and  $\tau_{B \rightarrow D}^{\text{dim}} \sim 300$  ns, with a ratio of about 75 for the transition from the global minimum to the excited disordered state with respect to the reverse transition. Thus, we observe a separation of timescales in the dimer, but not the monomer.

These results highlight the differences between the FES of the  $A\beta_{42}$  monomer and dimer. For the disordered funnel of the monomer, higher local minima are readily accessible and the time scales of interconversions between them are of similar



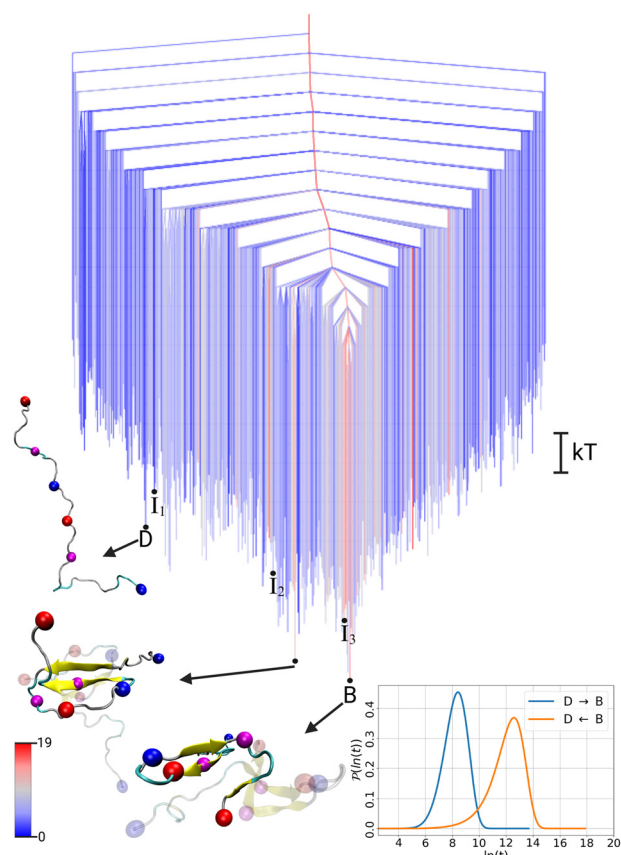


Fig. 2 Free energy disconnectivity graph for the FES of the  $A\beta_{42}$  dimer. The branches are colored according to the average number of residues in  $\beta$ -sheet conformation in the ensemble of structures belonging to the respective minimum, ranging from blue (no  $\beta$ -sheets) to red (19 residues involved in  $\beta$ -sheets). For further details, see the caption of Fig. 1. The cartoons of all indicated dimer structures are shown in Fig. S1 (ESI†).

magnitude, while transitions to excited states from the ordered global minimum of the dimer are relatively slow. Thus, while we characterise the monomer FES as a disordered funnel, the dimer FES is closer to the expected form for a folded protein. Our FES and timescale analysis of  $A\beta_{42}$  further confirms the predictions made for the energy landscapes of folded and disordered proteins based on the slope of the free energy funnels.<sup>21</sup> Chong and Ham determined the FESs of two folded proteins, which exhibited steep folding funnels, and of an IDP, which has a gentler slope and led to similar transition times to

and from the global minimum. A significant difference from our study is that the global minimum of the IDP they considered is a partially folded state, whereas at the global minimum of  $A\beta_{42}$  is essentially fully disordered.

To gain further insight into the structural transition from disordered state D to folded state B in the  $A\beta_{42}$  dimer, we extracted structural pathways associated with the fastest transition in the dimer FES (see ESI† for details). From the ensemble that includes all the possibilities  $D \rightarrow B$  we identified three characteristic pathways (Fig. S2 in ESI†), by associating structures from the MD simulations with the states in the kinetic model. These pathways include three intermediate states ( $I_1$ ,  $I_2$ ,  $I_3$ ), which are marked on the FES in Fig. 2. Here, the most important of these three pathways, shown in Fig. 3, is discussed along with the FPTs associated with the fastest transitions between sequential states.

The initial transition  $D \rightarrow I_1$  is characterised by a scissor motion, which closes the relatively extended conformation to a hairpin-like structure. The main driving force for this movement is the formation of a salt bridge between D23 and K28, which is then followed by the establishment of hydrophobic contacts between the strands on both sides of the turn, involving residues <sup>17</sup>LVFFA<sup>21</sup> and <sup>30</sup>AIIGLMV<sup>36</sup>. The formation of these intrapeptide contacts occurs cooperatively with the establishment of interpeptide contacts between the hydrophobic regions of both peptides (Fig. S3 with in-depth discussion in the ESI†). In state  $I_1$ , the most important hydrophobic contact is between F19 and L34, which is structurally suboptimal and therefore shifts to F19–I32 in state  $I_2$ . In fact, the intermediate  $I_1$  is  $\sim 1$  kT higher in energy than state D, and there is a significant barrier of  $\sim 9.4$  kT that needs to be overcome for the  $D \rightarrow I_1$  transition. This barrier is reflected in the four to nine times slower transition time for this first transition compared to the subsequent transitions. Nonetheless, the first passage time associated with the process  $\tau_{D \rightarrow I_1}^{\text{dim}} \sim 780$  ps is still very fast compared to the height of the energy barrier, which can be explained with the FPT probability distribution (Fig. S4 in ESI†). This distribution features both fast and slow relaxation modes, resulting in a mean first passage time that appears slow, which is also reflected in the high free energy barrier. Hence, analysis of the full FPT distribution once again reveals important details that are not evident from average rates.<sup>19</sup>

After reaching  $I_1$ , the system moves downhill in energy to the global minimum. After the positions of the hydrophobic contacts have been optimised in  $I_2$ , hydrogen-bonds form between

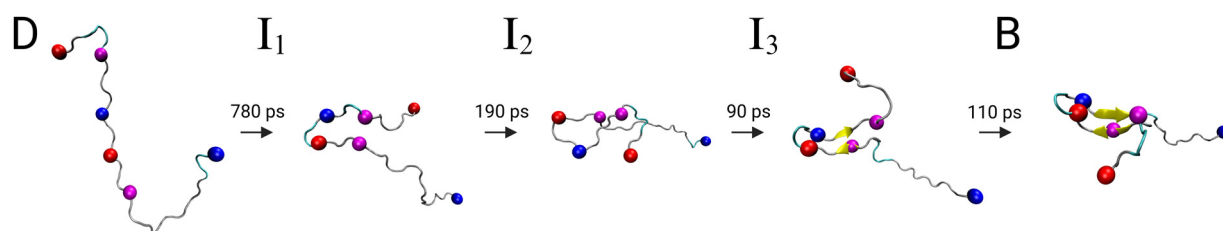


Fig. 3 Fastest pathway from the intrinsically disordered state (D) to the global minimum (B) in the dimer free energy surface. The spheres represent the centroids used in the DRID metric. The FPTs associated with the fastest transition between the states along the pathway are shown above the arrows, which were calculated separately for each transition and therefore do not add up to the total transition time associated with the full pathway, where all possible recrossing events are included. Fig. S1 (ESI†) shows the pathway again, but with both peptides within the dimer.





the strands, leading to a short antiparallel  $\beta$ -sheet in  $I_3$ . Inspecting the dimer FES, we see that the intermediate  $I_3$  is already located in the central folding funnel, close to the global minimum. The final transition  $I_3 \rightarrow B$ , further stabilizes the  $\beta$ -sheet and involves a reorientation of the C-terminal.

The elucidation of the pathway for structure formation during  $A\beta_{42}$  dimerisation confirms the relevance of the D23–K28 salt bridge for turn formation.<sup>22,23</sup> Our results reveal that the creation of this salt bridge initiates the turn, leading to  $\beta$ -hairpin formation. A recent study demonstrated that  $A\beta$  oligomerisation depends on this specific  $\beta$ -hairpin motif, as variants of the  $A\beta$  peptide that cannot adopt this hairpin structure do not produce oligomers.<sup>20</sup> In addition, this  $\beta$ -hairpin structure is crucial for mediating  $A\beta$  toxicity, as demonstrated by a study in which  $A\beta$  oligomers with a permanently present hairpin—achieved by introducing a disulfide bridge between residues 21 and 30—proved to be particularly toxic.<sup>24</sup> Our observations of the hydrophobic contact formation are consistent with a recent low-temperature solid-state NMR spectroscopy study of freeze-trapped  $A\beta_{40}$  oligomers.<sup>25</sup> Tycko and coworkers identified  $\beta$ -hairpin conformations with contacts within the peptide and between the two main hydrophobic segments of  $A\beta_{40}$  that developed within 1 ms. The primary hydrophobic contact was located between the aromatic sidechain of F19 and aliphatic sidechains in residues 30–35, which fully agrees with the structures in the folding funnel of the dimer FES determined here.

In conclusion, our study provides new insight into the structural dynamics of  $A\beta_{42}$  with implications for Alzheimer's disease pathology. Through analysis of free energy landscapes and first-passage time distributions, we have elucidated key aspects of  $A\beta_{42}$  conformational transitions from a disordered to a  $\beta$ -sheet forming state. Our results reveal distinctive features in the free energy landscapes of  $A\beta_{42}$  monomers and dimers, highlighting the impact of the molecular environment on structural stability and propensity for aggregation. Specifically, we observe that the FES of the monomer resembles a structurally inverted folding funnel, where disordered states are lowest in energy. In contrast, the lowest minima for the dimer are characterised by more ordered structures, predominantly consisting of  $\beta$ -hairpins, which form as  $A\beta_{42}$  folds upon binding to the hydrophobic region of another  $A\beta_{42}$  peptide. Thus, the increase in hydrophobicity in the environment of  $A\beta_{42}$  appears to be crucial for its folding during self-assembly. Our analysis of first-passage times indicates that the FES of the monomer is relatively flat, with excited states (higher free energy minima) more readily accessible compared to the dimer. This organisation reflects the disordered nature of the low-lying states for the monomeric form of  $A\beta_{42}$ .

Our pathway analysis identifies key intermediates and structural rearrangements involved in the transition from disordered to folded states, providing mechanistic insights into the initial aggregation steps for  $A\beta_{42}$ . Starting from an extended conformation, the peptide chain undergoes a scissor-like motion, establishing a salt bridge between residues D23 and K28, followed by the formation of hydrophobic contacts. Subsequently, these contacts reorganise, leading to  $\beta$ -hairpin structures. Our results should contribute to a deeper understanding of the molecular mechanisms underlying Alzheimer's disease pathology, and may inform the development of targeted therapeutic interventions.

The authors gratefully acknowledge computing time on the supercomputer JURECA at Forschungszentrum Jülich under grant IDP, financial support from the German Academic Exchange Service, and valuable discussions with Prof. Debayan Chakraborty.

## Data availability

The code for PATHSAMPLE can be found at <https://www.wales.ch.cam.ac.uk/software.html>. The data analysis scripts are available at <https://github.com/strodel-group/> under 'freeEnergyCalculation' and 'DRIDmetric'.

## Conflicts of interest

There are no conflicts to declare.

## References

- D. J. Selkoe and J. Hardy, *EMBO Mol. Med.*, 2016, **8**, 595–608.
- J. Roche, Y. Shen, J. H. Lee, J. Ying and A. Bax, *Biochemistry*, 2016, **55**, 762–775.
- A. Paul, S. Samantray, M. Anteghini, M. Khaled and B. Strodel, *Chem. Sci.*, 2021, **12**, 6652–6669.
- D. Willbold, B. Strodel, G. F. Schröder, W. Hoyer and H. Heise, *Chem. Rev.*, 2021, **121**, 8285–8307.
- U. Sengupta, A. N. Nilson and R. Kayed, *EBioMedicine*, 2016, **6**, 42–49.
- P. H. Nguyen, A. Ramamoorthy, B. R. Sahoo, J. Zheng, P. Faller, J. E. Straub, L. Dominguez, J.-E. Shea, N. V. Dokholyan, A. De Simone, B. Ma, R. Nussinov, S. Najafi, S. T. Ngo, A. Loquet, M. Chiricotto, P. Ganguly, J. McCarty, M. S. Li, C. Hall, Y. Wang, Y. Miller, S. Melchionna, B. Habenstein, S. Timr, J. Chen, B. Hnath, B. Strodel, R. Kaye, S. Lesné, G. Wei, F. Sterpone, A. J. Doig and P. Derreumaux, *Chem. Rev.*, 2021, **121**, 2545–2647.
- H. Fatafta, M. Khaled, B. Kav, O. O. Olubiyi and B. Strodel, *WIREs Comput. Mol. Sci.*, 2024, **14**, e1703.
- Y. Chebaro, A. J. Ballard, D. Chakraborty and D. J. Wales, *Sci. Rep.*, 2015, **5**, 10386.
- B. Strodel, *J. Mol. Biol.*, 2021, **433**, 167182.
- D. Granata, F. Baftizadeh, J. Habchi, C. Galvagnion, A. Simone, C. Camilloni, A. Laio and M. Vendruscolo, *Sci. Rep.*, 2015, **5**, 15449.
- J. A. Joseph and D. J. Wales, *J. Phys. Chem. B*, 2018, **122**, 11906–11921.
- J. Huang, S. Rauscher, G. Nawrocki, R. Ting, M. Feig, B. de Groot, H. Grubmüller and A. MacKerell, *Nat. Meth.*, 2017, **14**, 71–73.
- T. Zhou and A. Caflisch, *J. Chem. Theory Comput.*, 2012, **8**, 2930–2937.
- D. Chakraborty, J. E. Straub and D. Thirumalai, *Sci. Adv.*, 2023, **9**, eadd6921.
- E. J. Woods, D. Kannan, D. J. Sharpe, T. D. Swinburne and D. J. Wales, *Philos. Transact. A Math. Phys. Eng. Sci.*, 2023, **381**, 20220245.
- E. J. Woods and D. J. Wales, *Phys. Chem. Chem. Phys.*, 2024, **26**, 1640–1657.
- O. M. Becker and M. Karplus, *J. Chem. Phys.*, 1997, **106**, 1495–1517.
- D. J. Wales, M. A. Miller and T. R. Walsh, *Nature*, 1998, **394**, 758–760.
- D. J. Wales, *J. Phys. Chem. Lett.*, 2022, **13**, 6349–6358.
- M. Khaled, I. Rönnebäck, L. L. Ilag, A. Gräslund, B. Strodel and N. Österlund, *J. Am. Chem. Soc.*, 2023, **145**, 18340–18354.
- S.-H. Chong and S. Ham, *Sci. Rep.*, 2019, **9**, 14927.
- G. Reddy, J. E. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 11948–11953.
- M. Schäffler, S. Samantray and B. Strodel, *Int. J. Mol. Sci.*, 2023, **24**, 11238.
- A. Sandberg, L. Luheshi, S. Söllvander, T. Pereira de Barros, B. Macao, T. Knowles, H. Biverstäl, C. Lendel, F. Ekholm-Petterson, A. Dubnovitsky, L. Lannfelt, C. Dobson and T. Härd, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 15595–15600.
- J. Jeon, W.-M. Yau and R. Tycko, *Nat. Commun.*, 2023, **14**, 2964.



## ELECTRONIC SUPPLEMENTARY INFORMATION:

### The energy landscape of A $\beta$ <sub>42</sub>: a funnel to disorder for the monomer becomes a folding funnel for self-assembly

Moritz Schäffler,<sup>a,b</sup> David J. Wales,<sup>c</sup> and Birgit Strodel<sup>a,b,\*</sup>

#### 1 Methods

##### 1.1 Simulation Details

In this study, molecular dynamics (MD) simulations were conducted for the A $\beta$ <sub>1–42</sub> monomer and dimer. In both systems, A $\beta$ <sub>1–42</sub> was modeled with neutral histidine and no terminal capping groups, resulting in an overall peptide charge of 3–. The simulations were performed using the GROMACS simulation package.<sup>1</sup> As the monomer and dimer simulations where originally performed as part of different studies, some of the simulation settings differ slightly, but the most relevant settings are identical, in particular the force field parameters and ion concentration are the same. We are confident that the slight differences in the setup do not affect our key conclusions.

The CHARMM36m force field<sup>2</sup> was employed in all simulations. Previous research has revealed that the CHARMM36m force field provides an accurate representation of monomeric A $\beta$ <sup>3</sup> and is well suited for simulating amyloid aggregation phenomena.<sup>4</sup>

All system preparations followed the same protocol: the peptide(s) was(were) positioned within the simulation box with a minimum distance of 1.2 nm between any peptide atom and the simulation box faces or edges. The box was subsequently filled with TIP3P water molecules,<sup>5</sup> along with Na<sup>+</sup> and Cl<sup>–</sup> ions to achieve system neutralization and a physiological salt concentration of 150 mM. After system equilibration, each system was simulated under *NpT* conditions at 1 bar, using the Parrinello-Rahman pressure coupling scheme.<sup>6</sup> The dimer simulations were conducted at 298 K using the Nosé-Hoover thermostat,<sup>7,8</sup> while the monomer system was maintained at 300 K using a velocity rescaling thermostat.<sup>9</sup> In all simulations, periodic boundary conditions in all dimensions were applied, with the particle-mesh Ewald method<sup>10</sup> employed for the calculation of electrostatic interactions. The calculation of van der Waals and Coulomb interactions in real space were performed with a cutoff at 1.2 nm.

The total simulation time for the A $\beta$ <sub>1–42</sub> monomer accumulated to 6  $\mu$ s. For the dimer, three simulations were conducted,

each initiated from two extended A $\beta$ <sub>1–42</sub> monomers separated by at least 2 nm. Due to the initial peptide extended conformations, a relatively large simulation box was necessary to prevent self-interaction across periodic boundary conditions, leading to an extensive system predominantly composed of water. To conserve computational resources, after an initial 2  $\mu$ s simulation per dimer simulation, the peptide structures of the last MD frame were extracted and resolvated in a smaller simulation box, as at this point the two peptides had already formed a dimer. Subsequently, after a brief *NpT* equilibration, another 6  $\mu$ s per dimer system were collected, resulting in a cumulative simulation time of 3  $\times$  8  $\mu$ s for the A $\beta$ <sub>1–42</sub> dimer.

All MD simulations were run on the supercomputer JURECA.<sup>11</sup>

##### 1.2 Distance of Reciprocal Interatomic Distances Metric

In order to partition the sampled conformational space over the course of the MD simulation into discrete microstates, we used the distribution of reciprocal interatomic distance (DRID) metric<sup>12</sup> for subsequent structural clustering. Given the size of the A $\beta$ <sub>1–42</sub> monomer, structure-based clustering in Cartesian coordinates becomes a challenging task. One therefore usually relies on some form of dimensionality reduction, which should preserve as much of the kinetics and structural features as possible. It has been shown that the DRID metric is a good candidate to meet both of these criteria.<sup>12,13</sup> A key feature of the DRID metric is the use of the multiplicative inverse (reciprocal) distances, which highlights the difference in short distances, while not neglecting changes in large distances completely.

To apply the DRID metric, two essential atom sets are defined: a set of  $m$  centroids  $\mathcal{C}$  representing key structural elements, and a set of  $N$  reference atoms  $\mathcal{A}$  (excluding atoms that are covalently bound to the centroid). For a given structure, the distribution of reciprocal interatomic distances for each centroid  $i$  and the first three moments of that distribution ( $\mu_i, v_i, \xi_i$ ) are calculated, resulting in a  $3m$  dimensional vector for each structure (i.e., each frame of the MD trajectory). The moments are defined as follows,

$$\mu_i = \frac{1}{N-1-nb_i} \sum_j \frac{1}{d_{ij}}, \quad (1)$$

$$v_i = \left[ \frac{1}{N-1-nb_i} \sum_j \frac{1}{(d_{ij}-\mu_i)^2} \right]^{1/2}, \quad (2)$$

<sup>a</sup> Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

<sup>b</sup> Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52428 Jülich, Germany

<sup>c</sup> Yusuf Hamied Department of Chemistry, University of Cambridge, CB2 1EW Cambridge, U.K.

\* Correspondence: b.strodel@fz-juelich.de

$$\xi_i = \left[ \frac{1}{N-1-nb_i} \sum_i^N \frac{1}{(d_{ij}-\mu_i)^3} \right]^{1/3}, \quad (3)$$

where  $d_{ij}$  denotes the distance of atom  $a_j \in \mathcal{A}$  to centroid  $c_i \in \mathcal{C}$ ,  $nb_i$  is the number of covalent bonds of a centroid, and  $N$  the number of atoms in  $\mathcal{A}$ . The distance metric  $s_{jk}$  between a pair of conformations  $j$  and  $k$  in DRID space is defined as

$$s_{jk} = \frac{1}{3m} \sum_i^m \left[ (\mu_i^j - \mu_i^k)^2 + (v_i^j - v_i^k)^2 + (\xi_i^j - \xi_i^k)^2 \right]^{1/2}. \quad (4)$$

To group the structures into states, we performed regular space clustering in DRID space, as implemented in the PyEMMA python package,<sup>14</sup> using the  $s_{jk}$  distance metric.

For studying the conformational space of the  $\text{A}\beta_{1-42}$  peptide, we chose as centroids the  $C_\alpha$  atoms of structurally important residues, namely D1, F19, D23, K28, L34 and A42, resulting in an 18-dimensional DRID space. We chose D23 and K28, because they have been identified in previous studies to form a salt bridge in the  $\beta$ -hairpin state.<sup>15,16</sup> The residues F19 and L34 are members of the hydrophobic core regions and have been shown by NMR spectroscopy to form contacts in  $\text{A}\beta$  oligomers,<sup>17</sup> while the termini D1 and A42 were included to capture the overall compactness of the peptide. For clustering, we chose a cutoff of  $s_{jk}^c = 0.02 \text{ nm}^{-1}$ , which resulted in 447 states for the monomer and 511 states for the dimer.

To evaluate the robustness of the FES derived from projecting structures into the DRID space, we computed the FES for the dimer system using various DRID metrics. Specifically, we selected five times 6 random residues of the peptide chain as centroids for defining the DRID metric, and constructed five FESs. All FESs exhibited a consistent single funnel structure, similar to the FES presented in this manuscript (Fig. 2). To assess the predictability of our results, we calculated the overlap between the ensemble of states corresponding to the global minimum of the FES presented in Fig. 2 and those of the randomly generated FESs, yielding an average overlap of  $\sim 45\%$ . When the three most prominent states next to the global minimum but within the same basin are further included, there is an overlap of at least 77% between the current and the random FES. Moreover, the intrinsically disordered state consistently appeared as a high-energy excited state in all FESs. These findings underscore the robustness of the DRID metric, showing its reliability relatively independent of centroid selection.

### 1.3 Free Energy Calculation

The free energy surface (FES) of a protein determines its structural and dynamical properties and is therefore of great interest if one wants to understand the protein function. Here, we calculate the free energies associated with the states determined by structural clustering in DRID space, treating each state as a minimum in the FES.<sup>18</sup> The free energy of the minima  $F_i$  are calculated via their occupation probability  $p_i$ ,

$$F_i = -k_B T \log(p_i), \quad (5)$$

where  $k_B$  is the Boltzmann constant and  $T$  the temperature of the system. The rate matrix  $\mathbf{R}$ , representing the state-to-state rates  $r_{jk}$  between minima  $j$  and  $k$  observed in the MD simulations, is used to derive the transition state free energies  $F_{jk}$  via the Eyring–Polanyi formulation. The rate matrix was derived from the transition matrix of the MD trajectory in DRID space, which represents the corresponding right stochastic matrix. Assuming Markovian dynamics, these free energies can be calculated as follows:

$$F_{jk} = F_k - k_B T \log(k_{jk}) + k_B T \log\left(\frac{k_B T}{h}\right), \quad (6)$$

where  $h$  is the Planck constant. In a perfectly converged system, the transition state free energy for both interconversion rates between minima  $j$  and  $k$  should be the same, i.e.  $F_{jk} = F_{kj}$ . However, for a finite MD trajectory this equality is rarely achieved. To minimize the error we average over both rates,

$$F_{jk}^{\text{ts}} = \frac{F_{jk} + F_{kj}}{2}, \quad (7)$$

giving us an estimate for the transition state free energy  $F_{jk}^{\text{ts}}$  between minima  $j$  and  $k$ .

It is important to note that it is the rates that correspond to observable quantities, and the free energy barriers simply provide a convenient way to visualise the corresponding landscape. Equation (6) translates the rates into barriers on a log scale. Multiplying all the rates by a constant factor would preserve detailed balance and shift the connections in the disconnectivity graph uniformly, without affecting the organisation of the landscape. The relative free energies of the minima in the graph and the barriers between them reproduce the equilibrium distribution and the rates by construction. Hence we obtain insight into the organisation of the landscape in a representation that will faithfully reproduce the stationary distribution and the rates.

To assess the assumption of Markovian dynamics, we conducted a Chapman-Kolmogorov test on the kinetic network of the dimer system, following the standard PyEMMA protocol<sup>19</sup> with a discrete timestep of  $\tau = 20 \text{ ps}$ . The number of metastable states was set to five, representing the number of states on the pathway from the disordered state D to the global minimum B. Figure S5 illustrates the comparison between the estimated and predicted transition probabilities between metastable states for lagtimes up to 100 ps. The strong agreement between estimated and predicted probabilities supports the Markovian nature of the dynamics. Notably, the states involved in the pathway do not correspond to the metastable states identified by PCCA++ clustering<sup>20</sup>:

D : state 1

I<sub>1</sub> : state 3

I<sub>2</sub> : state 5

I<sub>3</sub> : state 5

B : state 5

However, this observation aligns with our findings, as states I<sub>2</sub>, I<sub>3</sub>,



and B are structurally similar, closely located within the same energy funnel, and exhibit relatively fast transitions, as confirmed by FPT analysis. Additionally, the highest transition probabilities are associated with the transitions between states  $1 \rightarrow 3$  and  $3 \rightarrow 5$ , reproducing the fastest pathway as determined by FPT analysis. The estimated and predicted transition probabilities as determined from a Chapman-Kolmogorov test for the kinetic network of the monomer system is presented in Fig. S6.

#### 1.4 First Passage Times

While the FES governs the structural and dynamical properties of a molecule, in experiment often the relaxation times associated with a given process are measured. Thus, studying the timescales associated with transitions between minima on the FES can bridge the gap between simulation and experiment, and also shed light on the processes associated with those timescales. Often quantified by the mean first passage time (MFPT), interconversion rates between minima offer insights into the average time taken for a system to transition between reactant and product states. It has been shown that studying the first passage time (FPT) distribution of a transition can reveal a interesting additional information.<sup>21</sup> Specifically, such analysis provides direct access to the organizational structure of the underlying energy landscape and facilitates the identification of distinct signatures linked to relaxation to different funnels in the FES.

For a given transition  $A \leftarrow B$  from reactant state A to product state B the first passage time probability distribution  $p(t)$  can be obtained by treating the product state as absorbing. Then, the master equation for the occupation probabilities  $P_\alpha(t)$  for the set of intermediate states I and reactant states I  $\cup$  B is

$$\begin{bmatrix} \dot{\mathbf{P}}_I(t) \\ \dot{\mathbf{P}}_B(t) \end{bmatrix} = \begin{bmatrix} \mathbf{K}_{II} - \mathbf{D}_I & \mathbf{K}_{IB} \\ \mathbf{K}_{BI} & \mathbf{K}_{BB} - \mathbf{D}_B \end{bmatrix} \begin{bmatrix} \mathbf{P}_I(t) \\ \mathbf{P}_B(t) \end{bmatrix} = \mathbf{M} \mathbf{P}_{I \cup B}(t), \quad (8)$$

where  $\mathbf{K}_{XY}$  is the rate matrix of transitions between connected states and  $\mathbf{D}_X$  a diagonal matrix containing the escape rates of each state in X, i.e.  $[\mathbf{D}_X]_{ii} = \sum_j \mathbf{K}_{ji}$ . Applying eigenvector decomposition to the formal solution of eq. 8 produces an analytic solution for the first passage time distribution

$$p(t) = \sum_l v_l e^{-v_l t} A_l, \quad (9)$$

here  $-v_l$  are the eigenvalues of  $\mathbf{M}$  and  $A_l$  are amplitudes, which depend on the eigenvectors of  $\mathbf{M}$ . The FPT can be represented by the probability distribution  $\mathcal{P}(y)$  for  $y = \log(t)$

$$\mathcal{P}(y) = \sum_l v_l e^{y - v_l \exp(y)} A_l, \quad (10)$$

which gives distinct peaks for different features and relaxation time scales in the FPT distribution.

## 2 Pathway Analysis

### 2.1 Extraction of Pathways

From the database of minima and transition states constituting the FES of the dimer, we extracted the shortest pathway connecting states D and state B using the PATHSAMPLE program.<sup>22</sup> This

pathway describes the fastest transition from a disordered structure (state D) to a  $\beta$ -hairpin structure (state B). The transition occurs through three intermediate states ( $I_1, I_2, I_3$ ), characterized by the establishment of a salt bridge between residues D23 and K28, followed by the formation of hydrophobic contacts and subsequent rearrangement of these contacts into a  $\beta$ -sheet. However, due to the dimensionality reduction from application of the DRID metric before clustering, the resulting minima represent an ensemble of configurations. While these structures are similar according to the DRID metric, this procedure leads to an ensemble of slightly different pathways. To identify potential pathways from this ensemble of candidates for further analysis, we have employed linear interpolation between the configurations to construct trajectories for the transition:

$$D \rightarrow I_1 \rightarrow I_2 \rightarrow I_3 \rightarrow B.$$

Given our focus on studying the pathway associated with the fastest transition rates, we only considered trajectories that do not involve chain crossing in the linear interpolation between states, which is an artefact. From this procedure, we inferred three pathways (Fig. S2), which exhibit slight differences in their final configuration but all begin with the formation of a salt bridge between D23 and K28, proceed with hydrophobic contact formation, and ultimately reorient into a  $\beta$ -sheet configuration.

The predominant pathway 1 is discussed in the main text. Here, we briefly discuss the other two pathways from D to B. Pathway 2 is different from pathway 1 primarily in two aspects. Firstly, the closing scissor motion from an extended to a closed structure is not fully accomplished via the transition to the first intermediate state  $I_1$ , but also involves the second intermediate state  $I_2$ . Nonetheless, it follows the same hierarchy of events, where first the salt bridge and then the hydrophobic contacts are formed. The second deviation from pathway 1 is the orientation of the C-terminus, which almost wraps around the rest of the chain. Examination of the full dimer structure of this member of the global minimum ensemble reveals a very stable complex, which is the predominant structure in one of the three MD trajectories. In contrast, pathway 3 exhibits  $\beta$ -hairpin formation only in the early stages of the transition, via the first intermediate state  $I_1$ . Subsequently, the hydrophobic contacts reorganise to form a short parallel  $\beta$ -sheet in  $I_1$ , resulting in a wide loop in the peptide central region. While this pathway differs significantly from pathways 1 and 2, the final structure constitutes only a very small fraction of the equilibrium population of the global minimum ( $< 1\%$ ).

### 2.2 First Passage Times

Besides calculating the FPT probability distribution for the overall transition, we also calculated the FPTs between the intermediate states corresponding to the fastest pathway between D and B. Fig. S4 shows the resulting FPT probability distributions for each intermediate transition. In each case, the FPT probability distribution indicates the presence of both fast and slow relaxation modes to the target state, leading to an overall slow MFPT. The two peaks correspond to two competing events: relaxation to the

global minimum and relaxation to the absorbing target state of that transition. The first, faster peak is associated with direct relaxation to the absorbing target state, while the slower peak corresponds to relaxation to the global minimum followed by eventual propagation to the absorbing target state. This is evident in the FPT probability distribution, where slow relaxation to the global minimum is generally more likely, except for the final transition  $I_3 \rightarrow B$ , where fast/direct relaxation to the global minimum is more probable compared to relaxation via intermediates. The last intermediate exhibits significant structural similarity to the global minimum, which is further underscored by the close proximity of  $I_3$  and  $B$  in the FES funnel. Consequently, the slow peak corresponding to relaxation via intermediates in the side funnel acts only as a minor kinetic trap. This latter process is not prominently featured in the FPT distribution for the  $D \rightarrow B$  transition, as it is not accessed significantly for the chosen starting point and thus has a negligible impact on the overall relaxation pathway.

The FPTs associated with the fastest mode according to the peak position of the fast relaxation mode are as follows:

$$\tau_{D \rightarrow I_1}^{\text{dim}} \sim 780 \text{ ps}$$

$$\tau_{I_1 \rightarrow I_2}^{\text{dim}} \sim 190 \text{ ps}$$

$$\tau_{I_2 \rightarrow I_3}^{\text{dim}} \sim 90 \text{ ps}$$

$$\tau_{I_3 \rightarrow B}^{\text{dim}} \sim 110 \text{ ps}$$

Since each FPT probability distribution was calculated separately, the respective target state was treated as an absorbing state. As a result, the individual estimates of the FPT associated with the fastest transitions do not add up to the estimate of the overall transition time between states  $D$  and  $B$  of  $\tau_{D \rightarrow B}^{\text{dim}} \sim 4 \text{ ns}$ , which include all possible recrossing event. The individual FPT distributions provide insights into the relative timescales of the intermediate transitions as well as the underlying structure of the FES.

### 2.3 Cooperative Folding

We demonstrated how the FES of the  $A\beta_{1-42}$  peptide is significantly altered in the presence of another  $A\beta_{1-42}$  peptide by employing the same DRID metric to calculate the FES. However, this approach only implicitly reveals the impact of their interaction. To more thoroughly analyse the cooperative effects driving the  $A\beta_{1-42}$  peptide towards a more folded,  $\beta$ -sheet-rich state rather than a disordered state, we studied the intra- and interpeptide contacts along the fastest pathway from state  $D$  to  $B$ .

Figure S3 displays the intra- and interpeptide contact maps for the ensemble of states belonging to  $D$  and  $B$  as well as for the three intermediates along that pathway. In the disordered state  $D$ , the two peptides show no interaction, as indicated by the absence of contacts between the two peptides. The intrapeptide contact map, on the other hand, reveals a slight tendency towards forming a hairpin structure, suggested by the faint contact trace perpendicular to the diagonal. This mirrors the behaviour of two  $A\beta_{1-42}$  monomers, as shown by the FES of the  $A\beta_{1-42}$  monomer in Fig. 1, and confirms that no stable  $\beta$ -hairpin is formed with-

out an interaction partner. In the first intermediate state  $I_1$ , the two  $A\beta_{1-42}$  peptides begin to form a complex, with one monomer contacting the hydrophobic C-terminal region of the other. In parallel, contacts begin to form within the peptide, in particular a salt bridge forms between residues  $D23$  and  $K28$ , and the hydrophobic core region around  $F19$  interacts with the hydrophobic C-terminal region around  $L34$ . At this point, the orientation of the interacting peptide segments within the peptide is mainly parallel and not antiparallel, as required for a hairpin structure. In the second intermediate state  $I_2$ , the interpeptide contact map reveals strong interactions between the hydrophobic regions of the two  $A\beta_{1-42}$  peptides, while the intramolecular  $D23$ - $K28$  salt bridge is broken to allow the hydrophobic intrapeptide contacts to reorganise. The binding of the hydrophobic regions between the peptides can be considered a hydrophobic platform essential for forming the proper intrapeptide contacts necessary for the  $\beta$ -hairpin structure. In the third intermediate state  $I_3$ , the  $D23$ - $K28$  salt bridge reforms, and the  $\beta$ -hairpin structure becomes established, as indicated by a strong contact trace perpendicular to the diagonal. Additionally, the dimer shows a high propensity for forming antiparallel contacts between the hydrophobic core regions or hydrophobic C-termini of the two peptides. Finally, the transition to the global minimum state  $B$  is characterised by further stabilisation of the  $\beta$ -hairpin and a stronger tendency for antiparallel alignment in the intermolecular hydrophobic contacts.

In summary, this juxtaposition of intra- and interpeptide interactions during the  $D \rightarrow B$  transition reveals that their formation occurs in a cooperative manner. The  $A\beta_{1-42}$  peptide gains the ability to fold into a stable  $\beta$ -hairpin by binding to the hydrophobic region offered by the other  $A\beta_{1-42}$  peptide. Therefore, the increase in hydrophobicity in the environment appears to be crucial for folding during self-assembly.

### Notes and references

- 1 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1**, 19–25.
- 2 J. Huang, S. Rauscher, G. Nawrocki, R. Ting, M. Feig, B. de Groot, H. Grubmüller and A. MacKerell, *Nature Meth.*, 2017, **14**, 71–73.
- 3 A. Paul, S. Samantray, M. Anteghini, M. Khaled and B. Strodel, *Chem. Sci.*, 2021, **12**, 6652–6669.
- 4 S. Samantray, F. Yin, B. Kav and B. Strodel, *J. Chem. Inf. Model.*, 2021, **60**, 6462–6475.
- 5 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 6 M. Parrinello and A. Rahman, *Mol. Phys.*, 1981, **52**, 7182–7190.
- 7 S. Nosé, *Mol. Phys.*, 1984, **52**, 255–268.
- 8 W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
- 9 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
- 10 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993.
- 11 D. Krause and P. Thörnig, *JLSRF*, 2018, **4**, A132.
- 12 T. Zhou and A. Caffisch, *J. Chem. Theory Comput.*, 2012, **8**, 2930–2937.
- 13 D. Chakraborty, J. E. Straub and D. Thirumalai, *Sci. Adv.*, 2023, **9**, eadd6921.
- 14 M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé, *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.
- 15 G. Reddy, J. E. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. U.S.A.*, 2009, **106**, 11948–11953.
- 16 M. Schäffler, S. Samantray and B. Strodel, *Int. J. Mol. Sci.*, 2023, **24**, 11238.
- 17 J. Jeon, W.-M. Yau and R. Tycko, *Nature Commun.*, 2023, **14**, 2964.
- 18 S. V. Krivov and M. Karplus, *J. Chem. Phys.*, 2002, **117**, 10894–10903.
- 19 M. K. Scherer, B. Trendelkamp-Schroer, F. Paul, G. Pérez-Hernández, M. Hoffmann, N. Plattner, C. Wehmeyer, J.-H. Prinz and F. Noé, *J. Chem. Theory Comput.*, 2015, **11**, 5525–5542.
- 20 S. Röblitz and M. Weber, *Adv. Data Anal. Classif.*, 2013, **7**, 147–179.
- 21 D. J. Wales, *J. Phys. Chem. Lett.*, 2022, **13**, 6349–6358.
- 22 PATHSAMPLE: A program for generating connected stationary point databases and extracting global kinetics, <http://www-wales.ch.cam.ac.uk/software.html>.

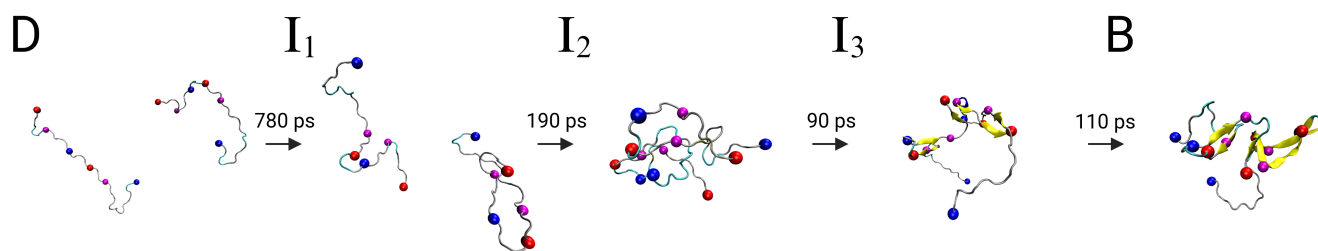


Fig. S1 Fastest pathway from intrinsically disordered state (D) to the global minimum (B) in the dimer free energy surface. Different to Fig. 3, here both peptides in each of the states are shown. The spheres represent the centroids used in the DRID metric: blue spheres for positively charged N-terminus and K28, red spheres for the negatively charged C-terminus and D23, and magenta spheres for the hydrophobic F19 and L34.

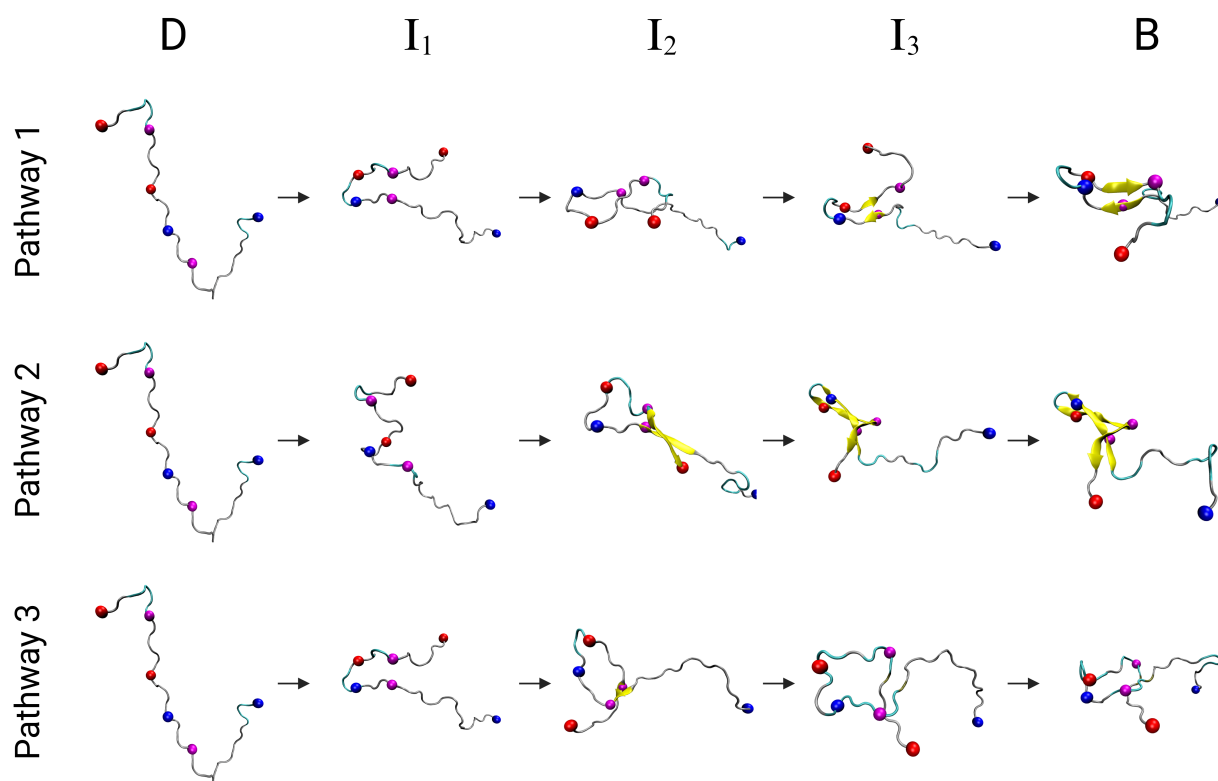


Fig. S2 The three most dominant pathways from the intrinsically disordered state (D) to the global minimum (B) in the free energy surface of the dimer. Only the peptide of the dimer for which the D  $\rightarrow$  B transition was analysed is shown.

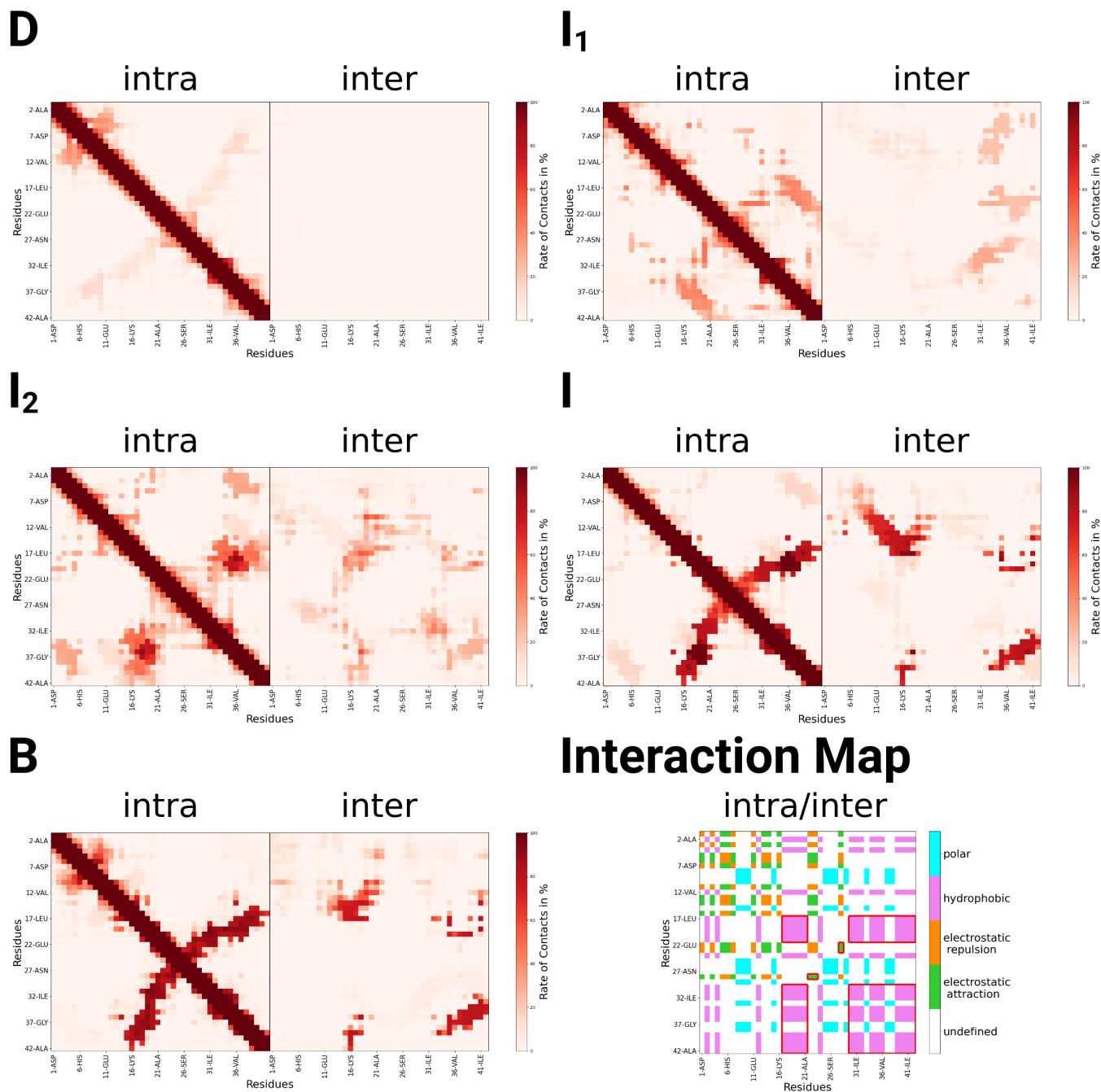


Fig. S3 Intra- and interpeptide contact maps along the D  $\rightarrow$  B transition for A $\beta$ <sub>1–42</sub> dimerisation. The contact maps were calculated for the ensemble of conformations belonging to the intrinsically disordered state (D), the global minimum (B), and the intermediate states of the fastest D  $\rightarrow$  B pathway (I<sub>1</sub>, I<sub>2</sub>, I<sub>3</sub>). Two residues were considered to be in contact if any pair of atoms of the two residues were within a distance of 6 Å. The lower right corner displays the physicochemical nature of all possible residue–residue interactions that could occur within an A $\beta$ <sub>1–42</sub> peptide or between A $\beta$ <sub>1–42</sub> peptides: electrostatic repulsion (orange) or attraction (green), polar (blue) and hydrophobic interactions (magenta). For direct comparison with the other panels of this figure, the regions of relevant hydrophobic interactions as well as the D23–K28 salt bridge are highlighted by red boxes.

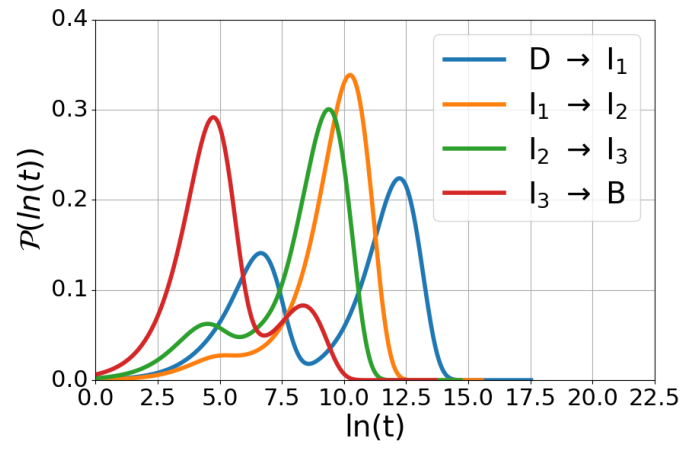


Fig. S4 Probability distribution of the first passage time  $t$ ,  $\mathcal{P}(\ln t)$ , for each intermediate transition of the pathway  $D \rightarrow I_1 \rightarrow I_2 \rightarrow I_3 \rightarrow B$ . The FPTs are calculated for each transition separately, treating the target state as an absorbing state.

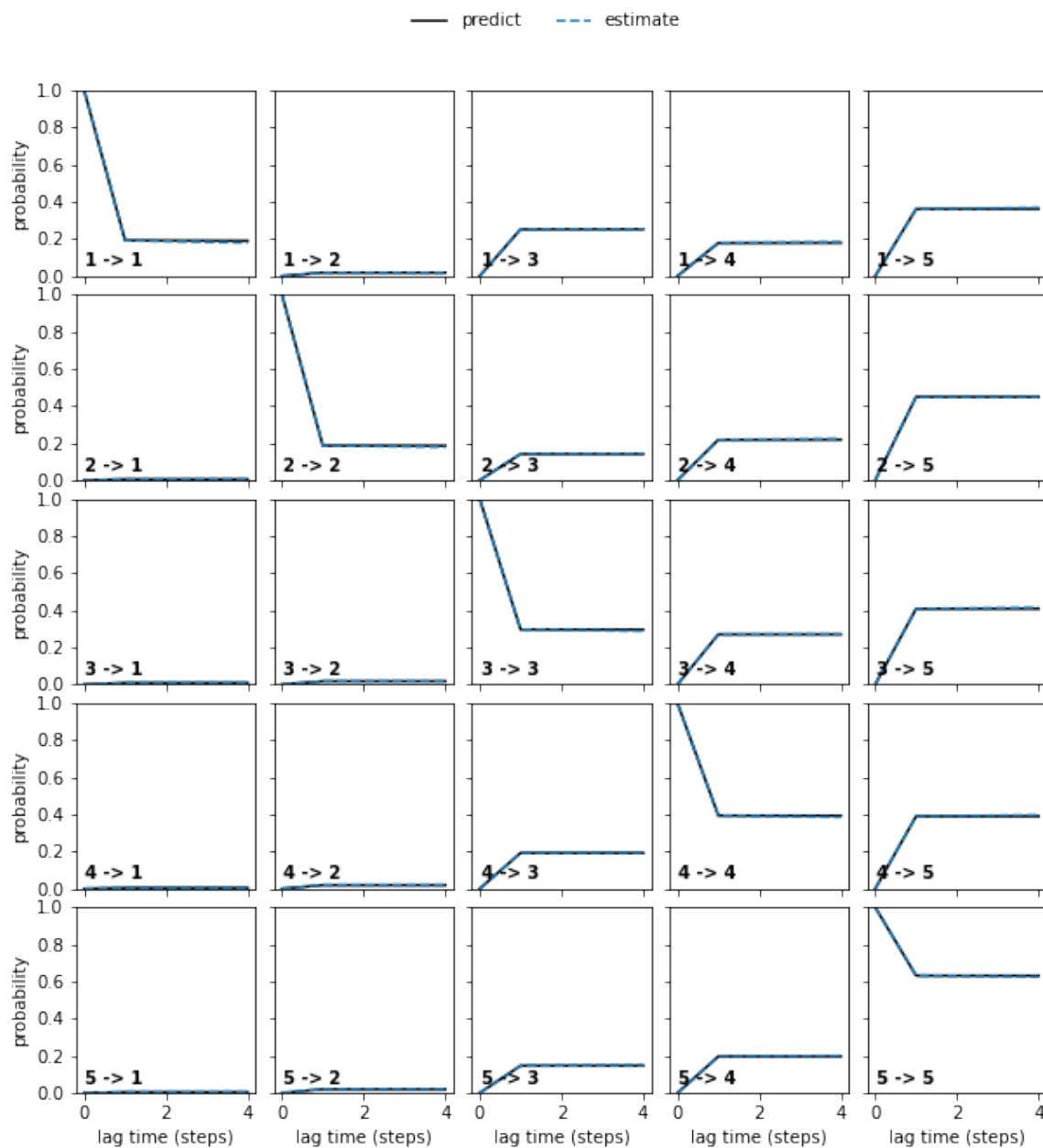


Fig. S5 Chapman-Kolmogorov test for the kinetic network of the dimer system, following a standard PyEMMA protocol with a discrete timestep of  $\tau=20$  ps. The number of metastable states as determined by PCCA++ was set to 5.

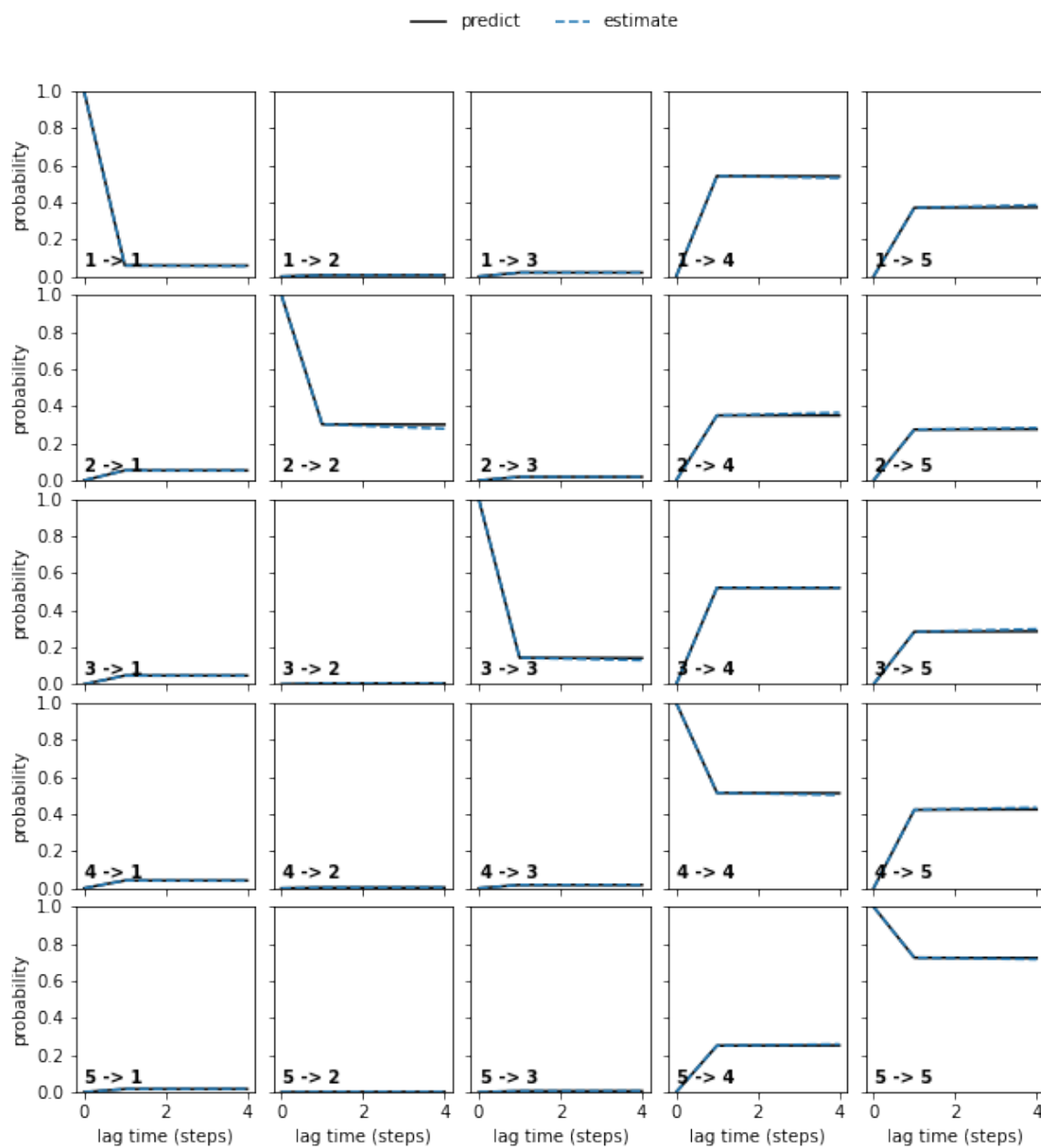


Fig. S6 Chapman-Kolmogorov test for the kinetic network of the monomer system, following a standard PyEMMA protocol with a discrete timestep of  $\tau=20$  ps. The number of metastable states as determined by PCCA++ was set to 5

# Photocontrolled Reversible Amyloid Fibril Formation of Parathyroid Hormone-Derived Peptides

André Paschold,<sup>○</sup> Moritz Schäffler,<sup>○</sup> Xincheng Miao, Luis Gardon, Stephanie Krüger, Henrike Heise, Merle I. S. Röhr, Maria Ott, Birgit Strodel,\* and Wolfgang H. Binder\*



Cite This: *Bioconjugate Chem.* 2024, 35, 981–995



Read Online

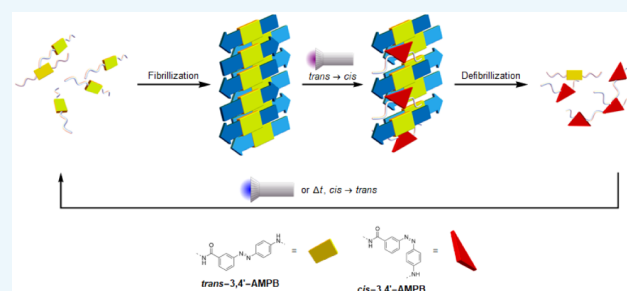
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

**ABSTRACT:** Peptide fibrillization is crucial in biological processes such as amyloid-related diseases and hormone storage, involving complex transitions between folded, unfolded, and aggregated states. We here employ light to induce reversible transitions between aggregated and nonaggregated states of a peptide, linked to the parathyroid hormone (PTH). The artificial light-switch 3-[[[(4-aminomethyl)phenyl]diazenyl]benzoic acid (AMPB) is embedded into a segment of PTH, the peptide PTH<sub>25–37</sub>, to control aggregation, revealing position-dependent effects. Through *in silico* design, synthesis, and experimental validation of 11 novel PTH<sub>25–37</sub>-derived peptides, we predict and confirm the amyloid-forming capabilities of the AMPB-containing peptides. Quantum-chemical studies shed light on the photoswitching mechanism. Solid-state NMR studies suggest that  $\beta$ -strands are aligned parallel in fibrils of PTH<sub>25–37</sub>, while in one of the AMPB-containing peptides,  $\beta$ -strands are antiparallel. Simulations further highlight the significance of  $\pi$ – $\pi$  interactions in the latter. This multifaceted approach enabled the identification of a peptide that can undergo repeated phototriggered transitions between fibrillated and defibrillated states, as demonstrated by different spectroscopic techniques. With this strategy, we unlock the potential to manipulate PTH to reversibly switch between active and inactive aggregated states, representing the first observation of a photostimulus-responsive hormone.



## INTRODUCTION

Modulating a protein's secondary structure stands as a pivotal strategy to define and harness its functionality.<sup>1</sup> While numerous protein structures are identified, predicted, and engineered, the concept of inducing conformational changes by external triggers to alter their biological activities remains rare. Temperature,<sup>2</sup> pH,<sup>3</sup> polarity,<sup>4</sup> or light offer avenues for such a dynamic control, in particular when applied to sensitive functional groups inserted inside the protein. Among these, light emerges as a particularly advantageous stimulus, providing precise temporal control across vast time scales, noninvasiveness, and compatibility with intricate matrices like living tissues.<sup>5</sup> Leveraging light-induced conformational changes has demonstrated success in various proteins, from transporter-proteins like rhodopsins to enzymes, showcasing potential applications in photopharmacology and diverse enzymatic processes.<sup>5,6</sup> Moreover, external photoswitches have successfully enabled photomodulation in a diverse array of enzymes,<sup>7</sup> modifying binding affinities,<sup>8</sup> facilitating peptide purification through photoaffinity,<sup>9</sup> and controlling secondary structure alterations.<sup>10</sup> Beyond the structure of individual peptides or proteins, there is a strong interplay between secondary structure and aggregation of peptides. This process is particularly pronounced in the context of amyloid

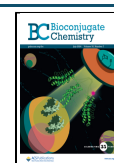
aggregation, where proteins form  $\beta$ -sheet-rich structures, leading to the formation of highly ordered fibrillar aggregates.<sup>11</sup> Such fibrils are a hallmark of several neurodegenerative diseases, including Alzheimer's and Parkinson's. Understanding and controlling amyloid aggregation can thus be crucial for developing therapies for diseases associated with these pathological protein assemblies. However, controlling peptide aggregation proves to be a significant challenge, given the complex processes involved. This challenge arises from intricate intrapeptide and interpeptide interactions, coupled with extensive conformational changes on a large scale, particularly when considering the regulation of primary and secondary nucleation preceding fibrillization. Therefore, the utilization of light to control amyloid assembly processes represents a groundbreaking advancement, enabling control over the bioavailability of hormones, such as the parathyroid hormone (PTH). PTH is reversibly stored in functional

Received: April 22, 2024

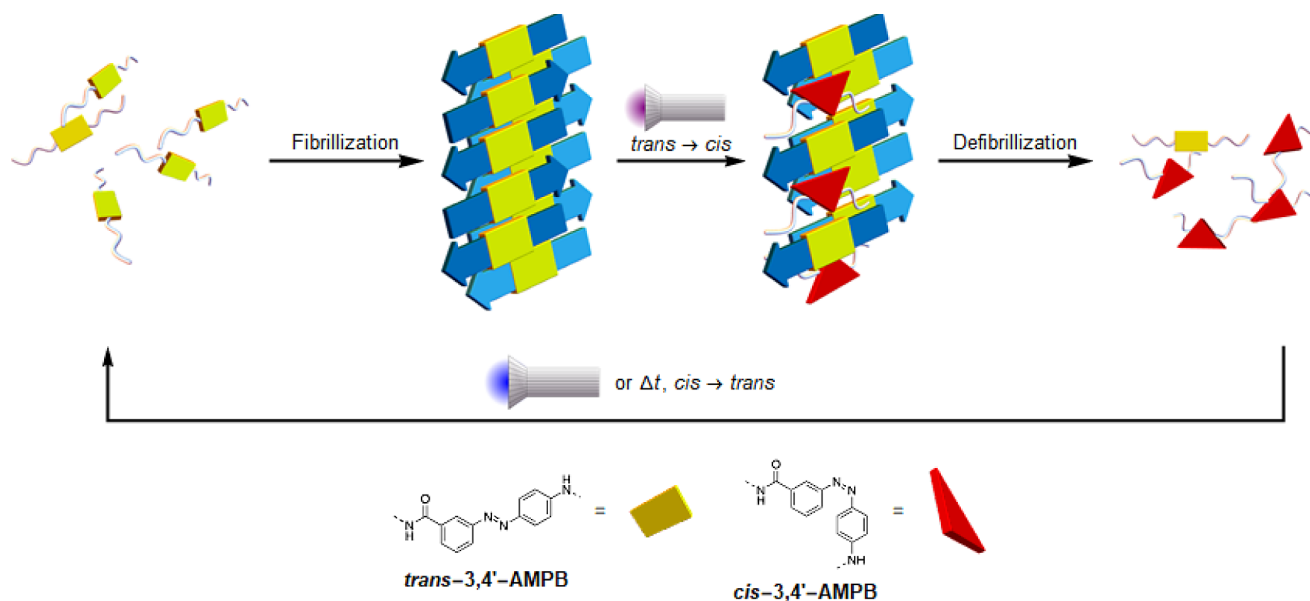
Revised: May 21, 2024

Accepted: May 24, 2024

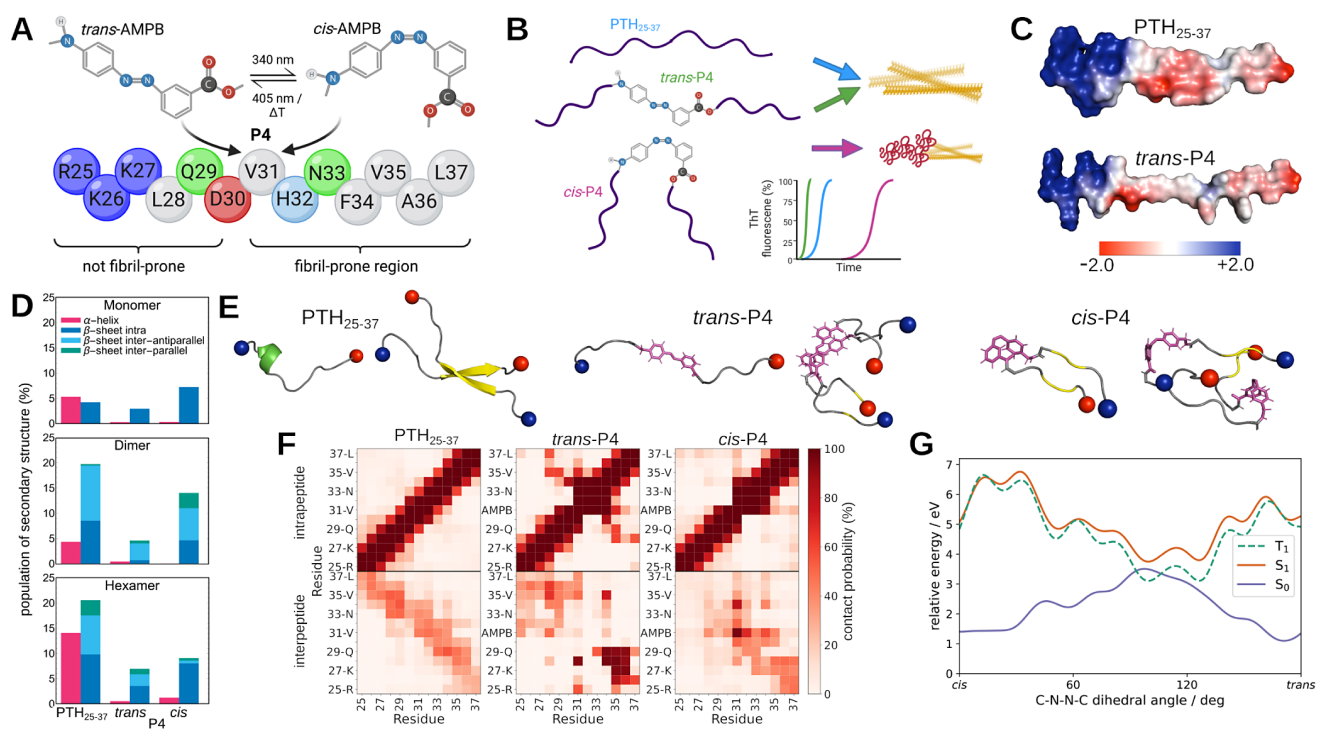
Published: June 12, 2024







**Figure 1.** Concept for a light-driven (de)-aggregation of the parathyroid-hormone (PTH). The artificial light-switch, 3-[(4-aminomethyl)-phenyl]diazenyl]benzoic acid (*cis/trans*-3,4'-AMPB) is embedded at various positions of a peptide-fragment of PTH, switching between the *cis*/*trans*-form, in this mode regulating the reversible assembly of the peptide into fibrils.



**Figure 2.** (A) Sequence of PTH<sub>25–37</sub>, where the fibril-forming region identified by bioinformatic analysis is labeled, and the chemical structure of AMPB, which replaces V31 in P4, is shown. The side chains of the amino acids, shown in blue and red, are positively and negatively charged at pH 7.4, respectively, with H32 shown in light blue as a borderline case, while green and gray indicate polar and hydrophobic amino acids, respectively. (B) PTH<sub>25–37</sub> and *trans*-P4 form fibrils, as shown in the ThT fluorescence cartoon below (blue and green line), while *cis*-P4 initially forms amorphous aggregates, which then transform into fibrils (red line in ThT cartoon). (C) Electrostatic potentials surface (values in kTe<sup>−1</sup> according to the color scale at the bottom) of PTH<sub>25–37</sub> and *trans*-P4. (D) Average simulated secondary structure population, divided into  $\alpha$ -helical, intrapeptide  $\beta$ -sheets, interpeptide parallel, and antiparallel  $\beta$ -sheets as indicated by the color code. Results are shown for simulations of monomers, dimers, and hexamers of PTH<sub>25–37</sub>, *trans*-P4, and *cis*-P4. (E) Representative snapshots of the monomer and dimer simulations, with the  $\alpha$ -helix shown in green, the  $\beta$ -sheets in yellow, the random coil in gray, the N- and C-termini as blue and red spheres, respectively, and AMPB in purple. (F) Contact matrix of dimer simulations for PTH<sub>25–37</sub>, *trans*-P4, and *cis*-P4. (G) Quantum chemical potential energy scan of the *trans*  $\rightarrow$  *cis* isomerization of the AMPB photoswitch in P4.

amyloid fibrils, as these fibrils, unlike, e.g., amyloid-A $\beta$  fibrils, can disintegrate again after aggregation. In this study, we pioneer the utilization of light to achieve a reversible transition between the aggregated and nonaggregated states of PTH (Figure 1), allowing the regeneration of PTH fibrils through precise light control, presenting a transformative advancement in the field.<sup>12</sup>

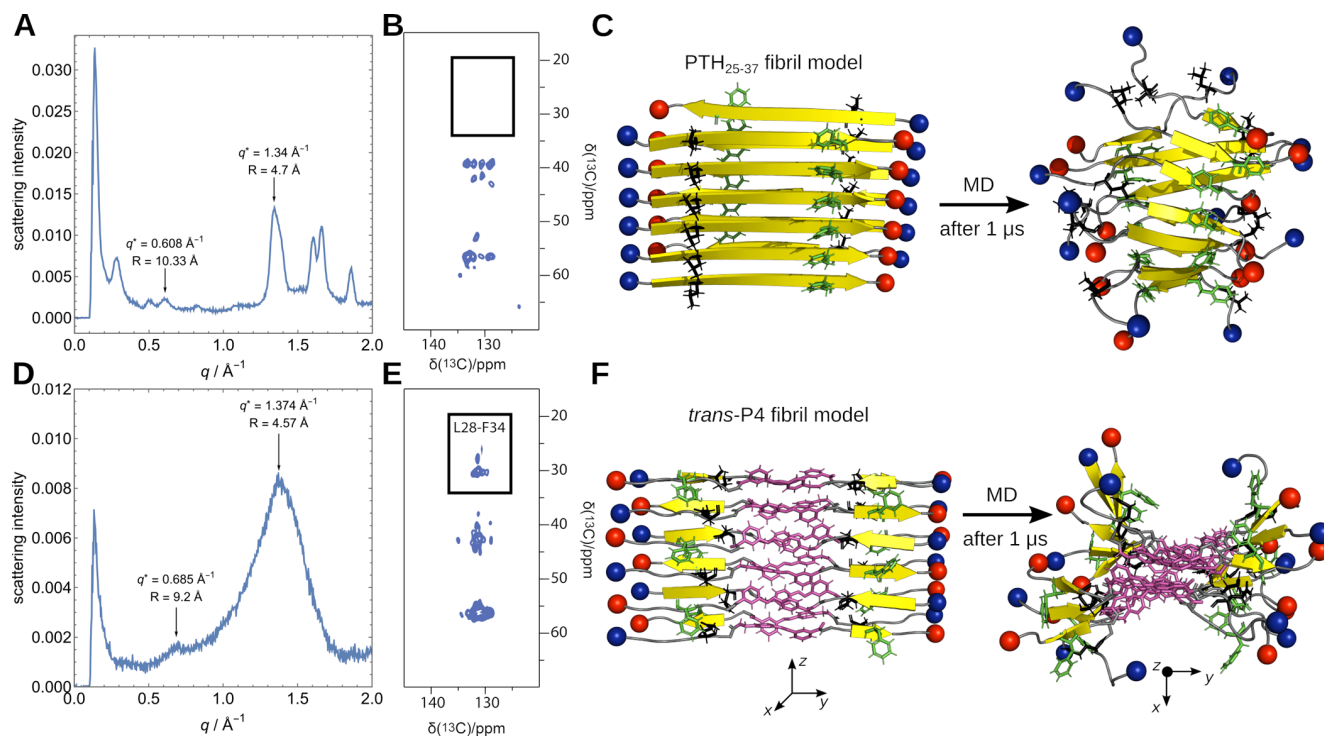
PTH, a reversibly fibrillating 84-amino acid hormone, is ubiquitously distributed in animals.<sup>13</sup> Responsible for regulating calcium and phosphate homeostasis,<sup>14</sup> mature PTH<sub>1–84</sub> is stored in functional amyloids before secretion,<sup>15</sup> wherein its N-terminal pro-sequence potentially prevents premature aggregation.<sup>16</sup> PTH-related diseases prompt the use of approved drugs, Natpara (PTH<sub>1–84</sub>) and Forteo (PTH<sub>1–34</sub>), addressing hormone imbalances.<sup>17</sup> While the physiological role of the N-terminal 34 amino acids of PTH<sub>1–84</sub> is well investigated, being crucial for the activation of G-protein coupled receptors of bone and kidney cells<sup>18</sup> and in the nervous system for calcium and phosphate homeostasis,<sup>19</sup> there remains a knowledge gap regarding the fibrillization process and the resulting fibrils. Current knowledge indicates that under physiological conditions, the thermodynamic stability of PTH<sub>1–84</sub> fibrils is low enough to allow dissociation upon dilution, with the fibril-forming sequence encompassing amino acids R25–L37.<sup>15</sup> Previous research explored the impact of the pro-sequence<sup>16</sup> and environmental factors on the fibrillization process<sup>20,21</sup> of PTH, however, with an only limited insight into the precise structural parameters controlling assembly and disassembly of the fibrils.<sup>22</sup> The aim of the current work is to gain deeper insights into the fibrillization of PTH<sub>25–37</sub>, coupled with the ability to reversibly control this process and understand the structural principles of the assembly process. We employ a synthetic approach for synthesizing PTH<sub>25–37</sub> peptides bearing the azobenzene photoswitch, 3-[(4-aminomethyl)phenyl]-diazanylbenzoic acid (AMPB), guided by bioinformatics to strategically place AMPB for optimal photocontrol, with biophysical techniques such as thioflavin T (ThT) fluorescence, CD spectroscopy, and transmission electron microscopy (TEM) to assess the peptide aggregation dynamics and aggregate morphology. This is further combined with molecular dynamics (MD) simulations to elucidate the impact of AMPB on the structure and dynamics of the designed peptides and their small oligomers (amounting to a total of 285  $\mu$ s simulation time), while wide-angle X-ray scattering (WAXS) and solid-state nuclear magnetic resonance (ssNMR) spectroscopy in conjunction with MD are employed to provide structure models for selected amyloid fibrils. Finally, quantum-chemical potential energy scans of both the ground ( $S_0$ ) and excited states ( $S_1$  and  $T_1$ ) reveal insights into the photo-switching mechanism of the AMPB group and its electronic interaction when integrated into PTH<sub>25–37</sub>, considering both the monomeric and dimeric peptide state. Providing a thorough understanding of the aggregation behavior of PTH<sub>25–37</sub> and its derivatives, we have engineered a peptide analogue with the unique capability of reversible light-induced switching of its aggregation state.

## RESULTS AND DISCUSSION

**Aggregation Characteristics of Unmodified PTH<sub>25–37</sub> and Engineered Peptide.** To understand the aggregation of PTH<sub>25–37</sub> and make informed design decisions, we began our analysis with a detailed examination of the original peptide PTH<sub>25–37</sub> and one engineered peptide (P4),<sup>22</sup> bearing the

photoswitch in the central part of the peptide (Figure 2A). Unmodified PTH<sub>25–37</sub> and *trans*-P4 form amyloid fibrils within 15 and 10 h, respectively, whereas *cis*-P4 initially forms amorphous aggregates and only forms amyloid fibrils after about 50 h (Figure 2B).<sup>22</sup> The increased rate of aggregation of *trans*-P4 compared to PTH<sub>25–37</sub> is assigned to the physicochemical properties of the amino acids and the phototrigger, AMPB, that make up the peptide (Figure 2A). PTH<sub>25–37</sub> has three positive charges (RKK) at the N-terminus, followed by a mixture of hydrophobic, polar and one negatively charged residue (D30) in the middle, while the C-terminal residues are mainly hydrophobic. This uneven distribution of physicochemical properties across the sequence is also reflected in the electrostatic potential surface, which shows a strongly positively charged N-terminus and a more hydrophobic C-terminal half with some negative charge accumulation beyond the first three residues (Figure 2C). The assumption therefore is that amyloid aggregation of PTH<sub>25–37</sub> is driven by the residues after the initial RKK sequence. To test this assumption, we used four aggregation predictors: PASTA,<sup>23</sup> AGGRESCAN,<sup>24</sup> AmyloGRAM,<sup>25</sup> and FoldAmyloid,<sup>26</sup> which show that the sequence <sup>32</sup>HNFVA<sup>37</sup>L is an aggregation hotspot and that the first five amino acids <sup>25</sup>RKKL<sup>29</sup>Q should not contribute to fibrillization. For <sup>30</sup>DV, a low tendency to aggregation was observed. In P4, one of the residues of the later sequence, V31, is replaced by AMPB. This increases the overall hydrophobicity of the peptide, while the azo group itself adds some positive charge to the electrostatic potential, which compensates for the predominant negative charge in the C-terminal part of the peptide. These two effects, therefore, explain the faster aggregation kinetics of *trans*-P4 compared to PTH<sub>25–37</sub>.

To understand the aggregation mechanisms and elucidate the structures that form during aggregation, namely first oligomers and finally fibrils, we performed MD simulations and ssNMR spectroscopy of both PTH<sub>25–37</sub> and P4, the latter simulated in the *trans*- and *cis*-states (see Table S1 for an overview of simulations performed). In this way, we can also develop a structural understanding of why the fibrillization of *cis*-P4 is slowed down and generally reduced. The monomers of either PTH<sub>25–37</sub>, *trans*-P4, and *cis*-P4 are mainly in a random coil state (Figure 2D). Nevertheless, small differences between the monomeric peptides can be observed. One of them is that PTH<sub>25–37</sub> shows a certain tendency to form an  $\alpha$ -helix in the N-terminal residues up to D30, which is lost when V31 is replaced by AMPB (Figure 2E). When the azo group is in the *cis*-state, this allows more intrapeptide contacts, including the formation of a  $\beta$ -hairpin, whereas in the *trans*-state, P4 is mostly in a fully elongated state. The dimer simulations showed an increase in  $\beta$ -sheets for all three peptides, but most for PTH<sub>25–37</sub>. This is a result of peptide aggregation, as there is a particular increase in  $\beta$ -sheets between the peptides (Figure 2D), which are mainly arranged antiparallel, as shown by the contact matrix (Figure 2F). In an antiparallel arrangement, the three positively charged N-terminal residues RKK can interact with the negative charge at the C-terminus of the neighboring peptide. Interestingly, the *trans*-P4 peptide, which we expected to aggregate the fastest based on the fibrillization data, forms fewer  $\beta$ -sheets at the oligomer level. Instead, the interpeptide interactions are dominated by contacts between the two AMPB groups, while the peptides are aligned antiparallel to each other. In contrast, although the *cis*-P4 peptide adopts  $\beta$ -sheet structures to some extent due to intrapeptide hairpins, it



**Figure 3.** (A, D) WAXS results for PTH<sub>25-37</sub> and the AMPB-modified PTH<sub>25-37</sub> peptide *trans*-P4. (B, E) Solid-state NMR (ssNMR) 2D  $^{13}\text{C}$ – $^{13}\text{C}$  proton driven spin diffusion spectra close to rotational resonance (mixing time 1s) of fibrils of the PTH<sub>25-37</sub> and *trans*-P4 peptide with uniformly  $^{13}\text{C}$ -labeled L28 and F34. (C, F) Fibril models of PTH<sub>25-37</sub> and *trans*-P4 constructed based on the ssNMR (left) and at the end of MD simulations (right), where  $\beta$ -sheets are shown in yellow, random coil in gray, the N- and C-termini as blue and red spheres, respectively, AMPB in purple, and the side chains of L28 and F34 are highlighted as black and green sticks, respectively. Note that in panel F, the structure on the right-hand side is rotated by 90° to better represent the fibrillar arrangement.

aggregates mainly randomly, as confirmed by the many interpeptide contacts, which is consistent with the amorphous aggregates observed *in vitro*. The distinct differences between peptide aggregation involving *cis*- and *trans*-AMPB allow fibril formation to be controlled at the molecular level by isomerization, providing a solid basis for photocontrol of amyloid formation.

In an effort to identify the nucleus of fibril formation, we also simulated the hexamer formation of PTH<sub>25-37</sub>, *trans*-P4, and *cis*-P4. However, this system size is still too small (or the simulation time too short) to observe the emergence of fibrillar structures. On the contrary, the hexamers of these peptides are less ordered than the dimers. This confirms the experimental observation that the propensity of PTH<sub>25-37</sub> to form amyloid is much lower than that of other peptides, such as that of A $\beta$ <sub>16-22</sub>. For the latter, we observe the formation of ordered hexamers when simulated under the same conditions as here,<sup>27</sup> while ThT experiments for this peptide show that fibrils are already present at the beginning of the measurements.<sup>28</sup> For both PTH<sub>25-37</sub> and P4, we even see a decrease in interpeptide  $\beta$ -sheets when we increase the system size to the hexamer, which can be explained by the increase in dimensionality of the conformational space, which allows for more interpeptide interactions and makes it less likely to see ordered aggregates on short time scales. Another interesting observation is that the amount of helix formed increased in the hexamer system of the PTH<sub>25-37</sub> compared to its monomer and dimer. This again reflects the helical propensity of this peptide, which can be stabilized by interpeptide interactions, a common observation

in intrinsically disordered peptides that can fold after binding to interaction partners.<sup>29</sup>

We further investigated the electronic interactions and the photoswitching mechanism of AMPB integrated into PTH<sub>25-37</sub>, both in the monomer and dimer states. To this end, we generated 34 switching trajectories using MD simulations to model the *trans*  $\rightarrow$  *cis* isomerization of the AMPB photoswitch along the CNNC dihedral angle. The resulting ensemble of pathways was individually analyzed using ONIOM-based QM/QM2<sup>30</sup> (NEVPT2<sup>31</sup>/xTB<sup>32</sup>) calculations, yielding potential energy scans for the ground and first excited triplet and singlet states (Figures 2G and S1). While the calculations clearly indicate that photoswitching is feasible, we also found that several scans exhibit structural barriers, that may impact the fluorescence wavelength and lifetime compared to the pure photoswitch (Figure S2).<sup>33</sup> Further elucidation through principal component analysis of the distance matrices between peptide residues (Figures S3 and S4) revealed critical structural motifs potentially responsible for the observed  $S_1$  barrier in *trans*  $\rightarrow$  *cis* isomerization paths. These motifs particularly involve configurations of R25 interacting with A36/L37 and H32/Q33, in both monomeric and dimeric forms (Figure S5).

**Structures of PTH<sub>25-37</sub> and P4 Fibrils.** The final state of aggregation, the amyloid fibrils, was characterized by wide-angle X-ray scattering (WAXS) and ssNMR, and the resulting data were used to generate structural models whose stability was evaluated in MD simulations. The WAXS measurement was performed with preformed fibrils of either PTH<sub>25-37</sub> or *trans*-P4. Due to the isotropic orientations of the fibrils and the



**Table 1. Designed PTH<sub>25–37</sub> Peptides Containing the AMPB Photoswitch (Azo) Inserted between Two Amino Acids (i) or an Amino Acid Exchanged by AMPB (e)<sup>a</sup>**

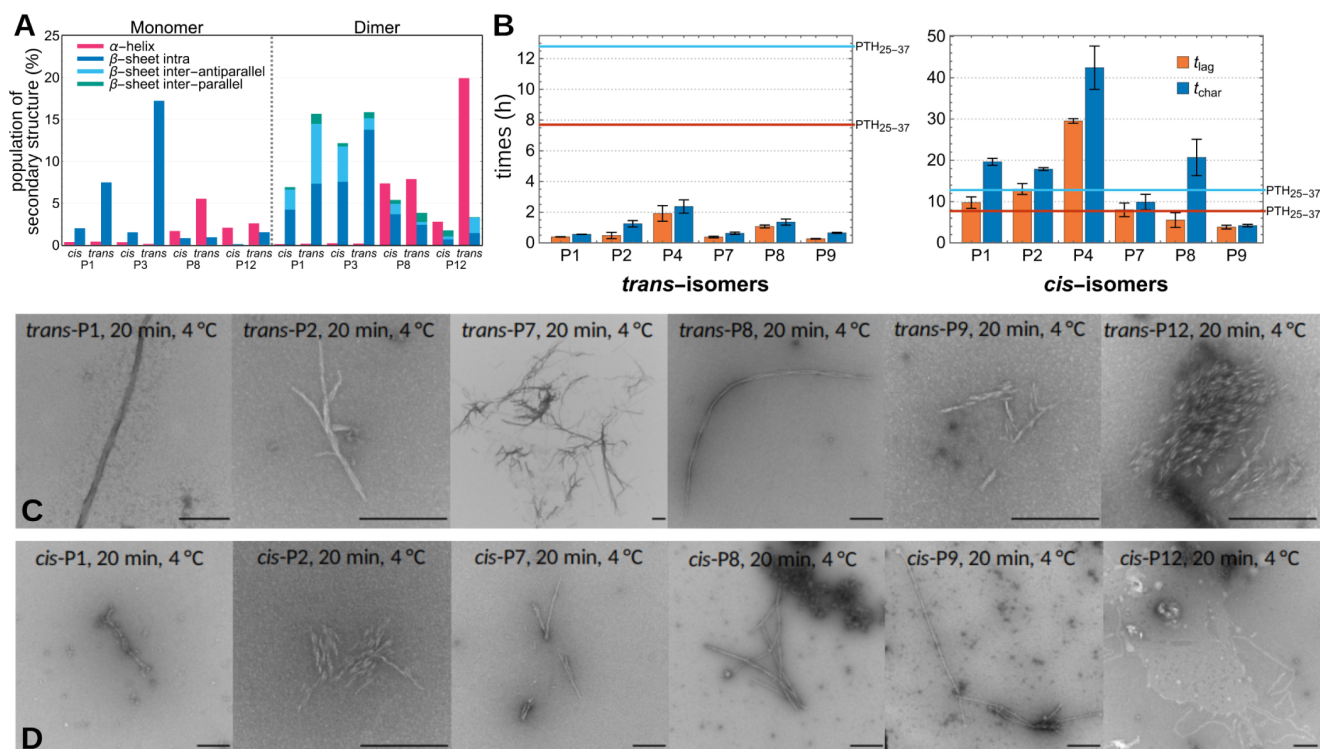
	peptide	primary sequence	modification	<i>t</i> <sub>1/2</sub> (cis) [h]
	PTH <sub>25–37</sub>	<sup>25</sup> RKKLQ <sup>30</sup> DVHNF <sup>35</sup> VAL	-	-
central	P1 (i)	<sup>25</sup> RKKLQ <sup>30</sup> D-Azo-VHNF <sup>35</sup> VAL	D30-Azo-V31	90
	P2 (i)	<sup>25</sup> RKKLQ <sup>30</sup> DV-Azo-HNF <sup>35</sup> VAL	V31-Azo-H32	90
	P3 (e)	<sup>25</sup> RKKLQ-Azo-VHNF <sup>35</sup> VAL	D30 → Azo	-
	P4 (e)	<sup>25</sup> RKKLQ <sup>30</sup> D-Azo-HNF <sup>35</sup> VAL	V31 → Azo	97
	P5 (e)	<sup>25</sup> RKKLQ-Azo-HNF <sup>35</sup> VAL	D30,V31 → Azo	-
	P6 (i)	<sup>25</sup> RKKLQ-Azo- <sup>30</sup> DVHNF <sup>35</sup> VAL	Q29-Azo-D30	-
N-terminal	P7 (i)	<sup>25</sup> RKKL-Azo-Q <sup>30</sup> DVHNF <sup>35</sup> VAL	L28-Azo-Q29	72
	P8 (i)	<sup>25</sup> RK-Azo-KLQ <sup>30</sup> DVHNF <sup>35</sup> VAL	K26-Azo-K27	89
	P9 (e)	<sup>25</sup> R-Azo-KLQ <sup>30</sup> DVHNF <sup>35</sup> VAL	K26 → Azo	86
C-terminal	P10 (i)	<sup>25</sup> RKKLQ <sup>30</sup> DVHN-Azo-F <sup>35</sup> VAL	N33-Azo-F34	-
	P11 (i)	<sup>25</sup> RKKLQ <sup>30</sup> DVHNF <sup>35</sup> V-Azo-AL	V35-Azo-A36	-
	P12 (e)	<sup>25</sup> RKKLQ <sup>30</sup> DVHNF-Azo-AL	V35 → Azo	63

<sup>a</sup>The corresponding peptides were prepared by on-resin synthesis (Merrifield-synthesis) using Fmoc-based building blocks. The synthesis of the AMPB photoswitch follows previously published methods.<sup>22</sup> The half-life time *t*<sub>1/2</sub> (hours) of the cis-form of the peptides was investigated via UV/vis spectroscopy (see the experimental section for further experimental details).

resulting isotropic scattering pattern (see Figure S6), the scattering intensities were angular averaged and are displayed in Figure 3 A, D. For PTH<sub>25–37</sub>, we observed a diffraction pattern typical for  $\beta$ -sheet-containing amyloid fibrils.<sup>34</sup> The reflection at 4.7 Å indicates a structural repeat corresponding to the distance between two  $\beta$ -strands within a sheet, whereas the reflection at 10.3 Å corresponds to the distance between two  $\beta$ -sheets in a fibril. A reflection at 9.4 Å, which would correspond to the repeating unit of two antiparallel  $\beta$ -strands within a  $\beta$ -sheet (i.e.,  $2 \times 4.7$  Å), is not observed, suggesting a parallel alignment of the strands within the  $\beta$ -sheets, which allows the hydrophobic, aggregation-prone residues on the C-terminal side of the peptide to lie adjacent to each other. To test this conclusion, we performed ssNMR measurements of synthesized PTH<sub>25–37</sub> with uniformly <sup>13</sup>C-labeled L28 and F34. In the 2D <sup>13</sup>C–<sup>13</sup>C spin diffusion measurements with longitudinal mixing times of 500 ms to 1 s, we did not observe any cross-peaks between these residues (Figures 3B, S7). This finding is indicative of a distance of >6 Å between these residues<sup>35</sup> and thus supports parallel  $\beta$ -sheets in PTH<sub>25–37</sub> fibrils. To answer the question of how two sheets of parallel stacked peptides could be arranged in the fibril,<sup>36</sup> we constructed four possible fibril models (with 6 peptides per sheet) consistent with the NMR distance data and tested their stability in MD simulations. Only one of these arrangements proved to be stable, even after 1  $\mu$ s MD. In this model, the  $\beta$ -sheets, which consist of parallel and in-register strands, are oriented antiparallel and their R25 side chains point inward (Figure 3C). This also agrees with our findings from the oligomer simulations, which revealed a preferred antiparallel arrangement between the PTH<sub>25–37</sub> peptides, as this allows the positive charges at the N-terminus to interact with the negatively charged C-terminus. In the fibril, this is realized via intersheet interactions, while within the sheets, the hydrophobic residues are adjacent to each other for optimal  $\beta$ -sheet stability. The fibril model at the end of the simulation confirms that the  $\beta$ -sheets are stable. The  $\beta$ -conformation is partially lost only at the terminal residues, which is due to the electrostatic repulsion between the three positive charges on the RKK residues, which also cause twisting of the fibril. The simulation-averaged distance between  $\beta$ -strands is 4.8 Å, and

between two sheets, it is 10.3 Å, in agreement with the WAXS data.

The WAXS signals for *trans*-P4 (Figure 3D) are much broader than for PTH<sub>25–37</sub>, indicating less structured fibrils for P4. Nevertheless, the peak for strand spacing at 4.6 Å dominates the signal, clearly indicating  $\beta$ -sheet formation. The next peak is at 9.2 Å, which corresponds to twice the interstrand distance. This indicates an antiparallel arrangement of the *trans*-P4 peptides within a sheet, as here the antiparallel unit consisting of two peptides is another repeat unit leading to a scattering signal. A signal for the intersheet distance is not visible, but could be hidden under the 9.2 Å signal, considering that the signal at 10.3 Å for the PTH<sub>25–37</sub> fibrils is also not strong. The conclusion that the *trans*-P4 peptides are antiparallel in the fibrils is supported by the ssNMR measurements showing an L28-F34 cross-peak in 2D <sup>13</sup>C–<sup>13</sup>C spin diffusion spectra, which is indicative of an inter-residual distance <6 Å (Figures 3E and S8). Based on this information and in addition to the results of our dimer and hexamer simulations of *trans*-P4, which revealed a strong interaction between the AMPB groups, we created several possible fibril structures and tested their stability in MD simulations. The structure that met all our experimental specifications and was also stable during the 1  $\mu$ s simulation is the one shown in Figure 3F. The structure at the end of the simulation shows a larger deviation from the idealized fibril model and with shorter  $\beta$ -sheets than observed for PTH<sub>25–37</sub>, which explains the broader WAXS signal. Stabilizing aspects are that the oppositely charged N- and C-terminal ends are directly adjacent within and between the  $\beta$ -sheets and that the central AMPB groups are aligned parallel to each other for  $\pi$ – $\pi$  stacking interactions. The antiparallel  $\beta$ -sheets themselves are less stable, however, as they are formed between the hydrophobic C-terminal half and the nonamyloid-prone N-terminal half on both sides of the AMPB group. Moreover, the AMPB group in the center of the peptide breaks the  $\beta$ -sheet structure, which explains that the P4 fibrils are shorter than those of PTH<sub>25–37</sub>. In addition, the  $\pi$ – $\pi$  stacking and  $\beta$ -sheet stacking have opposite spacing requirements with  $\approx 3.8$  Å and  $\approx 10$  Å, respectively, which can be clearly seen in the MD



**Figure 4.** (A) Simulated average secondary structure population, divided into  $\alpha$ -helical, intrapeptide  $\beta$ -sheets, and interpeptide parallel and antiparallel  $\beta$ -sheets as indicated by the color code. Results are shown for simulations of monomers and dimers of the *trans*- and *cis*-forms of P1, P3, P8, and P12. (B) Lag time  $t_{\text{lag}}$  (orange) and characteristic time  $t_{\text{char}}$  (blue) of the *trans*-forms (left) and *cis*-forms (right) of the peptides P1, P2, P4, P7, P8, and P9. The corresponding values of  $\text{PTH}_{25-37}$  are indicated as orange ( $t_{\text{lag}}$ ) and blue line ( $t_{\text{char}}$ ). (C) TEM images of the *trans*-isomers of P1, P2, P7, P8, P9, and P12 after 20 min at 4 °C. Scale bar = 250 nm. (D) TEM images of the *cis*-isomers of P1, P2, P7, P8, P9, and P12 after 20 min at 4 °C. Scale bar = 250 nm. TEM data for the photoswitching of P4 are presented in Figure S23.

snapshot shown in Figure 3F and further explains the broadness of the corresponding WAXS signal.

**Peptide Design.** Based on all simulations and structural investigations made for  $\text{PTH}_{25-37}$  and the P4 variant, we designed 11 novel  $\text{PTH}_{25-37}$ -derived peptides, wherein the amyloid-forming capabilities of the AMPB group in view of fibrillization were probed by placing the photoswitch at positions in the center of the peptide (P1–P6), at the N-terminal (P7–P9) or at the C-terminal part (P10–P12), with an amino acid either being exchanged by AMPB or AMPB being inserted between two amino acids (Table 1).

Our design strategy is based on the assumption that moving the AMPB group to the N-terminal side will enhance fibrillization by introducing the hydrophobic AMPB into the polar and charged N-terminal region, whereas moving it to the C-terminal side should impair amyloid formation by disrupting the amyloid-prone peptide region. In addition, we expect that the difference between the *cis*- and *trans*-forms of AMPB will become less important as it is shifted toward the termini, since hairpinning should no longer be possible in the *cis*-form. To refine these predictions, we performed MD simulations of the P1, P3, P8, and P12 variants, considering both the *cis*- and *trans*-configurations of each peptide as monomer and dimer (Figure 4A; P1 was also simulated as hexamer, see Figure S9). The simulations of P1 revealed a similar tendency for  $\beta$ -sheet formation as seen for P4 (Figure 2D), which makes sense given that AMPB is at the same position in both peptides, but instead of replacing V31 as in P4 it is added between D30 and V31, thereby extending the hydrophobic stretch on the C-terminal side of the peptide (Figure S10), which might lead to

a faster aggregation kinetics in the experiments. In P3, the AMPB group is shifted by one position toward the N-terminus compared to P4, replacing D30. This increases the hydrophobicity of the peptide in support of aggregation, while the removal of the negative charge increases the overall positive peptide charge, which could discourage aggregation. The simulations revealed that the increase in hydrophobicity prevails, as random dimerization with mainly intrapeptide instead of interpeptide  $\beta$ -sheets dominated in the simulations of *trans*-P3. P8 was simulated as a representative peptide in which AMPB is significantly shifted toward the N-terminus and inserted between K26 and K27. The electrostatic potential surface (Figure S10) shows that the insertion of the hydrophobic AMPB into the positively charged N-terminal region indeed significantly increases its hydrophobicity, promoting fibrillization. However, the secondary structure preferences are somewhat shifted away from  $\beta$ -sheet toward the formation of  $\alpha$ -helices in both *cis*- and *trans*-configurations and both as monomers and dimers, which was not seen for the P3 and P4 variants and may counteract amyloid fibrillization. Finally, we examined the P12 mutant in which the AMPB photoswitch is introduced into the aggregation-prone C-terminal region in the form of an exchange of V35. As predicted, the simulations confirm a drastic decrease in  $\beta$ -sheet formation compared to  $\text{PTH}_{25-37}$  and the other simulated peptide designs. As with P8, where AMPB is placed at the N-terminal side, its C-terminal position in P12 also leads to the formation of  $\alpha$ -helices, which could also counteract amyloid aggregation.

Next, we investigated the influence of the AMPB site experimentally. The designed peptides displayed varying solubility in a buffered aqueous solution (50 mM Na<sub>2</sub>HPO<sub>4</sub>, pH 7.4), ranging from 370  $\mu$ M for P4 to 7  $\mu$ M for P10 (Table S2), thus limiting the assay-conditions to those wherein sufficient critical concentrations could be reached. The critical concentration ( $c_{cr}$ ) of monomers in a fibril-forming system is the minimum concentration required to form fibrils in the dynamic equilibrium between the fibrils and the monomeric peptide form, which can be converted to a standard free energy of the fibril elongation reaction ( $\Delta G^0$ ). Since the  $c_{cr}$  of PTH<sub>25–37</sub> is 42  $\mu$ M at the chosen conditions, we decided to exclude the peptides with lower solubility, also motivated by the observation that no fibrillization was observed in P3 and P5, due to their low solubility. The peptides with small  $c_{cr}$  are the ones in which the hydrophobicity in the otherwise rather polar and positively charged peptide region was increased (P3, P5, P6), or in which the already hydrophobic C-terminal region was made more hydrophobic by placing the AMPB there (P10, P11). The reduced solubility of P3 is consistent with our MD simulation prediction of random aggregation. For the remaining peptides (P1, P2, P7, P8, P9, and P12, in addition to P4), we determined the photophysical characteristics of the isomerization reaction, such as the half-life time  $t_{1/2}$  of the thermodynamically less favorable *cis*-isomer, which is influenced by the position of the azobenzene switch (Table 1). To ensure probing of the fibrillization kinetics, it is important that the stability of the *cis*-isomer is significantly larger than the lag time of the fibrillization process. As both process are competing we strived to minimize the rate of *cis*  $\rightarrow$  *trans* isomerization to primarily investigate the behavior of the *cis*-isomer in the relevant time range. These half-life times range from 63 to 97 h, with the longest time observed for P4, where the central V31 was replaced by AMPB, and the shortest time, when the azo switch was shifted toward the C-terminus, replacing V35. The increased half-life time can be explained by the energy barrier for the *cis*  $\rightarrow$  *trans* isomerization in the ground state due to interaction between the residues at either side of the AMPB group (Figure 2G), which is also reflected in the *cis:trans* ratio. After synthesis, this ratio is between 3:97 and 6:94 for the freshly prepared peptides, subsequently increasing during the photo-induced *trans*  $\rightarrow$  *cis* isomerization to the photostationary state (PSS), reaching values of 90:10 to 82:18. Isomerization back under dark conditions reduces this ratio to 19:81 and 24:76 in the PSS of *cis*  $\rightarrow$  *trans* isomerization (Table S3). To exclude a photobleaching effect of the AMPB photoswitch, we repetitively conducted alternating *trans*  $\rightarrow$  *cis* and *cis*  $\rightarrow$  *trans* isomerizations over 5 cycles (Figure S11). When monitoring the absorption at two wavelengths (absorption maxima of *trans*- and *cis*-isomer, respectively) we did not observe a decline in the overall absorption, except for the first cycle, which can be explained by the fact that the initial peptides directly after synthesis possess an increased *trans*-content compared to the subsequent photostationary state. To assess the toxicity of the peptides with the incorporated photoswitch, we conducted cytotoxicity assays with the N-terminally modified P8, the centrally modified P4, and the C-terminally modified P12 (Figure S12), and we did not observe any toxicity toward 3T3 and NHDF cells.

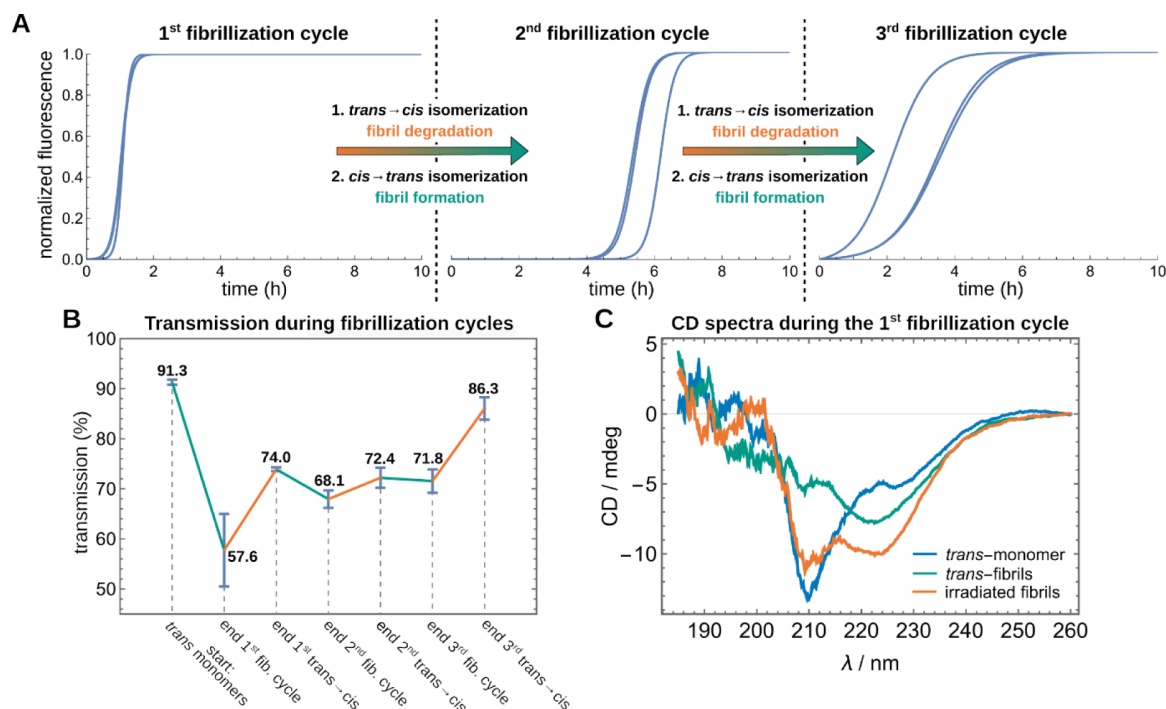
We then examined the fibrillization process of the peptides using the ThT fluorescence assay after an established protocol for PTH<sub>1–84</sub><sup>20</sup> (Figures S13–S20) and followed the formation of fibrils with TEM images after 20 min at 4 °C and after 1, 3,

24, and 96 h at 37 °C (Figures S21–27, original data for the photoswitching (TEM) for P4 are presented in Figure S23). The ThT fluorescence curves were fitted using eq. 1 and yielded the lag time  $t_{lag}$ , which corresponds to the onset of the fibril growth phase, and the characteristic time  $t_{char}$ , the point at which the fluorescence intensity reaches 50% of its maximum (Figures 4B, S13–20). Compared to PTH<sub>25–37</sub>, the fluorescence intensity was significantly lower, resulting in a poorer signal-to-noise ratio. We attribute this to fluorescence quenching by the azobenzene moiety, which is an already known property of this molecular building block.<sup>10 g,22</sup> Nevertheless, a sigmoidal fibrillization curve is clearly visible for the three replicate measurements of each peptide. Peptides P1, P2, P4, P7, P8 and P9 show typical fibrillization behavior with lag phase, growth phase, and stationary phase (Table S4). This observation is consistent with our prediction that the placement of AMPB in the middle or at the N-terminus of PTH<sub>25–37</sub> should not prevent the amyloid aggregation ability of the peptide. All *trans*-peptides exhibited, irrespective of the positioning of the photoswitch, a faster fibrillization process than PTH<sub>25–37</sub> with significantly shortened lag phases ( $t_{lag}$  of 8 h for PTH<sub>25–37</sub> vs 0–2 h for the peptide designs) and reduced  $t_{char}$  (Figure 4B). We assume that the increased hydrophobicity and the organizing effect of the AMPB group are the driving forces for the increased tendency of amyloid aggregation. TEM images of the fibrils were obtained for all peptides after only 20 min at 4 °C (Figure 4C). The *cis*-forms of peptides P1, P2, and P4, in which AMPB was inserted at the central position next to the aggregation hot-spot sequence H32-L37, showed slower fibrillization compared to PTH<sub>25–37</sub>, whereby for *cis*-P1 and *cis*-P2 the fibrillization occurred in a similar time range, while the fibrillization of *cis*-P4 was strongly delayed. In agreement with our predictions, the incorporation of the photoswitch in its *cis*-form at the N-terminal part had almost no influence on the fibrillization kinetics: *cis*-P7 and *cis*-P8 aggregated at a similar rate as PTH<sub>25–37</sub>, while the fibrillization of *cis*-P9 was even faster (due to the increased hydrophobicity of the peptide). The TEM images (Figures 4D, S21–27) confirm these observations. For the *cis*-isomers of P1 and P2, we mainly observed amorphous aggregates in the early images (after 20 min); the first objects, which could be clearly assigned to fibrillar aggregates, as with the *trans*-forms, were only detected after 24 h. For P4 (Figure S20) the difference was even more significant: in the *cis*-form, the first fibrillar aggregates were visible in the image only after 60 h. For the N-terminally modified peptides *cis*-P7, *cis*-P8, and *cis*-P9, however, fibrils were already visible after 20 min.

For the fibrillization behavior of the C-terminally modified P12, where V35 was replaced by AMPB, we found that it is entirely different. Unlike other peptides, ThT fluorescence curves for both *trans*- and *cis*-P12 show a linear increase after a lag phase of approximately 20 h, with no saturation observed even after 140 h, indicating reduced fibrillization. This also holds true for fibril morphologies, as both P12 isomers exhibit similar shorter fibrils in the TEM images, with *trans*-P12 forming fibril-like structures after 1 h and *cis*-P12 after 3 h.

In summary, our study revealed that the impact of the photoswitch varies depending on its placement: when positioned within the fibril-forming segment, such as in P12, fibril formation is significantly hindered, whereas outside this crucial region, the *trans*-isomers demonstrate faster fibril formation compared to PTH<sub>25–37</sub>. Notably, *cis*-isomers display a strongly hindered fibril formation or the formation of





**Figure 5.** Multiple fibrillization and degradation cycles of *trans*-P4 fibrils through photoisomerization. (A) Fitted ThT signal monitored by the fluorescence assay of three fibrillization cycles of *trans*-P4 comprising degradation of fibrils by photoinduced *trans* → *cis* isomerization and subsequent photoinduced *cis* → *trans* isomerization leading to renewed fibrillization. (B) Change in the transmission of a P4 sample over several switching cycles (cyan for fibrillization periods and orange during *trans* → *cis* isomerization). (C) CD spectra of monomeric *trans*-P4 (blue), *trans*-P4 fibrils (green), and the fibrils after *trans* → *cis* isomerization via irradiation with 340 nm for 5 h (orange).

amorphous aggregates, with the inhibition decreasing as the distance from the fibril-forming C-terminal region increases.

**Photocontrolled Multiple Switching of Fibril Formation.** By computationally analyzing the primary structure, modeling the peptide structures in their monomeric and aggregated forms, and studying the fibrillation process of PTH<sub>25–37</sub> and the designed peptides, we aimed to determine the most suitable position for the incorporation of the photoswitch into the peptide sequence to control fibril formation. Based on the results obtained, we applied several criteria to select the most promising candidate from the designed peptides (Table 1) to test whether amyloid formation can be reversibly switched on and off by light. After excluding peptides with insufficient solubility (P10 and P11) as well as the poorly aggregating P12, the focus was placed on the peptides P1, P2, and P4, which show accelerated fibrillization due to the central placement of the photoswitch in the *trans*-conformation, while the *cis*-form slows down fibrillization. Of these peptides, P4 exhibited both the best solubility and the largest difference in fibrillization for *trans*- and *cis*-P4. Therefore, we selected this peptide P4 to further test whether we could achieve reversible fibril formation (Figures 5, S19, S28). It should be noted that during the following sequence of experiments, the cuvettes were not changed; moreover, also no new peptides were added to the solution. Starting from the *trans*-isomer of P4, fibrils were formed in the first cycle at 37 °C in phosphate buffered aqueous solution and the aggregation kinetics followed via ThT fluorescence (Figure 5A, left). After the fibrillization passed over to the stationary phase, the solution was shaken another 20 h to affirm that the equilibrium has been established. Subsequently, the suspension was irradiated at a wavelength of 340 nm for 5 h under stirring,

after which UV/vis spectroscopy was used to verify that the *cis*-photostationary state of P4 was reached. Now, the solution was irradiated for 20 min with light of 405 nm wavelength to switch the *cis*-PSS to the *trans*-PSS to produce monomeric *trans*-peptides, which underwent another fibrillization cycle (Figure 5A, middle). The newly formed fibrils were again exposed to the photoisomerization treatment involving *trans* → *cis* isomerization to degrade the fibrils. The ThT-monitored fibrillization assay was repeated with a portion of the solution, and fibril formation was observed for a third time (Figure 5A, right). For the second and the third fibrillization cycle, a decreased starting concentration of the *trans*-isomer was observed (UV/vis spectra, Figure S29). As we could already exclude photobleaching, we suppose that aggregates adhere to the wall of the 96-well plates and are, therefore, not available for further fibrillization cycles. The fibrils formed during the fibrillation process can precipitate and thus disturb the measurement via scattering effects, causing noise in the signal after the first stationary phase. Because the *trans* → *cis* isomerization of the AMPB-switch is not reaching 100% due to the photostationary state, there certainly are fibrillar aggregates present in the solution, which, however, may interfere in the subsequent fibrillation processes and thus change the following fibrillation kinetics. Nonetheless, for each new fibrillation process monitored here, a lag phase, growth phase, and stationary phase were observed.

While monomers remain soluble, fibrils tend to precipitate, causing the solution to become turbid during the fibrillation assay and leading to a decrease in transmission. Observing the transmission across three switching cycles, it was evident that transmission decreases during the fibrillation process (Figure 5B, cyan lines). Degradation of the fibrils via *trans* → *cis*

isomerization, however, led to an increase in transmission, as released monomers dissolve back into the solution (Figure 5 B, orange lines). The incomplete recovery of the transmission during degradation results from the aforementioned observation that the *trans*  $\rightarrow$  *cis* isomerization only leads to a maximum *trans*:*cis* ratio of 18:82. Therefore, the degradation process of the fibrils is an only partial one, presumably as a full isomerization cannot be reached. The changes in secondary structure during fibrillization were probed by CD spectroscopy (Figures 5C, S28). In its monomeric form, *trans*-P4 displays a shoulder around 220–230 nm and a minimum around 210 nm. In the CD spectrum of the *trans*-P4 fibrils, a minimum is observable between 220 and 230 nm, while the minimum around 210 nm is nearly completely absent. Degradation of the fibrils by the *trans*  $\rightarrow$  *cis* isomerization leads to a CD spectrum where the minimum around 210 nm reappeared and the other minimum between 220 and 230 nm still exists. The minimum at 210 nm arises from the monomeric peptides, while the minimum between 220 and 230 nm is indicative for the fibril form. The observed CD spectrum is thus a superposition of both forms, the monomeric and the fibrillated form. From the CD spectra, we can, therefore, conclude that fibrillization from the *trans*-form and incomplete fibril degradation during *trans*  $\rightarrow$  *cis* isomerization can be observed.

Fibrillization of proteins are complex, often irreversible processes, which are characterized by a strong thermodynamic negative free energy, as e.g., in A $\beta$  fibrillization, and thus conventionally termed as “irreversible.”<sup>58</sup> This is often connected with a high kinetic barrier, usually preceding the fibrillization process, wherein nucleation is central to initiate that nucleation process. The PTH peptides studied here are “more reversibly” fibrillating peptides, wherein the thermodynamic stabilization of the amyloid state is less and, therefore, also the kinetic barriers.<sup>15,59</sup> Nonetheless, until now, it has not been possible to control the reversibility of this process, as accomplished here via the introduction of a photoswitch. However, in contrast to other reversibly (photo)-switchable systems, such as reported for small molecule assemblies,<sup>60</sup> adhesives,<sup>61</sup> or photoswitchable enzymes,<sup>62</sup> a full reversion of the aggregation of the current system is not reached, as expected, as the fibrillization and defibrillization processes depend on many factors, such as the kinetic barriers in either direction, the kind and number of nuclei present during the different switching cycles, or also the precipitation of fibrils.

## CONCLUSION

In this study, we have demonstrated precise control over reversible peptide fibrillization by strategically positioning a photoswitch within the central region of a fibril-forming peptide. Utilizing the artificial light switch 3-[(4-aminomethyl)phenyl]diazene]benzoic acid (AMPB) embedded in a peptide containing residues 25–37 of the parathyroid hormone (PTH), we investigated the impact of position on peptide aggregation. Through a comprehensive approach involving computational modeling, peptide synthesis, aggregation assays, and structural analyses, we elucidated key features governing the fibrillization of both unmodified and modified PTH<sub>25–37</sub> peptides. Notably, the *trans*-peptides with the modification positioned adjacent to the fibril-forming region in the center of the peptide (P1, P2, and P4) displayed enhanced fibrillization compared to unmodified PTH<sub>25–37</sub>, while the aggregation is slowed down for their *cis*-isomers. Peptides featuring the photoswitch in the nonamyloidogenic

N-terminal region behaved similarly to unmodified PTH<sub>25–37</sub> (P7, P8, and P9), whereas peptide P12, where the azobenzene unit replaced V35 in the amyloid-prone region, showed decreased fibrillization, largely unaffected by the photoswitch's isomer state. Our bioinformatics and simulation analysis uncovered that modification with AMPB typically boosts the peptides' hydrophobicity, thereby augmenting their tendency to aggregate. Furthermore, AMPB facilitates self-interaction among peptides through  $\pi$ – $\pi$  interactions, further enhancing their aggregation propensity. Structural investigations of P4 employing WAXS and ssNMR suggest that  $\beta$ -strands in amyloid fibrils of P4 are—in contrast to fibrils of unmodified PTH<sub>25–37</sub>—aligned antiparallel. Simulations suggest that AMPB might not fully adhere to the amyloid fold, owing to the distinct demands for interpeptide distances in  $\pi$ – $\pi$  interactions and fibril formation. This elucidates the diminished fibrillization observed when the photoswitch is positioned within the amyloid-forming segment of the peptide, as seen in P12. Importantly, our results showcase the potential of strategically placing the azo photoswitch, particularly exemplified by peptide P4, to control reversible amyloid aggregation. The phototriggered degradation of fibrils formed in the *trans*-state of AMPB enables repeated fibril formation, which in fact may allow to reversibly modulate the fibrillization of the PTH hormone, with the monomeric peptides released upon fibril degradation serving as the active form and the fibrillar structures acting as an inert peptide reservoir. Given the peptides' nontoxic nature (Figure S12), this light-triggered approach, therefore, presents a promising method for controlled drug delivery and release of such reversibly fibrillating peptides. We regard our system here as a switchable fibrillization system, where a phototrigger from the outside is able to induce fibrillization, as e.g., needed in many modern neuro-cellbiology systems.<sup>63</sup>

## EXPERIMENTAL SECTION

**Materials.** All technical solvents were distilled prior use; air- and moisture-sensitive reactions were carried out in flame-dried glassware under atmospheric pressure of nitrogen. 2-(6-Chloro-1-H-benzotriazole-1-yl)-1,1,3,3-tetramethylammonium hexafluorophosphate (HCTU), *N*-methyl-morpholine (NMM), *N,N*-dicyclohexylcarbodiimide (DIC), *N*-Hydroxy-benzotriazole (HOBt), trifluoroacetic acid, 4-amino-benzylamine, and oxone were purchased from Sigma-Aldrich. 9-Fluorenylmethyl-*N*-succinimidylcarbonate (Fmoc-OSu) was received from Fluorochem. 3-Aminobenzoic acid was purchased from Merck and was used without further purification.

**UV/Vis, CD, and Transmission Spectroscopy Measurements.** UV/vis-absorbance spectroscopy was measured on a JASCO V-660 absorbance spectrometer in a 1 cm quartz glass cuvette. For PTH<sub>25–37</sub>, the absorbance was measured at 205 nm with a molar extinction coefficient of 49,310 cm<sup>–1</sup> M<sup>–1</sup>; *trans*-azobenzene containing peptides were measured at 327 nm with a molar extinction coefficient of 13,000 cm<sup>–1</sup> M<sup>–1</sup>. CD spectroscopy was measured with a JASCO J-1500 CD Spectrometer in either a 1 mm. As buffered solution, a 50 mM aqueous Na<sub>2</sub>HPO<sub>4</sub> buffer solution was used with a pH value adjusted to 7.4. Transmission was measured with a Litesizer DLS 500 from Anton Paar using a 3 mm  $\times$  3 mm quartz glass cuvette. The irradiation wavelength was 658 nm. The equilibration time was 1 min, and the measurement time was 10 s. The temperature was maintained at 25  $^{\circ}$ C.



**Peptide Synthesis.** The 3,4'-AMPB photoswitch was synthesized in two steps according to our published procedure.<sup>22</sup> Solid-phase peptide synthesis was utilized on an automated peptide synthesizer MultiPep RS (Intavis AG, Koeln, Germany) using standard Fmoc-chemistry and preloaded resins. Standard coupling of all protected natural amino acids was performed as single couplings in dimethylformamid (DMF) using 5 equiv of amino acids, HCTU as coupling reagents, and 10 equiv of NMM as base for 1 h at room temperature. Special building groups, such as Fmoc-3,4'-AMPB, were coupled with 3 equiv using DIC and HOBT in DMF/ *N*-methyl-2-pyrrolidone (NMP) at room temperature and with gentle shaking in the dark overnight. The *N*-terminal Fmoc protecting group was removed by washing the resin with 20% piperidine for 20 min. The final side chain deprotection and cleavage from the resin employed a mixture of trifluoroacetic acid and water (90:10 Vol%) with gentle agitation for 2 h at room temperature. The crude peptides were purified to >95% purity using preparative RP-HPLC (Gilson, Limburg, Germany). For both analytical and preparative use, the mobile phase was a mixture of water (eluent A) and acetonitrile (eluent B), respectively, each containing 0.1% trifluoroacetic acid. Samples were eluted with a linear gradient from 5% B to 95% B in 15 min for analytical runs and in 90 min for preparative runs on a semipreparative PLRP-S column (Agilent Technologies, 300 × 25 mm, 8 μm). Finally, all peptides were characterized by analytical HPLC Dionex Ultimate 3000 (Thermo Scientific, Germany) using a PLRP-S column (Agilent Technologies, 150 × 4.6 mm, 3 μm) and MALDI-MS (Bruker Microflex LT, Bremen, Germany), which gave the expected  $[M + H]^+$  mass peaks. The full molecular characterizations are shown for all peptides in Figures S30–S47.

**Aggregation Kinetics.** The fibrillization process was investigated using a thioflavin T (ThT) monitored fluorescence assay following the established process for PTH<sub>1–84</sub>.<sup>20</sup> Therefore, the fluorescence intensity of ThT was measured. Lyophilized peptides were dissolved in buffered solution and kept on ice for the next steps. The sample solutions were centrifuged at 13 000 rpm for 10 s, and the concentration of the respective peptide was determined using UV/vis-absorbance spectroscopy. If required, the *cis*-isomer of the azobenzene containing peptides was produced as described below. The sample solutions were centrifuged at 10 000 rpm for 1 h at 4 °C, and the supernatant was transferred to another tube. The solutions were diluted with buffer, and ThT was added as a 1 mM stock solution to obtain a final concentration of 100 μM for the peptides, with a solubility above 100 μM, and 50 μM for ThT. For the other peptides, the solutions were solely diluted with the ThT stock solution, to achieve the highest possible concentration. For each sample, a total volume of 480 μL was prepared, and three aliquots with 150 μL were transferred to a medium binding 96-well plate (GREINER Bio-One 675 076). The plate was sealed with a microplate cover, and the fluorescence intensity was monitored at 37 °C using a BMG FLUOStar Omega multimode plate reader using fluorescence excitation and emission wavelengths at 460 and 485 nm, respectively. One measurement cycle lasted 5 min, consisting of double-orbital shaking for 150 s and incubating for 150 s. To describe the fibrillization process, the source data were fitted. As the description of fibrillar growth in terms of the molecular rate, kinetic can be considered as two main fibrillation processes.<sup>37</sup> Their contributions to the increase in

fluorescence  $\Delta F(t)$  can be analyzed by a function, which was derived by Dear et al.<sup>38</sup> and has been used to characterize the fibrillation kinetics of full-length PTH.<sup>20</sup>

$$\Delta F(t) = \left( 1 - \frac{1}{\frac{\kappa^3}{\lambda^3}(e^{\kappa t} - 1) + 1} \right) \Delta F_{pl} \quad (1)$$

$\Delta F_{pl}$  corresponds to the plateau value of the fluorescence in the stationary phase, while  $\lambda$  and  $\kappa$  are the rate constants of the primary and secondary nucleation processes.

**Determination of the Solubility Parameters.** To determine the maximal solubility of a peptide, small amounts (0.5 mg) of the respective peptide were added to 1 mL of buffered solution and shaken for 10 s until an insoluble precipitate was visible. The suspensions were centrifuged at 10 000 rpm for 1 h at 4 °C, and the supernatant was transferred to another tube. The remaining monomer concentration was determined using UV/vis-absorbance spectroscopy. To determine the critical concentration  $c_{cr}$ , two samples were prepared for each peptide according to the procedure for the ThT monitored fibrillization assay, one with ThT as a reference sample and the other one without ThT. Twenty hours after the reference sample reached the stationary phase, the fibril containing solutions from the sample without ThT were transferred to a tube and centrifuged at 10 000 rpm for 1 h at room temperature. The supernatant was transferred to another tube, and the concentration of the remaining monomers was determined using UV/vis-absorbance spectroscopy at 278 nm with a molar extinction coefficient of 3750 cm<sup>−1</sup> M<sup>−1</sup> for the azobenzene containing peptides and at 205 nm with a molar extinction coefficient of 49 310 cm<sup>−1</sup> M<sup>−1</sup> for PTH<sub>25–37</sub>.

**Photoisomerization.** The photoisomerization of the *trans*-azobenzene moiety in the peptides was performed by irradiating the dissolved peptides in a 1 cm quartz cuvette for 30 min with light of 340 nm wavelength using a 69.2 mW LED (Thorlabs, M340L5) equipped with a controller (Thorlabs, LEDD1B). The photoisomerization of the *cis*-azobenzene moiety in the peptides was performed by irradiating the dissolved peptides in a 1 cm quartz cuvette for 30 min with light of 405 nm wavelength using a 1.4 W LED (Thorlabs, M405L4) equipped with a controller (Thorlabs, LEDD1B).

**Photobleaching.** To test whether photobleaching occurs during the photoisomerization, a solution of 50 μM P4 was irradiated alternately for 30 min with light of 340 nm wavelength (*trans* → *cis* isomerization) and light of 405 nm wavelength (*cis* → *trans* isomerization). The absorption was measured after each isomerization step at two wavelengths: 295 nm (absorption maximum of the *cis*-isomer) and 327 nm (absorption maximum of the *trans*-isomer).

**Transmission Electron Microscopy.** TEM images were taken with an electron microscope (EM 900; Zeiss) at 80 kV acceleration voltage. For preparation, 5 μL of the peptide solution were added on Formvar/Cu grids (mesh 200). After 3 min of incubation, the grids were gently rinsed two times with water and then negatively stained using uranyl acetate (1%, w/v) for one minute.

**Fibril Degradation.** To investigate the fibril degradation through photoisomerization, 3 mL of a 100 μM *trans*-P4 in buffer was prepared according to the procedure for the ThT monitored fibrillization assay, except that ThT was excluded,

and aliquots of 150  $\mu\text{L}$  were transferred to a medium binding 96-well plate. As a reference sample, 480  $\mu\text{L}$  of a 100  $\mu\text{M}$  *trans*-P4 in buffer was prepared with ThT, and both approaches were incubated as described for the ThT monitored fibrillization assay. Twenty hours after the reference sample reached the stationary, the fibril containing solutions of the sample without ThT were collected in one tube. UV/vis-absorbance and CD were measured, and the fibril solution was transferred in a 1 cm quartz glass cuvette and treated according to the procedure for the *trans*  $\rightarrow$  *cis* isomerization, except that the irradiation time was prolonged to 5 h. Again UV/vis-absorbance and CD were measured. Subsequently, the *cis*-isomer was photoisomerized back to the *trans*-form. 456  $\mu\text{L}$  of the solution was mixed with 24  $\mu\text{L}$  of a 1 mM ThT stock solution and transferred as 150  $\mu\text{M}$  aliquots to a medium binding 96-well plate for a ThT monitored fibrillization assay. The remaining solution was transferred as well in 150  $\mu\text{L}$  aliquots to the 96-well plate and incubated under the same conditions. This procedure was repeated for every degradation cycle.

**Solid-State NMR Spectroscopy.** For the solid-state NMR spectroscopy, 6 mL of a 100  $\mu\text{M}$  solution of the respective peptides (PTH<sub>25–37</sub> and P4 with uniformly  $^{15}\text{N}/^{13}\text{C}$  labeled residues L28 and F34) was prepared according to the procedure for the ThT monitored fibrillization assay and transferred in 150  $\mu\text{L}$  aliquots to a medium binding 96-well plate. Twenty hours after the stationary phase was reached, the fibril containing solutions were collected in a tube and centrifuged at 60 000 rpm for 15 min at 10  $^{\circ}\text{C}$ . The supernatant was discarded, the remaining fibril pellet was frozen with liquid nitrogen, and the residual water was removed using lyophilization. The lyophilized peptides were packed into Bruker MAS rotors with outer diameters of 3.2 mm (PTH<sub>25–37</sub>) and 1.9 mm (P4), respectively. Solid-state NMR experiments were conducted on an 18.8 T (800 MHz  $^1\text{H}$  frequency) Bruker Avance III spectrometer equipped with a triple resonance HCN 3.2 mm MAS Efree probe and a 1.9 mm triple resonance probe. For determination of intermolecular long-range contacts between L28 and F34 resonances, proton-driven spin diffusion (PDSF) experiments with an MAS frequency of 20 kHz close to the rotational resonance condition between aromatic  $^{13}\text{C}$  resonances of F34 and aliphatic  $\text{C}_\gamma/\text{C}_\delta$  resonances of L28 were recorded with mixing times ranging from 50 ms to 1 s. For PTH<sub>25–37</sub>, additional PDSF and DQSQ spectra were recorded on a 14.1 T (600 MHz  $^1\text{H}$  frequency) Bruker Avance wide bore spectrometer equipped with a 3.2 mm MAS triple resonance  $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  probe. Typical radiofrequency field strengths were 91–100 kHz for  $^1\text{H}$ , and 55.6 kHz for  $^{13}\text{C}$  Spinal64  $^1\text{H}$  decoupling<sup>39</sup> (rf field of 85 kHz) was applied during  $^{13}\text{C}$  evolution and acquisition. The VT gas temperature was set to 263 K (thermocouple reported temperature); the sample temperature was estimated to be around 10–20 K higher due to frictional heating under MAS.

**WAXS Measurements.** The WAXS measurements were performed in transmission mode with a SAXSLAB laboratory setup (Retro-F) equipped with an AXO microfocus X-ray source. As a monochromator, the AXO multilayer X-ray optic (AXO Dresden GmbH, Dresden, Germany) was used for Cu-K $\alpha$  ( $\lambda = 0.154$  nm). The two-dimensional scattering patterns were recorded with a two-dimensional detector (PILATUS3 R 300 K, DECTRIS, Baden, Switzerland). The preformed fibrils were prepared as described for the measurement of the aggregation kinetics in a total volume of 1 mL. Twenty hours

after the fibrillization reached the stationary phase, the fibril suspension was ultracentrifuged at 60.000 g for 15 min, and the obtained pellet was transferred into a glass capillary and dried overnight. The scattering measurements were performed at room temperature in vacuum and corrected for background.

**Cytotoxicity Tests.** Cell viability was determined for *trans*-P4, *trans*-P8, and *trans*-P12 on NHDF as well as 3T3 fibroblasts with a resazurin reduction assay. Briefly, cells were seeded at desired cell densities in corresponding culture media supplemented with 10% FCS and penicillin/streptomycin in 96-well plates on day 0 and incubated overnight under standard cell culture conditions. On day 1, serial dilutions of the peptides in cell culture media were prepared. Then, medium (background and negative control–100% viability), TritonX in medium f.c. 0.025% (positive control–0% viability), and the peptide dilutions were added with one treatment per column ( $n = 8$ ). Incubation for 24 h or 96 h under standard cell culture conditions followed. On day of measurement, resazurin stock solution was added to a final concentration of 44  $\mu\text{M}$  into each well. After 2 h incubation, the resorufin fluorescence was measured with a Cytation 5 plate reader system. Means and standard deviation of each column were calculated. Experiments were repeated independently 3 times, and average values were evaluated (see Figure S12).

**MD Simulations.** For all MD simulations of the PTH<sub>25–37</sub> peptide, as well as the mutants P1, P3, P4, P8, and P12, we used the GROMACS simulation package.<sup>40</sup> Since the PTH<sub>25–37</sub> sequence classifies as an intrinsically disordered protein (IDP), we have used the CHARMM36m force field<sup>41</sup> to model protein interaction, which has previously been shown to be a suitable choice for IDPs.<sup>42</sup> For the azobenzene photoswitch (denoted AZO in our force field implementation), we parametrized the interactions following the cgenff standard protocol<sup>43</sup> and refined the parametrization using data from QM/MM simulations,<sup>44</sup> by fitting our parameters to reproduce the bond, angle, and dihedral angle distributions of the *cis*- and *trans*-state as obtained from QM/MM. The resulting force field parameters are available at [https://github.com/strodel-group/Charmm36m\\_Azobenzene-FF](https://github.com/strodel-group/Charmm36m_Azobenzene-FF). During the MD simulations of the modified PTH<sub>25–37</sub> peptides, the  $\angle\text{CNNC}$  dihedral angle was restricted to either the *cis*- or *trans*-states; thus, transitions between the two states of the AMPB photoswitch were not modeled. The MD simulations of all systems were prepared following the same protocol: first, the peptide(s) were placed in the simulation box, where in case of the dimer and hexamer simulations, the box size was always chosen to achieve a peptide concentration of 10 mM. Then, the box was filled with TIP3P water molecules,<sup>45</sup> as well as Na<sup>+</sup> and Cl<sup>−</sup> ions to neutralize the system and achieve a physiological salt concentration of 150 mM. After equilibration of the systems, a production run of 10  $\mu\text{s}$  per system (1  $\mu\text{s}$  for the fibril models) was carried out under  $NpT$  conditions at constant number of particles  $N$ , pressure  $p = 1$ , bar and temperature  $T = 300$  K. The pressure and temperature were regulated using the Parrinello–Rahman pressure coupling scheme<sup>46</sup> and Nosé–Hoover thermostat,<sup>47</sup> respectively. To exclude edge effects, periodic boundary conditions in all directions were applied, and the particle-mesh Ewald method<sup>48</sup> was used to calculate electrostatic interactions. For the calculation of van der Waals and Coulomb interaction in real space, a cutoff of 12 Å was applied. An overview over all simulations performed is given in Table S1, yielding 285  $\mu\text{s}$  of total simulation time across all systems. All MD simulations

were run on the high-performance cluster JURECA-DC.<sup>49</sup> Analysis of the MD data was performed with python using the MDAnalysis<sup>50</sup> package for reading of the MD trajectory and the calculation of distances between groups of atoms, while the MDTraj<sup>51</sup> package was used for secondary structure analysis. For visualization of the MD structures, the PyMOL<sup>52</sup> software was used. Additionally, PyMOL in combination with the APBS<sup>53</sup> plugin was used for the calculation and visualization of electrostatic potential surfaces.

**QM Calculations.** The photoswitching mechanism of AMPB integrated into PTH<sub>25–37</sub> was monitored in the ground and excited states along the  $\angle$ CNNC dihedral angle for the *cis*  $\leftrightarrow$  *trans*-transition of the azobenzene photoswitch. As a starting point for the QM calculations, we constructed pathways from *cis* to *trans* and *vice versa*, using MD simulations. To consider the structural ensemble along the pathway, we simulated 40 *cis*  $\leftrightarrow$  *trans* switching trajectories. We started the switching from equilibrated *cis*- and *trans*-P4 structures obtained from equidistant time steps of 250 ns from the 10  $\mu$ s MD simulations. The switching in the MD simulations was achieved by imposing a restraining potential on the  $\angle$ CNNC dihedral angle and changing it every 2 ns in increments of 10° between 0° and 180°. All other MD simulation parameters were the same as described above. The resulting trajectories were used as input for the subsequent QM calculations. The QM calculations were conducted using the ONIOM-based QM/QM2 method,<sup>30</sup> with NEVPT2(2,2)/def2-TZVP<sup>31,54</sup> for QM and xTB2<sup>32</sup> for QM2, as implemented in the ORCA program package.<sup>55</sup> The QM region was focused on the azobenzene core, including the peptide in  $\pi$ -conjugation with it. The solvent was implicitly modeled using the ALPB method.<sup>56</sup> Due to convergence issues, analysis was limited to 34 of the original 40 trajectories. For these paths, the energies of electronic states *S*<sub>0</sub>, *S*<sub>1</sub>, and *T*<sub>1</sub> were interpolated to obtain potential energy curves using Gaussian process regression from the scikit-learn package.<sup>57</sup> Some of the paths that have an energy barrier at dihedral angles around 170° were analyzed at the structural level. To this end, the inter-residue distances based on the centers of mass of the residues were calculated for all structures exhibiting a dihedral angle near 170° and then analyzed through dimensionality reduction via principal component analysis (PCA). The first three principal components were transferred back into the original distance matrix format and illustrated with three representative distance matrices from the corresponding cluster of structures. Representative structures for the distance matrices were also extracted and visualized.

## ■ ASSOCIATED CONTENT

### SI Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.bioconjchem.4c00188>.

**Supplementary figures** S1, QM potential energy curves for 34 *cis*  $\leftrightarrow$  *trans* isomerization paths; S2, fluorescence lifetime measurements; S3, PCA analysis of selected P4 structures; S4, principal components projected onto inter-residue distance space; S5, selected P4 snapshots; S6, WAXS data of P4 and PTH<sub>25–37</sub>; S7 ssNMR data for PTH<sub>25–37</sub> fibrils; S8, ssNMR data for *trans*-P4 fibrils; S9, summary of all MD simulation results; S10, electrostatic potential surface of PTH<sub>25–37</sub> and selected peptides; S11, absorption data to exclude photobleaching; S12,

cell viability assay of *trans*-P4, *trans*-P8, and *trans*-P12; S13–S20, ThT fluorescence data; S21–S27, TEM images of fibrils; S28, CD spectra during reversible fibrillization of P4; S29, UV–vis spectra of *trans*-P4 before each fibrillization cycle; S30–S47, <sup>1</sup>H NMR, HPLC traces, and MALDI data of the peptides P1–P12. **Supplementary tables** S1, list of all MD simulations performed; S2, sequence of PTH<sub>25–37</sub> and P1–P12 and their respective solubility; S3, photophysical properties of peptides; S4: fibrillization parameters of peptides (PDF)

## ■ AUTHOR INFORMATION

### Corresponding Authors

**Birgit Strodel** – Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf 40225, Germany; Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, Jülich 52425, Germany; [orcid.org/0000-0002-8734-7765](https://orcid.org/0000-0002-8734-7765); Email: [b.strodel@fz-juelich.de](mailto:b.strodel@fz-juelich.de)

**Wolfgang H. Binder** – Macromolecular Chemistry, Institute of Chemistry, Faculty of Natural Science II, Martin Luther University Halle Wittenberg, Halle 06120, Germany; [orcid.org/0000-0003-3834-5445](https://orcid.org/0000-0003-3834-5445); Email: [wolfgang.binder@chemie.uni-halle.de](mailto:wolfgang.binder@chemie.uni-halle.de)

### Authors

**André Paschold** – Macromolecular Chemistry, Institute of Chemistry, Faculty of Natural Science II, Martin Luther University Halle Wittenberg, Halle 06120, Germany

**Moritz Schäffler** – Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, Düsseldorf 40225, Germany; Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, Jülich 52425, Germany

**Xincheng Miao** – Center for Nanosystems Chemistry (CNC), Theodor-Boveri Weg, Universität Würzburg, Würzburg 97074, Germany

**Luis Gardon** – Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, Jülich 52425, Germany; Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

**Stephanie Krüger** – Biozentrum, Martin Luther University Halle-Wittenberg, Halle 06120, Germany

**Henrike Heise** – Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, Jülich 52425, Germany; Institut für Physikalische Biologie, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany; [orcid.org/0000-0002-9081-3894](https://orcid.org/0000-0002-9081-3894)

**Merle I. S. Röhr** – Center for Nanosystems Chemistry (CNC), Theodor-Boveri Weg, Universität Würzburg, Würzburg 97074, Germany; [orcid.org/0000-0002-2128-3498](https://orcid.org/0000-0002-2128-3498)

**Maria Ott** – Institute of Biophysics, Faculty of Natural Science I, Martin Luther University Halle-Wittenberg, Halle 06120, Germany; [orcid.org/0000-0002-4686-9516](https://orcid.org/0000-0002-4686-9516)

Complete contact information is available at: <https://pubs.acs.org/doi/10.1021/acs.bioconjchem.4c00188>

### Author Contributions

<sup>○</sup>A.P. and M.S. contributed equally



## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

WHB thanks the DFG project INST 271/444-1 FUGG for financial support; the DFG-Project BI1337/16-1; BI 1337/14-1 and the GRK 2670, W69000789, ProjectNr 436494874. MISR acknowledges funding by the Bavarian State Initiative "Solar Technologies Go Hybrid." MS and BS gratefully acknowledge computing time on the supercomputer JURECA at Forschungszentrum Jülich under grant no. IDP. H.H. and L.G. thank the DFG (INST 208/771-1 FUGG) for financial support. Access to the Jülich-Düsseldorf Biomolecular NMR Center jointly run by Forschungszentrum Jülich and HHU is acknowledged. We thank Dr. Sven Rothmund, from the Core Unit Peptide-Technologies of University Leipzig, for synthesizing the peptides and Julia Kollan and Dr. Henrike Lucas from the AG Mäder, Department of Pharmaceutical Technology at the Martin Luther University Halle-Wittenberg, for conducting the toxicological studies.

## ■ ABBREVIATIONS

AMPB, 3-[[4-(aminomethyl)phenyl]diazanyl]benzoic acid; CD, circular dichroism; MD, molecular dynamics; NMR, nuclear magnetic resonance; PSS, photostationary state; PTH, parathyroid hormone; ssNMR, solid state nuclear magnetic resonance; TEM, transmission electron microscopy; ThT, thioflavin T; UV/vis, ultraviolet/visible; WAXS, wide-angle X-ray scattering

## ■ REFERENCES

- (1) Li, D.; Liu, C. Conformational strains of pathogenic amyloid proteins in neurodegenerative diseases. *Nat. Rev. Neurosci.* **2022**, *23* (9), 523–534. (b) Alberstein, R. G.; Guo, A. B.; Kortemme, T. Design principles of protein switches. *Curr. Opin. Struct. Biol.* **2022**, *72*, 71–78. (c) Adamson, H.; Jeuken, L. J. Engineering protein switches for rapid diagnostic tests. *ACS Sens.* **2020**, *5* (10), 3001–3012. (d) Wang, H.; Feng, Z.; Xu, B. Assemblies of peptides in a complex environment and their applications. *Angew. Chem., Int. Ed.* **2019**, *131* (31), 10532–10541. (e) Bierlmeier, J.; Álvaro-Benito, M.; Scheffler, M.; Sturm, K.; Rehkopf, L.; Freund, C.; Schwarzer, D. Sortase-Mediated Multi-Fragment Assemblies by Ligation Site Switching. *Angew. Chem., Int. Ed.* **2022**, *61* (5), No. e202109032. (f) Shen, Y.; Wang, Y.; Hamley, I. W.; Qi, W.; Su, R.; He, Z. Chiral self-assembly of peptides: Toward the design of supramolecular polymers with enhanced chemical and biological functions. *Prog. Polym. Sci.* **2021**, *123*, 101469.
- (2) Li, Q.; Zhang, G.; Wu, Y.; Wang, Y.; Liang, Y.; Yang, X.; Qi, W.; Su, R.; He, Z. Control of peptide hydrogel formation and stability via heating treatment. *J. Colloid Interface Sci.* **2021**, *583*, 234–242. (b) Trzebicka, B.; Szveda, R.; Kosowski, D.; Szveda, D.; Otulakowski, Ł.; Haladjova, E.; Dworak, A. Thermoresponsive polymer-peptide/protein conjugates. *Prog. Polym. Sci.* **2017**, *68*, 35–76. (c) Hamley, I. W.; Dehsorkhi, A.; Castelletto, V.; Fuzeland, S.; Atkins, D.; Seitsonen, J.; Ruokolainen, J. Reversible helical unwinding transition of a self-assembling peptide amphiphile. *Soft Mater.* **2013**, *9* (39), 9290–9293.
- (3) Gontsarik, M.; Yagmur, A.; Ren, Q.; Maniura-Weber, K.; Salentinig, S. From structure to function: pH-switchable antimicrobial nano-self-assemblies. *ACS Appl. Mater. Interfaces* **2019**, *11* (3), 2821–2829. (b) Chen, H.; Cheng, J.; Cai, X.; Han, J.; Chen, X.; You, L.; Xiong, C.; Wang, S. pH-switchable antimicrobial supramolecular hydrogels for synergistically eliminating biofilm and promoting wound healing. *ACS Appl. Mater. Interfaces* **2022**, *14* (16), 18120–18132.
- (4) Fu, Y.; Xiao, K.; Zhang, X.; Du, C.; Chen, J. Peptide cleavage-mediated and environmentally friendly photocurrent polarity switching system for prostate-specific antigen assay. *Anal. Chem.* **2021**, *93* (2), 1076–1083.
- (5) Szymański, W.; Beierle, J. M.; Kistemaker, H. A.; Velema, W. A.; Feringa, B. L. Reversible photocontrol of biological systems by the incorporation of molecular photoswitches. *Chem. Rev.* **2013**, *113* (8), 6114–6178.
- (6) Liu, X.; Ramirez, S.; Pang, P. T.; Puryear, C. B.; Govindarajan, A.; Deisseroth, K.; Tonegawa, S. Optogenetic stimulation of a hippocampal engram activates fear memory recall. *Nature* **2012**, *484* (7394), 381–385. (b) Banghart, M.; Borges, K.; Isacoff, E.; Trauner, D.; Kramer, R. H. Light-activated ion channels for remote control of neuronal firing. *Nat. Neurosci.* **2004**, *7* (12), 1381–1386. (c) Hull, K.; Morstein, J.; Trauner, D. In Vivo Photopharmacology. *Chem. Rev.* **2018**, *118* (21), 10710–10747. (d) Peddie, V.; Abell, A. D. Photocontrol of Peptide Secondary Structure through Non-Azobenzene Photoswitches. *J. Photochem. Photobiol., C* **2019**, *40*, 1–20.
- (7) DuBay, K. H.; Iwan, K.; Osorio-Planes, L.; Geissler, P. L.; Groll, M.; Trauner, D.; Broichhagen, J. A predictive approach for the optical control of carbonic anhydrase II activity. *ACS Chem. Biol.* **2018**, *13* (3), 793–800. (b) Volgraf, M.; Gorostiza, P.; Numano, R.; Kramer, R. H.; Isacoff, E. Y.; Trauner, D. Allosteric control of an ionotropic glutamate receptor with an optical switch. *Nat. Chem. Biol.* **2006**, *2* (1), 47–52.
- (8) Jankovic, B.; Bozovic, O.; Hamm, P. Intrinsic dynamics of protein–peptide unbinding. *Biochemistry* **2021**, *60* (22), 1755–1763. (b) Yeoh, Y. Q.; Yu, J.; Polyak, S. W.; Horsley, J. R.; Abell, A. D. Photopharmacological Control of Cyclic Antimicrobial Peptides. *ChemBioChem* **2018**, *19* (24), 2591–2597. (c) Ventura, C. R.; Wiedman, G. R. Substituting azobenzene for proline in melittin to create photomelittin: A light-controlled membrane active peptide. *Biochim. Biophys. Acta, Biomembr.* **2021**, *1863* (12), 183759.
- (9) Prodromou, R.; Moore, B. D.; Chu, W.; Deal, H.; San Miguel, A.; Brown, A. C.; Daniele, M. A. A.; Pozdin, V. A.; Menegatti, S. Molecular Engineering of Cyclic Azobenzene-Peptide Hybrid Ligands for the Purification of Human Blood Factor VIII via Photo-Affinity Chromatography. *Adv. Funct. Mater.* **2023**, *33* (14), 2213881.
- (10) Dong, S. L.; Löweneck, M.; Schrader, T. E.; Schreier, W. J.; Zinth, W.; Moroder, L.; Renner, C. A Photocontrolled  $\beta$ -Hairpin Peptide. *Chem. Eur. J.* **2006**, *12* (4), 1114–1120. (b) Kumita, J. R.; Flint, D. G.; Smart, O. S.; Woolley, G. A. Photo-control of peptide helix content by an azobenzene cross-linker: steric interactions with underlying residues are not critical. *Protein Eng.* **2002**, *15* (7), 561–569. (c) Nuti, F.; Gellini, C.; Larregola, M.; Squillantini, L.; Chelli, R.; Salvi, P. R.; Lequin, O.; Pietrapertusa, G.; Papini, A. M. A photochromic azobenzene peptidomimetic of a  $\beta$ -turn model peptide structure as a conformational switch. *Front. Chem.* **2019**, *7*, 180. (d) Ulysse, L.; Cubillos, J.; Chmielewski, J. Photoregulation of cyclic peptide conformation. *J. Am. Chem. Soc.* **1995**, *117* (32), 8466–8467. (e) Cataldi, E.; Raschig, M.; Gutmann, M.; Geppert, P. T.; Ruopp, M.; Schock, M.; Gerwe, H.; Bertermann, R.; Meinel, L.; Finze, M. Amber light control of peptide secondary structure by a perfluoroaromatic azobenzene photoswitch. *ChemBioChem* **2023**, *24* (5), No. e202200570. (f) Parlato, R.; Volarić, J.; Lasorsa, A.; Kobauri, P.; Jain, G.; Miettinen, M. S.; Feringa, B. L.; Szymanski, W.; van der Wel, P. C. Photocontrol of the  $\beta$ -Hairpin Polypeptide Structure through an Optimized Azobenzene-Based Amino Acid Analogue. *J. Am. Chem. Soc.* **2024**, *146* (3), 2062–2071. (g) Doran, T. M.; Anderson, E. A.; Latchney, S. E.; Opanashuk, L. A.; Nilsson, B. L. An azobenzene photoswitch sheds light on turn nucleation in amyloid-beta self-assembly. *ACS Chem. Neurosci.* **2012**, *3* (3), 211–220.
- (11) Binder, W. H.; Smrzka, O. W. Self-Assembly of Fibers and Fibrils. *Angew. Chem., Int. Ed.* **2006**, *45* (44), 7324–7328.
- (12) Evgrafova, Z.; Rothmund, S.; Voigt, B.; Hause, G.; Balbach, J.; Binder, W. H. Synthesis and Aggregation of Polymer-Amyloid  $\beta$  Conjugates. *Macromol. Rapid Commun.* **2020**, *41* (1), 1900378. (b) Evgrafova, Z.; Voigt, B.; Roos, A. H.; Hause, G.; Hinderberger, D.; Balbach, J.; Binder, W. H. Modulation of amyloid  $\beta$  peptide aggregation by hydrophilic polymers. *Phys. Chem. Chem. Phys.* **2019**,

- 21 (37), 20999–21006. (c) Deike, S.; Rothmund, S.; Voigt, B.; Samantray, S.; Strodel, B.; Binder, W. H.  $\beta$ -Turn mimetic synthetic peptides as amyloid- $\beta$  aggregation inhibitors. *Bioorg. Chem.* **2020**, p 104012. (d) Funtan, S.; Evgrafova, Z.; Adler, J.; Huster, D.; Binder, W. H. Amyloid Beta Aggregation in the Presence of Temperature-Sensitive Polymers. *Polymers* **2016**, 8 (5), 178. (e) Sen, N.; Hause, G.; Binder, W. H. Membrane Anchored Polymers Modulate Amyloid Fibrillation. *Macromol. Rapid Commun.* **2021**, p 2100120. .
- (13) Zweifler, L. E.; Koh, A. J.; Daignault-Newton, S.; McCauley, L. K. Anabolic actions of PTH in murine models: two decades of insights. *J. Bone Miner. Res.* **2020**, 36 (10), 1979–1998. (b) Pinheiro, P. L.; Cardoso, J. C.; Power, D. M.; Canário, A. V. Functional characterization and evolution of PTH/PTHrP receptors: Insights from the chicken. *BMC Evol. Biol.* **2012**, 12 (1), 1–15. (c) Guerreiro, P. M.; Renfro, J. L.; Power, D. M.; Canario, A. V. The parathyroid hormone family of peptides: structure, tissue distribution, regulation, and potential functional roles in calcium and phosphate balance in fish. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **2007**, 292 (2), R679–R696. (d) Xie, J.; Sang, M.; Song, X.; Zhang, S.; Kim, D.; Veenstra, J. A.; Park, Y.; Li, B. A new neuropeptide insect parathyroid hormone iPTH in the red flour beetle *Tribolium castaneum*. *PLoS Genet.* **2020**, 16 (5), No. e1008772.
- (14) Martin, T. J.; Sims, N. A.; Seeman, E. Physiological and Pharmacological Roles of PTH and PTHrP in Bone using their Shared Receptor, PTH1R. *Endocr. Rev.* **2021**, 42 (4), 383–406. (b) Friedlander, G.; Amiel, C. Cellular mode of action of parathyroid hormone. *Adv. Nephrol. Necker Hosp.* **1994**, 23, 265–279.
- (15) Gopalswamy, M.; Kumar, A.; Adler, J.; Baumann, M.; Henze, M.; Kumar, S. T.; Fändrich, M.; Scheidt, H. A.; Huster, D.; Balbach, J. Structural characterization of amyloid fibrils from the human parathyroid hormone. *Biochim. Biophys. Acta Proteins Proteom.* **2015**, 1854 (4), 249–257.
- (16) Sachan, S.; Moya, C. G.; Voigt, B.; Köhn, M.; Balbach, J. The pro-sequence of parathyroid hormone prevents premature amyloid fibril formation. *FEBS Lett.* **2023**, 597, 995–1006.
- (17) Potts, J. T. Parathyroid hormone: past and present. *J. Endocrinol.* **2005**, 187 (3), 311–325.
- (18) Rendina-Ruedy, E.; Rosen, C. J. Parathyroid hormone (PTH) regulation of metabolic homeostasis: An old dog teaches us new tricks. *Mol. Metab.* **2022**, 60, 101480.
- (19) Wang, X.; Cheng, X.; Zhao, L.; Wang, Y.; Ye, C.; Zou, X.; Dai, A.; Cong, Z.; Chen, J.; Zhou, Q. Molecular insights into differentiated ligand recognition of the human parathyroid hormone receptor 2. *Proc. Natl. Acad. Sci. U. S. A.* **2021**, 118 (32), No. e2101279118.
- (20) Lauth, L. M.; Voigt, B.; Bhatia, T.; Machner, L.; Balbach, J.; Ott, M. Heparin promotes rapid fibrillation of the basic parathyroid hormone at physiological pH. *FEBS Lett.* **2022**, 596 (22), 2928–2939.
- (21) Evgrafova, Z.; Voigt, B.; Baumann, M.; Stephani, M.; Binder, W. H.; Balbach, J. Probing Polymer Chain Conformation and Fibril Formation of Peptide Conjugates. *ChemPhysChem* **2019**, 20 (2), 236–240.
- (22) Paschold, A.; Voigt, B.; Hause, G.; Kohlmann, T.; Rothmund, S.; Binder, W. H. Modulating the Fibrillization of Parathyroid-Hormone (PTH) Peptides: Azo-Switches as Reversible and Catalytic Entities. *Biomedicines* **2022**, 10 (7), 1512.
- (23) Walsh, I.; Seno, F.; Tosatto, S. C. E.; Trovato, A. PASTA 2.0: an improved server for protein aggregation prediction. *Nucleic Acids Res.* **2014**, 42 (W1), W301–W307.
- (24) Conchillo-Solé, O.; de Groot, N. S.; Avilés, F. X.; Vendrell, J.; Daura, X.; Ventura, S. AGGRESKAN: a server for the prediction and evaluation of "hot spots" of aggregation in polypeptides. *BMC Bioinform.* **2007**, 8 (1), 65.
- (25) Burdukiewicz, M.; Sobczyk, P.; Rödiger, S.; Duda-Madej, A.; Mackiewicz, P.; Kotulska, M. Amyloidogenic motifs revealed by n-gram analysis. *Sci. Rep.* **2017**, 7 (1), 12961.
- (26) Garbuzynski, S. O.; Lobanov, M. Y.; Galzitskaya, O. V. FoldAmyloid: a method of prediction of amyloidogenic regions from protein sequence. *Bioinformatics* **2010**, 26 (3), 326–332.
- (27) 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.* **2020**, 16 (12), 7825–7839. (b) Smorodina, E.; Kav, B.; Fatafta, H.; Strodel, B. Effects of ion type and concentration on the structure and aggregation of the amyloid peptide A $\beta$ 16–22. *Proteins: struct., Funct., Bioinf.* **2023**, Vol. 91 (1–14). .
- (28) Bera, S.; Arad, E.; Schnaider, L.; Shaham-Niv, S.; Castelletto, V.; Peretz, Y.; Zaguri, D.; Jelinek, R.; Gazit, E.; Hamley, I. W. Unravelling the role of amino acid sequence order in the assembly and function of the amyloid- $\beta$  core. *Chem. Commun.* **2019**, 55 (59), 8595–8598.
- (29) Dyson, H. J.; Wright, P. E. Coupling of folding and binding for unstructured proteins. *Curr. Opin. Struct. Biol.* **2002**, 12 (1), 54–60.
- (30) Mayhall, N. J.; Raghavachari, K.; Hratchian, H. P. ONIOM-based QM: QM electronic embedding method using Löwdin atomic charges: Energies and analytic gradients. *J. Chem. Phys.* **2010**, 132 (11), 114107.
- (31) Angeli, C.; Cimiraglia, R.; Malrieu, J.-P. N-electron valence state perturbation theory: a fast implementation of the strongly contracted variant. *Chem. Phys. Lett.* **2001**, 350 (3–4), 297–305.
- (32) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. *J. Chem. Theory Comput.* **2019**, 15 (3), 1652–1671.
- (33) Shimomura, M.; Kunitake, T. Fluorescence and photo-isomerization of azobenzene-containing bilayer membranes. *J. Am. Chem. Soc.* **1987**, 109 (17), 5175–5183.
- (34) Serpell, L. C.; Fraser, P. E.; Sunde, M. [34] X-Ray fiber diffraction of amyloid fibrils. In *Methods in Enzymology*; Elsevier, 1999; Vol. 309, pp. 526536. .
- (35) Petkova, A. T.; Yau, W.-M.; Tycko, R. Experimental constraints on quaternary structure in Alzheimer's  $\beta$ -amyloid fibrils. *Biochemistry* **2006**, 45 (2), 498–512.
- (36) Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; et al. Atomic structures of amyloid cross- $\beta$  spines reveal varied steric zippers. *Nature* **2007**, 447 (7143), 453–457. Riek, R. The three-dimensional structures of amyloids. *Cold Spring Harb Perspect. Biol.* **2017**, 9 (2), a023572.
- (37) Meisl, G.; Rajah, L.; Cohen, S. A.; Pfammatter, M.; Šarić, A.; Hellstrand, E.; Buell, A. K.; Aguzzi, A.; Linse, S.; Vendruscolo, M.; Dobson, C. M. Scaling behaviour and rate-determining steps in filamentous self-assembly. *Chem. Sci.* **2017**, 8 (10), 7087–7097.
- (38) Dear, A. J.; Michaels, T. C.; Meisl, G.; Klenerman, D.; Wu, S.; Perrett, S.; Linse, S.; Dobson, C. M.; Knowles, T. P. Kinetic diversity of amyloid oligomers. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, 117 (22), 12087–12094.
- (39) Fung, B.; Khitrin, A.; Ermolaev, K. An improved broadband decoupling sequence for liquid crystals and solids. *J. Magn. Reson.* **2000**, 142 (1), 97–101.
- (40) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, 1, 19–25.
- (41) Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; De Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: an improved force field for folded and intrinsically disordered proteins. *Nat. Methods* **2017**, 14 (1), 71–73.
- (42) Samantray, S.; Yin, F.; Kav, B.; Strodel, B. Different force fields give rise to different amyloid aggregation pathways in molecular dynamics simulations. *J. Chem. Inf. Model.* **2020**, 60 (12), 6462–6475.
- (43) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, 31 (4), 671–690.

- (44) Böckmann, M.; Peter, C.; Site, L. D.; Doltsinis, N. L.; Kremer, K.; Marx, D. Atomistic force field for azobenzene compounds adapted for QM/MM simulations with applications to liquids and liquid crystals. *J. Chem. Theory Comput.* **2007**, *3* (5), 1789–1802.
- (45) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79* (2), 926–935.
- (46) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52* (12), 7182–7190.
- (47) Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984**, *52* (2), 255–268. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31* (3), 1695.
- (48) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An  $N \cdot \log(N)$  method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089–10092.
- (49) Krause, D.; Thörnig, P. JURECA: modular supercomputer at Jülich supercomputing centre. *JLSRF* **2018**, *4*, A132–A132.
- (50) Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M. MDAnalysis: a Python package for the rapid analysis of molecular dynamics simulations. *Proceedings of the 15th python in science conference* Los Alamos National Laboratory 2016, 98, 105.
- (51) McGibbon, R. T.; Beauchamp, K. A.; Harrigan, M. P.; Klein, C.; Swails, J. M.; Hernández, C. X.; Schwantes, C. R.; Wang, L.-P.; Lane, T. J.; Pande, V. S. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **2015**, *109* (8), 1528–1532.
- (52) Schrödinger, L. *The PyMOL Molecular Graphics System, Version 1.8*; Cñiü, 2015.
- (53) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K.; Chun, M. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **2018**, *27* (1), 112–128.
- (54) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.* **2005**, *7* (18), 3297–3305. Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic generation of auxiliary basis sets. *J. Chem. Theory Comput.* **2017**, *13* (2), 554–562.
- (55) Neese, F. Software update: The ORCA program system Version 5.0. *WIREs Comput. Mol. Sci.* **2022**, *12* (5), No. e1606.
- (56) Sigalov, G.; Fenley, A.; Onufriev, A. Analytical electrostatics for biomolecules: Beyond the generalized Born approximation. *J. Chem. Phys.* **2006**, *124* (12), 124902.
- (57) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
- (58) Meisl, G.; Kirkegaard, J. B.; Arosio, P.; Michaels, T. C.; Vendruscolo, M.; Dobson, C. M.; Linse, S.; Knowles, T. P. Molecular mechanisms of protein aggregation from global fitting of kinetic models. *Nat. Prot.* **2016**, *11* (2), 252–272.
- (59) Voigt, B.; Bhatia, T.; Hesselbarth, J.; Baumann, M.; Schmidt, C.; Ott, M.; Balbach, J. The Prenucleation Equilibrium of the Parathyroid Hormone Determines the Critical Aggregation Concentration and Amyloid Fibril Nucleation Concentration and Amyloid Fibril Nucleation. *ChemPhysChem* **2023**, *24* (19), No. e202300439. Voigt, B.; Ott, M.; Balbach, J. A Competition of Secondary and Primary Nucleation Controls Amyloid Fibril Formation of the Parathyroid Hormone. *Macromol. Biosci.* **2023**, *23* (4), 2200525.
- (60) Maity, C.; Hendriksen, W. E.; van Esch, J. H.; Eelkema, R. Spatial structuring of a supramolecular hydrogel by using a visible-light triggered catalyst. *Angew. Chem., Int. Ed.* **2015**, *54* (3), 998–1001.
- (61) Li, J.; Wang, H.; Liu, B.; Chen, J.; Gu, J.; Lin, S. Photoinduced Contraction Fibers and Photoswitchable Adhesives Generated by Stretchable Supramolecular Gel. *Adv. Funct. Mater.* **2022**, *32* (30), 2201851.
- (62) Trauner, D.; Fialho, D. M. A Modular Platform for Covalent, Photoswitchable Control over Enzyme Activity. *Synfacts* **2023**, *19* (8), 0831.
- (63) Álvarez, Z.; Kolberg-Edelbrock, A. N.; Sasselli, I. R.; Ortega, J. A.; Qiu, R.; Syrgiannis, Z.; Mirau, P. A.; Chen, F.; Chin, S. M.; Weigand, S.; et al. Bioactive scaffolds with enhanced supramolecular motion promote recovery from spinal cord injury. *Sciences* **2021**, *374* (6569), 848–856.

# Supporting Information

## Photocontrolled reversible amyloid fibril formation of parathyroid hormone-derived peptides

André Paschold<sup>1,#</sup>, Moritz Schäffler<sup>2,3,#</sup>, Xincheng Miao<sup>4</sup>, Luis Gardon<sup>2,5</sup>, Stephanie Krüger<sup>6</sup>, Henrike Heise<sup>2,5</sup>, Merle Insa Silja Röhr<sup>4</sup>, Maria Ott<sup>7</sup>, Birgit Strodel<sup>2,3\*</sup>, Wolfgang H. Binder<sup>1\*</sup>

<sup>1</sup> Macromolecular Chemistry, Institute of Chemistry, Faculty of Natural Science II, Martin Luther University Halle Wittenberg, von-Danckelmann-Platz 4, 06120 Halle (Germany)

<sup>2</sup> Institute of Theoretical and Computational Chemistry, Heinrich Heine University Düsseldorf, 40225 Düsseldorf, Germany

<sup>3</sup> Institute of Biological Information Processing, Structural Biochemistry (IBI-7), Forschungszentrum Jülich, 52425 Jülich, Germany

<sup>4</sup> Center for Nanosystems Chemistry (CNC), Theodor-Boveri Weg, Universität Würzburg, 97074 Würzburg, Germany

<sup>5</sup> Institute of Physical Biology, Heinrich-Heine-Universität Düsseldorf, 40225 Düsseldorf, Germany

<sup>6</sup> Biozentrum, Martin Luther University Halle-Wittenberg, Weinberweg 22, 06120 Halle (Germany)

<sup>7</sup> Institute of Biophysics, Faculty of Natural Science I, Martin Luther University Halle-Wittenberg, Kurt-Mothes-Straße 3, 06120 Halle (Germany)

# These authors contributed equally to this work.

## Contents

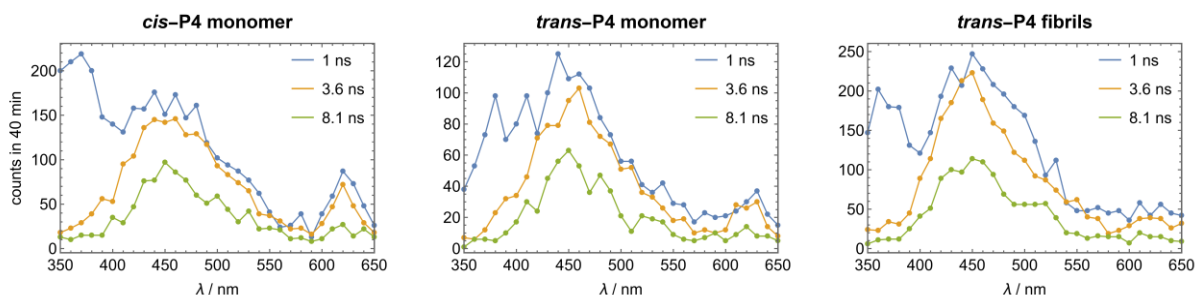
1. Supplementary figures.....	2
2. Supplementary tables .....	19
3. Peptide characterization after synthesis.....	21



# 1. Supplementary figures

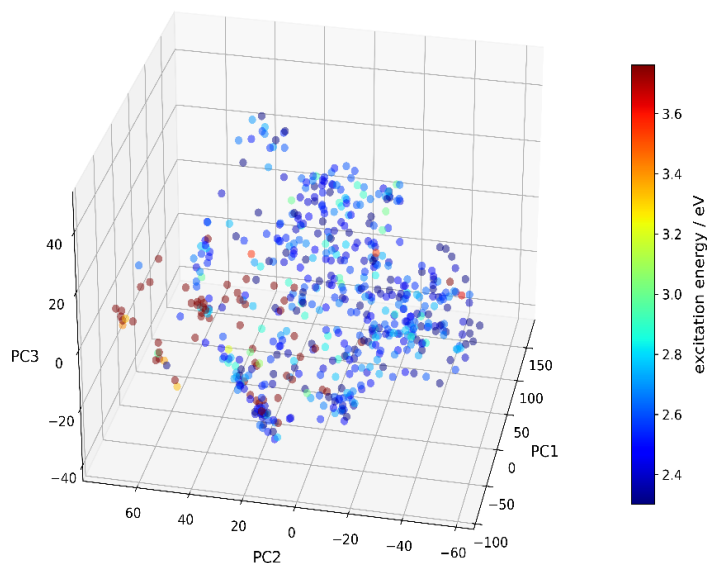


**Figure S1.** The smoothed potential energy curves from 34 converged *cis* ↔ *trans* isomerization paths. Single-point calculations are performed using NEVPT2(2,2)/xTB with implicit water through ALPB. The smoothing is carried out using Gaussian process regression. A subset of the paths has an energy barrier in the *trans*-*cis*-isomerization, notably paths 04, 10 and 21, with excitation energies before the barrier range from 2.5 to 3.0 eV. This correlates with the experimental fluorescence peak at 485 nm (**Figure S2**), albeit with an expected ipsochromic effect caused by the minimum active space considered in the calculations. We attribute the extended fluorescence lifetime, compared to azobenzene, to structures hindered by this barrier, preventing radiationless relaxation. These barriers predominantly arise from structures characterized by dihedral angles around 170°, which exhibit a heightened S0–S1 excitation energy in the scan.

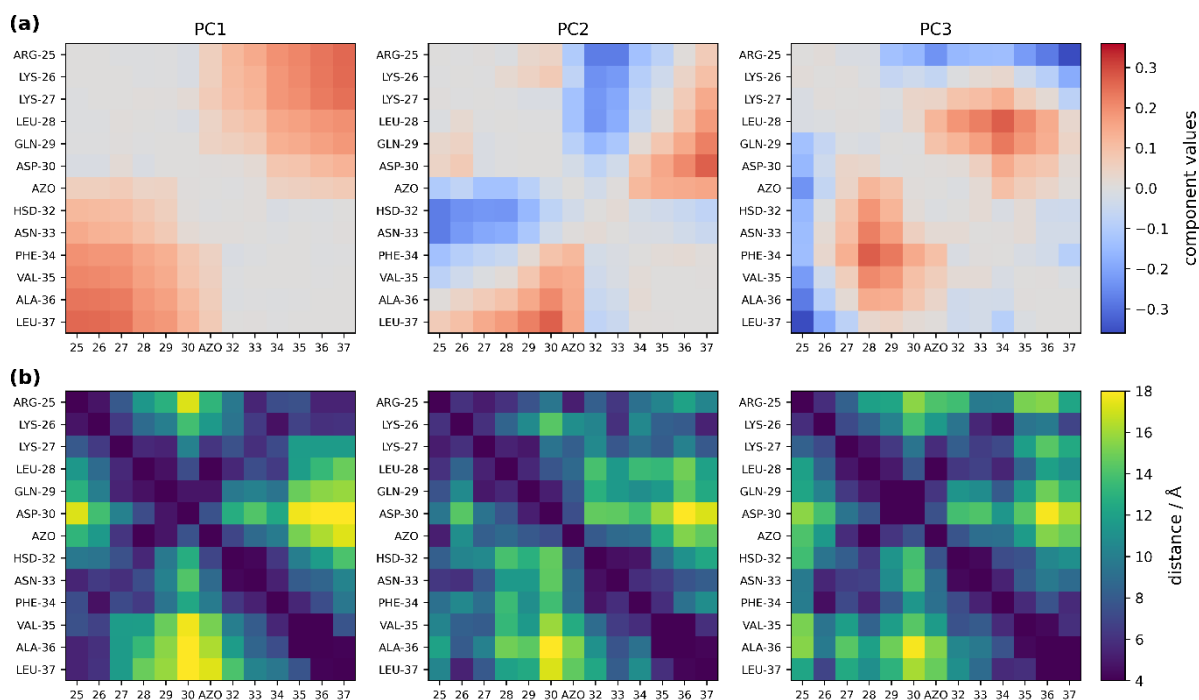


**Figure S2.** Fluorescence life-time measurements. Fluorescence spectra measured at various points in time after the excitation pulse.

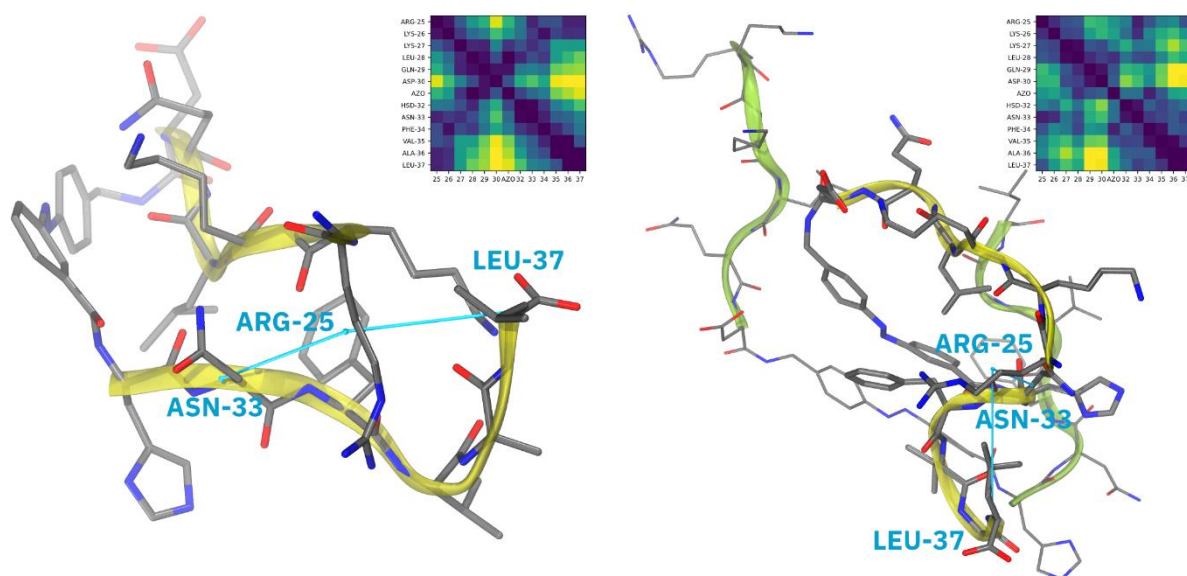




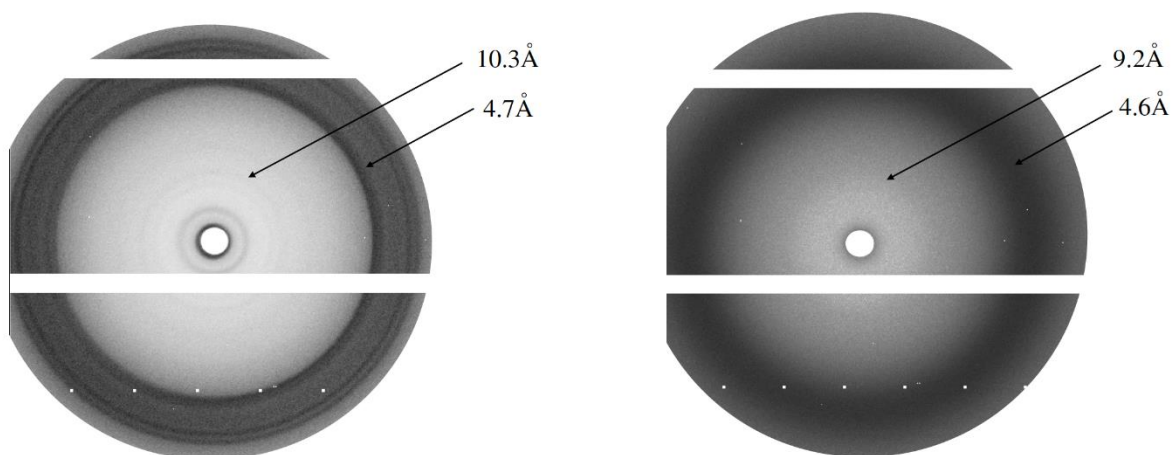
**Figure S3.** All geometries with dihedral angles near  $170^\circ$  projected on to the three-dimensional PCA space. The colors encode the S0–S1 excitation energies. The reduced three-dimensional space accounts for 86% of the total variance. Several clusters can be detected, notably one predominantly comprising structures with high excitation energies (visually represented in dark red in the lower left corner of the projection). This particular cluster is positioned at the negative extremity of the PC1-axis, advances towards the positive end of the PC2-axis and is situated approximately mid-way along the PC3-axis.



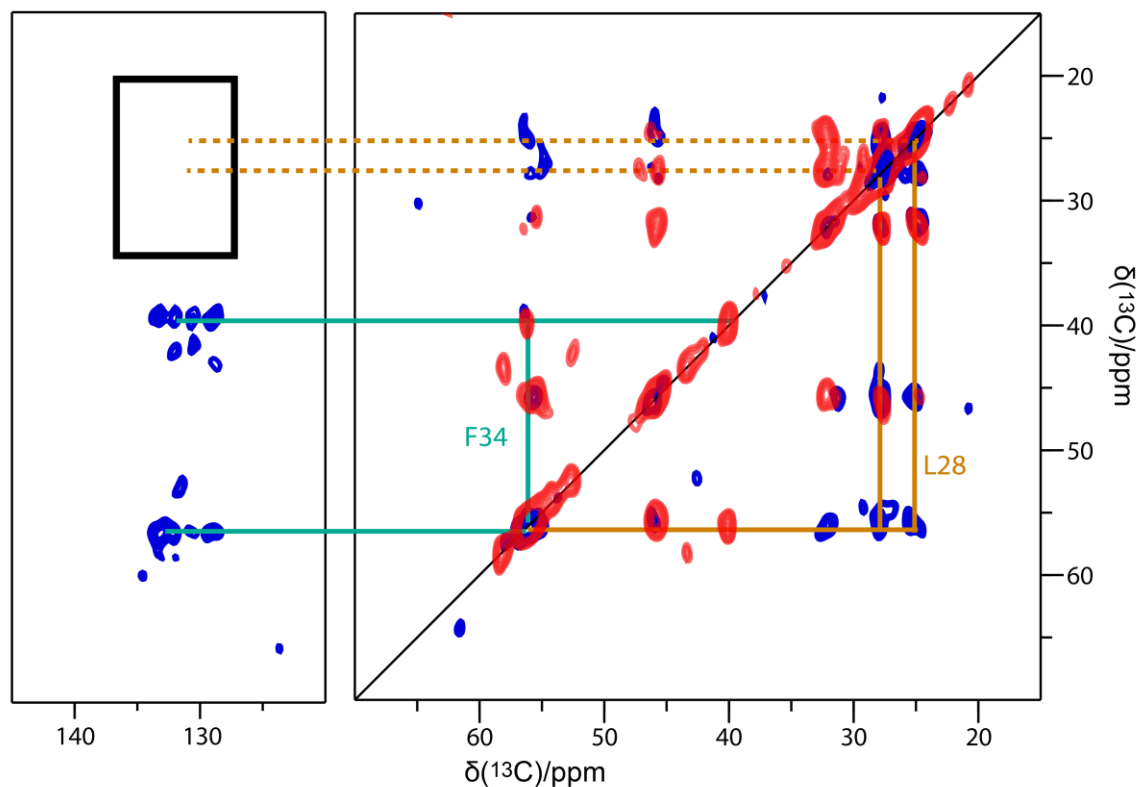
**Figure S4.** (a) The first three principal components represented in the distance matrix space. The first principal component signals a considerable separation between the residues R25 and A36/L37. Given this cluster's positioning at the PC1-axis's negative end, it implies that within these geometries, the distances between these residues are notably reduced. The second principal component reveals negative values for the separation between R25 and H32/N33, alongside positive distances between D30 and L37. The positive projection of this cluster on the PC2 suggests these distance matrix characteristics directly, without reversing the sign. With almost negligible projection on PC3, this component scarcely influences the distance matrices for structures within this cluster. (b) Distance matrices for clusters with high excitation energies. Despite originating from different paths, these matrices exhibit consistent patterns.



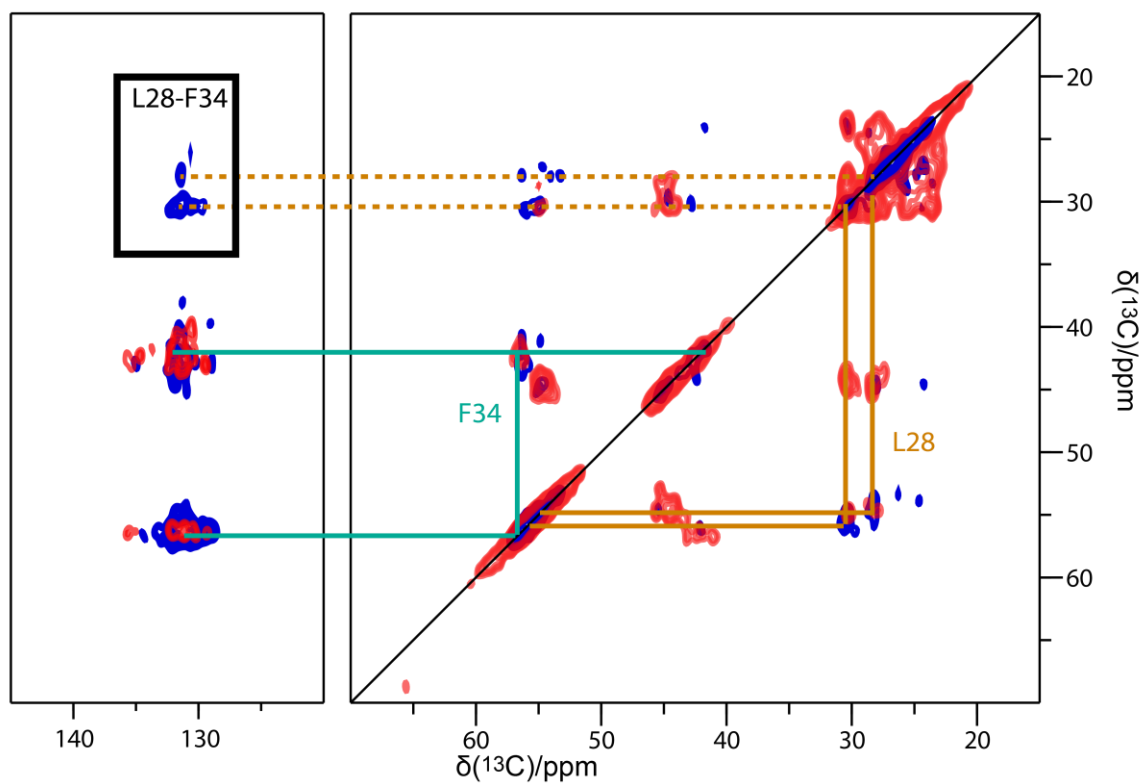
**Figure S5.** Selected monomer structure (left) with distance matrix shown in Figure S4b and dimer structure (right) with a similar distance matrix. The corresponding distance matrices are shown in the upper right corners. It can be seen that the simultaneous interactions of R25 with A36/L37 and H32/N33 potentially contribute to the S1-barrier during the *trans*→*cis* isomerization path, leading to longer fluorescence lifetimes. This structural motif is also found in the dimer, while their configurations stabilized by intermolecular interactions could also lead to an increased barrier.



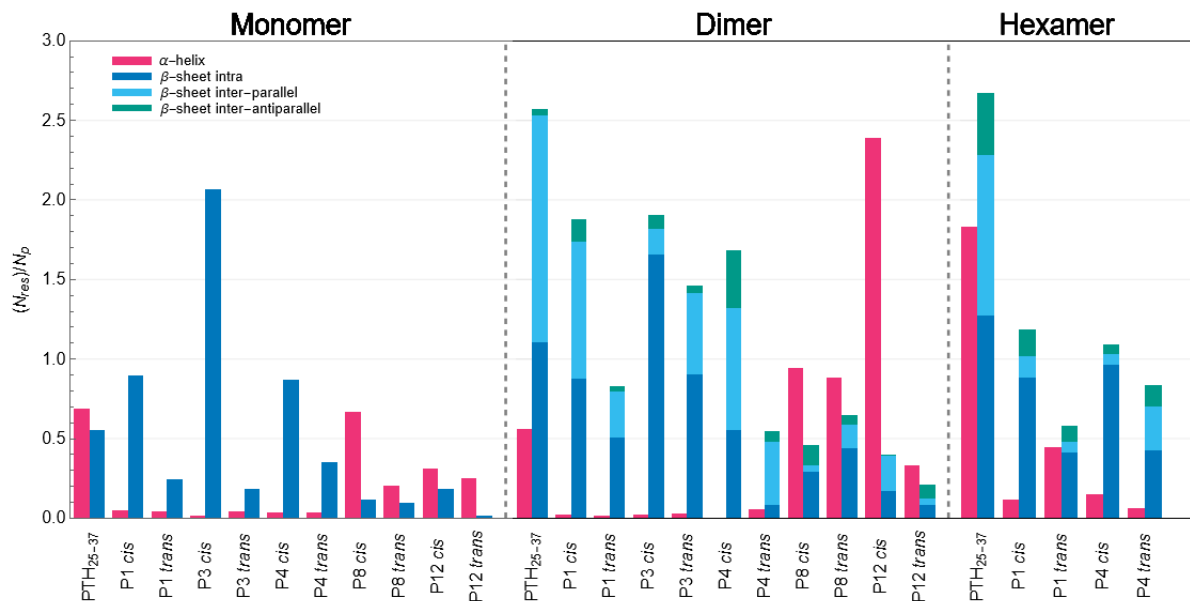
**Figure S6:** Wide-angle x-ray diffraction patterns from dried fibrillar samples of PTH<sub>25-37</sub> (left) and P4 (right).



**Figure S7.** 2D  $^{13}\text{C}$ ,  $^{13}\text{C}$  spin diffusion spectra of fibrils from PTH<sub>25-37</sub>, uniformly  $^{13}\text{C}$  labeled for L28 and F34, recorded under different conditions. Red: magnetic field strength 14.1 T (corresponding to 600 MHz  $^1\text{H}$  resonance frequency), spinning speed 11 kHz, mixing time 10 ms. Blue: magnetic fields strength of 18.8 T (corresponding to 800 MHz  $^1\text{H}$  resonance frequency), spinning frequency of 20 kHz, mixing time 1 s. Figure S6 shows the 2D  $^{13}\text{C}$ ,  $^{13}\text{C}$  spin diffusion spectra of fibrils from PTH<sub>25-37</sub>, uniformly  $^{13}\text{C}$  labeled for L28 and F34, recorded under different conditions. In the red spectrum, intraresidual cross-peaks between neighboring  $^{13}\text{C}$  sites of L28 and F34, respectively, are visible. At a longer mixing time of 1 s (blue), long-range correlations between all  $^{13}\text{C}$  spins within one residue are obtained. Spin systems of the labeled amino acids F34 (cyan) and L28 (brown) are marked by solid lines. Dashed lines show a possible contact between L28  $\text{C}_\gamma/\text{C}_\delta$  and an aromatic F34 carbon. The spinning speed of 20 kHz corresponds to the first order rotational resonance condition for resonances with a chemical shift difference of 100 ppm, leading to a recoupling of dipolar couplings between aromatic ring carbon atoms of F34 and aliphatic  $\text{C}_\gamma$  and  $\text{C}_\delta$  signals of L28, facilitating magnetization transfer between those residues if the distance between these residues would not exceed 6 Å.<sup>20</sup> The fact that no inter-residual cross-peaks between L28 and F34 can be observed is thus a strong indication against an antiparallel arrangement of  $\beta$ -strands within the  $\beta$ -sheet.



**Figure S8.** 2D  $^{13}\text{C}$ ,  $^{13}\text{C}$  spin diffusion spectra of fibrils from P4, uniformly  $^{13}\text{C}$  labeled for L28 and F34, recorded at a magnetic field strength of 18.8 T (corresponding to 800 MHz  $^1\text{H}$  resonance frequency) at a spinning speed of 20 kHz, corresponding to the first order rotational resonance condition for signals with a chemical shift difference of 100 ppm. Red: Mixing time of 50 ms. Blue: Mixing time of 1 s. See for more information in Fig. S6. Here, inter-residual cross-peaks between the aromatic ring signals of F and L C $\delta$  resonances are clearly visible for a mixing time of 1 s (blue spectrum).

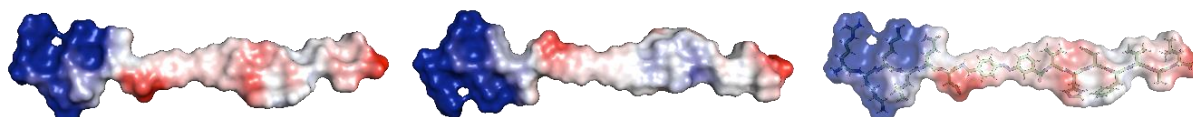


**Figure S9.** Summary of all MD simulation results. The average numbers of residues  $\langle N_{res} \rangle$  forming secondary structure elements, divided into  $\alpha$ -helical (magenta), intrapeptide  $\beta$ -sheets (blue), interpeptide parallel  $\beta$ -sheets (cyan) and interpeptide antiparallel  $\beta$ -sheets (green) are given. The average was taken over the number of frames of the simulation and normalized by the number of peptides  $N_P$  present in the corresponding simulation.

PTH<sub>25-37</sub>



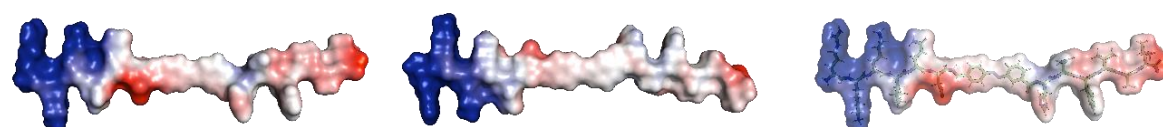
P1



P3



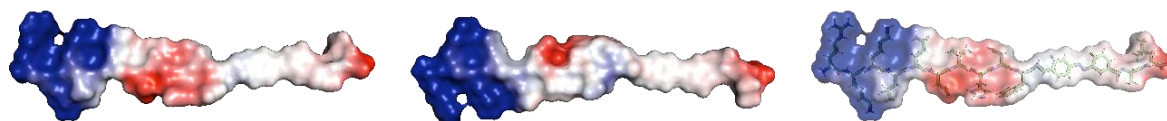
P4



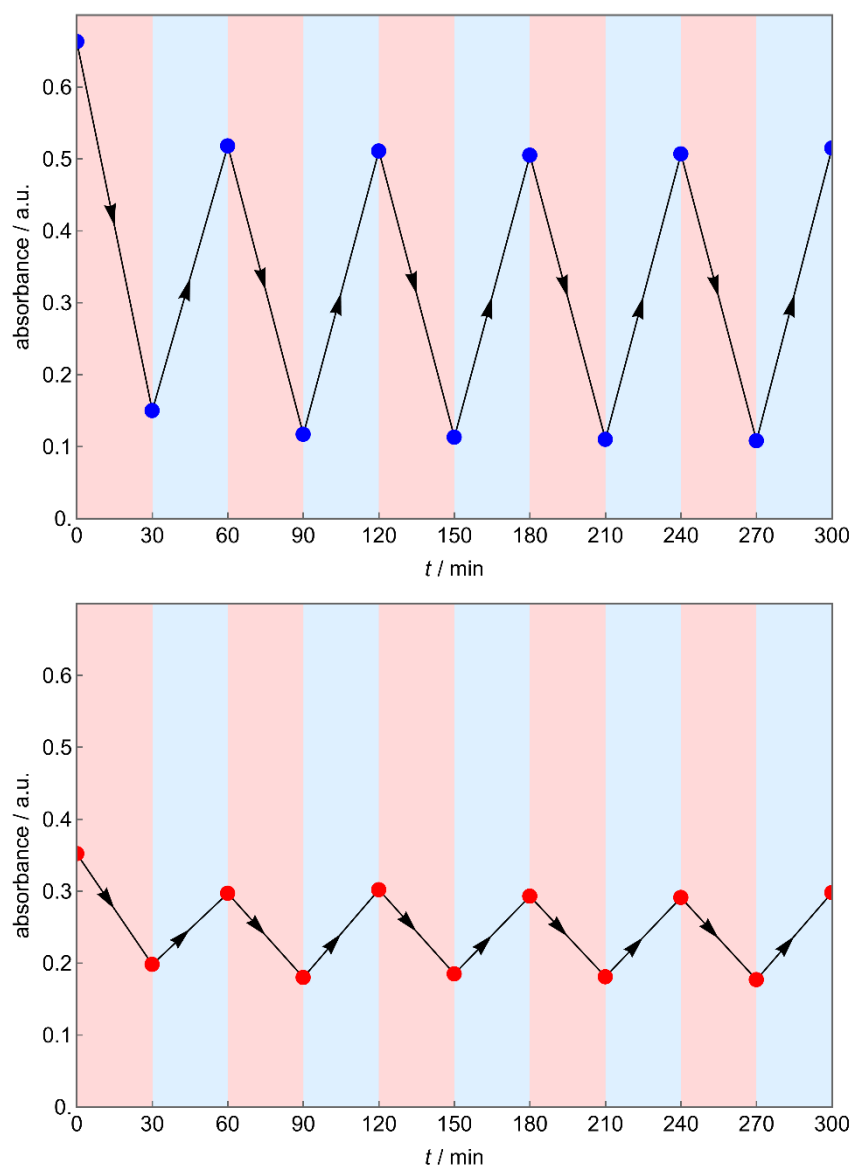
P8



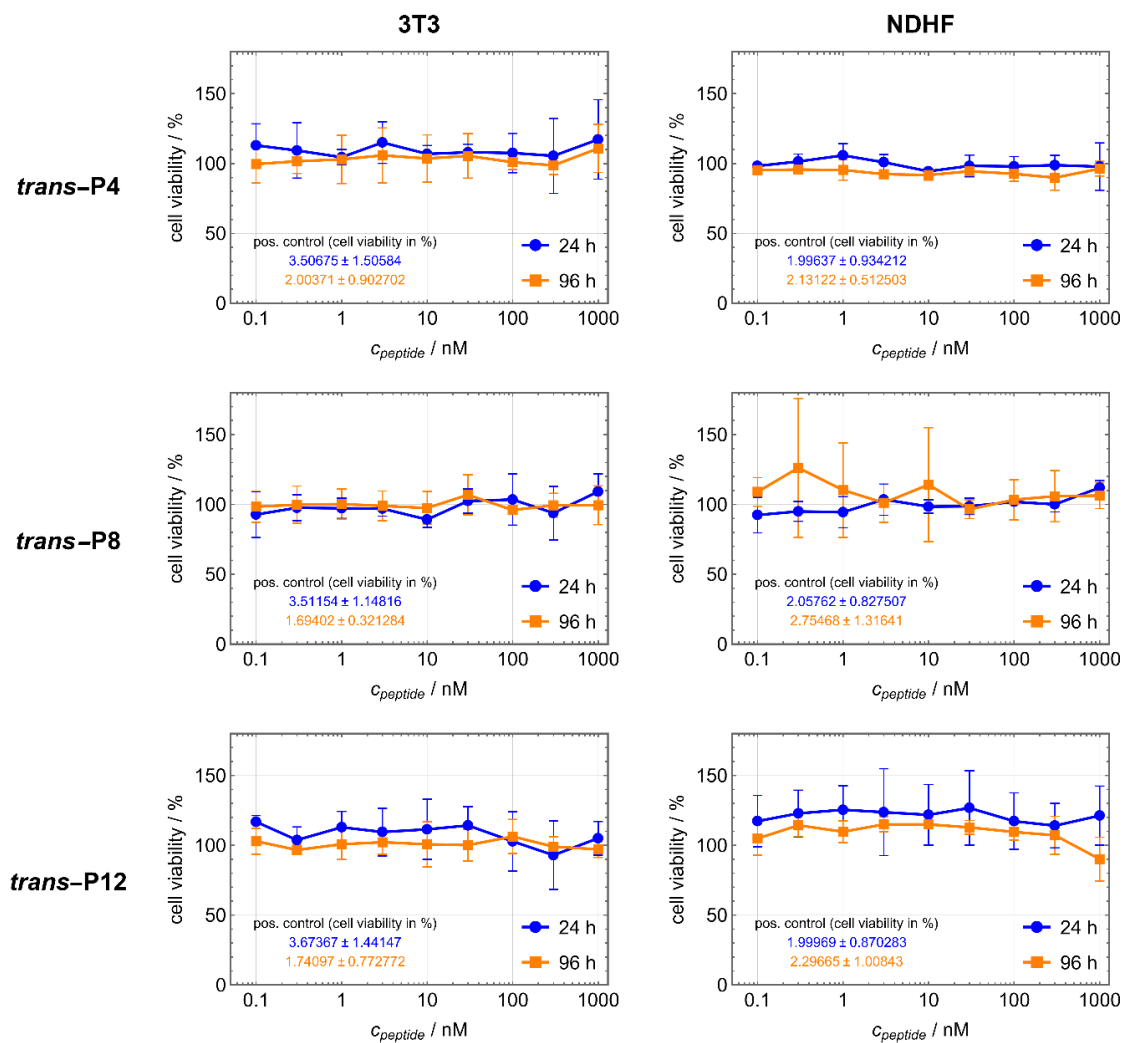
P12



**Figure S10.** Electrostatic potential surface of PTH<sub>25-37</sub> and the peptides P1, P3, P4, P8, and P12, with values according to the color map at the bottom, ranging from -2 (red) to +2 kTe<sup>-1</sup> (blue). The electrostatic potential mapped to the molecular surfaces was calculated using the Adaptive Poisson-Boltzmann Solver (APBS<sup>1</sup>) plugin for the pymol<sup>2</sup> software package. For each peptide, the two views that are rotated by 180° around the backbone axis are shown, as well as a transparent front view.

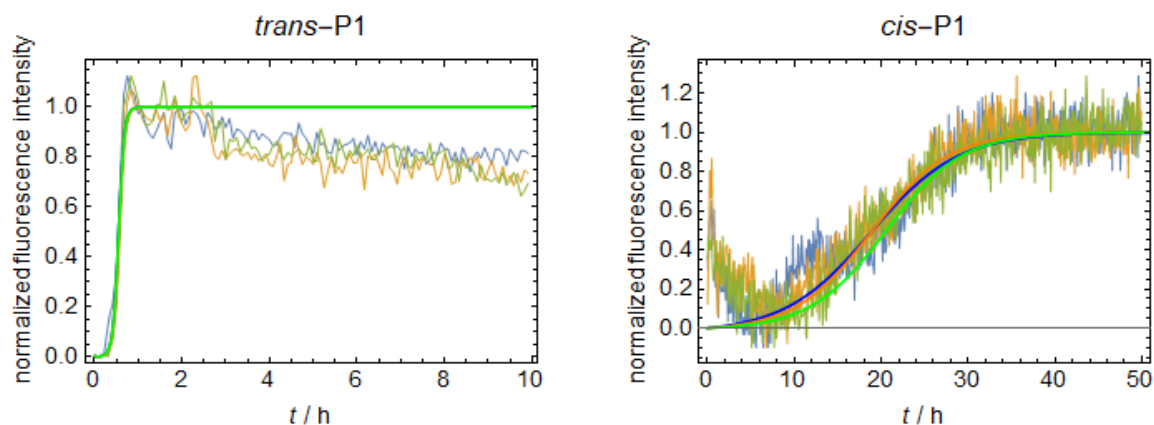


**Figure S11:** Control experiment to investigate photobleaching of the azobenzene unit during the photoisomerization. Absorption of P4 was measured after each isomerization step at 327 nm (top, blue dots) and 295 nm (bottom, red dots). Light red area corresponds to irradiation with light of 340 nm wavelength for 30 min to achieve *trans*→*cis* isomerization. Light blue area corresponds to irradiation with light of 405 nm wavelength for 30 min to achieve *cis*→*trans* isomerization.

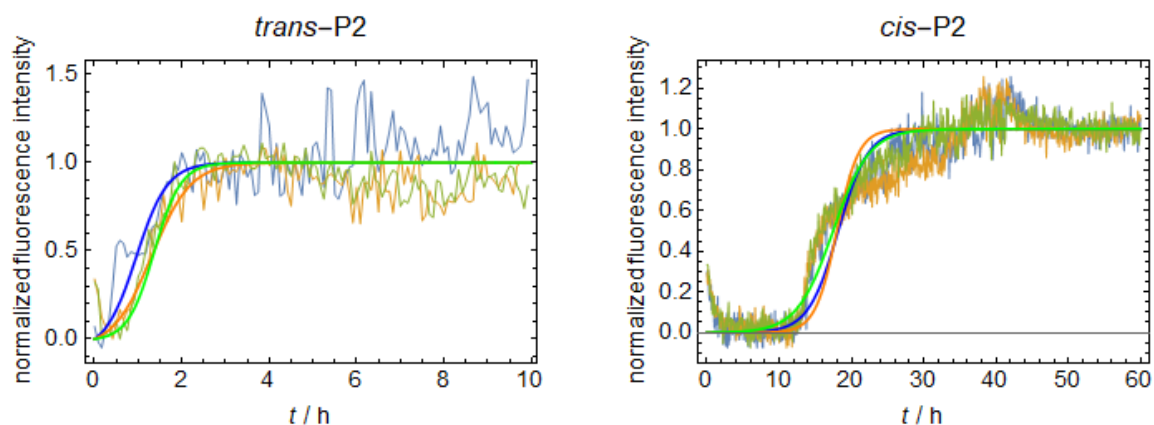


**Figure S12:** Cell viability assay of *trans*-P4, *trans*-P8, and *trans*-P12 on NDHF (human, adult, fibroblasts) and 3T3 (murine, embryonal, fibroblasts) cells after 24 h (blue) and 96 h (orange). 100% cell viability corresponds to the value of the negative, untreated control.

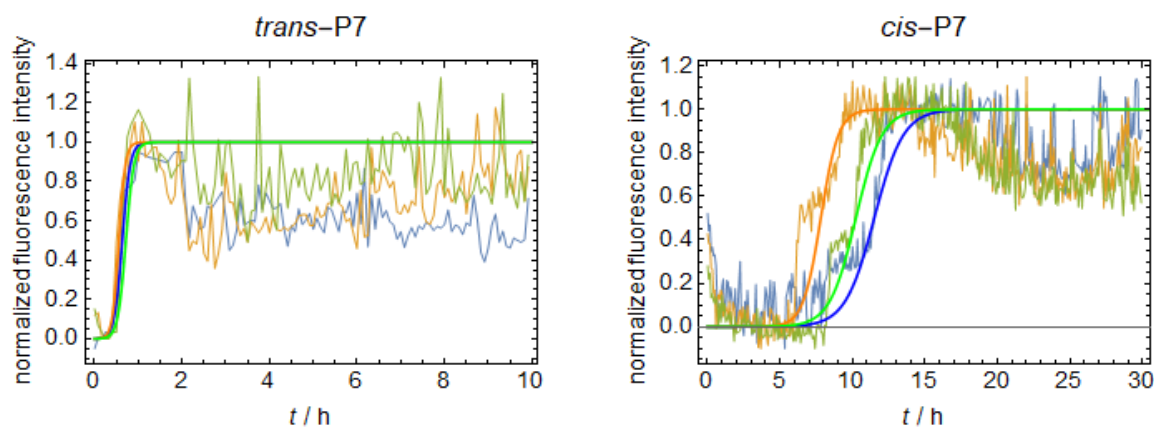




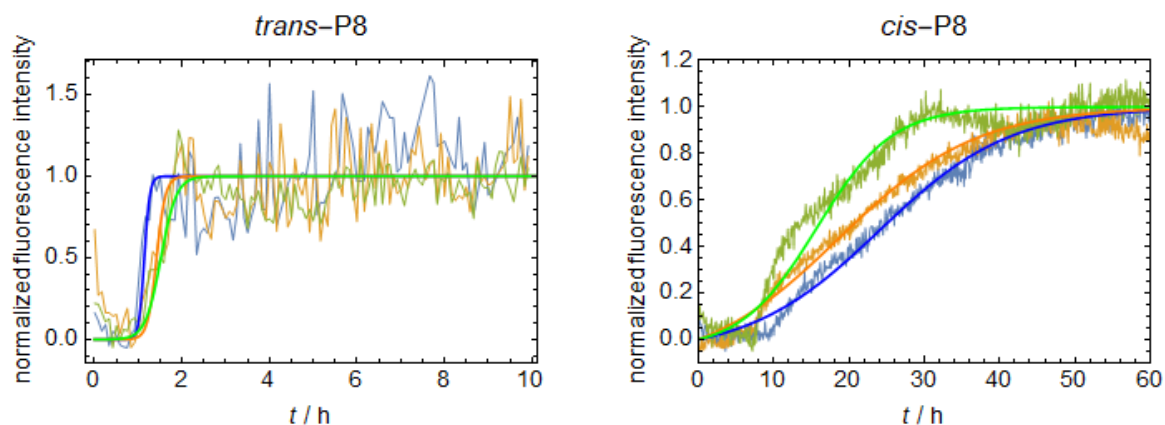
**Figure S13.** Fibrillization kinetics of 85  $\mu\text{M}$  P1 as *trans*- (left) and *cis*-isomer (right). Measured in a ThT-monitored fluorescence assay at 37  $^{\circ}\text{C}$  in 50  $\mu\text{M}$   $\text{Na}_2\text{HPO}_4$  buffered aqueous solution (pH 7.4).



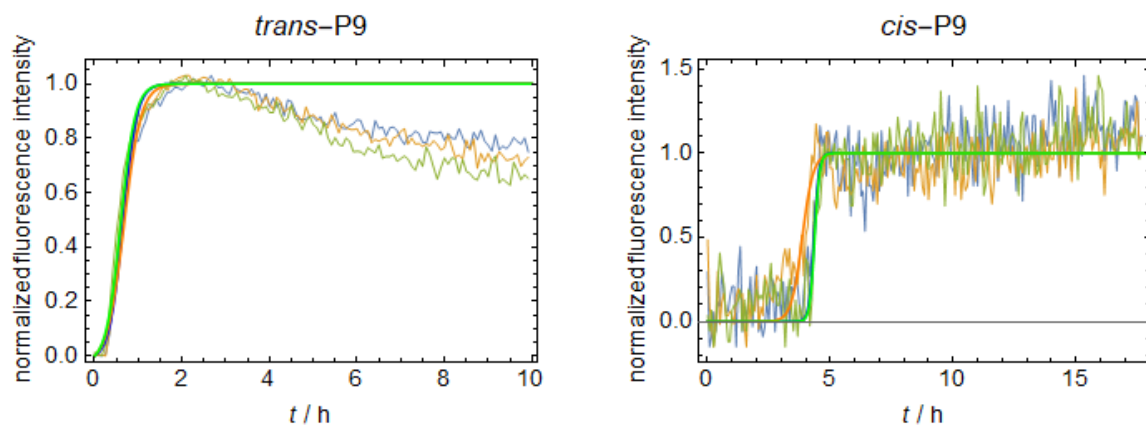
**Figure S14.** Fibrillization kinetics of 55  $\mu\text{M}$  P2 as *trans*- (left) and *cis*-isomer (right). Measured in a ThT-monitored fluorescence assay at 37  $^{\circ}\text{C}$  in 50  $\mu\text{M}$   $\text{Na}_2\text{HPO}_4$  buffered aqueous solution (pH 7.4).



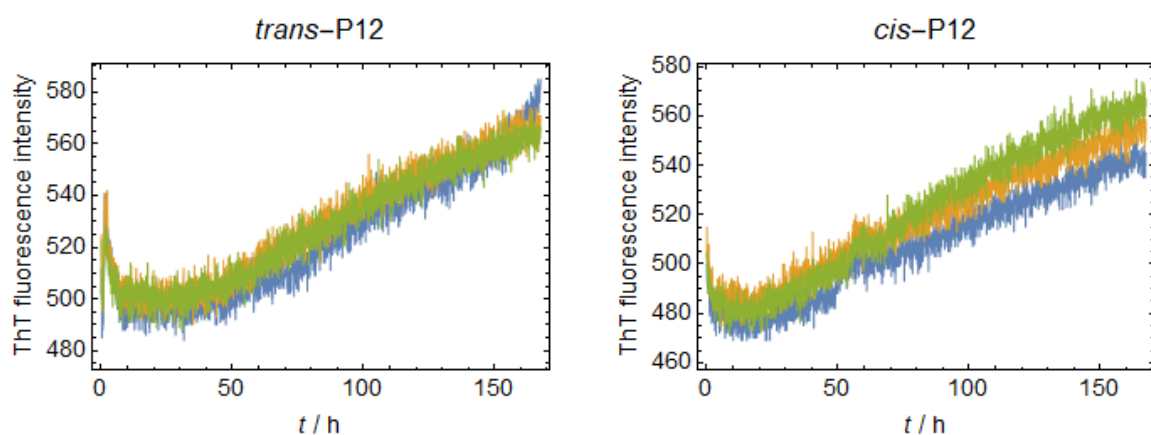
**Figure S15.** Fibrillization kinetics of 100  $\mu\text{M}$  P7 as *trans*- (left) and *cis*-isomer (right). Measured in a ThT-monitored fluorescence assay at 37  $^{\circ}\text{C}$  in 50  $\mu\text{M}$   $\text{Na}_2\text{HPO}_4$  buffered aqueous solution (pH 7.4).



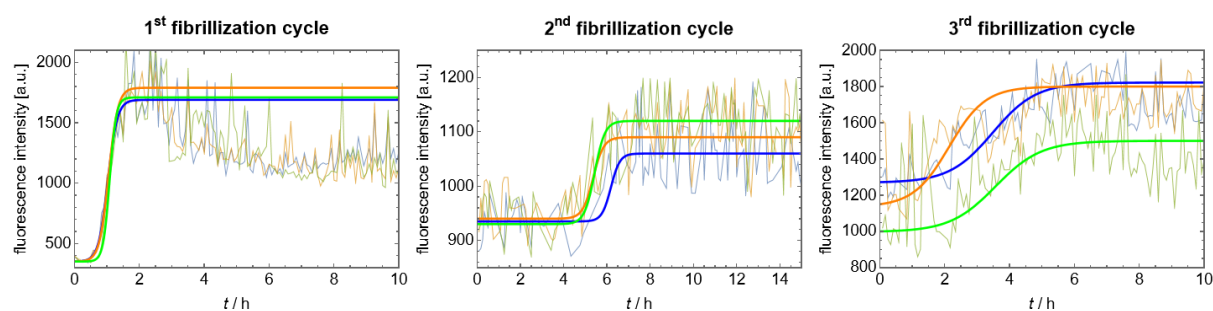
**Figure S16.** Fibrillization kinetic of 100  $\mu\text{M}$  P8 as *trans*- (left) and *cis*-isomer (right). Measured in a ThT-monitored fluorescence assay at 37  $^{\circ}\text{C}$  in 50  $\mu\text{M}$   $\text{Na}_2\text{HPO}_4$  buffered aqueous solution (pH 7.4).



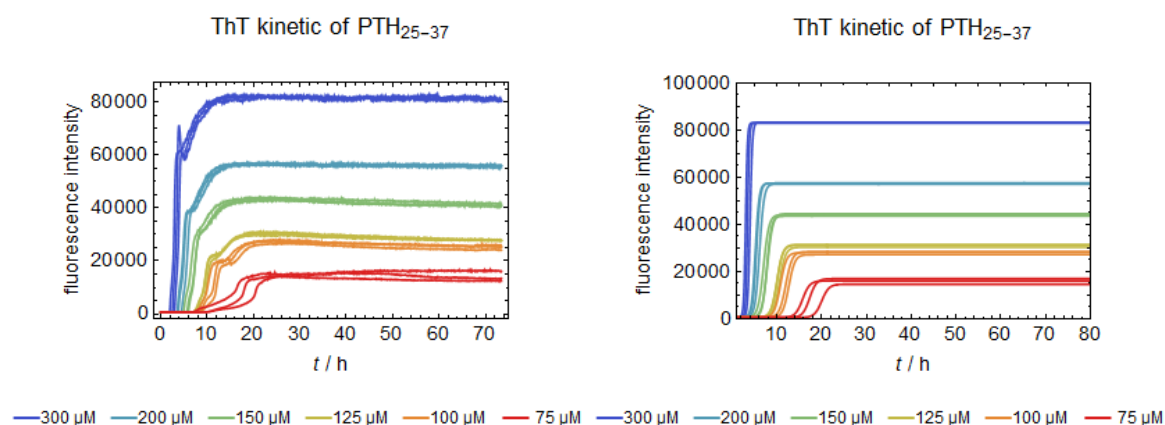
**Figure S17.** Fibrillization kinetics of 100  $\mu\text{M}$  P9 as *trans*- (left) and *cis*-isomer (right). Measured in a ThT-monitored fluorescence assay at 37  $^{\circ}\text{C}$  in 50  $\mu\text{M}$   $\text{Na}_2\text{HPO}_4$  buffered aqueous solution (pH 7.4).



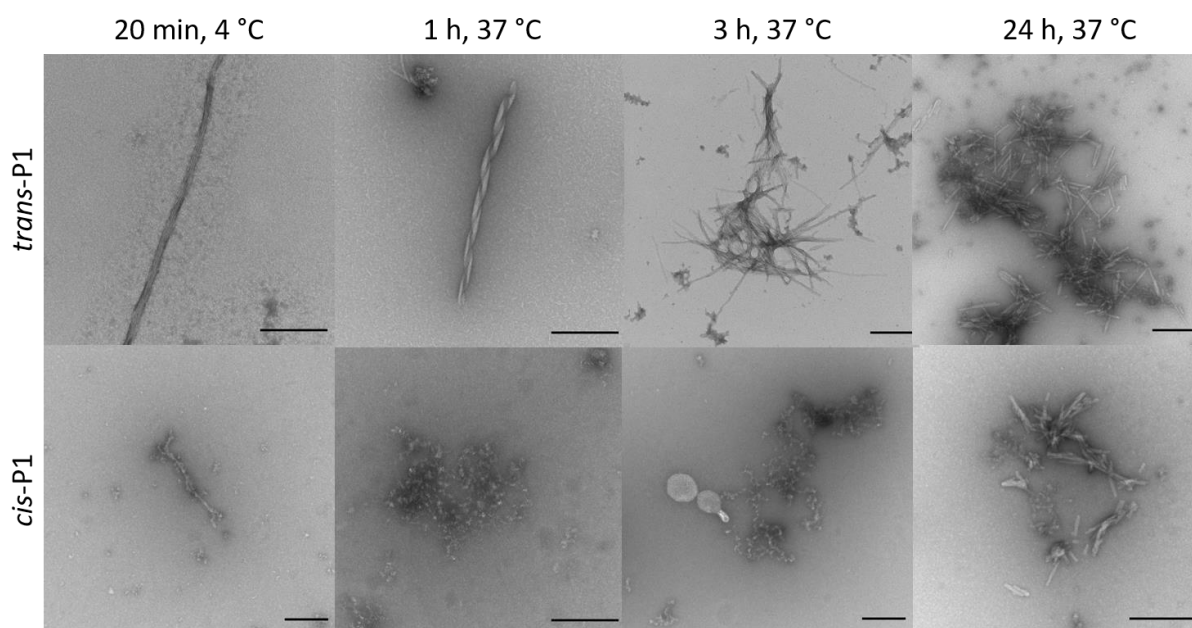
**Figure S18.** Fibrillization kinetics of 100  $\mu\text{M}$  P12 as *trans*- (left) and *cis*-isomer (right). Measured in a ThT-monitored fluorescence assay at 37  $^{\circ}\text{C}$  in 50  $\mu\text{M}$   $\text{Na}_2\text{HPO}_4$  buffered aqueous solution (pH 7.4).



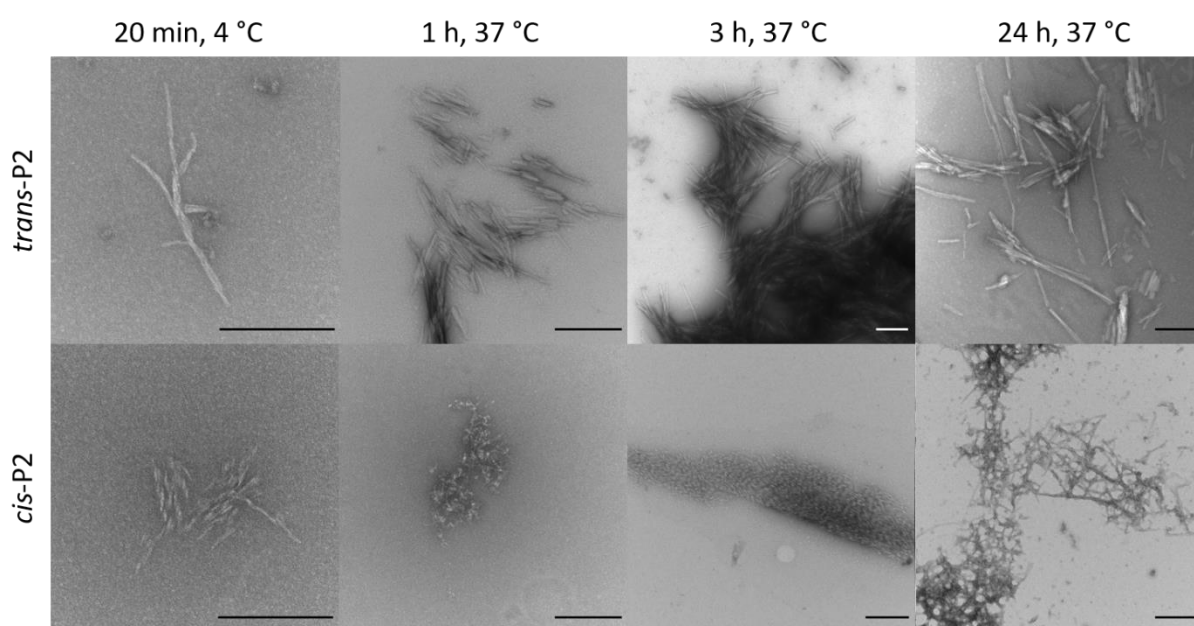
**Figure S19.** Fibrillization kinetics of P4 over three cycles of alternating fibrillization and fibril degradation through *trans*→*cis* isomerization measured in a ThT-monitored fibrillization assay at 37 °C in 50  $\mu$ M  $\text{Na}_2\text{HPO}_4$  buffered aqueous solution (pH 7.4).



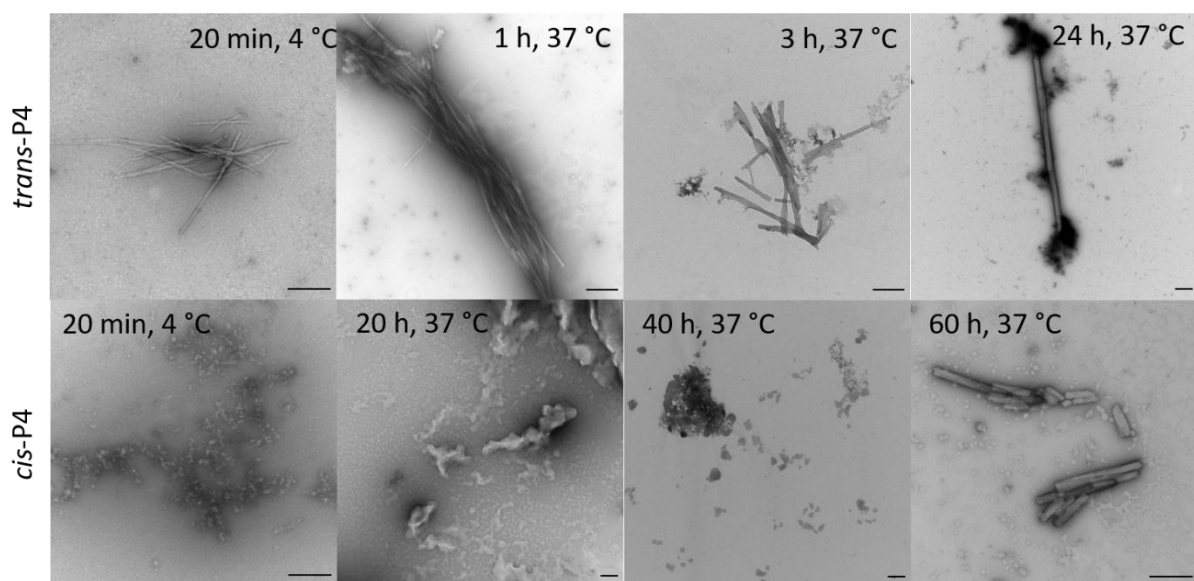
**Figure S20.** Fibrillization kinetics of PTH<sub>25-37</sub> in different concentrations measured in a ThT-monitored fluorescence assay at 37 °C in 50  $\mu$ M  $\text{Na}_2\text{HPO}_4$  buffered aqueous solution (pH 7.4); (left) raw data, (right) fitted data using Eq. (1).



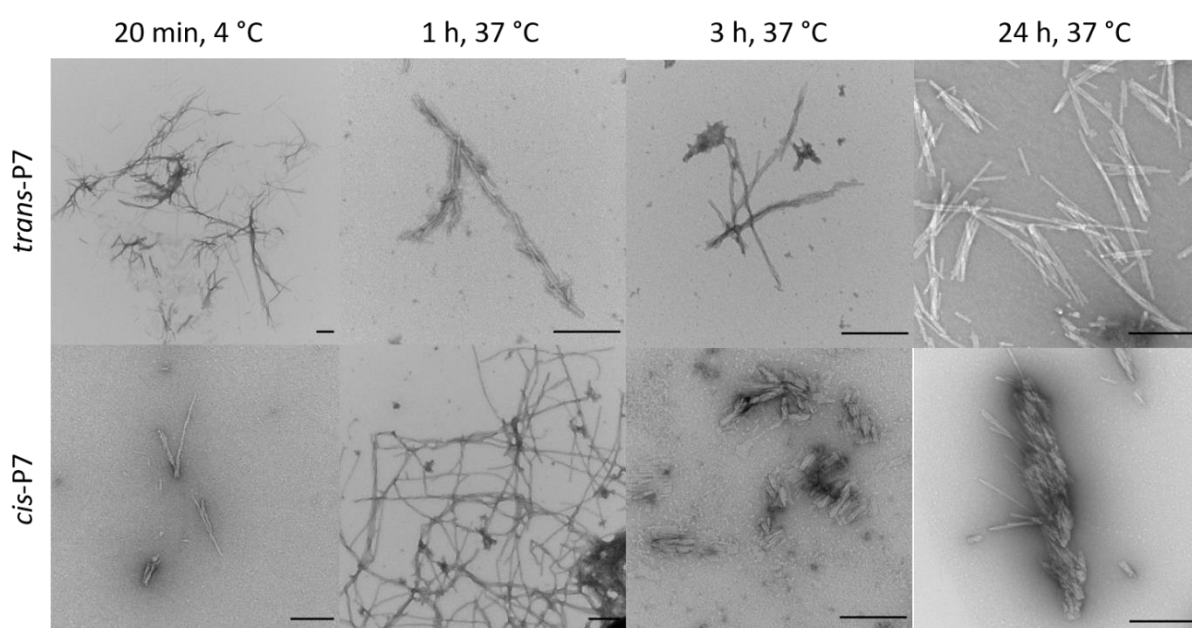
**Figure S21.** TEM images of P1 at different times and temperatures; scale bar = 250 nm.



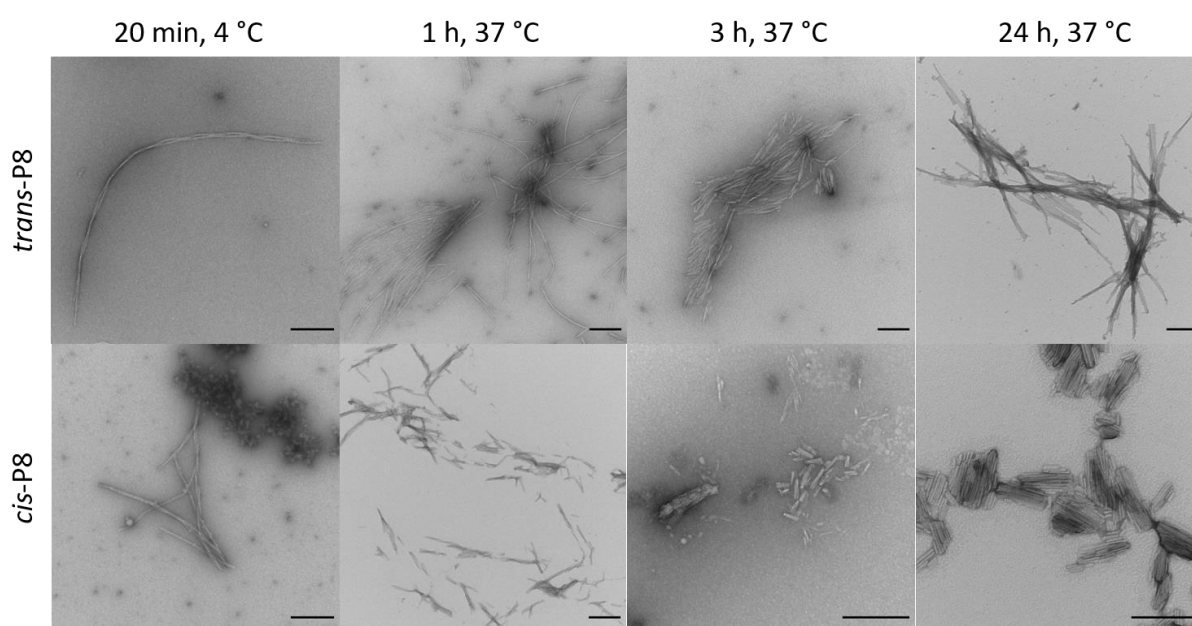
**Figure S22.** TEM images of P2 at different times and temperatures; scale bar = 250 nm.



**Figure S23.** TEM images of P4 at different times and temperatures; scale bar = 250 nm.

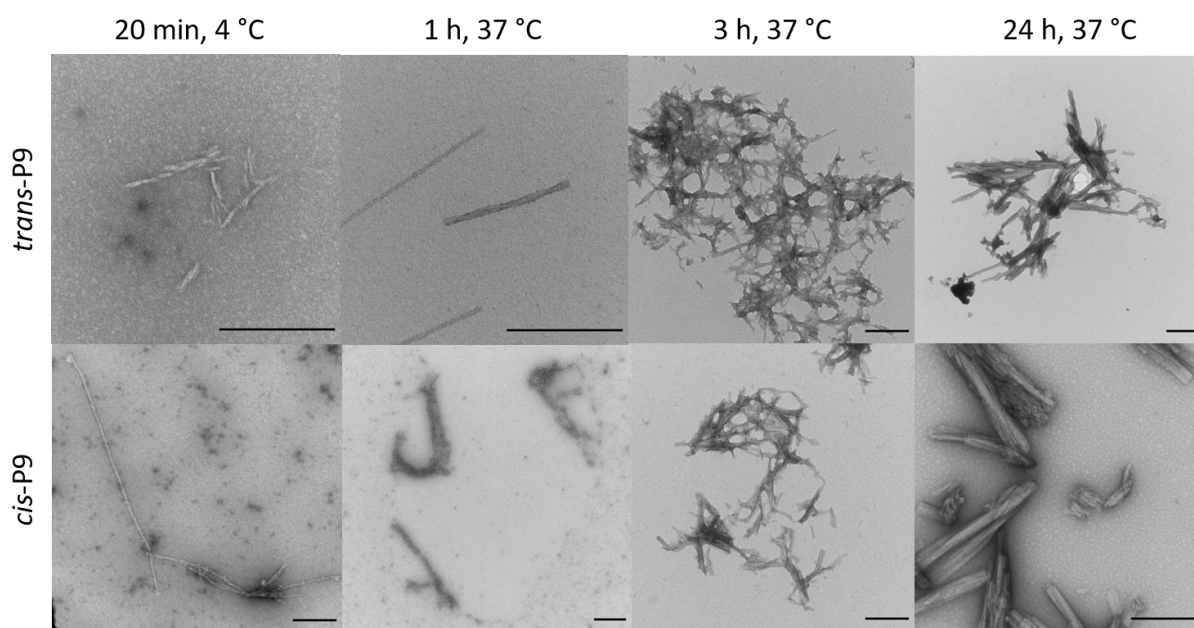


**Figure S24.** TEM images of P7 at different times and temperatures; scale bar = 250 nm.

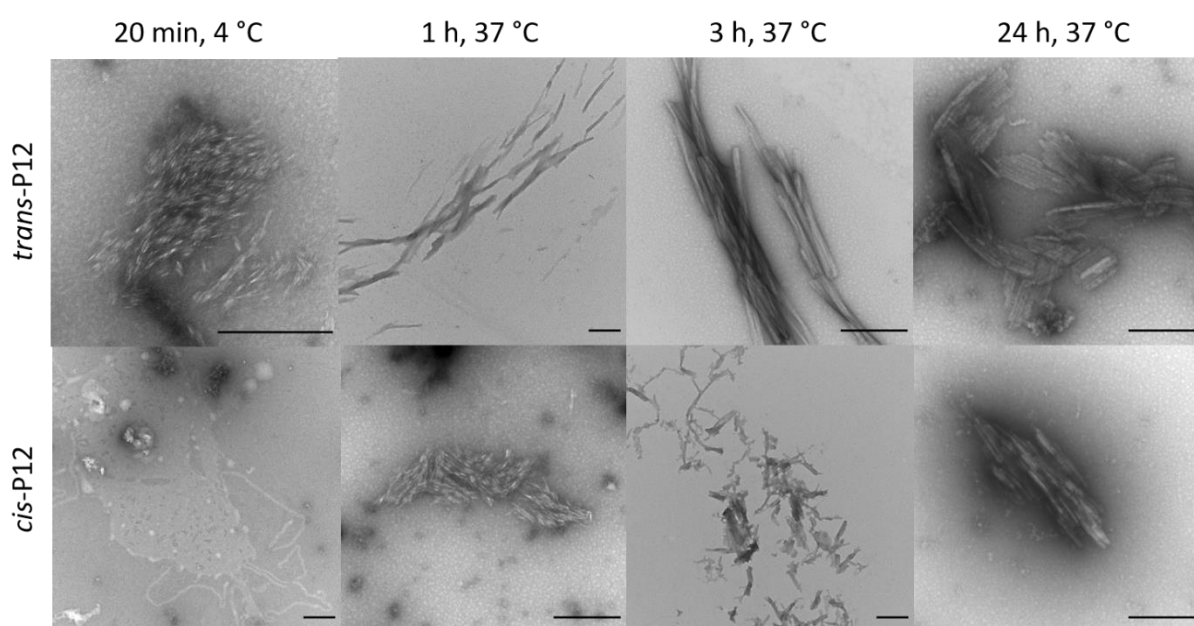


**Figure S25.** TEM images of P8 at different times and temperatures; scale bar = 250 nm.

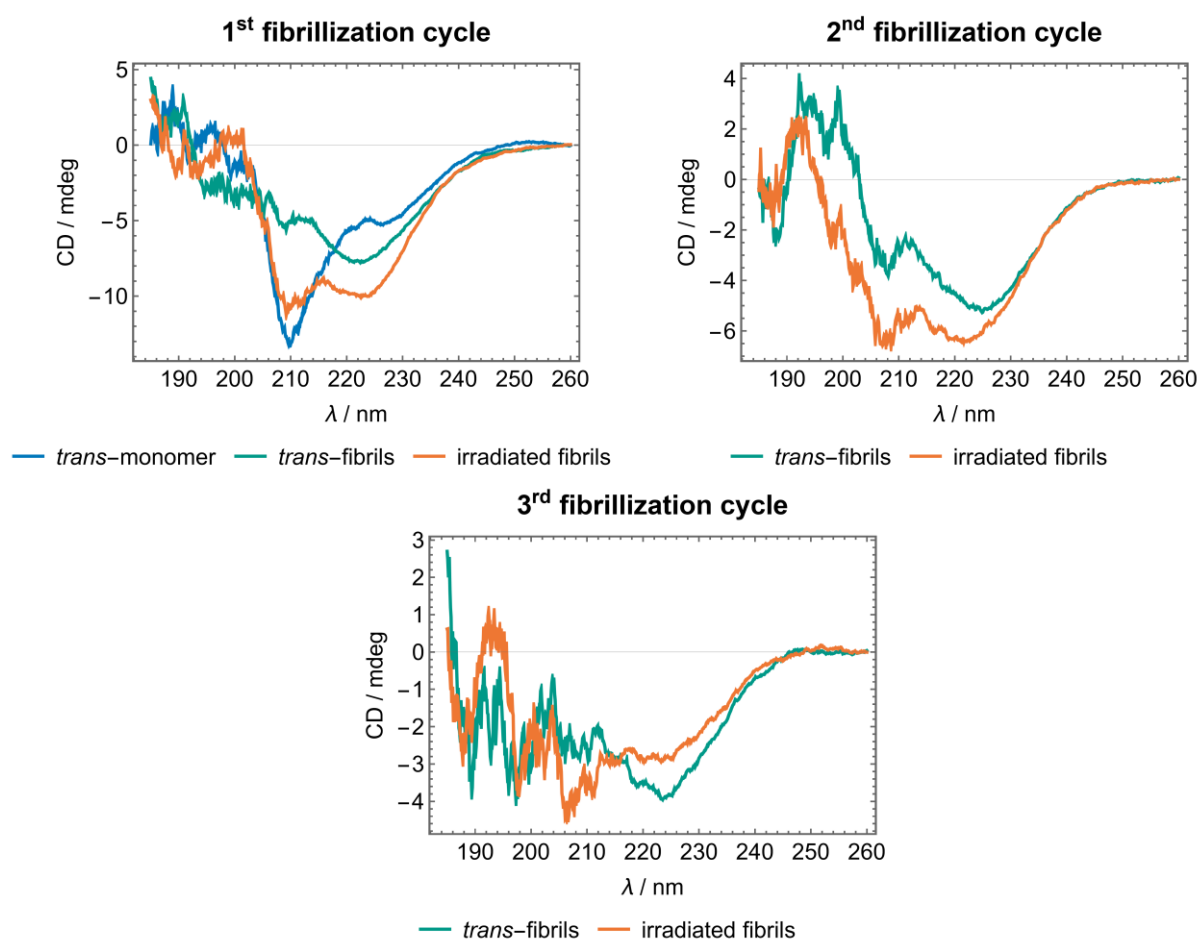




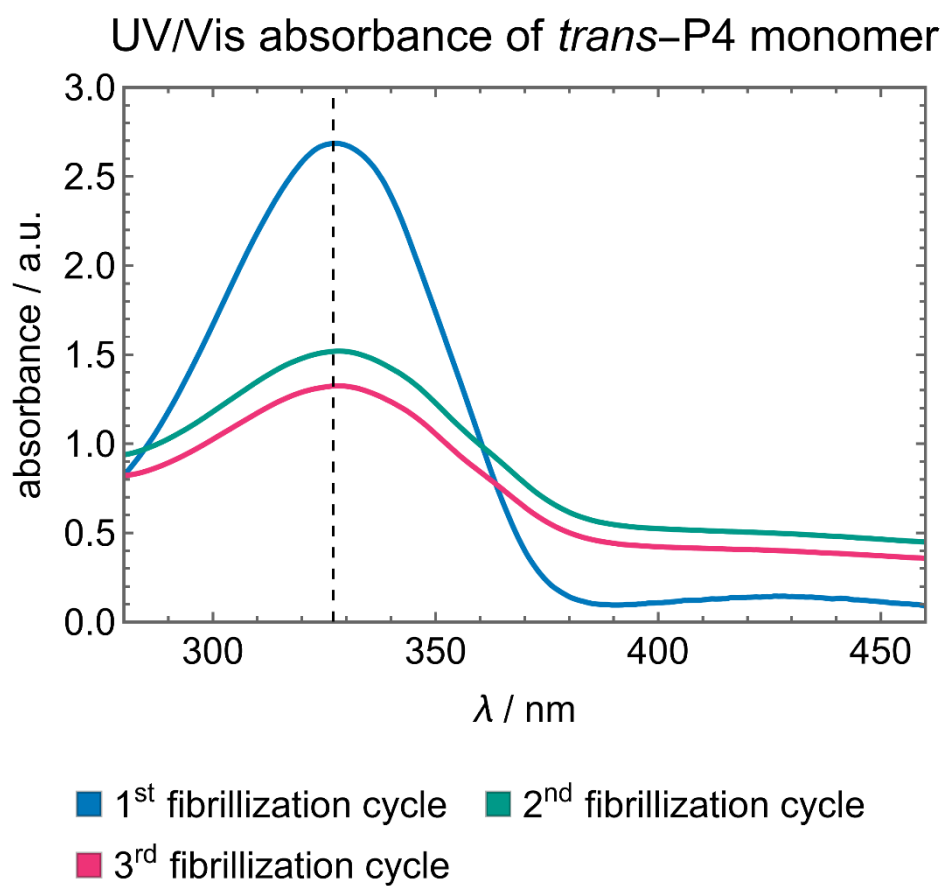
**Figure S26.** TEM images of P9 at different times and temperatures; scale bar = 250 nm.



**Figure S27.** TEM images of P12 at different times and temperatures; scale bar = 250 nm.



**Figure S28:** CD-spectra of the reversible fibrillization of P4 over three cycles (1<sup>st</sup>, 2<sup>nd</sup>, and 3<sup>rd</sup>). *trans*-Isomer (blue) was measured directly after dissolving the peptide, *trans*-fibrils (green) were measured 20 h after reference sample reached the stationary phase, and irradiated fibrils (orange) were measured directly after irradiating the fibrils with 340 nm for 5 h.



**Figure S29:** UV/Vis-spectra of *trans*-P4 monomer before each fibrillization cycle. Dashed line corresponds to the absorption maxima of the *trans*-isomer at 327 nm.



## 2. Supplementary tables

**Table S1.** All Simulations performed, with their respective simulation time

Peptide	Simulation	Simulation Time [ $\mu$ s]
PTH <sub>25-37</sub>	Monomer/Dimer/Hexamer	3 systems x 10
P1	Monomer/Dimer/Hexamer - <i>cis/trans</i>	6 systems x 10
P3	Monomer/Dimer - <i>cis/trans</i>	4 systems x 10
P4	Monomer/Dimer/Hexamer - <i>cis/trans</i>	6 systems x 10
P8	Monomer/Dimer - <i>cis/trans</i>	4 systems x 10
P12	Monomer/Dimer - <i>cis/trans</i>	4 systems x 10
PTH <sub>25-37</sub>	fibril	4 models x 1
P4	fibril	11 models x 1

**Table S2.** Sequence of PTH<sub>25-37</sub> and P1 – P12 and their respective solubility in buffered solution (50 mM aqueous Na<sub>2</sub>HPO<sub>4</sub>) with pH 7.4, the critical fibrillization concentration ( $c_{cr}$ ), and standard free energy of the fibrillization reaction  $\Delta G^0$ .

Peptide	Primary sequence	Solubility [ $\mu$ M]	$c_{cr}$ [ $\mu$ M]	$\Delta G^0$ [kJ/mol]
PTH <sub>25-37</sub>	<sup>25</sup> RKKLQ <sup>30</sup> DVHNF <sup>35</sup> VAL	>500	42	26
P1	<sup>25</sup> RKKLQ <sup>30</sup> D- <b>Azo</b> -VHNF <sup>35</sup> VAL	90	28	27
P2	<sup>25</sup> RKKLQ <sup>30</sup> DV- <b>Azo</b> -HNF <sup>35</sup> VAL	60	27	27
P3	<sup>25</sup> RKKLQ- <b>Azo</b> -VHNF <sup>35</sup> VAL	35	n.d.	n.d.
P4	<sup>25</sup> RKKLQ <sup>30</sup> D- <b>Azo</b> -HNF <sup>35</sup> VAL	370	23	28
P5	<sup>25</sup> RKKLQ- <b>Azo</b> -HNF <sup>35</sup> VAL	25	n.d.	n.d.
P6	<sup>25</sup> RKKLQ- <b>Azo</b> - <sup>30</sup> DVHNF <sup>35</sup> VAL	20	n.d.	n.d.
P7	<sup>25</sup> RKKL- <b>Azo</b> -Q <sup>30</sup> DVHNF <sup>35</sup> VAL	140	32	27
P8	<sup>25</sup> RK- <b>Azo</b> -KLQ <sup>30</sup> DVHNF <sup>35</sup> VAL	200	6	31
P9	<sup>25</sup> R- <b>Azo</b> -KLQ <sup>30</sup> DVHNF <sup>35</sup> VAL	130	26	27
P10	<sup>25</sup> RKKLQ <sup>30</sup> DVHN- <b>Azo</b> -F <sup>35</sup> VAL	7	n.d.	n.d.
P11	<sup>25</sup> RKKLQ <sup>30</sup> DVHNF <sup>35</sup> V- <b>Azo</b> -AL	20	n.d.	n.d.
P12	<sup>25</sup> RKKLQ <sup>30</sup> DVHNF- <b>Azo</b> -AL	135	-	-

n.d. – not determined

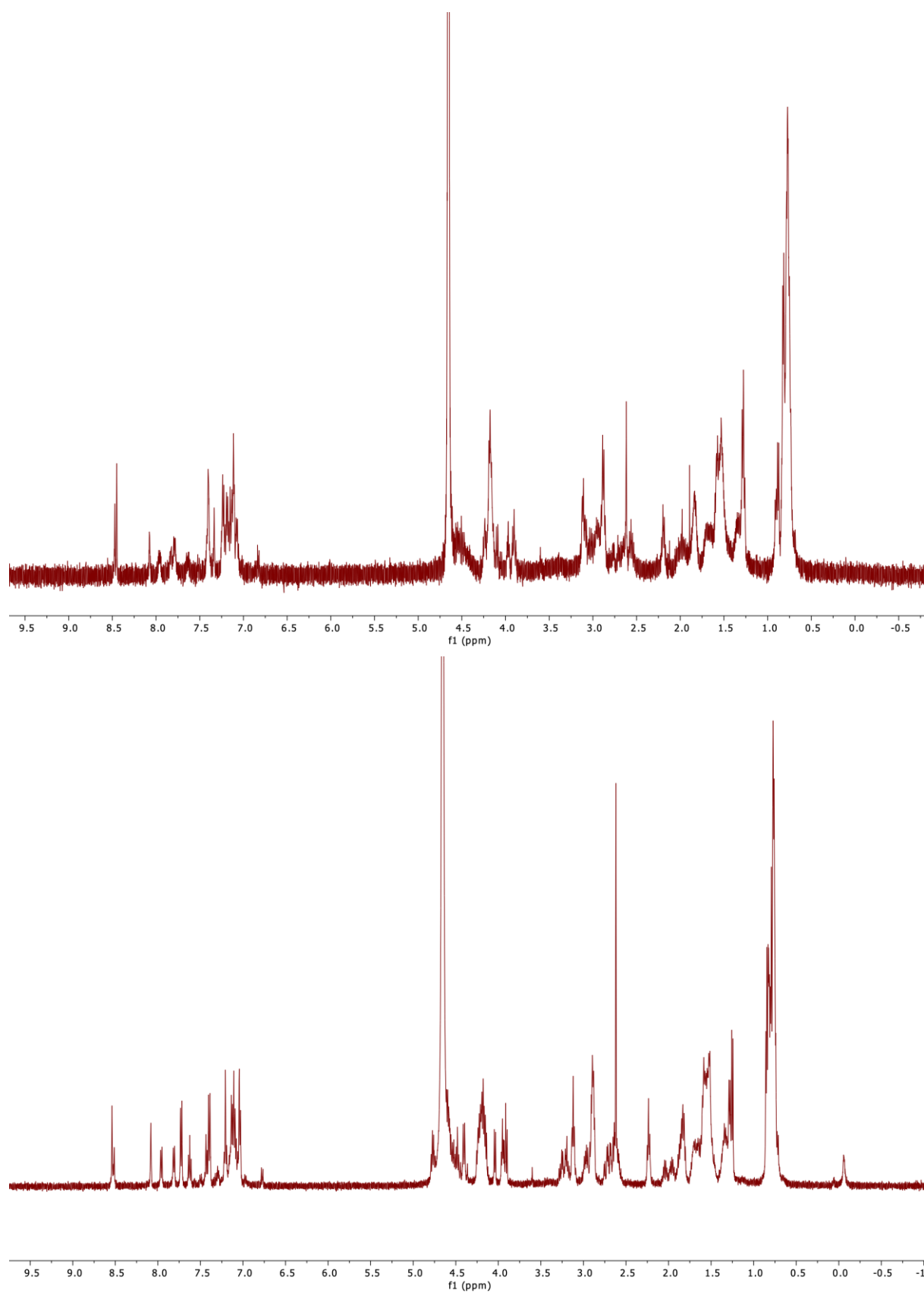
**Table S3.** Photophysical properties of P1, P2, P4, P7, P8, P9, P12 in buffered solution. *cis*-PSS – *cis*-photostationary state at 340 nm after 30 min. *trans*-PSS – *trans*-photostationary state at 405 nm after 20 min.

Peptide	$t_{1/2}$ ( <i>cis</i> , 37 °C) [h]	after synthesis	<i>cis</i> -PSS	<i>trans</i> -PSS
<i>cis:trans</i> ratio				
P1	90	5:95	86:14	19:81
P2	90	6:94	87:13	19:81
P4	97	3:97	82:18	24:76
P7	72	8:92	94:6	23:77
P8	89	6:94	85:15	23:77
P9	86	9:91	91:9	16:84
P12	63	4:96	90:10	19:81

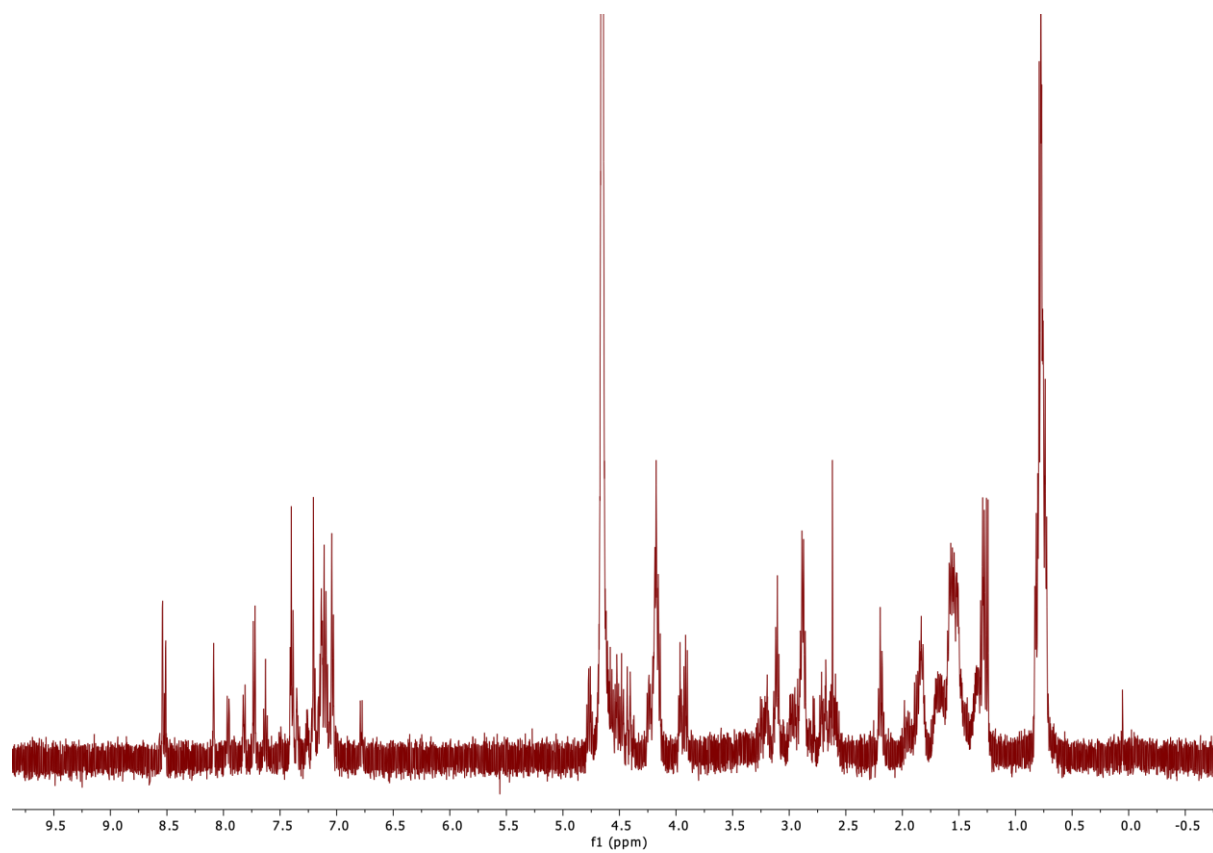
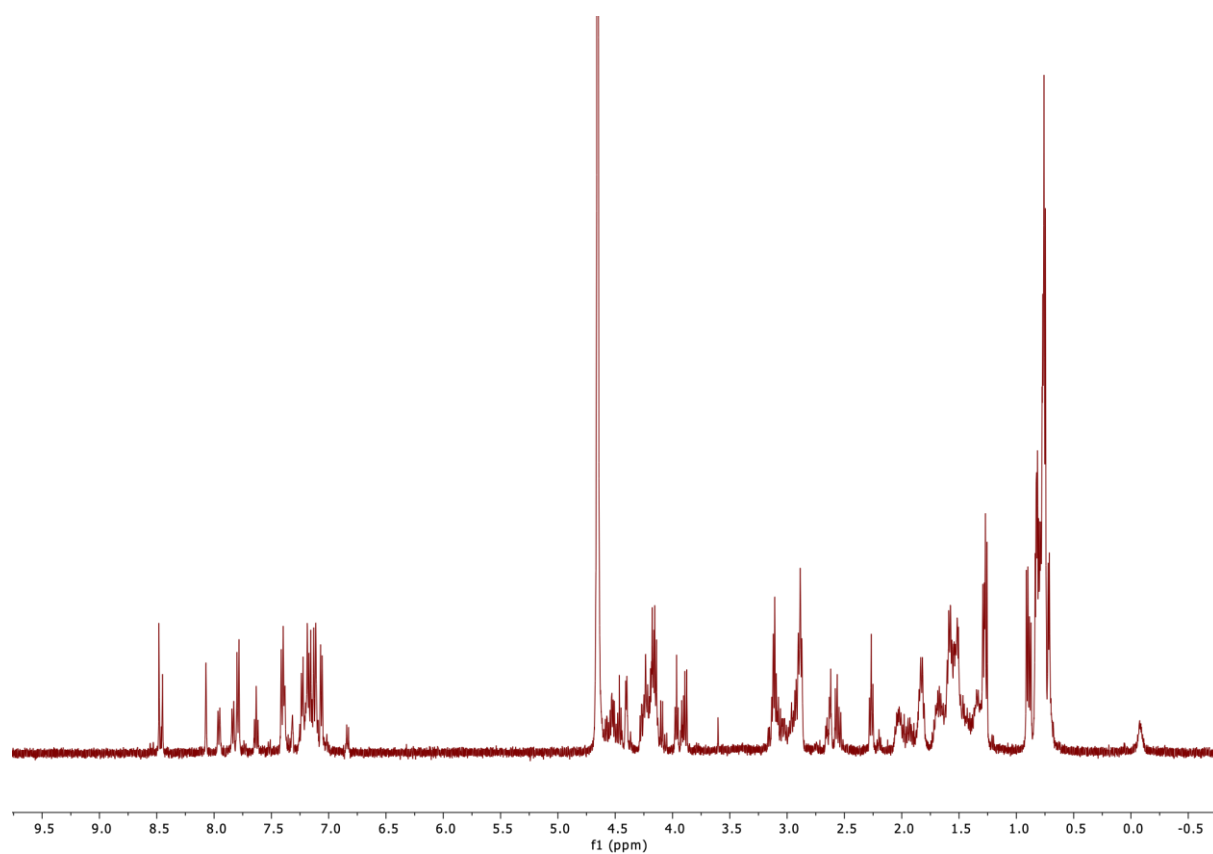
**Table S4.** Fibrillization parameters of PTH<sub>25-37</sub>, P1, P2, P4, P7, P8, and P9 in buffered solution.  $t_{lag}$  - lag time.  $t_{char}$  - characteristic time.

Peptide	$t_{lag}$ ( <i>trans</i> ) [h]	$t_{char}$ ( <i>trans</i> ) [h]	$t_{lag}$ ( <i>cis</i> ) [h]	$t_{char}$ ( <i>cis</i> ) [h]
PTH <sub>25-37</sub>	7.770 ± 1.111	12.797 ± 3.341	-	-
P1	0.392 ± 0.013	0.556 ± 0.008	9.759 ± 1.383	19.640 ± 0.842
P2	0.482 ± 0.207	1.247 ± 0.208	13.052 ± 1.336	17.867 ± 0.360
P4	1.917 ± 0.507	2.371 ± 0.433	29.525 ± 0.558	42.444 ± 5.251
P7	0.381 ± 0.056	0.627 ± 0.078	8.014 ± 1.643	9.875 ± 1.881
P8	1.069 ± 0.096	1.356 ± 0.197	5.524 ± 1.781	20.701 ± 4.415
P9	0.267 ± 0.019	0.653 ± 0.040	3.833 ± 0.458	4.188 ± 0.290

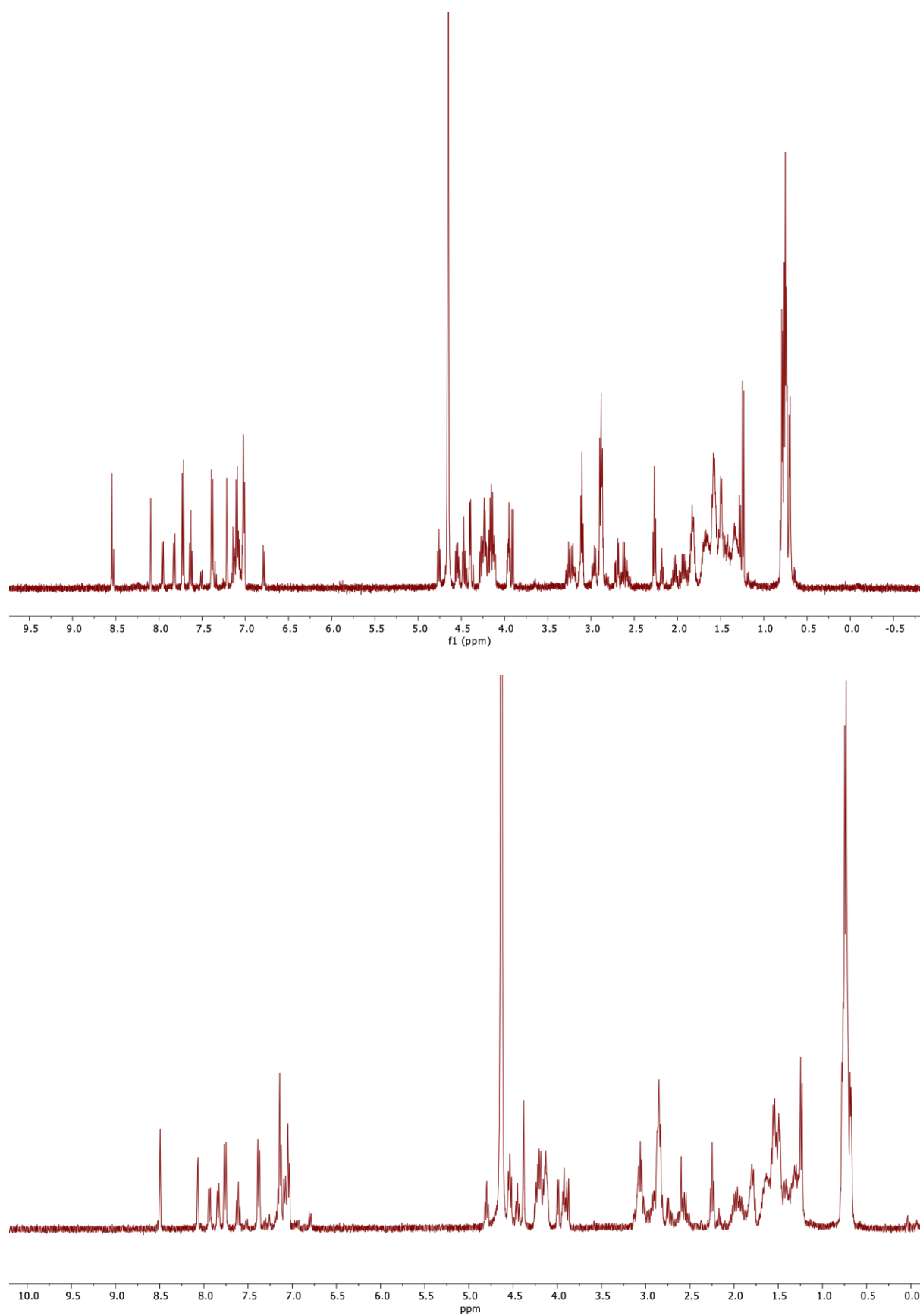
### 3. Peptide characterization after synthesis



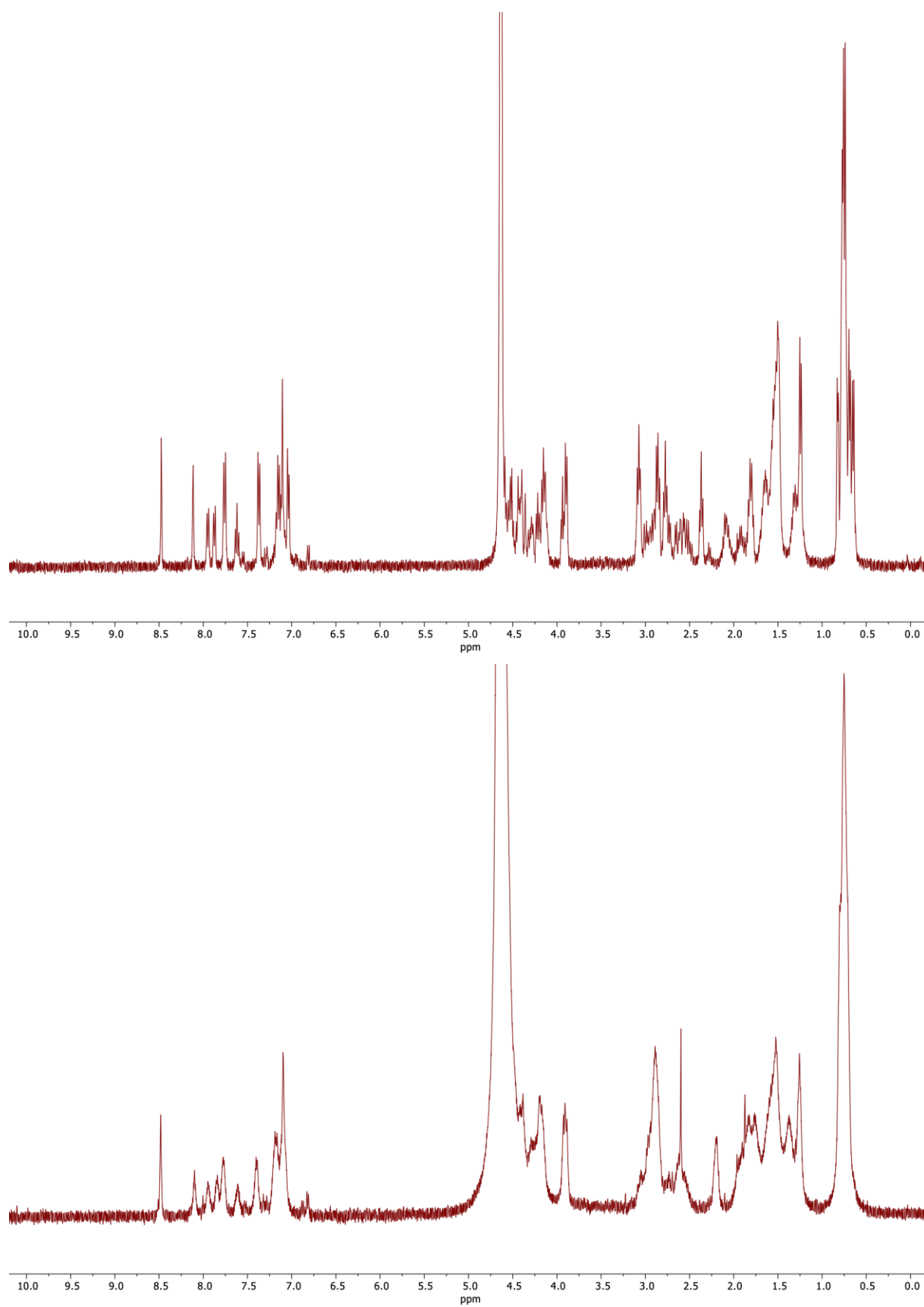
**Figure S30:**  $^1\text{H}$ -NMR spectra (500 MHz,  $\text{D}_2\text{O}$ ) of P1 (top, *trans*-isomer) and P2 (bottom, *trans*-isomer).<sup>3</sup>



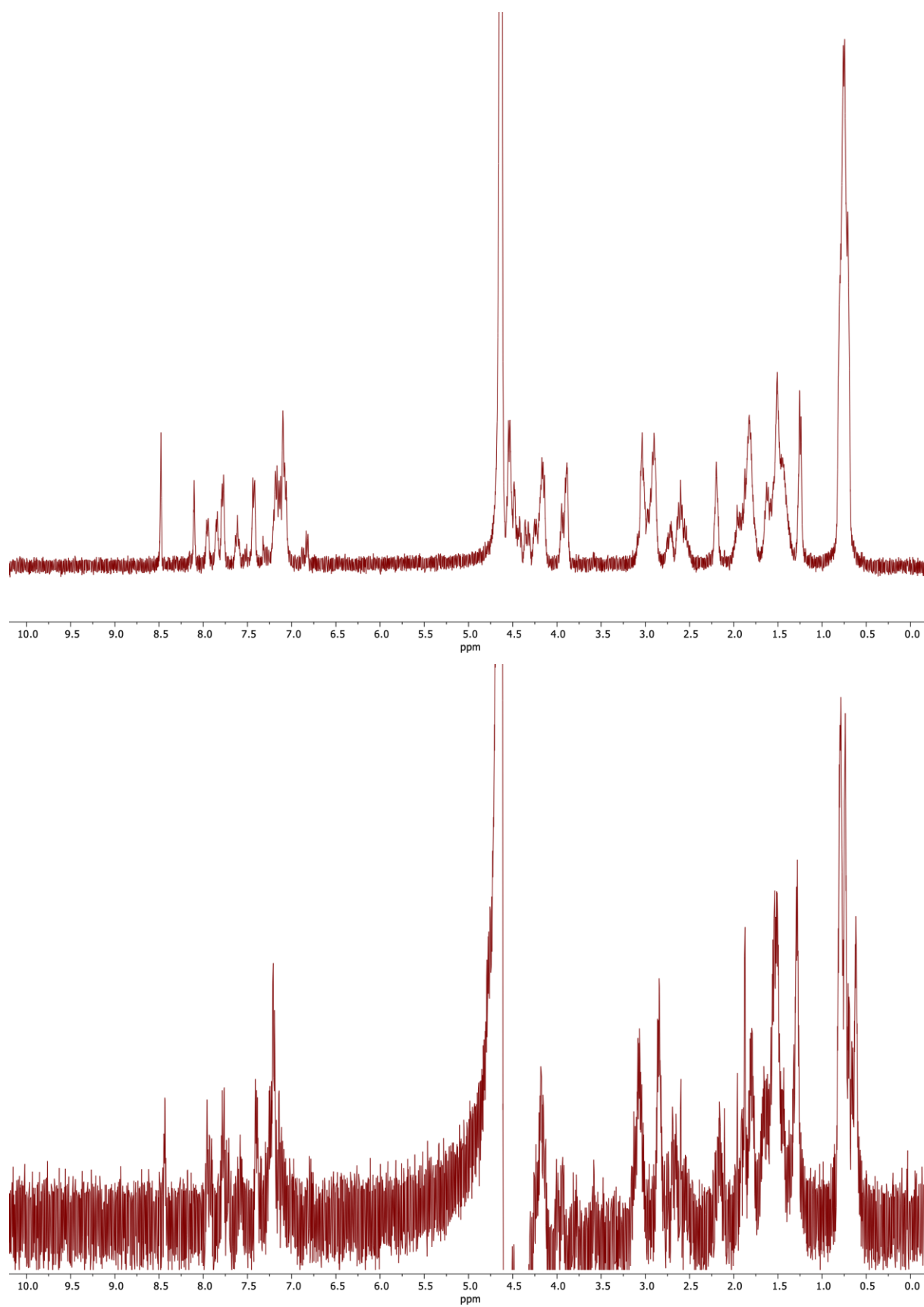
**Figure S31:** <sup>1</sup>H-NMR spectra (500 MHz, D<sub>2</sub>O) of P3 (top, *trans*-isomer) and P4 (bottom, *trans*-isomer).<sup>3</sup>



**Figure S32:**  $^1\text{H}$ -NMR spectra (500 MHz,  $\text{D}_2\text{O}$ ) of P5 (top, *trans*-isomer)<sup>3</sup> and P6 (bottom, *trans*-isomer).

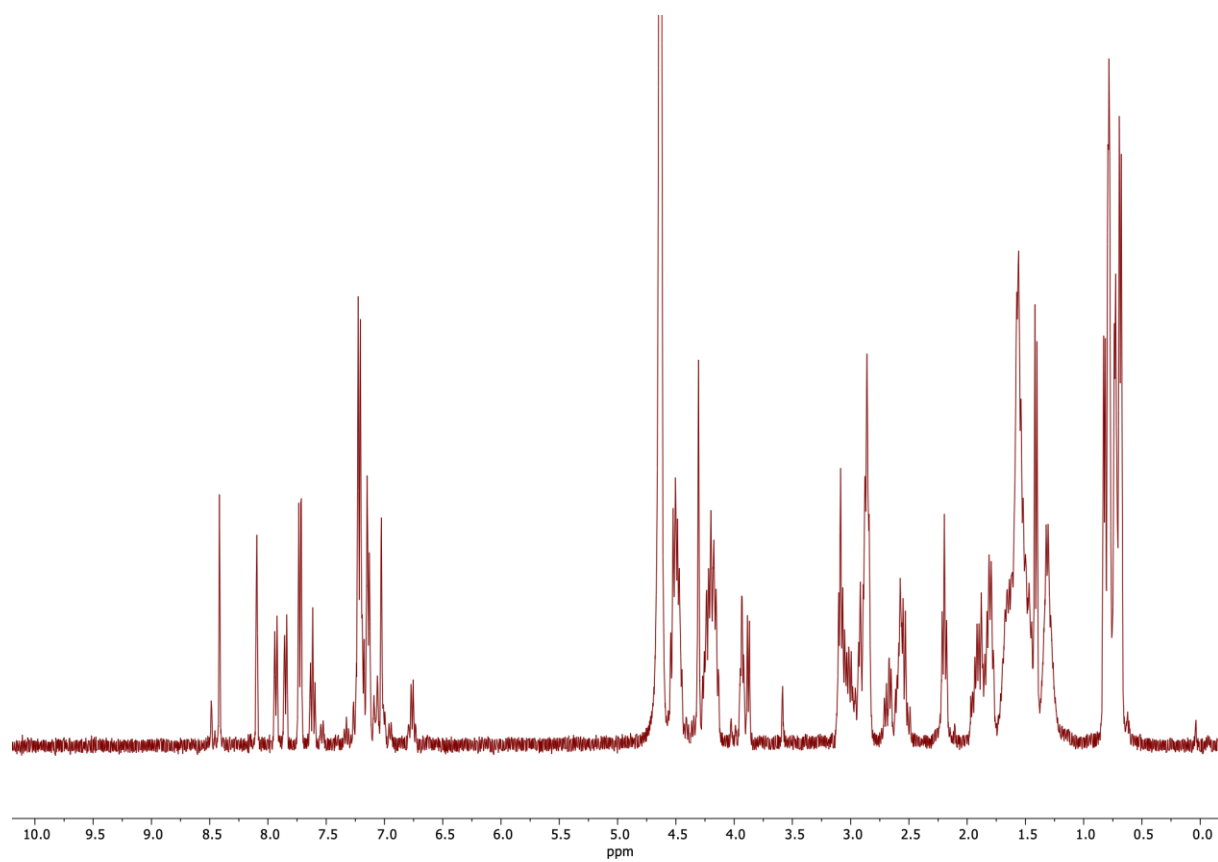
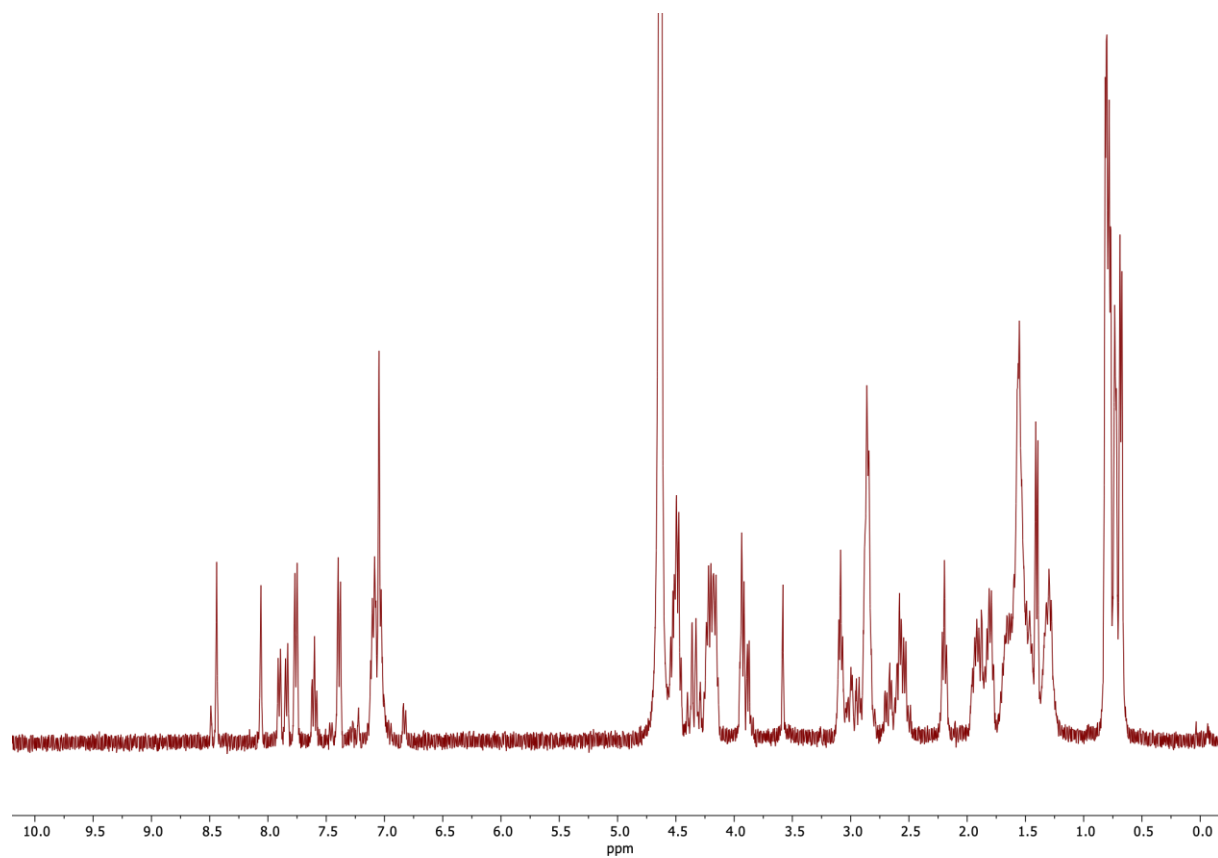


**Figure S33:**  $^1\text{H}$ -NMR spectra (500 MHz,  $\text{D}_2\text{O}$ ) of P7 (top, *trans*-isomer) and P8 (bottom, *trans*-isomer).



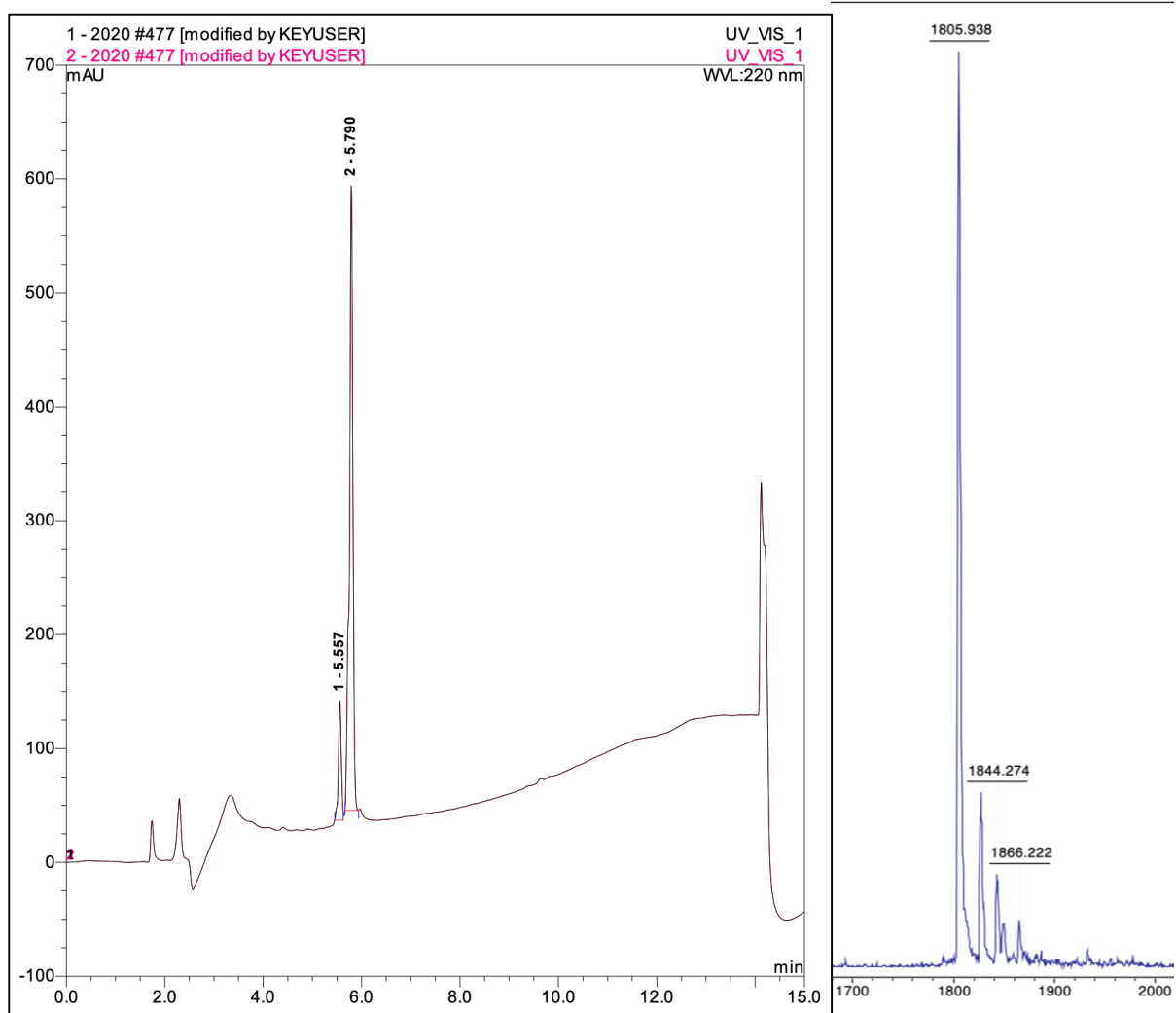
**Figure S34:**  $^1\text{H}$ -NMR spectra (500 MHz,  $\text{D}_2\text{O}$ ) of P9 (top, *trans*-isomer) and P10 (bottom, *trans*-isomer).

+

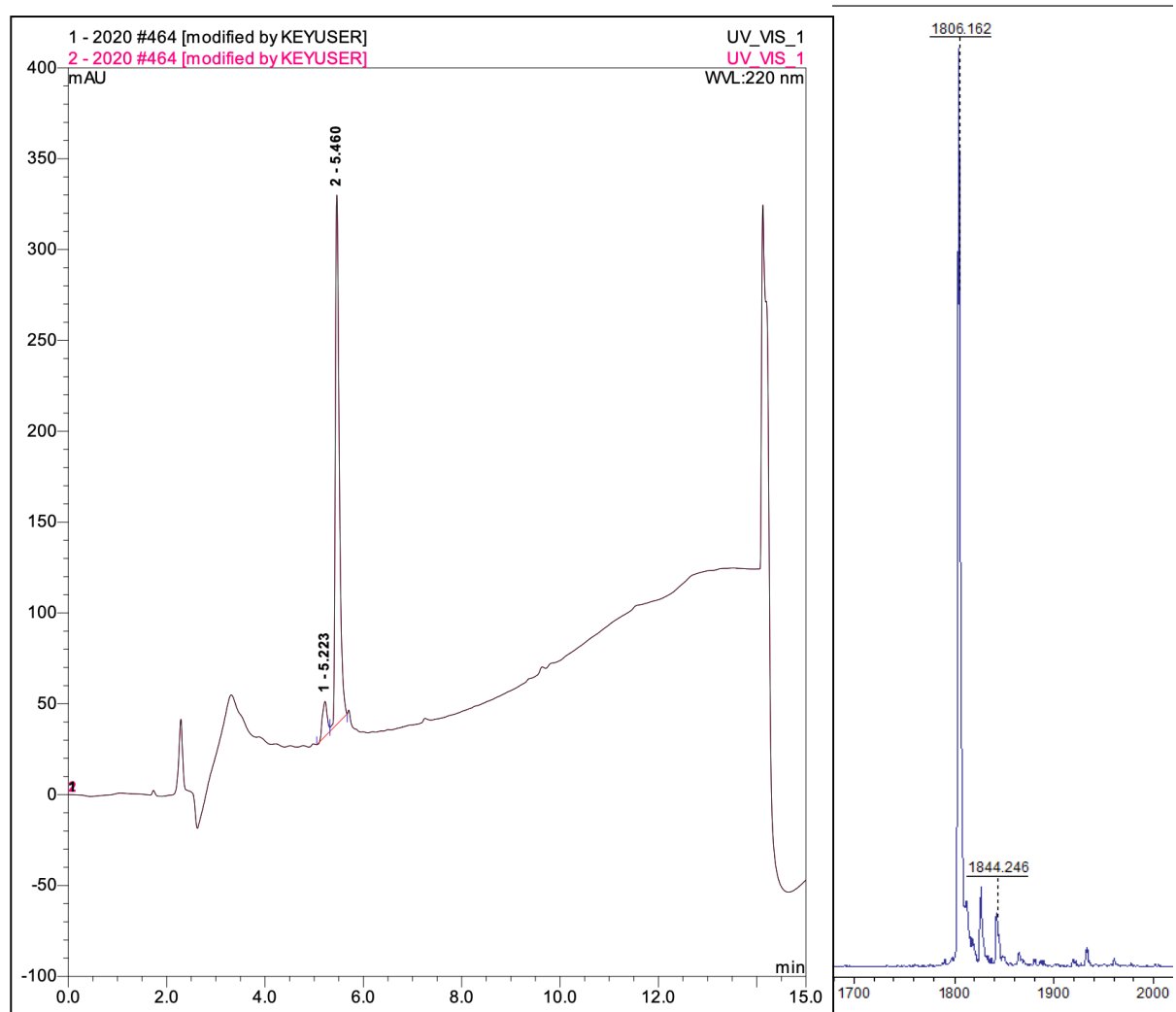


**Figure S35:** <sup>1</sup>H-NMR spectra (500 MHz, D<sub>2</sub>O) of P11 (top, *trans*-isomer) and P12 (bottom, *trans*-isomer).

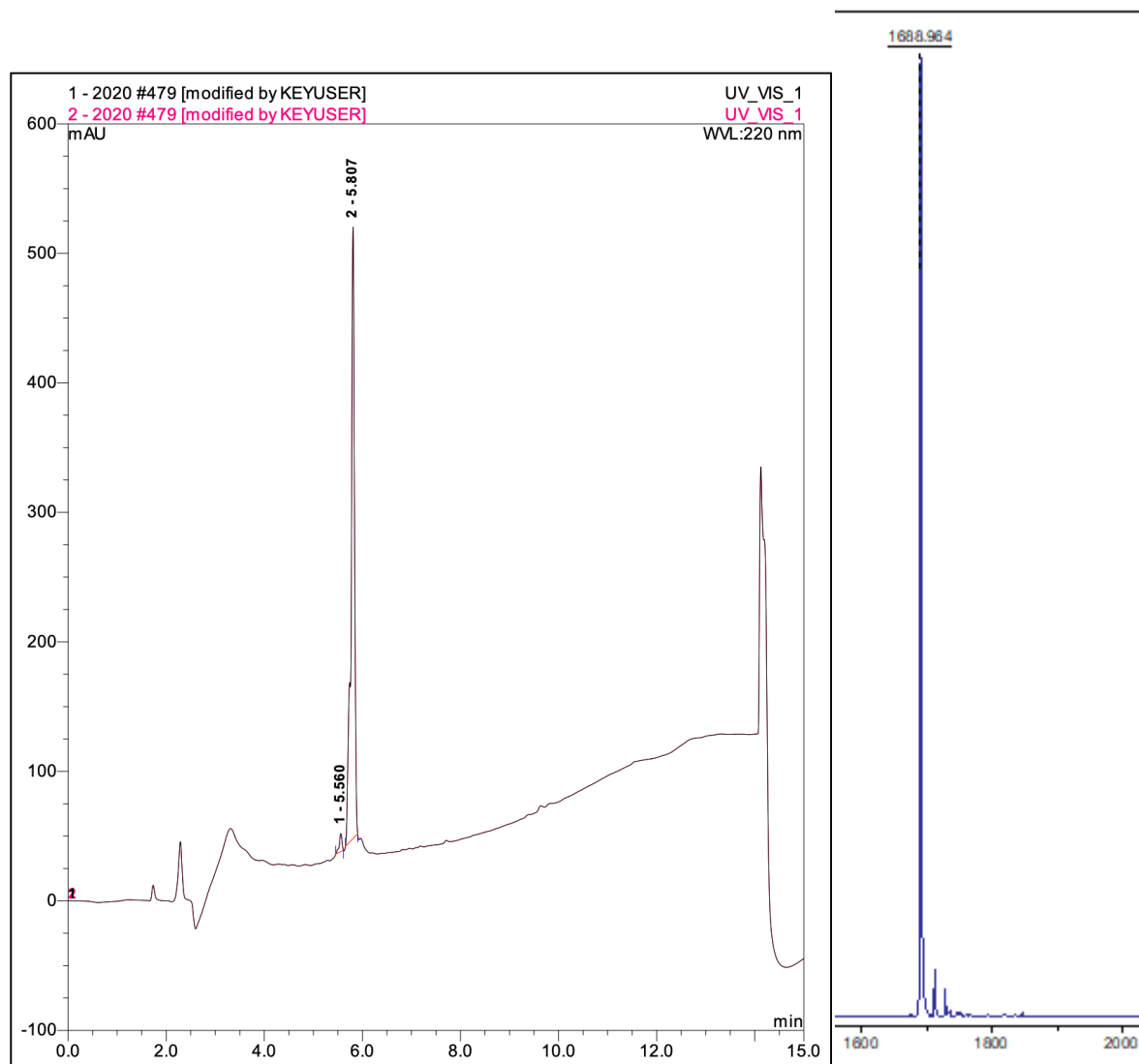




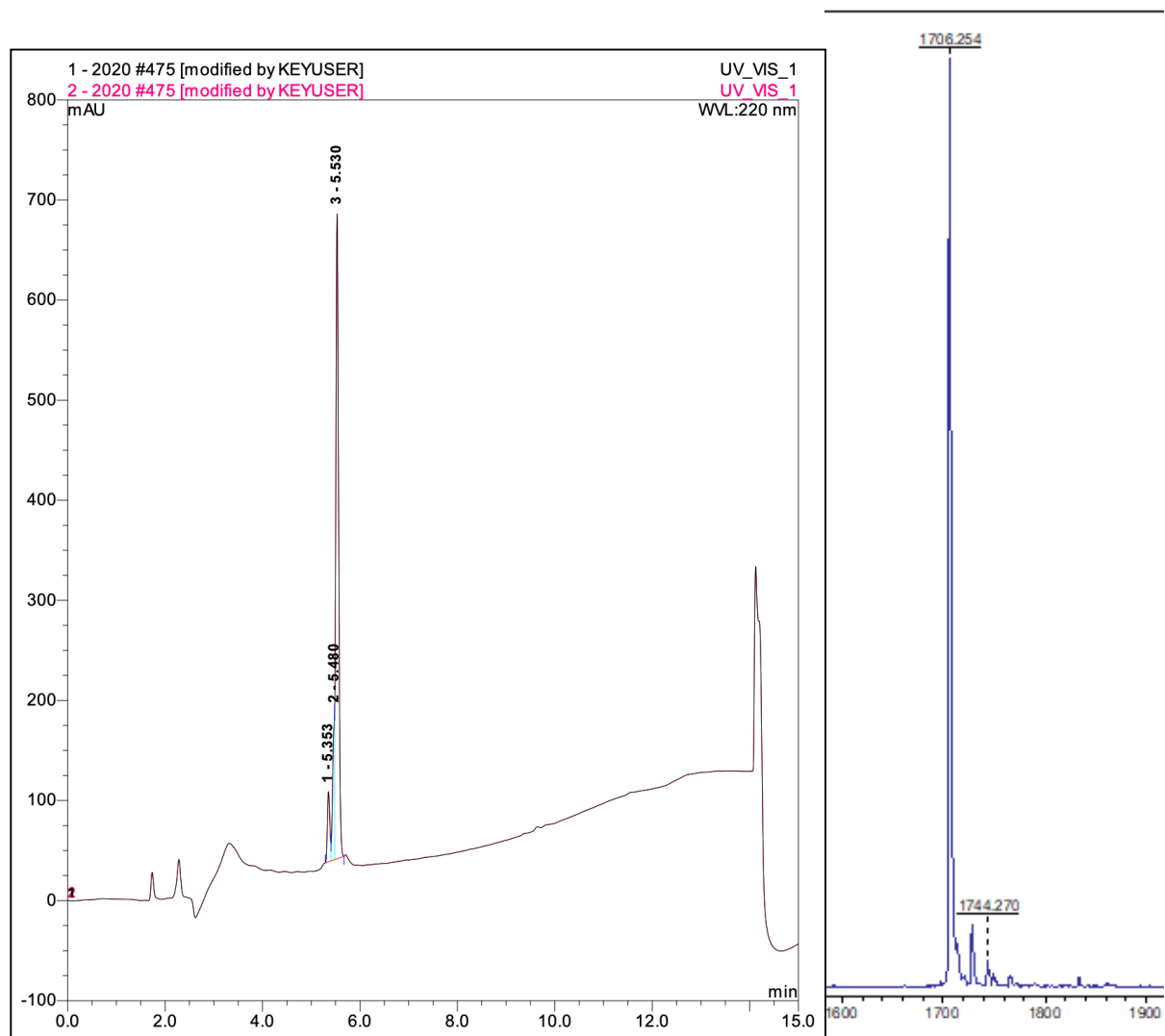
**Figure S36:** (left) HPLC-trace of P1 (*cis*- and *trans*-isomer between 5 and 6 min). (right) MALDI-spectrum of P1 ( $m/z$  calculated: 1805.00 found: 1805.938).<sup>3</sup>



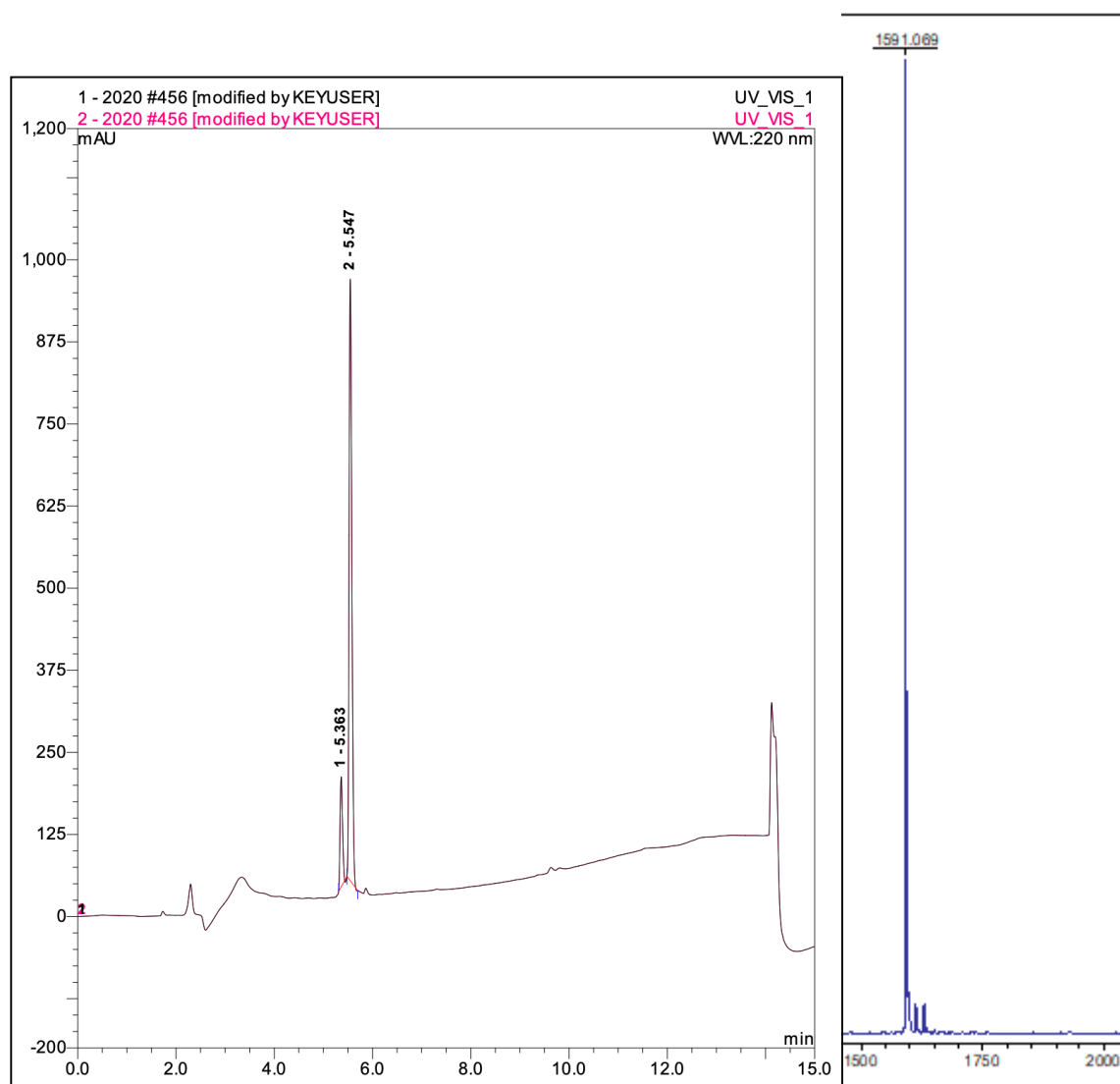
**Figure S37:** (left) HPLC-trace of P2 (*cis*- and *trans*-isomer between 5 and 6 min). (right) MALDI-spectrum of P2 ( $m/z$  calculated: 1805.00 found: 1806.162).<sup>3</sup>



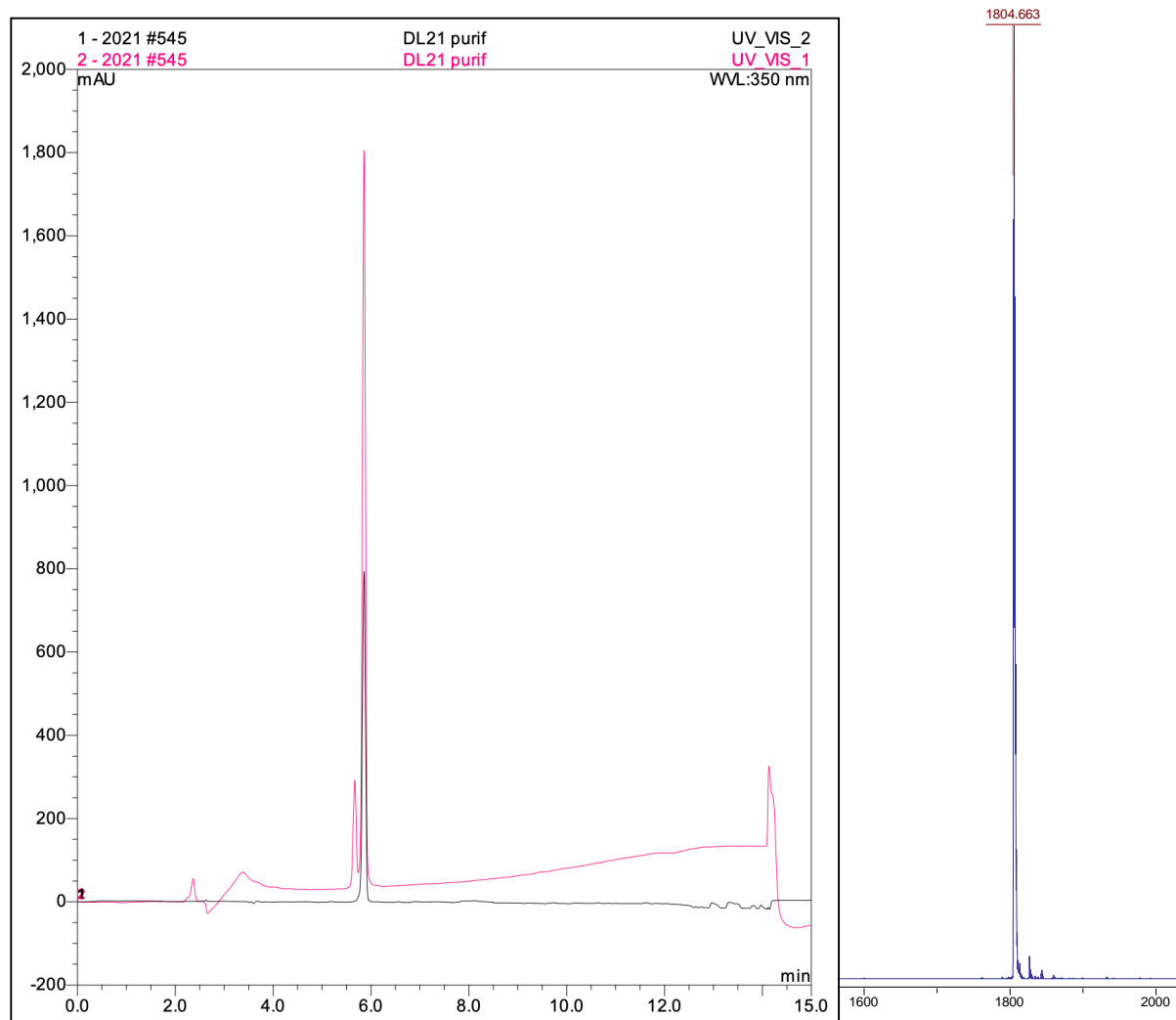
**Figure S38:** (left) HPLC-trace of P3 (*cis*- and *trans*-isomer between 5 and 6 min). (right) MALDI-spectrum of P3 ( $m/z$  calculated: 1689.97 found: 1688.964).<sup>3</sup>



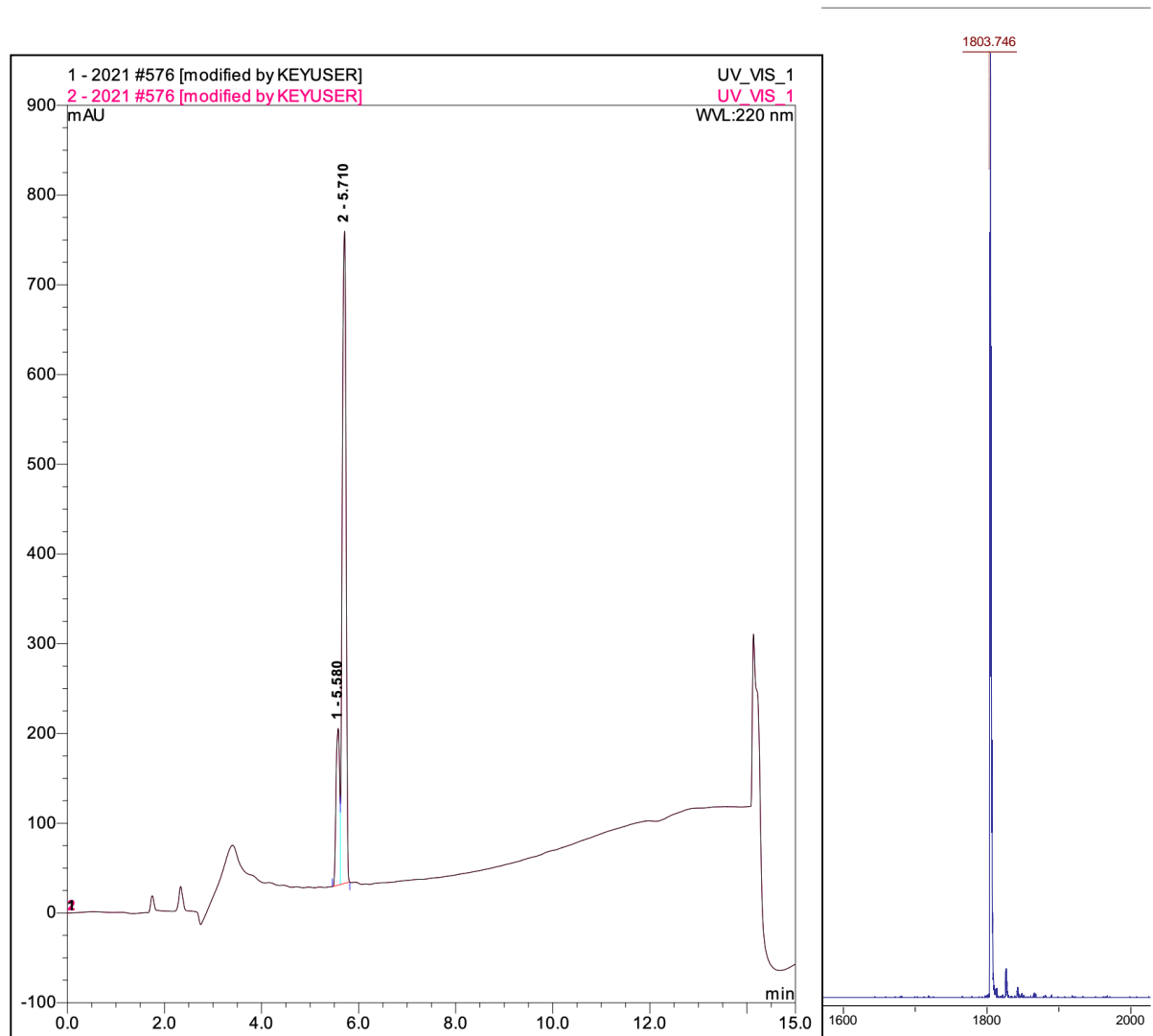
**Figure S39:** (left) HPLC-trace of P4 (*cis*- and *trans*-isomer between 5 and 6 min) (right) MALDI-spectrum of P4 ( $m/z$  calculated: 1705.93 found: 1706.254).<sup>3</sup>



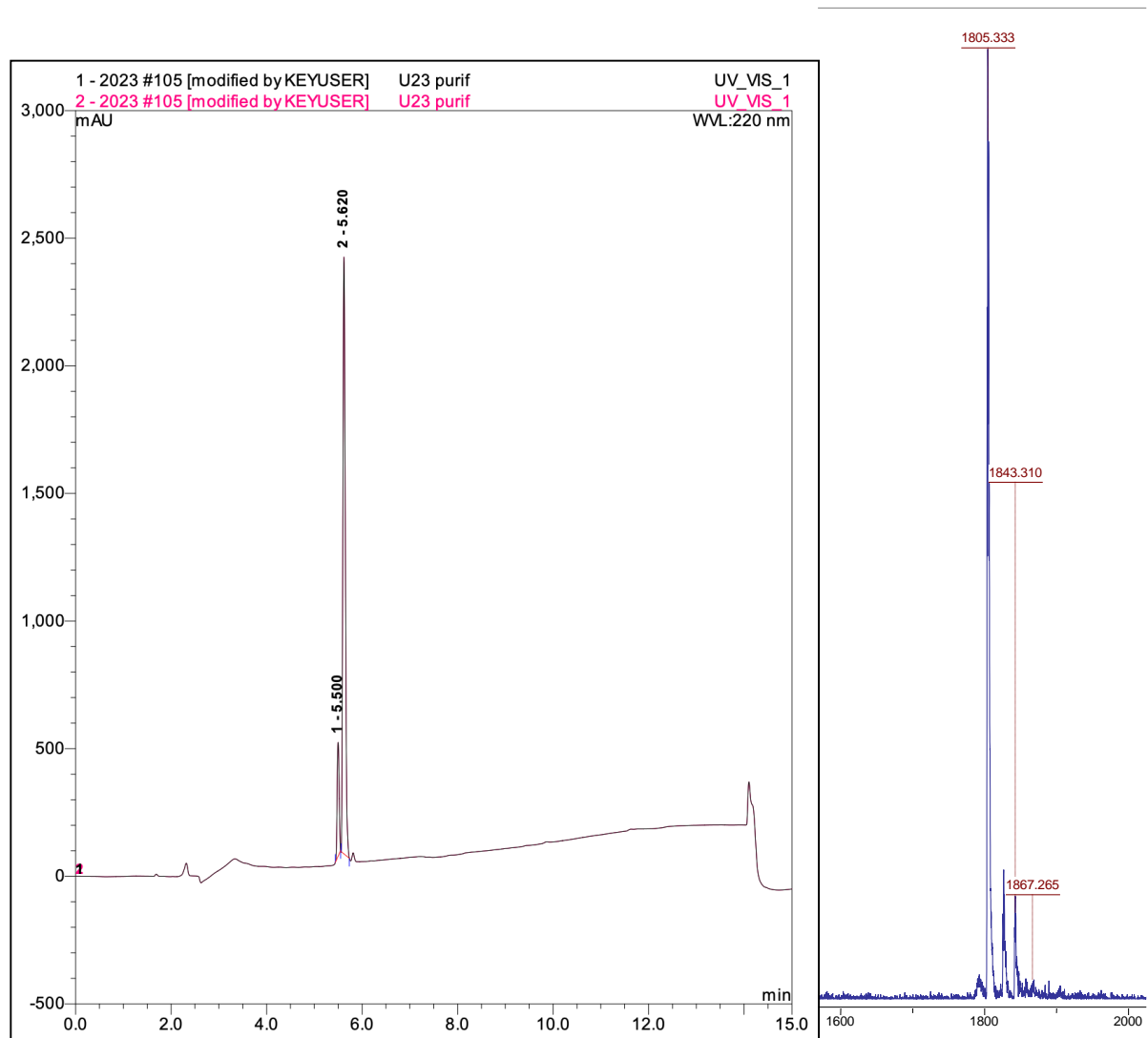
**Figure S40:** (left) HPLC-trace of P5 (*cis*- and *trans*-isomer between 5 and 6 min). (right) MALDI-spectrum of P5 ( $m/z$  calculated: 1590.90 found: 1591.069).<sup>3</sup>



**Figure S41:** (left) HPLC-trace of P6 (*cis*- and *trans*-isomer between 5 and 6 min). (right) MALDI-spectrum of P6 ( $m/z$  calculated: 1805.00 found: 1804.663).

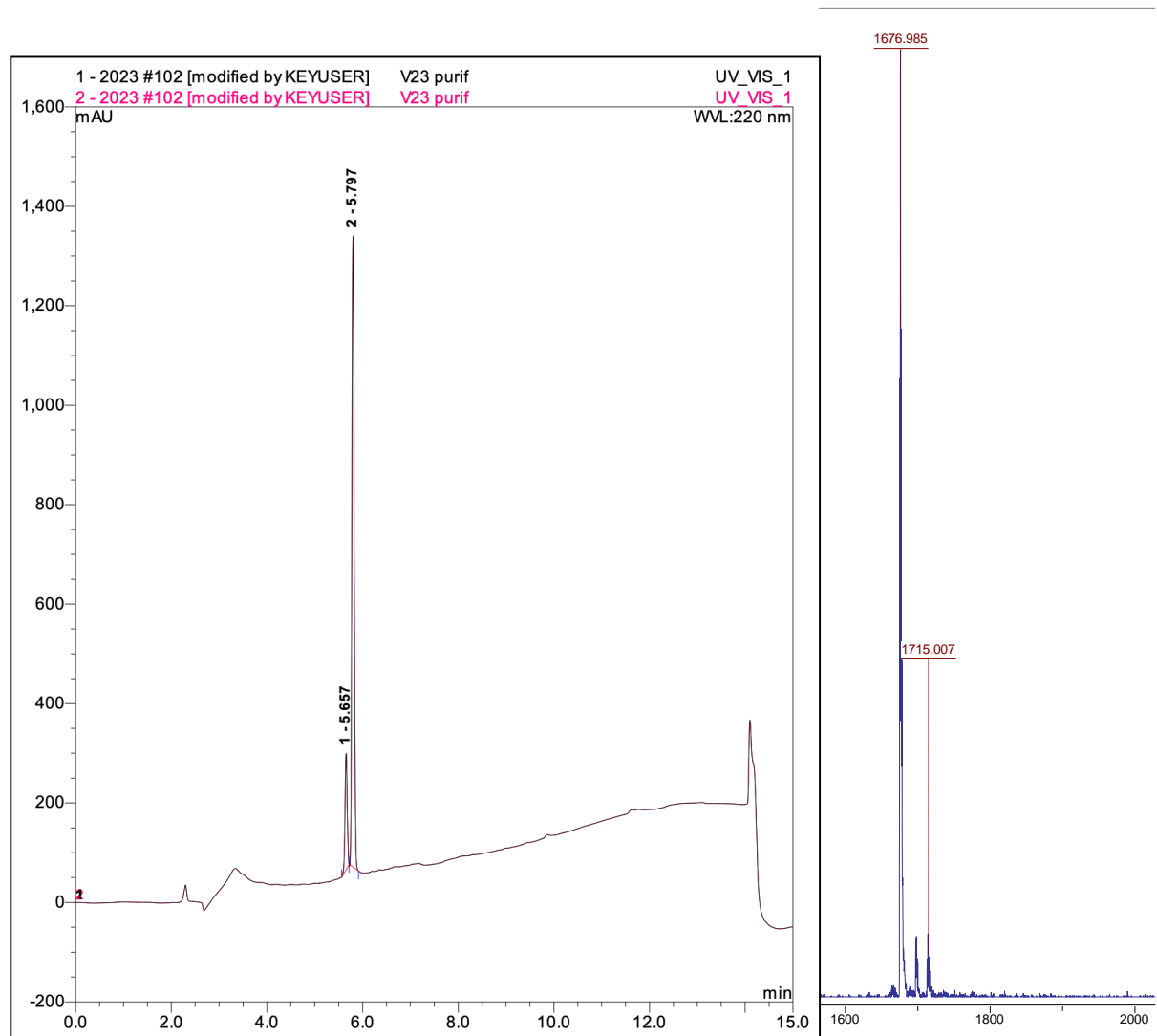


**Figure S42:** (left) HPLC-trace of P7 (*cis*- and *trans*-isomer between 5 and 6 min). (right) MALDI-spectrum of P7 ( $m/z$  calculated: 1805.00 found: 1803.746).

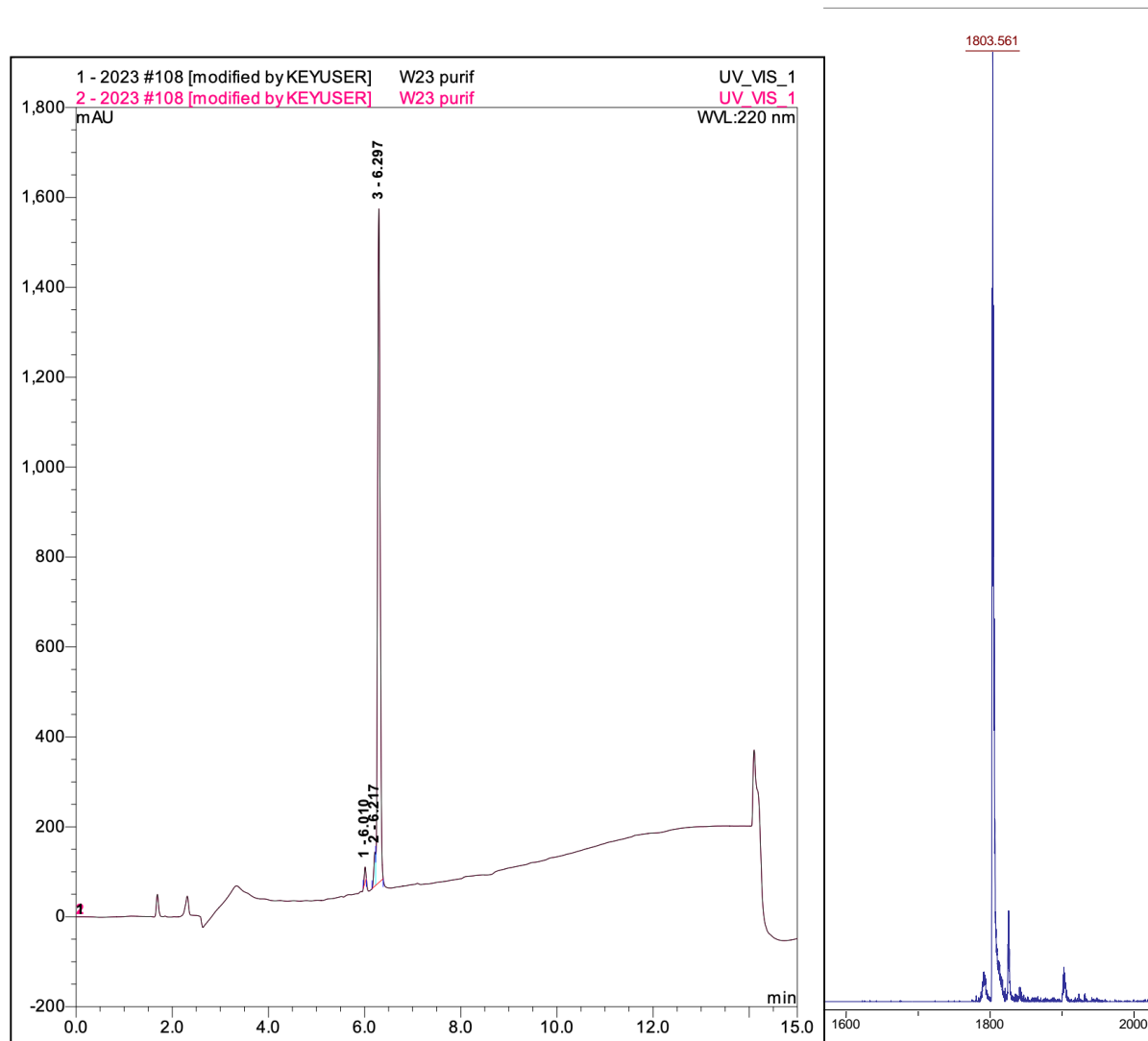


**Figure S43:** (left) HPLC-trace of P8 (*cis*- and *trans*-isomer between 5 and 6 min). (right) MALDI-spectrum of P8 ( $m/z$  calculated: 1805.00 found: 1805.333).

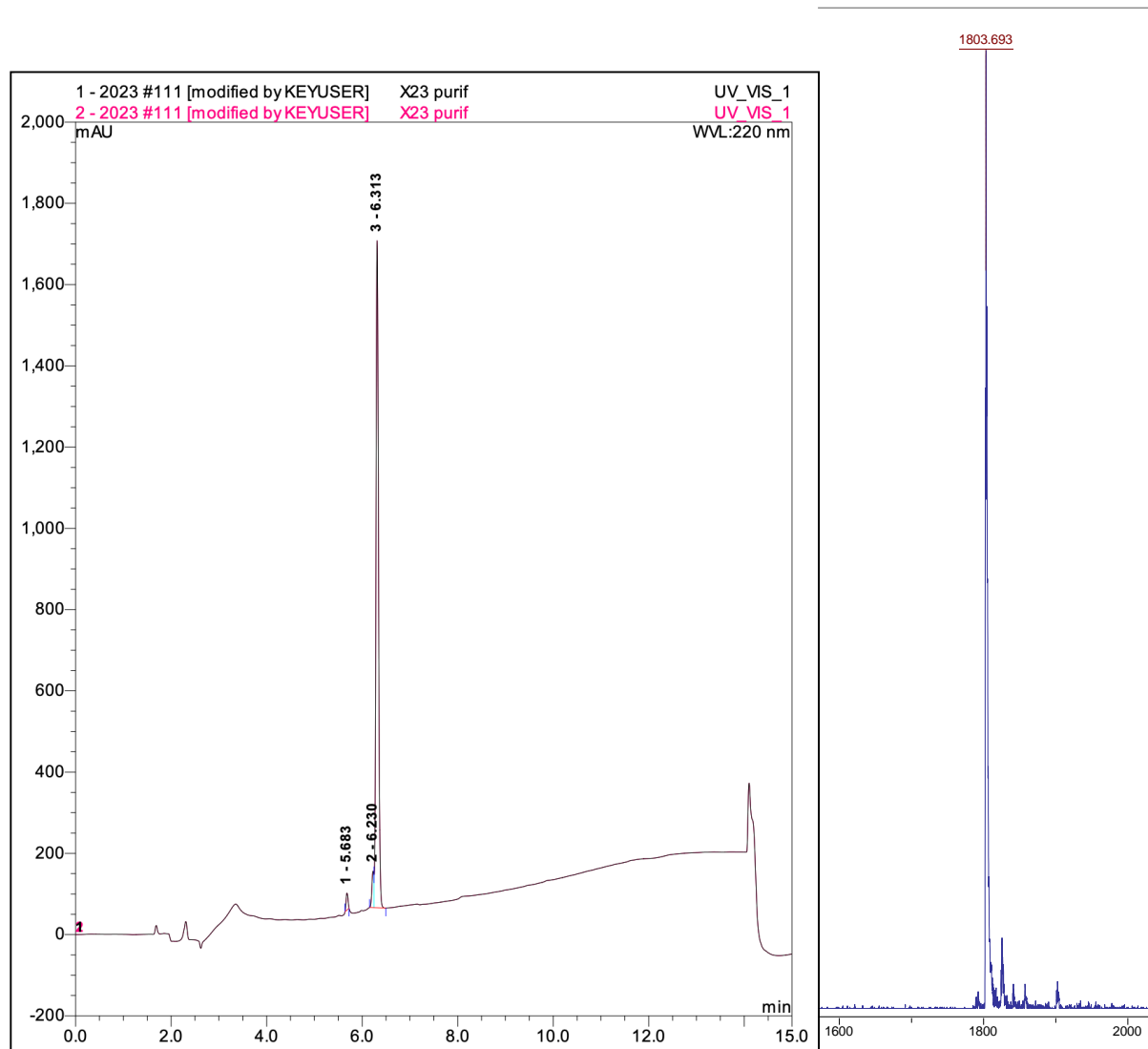




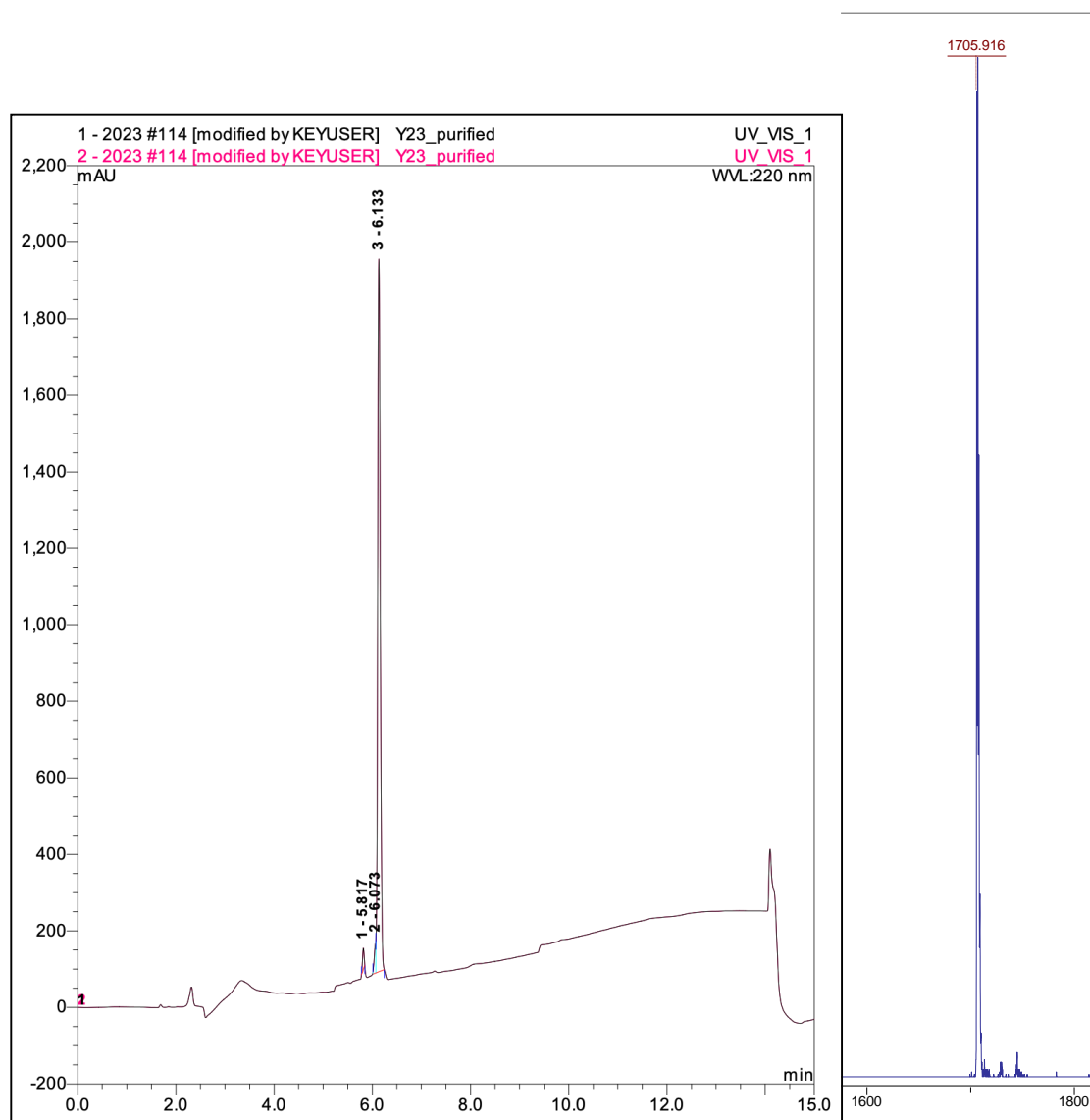
**Figure S44:** (left) HPLC-trace of P9 (*cis*- and *trans*-isomer between 5 and 6 min). (right) MALDI-spectrum of P9 ( $m/z$  calculated: 1676.90 found: 1676.985).



**Figure S45:** (left) HPLC-trace of P10 (*cis*- and *trans*-isomer between 6 and 7 min). (right) MALDI-spectrum of P10 ( $m/z$  calculated: 1805.00 found: 1803.561).



**Figure S46:** (left) HPLC-trace of P11 (*cis*- and *trans*-isomer between 6 and 7 min). (right) MALDI-spectrum of P11 ( $m/z$  calculated: 1805.00 found: 1803.693).



**Figure S47:** (left) HPLC-trace of P12 (*cis*- and *trans*-isomer between 5.5 and 6.5 min) (right) MALDI-spectrum of P12 (*m/z* calculated: 1705.93 found: 1705.916).

## References

- (1) Jurrus, E.; Engel, D.; Star, K.; Monson, K.; Brandi, J.; Felberg, L. E.; Brookes, D. H.; Wilson, L.; Chen, J.; Liles, K. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* **2018**, 27 (1), 112-128. DOI: 10.1002/pro.3280.
- (2) Schrödinger, L. The PyMOL Molecular Graphics System, Version 1.8. In *(No Title)*, 2015.
- (3) Paschold, A.; Voigt, B.; Hause, G.; Kohlmann, T.; Rothmund, S.; Binder, W. H. Modulating the Fibrillization of Parathyroid-Hormone (PTH) Peptides: Azo-Switches as Reversible and Catalytic Entities. *Biomedicines* **2022**, 10 (7), 1512. DOI: 10.3390/biomedicines10071512.