

# **Spontaneous aggregation of fibril-forming peptides studied by Molecular Dynamics simulations**

Dissertation

zur Erlangung des mathematisch-naturwissenschaftlichen Doktorgrades

"Doctor rerum naturalium"

der Georg-August-Universität Göttingen

vorgelegt von

Dirk Matthes

aus Eisenach

Göttingen 2011

Referent: Prof. Dr. Bernd Abel

Lehrstuhl für Physikalische Chemie und Reaktionsdynamik

Wilhelm-Oswald-Institut für Physikalische und Theoretische Chemie

Universität Leipzig

Koreferent: Prof. Dr. Bert de Groot

Computational and Biomolecular Dynamics Group

Abteilung für Theoretische und Computergestützte Biophysik

Max-Planck-Institut für Biophysikalische Chemie Göttingen

Tag der mündlichen Prüfung: 08.12.2011

## Abstract.

Polypeptides share the capacity to assemble into fibrillar aggregates with generic properties even though they can be otherwise unrelated in sequence or structure. Moreover, the relevance of peptide aggregation in amyloidogenic diseases brings it into the focus of interdisciplinary biophysical research. Advances have been made toward unraveling the structural characteristics of the fibrillar end-states using truncated segments of amyloidogenic proteins and peptides. However, detailed experimental knowledge of primary aggregates, soluble and nonfibrillar oligomers and the course of events in initial peptide assembly is still limited. Such low molecular weight oligomers are described to be transient and polymorphic intermediates in the self-assembly process to highly ordered amyloid fibers and were additionally found to exhibit a profound cytotoxicity.

In this thesis, computer simulations are used to reveal common principles for the primary aggregation stages of small peptide segments from the tau protein, insulin, the islet amyloid peptide and  $\alpha$ -synuclein. Starting from multiple peptide monomers the events of primary aggregate formation are analyzed in atomistic detail from the first diffusional encounter of peptides to contact formation and conformational reorganization of the initially formed aggregates.

The structural fluctuations of the aggregates are studied with a novel methodological framework to describe and unveil the critical variables that govern the processes of biomolecular peptide aggregation. The observed intermediates are structurally heterogeneous and early aggregate formation is likely under kinetic control. Mapping the conformational spectra attained by the diverse ensemble of small, oligomeric structures on collective coordinates highlights similarities and differences between the aggregates from several sequences. Qualitative evidence is provided that the early, polymorphic oligomers feature similar self-complementary sheet packing characteristics as it is proposed for fibrillar and crystalline aggregates. Furthermore, a detailed analysis of the forces driving the oligomerization reveals a common two-step process akin to a general condensation-ordering mechanism and thus provides a rational understanding of the molecular basis of peptide self-assembly. To elucidate the influence of external factors on the aggregation process simulations of model peptide aggregation in the vicinity of DMPC bilayers are carried out. Altogether the view is emphasized that solvent interactions at various stages of the aggregation process play a dominant role.

As an important prerequisite for this work different empirical models used in atomistic simulations have been validated by conducting a systematic study of secondary structure propensity in current molecular dynamics (MD) force fields.





## Acknowledgments.

Above all, I would like to thank my supervisor Bert de Groot for providing me with the opportunity to enter the field of computational biophysics in the first place. Ever since i stepped in his office, asking about the vacant Phd position, I felt to be part of a working atmosphere build on trust, freedom of decision-making and exchange. This was not only enjoyable, but also essential for me to do science in a creative and productive way. His presence, guidance and helpful suggestions were crucial for almost all parts of this thesis. So thank you!

For his efforts in the assembly of the group of people and infrastructure in the Department of Theoretical and Computational Biophysics, I owe Helmut Grubmüller many thanks. I consider myself extraordinarily lucky to have had the opportunity to work in this special scientific environment. I would also like to thank Bernd Abel for supporting me as part of the thesis committee.

Special thanks go to Ansgar Esztermann, Martin Fechner, Eveline Heinemann and Petra Lawecki for their administrative support. All of them made this thesis possible and my daily life with the workstation, clusters and paperwork less troublesome than it would have been otherwise. In particular, I want to thank Carsten Kutzner for his much appreciated help solving GROMACS related issues and for taking time to assemble the marvelous cover art illustration.

I want to express my deep gratitude to Lars Bock, who was there for me a lot at the beginning, getting myself familiar with all the little details. I thank him also for being the best lunch companion and conference buddy. It was also him with whom i enjoyed to connect on all sorts of non-science related topics. For his tireless aid with analysis, ideas and discussions I want to thank Vytautas Gapsys very much. He has helped me and the projects a great deal. Nothing less i owe to Venita Daebel for all the coffee and nail polish. Furthermore, I want to thank and mention some people, who were there with help, exchange or shared laughs along the way: Oliver Link, Jochen Hub, Gerrit Groenhof, Holger Bartels, Martin Vesper, Hadas Leonov, Martin Stumpe, Daniel Seeliger, Jürgen Haas, Ulrich Zachariae, Esteban Vöhringer-Martinez, Christian Blau and Katrin Sieferman.

I would like to thank Ulrike Gerischer and Hilkje Hänel for carefully proof reading the thesis and for helpful suggestions during the writing process. This thesis would also not exist without the unconditional and continuous support of Elke and Heinz throughout my studies, along with their interest in the progress and outcome of this work.

I acknowledge the physical chemistry graduate school (PCGG) and the Max-Planck Society for the funding during the duration of my thesis.



# Contents

---

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Amyloidogenic protein and peptide aggregation . . . . .	2
1.2	Model systems for the investigation of amyloidogenesis . . . . .	5
1.3	Aims and organization of the thesis . . . . .	9
<b>2</b>	<b>Theory and Concepts</b>	<b>11</b>
2.1	Principles and approximations of MD simulations . . . . .	11
2.1.1	Approximations . . . . .	11
2.1.2	Classical biomolecular force fields . . . . .	15
2.2	Integrating the equations of motion . . . . .	17
2.2.1	Time step, constraints and virtual sites . . . . .	17
2.2.2	Ensembles and temperature . . . . .	19
2.3	Trajectory analysis methods . . . . .	20
2.3.1	Principal Component Analysis . . . . .	20
2.3.2	Secondary structure and assignment . . . . .	22
<b>3</b>	<b>Secondary structure propensities in peptide folding simulations: A systematic comparison of molecular mechanics interaction schemes</b>	<b>25</b>
3.1	Introduction . . . . .	26
3.2	Methods . . . . .	27
3.3	Results . . . . .	32
3.3.1	Structural properties . . . . .	32
3.3.2	Sampling properties . . . . .	42
3.4	Discussion . . . . .	45

3.5	Conclusion . . . . .	47
<b>4</b>	<b>Conformational dynamics, assembly pathways and driving forces of steric zipper peptide oligomerization</b>	<b>49</b>
4.1	Mapping conformational dynamics and pathways of steric zipper peptide oligomerization . . . . .	51
4.1.1	Introduction . . . . .	51
4.1.2	Methods . . . . .	53
4.1.3	Results . . . . .	58
4.1.4	Discussion . . . . .	78
4.1.5	Conclusion . . . . .	82
4.2	Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations. . . . .	85
4.2.1	Introduction . . . . .	85
4.2.2	Methods . . . . .	88
4.2.3	Results . . . . .	96
4.2.4	Discussion . . . . .	123
4.2.5	Conclusion . . . . .	127
<b>5</b>	<b>Effect of phospholipid membranes on the self-assembly process of model peptide aggregates</b>	<b>129</b>
5.1	Introduction . . . . .	130
5.2	Methods . . . . .	131
5.3	Preliminary results and discussion . . . . .	134
5.4	Conclusion and outlook . . . . .	139
<b>6</b>	<b>Summary and Conclusions</b>	<b>143</b>
6.1	Secondary structure propensities of MD Force Fields . . . . .	144
6.2	Conformational dynamics, assembly pathways and driving forces of steric zipper peptide oligomerization . . . . .	144
6.3	Effect of phospholipid membranes on the self-assembly process of model peptide aggregates . . . . .	146
	<b>Bibliography</b>	<b>147</b>

## CHAPTER 1

# Introduction

---

The conformational states attainable to a newly synthesized polypeptide chain are governed by the underlying free-energy landscape of the system (Fig. 1.1) [1–5], ultimately determined by the sum of entropic and enthalpic contributions in a given environment.

The common folding paradigms assert that for small globular proteins the native fold represents a thermodynamically stable and kinetically accessible state. The native basin is degenerate and the ensemble of conformations at the global minimum of free-energy [3, 6–9], where the competition of energetically either favorable or unfavorable interactions determines the degree of frustration in the energy landscape. The current understanding of polypeptide folding behavior in terms of statistical ensembles highlights the existence of a substantial heterogeneity of states [1, 3, 10, 11]. The conformational possibilities and multiplicity of folding routes are governed by the large number of weak non-covalent interactions present in biological heteropolymers such as proteins [1, 3, 7, 12]. The intrinsic bonding capabilities of a particular polypeptide are usually satisfied with intramolecular interactions and therefore determine the interplay between amino acid sequence and structure [12]. More recently, it has been realized that a biologically relevant view of protein conformational ensembles has to capture the complexity of (functional) conformational change beyond the native state and must also include protein misfolding and oligomerization [9, 11, 13]. It has been proposed that the intermolecular interactions of the invariant polypeptide backbone result in an alternative generic conformational state, known as amyloid, competing with the configurational tendencies of an isolated polypeptide chain (Fig. 1.1) [4, 11, 12, 14–18]. Exposure of hydrophobic epitopes and regions of unstructured polypeptide backbone, as found in partially folded or misfolded states, is among the accepted causes of amyloidogenic aggregation [4, 11, 19, 20]. Often associated with the irreversible formation of this particular class of  $\beta$ -sheet aggregate structures is a toxic gain of function in vivo [4, 18, 21, 22].

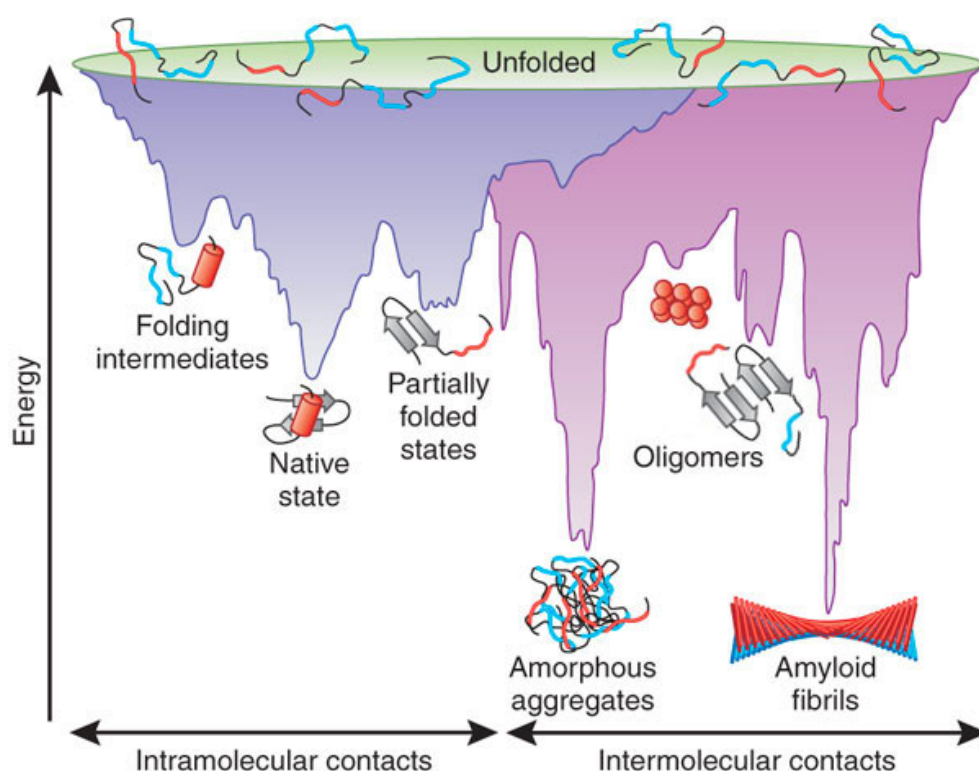


Figure 1.1: **Energy landscape illustration of protein folding and aggregation.** The purple regime of intramolecular contacts highlights the multitude of conformations available to a single polypeptide chain. The energy landscape 'funnels' to the native state. The regime of intermolecular contacts (pink) is characterized by an increased ruggedness and multiple deep minima. Polypeptide aggregates may explore amorphous aggregate or amyloid fibril conformations. Transient oligomeric aggregates can emerge as on- or off-pathway intermediates of amyloid fibril formation. Figure adapted from [11].

## 1.1 Amyloidogenic protein and peptide aggregation

**Amyloid fibrils.** Many natural polypeptide chains are able to form amyloid fibrils in vivo or in vitro [4, 14, 18, 23–33]. Current definitions of amyloid or amyloid fibrils can differ considerably depending on whether they are used in a biophysical or physiological context [34–37]. To circumvent confusion of terms, a structure based definition of amyloid fibrils and other polypeptide aggregates [35] will be used in the following. Regardless of the sequence or native fold, amyloid fibrils can be defined as self-assembled, elongated and unbranched (fibrillar) polypeptide aggregates with cross- $\beta$  conformation (see Figure 1.2) [28, 32, 38]. The cross- $\beta$  architecture, as revealed by X-ray fiber diffraction data, is described as stacked  $\beta$ -strands that run perpendicular to the fibril axis and exhibit extensive hydrogen bonding along the length of

the fibril [39–42]. To this end, A $\beta$  [43–45], amylin [46], fungal prion [47], and PrP peptide [48] fibrils have been characterized by the application of solid-state NMR (ssNMR) [49] and all of them were found to be composed of in-register parallel  $\beta$ -sheets. Most notably, an in-register parallel  $\beta$ -sheet potentially maximizes favorable interactions between hydrophobic, as well as polar side chains by aligning the residues with themselves. Amyloid fibrils are polymorphic structures and a wide range from thin, straight fibrils to wider, striated, twisted ribbons have been observed [31, 43, 44, 50, 51].

Analyses of experimentally observed aggregation kinetics suggest different mechanistic explanations for possible pathways and rate-limiting steps of amyloid fibril formation [52–56]. Recent reports suggest that there may be similarities in the assembly mechanism of amyloidogenic peptides and proteins, as the overall growth process exhibits the characteristics of a nucleated growth [52, 53, 57, 58]. In essence, the amyloidogenic self-association of peptides or proteins will initially populate high energy states, in which the sampled state of highest energy is termed the nucleus. Once the critical nucleus forms, either by addition of further assembly units or by stabilization through conformational change, the cooperative fibril formation process will be downhill in free-energy [29]. Nevertheless, it has not been possible to directly probe the nucleation event so far, thus leaving the details of the process not well understood.

**Amyloidosis - a conformational disease.** A number of human pathologies is associated with the deposition and accumulation of stable, ordered, filamentous aggregates of a specific protein or peptide in a variety of organs and tissues [59, 60]. These include neurodegenerative diseases like Alzheimer's disease (AD), Parkinson's disease (PD) and Huntington's disease (HD), type 2 diabetes, dialysis-related amyloidosis and familial systemic amyloidosis. Prion proteins that cause transmissible spongiform encephalopathies (TSEs) also form amyloid fibrils [61, 62]. The self-propagating variations in the molecular structure of amyloid fibrils and amyloid-like aggregates are believed to be responsible for multiple strains of mammalian prions and yeast prion phenotypes [31].

**Oligomers and nonfibrillar aggregates.** Recently, evidence has accumulated suggesting that, instead of mature amyloid fibers, soluble oligomers are the more pathogenic species and primary causative agents of several types of amyloid diseases [4, 63–70]. Oligomers were found to exhibit high levels of cytotoxicity in cell cultures and also have been found to localize in human tissue [64, 71]. Furthermore, the presence of oligomers correlates better

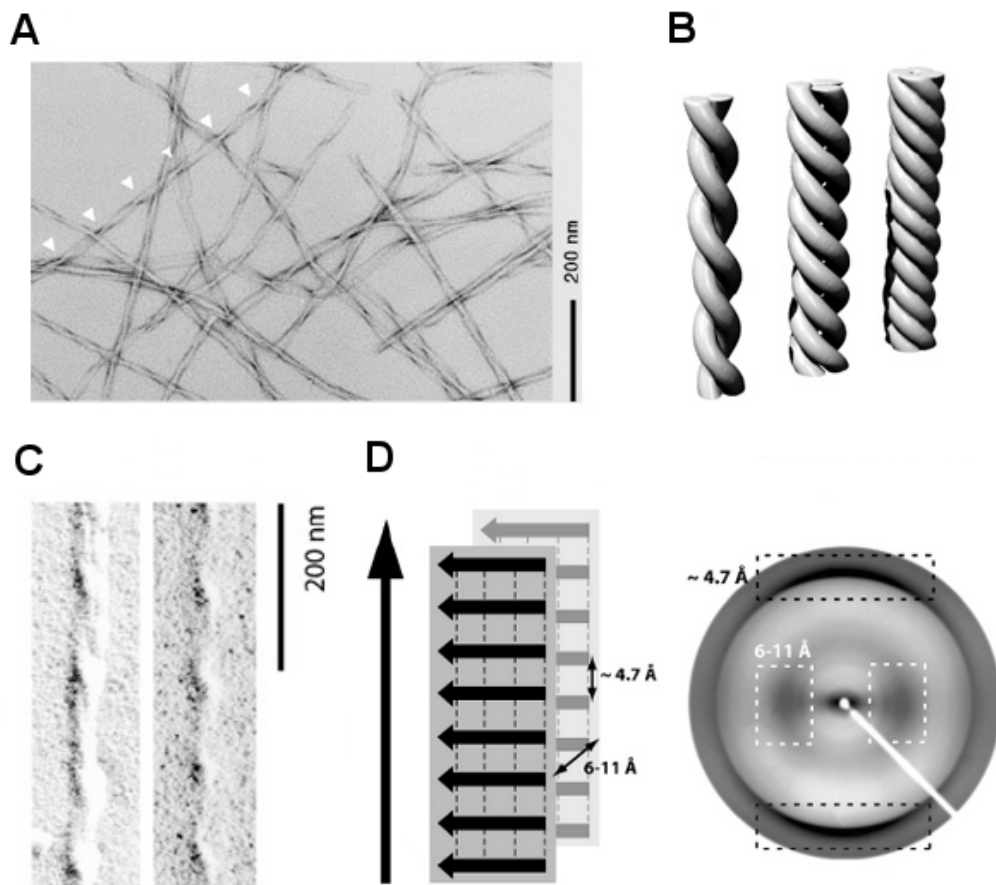


Figure 1.2: **Structure of amyloid fibrils.** Electron micrograph of long, unbranched  $A\beta(1-40)$  fibrils with characteristic twist. Twist crossovers at regular distances are indicated by white arrow heads (A). Schematic representation of fibrils composed of 2, 3 and 4 protofilaments (B). Left-handed fibril chirality of  $A\beta(1-40)$  amyloid fibrils observed with transmission electron microscopy (TEM) after platinum side shadowing (C). On the left: Schematic representation of cross- $\beta$  sheets architecture in a fibril. The black arrow indicates the orientation of the fibril main axis, backbone hydrogen bonds are represented by dashed lines. On the right: The typical fiber diffraction pattern with a meridional reflection at  $4.7\text{ \AA}$  (black dashed box) and an equatorial reflection at  $6-11\text{ \AA}$  (white dashed box), which correspond to the repetitive spacings of main chain and side chain atoms in the protofilament structure, respectively (D). Figure adapted from [18] and [35].

with the pathological changes than the insoluble fibrillar deposits do [72]. This indicates a possible origin for the relation between amyloid formation and cellular toxicity, namely the disruption of membrane integrity by oligomeric species. Several possible modes of membrane perturbations offer an explanation for cellular stress, e.g. through loss of chemical potential



and compartmentalization [13, 70, 73]. In addition, exposure or association with membranes was also found to be correlated with an increased rate of amyloid formation, leading to highly structured fibrillar states [74, 75]. A variety of morphologies have been described for oligomers, which are usually observed during the incubation of amyloidogenic peptide solutions. These include prefibrillar and fibrillar oligomers, annular protofibrils, among others [56, 63, 65–68, 71]. Molecular weights of the oligomeric precursor state are reported to span a range from a few ten to hundreds of kDa, corresponding from dimeric to multimeric aggregates, respectively. However, it is difficult to obtain the biochemical properties as well as structural information of the oligomeric species in experiments. Most notably the investigations are hindered by the transient, polymorphic, and noncrystalline behavior of the oligomers [66, 71, 76]. Nevertheless, several studies report that: (a) amyloid oligomers contain  $\beta$ -sheet rich structures [77, 78]; (b) oligomeric states are often heterogeneous and different sizes of the oligomers coexist in solution [79]; (c) antibodies recognize common structural features of oligomers formed from different amyloidogenic proteins [64]. Furthermore, given their qualitatively different morphologies from the characteristic appearance in TEM and AFM images, oligomeric precursor states and amyloid fibrils are surprisingly similar in molecular conformation and supramolecular structure [29, 67, 80]. In addition, crystal structures of peptide macrocyclics in an oligomeric (tetrameric) form were reported recently [81], providing additional structural constraints on this particular aggregation state. An interesting finding was that except for the observed sheet-to-sheet packing, they share most structural features of the fibrillar forms [82].

## 1.2 Model systems for the investigation of amyloidogenesis

To understand the properties of amyloid fibrils and to obtain a mechanistic interpretation of the multi-staged aggregation process detailed knowledge of the molecular structures of the involved species is inevitable. Studying structures of amyloid cross- $\beta$  spines at atomistic detail has been impeded by their inherently noncrystalline and insoluble nature, as well as their assembly from high-molecular-weight units [4]. Therefore, standard experimental approaches to structure determination of amyloid fibrils formed by natural proteins are not applicable or provide only limited information. In order to permit the study and systematically dissect the structural, physical, and chemical properties of the complex in vivo aggregation process small in vitro model systems were devised [83–90]. This approach helped to develop an understanding of the more general phenomenon of amyloid deposition by a controlled experimental access to the individual

contributions and delicate balance of interactions [90]. The use and biophysical characterization of model peptides for amyloidogenesis is motivated furthermore by the advantage to investigate sequence determinants and rationally study e.g. mutational effects on fibril formation [21, 85, 87].

Moreover, computational approaches have identified consensus aggregation-prone sequence motifs of amyloidogenic proteins, therefore leading to the idea that the amyloidogenicity of a sequence can be strongly localized [17, 89, 91, 92]. Indeed, experimental evidence is compiling that protein unfolding is necessary but not sufficient to promote aggregation. Furthermore, it was found that specific short stretches in a sequence can trigger self-assembly and mediate amyloid formation [92–96]. From experiments on amyloidogenic peptide segments many critical observations regarding the energetics and molecular structures of these systems have been derived [90, 97, 98]. Therefore, one can argue that short model peptides should be more suitable than full-length proteins to investigate those elements in sequences that favor aggregation.

Combined, these experimental findings underscore the notion of amyloid fibril formation being a universal property of the peptide backbone that depends on external factors and is modulated by sequence characteristics [4, 16, 21, 75, 99]. Structure based analysis of amyloidogenic sequence signatures predicted short segments and showed that hexapeptides are able to form amyloid-like fibrils [84, 88, 100]. The ability of these segments to even force a globular, non-fibrillizing protein into the amyloid state was demonstrated [101].

**Steric zipper peptides.** Crystal structures for a growing number of such minimal peptide sequences provided insight into what could be the general spine organization of amyloid fibrils [82, 100, 102–104]. A common motif, called a steric zipper, was revealed in all of the crystalline structures. The atomic structures show pairs of elongated  $\beta$ -sheets with parallel or antiparallel strand alignment. They are interdigitated such that a high complementarity packing of the side chains is achieved, leading to a tight and dry interface. Despite their fundamental similarity, the structures vary in their basic steric zipper motif. Alternative  $\beta$ -sheet packing arrangements of the same segment, as well as distinct  $\beta$ -sheets, built from different segments of a protein, have been found. It has been argued that this can help to understand the observed polymorphism of amyloid structures on a molecular basis [102]. It has been shown that crystalline and fibrillar amyloid polymorphs share fundamental structural characteristics such as the cross- $\beta$  diffraction pattern [18, 82, 103–105]. Furthermore, seeding experiments with crystals from fibril-forming short peptides reduce the lag time for the growth of the full-length parent protein fibrils [82, 88]. Nevertheless, the degree of order in the

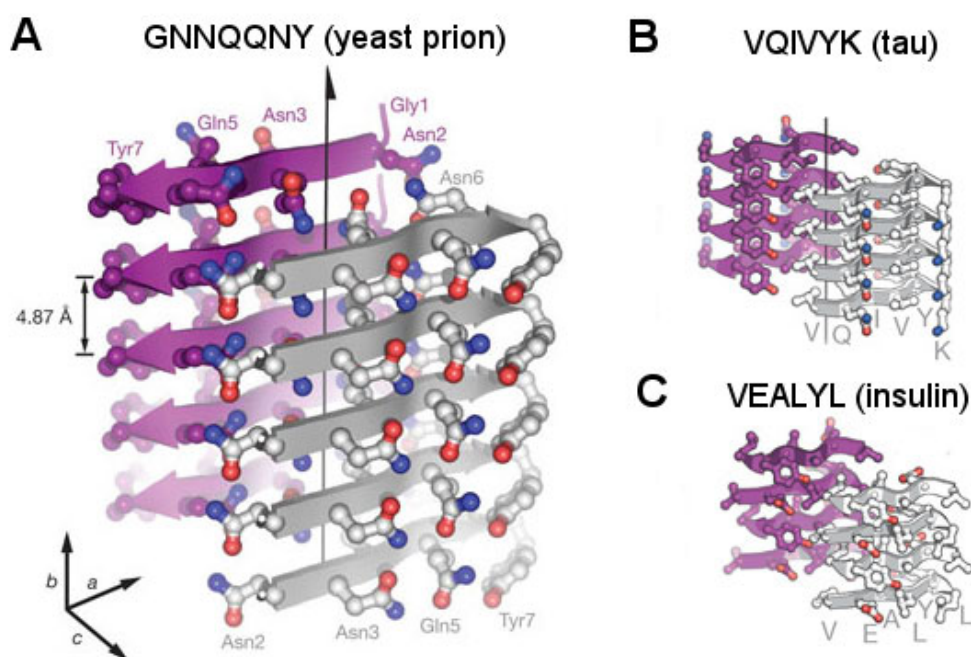


Figure 1.3: **Steric zipper structures of peptide segments from fibril-forming proteins.** Representative steric zipper, pair-of-sheets structures for prion Sup35 (class 1), protein tau (class 1) and insulin (class 7) are shown. The dry interface is between the two sheets, showing the front sheet in silver and the rear sheet in purple. Oxygen atoms are depicted in red and nitrogen atoms in blue, respectively. The steric zipper structures fall in different classes (annotated in parentheses) according to packing orientation of the sheets, as well as their strands. Figure adapted from [82] and [100].

crystalline conformations may not entirely represent those in the amyloid fibrils (e.g. the prominent twist in the fibrillar  $\beta$ -sheet arrangement), as indicated by ssNMR measurements on various crystals and fibrils [103–105]. Although crystals and fibrils often grow together in the same solution [82], the crystallization conditions usually involve several chemical additives to promote crystal formation.

Studying small peptides as simplified model systems for amyloidogenic protein aggregation has led to insights into the underlying universal features of fibril formation and provided high-resolution structures of the fibrillar state.

Theory and computation have facilitated the current understanding of the fundamental biophysical aspects and molecular events in the early stages of amyloidogenic peptide aggregation [106, 107]. Using small peptide fragments and experimentally acquired structural knowledge computer simulation techniques have provided insight into several questions raised, concerning conformational dynamics and thermodynamics of amyloidogenic peptides as well

as aggregation kinetics of oligomeric structures [106, 108–127]. MD simulations appear to be particularly suited to probe the formation of oligomeric species in atomistic detail and inform on the transition pathways between them, on timescales still not amenable to experiment. For example, based on simulations of A $\beta$ <sub>16–22</sub> peptides [124] a dock-lock mechanism has been proposed to explain the commonly implied nucleated growth process of oligomers and fibrils [52, 53, 57, 128]. The authors provide elaborate insight on how monomeric peptides add (dock) to preformed amyloid seed structures in a diffusion-limited process and integrate (lock) by undergoing a substantial conformational conversion [118, 124, 129–131]. Furthermore, the properties of small multimeric aggregates (dimers to decamers) of various amyloidogenic peptide sequences have been studied by atomistic simulations and described as partially ordered, nematic structures, which are subject to rapid fluctuations and large conformational rearrangements [106, 108, 111, 115, 120, 121, 131–133]. The obtained oligomer ensembles were distinct from the monomeric form [108, 115, 130, 134] due to conformational changes associated with an emerging  $\beta$ -sheet structure. These structural transitions come to the expense of intra-peptide interactions [108, 120, 126] and are accompanied by the desolvation of nonpolar surface [106, 108, 114, 126].

## 1.3 Aims and organization of the thesis

The present thesis is concerned with the spontaneous aggregation process of short peptides studied by means of atomistic molecular dynamics (MD) simulations. The principal goal of this work is to characterize the aggregation behavior of so-called steric zipper peptides as a model system for biomolecular aggregation at atomistic detail. This thesis furthermore aims at unveiling the energetic and structural determinants which drive the formation of amyloidogenic peptide assemblies and stabilize the formed aggregates. More specifically, the work focuses on aggregation-prone peptide segments from the tau protein (with the sequence: VQIVYK), insulin (VEALYL),  $\alpha$ -synuclein (GVATVA) and the islet amyloid peptide (NNFGAIL) in order to elucidate the key interactions and mechanistic pathways in the primary aggregation phase as well as in the oligomer formation process.

The thesis is organized in the following way: The principles of molecular dynamics simulations are described in Chapter 2. In addition, methods to derive observables relevant to the work are introduced briefly.

Chapter 3 deals with the validation of different empirical models used in atomistic simulations by investigating the formation propensity of dominant secondary structure elements of popular MD force fields. The systematic study of the relative stabilities of helical and extended conformations in various model peptides has been conducted as a prerequisite for the further work since secondary structure formation has particular relevance for simulations of intermolecular protein and peptide interactions. Consequently, the question of which force field to choose is one of the most important factors in an MD study of peptide aggregation.

Chapter 4 addresses several aspects of steric zipper peptide oligomerization. Novel mapping techniques are presented and employed to reveal the conformational states as well as dynamics and assembly pathways of spontaneous model peptide aggregation. The first part focuses on the initial stage of aggregation, the second part is dedicated to the reorganization dynamics within the oligomeric aggregate states. First, the association and  $\beta$ -sheet formation of the VQIVYK and VEALYL peptides are monitored in multiple simulations, in each case starting from 10 separated monomers with random initial conformations. A detailed analysis of the structure and dynamics of the early oligomerization process is presented. Specifically it was investigated if and how intermediates, which were found along the aggregation pathway, affect the structural evolution of larger oligomers. The second part of chapter 4 deals with the assembly and structural ordering of tau, insulin and  $\alpha$ -synuclein peptide oligomers on the microsecond time scale. The three different peptide systems were probed for common driving forces and structural determinants of the aggregation process.

Here, crystalline conformations of steric zipper peptides serve as an excellent reference for the amyloid fibril cross- $\beta$  spine organization. Hence, a particular aspect of this study was to provide qualitative evidence to which extent early oligomers resemble the proposed fibrillar end-states of amyloidogenic aggregation. The evaluation of the individual energetic contributions of the various intermolecular interactions provide a rational understanding of the forces driving the oligomerization and reveal the molecular basis of the peptide self-assembly process.

Chapter 5 shows that simulations of model peptide aggregation in the vicinity of DMPC bilayers help to elucidate the possible influence of external factors on the aggregation process. Preliminary results and discussions are presented.

A comprehensive overview of the main conclusions of the thesis and future perspectives are summarized in chapter 6.

# Theory and Concepts

---

The present thesis is concerned with the application of molecular dynamics (MD) simulations. These are used to study the spontaneous (and induced) aggregation of short, amyloidogenic model peptides. The following chapter outlines the principles and approximations of the general simulation methodology. In addition, the methods, used to derive observables from MD simulations, are briefly introduced.

## 2.1 Principles and approximations of MD simulations

MD is a deterministic simulation method to describe the dynamics of many-particle systems. Governed by the system Hamiltonian, an MD simulation generates a statistical amount (ensemble) of configurations, i.e. coordinates and velocities of all particles in the system [135, 136]. The theoretical study of biomolecular systems with MD simulations allows for obtaining detailed information, e.g. atomic fluctuations and conformational changes, which are not readily accessible otherwise. This has been an integral part of the successful application of MD simulations in the past and continues to facilitate the understanding and interpretation of experimental data [137–141]. Recently, large macromolecular systems such as the ribosome [142] or viral capsids [143] have been studied with MD simulations. Moreover, folding simulations of smaller systems in the 100  $\mu$ s and ms time scale have been performed [141, 144, 145].

### 2.1.1 Approximations

Classical MD simulations rest on the validity of the following three approximations [136]:

(1) the separation of electronic and nuclear degrees of freedom (Born-Oppenheimer approximation), (2) the assumptions that the nuclear motions can be described by classical Newtonian dynamics, and (3) the application of a classical potential energy function (force field). Conversely, these assumptions give rise to certain limitations of the MD method, which will be discussed in more detail in 2.1.2.

### (1) Born-Oppenheimer approximation

Solving the time-dependent Schrödinger equation is required to describe the dynamics of a physical system of many particles in an exact, quantum-mechanical manner.

$$i\hbar \frac{\partial}{\partial t} \psi(\mathbf{r}, \mathbf{R}) = H \psi(\mathbf{r}, \mathbf{R})$$

Where  $H$  denotes the Hamiltonian of the system,  $\hbar$  the Planck constant divided by  $2\pi$  and  $\psi$  the wave function with  $\mathbf{r}$  and  $\mathbf{R}$  being the Cartesian position vectors of the nuclei and electrons, respectively.

However, evaluating the interactions of larger molecules on this level of theory is not feasible. To simplify the task, in a first good approximation one can make use of the fact that electron dynamics are much faster than the dynamics of the nuclei, because of their orders of magnitude lower mass. Therefore, the relaxation of the electrons within the ground state is fast with respect to the nuclear motion. The Born-Oppenheimer approximation [146] then states that the total wave function can be separated into the nuclear  $\psi_n$  and the electronic wave function  $\psi_e$ ,

$$\psi(\mathbf{r}, \mathbf{R}) = \psi_n(\mathbf{r}) \psi_{e;\mathbf{r}}(\mathbf{R}).$$

The electronic wave function  $\psi_{e;\mathbf{r}}(\mathbf{R})$  no longer depends on the motions of the nuclei. Thus, the dynamics of the electrons are described by the time-independent Schrödinger equation for given positions of the nuclei  $\mathbf{r} = (\mathbf{r}_1, \dots, \mathbf{r}_N)$ . The nuclear dynamics are influenced by the electronic degrees of freedom in the form of a potential energy surface  $V_e(\mathbf{r})$ .

### (2) Classical description of nuclear dynamics

From the gradient of this potential energy surface, the force on particle  $i$  is obtained as function of the configuration  $\mathbf{r}_N$ .

$$-\nabla V_e(\mathbf{r}_1, \dots, \mathbf{r}_N) = m_i \frac{d^2}{dt^2} \mathbf{r}_i$$

The classical description of the nuclear motions in the potential by Newton's second law is the second principal approximation of MD simulations.

$$\mathbf{F}_i = m_i \mathbf{a}_i$$

Knowledge of the force  $\mathbf{F}_i$ , which is exerted on particle  $i$  with mass  $m_i$ , then gives the acceleration  $\mathbf{a}_i$  of the particle in the system.



### (3) Force field

Still further simplification is necessary for sufficient computational efficiency such that  $V_e(\mathbf{r})$  is approximated by a sum of simple potential energy terms [137, 147].

$$\begin{aligned} V(\mathbf{r}) &= V_{\text{bonded}}(\mathbf{r}) + V_{\text{nonbonded}}(\mathbf{r}) \\ &= \sum_{\text{bonds } i} \frac{k_i}{2} (b_i - b_{i,0})^2 \\ &\quad + \sum_{\text{angles } i} \frac{f_i}{2} (\rho_i - \rho_{i,0})^2 \\ &\quad + \sum_{\text{dihedrals } i} \frac{V_i}{2} [1 + \cos(n\varphi_i - \varphi_{i,0})] \\ &\quad + \sum_{\text{impropers } i} \kappa_i (\xi_i - \xi_{i,0})^2 \\ &\quad + \sum_{\text{pairs } i,j} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \end{aligned}$$

This level of abstraction is often based on physicochemical knowledge [12]. Here, the particles are described as spherically symmetric point masses with a partial charge, connected into molecules through covalent bonds. The interactions between the particles is split in two contributions, namely bonded ( $V_{\text{bonded}}$ ) and nonbonded ( $V_{\text{nonbonded}}$ ) interactions. A graphical abstract of the individual components is shown in Figure 2.1.

Bonds describe the connectivity and the bond stretching is modeled using squared harmonic potentials. Bending of bond angles is represented using harmonic potentials as well. The dihedral angle torsional potentials consists of cosine functions with periodicity  $n$ , where barriers between different conformers are defined by a potential  $V_i$ . Finally, improper dihedrals are important for the structural invariance of specific geometries. To keep for example chiral tetrahedral groups or planar groups in the correct conformation, these additional dihedral-angle potentials are used. Van der Waals interactions, the attractive dispersion interactions between atoms, as well as the short-range repulsion due to the Pauli principle are described via a Lennard-Jones type potential and are part of the nonbonded interactions. Since also interactions between the partial charges  $q_i$  of the particles have to be considered, electrostatic interactions are treated according to Coulomb's law. All nonbonded interactions are pair-additive and centro-symmetric, i.e. they depend only on distance.

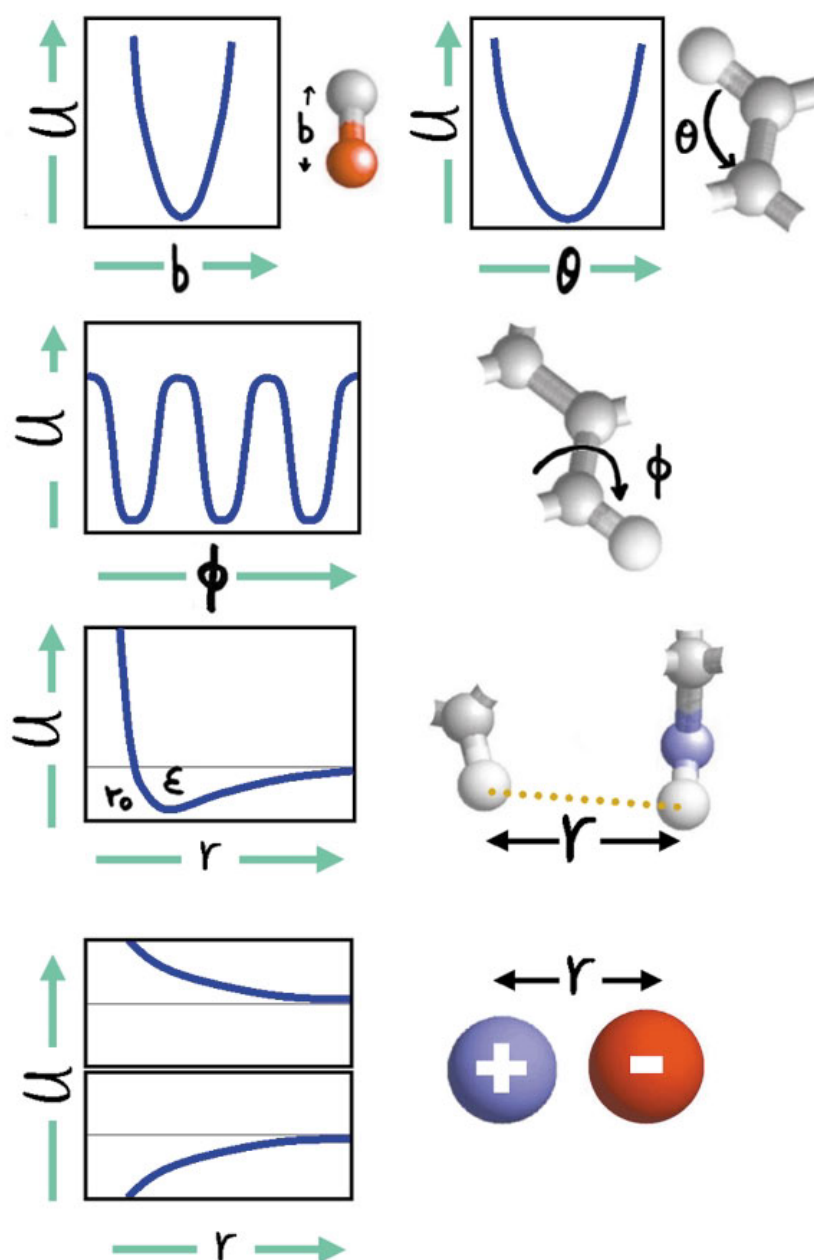


Figure 2.1: **Potential energy terms of classical force fields.** The illustration depicts the individual potential energy terms of a typical biomolecular force field with empirical energy functions. Schematic representations of interaction type and potential form are given for bond stretching, bond angle bending, bond torsions, van der Waals forces and electrostatics, respectively. Figure adapted from [137].

### 2.1.2 Classical biomolecular force fields

MD simulations make use of simple functional forms to calculate conformation-energy relationships at an atomistic level of detail for a molecular species of interest. The entity of both, the interaction energy functions and the set of parameters to be included, is then termed a force field. Force fields have been developed for a wide range of molecular systems. From the various force fields commonly applied to the different classes of biomolecules (proteins, nucleic acids, lipids and carbohydrates); we will briefly introduce aspects of protein and lipid force fields, as they are most relevant to this thesis. There exists exhaustive literature on empirical force fields for simulations of biological macromolecules covering applications, improvements and future perspectives on the subject matter [137–139, 147–153].

**Force field variants and philosophy of parameterization.** The majority of currently used biomolecular force fields has common terms for their potential energy functions. For simulations around room temperature and therefore close to the minimum-energy or equilibrium values, these so called Class 1 force fields adequately and successfully describe the behavior of canonical biomolecules. Among others, the most popular Class 1 force fields used today are in alphabetical order: AMBER (Assisted Model Building with Energy Refinement) [154–156], CHARMM (Chemistry at Harvard Molecular Mechanics) [157, 158], GROMOS (GROningen MOlecular Simulation) [159–162] and OPLS (Optimized Potentials for Liquid Simulations) [163, 164]. The various force fields differ in their philosophy of how to specify and calibrate the set of parameters for the functional formulation of the bonded and nonbonded interactions [148]. Thus, considerable differences in the parameter values are possible. This has implications on the applicability of a given force field to a certain problem at hand. Usually, a combination of experimental data and high level QM calculations on small molecules is used to derive the force field parameters. In contrast to approaches where mostly QM target data was used for the optimization [155], GROMOS force fields, which were mostly used in this thesis, rely heavily on condensed phase experimental data. Furthermore, they are united atom force fields and treat only the heavy atoms and polar hydrogens explicitly. Hydrogen atoms bound to aliphatic carbon atoms are treated implicitly and the nonbonded parameters of these carbon atoms adjusted accordingly. Current AMBER, CHARMM and OPLS force field generations are using all-atom representations of the simulation systems.

Most of the above mentioned force field variants do not contain parameters for lipid and detergent molecules. In particular the structure and dynamics of lipid bilayers represent a significant challenge for empirical force fields because of their complex phase behavior at physiological

temperature and the absence of high-resolution experimental data for the force field calibration. Due to their long hydrocarbon chains, lipid molecules feature a large number of aliphatic hydrogens. Employing a united-atom model therefore leads to a significant gain in computational performance. Accordingly, a large number of MD studies is using united atom models, such as the one by Berger et al. [165].

Since the pioneering MD simulation on BPTI in 1977 [166], studies of protein and peptide dynamics have found widespread application and are a routinely performed task [137]. Moreover, the different models of AMBER, CHARMM, GROMOS96 and OPLS protein force fields seem to have reached converging results for small globular proteins. Improved agreement with experimental solution state data, as well as low deviations from the crystal structures are reportedly achieved [167, 168].

**Remaining challenges.** The quality of the force fields along with their proper implementation may be considered the most important determinant of the accuracy of an MD simulation aside from the sampling problem [149, 169]. Judgment of force field quality is furthermore mainly drawn from the ability to reproduce or predict experimentally accessible properties. Despite the steadily increasing timescales accessible to classical simulations, the obtained results will likely continue to be affected by the choice of force field, solvation model and electrostatic interaction treatment method. Although significant progress has been made, a consensus force field that e.g. provides a correct balance of protein secondary structures is still elusive [167, 170, 171]. An important remaining issue along those lines is the treatment of the conformational energies associated with the torsional degrees of freedom. The latest efforts to improve the accuracy of the popular and commonly used protein force fields AMBER [155, 156, 167, 172], CHARMM [158], GROMOS96 [161] and OPLS [164] focused on refining parameters for the torsional potentials of the polypeptide backbone in order to balance the conformational equilibrium between extended and helical structures. Ultimately, the accuracy of a classical force field on the one hand is determined by the compromise to apply simple energy functions, which allow easy and consistent parameterization and have low computational demand at the same time. For example, the accurate description of chemical bond formation and breaking is not possible with harmonic functions for bond stretching. Furthermore, the classical formulation does not represent electronic reorganizations or quantum effects at very low temperatures [135]. In this regard, polarizable force fields, which explicitly include electronic polarization are considered a major improvement to treat nonbonded interactions in biomolecules in the foreseeable future [173].

## 2.2 Integrating the equations of motion

In MD simulations the potential energy function of the atomic positions is solved by numerical integration of Newton's equations of motion [135, 136]. The equations of motion are deterministic, in the sense that from a set of initial coordinates and an initial distribution of velocities a trajectory can be calculated that predicts the state of the system (positions, velocities and accelerations of the particles) at all other times. The required initial coordinates may be obtained from structures of biomolecules, which were solved experimentally by X-ray crystallography or NMR spectroscopy.

There has been a development of numerous specific MD integration algorithms most importantly aimed at computational efficiency and high accuracy (conservation of energy and momentum). The mathematically equivalent Verlet-type algorithms (the Verlet, velocity-Verlet, and the leap-frog algorithm) are the most commonly used in classical MD today [174]. The integration in GROMACS [175–177] is performed using the leap-frog algorithm [178] and can be summarized in the following scheme.

$$\mathbf{v}\left(t + \frac{\Delta t}{2}\right) = \mathbf{v}\left(t - \frac{\Delta t}{2}\right) + \frac{\mathbf{F}(t)}{m}\Delta t$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{\Delta t}{2}\right)\Delta t$$

Here  $\Delta t$  denotes the time step,  $\mathbf{r}(t)$  the particle's coordinate vector and  $\mathbf{v}(t)$  the respective velocities.

### 2.2.1 Time step, constraints and virtual sites

**Integration time step.** The integration time step size depends on the type of propagation algorithm, as well as the steepness of the potential. The maximal time step is in general inversely proportional to the frequency of the fastest degree of freedom present in the system. For an accurate and stable propagation, the value of the time step must be an order of magnitude smaller than the fastest motion. Exceeding this length of the time step will lead to a rapid accumulation of errors in the numerical integration and will ultimately break down the propagation, as indicated by a drift in the total energy.

The dynamics of complex biomolecular systems take place on different time scales at physiological temperature [9, 179]. The fastest motions proceed on a femtosecond time scale, e.g. intra-molecular fluctuations of bonds, as well as angular and torsional oscillations.

Side chain rotations or loop and domain motions occur on a picosecond and nano- to millisecond time scale, respectively. The biologically relevant conformational changes of biomolecules happen mostly on longer time scales, which range from at least several nanoseconds to often beyond even seconds [9].

**Constrained bond dynamics.** The fast bond vibrations require a small integration step size and thereby severely limit the accessible simulation time. A common way to alleviate this problem is to apply constraint algorithms, like SHAKE [180] or LINCS (LINear Constraint Solver) [181, 182] that implement geometric (holonomic) constraints, while advancing the particle coordinates. Replacing bonds with constraints is a usual procedure in simulations with the GROMACS software package, i.e. fixing the distances between bonded atoms to their equilibrium values. Thereby, the bond vibrations are removed and an increased time step of up to 2 fs can be applied. Furthermore, it has been argued that a constraint is a more faithful representation of the quantum-mechanical ground state of the bond-stretching vibration than the description in terms of a harmonic oscillator [183].

**Virtual interaction sites.** Most of the fastest motions in a biomolecular simulation necessarily involve hydrogen atoms because of their small mass. After constraining all bond-lengths, the next shortest oscillation period in a simulation is therefore the hydrogen bond-angle vibration with 13 fs [179].

Yet another way to achieve an effective increase of the integration time step is thus to define all hydrogen atoms as virtual interaction sites. The virtual site's position is reconstructed from three predefined and nearby heavy atoms that have a fixed orientation with respect to each other, thereby removing all internal high-frequency degrees of freedom. Since only the heavy atom positions are integrated, all forces acting on the hydrogen atom will be redistributed over these particular atoms. However, a slightly different virtual site construction approach is required for hydroxyl or amine groups, since there the rotational freedom has to be persevered. By disregarding these very fast oscillations, the next shortest periods are around 20 fs, which in turn allow for a maximum time step of 5 fs, while still integrating with reasonable accuracy [175, 179].

Note that for the popular explicit solvent models (SPC [184], TIP3P [185], TIP4P [186]) a completely rigid valence geometry is used as a good approximation, therefore the libration of the water molecule with a frequency of 28 fs corresponds to the fastest degree of freedom [179].

### 2.2.2 Ensembles and temperature

The MD algorithm presented so far describes the dynamics of an isolated system solving Newton's equation of motion and therefore should in principle generate a constant NVE (particle number, volume, energy) or microcanonical ensemble of conformations [136]. However, biomolecular processes occur in systems which are in thermal (T) and mechanical (P for pressure) equilibrium with their environment. Consequently, the sampling of canonical (NVT) or isothermal-isobaric (NPT) ensembles is more suitable. In order to achieve these thermodynamic conditions in the simulation system, the solution of the equations of motion has to be modified.

Algorithms for constant temperature MD are called thermostats. Several approaches of temperature control have been presented, while each of them has certain advantages, some of them are more sophisticated and rigorous than others [187–190]. The Berendsen thermostat [187] affects the heat flow by rescaling all particle velocities to adjust the instantaneous kinetic energy of the system to the desired temperature. However, the weak coupling (first-order kinetics) to an external heat bath with given reference temperature suppresses the fluctuations of the kinetic energy and therefore does not generate proper canonical ensembles [191]. The velocity rescaling thermostat [188] is similar to the Berendsen thermostat in the sense that it imposes an exponential relaxation of temperature deviations on the system and will not produce oscillations. At the same time this algorithm generates a correct ensemble, while an additional stochastic term ensures that the correct kinetic energy distribution is obtained.

For constant pressure MD similar algorithms have been developed, such as the weak coupling scheme of Berendsen [187] but they will not be discussed here.

## 2.3 Trajectory analysis methods

### 2.3.1 Principal Component Analysis

The basic ideas of principal component analysis (PCA) [192] and its application to conformational ensembles of biomolecules are shortly reviewed here.

Mathematically, PCA is defined as an orthogonal linear transformation that re-expresses a set of data in a new coordinate system such that the greatest variance by any projection of the data lies on a linear combination of the original coordinate basis. These principal components are orthogonal to each other and describe the data successively best. Most importantly, PCA can be used for dimensionality reduction in a data set since it retains those aspects of the data that contribute most to its variance. Therefore, PCA is valuable to identify patterns in high-dimensional data sets and able to reveal a hidden underlying structure, as well as trends or highlight similarities and differences, if present. Consequently, PCA has found widespread applications in various fields, for example neuroscience, data compression, pattern recognition, data visualization, and image processing [193].

Briefly, when a PCA is performed the covariance matrix for a set of data with given dimensions  $n$  is constructed and diagonalized. This is done by calculating the covariance value between any two dimensions in the data set, after the mean from each of the data dimensions was subtracted.

$$C^{m \times n} = (c_{i,j}, c_{i,j} = \text{cov}(\text{dim}_i, \text{dim}_j))$$

Solving the eigenvalue problem for  $C$  yields a set of  $n$  eigenvalues and eigenvectors, where the eigenvector with the highest eigenvalue is the first principal component of the data set. These perpendicular eigenvectors can now be used to re-express the data.

PCA has been shown to be particularly suited for the theoretical description of collective motions of internal protein dynamics [194, 195]. That is, the detection of relevant degrees of freedom of biomolecules from trajectories or structure ensembles obtained e.g. with MD simulations or experimental structure determination techniques. Termed as quasi-harmonic analysis, PCA was first applied to protein dynamics in order to estimate the configurational entropy of macromolecules [196]. Later, PCA-based analysis has shown that protein dynamics are governed by a few collective degrees of freedom, covering most of the positional fluctuations in the sampled configurational space [194, 197]. It was found that as much as 90% of the atomic displacement occurs along only a small subset (5-10%) of collective motions of one or various parts of the biomolecule under investigation [194]. Furthermore, it was shown that these essential degrees of freedom or motions also describe most of the functional dynamics of



proteins [194, 197]. However, one should keep in mind that PCA by construction only takes linear correlations between atomic displacements into account, potentially overlooking nonlinear correlations.

In the following I will sketch the formalism of a PCA applied to an ensemble of protein structures, e.g. derived from an MD simulation trajectory. For a general derivation and detailed mathematical treatment of PCA I refer to [193, 194, 198, 199]. Applying PCA to a structure ensemble of a biomolecule can be regarded as a multi-dimensional linear least squares fit procedure in configuration space. As a first step, the overall rotation and translation of each configuration of the ensemble has to be removed, since one is only interested in the internal dynamics. Therefore, a superposition of every configuration of the ensemble to a common reference structure is carried out. The Cartesian coordinate vector  $\mathbf{x} \in \mathbb{R}^{3N}$  then describes members of the fitted structure ensemble for a molecule consisting of  $N$  atoms. The construction and diagonalization of the variance-covariance matrix of positional fluctuations of the atoms then reads as follows.

$$C = \langle (\mathbf{x} - \langle \mathbf{x} \rangle) (\mathbf{x} - \langle \mathbf{x} \rangle)^T \rangle$$

Here  $\langle \mathbf{x} \rangle$  denotes the ensemble average of  $\mathbf{x}$ . The principal components are retrieved by transforming the symmetric matrix  $C \in \mathbb{R}^{3N} \times \mathbb{R}^{3N}$  by an orthogonal coordinate transformation  $T$  into a diagonal matrix  $\Lambda$ , in which the eigenvalues  $\lambda_i$  of  $C$  enter as the diagonal elements.

$$\Lambda = T^T C T \quad \text{or} \quad C = T \Lambda T^T$$

Consequently, the  $i$ th column of  $T$  then contains the normalized eigenvectors (principal component)  $\mu_i$  of  $C$  corresponding to the eigenvalues  $\lambda_i$ . Then the original configurations of the structural ensemble can be projected onto each of the principal components to yield the principal coordinates  $p_i$ .

$$p_i = \mu_i \cdot (\mathbf{x} - \langle \mathbf{x} \rangle)$$

The contribution of each principal component to the total fluctuation can now be read from the mean square positional fluctuation along the respective eigenvectors. Accordingly, the variance  $\langle p_i^2 \rangle$  equals the eigenvalue  $\lambda_i$ .

For visualization purposes one can transform these projections back to the cartesian coordinate space.

$$\mathbf{x}_i^p = p_i \cdot \mu_i + \langle \mathbf{x} \rangle$$

### 2.3.2 Secondary structure and assignment

The peptide backbone of each amino acid residue has three conformational degrees of freedom, namely the main chain torsions  $\omega$  (peptide bond),  $\varphi$  (C-N-C $_{\alpha}$ -C) and  $\psi$  (N-C $_{\alpha}$ -C-N). The planarity of the peptide bond (partial double bond character) restricts  $\omega$  to 180 degrees. For a cis peptide bond in proline residues one can find also 0 degrees. The accessible conformational space even for a small protein will therefore be enormously large, let alone for proteins of tens or even hundreds of residues [2].

However, the local (secondary) structure of the peptide main chain is very ordered in the folded state and organized in repeating patterns [12]. This is due to a tight local packing and the extensive hydrogen bond formation between (intra-)backbone donor (NH) and acceptor (C=O) atoms. The different amino acid propensities for certain secondary structure conformations indicate the important role of the side chains in defining the regular secondary structure motifs (e.g.  $\beta$ -sheets are rich in valine, isoleucine and poor in glycine and proline) [200].

As in many other biological processes, the main structural changes in peptide folding and self-association occur on the level of secondary structure, with a particular prevalence of  $\beta$ -sheet conformations in the latter case [4, 201]. Given the atomic coordinates of a protein or peptide structure one seeks to obtain a consistent assignment of secondary structure elements in an automated fashion. I therefore briefly review the basic principles of the two popular assignment methods DSSP [202] and STRIDE [203], which were used for secondary structure classification in this thesis.

**DSSP (Dictionary of Secondary Structure of Proteins).** The DSSP algorithm [202] assigns secondary structure elements purely based on calculations of backbone-backbone hydrogen bond energetics. The hydrogen bond coulomb energy  $E$  is approximated by the term

$$E = q_1 q_2 \left\{ \frac{1}{r_{ON}} + \frac{1}{r_{CH}} - \frac{1}{r_{OH}} - \frac{1}{r_{CN}} \right\} \cdot f,$$

where the partial charges  $q_1 = 0.42$ ,  $q_2 = 0.20$  and the constant  $f = 1390$  kJ/mol. To identify a hydrogen bond, a cutoff with an energy  $E$  of less than -2.1 kJ/mol is applied.

Eight types of secondary structure are recognized by DSSP depending on the pattern of hydrogen bonds. A repetitive sequence of hydrogen bonds in which the donor residue is three, four, or five residues later in the backbone define a  $3_{10}$ -,  $\alpha$ - and  $\pi$ -helix, respectively. Hydrogen bond pairs in  $\beta$ -sheet structures are classified as parallel and antiparallel bridges;

extended (repeating) sets of hydrogen bond pairs of the same type are equivalent to a  $\beta$ -sheet. Remaining types are turn (featuring a hydrogen bond typical of a helix) and bend motifs for a region of high curvature.

**STRIDE (STRuctural IDentification).** The assignment of individual secondary structural elements in STRIDE [203] are based on a more complex expression of hydrogen bond energy and in addition to DSSP, on empirical  $\varphi$ - $\psi$  torsional angle criteria.

The term for the total hydrogen bond energy  $E_{hb}$  reads

$$E_{hb} = E_r \cdot E_t \cdot E_p,$$

with  $E_r$  being a distance-dependent 8-6 potential similar to a Lennard-Jones potential with optimal distances of 3 Å (NO) for the backbone hydrogen bond.  $E_t$  and  $E_p$  are angular dependent factors, which define the optimized hydrogen bond geometry. The terminal residues are assigned with secondary structure through reliance on torsional angles. The individual secondary structural elements are mapped into the same classes (helix, sheet, coil) as those reported by DSSP.



## **Secondary structure propensities in peptide folding simulations: A systematic comparison of molecular mechanics interaction schemes**

---

- [D. Matthes](#) and B. L. de Groot, Biophys. J. 2009, 97, 599-608.

### **Summary**

A systematic study directed toward the secondary structure propensity and sampling behavior in peptide folding simulations with eight different molecular dynamics force field variants in explicit solvent is presented. It reports on the combinational result of force field, water model, and electrostatic interaction schemes and compare to available experimental characterization of five studied model peptides in terms of reproduced structure and dynamics. The total simulation time exceeded 18  $\mu$ s and included simulations that started from both folded and extended conformations. Despite remaining sampling issues, a number of distinct trends in the folding behavior of the peptides emerged. Pronounced differences in the propensity of finding prominent secondary structure motifs in the different applied force fields suggest that problems point in particular to the balance of the relative stabilities of helical and extended conformations.

## 3.1 Introduction

Molecular dynamics (MD) simulations are routinely utilized to study the folding dynamics of peptides and small proteins as well as biomolecular aggregation. The critical constituents of such molecular mechanics studies are the validity of the underlying physical models together with the assumptions of classical dynamics and a sufficient sampling of the conformational space. In order to verify and validate simulation results, a careful comparison of the simulation outcome directly to experimental data is mandatory (e.g., obtained by NMR, CD or infrared spectroscopy) [204]. Comprehensive reports on applications, improvements and remaining challenges of empirical force field based simulation methods, the choice of water model and electrostatic interaction schemes to study biomolecular systems have been discussed in the literature [147–152].

Within the framework of MD force fields, particular importance is directed to the consistent and proper parameterization of the atomistic interactions, with the functional formulation of the bonded and nonbonded forces often similar among nonpolarizable MD schemes. The latest efforts to improve the accuracy of the popular and commonly used force fields AMBER [155, 156], CHARMM [158], GROMOS96 [161] and OPLS [164] mainly focused on refining parameters for the torsional potentials of the protein backbone in order to balance the conformational equilibrium between extended and helical structures.

A recent comparative study using selected variants of the AMBER, CHARMM, GROMOS96 and OPLS force fields reported on converging results for folded proteins between the different compared models. It was suggested that there is an apparent consensus view of protein dynamics [168]. In that study simulations of relatively short lengths were performed and the natively folded state was used as starting point, possibly biasing the results [168].

For folding simulations such a systematic test has not been carried out so far, although with growing computer power several approaches towards the *in silico* folding problem for peptides and small proteins, both using an implicit or explicit representation of the solvent environment, have been presented [108, 205–210]. Given an efficient sampling of conformational space and access to sufficient simulation timescales, one should expect to sample conformational ensembles close to the natively most populated states in solution, even when starting from peptide conformations away from the native structure. Hence, the application of biomolecular simulations offers the unique opportunity to study and predict complex processes in detail that underlie the protein folding thermodynamics and kinetics. For instance, the early events of peptide and protein folding, marked by established and stabilized secondary structure motifs [201].

A realistic preferential formation and representation of secondary structure is therefore a critical prerequisite for the successful study of *in silico* folding and aggregation. Thus the question of overall peptide folding representation in different force fields prompted us to investigate the folding behavior and secondary structure formation at the microsecond timescale of a number of prototypic peptides in different MD force fields.

Here, the results of peptide folding and secondary structure formation for five model peptides (two  $\beta$ -hairpins, two  $\alpha$ -helical peptides and the Trp-cage) in five state of the art force fields and different schemes for calculating electrostatic interactions are presented. Extensive MD simulations in explicit water, starting from both extended and prefolded structures are presented that address the folding thermodynamics and sampling characteristics of the different interaction schemes.

## 3.2 Methods

### Model peptides

MD simulations of five isolated peptides were performed, which adopt different well-defined, stable secondary or tertiary structures in solution (Table 3.1). The chosen peptides are considered as minimalistic model systems to probe the different force fields for either  $\alpha$ -helical or  $\beta$ -sheet folding propensity.

**Chignolin.** The 10 residue  $\beta$ -hairpin peptide Chignolin was designed by statistical considerations [211] and characterized by nuclear magnetic resonance (NMR) and circular dichroism (CD) experiments in solution. The molar fraction of folded peptide at 300 K was determined to approximately 60 %. Chignolin was also subject of numerous computational studies and the NMR-structure was reproduced by MD from the extended state [212–214].

**Mbh12.** The sequence for the Mbh12 peptide stems from a combinatorial approach to de novo design a stable  $\beta$ -hairpin fold in a linear peptide [91]. CD and NMR experiments identified a percentage of  $\beta$ -hairpin structure higher than 66% at 278 K.

**Trp-cage (Tc5b).** The Trp-cage is a 20 residue peptide designed and first characterized by Neidigh et al. [215]. The sequence was optimized by mutation and folds into a

compact structure consisting of a N-terminal  $\alpha$ -helix, a  $3_{10}$ -helix and a hydrophobic core formed by Trp, Tyr and Pro residues. The Tc5b mutant used in this study was found to be folded to  $> 95\%$  in aqueous solution and melts at 315 K as determined by CD and NMR [215]. The Trp-cage is considered a model system for folding simulations, displaying two-state folding properties and a folding time in the microsecond regime [216]. Tc5b has been thoroughly studied by molecular dynamics and numerous reports of folding simulations since then have contributed to the understanding of structural features, which govern fold stabilization and facilitation of the fast folding [207, 209, 217, 218].

**Fs<sub>21</sub>.** Originally designed by Lockhart et al. to assess electrostatic interactions in  $\alpha$ -helices, the 21 residue polyalanine peptide Fs-NH<sub>2</sub> became a popular model system for helix-coil transition and is well studied, both experimentally [219–223] and by MD simulations [210, 224–227]. Based on CD spectra signatures the peptide was reported to be  $> 90\%$  helical in aqueous solution at 273 K [219]. Controversial data is available concerning the melting temperature. The melting point was reported at 308 K [219] and 303 K [221] in studies using CD, while infrared spectroscopy experiments observed a larger melting temperature of 334 K, but cannot distinguish between  $\alpha$ - and  $3_{10}$ -helix [220]. Note that depending on the experimental setup different N-terminal capping groups were used.

**Agd1 (de novo).** It was decided to include another helical peptide and constructed a short peptide sequence with a high helical propensity using the AGADIR prediction algorithm [201], which is based on the empirical analysis of experimental data. A 13 amino acids long peptide sequence (here termed Agd1) was derived, whose predicted helical content with protected N- and C-termini under conditions of pH 7 and 300 K is calculated to be 50 %.

Table 3.1: Experimental characterization of the model peptides.

Peptide	Secondary structure (Experimental conditions)	Experimental technique	Reference
Chignolin	$\beta$ -hairpin: 60 % (300 K, pH 5.5)	CD & NMR	Honda et al. [211]
Mbh12	$\beta$ -hairpin: $66 \pm 4\%$ (278 K, pH 5)	CD & NMR	Pastor et al. [91]
Trp-cage	$\alpha$ -helix: 30 % (300 K, pH 7)	CD & NMR	Neidigh et al. [215]
Fs <sub>21</sub>	$\alpha$ -helix: 90 % (Fs-NH <sub>2</sub> , 273 K, pH 7)	CD	Lockhart et al. [219]
	$\alpha$ -helix: 55 % (278 K)	Raman	Asher et al. [222, 223]
	$\alpha$ -helix: 50 % (MABA-Fs-NH <sub>2</sub> , 300 K, pH 7)	CD	Thompson et al. [221]
	helical: 68 % (300 K, pH 7)	AGADIR prediction algorithm [201]	
Agd1	helical: 50 % (300 K, pH 7)	AGADIR prediction algorithm [201]	



## Setup and simulation procedure

The simulations were categorized according to the name of the peptide and the starting model. An overview of the simulated peptide systems is given in Table 3.2.

Table 3.2: Summary of performed simulations with the respective number of residues ( $N_{res}$ ) and the expected net charge in water at neutral pH ( $N_{cha}$ ) for each model peptide.

Peptide	Sequence ( $N_{res}/N_{cha}$ )	Simulation name	Starting structure	Simulation length [ns]
Chignolin	$\text{NH}_3^+ \text{-GYDPETGTWG-CO}_2^-$ (10/2-)	CHI CHI.REF	extended $\beta$ -hairpin (PDB: 1uao)	8 x 250 8 x 250
Mbh12	$\text{NH}_3^+ \text{-RGKWTYNGITYEGR-CO}_2^-$ (14/2+)	MBH MBH.REF	extended $\beta$ -hairpin (PDB: 1k43)	8 x 250 8 x 250
Trp-cage (Tc5b)	$\text{NH}_3^+ \text{-NLYIQWLKDGGPSSGRPPPS-CO}_2^-$ (20/1-)	TC5B.REF	$\alpha$ -helix/turn (PDB: 1l2y)	8 x 250 8 x 3 x 30
Fs <sub>21</sub>	Ace-A <sub>5</sub> [AAARA] <sub>3</sub> A-NME (21/3+)	FS21 FS21.REF	extended $\alpha$ -helix	8 x 250 8 x 250
Agd1 (de novo)	Ace-EVLMKVLMEIYLK-NH <sub>2</sub> (13/0)	AGD1 AGD1.REF	extended $\alpha$ -helix	8 x 250 8 x 250

To overcome limited sampling and possible bias imposed by the starting structure, each peptide was simulated starting from both a folded and an unfolded (extended) conformation. The initially extended structures were obtained by constructing the respective peptide chain with PyMOL [228] by imposing an all-trans geometry to every backbone dihedral. The folded conformations were either obtained from the Protein Data Bank (first model of the respective NMR-ensembles) 1uao (Chignolin); 1k43 (Mbh12); 1l2y (Tc5b) or in case of the helical peptides (Fs<sub>21</sub>, Agd1) by building an  $\alpha$ -helical conformation. The Fs<sub>21</sub> and Agd1 peptides were capped with acetyl groups at the N-terminus. To cap the C-terminal site, *N*-methyl- (NME) and amino groups were used, respectively. The terminal residues of the Chignolin and Mbh12 peptides were considered charged. The protonation state of the peptides was according to the one in solution at pH 7, counter-ions ( $\text{Na}^+$ ,  $\text{Cl}^-$ ) were added to adjust for excess charges. The MD simulations of the respective peptide monomers and subsequent analysis were carried out using the GROMACS software package (version 3.3.1) [175, 176, 229]. Each of the production runs after equilibration was 250 ns long. In case of the Tc5b peptide, three additional 30 ns runs per force field were carried out.

**Force field settings.** The focus was on the comparison of force field variants as implemented in the GROMACS simulation software suite: GROMOS96 43A1 [159, 160], GROMOS96 53A6 [161, 162], OPLS-AA/L [163, 164], AMBER03 (ff03) [155] and AMBER99SB (ff99SB) [154, 156]. The input parameters were chosen according to the original publications of the developers to ensure a systematic comparison between the tested force fields.

The AMBER99SB force field (ff99SB) [156] is based on the AMBER99 force field [154], with new parameters for backbone dihedrals to achieve a better balance of secondary structure elements. The AMBER03 force field (ff03) [155] is also a variant of the AMBER99 potential with rederived charges and main-chain torsion potentials, introducing unique main-chain charges for each amino acid. The OPLS-AA/L force field (OPLS) [163] was parameterized for liquid state thermodynamics and later improved by refitting the torsion/backbone parameters from quantum chemical calculations [164]. The GROMOS96 (G96) united atom force field is continuously improved and refined. Popular examples are the G96 43A1 [159, 160] and the G96 53A6 force field [161, 162], with the latter being the newest parameter set with adjusted partial charges to better reproduce hydration free enthalpies in water. All simulations were carried out using electrostatic schemes as originally used for development and in addition, the Particle Mesh Ewald (PME) [230, 231] method for comparison.

The long-range and slowly decaying coulomb potential still poses the most challenging problem to solve in MD simulations, both in terms of accuracy and computational effort. Several methods have been introduced to make the calculation of the nonbonded interactions in biomolecular systems feasible. Cutoff methods simply neglect the contributions from the electrostatic interactions beyond a certain cutoff distance. Satisfactory accuracy without making severe errors is not expected, even for large cutoff distances. With the continuum based reaction-field (RF) [232] approach, charge-charge interactions are calculated explicitly within a sphere of a given radius and beyond by assuming a homogeneous medium with a certain dielectric constant. The PME method derives the electrostatic forces in a periodic system by splitting the potential into two separate summation terms. These terms describe the short-range (direct space) and long-range (reciprocal space) part of the coulomb potential. Both rapidly converge and therefore can be truncated at a cutoff without losing much accuracy. For an efficient calculation of the long-range part, the continuous charge positions are substituted by a mesh-based charge density that a fast fourier transformation can be used to perform the necessary transformation to reciprocal space.

## 3.2 Methods

The MD simulations for each peptide system were performed using eight combinations of force fields, water models and methods for treating the long-range electrostatic interactions. The cutoff distances for the nonbonded interactions which were used with the different force fields are summarized in Table 3.3.

Table 3.3: Force field setting and treatment of nonbonded interactions.

System	Force field	Electrostatic scheme	VdW scheme	Water model
ff03-PME	AMBER03 [155]	PME (1.0 nm)	Cutoff (0.8 nm)	TIP3P
ff99SB-PME	AMBER99SB [156]	PME (1.0 nm)	Cutoff (0.8 nm)	TIP3P
43A1-RF	GROMOS96 43A1 [159, 160]	Reaction-Field (1.4 nm)	Cutoff (1.4 nm)	SPC
43A1-PME	GROMOS96 43A1	PME (0.9 nm)	Cutoff (1.4 nm)	SPC
53A6-RF	GROMOS96 53A6 [161, 162]	Reaction-Field (1.4 nm)	Cutoff (1.4 nm)	SPC
53A6-PME	GROMOS96 53A6	PME (0.9 nm)	Cutoff (1.4 nm)	SPC
OPLS-Cutoff	OPLS-AA/L [163, 164]	Cutoff (1.4 nm)	Cutoff (1.4 nm)	TIP4P
OPLS-PME	OPLS-AA/L	PME (0.9 nm)	Cutoff (1.4 nm)	TIP4P

All simulations were carried out with electrostatic schemes as originally used for force field development. In addition, the Particle Mesh Ewald (PME) [230] method was used for comparison, as it is nowadays common practice to apply PME in conjunction with force fields like OPLS and GROMOS96 (G96) originally designed using cutoff and reaction-field (RF) [232], respectively. In this study the electrostatic interactions with PME were calculated with a grid spacing of 0.12 nm. The relative tolerance at the cutoff was set at  $10^{-6}$ , electrostatic interactions for a distance smaller than the real space cutoff were calculated explicitly. For the calculations with reaction-field (G96 force fields) the relative dielectric permittivity outside the cutoff sphere was set to  $\epsilon=54$  (SPC water) [233].

To set up the simulation system, each peptide was placed in a periodic truncated octahedral box solvated with explicit water. The distance between solute and box was chosen to be at least 1.5 nm on all sides.

The SPC water model [184] was used for the simulations with the G96 43A1/53A6 force fields, the TIP4P solvent model [186] was applied when using the OPLS-AA/L force field and TIP3P [185] for AMBER03 and AMBER99SB. The system was subsequently energy minimized using steepest descent. Initial velocities were taken from a Maxwell distribution at 300 K. Rigid bond constraints were chosen, providing a reasonable representation of the covalent bond geometry of the studied timescales [183]. All protein bonds were constrained with the SHAKE algorithm [180] for simulations with G96 43A1/53A6 and OPLS-AA/L, and LINCS [181] when using AMBER03 and AMBER99SB.

An integration time step of 2 fs was chosen. Water was constrained using SETTLE [234].

Neighbor lists for nonbonded interactions were updated every 5 steps for the G96 and OPLS, and every 10 steps for simulations with the AMBER force fields. Berendsen coupling algorithms [187] were applied to the simulation system. The temperature was kept constant by weakly ( $\tau=0.1$  ps) coupling the system to a temperature bath of 300 K. Likewise, the pressure was kept constant by coupling the system to a pressure bath of 1 bar ( $\tau=1$  ps).

## Analysis

For assessment of secondary structure type and content, the DSSP definition introduced by Kabsch and Sander [202] was used. The data was averaged over both simulations starting from the extended and the reference structure. The first 10 % of the 250 ns and the first 50 % for the 30 ns runs of each trajectory were omitted to reduce the bias of the respective starting model.

**Nuclear Overhauser Enhancement (NOE).** The NOE distance restraint sets for Chignolin, Mbh12 and Tc5b (available from the PDB) were used to calculate ensemble sum-averaged violations ( $r^{-6}$ ) of NOE distances in the MD ensembles. The violations were calculated based on a set of 250 conformations (one snapshot per ns) taken from each trajectory, respectively. To account for the different representations of the peptides in the used force fields, each of the 250 representative structures was converted to a OPLS topology with the pdb2gmx program prior to analysis, adding explicit proton positions to aliphatic carbon atoms.

**Principal Component Analysis (PCA).** The PCA [194] was carried out over the combined trajectories starting from different initial structures (extended and reference) for each peptide in each force field test setting. The covariance matrix of atomic displacement was calculated and diagonalized for the coordinates of main-chain and  $C_{\beta}$  atoms. All structures were superimposed to the respective reference conformation prior to analysis.

## 3.3 Results

### 3.3.1 Structural properties

The conformations sampled during the simulations were compared to experimental data from CD and NMR measurements.

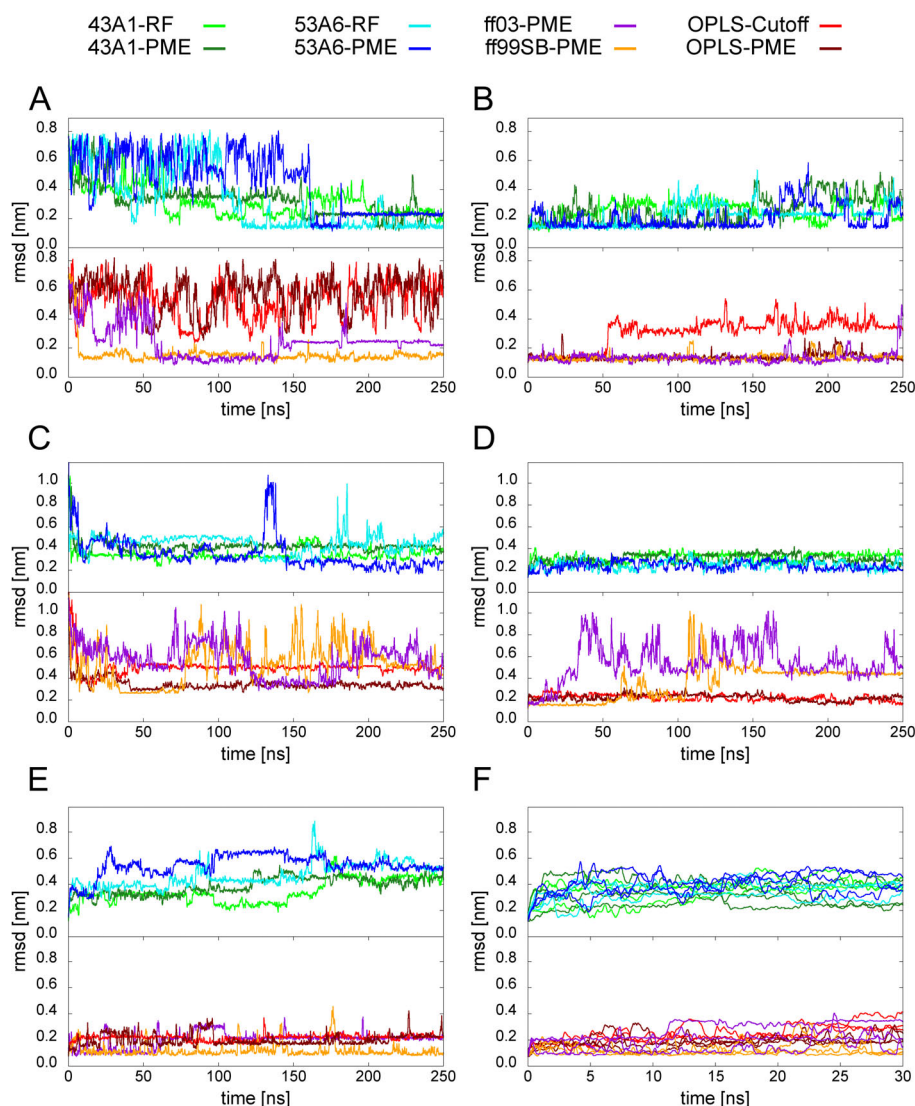


Figure 3.1: **RMSD to reference structures.** RMSD of the main-chain and  $C_{\beta}$  atoms with respect to the NMR reference structure as function of the simulation time for Chignolin, Mbh12 and Tc5b. RMSD curves for the simulations CHI (A), CHI.REF (B), MBH (C), MBH.REF (D), TC5B.REF - 250 ns (E) and TC5B.REF - 30 ns (F) were smoothed.

**Root Mean-Square Deviation (RMSD).** An analysis of the root mean-square deviations (RMSD) for the peptides with experimentally determined native state (Chignolin, Mbh12 and Tc5b) was performed. Fig. 3.1 shows the RMSD for the main-chain and  $C_{\beta}$  atom coordinates of these three peptides to the respective NMR reference structure as function of simulation time. A low RMSD ( $< 0.2$  nm) denotes a conformational state which is close to the one observed in the experiment.

The focus will be on the RMSDs for the  $\beta$ -hairpin-peptides (Chignolin, Mbh12). For the simulations starting from the folded reference state of Chignolin (CHI.REF), only small structural deviations over the total range of simulation time in most of the different force fields were found, indicating a stable fold. Particularly low deviations from the reference structure are found with ff99SB, OPLS-PME for the CHI.REF simulations. The hairpin structure is partly lost for the simulations with ff03, 43A1 as well as OPLS-Cutoff. Different results were obtained for the simulations started from an extended peptide chain (CHI). The Chignolin peptide is folded to a structure very close to the experimentally determined one, for example with the ff99SB and ff03, respectively (Fig. 3.1A and 3.1B). A high structural stability from the prefolded state and fast sampling towards the reference therefore led to the overall smallest RMSDs in both AMBER force fields, irrespective of the initial configuration. Also in all G96 simulations the correct folded state was adopted (Fig. 3.1A and 3.1B). The OPLS force fields with both PME or cutoff, however did not sample the folded configuration within 250 ns. This is in contrast to the result that the hairpin structure was found to be stable for the combination of OPLS and PME.

A different scenario arises from the simulations starting from the extended chain of the Mbh12 peptide (MBH). Hairpin formation took place rather fast in all studied G96 and OPLS force field variants. The peptide conformations were also stable from the reference structure for these force fields. The MBH simulations in both AMBER force fields, however, did not sample stable hairpin structures, and the folded state was only transiently visited. Instead, partly  $\alpha$ -helical conformations were sampled, which resulted in large RMSDs. Interestingly, this was also observed in the simulations starting from the NMR reference structure. The hairpin peptide did not unfold in any of the other MBH.REF simulations (Fig. 3.1C and 3.1D). The Tc5b simulations were started only from the reference NMR structure in the different force field variants. In addition to one simulation of 250 ns, three independent runs of 30 ns were conducted in order to check for statistical significance of the results (Fig. 3.1E and 3.1F). Reproducible low RMSDs of around 0.12 nm were found for the structural ensembles captured by the ff99SB force field. In contrast, large RMSDs were observed in the simulations using the G96 force fields. Furthermore, substantial unfolding events occurred already within the first 30 ns in most of the G96 trajectories (Fig. 3.1F).

**Nuclear Overhauser Enhancement (NOE).** An additional structural comparison to experimental data for both  $\beta$ -hairpin peptides and the Trp-cage can be drawn based on NOEs between proton pairs or groups of protons obtained from NMR experiments.

Table 3.4: Sum of NOE distance violations as ensemble average over 250 structures.

System	sum of NOE violations [nm]				
	CHI	CHI.REF	MBH	MBH.REF	TC5B.REF
ff03-PME	0.99	0.72	7.83	3.87	3.10
ff99SB-PME	0.66	0.64	7.61	1.17	1.95
43A1-RF	1.30	1.17	3.04	1.16	6.09
43A1-PME	0.83	0.59	9.19	1.05	4.54
53A6-RF	0.98	0.72	7.09	0.83	7.25
53A6-PME	2.69	0.67	1.22	0.87	10.03
OPLS-Cutoff	5.32	1.34	10.27	1.11	3.21
OPLS-PME	12.04	0.61	5.29	0.95	2.73

Measured NOEs are usually translated into proton-proton distance ranges. Exceeding the upper limit of such a distance range was counted as NOE violation. For the studied peptides the NOE restraints included short-range restraints between atoms on neighboring residues and intra-residual atoms, medium- and long-range restraints. The violations as an ensemble averaged sum were determined. Results are shown in Table 3.4. As a consequence of slow or incomplete sampling towards the native structure, in general a larger total violation for the simulated ensembles starting from the extended conformations was found, as compared to the simulations starting from the prefolded state.

From the calculated interproton distances marginal deviations were observed from the experimental structures for all hairpin peptides with the 53A6-PME and the 43A1-RF force field variants, irrespective of the peptide's starting conformation. Except for the 53A6-PME, all force fields showed rather large structural deviations for the Mbh12 peptide when starting from the extended state. The partially helical Trp-cage, however, is not represented correctly in any of the G96 force fields, especially the 53A6 versions yielded large NOE violations. The ff99SB force field showed fewest violations of distance restraints for the Trp-cage simulations with an ensemble sum-average of less than 2 nm, corresponding to a stable fold over the simulated 250 ns (Table 3.4). This is in line with low values of summed NOE violations found with ff99SB, OPLS-PME and to some extent with ff03 for both, hairpin and helical peptides, in trajectories starting from the NMR reference structure (Table 3.4).

**Secondary structure propensity.** The average populations of various secondary structure elements were calculated as listed for the individual simulations in Table 3.5, 3.6 and 3.7. Each listed fraction represents a mean over 450 ns for the peptides Chignolin, Mbh12, Fs<sub>21</sub> and Agd1 and 270 ns for the Tc5b in each specified force field configuration, respectively.

The highest fractional population of turn and  $\beta$ -sheet conformations were found in the



Table 3.5: Averaged secondary structure content obtained from a DSSP analysis of the hairpin peptide simulation.

System	$\alpha$ -helix		$3_{10}$ -helix		$\beta$ -sheet		$\beta$ -bridge		turn	
	CHI	MBH	CHI	MBH	CHI	MBH	CHI	MBH	CHI	MBH
ff03-PME	0.0	5.6	0.2	3.7	23.0	0.8	5.3	0.5	33.4	12.6
ff99SB-PME	0.0	0.4	0.1	0.9	26.2	16.9	5.9	1.3	37.1	14.5
43A1-RF	0.4	0.0	0.1	0.2	10.8	27.2	7.0	2.6	18.0	4.8
43A1-PME	0.1	0.0	0.3	0.1	8.8	45.9	0.3	2.1	22.6	10.1
53A6-RF	0.1	0.0	0.1	0.1	19.6	37.6	4.7	2.0	25.7	8.1
53A6-PME	0.0	0.0	0.2	0.0	15.5	34.3	4.6	2.0	24.7	5.1
OPLS-Cutoff	0.0	0.0	1.9	0.0	1.7	27.0	2.1	1.4	15.8	14.0
OPLS-PME	0.0	0.0	0.0	0.0	11.6	32.3	3.6	3.0	18.6	5.7

Table 3.6: Averaged secondary structure content ( $\alpha$ -,  $3_{10}$ - and  $\pi$ -helix) obtained from a DSSP analysis of the helical peptide simulation.

System	$\alpha$ -helix			$3_{10}$ -helix			$\pi$ -helix		
	TC5B	FS21	AGD1	TC5B	FS21	AGD1	TC5B	FS21	AGD1
ff03-PME	33.8	74.4	32.5	15.0	0.8	9.2	0.0	0.0	0.0
ff99SB-PME	34.2	0.6	5.7	15.1	2.5	6.8	0.0	0.0	0.0
43A1-RF	26.8	7.8	25.7	0.4	0.6	0.0	0.2	0.3	20.0
43A1-PME	25.9	16.8	3.6	0.1	0.3	0.1	0.0	0.0	25.8
53A6-RF	7.1	0.3	2.7	0.3	0.1	0.1	0.1	0.0	1.8
53A6-PME	3.6	0.1	0.3	0.0	0.0	0.0	0.5	0.0	0.1
OPLS-Cutoff	22.5	14.5	14.0	1.2	4.5	4.6	0.0	0.0	0.1
OPLS-PME	20.7	13.4	2.7	1.7	3.2	4.3	0.0	0.1	0.0

simulations of the hairpin peptides with ff99SB, ff03 for Chignolin and with 43A1-PME, both 53A6 sets and OPLS-PME for Mbh12 (Table 3.5). These results are compatible with stable hairpin-structures, as found in solution. A significantly low presence of extended  $\beta$ -structures is observed in the conformational ensembles produced by OPLS used with cutoff for the Chignolin peptide. The same is true for the simulations of Mbh12 with ff03. Moreover, in ff03 a notable extent of  $\alpha$ - and  $3_{10}$ -helix structures was sampled instead, which was not found for the other hairpin peptides and appears to be in contrast to the experimental observations. The secondary structure analysis of the helical peptides yielded a picture with larger differences between the occupied populations (Table 3.6 and 3.7).

Several differences in the secondary structure content for Tc5b (Trp-cage) were observed in the various tested force fields. The highest helical propensity was obtained with ff99SB and ff03. Both force fields yield the same percentage of  $\alpha$ - and  $3_{10}$ -helix, preserving the content of the reference state. In contrast, the 43A1 and OPLS force fields sampled no or only very few  $3_{10}$ -helix structures, but maintained  $\alpha$ -helical structures. A low occurrence of helical content



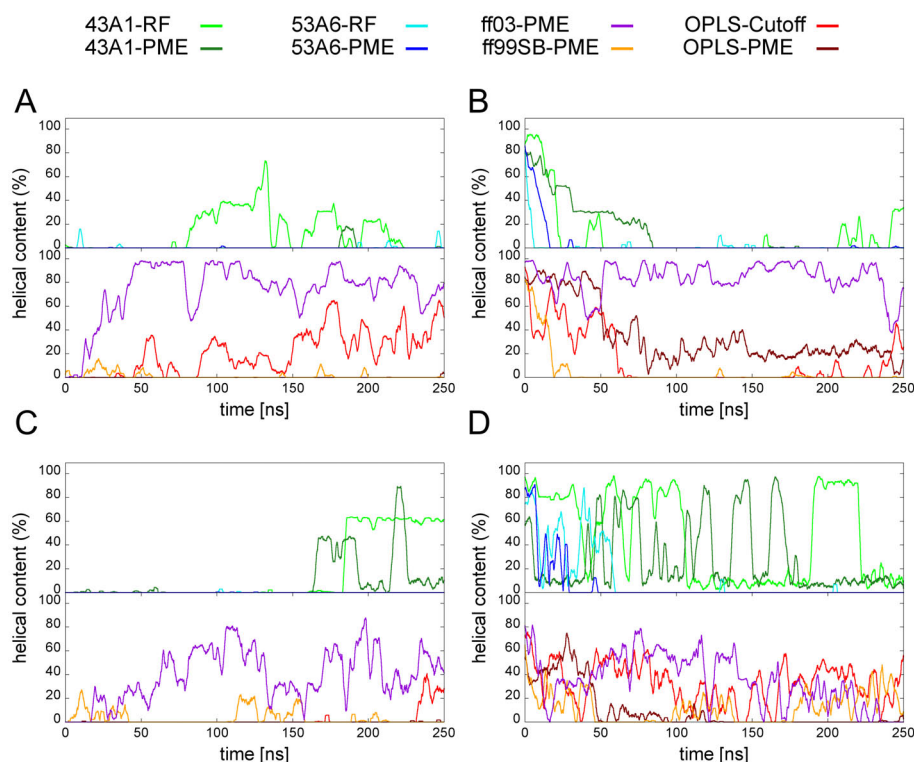


Figure 3.2: **Content of helical secondary structure.** Percentage of  $\alpha$ -helical segments present in the trajectories of Fs<sub>21</sub> and Agd1 peptides as function of simulation time. Helical content curves for the simulations FS21 (A), FS21.REF (B), AGD1 (C) and AGD1.REF (D) were smoothed.

is found with the 53A6 force field variants, which also sampled the overall highest amount of  $\beta$ -structures for the Trp-cage among all tested force fields.

The content of  $\alpha$ -helical conformations in the simulations was identified as a function of time (Fig. 3.2) to better assess the representation of the mainly helical peptides which were also simulated from the extended conformation (Fs<sub>21</sub> and Agd1). In the respective combinations of force field and electrostatic scheme a distinct trend with a marked difference between the two different initial conformations for the Fs<sub>21</sub> peptide was observed (Fig. 3.2A and 3.2B). Starting from the completely  $\alpha$ -helical reference state of Fs<sub>21</sub> the helix content in all systems except for the simulations with ff03 and OPLS-PME vanished rapidly (Fig. 3.2A). Moreover, the helical content was found to be substantially higher with ff03 than in any of the other force fields. The Fs<sub>21</sub> peptide did not visit any helical conformation with ff99SB, but rather sampled coil structures. The simulations with 43A1 showed a tendency to nonlocal  $\beta$ -bridge/sheet contacts between the terminal sites of the peptide. The preference of coil and extended structures over

Table 3.7: Averaged secondary structure content ( $\beta$ -sheet/-bridge and turn) obtained from a DSSP analysis of the helical peptide simulation.

System	$\beta$ -sheet/-bridge			turn		
	TC5B	FS21	AGD1	TC5B	FS21	AGD1
ff03-PME	0.0	0.0	0.0	9.9	4.7	14.0
ff99SB-PME	0.0	8.7	0.2	10.4	8.1	10.4
43A1-RF	3.1	14.2	3.6	15.6	10.0	8.4
43A1-PME	3.5	16.6	2.9	9.8	7.7	12.6
53A6-RF	5.4	18.4	3.5	5.3	5.4	7.8
53A6-PME	9.7	36.3	16.1	4.3	5.2	3.2
OPLS-Cutoff	0.2	3.4	0.0	17.9	13.3	18.7
OPLS-PME	0.2	4.3	6.4	19.9	12.6	11.9

helical ones was less pronounced when the reaction-field approach for electrostatics was used. Most notable for the simulations of  $Fs_{21}$  is an observed loss or complete absence of helical content in most of the G96 simulations. Particularly low fractions of  $\alpha$ -helix and a preferential conversion to extended conformations were found when using the 53A6 force field, regardless of the applied model for electrostatic treatment and starting structure. In fact  $\beta$ -sheet was the predominant secondary structure for the  $Fs_{21}$  peptide with 53A6-PME. Here, the whole peptide was readily folded into stable hairpin-like structures.

Starting from the extended conformation the formation of an  $\alpha$ -helix took place only in the simulations with the ff03, OPLS-Cutoff and transiently with the 43A1-RF force field (Fig. 3.2 B). Sampling from the extended peptide chain did not converge within 250 ns with OPLS-PME. The comparison to the experimental findings (Table 3.1) indicate that for the conditions used in the simulations close to the melting temperature of the  $Fs_{21}$  peptide around 300 K, neither a high helicity, as found with ff03 nor the completely extended structures sampled with the G96 force fields are reasonable. Rather a helical content of around 50 % would be consistent with the available CD data, therefore suggesting that only the FS21.REF simulation with OPLS and PME is reasonably well in line with the helical content found in the experiment.

The relative preference towards a certain secondary structure element is less distinct for the Agd1 peptide in the different force fields, although the overall trends are similar to the results obtained for  $Fs_{21}$  (Fig. 3.2C and 3.2D ). Helix formation proceeded fast with ff03, while no helix formation and fast unfolding of the helix were observed with the 53A6 force field. Both 43A1 force fields preserved the most helical content starting from the reference state of the Agd1 peptide. Also, a significant amount of  $\pi$ -helix is observed, especially in combination with PME which was not observed in simulations with other force fields or any of the other peptides. The repeated interconversion of  $\alpha$ - to the  $\pi$ -helix can be seen as oscillation in the helical content plot

(Fig. 3.2D). The helical content predicted with AGADIR for Agd1 (50 % at 300 K) matched best with the ff03 and 43A1 simulations, while it was underestimated with all other remaining force fields.

In general, for simulations where force fields with different methods for the electrostatic interactions are used, distinct variations are observed. Particularly, the balance between the sampling of  $\alpha$ -helix and  $\beta$ -sheet is affected. As can be seen from the analysis, the OPLS force field populates more sheet and less helix in all simulations with PME than with a simple cutoff for the electrostatics. For the 53A6 force field the relative populations of  $\alpha$ -helix and  $\beta$ -sheet are similarly affected. The differences are less distinct and systematic for the 43A1 model.

**Backbone dihedral analysis.** To further assess these discrepancies a dihedral analysis was performed, obtaining more direct information on the adopted local backbone conformations. It was chosen to evaluate the torsional sampling of alanine residues, which are only present in Fs<sub>21</sub> among the tested model peptides, and glycine and proline, as these are the most abundant in the simulated model peptide sequences. Histogram analysis was used to calculate normalized density plots of dihedral pairs ( $\varphi/\psi$ ) over all trajectories produced by each force field, respectively. A qualitative comparison for occupancy and relative sampling of  $\varphi/\psi$  dihedral pairs was made. The major low-energy regions in the Ramachandran plot, which are characteristic for prominent secondary structure elements are shown in Fig. 3.3: right-handed  $\alpha$ -helix ( $\alpha_R$ :  $-70^\circ \pm 30^\circ$ ,  $-50^\circ \pm 30^\circ$ ), left-handed  $\alpha$ -helix ( $\alpha_L$ :  $+50^\circ \pm 30^\circ$ ,  $+50^\circ \pm 30^\circ$ ), polyproline II (PP<sub>II</sub>:  $-70^\circ \pm 30^\circ$ ,  $+150^\circ \pm 30^\circ$ ) and extended  $\beta$  basin ( $\beta$ :  $-150^\circ \pm 30^\circ$ ,  $+150^\circ \pm 30^\circ$ ) [235–237]. For the alanine aminoacids of the Fs<sub>21</sub> peptide all force fields sampled the principal regions, except for ff03, where the  $\alpha_L$  basin is not visited and the  $\beta$  region only to very small extent (Fig. 3.4). The  $\alpha_L$  basin is visited most with the ff99SB force field and the relative  $\varphi/\psi$  distributions are comparable to the Ramachandran map obtained from statistical analysis of alanine backbone conformations in 500 high-resolution protein structures (Fig. 3.3) [237]. In the OPLS trajectories similar favorable regions were sampled, except for the dihedral pairs in the range of  $-125^\circ \pm 30^\circ$ ,  $+75^\circ \pm 30^\circ$ , which are not present in the ff99SB data. In all simulations carried out with the GROMOS96 force fields, the  $\alpha_L$  is visible as a broad feature over a large  $\psi$  range. The  $\beta$  and PP<sub>II</sub> regions appear as four dense states, which are particularly high populated in 53A6 and deviate from the empirical dihedral distribution from protein structure validation [237].

The histogram plots for the achiral glycine residues are asymmetric because of the influence of chiral centers in neighboring residues, and left-handed helix regions are therefore more densely

### 3.3 Results

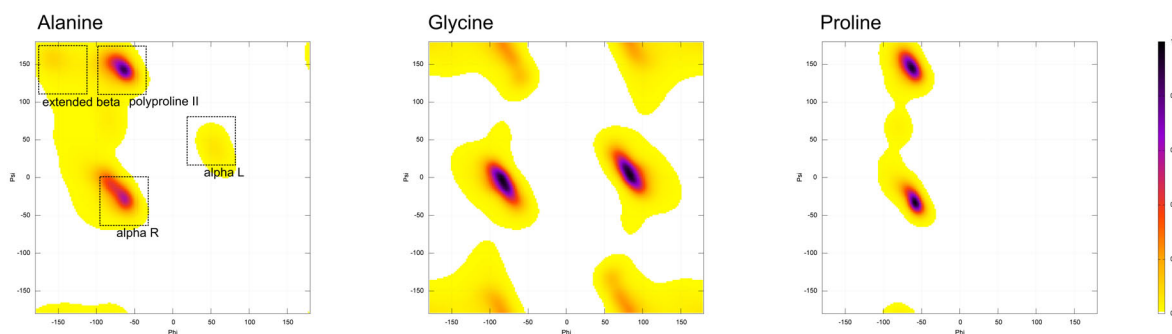


Figure 3.3: **PDB dihedral angle statistics.** Normalized  $\varphi/\psi$  histogram plot for alanine, glycine and proline residues obtained from a dihedral analysis of 500 high-resolution protein structures [237].

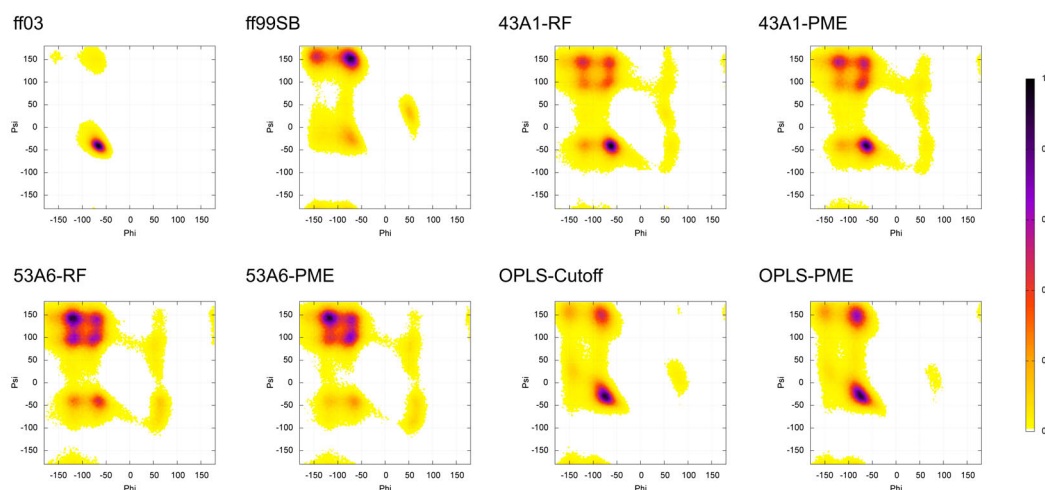


Figure 3.4: **Dihedral angle distribution for alanine residues.** Normalized  $\varphi/\psi$  histogram plot for alanine residues obtained in the simulations with the various force fields.

populated. Moreover, a large number of dihedral pairs is allowed without the steric hindrance of a  $C_\beta$ , compared to all other amino acids. Markedly, the extended  $\beta$ - and  $\alpha_L$ -regions are strongly sampled in the OPLS simulations, but within a very narrow range of  $\varphi$ -values (Fig. 3.5). In contrast, for all GROMOS96 trajectories a very broad sampling of  $\varphi/\psi$  pairs in the glycine Ramachandran plots was found, which are decomposed in several basins and deviate from empirical distributions obtained from a dihedral statistics as derived from the protein data bank (Fig. 3.3) [237]. For proline residues the conformational preferences are restricted to the  $\alpha_R$  and the PP<sub>II</sub> regions due to steric constraints. Here, the obtained normalized Ramachandran plots are more similar among the compared force fields (Fig. 3.6). The occupancy of the PP<sub>II</sub>

### 3.3 Results

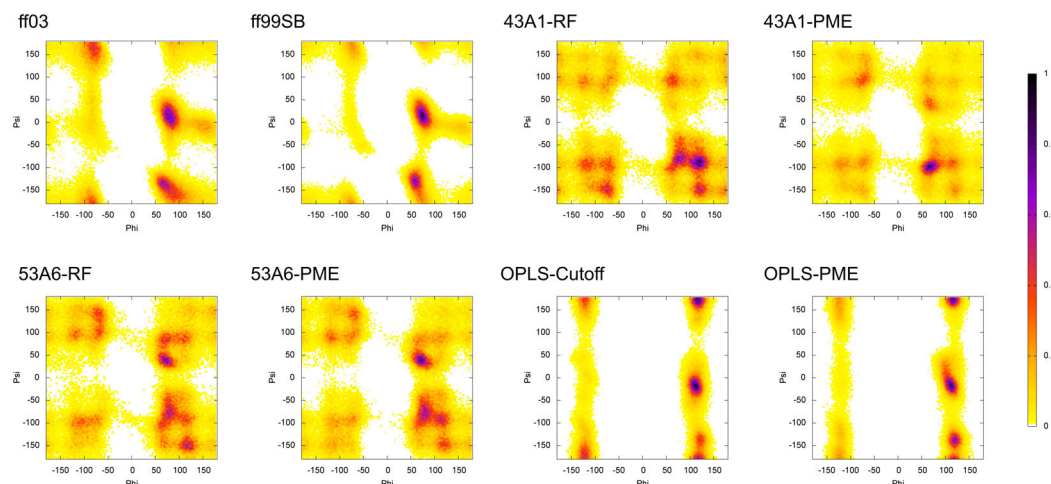


Figure 3.5: **Dihedral angle distribution for glycine residues.** Normalized  $\varphi/\psi$  histogram plot for glycine residues obtained in the simulations with the various force fields.

region is high for all simulations conducted with the GROMOS96 and OPLS force fields. A more balanced relative sampling is found with ff03 and ff99SB, which is in agreement with the empirical PDB statistics for proline residues (Fig. 3.3) [237].

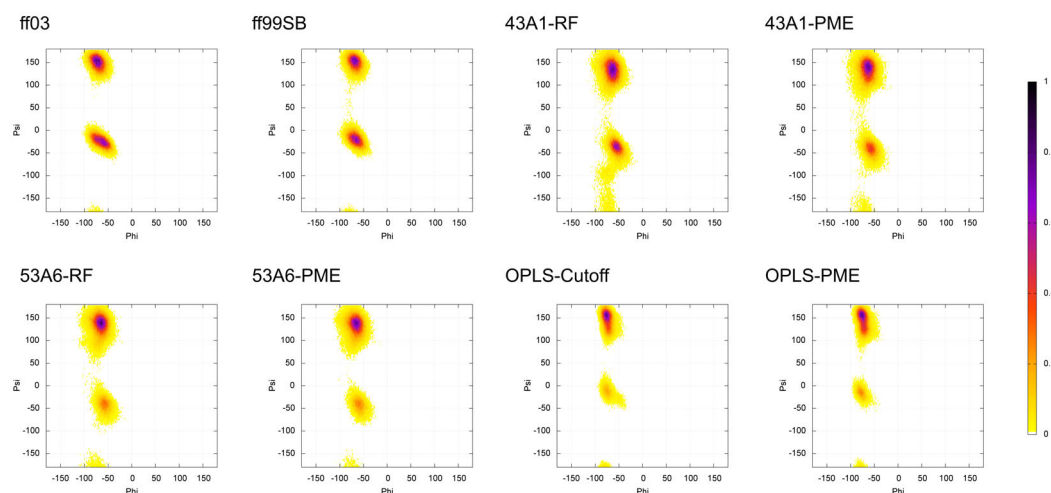
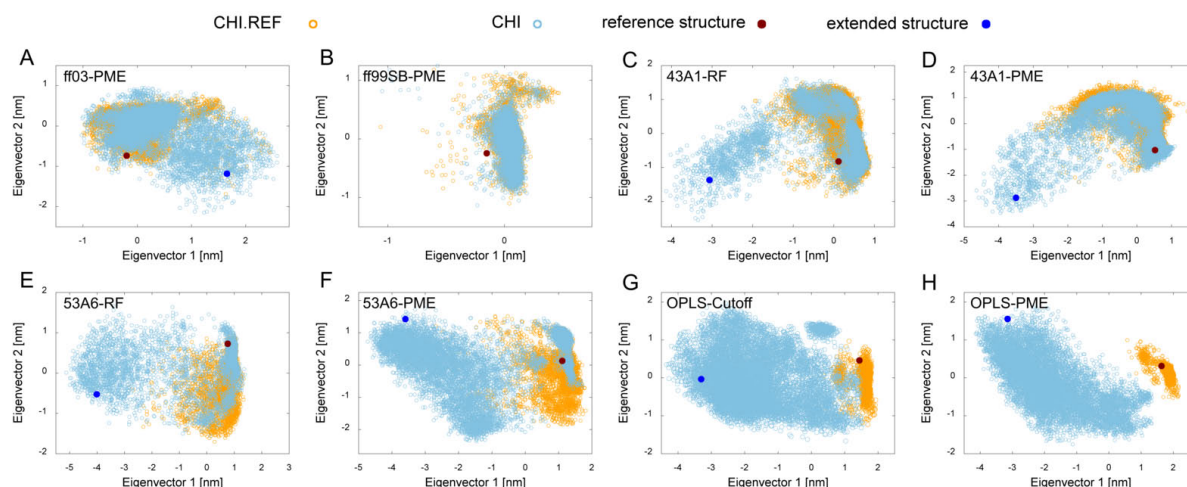


Figure 3.6: **Dihedral angle distribution for proline residues.** Normalized  $\varphi/\psi$  histogram plot for proline residues obtained in the simulations with the various force fields.

In summary, the results highlight the preferential and relative sampling of extended versus helical structures for the model peptides particular in the simulations carried out with the G96 53A6 force field. The apparent differences in structural representation of the polyalanine peptide

### 3.3 Results



**Figure 3.7: Projection of Chignolin conformations onto principal component space.** Comparison of the conformational space sampled in the simulations of the Chignolin peptide with ff03 (A), ff99SB (B), 43A1-RF (C), 43A1-PME (D), 53A6-RF (E), 53A6-PME (F), OPLS-Cutoff (G) and OPLS-PME (H). The individual trajectories were projected onto the first two common eigenvectors obtained from the principal component analysis (PCA) of the combined simulation runs for each force field, respectively. The location of the NMR reference structure and the completely extended starting conformation is indicated by a dot. Snapshots from each trajectory are shown with open circles.

$F_{s21}$  indicated that it might prove insightful to examine the backbone conformations in terms of sampled distributions of torsion angles  $\varphi$  (C-N-C $_{\alpha}$ -C) and  $\psi$  (N-C $_{\alpha}$ -C-N).

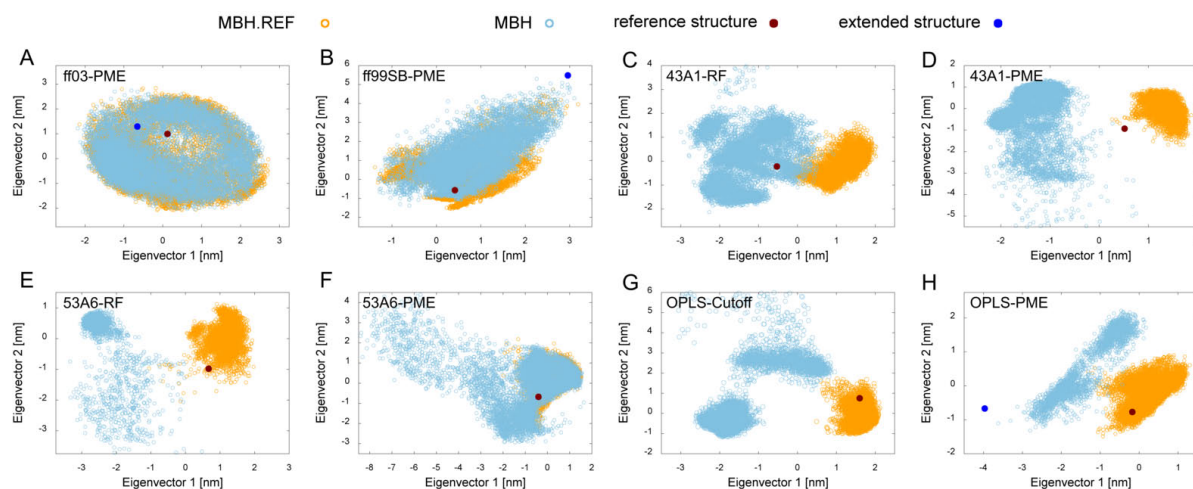
#### 3.3.2 Sampling properties

The conformational space explored by the peptides in each simulation was quantified using PCA (see Methods). Figs. 3.7 and 3.8 show the projections of the trajectories onto the plane spanned by the first two common eigenvectors obtained from the PCA for the two hairpin peptides, respectively. The covered area in the projections represents the sampled conformational space in the given force field and with respect to the reference structure. Each point is a low dimensional representation of a conformation sampled in the trajectory.

The PCA projections of the Chignolin peptide simulations with the different force fields shown in Fig. 3.7 exhibit mostly overlaying phase space regions close to the reference structure. This denotes that the folded state was successfully found from the extended structure. Starting from the prefolded hairpin structure of Chignolin, mostly conformations close to the reference structure were sampled, compatible to a stable hairpin in solution. Interestingly, irrespective



### 3.3 Results



**Figure 3.8: Projection of Mbh12 conformations onto principal component space.** Conformational space sampled in the simulations of the Mbh12 peptide. Projections of the NMR reference structure, extended starting conformation and snapshots from each trajectory are denoted according to Fig. 3.7.

of the initial structure almost the same conformational space is sampled by the ff99SB force field, indicative of a converged ensemble. Both independent trajectories (CHI and CHI.REF) cover the same area in the projection close to the NMR reference state (Fig. 3.7B). For ff03, 43A1-RF and 43A1-PME strongly overlapping areas in the projections of the simulations were found, irrespective of the starting conformations as well (Fig. 3.7A, Fig. 3.7C and Fig. 3.7D). The 53A6 force field variants also found stable structures of the Chignolin peptide, sampling regions close to the reference (Fig. 3.7E and 3.7F). In particular, the conformational space around the NMR reference was sampled exhaustively with 53A6-PME for both initial conditions, corresponding to the lowest averaged RMSDs and sum-averaged NOE distance violations among all tested force fields (Fig. 3.7F). The overall trends of the conformational sampling in the simulations of Chignolin with OPLS-PME (Fig. 3.7H) and OPLS-Cutoff (Fig. 3.7G) are similar. The simulation ensembles, when started from the extended hairpin peptide (CHI) do not contain defined structures close to the reference state. A broad area corresponding to predominantly unfolded conformations was sampled with only few structures approaching the reference state. The explored conformational space in the CHI.REF simulations, however, was only very small and close to the reference. This substantial difference in the extent of visited phase space results in well separated regions for the simulations with the different initial structures, especially with OPLS-PME (Fig. 3.7H).

The PCA of the trajectories for Mbh12 yielded a more non-uniform picture concerning the

sampling behavior of different force fields. For ff03 both trajectories of the hairpin peptide sampled the same regions of conformational space in the PCA projection, as seen before in the case of Chignolin. Most of the structures did not match with the NMR reference (Fig. 3.8A). This is in line with the observation that the Mbh12 hairpin was least stable in ff03 among all force fields. Fig. 3.8B shows an overlay of sampled phase space for both production runs with ff99SB, with structural ensembles close to the reference. A refolding event of the hairpin during the MBH.REF simulation resulted in a stable off-register structure, which is observed as strongly sampled sub-state. As shown in Fig. 3.8F the simulations of Mbh12 carried out with 53A6-PME yielded nearly identical conformational ensembles in the projection, extensively populating conformations close to the NMR reference state. In contrast, the sampled conformational space in the MBH and MBH.REF trajectories are separated for all other force fields. The 43A1-RF simulation starting from the extended peptide chain populated several sub-states in the PCA projection en-route to hairpin formation (Fig. 3.8C). Markedly separated areas were found when comparing this projection to the one obtained from the 43A1-PME trajectories (Fig. 3.8D). The projected simulation ensemble of Mbh12 with 53A6-RF were, contrary to one obtained with 53A6-PME, separated. The explored conformational space in the MBH simulation with the 53A6-RF force field appears very small and dense (Fig. 3.8E). The projections of the OPLS trajectories resemble the ones of the Chignolin peptide. The simulations from the extended chain (MBH) sampled mainly two states, which are separated in conformational space from the reference. Starting from the hairpin (MBH.REF), mostly folded conformations were gauged during the simulation as shown in Fig. 3.8G and 3.8H.

Pronounced differences in conformational space sampling could also be found among the probed force fields for the simulations of the Trp-cage. The PCA projections are shown in Fig. 3.9.

In the PCA projections only one broadly and densely sampled area was found for all four ff99SB simulations, which is explained by only minor conformational changes (Fig. 3.9B). In the projections of the trajectories obtained with both OPLS force field variants are mainly two visited states visible (see Fig. 3.9G and 3.9H). A transition path to a more populated cluster of structures away from the reference was seen. The use of PME apparently decreased the frequency of larger conformational transitions, as only the 250 ns long TC5B.REF simulation with OPLS-PME visited this phase region (Fig. 3.9H). In the Tc5b simulations with ff03, four states are sampled. One of them is exclusively visited in the 30 ns long TC5B.REF.3 simulation (Fig. 3.9A, red point cloud), which also showed an increased RMSD over the last 15 ns. The simulations of the Trp-cage with the G96 force fields showed a overall similar trend in the PCA



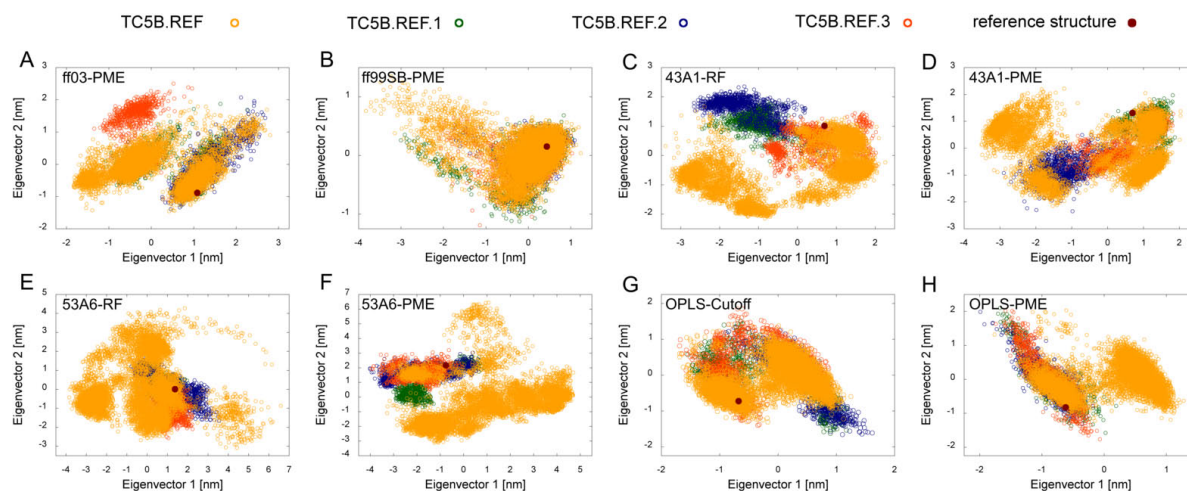


Figure 3.9: **Projection of Trp-cage conformations onto principal component space.** Conformational space sampled in the simulations of the Trp-cage (Tc5b). The reference structure is indicated by the dark-red dot. Snapshots from the 250 ns trajectory are shown with orange open circles (TC5B.REF). Conformations of the three 30 ns simulations are depicted by green (TC5B.REF.1), dark-blue (TC5B.REF.2) and orange-red circles (TC5B.REF.3).

projections (Fig. 3.9C-3.9F). Although partially conformations close to the reference structure were visited, the simulations quickly explored different areas of the phase space, with most of them distinct from the NMR reference state. The projected conformational space is decomposed in many states and different for each of the four simulations. Moreover, the conformations with very little native contacts were already sampled within 30 ns in most of the cases, as indicated by the several short runs. In particular, conformational ensembles of the Trp-cage for both 53A6 variants evolved around the reference structure as a diffuse cloud of points (Fig. 3.9E and 3.9F).

## 3.4 Discussion

As seen from the NOE and PCA analysis, most of the simulations starting from the extended and folded structure showed differences in the sampled conformational ensembles indicating incomplete convergence. The only exceptions are the simulations with both tested AMBER force fields, in which the conformational space was quickly explored and did not critically depend on the initial structure. The ff03 force field compared favorably to available experimental data in most of the simulations carried out. The only exceptions are a particular high percentage of helical content for the polyalanine peptide Fs<sub>21</sub> and the unfolding of the Mbh12 hairpin,

indicating a slight preference of helix over extended  $\beta$ -sheet conformations with ff03. This is in line with the results from Hornak et al. [156] for trialanine peptides and is known from previous AMBER force fields [226, 238].

A fast sampling towards the reference state of the Chignolin peptide and particularly low RMSDs for the Trp-cage were found for the ff99SB force field. The analysis of secondary structure content and the low interproton distance violations suggest that the simulations with ff99SB agree well to available experimental data. The content of  $\alpha$ -helix for the mainly helical peptides (Fs<sub>21</sub>, Agd1) was likely underestimated. The GROMOS96 force fields revealed an underestimation of propensity for sampling  $\alpha$ -helical conformations in the simulations. While the results for the 43A1 version were well balanced for most of the model peptides, a significant disparity was the considerable amount of  $\pi$ -helix, which was not found in the other force fields. Different methods for electrostatics (RF, PME) in conjunction with the 43A1 force field were found to have little effect on the folding characteristics of the model peptides. The simulations using PME, showed a slightly better structural representation, but sampling might be more efficient with the reaction-field approach [239]. With the 53A6 force field variants  $\beta$ -sheet formation was most abundant and persistent, when comparing the results from all simulations. The structural representation of the hairpin peptides was in very close agreement with NOE data. Preformed  $\alpha$ -helical conformations were, however, least stable with 53A6 in all helical model peptides. Moreover, using PME appears to favor extended conformations even more. In accordance with the present results, reproducible unfolding of  $\alpha$ -helical peptides was reported recently by Cao et al. in a study presenting refined terms of conformational backbone description for 53A6 [240]. Another study also found reduced structural variety and less covered phase space in the simulations with 53A6 compared to earlier GROMOS96 variants, which was suggested to be due to the increased partial charges on the backbone carbonyls and amides [161, 241].

The combination of OPLS force field and PME for electrostatics provided superior results compared to the OPLS with a straight cutoff, and most of the other force field models, when representing the folded state. However, the chosen initial conformations affected the sampled structural ensembles considerably. A dependence on the initial conditions especially for the OPLS force field was apparent suggesting slower conformational sampling than observed in the other studied force fields. The sampling differences between the different force fields indicate a different roughness of the respective energy landscapes, affecting the convergence of the simulations. For the helical peptides the least overall concurrent sampled secondary structure with the different force fields was found, as the helical systems were simulated near their melting temperature. In this regime small shifts in the free energy involve large changes in the population

of folded and unfolded conformations and thus slightly different melting temperatures in the different force fields might contribute to the found pronounced differences in sampled helical propensity among the simulations. In the large set of simulations presented here, the CHARMM force field [157, 158] was not included, still lacking a validated port for the GROMACS software package. Assessing the performance of the CHARMM force field in the context of the discussed simulation results would be an interesting aspect to address in future studies.

## 3.5 Conclusion

Peptide folding with MD simulations is inherently dependent on the accuracy of the applied force field. The observed similarities between different force fields support the consensus view of biomolecular dynamics [168], but remaining differences emphasize the importance of continuous force field development and refinement. The folding behavior of various peptide sequences in different MD force fields revealed significant and systematic differences in the stability and formation propensity of dominant secondary structure elements. The observations suggest that in particular the relative stabilities of helical and extended conformations depend on a subtle balance of force field parameters. In addition, indications for different sampling characteristics of the respective force fields were found, affecting both the kinetics and convergence of the simulations. It is likely that the discussed deviations in the structural representations are less critical in protein simulations, when starting from the folded native state. However, a significant bias for peptide and protein simulations on long timescales is expected [210]. This also indicates the relevance for simulations of folding intermediates, natively unfolded proteins, and studies on peptide aggregation. Concerning the treatment of electrostatics, it is nowadays common practice to apply PME for biomolecular simulations. However, care is required when employing PME in conjunction with force fields and water models which were originally developed using cutoff or reaction-field (OPLS, GROMOS96) [148]. In terms of secondary structure propensities, the peptides studied here revealed a tendency towards sampling  $\beta$ -hairpin structures when employing PME combined with the OPLS and GROMOS96 force fields.

In summary, for folding studies a force field bias can not be excluded and from the current perspective there is no single best fit solution for peptide folding simulations with today's nonpolarizable force fields. Rather a multiple force field or consensus approach is suggested: if computationally feasible to simulate, using more than one suitable force field to address the particular question at hand, and whenever possible, to compare the simulation results to direct experimental data.



## CHAPTER 4

# Conformational dynamics, assembly pathways and driving forces of steric zipper peptide oligomerization

---

- [D. Matthes](#), V. Gapsys, V. Daebel and B. L. de Groot, PLoS ONE, 2011, 6(5): e19129.
- [D. Matthes\\*](#), V. Gapsys\* and B. L. de Groot, J. Mol. Biol., 2012, (DOI: 10.1016/j.jmb.2012.02.004), \*contributed equally

## Individual contributions of others

- PLoS ONE, 2011, 6(5): e19129: The manuscript was written together with B. L. de Groot. V. Gapsys contributed parts of the data analysis with the idea and implementation of the relabeling scheme, as well as by performing the PCA and conformational clustering. V. Daebel did proof reading of the manuscript and contributed to discussions throughout the writing process.
- J. Mol. Biol., 2012, (DOI: 10.1016/j.jmb.2012.02.004): The manuscript was written together with B. L. de Groot. V. Gapsys designed and performed larger portions of the data analysis (PCA, nearest neighbor searching, conformational clustering and FMA).

---

## Summary

The process of protein misfolding and self-assembly into various, polymorphic aggregates is associated with a number of important neurodegenerative diseases. Only recently, crystal structures of several short peptides have provided detailed structural insights into  $\beta$ -sheet rich aggregates, known as amyloid fibrils. Knowledge about early events of the formation and interconversion of small oligomeric states, an inevitable step in the cascade of peptide self-assembly, however, remains still limited.

Molecular dynamics simulations in explicit solvent have been employed to study the spontaneous aggregation process of steric zipper peptides from the tau protein, insulin and  $\alpha$ -synuclein in atomistic detail on the microsecond timescale. Starting from separated chains with random conformations, a rapid formation of structurally heterogeneous,  $\beta$ -sheet rich oligomers, emerging from multiple bimolecular association steps and diverse assembly pathways was found.

The study provides evidence that aggregate intermediates as small as dimers can be kinetically trapped and thus affect the structural evolution of larger oligomers. Alternative aggregate structures are found for both peptide sequences in the different independent simulations, some of which feature characteristics of the known steric zipper conformation (e.g.  $\beta$ -sheet bilayers with a dry interface). The final aggregates interconvert with topologically distinct oligomeric states exclusively via internal rearrangements. The peptide oligomerization was analyzed through the perspective of a minimal oligomer, i.e. the dimer. Thereby all observed multimeric aggregates can be consistently mapped onto a space of reduced dimensionality. The conformational mapping reveals heterogeneous association and reorganization dynamics which are governed by the characteristics of peptide sequence and oligomer size.

Detailed analysis of the forces driving the oligomerization reveals a common two-step process akin to a general condensation-ordering mechanism and thus provides a rational understanding of the molecular basis of peptide self-assembly. Our results suggest that the initial formation of partially ordered peptide oligomers is governed by the solvation free-energy, whereas the dynamical ordering and emergence of  $\beta$ -sheets are mainly driven by optimized inter-peptide interactions in the collapsed state. A mapping technique based on collective coordinates is employed to highlight similarities and differences in the conformational ensemble of small oligomer structures. Elucidating the dynamical and polymorphic  $\beta$ -sheet oligomer conformations at atomistic detail furthermore suggests complementary sheet packing characteristics similar to steric zipper structures, but with a larger heterogeneity in the strand alignment pattern and sheet-to-sheet arrangements compared to the cross- $\beta$  motif found in the fibrillar or crystalline states.

## 4.1 Mapping conformational dynamics and pathways of steric zipper peptide oligomerization

### 4.1.1 Introduction

The assembly of polypeptides and proteins into  $\beta$ -sheet rich aggregates termed amyloid plaques and fibrils is known to be associated with several severe diseases in vivo [4, 29, 59, 242, 243]. The morphology and molecular structure of mature amyloid fibrils have been investigated extensively in vitro as many polypeptide chains form fibrillar aggregates with cross- $\beta$  conformation [35]. The application of a wide range of different diffraction and spectroscopy techniques [43, 45, 47, 55, 244–247] has helped to elucidate various aspects of fibril architecture and has confirmed molecular polymorphism on several structural levels of filament organization [35, 43, 51, 66, 248]. However, the mechanistic details and underlying energetics of amyloid fibril formation and the appreciable multitude of conformational states, starting from entities as small as monomers are not well understood [29, 56, 59, 242]. Specifically, the initial and early stages, marked by the population of various soluble oligomeric species still await a consistent and unified description [53, 54, 56, 66, 67]. Recent advances in the field have contributed to the discovery of a rich variety of amyloid precursors and oligomeric states [63–65, 68], yet the structurally heterogeneous and transient nature of these aggregates still does not permit precise experimental characterization.

Analyses of experimentally observed aggregation kinetics suggest different mechanistic explanations for possible pathways and rate-limiting steps of amyloid fibril formation [52–56]. But it has not been possible to directly probe the nucleation event in amyloid aggregation so far. While it remains controversial if all observed types of oligomers are indeed on-pathway intermediates [29, 66, 67], an increasing amount of experimental evidence points to soluble protein oligomers as the primary cause of cell impairment and dysfunction in the pathogenesis of neurodegenerative diseases and various amyloidoses [4, 63–69].

A mechanistic interpretation of the multi-staged aggregation process and a structural characterization of key intermediates therefore pose an essential challenge, also with regard to the identification and application of therapeutic strategies interfering with nonnative peptide and protein aggregation. Computer simulation techniques have provided insight into several raised questions, such as thermodynamics or conformational dynamics of early oligomer formation [106, 108–119] and additionally have made valuable mechanistic propositions [120–127].

In vitro studies established segments of amyloidogenic proteins and de novo designed pep-

tides as suitable model systems to investigate sequence determinants of fibril formation (e.g. mutational effects) [83–88]. These short peptides were shown to be capable of forming amyloid-like fibrils, yet they are sufficiently small to allow systematic and controlled experimental access to their biophysical properties and to detailed structural models [97, 98]. Recently obtained crystal structures of such minimal peptide sequences provided insight into what could be the general spine organization of amyloid fibrils [82, 100, 102–104]. A common steric zipper motif was revealed for a number of peptides in the crystalline state, where two elongated sheets of peptide strands are arranged such that a complementarity packing of the side chains leads to a tight and dry interface. Combined, these experimental findings underscore the notion of amyloid fibril formation being a universal property of the peptide backbone depending on external factors and modulated by sequence characteristics [4, 16]. Moreover, it was shown that these specific short stretches can trigger self-assembly and mediate amyloid formation [93, 94], therefore leading to the idea that the amyloidogenicity of a sequence can be strongly localized. The ability of these segments to even force a globular, non-fibrillizing protein into the amyloid state was demonstrated [101].

In the present study the structure and dynamics of spontaneously assembled oligomers for two different steric zipper peptides (PHF6, IB12) was explored using atomistic molecular dynamics simulations in explicit water.

The PHF6 (<sup>306</sup>VQIVYK<sup>311</sup>) peptide is a segment from the microtubule-associated tau protein and has been shown to be sufficient for in vitro polymerization to filamentous structures and microcrystals [82, 86, 95, 96]. The PHF6 motif is located in the repeat regions of the microtubule-binding domain of tau and has been suggested to play a prominent role in the formation of paired helical filaments (PHFs) and is also part of the PHF core composed of cross- $\beta$  structure [95, 96, 246]. The natively unfolded and highly soluble full-length protein functions in assembly and stabilization of microtubules [96] and self-associates into PHFs, when hyperphosphorylated [96, 246]. Pathological accumulation of such tau aggregates into neurofibrillary tangles is a characteristic signature of Alzheimer's disease and other tauopathies [246].

The hexapeptide <sup>12</sup>VEALYL<sup>17</sup>, here referred to as IB12, is a segment from the B chain of the peptide hormone insulin and has been found to form amyloid-like fibrils, as well as microcrystalline aggregates with typical cross- $\beta$  diffraction pattern [82, 88]. The IB12 peptide is proposed to have importance in full-length insulin misfolding and aggregation [249]. Insulin has been studied as a model system for aggregation and generally under conditions (elevated temperatures, low pH) which favor the monomeric and partially unfolded state [250–252]. Fibrillation proceeds via oligomeric, nonnative intermediates, during which



## 4.1 Mapping conformational dynamics and pathways of steric zipper peptide oligomerization

insulin is subject to major structural alterations from a predominantly  $\alpha$ -helical to  $\beta$ -sheet rich conformation [245, 250–252]. According to several studies [82, 249, 251, 253] critical intermolecular interactions have been attributed to the IB12 segment of the B chain, which is also most likely incorporated in the extended  $\beta$ -strands that make up the core region of insulin fibrils [254, 255].

Here a quantitative description of the spontaneous steric zipper peptide aggregation and oligomer growth process is presented based on multiple submicrosecond molecular dynamics simulations. Two different short segments of amyloidogenic proteins are used to study and unveil the critical variables that govern the kinetics of the initial biomolecular aggregation stages. From the observed conformational dynamics during the formation of small oligomers common mechanistic steps at molecular detail are elucidated. In order to do so, a novel low-dimensional mapping procedure is introduced to visualize the heterogeneous oligomerization pathways, allowing the identification of common encounter complexes and intermediates.

### 4.1.2 Methods

#### Simulation setup and procedure.

**Simulated Systems.** An overview of the simulated peptide systems is given in Table 4.1. The simulations were categorized according to the name of the peptide and the starting configuration and have a length of 4.8  $\mu$ s altogether.

Table 4.1: Summary of performed simulations.

system (ID)	starting configuration	conc. (mM)	time (ns/run)	no. of sim.
VQIVYK (PHF6)	10 monomeric peptides, random	16.6	300	8
VEALYL (IB12)	10 monomeric peptides, random	16.6	300	8

conc.: peptide concentration.

no. of sim.: number of independent simulations.

**Initial conformations.** Conformational ensembles containing 1000 peptide structures each were generated with CONCOORD [256] based on the atomic coordinates of the PDB crystal structures: 2ON9 (VQIVYK; PHF6) and 2OMQ (VEALYL; IB12) [82], respectively. Only topological constraints were defined, resulting in random starting

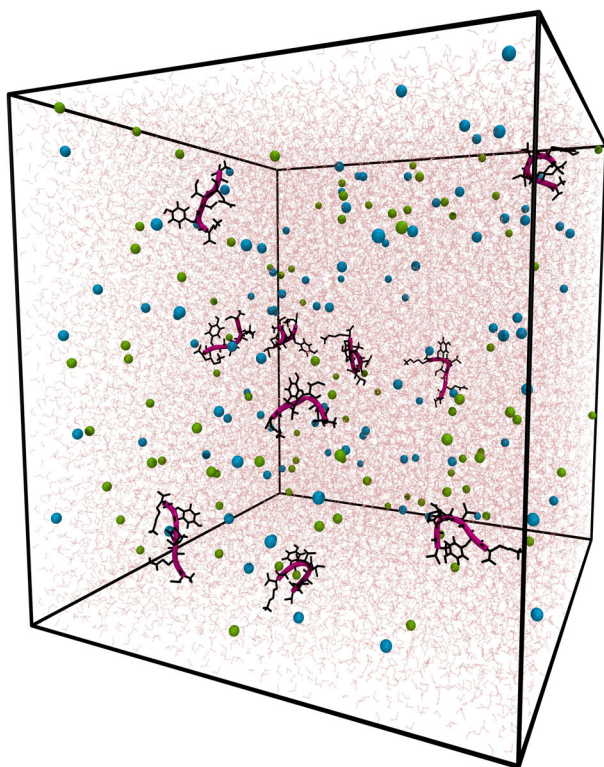


Figure 4.1: **Typical starting configuration: A cubic simulation box with 10 monomeric peptides, ions and explicit water molecules.** The peptide backbones are depicted in cartoon representation and side chain atoms as sticks, sodium and chloride ions as spheres and water molecules as transparent sticks.

configurations for the simulations. The individual simulations were set up according to the following protocol: 10 different peptide conformations were randomly chosen from the pre-generated structure ensemble and placed randomly in position and orientation in a cubic box ( $1000 \text{ nm}^3$ ) to result in a concentration of  $16.6 \text{ mM}$ . This procedure was applied to reduce the bias from the individual peptide's starting position and configuration and to ensure a fully monomeric starting configuration for each of the conducted simulation runs, respectively.

Subsequently all systems were solvated with explicit water molecules. The protonation state of the peptides was at pH 7 for all simulations of the PHF6 peptides according to the one in solution. For the IB12 peptide simulations and consistent with experimental conditions at pH 2.5 [82, 88], the C-terminus and the glutamate side chains were assumed to be protonated. Counter-ions ( $\text{Na}^+$ ,  $\text{Cl}^-$ ) were added to yield an appropriate ionic strength ( $0.15 \text{ M}$ ) and to neutralize the net system charge. After the system preparation an energy minimization using steepest descent was performed. The simulation systems were each comprised of roughly  $100 \times 10^3$  atoms. A typical simulation box is shown in Figure 4.1.

**MD setup.** All MD simulations were carried out using the GROMACS software package (version 4.0) [175–177]. The Berendsen coupling algorithm [187] was applied to keep the pressure constant by coupling the system to a pressure bath of 1 bar ( $\tau = 1$  ps). Velocity rescale [188] was applied for temperature coupling to a temperature bath of 310 K. Initial velocities were taken from a Maxwellian distribution at 310 K. All protein bonds were constrained with the P-Lincs algorithm [182]. Virtual interaction sites of all hydrogen atoms were introduced, thereby removing all internal vibrational degrees of freedom. This allowed us to use an integration time step of 5 fs while maintaining energy conservation [175, 179]. Neighbor lists for nonbonded interactions were updated every 5 steps. For production runs the GROMOS96 43A1 [159, 160] force field and the SPC water model [184] were used. Water molecules were constrained using SETTLE [234]. The short-ranged nonbonded interactions, namely van der Waals and electrostatic were cut-off at 1.4 nm and 0.9 nm, respectively. All simulations were carried out using periodic boundary conditions and the Particle Mesh Ewald (PME) [230, 231] method. The electrostatic interactions with PME were calculated at every step with a grid spacing of 0.12 nm. The relative tolerance at the cut-off was set at  $10^{-6}$ , electrostatic interactions for a distance smaller than the real space cut-off were calculated explicitly.

### Analysis

Samples for analysis were taken every 2.5 ps from the collected trajectories.

**Definitions of general and  $\beta$ -sheet aggregates.** In order to quantitatively probe the association state of the peptides at any given time a hierarchical classification of the formed aggregates and their numerous conformations was pursued:

- i) Pairwise inter-peptide contact analysis was used to identify the individual aggregates, defined as general peptide assemblies: Peptides which shared an inter-chain residue contact were assumed to be within the same aggregate. For any two peptides  $i$ ,  $j$  an inter-chain contact was considered to be formed if any heavy atom of peptide  $i$  was within a distance of 0.45 nm from any heavy atom of peptide  $j$ .
- ii) To trace not only the general peptide association but also any sort of transition to ordered species, assemblies of peptides aggregated into intermolecular  $\beta$ -sheets were defined: Any two peptides which shared two consecutive inter-chain  $\beta$ -sheet contacts as defined by the DSSP algorithm [202] were considered to constitute a  $\beta$ -sheet aggregate, a subpopulation of the above considered general peptide aggregates.

**Orientational order parameter P2.** The nematic order parameter (P2) of the system yields information about the extent of alignment and relative orientation of the individual peptides. A suitable molecular vector ( $\hat{\mathbf{z}}_i$ ) was defined, here the unit vector linking the C $_{\alpha}$ -atoms of the second to fifth residue of the  $i$ -th peptide.

$$P_2 = \frac{1}{N} \sum_{i=1}^N \frac{3}{2} (\hat{\mathbf{z}}_i \cdot \hat{\mathbf{d}})^2 - \frac{1}{2}$$

The order parameter P2 was calculated using the WORDOM program package [257].

**Oligomer decomposition and dimer relabeling.** The pathways of PHF6 and IB12 peptide aggregation were mapped and analyzed through the perspective of a minimal oligomer, i.e. the peptide dimer. All aggregates consisting of three, four or ten peptides were separately decomposed into smaller dimeric subsystems based on a minimal distance criterion. Afterwards, these decomposed structures were consistently relabeled, clustered and transition pathways between clusters constructed: The decomposition of higher oligomers into dimeric structures was performed separately for trajectories which contained all trimeric, tetrameric and decameric aggregates observed in all simulations for both, PHF6 and IB12, peptides. For each trajectory frame a neighbor list of peptides was generated for each molecule. Peptide  $A$  was added to the neighbor list of peptide  $B$  if at least one heavy atom of  $A$  and  $B$  were less than 0.45 nm apart. That way all possible combinations of dimers for each molecule and the molecules in its neighbor list were saved. Only unique dimer combinations were considered by removing identical and permuted combinations, e.g. dimers  $AB$  and  $BA$  were treated as equivalent and hence one of them discarded. Thereby a conformational ensemble of dimeric structures was obtained, constructed from a specific oligomeric state (trimer, tetramer and decamer).

Both, the clustering of the dimeric structures and mapping them onto principal components, requires superposition to a reference structure. In order to fit structures consistently, peptides forming dimers were labeled such that similar molecules were given the same chain identifier (ID).

In the first step of this relabeling procedure, peptides in a dimeric structure were assigned with individual IDs and fit to a reference structure. In the second step, IDs for the peptides were interchanged and fitting was performed again. Labels that resulted in a lower root mean square deviation (RMSD) between the dimer of interest and the reference structure were kept. Main-chain and C $_{\beta}$  atoms were used for fitting and the RMSD calculation. To prevent the

dependence of labeling on the reference structure, an iterative relabeling scheme was applied. Initially, relabeling was performed using a peptide dimer built from the respective crystal structure conformation as a reference. The average structure of the relabeled trajectory was calculated. The relabeling procedure was repeated using this average structure as a reference. The cycle was repeated until all structures in the trajectory were labeled identically in two subsequent iterations.

**PCA and k-means clustering.** Principal component analysis (PCA) [194] was carried out over the conformations of the spontaneously formed PHF6 and IB12 dimers, respectively. The covariance matrix of atomic displacement was constructed and diagonalized for the coordinates of main-chain and  $C_\beta$  atoms. All structures were superimposed to the respective average structure calculated over all the dimer conformations prior to analysis. The conformational ensemble of spontaneously formed PHF6 and IB12 dimers were projected onto the first three eigenvectors to obtain a mapping into a space of reduced dimensionality, respectively.

For the conformational clustering of dimeric structures, coordinates of the main-chain and  $C_\beta$  atoms were extracted from each peptide dimer. RMSD values between dimeric structures were used as a distance measure for k-means clustering. The implementation of the k-means Hartigan-Wong algorithm [258] in the statistical software package R [259] was used for clustering. Cluster centers were selected according to the global k-means algorithm [260], the optimal number of clusters was determined by the Krzhanowski-Lai criterion [261]. Structures with the lowest RMSD value to the geometrical centers of the clusters were selected as representatives. The selection of cluster centers and numbers was performed only for the spontaneously formed dimer conformations of PHF6 and IB12, respectively. The decomposed trajectories of trimers and tetramers were clustered using these respective cluster centers and numbers.

**Pathway mapping.** Transition networks were constructed based on the conformational clustering. Nodes represent clusters and edges the transitions between them. The size of a node was determined by counting occurrences of a structure in a certain cluster. Transitions between two clusters were counted if dimer conformations traversed to another cluster in the subsequent time step (frame). If a dimer conformation disassembled or formed a larger aggregate for at least one time step, a transition out of the node (cluster) was counted. If a dimer conformation was formed, a

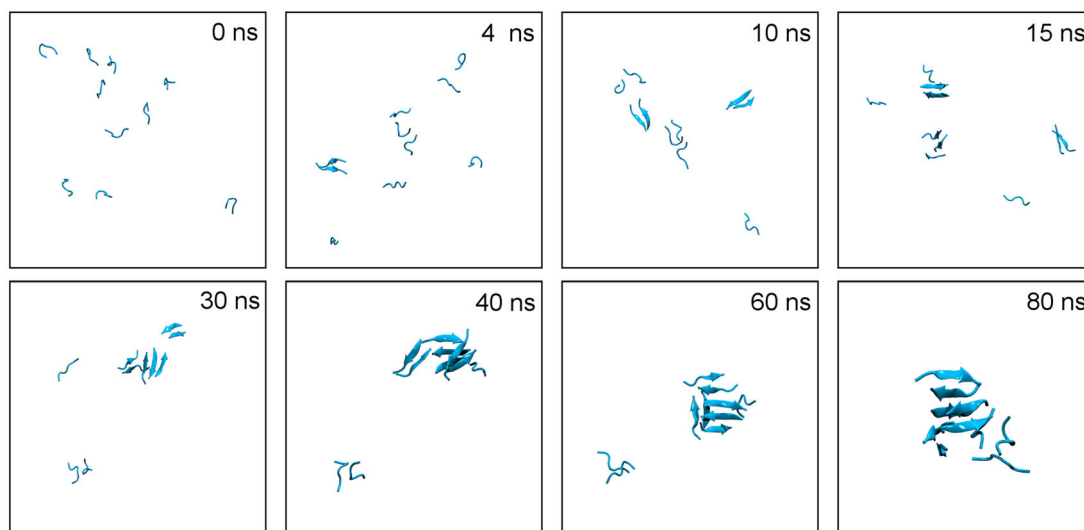


Figure 4.2: **Illustration of primary PHF6 peptide aggregation events.** Snapshots of conformations observed during the first 80 ns of a representative simulation of PHF6 oligomerization are shown (water and ions omitted for clarity). Starting from the initial state, the 10 monomeric peptides with random conformations rapidly assemble and start to form small sized aggregates ( $n \leq 4$ ) after a short lag time. Towards the end of the simulation an aggregated phase is observed, characterized by the presence of one single, large oligomer ( $n = 10$ ).

transition into the node (cluster) was counted. Pathway network construction was performed with Cytoscape [262]. Node and edge sizes were normalized for each network separately.

**Visualization.** Molecular images in the main text and PCA projections were rendered using VMD [263] and Tachyon [264]. The pathway networks were visualized using the Cytoscape package [262].

### 4.1.3 Results

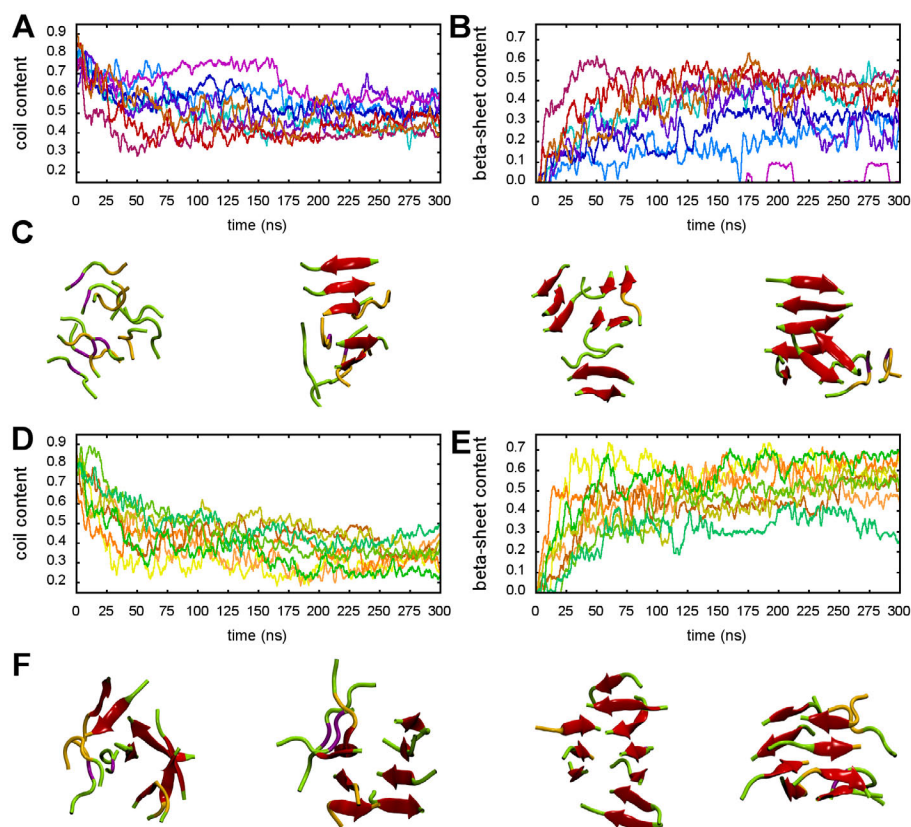
**Secondary structure and topology of spontaneously assembled steric zipper peptide oligomers.** In the present study the primary aggregation events were monitored for the two different peptides PHF6 and IB12, with eight independent simulations each (see Table 4.1). The simulations started from 10 separated peptide chains with random conformations, positions and orientations, respectively (Fig. 4.1). Representative snapshots from one of the PHF6 simulations are shown in Figure 4.2. They illustrate a typical, spontaneous self-assembly into  $\beta$ -sheet oligomers as frequently observed in the 300 ns simulations.

A secondary structure content analysis reveals that the peptide aggregation was accompanied by a conversion from random coil to extended  $\beta$ -sheets as the dominant structural motif in the oligomeric state (Fig. 4.3). As simulation time progresses, a rapidly increasing number of inter-molecular hydrogen bonds was found, together with the spontaneous formation of  $\beta$ -sheet rich oligomers. Although  $\beta$ -sheet formation was a general feature, differences in the  $\beta$ -sheet content were observed for PHF6 and IB12 simulations. While in most of the PHF6 simulations more than 40% of the high initial random coil content was retained, for the IB12 peptides a lower coil fraction was found after 300 ns. In the course of the IB12 simulations a near monotonically increasing  $\beta$ -sheet content was observed (see Fig. 4.3E). On average more than half of all IB12 peptide residues were found in extended backbone conformation. In contrast, smaller fractions and larger fluctuations in the amount of  $\beta$ -sheets were observed for the PHF6 aggregates, as shown in Figure 4.3B. One out of the eight PHF6 trajectories even yielded an almost amorphous decameric peptide aggregate, exhibiting strong disorder.

Nevertheless, the majority of all the observed aggregates did organize in  $\beta$ -sheet rich structures, although topologically quite diverse (see Fig. 4.3C and 4.3F) when comparing individual simulations of the same sequence, as well as between the studied PHF6 and IB12 systems.

Among the most frequently sampled motifs in these heterogeneous structures were two or more opposing smaller  $\beta$ -sheets stacked on top of each other. In addition, mostly single sheets of two to five strands facing a residual portion of disordered peptide chains were observed. With all PHF6 or IB12 peptides assembled into a single aggregate,  $\beta$ -sheets with a perpendicular arrangement, as well as incomplete and distorted barrel-like orientations were formed transiently. A number of the spontaneously formed decameric peptide aggregates featured bilayers of well aligned and tightly laminated  $\beta$ -sheets. The resulting dry sheet interfaces constitute a structural characteristic of steric zipper cross- $\beta$  spines [82], although the observed oligomeric aggregates lacked the regular strand arrangement and complete side chain interdigitation as found in the crystalline conformation. The described conformationally distinct oligomeric states with their relatively stable  $\beta$ -sheet rich subdomain conformations were found to interconvert between various forms of  $\beta$ -sheet aggregates. Indeed, most of the oligomers displayed orientational disorder and were rather dynamic due to an ongoing intra-sheet and inter-sheet side chain repacking. As a result  $\beta$ -sheets were usually bent and twisted to some extent. Irregularly congregated sheets and weakly attached edge-strands were partially prone to break, shift, flip and reform during the simulations.



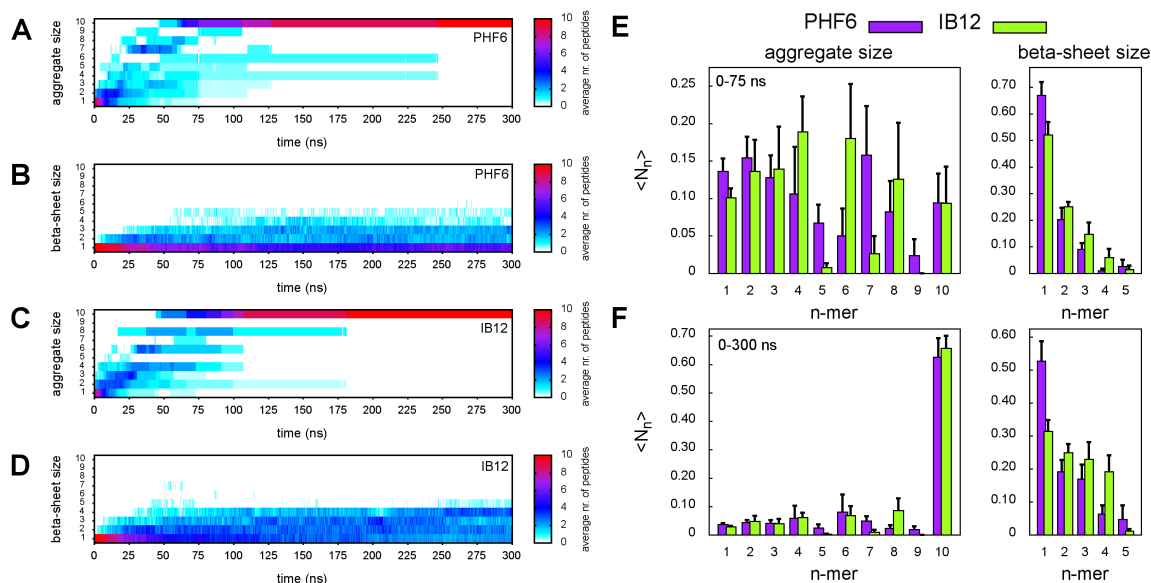


**Figure 4.3: Analysis of the secondary structure evolution for PHF6 and IB12 oligomerization.** The development of secondary structure elements according to DSSP is shown for all independent PHF6 (A, B) and IB12 (D, E) simulations. For PHF6 oligomerization a substantial conversion from random coil (A) to extended  $\beta$ -sheet (B) peptide structure is revealed. A set of representative PHF6 aggregate end-structures (300 ns) are depicted in cartoon representation, showing disordered, as well as intermediate and ordered  $\beta$ -sheet assemblies consisting of several stacked  $\beta$ -sheet subdomains (C). The colors encode  $\beta$ -sheet (red),  $\beta$ -bridge (purple), random coil (green) and bend (yellow) secondary structure elements. The self-association of IB12 peptides and formation of larger aggregates was accompanied by an almost complete conversion from random coil (D) to extended  $\beta$ -sheet (E) peptide structure in all simulation runs. A set of representative IB12 end-structures (300 ns) in cartoon representation shows  $\beta$ -sheet rich oligomers, which were predominantly composed of bilayers of aligned  $\beta$ -sheets. Barrel-like conformations with dry  $\beta$ -sheet interfaces were found to a less extent and exhibited residual disorder (F).

**Aggregate and  $\beta$ -sheet size distributions.** The aggregate size distribution evolving during the series of eight independent trajectories per peptide system was followed. In order to quantitatively probe the aggregation state of the peptides at any given time



## 4.1 Mapping conformational dynamics and pathways of steric zipper peptide oligomerization



**Figure 4.4: Time series and populations of aggregate and  $\beta$ -sheet size distributions.** Averaged populations of general aggregate and  $\beta$ -sheet size  $n$  at simulation time  $t$  are shown for PHF6 (A, B) and IB12 (C, D), respectively. Dark-blue, purple and red colors indicate a high averaged abundance of a particular aggregate or  $\beta$ -sheet size. The normalized aggregate and  $\beta$ -sheet size probability distributions for two time windows during the spontaneous aggregation simulations of both peptides are shown (E, F): from  $t = 0-75$  ns, which approximately corresponds to the early association and growth phase, after which the pool of monomers was found to be depleted in all simulations (E); for  $t = 0-300$  ns, where the distribution was dominated by the final aggregated phase and mainly characterized by a peptide reorganization process within the single, large aggregates. At this stage, stable but partially disordered oligomers with  $\beta$ -sheet subdomains of intermediate size emerged (F).

a hierarchical classification of the formed aggregates and their numerous conformations according to the criteria of general and  $\beta$ -sheet peptide association was pursued (see Methods).

The averaged population of a specific aggregate size  $n$  at a time  $t$  for the PHF6 and IB12 simulations is depicted in Figure 4.4A and 4.4C, respectively. The spontaneous oligomer formation was initiated by a rapid clustering of the peptides in all of the simulations. Starting from monomers, common association states were visited in the early phases of oligomerization. The peptide molecules were found to be dynamic and reversibly associating and dissociating, eventually ending up in the decameric state as indicated by the averaged occupations over all trajectories. The decamer represented a stable aggregation end-product in all simulations, as no substantial dissociation events of one or more peptides were observed.

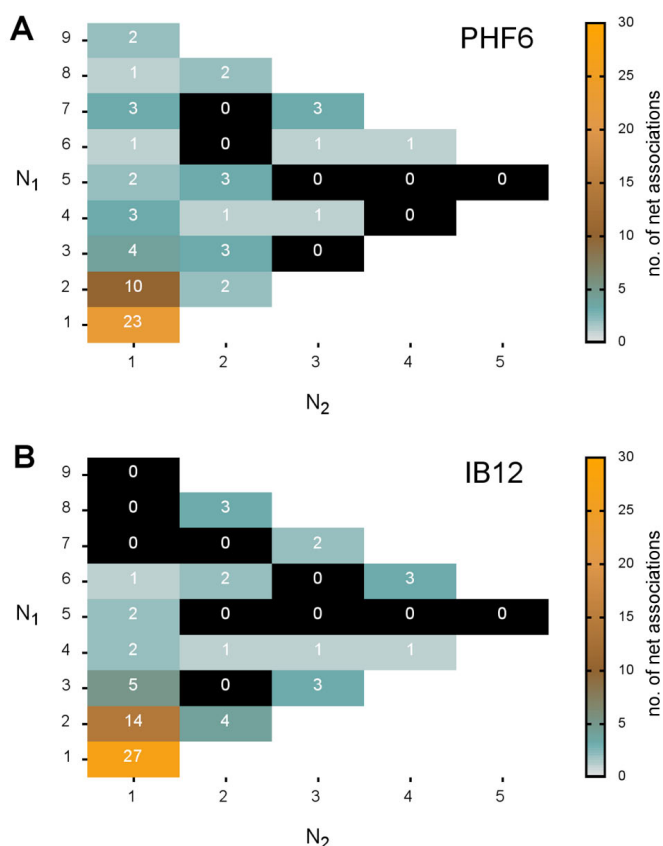
The spontaneous peptide aggregation and oligomer formation process can therefore be characterized by two principal phases: an early association and growth phase where the assembly of peptides proceeded rather fast, and an aggregated phase where no further aggregation or disaggregation events took place, dominated by internal reorganizations.

Aggregates up to trimers were formed on average within a few nanoseconds (Fig. 4.4A and 4.4C). Higher order oligomers ( $3 \leq n \leq 8$ ) were formed within tens of nanoseconds. Interestingly, the IB12 peptide showed a rather uniform distribution of aggregate sizes over time (indicated by dark blue regions). In contrast, for PHF6 a broad and divergent distribution of association states was found. The aggregate populations of the individual simulations highlight the dynamical and complex assembly process into the decamers, suggesting that multiple assembly pathways exist.

From the computed normalized aggregate size probability distribution (Fig. 4.4E and 4.4F), it was found that for PHF6 all possible oligomer sizes were populated at some point. For IB12, odd-numbered aggregate sizes larger than the trimer ( $n = 5, 7$  and  $9$ ) were not or only briefly visited. Furthermore, the spectrum of IB12 peptide aggregate sizes in the association phase showed a preferential population of 4-, 6- or 8-mers compared to PHF6 simulations.

The early phase of peptide aggregation was most prominently marked by the burial of a large fraction of hydrophobic solvent accessible surface area (hSAS) in all performed simulation runs. Half of the initial hSAS was buried upon peptide self-association from monomers into multimeric assemblies and along with the formation of larger aggregates. This demonstrates that the initiation of peptide oligomerization was primarily associated with the reduction of nonpolar peptide surface, predominantly of the side chains. The hSAS values converged around 100 ns for both simulated peptide systems, respectively and displayed only little fluctuation afterwards.

The onset of  $\beta$ -sheet formation (Fig. 4.4B and 4.4D) was slightly delayed in comparison to the general peptide assembly, suggesting that  $\beta$ -sheet formation was not the primary driving force for peptide aggregation. While dimeric or trimeric  $\beta$ -sheets formed relatively fast, noticeable build-up of  $\beta$ -sheet assemblies composed of up to four or five  $\beta$ -strands took place on timescales beyond 50 ns (Fig. 4.4E and 4.4F). However, the oligomers did not grow in one single sheet, as already discussed for representative snapshots from the simulations. They rather showed a tendency to be composed of at least two smaller sheets stacked on top of each other. The average size of these ordered subdomains was found to fluctuate strongly - consistent with frequent formation and breaking of backbone hydrogen bonds, even in the aggregate interior, reflecting the conformational plasticity of the observed oligomers.

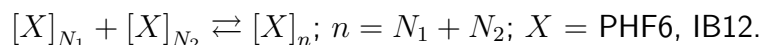


**Figure 4.5: Oligomer formation and growth proceeds by bimolecular associations.** The scheme highlights all possible bimolecular reactions up to the decamer (non-white boxes) and their respective number of occurrence in all of the PHF6 (A) and IB12 (B) simulations. The values in the boxes indicate the number of net association reactions between aggregates of size  $N_1$  and  $N_2$  in the general form:  $[X]_{N_1} + [X]_{N_2} \rightarrow [X]_n$ ;  $n = N_1 + N_2$ ;  $X = \text{PHF6, IB12}$  (i.e. the overall productive oligomer formation and growth steps).

The aggregated phase started with the emergence of stable, but partially disordered decamers. All observed oligomeric end-states showed a low nonpolar surface area and compact arrangements, although the aggregates were still subject to structural fluctuations and reorganizations. This is in accordance with the overall higher abundance of three- and four-stranded  $\beta$ -sheets (Fig. 4.4F), which was nearly twice as high compared to the initial stages of aggregate formation. While overall significant fractions of  $\beta$ -sheets composed of up to five strands were found, the propensity of sampling  $\beta$ -sheet sizes with two to four strands was on average higher in IB12 (2: 25%, 3: 23% and 4: 19%) versus PHF6 (19%, 17% and 6%) oligomers.

**Growth of aggregates proceeds via bimolecular association reactions.** To further quantify and compare the spontaneous aggregation behavior of the PHF6 and IB12 peptide systems, every individual association event in the simulations was traced, starting in all cases from an ensemble of monomeric conformations. To exclude brief, unreactive collisions of peptide molecules from the analysis, only association or dissociation of aggregates which were stable for at least 10 ps were considered.

It was found that oligomer formation and growth proceeded exclusively by bimolecular association reactions, no trimolecular or higher order association reactions were observed. The observed association events thus can be represented in the general form:



The net associations observed in all PHF6 and IB12 simulations are summarized in Figure 4.5A and B, respectively. After the initiation step of the pairing of two monomers and irrespective of the investigated sequence, aggregate growth was found to proceed heterogeneously, as already seen in the series of time dependent oligomer size distributions (Fig. 4.4A and 4.4C). However, two principal reaction types according to basic kinetic models of nonnative protein aggregation could be distinguished [56]: aggregate growth by (chain) polymerization and condensation. Polymerization type reactions cover additions of one or more monomers, whereas condensation reactions involve any other aggregate-aggregate association step that does not directly consume monomers. In the case of PHF6, decameric oligomers and aggregates of all sizes were found to grow by parallel routes, either by adding one monomer at a time (first column in Fig. 4.5A;  $N_2 = 1$ ) or through condensation reactions ( $N_2 \geq 2$ ). For the IB12 assembly a less diverse set of bimolecular association reactions and more of a multi-staged process was observed. Here, the mutual fusion of aggregates (condensation) was the dominant pathway for the formation of larger aggregates. Especially dimeric or trimeric IB12 aggregates were found to condense preferably, as well as dimers with tetra-, hexa-, and octamers. These prevalent association reactions were also reflected in the marginal population of odd-numbered general aggregate sizes for IB12, in contrast to the PHF6 simulations (Fig. 4.4E). Moreover, smaller fluctuations and a larger irreversibility was found for IB12 peptide association in comparison to PHF6, when examining all the individual association and dissociation events. It was found that the efficient assembly into dimers and trimers leads to a fast and irreversible depletion of available IB12 monomers. This offers a plausible explanation why IB12 aggregates larger than heptamers exclusively were found to grow by condensation type reactions. PHF6 peptides were found to associate less efficiently or to dissociate after transient contact formation, which gave rise to more complex assembly pathways all the way to the decamer.

**Conformational properties of oligomeric aggregates with various size.** In order to further access the molecular details of assembly, as well as aggregate structure and dynamics, several essential features of the spontaneously formed oligomers as a function of aggregate order  $n$  were examined (Fig. 4.6). The analysis was based on all performed

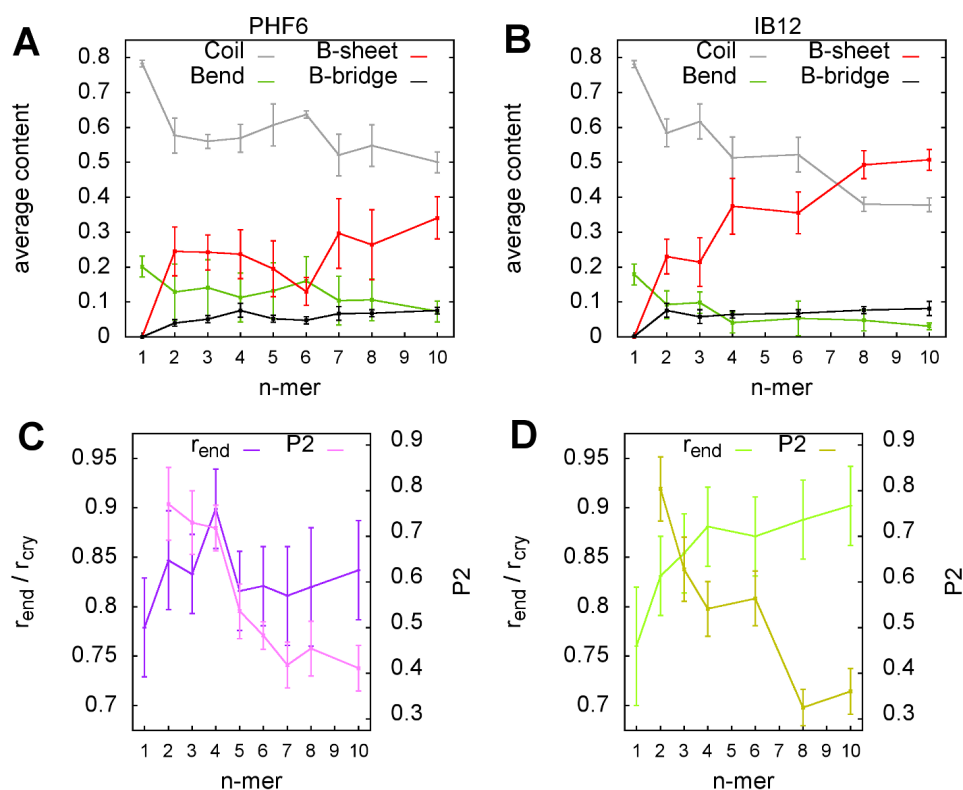


Figure 4.6: **Structural analysis of PHF6 and IB12 oligomers as a function of the number of peptides contained.** Secondary structure propensity according to DSSP based on all simulations for PHF6 (A) and IB12 (B) aggregates. An increase in  $\beta$ -sheet structure with aggregate order  $n$  was observed for both peptide systems. However, the secondary structure content observed for aggregates larger than trimers differed significantly for PHF6 with respect to IB12. The average intra peptide distance between the N- and C-terminal  $C_{\alpha}$ -atoms and the nematic order P2 is shown for PHF6 (C) and IB12 (D) oligomers. The calculated peptide chain extension ( $r_{\text{end}}$ ) was normalized to the respective distance in the conformation found in the crystal structure ( $r_{\text{cry}}$ ). The analysis reveals that the individual PHF6 and IB12 peptide chains extended with the order of the aggregate  $n$  (C and D). The nematic order of the aggregates decreased generally with size. Large PHF6 oligomers (C) had on average a higher orientational order compared to IB12 aggregates (D) of the same size, yet they aggregated in less extended chain conformations.

simulation runs, but oligomer sizes for which statistics were insufficient (less than 50 ns of cumulative occurrence; PHF6:  $n = 9$ ; IB12:  $n = 5, 7$  and  $9$ ) were excluded.

First of all, certain conformational quantities seem to depend on the aggregate order, but to a different extent for both investigated peptide systems. This can be seen for the formation of secondary structure elements as shown in Figure 4.6A and 4.6B. No significant secondary

structure was found for the peptide monomers, as they adopted predominantly coiled (roughly 80 %) and bend conformations regardless of the sequence. Starting with dimers, a significant conversion from coil to  $\beta$ -sheet structure was observed. While PHF6 oligomers of intermediate size maintained the average  $\beta$ -sheet content of the dimeric aggregates (around 25 %), a marked increase in  $\beta$ -sheet conformations from IB12 dimers to trimers and tetramers was detected. The decameric aggregates exhibit the highest  $\beta$ -sheet content among all aggregate sizes of both studied peptide sequences. However, the IB12 decamers showed a higher and more uniform amount of  $\beta$ -sheet structure in comparison to PHF6 assemblies of this particular size.

In terms of structural content per residue as assigned by DSSP [202], it was found that for both peptides the central hydrophobic aliphatic residues Ile3 (PHF6) and Ala3 (IB12), as well as the respective flanking residues have the highest probability to adopt  $\beta$ -sheet conformation. This feature was conserved from the dimers to the decamers. Interestingly, PHF6 and IB12 peptides sampled extended conformations at position five less frequently, which is in both cases a tyrosine residue.

It was found that the end-to-end distances of the peptides were shifted gradually towards the ones of more stretched conformers, as the individual strands organized in larger aggregates (Fig. 4.6C and 4.6D). Especially the IB12 peptide chains were driven to extended conformations due to the presence of inter-peptide interactions, along with the appreciable change in secondary structure. This feature was found to be less distinct for PHF6 peptides, consistent with the persistent sampling of random coil and bend structures.

To assess the alignment order of specific aggregates, the nematic order parameter  $P_2$ , which discriminates between uniaxial, ordered or disordered (amorphous) conformations were used (see Methods). Here  $P_2$  values larger than 0.5 indicate the propensity to be in an ordered, well aligned state. Figures 4.6C and 4.6D show the averaged nematic order of aggregates with size  $n$ . The initially very high orientational order in the spontaneously formed peptide dimers ( $P_2 \approx 0.8$ ) decreased prominently for IB12 trimers and tetramers ( $P_2 \approx 0.55-0.6$ ), whereas PHF6 tetramers still exhibited high nematic order. The averaged orientational order was lowest in the IB12 decamers ( $P_2 \approx 0.35$ ) despite the fact that these aggregates contained the highest amount of  $\beta$ -sheet structure and sampled highly extended conformations (Fig. 4.6D). This can be attributed on one hand to the strong fluctuations due to conformational reorganizations present in all of the aggregates. On the other hand this is due to the degree of disorder caused by lateral stacking and twisting in the larger  $\beta$ -sheet assemblies.

A mixture of parallel and anti-parallel peptide strand alignments was found within  $\beta$ -sheets of all sizes for both, PHF6 and IB12 aggregates. Strands aligned

preferably anti-parallel in IB12  $\beta$ -sheet dimers, whereas parallel orientations were more prominent in larger IB12  $\beta$ -sheet aggregates. A general preference for anti-parallel orientations was persistently found for PHF6  $\beta$ -sheets of all observed sizes. The patterns of peptide strand registry were found to be heterogeneous in both peptide systems.

**PHF6 and IB12 peptides form anti-parallel  $\beta$ -sheet dimers.** As seen from the secondary structure analysis, peptide dimers were critical intermediates in PHF6 and IB12 oligomerization. This particular observation was made with regard to the significant conversion from random coil to  $\beta$ -sheet structure. Furthermore, peptide dimerization was the primary step of the aggregation process. It is therefore of interest to investigate the influence of the specific dimer conformation on the formation of larger oligomeric structures. This was done by analyzing the relevant participation of dimers in the discussed reaction types and association pathways. A detailed analysis of the conformational characteristics and dynamics of the dimeric aggregates was therefore carried out.

A projection of the spontaneously formed PHF6 and IB12 dimers collected from all simulations onto two observables ( $R_g$ , RMSD) is shown in Figure 4.7 together with representative structures of frequently visited conformational states. Overall, both ensembles display a similar landscape featuring a variety of well-aligned anti-parallel, less ordered parallel, orthogonal, compact as well as largely unstructured chain conformations.

From the normalized sampling probability for PHF6 dimers in Fig. 4.7A, it was found that one major and several minor dimer conformations were populated. The statistically most significant structure was an ordered, anti-parallel  $\beta$ -sheet, which deviates 0.62 nm in RMSD from the known parallel crystal structure arrangement. Another frequently sampled conformation did not have extended  $\beta$ -strands, but rather showed disordered, but compact peptide chains, which were stabilized by an isolated  $\beta$ -bridge (0.53 nm RMSD). Parallel, out-of-register  $\beta$ -sheet dimers of intermediate order appear at 0.375 nm RMSD. Parallel, in-register dimers were not formed spontaneously.

Figure 4.7B shows the projection of the spontaneously formed IB12 dimers. An anti-parallel  $\beta$ -sheet with out-of-register conformation (0.22 nm RMSD) was found as the most abundant structure. A collapsed, disordered state (0.3 nm RMSD) stabilized by a single backbone hydrogen bond pair and contacts of the N-terminal side chains were also found with a high probability. Disordered or partially ordered, parallel  $\beta$ -sheet structures were sampled as well, but to a smaller extent. A significant portion of less compact structures with a large radius of gyration was seen in comparison to PHF6 dimers. A notable finding was the prevalent anti-parallel strand align-



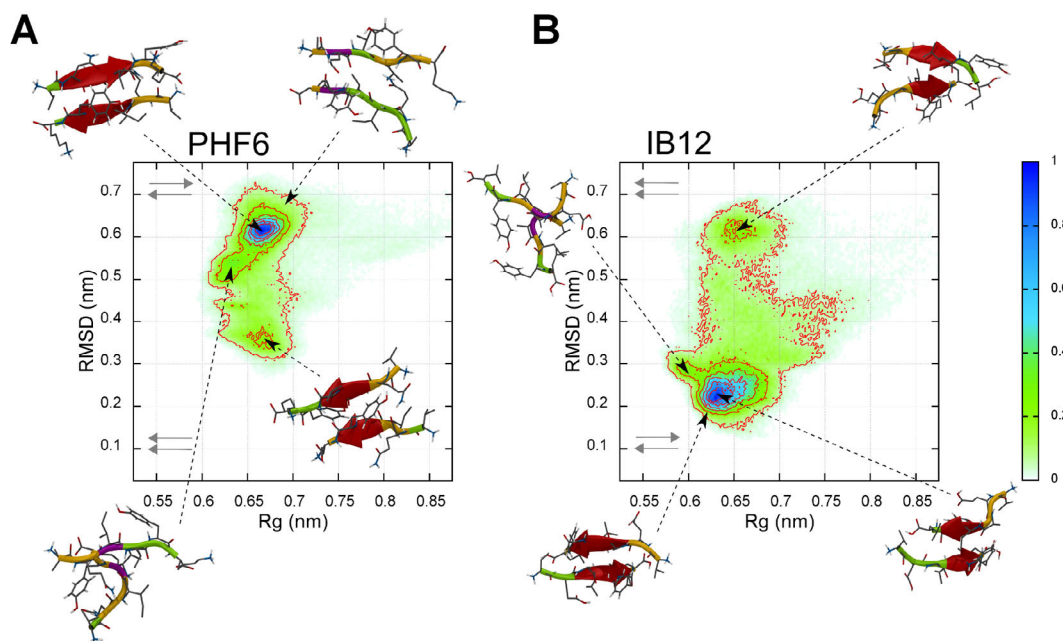
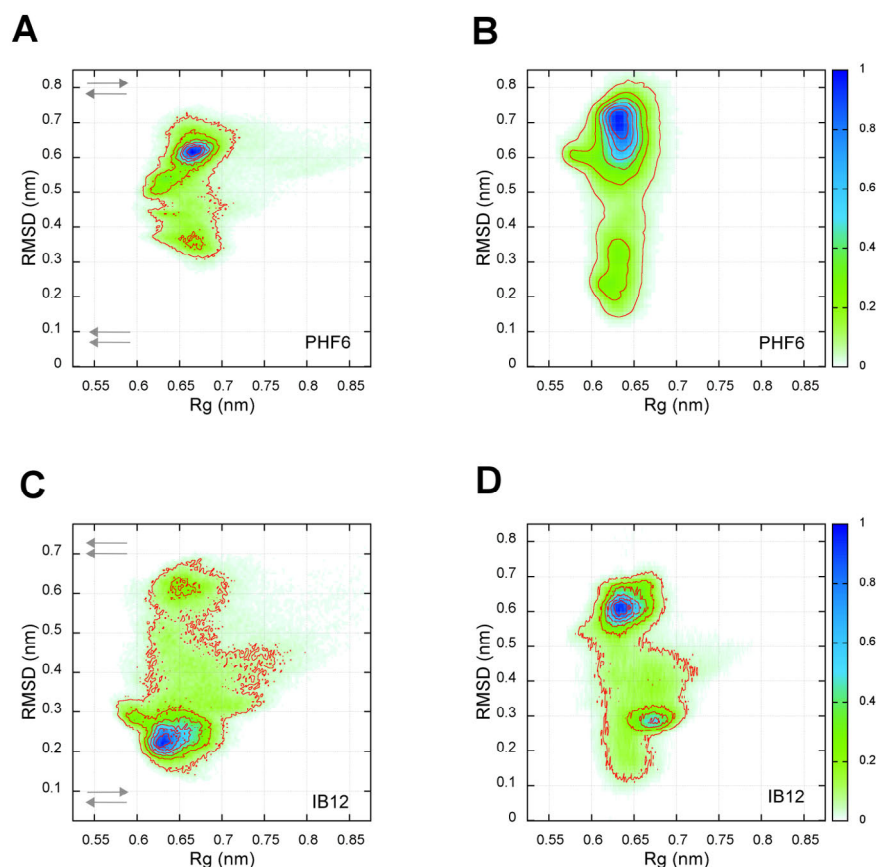


Figure 4.7: **Ensemble of spontaneously formed PHF6 and IB12 peptide dimers.** A projection of the spontaneously formed dimer ensemble as a function of radius of gyration ( $R_g$ ) and the  $C_\alpha$  root-mean-square deviation (RMSD) to the known crystal structure conformation of the PHF6 and IB12 peptides (2ON9, 2OMQ) is depicted, respectively. Corresponding alignment states are schematically highlighted by gray arrows in the upper and lower left part of the projections. Representative structures for peptide dimers with a normalized probability higher than 0.3 are shown for PHF6 (left) and IB12 (right) dimers. The normalized frequency of occurrence scale is given on the right.

ment for the dimer conformations of both peptide sequences. The alignment corresponded well to the native filament pattern of the IB12 crystal structure. However, the spontaneous aggregation simulations of the PHF6 peptides did not sample the formation of ordered, parallel dimers, which approached the respective crystalline reference state with a RMSD of less than 0.3 nm.

To systematically test the stability of the fibril-like dimer peptide arrangement a total of 20 short simulations (15 ns) of isolated dimers per peptide sequence was carried out (data not shown). This particularly addressed the question whether the parallel PHF6 dimer was structurally stable on the nanosecond timescale. Two exemplary simulations were prolonged up to 1  $\mu$ s to probe the dimer dynamics beyond the nanosecond timescale (Fig. 4.8). Two preformed  $\beta$ -sheets in an ordered, either parallel or anti-parallel arrangement were chosen as initial conformations and representative dimeric states. One of them was the experimentally





**Figure 4.8: Ensemble of spontaneously formed and isolated PHF6 and IB12 peptide dimers.** Projections of various dimer ensembles as a function of radius of gyration ( $R_g$ ) and the  $C_\alpha$  root-mean-square deviation (RMSD) to the known crystal structure conformation of the PHF6 and IB12 peptides (2ON9, 2OMQ) are depicted, respectively. Spontaneously formed dimers (A - PHF6 and C - IB12) and isolated dimer conformations, which were simulated additionally on long timescales (B - PHF6 and D - IB12) are shown. The projections of the isolated dimer conformations were obtained from two 1  $\mu$ s long simulations, respectively. The normalized frequency of occurrence scale is given on the right.

determined crystal structure conformation, the other one was an ordered  $\beta$ -sheet dimer extracted from the spontaneous aggregation simulations with opposite strand polarity.

In summary, the ensembles of spontaneously formed PHF6 and IB12 dimer structures were relatively heterogeneous with predominantly anti-parallel chain conformations. However, from multiple validation simulations of isolated dimer structures it was found that parallel PHF6 and IB12 dimers are stable as well. For the isolated anti-parallel IB12 reference dimer structure only a limited kinetic stability was revealed. The preformed PHF6 and IB12  $\beta$ -sheet dimer

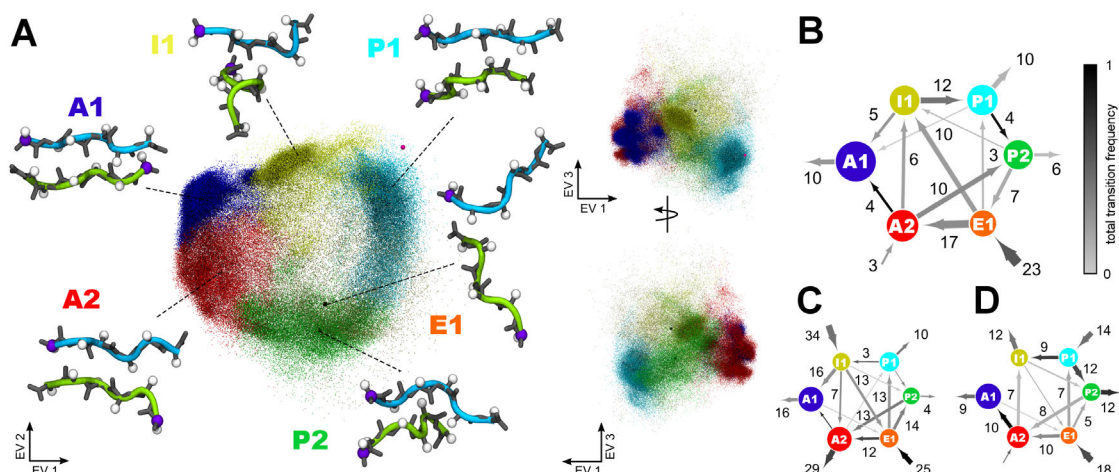
conformations can inter-convert between different alignment states on the sub-microsecond timescale, which is an order of magnitude longer than the average life time of dimers in the spontaneous aggregation simulations.

In addition to simulations with GROMOS96 43A1, the AMBER99SB and CHARMM27 force fields were included to test the structural properties of preformed and encountering PHF6 and IB12 dimer aggregates. All these force fields are frequently used and have been shown to perform particularly well in peptide aggregation and folding simulations [115, 121, 171]. The different molecular mechanics force fields were compared in their ability to characterize and preserve the isolated dimer structures (data not shown). To that end way the findings for the simulations of spontaneous peptide aggregation have been evaluated and validated with a consensus force field approach [171]. In the simulations of preformed, anti-parallel PHF6 dimers, as well as parallel IB12 dimers, a similar behavior and stability in all the tested force field variants was observed. Therefore it appears that the choice of force field is only a minor concern with respect to stability of preformed  $\beta$ -sheets.

**Conformational mapping on collective coordinates yields distinct conformational states of PHF6 and IB12 dimers.** To address the question how the conformational dynamics of the peptide dimerization plays a role in the selective population of the anti-parallel conformations and how this can translate to structural features of larger aggregates, a novel method of conformational mapping is introduced (see Methods).

With a principal component analysis (PCA) collective coordinates were obtained to describe a peptide aggregate by means of its large-scale structural fluctuations in a space of reduced dimensionality. Thereby the conformational changes associated to, for example initial peptide encounter complexes, but also conformational transitions within formed aggregates was determined. To be able to map multimeric aggregates larger than dimers, oligomeric structures were described in terms of minimal, dimeric units. The application of a clustering procedure then allows for a consistent description of conformational distributions and association modes. With this approach a fundamental challenge is addressed, which is the determination of a way to analyze the structural aspects of the large and manifold ensemble of multimeric aggregate conformations sampled in the simulations.

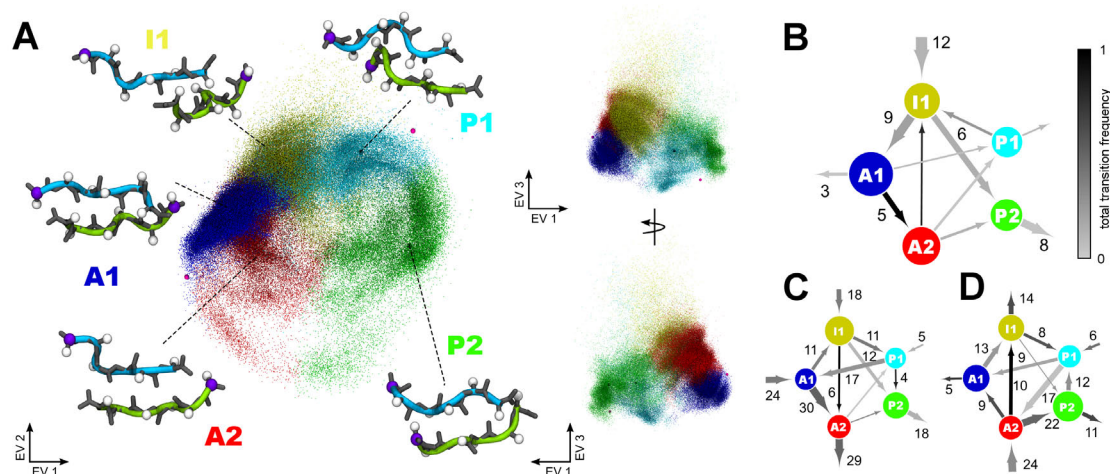
First, the main structural features of the dimeric states as revealed by the PCA and the conformational clustering will be presented. Figures 4.9A and 4.10A show a three-dimensional representation of the conformational space of the spontaneously formed PHF6 and IB12 peptide dimers, projected onto the first three principal components - eigenvectors - EV1, EV2 and EV3,



**Figure 4.9: Mapping of PHF6 aggregate sizes and transitions between individual conformational clusters.** On the left side of the panel: Visualization of PHF6 dimer conformations in a three-dimensional principal subspace of collective fluctuations (A). Each dot corresponds to a structure of the simulation projected on the first three eigenvectors as obtained from a principal component analysis. The projection is viewed from three different angles. The colors indicate the assignment of a PHF6 dimer structure to one of six conformational states: A1, A2, P1, P2, I1, E1. The states were labeled with a capital letter according to the overall dimer arrangement within the respective cluster: A - anti-parallel dimer; P - parallel dimer; I - intermediate ordered or irregular dimer structure, E - encounter complex. Cluster centers are represented as black spheres, parallel and anti-parallel reference dimer conformations are highlighted as magenta spheres. The center structures from each cluster are shown as main-chain (gray sticks) and  $C_\beta$  atoms (white spheres), as well as the backbone in a cartoon representation. The N-terminal  $C_\alpha$  atom of each chain is shown in purple. On the right side of the panel: Transition networks illustrate the net transitions between the clusters as observed for dimers (B), trimers (C) and tetramers (D). The node colors match the clusters in the projection, the node size is consistent with the total number of assigned structures. The width and direction of the node edges corresponds to the number of net transitions. The total transition frequency between two nodes is color coded according to the gray scale on the top right corner.

respectively. The conformational clusters were labeled according to definitions given in Figures 4.9 and 4.10, respectively.

Surprisingly, the obtained collective coordinates of the PHF6 and IB12 dimer structure ensembles were found to be similar. From both the projections, one can identify anti-parallel (cluster A1 and A2) and parallel (cluster P1 and P2) dimer conformations, which are discriminated along EV1. A twisting mode is described by EV2 and alters the relative chain orientations, respectively. In the case of the mapped PHF6 dimers (Fig. 4.9A) several dense



**Figure 4.10: Mapping of IB12 aggregate sizes and transitions between individual conformational clusters.** On the left side of the panel: Visualization of IB12 dimer conformations in a three-dimensional principal subspace of collective fluctuations (A). Each dot corresponds to a structure of the simulation projected on the first three eigenvectors as obtained from a principal component analysis. The projection is viewed from three different angles. The colors indicate the assignment of an IB12 dimer structure to one of five conformational states: A1, A2, P1, P2, I1. The states were labeled with a capital letter according to the overall dimer arrangement within the respective cluster: A - anti-parallel dimer; P - parallel dimer; I - intermediate ordered or irregular dimer structure. Cluster centers are represented as black spheres, parallel and anti-parallel reference dimer conformations are highlighted as magenta spheres. The center structures from each cluster are shown as main-chain (gray sticks) and  $C_{\beta}$  atoms (white spheres), as well as the backbone in a cartoon representation. The N-terminal  $C_{\alpha}$  atom of each chain is shown in purple. On the right side of the panel: Transition networks illustrate the net transitions between the clusters as observed for dimers (B), trimers (C) and tetramers (D). The node colors match the clusters in the projection, the node size is consistent with the total number of assigned structures. The width and direction of the node edges corresponds to the number of net transitions. The total transition frequency between two nodes is color coded according to the gray scale on the top right corner.

regions in cluster A1 and A2, as well as P1 and I1 were evident. Interestingly, the center structures of cluster A1, I1 and P1 correspond well to the representative ordered anti-parallel, partially ordered parallel and disordered structures in Figure 4.7A. Especially sparse and diffuse were the mapped structures along EV3 in the upper part of both projections. Here, the center of mass distance between the two peptide chains was large, and indeed a chain separation mode was the one mostly described by the third eigenvector. It can also be seen that for PHF6 around the reference parallel dimer (magenta sphere in the upper right of

Figure 4.9A) no sampling density exists. For the mapped conformational space of the IB12 peptide dimers (Fig. 4.10A), an important difference to PHF6 was the identification of only five clusters as compared to six found for PHF6. Furthermore, less symmetry with respect to the mapping of the IB12 structures along EV2 was seen. Again, cluster and densely populated regions matched the prominently sampled states in Figure 4.7B (A1, A2, P1, I1). In the spontaneous aggregation of IB12 peptides neither fully ordered parallel nor anti-parallel in-register structures were sampled (magenta spheres in Fig. 4.10A). Averaged structural features of the spontaneously formed dimers were extracted based on the conformational clustering and are summarized in Tables 4.2 and 4.3. To compare the distinct structural properties of the clustered aggregate structures to one another, the hydrogen bond energy according to Espinosa *et al.* [265] was calculated and decomposed it into several contributions. In addition, the hydrophobic solvent accessible surface, radius of gyration and average number of residues in  $\beta$ -sheet conformation for each cluster were analyzed. The cluster A1 was the most populated state for the two studied peptides. A1 has the highest amount of  $\beta$ -sheet, the lowest total hydrogen bond energy for the dimer complex and also the most reduced hydrophobic surface area, respectively. Inter side chain hydrogen bonds are negligible in the case of IB12 dimers, while there were substantial energetic contributions for PHF6, especially in the parallel dimers of cluster P1. To obtain the extent and pattern of inter-peptide contact formation present in the individual conformational clusters, the average probability of inter-peptide residue pair contacts was calculated. This was achieved by averaging over all the structures of a cluster, respectively. The color-coded contact maps are shown in the Fig. 4.11 and 4.12 and reflect the overall arrangement of the peptide chains (alignment, registry) in the respective conformational cluster. It was found that the central hydrophobic residues formed contacts with high probability, both in PHF6 (Ile3, Val4) and IB12 (Ala3, Leu4) dimers. Interestingly, the averaged contact maps for the individual IB12 dimer clusters show that the hydrophobic C-terminus (Leu6) was involved in many interchain contacts. In the case of PHF6 dimers, fewer contacts were observed involving the C-terminal end, while the Gln2 residue was found to participate prominently in contact formation (e.g. in cluster A2, P2 and E1).

**Kinetics of dimerization are sequence dependent.** The obtained ensemble of spontaneously formed dimer structures cover every step from diffusional encounter of two peptide molecules to primary contact formation and conformational reorganization of the initially formed aggregates. In order to analyze the kinetics of dimerization in more detail, transition networks between the identified clusters of peptide conformations were constructed

Table 4.2: Averaged structural features of the conformational clusters of the PHF6 dimers.

cluster	f	$N_{\beta}$	$hbE_{tot}$	$hbE_{amid}$	$hbE_{sc-sc}$	$hbE_{mc-sc}$	hSAS	$R_g$
A1	0.34	2.4	-104.72	-80.26	-2.69	-6.26	9.63	0.671
P1	0.15	1.1	-78.54	-52.40	-3.61	-7.48	9.90	0.667
P2	0.13	1.1	-68.21	-46.94	-1.99	-7.55	9.99	0.668
I1	0.14	1.1	-79.60	-54.46	-1.70	-5.84	9.83	0.655
E1	0.07	1.0	-63.30	-40.80	-1.81	-7.27	10.96	0.832
A2	0.17	0.7	-64.02	-37.78	-1.33	-7.77	10.07	0.677

f: Fraction of dimer structures assigned to this conformational cluster.

$N_{\beta}$ : Number of residues in  $\beta$ -sheet conformation (per chain).

$hbE_{tot}$ : Total hydrogen-bond energy of the dimer complex (kJ/mol).

$hbE_{amid}$ : NH - CO backbone hydrogen-bond energy (kJ/mol).

$hbE_{sc-sc}$ : side chain - side chain hydrogen-bond energy (kJ/mol).

$hbE_{mc-sc}$ : main chain - side chain hydrogen-bond energy (kJ/mol).

hSAS: hydrophobic solvent accessible surface area (nm<sup>2</sup>).

$R_g$ : Radius of gyration (nm).

(see Methods). Thereby mechanistic insight into common and alternative pathways was gained. Here, the focus was on the pathway analysis of assembly into dimers, trimers and tetramers (Fig. 4.9B-D and 4.10B-D), as these aggregates were involved in the primary aggregation stage, covered most of the initial association steps and yielded adequate statistics. First of all the focus will be on the respective dimerization events. From Figure 4.9B it is seen that encountering PHF6 peptides frequently attached loosely or transiently via structures similar to E1. This was predominantly facilitated by the contact formation of the N-terminal Gln2 residue of one peptide to the C-terminal Lys6 of the other (Fig. 4.11). The dimeric aggregates from this generic encounter complex ensemble (E1) were, however, not stable and evolved further via two main pathways. Either a sequential zipping up of the extended peptides to result in A2 type structures or a collapse of the peptide chains towards strongly disordered I1 conformations was identified. Here, the compaction mainly originated from the burial of the central hydrophobic parts (Ile3, Val4). Access to the ordered anti-parallel dimer conformations of cluster A1 was found to be possible from both of these configuration types and accompanied by increasing backbone hydrogen bond interactions (see Table 4.2). However, dimer structures from the I1 cluster were found to reorganize predominantly to parallel dimers (P1, P2).

The IB12 dimer ensemble lacks a discrete encounter complex cluster with well separated



Table 4.3: Averaged structural features of the conformational clusters of the IB12 dimers.

cluster	f	$N_{\beta}$	$hbE_{tot}$	$hbE_{amid}$	$hbE_{sc-sc}$	$hbE_{mc-sc}$	hSAS	$R_g$
A1	0.35	2.1	-99.75	-85.55	-0.76	-1.07	9.87	0.644
P1	0.10	0.9	-59.62	-47.42	-0.64	-2.96	10.20	0.674
P2	0.19	0.5	-67.32	-51.70	-0.20	-2.16	10.36	0.718
I1	0.12	1.4	-70.77	-58.26	-0.28	-4.44	10.11	0.663
A2	0.24	1.2	-77.87	-67.35	-0.61	-2.20	10.13	0.660

f: Fraction of dimer structures assigned to this conformational cluster.

$N_{\beta}$ : Number of residues in  $\beta$ -sheet conformation (per chain).

$hbE_{tot}$ : Total hydrogen-bond energy of the dimer complex (kJ/mol).

$hbE_{amid}$ : NH - CO backbone hydrogen-bond energy (kJ/mol).

$hbE_{sc-sc}$ : side chain - side chain hydrogen-bond energy (kJ/mol).

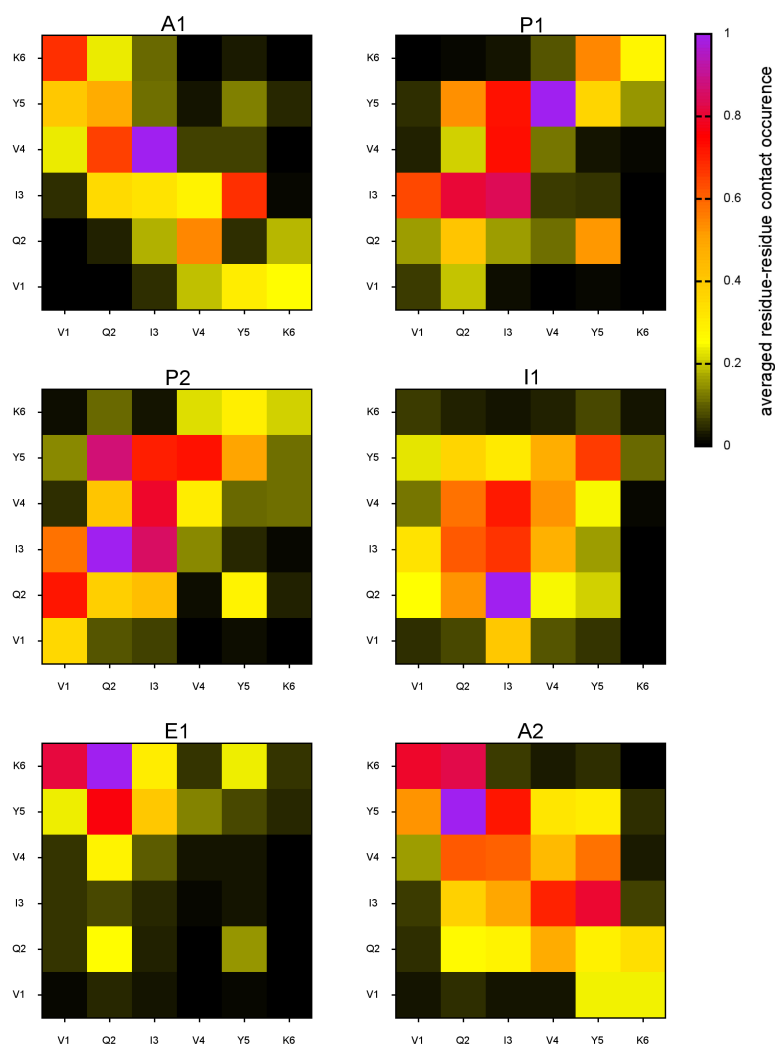
$hbE_{mc-sc}$ : main chain - side chain hydrogen-bond energy (kJ/mol).

hSAS: hydrophobic solvent accessible surface area (nm<sup>2</sup>).

$R_g$ : Radius of gyration (nm).

peptide chains, as can be seen in Figure 4.10B. A common first conformational state of peptide association exists in the disordered, but compact cluster I1 type structures, which were stabilized by packing interactions of the hydrophobic C-terminal Leu4, Tyr5 and Leu6 residues (Fig. 4.12). From this pool of conformations transitions to ordered conformations in the large cluster A1 and out-of-register A2 dimers occurred, as well as to the P2 cluster. A direct interconversion of P1 and P2 structures was not seen, while structure transitions between clusters A1 and A2 were observed with a high frequency.

To summarize the observations made for the dimers so far, the pathway analysis indicates that the peptide dimerization is facilitated through the formation of specific key residue contacts between the two strands, respectively. From less directional, early encounter conformations a general ordering transition with increased backbone hydrogen bond interactions was observed. Both peptide systems differed in the dynamics and complexity of association pathways. For IB12 peptides the initial hydrophobic collapse brought the peptide chains together rapidly. These less ordered dimer conformations are characterized by a large number of mutual contacts and a significant portion of backbone hydrogen bonds. From there a reorganization generally took place that further maximized the inter-peptide interactions. In the case of PHF6, the observed dominant association pathway started from a generic encounter complex with specific side chain contacts. Here, a sequential gain of interactions was found, where residue contacts and backbone hydrogen bonds formed throughout the



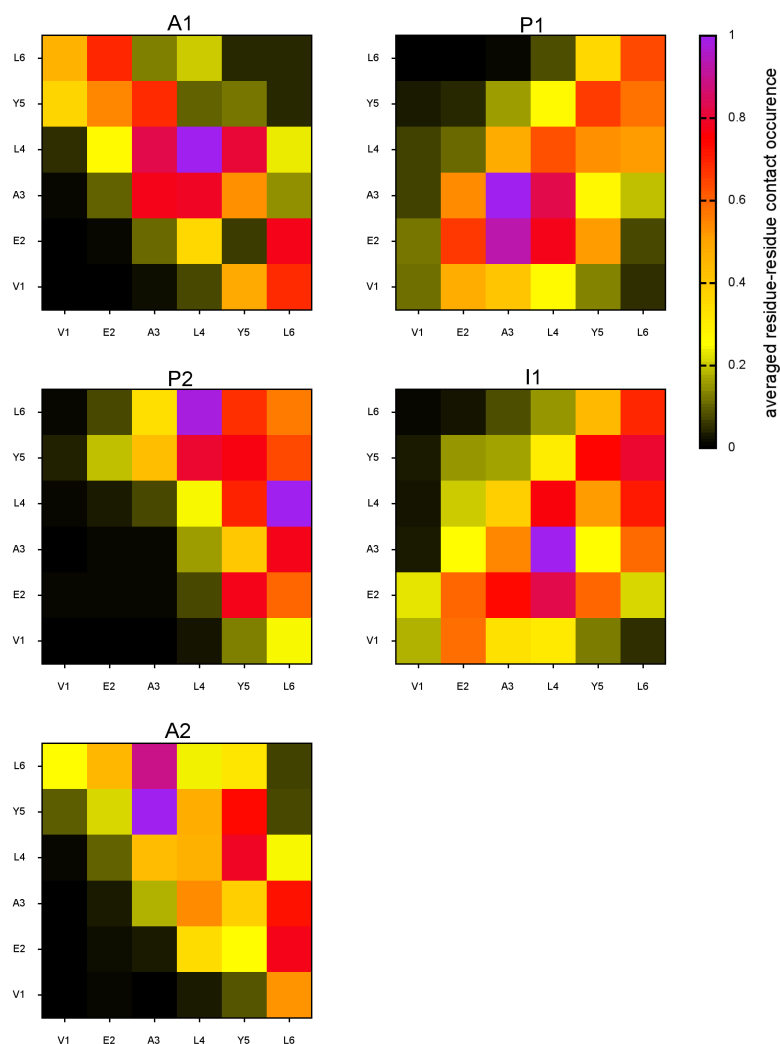
**Figure 4.11: Residue-Residue contact map for individual PHF6 dimer clusters.** The calculation was performed separately for the dimer structures of each of the six identified conformational states shown in Figure 7. The map is colored by the average occurrence of inter-peptide residue pairs, which share at least one heavy atom contact. The scale is given on the right top.

molecules in a zipper-like fashion. In addition, a less ordered collapse of peptide chains prior to conformational rearrangement was found.

**Dimeric versus multimeric conformational dynamics.** The trimer and tetramer ensembles were probed through the perspective of dimeric structures to assess the changes to the dynamical behavior of these higher order oligomers with respect to the actual dimers (see Methods). Here, more complex dynamics were found, reflected in the different transition patterns among the conformational cluster types, as well as in and out of the network (Fig. 4.9C-D and 4.10C-D). In the following a number of mechanistically relevant findings obtained from these complex transition networks will be expanded on.

For PHF6 it was found that the pathway identified for the dimerization (from cluster E1





**Figure 4.12: Residue-Residue contact map for individual IB12 dimer clusters.** The calculation was performed separately for the dimer structures of each of the five identified conformational states shown in Figure 8. The map is colored by the average occurrence of inter-peptide residue pairs, which share at least one heavy atom contact. The scale is given on the right top.

to anti-parallel conformers) persisted for trimers and tetramers. Transitions from E1 to parallel structures (P1, P2) were now seen as well. In particular P2 type conformations were often found to be prone to rearrange to A2. Ordered anti-parallel dimers (A1, A2) were the largest conformational ensemble and also dominant end-conformations (exit nodes). However, the described conformational reorganization and ordering transitions were slower or less often sampled. Evidence comes from a stall at the intermediate locked and partially zipped A2 structures in the trimer and the collapsed and disordered structures (I1), which were mainly found as the initial or end-conformations.

In the case of IB12 trimer and tetramer assembly dynamics mainly involved an ordered, preformed complex (A1), a seed structure onto which incoming peptides collapsed and added up to in a parallel fashion. The preformed  $\beta$ -sheet dimers acted as suitable templates and parallel

$\beta$ -sheets formed easily in many instances. This demonstrates a directional strand assembly different from the dimerization case. The P2 and P1 clusters were the largest sampled clusters for the aggregates of order 3 and 4. In the process of accommodating a free monomer or dimer, the preformed structured oligomer underwent large fluctuations, which is linked to orientational disordering of the individual peptides (see Fig. 4.6B). This is in line with the sizable population of the cluster A2, P2 and especially I1. In contrast to the PHF6, rearrangement from parallel to anti-parallel IB12 alignment states was found and did not involve intermediate structures of type A2 or I1. This pathway was especially frequently seen in the tetramer. Similarly, reptation out-of-register transitions to go from A1 to A2 were found. This suggests that the possibility to transit among different conformation types within an aggregate was dependent on the oligomer size and characteristics. The increased internal fluctuations of the aggregates and flexibility of the peptide molecules was more pronounced for IB12 compared to PHF6. A likely reason is the larger hydrophobic patch in the IB12 sequence, and in turn the larger influence of weak dispersion interactions on aggregate energetics and dynamics.

### 4.1.4 Discussion

**Oligomer structure and dynamics of association and reorganization are heterogeneous.** The observed early oligomers were found to be partially ordered structures rather than completely extended peptide chain conformations [108, 113, 115, 124, 266, 267] and lack a uniform strand registry and alignment. Densely packed oligomers with residual orientational disorder, but predominantly high  $\beta$ -sheet content were found. Although topologically quite diverse and heterogeneous structures emerged from the simulations after 300 ns, some of these aggregates displayed structural characteristics of the crystalline conformation, namely stacked  $\beta$ -sheet bilayers with steric zipper-like interfaces [82, 100]. The aggregates were found to grow via multiple and diverse bimolecular association reactions. The work demonstrates that the formed aggregates are dynamic and undergo substantial conformational reorganization during the growth and accommodation of monomers and small intermediates. Moreover, the characterization of multichain configurations by decomposition and mapping of dimeric structures reveal differences in conformational ensembles of oligomers of different orders and an assembly process that was heterogeneous at the molecular level [106, 130]. This indicates that association and reorganization pathways are dependent on the oligomer size and characteristics. The interconversions between many distinct oligomeric states after the growth stage occurred without full or partial detachment of the peptide strands from the aggregate. Multiple internal reorganization pathways were found, which involved

sliding or reptation of individual strands and  $\beta$ -sheet subdomains relative to each other. All the observed oligomeric states therefore exhibit compact arrangements with a low radius of gyration [266]. These findings are consistent with a proposed aggregation mechanism that has been observed in simulations of various amyloidogenic peptides [106, 115, 122, 122, 268] and has been also found experimentally [55], prevalent at high concentrations.

**Hydrophobic character of peptide and  $\beta$ -sheet content of aggregates are correlated.** The assembly and ordering dynamics of the early oligomers are likely governed by the burial of hydrophobic side chains and intermolecular hydrogen bonding. The peptide self-assembly was found to be primarily accompanied by the desolvation of hydrophobic parts of the molecules. This is in line with the observations from various experimental and theoretical studies on low molecular weight oligomers [53, 106, 108, 114, 124, 125, 131, 266, 269, 270].

From both investigated peptide sequences, IB12 is the more hydrophobic, featuring a patch of adjacent hydrophobic residues. Interestingly, a decreased proportion of orientationally well-aligned IB12 peptide aggregates was found, while the amount of  $\beta$ -sheet structure was increased with respect to PHF6 aggregates of all observed sizes [114, 133]. At the same time it was observed that the mutual association of smaller aggregates (condensation) was the preferred growth pathway for larger IB12 oligomers. Here, an elevated level of backbone hydrogen bonding is directly correlated with a marked increase in hydrophobic burial, as well as the onset and magnitude of single  $\beta$ -sheet formation and  $\beta$ -sheet stacking. The critical importance of nonpolar surface burial in peptide self-assembly has been specifically attributed to single  $\beta$ -sheet layer stability and the lateral lamination of  $\beta$ -sheet [47, 119, 270–272].

**Kinetics impact the primary steps of amyloidogenic peptide assembly.** The structural characteristics of the spontaneously assembled aggregates were probed and a rich variety of intermediates observed. A striking observation was the preferential anti-parallel inter-strand orientation in both the PHF6 and IB12 dimers, which could be directly traced back to key residue interactions in commonly observed encounter complexes, respectively. It is suggested that basic features like the  $\beta$ -strand alignment (parallel vs. anti-parallel) can be kinetically determined at the early stages of assembly for aggregates as small as dimers [123].

In Fig. 4.8 the projections of the spontaneously formed dimers together with the isolated dimers are shown, which were additionally simulated on long timescales. The conformational distributions are comparable for the different PHF6 ensembles, whereas the IB12 ensembles differ substantially. The isolated IB12 dimers that were allowed to relax for 1  $\mu$ s, were mainly

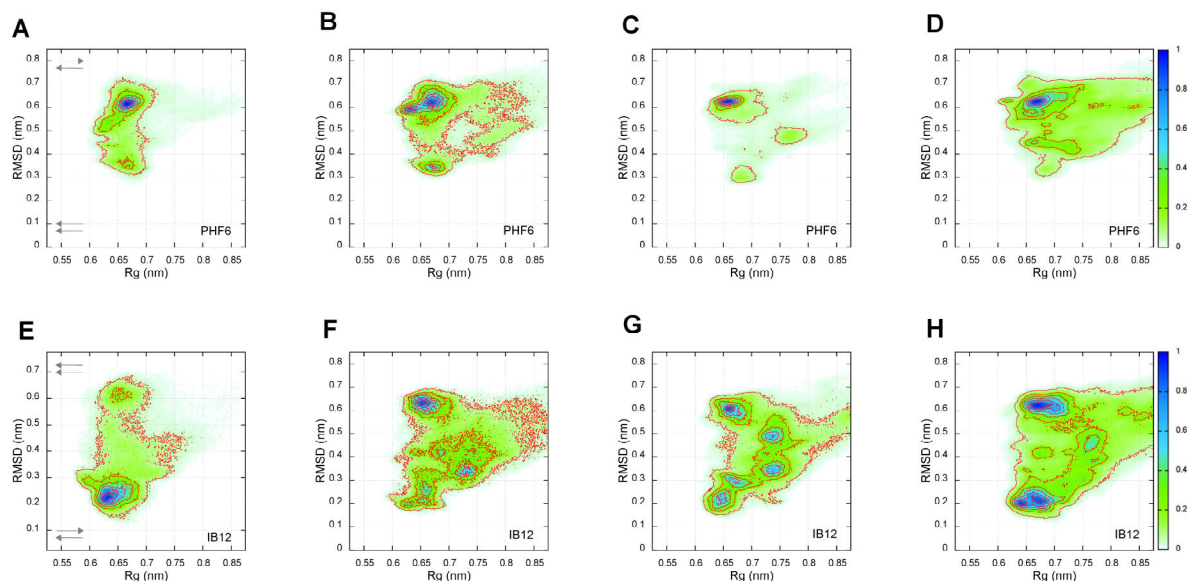
parallel in contrast to the spontaneously formed anti-parallel ones. It was concluded that kinetic trapping was in particular relevant for the described anti-parallel IB12 dimer formation, which was in turn mainly driven by a hydrophobic effect. As seen from the transition networks and in the isolated dimer simulations, the two alignment states can inter-convert in the dimer nevertheless, as the barriers separating these states are rather small. For larger aggregates this interconversion is limited to edge strands. Hence, the non-equilibrium situation in the present study did not allow for structural relaxation of the spontaneously formed aggregates. The rapid oligomer growth therefore is the main factor for the observed kinetic control of dimer interfaces. Preformed  $\beta$ -sheet dimers were found to be involved in the addition of isolated monomeric peptides in solution, as well as in the growth of larger aggregates. Do therefore kinetically trapped small aggregates determine the structural evolution of larger oligomers, as the presence of a number of alternative end-structures in the decamers suggest?

To answer this question it was investigated if the preferentially anti-parallel strand alignment was also observed in the higher order oligomers due to kinetic trapping. Various dimer ensembles were compared and the projections of combined dimers obtained by decomposing trimer, tetramer and decamer aggregates into dimers are shown (see Fig. 4.13).

The decomposed multimeric PHF6 aggregates display a similar landscape to the spontaneously formed 'true' dimers. In the case of IB12, a mixed strand alignment pattern for trimers, tetramers and decamers is observed, which is different from the mainly anti-parallel 'true' dimers, but also not the same as found for isolated, relaxed dimers. In fact, the results support the idea that a structure can be selected kinetically during early stages of assembly, where the nucleation barrier and hence production rate determines the abundance of the different aggregate morphologies and structural forms [117, 127]. This suggests that a rich structural heterogeneity or polymorphism on the dimer level can translate to the ensemble of oligomer conformations [106, 267], even at concentrations far below to the one investigated here [123]. Experimental evidence shows that alternative states of peptides in the condensed phase encompass a broad and diverse spectrum of oligomers and protofibrils, which themselves are polymorphic [66, 269]. Additionally, polymorphic forms of mature amyloid fibrils are well documented and can originate from variations in filament architecture and organization [35, 43, 248].

The study of short amyloidogenic peptides supports the notion that alternative packing schemes of highly ordered steric zipper conformations in the crystal structures and fibrils serve as a basis for molecular polymorphism [82, 102–104, 119, 269]. A different strand alignment pattern in the simulations of spontaneously assembled PHF6 and IB12 oligomers was found than experimentally determined in the x-ray structures by Nelson and co-workers [82],

#### 4.1 Mapping conformational dynamics and pathways of steric zipper peptide oligomerization



**Figure 4.13: Ensemble of PHF6 and IB12 peptide dimers derived from different aggregation states.** Projections of various dimer ensembles as a function of radius of gyration ( $R_g$ ) and the  $C_\alpha$  root-mean-square deviation (RMSD) to the known crystal structure conformation of the PHF6 and IB12 peptides (2ON9, 2OMQ) are depicted, respectively. Spontaneously formed dimers (A - PHF6 and E - IB12) and ensemble of trimers, tetramers and decamers (B, C, D - PHF6 and F, G, H - IB12) are shown, respectively. Ensembles of higher order oligomers were obtained by decomposition of the respective multimer into dimers (see Methods). The normalized frequency of occurrence scale is given on the right.

although the peptide concentration and pH are in accord with the crystallization conditions. It is assumed that the protonation state of the C-terminus primarily affects the direction of strand alignment in the short steric zipper peptides. Since only static protonation states are considered throughout the simulations, a strong influence on PHF6 and IB12 peptide assembly is expected. As a result, the aggregate conformations might be strongly affected by a dominant electrostatic effect of the termini. That is because the anti-parallel/in-register state was associated with the strongest attraction between the charged terminal residues. It is suggested that the packing of these peptides in the crystal is determined by a delicate balance of different factors [106, 119] (e.g. electrostatic contacts between symmetry mates, presence of solvent and counter ions), which implies alternative or co-existing  $\beta$ -sheet bilayer conformers of similar stability. Furthermore, possible hierarchical or sequential assembly scenarios [104, 119, 273] can also affect the steric zipper structure selection and therefore render an extrapolation from the early oligomers in the simulations to the crystalline or fibrillar end-product challenging. It is

important to note that the present findings, however, do not rule out that the spine architecture of VQIVYK and VEALYL fibrils may resemble the steric zipper motifs identified by x-ray microcrystallography [82]. The spontaneously formed oligomers sample essential structural features of the steric zipper conformation: stacked  $\beta$ -sheets with a dry interface, displaying tight side chain interdigitation. It is therefore conceivable that the observed structural transitions at the decamer level are indicative of on-pathway sampling to the mature fibrils.

**Spontaneous oligomer formation occurs fast.** The various spontaneously formed multimeric aggregates demonstrate that, the initially monomeric peptide molecules interact without encountering any major barriers. Furthermore, it was seen that stable PHF6 and IB12 oligomer formation by association of smaller intermediates and remaining free monomers progressed rapidly, leading in all cases to one big aggregate. This indicates a fast oligomer growth up to a size of 10 peptide chains. This suggests that the primary aggregation and pre-nucleation stage is primarily an energetically downhill process. A recent Monte Carlo study on AcPHF6 oligomerization reports a similar scenario [116]. There the nucleation of a fibril competent species for further growth was not required until aggregates became larger than at least decamers.

The early and dynamic oligomers of minimal steric zipper peptides observed in the study were found to be pre-structured and did not condense in a fluctuating micelle-like arrangement, which are held together mainly by weak dispersion interactions as proposed for assemblies of longer peptides [52, 53, 80]. The conformational reorganizations necessary to access the highly ordered fibrillar state with a sterically complementary  $\beta$ -sheet interface mainly concern the peptide strand alignment and repacking of the side chains. The nucleation of a growth-competent steric zipper oligomer species might therefore be strongly disfavored by entropic arguments, thereby explaining the gap between the fast oligomerization as observed here and the known slow kinetics of in vitro fibril formation. This work has been focused on the primary aggregation events. The oligomeric end-states of the simulations show interesting structural reorganization dynamics that warrants further analysis. This will be the subject of a following study.

### 4.1.5 Conclusion

Atomistic MD simulations of the unbiased spontaneous aggregation process of PHF6 and IB12 steric zipper peptides from unstructured monomers to  $\beta$ -sheet rich oligomeric assemblies have been presented. The current study and detailed analyses of eight independent simulations

for a total combined time of 2.4  $\mu$ s per peptide system highlights several findings that in particular address the course of primary events in peptide assembly. First, a rapid formation of a heterogeneous ensemble of  $\beta$ -sheet rich oligomer structures was observed, where kinetically trapped aggregate intermediates affected the structural evolution of larger assemblies. Second, oligomerization was found to proceed via a combination of polymerization and condensation mechanisms. Finally, it was revealed that the observed diverse association and reorganization dynamics are governed by the characteristics of peptide sequence and oligomer size.

A thorough characterization of the heterogeneity in molecular dynamics and structures of the low molecular weight oligomers may hold the key to understand the profound differences in macroscopic fibril growth kinetics [52, 273–275] and the observed rich structural diversity of aggregate states [80, 102–104, 248, 269]. Specifically, one may speculate that the structural relations between crystalline and fibrillar polymorphs of amyloidgenic peptides may be only resolved by tracing and determining the characteristics of the oligomeric conformational states from which either of the species originate.





## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

### 4.2.1 Introduction

The deposition and accumulation of stable, filamentous aggregates of a specific protein or peptide in a variety of tissues are known to be associated with a number of human pathologies [59]. These aggregates are generally termed amyloid fibrils [62] and formed by many natural polypeptides [18, 33]. In addition, truncated parts of such protein sequences [88, 100] and de novo designed peptides [21, 90] were shown to assemble into amyloid-like fibrils in vitro. Thus, it has been proposed that the intermolecular interactions of the invariant polypeptide backbone result in the amyloid structure as an alternative generic conformational state with the kinetics of the process being dependent on the side chains and external factors, such as pH, temperature or ionic strength [11, 16–18, 21, 276]. Regardless of the sequence or native fold, the commonly formed amyloid fibrils are defined as self-assembled, elongated and unbranched (fibrillar) polypeptide aggregates with a cross- $\beta$  conformation [32]. The cross- $\beta$  architecture, as revealed by X-ray fiber diffraction, is described as stacked  $\beta$ -strands that run perpendicular to the fiber axis with extensive hydrogen bonding along the length of the fiber [41]. Moreover, a growing number of crystal structures of short model peptide sequences revealed a common steric zipper motif [82, 100, 102, 104, 105]. The atomic structures of the crystalline conformers show pairs of elongated  $\beta$ -sheets with parallel or anti-parallel strand alignment. The opposing sheets are interdigitated such that a highly complementary packing of the side chains is achieved, yielding a tight and dry interface. Despite their fundamental similarity, the structures vary in their basic steric zipper motif, a feature that rationalizes the observed polymorphism of the self-propagating amyloid structures on a molecular basis [102]. It has been shown that crystalline and fibrillar amyloid polymorphs share structural characteristics such as the cross- $\beta$  diffraction pattern and therefore offer a plausible paradigm for the general spine organization of amyloid fibrils [18, 82, 104, 105]. Nevertheless, the degree of order in the crystal structure may not fully represent the one in the fibrillar form, as indicated by solid-state NMR experiments on various crystals and fibrils [104, 105]. Exposure of hydrophobic epitopes and regions of unstructured polypeptide backbone, such as found in partially folded or misfolded states, is among the accepted causes of amyloidogenic aggregation [4, 11, 19], which is the irreversible formation of the  $\beta$ -sheet-rich amyloid structures [4, 18, 21]. Consensus aggregation-prone sequence patterns

of amyloidogenic proteins have been identified, and they demonstrate that protein unfolding is necessary but not sufficient to promote aggregation [17, 91, 92]. In fact, experimental evidence is accumulating that short amyloidogenic signatures in natural protein sequences can facilitate self-assembly [92, 94, 95]. The multi-staged aggregation process is canonically described as the conversion of isolated peptide monomers in solution to soluble oligomeric assemblies, and the final, fibrillar aggregates via a nucleated growth process [52, 53]. It is likely that even more intermediate states have to be considered, and that the conformational transitions between all of them are associated with different barrier heights [277–279]. In order to obtain and understand the full picture, detailed knowledge of the molecular structures of the involved species is indispensable. Despite the considerable progress in characterizing the fibrillar end-states, it is still difficult to obtain the biochemical and precise structural information for the oligomeric species in experiments. Heterogeneous oligomeric aggregates of different sizes are usually observed during the incubation of amyloidogenic peptide solutions [79] and discussed as either on- [252, 280] or off-pathway [67, 281] intermediates to the fibrils. Pre-fibrillar and fibrillar oligomers, as well as annular protofibrils have been described among a variety of morphologies [63, 66, 67, 72]. However, thorough investigations are hindered by either the transient or polymorphic, and non-crystalline behavior of the oligomers [66, 76, 278]. Critical observations regarding the structural properties of oligomeric aggregates have been derived recently from experiments on short amyloidogenic peptides [90, 132, 270, 282, 283]. Although some key aspects are not entirely understood, several studies report on the general consent that: (a) Given their qualitatively different morphologies from the characteristic appearance in transmission electron and atomic force microscopic images, oligomeric precursor states and amyloid fibrils are surprisingly similar in molecular conformation and supramolecular structure [29, 67, 80]. (b) Specific binding to antibodies alludes to common structural features shared by oligomers from different amyloidogenic proteins [72, 284]. (c) The oligomeric aggregates assume  $\beta$ -sheet-rich conformations [77, 279]. (d) Soluble amyloid intermediates are established as the primary pathogenic agents in several types of neurological amyloid diseases [4, 63, 66, 67, 70, 72]. In addition, computational studies have facilitated the current understanding of molecular determinants and events in the early stages of amyloidogenic peptide aggregation. Molecular dynamics (MD) simulations appear to be particularly suited to probe the formation of oligomeric species in atomistic detail, unraveling the transition pathways on timescales not amenable to experiments. The properties of small multimeric aggregates (dimers to decamers) of various amyloidogenic peptide sequences have been studied by atomistic simulations and described as partially ordered, nematic structures, which are subject to rapid fluctuations and large conformational rearrange-

#### 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

---

ments [106, 108, 115, 120, 124, 129, 131–133, 270, 285]. The obtained oligomer ensembles are described as distinct from the monomeric form [108, 115, 130] due to the conformational changes associated with emerging  $\beta$ -sheet structure [124, 129]. These structural transitions were found to be accompanied by a loss of intra-peptide interactions and conformational entropy [108, 120]. The desolvation of nonpolar surface and formation of inter-peptide backbone hydrogen bonds was found concomitantly [106, 108, 286, 287]. Dynamical reorganization via sliding, reptation or flipping of individual strands as opposed to repeated dissociation and annealing has been observed in simulations [106, 115, 122, 285] and has found experimental validation [55]. Furthermore, the initial stages of assembly are reported to be likely under kinetic control, and a multiplicity of association and interconversion pathways gives rise to polymorphic aggregate structures [123, 127, 285, 287]. In the context of the clearly very complex underlying free-energy landscape, the prominent and crucial role of water in the aggregation process has been highlighted [130, 270, 286–289].

Here, we perform unbiased, atomistic simulations of steric zipper peptide oligomerization in explicit solvent as model systems for amyloidogenic aggregation. The studied peptides are short segments identified from the fibril-forming proteins tau (<sup>306</sup>VQIVYK<sup>311</sup>, referred to as PHF6) [82, 95], insulin (<sup>12</sup>VEALYL<sup>17</sup>, referred to as IB12) [82, 88] and  $\alpha$ -synuclein (<sup>51</sup>GVATVA<sup>56</sup>, referred to as AS51) [82]. They have been found to be essential in fibril formation of the full-length proteins and adopt  $\beta$ -strand conformations in the fibrillar aggregate structures, respectively [33, 82, 95, 101, 255]. Moreover, the hexa-peptides PHF6, IB12 and AS51 were shown to spontaneously polymerize in solution, yielding amyloid-like fibrils, microcrystals or both [82]. Fibrillization assays of these peptides monitored with, for example, ThT-fluorescence measurements, show a common lag-time and imply a typical nucleation-dependent growth process [82, 255, 290, 291]. The aim of this work is to facilitate the understanding of the early events in spontaneous peptide oligomerization on a molecular level. To directly assess the multifaceted conformational ensemble of oligomers and the sampled structural transitions, a novel mapping technique is presented. Key structural elements of the spontaneously assembled oligomeric states are compared with the known X-ray crystallography structures of various amyloidogenic peptides, thereby contributing to the structural and dynamical characterization of low-molecular weight peptide oligomers. Furthermore, addressed are the following questions: What drives the initially monomeric peptides to form oligomeric assemblies? What determines their subsequent structural ordering and stabilizes the formed aggregates? The comparison of aggregation pathways and structures of different peptide sequences and concentrations allows for insight into possible common mechanistic steps in amyloidogenic peptide aggregation.

## 4.2.2 Methods

### Simulation setup and procedure.

**MD Simulations.** All MD simulations were carried out using the GROMACS software package (version 4.0) [177]. The Berendsen coupling algorithm [187] was applied to keep the pressure constant by coupling the system to a pressure bath of 1 bar ( $\tau = 1$  ps). Velocity rescale [188] was applied for temperature coupling to a temperature bath of 310 K. Initial velocities were sampled from a Maxwellian distribution at 310 K. All protein bonds were constrained with the P-Lincs algorithm [182]. All the hydrogens were replaced by virtual interaction sites, and therefore, all internal vibrational degrees of freedom of the hydrogen atoms were removed [175, 179]. An integration time step of 5 fs was used. Neighbor lists for nonbonded interactions were updated every 5 steps. For production runs the GROMOS96 43A1 [159] force field and the SPC water model [184] were used. Water molecules were constrained using SETTLE [234]. The short-ranged van der Waals and electrostatic interactions were cutoff at 1.4 nm and 0.9 nm, respectively. All simulations were carried out using periodic boundary conditions (PBC) and the Particle Mesh Ewald (PME) [230, 231] method. The electrostatic interactions with PME were calculated at every step with a grid spacing of 0.12 nm. The relative tolerance at the cut-off was set to  $10^{-6}$ .

**Simulation Setup and Procedure.** An overview of the simulated peptide systems, simulation lengths and sampling intervals (subscript) is given in Table 4.4. The simulations were labeled according to the name (sequence) of the peptide and the starting configuration. The total simulation time of all trajectories adds up to 42  $\mu s$  for this study.

The individual simulations (M10) with 10 initially monomeric PHF6 (VQIVYK), IB12 (VEALYL), AS51 (GVATVA) and mutant peptides (GVATVA, GVAVTA, GVTTVA, GVATFA, GAATAA) were set up according to the protocol described in Ref. [285]: 1000 distinct peptide conformations were pre-generated with CONCOORD [256]. From this structure ensemble, 10 peptide structures were chosen at random and in random position and orientation in a cubic box ( $1000 \text{ nm}^3$ ) to result in a concentration of 16.6 mM if not stated otherwise. This procedure was applied to ensure a fully monomeric configuration and to reduce the bias of similar initial starting structures for each of the conducted simulation runs. Additional M10 simulations with different peptide concentrations (3.3, 8.3 and 83 mM) were realized by adjusting the box size ( $5000, 2000, 200 \text{ nm}^3$ ).

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

Table 4.4: **Summary of performed simulations and initial conformations.**

Sequence	System ID	Starting peptide configuration	#Trj	Simulation time ( $\mu s$ ) and sampling interval ( $ps$ , subscript)
VQIVYK (PHF6)	M10	random, monomeric configuration	8	2.50 <sub>50</sub> , 2.15 <sub>50</sub> , 2.01 <sub>50</sub> , 2.00 <sub>50</sub> , 1.75 <sub>400</sub> , 0.65 <sub>50</sub> , 0.33 <sub>50</sub> , 0.31 <sub>50</sub>
	SH10	$\beta$ -sheet, parallel strands	2	0.70 <sub>50</sub> , 0.50 <sub>50</sub>
	SH5-5	$\beta$ -sandwich, parallel strands	1	1.00 <sub>50</sub>
VEALYL (IB12)	M10	random, monomeric configuration	8	1.92 <sub>50</sub> , 1.80 <sub>500</sub> , 1.80 <sub>500</sub> , 1.48 <sub>50</sub> , 0.36 <sub>50</sub> , 0.32 <sub>50</sub> , 0.30 <sub>50</sub> , 0.30 <sub>50</sub>
	M10 (3.3 mM)		1	0.17 <sub>50</sub>
	M10 (8.3 mM)		2	0.39 <sub>50</sub> , 0.37 <sub>50</sub>
	M10 (83 mM)		3	1.00 <sub>50</sub> , 0.50 <sub>50</sub> , 0.50 <sub>50</sub>
	SH10	$\beta$ -sheet, anti-parallel strands	2	1.00 <sub>50</sub> , 0.75 <sub>50</sub>
	SH5-5	$\beta$ -sandwich, anti-parallel strands	1	1.03 <sub>50</sub>
GVATVA (AS51)	M10	random, monomeric configuration	4	2.44 <sub>500</sub> , 2.00 <sub>500</sub> , 1.87 <sub>500</sub> , 1.77 <sub>50</sub>
GVATAV	M10	random, monomeric configuration	1	1.25 <sub>50</sub>
GVAVTA	M10	monomeric configuration	1	1.25 <sub>50</sub>
GVTTVA	M10		1	1.25 <sub>50</sub>
GVATFA	M10		1	1.25 <sub>50</sub>
GAATAA	M10		1	1.25 <sub>50</sub>

For the initial configuration of the reference simulations, the atomic coordinates of the crystal structures PDB ID: 2ON9 (PHF6: VQIVYK) and PDB ID: 2OMQ (IB12: VEALYL) [82] were used to model the ordered steric zipper aggregate conformations (SH5-5: two 5-stranded  $\beta$ -sheets in a sandwich and SH10: single, 10-stranded  $\beta$ -sheet). The peptide concentration and box size matched the ones of the M10 aggregation setup. There were no crystalline coordinates available for the GVATVA peptide [82]. Subsequently, all systems were solvated with explicit water molecules. The protonation state of the peptides was according to the one in solution at a pH of 7 for all simulations of the PHF6, as well as the AS51 peptides. For the IB12 peptide simulations and according to the crystallization conditions at pH 2.5, the C-terminus and glutamate side chains were assumed to be protonated [82, 88]. Counter-ions ( $Na^+$ ,  $Cl^-$ ) were added to yield an appropriate ionic strength (150 mM) and to neutralize the net system charge. The simulation systems comprised roughly 100000 atoms. After the system preparation an energy minimization using steepest descent was performed.

## Analysis

From the individual simulation trajectories, samples were taken for analysis every 50, 400 or 500 ps (Table 4.4). After pooling all independent simulations, the resulting total number of collected configurations was for PHF6: M10, 203351; SH5-5 and SH10, 45699; for IB12: M10, 100851; SH5-5 and SH10, 55604; M10 (3.3 mM), 3401; M10 (8.3 mM), 15202; M10 (83 mM), 40003 and for AS51: M10, 47960.

**Mapping the aggregation dynamics by projections onto collective coordinates.** The used mapping procedure enables us to describe an initially high dimensional problem in a reduced (low dimensional) collective coordinate space. The re-expression in terms of collective coordinates (principal components) preserves the trends and variance of the full-dimensional data as well as highlights similarities and differences. In this formulation one is able to map the multimeric aggregates found along the aggregation pathway and discriminate structures with different properties. In addition, the projection onto collective coordinates allows for a direct visualization of common sampling routes and probabilities.

**Calculation of observables.** A total of 25 different observables were calculated from the Cartesian coordinates of the configurations sampled from the individual trajectories. The observables encompass diverse metrics that adequately capture the topological (i), structural (ii) and energetic (iii) properties of amyloidogenic  $\beta$ -aggregation. Some of them are established descriptors for simulated peptide assembly and were used in previous studies [131, 133]. A detailed description is provided below.

**(i) Topological properties:** To describe the peptide assembly state at any given time and to ensure that all assemblies with different multichain topologies and configuration types are discriminated, a specific notation was used. The aggregation state was expressed in terms of a Connectivity Length ( $cl$ ) [133, 292]. The  $cl$  was defined to be the sum over the square roots of the determined individual aggregate sizes, with  $k$  as the total number of aggregates and  $n$  as the number of peptides in aggregate  $x$ . A normalized connectivity length ( $CL$ ) was used since this can be read more intuitively, with  $N=10$  for all simulations, which is also the largest possible aggregate size.

$$cl = \sum_{x=1}^k \sqrt{n_x} \quad CL = \left(1 - \frac{cl - \sqrt{N}}{N - \sqrt{N}}\right)$$

The CL values for the aggregation configuration types range from 0 for fully monomeric, to 1 for the state where all peptides were found in the same aggregate. A hierarchical categorization of the peptide aggregate sizes was pursued to monitor the assembly state of the peptides. General peptide assemblies or aggregates were defined as follows: pairwise contact analysis is used to identify the individual aggregates. Peptides that share an inter-chain residue contact are then counted to be within the same aggregate. For any two residues  $i, j$  an inter-chain contact is considered formed if any heavy atom of residue  $i$  is within a cutoff of 0.45 nm from any heavy atom of residue  $j$ . Assemblies of peptides aggregated into intermolecular  $\beta$ -sheets were defined as follows: two peptides that share two consecutive inter-chain  $\beta$ -sheet contacts as defined by the DSSP definition [202] are counted within the same aggregate. The identified  $\beta$ -sheet aggregates are a subpopulation of the above considered general aggregates. These two metrics comprise the first two observables that are considered. **(1)**: Assembly state of general aggregates according to van der Waals contacts (CLg); **(2)**: Assembly state of  $\beta$ -sheet aggregates according to  $\beta$ -sheet contacts (CLs). Instructive descriptors for different orientational properties of the system are the polar (P1) and the nematic (P2) order parameters, commonly used to analyze properties of anisotropic fluids such as liquid crystals. Several studies highlight the useful and complementary information of these order parameters in the context of peptide aggregation [124, 131]. A suitable molecular vector ( $\hat{\mathbf{z}}_i$ ) was defined, here the unit vector linking the  $C_\alpha$ -atoms of residues 2 and 5 of each peptide. The choice was based on the observation that the terminal residues were not found to have high  $\beta$ -sheet propensities. The polarity of the system is described by P1, distinguishing between parallel, anti-parallel or mixed strand (molecular vector) orientations.

$$\overline{P_1} = \frac{1}{N} \sum_{i=1}^N \frac{3}{2} \hat{\mathbf{z}}_i \cdot \hat{\mathbf{d}} \quad \overline{P_2} = \frac{1}{N} \sum_{i=1}^N \frac{3}{2} (\hat{\mathbf{z}}_i \cdot \hat{\mathbf{d}})^2 - \frac{1}{2}$$

The orientational or nematic order parameter P2 of the system discriminates between ordered and disordered peptide conformations: the extent of alignment and relative orientation of the individual peptides described by their molecular vector. The order parameter P1 and P2 were calculated using the wordom program package [257]. **(3)**: Polar order (P1); **(4)**: Nematic order (P2).

**(ii) Structural properties:** A heavy-atom contact analysis was used to calculate the total number of inter-strand contacts present between either polar or nonpolar peptide atoms. An atom was considered nonpolar (or hydrophobic) if the absolute value of the partial charge was below 0.2. Otherwise this atom was defined as polar. A contact was considered formed



if any heavy atom was within a cut-off of 0.45 nm from any other heavy atom of the same type (polar/polar) of any other strand. **(5)**: Total number of polar atom contacts (Npa); **(6)**: Total number of nonpolar atom contacts (Nna). **(7)**: Solvent-inaccessible molecular surface (SiMS). The total SiMS area was calculated using a solvent probe radius of 1.0 nm. Similar to the above described atom contact analysis, the sum of all peptide heavy atoms of either polar or nonpolar type in contact with solvent molecules was calculated. The cut-off criterion was chosen to be 0.35 nm, reflecting the approximate contact distance of atoms on the peptide surface to the water molecules of the first solvation shell. **(8)**: Total number of polar atom contacts to the solvent (Npas), that is, number of water molecules in contact with polar peptide atoms; **(9)**: Total number of nonpolar atom contacts to the solvent (Nnas). For an assessment of secondary structure type and content, the DSSP definition introduced by Kabsch and Sander [202] was used. The populations of various secondary structure elements were calculated (random coil,  $\beta$ -sheet,  $\beta$ -bridge, bend, turn) for the individual simulations as a function of simulation time. **(10)**: Random-coil content (Cc); **(11)**:  $\beta$ -sheet content (Bsc); **(12)**:  $\beta$ -bridge content (Bbc); **(13)**: Bend content (Bc); **(14)**: Turn content (Tc). **(15)**: Average peptide chain extension (Ext). The peptide extension was calculated by averaging the cumulative intra-peptide distances between the N- and C-terminal  $C_\alpha$ -atoms of each of the 10 chains.

**(iii) Energetic properties:** The sum of the potential energy terms of different groups of atoms was calculated to identify their contributions to the assembly process. The short- and long-range van der Waals, as well as short-range electrostatic interaction energies of the force field were considered separately. In total, ten potential energy terms were analyzed. **(16)**: Sum of coulombic inter- main chain interactions (Inter-mc-Ecoul); **(17)**: Sum of van der Waals inter-main chain interactions (Inter-mc-EvdW); **(18)**: Sum of coulombic inter-side chain interactions (Inter-sc-Ecoul); **(19)**: Sum of van der Waals inter-side chain interactions (Inter-sc-EvdW); **(20)**: Sum of coulombic intra-main chain interactions (Intra-mc-Ecoul); **(21)**: Sum of van der Waals intra-main chain interactions (Intra-mc-EvdW); **(22)**: Sum of coulombic main chain interactions with the solvent (Sol-mc-Ecoul); **(23)**: Sum of van der Waals main chain interactions with the solvent (Sol-mc-EvdW); **(24)**: Sum of coulombic side chain interactions with the solvent (Sol-sc-Ecoul); **(25)**: Sum of van der Waals side chain interactions with the solvent (Sol-sc-EvdW).

**Dimensionality reduction and projection of the observable data.** The obtained observable data sets were concatenated (see below for details) and normalized



## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

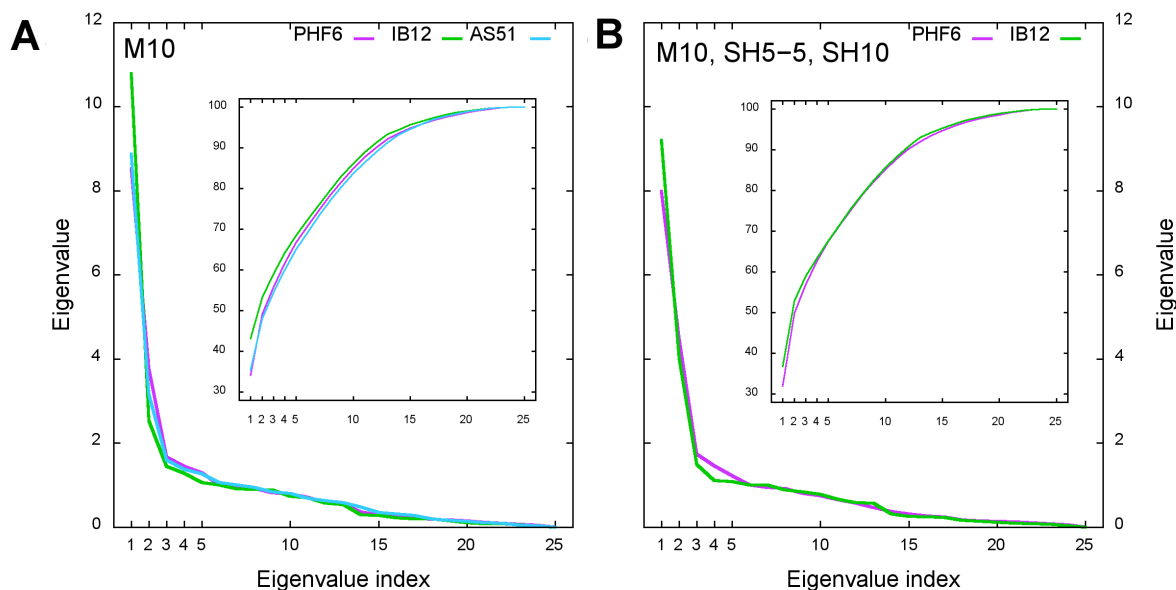


Figure 4.14: **PCA eigenvalue spectra of the collective coordinate mapping.** The eigenvalue spectra for the different observable data sets on which the PCA was carried out: starting from initially monomeric peptide conformations (A) (see Fig. 4.17 and Fig. 4.21); considering also the simulations with modeled reference structures (B) (Fig. 4.24). The insets illustrate the cumulative variance explained by the eigenvectors.

(z-scored) prior to further analysis. Each set of a given observable ( $x$ ) was normalized separately, such that the mean ( $\mu$ ) was set to 0 and the standard deviation ( $\sigma$ ) to unity.

$$z = \frac{x - \mu}{\sigma}$$

This preprocessing step was carried out to circumvent issues when comparing observables with different units and dimensions, which might otherwise affect the outcome of the subsequent principal component analysis (PCA). The PCA involved the diagonalization of the covariance matrix of the data set composed of the 25 described observables. The dimensionality reduction of the observable data was done by projecting the sets onto its respective principal components (the first three eigenvectors of the covariance matrix). PCA was performed for the 25-dimensional observable data sets of the PHF6, IB12 and AS51 peptide simulations separately if not stated otherwise, using the statistical software package R [259]. The subspace of the first three eigenvectors typically accounts for more than 55% of the variance (see Fig. 4.14).

### **Estimating the extent of overlap between mapped ensembles and identification of conformations from the overlap region.**

To examine whether the M10 and SH5-5, as well as SH10 simulations access the same configurational states as described by the various observables, a step-wise approach was chosen. Firstly, the extent of overlap in the multi-dimensional observable space was determined. Secondly, the characteristics of the overlapping states were analyzed.

**Identification of the overlap region.** The observable data sets obtained from M10, SH5-5 and SH10 simulations were concatenated for the PHF6 and IB12 systems, respectively. On the concatenated data sets, PCA was performed and subsequently a projection onto the first three eigenvectors of the covariance matrix was carried out.

Determining the extent of overlap between the point clouds of spontaneous (M10) and reference (SH5-5 and SH10) simulations in the multi-dimensional observable configuration space was approached as a classification problem, whose solution also allowed the direct identification of the structures from the ensemble overlap regions. For each point in the observable data set, k-nearest neighbors were identified employing the k-NN algorithm [293,294] implemented in R [259]. Hereby the number of points for the classification are chosen such that  $N$  (number of neighbors) equals the square root of the number of samples in the data set ( $N_{\text{PHF6}}$ : 499 and  $N_{\text{IB12}}$ : 395). Each point was assigned a label of that ensemble (spontaneous or reference), the more members of which there were among its k-nearest neighbors. All those points of the data set that were assigned to the ensemble different from which they originally belonged to, were considered to be in the overlap region of the two ensemble of points. For every point identified in this way, the closest point from the other ensemble not yet in the list of structures from the overlap region was added to the list. With this approach, 8194 nearest-neighbor conformations from the spontaneous and reference point clouds of PHF6 simulations and 10626 for IB12 were found, respectively.

**PCA and clustering of conformations from the overlap region.** The identified data points from the overlapping regions of the ensembles were mapped back to their cartesian coordinates. Since all the points correspond to decameric oligomers, a relabeling procedure described in [285] was performed in order to reduce permutation redundancy of the aggregates. The redundancy arises due to permutation of the labels (chain identifiers) in structures with equivalent conformations. The reference structure for relabeling was chosen to

be the one closest to the center of the overlapping region in the observable space. In every step of the relabeling procedure  $10!$  label permutations were tested. For each permutation a fitting of the main chain and  $C\beta$  atoms onto the reference structure was performed. The labeling that yielded the smallest RMSD to the reference structure was assigned to the aggregate.

Subsequently, a conformational clustering was carried out. PCA was performed using the Cartesian coordinates of the relabeled and superimposed structures, after a least-squares fit onto the main chain and  $C\beta$  atoms. Projections onto the first 100 principal components were used for the k-means clustering, following the results of Ding and He [295] showing that principal components are the solutions for the k-means clustering problem. K-means clustering was performed using the Hartigan-Wong algorithm [258] as implemented in R [259]. The Krzhanowski-Lai criterion [261] was used to choose the number of conformational clusters and the global k-means algorithm [260] was used to determine the respective cluster centers.

**FMA analysis with partial least squares.** *Calculation of sampling density as external variable for the aggregation progress.* For the density estimate in the full 25-dimensional observable configuration space, again the k-nearest neighbors (k-NN) method was used. For each data point the distance to k-nearest-neighbors was calculated [296]. Distances for all k-NN were summed up and inverted to yield a measure for the density for each data point. Densities for the PHF6, IB12 and AS51 M10 simulations were calculated from concatenated and commonly z-scored observable data set. The number of nearest neighbors to be considered for the k-NN distance calculations was selected to be 593, which was the square root of the total number of data points in the concatenated set of PHF6, IB12 and AS51 configurations.

*FMA model building and cross-validation.* In order to elucidate the main forces driving the aggregation process, the functional mode analysis (FMA) [297] using the partial least squares (PLS) algorithm [298, 299] was employed. FMA is a technique to construct a linear multiple regression model, which maximizes the correlation between the observables and an external variable, while at the same time maximizing the variances in both, the variable and the observables. Here, 28 independent observables grouped into descriptors (1) and energetic properties (2) were used to predict the sampling density in the configuration space. In addition to the 25 described observables, the solvation free energy as defined by Eisenberg and McLachlan [300] (dGSol) and the sum of coulombic, as well as van der Waals main chain to side chain interactions (Inter-mc-sc-Ecoul and Inter-mc-sc-EvdW) were used.

For each peptide system, PHF6, IB12 and AS51, a separate FMA model was built for the association and the decameric phase. In order to probe the formation of ordered aggregates in the decameric phase, an additional model was built where the sampling density was weighted by the factor  $\exp(\text{P2-Bsc})$  prior to the FMA.

A cross-validation procedure was applied to estimate the predictive power of each FMA model. Every data set was divided into eight equal parts. One part at a time was left aside for the cross-validation, while an FMA model was built on the rest of the data. This process was repeated such that each part of the data set was used for the cross-validation once. The predicted FMA models were concatenated yielding a completely cross-validated data set. The final FMA vectors, as well as the training sets in Fig. 4.27, were built from the full data sets, respectively. To estimate the quality of the built FMA models, the correlation coefficients for both, the training and testing parts, were calculated for the FMA models based on the descriptor observables. In order to avoid overfitting, the number of components for the FMA model construction was selected such that with a further increase in the number of components (8), only minor changes in the correlation coefficient for the testing data set were observed.

*Regression coefficients of the ensemble-weighted FMA model.* The influence of the observables on the changes in sampling density was analyzed using an ensemble-weighted FMA [297] model, where the contribution of each observable parameter was weighted according to its variance in the data set at hand. This approach allowed to investigate which of the observed aggregate features positively or negatively contributed to the sampling density in the observable configuration space, that is, the peptide aggregation.

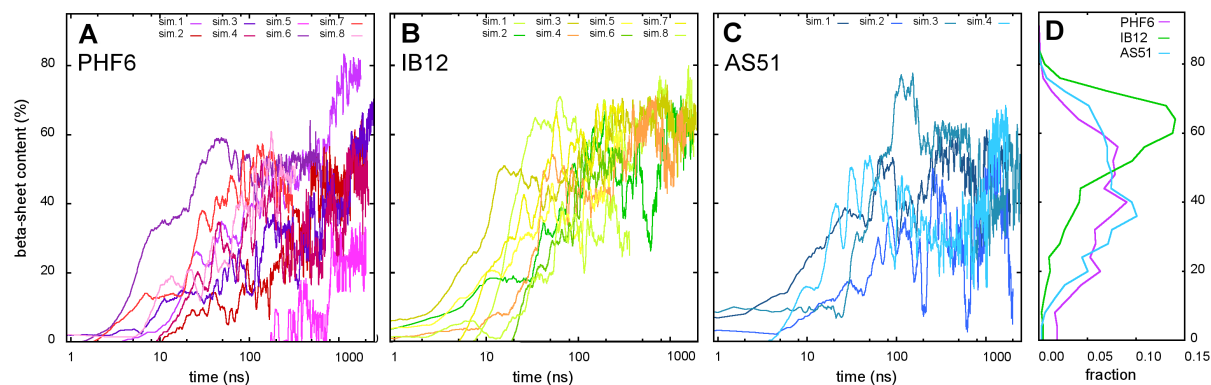
### 4.2.3 Results

#### **Spontaneous aggregation of steric zipper peptides results in $\beta$ -sheet rich decamers.**

Multiple unbiased MD simulations for the PHF6, IB12 and AS51 peptide systems were initiated from randomized and fully dispersed monomeric conformations (M10, see Table 4.4). Visual inspection of the PHF6, IB12 and AS51 simulation trajectories suggested common characteristics in the aggregation process. Regardless of sequence, a successive association of all 10 peptide chains as well as a gradual increase in  $\beta$ -sheet structure was found. While the onset of  $\beta$ -sheet formation was fast, usually within 10 ns, fluctuations in the content of  $\beta$ -sheet conformations of the decameric peptide aggregates were observed on the microsecond timescale (Fig. 4.15).

In contrast to the PHF6 and AS51 simulations, a near-monotonical increase of  $\beta$ -sheet

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.



**Figure 4.15: Time dependence and probability of  $\beta$ -sheet structure content.** The change in  $\beta$ -sheet conformation with simulation time is shown on a log scale for the PHF6 (A), IB12 (B) and AS51 (C) trajectories. The normalized abundance of a certain  $\beta$ -sheet content is given as histogram (D). The colors match the shades of the individual time traces for PHF6 (purple), IB12 (green), AS51 (blue).

conformation was found for all IB12 trajectories (Fig. 4.15B). The most abundant IB12 peptide aggregate conformers were found to have a  $\beta$ -sheet content of around 65%, as seen from a combined histogram of all simulations (Fig. 4.15D). The overall fraction of such  $\beta$ -sheet-rich IB12 structures was more than four times higher compared to PHF6 and three times as high as for the AS51 aggregates (they sampled mostly a content of 40-50%). The set of independent PHF6 simulations showed large variations with respect to the level of  $\beta$ -sheet content (Fig. 4.15A). Notably, a significant proportion of PHF6 peptides were found with a low or without any  $\beta$ -sheet content in several simulations. Multiple AS51 simulations displayed a substantial, repeated loss and recovery of  $\beta$ -sheet structure, resulting in a high standard deviation for the  $\beta$ -sheet content (Fig. 4.15C).

The PHF6, IB12 and AS51 peptide chains associated in general to fairly ordered structures, as the increase in  $\beta$ -sheet structure with simulation time indicates. Nevertheless, as shown below, the 10 chains populated a multitude of different aggregate configuration types and heterogeneous topologies throughout the simulations, respectively.

**Mapping peptide aggregates onto collective coordinates yields immediate insight into common association pathways and diversity of structures.** It is not straightforward to characterize the process of peptide oligomerization comprehensively using just one observable (e.g., the  $\beta$ -sheet content). In fact, multiple metrics are necessary to discriminate the aggregate conformation ensembles in a meaningful way or to examine specific structural properties. Although every chosen observable might provide its own information

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

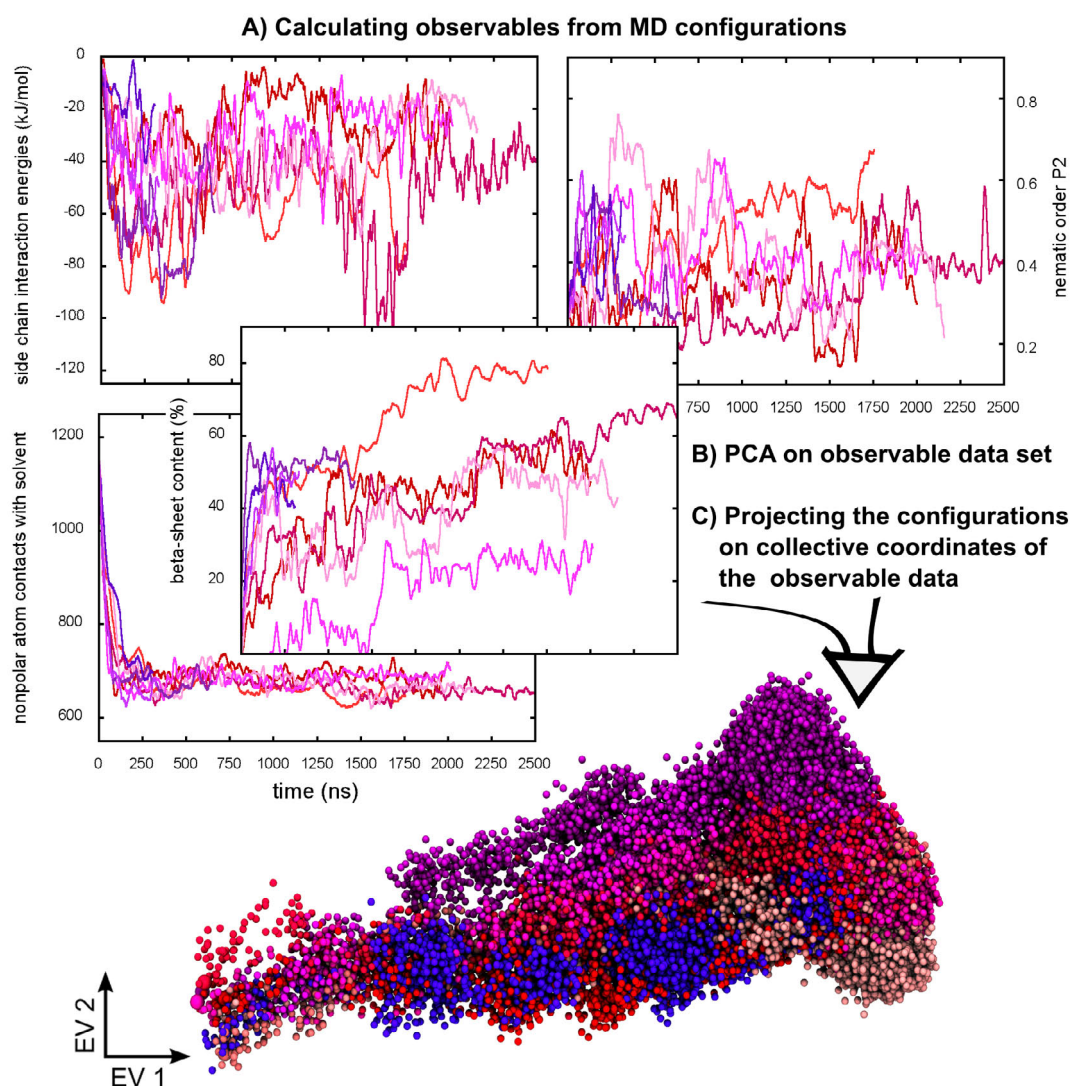


Figure 4.16: **Schematic of the collective coordinate mapping procedure.** First, from multiple independent simulations a set of observables (topological, structural and energetic descriptors) is calculated. Subsequently, a matrix is built, where each row represents the different observable values for a given MD conformation. A PCA on this multi-dimensional data set yields collective coordinates (first three eigenvectors of the covariance matrix with the largest eigenvalues) onto which the original data set can be projected. The resulting low-dimensional representation of the full data set allows a direct and intuitive mapping of the conformational ensembles, that is, the aggregation configuration space.

content, it is often not possible to compare them simultaneously. Yet another complication is the need to define or select the most suitable observable in the first place.

Here, we present a collective coordinate approach to describe the oligomeric structures and



their transitions sampled in the different aggregation simulations (see Fig. 4.16 and the Methods section for a detailed description). In order to probe amyloidogenic  $\beta$ -aggregation adequately, 25 observables were chosen as topological, structural and energetic descriptors of the sampled configurations.

Through the use of this set of measures instead of the Cartesian coordinates it is possible to apply a dimensionality reduction step using principal component analysis (PCA). The first three eigenvectors of the covariance matrix constructed from the observable data then represents a newly identified basis for the subsequent analysis. Thereafter, all the configurations from the simulations were projected onto these collective coordinates to obtain a low-dimensional representation of the sampled phase space and hence to identify underlying collective trends.

The projections of the combined simulation data for the PHF6, IB12 and AS51 peptide systems are shown in Fig. 4.17, where each sphere represents a simulation configuration snapshot. The mapping procedure discriminates structures with different features, therefore allowing the direct assessment of the multimeric aggregates found along the aggregation pathway. From each projection map the common structural and energetic properties can be read directly together with the associated oligomer structures and configurations. A visual inspection of the representative structures in the various regions of the projections indicates that for all three peptide systems, the first collective coordinate (EV1) maps the conformational conversion from monomers to oligomeric aggregates (i.e. the general association state). Starting from initially dispersed peptides ('S' in Fig. 4.17), a rapid and concerted change in several observables was observed as the ten peptide chains began to collide and aggregate with one another. The corresponding sampling along EV1 of the projections in the collective coordinate space as a function of simulation time is shown in Fig. 4.18, indicating a convergence within 100 ns. The inter-peptide atom contacts, as well as the molecular surface burial increased, while the number of solvent molecules in contact with the peptides decreased. Seen from an energetic point of view, the first collective coordinate coincides with favorable coulombic inter-peptide main chain interactions, whereas those with solvent get less favorable. The same holds for the van der Waals interactions of the side chains between the peptides (increase) and interactions with water (decrease). The change in the other observables was explained to a smaller extent by the first collective coordinate.

The commonly observed general association for the PHF6, IB12 and AS51 peptides resulted in decameric, fully assembled oligomers as the most abundant general aggregate configuration type in all of the individual trajectories. The formed decamers were found to be stable assemblies as dissociation events were negligible (AS51) or not observed (PHF6, IB12). Note

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

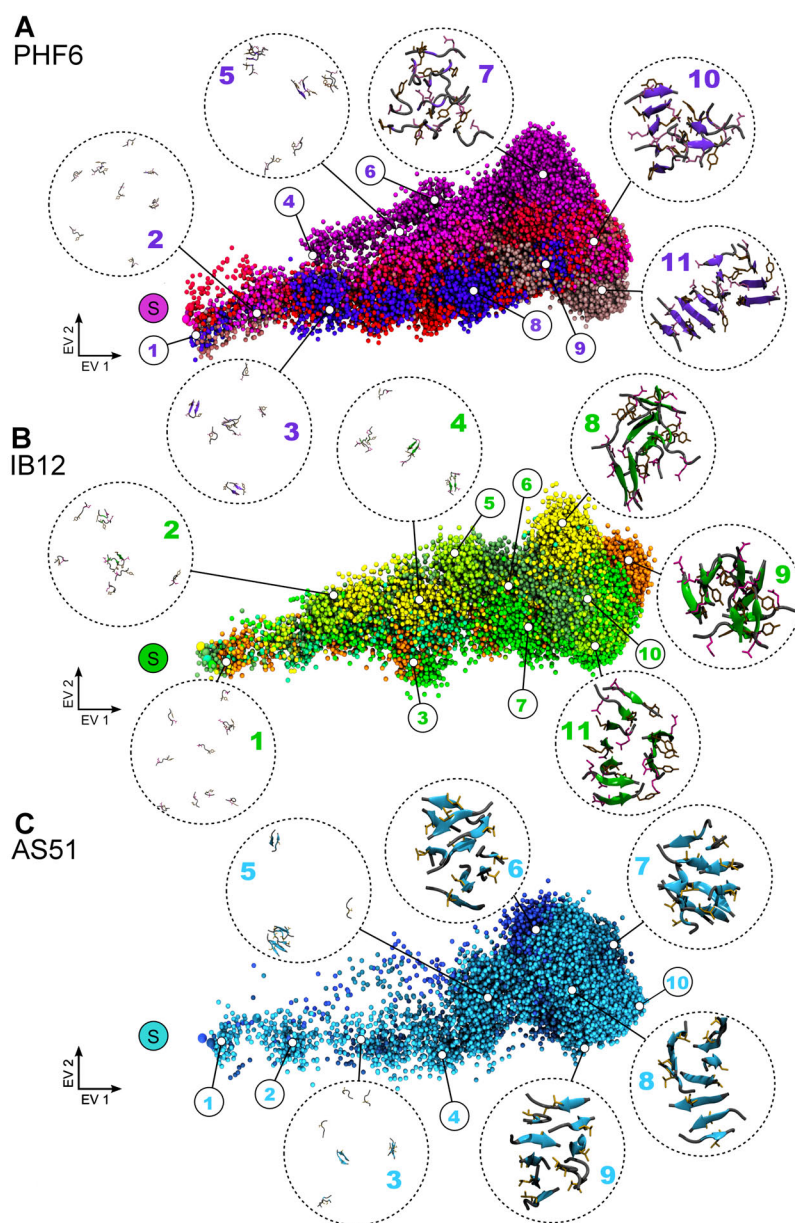


Figure 4.17: **Projection of conformational ensembles and representative structures.** Shown for PHF6 (A), IB12 (B) and AS51 (C), are all the sampled conformations projected in a three-dimensional collective coordinate space. Each sphere in the projection represents one trajectory configuration mapped into this space. The color shades encode the different independent simulations for each of the peptide sequences.

that due to the finite system size, decameric oligomers could not grow further in the present setup. En route to the decamer, intermediate aggregate sizes ranging from dimers to nonamers



## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

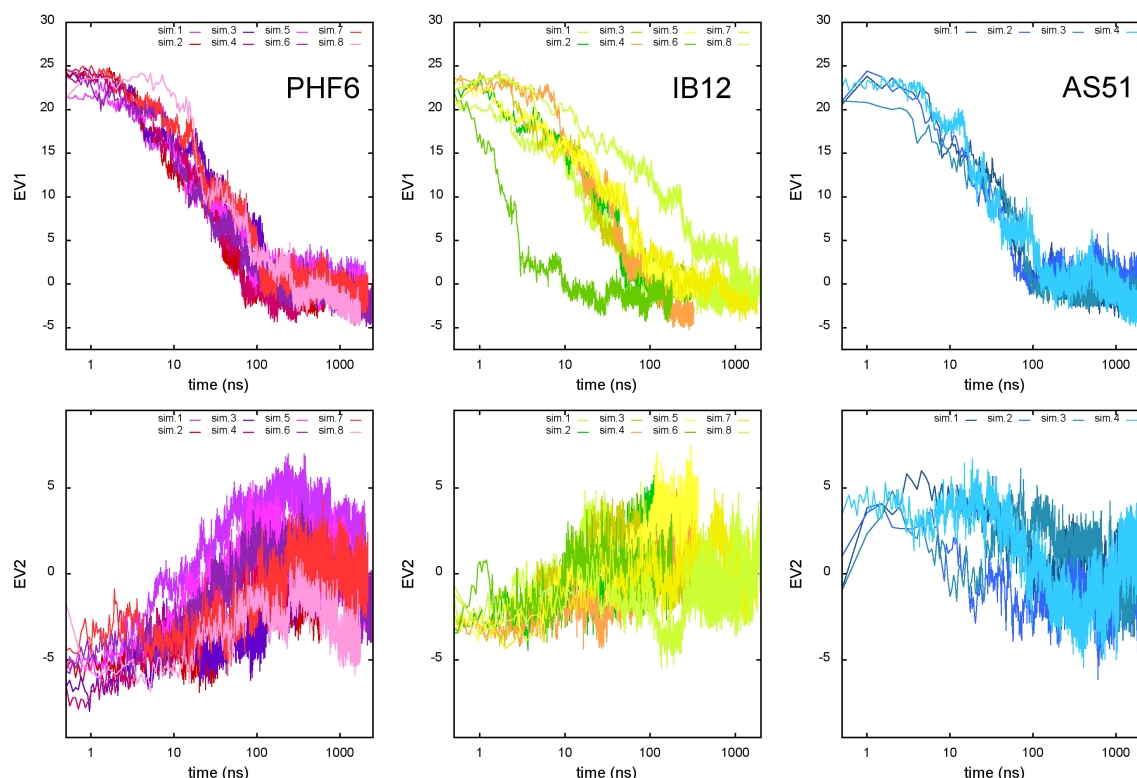


Figure 4.18: **Time evolution in the collective coordinate space along EV1 and EV2.** The displacement along EV1 and EV2 in the projections of the collective coordinate space is shown as a function of simulation time (log scale) for the PHF6, IB12 and AS51 M10 trajectories. The colors match the shades of the individual time traces for PHF6 (purple), IB12 (green), AS51 (blue).

were transiently formed, as previously reported [285]. The successive assembly of early and intermediate configurations proceeded via monomer addition, as well as condensation of primarily dimeric and trimeric precursor states [285]. Overall, the same association tendency was found for the three different peptide systems. The simulations suggest that assembly toward the final oligomeric state occurred the fastest for AS51. Similar to PHF6, in AS51 simulations, all the aggregate sizes have been sampled, however the AS51 aggregates of intermediate size ( $n=3-6$ ), had on average a shorter lifetime, that is, were consumed more rapidly into larger oligomers. In addition, AS51 simulations did explore only a small portion of all the different possible association types. The temporal evolution of the aggregate sizes can be appreciated from Fig. 4.20, where the average size distributions of PHF6, IB12 and AS51 are shown for two time windows (0-300 ns, 300 ns-end of sim.).

The sequence of events in the oligomerization process can be understood by looking at the

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

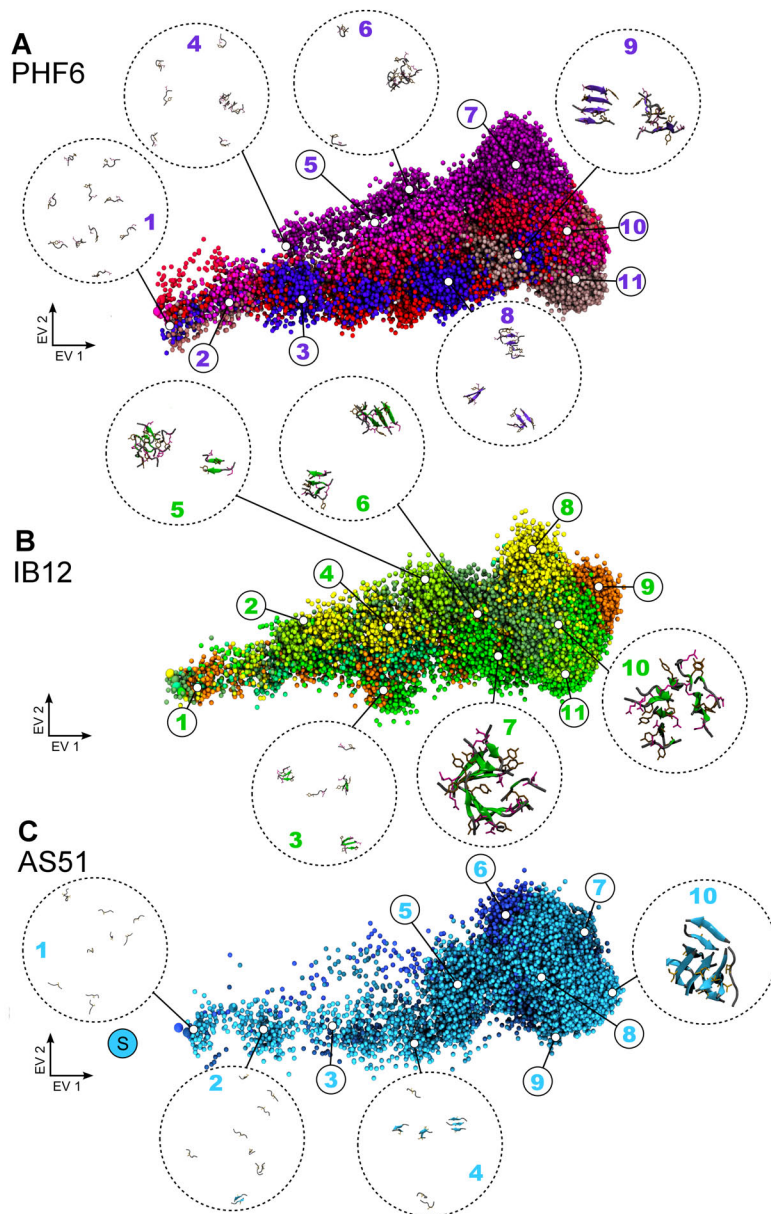


Figure 4.19: **Representative structures of projected conformational ensembles.** The chosen representative structures are shown in cartoon representation and enclosed by circles. The different colors and index numbers encode the location of the respective structures in the same way as in Fig. 4.17. A subset of residue side chains is shown in stick representation, namely: PHF6 (Gln, Tyr), IB12 (Glu, Tyr) and AS51 (Thr).

representative structures and conformations mapped close to them in Fig. 4.17.

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

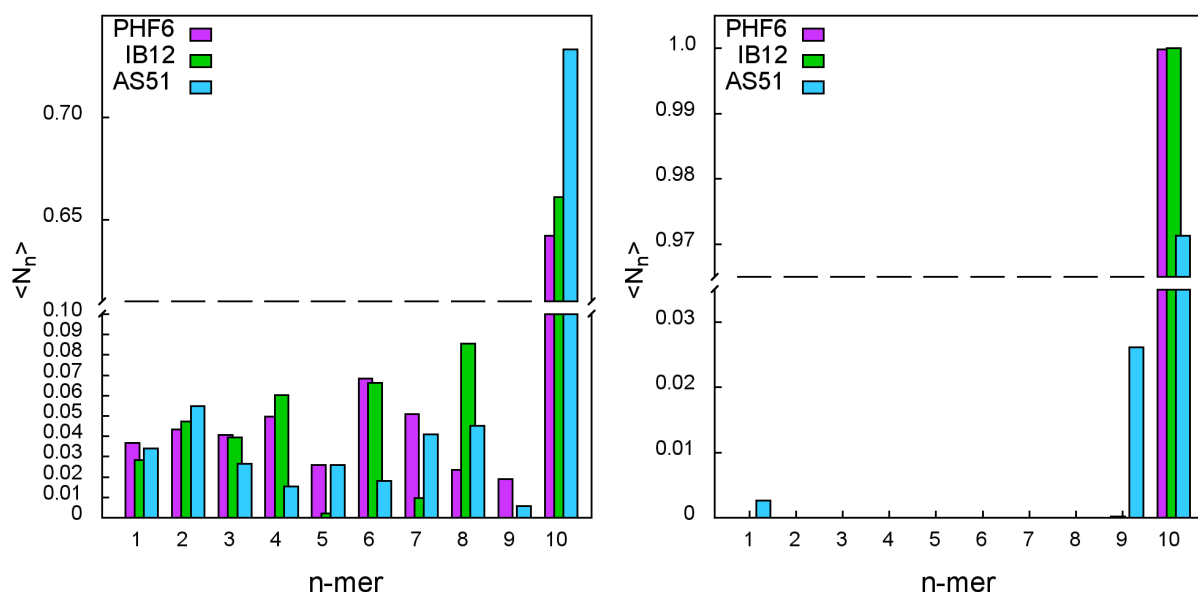


Figure 4.20: **Aggregate size distributions.** The normalized aggregate size probability distributions are shown for the PHF6, IB12 and AS51 simulations. Analysis was carried out for two time windows: 0-300 ns (left); 300 ns to the respective end of the simulations (right).

Multiple conformations for oligomers of intermediate size were present, such as two- to four-stranded  $\beta$ -sheets in extended and untwisted conformations (Fig. 4.17A, structures 8 and 9; Fig. 4.17B, 2,3 and 6; Fig. 4.17C, 4 and 5), while others appear bent and collapsed to compact structures (Fig. 4.17A, 4-6; Fig. 4.17B, 1,2 and 5; Fig. 4.17C, 3). No apparent systematic pattern in strand alignment and registry was found for the diverse mixture of extended and collapsed structures of the different peptide systems.

The individual sampling routes and distribution of states along the first two collective coordinates in the projections of the independent PHF6, IB12 and AS51 simulations can be directly inferred from Fig. 4.17 and Fig. 4.18. For example, the PHF6 simulations explored common regions in the projections, but also populated distinct configurations as seen for the purple colored snapshots, which represent mostly collapsed aggregates without significant  $\beta$ -sheet content. These particular conformations diverged very early from the common sampling routes, suggesting an alternative aggregation pathway for PHF6. In multiple simulations the stepwise assembly of IB12 oligomers was found to proceed as a single  $\beta$ -sheet up to the pentamer, while larger aggregate sizes exhibited lateral growth at the already established sheet surface and eventually  $\beta$ -sandwich structure formation (Fig. 4.17B, 5, 6 and 11).

As outlined before, the assembly of stable, decameric oligomers was observed in all simula-

tions, however, the  $\beta$ -sheet aggregates sampled by each sequence, as well as in the independent trajectories for each peptide system, were different in topology and abundance. PHF6 and AS51 aggregates were found with very similar types of  $\beta$ -sheet configurations within the fully assembled, decameric state. Especially, two- and three-stranded  $\beta$ -sheets in conjunction with a significant amount of disordered chains (PHF6 > AS51) occurred with a high probability. The most common sheet topologies were two  $\beta$ -sheet dimers and six disordered chains (here written in the following notation: [2x2+6x1]), as well as [3x2+4x1] and [3+2+5x1]. In contrast, the IB12 decamers were found frequently to be composed of larger, four- and five-stranded  $\beta$ -sheets, and configuration types such as [5+4+1] and [4+3+2+1]. Interestingly, the total number of observed  $\beta$ -sheet configuration types were roughly the same for all the peptide systems (PHF6: 24; IB12: 24; AS51: 26). The wealth of distinct conformational basins in the decameric state for PHF6, IB12 and AS51 ranged from amorphous to ordered  $\beta$ -sheet assemblies and can be readily identified within each projection (Fig. 4.17A-C). Specifically, the second collective coordinate (EV2) describes the variance in  $\beta$ -sheet content (small or large number of  $\beta$ -sheets formed) and  $\beta$ -sheet aggregate configuration types (small or large, intact sheets). Most of the decameric oligomers were found to be either  $\beta$ -sandwich-like structures (Fig. 4.17A, 10; Fig. 4.17C, 9) or single, larger  $\beta$ -sheets facing smaller sheets (Fig. 4.17A, 11; Fig. 4.17B, 11) with multiple edge strands exposed to the solvent. In particular for IB12 and AS51 oligomers, oval  $\beta$ -sheet as well as bent barrel-like structures (Fig. 4.17B, 10; Fig. 4.17C, 7,8 and 10) and orthogonal sheets were frequently sampled (Fig. 4.17B, 9). Aggregates with multiple smaller sheets positioned on top of each other (Fig. 4.17B, 8; Fig. 4.17C, 6), globular structures and amorphous assemblies with no specific packing order (Fig. 4.17A, 7) were sampled as well.

In addition to the supramolecular organization of the individual aggregates, the position and extent of regular packing of the side chains was a prominent feature observed for the different oligomer structures. A coarse, but apparent classification of the side chain packing distribution in the PHF6 and IB12 oligomers could be derived by mapping the side chain solvent accessibility for all the aggregate configurations. From this a preferential packing of the bulky Tyr residues to the interior (Fig. 4.17A, 7; Fig. 4.17B, 9), as well as the accumulation of Tyr residue side chains on the outside of the oligomeric aggregates (Fig. 4.17A, 9 and 10; Fig. 4.17B, 7,10 and 11) was seen. Specifically, for all the decameric PHF6 conformations located in the upper and rightmost area of the projection in Fig. 4.17A (purple spheres, around structure 7), most of the Tyr residues were found to be strongly desolvated, whereas the Gln residues were uniformly oriented to the solvent. In the projection of the IB12 aggregates a similar region could be identified

(Fig. 4.17b; orange spheres, around structure 9), where the interior of the oligomers was found to be occupied mostly by Tyr residues, whereas the protonated Glu side chains were exposed on the aggregate surface. Interestingly, IB12 aggregate configurations that cluster around structure 11 of the projection in Fig. 4.17B (green spheres) showed the exact opposite solvent exposure characteristics. The respective oligomer conformations were in all cases stabilized by a transient hydrogen bonding network between the hydrophilic side chains.

#### **Sampling density identifies prominently visited aggregate conformations.**

In order to investigate the prominently visited structures in each of the aggregate configuration ensembles, the sampling density in the full-dimensional observable space was determined using a k-nearest neighbor (k-NN) approach (see Methods). Despite the varying sampling routes in the independent simulations, the highest densities lie without exception in the ordered region of the decamers, although the highest density appears more localized for the IB12 and AS51 peptide systems (Fig. 4.21). In the densest sampled regions of the IB12 and AS51 configuration space, oligomers with an established sheet to sheet interface were identified. For AS51 an almost closed, flat  $\beta$ -barrel-like structure was found, while IB12 oligomers displayed a buckled  $\beta$ -sandwich aggregate architecture (Fig. 4.21). The ensemble of PHF6 oligomers extracted from the highest density region was associated with a lower  $\beta$ -sheet content, compared to the IB12 and AS51 aggregates. Interestingly, the rather distorted and less compact arrangement of smaller and twisted  $\beta$ -sheets featured numerous inter-peptide Gln side chain to main chain contacts (Fig. 4.21A).

**Concentration dependence of the oligomerization process.** To address the concentration dependence of the oligomer growth process, additional M10 simulations of the IB12 peptide system in a range of concentrations above (83 mM) and below (3.3 mM, 8.3 mM) the initially simulated 16.6 mM have been carried out (see Table 4.4). As before, the results are presented in the form of a collective coordinate mapping (Fig. 4.22). The projections clearly show that the systems with lower IB12 monomer concentration follow the same sampling routes toward the oligomeric states of higher order, as discussed above (16.6 mM, M10 simulations).

The lag-time for the formation of decameric aggregates was found to be prolonged significantly due to the reduced probability of diffusional encounter of the peptide aggregates in comparison to the higher concentrations. No apparent concentration dependence on the stability of the formed aggregates was observed. In contrast, the highly saturated simulation systems

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

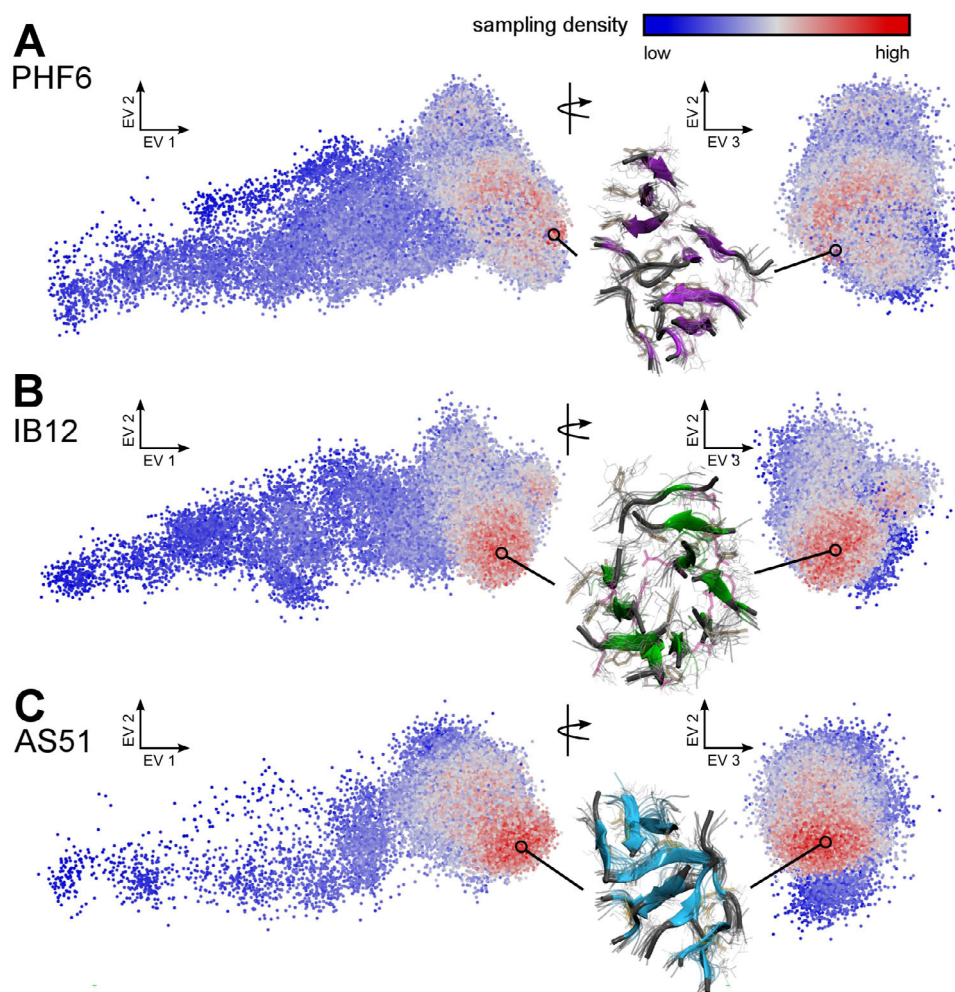


Figure 4.21: **Projection of conformational ensembles and their sampling density.** The sampled conformations in the PHF6 (A), IB12 (B) and AS51 (C) trajectories are projected in the same collective coordinates space, as before where each configuration is represented by a point. The projection is viewed from two different angles, and the sampling density for all simulations is indicated by a color gradient (red - high; blue - low density). A superposition of the structure ensemble (10 oligomer conformations) is shown for each peptide system, corresponding to the respective region of highest sampling density.

(83 mM) showed a prominent excursion via initially isotropic and amorphous aggregates caused by the near instantaneous collapse of the peptides, resulting from their small initial separation. For these disordered oligomer aggregates a subsequent conformational transition toward  $\beta$ -strand structure was observed (Fig. 4.22).

Overall, the assembly pathways taken toward the decameric state were found to depend on the IB12 monomer concentration as sketched by the gray arrows in Fig. 4.22. However, the



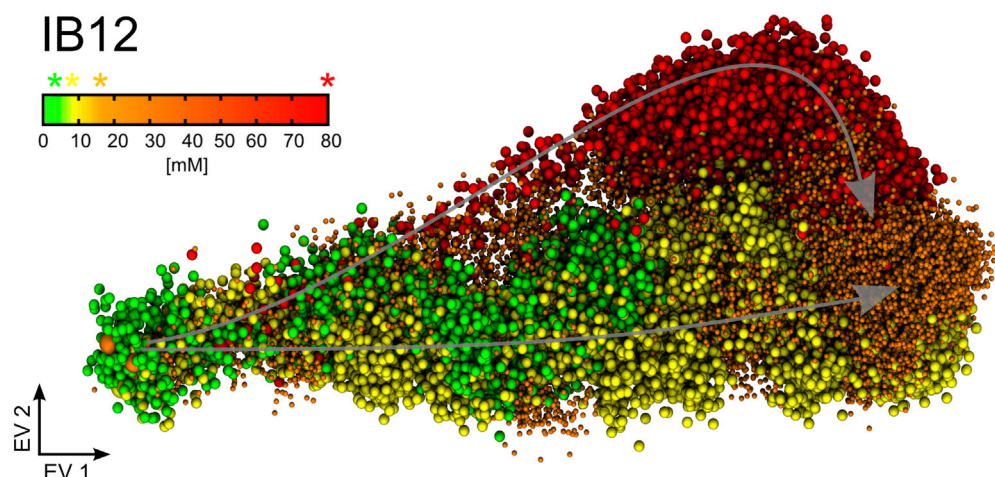


Figure 4.22: **Projection of IB12 conformational ensembles simulated at different concentrations.** Shown are the IB12 conformations sampled at 3.3, 8.3, 16.6 and 83 mM initial concentration and projected in a three-dimensional collective coordinate space. Each sphere in the projection represents one trajectory configuration mapped into this space. All conformations from simulations with the same concentration are colored according to the shown scale. For comparison the IB12 aggregate conformations from Fig. 4.17 (16.6 mM) are shown as smaller orange spheres. The gray arrows indicate sampling along the two prominent aggregation pathways observed: via disordered intermediates (upper arrow; red spheres) or ordered intermediates (lower arrow; green, yellow and orange spheres).

final structures of each of the simulations approached ordered aggregate conformations with consensus steric zipper-like structural features, irrespective of the initial concentration.

**Spontaneously formed aggregates and steric zipper oligomers converge to similar structures.** Some of the observed oligomers have topological resemblance to a cross- $\beta$  sheet motif, the basic structural element of peptides found in the fibrillar [18, 105] and crystalline states [82, 104], which is characterized by an tightly packed pair of sheets, an ordered arrangement of strands and interdigitated side chains. To assess if and to what extent the spontaneously formed oligomers relate or even converge to aggregate structures with such an ideal sheet packing motif, both in terms of structural and energetic properties, additional simulations were set up for the PHF6 and IB12 peptide systems (see Methods). The two starting configurations were modeled based on the available crystal structure coordinates: a  $\beta$ -sheet composed of 10 strands (SH10) with extensive backbone hydrogen bonding and a

planar  $\beta$ -sandwich of two facing  $\beta$ -sheets with 5 strands each (SH5-5), making up a steric zipper interface. Compared to the single  $\beta$ -sheet, the SH5-5 configuration featured four instead of two edge strands and less solvent-exposed side chain surface due to the different packing. Note that the strands were organized in parallel within the PHF6 aggregates, whereas the strands were aligned in anti-parallel fashion in the case of IB12. The trajectories that used the crystallographic conformations as initial structures will be referred to as reference simulations throughout. The configuration ensemble sampled in the reference simulations was investigated based on the same observables as before and evaluated together with the data set from the spontaneously formed oligomers. The low dimensional projections of the individual structures on the collective coordinates of the concatenated ensemble (spontaneous and reference) are shown in Fig. 4.24A and Fig. 4.24B. The collective coordinates (and their components) did not change significantly compared to Fig. 4.17, such that the shape of the projected spontaneous structure configurations of PHF6 and IB12 was only slightly altered. The starting structures ('R' in Fig. 4.24A and Fig. 4.24B) and most of the sampled configurations of the reference simulations appeared in regions not accessed by the M10 simulations. As expected, the reference configurations initially clustered around the regions of the projection corresponding to decameric oligomers with very high  $\beta$ -sheet content as well as ordered, intact and large sheets.

In the collective coordinate representation it can be seen that the PHF6 and IB12 reference simulations (SH10 and SH5-5) explored the space mostly along the second and third (perpendicular to the paper plane) PCA vectors toward the configurations spontaneously formed in the M10 simulations (Fig. 4.24A and Fig. 4.24B). The PHF6 reference structures lost their initially high  $\beta$ -sheet content and diverged quickly from aggregate topologies with large sheets and extended peptide chains. Furthermore, directly readable from the collective coordinates is a gradual reduction in nonpolar PHF6 atom contacts with the solvent. In comparison, through examination of the sampling of the IB12 reference structures in the collective coordinate map, in particular a loss in favorable coulombic interactions between the peptide main chains is found, while van der Waals interactions among the peptide main chain and side chain atoms are increased (also partially described by a higher number of nonpolar atom contacts).

The projections in Fig. 4.24 and the shown representative structures clearly illustrate that the two sets of reference structures relaxed differently in the explicit solvent environment as well as when comparing the different independent PHF6 and IB12 reference simulations. In particular, the single 10-stranded PHF6 and IB12  $\beta$ -sheets evolved toward different supramolecular organizations as also seen from the time evolution of the radius of gyration ( $R_g$ ) in Fig. 4.23.



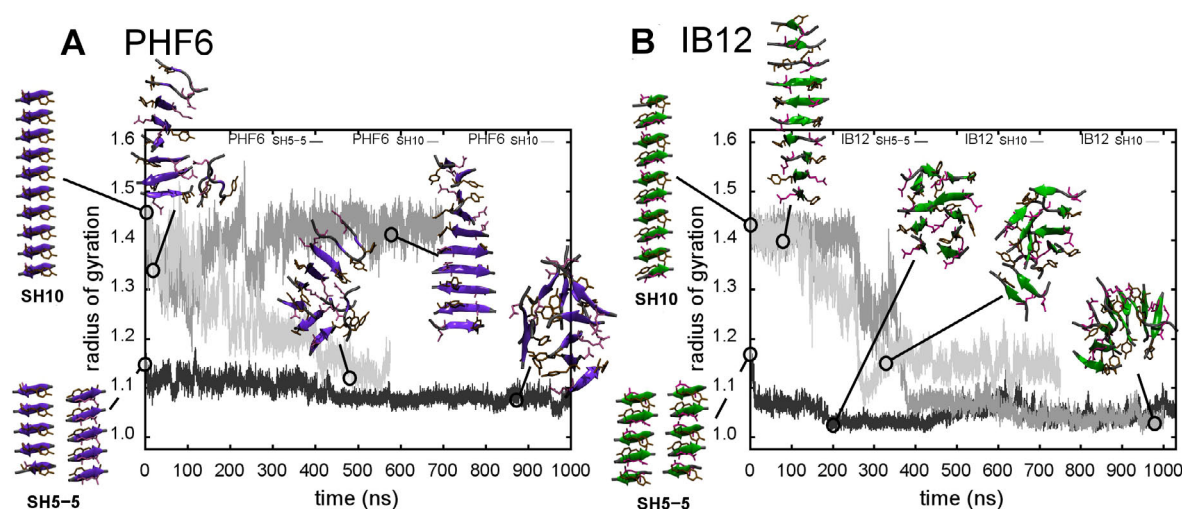


Figure 4.23: **Time evolution of the radius of gyration for the single and double layer sheet oligomers.** Time evolution of the radius of gyration is shown for the differently modeled reference oligomers: single  $\beta$ -sheet (SH10) and  $\beta$ -sandwich (SH5-5). Various representative structures are shown and illustrate the observed conformational changes with simulation time. The peptide backbone is rendered as cartoon and shown in purple (PHF6) and green color (IB12), respectively. A subset of side chains is depicted in stick representation: PHF6 (Gln, Tyr) and IB12 (Glu, Tyr). Representative structures are shown for the different initial conditions.

The IB12 SH10 state reproducibly showed a prominent twisting and compactation on a time scale of 100-300 ns (Fig. 4.24B, 1-3). Interestingly, both IB12 SH10 simulations showed a break up roughly in the middle of the single  $\beta$ -sheet, after which the oligomer reorganized to more compact conformations very similar to a  $\beta$ -sandwich, transiently sampling orthogonal packed  $\beta$ -sheets (Fig. 4.24B, 2). In contrast, the observed conformational rearrangements in the single PHF6  $\beta$ -sheet were mostly the result of strand reorientations and a overall  $\beta$ -sheet twisting (Fig. 4.24A, 1). The initially parallel PHF6 strands flipped mainly on the open edges but also within the  $\beta$ -sheet, after a partial break up and loss of  $\beta$ -sheet structure occurred. A stable  $\beta$ -sheet conformation was reformed in one PHF6 simulation, while in the other the elongated sheet structure was only partially recovered and stabilized by two adjacent chains (Fig. 4.24A, 1).

Only a small decrease in  $R_g$  was observed for the PHF6 and IB12 simulations starting from the SH5-5 configurations. Given the comparable, but minor extent of compactation seen for both PHF6 and IB12 SH5-5 reference simulations, the structural integrity of the respective oligomeric states differed substantially. The IB12  $\beta$ -sandwich oligomer twisted from the planar starting configuration and a partial loosening as well

as a migration and sliding of edge strands was observed. However, the tight and complementary  $\beta$ -sheet interface was essentially preserved on the microsecond time scale. In the case of PHF6, a complete disarray of the SH5-5 configuration to a more globular organization of the oligomer was found (Fig. 4.24A, 2 and 3), where smaller sheets are packed together and with the  $\beta$ -strands rearranged again toward a predominant anti-parallel alignment.

#### **Identification of commonly sampled conformations and extent of overlap.**

The projections in Fig. 4.24A and 4.24B suggest a number of commonly sampled conformational states in the spontaneous aggregation and in the reference simulations for PHF6 and IB12 oligomers, respectively. The simulations of the single  $\beta$ -sheets are shown in light gray and silver,  $\beta$ -sandwich simulations in dark gray spheres, respectively. The larger spheres marked with 'R' indicate the location of the initial reference configurations. Representative structures of the reference simulations are shown for each map as cartoon drawing. Here we examine whether conformations from the [10x1] simulations access the same configurational states as the [10] or [5+5] simulations and which features these aggregates share. The ensemble of similar structures from both pools of simulations was investigated by using a classification scheme, searching for k-nearest neighbors in the space of observable configurations (see Methods). The number of identified overlapping structures was found to be: 8194 oligomer structures for PHF6 (3.2% of all configurations) and 10626 in the case of IB12 (6.8%). To analyze the obtained configurations with high structural detail, a PCA based on the corresponding cartesian structure coordinates was performed (see Methods). Afterwards, similar conformations were grouped by applying the k-means clustering algorithm.

The results are shown in Fig. 4.24C and 4.24D. Here, commonly sampled oligomer conformations from spontaneous and reference simulations are shown in a projection onto the first two principal components obtained by PCA in cartesian coordinate space. The six conformational clusters identified for PHF6 oligomer structures are shown in different colors, containing mixed (1, cyan), and conformations from spontaneous (4), and reference (2, 3, 5, 6) simulations only (4.24C). In the same way, in (4.24D) the three clusters identified for IB12 oligomers, containing mixed (1, orange), and conformations from spontaneous (2), and reference (3) simulations only are shown (4.24D). The location of the center structures of the three largest clusters is indicated by a black dot. For each of these clusters, a superposition of the center structures and the nine closest oligomer conformations are shown in cartoon representation. As before, Gln, Glu and Tyr side chains are shown in stick representation, respectively.

For both, PHF6 and IB12 simulations, the largest identified conformational cluster (no. 1;

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

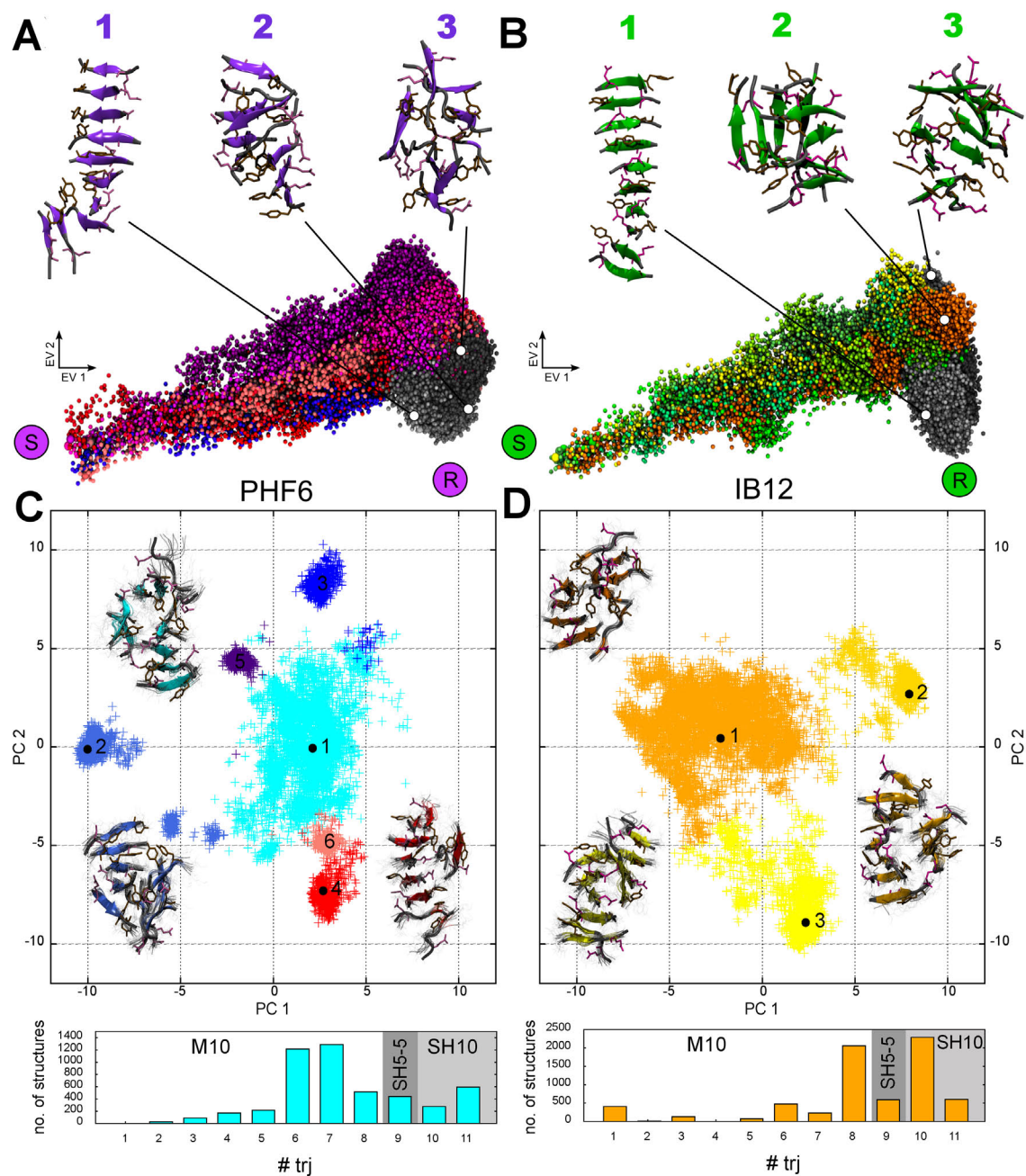


Figure 4.24: **Projection and clustering of spontaneously assembled and reference oligomer ensembles.**

PHF6: cyan and IB12: orange) contains decameric structures from the M10, SH5-5 and SH10 simulations. In each case more than one trajectory contributed significantly to the number of structures in the cluster, confirming structural overlap in cartesian space (Fig. 4.24). From the

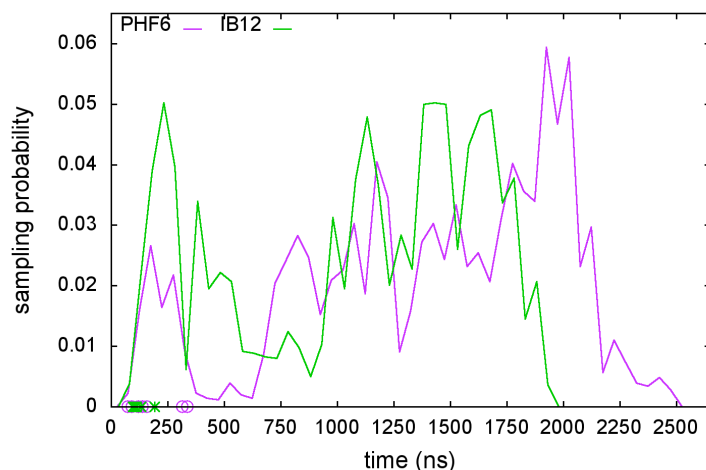


Figure 4.25: **Sampling of overlap oligomer conformations.**

The sampling probability of the mixed conformational cluster 1 in Fig. 4.24 for PHF6 (cyan) and IB12 (orange) is shown as a function of simulation time. The analysis was carried out over the pool of all 8 simulations with initially separated peptides and with a block average over 50 ns. The asterisk and open circle symbols indicate the first sampling in the different independent simulations (7 out of 8) for PHF6 and IB12, respectively.

initially monomeric state, 7 out of 8 PHF6 and IB12 trajectories sampled this overlap region (cluster no. 1). In Fig. 4.25 a histogram is shown, reporting the probability to find PHF6 and IB12 conformations in cluster 1 at a given time. Interestingly, transitions toward these particular conformations occurred rapidly, within 150 ns for PHF6 and within 120 ns for IB12.

The other clusters are identified as dense and distinctly separated states in the 2D PCA projection and comprise structures from either the spontaneous or the reference conformational ensembles. In comparison to IB12, more than one conformational cluster with only reference structures was found for PHF6. This can be explained by the larger structural diversity observed for the different PHF6 reference simulations, although in total fewer overlap conformations were identified than for IB12.

As one can see from the respective center structures shown in Fig. 4.24C and 4.24D, the main difference between clusters concerns the packing arrangement and size of the  $\beta$ -sheets. All identified PHF6 and IB12 oligomers were  $\beta$ -sheet-rich structures with extended peptide conformations. In detail, for cluster 1 (mixed) of PHF6, a fairly irregular and disordered packing was found, with small and twisted sheets positioned side by side and on top, while cluster 2 (reference structures only) showed an orthogonal  $\beta$ -sandwich structure (Fig. 4.24C). Finally, cluster 4 (spontaneous structures only) consists of a large twisted and bent sheet of eight strands with mixed alignment, facing two disordered peptide chains. The topology of the latter PHF6 structures is comparable to the respective spontaneously assembled IB12 oligomers (cluster 2), showing a similar elongated, twisted sheet with exposed edge strands and a smaller stabilizing sheet on the side (Fig. 4.24D). A preferential burial of Tyr and simultaneous exposure of Gln

and Glu residue side chains was found for the PHF6 and IB12 oligomers in cluster 2 (reference structures only) and cluster 1 (mixed). The contrary scenario, where the Tyr residues point to the solvent was found for the IB12 cluster 2 (spontaneous) and 3 (reference). These findings furthermore suggest that topologically similar oligomer populations can show quite different solvent accessibilities on the residue side chain level, similar to what has been reported in Fig. 4.17A and Fig. 4.17B.

In summary, it could be deduced that the highest sampling density of the M10 simulations lies near the overlap region (Fig. 4.24). Moreover, it was shown that spontaneous and reference simulations of PHF6 and IB12 peptides sampled an ensemble of structures with the common characteristics of an established sheet-to-sheet interface. Although spontaneous assembly toward these particular structures was fast, the formed oligomers as well as the modeled reference states reorganized considerably throughout the simulations, therefore complete convergence to highly ordered cross- $\beta$  structures with steric zipper was not reached.

**Elucidating the driving forces of the oligomerization process.** The collective coordinate mapping illustrates that the interactions of peptide and solvent molecules are among the most important driving forces underlying the oligomerization process as well as an important determinant for the conformational dynamics in the assembled state. In order to quantify the peptide-solvent interactions in the spontaneous aggregation of the PHF6, IB12 and AS51 peptides, the free-energy difference of transfer from the initially monomeric to the assembled oligomeric state was estimated. Similar to protein folding processes, where the solvation free-energies are lower for folded than for unfolded conformations [300], a prominent reduction is expected for the aggregation process. Indeed, the computed atomic solvation free-energies [300] decreased by more than half compared to the initial value for all the simulations starting from monomers.

The relation between the solvation free-energy and the density of sampling in the collective coordinate space shows an interesting nonlinear behavior (Fig. 4.26) for all three peptide systems. The different degree of correlation apparently arises from the two different aggregation phases, which in turn directly reflect the observed two principal stages of the oligomerization process: the assembly from monomers to decamers, where a high correlation of the solvation free-energy with the general assembly was found and the subsequent structural transitions within the aggregated state, which show no apparent correlation. This suggests that the solvation free-energy drives the peptides together in the initial aggregation phase, but does not play a prominent role in the further development of the oligomers. A functional mode analysis [297]



#### 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

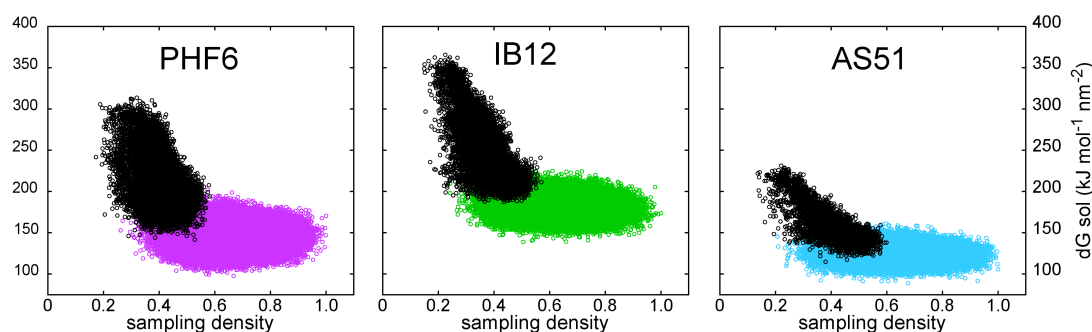


Figure 4.26: **Correlation of the solvation free-energy to the sampling density.** The correlation of the sampling density with solvation free-energy is shown for each peptide system. Notably, the sampling density in the initial association phase (black) and the subsequent structural ordering in the decameric state (colored) is correlated to a different extent.

(FMA, see Methods) was carried out to examine additional structural determinants and main forces driving the peptide aggregation. In general, the FMA attempts to correlate a suitable variable with a linear combination of a set of observables. Here, we select the sampling density in aggregate configuration space, i.e. the degree of aggregation, as an indicator for the aggregation progress. The idea of the FMA approach is then to construct a linear multiple regression model for the aggregation process based on the sampling density, as exemplified above in one dimension for the solvation free-energy.

For a model with adequate predictive power, this allows an assessment of the most prominent factors that govern the peptide oligomerization on a molecular level. Motivated by the observed difference in correlation to the solvation free-energy and corresponding to the fast association to stable decamers as the predominant species, as well as convergence along EV1 in the collective coordinate projections (Fig. 4.17 and 4.18), we chose to split the analysis. Thus, the peptide association events and the decameric phase were investigated separately, as they apparently follow two different mechanisms. It is important to note that, although the analysis is based on the collective trends in the data of multiple trajectories, they may not be fully converged. Therefore, extending the simulations could change the observed densities, especially in the decameric phase.

To distinguish mere descriptors for the aggregation process (group 1) from putative true driving forces (group 2) the PHF6, IB12 and AS51 FMA models were split into two separate groups of observables. In the second group, the pairwise additive force field energy terms are included, which model the principal interactions between the molecules (hydrogen bonds, hydrophobic effect, etc.).

Figure 4.27A-C depict the FMA models for the three peptide systems in the association

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

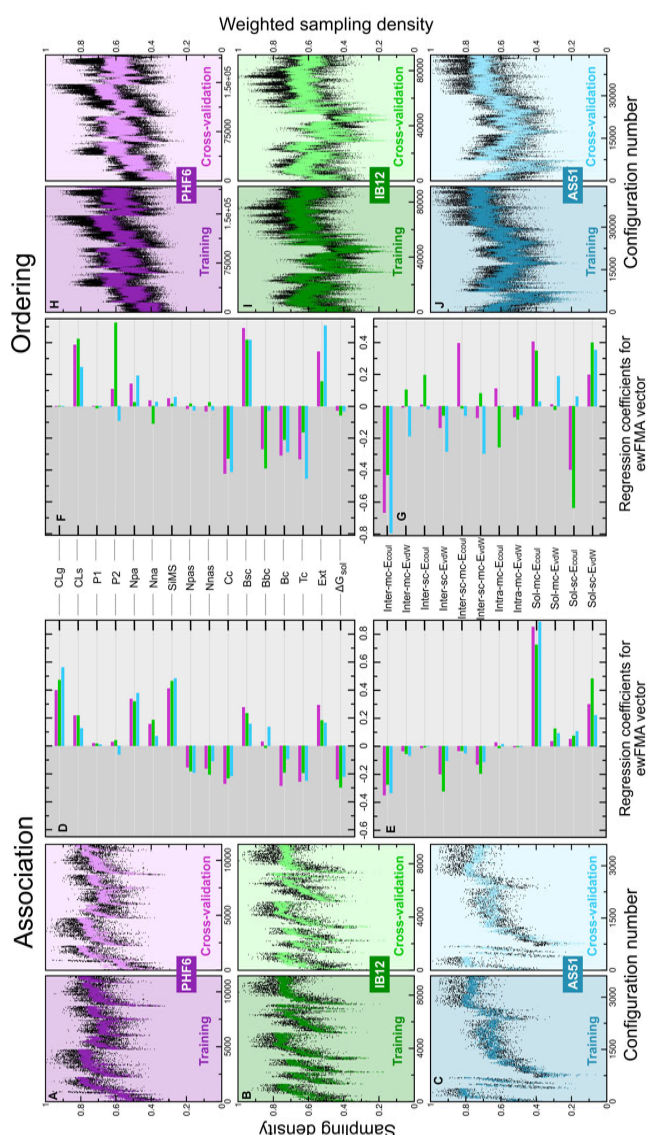


Figure 4.27: **Functional mode analysis elucidates the aggregation driving forces.** FMA models were built to predict the sampling densities in the aggregate configuration space of the (A; H) PHF6, (B; I) IB12 and (C; J) AS51 peptide systems, both for the peptide association (left) and the structural ordering in the decameric phase (right), respectively. The observables in terms of aggregate descriptors (D; F) and energetic components (E; G) were evaluated separately and had either positive or negative regression coefficients with the increase in the sampling density.

stage. Abbreviations and description of the observables are given in the Methods section.

The respective models were built on the full simulation data available, whereas in a second step an iterative procedure was applied to test and cross-validate them against a completely

independent set of data (see Methods). Here, the predictive power of the FMA models was assessed by cross-validation, using an independent data set not used for model training. The contribution of each observable parameter to the changes in sampling density was analyzed using an ensemble weighted FMA [297] model. The cross-validation correlation coefficients for the FMA descriptor models are shown in Table 4.5. For the association phase the diagonal elements in Table 4.5 indicate that the created models are capable of predicting the progress of peptide assembly for the system used to train the algorithm (correlation coefficients > 0.7). Interestingly, the off-diagonal correlation coefficients reveal that the models trained on the data of one peptide system were able to mutually predict the densities of the other peptide systems.

Table 4.5: **Correlation coefficients for the PHF6, IB12 and AS51 FMA models of association and ordering in the decameric phase.**

Association	PHF6		IB12		AS51	
	Train	Test	Train	Test	Train	Test
PHF6	0.74	0.74		0.79		0.76
IB12		0.71	0.81	0.82		0.77
AS51		0.60		0.77	0.88	0.87
Decamer (weighted)	PHF6		IB12		AS51	
	Train	Test	Train	Test	Train	Test
PHF6	0.74	0.74		0.74		0.76
IB12		0.69	0.81	0.81		0.76
AS51		0.70		0.66	0.81	0.81

Row-wise: Sampling density of the peptide system used for FMA model building.

Column-wise: Sampling density of the peptide system used for FMA model validation.

After validating the FMA models for the peptide association phase, the individual model components were investigated to probe the contributions of specific interactions to the aggregation mechanism. In the subsequent analysis, positive coefficient values indicate that an observable was positively correlated with the configurational space density of the peptide aggregates (i.e., the conversion from monomeric to decameric structures). For the energy terms, negative regression coefficient values indicate favorable interactions driving the assembly, that is, favorable changes in an energy term were negatively correlated with the density.

In the following, the main findings for the association process are summarized (Fig. 4.27D and 4.27E). For example the AS51 FMA model for the association phase correctly indicates the aforementioned slightly faster aggregation to higher-order oligomers, as can be seen from the



stronger influence of CLg (the assembly to general aggregates) compared to PHF6 and IB12. The oligomerization of all peptide systems is accompanied by a reduced solvent accessible surface area and the preferential burial of hydrophobic groups. This is consistent with the finding that water molecules were excluded from the peptide aggregate interior upon assembly from the initially fully solvated monomers and can be seen from the negative correlation for the number of polar and nonpolar peptide contacts to water molecules ( $N_{pas}$ ,  $N_{nas}$ ). While the exchange of solvent molecules with the peptide surface was on a picosecond timescale, the total extent of water release from the peptide hydration layer into the bulk was in general about 40% in the association phase and directly correlated with the decrease in the solvation free-energy. The importance of the interaction with the solvent molecules for the aggregation process is illustrated furthermore by the strong positive correlation between the sampling density and the formation of large, solvent-inaccessible interfaces between the peptides (SiMS). The secondary structure conversion from predominantly coil and turn to  $\beta$ -sheet conformations, as well as to more extended peptide chains was found to be correlated with the peptide association to a similar degree. Interestingly, neither the polar (P1), nor the nematic order (P2) of the peptide aggregates was a necessary components to predict the sampling density in the association phase, as they have relatively low correlation for either of the peptide systems.

All the energetic properties concerning the interactions of the peptides with the solvent have positive regression coefficients, i.e. the aggregation process is accompanied by an increasing energy between the peptides and water. Here, the coulombic interactions of the water molecules with the peptide main chain groups (amides and carbonyls) showed the most prominent loss. The van der Waals interactions between the water and peptide side chains were reduced significantly as well. A concurrent gain in inter-peptide interactions was found mainly from contributions of coulombic main chain and van der Waals side chain interactions. These two mostly account for the hydrogen bond formation and hydrophobic collapse as the peptides rapidly oligomerize in the early association phase. In addition, mixed inter-peptide main chain and side chain interactions were observed frequently as the peptides formed oligomeric aggregates with a substantial degree of nonspecific contacts and disorder. Although not included in the FMA model analysis because of the large-scale fluctuations, the increase in solvent-solvent interactions again stresses the important role of the solvent in the aggregation process.

The FMA model cross-validation and component analysis shows that common driving forces and structural determinants are at play in the initial association stage of all three peptide systems. This suggests a largely sequence independent mechanism based on the generic physico-chemical properties of the peptide molecules.

Table 4.6: **Correlation coefficients for the PHF6, IB12 and AS51 FMA models of the decameric phase.**

Decamer	PHF6		IB12		AS51	
	Train	Test	Train	Test	Train	Test
PHF6	0.49	0.49		0.43		0.76
IB12		0.44	0.69	0.69		0.71
AS51		0.47		0.40	0.78	0.77

Row-wise: Sampling density of the peptide system used for FMA model building.

Column-wise: Sampling density of the peptide system used for FMA model validation.

In comparison, the correlation coefficients of the FMA models for the decameric phase were in general substantially smaller, while the trends in the predictive power of the models are preserved ( $AS51 > IB12 > PHF6$ , see Table 4.6). In particular the PHF6 sampling density in the decamer appears to be the most difficult to predict for any of the three models. The rather poor predictive power of the PHF6 FMA model suggests that the sampled aggregate configurations and structural transitions are not described well by just a single model and this is probably due to the less pronounced gradient in the sampling density, as also seen from Fig. 4.21A.

In order to probe the formation of particular ordered aggregates in the decameric phase featuring cross- $\beta$  characteristics, an additional FMA model was built on the basis of a weighted sampling density. Oligomeric aggregates with high nematic order (P2) and  $\beta$ -sheet content, which are presumably similar to fibrillar  $\beta$ -sandwich aggregates, therefore received a higher weight prior to the FMA (weighting factor:  $\exp(P2 \cdot Bsc)$ ). From the newly obtained, weighted sampling density (Fig. 4.28) a striking result is the more localized density for PHF6 and IB12, while the AS51 sampling density remains virtually unchanged. This is due to the very small fraction of AS51 structures which fulfill both of the structural order criteria. Common for all the three peptide systems, the structural ordering in the initially collapsed decameric peptide assemblies was found to be positively correlated with the sheet configuration type of the aggregates (CLs,  $\beta$ -sheet aggregation to large sheets) (Fig. 4.27H-J). The common features of the most probable ordered oligomeric states were found to be the favorable interactions between coulombic inter-peptide main chain atoms (extensive hydrogen bonding), as well as van der Waals interactions from tight inter-peptide side chain packing. In detail, the ordering in the decameric phase was accompanied by an increase in  $\beta$ -sheet structure (Bsc) and peptide chain

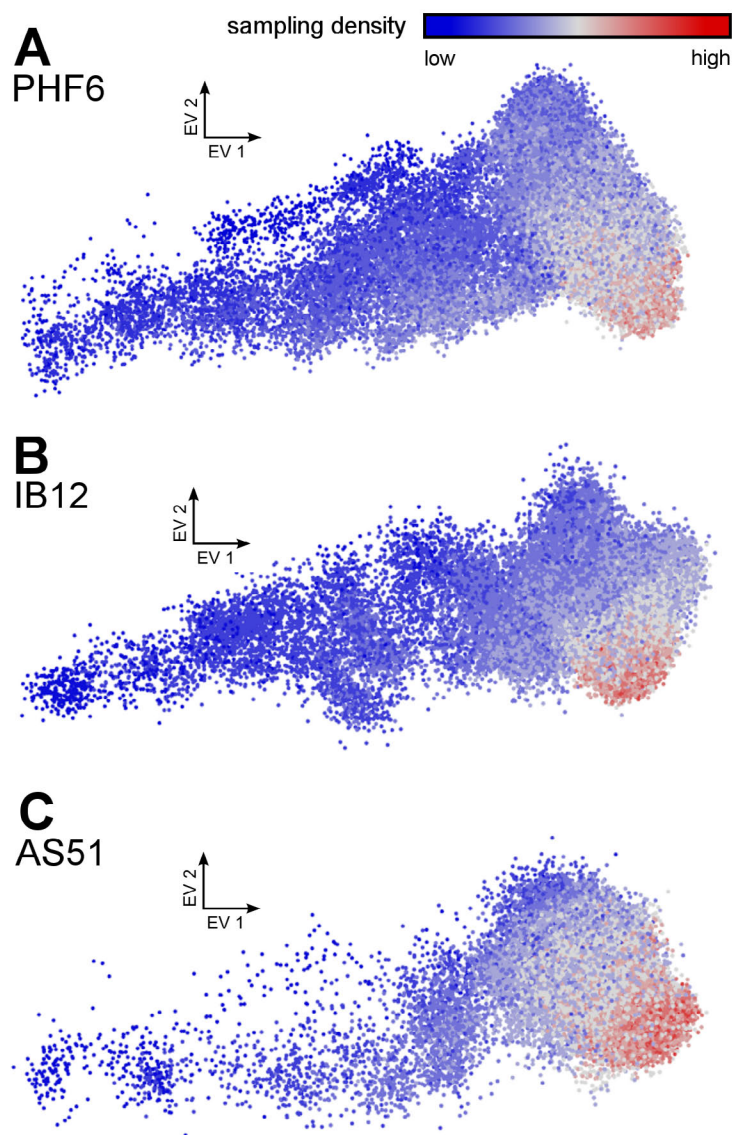


Figure 4.28: **Projection of conformational ensembles and sampling density.** The sampled conformations in the PHF6 (A), IB12 (B) and AS51 (C) trajectories are projected in the same collective coordinates space, as before where each configuration is represented by a point. The projection is viewed from two different angles and the sampling density for all simulations is indicated by a color gradient (red - high; blue - low density).

extension (Ext), while formation of extended  $\beta$ -strands occurs at the expense of any other secondary structure elements (coil, bend, turn). The conformational dynamics in the assembled state are governed by subtle, sequence dependent differences as already seen in the different mapped structure ensembles. For example, the formation of  $\beta$ -sheets was found to be less

prominent in the AS51 oligomers, whereas the sampling density in the configurational space was found to be positively correlated with the  $\beta$ -bridge content (Bbc). This is also expressed in the smaller correlation of the sampling density to the CLs observable and explained by the prominent sampling of closed, barrel-like sheet structures with low nematic order (P2). Moreover, there were notable contributions from interactions between main chain and side chain atoms, which stabilized the AS51 aggregates, opposed to a decrease in these quantities for PHF6 (coulombic) and IB12 (van der Waals) aggregates. The observed decrease in coulombic interactions between the peptide main chain and side chains in the PHF6 decamers is due to the reduced number of Gln side chain hydrogen bonds to the backbone. Thus offering an explanation for the structural stability of oligomer structures with only small  $\beta$ -sheet content and their slow conversion to more ordered ones. The loss of these nonspecific interactions allowed for the growth of double layer sheets with interdigitated side chains. The structural ordering in IB12 oligomers occurred mainly by orientational reorganizations to parallel pairs of sheets, hence explaining the high correlation with P2. Thereby a more efficient packing of adjacent  $\beta$ -sheets was achieved, leading to a larger interface burial (SiMS) and a slightly higher solvent contact. These sheet rearrangement are indicative of transitions to IB12 oligomer structures with a dry and wet sheet interface, as well as solvent exposed edge strands. While the loss in van der Waals peptide side chain to solvent interactions was commonly found to decrease further for the decameric states of PHF6, IB12 and AS51, this was not the case for the coulombic part, which was favorable for the aggregate ordering of the more hydrophilic PHF6 and IB12 peptides.

Overall, a picture emerges where a hydrophobic solvent effect appears to drive the initial clustering of the peptides into collapsed and partially ordered aggregates. This is followed by a phase of structural ordering in the fully assembled state, primarily characterized by a further build-up of  $\beta$ -sheets and, in particular, by the formation of dry inter-sheet interfaces. The FMA results corroborate the obtained structural information on the respective oligomer structures and rationalize the energetic contributions and sequence dependent differences on a molecular level.

### **Effect of mutations on oligomer topology and conformational order.**

Based on the analysis of structural determinants and driving forces underlying the primary aggregation steps we hypothesize that a mutational study might give additional insight in the aggregation mechanism of amyloidogenic peptide model systems. Several mutations were chosen to probe the residue specific effect on the interactions driving the formation of ordered,

## 4.2 Driving forces and structural determinants of steric zipper peptide oligomer formation elucidated by atomistic simulations.

$\beta$ -sheet-rich aggregates, similar to what has been done previously in hexa-peptide fibrilization assays [87, 270]. In particular, we tested the impact of side chain mutations on the aggregation kinetics and sheet-to-sheet packing organization in steric zipper peptide oligomers. The AS51 peptide (GVATVA) was selected for the mutations since it aggregates into oligomers with the least stabilization through  $\beta$ -sandwich structure-like interfaces and therefore is considered a suitable candidate for the predictions to directly test our simulation hypothesis. In this context, exchanging Val, Ala and Thr residues in neighboring positions in the peptides should not affect the  $\beta$ -strand formation propensity strongly, whereas shifting Ala and Val residue side chains to the same side of the strand, might lead to different sheet interfaces upon assembly (GVAT**AV**, GVA**VT**A). Introducing an additional site with hydrogen bonding capability (Ala3Thr) should increase disordered inter-strand main chain-to-side chain interactions (GV**TT**V**A**). The presence of additional Ala residues is expected to reduce the inter-strand side chain stabilizing contacts, on the one hand leading to more random-coil conformations and on the other hand weaker inter-sheet packing interactions (GA**ATAA**). In contrast, a mutation to Phe at position 5 is predicted to have a high  $\beta$ -sheet propensity and the potential to form a large number of inter-peptide interactions. However, replacing Val5 with the bulky and strongly hydrophobic Phe residue, might alter the initial assembly of the oligomers as well as the otherwise sterically less demanding sheet interface to a large extent (GVAT**FA**).

Table 4.7: **Summary of structural and energetic properties for the AS51 peptide mutant simulations.**

System	$\beta$ -sheet	P2	$E_{coul+vdW}$ mc	$E_{coul+vdW}$ sc	$E_{coul+vdW}$ mc-sc
GVATVA	$0.44 \pm 0.11$	$0.26 \pm 0.07$	$-806 \pm 80$	$-208 \pm 17$	$-349 \pm 54$
GVAT <b>AV</b>	$0.46 \pm 0.11$	$0.49 \pm 0.11$	$-859 \pm 69$	$-192 \pm 19$	$-269 \pm 27$
GVA <b>VT</b> A	$0.55 \pm 0.09$	$0.22 \pm 0.06$	$-817 \pm 65$	$-207 \pm 14$	$-360 \pm 40$
GV <b>TT</b> V <b>A</b>	$0.30 \pm 0.09$	$0.22 \pm 0.07$	$-920 \pm 66$	$-237 \pm 26$	$-402 \pm 51$
GVAT <b>FA</b>	$0.53 \pm 0.13$	$0.41 \pm 0.09$	$-798 \pm 66$	$-330 \pm 22$	$-361 \pm 33$
GA <b>ATAA</b>	$0.30 \pm 0.11$	$0.34 \pm 0.11$	$-916 \pm 87$	$-118 \pm 13$	$-296 \pm 33$

Mutations in the AS51 peptide are indicated in boldface. The average and standard deviation were calculated for various observables over a 500 ns interval for each simulation (0.75-1.25  $\mu$ s):  $\beta$ -sheet content, orientational order P2 and the sum of several coulombic and van der Waals potential energy terms ( $E_{coul+vdW}$ ) describing the interactions. Abbreviations and description of the observables are given in the Methods section.

The spontaneous oligomerization was studied for each peptide mutant with an additional

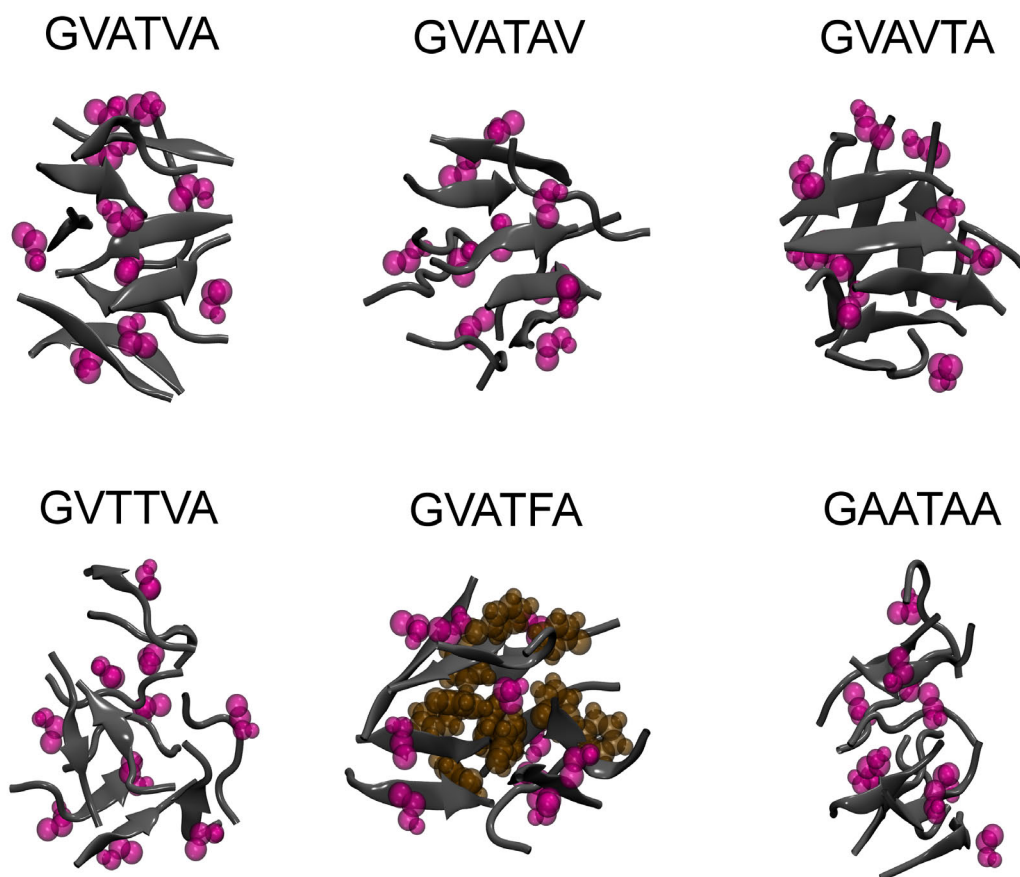


Figure 4.29: **Representative structures of mutational study.** Representative oligomer structures for each AS51 peptide mutant simulation with the highest nematic order, as well as  $\beta$ -sheet content extracted from the last 500 ns. The peptide backbone is rendered as cartoon and shown in dark gray. A subset of side chains is depicted in transparent sphere representation: Thr (magenta) and Phe (brown).

1.25  $\mu$ s long simulation. Indeed, a number of predicted effects take place upon mutation, illustrated by the set of preliminary results summarized in Table 4.7 and the representative oligomer structures shown in Fig 4.29. As expected for the GVTTVA peptide, ordering transitions were impeded by strong main chain-to-side chain interactions, such that the oligomer structures were mainly composed of dimer subunits packed together. The GAATAA peptide aggregates were indeed found to be mostly disordered, with a low  $\beta$ -sheet content and additionally no apparent interface formation. In stark contrast, the interior of the  $\beta$ -sheet-rich GVATFA decamers was found to be occluded by the introduced Phe side chains. The sheet topology in these oligomers

were rather wedge-shaped and did not resemble a  $\beta$ -sandwich. The interface characteristics in the GVATAV and GVAVTA, in comparison to the GVATVA peptide aggregates, emphasize the subtle effects of side chain complementarity in small  $\beta$ -sheet oligomers. While the GVAVTA oligomers formed two orthogonal and stably packed  $\beta$ -sheets, the GVATAV decamers had in addition to a high  $\beta$ -sheet content also a high nematic order. This was found to be due to the parallel arrangement of the sheets with defined inter-sheet interface leading to weaker main chain-to-side chain contacts and a high average number of exposed edge strands. We conclude that these mutations therefore also offer an attractive means to probe the aggregation kinetics of such peptides experimentally.

#### 4.2.4 Discussion

The application of a collective coordinate mapping allowed for the detailed characterization of the diverse conformational ensembles for the simulated oligomerization process of three steric zipper peptides. This mapping procedure enabled the visualization of the structural inter-relationships of the sampled multimeric aggregates and thus provided a comprehensive overview of the global trends in the large-scale molecular changes accompanying the studied assembly process. At the same time, the discriminative power of the mapping allows a direct identification of appropriate reaction coordinates for the projection of high-dimensional configuration spaces [301, 302].

The investigation of the oligomerization process and the underlying driving forces revealed two principal phases and illustrates the delicate balance between inter-peptide and peptide-solvent as well as solvent-solvent interactions. The first step, the association from monomers to oligomers, was commonly found to be correlated with the reduction in solvation free-energy, leading to a hydrophobic collapse of the peptides and minimized peptide surface area in the aggregates. The primary cause of the loss in peptide-solvent interactions was determined to be the poor solvation properties of water for the polypeptide backbone and a hydrophobic dewetting of the peptide side chains. A similar preference of peptide-peptide over peptide-solvent interfaces has been discussed for the aggregation of strongly hydrophilic polyglutamine chains [90, 303]. In addition to a gain in water-water hydrogen bonds, the release of water molecules into bulk solvent is postulated as a universal driving force to further facilitate the aggregation process by an increase in solvent entropy [120, 289, 304], opposing the loss of conformational entropy in the peptide aggregates.

The observed concurrent increase in inter-peptide interactions is in accordance with previous simulations, where the partitioning of the residue side chains from the aqueous phase preceded



the inter-peptide hydrogen bond formation [123]. Here, the emerging ensemble of early steric zipper peptide oligomers was characterized by stable and compact, but only partially ordered aggregate structures, due to frequently formed inter-peptide main chain to side chain contacts. For the PHF6 peptide system this scenario was particularly predominant, as the Gln side chains function as a good solvent for the peptide backbone and nonspecific hydrogen bond formation led to stronger and persisting aggregate disorder. Moreover, the subsequent conformational rearrangements in the decameric structures were marked by the development of more directional nonbonded inter-peptide interactions in form of extensive hydrogen-bonds between peptide strands, resulting in a significant amount of  $\beta$ -sheet structure. In line with previous findings [303], we observe that the condensed peptide aggregates with dry interior, where peptides are primarily solvated by other chains, promote the backbone-driven  $\beta$ -sheet structure formation and contribute to the overall stability of the aggregates.

The formed decameric structures were subject to fluctuations on the microsecond timescale and despite the rich structural diversity, a notable finding was the common formation of a dry inter-sheet interface. The ordered peptide arrangements originated from stronger inter-strand interactions in the  $\beta$ -sheets and further optimized inter-sheet packing through interdigitating side chains. Although multiple sampling routes, involving the presences of various intermediate conformations [285], were taken on the way toward the decamers, the difference between the ordered PHF6, IB12 and AS51 oligomers in the decameric phase was mostly related to the sheet topology ( $\beta$ -sheet size and packing). For the hydrophobic and less bulky AS51 peptide, the largest fluctuation in  $\beta$ -sheet content and a more continuous hydrogen bonding pattern were found. Hence, the conformational characteristics of the most abundant AS51 aggregates were mostly  $\beta$ -barrel-like structures with a dry interior. In contrast to AS51, the PHF6 and IB12 peptides spontaneously formed frayed and twisted  $\beta$ -sandwich oligomers with a significant amount of solvent-exposed edge strands.

Some of the observed oligomers achieved a close structural resemblance to the basic cross- $\beta$  element proposed for the fibrils [82, 105]. However, they were not fully ordered compared to the steric zipper motif in the crystalline state, which displays two planar  $\beta$ -sheets positioned face-to-face [82]. In particular, the high polar and nematic  $\beta$ -strand order remains an important difference to the established interfaces for the small and polymorphic oligomers, found here. Rather, additional degrees of freedom, such as sheet-to-sheet angle and surface curvature, strand alignment and residual coil and bend conformations characterized the oligomeric states. Compelling evidence for a larger heterogeneity in sheet pairing arrangements in non-fibrillar aggregates comes from recent crystal structures of a minimal, tetrameric assembly unit of

macrocyclic peptides [81]. There,  $\beta$ -sheet dimers with steric zipper analogous interfaces were found. This furthermore supports the notion that the formation of complementary and dry quaternary inter-sheet contacts is a common feature of oligomers and fibrils [81, 283], thereby extending the view on the structural organization of oligomers at a pre-nucleation stage. Further growth of pre-fibrillar intermediates beyond a critical size [52, 53, 277] is proposed to proceed by replicating the principal zipper structure via addition of  $\beta$ -strands [81, 124, 129, 130]. In our simulations, the reference IB12 cross- $\beta$  oligomers with the crystal packing were at most metastable for the studied size of 10 chains, and relaxed to conformations similar to the spontaneously formed ones. Interestingly, the SH5-5 construct of the PHF6 peptide with a purely hydrophobic crystal structure interface was found not to be stable in our simulations and by others, where even larger steric zipper assemblies were probed [305]. Thus, the presence of polar and aromatic residues likely provides a stronger contribution than hydrophobic ones to the interface stability of opposing  $\beta$ -sheets in small oligomeric aggregates [133, 286, 305]. This was confirmed by our simulations of the AS51 peptide mutants, which showed an overall enhanced propensity to sample stably packed  $\beta$ -sheets when a Phe residue side chain was introduced into the sequence. The preferential anti-parallel strand alignment as often observed for short peptides [270, 285, 287], offers an additional explanation for the destabilization and distortion of the all-parallel PHF6 steric zipper motif due to the less favorable intra-sheet interactions.

The observed aggregation kinetics and obtained conformational ensembles are inherently dependent on the accuracy of the applied force field, therefore a force field effect cannot be fully excluded. We chose the GROMOS96 43A1 force field for the current study for the following reasons: (1) previous simulations on model peptide folding revealed a balanced secondary structure propensity with no particular bias toward  $\beta$ -sheet formation [171]. (2) This force field is known to spontaneously sample amyloid-like  $\beta$ -strand structures [124, 285, 306], consistent with experimental findings [307] and to stabilize preformed pairs of cross- $\beta$  sheets [129, 308]. (3) The GROMOS96 force fields are primarily calibrated against thermodynamic experimental data and therefore achieve an adequate accuracy for the free-energy of solvation for the neutral analogs of the nonpolar (Ala, Val, Leu, Ile) and aromatic (Tyr) amino acids [309], which mainly constitute the hydrophobic peptides (VQIVYK, VEALYL, GVATVA) investigated here. In addition, we observed in the present study that spontaneous and reference simulations converge to similar conformational ensembles compatible with the relaxed crystallographic steric zipper structures.

The detailed structure of the amyloid state is not entirely determined by sequence [43] and thus hinders the straightforward identification of the sequence-specific driving forces [16, 91,

92, 277]. The partitioning into multiple aggregation pathways, where the formed oligomers exhibit different extents of  $\beta$ -sheet secondary structure, is proposed to be kinetic in nature [123, 127, 285], as experimentally determined, for example, A $\beta$  [72, 284, 310] and PrP oligomerization [279]. The promiscuous polymorphism found for amyloid aggregates [43, 104, 105] may also be explained by the thermodynamic selection of the most stable steric zipper motif in the nucleus [119, 277]. The observed initial hydrophobic collapse of the peptides is an early event in the oligomerization and therefore not rate-limiting for the nucleation, as deduced from simulations [120] and experiments [53, 277]. In line with the present findings this might suggest a mechanistic continuum for amyloidogenic aggregation with the principal characteristics of a condensation-ordering mechanism. The determining conformational conversion step occurs in the oligomeric state directed by specific structural predilections of sheet packing interactions and is largely independent of the monomer concentration [53, 277, 310–312].

The overall registry and alignment of the individual strands were found to be of less importance in the initial association phase for the small steric zipper peptides. This lack of homogeneous strand patterning in the  $\beta$ -sheet structure of early oligomers may be interesting to probe by solid-state NMR experiments, in particular since a regular strand and sheet packing is probably one of the key requirements for the nucleation and selection of a particular zipper structure and therefore also dependent on the sequence. In this context, it is expected that the formation of ordered pre-nucleation intermediates with intra- and inter-sheet  $\beta$ -structure elements similar to the fibrillar state critically depend on the steric complementarity of the sheet interfaces in the oligomeric species, that is the fine detail of matching and properly packed side chains [81, 276, 287].

The observed structural ordering and conformational sampling in the ensemble of collapsed peptide aggregates were not found to correlate well with the solvation free-energy. It has been suggested along these lines that the surface geometries at the molecular level influence solvent-mediated forces, such as dewetting transitions [304, 313] and thus might govern the formation of the high-energy nucleus. The oligomer topologies with sheet pairing angles very different from parallel  $\beta$ -sandwich structures might be indicative of off-pathway oligomers [81]. For these aggregates the expected height of the nucleation barrier to a cross- $\beta$  transition state is large. This would be consistent with the observed lag time in amyloid formation and the observation that pre-fibrillar oligomers can acquire significant  $\beta$ -sheet structure as determined by experiments [290, 291, 310]. Similar effects could originate from the found positional preference of certain side chains, inside or outside relative to the oligomer surface. With reference to a free-energy landscape point of view, it has been argued that small toxic oligomers must exist as

local free-energy minima with significant barriers to amyloid formation, although single amino acid mutations that increase the rate of disease progression often increase the rate of amyloid formation [278].

Finally, the oligomerization of larger peptides and full-length proteins might obey similar principles but usually needs to be accompanied by a substantial change or loss of intra-peptide interactions. Moreover, aggregation is facilitated by transient misfolding or unfolding events of particular aggregation-prone segments, which populate amyloid-competent conformers, such as hairpin motifs or intra-molecular  $\beta$ -strand stacks in natively disordered peptides [106, 132, 284, 314] or accessible self-complementary stretches on protein surfaces [92, 94, 101]. Recent findings furthermore suggest that while oligomer formation might be facilitated by small segments and their local properties, the eventual maturation toward fibrillar structure will involve a remodeling stage to further incorporate and accommodate residues into the  $\beta$ -strands [255, 315].

#### 4.2.5 Conclusion

In the present work the oligomerization of three different amyloidogenic peptide sequences was studied with atomistic MD simulations in explicit solvent environment. The initial stages of the aggregation process were characterized by two distinct phases and the conformational conversions from random assembly to  $\beta$ -sheet oligomers resembled the previously proposed condensation-ordering mechanism. As a first step, the peptides assembled via various intermediates to partially ordered aggregates, thereby creating desolvated interfaces between the chains. In a second step, sequence-dependent conformational reorganizations toward  $\beta$ -sheet-rich structures took place in the collapsed oligomeric state. The kinetics and stability of aggregates with  $\beta$ -sandwich structure motifs were found to exhibit a profound dependence on the hydrophobic character, steric constraints and positioning of the side chains participating in the sheet-to-sheet interfaces. This study furthermore provides qualitative evidence that early steric zipper peptide oligomers feature similar self-complementary sheet packing characteristics as it is proposed for the fibrillar aggregation end-states.

The elucidation of the energetic and structural determinants of amyloidogenic aggregation poses an essential challenge to biophysical studies and still needs to rely on the study of simplified model systems. While structural models of short peptides become more readily available, an thorough understanding of the dynamical transitions between the canonically defined states along the amyloidogenic aggregation pathway is still missing. Here, we found that solvent-mediated interactions, such as the prominent reduction in solvation free-energy drive the primary peptide oligomerization steps. After the initial collapse, the onset of an order-

ing process was observed, mainly driven by extensive backbone hydrogen bond formation and  $\beta$ -sheet lamination.

The overall dominant solvent effects and the observed conformational changes in the metastable oligomer species hint at a causal relation, also crucial for other molecular recognition processes [316]. Therefore, the view is emphasized that biomolecular aggregation of peptides and proteins is governed by the fine chemical details of peptide-solvent interactions and water-structure at various stages of the self-assembly process [106, 108, 120, 130, 289, 304, 313].

## Effect of phospholipid membranes on the self-assembly process of model peptide aggregates

---

### Summary

The influence of a phospholipid bilayer on the primary steps of peptide aggregate self-assembly for two model peptides,  $\alpha$ -synuclein and the islet amyloid peptide, has been studied and compared to a situation with peptides in aqueous solution.

In contrast to the simulations in aqueous solution, we found that aggregate growth did not occur on the lipid bilayer surface. Upon partitioning of the peptide molecules into the membrane water interface, the diffusion slowed down, essentially immobilizing the molecules. Moreover, the conformational fluctuations of the peptide chains were significantly reduced. Although partially aggregated, the adopted structures on the bilayer surface were not compatible with extended  $\beta$ -sheet conformations, corroborating the observation of a competition between adsorption on the model membrane and peptide aggregation in solution. The preliminary data suggests that the strongly altered kinetics of the aggregation process in comparison to the aqueous phase are caused by the favorable coulombic interactions of polar parts of the peptides and the lipid head group region as well as the strong desolvation of the peptide main and side chain atoms on the bilayer interface. The detailed investigation of these observations is expected to yield further insight into the interactions of peptide and phospholipid interfaces, as well as their effect on the early steps of the aggregation process.

## 5.1 Introduction

The complex structural and dynamical properties of phospholipid bilayers play a fundamental role in cells and organelles. Cellular membranes exhibit selective permeability to ions and small molecules, thereby maintaining chemical and biochemical potentials across different compartments and act as matrices to support embedded peptides and proteins [317].

Many amyloidogenic peptides, such as the beta amyloid peptide ( $A\beta$ ),  $\alpha$ -synuclein and the islet amyloid polypeptide (IAPP) are soluble in aqueous and membrane environments [13, 70]. Although partially unstructured or unfolded in solution, helix formation is induced upon membrane binding [74, 318].

Membrane exposure or association was also found to be correlated with amyloid formation suggesting a structural effect on the aggregation process, leading to highly structured fibrillar aggregates [13]. Furthermore, oligomeric intermediates were reported to localize on biological membranes [64, 71]. This offers a plausible answer to the question how amyloid formation and cellular toxicity are related, namely by oligomeric states which may disrupt membrane integrity. Nevertheless, several possible membrane perturbation scenarios are discussed and all of them offer an explanation for cellular stress, e.g. through loss of chemical potential and compartmentalization. Although the exact mechanism and modes of toxicity are still under debate [13, 73], in vitro studies have shown the ability of amyloid intermediates to render model membranes permeable to ions and dye molecules [13].

As outline above, an important aspect of lipid bilayers is their potential to act as a catalyst for amyloid formation.  $A\beta$ ,  $\alpha$ -synuclein and IAPP bind readily to lipid bilayers. The timescales of association to the membrane surface are fast compared to the lag time of the nucleation-dependent assembly process to peptide fibers [319–322]. It is uniformly and consistently found that the presence of lipid surfaces dramatically accelerates the conversion to the  $\beta$ -sheet rich amyloid aggregates. These findings suggest that the organization of intermediate, membrane-bound states is relevant to stabilize early, oligomeric aggregate species (pre-nucleus), which are high in energy and therefore unfavorable [29].

In the present study the question is addressed whether a lipid bilayer interface can act as a mediator for the primary steps of peptide aggregate self-assembly. From the various types of phospholipids, the most abundant in neural membranes are the ones with zwitterionic phosphatidylcholine (PC) head groups [323]. To identify the influence of a lipid bilayer, the spontaneous peptide aggregation process in the absence and presence of a dipalmitoylphosphatidylcholine (DMPC) model membrane for two different model peptides (AS51, IA21) was compared using molecular dynamics simulations.



The AS51 (<sup>51</sup>GVATVA<sup>56</sup>) peptide is one of several segments of  $\alpha$ -synuclein which have been found to be essential in fibril formation [33]. The physiological function of full length  $\alpha$ -synuclein is not fully understood, but likely involves lipid mediated interactions [13, 74, 318, 322, 324]. In solution,  $\alpha$ -synuclein adopts mostly unstructured conformations at neutral pH [74]. It was shown that insertion of the AS51 segment in the globular protein RNase is sufficient to enforce and induce the formation of  $\beta$ -sheets and fibrils [101], while deletion of the same segment from the sequence of  $\alpha$ -synuclein renders the peptide incapable of forming fibrils [33]. An increase in  $\beta$ -sheet structure is correlated with the neurotoxicity of  $\alpha$ -synuclein [322, 325, 326]. In addition, early onset of familial Parkinson's disease and an increase in the aggregation propensity was found for the A53T  $\alpha$ -synuclein mutant [327, 328].

Human IAPP is associated with type II diabetes, as fibrillar extracellular deposits are found in pancreatic extracts of nearly all diabetes type II patients. This is furthermore corroborated by the significant cytotoxicity which IAPP displays to cultured islet cells from the pancreas (where it is expressed, stored and secreted in vivo) [329–331]. The fibrillar form of the 37-residue long IAPP has been studied intensively, and several amyloidogenic regions within IAPP have been identified [46, 50, 332, 333]. One of the segments from the IAPP sequence which was reported to be capable of forming amyloid-like fibrils and steric zipper microcrystals is the hepta-peptide IA21 (<sup>21</sup>NNFGAIL<sup>27</sup>) [102, 333]. The IA21 segment is also part of the putative spine of the IAPP fibrils under physiological conditions [83, 334, 335].

Both peptides were simulated at a concentration close to the crystallization conditions, in the presence and absence of a DMPC lipid bilayer.

## 5.2 Methods

### MD Simulation Protocol

All MD simulations were carried out using the GROMACS software package (version 4.5) [175–177, 229]. All protein bonds were constrained with the P-Lincs algorithm [182] and virtual interaction sites of all hydrogen atoms were introduced, thereby removing all internal vibrational degrees of freedom [175, 179]. Water molecules were constrained using SETTLE [234].

The simulations were run using a 4 fs integration time step. Neighbor lists for non-bonded interactions were updated every 5 steps. The short-ranged Lennard-Jones and coulombic interactions were cut-off at 1.4 nm and 0.9 nm, respectively. Electrostatic interactions for a distance larger than the (real space) cut off were calculated with

the Particle Mesh Ewald (PME) [230, 231] method at every step with a grid spacing of 0.12 nm. The relative tolerance at the cut-off was set at  $10^{-6}$ . The simulations were carried out using periodic boundary conditions. The plane of the DMPC bilayer defined the xy-plane of the box and the membrane normal the z direction, accordingly. A semi-isotropic pressure coupling (xy and z) was applied to keep the pressure constant by coupling the system to a pressure bath of 1 bar ( $\tau = 1$  ps). The Berendsen coupling algorithm [187] was used. Velocity rescale [188] was applied for temperature coupling to a temperature bath of 323 K. Initial velocities were taken from a Maxwellian distribution at 323 K.

**Force field parameters.** For the production runs the GROMOS96 45A3 [336] protein force field and the SPC water model [184] were used. The interactions of the united-atom DMPC molecules were described by the Berger lipid force field parameters [165, 337]. The set of nonbonded parameters ( $\epsilon$ ,  $\sigma$ ) for interactions between protein (GROMOS96 45A3) and lipid (Berger) atoms were derived by applying the standard combination rules.

## Initial Conformations

For each peptide sequence (GVATVA: AS51; NNFGAIL: IA21) conformational ensembles containing 1000 structures were generated with CONCOORD [256]. Only topological constraints were defined, resulting in random starting configurations.

The individual simulations without the DMPC bilayer were set up according to the protocol reported in chapter 4. The initial conditions for the simulations with a DMPC bilayer were set up as follows: 10 peptide conformations were randomly chosen from the pre-generated structure ensemble and placed in a cubic box randomly in position and orientation (dimensions: 11.5 nm  $\times$  11.5 nm  $\times$  11.5 nm; resulting peptide concentration: 16.6 mM). The box contained a pre-equilibrated, symmetric bilayer patch consisting of 384 DMPC lipids. The peptide molecules were placed such that they were always more than 2 nm away from either sides of the lipid bilayer. This procedure was applied to reduce the bias from the individual peptide's starting position and configuration and to ensure a fully monomeric starting configuration in each of the conducted simulation runs. The protonation state for the different systems was according to a pH of 7. Subsequently, all boxes were solvated with explicit water molecules and counter ions ( $\text{Na}^+$ ,  $\text{Cl}^-$ ) were added to yield an appropriate ionic strength (0.05 mM) and to neutralize the net system charge. After the system preparation, an energy minimization using steepest descent was performed.

## Simulated Systems

An overview of the simulated peptide systems and the respective simulation lengths is given in Table 5.1. The AS51 and IA21 peptides were simulated with and without the presence of a model DMPC bilayer starting with monomeric and random peptide configurations, respectively. Additionally, one simulation for each sequence was set up with small, preformed peptide aggregates. These spontaneously assembled dimeric and trimeric structures were selected randomly from prior simulations without DMPC lipids.

Table 5.1: Summary of performed simulations

system ID	no. of molecules peptide / lipid / water	simulation time (ns)
AS51.monomer	1 / 0 / 5484	1x1000
AS51.monomeric	10 / 0 / 32869	4x500
AS51-DMPC.monomeric	10 / 384 / 36441	1x500
AS51-DMPC.preformed	10 / 384 / 36429	1x500
IA21.monomer	1 / 0 / 5549	1x1000
IA21.monomeric	10 / 0 / 32646	1x500
IA21-DMPC.monomeric	10 / 384 / 36307	1x500
IA21-DMPC.preformed	10 / 384 / 36310	1x500

## Analysis

Samples for analysis were taken every 100 ps from the collected trajectories.

**Definitions of general aggregates.** In order to quantitatively probe the association state of the peptides at any given time, the formed aggregates were classified as follows: Pairwise inter-peptide contact analysis was used to identify the individual aggregates, defined as general peptide assemblies: Peptides which shared an inter-chain residue contact were assumed to be within the same aggregate. For any two peptides  $i$ ,  $j$  an inter-chain contact was considered to be formed if any heavy atom of peptide  $i$  was within a distance of 0.45 nm from any heavy atom of peptide  $j$ .

**Root mean square fluctuation (RMSF).** The average RMSF of the backbone and side chain atom positions was calculated for each residue and each of the ten chains, after fitting to the average structure over all conformations in the trajectory, respectively.

## 5.3 Preliminary results and discussion

The spontaneous peptide aggregation in aqueous phase has been described in detail among others for the AS51 peptide in chapter 4, therefore the essential differences found in the simulations with model membrane interface will be reported and discussed here.

### Adsorption on DMPC bilayers brings aggregate growth to a halt

In order to measure and compare the aggregation behavior of the lipid bilayer with the aqueous solution systems, the aggregate size distributions were calculated over time (see Figure 5.1). From the observed aggregate sizes as well as from a visual inspection of the trajectories, a clear trend was visible for both the AS51 and IA21 peptide systems. While the ten, initially monomeric peptides aggregate fast and irreversibly to decameric structures in solution (Fig. 5.1A and 5.1B), formation of aggregates larger than trimers was not found in the simulations with a DMPC bilayer (Fig. 5.1C and 5.1D). Here, the peptide molecules initially placed in solution and at least 2 nm away from the DMPC water interface bound rapidly to the lipids within 10-25 ns of simulation time (Figure 5.2). The diffusive (lateral) motion of the adsorbed and only partially aggregated molecules was greatly reduced compared to the ones in solution. The desorption of one or more peptide chains from the interface was not found in the timescale of

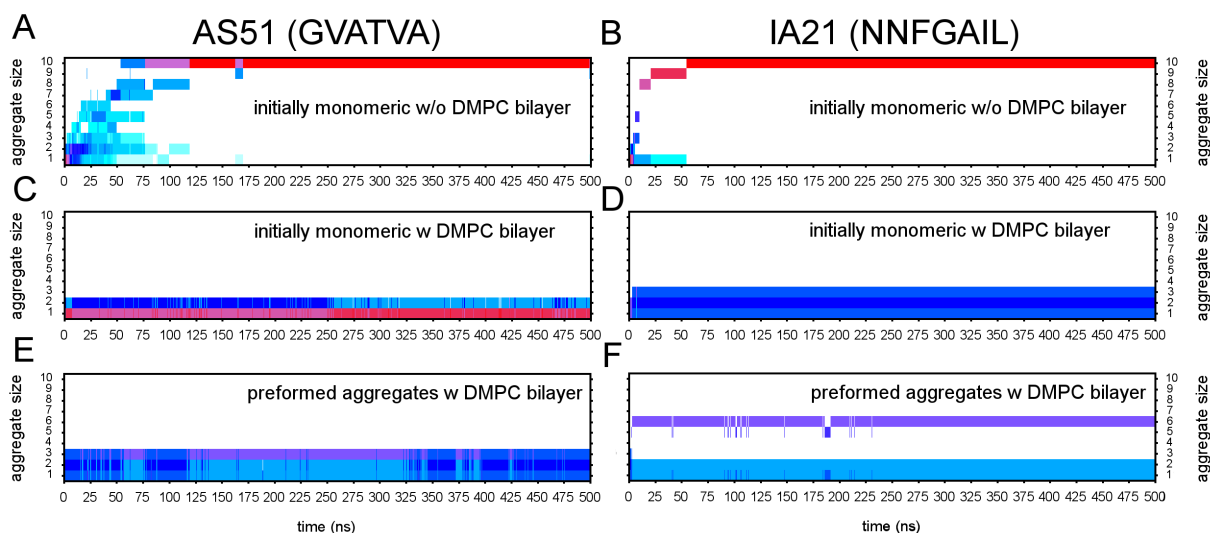


Figure 5.1: **Aggregate size distributions.** Averaged populations of general aggregate size  $n$  at time  $t$  are shown for simulations of AS51 and IA21 peptides in solution and near a DMPC bilayer, respectively. Darkblue, purple and red colors indicate a high averaged abundance of a particular aggregate size.

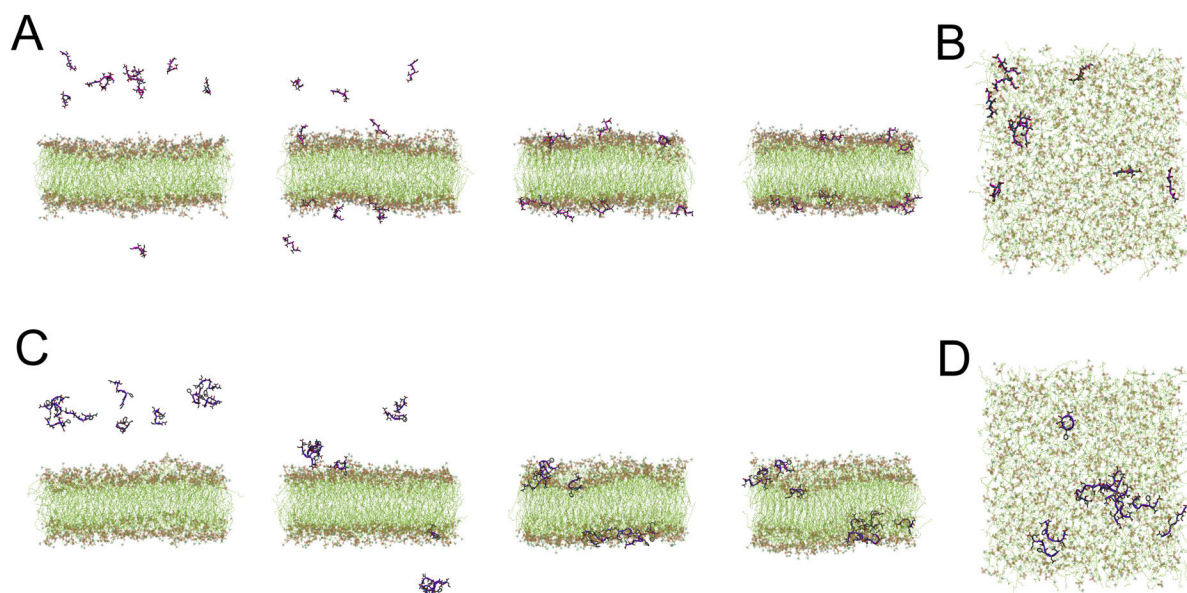


Figure 5.2: **Representative structures for peptide aggregation on DMPC bilayers.** Selected snapshots of the AS51 and IA21 peptide simulations in the vicinity of a DMPC bilayer are shown at 0, 5, 50 and 500 ns, respectively (A, C). In addition, a top view on the organization of peptide aggregates on the bilayer surface at 300 ns for the simulations with preformed aggregates (B, D) is presented. The peptide molecules are drawn as sticks and the peptide backbone is visualized by a cartoon representation (AS51: magenta; IA21: purple). DMPC molecules are shown as transparent sticks and the atoms of lipid head group region are shown by transparent space filling spheres. The lipid atoms are colored in green (carbon), blue (nitrogen) and orange (oxygen). Water molecules are not shown. Molecular images are rendered using the VMD graphics software [263].

the simulations. Once the peptides absorbed flat on the DMPC membrane surface, no further aggregate growth was observed (Figure 5.1). Hence, peptide assembly took place exclusively in solution, where IA21 aggregated faster (higher abundance of aggregates larger than dimers and trimers) in comparison to AS51 (Figure 5.2). Starting from preformed aggregates did not change this behavior (Figure 5.1E and 5.1F). The partitioning of peptide molecules into the center of the membrane was not observed since this transfer process is probably associated with high free energy barriers because of the charged termini and polar residues of the peptides [338].

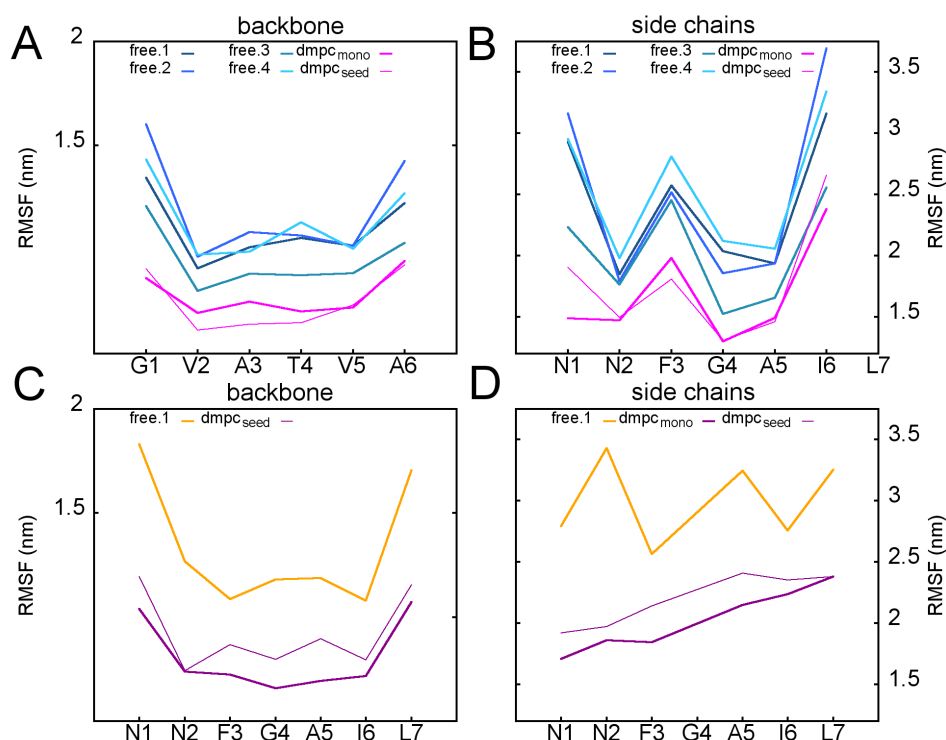


Figure 5.3: **RMSF of backbone and side chain atom positions.** The average RMSF for the AS51 peptide aggregates is shown in the top row: backbone (A) and side chain atoms (B). The bottom row shows the average RMSF for IA21 backbone (C) side chain atoms (D). Note that glycine (G4) does not have a side chain heavy atom, therefore no RMSF was calculated. The blue and orange colored lines match the simulations without, the magenta and purples lines the simulations with DMPC, respectively.

## Structural properties of peptide molecules are altered upon adsorption to the DMPC lipid bilayer

As can be seen from the representative structures in Figure 5.2, the peptide conformations in monomeric as well as partially aggregated form are disordered, both in water and at the interface. Investigating the properties of the peptide molecules in aqueous solution and the quasi instantly adsorbed ones more closely, revealed a number of important differences.

First, the intra peptide end-to-end distance distribution, a global structural measure, was found to be significantly sharper for the peptide chains adsorbed to the DMPC bilayer. Second, the distribution of sampled end-to-end distances peaked around intermediate extensions of the chains ( $\approx 1.0$ - $1.3$  nm). Larger distances ( $\geq 1.4$  nm) between the peptide's terminal ends were only sampled by the fully aggregated structures in solution. Third, in comparison to a free monomer in solution, which was found in a wide range of different and frequently interconverting

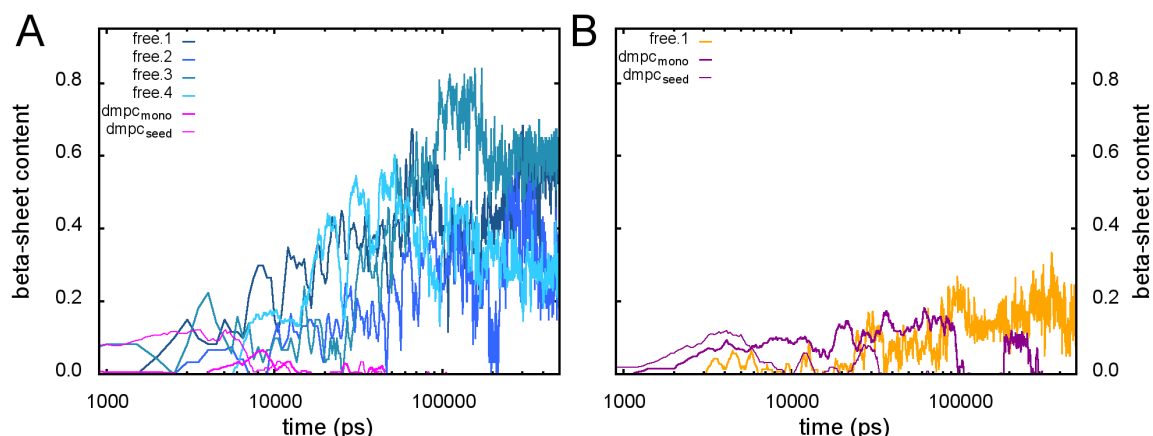


Figure 5.4: **Analysis of  $\beta$ -sheet content.** The development of  $\beta$ -sheet secondary structure elements according to DSSP shown for the all the AS51 (A) and IA21 (B) simulations. Peptide oligomerization in solution was found to be accompanied with the formation of a substantial amount of extended  $\beta$ -sheet structure. The prominent adsorption of the peptides to the DMPC water interface, however did impede  $\beta$ -sheet formation.

conformations, the structures adopted by the adsorbed peptide molecules appear to be even more restricted in their conformational sampling over the course of the simulations (data not shown). To further investigate the flexibility of free and adsorbed peptide molecules, the average RMSF per residue was calculated for backbone and side chain atom positions (Figure 5.3). The conformational fluctuations were found to be prominently reduced for the adsorbed AS51 and IA21 peptides, in particular for the polar asparagine and threonine side chains (Fig. 5.3C and 5.3D). From this, one can conclude that the partitioning of individual side chains into the DMPC head group moieties lead to an overall less fluctuating character of the peptides compared to the situation in a fully assembled decameric aggregate. In addition, the nonpolar residues retained more conformational freedom than the polar ones.

Finally, the  $\beta$ -sheet content did not increase or even decreased (for simulations with pre-formed aggregates) on the DMPC bilayers (Figure 5.4). Instead, mostly coil and to some extent bend (AS51), as well as turn (IA21) secondary structure was sampled. This observation was in stark contrast to the simulations of peptides in solution, where  $\beta$ -sheet formation was observed reproducibly for a 30-70% of all the peptide residues (Figure 5.4).



## Peptide lipid versus peptide peptide association leads to a different degree of solvent-exposed peptide surface

Peptide adsorption was observed to take place fast and irreversibly, while complete insertion into the hydrophobic core of the lipid phase was not observed. The above results suggest a competition between adsorption on the DMPC water interface and peptide self-assembly, as found in solution.

As reported in chapter 4, peptide aggregation in water was associated with the desolvation of a large fraction of the nonpolar molecular surface area by sequestering the respective parts in the interior of the aggregates. However, adsorption of peptides on the surface of the DMPC lipid bilayer in monomeric or partially aggregated form also resulted in a significant reduction of hydrophobic solvent accessible surface area (hSAS). A comparison of nonpolar main and side chain solvation is shown in Figure 5.5 for the simulations with and without DMPC bilayer.

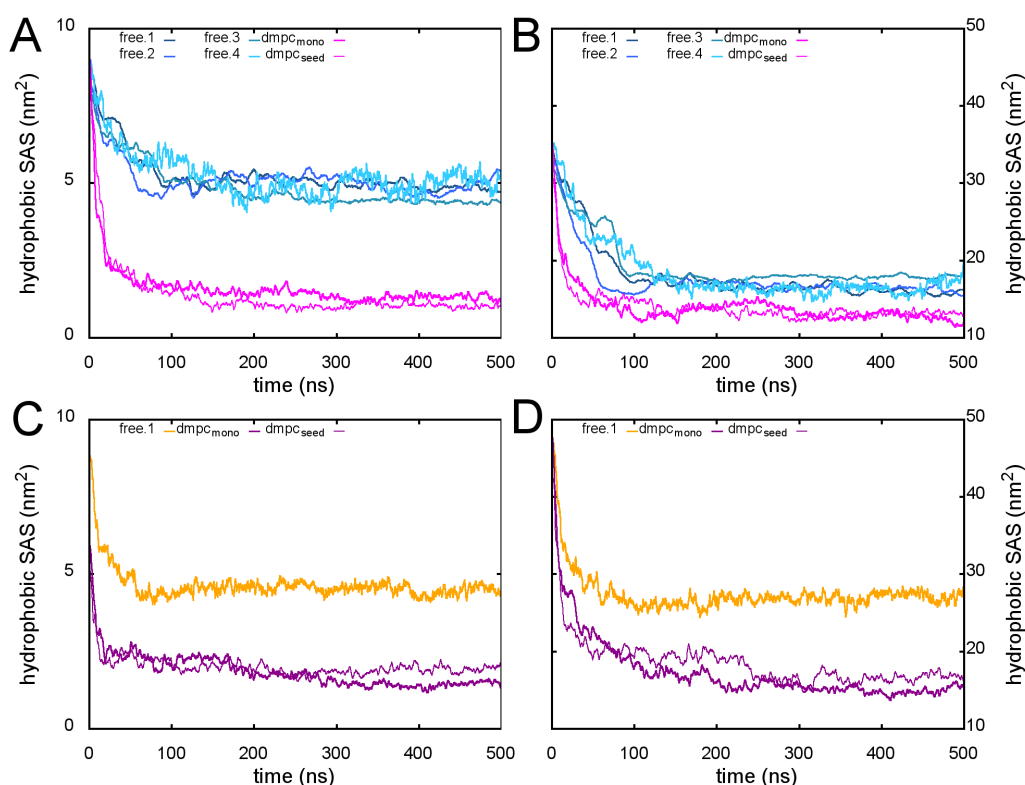


Figure 5.5: **Development of hydrophobic solvent accessible surface area.** The hSAS of the AS51 and IA21 main (A, C) and side chain atoms (B, D) is shown as function of simulation time, respectively.



Surprisingly, the peptide's side chains as well as their main chains were much less hydrated in the case of tight adsorption to the DMPC membrane. Although not assembled into a decameric oligomer, as found in all simulations without DMPC bilayer, a dramatically decreased hydrophobic and hydrophilic SAS was observed. Moreover, the decrease in main and side chain hSAS occurred almost instantly as it coincided with the rapid adsorption of the peptides.

In detail, the strongly hydrophobic AS51 peptides experienced a loss of more than half of the main chain hSAS compared to the oligomeric aggregate. The side chain hSAS of the peptides on the DMPC surface was found to be reduced by 25% compared to the oligomeric aggregates formed in aqueous phase and absence of DMPC bilayer. The IA21 peptide sequence is amphiphilic, with a hydrophilic N-terminal and a strongly hydrophobic patch at the C-terminal site. The burial of main and side chain hSAS was found to be very efficient, as the solvent exposure was reduced by more than 50% compared to the oligomeric aggregates found without DMPC membrane. The hydrophilic parts of the side chains were almost completely deprived from contact to the solvent. The solvation free energy as defined by Eisenberg and McLachlan [300] is an estimate for both of these contributions and showed an overall decrease by 40 and 50 kJ per mol and nm<sup>2</sup>, for GVATVA and NNFGAIL, respectively.

In general, the hSAS converged faster for the peptide main chain (10-25 ns) compared to hSAS of the side chain atoms (100-150 ns). Again, no difference between the simulations of initially monomeric peptides or preformed aggregates was observed.

## 5.4 Conclusion and outlook

The preferential adsorption to the lipid bilayer surface over peptide aggregate self-assembly was the key difference observed in the simulations of two model peptides with and without DMPC model membrane.

In both scenarios, the initial events of peptide adsorption were accompanied by a prominent reduction of nonpolar solvent accessible surface area. However, the large surface presented by the DMPC lipids accommodated the peptide side chains seemingly more efficiently than the small and compact oligomeric aggregates. In particular the partitioning of aliphatic side chains (Ala, Val, Leu, Ile) to the interfacial region and center of the lipid bilayer is energetically favorable as indicated by MD simulations with an umbrella sampling protocol [338]. These findings suggest a competition between peptide aggregate self-assembly and adsorption to the model membrane. Moreover, the solvent-exposed hydrophobic peptide surface in the presence of a DMPC bilayer was strongly decreased in comparison to the observed peptide oligomerization

without a DMPC bilayer. This result confirms the role of hSAS reduction (i.e. the hydrophobic effect) as a driving force for peptide aggregation.

Polar and charged head groups constitute the interfacial region of the DMPC lipids to the solvent and exhibited favorable coulombic interactions to the polar peptide residues, as well as the peptide main chain. Accordingly, the adsorption had profound effects on the structure and hydration of the peptide molecules. The high affinity binding to the DMPC water interface decreased the conformational freedom of the polypeptide chains and narrowed the accessible distribution of conformations. The adsorbed peptides sampled mostly turn and bend states, in contrast to the elongated, strand-like conformations found in the aggregates without bilayer. The strong interactions between peptides and bilayer interface precluded extensive inter-peptide interactions and might explain why adsorption was favored over self-assembly of monomers and small intermediate aggregate species. Furthermore, the molecular kinetics and rates of conformational rearrangement of the peptides at the lipid water interface slowed down, relative to those in water.

All this may cause diverging pathways of peptide aggregation in water and near a phospholipid membrane. The observed peptide lipid interactions should have a concentration-dependent effect on the aggregation mechanism for early and small multimeric aggregates. At large peptide concentrations little to no effect is expected, since the spontaneous aggregation does occur in the aqueous phase as reported, even with the membrane interface being saturated by bound peptides. At very low concentrations the adsorption of monomeric or partially aggregated species to the interface might slow down the aggregation process, by effectively lowering the peptide concentration in solution even further. The highly favorable binding should hinder conformational rearrangements resulting in extensive peptide-peptide interactions furthermore. The effective enhancement of peptide concentration at the water lipid interface might play a role for peptide aggregation under intermediary concentrations.

Ultimately, the fine interplay between density (ratio of peptide to lipid molecules), organization and structure of membrane-bound oligomeric aggregates will be crucial for the subsequent assembly to amyloid fibrils [13]. Despite initial efforts in the characterization of the lipid-catalyzed amyloid formation [13], further efforts, both from the experimental and the theoretical side are required.

The effect of the lipid head groups and peptide concentration could be further investigated following our presented computational approach. It would be interesting to compare different phospholipid types (e.g. anionic PS lipids) in this context [339]. In order to disentangle the effect of electrostatic interactions completely, it could be also interesting to probe the peptide

aggregation process on a purely hydrophobic surface e.g. near an octane slab [340]. Lastly, the question to what extent the properties of the lipid bilayer itself are altered locally or globally by the absorbed aggregates remains to be studied in more detail. Of particular interest in this regard could be to test if and to what extent conformational transitions in already assembled decamers with high  $\beta$ -sheet content occur once they interact with a DMPC bilayer surface, and in addition, how this affects the conformational properties of the membrane.



# Summary and Conclusions

---

The cellular environment is a crowded and dense setting in which peptides and proteins are subjected to a complex apparatus and network of regulatory mechanisms. The performed operations include supervision of protein folding, prevention of aggregation, activation of post-translational modifications, assistance in transport to and initiation of degradation. Failures in these mechanisms can result in the formation of a variety of harmful proteinaceous assemblies, in all likelihood the causative agents of several neurodegenerative disorders subsumed as amyloidosis.

The primary aggregates, such as low-molecular weight oligomers, and the course of events in initial peptide assembly free in solution and on membrane interfaces, however, still lack a detailed characterization. Understanding the molecular basis of nonfibrillar oligomer formation harbors the potential to decipher an important, yet still obscure part of amyloidogenic peptide and protein aggregation. The elucidation of the energetic and structural determinants of amyloidogenic aggregation poses an essential challenge to biophysical studies and still needs to rely on the study of simplified model systems. Although it is not straightforward to establish a link between the amyloid phenomena in vivo to in vitro studies, significant advances in the use of model systems have led to the determination of reliable structural models for amyloid fibril formation. Moreover, concepts and insight from theoretical and simulation studies help to describe and understand amyloidogenic aggregation on a molecular level.

In the present thesis, the early stages of peptide oligomerization have been studied by means of MD simulations, as an important complementary tool to experimental studies in order to understand the main forces and conformational transitions which govern oligomeric aggregate formation.

Major findings and conclusions of this thesis are summarized in the following.

### 6.1 Secondary structure propensities of MD Force Fields

The conformational ensembles obtained with MD simulations of e.g. peptide folding and aggregation are inherently dependent on the accuracy of the applied force field. The study of the folding behavior of various peptide sequences in different MD force fields revealed significant and systematic differences in the stability and formation propensity of dominant secondary structure elements. The observations made suggest that in particular the relative stabilities of helical and extended conformations depend on a subtle balance of force field parameters. Indications for different sampling characteristics of the respective force fields, affecting both the kinetics and convergence of the simulations, were also found. It is likely that the discussed deviations in the structural representations are less critical in protein simulations, when studying the native state dynamics.

A particular finding concerns the treatment of electrostatics in biomolecular simulations. It is a common practice nowadays to apply Particle-Mesh Ewald (PME) methods. However, care is required when employing PME in conjunction with force fields and water models which were originally developed using cut-off or reaction-field (OPLS, GROMOS96). In terms of secondary structure propensities the peptides studied revealed a tendency towards sampling  $\beta$ -hairpin structures when employing PME combined with the OPLS and GROMOS96 force fields. The observed similarities between the different force fields support the notion of converging results for biomolecular systems, but remaining differences emphasize the importance of continuous force field development and refinement. Thus a force field bias for folding studies can not be excluded. From the current perspective there is no single best fit solution for peptide folding simulations with today's non-polarizable force fields, rather a multiple force field or consensus approach is suggested: If computationally feasible to simulate, using more than one suitable force field to address the particular question at hand, and whenever possible, to compare the simulation results to direct experimental data.

### 6.2 Conformational dynamics, assembly pathways and driving forces of steric zipper peptide oligomerization

Atomistic MD simulations of the unbiased spontaneous aggregation process of PHF6, IB12 and AS51 steric zipper peptides from unstructured monomers to  $\beta$ -sheet rich oligomeric assemblies have been reported. The detailed analyses of multiple independent simulation runs highlights several findings which address the course of events in the primary steps of peptide self-assembly

in particular.

In general the oligomerization was found to take place as a common two-step process akin to a general condensation-ordering mechanism. First, a rapid formation of a heterogeneous ensemble of  $\beta$ -sheet rich oligomer structures was observed, where kinetically trapped aggregate intermediates affect the structural evolution of larger assemblies. The oligomeric precursor states were found to associate via a combination of polymerization and condensation mechanisms governed by the characteristics of peptide sequence and oligomer size. The peptides assembled to partially ordered aggregates and thereby created desolvated interfaces between the chains. Second, sequence-dependent conformational reorganizations towards  $\beta$ -sheet-rich structures took place in the collapsed oligomeric state. The kinetics and stability of aggregates with  $\beta$ -sandwich structure motifs was found to exhibit a profound dependence on the hydrophobic character, steric constraints and positioning of the side chains participating in the sheet-to-sheet interfaces. Furthermore, elucidating the  $\beta$ -sheet oligomer conformations at atomistic detail provides qualitative evidence that early oligomers feature similar self-complementary sheet packing characteristics as the fibrillar aggregation end-states.

The detailed analysis of the forces driving the oligomerization revealed that solvent-mediated interactions, such as the prominent reduction in solvation free energy, drive the primary peptide oligomerization steps. After the initial collapse the onset of an ordering process was mainly driven by extensive backbone hydrogen bond formation and lamination of  $\beta$ -sheets.

A thorough characterization of the heterogeneity in molecular dynamics and structures of the low molecular weight oligomers may hold the key for understanding the profound differences in macroscopic fibril growth kinetics [52, 273–275] as well as the observed rich structural diversity of aggregate states [80, 102–104, 248, 269]. Specifically the structural relations between crystalline and fibrillar polymorphs of amyloidgenic peptides may be only resolved by tracing and determining the characteristics of the oligomeric conformational states from which either of the species originate. The overall dominant solvent effects and the observed conformational changes in the metastable oligomer species hint at a causal relation, which is also crucial for other molecular recognition processes such as protein ligand binding and membrane fusion dynamics [316, 341]. For these reasons the view is emphasized that biomolecular aggregation of peptides and proteins is governed by the fine chemical details of peptide-solvent interactions and water-structure at various stages of the self-assembly process [106, 108, 120, 130, 289, 304, 313].

## 6.3 Effect of phospholipid membranes on the self-assembly process of model peptide aggregates

The influence of a DMPC bilayer on the self-assembly process with two model peptides from  $\alpha$ -synuclein and the islet amyloid peptide has been studied and compared to a situation of peptides in aqueous solvent only.

A key observation in contrast to the simulations in the aqueous phase was the finding that peptide aggregate growth did not occur on the membrane surface. Upon adsorption of the peptide molecules to the DMPC bilayer the diffusion slowed down, essentially immobilizing the molecules on the surface. Moreover, the conformational fluctuations of the peptide chains were significantly reduced. Although partially aggregated, the adopted structures on the DMPC surface were not compatible with extended  $\beta$ -sheet conformations. This finding corroborates the notion that a competition between adsorption on the model membrane and peptide aggregation in solution exists. Our preliminary data suggests that the favorable coulombic interactions of hydrophilic parts of the peptides and the head group region of the lipids as well as the strong desolvation of the peptide main and side chains are responsible for the strongly altered kinetics of the aggregation process in the presence of a model membrane surface in comparison to the situation in aqueous phase. The favorable coulombic interactions of hydrophilic parts of the peptides and the head group region of the lipids as well as the strong desolvation of the peptide main and side chains. A detailed investigation of the interactions of the peptides and phospholipid interfaces is expected to yield further insight into the observed impact on the early steps of the aggregation process.

The effect of the lipid head group and peptide concentration could be further investigated following the presented computational approach. It would be interesting to compare different phospholipid types (e.g. anionic PS lipids) in this context [339]. In order to disentangle the effect of electrostatic interactions completely, it could be also interesting to probe the peptide aggregation process on a purely hydrophobic surface, e.g. an octane slab [340]. Last but not least, the question to what extent the properties of the lipid bilayer itself are altered locally or globally by the absorbed peptide aggregates remains to be studied in more detail. Of particular interest in this regard could be to test if and how conformational transitions in already assembled oligomers with high  $\beta$ -sheet content occur once interacting with a DMPC bilayer, and, in addition, how the peptide oligomer interaction affects the conformational properties of the membrane.



# Bibliography

---

- [1] C. M. Dobson and M. Karplus, *Curr. Opin. Struct. Biol.*, **1999**, 9, 92–101.
- [2] C. Levinthal, *J. Chim. Phys.*, **1968**, 65, 44–45.
- [3] K. A. Dill and H. S. Chan, *Nat. Struct. Mol. Biol.*, **1997**, 4, 10–19.
- [4] C. M. Dobson, *Nature*, **2003**, 426, 884–890.
- [5] J. Onuchic and P. Wolynes, *Curr. Opin. Struct. Biol.*, **2004**, 14, 70–75.
- [6] C. Anfinsen, *Science*, **1973**, 181, 223–230.
- [7] R. L. Baldwin, *Nature*, **1994**, 369, 183–184.
- [8] K. A. Dill, S. B. Ozkan, M. S. Shell and T. R. Weikl, *Annu. Rev. Biophys.*, **2008**, 37, 289–316.
- [9] K. Henzler-Wildman and D. Kern, *Nature*, **2007**, 450, 964–972.
- [10] M. Vendruscolo, E. Paci, M. Karplus and C. M. Dobson, *Proc. Natl. Acad. Sci. USA*, **2003**, 100, 14 817–14 821.
- [11] F. U. Hartl and M. Hayer-Hartl, *Nat. Struct. Mol. Biol.*, **2009**, 16, 574–581.
- [12] H. S. Chan and K. A. Dill, *Annu. Rev. Biophys. Biophys. Chem.*, **1991**, 20, 447–490.
- [13] J. Hebda and A. Miranker, *Annu. Rev. Biophys.*, **2009**, 38, 125–152.
- [14] V. d. Vigneaud, R. Sifferd and R. Sealock, *J. Biol. Chem.*, **1933**, 102, 521–533.

- [15] D. F. Waugh, *J. Am. Chem. Soc.*, **1946**, *68*, 247–250.
- [16] M. Fändrich and C. M. Dobson, *EMBO J.*, **2002**, *21*, 5682–5690.
- [17] A. P. P. Gian Gaetano Tartaglia and, S. Campioni, C. M. Dobson, F. Chiti and M. Vendruscolo, *J. Mol. Biol.*, **2008**, *380*, 425–436.
- [18] J. Greenwald and R. Riek, *Structure*, **2010**, *18*, 1244.
- [19] V. N. Uversky and A. L. Fink, *Biochim. Biophys. Acta.*, **2004**, *1698*, 131–153.
- [20] D. J. Selkoe, *Nature*, **2003**, *426*, 900–904.
- [21] M. T. Pastor, N. Kümmerer, V. Schubert, A. Esteras-Chopo, C. G. Dotti, M. Lopez de la Paz and L. Serrano, *J. Mol. Biol.*, **2008**, *375*, 695–707.
- [22] M. Stefani and C. Dobson, *J. Mol. Med.*, **2003**, *81*, 678–699.
- [23] A. S. Cohen and E. Calkins, *Nature*, **1959**, *183*, 1202–1203.
- [24] W. Eaton and J. Hofrichter, *Advances in Protein Chemistry*, **1990**, *40*, 63–279.
- [25] F. Chiti, P. Webster, N. Taddei, A. Clark, M. Stefani, G. Ramponi and C. Dobson, *Proc. Natl. Acad. Sci. USA*, **1999**, *96*, 3590–3594.
- [26] M. Fändrich, M. A. Fletcher and C. M. Dobson, *Nature*, **2001**, *410*, 165–166.
- [27] F. Chiti, N. Taddei, F. Baroni, C. Capanni, M. Stefani, G. Ramponi and C. M. Dobson, *Nat. Struct. Mol. Biol.*, **2002**, *9*, 137–143.
- [28] J. D. Sipe, *Amyloid Proteins. The Beta Sheet Conformation and Disease.*, WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, 2005.
- [29] F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.*, **2006**, *75*, 333–366.
- [30] F. Cohen and J. Kelly, *Nature*, **2003**, *426*, 905–909.
- [31] B. H. Toyama, M. J. S. Kelly, J. D. Gross and J. S. Weissman, *Nature*, **2007**, *449*, 233–238.
- [32] R. Nelson and D. Eisenberg, *Curr. Opin. Struct. Biol.*, **2006**, *16*, 260–265.

- [33] M. Vilar, H.-T. Chou, T. Lührs, S. K. Maji, D. Riek-Loher, R. Verel, G. Manning, H. Stahlberg and R. Riek, *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 8637–8642.
- [34] R. Virchow, *Virchows Arch.*, **1854**, *6*, 415–426.
- [35] M. Fändrich, *Cell. Mol. Life Sci.*, **2007**, *64*, 2066–2078.
- [36] P. Westermarck, M. D. Benson, J. N. Buxbaum, A. S. Cohen, B. Frangione, S.-I. Ikeda, C. L. Masters, G. Merlini, M. J. Saraiva and J. D. Sipe, *Amyloid: J. Protein Folding Disord.*, **2005**, *12*, 1–4.
- [37] F. Rousseau, J. Schymkowitz and L. Serrano, *Curr. Opin. Struct. Biol.*, **2006**, *16*, 118–126.
- [38] J. D. Sipe and A. S. Cohen, *J. Struct. Biol.*, **2000**, *130*, 88–98.
- [39] D. A. Kirschner, C. Abraham and D. J. Selkoe, *Proc. Natl. Acad. Sci. USA*, **1986**, *83*, 503–507.
- [40] L. Serpell, P. Fraser and M. Sunde, *Methods in Enzymology*, **1999**, *309*, 526–36.
- [41] M. Sunde and C. Blake, *Q. Rev. Biophys.*, **1998**, *31*, 1–39.
- [42] O. S. Makin, E. Atkins, P. Sikorski, J. Johansson and L. C. Serpell, *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 315–320.
- [43] A. T. Petkova, R. D. Leapman, Z. Guo, W.-M. Yau, M. P. Mattson and R. Tycko, *Science*, **2005**, *307*, 262–265.
- [44] A. Paravastu, R. Leapman, W. Yau and R. Tycko, *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 18 349–18 354.
- [45] T. Lührs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Dobeli, D. Schubert and R. Riek, *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 17 342–17 347.
- [46] S. Luca, W. Yau, R. Leapman and R. Tycko, *Biochemistry*, **2007**, *46*, 13 505–13 522.
- [47] C. Wasmer, A. Lange, H. V. Melckebeke, A. Siemer, R. Riek and B. Meier, *Science*, **2008**, *319*, 1523–1526.
- [48] J. Helmus, K. Surewicz, P. Nadaud, W. Surewicz and C. Jaronec, *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 6284–6289.

- [49] R. Tycko, *Annu. Rev. Phys. Chem.*, **2011**, 62, 279–299.
- [50] C. S. Goldsbury, G. J. S. Cooper, K. N. Goldie, S. A. Müller, E. L. Saafi, W. T. M. Gruijters, M. P. Misur, A. Engel, U. Aebi and J. Kistler, *J. Struct. Biol.*, **1997**, 119, 17–27.
- [51] H. Heise, W. Hoyer, S. Becker, O. C. Andronesi, D. Riedel and M. Baldus, *Proc. Natl. Acad. Sci. USA*, **2005**, 102, 15 871–15 876.
- [52] A. Lomakin, D. S. Chung, G. B. Benedek, D. A. Kirschner and D. B. Teplow, *Proc. Natl. Acad. Sci. USA*, **1996**, 93, 1125–1129.
- [53] T. R. Serio, A. G. Cashikar, A. S. Kowal, G. J. Sawicki, J. J. Moslehi, L. Serpell, M. F. Arnsdorf and S. L. Lindquist, *Science*, **2000**, 289, 1317–1321.
- [54] A. J. Modler, H. Fabian, F. Sokolowski, G. Lutsch, K. Gast and G. Damaschun, *Amyloid: J. Protein Folding Disord.*, **2004**, 11, 215–231.
- [55] S. A. Petty and S. M. Decatur, *Proc. Natl. Acad. Sci. USA*, **2005**, 102, 14 272–14 277.
- [56] C. J. Roberts, *Biotechnol. Bioeng.*, **2007**, 98, 927–938.
- [57] J. T. Jarrett and P. T. Lansbury, *Biochemistry*, **1992**, 31, 12 345–12 352.
- [58] M. Fändrich, *J. Mol. Biol.*, **2007**, 365, 1266–1270.
- [59] J. D. Sipe, *Annu. Rev. Biochem.*, **1992**, 61, 947–975.
- [60] J. Sacchettini and J. Kelly, *Nat. Rev. Drug Discov.*, **2002**, 1, 267–275.
- [61] S. Prusiner, *Science*, **1991**, 252, 1515–1522.
- [62] J. Collinge, *J. Neurol. Neurosurg. Psychiatry*, **2005**, 76, 906–919.
- [63] M. Bucciattini, E. Giannoni, F. Chiti, F. Baroni, L. Formigli, J. Zurdo, N. Taddei, G. Ramponi, C. M. Dobson and M. Stefani, *Nature*, **2002**, 416, 507–511.
- [64] R. Kaye, E. Head, J. L. Thompson, T. M. McIntire, S. C. Milton, C. W. Cotman and C. G. Glabe, *Science*, **2003**, 300, 486–489.
- [65] M. Bucciattini, G. Calloni, F. Chiti, L. Formigli, D. Nosi, C. M. Dobson and M. Stefani, *J. Biol. Chem.*, **2004**, 279, 31 374–31 382.

- [66] R. Kodali and R. Wetzel, *Curr. Opin. Struct. Biol.*, **2007**, *17*, 48–57.
- [67] C. G. Glabe, *J. Biol. Chem.*, **2008**, *283*, 29 639–29 643.
- [68] S. Campioni, B. Mannini, M. Zampagni, A. Pensalfini, C. Parrini, E. Evangelisti, A. Relini, M. Stefani, C. M. Dobson, C. Cecchi and F. Chiti, *Nat. Chem. Biol.*, **2010**, *6*, 140–147.
- [69] A. Sandberg, L. M. Luheshi, S. Söllvander, T. Pereira de Barros, B. Macao, T. P. J. Knowles, H. Biverstal, C. Lendel, F. Ekholm-Petterson, A. Dubnovitsky, L. Lannfelt, C. M. Dobson and T. Härd, *Proc. Natl. Acad. Sci. USA*, **2010**, *107*, 15 595–15 600.
- [70] H. Lashuel and P. T. Lansbury, *Q. Rev. Biophys.*, **2006**, *39*, 167–201.
- [71] A. Quist, I. Doudevski, H. Lin, R. Azimova, D. Ng, B. Frangione, B. Kagan, J. Ghiso, and R. Lal, *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 10 427–10 432.
- [72] C. A. Lasagna-Reeves, C. G. Glabe and R. Kayed, *J. Biol. Chem.*, **2011**, *286*, 22 122–22 130.
- [73] J. Killian and T. Nyholm, *Curr. Opin. Struct. Biol.*, **2006**, *279*, 473–479.
- [74] E. R. Middleton and E. Rhoades, *Biophys. J.*, **2010**, *99*, 2279–2288.
- [75] J. McLaurin, D. Yang, C. Yip and P. Fraser, *J. Struct. Biol.*, **2000**, *130*, 259–270.
- [76] D. Smith, S. Radford and A. Ashcroft, *Proc. Natl. Acad. Sci. USA*, **2010**, *107*, 6794–6798.
- [77] S. Chimon, M. A. Shaibat, C. R. Jones, D. C. Calero, B. Aizezi and Y. Ishii, *Nat. Struct. Mol. Biol.*, **2007**, *14*, 1157–1164.
- [78] K. Ono, M. Condrón and D. Teplow, *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 14 745–14 750.
- [79] S. L. Bernstein, N. F. Dupuis, N. D. Lazo, T. Wytténbach, M. M. Condrón, G. Bitan, D. B. Teplow, J.-E. Shea, B. T. Ruotolo, C. V. Robinson and M. T. Bowers, *Nat. Chem.*, **2009**, *1*, 326–331.
- [80] M. Ahmed, J. Davis, D. Aucoin, T. Sato, S. Ahuja, S. Aimoto, J. Elliott, W. E. van Nostrand and S. O. Smith, *Nat. Struct. Mol. Biol.*, **2010**, *17*, 561–567.

## BIBLIOGRAPHY

---

- [81] Liu, M. Sawaya, P. Cheng, J. Zheng, J. Nowick and D. Eisenberg, *J. Am. Chem. Soc.*, **2011**, *133*, 6736–6744.
- [82] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. W. Wiltzius, H. T. McFarlane, A. O. Madsen, C. Riek and D. Eisenberg, *Nature*, **2007**, *447*, 453–457.
- [83] K. Tenidis, M. Waldner, J. Bernhagen, W. Fischle, M. Bergmann, M. Weber, M.-L. Merkle, W. Voelter, H. Brunner and A. Kapurniotu, *J. Mol. Biol.*, **2000**, *295*, 1055–1071.
- [84] M. Balbirnie, R. Grothe and D. S. Eisenberg, *Proc. Natl. Acad. Sci. USA*, **2001**, *98*, 2375–2380.
- [85] M. Lopez de la Paz, K. Goldie, J. Zurdo, E. Lacroix, C. M. Dobson, A. Hoenger and L. Serrano, *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 16 052–16 057.
- [86] W. J. Goux, L. Kopplin, A. D. Nguyen, K. Leak, M. Rutkofsky, V. D. Shanmuganandam, D. Sharma, H. Inouye and D. A. Kirschner, *J. Biol. Chem.*, **2004**, *279*, 26 868–26 875.
- [87] M. Lopez de la Paz and L. Serrano, *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 87–92.
- [88] M. I. Ivanova, M. J. Thompson and D. Eisenberg, *Proc. Natl. Acad. Sci. USA*, **2006**, *103*, 4079–4082.
- [89] M. J. Thompson, S. A. Sievers, J. Karanicolas, M. I. Ivanova, D. Baker and D. Eisenberg, *Proc. Natl. Acad. Sci. USA*, **2006**, *103*, 4074–4078.
- [90] R. M. Murphy, *Annu. Rev. Biomed. Eng.*, **2002**, *4*, 155–174.
- [91] M. T. Pastor, M. Lopez de la Paz, E. Lacroix, L. Serrano and E. Perez-Paya, *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 614–619.
- [92] L. Goldschmidt, P. K. Teng, R. Riek and D. Eisenberg, *Proc. Natl. Acad. Sci. USA*, **2010**, *107*, 3487–3492.
- [93] A. Esteras-Chopo, L. Serrano and M. Lopez de la Paz, *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 16 672–16 677.

## BIBLIOGRAPHY

---

- [94] S. Ventura, J. Zurdo, S. Narayanan, M. Parreno, R. Mangues, B. Reif, F. Chiti, E. Giannoni, C. M. Dobson, F. X. Aviles and L. Serrano, *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 7258–7263.
- [95] M. von Bergen, P. Friedhoff, J. Biernat, J. Heberle, E.-M. Mandelkow and E. Mandelkow, *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 5129–5134.
- [96] M. von Bergen, S. Barghorn, L. Li, A. Marx, J. Biernat, E.-M. Mandelkow and E. Mandelkow, *J. Biol. Chem.*, **2001**, *276*, 48 165–48 174.
- [97] J. J. Balbach, Y. Ishii, O. N. Antzutkin, R. D. Leapman, N. W. Rizzo, F. Dyda, J. Reed, and R. Tycko, *Biochemistry*, **2000**, *39*, 13 748–13 759.
- [98] C. P. Jaronec, C. E. MacPhee, V. S. Bajaj, M. T. McMahon, C. M. Dobson and R. G. Griffin, *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 711–716.
- [99] J. Pedersen and D. Otzen, *Protein Sci.*, **2008**, *17*, 2–10.
- [100] R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe and D. Eisenberg, *Nature*, **2005**, *435*, 773–778.
- [101] P. K. Teng and D. Eisenberg, *Protein Eng. Des. Sel.*, **2009**, *22*, 531–536.
- [102] J. J. W. Wiltzius, M. Landau, R. Nelson, M. R. Sawaya, M. I. Apostol, L. Goldschmidt, A. B. Soriaga, D. Cascio, K. Rajashankar and D. Eisenberg, *Nat. Struct. Mol. Biol.*, **2009**, *16*, 973–978.
- [103] P. C. A. van der Wel, J. R. Lewandowski and R. G. Griffin, *J. Am. Chem. Soc.*, **2007**, *129*, 5117–5130.
- [104] K. Marshall and L. Serpell, *Soft Matter*, **2010**, *6*, 2110–2114.
- [105] P. C. van der Wel, J. R. Lewandowski and R. G. Griffin, *Biochemistry*, **2010**, *49*, 9457–9469.
- [106] J. E. Straub and D. Thirumalai, *Curr. Opin. Struct. Biol.*, **2010**, *20*, 1–9.
- [107] J. E. Straub and D. Thirumalai, *Annu. Rev. Phys. Chem.*, **2011**, *62*, 437–463.
- [108] C. Wu, H. Lei and Y. Duan, *Biophys. J.*, **2004**, *87*, 3000–3009.

- [109] H. D. Nguyen and C. K. Hall, *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 16 180–16 185.
- [110] H. Tsai, M. Reches, C. Tsai, K. Gunasekaran, E. Gazit and R. Nussinov, *Proc. Natl. Acad. Sci. USA*, **2010**, *132*, 4266–4280.
- [111] B. Ma and R. Nussinov, *Curr. Opin. Chem. Biol.*, **2006**, *10*, 445–452.
- [112] R. Pellarin and A. Caflisch, *J. Mol. Biol.*, **2006**, *360*, 882–892.
- [113] Z. Zhang, H. Chen, H. Bai and L. Lai, *Biophys. J.*, **2007**, *93*, 1484–1492.
- [114] M. Cheon, I. Chang, S. Mohanty, L. M. Luheshi, C. M. Dobson, M. Vendruscolo and G. Favrin, *PLoS Comput. Biol.*, **2007**, *3*, e173.
- [115] B. Strodel, C. Whittleston and D. Wales, *J. Am. Chem. Soc.*, **2007**, *129*, 16 005–16 014.
- [116] D.-W. Li, S. Mohanty, A. Irbäck and S. Huo, *PLoS Comput. Biol.*, **2008**, *4*, e1000 238.
- [117] G. Bellesia and J.-E. Shea, *J. Chem. Phys.*, **2009**, *131*, 111 102–111 105.
- [118] E. P. O'Brien, Y. Okamoto, J. E. Straub, B. R. Brooks and D. Thirumalai, *J. Phys. Chem. B*, **2009**, *113*, 14 421–14 430.
- [119] J. Park, B. Kahn and W. Hwang, *PLoS Comput. Biol.*, **2009**, *5*, e1000 492.
- [120] D. K. Klimov and D. Thirumalai, *Structure*, **2003**, *11*, 295–307.
- [121] J. Gsponer, U. Haberthur and A. Caflisch, *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 5154–5159.
- [122] S. Santini, G. Wei, N. Mousseau and P. Derreumaux, *Structure*, **2004**, *12*, 1245–1255.
- [123] W. Hwang, S. Zhang, R. D. Kamm and M. Karplus, *Proc. Natl. Acad. Sci. USA*, **2004**, *101*, 12 916–12 921.
- [124] P. H. Nguyen, M. S. Li, G. Stock, J. E. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 111–116.
- [125] N. W. Kelley, V. Vishal, G. A. Krafft and V. S. Pande, *J. Chem. Phys.*, **2008**, *129*, 214 707–214 717.
- [126] B. Urbanc, M. Betnel, L. Cruz, G. Bitan and D. B. Teplow, *J. Am. Chem. Soc.*, **2010**, *132*, 4266–4280.



## BIBLIOGRAPHY

---

- [127] R. Pellarin, P. Schuetz, E. Guarnera and A. Caflisch, *J. Am. Chem. Soc.*, **2010**, *132*, 14 960–14 970.
- [128] W. P. Esler, E. R. Stimson, J. M. Jennings, H. V. Vinters, J. R. Ghilardi, J. P. Lee, P. W. Mantyh and J. E. Maggio, *Biochemistry*, **2000**, *39*, 6288–6295.
- [129] J. Wallace and J. Shen, *Biochemistry*, **2010**, *49*, 5290–5298.
- [130] G. Reddy, J. E. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 11 948–11 953.
- [131] M. Cecchini, F. Rao, M. Seeber and A. Caflisch, *J. Chem. Phys.*, **2004**.
- [132] M. Grabenauer, C. Wu, P. Soto, J.-E. Shea, and M. T. Bowers, *J. Am. Chem. Soc.*, **2010**, *132*, 532–539.
- [133] Y. Lu, P. Derreumaux, Z. Guo, N. Mousseau and G. Wei, *Proteins: Struct., Funct., Bioinf.*, **2009**, *75*, 954–963.
- [134] S. Kim, T. Takeda and D. K. Klimov, *Biophys. J.*, **2010**, *99*, 1949–1958.
- [135] A. R. Leach, *Molecular Modelling: Principles and Applications*, Pearson Education EMA, 2001.
- [136] H. J. C. Berendsen, *Simulating the Physical World, Hierarchical Modeling from Quantum Mechanics to Fluid Dynamics*, Cambridge University Press, 2007.
- [137] M. Levitt, *Nat. Struct. Biol.*, **2001**, *8*, 392–393.
- [138] W. Wang, O. Donini and K. P. Reyes, CM and, *Annu. Rev. Biophys. Biomol. Struct.*, **2001**, *30*, 211–243.
- [139] H. A. Scheraga, M. Khalili and A. Liwo, *Annu. Rev. Phys. Chem.*, **2007**, *58*, 57–83.
- [140] E. H. Lee, J. Hsin, M. Sotomayor, G. Comellas and K. Schulten, *Structure*, **2009**, *17*, 1295–1306.
- [141] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen, S. Piana, R. O. Dror, M. P. Eastwood, J. A. Bank, J. M. Jumper, J. K. Salmon, Y. Shan and W. Wriggers, *Science*, **2010**, *330*, 341–346.

## BIBLIOGRAPHY

---

- [142] K. Y. Sanbonmatsu, S. Joseph and C.-S. Tung, *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 15 854–15 859.
- [143] M. Zink and H. Grubmüller, *Biophys. J.*, **2009**, *96*, 1350–1363.
- [144] P. L. Freddolino and K. Schulten, *Biophys. J.*, **2009**, *97*, 2338–2347.
- [145] V. A. Voelz, G. R. Bowman, K. Beauchamp and V. S. Pande, *J. Am. Chem. Soc.*, **2010**, *132*, 1526–1528.
- [146] M. Born and R. Oppenheimer, *Annalen der Physik*, **1927**, *84*, 457–484.
- [147] W. L. Jorgensen and J. Tirado-Rives, *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 6665–6670.
- [148] A. D. Mackerell, M. Feig and C. L. Brooks III, *J. Comput. Chem.*, **2004**, *25*, 1400–1415.
- [149] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glättli, P. H. Hünenberger, M. A. Kastenholtz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. A. van der Vegt and H. B. Yu, *Angew. Chem. Int. Ed.*, **2006**, *45*, 4064–4092.
- [150] E. J. Sorin, Y. M. Rhee, M. R. Shirts and V. S. Pande, *J. Mol. Biol.*, **2006**, *356*, 248–256.
- [151] B. Hess and N. F. A. van der Vegt, *J. Phys. Chem. B*, **2006**, *110*, 17 616–17 626.
- [152] M. M. Reif, V. Krutler, M. A. Kastenholtz, X. Daura and P. H. Hünenberger, *J. Phys. Chem. B*, **2009**, *113*, 3112–3128.
- [153] M. Karplus and J. A. McCammon, *Nat. Struct. Mol. Biol.*, **2002**, *9*, 646–652.
- [154] J. Wang, R. Cieplak and P. Kollman, *J. Comput. Chem.*, **2000**, *21*, 1049–1074.
- [155] Y. Duan, C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang and P. Kollman, *J. Comput. Chem.*, **2003**, *24*, 1999–2012.
- [156] V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Struct., Funct., Bioinf.*, **2006**, *65*, 712–725.
- [157] B. Brooks, R. Bruccoleri, B. Olafson, D. States, S. Swaminathan and M. Karplus, *J. Comput. Chem.*, **1983**, *4*, 187–217.

- [158] M. Feig, A. MacKerell and C. Brooks, *J. Phys. Chem. B*, **2003**, *107*, 2831–2836.
- [159] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott and I. G. Tironi, *Biomolecular Simulation: The GROMOS96 manual and user guide*, Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996.
- [160] W. Scott, P. Hünenberger, I. Tironi, A. Mark, S. Billeter, J. Fennen, A. Torda, T. Huber, P. Krüger and W. van Gunsteren, *J. Phys. Chem. A*, **1999**, *103*, 3596–3607.
- [161] C. Oostenbrink, A. Villa, A. E. Mark and W. F. V. Gunsteren, *J. Comput. Chem.*, **2004**, *25*, 1656–1676.
- [162] C. Oostenbrink, T. A. Soares, N. F. A. van der Vegt and W. F. van Gunsteren, *Eur. Biophys. J.*, **2005**, *34*, 273–284.
- [163] W. Jorgensen, D. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, **1996**, *118*, 11 225–11 236.
- [164] G. Kaminski, R. Friesner, J. Tirado-Rives and W. Jorgensen, *J. Phys. Chem. B*, **2001**, *105*, 6474–6487.
- [165] O. Berger, O. Edholm and F. Jähnig, *Biophys. J.*, **1997**, *72*, 2002–2013.
- [166] J. A. McCammon, B. R. Gelin and M. Karplus, *Nature*, **1977**, *267*, 585–590.
- [167] R. B. Best and G. Hummer, *J. Phys. Chem. B*, **2009**, *113*, 9004–9015.
- [168] M. Rueda, C. Ferrer-Costa, T. Meyer, A. Perez, J. Camps, A. Hospital, J. L. Gelpi and M. Orozco, *Proc. Natl. Acad. Sci. USA*, **2007**, *104*, 796–801.
- [169] A. Grossfield and D. Zuckerman, *Annu. Rep. Comput. Chem.*, **2009**, *5*, 23–48.
- [170] R. B. Best, N.-V. Buchete and G. Hummer, *Biophys. J.*, **2008**, *95*, L07–L09.
- [171] D. Matthes and B. de Groot, *Biophys. J.*, **2009**, *97*, 599–608.
- [172] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.*, **2010**, *78*, 1950–1958.
- [173] V. M. Anisimov, G. Lamoureux, I. V. Vorobyov, N. Huang, B. Roux and A. D. MacKerell Jr., *J. Chem. Theory Comput.*, **2005**, *1*, 153–168.

- [174] L. Verlet, *Phys. Rev.*, **1967**, *159*, 98–103.
- [175] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, **2005**, *26*, 1701–1718.
- [176] C. Kutzner, D. van der Spoel, M. Fechner, E. Lindahl, U. W. Schmitt, B. L. de Groot and H. Grubmüller, *J. Comput. Chem.*, **2007**, *28*, 2075–2084.
- [177] B. Hess, C. Kutzner, D. van Der Spoel and E. Lindahl, *J. Chem. Theory Comput.*, **2008**, *4*, 435–447.
- [178] R. Hockney, S. Goel and J. Eastwood, *J. Comput. Phys.*, **1974**, *14*, 148–158.
- [179] K. A. Feenstra, B. Hess and H. J. C. Berendsen, *J. Comput. Chem.*, **1999**, *20*, 786–798.
- [180] J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, **1977**, *23*, 327–341.
- [181] B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, **1997**, *18*, 1463–1472.
- [182] B. Hess, *J. Chem. Theory Comput.*, **2008**, *4*, 116–122.
- [183] M. Christen, C. D. Christ and W. F. van Gunsteren, *ChemPhysChem*, **2007**, *8*, 1557–1564.
- [184] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren and J. Hermans, *Interaction models for water in relation to protein hydration*, D. Reidel Publishing Company, 1981 331–342, 331–342.
- [185] M. W. Mahoney and W. L. Jorgensen, *J. Chem. Phys.*, **2000**, *112*, 8910–8922.
- [186] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, **1983**, *79*, 926.
- [187] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, **1984**, *81*, 3684–3690.
- [188] G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, **2007**, *126*, 014 101.
- [189] S. Nose, *Mol. Phys.*, **1984**, *52*, 255–268.
- [190] W. Hoover, *Phys. Rev. A*, **1985**, *31*, 1695–1697.

## BIBLIOGRAPHY

---

- [191] T. Morishita, *J. Chem. Phys.*, **2000**, *113*, 2976–2982.
- [192] K. Pearson, *Philosophical Magazine*, **1901**, *2*, 559–572.
- [193] R. O. Duda, P. E. Hart and D. G. Stork., *Pattern Classification*, Wiley-Interscience, 2000.
- [194] A. Amadei, A. B. M. Linssen and H. J. C. Berendsen, *Proteins: Struct., Funct., Genet.*, **1993**, *17*, 412–425.
- [195] S. Hayward and N. Go, *Annu. Rev. Phys. Chem.*, **1995**, *46*, 223–250.
- [196] M. Karplus and J. N. Kushick, *Macromolecules*, **1981**, *14*, 325–332.
- [197] A. Kitao and N. Go, *Curr. Opin. Struct. Biol.*, **1999**, *9*, 164–169.
- [198] A. Kitao, F. Hirata and N. Go, *Chem. Phys.*, **1991**, *158*, 447–472.
- [199] S. Hayward and B. L. de Groot, *Methods in molecular biology, Molecular modelling of proteins*, Humana Press, 2008 .
- [200] D. L. Minor and P. S. Kim, *Nature*, **1994**, *367*, 660–663.
- [201] V. Munoz and L. Serrano, *Struct. Biol.*, **1994**, *1*, 399–409.
- [202] W. Kabsch and C. Sander, *Biopolymers*, **1983**, *12*, 2577–2637.
- [203] D. Frishman and P. Argos, *Proteins: Struct., Funct., Genet.*, **1995**, *23*, 566–579.
- [204] W. F. van Gunsteren, J. Dolenc and A. E. Mark, *Curr. Opin. Struct. Biol.*, **2008**, *18*, 149–153.
- [205] P. Ferrara, J. Apostolakis and A. Caflisch, *J. Phys. Chem. B*, **2000**, *104*, 5000–5010.
- [206] A. R. Fersht and V. Daggett, *Cell*, **2002**, *108*, 573–582.
- [207] C. Simmerling, B. Strockbine and A. E. Roitberg, *J. Am. Chem. Soc.*, **2002**, *124*, 11 258–11 259.
- [208] C. D. Snow, H. Nguyen, V. S. Pande and M. Gruebele, *Nature*, **2002**, *420*, 102–106.
- [209] C. D. Snow, B. Zagrovic and V. S. Pande, *J. Am. Chem. Soc.*, **2002**, *124*, 14 548–14 549.
- [210] S. Gnanakaran, H. Nymeyer, J. Portman, K. Y. Sanbonmatsu and A. E. Garcia, *Curr. Opin. Struct. Biol.*, **2003**, *13*, 168–174.

## BIBLIOGRAPHY

---

- [211] S. Honda, K. Yamasaki, Y. Sawada and H. Morii, *Structure*, **2004**, *12*, 1507–1518.
- [212] M. M. Seibert, A. Patriksson, B. Hess and D. van der Spoel, *J. Mol. Biol.*, **2005**, *354*, 173–183.
- [213] D. van der Spoel and M. M. Seibert, *Phys. Rev. Lett.*, **2006**, *96*, 238 102.
- [214] W. Xu, T. Lai, Y. Yang and Y. Mu, *J. Chem. Phys.*, **2008**, *128*, 175 105–175 110.
- [215] J. W. Neidigh, R. M. Fesinmeyer and N. H. Andersen, *Nat. Struct. Mol. Biol.*, **2002**, *9*, 425–430.
- [216] L. Qiu, S. Pabit, A. Roitberg and S. Hagen, *J. Am. Chem. Soc.*, **2002**, *124*, 12 952–12 953.
- [217] R. Zhou, *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 13 280–13 285.
- [218] J. Ulmschneider, M. Ulmschneider and A. Di Nola, *J. Phys. Chem. B*, **2006**, *110*, 16 733–16 742.
- [219] D. J. Lockhart and P. S. Kim, *Science*, **1992**, *257*, 947–951.
- [220] S. Williams, T. Causgrove, R. Gilmanishin, K. Fang, R. Callender, W. Woodruff and R. Dyer, *Biochemistry*, **1996**, *35*, 691–697.
- [221] P. Thompson, W. Eaton and J. Hofrichter, *Biochemistry*, **1997**, *36*, 9200–9210.
- [222] I. Lednev, A. Karnoup, M. Sparrow and S. Asher, *J. Am. Chem. Soc.*, **1999**, *121*, 8074–8086.
- [223] S. Asher, A. Mikhonin and S. Bykov, *J. Am. Chem. Soc.*, **2004**, *126*, 8433–8440.
- [224] A. E. Garcia and K. Y. Sanbonmatsu, *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 2782–2787.
- [225] H. Nymeyer and A. E. Garcia, *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 13 934–13 939.
- [226] E. J. Sorin and V. S. Pande, *Biophys. J.*, **2005**, *88*, 2472–2493.
- [227] E. J. Sorin and V. S. Pande, *J. Comput. Chem.*, **2005**, *26*, 682–690.
- [228] W. DeLano, “The PyMOL Molecular Graphics System”, <http://www.pymol.org>, **2002**.
- [229] E. Lindahl, B. Hess and D. van der Spoel, *J. Mol. Mod.*, **2001**, *7*, 306–317.

## BIBLIOGRAPHY

---

- [230] T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, **1993**, *98*, 10 089–10 092.
- [231] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, **1995**, *103*, 8577–8593.
- [232] I. G. Tironi, R. Sperb, P. E. Smith and W. F. van Gunsteren, *J. Chem. Phys.*, **1995**, *102*, 5451–5459.
- [233] P. E. Smith and W. F. van Gunsteren, *J. Chem. Phys.*, **1994**, *100*, 3169–3174.
- [234] S. Miyamoto and P. A. Kollman, *J. Comput. Chem.*, **1992**, *13*, 952–962.
- [235] G. N. Ramachandran, C. Ramakrishnan and V. Sasisekharan, *J. Mol. Biol.*, **1963**, *7*, 95–99.
- [236] C. Ramakrishnan and G. G. N. Ramachandran, *Biophys. J.*, **1965**, *5*, 909–933.
- [237] S. C. Lovell, I. Davis, W. Arendall III, P. I. W. de Bakker and J. M. Word, *Proteins: Struct., Funct., Bioinf.*, **2003**, *50*, 437–450.
- [238] T. Wang and R. C. Wade, *J. Chem. Theory Comput.*, **2006**, *2*, 140–148.
- [239] L. Monticelli, C. Simoes, L. Belvisi and G. Colombo, *J. Phys.: Condens. Matter*, **2006**, *18*, S329–S345.
- [240] Z. Cao, Z. Lin, J. Wang and H. Liu, *J. Comput. Chem.*, **2008**, *4*, 645–660.
- [241] B. Zagrovic, Z. Gattin, J. K.-C. Lau, M. Huber and W. F. van Gunsteren, *Eur. Biophys. J.*, **2008**, *37*, 903–912.
- [242] J. W. Kelly, *Nat. Struct. Mol. Biol.*, **2000**, *7*, 824–826.
- [243] F. Chiti and C. M. Dobson, *Nat. Chem. Biol.*, **2009**, *5*, 15–22.
- [244] M. Sunde, L. C. Serpell, M. Bartlama, P. E. Frasera, M. B. Pepysa and C. C. F. Blake, *J. Mol. Biol.*, **1997**, *273*, 729–739.
- [245] J. L. Jimenez, E. J. Nettleton, M. Bouchard, C. V. Robinson, C. M. Dobson and H. R. Saibil, *Proc. Natl. Acad. Sci. USA*, **2002**, *99*, 9196–9201.
- [246] J. Berriman, L. C. Serpell, K. A. Oberg, A. L. Fink, M. Goedert and R. A. Crowther, *Proc. Natl. Acad. Sci. USA*, **2003**, *100*, 9034–9038.

## BIBLIOGRAPHY

---

- [247] G. Zandomenoghi, M. R. Krebs, M. G. McCammon and M. Fändrich, *Protein Sci.*, **2004**, *13*, 3314–3321.
- [248] J. Meinhardt, C. Sachse, P. Hortschansky, N. Grigorieff and M. Fändrich, *J. Mol. Biol.*, **2009**, *386*, 869–877.
- [249] T. J. Gibson and R. M. Murphy, *Protein Sci.*, **2006**, *15*, 1133–1141.
- [250] J. Brange, L. Andersen, E. Laursen, G. Meyn and E. Rasmussen, *J. Pharm. Sci.*, **1997**, *86*, 517–25.
- [251] L. Nielsen, R. Khurana, A. Coats, S. Frokjaer, J. Brange, S. Vyas, V. Uversky and A. Fink, *Biochemistry*, **2001**, *40*, 6036–6046.
- [252] A. Ahmad, V. N. Uversky, D. Hong and A. L. Fink, *J. Biol. Chem.*, **2005**, *280*, 42 669–42 675.
- [253] J. Haas, E. Vöhringer-Martinez, A. Bögehold, D. Matthes, U. Hensen, A. Pelah, B. Abel and H. Grubmüller, *ChemBioChem*, **2009**, *10*, 1816–1822.
- [254] P. Tito, E. J. Nettleton and C. V. Robinson, *J. Mol. Biol.*, **2000**, *303*, 267–278.
- [255] M. I. Ivanova, S. A. Sievers, M. R. Sawaya, J. S. Wall and D. Eisenberg, *Proc. Natl. Acad. Sci. USA*, **2009**, *106*, 18 990–18 995.
- [256] B. de Groot, D. van Aalten, R. Scheek, A. Amadei, G. Vriend and H. Berendsen, *Proteins: Struct., Funct., Genet.*, **1997**, *29*, 140–251.
- [257] M. Seeber, M. Cecchini, F. Rao, G. Settanni and A. Caflisch, *Bioinf.*, **2007**, *23*, 2625–2627.
- [258] J. Hartigan and M. Wong, *Applied Statistics*, **1979**, *28*, 100–108.
- [259] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, **2010**.
- [260] A. Likas, N. Vlassis et al., *Pattern Recognition*, **2003**, *36*, 451–461.
- [261] W. Krzanowski and Y. Lai, *Biometrics*, **1988**, *44*, 23–34.
- [262] P. Shannon, A. Markiel, O. Ozier, N. Baliga, J. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome research*, **2003**, *13*, 2498.



- [263] W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics*, **1996**, *14*, 33–38.
- [264] J. Stone, *An efficient library for Parallel Ray Tracing and Animation*, Diplomarbeit, Computer Science Department, University of Missouri-Rolla, **1998**.
- [265] E. Espinosa, E. Molins and C. Lecomte, *Chem. Phys. Lett.*, **1998**, *285*, 170–173.
- [266] P. Jiang, W. Xu and Y. Mu, *PLoS Comput. Biol.*, **2009**, *5*, e1000357.
- [267] G. Wei, A. I. Jewett and J.-E. Shea, *Phys. Chem. Chem. Phys.*, **2010**, *12*, 3622–3629.
- [268] D.-W. Li, L. Han and S. Huo, *J. Phys. Chem. B*, **2007**, *111*, 5425–5433.
- [269] R. Kodali, A. D. Williams, S. Chemuru and R. Wetzol, *J. Mol. Biol.*, **2010**, *401*, 503–517.
- [270] M. Lopez de la Paz, G. M. de Mori, L. Serrano and G. Colombo, *J. Mol. Biol.*, **2005**, *349*, 583–596.
- [271] S. Yan, G. Gawlak, K. Makabe, V. Tereshko, A. Koide and S. Koide, *J. Mol. Biol.*, **2007**, *368*, 230–243.
- [272] N.-V. Buchete, R. Tycko and G. Hummer, *J. Mol. Biol.*, **2005**, *353*, 804–821.
- [273] W. Xue, S. Homans and S. Radford, *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 8926–8931.
- [274] P. Hortschansky, V. Schroeckh, T. Christopeit, G. Zandomenighi and M. Fändrich, *Protein Sci.*, **2005**, *14*, 1753–1759.
- [275] B. Koo, J. Hebda and A. Miranker, *Protein Eng. Des. Sel.*, **2008**, *21*, 147–154.
- [276] A. Ruschak and A. Miranker, *J. Mol. Biol.*, **2009**, *393*, 214–226.
- [277] J. D. Schmit, K. Ghosh and K. Dill, *Biophys. J.*, **2011**, *100*, 450–458.
- [278] X. He, J. Giurleo and D. Talaga, *J. Mol. Biol.*, **2010**, *395*, 134–154.
- [279] K. Cho, Y. Huang, S. Yu, S. Yin, M. Plomp, S. Qiu, R. Lakshminarayanan, J. Moradian-Oldak, M. Sy and J. De Yoreo, *J. Am. Chem. Soc.*, **2011**, *133*, 8586–8593.
- [280] E. Frare, M. F. Mossuto, P. P. de Laureto, S. Tolin, L. Menzer, M. Dumoulin, C. M. Dobson and A. Fontana, *J. Mol. Biol.*, **2009**, *387*, 17–27.

## BIBLIOGRAPHY

---

- [281] J. Kaylor, N. Bodner, S. Edridge, G. Yamin, D. Hong and A. Fink, *J. Mol. Biol.*, **2005**, 353, 357–372.
- [282] Y. Liang, D. G. Lynn and K. M. Berland, *J. Am. Chem. Soc.*, **2010**, 132, 6306–6308.
- [283] P. Walsh, P. Neudecker and S. Sharpe, *J. Am. Chem. Soc.*, **2010**, 132, 7684–7695.
- [284] A. Sandberg, L. Luheshi, S. Sölvander, T. Pereira de Barros, B. Macao, T. Knowles, H. Biverstal, C. Lendel, F. Ekholm-Petterson, A. Dubnovitsky, L. Lannfelt, C. Dobson and T. Härd, *Proc. Natl. Acad. Sci. USA*, **2010**, 107, 15 595–15 600.
- [285] D. Matthes, V. Gapsys, V. Daebel and B. L. de Groot, *PLoS ONE*, **2011**, 6, e19 129.
- [286] J. Gsponer and M. Vendruscolo, *Protein & Peptide Letters*, **2006**, 13, 287–293.
- [287] J. Nasica-Labouze, M. Meli, P. Derreumaux, G. Colombo and N. Mousseau, *PLoS Comput. Biol.*, **2011**, 7, e1002 051.
- [288] D. Klimov, J. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, **2004**, 101, 14 760–14 765.
- [289] G. Reddy, J. Straub and D. Thirumalai, *Proc. Natl. Acad. Sci. USA*, **2010**, 107, 21 459–21 464.
- [290] F. Quijano, D. Morrow, B. Wise, F. Brancia and W. Goux, *Biochemistry*, **2006**, 45, 4638–4652.
- [291] M. Inoue, A. Hirata, K. Tainaka, T. Morii and T. Konno, *Biochemistry*, **2008**, 47, 11 847–11 857.
- [292] B. Vekhter and R. S. Berry, *J. Chem. Phys.*, **1999**, 110, 2195–2201.
- [293] B. Ripley, *Pattern recognition and neural networks*, Cambridge University Press, 1996.
- [294] W. Venables and B. Ripley, *Modern applied statistics with S*, Springer verlag, 2002.
- [295] C. Ding and X. He, in “Proceedings of the twenty-first international conference on Machine learning”, ACM, 2004 .
- [296] S. Arya, D. Mount, N. Netanyahu, R. Silverman and A. Wu, *J. ACM*, **1998**, 45, 891–923.
- [297] J. Hub and B. de Groot, *PLoS Comput. Biol.*, **2009**, 5, e1000 480.

- [298] H. Wold, *Multivariate analysis*, **1966**, 1, 391–420.
- [299] H. Wold, *Quantitative Sociology: International perspectives on mathematical and statistical model building*, **1975**, 307–357.
- [300] D. Eisenberg and A. D. McLachlan, *Nature*, **1986**, 319, 199–203.
- [301] D. Gfeller, P. De Los Rios, A. Caflisch and F. Rao, *Proc. Natl. Acad. Sci. USA*, **2007**, 104, 1817–1822.
- [302] H. Stamati, C. Clementi and L. Kavraki, *Proteins*, **2010**, 78, 223–235.
- [303] A. Vitalis, N. Lyle and R. Pappu, *Biophys. J.*, **2009**, 97, 303–311.
- [304] M. Krone, L. Hua, P. Soto, R. Zhou, B. Berne and J. Shea, *J. Am. Chem. Soc.*, **2008**, 130, 11066–11072.
- [305] L. Vitagliano, F. Stanzione, A. De Simone and L. Esposito, *Biopolymers*, **2009**, 91, 1161–1171.
- [306] P. Nguyen, M. Lid and P. Derreumaux, *Phys. Chem. Chem. Phys.*, **2011**, 13, 9778–9788.
- [307] N. Sgourakis, Y. Yan, S. McCallum, C. Wang and A. Garcia, *J. Mol. Biol.*, **2007**, 368, 1448–1457.
- [308] X. Periole, A. Rampioni, M. Vendruscolo and A. Mark, *J. Phys. Chem. B*, **2009**, 113, 1728–1737.
- [309] A. Villa and A. E. Mark, *J. Comput. Chem.*, **2002**, 23, 548–553.
- [310] J. Lee, E. K. Culyba, E. T. Powers and J. W. Kelly, *Nat. Chem. Biol.*, **2011**, 7, 602–609.
- [311] C. Bleiholder, N. Dupuis, T. Wyttenbach and M. Bowers, *Nat. Chem.*, **2011**, 3, 172–177.
- [312] S. Auer, F. Meersman, C. M. Dobson and M. Vendruscolo, *PLoS Comput. Biol.*, **2008**, 11, e1000222.
- [313] K. Ravikumar and W. Hwang, *J. Am. Chem. Soc.*, **2011**, 133, 11766–11773.
- [314] J.-E. Suk, S. B. Lokappa and T. S. Ulmer, *Biochemistry*, **2010**, 49, 1533–1540.
- [315] H. Scheidt, I. Morgado, S. Rothmund, D. Huster and M. Fändrich, *Angew. Chem. Int. Ed.*, **2011**, 50, 2837–2840.

- [316] G. Hummer, *Nat. Chem.*, **2010**, 2, 906–907.
- [317] J. F. Nagle and S. Tristram-Nagle, *Biochim. Biophys. Acta.*, **2000**, 1469, 159–195.
- [318] T. Ulmer, A. Bax, N. B. Cole and R. L. Nussbaum, *J. Biol. Chem.*, **2005**, 280, 9595–9603.
- [319] J. Knight and A. Miranker, *J. Mol. Biol.*, **2004**, 341, 1175–1187.
- [320] M. Engel, L. Khemtёмourian, C. Kleijer, H. Meeldijk, J. Jacobs, A. Verkleij, B. de Kruijff, J. Killian and J. Höppener, *Proc. Natl. Acad. Sci. USA*, **2008**, 105, 6033–6038.
- [321] M. Zhu and A. L. Fink, *J. Biol. Chem.*, **2003**, 278, 16 873–16 877.
- [322] A. L. Fink, *Acc. Chem Res.*, **2006**, 39, 628–634.
- [323] W. Norton, T. Abe, S. Poduslo and G. DeVries, *J. Neurosci. Res.*, **1975**, 1, 54–63.
- [324] K. Beyer, *Cell Biochem. Biophys.*, **2007**, 47, 285–299.
- [325] L. C. Serpell, J. Berriman, R. Jakes, M. Goedert and R. A. Crowther, *Proc. Natl. Acad. Sci. USA*, **2000**, 97, 4897–4902.
- [326] M. Sandal, F. Valle, I. Tessari, S. Mammi, E. Bergantino, F. Musiani, M. Brucale, L. Bubacco and B. Samori, *PLoS Biol.*, **2008**, 6, e6.
- [327] K. A. Conway, J. D. Harper and P. T. Lansbury, *Nat. Med.*, **1998**, 4, 1318–1320.
- [328] L. Narhi, S. J. Wood, S. Steavenson, Y. Jiang, G. M. Wu, D. Anafi, S. A. Kaufman, F. Martin, K. Sitney, P. Denis, J.-C. Louis, J. Wypych, A. L. Biere and M. Citron, *J. Biol. Chem.*, **1999**, 274, 9843–9846.
- [329] A. Kapurniotu, *Biopolymers*, **2001**, 60, 438–459.
- [330] G. Cooper, *Endocr. Rev.*, **1994**, 15, 163–201.
- [331] P. Westermarck, C. Wernstedt, E. Wilander, D. Hayden, T. O'Brien and K. Johnson, *Proc. Natl. Acad. Sci. USA*, **1987**, 84, 3881–3885.
- [332] A. Kajava, U. Aebi and A. Steven, *J. Mol. Biol.*, **2005**, 348, 247–252.
- [333] J. J. Wiltzius, S. A. Sievers, M. R. Sawaya, D. Cascio, D. Popov, C. Riekel and D. Eisenberg, *Protein Sci.*, **2008**, 17, 1467–1474.

## BIBLIOGRAPHY

---

- [334] P. Westermarck, U. Engstrom, K. Johnson, G. Westermarck and C. Betsholtz, *Proc. Natl. Acad. Sci. USA*, **1990**, *87*, 5036–5040.
- [335] C. Goldsbury, K. Goldie, J. Pellaud, J. Seelig, P. Frey, S. Muller, J. Kistler, G. Cooper and U. Aebi, *J. Struct. Biol.*, **2000**, *130*, 352–362.
- [336] L. D. Schuler, X. Daura and W. F. van Gunsteren, *J. Comput. Chem.*, **2001**, *22*, 1205–1218.
- [337] E. Lindahl and O. Edholm, *Biophys. J.*, **2000**, *79*, 426–433.
- [338] J. L. MacCallum, W. D. Bennett and D. P. Tieleman, *J. Gen. Physiol.*, **2007**, *129*, 371–377.
- [339] C. H. Davis and M. L. Berkowitz, *Biophys. J.*, **2009**, *96*, 785–797.
- [340] A. Nikolic, S. Baud, S. Rauscher and R. Pomes, *Proteins: Struct., Funct., Bioinf.*, **2011**, *79*, 1–22.
- [341] P. Kasson, E. Lindahl and V. Pande, *J. Am. Chem. Soc.*, **2011**, *133*, 3812–3815.