Copyright Warning & Restrictions

The copyright law of the United States (Title 17, United States Code) governs the making of photocopies or other reproductions of copyrighted material.

Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be "used for any purpose other than private study, scholarship, or research." If a, user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of "fair use" that user may be liable for copyright infringement,

This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve violation of copyright law.

Please Note: The author retains the copyright while the New Jersey Institute of Technology reserves the right to distribute this thesis or dissertation

Printing note: If you do not wish to print this page, then select "Pages from: first page # to: last page #" on the print dialog screen

ABSTRACT

AMYLOID PROTEINS AND FIBRILS STABILITY

by
Farbod Mahmoudinobar

Compared to globular proteins that have a stable native structure, intrinsically disordered peptides (IDP) sample an ensemble of structures without folding into a native conformation. One example of IDP is the amyloid-β (Aβ) protein which is the main constituent of senile plaques in the brain of Alzheimer's patients. Understanding the process by which IDPs undergo structural changes to form oligomers that eventually aggregate into senile plaques/amyloid fibrils may significantly advance the development of novel therapeutic methods to treat neurodegenerative diseases, for which there is no cure to date. This dissertation has two main objectives. The first one is to investigate and identify structural conformations of Aβ monomer which are precursor to aggregation. The second objective is to understand the underlying mechanisms of amyloid fibril stability using atomistic molecular dynamics simulations in explicit water.

The aggregation of Aβ peptides into amyloid fibrils in Alzheimer's patients depends on the spectrum of conformations adopted by monomers of this peptides. These conformations are strongly affected by properties of the aqueous environment. In the first part of this dissertation, conformations of Aβ in environments that promote and inhibit fibril formation are studied. Micro-second Replica Exchange Molecular Dynamics (REMD) simulations are performed for that purpose. A comparative study of the set of conformations in each environment is performed using contact maps, cluster analysis and by studying the network of the backbone hydrogen bonds of Aβ. A specific in-register strand-loop-strand conformation is found in the environment that promotes fibril formation, which is not observed in environments that inhibit fibril formation. It is proposed here that this conformation

may act as intermediate structure in fibril formation. Inhibiting the formation of this conformation might be helpful in developing drugs for Alzheimer's disease.

In the second part of this dissertation, the molecular mechanisms of amyloid fibril stability are investigated using a thermodynamic framework. Understanding the atomic interactions responsible for fibril stability may be useful in designing novel therapeutic methods to disrupt fibrils and plaques in neurodegenerative diseases. However, despite numerous studies on amyloid fibrils, a thorough understanding of fibril stability is still missing. A combination of enhanced sampling methods is used to simulate all-atom models in explicit solvent in order to investigate the stability of non-polar ($A\beta_{16-21}$) and polar ($IAPP_{28-33}$) amyloid fibrils. Umbrella sampling is performed jointly with replica exchange molecular dynamics to determine the free energy of peptide addition to a pre-formed amyloid fibril at different temperatures. Results from these simulations show that the non-polar fibril becomes more stable with increasing temperature and its stability is dominated by entropy. In contrast, the polar fibril becomes less stable as temperature increases while it is stabilized by enthalpy. These findings suggest that the stability of fibrils can be customized by the choice of amino acid sequence in the dry core of the amyloid fibrils, e.g., proteins can be modified to transition between fibril and monomer state at a designated temperature. Such fine-tuned amyloid fibrils can be used as scaffolds for drug delivery and other biomaterials.

AMYLOID PROTEINS AND FIBRILS STABILITY

by
Farbod Mahmoudinobar

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology and
Rutgers, The State University of New Jersey – Newark
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Applied Physics

Department of Physics, NJIT

December 2019

APPROVAL PAGE

AMYLOID PROTEINS AND FIBRILS STABILITY

Farbod Mahmoudinobar

Dr. Cristiano L. Dias, Dissertation Advisor — Date
Associate Professor of Physics, NJIT

Dr. Treena L. Arinzeh, Committee Member — Date
Professor of Biomedical Engineering, NJIT

Dr. Edgardo T. Farinas, Committee Member — Date
Associate Professor of Chemistry and Environmental Sciences, NJIT

Dr. Gordon A. Thomas, Committee Member — Date
Professor of Physics and Biomedical Engineering, NJIT

Dr. Tao Zhou, Committee Member — Date
Associate Professor of Physics, NJIT

BIOGRAPHICAL SKETCH

Author: Farbod Mahmoudinobar

Degree: Doctor of Philosophy

Date: December 2019

Undergraduate and Graduate Education:

- Doctor of Philosophy in Applied Physics,
  New Jersey Institute of Technology, Newark, NJ, 2019
- Master of Science in Biomedical Engineering,
  Amirkabir University of Technology, Tehran, Iran, 2013
- Bachelor degree of Science in Biomedical Engineering,
  Amirkabir University of Technology, Tehran, Iran, 2010

Major: Applied Physics

Publications and Presentations:

F. Mahmoudinobar, Z. Su and C. L. Dias, "Thermodynamic stability of amyloid fibrils: effects of the hydropathy of the peptide sequence", *Journal of Chemical Theory and Computation*, 15(6):3868-3874, 2019.

F. Mahmoudinobar and C. L. Dias, "GRADE: A code to determine clathrate hydrate structures", *Computer Physics Communications*, 244:385 − 391, 2019.

R. Omidvar, M. Tafazzoli-Shadpour, F. Mahmoudinobar, S. Azadi, and MM. Khani, "Quantifying effects of cyclic stretch on cell–collagen substrate adhesiveness of vascular endothelial cells", *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 232(5), pp.531-541, 2018

Z. Su, F. Mahmoudinobar and C. L. Dias, "Effects of TMAO on the conformation of peptides and its implications for small proteins," *Physical Review Letters*, 119:108102, 2017.

F. Mahmoudinobar, C. L. Dias and R. Zangi, "Role of side chain interactions on the formation of α-helices in model peptides", *Physical Review E*, 91, 032710, 2015

Srinivasa R. Jampani, F. Mahmoudinobar, Z. Su and C. L. Dias, "Thermodynamics of $A\beta_{16-21}$ dissociation from a fibril: Enthalpy, entropy, and volumetric properties," *Proteins: Structure, Function, and Bioinformatics*, 83:1963, 2015.

F. Mahmoudinobar, American Physical Society (APS) March Meeting, Boston, MI, 2019 (Oral presentation)

F. Mahmoudinobar, Fourth Annual Rutgers Brain Health Institute Symposium, Nokia-Bell Labs, Murray Hill, NJ 2018 (Poster)

F. Mahmoudinobar, Molecular Simulations Mini-workshop, NJIT, Newark, NJ, 2018 (Oral Presentation)

F. Mahmoudinobar, Biophysical Society (BPS) 62nd Annual Meeting, San Francisco, CA, 2018 (Poster)

F. Mahmoudinobar, Annual meeting of American Physical Society (APS) Mid-Atlantic section, Newark, NJ, 2017 (Oral Presentation)

F. Mahmoudinobar, Biophysical Society (BPS) 61st Annual Meeting, New Orleans, LA, 2017 (Poster)

F. Mahmoudinobar, Annual Research Computing Fair, Rutgers University, Newark, NJ, 2017 (Poster)

F. Mahmoudinobar, Second Annual Symposium on Research and Teaching Using IST ARCS-Managed High-Performance Computing (HPC), Big Data (BD), and Database Resources, NJIT, Newark, NJ, 2017 (Oral Presentation)

F. Mahmoudinobar, American Physical Society (APS) March Meeting, Baltimore, MD, 2016 (Oral Presentation)

F. Mahmoudinobar, Johnson & Johnson Engineering Showcase, New Brunswick, NJ, 2015 (Poster)

F. Mahmoudinobar, Biophysical Society (BPS) 59th Annual Meeting, Baltimore, MD, 2015 (Poster)

F. Mahmoudinobar, Dana Knox Student Research Showcase, New Jersey Institute of Technology, Newark, NJ, 2014-2018 (Poster)

To my parents, *Behzad* and *Laleh*.

If civilization is to survive, we must cultivate the science of human relationships - the ability of all peoples, of all kinds, to live together, in the same world at peace.

Franklin D. Roosevelt

ACKNOWLEDGMENT

Throughout the course of my Ph.D., I have received a great deal of support and guidance. It is a pleasure for me to acknowledge all those who have supported me in one way or another during the past six years.

First and foremost, I want to offer my sincerest gratitude to my adviser and mentor, Dr. Cristiano L. Dias, who has supervised and guided me on the path to becoming a researcher. I am inspired by his passion for science, extensive knowledge and relentless support. With his valuable insights and continuous guidance he was my role model during my years at NJIT. I am forever indebted to him for being such an outstanding mentor.

I would like to acknowledge Dr. Gordon A. Thomas, Dr. Tao Zhou, Dr. Treena L. Arinzeh and Dr. Edgardo T. Farinas for taking the time and effort to serve as my Ph.D. committee members. They have provided valuable feedback which assisted me in completion of this dissertation. I am grateful to my labmate, colleague and friend, Dr. Zhaoqian Su, with whom I had frequent discussions and passionate conversations. A special thanks goes to my classmates, Carley Nicoletti and David Apigo. I cherish our friendship as well as the many study sessions we had together, without them passing the qualifying exam would have been a harder challenge. In addition, I would like to thank everyone in Physics Department at NJIT, in particular, Christine A. Oertel, Cindy Montalvo-Harden and Leslie M. Williams for their administrative help.

I want to dedicate this work to my incredible family, Behzad Mahmoudinobar, my mother Laleh and my brother Farzad. These wonderful people have provided both moral and practical support every day of my doctorate degree. Last but definitely not the least, I want to thank my girlfriend, Sharareh Jalali for her unconditional support, encouragement and being an unwavering inspiration to me.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

Chapter **Page**

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

Proteins are the second most abundant molecules in human body. Some proteins (antibodies) bind to viruses to mark them for destruction, and others (enzymes) function as catalysts of chemical reactions [5]. In order to perform these functions, they need to fold into a specific native structure. As a result of weak non-covalent bonds, proteins tend to fold into a stable three-dimensional conformation which minimizes the free energy of the system [220]. Folding-misfolding process can, however, be disrupted by mutations and other biological factors. This leads to protein misfolding which may result in unwanted aggregation in the body.

The aggregates formed by some proteins are associated with neurodegenerative diseases including Alzheimer's, Parkinson's and Huntington's [13, 26, 55]. The main constituents of the aforementioned fibrils are amyloid-β, α-synuclein and polyglutamine proteins, which are a class of proteins called intrinsically disordered peptides (IDP) which lack a native structure [13, 26, 55]. Because of their role in neurodegenerative diseases, protein aggregation has been the subject of numerous studies in recent years. Understanding the process in which IDPs undergo structural changes, form oligomers and eventually result in fibril, is a topic of great interest to the biophysical community. The main focus of this dissertation is to understand the atomic forces responsible for the stability of amyloid fibrils as well as investigating the effect of small molecules on conformations of amyloid proteins. This knowledge will play a crucial role in designing and manufacturing new methods to inhibit or cure amyloid diseases.

This dissertation is organized into three chapters and two Appendices. In Chapter 1, an introduction to proteins structure, protein folding and thermodynamics,

amyloid fibril formation and molecular dynamics simulations are discussed. Chapter 2 describes the effect of small compounds on the structure of Aβ_{42} monomers and introduces a newly found intermediate structure which may be a precursor to aggregation. Inhibiting the formation of this structure might be helpful in developing drugs for Alzheimer's disease. Chapter 3 provides a thermodynamical analysis of the stability of amyloid fibrils in which we show that polar amyloid fibrils are stabilized by enthalpy and non-polar amyloid fibrils are stabilized by entropic energy [154]. Chapter 4 provides a summary of this dissertation and proposes future research direction. In Appendix A, the effect of side chain interactions on the formation of protein secondary structures is investigated and length scales promoting α-helices are determined. Appendix B describes an open-source C++ program which analyzes the trajectory of a molecular dynamics simulation and determines the evolution of water clathrates in the system. Clathrates are hydrogen bonded network of water molecules that form in water solutions at low temperatures and high pressure and have a critical role in designing and manufacturing anti-freeze proteins as well as some applications in gas and oil industry.

1.1 Proteins

Proteins are complex molecules made by amino acids connected to each other through covalent bonds known as the peptide bond. There are 20 types of amino acids which are made of four groups connected to a central C_α atom as shown in Figure 1.1a. These groups are the amino group, carboxyl group, hydrogen atom and side chain. The amino group, C_α and carboxyl group make the backbone of the protein. What differentiates the amino acids from each other, is their side chain group which can be hydrophobic, polar or charged. Figure 1.1b shows these amino acids connected to each other through peptide bonds.

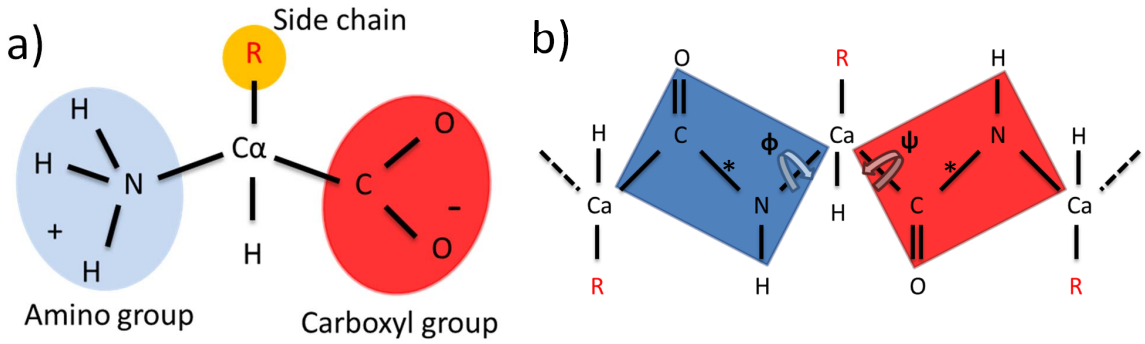


Figure 1.1 Proteins and amino acids. (a) Chemical structure of an amino acid. (b) Torsion angles ϕ and ψ in a polypeptide chain. Peptide bonds connecting the two amino acids are marked with an asterisk.

Torsion angles of backbone atoms are commonly used to describe conformations of the polypeptide chain. Rotations around $N - C_\alpha$, $C_\alpha - C$ and $C - N$ are known as phi (ϕ), psi (ψ) and omega (ω), respectively as shown in Figure 1.1. The ω torsion angle is fixed at 180 degrees due to the partial double-bond character of the peptide bond. Hence, knowing the ϕ and ψ angles of all amino acids in a polypeptide chain defines the three-dimensional structure of the backbone. Because of steric constraints, only specific combinations of these angles are allowed in proteins [29, 220]. However, folding of a protein is further constrained by many other interactions and noncovalent bonds between peptide backbone and side chains which will be discussed in following

sections. The angles ϕ and ψ can be plotted against each other in a diagram called Ramachandran plot. Secondary structures of a protein appear in different regions of the Ramachandran plot. For example, the ($\phi \sim -100$, $\psi \sim 140$) region corresponds to β strands and ($\phi \sim -100$, $\psi \sim -30$) corresponds to α helices. An example of Ramachandran plot is shown in Figure 1.2 illustrating the allowed values of ϕ and ψ angles.

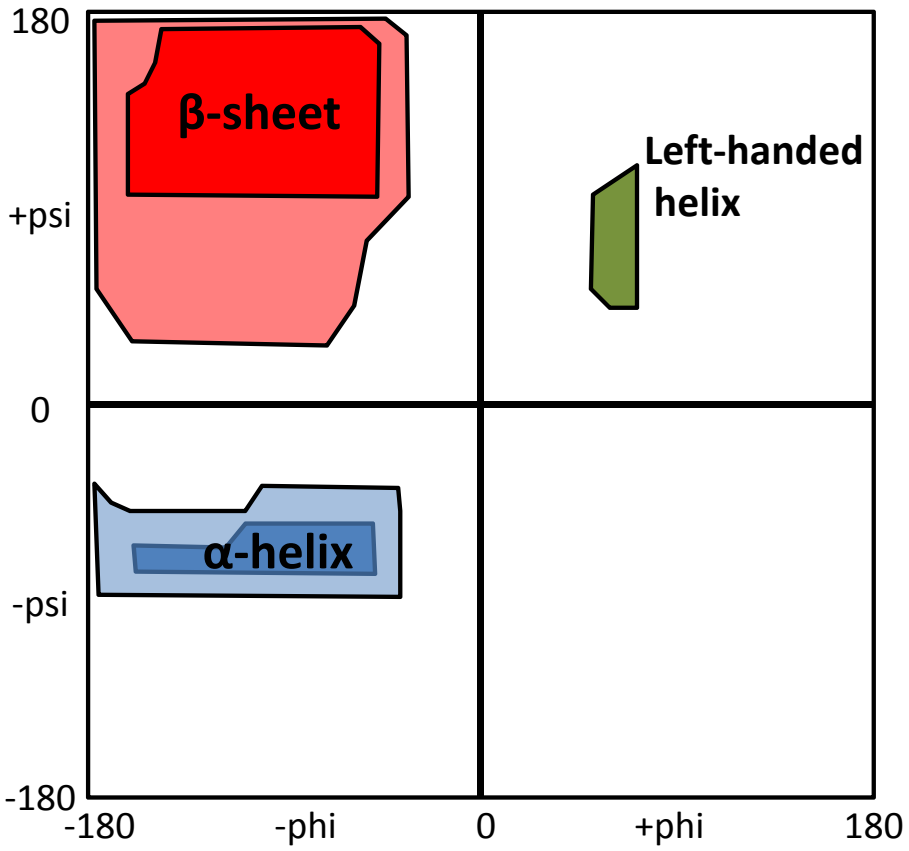


Figure 1.2 Ramachandran plot showing allowed conformations of a Phenyl alanine in a polypeptide. Red, blue and green areas correspond approximately to β sheets, right-handed α helices and close to left-handed α helices.

1.1.1 Protein Structure

In biological studies, protein structure is divided into four levels: primary, secondary, tertiary and quaternary structures. The sequence of amino acids comprising a protein

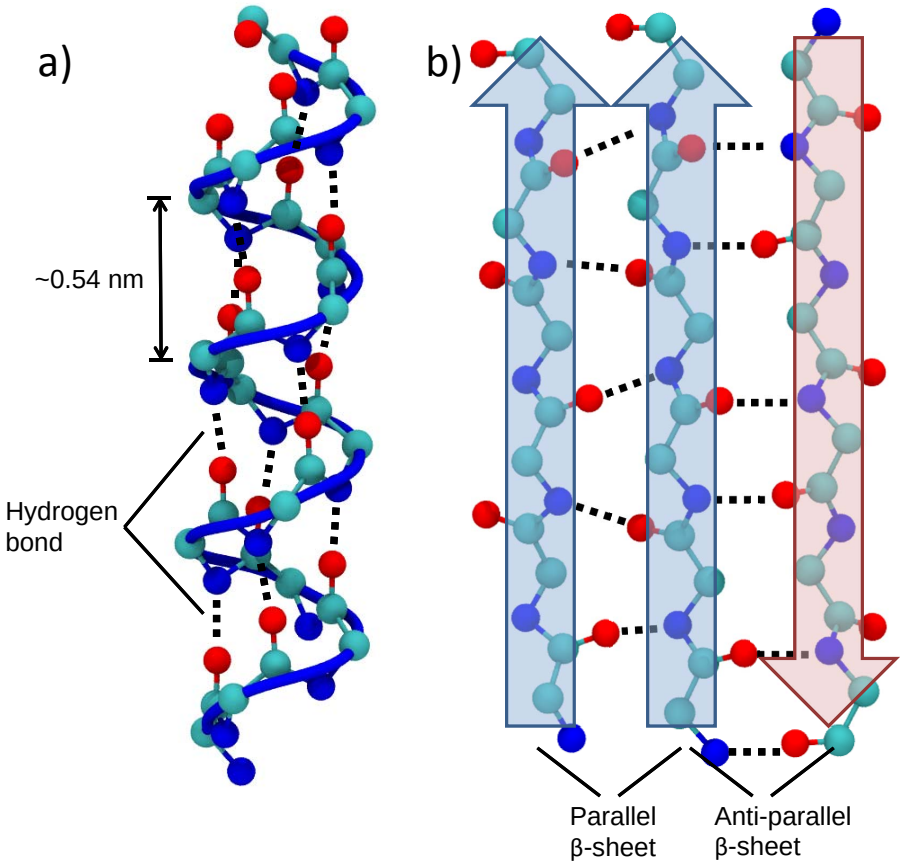


Figure 1.3 Protein secondary structures. (a)α-helix and (b)β-sheet are the most common secondary structures. N, C and O atoms are shown in blue, cyan and red colors respectively. Dashed lines represent hydrogen bonds and hydrogen atoms are removed for simplicity. α-helix and β-sheets are shown with tube and arrows.

molecule is the primary structure. The two most important secondary structures in proteins are α-helices and β-sheets [220]. In an α-helix, the amino group of an amino acid i forms a hydrogen bond with the carbonyl group of amino acid i+4. Structurally, an α-helix has a pitch height of $\sim$ 0.54 nm which includes $\sim$ 3.6 residues in a complete turn as shown in Figure 1.3a. The β-sheet is made of fully extended regions called β-strands. Beta strands can hydrogen bond with each other in two ways to form a sheet. The amino acids in the aligned β-strands can run in the same biochemical direction (from N to C terminal), in which the sheet is called parallel, or the amino

acids in successive strands can have alternating directions, in which the sheet is called anti-parallel [29, 220] as shown in Figure 1.3.

The three-dimensional conformation of a protein is called tertiary structure and if a particular protein molecule is formed by more than one chain, the whole complex is called quaternary structure.

1.2 Molecular Interactions

Bonds between two particles can be defined by a *pair potential*, $u(r)$, which describes the energy as a function of distance (r) between the two particles. The derivative of pair potential is the force, $f(r)$, between two particles [67]:

$$f(r) = -\frac{du(r)}{dr} \tag{1.1}$$

In this section, a brief description of the important bonds in protein chemistry is provided.

1.2.1 Covalent Bonds

Covalent bonds are formed when two atoms share electrons with each other. It can be a single bond (two electrons are shared), double bond (two pairs of electrons are shared) or a triple bond (three pairs of shared electron). Covalent bonds are 30-300 times stronger than non-covalent bonds. The energy to break them varies from $\sim$125 to over 800 kJ/mol [5, 88]. Since amino acids are made of covalently bonded atoms, at normal body temperature ($\sim$2.5 k_BT) the amino acid sequence will not break up due to thermal energy.

1.2.2 Non-Covalent Bonds

Hydrogen Bonds are important in stabilizing protein secondary structures. This bond ($8 - 30$ kJ/mol) is formed when a hydrogen atom, which is covalently bonded

to an electronegative atom, is close to a second electronegative atom [159]. Hydrogen bonds in proteins are formed between Nitrogen and Oxygen atoms whereas in water molecules, they are formed between Oxygen atoms of neighboring water molecules. The latter is illustrated in Figure 1.4. Thermal energy can easily break up hydrogen bonds. Changes in secondary structures of proteins happen as a result of this phenomenon. Hydrogen bonds are directional and short-ranged as they can only form at distances of ~ 3.5 Å and over a restricted range of angles ($< 30°$).

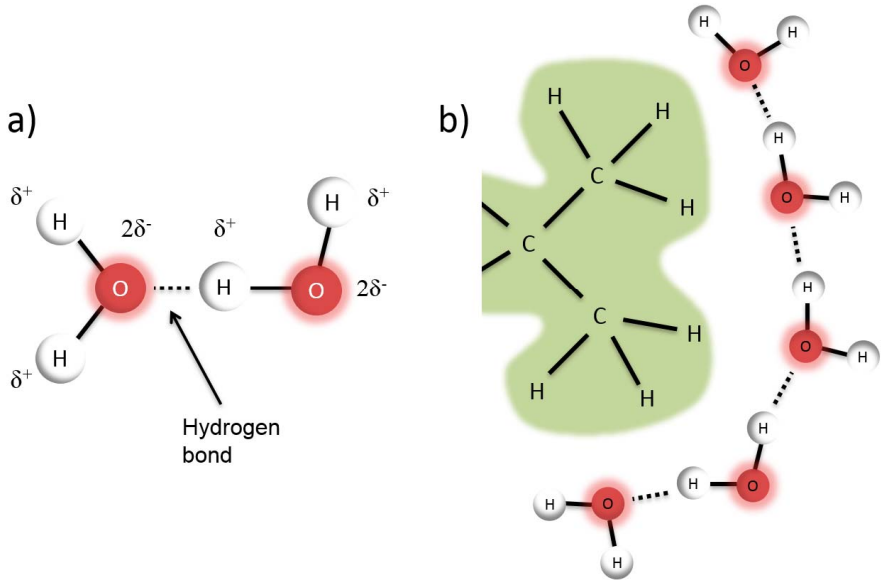


Figure 1.4 Hydrogen bond and hydrophobic effect. (a) Polarized water molecules forming hydrogen bond and, (b) solvation of a hydrophobic protein in water. Water molecules try to keep their hydrogen bonded network.

Electrostatic Bonds are due to interactions between charged particles. In vacuum, the energy of such interaction between two particles with charges q_1 and q_2 is:

$$u_e(r) = \frac{1}{4\pi\epsilon_0}\frac{q_1 q_2}{r}, \tag{1.2}$$

where ϵ_0 is the permittivity of vacuum and r is the distance between two charges. Electrostatic bonds are long-range interactions and their energy diminish with distance between particles as r^{-1}. In proteins the interaction between two oppositely charged residues that are sufficiently close to each other to experience electrostatic attraction, is called a *salt bridge* [27]. For example, in amyloid-β protein residues Asp23 and Lys28 form a salt bridge which is believed to have significant influence on the structural conformations of Aβ [207, 255].

Van der Waals Bonds The basis of van der Waals interaction is that the distribution of the electron cloud of an atom changes when two atoms are close to each other. This phenomenon induces a transient dipole that leads to a net attraction between these atoms. Energies associated with van der Waals interactions are small, typically in the order of 2-4 kJ/mol. The attractive part of the Lennard-Jones potential is used to describe the van der Waals interactions as a function of distance.

Hydrophobic Effect One of the most important driving forces in protein folding is the hydrophobic effect ($\sim$ 5-10 kJ/mol [29]), which loosely speaking, can be described as the tendency of nonpolar residues to avoid water molecules and interact with each other. This interaction emerges from the lower entropy of more ordered water molecules surrounding a nonpolar molecule compared to the higher entropy of bulk water. As a result, nonpolar molecules in water tend to aggregate in order to reduce their exposure to water, thus, minimizing the number of ordered water molecules. The hydrophobic effect in proteins makes the nonpolar residues to collapse and transition from extended structures to a globular conformation as shown in Figure 1.5.

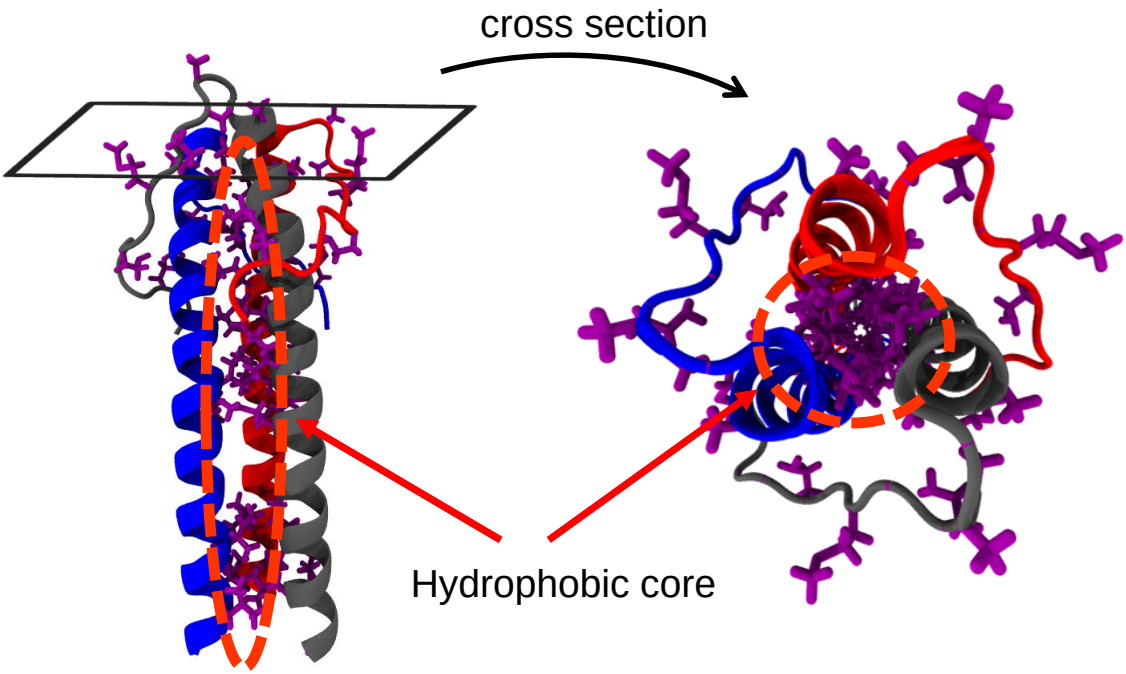


Figure 1.5 Part of the structure of the adaptor protein CIN85 (PDB:2N64) and its nonpolar residues. CIN85 protein regulates signaling from a number of cell surface receptors, including growth factor receptor and antigen receptors on lymphocytes. Hydrophobic residues are shown in magenta using Licorice representation, and three chains of the molecule are illustrated in blue, gray and red. Hydrophobic core of the protein is shown by dashed lines.
source: [134]

1.3 Thermodynamics of Protein Folding

Proteins can adopt ordered structures known as folded (native) state or disordered conformations called unfolded state. The folded state is more stable and has a lower Gibbs free energy whereas the unfolded state has more extended structures and higher free energy. Native and unfolded states exist in thermodynamic equilibrium at a given temperature, wherein the dissociation constant K can be described by the the ratio of the population of unfolded [U] and native [N] states [7]:

$$K = \frac{[U]}{[N]}. \tag{1.1}$$

K values greater than one imply that the equilibrium will favor the unfolded state and values less than one imply that the equilibrium favors the native state. At equilibrium, there is a relationship between the equilibrium constant and the change in free energy which is given by the equation:

$$\Delta G = -RT \ln K, \tag{1.2}$$

where R is the universal gas constant and T is the absolute temperature in Kelvin. To put it simply, the equation 1.2 describes the dissociation constant in terms of the free energy. For example, if the dissociation constant is equal to one, the change in free energy is zero and the population of unfolded and folded states are the same. If dissociation constant is greater than one, the change in free energy is negative, the unfolding process is favorable and it will be accomplished spontaneously. At a given temperature (T) the Gibbs free energy between folded and unfolded states is given by [67, 97, 199, 245, 280]:

$$\Delta G = \Delta H - T\Delta S. \tag{1.3}$$

The temperature dependence of ΔH and ΔS can be described in terms of their values at the transition temperature (T_o). Whenever the change in heat capacity is constant they become:

$$\Delta H(T) = \Delta H_o + \int_{T_o}^{T} \Delta C_{op}\, dT = \Delta H_o + \Delta C_{op}(T - T_o), \tag{1.4}$$

$$\Delta S(T) = \Delta S_o + \int_{T_o}^{T} \frac{\Delta C_{op}(T)}{T}\, dT = \Delta S_o + \Delta C_{op}\ ln\left(\frac{T}{T_o}\right), \tag{1.5}$$

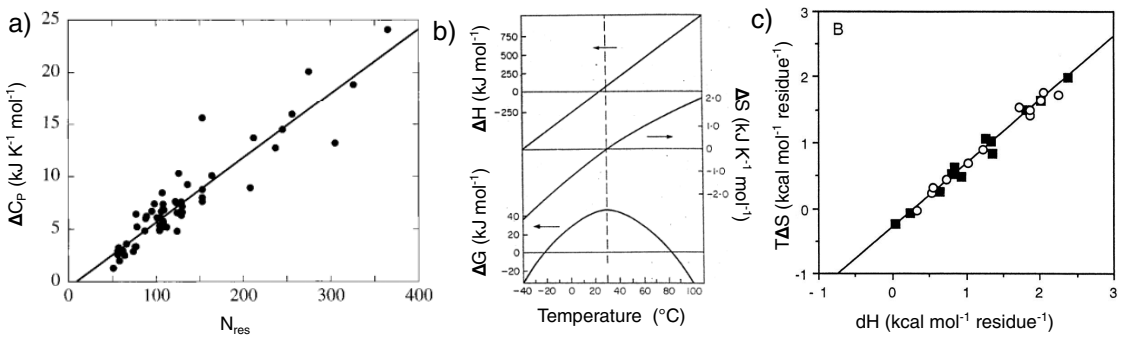


Figure 1.6 Thermodynamics of protein folding and unfolding transitions. (a) Relation between changes in heat capacity of unfolding (ΔC_p) and the number of residues (N_{res}) of various globular protein. (b) Changes in Gibbs free energy (ΔG), enthalpy (ΔH) and entropy (ΔS) upon unfolding of metmyoglobin in 100 mM-glycine buffer (pH 10.0) as a function of temperature. (c) Compensation of enthalpy and entropy upon folding of various globular proteins.
source: [198, 213, 235]

where ΔH_o, ΔS_o and ΔC_{op} are changes in enthalpy, entropy and heat capacity at the transition temperature T_o and constant pressure. Equations 1.3 to 1.5 lead to the temperature dependence of the Gibbs free energy:

$$\Delta G(T) = \Delta H_o - T\Delta S_o + \Delta C_{op}\left[(T - T_o) - Tln\left(\frac{T}{T_o}\right)\right]. \qquad (1.6)$$

The unfolding of globular proteins is characterized by a positive and large ΔC_{op} which has been associated with the solvation of non-polar residues that are buried in the native state and become exposed to the solvent upon unfolding [53,59,62,160,190]. Accordingly, the magnitude of ΔC_{op} was shown to increase with the number of non-polar residues in the system (See Figure 1.6a) [213]. In globular proteins ΔG results from the sum of large opposing enthalpic and entropic terms [235]. For example, above the ambient temperature (T> 30°C for metmyoglobin in Figure 1.6b) the entropic energy ($-T\Delta S$) favors the unfolded state due mostly to configurational entropy while the enthalpy (ΔH) favors the native state. These opposing contributions differ only

slightly in magnitude (See Figure 1.6c) which accounts for the marginal stability that is characteristic of globular proteins, i.e., $\Delta G \sim 40$ kJ.mol^{-1} for metmyoglobin in Figure 1.6b [198, 199].

Although thermodynamics is commonly used to study protein folding, the equilibrium thermodynamic quantities of amyloid fibrils are not easily accessible experimentally and they remain mostly unknown. In chapter 3 of this dissertation, we study the stability of amyloid fibrils using computer simulations and thermodynamic methods similar to the ones used to study protein folding.

1.4 Proteins and Alzheimer's Disease

Alzheimer's disease (AD) accounts for %60-%80 of all dementia cases. It is the sixth leading cause of death in the United States and it is estimated to cost over $1 trillion by the year 2050 [274]. Currently, there are treatments for AD symptoms which temporarily delay the progression of the disease, however, there is no cure for it. AD is associated with the accumulation of a specific type of plaque, called amyloid plaques which is mostly made up of a protein named amyloid-β (Aβ). This protein can vary in length from 39-43 amino acids and is derived from cleavage of the transmembrane amyloid precursor protein (APP) [95, 230]. Due to the role of the 42-residue long Aβ (Aβ_{42}) and its fibrils in the pathology of Alzheimer's disease, it has been the subject of many studies aimed at unraveling the self assembly process of this peptide.

Notice that any protein sequence is expected to form amyloid fibrils under appropriate conditions as a result of backbone interactions [79]. However a protein's ability to form amyloid structures can vary to a great extent depending on the inherent chemical-physical properties of its sequence. The amino acid sequence affects the propensity to form fibrils by favoring or hindering conformations along the pathway of fibril formation. It is crucial to understand the initial stages of this process in order to find structures which are prone to fibril formation and enhance

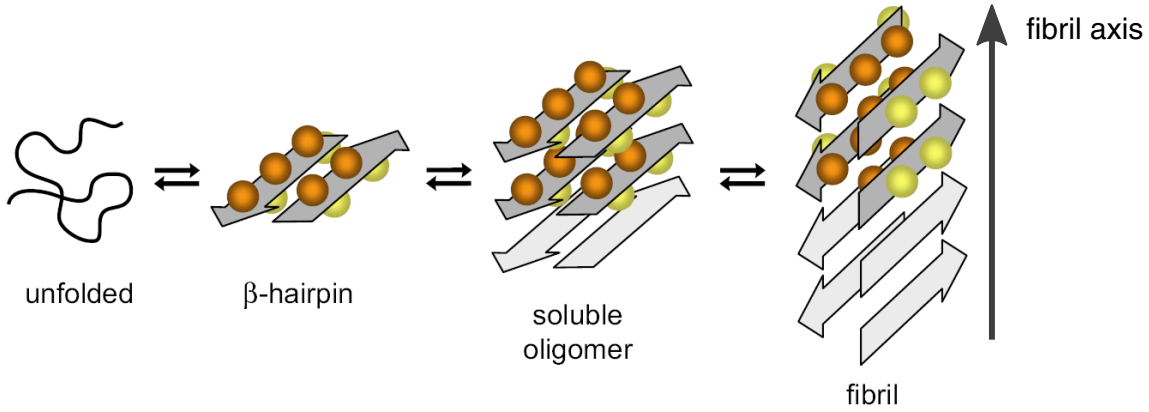


Figure 1.7 Hypothetical aggregation mechanism involving Aβ β-hairpin. Arrows represent β-strands, hydrophobic side chains are shown in orange and polar side chains are in yellow.
source: [103]

the protein self assembly. However, because proteins involved in neurodegenerative diseases, including amyloid-β, are inherently disordered and lack a stable structure, they are often hard to study experimentally. For this reason, computer simulations and all-atom molecular dynamics simulations have been employed in recent years to enhance our understanding of protein self assembly and fibril formation.

Mutagenesis studies have provided important insights into the properties of the amino acid sequences that are prone to from fibrils [16, 42]. This is important to predict proteins that can cause amyloid diseases [35,258]. Increasing the hydrophobic nature of protein sequences has been shown to increase the fibrillization rate. Aromatic side chain groups were found to occur frequently in fibril forming peptides but are not essential for aggregation [8, 43, 51, 57, 162]. Charge-charge interaction can be the dominant force of stabilization in small peptides with opposite charges [17] but they can also inhibit fibrillization in large sequences [225]. The length of the amino acid sequence also affects fibril formation wherein the truncation of a single residue from a fibril forming peptide was shown to prevent fibrillization [140]. In addition to these general properties of fibril-prone sequences, specific amino acid

patterns have also been investigated. In particular, alternating hydrophobic and hydrophilic residues, called amphipathic peptides, are prone to self-assemble due to burial of hydrophobic core and exposure of hydrophilic residues to the solvent [28,139]. Packing of side chains with complementary van der Waals volume plays a major role in self-assembly of these amphipathic peptides [140]. Sequences containing FF (Phe-Phe) or GxxxG (where x is a random amino acid) have been shown to promote fibril formation in proteins by forming a series of ridges and grooves on peptide chains which can dovetail into one another [4,148,158,205]. Short amyloid-inspired peptides containing these and similar motifs have been investigated to gain more knowledge about effects of protein sequence on aggregation [14, 17, 157, 162, 259].

Studies suggest that the mutation of fibrils has been shown to emerge from a hydrophobic collapse during which the formation of turns plays a major structural role [144]. Flexible turn sequences can be a key element inducing fibril formation while a more rigid motif connecting two KFFE sequences has been shown to prevent self-assembly [102]. Another structural feature that plays an important role in fibril formation is β-hairpin (See Figure 1.7). It has been reported that forming a β-hairpin in monomeric peptides can prevent amyloid fibril formation [102,103]. However, there are other studies suggesting the opposite [17, 139].

A number of computational studies have tried to characterize the structure of amyloid proteins and unravel the link between structural properties of proteins and aggregation pathways. Most of these studies have focused on the effect of the amino acid sequence of short peptides in fibril formation. However, due to numerous factors affecting the simulations of IDPs, including force field parameters, lack of sampling and simulation properties(temperature, concentration, PH ...) there is not a consensus on the effect of amino acid sequence on structure of Aβ monomers. In Chapter 2 of this dissertation, we focus on effect of aqueous environments that are either favorable or unfavorable to fibril formation on the structural changes of

full-length Aβ_{42} monomers. Using micro-second long molecular dynamics simulation, we found unique intermediate structures in the fibril formation pathway which might be favorable to fibril formation.

1.5 Molecular Dynamics Simulation

In this dissertation, we use computer simulations to discover Aβ structures that are prone to aggregate and to understand the molecular interactions accounting for the stability of amyloid fibrils. The main method used in this dissertation is molecular dynamics (MD) simulation in which the trajectory of all the atoms in the system is determined. Newton's second law is used to understand how the net force on each atom affects its motion as a function of time. This method is explained briefly in this section.

1.5.1 Equation of Motion and Interactions

The equation of motion used in molecular dynamics simulations is Newton's second law:

$$m_i \ddot{\vec{r}_i} = \vec{f_i}, \tag{1.1}$$

in which m_i, $\vec{r_i}$ and $\vec{f_i}$ are the mass, displacement and the net force on atom i. The net force on atom i is determined from the sum of all the bonded (covalent) and nonbonded (noncovalent) interactions between atom i and the rest of the system. These interactions are described by the potential energy ($u(r)$). The derivative of the potential energy gives the forces on each atom:

$$\vec{f} = -\vec{\nabla} u(r). \tag{1.2}$$

At each time step of MD simulation, the net force is calculated from the potential energy to determine the trajectory of all the atoms.

1.5.2 Force Fields

Force fields in molecular dynamics simulations are the set of equations (called the potential functions) and their parameters which are used to describe the interactions

between atoms of a system. There are two types of force fields:

i) *all-atom*, in which parameters are provided for every single atom in the system,

ii) *united atom*, in which several atoms are grouped into one bigger particle. This is also known as *coarse grained* model.

For all-atom models, the force field incorporates a basic form of potential energy functions:

$$u(r) = \sum_{bonds} k_b(b-b_o)^2 + \sum_{angles} k_\theta(\theta-\theta_o)^2 + \sum_{torsions} k_\phi[cos(n\phi+\delta)+1] + \sum_{\substack{nonbonded \\ pairs}} \left[\frac{q_i q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}}\right]. \tag{1.3}$$

Sums over bonds (b), angles (θ) and torsions (ϕ) in equation 1.3 describe oscillations of a pair of covalently bonded atoms about the equilibrium bond length (b_o), oscillations of three covalently bonded atoms about an equilibrium bond angle (θ_o) and torsional rotation of four covalently bonded atoms about a central bond, respectively. Schematic of these potentials and their corresponding particle conformations are shown in Figure 1.8. k_b, k_θ, k_ϕ, n and δ are constants defining these oscillations. The final summation which uses partial atomic charges q_i and q_j, pair distance r_{ij} and constants A_{ij} and C_{ij}, describes non-bonded energy terms (electrostatics and Lennard-Jones) between atom pairs. Depending on the force filed, the terms as well as the constants used in equation 1.3 may vary. The combination of potential energy functions (as in equation 1.3) and all the parameters used in it ($k_b, b_o, k_\theta, \theta_o$, etc.) constitutes a force field.

In this dissertation, AMBER99SB-ILDN [146] force field was chosen for simulations involving amyloid-β protein because of its good agreement with NMR structure data [146, 196].

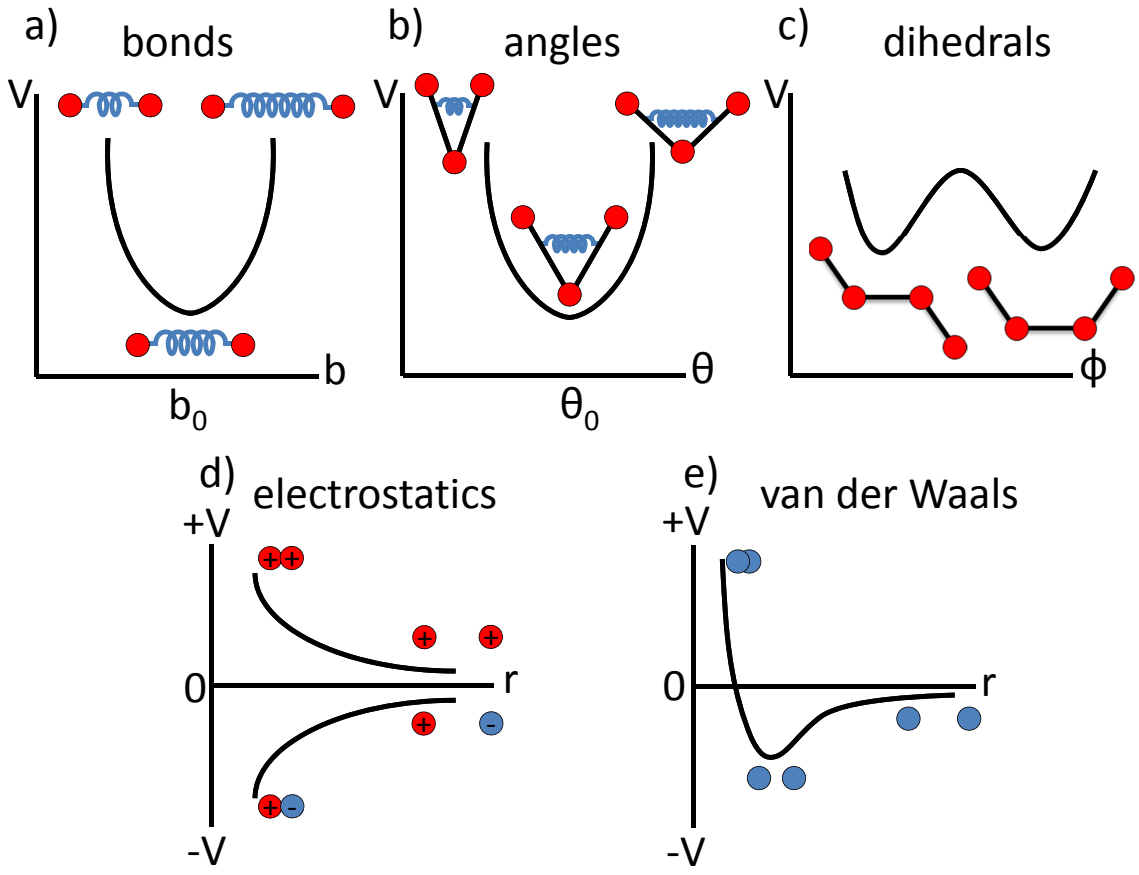


Figure 1.8 Schematic of force field potential functions. (a) Spring potential between pair of covalently bonded particles with equilibrium bond length b_0, (b) spring angle potential between three particles forming an equilibrium angle θ_0, (c) dihedral potential between four particles, (d) electrostatic potential between two charged particles, and (e) van der Waals potentials between two particles.

1.5.3 Periodic Boundary Conditions

A problem with simulating finite systems, is the boundary effect. In small systems particles have too many unwanted interactions with the boundaries of the system and as a result, the trajectory of the system is influenced by the interactions between particles and their boundaries. These interactions must be avoided if we intend to model a bulk system. Applying periodic boundary condition is the way to the avoid boundary effect in molecular dynamics simulations.

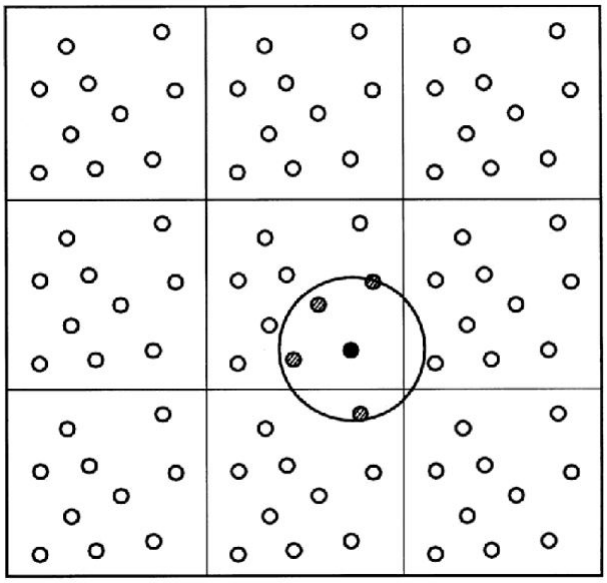


Figure 1.9 Periodic boundary condition in two dimensions. The unit cell is replicated infinitely in both directions.
source: [15]

To apply periodic boundary conditions, the simulation box is replicated in every direction throughout the space to form an infinite lattice as shown in Figure 1.9. During the simulation, as a particle in the original box moves, all of its periodic images also move in their periodic box in exactly the same way. Whenever a particle leaves the original box from one side, one of its images will enter the original box through the opposite side. By implementing this method, particles no longer interact with the boundaries and the problem is overcome [6]. Periodic boundary conditions is implemented in molecular dynamics simulations throughout this dissertation.

1.6 Enhanced Sampling Methods in Computer Simulations

Free energy difference is the driving force of chemical reactions and its calculation is an essential task in computational studies of proteins. Biological molecules are known to have rough free energy landscapes with many local minima that are separated by high energy barriers [187]. Due to these energy barriers, transitions between free

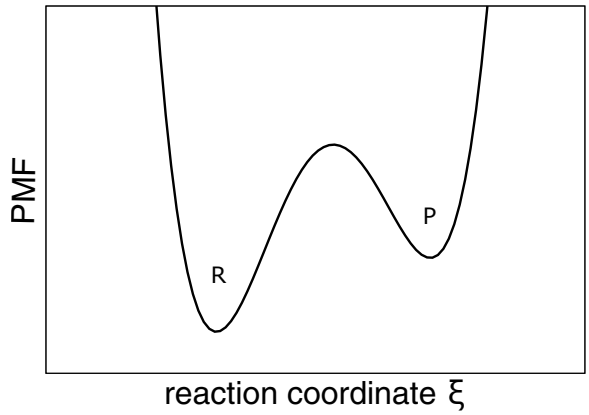


Figure 1.10 Local minima on potential of mean force as a function of reaction coordinate. The protein can get stuck in either of the two local free energy minima (denoted as R or P) and slow down the sampling process. Umbrella sampling is an efficient method to overcome such sampling obstacles.

energy minima may require a significant amount of simulation time. To avoid this problem, a number of methods have been developed to accelerate the configuration sampling in order to calculate the free energies [31, 249, 261]. Among these methods are *umbrella sampling*, which forces the system to sample unstable conformations that are along a reaction coordinate by introducing a bias potential. Another method is *replica exchange molecular dynamics* (REMD), which speeds up the sampling by exchanging conformations of a system with conformations of the exact system at a different temperature. This method is especially beneficial for simulations in which the conformations are separated by high energy barriers.

In this dissertation, we have employed umbrella sampling and replica exchange molecular dynamics either individually or in combination with each other to further enhance the sampling. To the best of our knowledge, employing both methods together in an all-atom molecular dynamics simulation of amyloid fibrils with explicit water solvent has never been performed before.

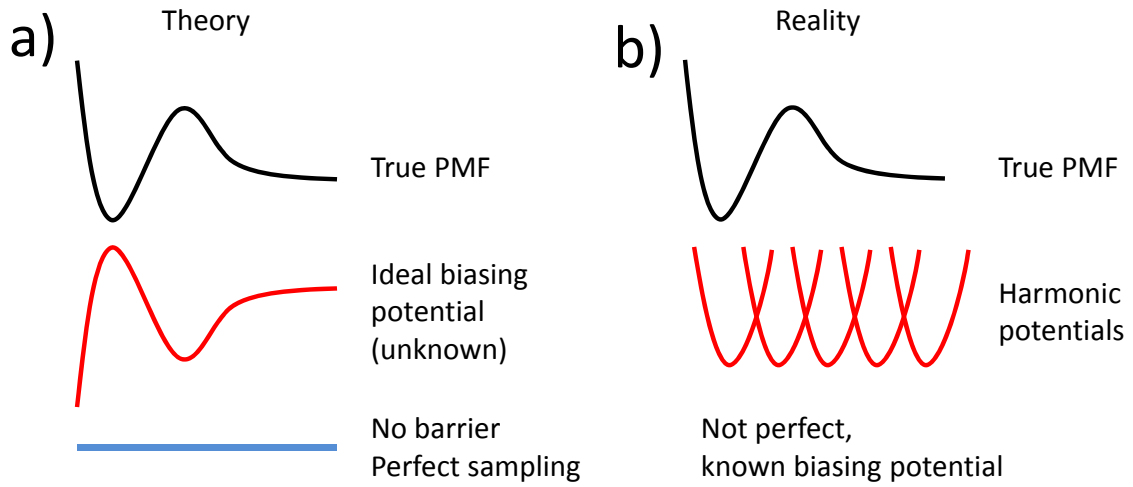


Figure 1.11 Concept of umbrella sampling method in simulations. (a) Ideal situation in which the biasing potential completely neutralizes the PMF and sampling will be perfect. (b) Schematic of umbrella sampling in practice, in which we apply a series of harmonic potential and force the system to sample the unfavorable conformations.

1.6.1 Umbrella Sampling

The calculation of free energy differences is an important task in computational biology. Umbrella sampling is one of the methods that provide the free energy along a reaction coordinate. In this method, a reaction coordinate which distinguishes two thermodynamic states of the system is chosen. Then a series of simulations (called *windows*) are performed along this reaction coordinate with a bias potential applied to them as shown in Figure 1.11. The bias potential forces each simulation to sample a specific range of the reaction coordinate and ensures sufficient sampling at every window.

The statistics of unbiased simulation can be obtained from the biased simulations because the biasing potential is known. To study the interactions between two proteins, umbrella sampling has been used to calculate the free energy as a function of separation between proteins [72, 122, 170]. The formalism of recovering unbiased free energy differences from biased simulations is discussed briefly in this section. The

bias potential $v_i(r)$ at window i is an additional energy term that depends on the configuration of the system defined by r. The potential energy of the biased system, $E^b(r)$, is:

$$E^b(r) = E^u(r) + v_i(r), \tag{1.1}$$

in which the superscript u denotes unbiased potential energy. In order to obtain the unbiased free energy, we need the unbiased probability distribution of the system which is:

$$P_i^u(\xi) = \frac{\int exp[-\beta E(r)]\ \delta[\xi(r) - \xi']d^N r}{\int exp[-\beta E(r)]d^N r}, \tag{1.2}$$

where $\beta = 1/(k_B T)$, k_B being Boltzman constant, $\xi(r)$ is the reaction coordinate of the system for a given configuration defined by r, ξ' is the reference reaction coordinate and N denotes the number of degrees of freedom of the system. Umbrella sampling simulation of the biased system provides the biased distribution along the reaction coordinate. Assuming the simulation is ergodic, i.e., every point in the phase space is sampled during the simulation, the biased distribution along the reaction coordinate is:

$$P_i^b(\xi) = \frac{\int exp\{-\beta[E(r) + v_i(\xi(r))]\}\ \delta[\xi(r) - \xi']d^N r}{\int exp\{-\beta[E(r) + v_i(\xi(r))]\}d^N r}. \tag{1.3}$$

Because the bias depends only on ξ and the integration in the enumerator is performed over all degrees of freedom except ξ,

$$P_i^b(\xi) = exp[-\beta v_i(\xi')]\ \times \frac{\int exp[-\beta E(r)]\ \delta[\xi(r) - \xi']d^N r}{\int exp\{-\beta[E(r) + v_i(\xi(r))]\}d^N r}. \tag{1.4}$$

Using equation 1.2, we get,

$$
\begin{aligned}
P_i^u(\xi) &= P_i^b(\xi)\ exp[\beta v_i(\xi')] \times \frac{\int exp\{-\beta[E(r) + v_i(\xi(r))]\} d^N r}{\int exp[-\beta E(r)] d^N r} \\
&= P_i^b(\xi)\ exp[\beta v_i(\xi')]\ \langle exp[-\beta v_i(\xi)]\rangle,
\end{aligned}
\tag{1.5}
$$

where $\langle ... \rangle$ represents the expected value, $P_i^b(\xi)$ is obtained from umbrella sampling using biased system and $v_i(\xi')$ is given analytically. By defining $F_i = -(1/\beta)\ln\langle exp[-\beta v_i(\xi)]\rangle$ which is a constant, Gibb's Free energy of the unbiased system at window i is:

$$
G_i^u(\xi) = -(1/\beta)\ \ln\ P_i^u(\xi) = -(1/\beta)\ \ln\ P_i^b(\xi) - v_i(\xi) + F_i. \tag{1.6}
$$

To ensure sampling in all regions of the reaction coordinate, the range of interest of ξ is divided into a number of windows. A bias potential is applied to each window to keep the system close to the reference point ξ' of the respective window. Harmonic potential with spring constant K is commonly used as bias potential in MD simulations:

$$
v_i(\xi) = \frac{1}{2} K (\xi - \xi_i')^2. \tag{1.7}
$$

In Chapter 3 of this dissertation, we have used this harmonic potential functions in order to calculate free energy of amyloid fibril dissociation and used equation 1.6 to determine the unbiased free energy from the biased one.

1.6.2 Replica Exchange Molecular Dynamics

Molecular dynamics simulations are generally carried out at a given temperature, using a thermostat to keep the average temperature of the system constant. An initial configuration of the system is chosen arbitrarily and is equilibrated for a short time to avoid any steric clashes between the atoms. The simulation is then run for a long time to calculate molecular properties of the system. However, the system of interest may have many potential energy minima, which are separated by relatively

high barriers and are difficult to cross at ambient temperatures during the simulation time (See Figure 1.10). It means that results of a computer simulation are confounded by the choice of initial conditions. This is the case for molecular dynamics simulation of proteins in solutions which need mili seconds to form a specific conformation.

Several methods have been devised in the past to overcome this problem. A well known method that produces efficient sampling in systems with slow equilibration is known as *replica exchange*. In replica exchange simulation, multiple replicas of the same system are simulated at different temperatures and exchanges between replicas are attempted at regular intervals. The exchange between low and high temperatures allows the lower temperature system, which may have been stuck in a free energy minima, to escape the barrier and sample a broader range of phase space at a higher temperature as seen in Figure 1.12. To have an efficient number of exchanges between the systems, the neighboring replicas require sufficient overlap between their potential energies. This results in a high number of replicas within the intended temperature range which makes REMD computationally intensive for all-atom models [204].

In this dissertation we use the most commonly used replica exchange molecular dynamics method of Sugita and Okamoto (1999). In this method, M replicas $S_1, S_2, ..., S_M$ of a system, at different temperatures $T_1, T_2, ..., T_M$ and thermal energies $\beta_1, \beta_2, ..., \beta_M$ are produced. At a given time, the ensemble of systems consists of M configurations with potential energies $E_1, E_2, ..., E_M$. The probability of finding the system in this state is:

$$P \propto exp[-(\beta_1 E_1 + \beta_2 E_2 + ... + \beta_M E_M)]. \tag{1.8}$$

Consider the exchange between systems S_i at β_i and S_j at β_j. The probability of finding these systems before the exchange is:

$$P_{before} \propto exp[-(\beta_i E_i + \beta_j E_j)], \tag{1.9}$$

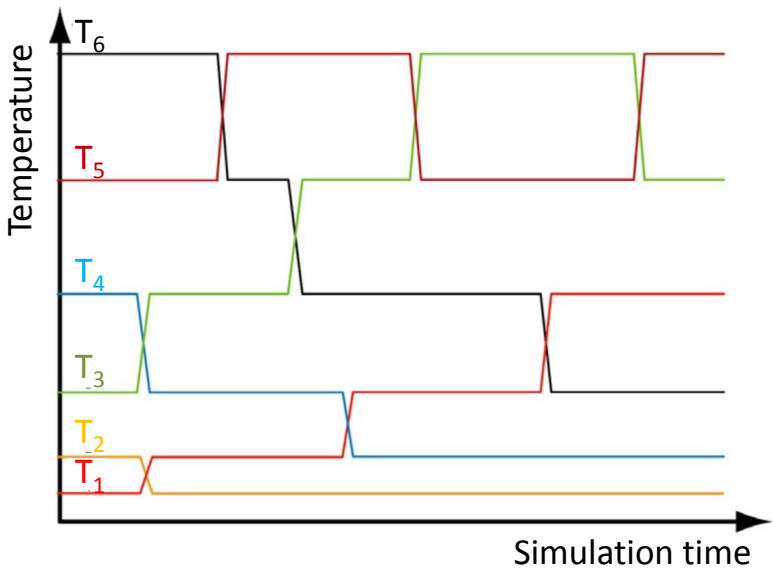


Figure 1.12 Exchanges between neighboring replicas at different temperatures. Each replica starts at a specific temperature and after regular time intervals, exchanges between neighboring replicas are attempted.
source: [173]

If the exchange is allowed between the systems, such that S_i and S_j will be at β_j and β_i respectively, the probability of the system after the exchange is:

$$P_{after} \propto exp[-(\beta_j E_i + \beta_i E_j)]. \tag{1.10}$$

The ratio of the probabilities after and before the exchange will be:

$$\frac{P_{after}}{P_{before}} = e^{-\Delta}, \tag{1.11}$$

where

$$\Delta = (\beta_i - \beta_j)(E_i - E_j). \tag{1.12}$$

The transition probabilities $p_{\rightarrow}$ (transition from "before" to "after" the exchange) and $p_{\leftarrow}$ (transition from "after" to "before" the exchange) should fulfill

the equilibrium criteria:

$$p_{\rightarrow}\ P_{before} = p_{\leftarrow}\ P_{after}. \tag{1.13}$$

Thus it follows that

$$\frac{p_{\rightarrow}}{p_{\leftarrow}} = e^{-\Delta}. \tag{1.14}$$

The probability of exchange between two states i and j is accomplished by the Metropolis criteria:

$$P_{i \leftrightarrow j} = min\left(1,\ exp\left[(\beta_i - \beta_j)(E_i - E_j)\right]\right), \tag{1.15}$$

In order to ensure that REMD simulations are efficient, the set of temperatures in replica exchange should be chosen in a way that sufficient exchanges between the replicas occur over the trajectory. An acceptance ratio of 20% is considered reasonable [18]. In Chapters 2 and 3 of this dissertation, we employed replica exchange molecular dynamics to enhance the sampling in our systems.

CHAPTER 2

EFFECTS OF SMALL COMPOUNDS ON STRUCTURE OF Aβ_{42} MONOMERS

2.1 Introduction

The amyloid beta (Aβ) protein is the main constituent of senile plaques in the brain of Alzheimer's patients. It is an intrinsically disordered protein with a high propensity to aggregate into soluble oligomers that can nucleate into amyloid fibrils. More than 15 point mutations in the Aβ sequence have been found to be pathological accounting for autosomal dominant Alzheimer's disease with an early onset–see Figure 2.1. Most of these mutations enhance the aggregation rate of Aβ compared to wild type. In contrast, point mutations that account for a slower aggregation rate of Aβ have been found to have a protective effect delaying the onset of the disease. This points to a relationship between Aβ aggregation and Alzheimer's which has been the subject of extensive studies. In particular, it has been hypothesized that within the spectrum of structures adopted by Aβ monomers, some conformations are more prone to aggregation than others. These aggregation-prone conformations are expected to play an important role in determining the rate of Aβ oligomerization and, thus, the progression of the disease. Inhibiting the formation of these aggregation-prone structures could become part of a treatment to slow down the progression of Alzheimer's.

In the brain of healthy and Alzheimer's individuals, Aβ emerges from cleavage of the transmembrane "amyloid precursor protein". This process can proceed through different cleavage pathways accounting for Aβ proteins with lengths varying from 34 to 43 residues. The most abundant Aβ proteins are 40 and 42 residues long with the latter (i.e., Aβ_{42}) exhibiting a faster aggregation rate and increased toxicity. This increased propensity of Aβ_{42} to aggregate has been related to the hydrophobic nature

of its additional residues 41 and 42. Accordingly, the C-terminus and the central region of Aβ are highly non-polar whereas the N-terminus of this peptide contains negative and positive amino acids–see Figure 2.1. Notice that a similar segregation of charged and non-polar residues is also observed in many antimicrobial peptides [241] which may explains the antimicrobial activity of Aβ against various bacteria including *E. Coli* and *S. aureus*.

The intrinsically disordered nature of Aβ implies that, at physiological condition, monomers of this protein sample an ensemble of conformations without adopting a stable native structure. Solution NMR has shown that this ensemble for Aβ_{42} consists mainly of collapsed coil structures [215]. Measurements using backbone H_α, C_α, and C_β chemical shifts have reported [215, 216] a high propensity to form β-strands for residues 17-21, 31-36, and 39-41 whereas residues 7-11 and residues 20-26 exhibit a high propensity to form turns–see Figure 2.1. Consistent with these experiments, CD measurements have reported an overall β content for Aβ_{40} monomers of 24% [215]. Using Fourier transform infrared (FTIR) spectroscopy, these β-strands were found to form mostly anti-parallel β-sheets for monomeric and oligomeric states of Aβ [215]. Since amyloid fibrils are made of parallel β-sheets, a structural rearrangement has to take place during the nucleation of oligomers into amyloid fibrils.

In addition to experiments, computer simulations are also providing insights into the structure of Aβ monomers [234]. Earlier all-atom molecular dynamics simulations in explicit solvent reported a significantly lower frequency of β-strands compared to experiments. Recently, this discrepancies was shown to emerge in simulations where the conformation of Aβ was not sampled sufficient. As a matter of fact, it was shown that, in order to obtain an equilibrium structural ensembles of Aβ monomers, extensive replica exchange molecular dynamics (REMD) simulations with over 400 ns of simulation times are necessary [216]. This requires substantial computational resources which have only recently become available to the scientific community. In

these equilibrium molecular dynamics ensembles of Aβ, β-strands are formed within the same residues as experiments and with a frequency close to the one measured experimentally. This result was reproduced using three force fields showing that equilibrium Aβ ensembles are mostly robust against changes in the force field [216]. Changes in the latter was shown to affect mainly the flexibility of loop regions as well as electrostatic interactions between charged side chains. This finding is reassuring and suggests that, when simulations are performed long enough to produce equilibrium ensembles, they can complement experiments by providing atomic level insights.

Changes in the properties of the solution in which Aβ is immersed has a strong effect on its aggregation rate and toxicity. These changes can be produced by adding small molecules to the solution [248, 270]. Metal ions, which affect electrostatic interaction in proteins, can significantly increase the rate of amyloid fibril formation [129]. Similarly, amyloid fibrils can be formed 3-4 times faster when NaCl is added to the solution at a concentration of 150 mM, i.e., at physiological conditions [2]. Small molecules can also inhibit the formation of amyloid fibrils. Some of these molecules contain aromatic rings, which is the case of 4-Aminophenol (4AP). The latter inhibit Aβ_{42} oligomerization in a concentration dependent manner [56]. Inositol isomers, which is a sugar molecule, also inhibits fibril formation by driving Aβ peptides into an unstructured macroscopic aggregate [166, 167, 240]. It has been hypothesized that changes in the aggregation rate of Aβ are produced by small biases in the conformation sampled by its monomer either favoring or inhibiting aggregation-prone structures [215].

Here, we perform REMD simulations to compute equilibrium ensembles of Aβ conformations in pure water as well as in aqueous solutions containing small inhibitors (i.e., 4AP and inositol) or promoters (i.e., NaCl) of amyloid fibril formation. A comparative study of conformations in these different ensembles is performed

to determine Aβ structures that are more frequently populated in the presence of NaCl while inhibited in the presence of 4AP or inositol. We find that the content of secondary structures does not change significantly in these different ensembles. However, there are significant differences in how the different Aβ regions interact with each other in the different environment. Using cluster analysis, we find a specific strand-loop-strand conformation that is highly populated in aqueous solution containing NaCl but is absent in solutions containing inhibitors. A structural analysis of this conformation reveals the presence of a β-sheet between the C-terminus and the hydrophobic-core region which provides the proper alignment for the formation of amyloid fibrils determined using solid-state NMR, e.g., PDB ID: 2BEG [149]. In this strand-loop-strand conformation, residues G38 and V40 of the C-terminus is aligned with residues L17 and F19, which constitute the hydrophobic core of the amyloid fibril structure. We propose that this strand-loop-strand conformation is an aggregation-prone structure of Aβ and, thus, an intermediate structure in fibril formation. Inhibiting its formation might slow down protein aggregation and reduce toxicity and related to Alzheimer's disease.

2.2 Methods

Conformations of Aβ_{42} monomers in pure water and aqueous solutions containing $\sim$ 0.2 M of NaCl, inositol, or 4-Aminophenol (4AP) are studied. An helical conformation of Aβ_{42} (PDBID: 1IYT) is used as its initial structure in the simulation [50]. For simulations performed in pure water, the peptide is placed in a cubic box of 6 nm length containing 6,305 TIP3P water molecules. The net charge of the system was neutralized by adding 3 Na ions. Details of the simulation setup in the other aqueous solutions are summarized in Table 2.1. The AMBER99SB-ILDN force field is used in all simulations because it provides reasonable agreement with NMR data of small proteins [147, 196]. The energy of the system is minimized and the system is

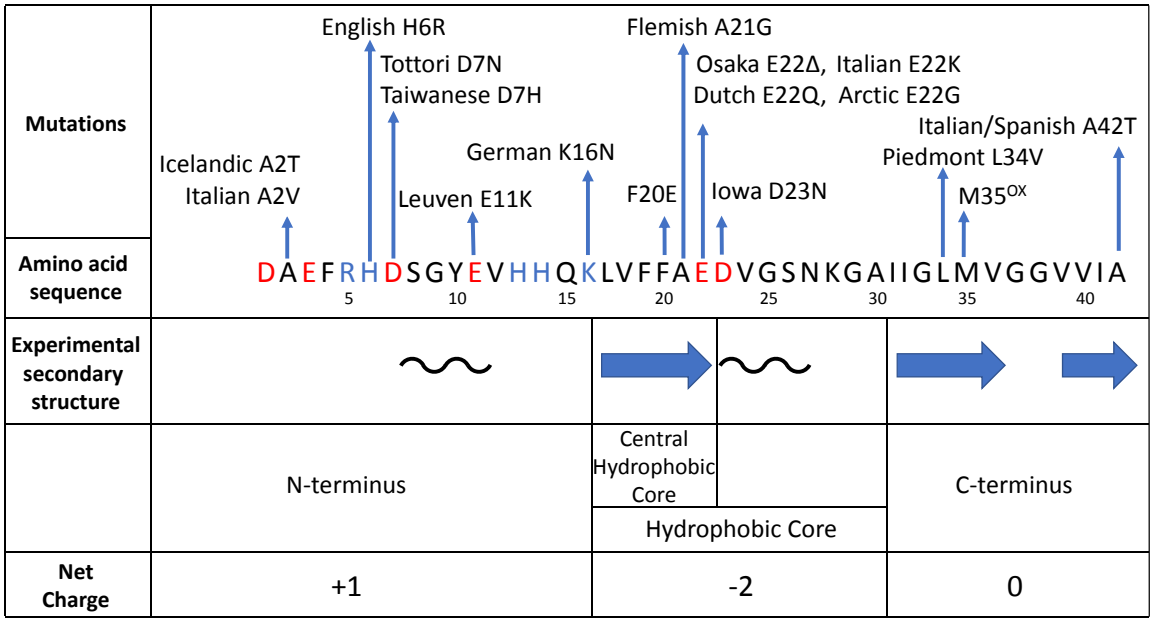


Figure 2.1 Amino acid sequence of Aβ_{42} and some of its point mutations associated with early onset Alzheimer's. Mutations are typically named after the geographic location in which they were first identified. Red and blue colors in the sequence of Aβ_{42} is used to represent negatively and positively charged residues, respectively.Experimentally identified turns and β-sheets are marked using curled black lines and blue arrows, respectively. Residues comprising the hydrophobic core, N- and C-terminus of Aβ_{42} are highlighted as well as the net charge of these regions. *source: [9, 39, 49, 99, 119, 182, 183, 217, 279]*

equilibrated in the NPT ensemble (300 K and 1 atm) for 500 ps. The final structure of this equilibriation period is used in our constant pressure Replica Exchange Molecular Dynamics (REMD) simulations. In these simulations, the system was coupled to a Velocity Rescaling thermostat (τ_T=0.1 ps) to maintain constant average temperature and the pressure was fixed using Berendsen barostat (τ_P=0.1 ps). Periodic boundary conditions were applied and the particle-mesh Ewald (PME) method was used to treat long-range electrostatic interactions. A 1.0 nm cut-off distance was used for van der Waals interactions. Covalent bonds were constrained using the LINCS algorithm, and an integration time-step of 2 fs was used together with the leap-frog integrator. Simulations were performed using GROMACS version 5.1.

REMD simulation details. We use a total of 64 replicas with temperatures distributed exponentially in the 310-500 K range [191]. Swaps between replicas were attempted every 4 ps which resulted in an average exchange probability of approximately 20%. Each system was simulated for 700 ns/replica, with the exception of simulations performed in pure water which converged faster and were simulated for 600 ns/replica. Convergence of the simulations were assessed by computing the secondary structure content of Aβ. We used the DSSP algorithm to determine secondary structures of a given conformation. The overall secondary structure content of Aβ averaged over windows of 50 ns is shown in Figure 2.2. Secondary structures converge after 400 ns and 500 ns for simulations performed in pure water and the other aqueous solutions, respectively. The last 200 ns of each ensemble at 320 K is used for all analysis and this time window is referred to as *production run*. Fluctuations in β-sheet content during the production run is less than %5–see Figure 2.2. Notice that β-sheet is the slowest secondary structure to converge. Results in Figure 2.2 show the need of performing long REMD simulations.

Force field of Small Compounds. Parameters for inositol and 4AP were obtained using AmberTools17 as follows: the electrostatic potential of inositol and 4AP are obtained at the HF/6-31G* level after a geometry optimization at the AM1 level of theory. Partial charges are derived by fitting the electrostatic potential using the Bond-Charge Corrections (BCC) method, and other force field parameters of the molecules are taken from the GAFF [269] parameter set. Parameters of NaCl force field used in this study are based on the work of Joung et al. [118].

J-coupling calculation. Spin-spin splitting occurs between nonequivalent nuclei which are connected by 1-3 covalent bonds. This splitting is is referred to as J-coupling and it provides information about the conformation of the chemical groups in the protein [218]. For example, the three-bond J-coupling constant, $^3J_{HNHA}$, between amide proton and the alpha proton depends on the ϕ dihedral angle in the peptide

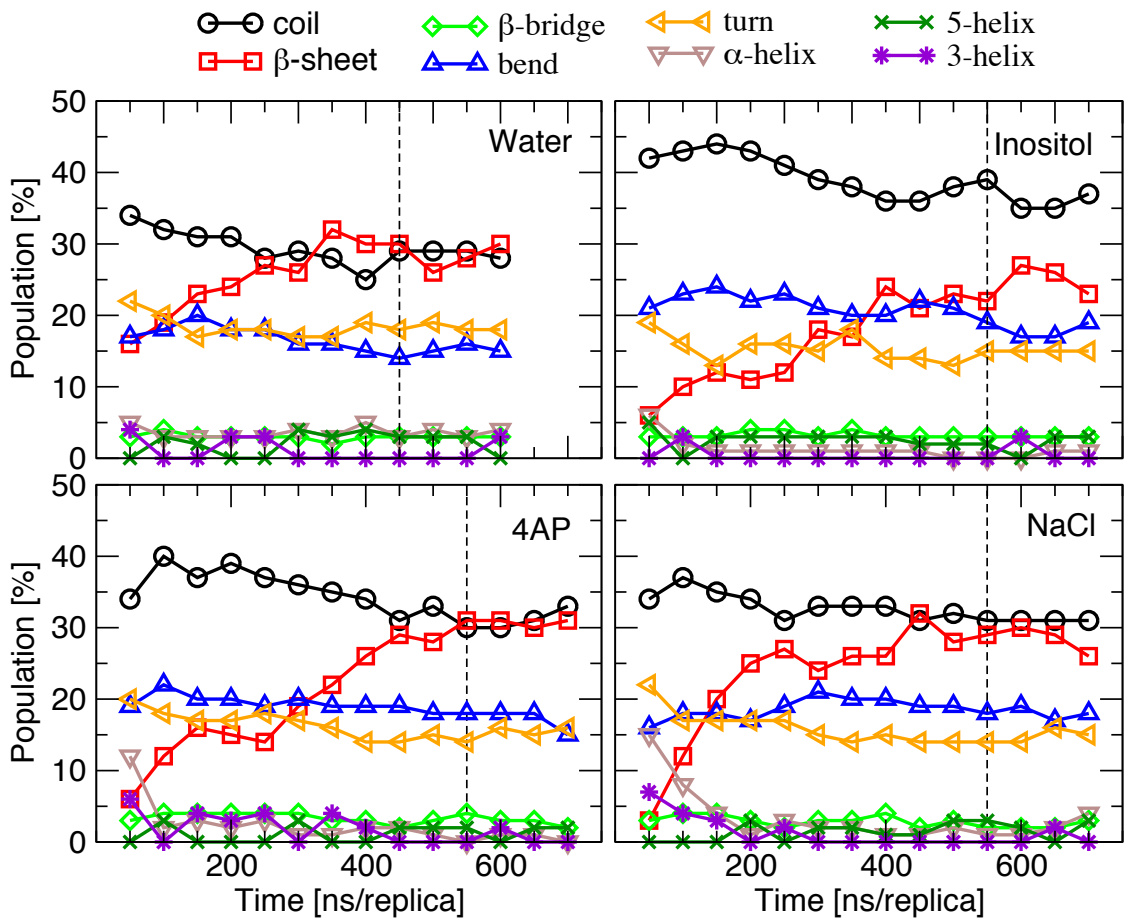


Figure 2.2 Convergence of REMD simulations at 320 K. Content of secondary structures for each system as determined by DSSP over 50 ns/replica windows. Dashed lines show the threshold where the convergence is achieved in ensembles.

and holds values >9 Hz for β-sheets and <4 Hz for α-helices. In order to assess the validity of structures found in REMD simulations, we compare the experimental and calculated values of the $^3J_{HNHA}$ coupling constant which is calculated using the Karplus equation [218] from our MD simulations of Aβ_{42} in pure water:

$$^3J_{HNHA} = A\cos^2(\phi - 60) + B\cos(\phi - 60) + C, \tag{2.1}$$

where A, B and C constants are determined by different groups [215, 266] and ϕ is the peptide dihedral angle. The two sets of Karplus constants used in this study are

Table 2.1 Details of Simulations Performed in This Study

Aβ_{42} monomer	Time per replica	Number of water molecules	Number of solute molecules	Concentration of solute
Pure water	700 ns	6305	0	0 M
NaCl	700 ns	6305	23	0.196 M
Inositol	700 ns	6305	23	0.193 M
4AP	700 ns	6305	23	0.192 M

determined by Vuister *et al.* [266] and Rosenman *et al.* [215] and are shown in Table 2.2. Note that Vuister's parameter set are determined by fitting to crystallographically well-defined structures whereas Rosenman's parameter are calculated to minimize the difference between Aβ_{42} experimental and simulation-derived J-couplings. Production run of Aβ_{42} in pure water ensemble was used to calculate the average value of peptide dihedral angles and the results were compared with experimental $^{3}J_{HNHA}$ values obtained by Garcia *et al.* [215, 234].

Correlations of the MD-derived and experimental J-coupling values are examined by the Pearson Correlation Coefficient (PCC) [215, 234]:

$$\mathrm{PCC} = \frac{\Sigma_{i=1}^{n}[(x_i - \overline{x})(y_i - \overline{y})]}{(n-1)\sigma_x\sigma_y}, \tag{2.2}$$

for data sets x and y with size n and standard deviations of σ_x and σ_y respectively. PCC values range between -1 and 1 where PCC = 0 means no correlation and PCC = ± 1 shows perfect correlation between data sets.

2.3 Results

Secondary Structures. To understand the structural differences in Aβ_{42} monomer introduced by small compounds, in Figure 2.3 we calculated the (a) β-sheet and (b) turn content of the four systems over all residues in the production run. All of

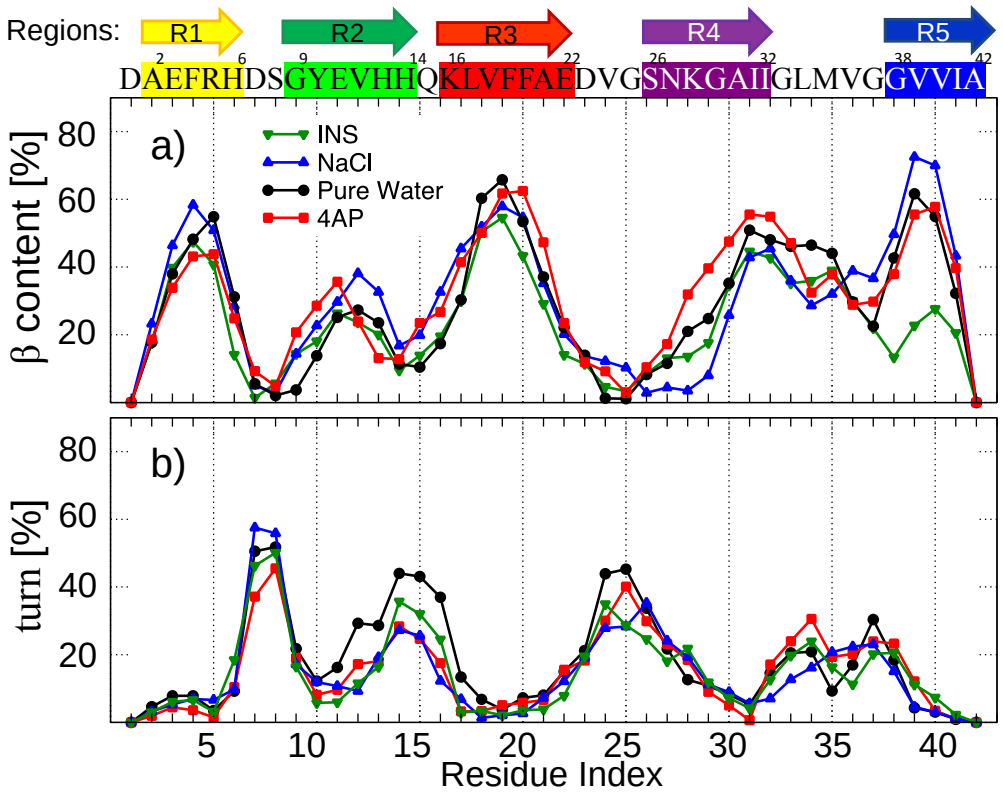


Figure 2.3 Residue-wise secondary structure content at 320K for Aβ_{42} in Inositol, NaCl, pure water and 4AP averaged over production run. The amino acid sequence of Aβ_{42} and five regions forming highest amount of β-sheet are shown on the top as well as (a) β-sheet and (b) turn content.

Aβ_{42} ensembles show a similar overall pattern with peaks at N-terminal (R1-R2), central hydrophobic cluster (R3) and in C-terminal (R4-R5) and residues are mostly found in β-sheet or turn conformation. Differences among ensembles arise mostly in the region containing residues 22 to 30 (Turn Region) and C-terminal. In the latter region, NaCl increases β-content whereas Inositol shows a significant decrease in β-content. This decrease in β-strand at C-terminal has been observed in monomeric structures of the Alzheimer's protective mutation (A2T) [54]. In the turn region the decrease in β-sheet in NaCl is compensated by an increase in turn as observed in

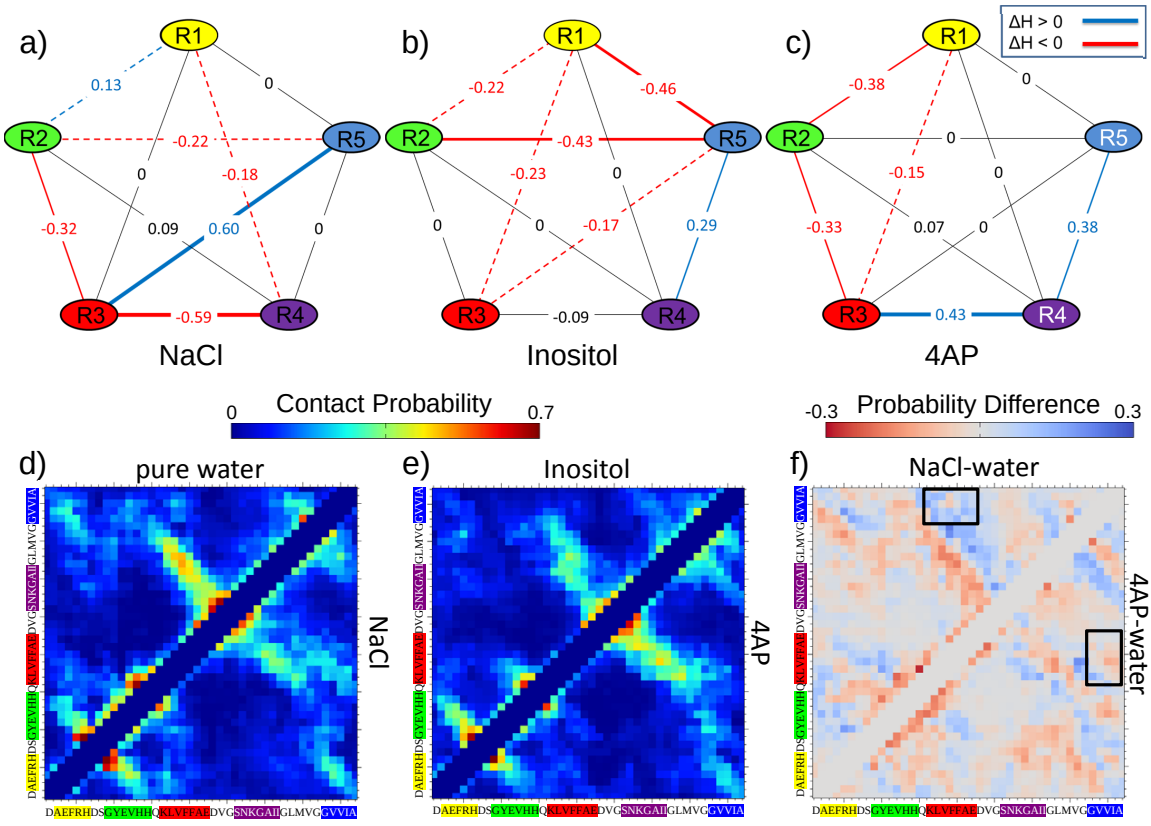


Figure 2.4 Aβ_{42} interactions analyzed by hydrogen bonds and contact maps. Change in average number of hydrogen bonds in systems with (a) NaCl, (b) Inositol and (c) 4AP relative to pure water over the production run between regions introduced in Fig2.3. Dashed, thin and thick lines represent trivial, moderate and significant changes in hydrogen bonds respectively. Blue lines shows increase in hydrogen bonds relative to pure water and red lines shows decrease in hydrogen bonds compared to pure water. (d-e) Contact probability maps of C_α atoms over the production run for each of the ensembles are illustrated. Two residues are in contact when their C_α-C_α distance is less than 8 Å. Only non-sequential contacts ($|i-j|>3$) are shown in the maps. (f) Difference between contact probability maps of NaCl (upper corner) and 4AP (lower corner) from pure water ensemble. Blue and red colors represent increase and decrease in contacts compared to pure water ensemble.

Table 2.2 Karplus Data Set Parameters Used to Compare Experiments with Simulation Derived Couplings

	Coupling	A	B	C
Vuister *et al* [266]	$^3J_{HNHA}$	6.51	-1.76	1.60
Rosenman *et al* [215]	$^3J_{HNHA}$	6.88	-6.50	-3.53

Figure 2.3b. 4AP shows higher β-sheet propensity in turn region whereas Inositol shows no significant difference with water. Another region with high β-content is R2 region with more β-sheet structures in NaCl and 4AP compared with pure water and Inositol. Consequently, a decrease in turn is observed for these systems in R2 region. The data presented in Figure 2.3 show the compensation of β-sheet and turn content and indicate that small molecules have no significant effect on the secondary structures of Aβ_{42} monomer, however interactions among the regions in Aβ_{42} are significantly affected by these compounds as discussed in the following sections.

Hydrogen Bonding regions. To understand how the small compounds used in this study affect the intra-peptide interactions of Aβ_{42}, in Figure 2.4a-c we show the difference in average number hydrogen bonds in each ensemble relative to pure water between regions of Aβ_{42} monomer. These regions are introduced in Figure 2.3 and are chosen due to their higher β-sheet content. In Figure 2.4a-c, dashed, thin and thick lines represent trivial, moderate and significant changes whereas blue and red colors show increase or decrease in hydrogen bonds relative to pure water respectively. For Aβ_{42} in NaCl, Figure 2.4a shows diminishing hydrogen bonds between regions R2 and R3 but regions R3 and R5 form considerably higher number of hydrogen bonds compared to water which can be explained by higher β-sheet content

in these regions in NaCl solution. Inositol mainly disrupts hydrogen bonds between Aβ_{42} residues and has the most effect on interactions between N and C-termini (R1-R5 and R2-R5 hydrogen bonds)-See Figure 2.4b. This behavior is inline with Inositol's lower β content in C-terminal resulting in lower overall hydrogen bonds. 4AP enhances hydrogen bonds in C-terminal regions (R3-R4 and R4-R5) while disrupting N-terminal interactions (R2-R3, R1-R4 and R1-R2) as shown in Figure2.4c. This can be related to its slight increase of β-sheet in R4 and reduced β-sheet in R1 region, however further information is needed to clearly understand the decline in hydrogen bonds caused by 4AP.

Contact Probability Maps. To investigate the tertiary structures of Aβ_{42}, contact probability maps in the production run are computed for the four systems in Figure 2.4d,e. Two residues are assumed to be in contact when the distance between their C_α-C_α atoms is less than 8 Å. To analyze these maps we use the five regions introduced in Figure2.3. In pure water, Figure 2.4d-upper corner, four sets of contacts are observed: I) the predominant anti-diagonal contacts between regions R3 and R4 which is indicative of an anti-parallel β-hairpin, II) the anti-parallel N-terminal contacts of R1 and R2, III) the C-terminal contacts between R4 and R5 region and IV) dispersed N and C-terminal contacts that span the residues 1-13 and 31-42. Despite having ample intra-peptide contacts, residues 20-30 show no contact with N and C-termini due to their high turn propensity (See Figure2.3b) which connects two β-containing regions existing in all ensembles. Figure 2.4d-lower corner illustrates the higher intra-peptide interactions with higher anti-parallel N-terminal contacts between R1 and R2 regions in NaCl ensemble. Lower contacts in R3-R4 regions in NaCl is compensated by a major peak in R3-R5 region, which together with its higher β-content and hydrogen bonding is indicative of anti-parallel β-sheets. The dispersed N and C-terminal interactions exist in NaCl as the value of hydrogen bonds between R1-R5 has no significant change. Inositol exhibits roughly the same contacts in N-terminal region

Table 2.3 PCC and RMSD between experimental and calculated J-coupling values

Parameter Set	Vuister [266]		Rosenman [215]		Rosenman optimized	
	PCC	RMSD (Hz)	PCC	RMSD (Hz)	PCC	RMSD (Hz)
Water	0.76	1.00	0.78	1.06	0.78	0.02

as pure water, but disrupts the contacts in R3-R4 and C-terminal regions as shown in Figure 2.4e-upper corner. The R1-R5 contacts are also diminished, aligned with their significant decrease in hydrogen bonding and β-content. Contact probability map of 4AP, Figure 2.4e-lower corner, show the anti-diagonal R3-R4 contacts are more abundant than other ensembles which is due to their higher hydrogen bonds in these regions. 4AP also illustrates higher interactions between R2-R5 which is less observed in the other ensembles.

The contact difference map of NaCl and water, Figure 2.4f-upper corner, shows a register shift toward C-terminal in anti-diagonal contacts of R3-R4 (manifesting as red-blue stripes) and another shift in termini region representing further interactions between R1-R5. The contact difference map of 4AP, 2.4f-lower corner, represents the higher R2-R5 interactions with more prevalent blue areas in C-terminal region and a reside shift in R3-R4 regions compared to pure water ensemble. The biggest increase in hydrogen bonds observed among all ensembles is in R3-R5 region of NaCl (black box on contact difference map) which is not observed in other ensembles. On the other hand, the decline in the interactions between regions R3-R4 and R3-R5 has been reported in other computational studies of Aβ_{42} inhibitors and is associated with decrease in Aβ_{42} dimerization and aggregation [54, 281].

Cluster Analysis. To gain insights into the Aβ_{42} structures favored by each molecule, representatives of the five highest-populated clusters are depicted in Figure 2.6. Cluster analysis was performed using gromos method with 3.5 Å RMSD cut-off

over the backbone atoms. Inspection of clusters with different cut-off values, lead to the choice of 3.5 Å cut-off which provides well-separated and highly diverse clusters with least number of shared structures between neighboring clusters. Top 10 clusters produced by this analysis, encompass 35-40% of the structures in the production run of each ensemble, which shows the convergence of REMD simulations and diversity of simulated ensembles. The Turn region (residues 22-30) shown in cyan in Figure 2.6, mostly adopts coil or bend structure while the rest of the peptide forms β-strands with occasional α-helices. Although the clusters are rich in β content, each cluster shows unique characteristics.

In pure water, dominant interactions in cluster S1 are between N- and C-terminal, more specifically, R1-R2-R5 regions which form a sheet of three anti-parallel β-strands. The CHC region (R3) either interacts with R5, forming anti-parallel and parallel β-hairpins or it is isolated by adopting an α-helix conformation. Aβ_{42} in pure water forms mostly 3-stranded β-sheet conformations with exception of less collapsed S5 in which R3 and turn region form a single hairpin.

In NaCl, anti-parallel β-hairpins between region R1-R5 are observed similar to pure water but in separate hairpin conformations rather than 3-stranded ones (Figure 2.6) which is in agreement with higher β content observed in the R1 and R5 regions. The CHC region is more active compared to pure water by forming anti-parallel β strands with R5 and R4 regions due to higher hydrogen bonding propensity between these parts. A characteristic feature of NaCl ensemble is double anti-parallel β-hairpins as seen in S1 and S3. In cluster S1 a β-hairpin is observed between R3 and R4 (red and purple arrows in Figure 2.6), corresponding to the strand-loop-strand (SLS) structure which has been suggested to be an important intermediate structure to fibril elongation [12, 21, 93, 94, 149, 195, 260]. In this conformation, residues 17-23 and 27 to 33 -the two strands- are connected by a loop region while hydrophobic residues are buried inside the hairpin. Another interesting hairpin conformation that

occurs in NaCl involves regions R3 and R5, depicted in clusters S3 and S4. These structures which we call strand-loop-loop-strand (SLLS), involve residues 17-22 and 37-42 in β-strand and residues 23-36 in loop conformation. This structure happens more frequently in NaCl ensemble rather than SLS.

Inositol adopts less ordered structures with β-hairpins forming between R3-R4 regions in most of the clusters. R1 and R2 regions interact closely with each other and the peptide forms barrel-like conformations with a solvent accessible core. The regions interact less with each other which is expected from the decrease in hydrogen bonds and lower C-terminal secondary structures of the ensemble. SLS structures in Inositol are not fully developed and are limited to only 2-3 residues in strand conformation.

Despite having high β content, structures of 4AP clusters have a different characteristic compared to other ensembles: they adopt ordered structures with only anti-parallel conformations and mostly sequential β-hairpins (i.e., R1-R2, R2-R3, R3-R4 or R4-R5) with the occasional hairpins between N- and C-terminal regions. As observed in the clusters and in-line with changes in hydrogen bonding, the N-terminal contacts, e.x., R1-R2 and R2-R3 regions, are less frequent -except for S2- and interactions among residues are shifted toward the middle and C-terminal of Aβ_{42} compared to other ensembles.

Comparison to Experiment. Identifying the ensemble of Aβ_{42} structures that can promote the aggregation is beneficial in developing new drug molecules for AD. In order to measure the validity of structures identified by REMD simulations, we have calculated the J-coupling of pure water ensemble and compared our results with experimental data obtained from NMR spectroscopy [215] in Figure 2.5. The calculated J-couplings from simulations using Vuister parameters [266] show a good match with the experimental values between residues 3-4, 18-19 and the C-terminal as observed in Figure 2.5a. The higher β-content in central hydrophobic core (R3)

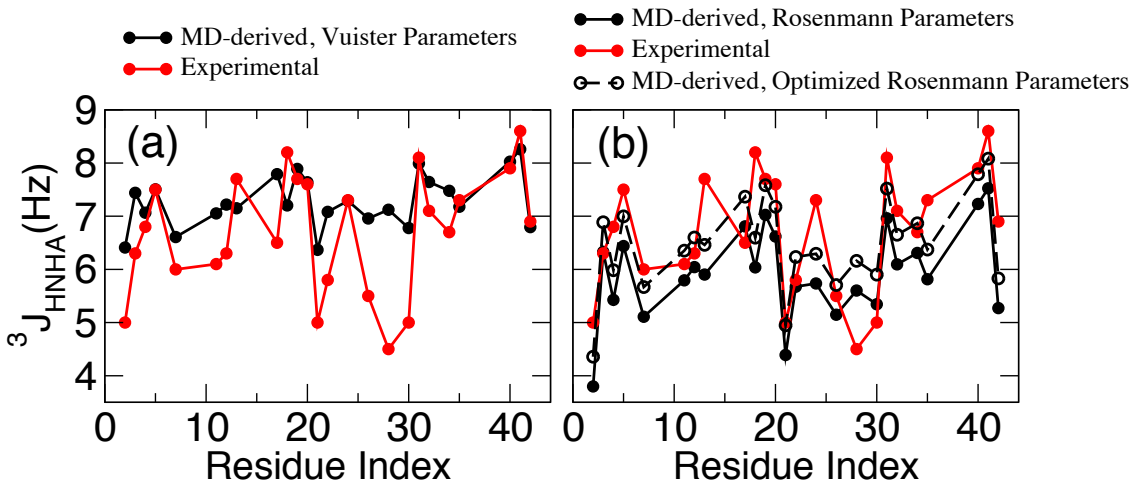


Figure 2.5 Comparison of experimental and MD-derived $^3J_{HNHA}$ values over the production run of pure water ensemble at 320 K. (a) MD-derived J-coupling values calculated using Vuister parameters [266] (black) and experimental values (red) obtained from ref. [215]. (b) MD-derived J-coupling values calculated using Rosenman parameters (solid black), experimental data obtained from ref. [215] (red) and optimized J-couplings calculated from simulations (dashed black) as explained in the text.

is also reflected in these results as a peak at residues 17-20 which is consistent with experimental data. The discrepancies between simulated and experimental values observed at residues 20-30 may be due to the fact that Vuister parameters in Karplus equation are not optimized for Aβ_{42} protein. Indeed, better agreement between simulated and experimental results have been observed using a set of values for A, B and C parameters in Equation 2.1 which are optimized for Aβ_{42} [215]. In Figure 2.5b, we show the MD-derived $^3J_{HNHA}$ values for pure water ensemble using parameters obtained by Rosenman *et al.* [215] in solid black line which provides a better agreement between experimental and our simulated values. To better visualize the consistency between experimental J-couplings and our MD-derived data using Rosenman parameters, the optimized values which minimize the average difference between the two data sets are shown in dashed black line in Figure 2.5b.

Table 2.3 shows the correlation and RMSD between the MD-derived J-couplings and experimental values using different Karplus parameter sets. For both data set

parameters, good agreement between experiment and simulation is achieved with PCC values greater than 0.7 which shows a strong correlation between the two data sets. Previous computational studies on Aβ_{42} using OPLS-AA and TIP3P water model reported PCC values below 0.58 [215]. This provides further proof that the combination of AMBER99SB-ILDN force field and TIP3P water describes the experimental properties of Aβ_{42} better than the other combinations of force field and water model. It is also observed that the RMSD between simulation and experiment are significantly decreased with optimized values of the Rosenman parameters whereas the PCC remains consistent.

2.4 Discussion

Decreasing the aggregation rate of Aβ plays an important role in treatment of Alzheimer's disease. This can be accomplished by inhibiting the aggregation-prone conformations of Aβ. However, identifying such structures have proved to be a difficult task due to numerous ways in which Aβ aggregates. Existence of different aggregation pathways may be related to the monomeric state of this peptide which forms an ensemble of structures instead of folding into a native conformation. Hence, investigating the structures adopted by Aβ and determining the parameters that affect the monomeric state of the peptide is critical in finding a cure for Alzheimer's.

Here, we have studied the effect of aqueous solutions on monomeric structures of Aβ_{42}. Our results illustrate that the structures sampled by Aβ_{42} strongly depend on the solution of the monomer. More specifically NaCl which increases the aggregation rate of Aβ_{42}, enhances the interactions between regions R3-R5 and disrupts the R3-R4 interactions (shown in Figure 2.4a-c), resulting in a unique conformation called iSLS. The iSLS structure is not found in the five most populated clusters of the aggregation inhibiting molecules, i.e., 4AP and Inositol, whereas the higher R3-R4 interactions in these solutions favors the formation of SLS structure. We hypothesize that the iSLS

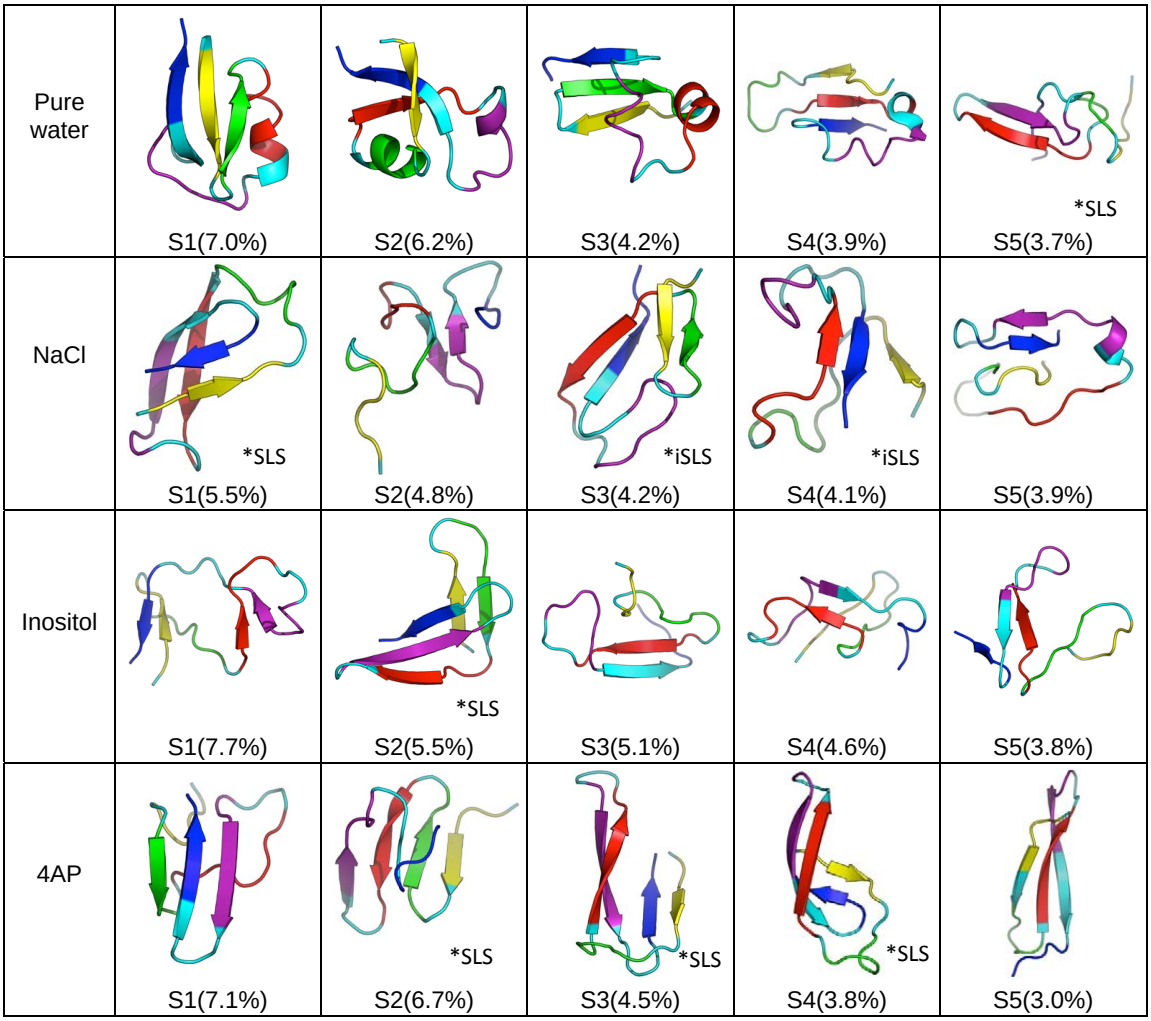


Figure 2.6 Representative structures of the five most populated clusters from gromos method cluster analysis using backbone RMSD cut-off of 0.35 nm. Arrows correspond to residues in β-strand conformation. Residues are colored according to regions introduced in Figure 2.3. Regions R1 through R5 are shown in colors as follows: residues 2-6: yellow, residues 9-14: green, residues 17-23: red, residues 27-33: purple, residues 37-42: blue and the rest are in cyan.

conformation may be another intermediate structure to Aβ_{42} aggregation because: (i) its formation is enhanced by NaCl which promotes Aβ_{42} aggregation while it does not from in ensembles with aggregation inhibiting molecules and, (ii) visual analysis of NMR-derived fibril structure of Aβ_{42} (PDBID: 2BEG [149]) suggests that side chain of residues G38 and V40 (region R5) form a hydrophobic core with side chains of residues L17 and F19 (region R3). This suggests that formation of iSLS structure may facilitate the transition from monomers to more ordered aggregates since residues that form hydrophobic core of the fibril are already closer to each other.

While we have identified the effect of various aqueous solutions on structure of Aβ_{42} and identified a possible aggregation-prone conformation, limitations of this study should also be mentioned. The results obtained by computer simulations depend strongly on the accuracy of their force fields. Among the widely used force fields for biomolecular simulations and various water models, we have chosen AMBER99SB-ILDN and the TIP3P water model because it has been proven to reproduce several structural quantities measured experimentally [147,196]. Accordingly, we have validated structures obtained from our REMD simulations with experimentally derived J-couplings (Figure 2.5) and found strong correlations (PCC > 0.75) with experiments. However, further investigation with different sets of force fields and water models are necessary to eliminate the bias of force field parameters on the structures of Aβ. The extensive computational resources needed for that purpose (more than$\sim 45\ \mu s$ per force field per aqueous solution) has stopped us from accomplishing this task. It is also desirable to study a broader set of small compounds to identify other possible structures that promote Aβ_{42} aggregation. In this study, we have focused our attention to the three molecules for which experimental evidence were available [2, 56, 166, 167, 240].

CHAPTER 3

THERMODYNAMIC STABILITY OF POLAR AND NON-POLAR AMYLOID FIBRILS

3.1 Introduction

Peptide self-assembly into cross-β fibril structures has important implications for plaque formation in amyloid diseases that include Alzheimer's and Parkinson's [41,69]. Accordingly, this phenomenon has been the subject of intensive studies to provide insights into the critical interactions that need to be targeted by drugs to avoid plaque formation. Interactions between backbone atoms, which are common to all peptides, play an important role in accounting for the superior mechanical strength of fibrils [127, 131] and they may explain the universal nature of fibrils that can form from seemingly unrelated amino acid sequences given the right conditions [78, 80]. Side chain interactions modulate the rate of fibrillization which increases with the hydrophobicity and the β-sheet propensity of the peptide sequence [42]. These interactions may also play an important role in accounting for the thermodynamic stability of cross-β structures as shown in alanine scanning mutagenesis experiments [272]. In such experiments, the free-energy (ΔG) to add an Aβ_{1-40} peptide to a fibril changed by up to $\sim$2 kcal/mol when a single residue was mutated to alanine. It is important to note that equivalent ΔG values can emerge from different combinations of enthalpy (ΔH) and entropic energy (-TΔS), i.e., $\Delta G = \Delta H - T\Delta S$. Knowledge of ΔH and $-$TΔS can provide insights into the stabilizing mechanisms of fibrils since hydrophobic interactions of small non-polar side chains are mainly related to ΔS, whereas ΔH emerges mainly from direct pairwise interactions, e.g., van der Waals and electrostatic interactions [36, 136, 277]. Thus, thermodynamics provides a framework to quantify fibril stability and the interactions accounting for it.

Albeit commonly used to study protein folding, equilibrium thermodynamic quantities of mature amyloid fibrils are not easily accessible experimentally and they

remain largely unknown [71,107,171]. Only recently have experiments shown that for some protein sequences, fibrils grow to an equilibrium state in which they coexist with dissolved proteins [186]. The threshold concentration of proteins dissolved in solution below which fibril nucleation cannot occur [121] has been explored to measure ΔG and to discover effects of individual amino acids on the stability of fibrils [70, 232, 233, 271]. Studies of the temperature dependence of this equilibrium can also be used to compute other thermodynamic quantities, e.g., ΔH, ΔS, and changes in heat capacity ΔC_p [38, 107, 253, 262]. Studies of β-sheet association provide evidence that the molecular mechanisms accounting for fibril stability depend on the peptide sequence [133, 208, 256]. In particular, the enthalpically unfavorable desolvation of preformed β-sheets made from polar peptides (Sup35) could be the rate limiting process of their association, whereas entropic effects related to hydrophobic interactions could favor the association of β-sheets made from non-polar peptides. It is important to note that while effects of temperature on ΔG are not well understood, higher temperatures have been shown to affect the kinetics of some amyloid fibrils by significantly increasing their nucleation and growth rates [263].

The spontaneous addition of peptides to fibrils has been studied through computer simulations. This process was shown to occur in at least two steps wherein peptides dock into fibrils before locking into them via nonspecific hydrogen bonds [33, 77, 86, 181, 206, 226, 252, 253]. To compute ΔG, different simulation setups based on the dissociation of peptides from known fibril structures have been used [141, 203, 215, 253]. Extensive sampling is required to account for equilibrium ΔG where $\Delta G \equiv -\Delta G_{\text{dissociation}} = \Delta G_{\text{association}}$. Recently, ΔG values computed from simulations have been shown to be in reasonable agreement with experiments for the Aβ_{9-40} fibril [227]. These ΔG emerge from a favorable entropic contribution and a small non-favorable enthalpy. Rationalization of these results require an

understanding of how the amino acid sequence and fibril structure account for equilibrium thermodynamic quantities.

In this dissertation, we compute potentials of mean force (PMF) to add monomeric peptides to non-polar and polar fibrils. PMF(ξ) corresponds to the free-energy to bring a peptide from non-interacting distances to a distance ξ from the fibril. This quantity is computed at different temperatures in order to provide estimates for ΔH and $-$TΔS. Simulations are performed by combining two enhanced sampling methods, i.e., Replica Exchange Molecular Dynamics (REMD) and Umbrella Sampling (US). REMD is used to improve sampling of the different US windows, thus, providing equilibrium ensembles of peptide structures around the fibril at different temperatures. The weighted histogram analysis method (WHAM) is used to compute PMF(ξ) at different temperatures from the different ensembles. We anticipate that the methodology used here will become popular as the intensive computational resources required to perform the simulations are becoming more widely available to researchers. The combination of REMD and umbrella sampling presents the advantage of being easily generalized to other systems as it requires little prior knowledge of the system being study, and it provides equilibrium conformational ensembles at different temperatures.

We find that the non-polar fibril studied in this dissertation becomes more stable with increasing temperature. At first sight, this result is counter-intuitive as solid materials tend to become less stable with increasing temperature, and not the opposite. We show that hydrophobic interactions in the core of non-polar fibrils are responsible for this non-conventional dependence of stability on temperature. In contrast, the polar fibril becomes less stable with increasing temperature. Thus, our results suggest that the stability of fibrils can be tuned by carefully choosing the amino acid sequence in the dry core of the fibril. Accordingly, one may envisage fibrils being used as thermosensors that will fall apart whenever a given temperature T_c is reached.

To the best of our knowledge, this is the first computational study to investigate the effect of temperature on the thermodynamic stability of amyloid-like fibrils using all-atom molecules and explicit solvent. Moreover, the combination of REMD and US allows us to compute enthalpic and entropic energies with small uncertainties for both non-polar and polar fibrils. We find that the former is stabilized by entropic energy and the latter by enthalpy. This suggests, that similarly to the thermodynamic theory of protein folding, it may be possible to develop a thermodynamic theory for fibril growth wherein the addition of a peptide to a fibril accounts for specific changes in enthalpy, entropic energy, and heat capacity. However, this will require the study of other fibrils from different amino acid sequences as well as polymorphic fibril structures. This study provides a proof of concept in that direction and shows a new methodology that can be used for that purpose.

3.2 Methods

All simulations are performed using GROMACS 4.6 with AMBER99sb-ILDN force-field and TIP3P water [101]. The initial structure of the non-polar Aβ_{16-21} fibril (sequence KLVFFA) is based on the PDB entry 3OW9 which contains twelve anti-parallel peptides [47]. To reduce the computational cost of the simulation while allowing the free peptide to interact with the solvent accessible sides of the fibril, we retain six of the twelve peptides in the simulation box–see Figure 3.3a. The polar IAPP$_{28-33}$ fibril (sequence SSTNVG) was constructed by extending the PDB entry 3DG1 [273] to account for a cross-β structure made of six chains with parallel β-sheets as determined experimentally [175,273]–see Figure 3.3b. Simulation boxes of Aβ_{16-21} and IAPP$_{28-33}$ fibrils are solvated with 7379 and 7300 water molecules and the net charge of the Aβ_{16-21} system is neutralized by adding six Cl^- ions. These initial fibrils were relaxed for 2 ns at 300 K and 1 atm to remove unrealistic contacts.

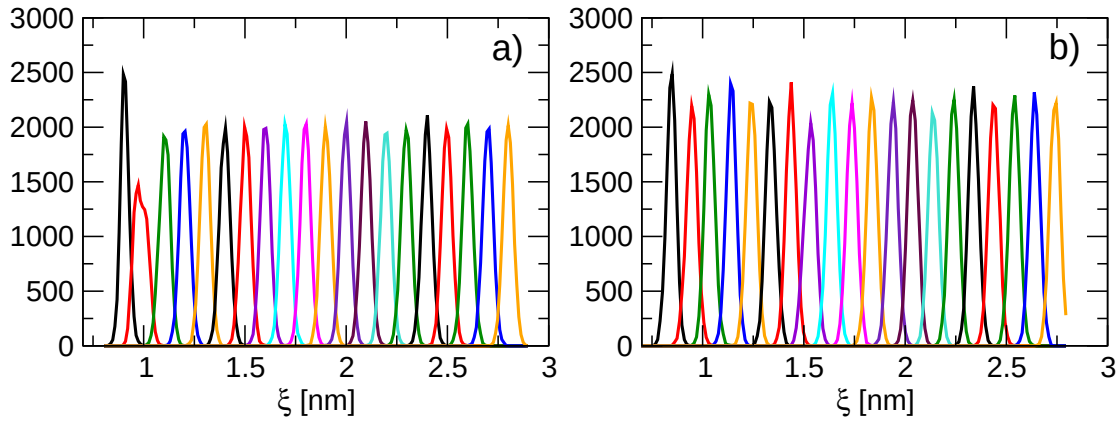


Figure 3.1 Histograms of replica exchange MD simulations for (a) Aβ_{16-21} and (b) IAPP$_{28-33}$ at 320 K showing reasonable overlap between neighboring replicas. Colors represent individual replicas and y-axis shows the number of occurrence during the last 25 ns of simulation.

Simulations in this work are performed using the leap-frog algorithm with a time-step of 2 fs to integrate the equations of motion. The neighbor list is updated every 10 steps and bonded interactions are constrained using the LINCS algorithm. A Lennard-Jones cutoff of 1.4 nm is used. Electrostatics is treated using the Smooth Particle Mesh Ewald method with a grid spacing of 0.12 nm and a 1.4 nm real-space cutoff. Temperature is controlled using the velocity-rescale thermostat (τ_T=0.1 ps) and the pressure is set to 1 atm using the Berendsen-coupling barostat (τ_P=1.0 ps). Each of the restrained ξ-distances for Abeta$_{16-21}$ and IAPP$_{28-33}$ are sampled using REMD for 75 ns. Thirty-two replicas are chosen in the temperature range 290 K to 373 K such as to account for an average exchange rate of approximately 20 %. Exchanges between neighboring replicas are attempted at every 750 steps. In Figure 3.1, we show the overlap between replicas for both ensembles as a function of ξ. Figure 3.2 illustrates the convergence of simulations by calculating the PMF over 10 ns time intervals. Aβ_{16-21} and IAPP$_{28-33}$ systems reach equilibrium after 30 ns and 40 ns of simulation time.

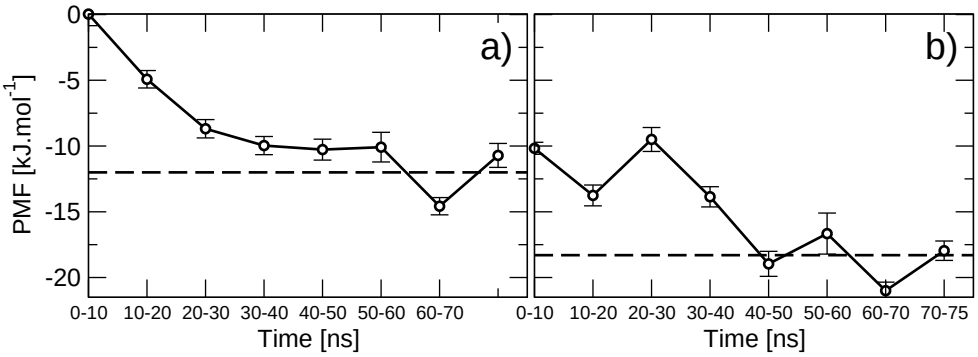


Figure 3.2 PMF at the global minimum averaged over time intervals of 10 ns for (a) Aβ_{16-21} and (b) IAPP$_{28-33}$ at 320 K. Dashed lines correspond to the average over the last 25 ns, i.e., average over 50-75 ns. Notice that the PMF averaged over 40-50 ns is already very close to the dashed line.

A two step Umbrella Sampling protocol [141] combined with replica exchange molecular dynamics is used to study fibril dissociation at different temperatures. Notice that US (without REMD) is commonly used to study free-energies to add peptides to a fixed-main chain fibril [108, 142, 203, 221, 227]. The combination of US and REMD enhances sampling of the phase space and it provides estimates of the PMF at different temperatures which we use to compute enthalpy and entropic energy. In these simulations, heavy atoms of five chains of the initial fibril are restrained to their initial positions by a spring with constant 1000 kJ mol^{-1} nm^{-2}.

In the first US step, the C$_\alpha$ atom of the N-terminal of the "free" peptide is pulled away from the center-of-mass of its closest three chains along the one-dimensional direction of the fibril axis (ξ_z). From these 10 ns steered molecular dynamics simulations, 20 and 19 configurations for Aβ_{16-21} and IAPP$_{28-33}$, respectively, were extracted along the dissociation pathway. The ξ_z distance between peptide and fibril in these extracted configurations were in the range of 0.9–2.8 nm for Aβ_{16-21} and 0.94–2.74 nm for IAPP$_{28-33}$ with 0.1 nm increments for both systems. The goal is to

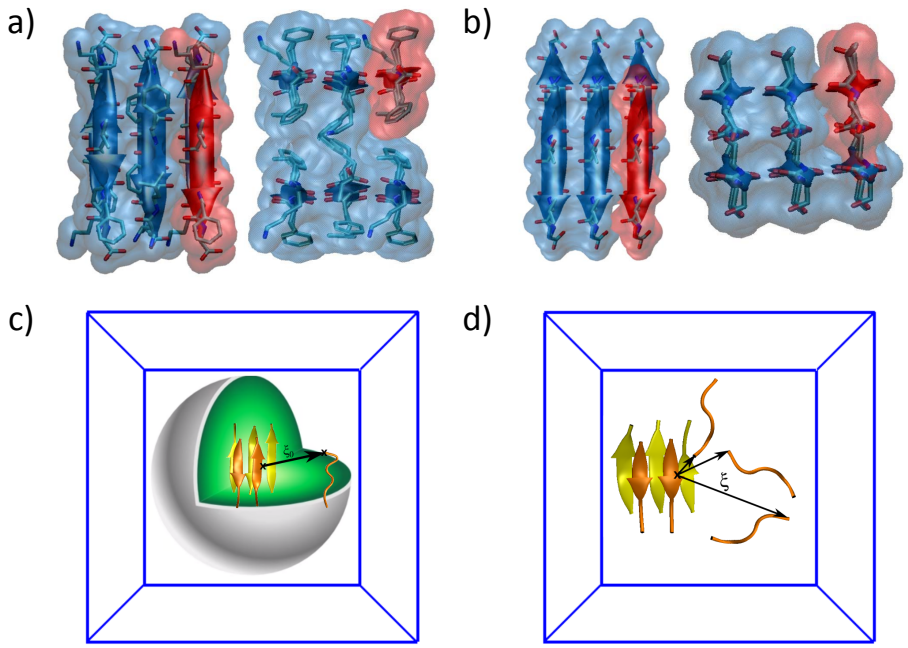


Figure 3.3 Schematic representation of the simulation setup showing the reaction coordinate ξ for (a) the anti-parallel Aβ_{16-21} fibril and free peptide overlapped on a sphere with radius ξ, illustrating the three-dimensional surface (grey) on which the free peptide can move, and (b) the parallel IAPP$_{28-33}$ fibril with free peptide at three different ξ values. The two β-sheets are shown in yellow and orange. Water molecules are not shown for clarity. Top view and cross-section of (c) Aβ_{16-21} and (d) IAPP$_{28-33}$ fibrils. Van der Waals surfaces of the fibril and free peptide are shown in blue and red, respectively. Residues forming the dry core of Aβ_{16-21} (F and L) and IAPP$_{28-33}$ (S and N) fibrils are highlighted.

use these structures as initial configurations to sample the system along the reaction coordinate.

We define the distance ξ between the C_α atom of the N-terminal of the "free" peptide and the three closest chains of the fibril as our reaction coordinate for the second step of umbrella sampling–see SI. The selection of ξ in this way limits the conformations of the free peptide to move on a sphere with radius ξ at each window–see Figure 3.3c. To avoid potential biases introduced by the initial structure of the "free" peptide, we removed all the water from the simulation box and performed a manual random rotation of the peptide around its C_α atom of N-terminal residue, avoiding steric collisions with the fibril. This rotation was performed once for each

window of both systems while keeping the distance ξ fixed. Examples of rotations for three ξ values are shown in Figure 3.3d. The configurations with rotated free peptide were solvated and equilibrated for 2 ns with position restraints on heavy atoms.

In the second US step, a spring with constant of 4000 kJ mol^{-1} nm^{-2} is used to restrain initial ξ distances of the 20 Aβ_{16-21} and 19 IAPP$_{28-33}$ windows. These systems are simulated using REMD [229] for 75 ns each. The last 25 ns of the trajectories is used to compute PMF(ξ) at each of the 32-temperatures (ranging from 290 K to 373 K) using the weighted histogram analysis method [135]. Details of the simulation and analysis of the convergence of the PMF are provided in the Supporting Information (SI). Notice that the PMF increases with $-k_bT\log(\xi^2)$ due to the three-dimensional nature of ξ. We subtract this dependence of the PMF on ξ and the PMF is shifted to zero at $\xi = 2.7$ nm.

3.3 Results

3.3.1 Reaction Coordinate and Cluster Analysis of Locked States

Different reaction coordinates ξ can be used to describe the pathway of peptide dissociation from fibrils. A commonly used reaction coordinate is the distance ξ_z along the fibril axis between center-of-mass of the fibril and center-of-mass of the free-peptide. In simulations that make use of ξ_z, the peptide starts locked into the fibril and it is pulled away from the fibril using steered molecular dynamics. From the trajectory of this steered simulation, conformations along ξ_z are selected and used as starting conformations for umbrella sampling. Because the initial structure in these simulations is the locked state, this protocol may account for PMF that are biased towards the fibril state.

In our simulations, initial conditions for each window are prepared to avoid biases towards the fibril state. This was achieved by manually rotating peptide conformations selected from the steered simulation-see discussion in main manuscript.

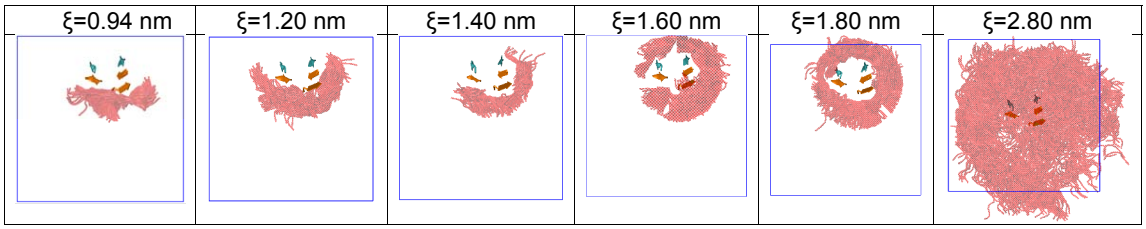


Figure 3.4 Conformations sampled by the Aβ_{16-21} peptide at six different ξ values. Overlaid structures of the free peptide are represented in red. The five peptides forming the fibril are represented in a cartoon representation. ξ corresponds to the distance between the center-of-mass of the three peptides of the fibril that are colored in orange and the Ca atom of the N-terminal of the free-peptide.

This rotation was performed without changing significantly the value of the reaction coordinate. Notice that, when using ξ_z as the reaction coordinate, it is not possible to produce a significant rotation of the peptide without having it overlap with the fibril. However, such a rotation is possible using other reaction coordinates. In our simulations, we use the distance ξ between center-of-mass of the three closest chains in the fibril and the C_α atom of the N-terminal of the free-peptide as our reaction coordinate. Using this reaction coordinate, it is possible to rotate the peptide while keeping ξ fixed. Moreover, independently of the initial structure, we find that the peptide samples the locked state in simulations close to ξ=0.94 nm. Notice that an alternative reaction coordinate is the distance ξ' between center-of-mass of the three closest peptides in the fibril and the C_α atom of the C-terminal of the free-peptide.

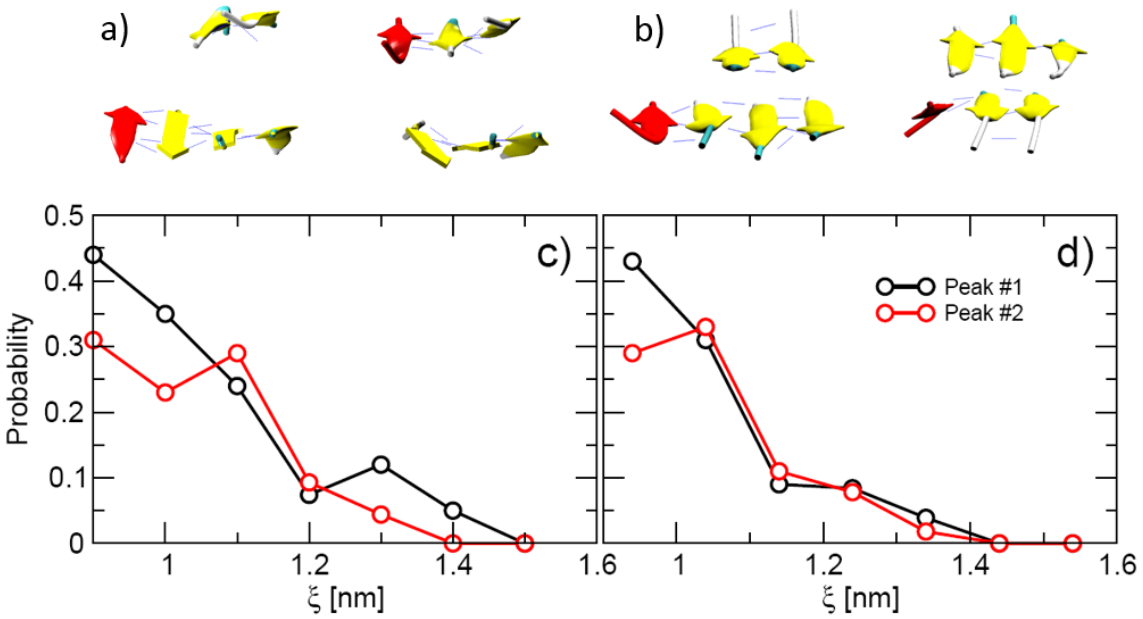


Figure 3.5 Locked states of Aβ_{16-21} and IAPP$_{28-33}$ fibrils. In the locked state, the free peptide can be incorporated into the β-sheet of the fibril made of three peptides (a-b, left), or into the β-sheet that contains two peptides (a-b, right). These locked conformations are represented in panels (a) and (b) for the Aβ_{16-21} and IAPP$_{28-33}$ peptides, respectively. The probability of peptide being in β-sheet made of three peptides (black) and being in β-sheet containing two peptides (red) is shown in panels (c-d). Cluster analysis (using a RMSD cutoff of 2.0 Å for Cα atoms) was used to determine the probability of finding peptides in the two locked states. This analysis was performed for conformations at 320 K.

Since our choice of ξ is not commonly used in the literature, we show in Figure 3.4 the configurational space of the peptide of Aβ_{16-21} at six different ξ values: At ξ=0.94 nm, the peptide interacts only with the tip of the fibril. Moreover, in Figure 3.5 we show through cluster analysis that approximately 70% of the conformations sampled by the peptide at ξ=0.94 nm correspond to locked states. As ξ increases from 1.20 nm to 1.80 nm, the peptide interacts with the solvent exposed surface area of the fibril.

At ξ=2.80 nm, the peptide is detached from the fibril. Thus, ξ describes a pathway in which the peptide evolves from a state in which it is interacting with

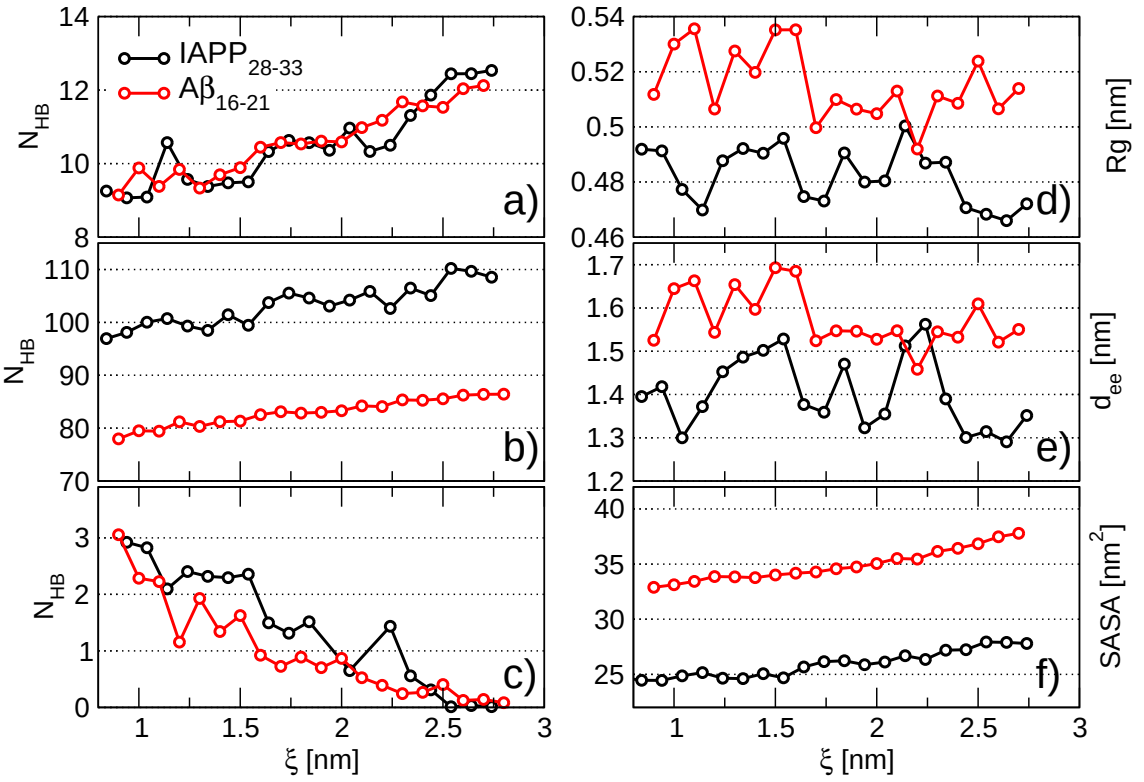


Figure 3.6 Structural properties of Aβ_{16-21} (red) and IAPP$_{28-33}$ (black) along the reaction coordinate at 320 K. Number of hydrogen bonds between (a) "free-peptide"-water, (b) protein (i.e., all six peptides)-water and (c) main chain atoms of fibril (i.e., five peptides)- "free-peptide". (d) Radius of gyration, i.e., Rg, and (e) end-to-end distance, i.e., dee, of the "free-peptide". (f) Solvent Accessible Surface Area (SASA) of protein (i.e., all six peptides).

the tip of the fibril in a locked configuration to a state where it is detached from the fibril. In between these two states, e.g., at ξ=1.80 nm, the peptide is found to interact significantly with the side chains of the fibrils that are exposed to the solvent–see Figure 4d of the manuscript.

3.3.2 Structural Quantities Upon Fibril Dissociation

Figure 3.15 shows that the molecular interactions stabilizing Aβ_{16-21} and IAPP$_{28-33}$ fibrils are significantly different from each other. In particular, the former is stabilized by entropic energy while the latter is stabilized by enthalpy. In Figure 3.6, we show that structural quantities measured upon peptide dissociation from Aβ_{16-21}

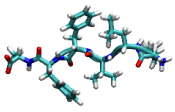
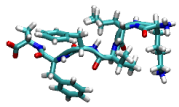
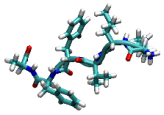
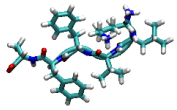
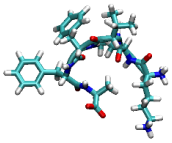
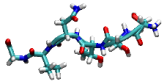
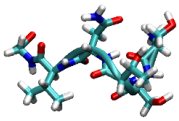
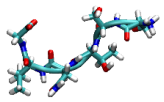
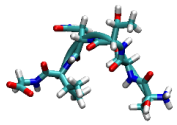
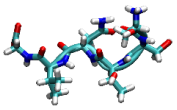
	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
$A\beta_{16\text{-}21}$	 10.7 % Rg=0.64 nm	 8.01 % Rg=0.57 nm	 6.27 % Rg=0.60 nm	 4.79 % Rg=0.57 nm	 3.94 % Rg=0.53 nm
$IAPP_{28\text{-}33}$	 15.3 % Rg=0.56 nm	 13.01 % Rg=0.45 nm	 9.53 % Rg=0.53 nm	 7.67 % Rg=0.46 nm	 6.97 % Rg=0.53 nm

Figure 3.7 Five most populated clusters of single peptide simulations in explicit water for 100 ns.

and $IAPP_{28-33}$ fibrils do not provide much insight into these different molecular interactions. This justifies the need to compute thermodynamic quantities in this work.

It is interesting to notice the weak dependence of Rg and dee on ξ-see Figure 3.6(d-e). In particular, one would expect the $A\beta_{16-21}$ peptide to become significantly more compact upon dissociation due to the non-polar nature of its side chains. However, its Rg decreases by less than 0.03 nm upon dissociation. To provide some insights into this weak dependence of Rg and dee on ξ, we performed two additional 100 ns simulations at 330 K consisting of a single $A\beta_{16-21}$ peptide in pure water and a single $IAPP_{28-33}$ peptide in pure water. The structure of peptides in these new simulations is studied in Figure 3.7 through cluster analysis. This figure shows that structures of the five most popular clusters of $A\beta_{16-21}$ or $IAPP_{28-33}$ peptides in pure water are significantly extended. The small size of these peptides as well as the bulkiness of the side chains may explain their extended nature. Accordingly, the $A\beta_{16-21}$ peptide, which has bulkier side chains than $IAPP_{28-33}$, is more extended-see Figure 3.7.

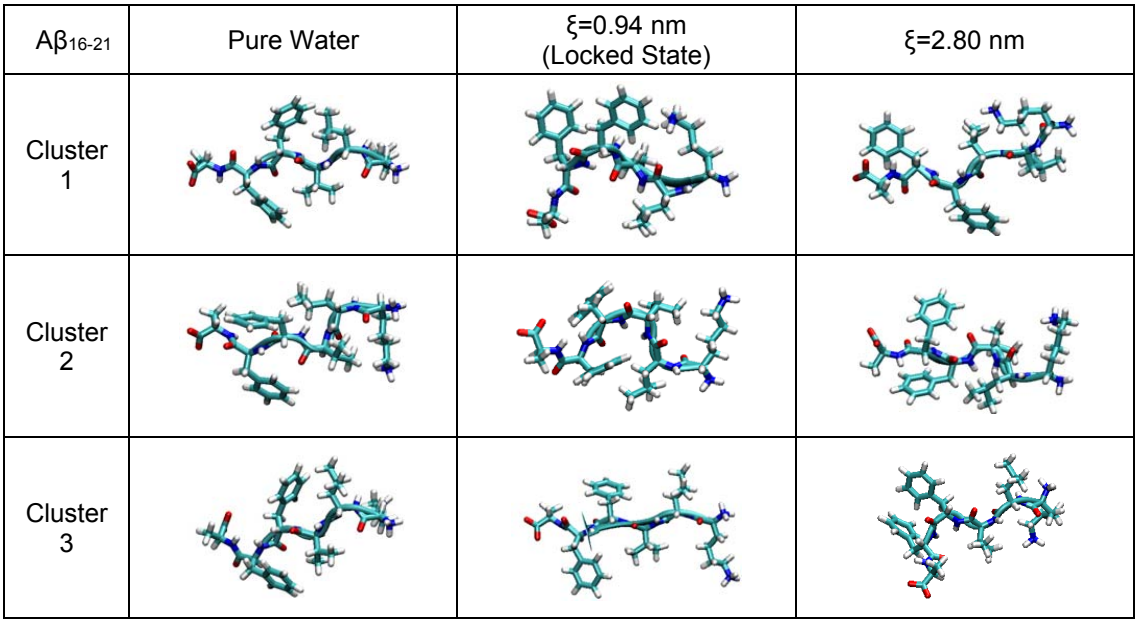


Figure 3.8 Cluster analysis of the Aβ_{16-21} in pure water, free peptide in locked state and free peptide at ξ=2.80 nm. Three most populated clusters for each system are shown and in all of them, the peptide is in an extended conformation.

Figure 3.8 provides a direct comparison of the structure of the Aβ_{16-21} peptide in pure water, in the locked state (i.e., ξ=0.94 nm), and at ξ=2.80 nm. Structures are shown for the three most populated clusters. In all three systems, the peptide is considerably extended.

3.3.3 Effect of Temperature Range on Computing Thermodynamic Quantities

In Figure 3.15 we show the dependence of ΔH and -TΔS on the reference temperature T_0. These quantities were computed by fitting the temperature dependence of the PMF to equation 3.1. Thirty-two PMF, computed in the temperature range 290 K to 373 K, were used in these fits. In Tables 3.1 and 3.2, we show that the use of a smaller temperature range already provides good estimates for ΔH and -TΔS. This smaller temperature range resembles more closely to the range of temperatures used experimentally, i.e., 273 K to 333 K, in references [107, 128].

Table 3.1 Effects of the Temperature Range in Computing ΔH, -TΔS, and ΔC for $A\beta_{16-21}$ at Physiological Conditions, i.e., 309.98 K.

$A\beta_{16-21}$				
Temperature range (K)	PMF kJ/mol	ΔH kJ/mol	-TΔS kJ/mol	ΔC kJ/mol/K
309.98 ± 10	-9.52	18.10 ± 12.12	-27.62 ± 12.12	9.2522 ± 4.18
309.98 ± 15	-9.89	44.10 ± 8.94	-54.00 ± 8.94	3.16 ± 2.11
309.98 ± 20	-10.18	46.14 ± 5.78	-56.33 ± 5.78	0.66 ± 1.03
290—373	-10.57	43.65 ± 4.07	-54.22 ± 4.05	-0.85 ± 0.18

Table 3.2 Effects of the Temperature Range in Computing ΔH, -TΔS, and ΔC for $IAPP_{28-33}$ at Physiological Conditions, i.e., 309.98 K.

$IAPP_{28-33}$				
Temperature range (K)	PMF kJ/mol	ΔH kJ/mol	-TΔS kJ/mol	ΔC kJ/mol/K
309.98 ± 10	-18.51	4.69 ± 6.87	-23.20 ± 6.87	-0.94 ± 2.37
309.98 ± 15	-18.57	-14.20 ± 6.65	-4.37 ± 6.65	-2.71 ± 1.57
290—373	-18.45	-14.76 ± 3.58	-3.69 ± 3.56	-1.44 ± 0.16

Note: The Use of a Temperature Range of 15 K Provides a Reasonable Estimate for ΔH and -TΔS in Tables 3.1 and 3.2.

3.3.4 Umbrella Sampling of $A\beta_{16-21}$ With and Without REMD

In Figure 3.9, we compare histograms and PMF of simulations performed using Umbrella Sampling alone, i.e., US without REMD (left panels), and Umbrella Sampling with REMD (right panels) at 330 K. Trajectories for US + REMD are from simulations discussed in the main manuscript. To compute the PMF from simulations using US alone, we used the same number of windows, spring constant, simulation time (i.e., 75 ns), and initial conditions for each window as in our US + REMD simulations. Notice that in our simulations, initial conditions for each window are prepared to avoid biases towards the fibril state whereas the free-peptide in most studies in the literature is initially locked into the fibril. Panels (a-b) show that there is enough overlap between neighboring histograms to allow the PMF to be estimated using the weighted histogram analysis method (WHAM) for both US and US+REMD. However, the PMF computed using US alone is far from being smooth

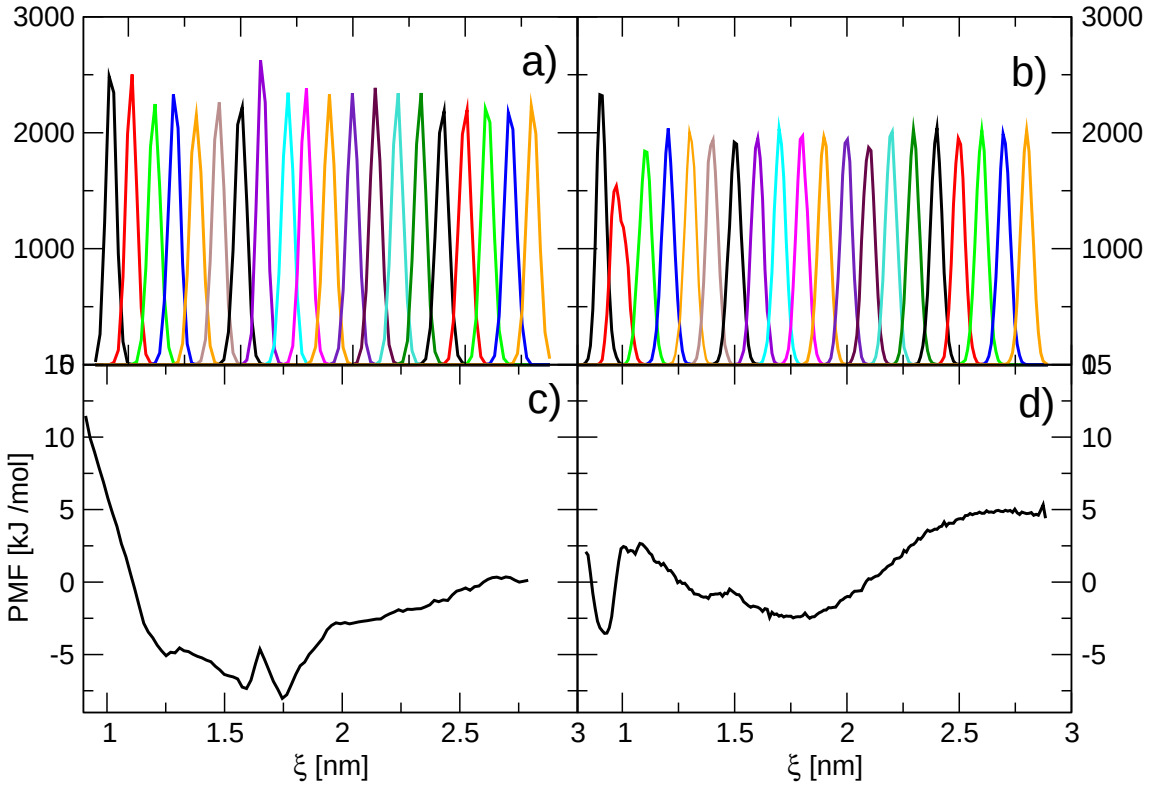


Figure 3.9 Histograms of Umbrella Sampling simulation (a) without and (b) with REMD. PMF computed from simulations performed (c) without and (d) with REMD.

and the ground state is not the locked state which occurs at ξ=1.0 nm. This suggests sampling problems in simulations performed with US alone.

To provide further insight into how REMD improves sampling, we extended the trajectory of the Umbrella Sampling simulation at ξ=1.80 nm to 300 ns. In the first row of Figure 3.10, we show overlaid conformations of the free-peptide for different time intervals. These conformations are compared with the ones from REMD in the second row of Figure 3.10. This figure shows that the free-peptide samples a larger phase space in 75 ns US + REMD simulation than in 300 ns of US (without REMD).

In the 75 ns simulation using US + REMD, the peptide samples the space of a cylindrical shell around the fibril. In the 300 ns of US (without REMD) simulations, the peptide samples approximately half of a cylindrical shell. Thus, the time required to sample the whole cylindrical shell may be estimated to be 600 ns, i.e., 8 times more

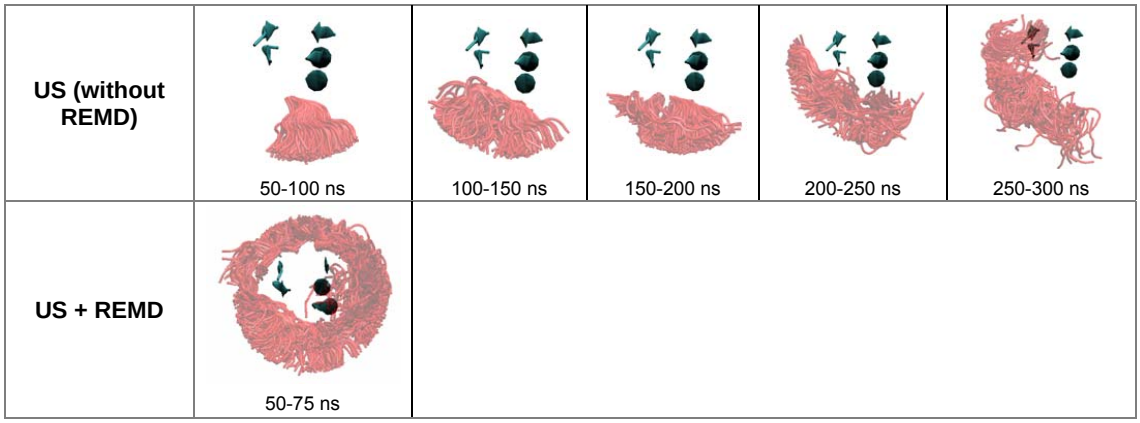


Figure 3.10 Sampling of normal MD simulation at different time intervals in a 300 ns simulation compared with REMD of 75 ns, both at ξ=1.80 nm and T=330K.

than the time require to sample the same space using US + REMD. This implies that using the computational resources required to estimate the PMF at 32 temperatures with US + REMD, one would only be able to compute the PMF at 4 temperatures using US (without REMD).

3.3.5 Error Analysis

In Figure 3.15 we computed ΔH and -TΔS by fitting the temperature dependence of the PMF to equation 3.1 via a least mean square fit. Estimates of the error of these fits are also provided. Here, we provide justification for the small errors reported in Figure 3.15. Notice that PMFs computed at the 32 temperatures in Figure 3.14 e-f are subjected to errors. These errors of 0.7 kJ/mol have been estimated using bootstrap analysis. To justify the small errors reported in Figure 3 for ΔH and -TΔS, we added random numbers within $\pm$ 1Standard Deviation to the PMF(ξ_0) at all 32 temperatures. For this new realization of the temperature dependence of the PMF, we computed ΔH, -TΔS and ΔC using least mean square fit to equation 3.1. This procedure was repeated 20 times which allowed us to compute the mean and standard deviation of ΔH, -TΔS and ΔC. These quantities are reported in Table 3.3.

Table 3.3 Mean and Standard Deviation of 20 Calculations After Adding a Random Error to All PMF Values at ξ=ξ_0 for Aβ and IAPP Fibrils.

	$A\beta_{16\text{-}21}$		$IAPP_{28\text{-}33}$	
T_0=309.98 K	Mean	SD	Mean	SD
ΔH (kJ/mol)	42.95	1.74	-14.57	2.00
-TΔS (kJ/mol)	-53.46	1.71	-3.91	1.96
ΔC (kJ/mol /K)	-0.79	0.058	-1.47	0.069

The mean value of ΔH, -TΔS and ΔC should converge to the reported values in Figure 3.15 when the number of repeats goes to infinity. For the 20 repeats performed here, mean values of ΔH, -TΔS and ΔC are already very close to the values reported in Figure 3.15. Notice that the standard deviation of these quantities is smaller than the error estimates in Figure 3.15.

3.3.6 Side Chain Analysis of Lysine in Aβ_{16-21}

All the residues of Aβ_{16-21} are non-polar except for Lysine (K) which is positively charged. This residue is made of non-polar and charged moieties-see inset of Figure 3.11. Characteristic conformations of the peptide indicate that the charged moiety of Lysine is mostly exposed to the solvent while its non-polar moiety is closer to the fibril allowing it to interact with exposed side chains of the fibril-see Figure 3.11. This is quantified in Figure 3.12 and it suggests that despite being a charged residue, lysine interacts with the fibril mainly via its non-polar moiety. This justifies our treatment of Aβ_{16-21} as being a non-polar peptide in the main manuscript

3.3.7 Side Chain Analysis of Asparagine in IAPP$_{28-33}$

Asparagine (N) is commonly found in peptides that are prone to form fibrils. In Figure 3.13, we investigate how it contributes to fibril stability by computing the fraction of hydrogen bonds between asparagine's side chain of the free peptide and

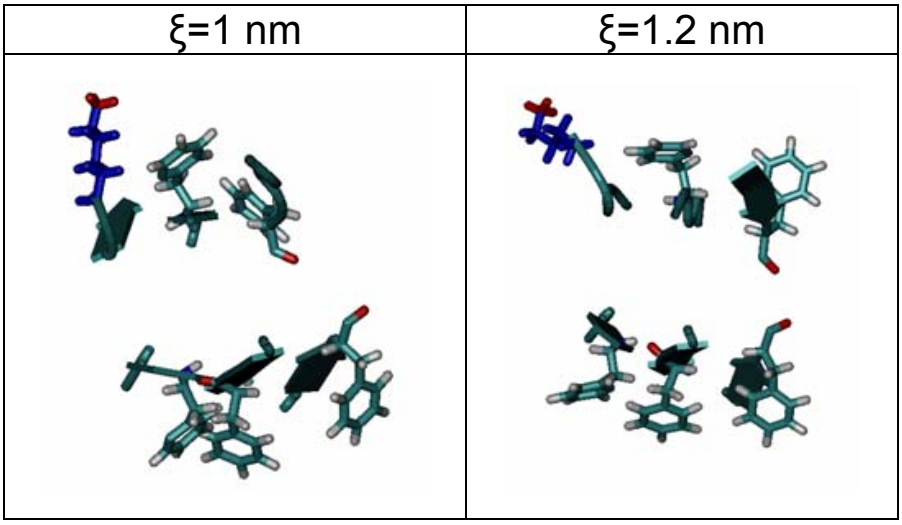


Figure 3.11 Characteristic conformation of the peptide-fibril system at ξ=1 nm and ξ=1.2 nm. Side chain atoms of lysine are shown in blue and red for its non-polar and charged moiety, respectively.

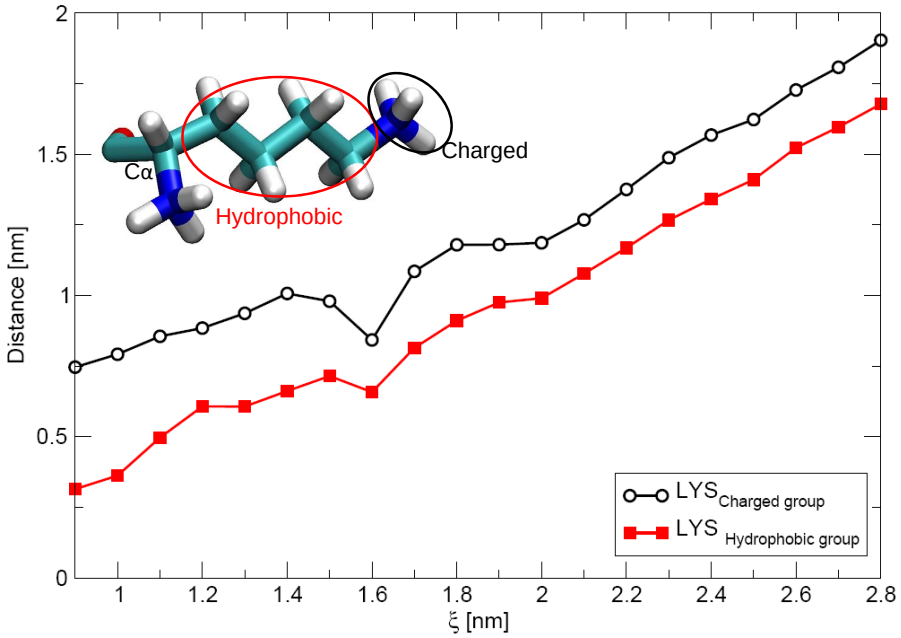


Figure 3.12 Distance between COM of charged and hydrophobic groups of Lys and COM of fibril backbone. This graph shows that the hydrophobic group of Lys interacts more closely with the fibril, while the charged group of Lys (NH3) points away from fibril backbone, interacting with water molecules.

side chains of different residues of the fibril. At short ξ distances, asparagine forms hydrogen bonds mainly with serine at position 29 (S29) and asparagine. In panel

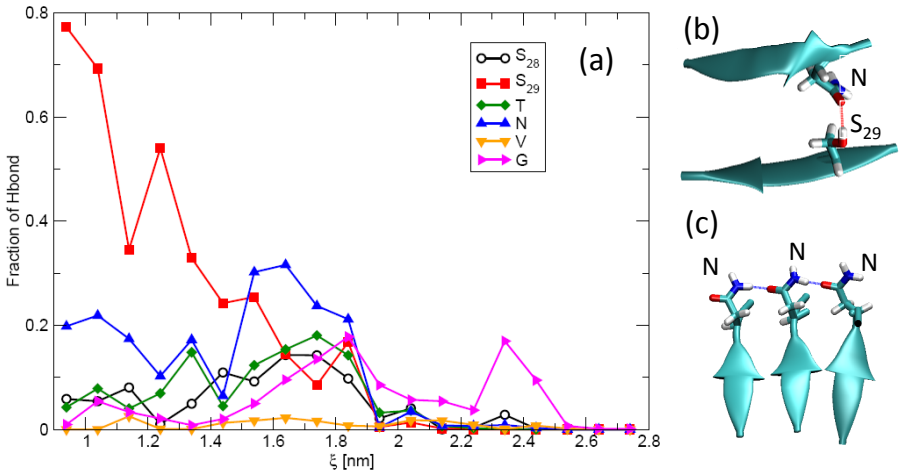


Figure 3.13 (a) Fraction of side chain hydrogen bond between Asn of the free peptide and other amino acids of the fibril. (b) Inter-β-sheet hydrogen bonds between N of the free peptide (peptide on top) and S29 (peptide at the bottom). (c) Intra β-sheet hydrogen bonds involving N of neighboring peptides. The right chain corresponds to the free peptide.

(b), we show that N-S29 hydrogen bonds are possible between peptides in opposing beta-sheets. In panel (c), we show that N-N hydrogen bonds occur between peptides within the same beta-sheet.

In Figure 3.14c-d, we study the PMF of Aβ_{16-21} and IAPP$_{28-33}$ fibrils along the reaction coordinate ξ at temperatures 309K, 330K and 360K in black, red and blue color respectively. Global minima of these PMF (dashed lines in Figure 3.14c-d) correspond to configurations in which the free peptide is locked to the fibril, i.e., the free peptide adopts a β-strand structure that is hydrogen bonded to the fibril. Cluster analysis (see Figure 3.5) shows that locked states are accounted for by two main structures: one where β-sheets are made of four and two peptides (Figure 3.14a) and the other where both β-sheets of the fibril are made of three peptides each (Figure 3.14b). The latter structure represents an "ideal fibril" where the peptide is hydrogen bonded to one β-sheet of the fibril and its side chains interact with the opposing β-sheet. Locked structures make up $\sim$70% of the conformations sampled by the peptide at $\xi_o = 0.94$ nm for Aβ_{16-21} and $\xi_o = 1.00$ nm for IAPP$_{28-33}$. At large distances ($\xi > 2.5$ nm), the peptide does not interact with the fibril.

The temperature dependence of the PMF at ξ_o, i.e., the locked state, is shown in panels e and f of Figure 3.14 for Aβ_{16-21} and IAPP$_{28-33}$ fibrils, respectively. These panels show that the non-polar fibril becomes more stable with increasing temperature while the opposite is observed for the polar fibril. At first sight, the temperature dependence of the Aβ_{16-21} fibril appears counterintuitive as thermal fluctuations reduce the stability of conventional material. However, this behavior is consistent with hydrophobic interactions becoming stronger with increasing temperature [66, 123]. In Figure 3.12, we discuss how the positively charged lysine in Aβ_{16-21} interacts with the fibril. Further insights into the role of hydrophobic interactions can be obtained by decomposing the PMF into enthalpy $\Delta H_o(\xi)$ and entropic energy $-T_o\Delta S_o(\xi)$ at the reference temperature T_o. These quantities, as well as the heat capacity $\Delta C_{po}(\xi)$ to add a peptide to a fibril can be obtained by fitting the temperature dependence of

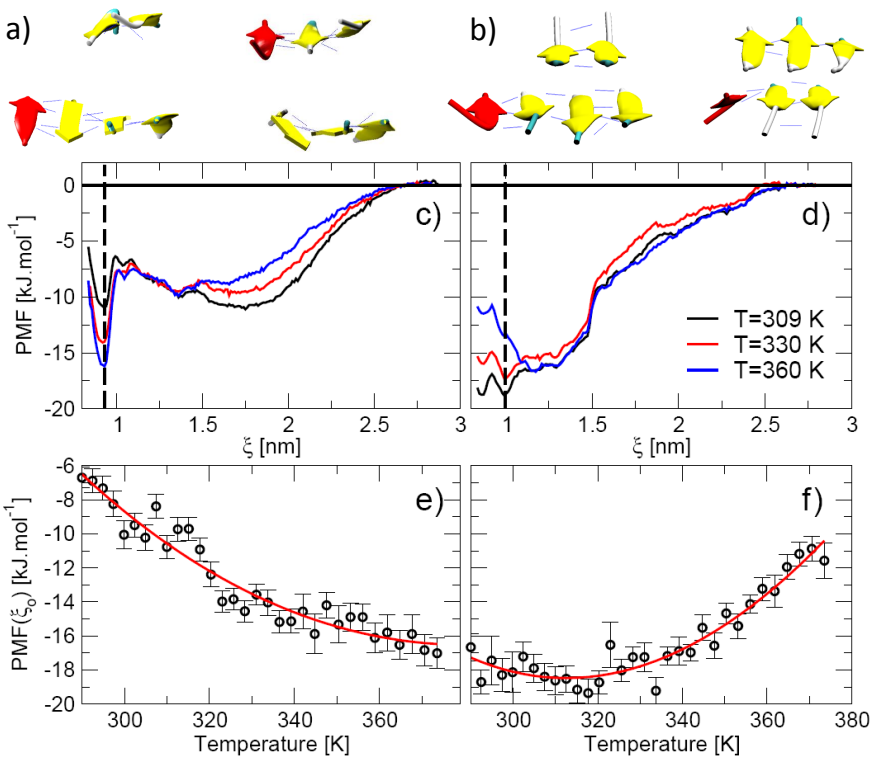


Figure 3.14 Characteristic configurations of the peptide (in red) locked into (a) Aβ_{16-21} and (b) IAPP$_{28-33}$ fibrils. Backbone hydrogen bonds are shown in dotted blue lines and β-sheet structures in the fibril are represented in yellow. PMF to add a peptide to the fibril along the reaction coordinate ξ for (c) Aβ_{16-21} and (d) IAPP$_{28-33}$ fibrils at 309K (black), 330K (red) and 360K (blue). Temperature dependence of the PMF at locked state for (e) Aβ_{16-21} and (f) IAPP$_{28-33}$ fibrils. Red lines correspond to fits of the PMF at ξ_o to equation 3.1 and error bars were computed using bootstrap analysis.

the PMF to $[63, 97, 200, 245, 280]$:

$$\begin{aligned} \mathrm{PMF}(\xi, T) = \; & \Delta \mathrm{H_o}(\xi) - T\Delta \mathrm{S_o}(\xi) \\ & + \Delta \mathrm{C_{po}}(\xi) \left[(T - T_\mathrm{o}) - T \log\left(\frac{T}{T_\mathrm{o}}\right) \right]. \end{aligned} \tag{3.1}$$

Lines in Figure 3.14e-f correspond to fits of the PMF at ξ_o to equation 3.1. These fits provide a good description of our simulation data as well as numerical estimates for $\Delta \mathrm{H_o}(\xi_o)$, $-\mathrm{T}\Delta \mathrm{S_o}(\xi_o)$, and $\Delta \mathrm{C_{po}}(\xi_o)$. The temperature dependence of $\Delta \mathrm{H_o}(\xi_o)$ and $-\mathrm{T_o}\Delta \mathrm{S_o}(\xi_o)$ are shown in Figure 3.15 and show that for the Aβ_{16-21} fibril, $-\mathrm{T_o}\Delta \mathrm{S_o}(\xi_o)$ favors the fibrillar state while $\Delta \mathrm{H_o}(\xi_o)$ opposes it at all temperatures,

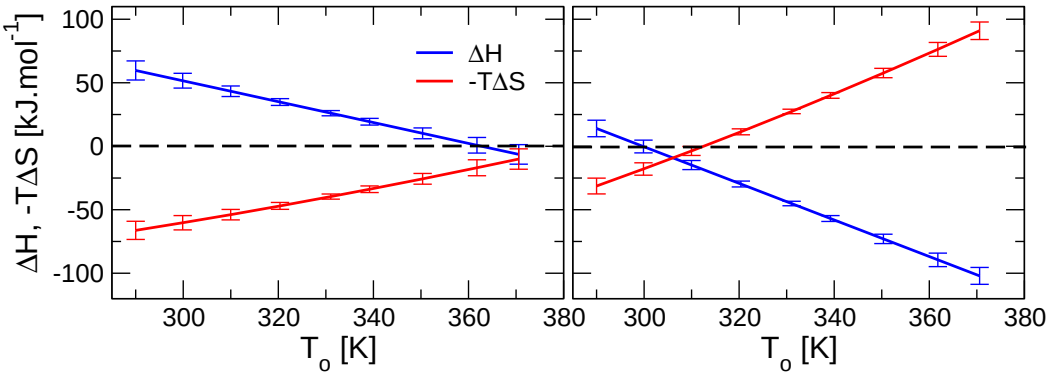


Figure 3.15 Temperature dependence of the enthalpy ΔH_o and entropic energy $-T_o\Delta S_o$ for (a) Aβ_{16-21} and (b) IAPP$_{28-33}$ fibrils computed at ξ_o. These quantities were obtained by fitting the PMF to equation 3.1 at different reference temperature T_o. Error bars were obtained from fits to equation 3.1.

including physiological conditions, i.e., 310 K. The dominant $-T_o\Delta S_o$ arises from hydrophobic interactions in the core of Aβ_{16-21}. [36, 65, 81, 174] These interactions emerge because water in the vicinity of non-polar residues are released into the bulk solution with increased entropy when non-polar molecules approach each other. The stability of the IAPP$_{28-33}$ fibril is favored by enthalpy while it is opposed by entropy at temperatures above 306 K. Direct interactions involving atoms of peptide, fibril, and water molecules can rationalize this enthalpic stabilization [36]. The configurational entropy of the peptide which becomes smaller when it binds to the fibril may explain the unfavorable entropic energy of binding of IAPP$_{28-33}$. At physiological conditions, the entropic component is negligible and the stability of IAPP$_{28-33}$ is dominated by enthalpy.

ΔC_{po} measures the curvature of the temperature dependence of the PMF—see equation 3.1. For protein folding, this quantity is invariably negative and it has been related to the burial of non-polar residues away from water [73,123,213]. Accordingly, ΔC_{po} for protein folding is often written as the sum of negative and positive terms accounting for the desolvation of non-polar $\Delta C_{po}^{nonpolar}$ and polar ΔC_{po}^{polar} residues, respectively [213]. In contrast to this decomposition, a recent study reported ΔC_{po}^{polar} to be negative for a large class of polar compounds [228] while experimental studies

have called for a reevaluation of the additive interpretation of ΔC_{po} [32,155]. Amyloid fibrils may constitute ideal systems to study these questions as the hydropathy of their dry core can be tuned by varying the peptide sequence. In our simulations, we find that ΔC_{po} is negative for both Aβ_{16-21} ($\Delta C_{po} = -0.85$ kJ/mol/K) and IAPP$_{28-33}$ (-1.44 kJ/mol/K) fibrils. This suggests that the burial into the fibril core of both non-polar side chains of Aβ_{16-21} and polar side chains of IAPP$_{28-33}$ account for a negative change in heat capacity. Simulations of other polar fibrils are needed to validate these results.

The lowest temperature probed by our simulations is 290 K. However, equation 3.1 can be used to extrapolate PMF(ξ_o) to lower temperatures. In particular, the extrapolated PMF(ξ_o) at 265.4 K for Aβ_{16-21} and 227.1 K for IAPP$_{28-33}$ is zero implying that fibrils become unstable at these temperatures. We anticipate that experimental studies of amyloid fibrils at high pressure and/or in solutions containing cosolvents may be used to explore the dissociation of non-polar fibrils at temperatures below 273 K. Pressure and cosolvents account for a reduction in the freezing point of water, enabling studies at temperatures below 273 K [245]. Experimental evidence that fibrils can dissociate at low temperatures has been provided for α-synuclein [106, 128, 168].

Figure 3.16 depicts thermodynamic quantities computed along the reaction coordinate, ξ, for Aβ_{16-21} and IAPP$_{28-33}$ fibrils at 310 K and 320 K. Panel (a) in Figure 3.16 shows that the first and second minima of the PMF of the Aβ_{16-21} fibril are favored by entropy and enthalpy, respectively. At the first minimum, non-polar surfaces of L_{17} and F_{19} side chains are buried in the fibril core (see Figure 3.3c) accounting for the release of shell water into the bulk and the dominant entropic component of the free-energy. At the second minimum, the peptide interacts with the side of the fibril, maximizing the number of van der Waals and electrostatic interactions, thus explaining the dominance of the enthalpic component. Accordingly,

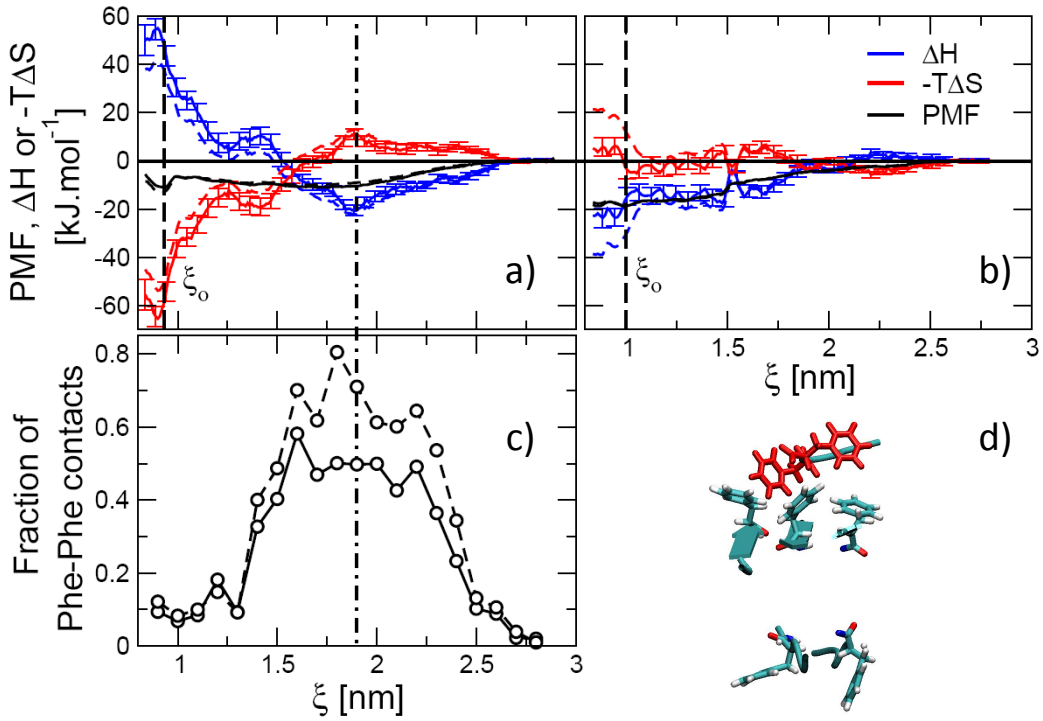


Figure 3.16 Changes in enthalpy ΔH and entropic energy -TΔS as a function of ξ for (a) Aβ_{16-21} and (b) IAPP$_{28-33}$ fibrils at 310 K (full lines) and 320 K (dashed lines). Error bars in these figures are obtained from fits to equation 3.1. (c) Fraction of contacts between phenylalanine residues of the free-peptide and the fibril at 310 K (full line) and 320 K (dashed line). Residues are considered to be in contact whenever their atoms are at a distance smaller than 0.25 nm from each other. (d) Characteristic configuration of fibril (cyan) and peptide (red) at $\xi = 1.8$ nm.

Figure 3.16c shows that the number of contacts between phenylalanine side chains of the peptide and the fibril is a maximum at the second minimum. A sample conformation in which phenylalanine side chains of the peptide and the fibril are in contact is shown in Figure 3.16d. In contrast, the dry core of IAPP$_{28-33}$ at locked states is made of polar residues (see Figure 3.3d) that form electrostatic, van der Waals, and hydrogen bonds with the fibril, accounting for the enthalpically dominant component of the PMF. Accordingly, in Figure 3.13 we show that residue N$_{31}$ of the peptide forms hydrogen bonds with residues N$_{31}$ and S$_{29}$ of the fibril contributing to intra and inter β-sheet stability. Notice that the magnitudes of ΔH(ξ) and $-$TΔS(ξ)

in Figure 3.16a-b are 2–4 times larger than the magnitude of the PMF(ξ). However, since one of these quantities is negative and the other is positive, i.e., enthalpy and entropic energy compensate each other, the magnitude of the PMF is only of the order of 15-20 kJ.mol^{-1}. [235]

3.4 Conclusion and Discussion

In summary, our results support the idea that the molecular mechanisms stabilizing cross-β structures are strongly related to the amino acids that are buried in the fibril core. This is in line with scanning mutagenesis experiments in which ΔG is shown to depend significantly on individual residues that are being mutated [141,272], bioinformatics estimates of ΔG based on structural complementarity of side chains forming the β-sheet [257], thermodynamic efforts to design new amino acid sequences that form fibrils [70, 232, 233], and thermodynamic studies of coarse-grained models [209, 211]. We find that non-polar fibrils are stabilized by entropy and destabilized by enthalpy while the opposite trend is observed for polar fibrils. This suggests that non-polar fibrils are stabilized by hydrophobic interactions, which are characterized by an increase in the entropy of water molecules, whereas the stability of polar fibrils emerges from van der Waals and electrostatic bonds including hydrogen bonds. Notice that in implicit water simulations, enthalpy and entropy were found to contribute equally to the free-energy of a non-polar fibril [184] suggesting that an explicit treatment of water may be required to account for thermodynamic properties.

Limitations of the present work should also be noted. In particular, our simulations are based on fibril models with fixed main-chains. They provide a template that corresponds to a deep free-energy minimum of the system as the peptide is found locked to the fibril in approximately 70% of the time for small ξ distances. However, more relaxed conformational restrictions of the fibril will be necessary to explore, for example, conformational fluctuations at the fibril end and how it affects

transient states leading to the locked state [108]. Also, it is desirable to reproduce results from this work with different force-fields. The extensive computational resources required to simulate 48 μs for Aβ_{16-21} and 45.6 μs for IAPP$_{28-33}$ in boxes containing $\sim$ 7,300 water molecules has so far prevented us from doing so. However, strengthening of hydrophobic interactions and weakening of direct interactions with increasing temperature, which gives rise to the entropic and enthalpic stability of non-polar and polar fibrils, is robustly reproduced by different force-fields. This provides evidence that qualitative results from this work are independent of the force-field.

CHAPTER 4

SUMMARY AND FUTURE WORK

4.1 Conclusion

In this dissertation, we addressed two major problems related to the aggregation of amyloid proteins using molecular dynamics simulations. First, we studied effects of aqueous solutions on the conformations of Aβ monomer and identified a potential precursor structure to fibril formation. Understanding the aggregation of intrinsically disordered peptides are a fundamental step towards finding a cure for amyloid diseases including Alzheimer's. Second, we determined the molecular interactions that stabilize amyloid fibril fragments in water. This knowledge is important to develop novel methods to disrupt the structure of fibrils in amyloid diseases. To address these problems, we performed all atom molecular dynamics simulations with explicit water model using state of the art force fields to accurately model the systems. Further, to ensure that our simulations have adequately sampled the free energy landscape, we have enhanced the sampling of our models through the combination of replica exchange molecular dynamics and umbrella sampling. Employing both methods together, significantly improves the accuracy of determining protein interactions.

In Chapter 2 of this dissertation, we studied the effects of an aggregation prone molecule and two aggregation inhibitor compounds on the conformations of the Aβ_{42} monomer. The sampling in the simulations are enhanced using REMD with 64 replicas in temperatures ranging between 310 K and 500 K. Our results show that the set of structures sampled by Aβ_{42} is strongly dependent on its aqueous solution. Despite small changes in secondary structures, the intrapeptide interactions are completely distinct in the four systems. Among other differences, our simulations demonstrate that the aggregation prone molecule promotes a unique structure which

does not exist in the other Aβ_{42} systems. We propose that such unique conformations are intermediate structures in the fibril formation pathway. Thus inhibiting the formation of this structure, might be beneficial in developing drugs for Alzheimer's disease.

In Chapter 3 of this dissertation, we investigated the molecular interactions that stabilize amyloid fibrils. Two fibrils made from the nonpolar protein Aβ_{16-21} and the polar protein IAPP$_{28-33}$ were studied. To understand the underlying mechanisms of their stability, we determined the free energy of peptide addition to the fibril using umbrella sampling at different temperatures. To further enhance our sampling, at each window of umbrella sampling we performed replica exchange molecular dynamics with 32 replicas in temperatures ranging from 290 K to 373 K. Potential of mean force is calculated for each system at different temperatures to provide estimates for enthalpy and entropy. We find that the nonpolar fibril becomes more stable with increasing temperatures and is stabilized by entropic energy. In contrast, the polar fibril becomes less stable with increasing temperatures and is stabilized by enthalpy. Our results show that thee nature of side chains in the dry core of amyloid fibrils plays a dominant role in accounting for their thermodynamic stability. These findings suggest that the stability of fibrils can be customized by the choice of amino acid sequence in the dry core of the amyloid fibrils. Such fine-tuned amyloid fibrils can be used as scaffolds for drug delivery and other biomaterials.

In Appendix A, we studied the role of side chain interactions on the formation of α-helices using all-atom molecular dynamics simulations of polyalanine-like peptides. We modified both the equilibrium Lennard-Jones distance, σ, and the well-depth, ϵ. In addition, we computed potentials of mean force for the interaction of methane–like molecules that represent side-chain group of alanine and identified the values of the Lennard-Jones parameters that promote α-helix formation in polyalanine peptides. These findings highlight the limitations of two-bead coarse-grained models to account

for side chain interactions in α-helices and they are important in development of coarse-grained models.

In Appendix B of this dissertation, we describe a C++ program that identifies clathrate hydrate structures in a molecular dynamics simulation. These structures are commonly found in natural gases and they are relevant to understand anti-freezing proteins as well as the hydrophobic effect. As shown in Chapter 3 of this dissertation, the hydrophobic effect can be an important force accounting for fibril formation. The program identifies 5^{12}, $6^2 5^{12}$ and $6^4 5^{12}$ cages as well as the four-body order parameter (F_4). The program also generates three-dimensional structure of clathrates which can be easily visualized using software such as VMD (Visual Molecular Dynamics). These are main quantities used to quantify clathrate formation in molecular dynamics simulations. To the best of our knowledge, there is no open-source code to analyze the aforementioned parameters in all-atom molecular dynamics simulations. We anticipate that the freely available source code will enable research groups to easily analyze their simulation results. The code can also be modified to allow the investigation of other order parameters to quantify water structures.

4.2 Proposed Future Work

The aggregation of the 42-residue Aβ protein (Aβ_{42}) into larger deposits and plaques has been associated with Alzheimer's diseases (AD). Along its aggregation pathway, low molecular weight oligomers of Aβ has been identified to be the primary toxic aggregates [164, 231]. Understanding the molecular mechanisms of how these oligomers from is critical in finding a cure for Alzheimer's disease. Several aspects of Aβ aggregation has been investigated in recent years, e.g., its nucleation dependent aggregation which starts with a lag phase prior to fibril growth and the number of monomers associated in the nucleus (nucleation number, n^*) [85, 96, 112]. However, a thorough understanding of the pre-nucleation phase that affects the aggregation pathway of Aβ are still missing.

In this dissertation, we studied the effects of different molecules on the structure of Aβ_{42} monomer which is among the earliest steps in the aggregation pathway. We found a unique structure that is enhanced by the aggregation prone molecule. Further research can be performed using other aggregation prone compounds in order to find an ensemble of intermediate structures along the pathway of fibril formation. Knowledge of an ensemble of intermediate structures can provide a more clear understanding of the underlying molecular interactions that promote Aβ_{42} aggregation [45]. Another extension to our current research is to assess the propensity of these unique structures to form fibrils. This can provide proof for the estimates of the nucleation number, n^*, of Aβ which has been predicted to range between 6 and 14 [85, 265].

In Chapter 3, we provided insights on the stability of two amyloid fibril fragments and determined the thermodynamic parameters that stabilize polar and nonpolar fibrils. However, developing a thermodynamic theory for fibril growth, similar to the thermodynamic theory of protein folding, requires the study of other fibrils from different amino acid sequences. Results in this study are based on fibril

models with fixed main chains. For future work, it is proposed to employ more relaxed conformational restrictions of the fibril in order to explore conformational fluctuations at the fibril ends and how it affects transient states leading to the locked state.

APPENDIX A

ROLE OF SIDE CHAIN INTERACTIONS ON THE FORMATION OF α-HELICES IN MODEL PEPTIDES

A.1 Introduction

α-helices and β-sheets are the main building blocks of protein structures serving as a template for almost 50% of all residues [29]. These motifs are also present in the structure of intrinsically disordered and amyloid peptides in which stacking of β-sheets has been related to diseases like Alzheimer's and Parkinson's [111, 149, 223]. Due to their ubiquitous presence, α-helices and β-sheets have been the subject of numerous studies aiming to understand the molecular forces driving their formation. However, this remains a topic of debate as intra-backbone hydrogen bonds (which were initially thought to account for α-helices and β-sheets [192–194]) were not found to be significantly favorable to drive this process in aqueous solution because they require the unfavorable breakage of backbone-water hydrogen bonds [66, 115, 124, 174, 224, 247]. As a results, most algorithms designed to predict the propensity of secondary structures are knowledge-based [44]. The aim of the current research is to provide insights into the forces driving the formation of α-helices and, in particular, the role played by the effective interactions between the side-chains.

Since the seminal work of Kauzmann, hydrophobic interactions are believed to be the main ingredients determining native protein structures [66, 124, 143]. They emerge because non-polar regions of proteins or peptides tend to minimize their solvent exposed area accounting for the globular shape that characterizes the native state. The importance of these hydrophobic interactions can be inferred from the positive curvature of the Gibbs free-energy of unfolding with respect to temperature [10,124,199]. This is typical of non-polar solvation [73,189,236,237] and it rationalizes cold denaturation of proteins [60, 62, 199]. Also, the diversity of native structures

can only be encoded in the amino acid sequence (not in the backbone), suggesting that side-chain properties and, in particular, the burial of non-polar residues in the dry protein core is responsible for folding. Accordingly, after the first protein structure was resolved experimentally, its dry core was observed to be made mostly of non-polar residues [125, 126]. Secondary structures form during folding because the polar backbone is also buried in the dry protein core accounting for an enthalpic penalty that can be minimized through the formation intra-backbone hydrogen bonds. This favors internal organizations within the collapsed state such that α-helices and β-sheets emerge in proteins to avoid the enthalpic penalty of burying the backbone in the dry core. In contrast to this process, peptide structures do not exhibit a dry core suggesting that the mechanism for forming secondary structures could differ from the one in globular proteins. This is supported by experiments showing that destabilization of α-helices by co-solvents that form hydrogen bonds correlate with the strength of these bonds for peptides but not for proteins [66]. In addition, very small concentrations of surfactants is sufficient to unfold proteins efficiently [254] whereas they do not destabilize helices [151]. Despite these insights, it is still not clear what drives α-helices and β-sheets in peptides.

Propensities to form α-helices were first attributed to the restriction of the configurational entropy of side-chains upon folding [185, 188]. However, a poor correlation between the reduction in the side-chain entropy and helix propensity was found [25] putting into question the validity of this argument [156]. The possibility that helix propensities are modulated by energy was first proposed by Luo and Baldwin who used thermal unfolding curves of five nonpolar amino-acids in water/trifluoroethanol mixtures [150]. This was then extended by Makhatadze and coworkers who used calorimetric measurements of folding a model host peptide in which the helix formation is induced by metal binding [156, 210]. This concept has been further developed in the prediction of the helical behavior of peptides. Here,

experimental data was used to parameterize empirically a set of energy contributions for every side-chain [172]. In fact, modulating side-chain interactions has been exploited in designing very short helical peptides in solution such as in the 5-mer peptide $WAAAH^+$ where strong cation-π interaction are established between $i-i+4$ neighbors [145]. In a computational study, side-chain interactions were reported to play an important role in transitions from α-helix to β-sheet in a short polyleucine peptide [46]. Furthermore, in the AGADIR algorithm to predict α-helix content [172], the inclusion of side-chain interactions was particularly relevant since, without them, natural amino-acid sequences tend to lack measurable helix content in water [11].

It should be noted that when taking into account side-chain contributions in promoting secondary structures, the important quantity to consider is the *effective* interaction of the side-chain with the other groups of the peptide. This is particularly relevant in aqueous solutions where water increases the complexity of energy landscapes of molecular interactions. For example, the interaction between methane molecules in water (which are often used as a model for the interaction of non-polar side-chains [37, 40, 104, 202]) is characterized by a global and a local minimum at short ($\sim$ 3.8 Å) and intermediate ($\sim$ 7 Å) distances, respectively, separated by an energy barrier related to desolvation effects (at $\sim$ 5.7 Å) [197]. These features affect short range structures in proteins and peptides. For example, in peptides made from aliphatic amino-acids it was shown that distances between $C_\beta - C_\beta$ atoms of residues $i - i + 3$ and $i - i + 4$ coincide with the position of desolvation barriers while $C_\beta - C_\beta$ distances of residues $i - i + 2$ in β-sheets coincide with the local minimum [60,65,176,177]. This was shown to play an important role in the propensity of secondary structures studied computationally with implicit water models [65].

In the current dissertation, we study how the effective potential between side-chains affects the probability to form α-helices using all-atom molecular dynamics

simulations in explicit water. To that purpose we use poly-alanine–like peptides described by the OPLS-AA force-field where the Lennard-Jones interactions between C_β atoms is modified. We change both the equilibrium Lennard-Jones distance, σ, and the well-depth, ϵ. In addition, we also compute potentials of mean force for the interaction of methane–like molecules that represent side-chain groups in the poly-alanine simulation. We find a good correlation between the propensity to induce α-helical conformations in peptides and the effective interactions between the side-chains (computed using PMF(s) of methane-like molecules). In particular, greater propensities are observed when PMF(s) of methane–like molecules exhibit minimum at distances corresponding to C_β distances in α-helices.

A.2 Methods

We model 9- and 12-residue homogeneous (uncharged) peptides deprotonated at the N-terminal and protonated at the C- terminal using the OPLS-AA force field. Peptides are poly-alanine chains in which the σ parameter of the Lennard-Jones (LJ) potential between C_β atoms is varied systematically from 0.27 nm to 0.57 nm in steps of 0.10 nm in the different simulations. Since the original LJ parametrization of poly-alanine is σ=0.35 nm and ϵ=0.276144 kJ/mol, we also perform simulations using these values as well as σ=0.45 nm and ϵ=0.276144 kJ/mol. All other atoms of the peptides were represented by the OPLS-AA force field [117, 120, 165, 212]. The interaction between C_β and any other atom in the system (excluding other C_β's) corresponds to that of poly-alanine.

The simulation box consists of a 9-mer or a 12-mer peptide solvated in 1535 or 1681 TIP4P water molecules [116], respectively. Bond distances and angles within water molecules are constrained using the SETTLE algorithm [169] whereas covalent bonds within the peptide are constrained using the LINCS algorithm [100]. Starting with fully extended poly-alanine configurations (generated using the program

WHATIF [264]), the system is relaxed by a 100 ps molecular dynamics simulations (with $\epsilon_{C_\beta - C_\beta}$=0.0 kJ/mol) yielding extended random coil conformations. These structures are used as starting configurations in our simulations. Within each peptide length, we use the same starting configurations for simulations with the different σ and ϵ parameters.

Each trajectory was propagated for 400 ns whereas results for ϵ=0.47 nm and ϵ = 1 kJ/mol were taken from a previous study [278]. Atomic positions of the peptide were saved every 10 ps and they were used in all analyses. Note that a recent computational study of hepta-alanine modeled using the OPLS-AA force-field and compared with NMR-derived J-coupling constants reported convergence of the value of χ^2 within the first 250 ns of the trajectory [84].

The molecular dynamics package GROMACS version 4.5.4 [101] was used to perform all simulations using a time step of 2 fs. Electrostatic forces were evaluated using the Particle-Mesh Ewald method [52] (real-space cut-off of 1.2 nm, grid spacing of 0.12 nm, and quadratic interpolation) while a cutoff of 1.2 nm was used for LJ forces (with long range dispersion correction for the energy and pressure). The entire system was maintained at a constant temperature of 300 K using the velocity rescaling thermostat [30] and a coupling time of 0.1 ps. Pressure was kept fixed at 1.0 bar using the Berendsen thermostat [19] with a compressibility of $5{\cdot}10^{-5}$ 1/bar and a coupling time of 1.0 ps.

To estimate the effective interactions between two C_β atoms within the peptide chain we calculated the potential of mean force (PMF) between methane-like molecules in solution. PMF(s) were computed using the λ-coupling approach with a series of 51 λ-points from 0.00 to 1.00. At every λ-point the distance between carbon atoms of the two methane-like molecules was constrained to a specified value and the system simulated for 4.0 ns. The average force (over 3.5 ns data collection step) needed satisfy this constraint, thus, $\langle \partial \mathcal{H} / \partial \lambda \rangle$, was then integrated as a function of λ

to obtain the PMF. As the PMF represents only relative values, the resulting curve was shifted such that the value at λ=1.00 was equal to zero. In these free energy simulations, the carbon-hydrogen bonds were described by harmonic potential, and therefore, the time step of the simulations was reduced to 0.001 ps. Because, each OPLS-AA hydrogen carried a charge of +0.06 e the carbon atom of the methane-like molecule was assigned a charge of -0.24 e.

A.3 Results

A.3.1 Side-Chain Interactions

In Figure A.1 we study how the σ parameter of the LJ interaction between C_β atoms affects the formation of secondary structures using fixed $\epsilon = 1$ kJ/mol. For all values of σ, residues of the peptide spend most of their time in coil conformations. This is particularly striking for $\sigma = 0.57$ nm (Figure A.1d) in which case ordered structures (α-helix and β-sheet) are not observed in the timeframe of the simulation and the peptide spends 94 % of the time in coil conformations. Ordered secondary structures, mostly α-helices and turns, are only observed for σ values smaller than 0.57 nm. From Figure A.1, it seems that turns are more prominent for $\sigma = 0.37$ nm (panel b), whereas, populations of α-helix seem to occur more frequently for $\sigma = 0.47$ nm (panel c). These observations are quantified in Figure A.2 where we show the average content of secondary structures as a function of σ for different ϵ values. It confirms that the largest content of turn occurs for $\sigma = 0.37$ nm, whereas, α-helix content peaks at $\sigma = 0.47$ nm. In our simulations, the formation of β-sheet structures is negligible with a maximum content of ~ 1 % occurring at $\sigma = 0.37$ nm.

However, the lack of β-sheet content observed in this study is expected because the size of our peptides are relatively short rendering the penalty for loop formation large relative to the stabilization obtained from the interaction between the strands. Moreover, to induce β-sheet (β-hairpin) structures in short peptides, segments with

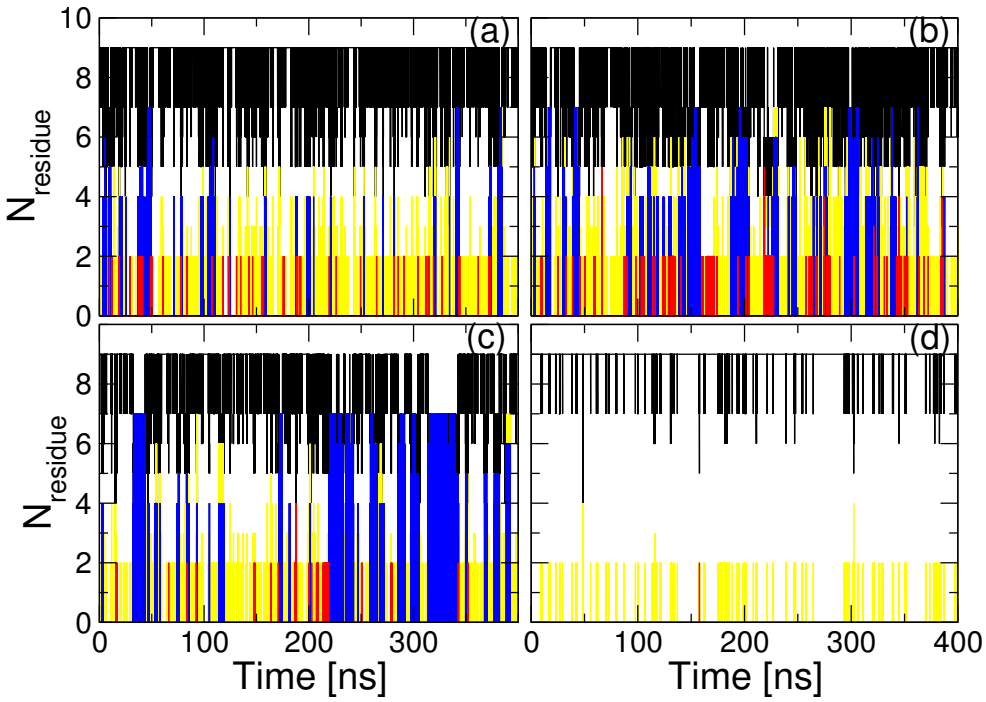


Figure A.1 Time evolution of secondary structure content for $\epsilon = 1$ kJ/mol, N = 9, and σ values of (a) 0.27 nm, (b) 0.37 nm, (c) 0.47 nm, and (d) 0.57 nm. $N_{residue}$ is the number of residues assuming coil (black), β-sheet (red), turn (yellow), and α-helix (blue) structures.

specific sequences can be incorporated in the middle of the chain (for example, sequences containing proline) in order to promote a loop structure [246].

In Figure A.2 we also study how the strength of the ϵ parameter of the LJ interaction affects secondary structure formation. For σ values smaller than 0.57 nm, we observe an increase in α-helix content with increasing ϵ. This is in agreement with previous studies showing that interactions between side-chains can modulate α-helix formation [150, 156, 210, 278]. The population of turns has a more complex behavior in the parameter space we are exploring. At fixed $\sigma = 0.37$ nm, the content of turns increases with increasing ϵ, whereas, at fixed $\sigma = 0.47$ nm the percentage of turn structures drops drastically from ~ 10 % at $\epsilon = 1$ kJ/mol to zero at $\epsilon = 2$ kJ/mol. This abrupt reduction in the content of turns coincides with a large increase (~ 15 %) in the population of α-helices suggesting that increasing ϵ from 1 kJ/mol to 2 kJ/mol

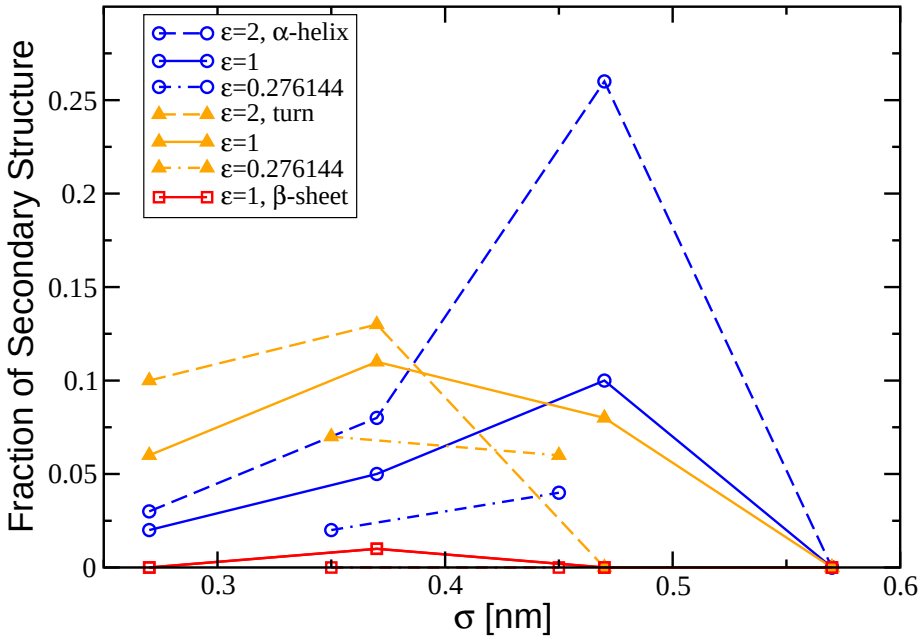


Figure A.2 Content of secondary structure elements as a function of σ for several values of ϵ. All plots are for the 9-residue long peptides (N=9) averaged over the entire 400 ns trajectory.

at fixed $\sigma = 0.47$ nm could trigger a turn-to-helix transition. Note that according to DSSP definitions, a residue is assigned to a turn whenever the CO– group of residue i forms a hydrogen bond with the NH– group of residue $i+n$ where n=3,4, or 5 while a residue is assigned to an α-helix whenever two (or more) consecutive residues form turns with n=4. Thus, a structure determined as a turn according to the definitions of DSSP might not be too different from an α-helix, and thus, an increase in ϵ at fixed $\sigma = 0.47$ nm (which is a length-scale that favors C_β–C_β interactions in α-helix) could indeed trigger turn-to-helix transitions.

A.3.2 Peptide Length

Experimentally, the formation of secondary structures in poly-alanine has been shown to depend on peptide length. In particular, α-helix content was reported to increase

with peptide length up to N = 19 whereas larger chains (19 > N > 25) were shown to aggregate into oligomers [20]. End effects in α-helices, involving the first and last three residues (in addition to capping groups at the N– and C– termini), contribute to this dependence because only one intra-backbone hydrogen bond is formed in these residues as opposed to two for residues in the middle of the helix. Similarly, side-chain atoms form less interactions in these residues compared to the ones in middle of the helix. Thus, end effects are more pronounced in short compared to long peptides accounting for 2/3 and 1/2 of the residues in 9- and 12-mer chains, respectively.

To test how changes in side-chain interactions affect peptides of different lengths, we perform a set of simulations using peptides made of 12 residues (N=12). In Table A.1, they are compared with the 9-mer simulations performed in the previous section. Turn content is not strongly affected by peptide length: changing peptide

Table A.1 Secondary Structure Content Averaged over 400 ns Using $\epsilon = 1.0$ kJ/mol

	$\sigma = 0.37$ **nm**			$\sigma =$ **0.47 nm**		
N	**% helix**	**% turn**	**% coil**	**% helix**	**% turn**	**% coil**
9	5	11	63	10	8	65
12	7	11	56	44	9	38

length from 9 to 12 residues increases the percentage of turn by 0 % for $\sigma = 0.37$ nm and by 12 % for $\sigma = 0.47$ nm. In contrast, α-helix content is strongly dependent on peptide length; for $\sigma = 0.37$ nm, the percentage of α-helical structures increases by 40 % with increasing N and for $\sigma = 0.47$ nm it increases by 340 %.

Time dependence of assigned secondary structures based on DSSP to residues along the amino acid sequence is shown in Figure A.3. For $\sigma = 0.37$ nm (panel a), five α-helix nucleating events involving at least four residues are observed within the simulation time. For $\sigma = 0.47$ nm (panel b), there is only one main α-helix

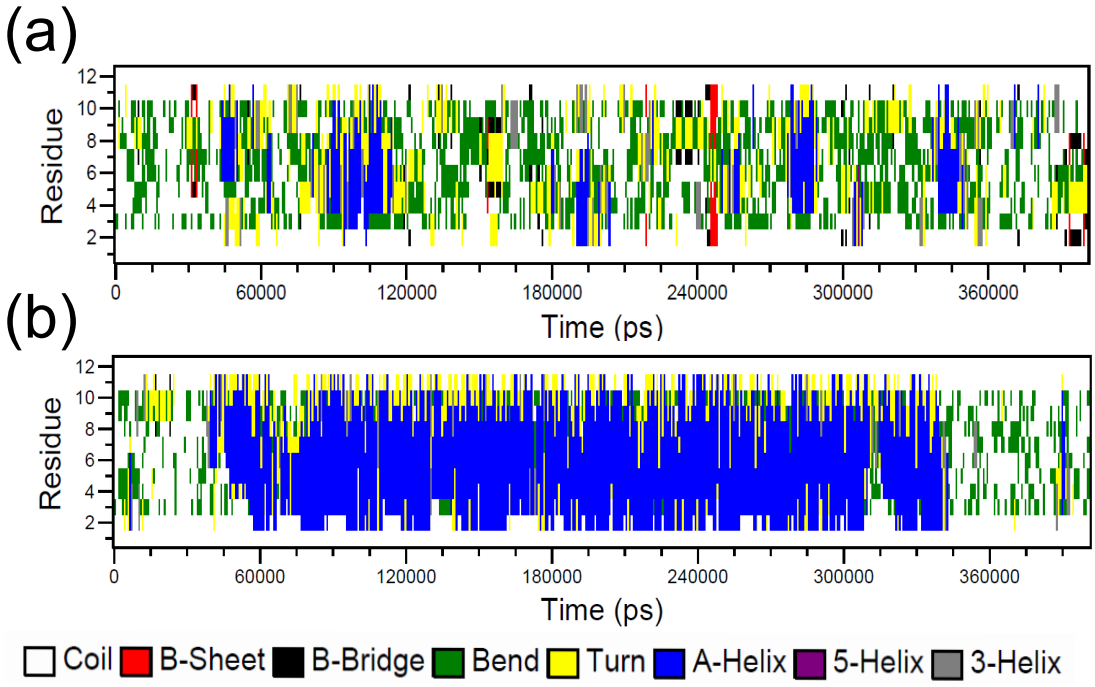


Figure A.3 Time evolution of secondary structure content along the amino acid sequence (y-axis) for $\epsilon = 1$ kJ/mol. Panels (a) and (b) correspond to $\sigma = 0.37$ nm and 0.47 nm, respectively.

nucleating event with a life-time of $\sim$ 300 ns, albeit several structural transitions involving a complete and partial α-helices are observed.

Notice that in an α-helix, the number and the energy of intra-backbone interactions change by the same amount when the peptide length increase from 9 to 12 residues for both $\sigma = 0.37$ nm and 0.47 nm. The number of side-chain interactions also increases by the same amount for $\sigma = 0.37$ nm and 0.47 nm but not the energy of these interactions. Thus, in our simulations (see Table A.1) the 40 % increase in α-helix content (when the length of the peptide increases from 9 to 12 residues) for $\sigma = 0.37$ nm compared to the 340 % increase for $\sigma = 0.47$ nm can only be accounted for by studying the interaction between side-chains. This is the purpose of the next section.

A.3.3 Methane-Like Dimers

In order to explain the propensity of α-helix formation in our simulations, we note that for a fixed ϵ, simulations differ in the σ parameter of the C_β–C_β LJ potential. This accounts for different effective potentials between side-chains of the peptide. Here, we assume that the effective interaction between side-chains (–CH_3 groups) can be approximated by the PMF between two methane-like (CH_4) molecules in solution. Main contributions to this PMF are the direct LJ interactions between C_β atoms and entropic contributions of surrounding water molecules. LJ interactions between carbon atoms are favorable over a wide range of distances (see Figure A.4) whereas the effective PMF between methane molecules is negative only for well defined positions corresponding to "contact minimum" (CM) and "solvent separated minimum" (SSM)—see Figure A.4. As a consequence, the effective potential between side-chains can promote a particular secondary structure only if CM and SSM distances are consistent with side-chain distances of this particular structure.

In Figure A.6a we show PMF(s) for the interaction of methane-like molecules that mimic side-chains in our poly-alanine simulations. To understand how CM and SSM affect α-helix formation, we show in Figure A.6b distributions of $C_\beta - C_\beta$ distances when peptides are in an helical conformation for our 9-mer and 12-mer simulations [1]. Two prominent peaks emerge at 0.54 nm and 0.72 nm. The first peak in Figure A.6b at r=0.54 nm is the result of interactions between residues $i - i + 1$ (0.529 nm for a perfect α-helix) and $i - i + 3$ (0.548 nm for a perfect α-helix)—see Figure A.5. These two inter-neighbor distances appear as a single peak due to thermal fluctuations.

[1]Peptides are defined to be in an α-helical conformation whenever the root-mean-square deviation (RMSD) of backbone atoms relative to an ideal α-helix is smaller than 0.13 nm and 0.12 nm for 9-mer and 12-mer peptides, respectively. These cut-off values for the RMSD correspond to the minimum separating the α-helical peak in the RMSD distribution from other conformations. A similar criteria was used in a previous study [278].

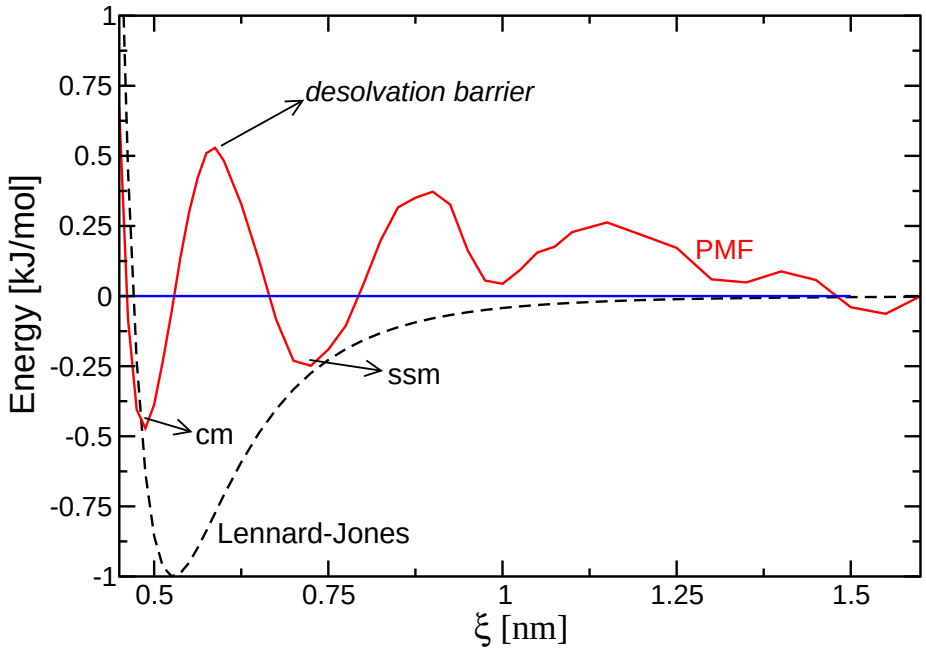


Figure A.4 Potential of mean force between methane-like particles and C_β-C_β LJ interaction using $\epsilon = 1.0$ kJ/mol and $\sigma = 0.47$ nm.

The distance corresponding to the first peak is projected onto the different PMF(s) using a dotted line in Figure A.6a. It indicates that for most of the LJ parameters studied here, side-chain interactions between i−i+1 and i−i+3 contribute positively to the energy of α-helix formation. The exception being $\sigma = 0.47$ nm where the first peak distance is close to the global minimum of the PMF. Accordingly, simulations using σ=0.47 nm show the largest α-helical content—see Figure A.2 and Table A.1. For σ=0.27 nm and σ=0.37 nm, the first peak in the distribution of the $C_\beta - C_\beta$ distances falls within the desolvation barrier. As a result, $i - i + 1$ and $i - i + 3$ neighbors contribute unfavorably to α-helices which could explain the low helical content in the corresponding simulations in Figure A.2. Furthermore, for σ=0.57 nm the position of the first peak in Figure A.6b corresponds to hard-core

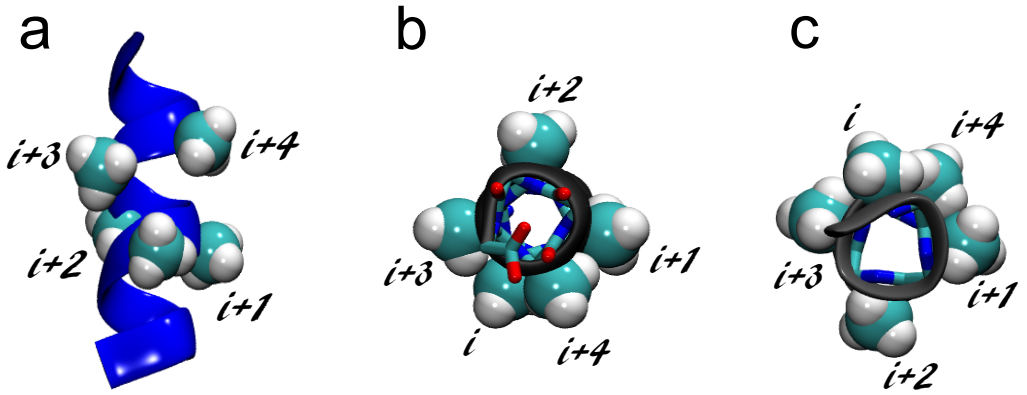


Figure A.5 Representation of an ideal α-helical structure viewed from (a) the side, (b) C-terminal, and (c) N-terminal (c).

repulsion in the PMF. This steric clash is very important because in this case the formation of any helical structure is completely destroyed.

The distribution of C_β–C_β distances shown in Figure A.6b also displays a shoulder (or a small maximum) at r=0.63 nm, due to $i-i+4$ neighbors. For $\sigma < 0.57$ nm, intensities of the different PMF(s) at r=0.63 nm fall within the desolvation barrier and thus $i-i+4$ neighbors are not likely to contribute favorably to α-helix formation. For σ=0.57 nm, $i-i+4$ neighbors contribute favorably to α-helix formation. However, as mentioned previously, this value of σ can not accommodate $i-i+3$ and $i-i+1$ neighbors in an α-helix because of its hard-core repulsion.

The distribution of C_β–C_β distances exhibits a third peak at r=0.72 nm due to $i-i+2$ neighbors. PMF(s) of methane-like molecules (Figure A.6a) at this distance correspond to SSM and, therefore, they contribute favorably to the formation of α-helices. However, PMF(s) of methane-like particles in pure water (as the ones computed in Figure A.6a) are not good models to describe interactions between $i-i+2$ neighbors because these two residues are at opposites sides of an α-helix (see Figure A.5) and, thus, are not separated by water. We speculate that $i-i+2$ neighbors are not likely to contribute significantly to α-helix formation since direct

LJ and electrostatic interactions between these two residues at 0.72 nm are small—see Figure A.6.

To provide a quantitative framework for the observed correlation between the formation of α-helix and PMF(s) between methane-like molecules, we compute the average energy of side-chain interactions in our eight simulations. Mathematically, we define side-chain energies as:

$$E_{\text{sidechain}} = \frac{1}{N} \sum_{k} \sum_{i,j} \text{PMF}(\xi_{ij}), \tag{1.1}$$

where the first sum is over the N helical frames in the trajectory and the second sum is over all residue pairs $i - j$ in one frame taken without double counting. Figure A.7 shows that LJ parameters that increase the attractive energy between the side-chains when the peptide assumes an α-helix conformation correlate with a larger population of this secondary structure. This is valid for both peptide lengths studied here.

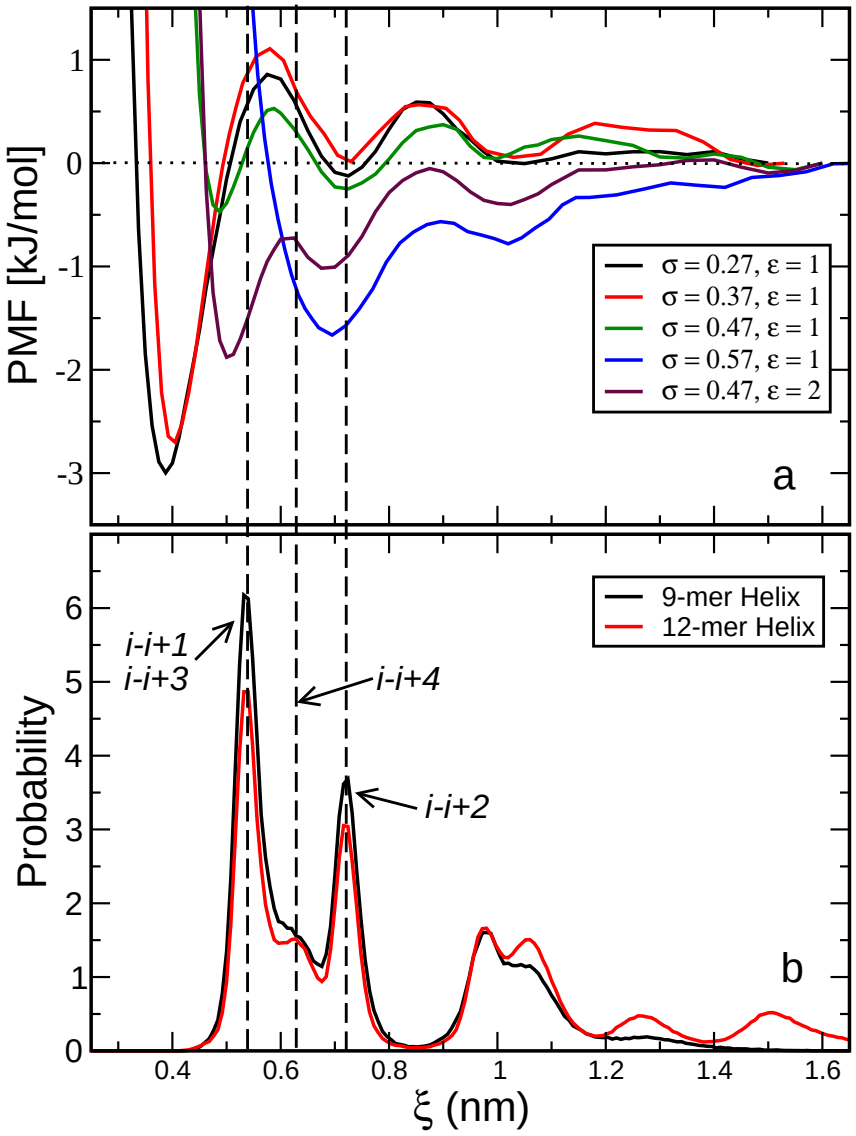


Figure A.6 (a) The potential of mean force between two methane-like molecules with different values of the $C_\beta - C_\beta$ LJ σ parameter ($\epsilon_{C_\beta - C_\beta}$=1.0 kJ/mol), solvated in aqueous solution, as a function of the distance between these two central atoms. (b) The normalized distribution of the distance between the C_β atoms in α-helical conformations in 9-mer (σ=0.47 nm and ϵ=2.0 kJ/mol) and 12-mer (σ=0.47 nm and ϵ=1.0 kJ/mol) peptides. The frames of the trajectories in an α helical conformation were determined by a RMSD (with respect to a perfect helix) cutoff value of 0.12 nm which corresponds to the location of the first minimum of the RMSD histogram.

A.4 Conclusion

To study the role of side-chain interactions in α-helix formation, we performed extensive all-atom molecular dynamics simulations of modified poly-alanine peptides in explicit water. We varied the distance and depth of the $C_\beta - C_\beta$ LJ interaction and we identified length-scales that promote α-helices. Previous studies on the role of side-chain interactions in secondary structure formation were either limited to implicit water model [65, 68, 179, 180], restrained peptide simulations [60] or they did not take into account effects due to different length-scales [278]. To rationalize variations in α-helix content observed in our simulations we computed effective interactions, i.e., PMF(s), between methane-like particles that mimic side-chains in our modified poly-alanine peptides. Contact-minimum, desolvation barrier, and solvent-separated-minimum of computed PMF(s) when superposed to distances between $i-i+1$, $i-i+3$, and $i-i+4$ neighbors provided a qualitative explanation for the observed α-helix content in our simulations.

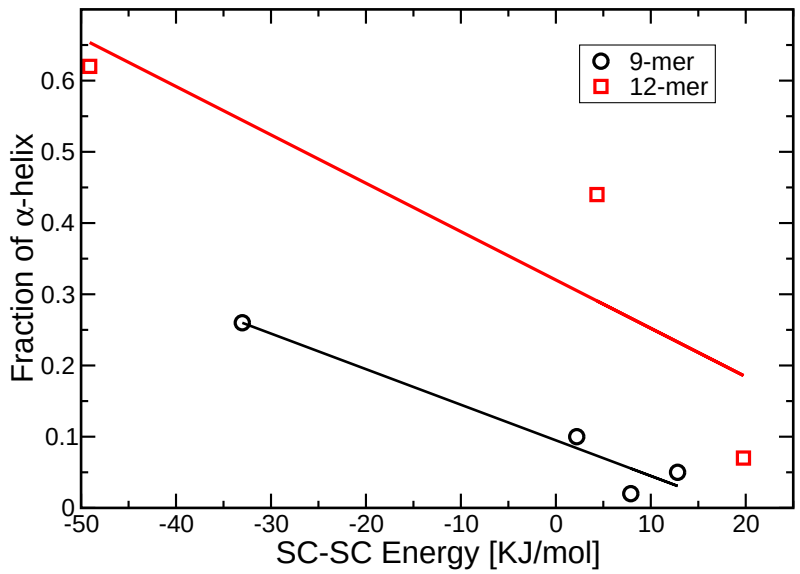


Figure A.7 Correlation between side-chain–side-chain (SC-SC) energy and the fraction of α-helix for 9-mer and 12-mer peptides. Lines are a guide to the eye.

One implication of our findings is for the development of coarse-grained models. We show the importance of using potentials for side-chain interactions that have solvent effects embedded into them, e.g., desovation barriers. Our results also highlight limitations of two-beads coarse-grained models to account for side-chain interactions in α-helices. In these structures, beads representing side-chain atoms can only form favorable contact if their size is defined by $\sigma = 0.47$ nm. This restricts the variety of amino acids that can be studied with these type of models.

Based on the PMF analysis between methane-like particles, side-chain interactions in poly-alanine (which is defined by $\sigma = 0.35$ nm) are unfavorable to α-helices due to the formation of desolvation configuration between side-chains. This result is consistent with the finding from an early explicit-water simulation of the interaction between water and an alanine-based α-helix that water does not penetrate much between spatially adjacent C_β groups along the helix axis [82]. Nevertheless, alanine has one of the highest helix-forming tendency among natural amino acids. If side-chain interactions were to play a role in alanine's helix-forming tendency, it would imply that side-chain interactions in unfolded conformations of alanine would be even more unfavorable than in α-helix conformations. However, for this small amino acid it is likely that side-chain interactions are not an important factor in helix formation [161]. In contrast side-chain interactions for amino acids defined by $\sigma = 0.47$ nm (this could mimic leucine), are favorable to α-helices since $i - i + 3$ and $i - i + 1$ neighbors are in close contact (contact minimum in the PMF). Thus, side-chain interactions could play a multifaceted role which can either promote or penalize α-helices. Notice that applied pressure and temperature can change the PMF affecting the stability of side-chain interactions and thereby the propensity for secondary-structure formation [60]. In particular, we can speculate that since applied pressure increases the desolvation barrier, it might penalize α-helices. Further investigation are, however, needed to unravel the effect of pressure and temperature.

APPENDIX B

GRADE: A CODE TO DETERMINE CLATHRATE HYDRATE STRUCTURES

B.1 Introduction

Gas clathrates are solid compounds formed by natural gas molecules trapped within solid cage-like water structure [242]. These structures are found on the seabed, and in ocean/lake sediments where conditions of high pressure and low temperature are favorable to their formation [74, 87, 132, 243, 250]. As the extraction of gas clathrates is becoming economically viable, these natural deposits are expected to account for an important fraction of the world's energy supply [48, 219]. However, the existence of these natural deposits is also a source of concern as the release of large quantities of methane during earthquakes and tsunamis could contribute significantly to global climate. Moreover, methane clathrate represents a dangerous problem for the gas industry due to potential plug formation in pipelines [34, 242, 244]. Accordingly, extensive efforts are being dedicated to understand the formation, stability, and inhibition of gas clathrates [76, 275].

Nucleation of gas clathrates occurs at nanometer length scale and micro-second timescale, which are not easily accessible experimentally but can be probed using all-atom computer simulations [58]. The latter is providing insights into the sequence of molecular events leading to the nucleation of clathrate structures [22, 89, 109, 130, 267, 268]. In particular, simulations have shown that an isolated water cage hosting a gas molecule in aqueous solution is short-lived [91, 92, 201] and clathrate nucleation requires local concentration of guest molecules at supersaturated conditions, known as "blobs" [137,201,276] The formation of these blobs in water is the rate limiting step of clathrate nucleation which prompts water molecules to freeze into cages hosting gas molecules. Initially, cages are stacked together in an amorphous manner before

relaxing into the crystalline clathrate phase [110]. This sequence of events, i.e., blobs $\rightarrow$ amorphous cages $\rightarrow$ clathrate phase, is known as the multistep hypothesis (MSH) of clathrate formation.

The formation of water clathrates also plays a role in biological systems. In particular, water molecules around non-polar side chains in proteins are more structured than in the bulk solution with a tendency to form incomplete cage-like structures [60, 62, 63, 81, 239]. In antifreeze proteins, e.g., the Antarctic bacterial antifreeze protein, the more structured water molecules around non-polar side chains mediate binding of ice surfaces to the protein inhibiting ice propagation [83]. Also, semi-clathrate water structures have been shown to contribute to the stabilization of antifreeze protein folds [251]. Thus, understanding the formation of clathrates in biological systems is important to understand protein structure formation and to discover new antifreezing mechanisms.

The aim of this dissertation is to present an open source computer code that analyzes atomic positions of oxygen atoms of water to compute the number of cages and account for their three-dimensional structures. The latter can be used for visual inspection using software such as VMD (Visual Molecular Dynamics) [105]. We named the code *GRADE* which stands for "cage" in Portuguese. The number of cages is commonly used as an order parameter to estimate the level of clathrate structures that have been formed in a simulation. Moreover, visual inspection of the time evolution of these structures is essential to provide insights into the sequence of molecular events leading to clathrate formation. Codes to analyze water structures have been developed by other labs including Molinero's group who developed CHILL+ to compute different order parameters that distinguishes between hydrates, cubic ice, hexagonal ice, and liquid water [178], as well as a cage analysis code [109]. Bi *et al.* [22], Rodger *et al.* [98], Walsh *et al.* [268] and Guo *et al.* [90] have also written codes to identify water structures. Here we introduce an open-source C++ code, called

GRADE, which employs a hierarchical algorithm to identify the evolution of rings, cups and cages in molecular dynamics simulations of water molecules and computes the four-body order parameter F_4. This code is freely available on GitHub [152] and we anticipate that *GRADE* will serve as a C++ template code by the community which can be modified by research groups to account for their specific needs. Also, users that have significantly improved *GRADE* and would like to see their changes appended to new versions of the code are encouraged to contact the authors of this manuscript.

B.2 Methods

The aim of *GRADE* is to identify water structures, known as *cages*, that are the building blocks of gas clathrates. The current version of *GRADE* identifies 5^{12}, $6^2 5^{12}$ and $6^4 5^{12}$ *cages* that are made of 20, 24 and 28 water molecules, respectively. *GRADE* uses a hierarchical algorithm in which first-neighbors of all water molecules are identified followed by rings, cups, and cages. It computes all *rings* made of five and six first-neighbor water molecules [163]. *Cups*, i.e., half cages, are computed based on the connectivity of these rings to one another and *cages* are computed based on how cups are bound to each other. This approach was also used by other codes to identify water cages in solution [109] as well as to define order parameters to enable the development of new algorithm to simulate hydrate nucleation rates [22,23]. Hierarchical based algorithms have also been developed to identify polymorphic clathrate structures by Lauricella *et al.* [138]. Notice that the number of operations required to compute cages is very large and, thus, it can be very time consuming, specially for large systems and/or long trajectories. One advantage of *GRADE* is that it has built up schemes to speed up the calculation of cages. Below, we describe in detail how different quantities are computed within *GRADE* as well as the speed up schemes.

Table B.1 Molecular Dynamics Simulations Performed in This Work.

	Number of Simulations	Temperature (K)	Pressure (bar)	Number of water	Number of methane
A	1	298	1	500	0
B	4	270	500	3300	200

Two molecular dynamics setups are used to illustrate the different quantities computed by *GRADE*: one where water is in equilibrium in the liquid state and the other where water initially in the liquid state evolves towards forming 5^{12}, $6^2 5^{12}$ and $6^4 5^{12}$ cages around methane molecules in solution. Details of these setups are given in Table B.1. Simulations were performed using the open source software GROMACS [3] version 5.1 with the leap-frog algorithm to integrate the equation of motion in the NPT ensemble. Temperature was controlled using the v-rescale thermostat (τ_T = 1 ps), and pressure was kept constant using the Parrinello-Rahman barostat (τ_P = 0.1 ps). Periodic boundary conditions were applied in all directions. A cutoff of 1.0 nm was used to account for short-range nonbonded interactions. Long-range electrostatics were calculated using the Particle Mesh Ewald (PME) algorithm with grid spacing of 0.16 nm and 1.0 nm real-space cutoff. The TIP4P-ice model [1] was used to mimic water, and methane molecules were modeled using a united atom representation with 6–12 Lennard-Jones interactions (σ = 0.373 nm and ϵ =1.234 kJ/mol) [222]. Simulations A and B described in Table B.1 lasted for 5 ns and 1 μs each, respectively.

B.2.1 First-Neighbors

In Figure B.1, we show the oxygen-oxygen radial distribution function (RDF) of liquid water computed from simulation A in Table B.1. First and second peaks of this RDF

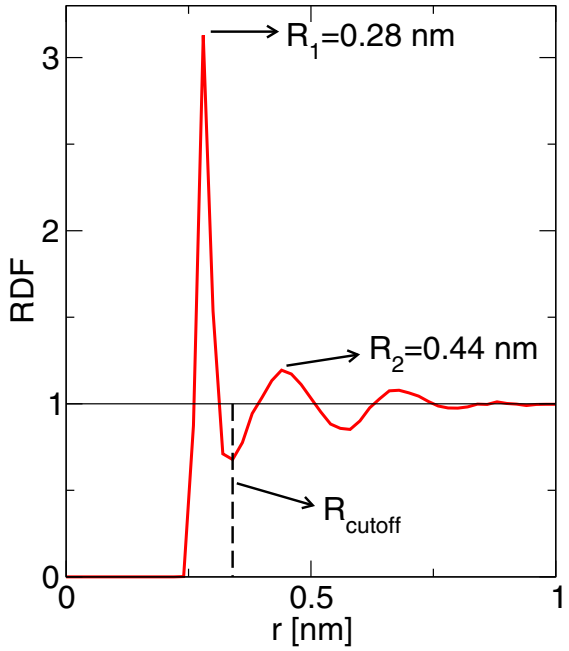


Figure B.1 Oxygen-Oxygen radial distribution function (RDF) of TIP4P/ice at 298 K and 1 atm. The minimum between first and second peaks of the RDF, i.e., R_{cutoff}, is used to define first neighbors of water molecules.

are located at distances R_1=0.28 nm and R_2=0.44 nm, respectively. The minimum between these peaks, i.e., R_{cutoff}, is commonly used as the cut-off radius to define first-neighbors. In other words, if the distance between oxygen atoms of water molecules i and j is smaller than R_{cutoff}, then these two water molecules are considered to be first-neighbors of each other. The default distance cut-off of *GRADE* is $R_{cutoff} = 0.35$ nm. This quantity can, however, be controlled using the flag '`-r`' to define alternative R_{cutoff} values, e.g., to compute cages in simulations of coarse-grained models where bond-lengths are given in normalized units [24, 61, 64, 153].

B.2.2 Rings

Rings are closed structures obtained by connecting first-neighbor water molecules to each other. For example, molecules i, j, k, l, m form a ring if: j is first-neighbor of i and k, l is a first-neighbor of k and m, and m is first-neighbor of i. The size of a ring corresponds to the number of water molecules in the loop. Thus, the size of

the ring formed by molecules i, j, k, l, and m is 5. Using a depth-first algorithm, *GRADE* searches for all five- and six-folded rings in the simulation box. The algorithm starts at one water molecule and explores as far as possible along each of the branches formed by first-neighbor water molecules before backtracking. An example is shown in Figure B.2a where the five-folded ring shown in red is found starting from molecule label 1. The branch formed by molecules 1 and 2 do not lead to the formation of a ring and, thus, the system is backtracked to explore the branch 1,3,4. Similarly, the latter branch does not lead to the formation of rings and the system goes on to explore branch 1,3,5,6 and so on. All the iterations required to find the five-folded ring are shown in panel b of Figure B.2. When using these rings to identify cups, *GRADE* spends a significant amount of time on loops going over all five- and six-folded rings. To reduce this performance bottleneck, *GRADE* ignores deformed rings that cannot account for stable cups or cages. In particular, *GRADE* identifies two types of deformations related to the convexity and planarity of ring structures.

Non-convexity. Five- and six-folded loop structures that are non-convex are excluded from the list of rings identified by *GRADE*. For a five-folded ring to be non-convex, the distance between vertices i and $i+2$ has to be smaller than $d_5 = 1.6 \times a$, where a is the distance between neighboring vertices and d_5 is the distance between vertices i and $i+2$ in a regular pentagon–see Figure B.3a. Thus, to flag non-convex five-folded loops, we impose that distances between oxygen atoms of water molecules i and $i+2$ have to be greater than $(1.6 \times R_{\text{cutoff}}) - \delta_1$, where $\delta_1 > 0$ is an adjustable parameter determining the degree of convexity to be tolerated. Similarly, we impose that distances between vertices i and i+3 in six-folded rings have to be greater than $(2 \times R_{\text{cutoff}}) - \delta_2$, with $\delta_2 > 0$–see Figure B.3b.

In Figs. B.3c, we show the dependence of the number of five- and six- folded rings on δ_1 and δ_2. As rings are allowed to exhibit more deformations, i.e., with increasing δ_1 and/or δ_2, the number of rings increase. To find reasonable values for δ_1

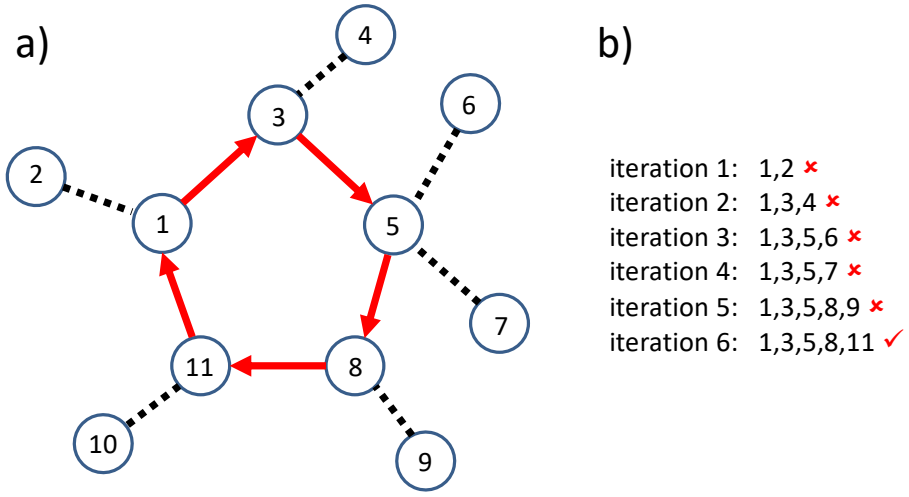


Figure B.2 Schematic representation of the depth-first search algorithm to find rings. (a) Eleven water molecules are represented by nodes and first-neighbor nodes are connected to each other by dashed lines. Red arrows are used to depict first-neighbor nodes that are forming a five-folded ring. (b) The six iterations required to find the five-folded ring in panel a starting from node 1 are listed. It is important to note that the same ring can be found by moving in a counter-clockwise direction from node "1" or starting from any other node in the ring.

and δ_2 to be used by *GRADE*, the dependence of the number of cages (defined later in the text) on these parameters is shown in Figure B.3d. In this figure, the number of $6^2 5^{12}$ and $6^4 5^{12}$ cages are computed by fixing δ_1 to its maximum (0.7) and by varying δ_2. The number of cages increases sharply with increasing δ_1 and δ_2 before reaching a plateau at $\delta_1 = 0.18$ nm and $\delta_2 = 0.26$ nm. Notice that the plateau is reached while the number of rings in Figure B.3c is still increasing. This implies that many rings found by *GRADE* when using $\delta_1 > 0.18$ nm and $\delta_2 > 0.26$ nm do not contribute to the formation of rings as they are too deformed. These rings can, therefore, be ignored and *GRADE* uses $\delta_1 = 0.18$ nm and $\delta_2 = 0.26$ nm as a default. Values of δ_1 and δ_2 can be modified using input flag option '`-d1`' and '`-d2`'.

Non-planarity. Some convex five- and six- folded rings in the simulation still exhibit high levels of deformity with some water molecules abnormally projected out of the plane formed by the other water molecules. Figure B.4 shows examples of

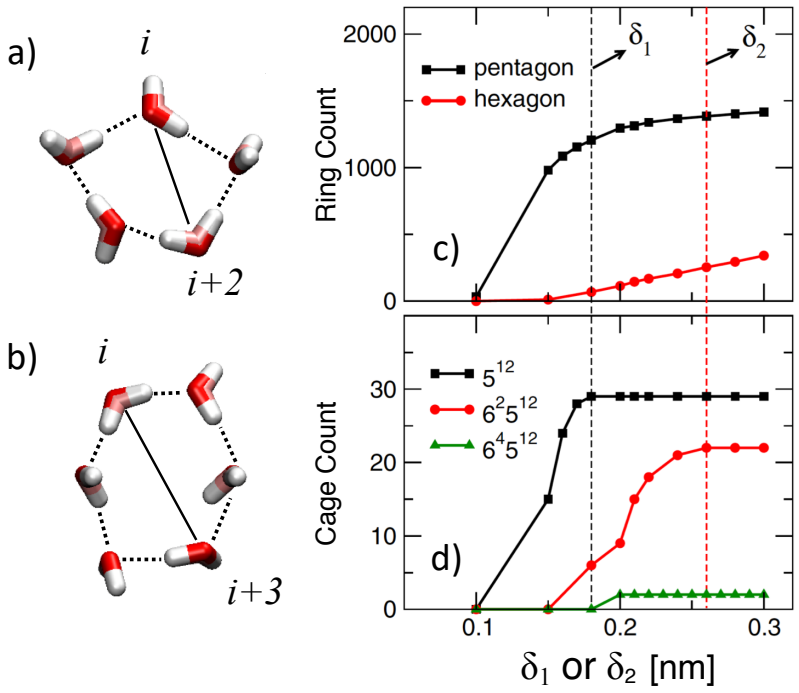


Figure B.3 Non-convexity condition and its effect on identified rings and cages. (a-b) Distance between vertices i and i+n (n=2, 3) for five- and six-folded rings calculated to exclude convex loops. Licorice representation is used to depict water molecules and hydrogen bonds are illustrated by dashed lines. (c-d) Number of loops and cages identified by *GRADE* as a function of δ_1 and δ_2. Black and red dashed lines show the recommended default values for δ_1 and δ_2, respectively.

deformed (panels *a* and *b*) and non-deformed (panels *c* and *d*, top) five- and six-folded rings. Since stable cages cannot be constructed with highly deformed rings, they do not need to be included in the list of identified rings. *GRADE* identifies this type of deformity by dividing five- and six- folded rings into three and four planes, respectively, as shown in Figure B.4c,d (bottom). Angles between these planes (i.e., θ_{P1-P2}, θ_{P2-P3} and θ_{P3-P4}) are computed and are required to be smaller than a given cutoff value, i.e., θ_{cutoff}. In Figure B.5a, we show that the number of five- and six-folded rings increases with increasing θ_{cutoff} as rings found by *GRADE* are allowed to be more deformed. In Figure B.5b, we show the dependence of the number of cages (defined using $\delta_1 = 0.18$ nm and $\delta_2 = 0.26$ nm) on θ_{cutoff}. The number of cages increases with θ_{cutoff} up to 45° where it reaches a plateau. This implies that there are many rings defined by a cutoff angle greater than 45° but these deformed rings do not contribute to the formation of cages. Based on this analysis, *GRADE* uses 45° as its default cut-off angle as this value of θ_{cutoff} reduces significantly the number of rings without affecting the number of cages found. However, the default cutoff angle can be modified using the input flag option '`-theta`'.

In Figure B.5c, we show that a large fraction of the rings found by *GRADE* are either non-convex (15% and 27% for five- and six-folded rings, respectively) or exhibit deformities related to planarity (7% and 16% for five- and six-folded rings, respectively). By ignoring these rings the run time of *GRADE* can be reduced by more than 80%. However, we recommend for users to test the effect of δ_1, δ_2, and θ_{cutoff} on their particular systems.

B.2.3 Cups

To define cups and cages, it is convenient to introduce a few new concepts. Two rings are considered *neighbors* to each other if they have two vertices (one edge) in common. The number of neighboring rings of a reference ring i is known as the *coordination*

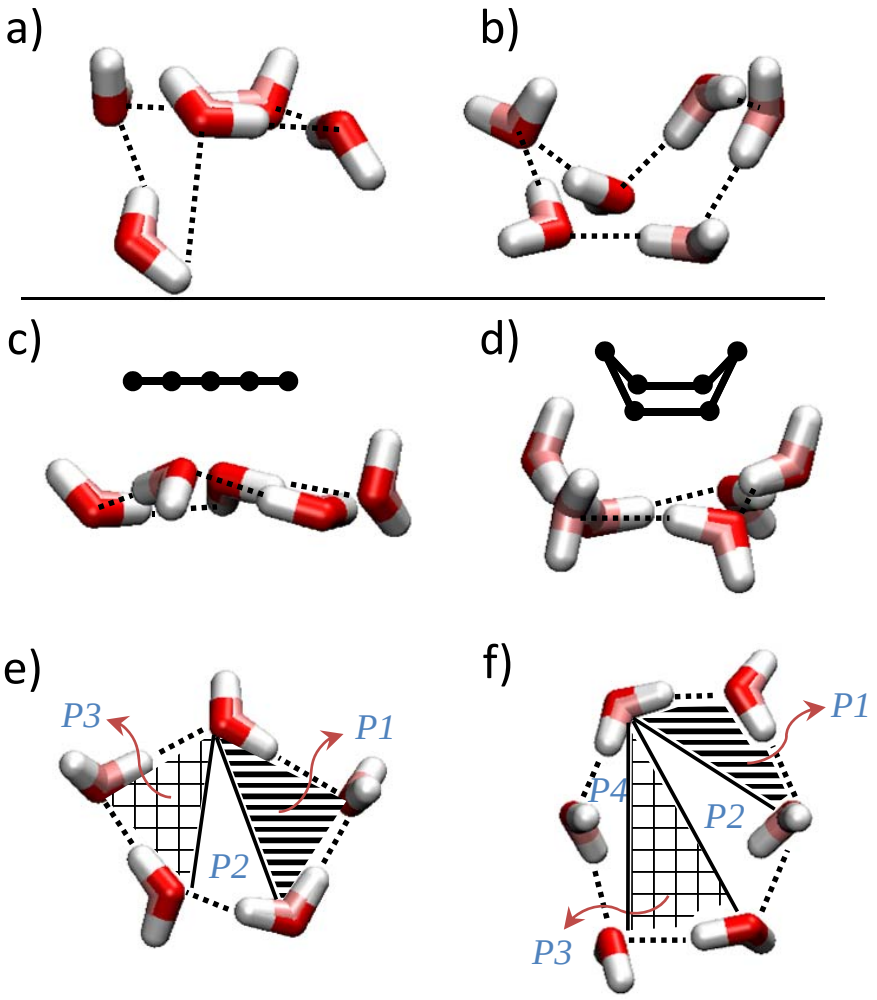


Figure B.4 Non-planar ring deformity in five- and six-folded rings. (a-b) Examples of deformities due to non-planarity of vertices in five and six folded rings. (c-d) Side view of five and six folded rings, showing their planar and *boat* conformation. (e-f) Division of pentagon and hexagon structure into three and four planes in order to determine out of plane vertices.

number of i. Thus, a ring is *fully-coordinated* if it has the same number of neighbor rings as its size. Rings that are neighbors of a fully-coordinated ring are called *lateral* rings.

The set of water molecules comprising fully-coordinated and lateral rings form a cup when each of the lateral rings are neighbors to two other lateral rings. To illustrate this definition, Figure B.6 depicts cups with their fully-coordinated five and six-folded rings in red and their lateral rings in black. Using five and six folded rings, the two most common types of cups that can be formed are:

- 5^6 cup in which the size of the fully coordinated and lateral rings are five. The upper index 6 stands for the number of five-folded rings.
- $6^1 5^6$ cup in which the fully coordinated ring is a six-folded ring and the lateral rings are five-folded. Upper indices stand for the number of six- and five-folded rings.

GRADE finds shared edges between pentagon-pentagon and pentagon-hexagon rings to determine 5^6 and $6^1 5^6$ cups. All cups are found and assigned a unique identifying number to avoid double counting.

B.2.4 Cages

The set of molecules comprised by two cups forms a 5^{12} or $6^2 5^{12}$ cage when each lateral ring of one cup is neighbor to two lateral rings of the other cup as depicted in Figure B.6a-b. A $6^4 5^{12}$ cage is formed by the set of molecules in four $6^1 5^6$ cups in which each pair of these cups shares two five-folded rings with each other. The three types of cages that *GRADE* detects are:

- 5^{12} cages, formed by two 5^6 cups and comprising 20 water molecules.
- $6^2 5^{12}$ cages, formed by two $6^1 5^6$ cups and containing 24 water molecules.
- $6^4 5^{12}$ cages, formed by four $6^1 5^6$ cups and containing 28 water molecules.

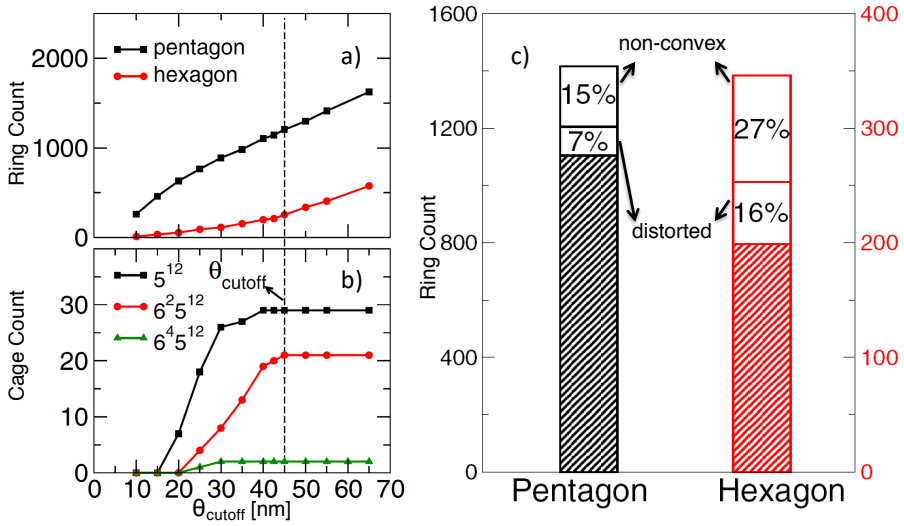


Figure B.5 Effects of deformities on identified structures. (a-b) dependence of θ_{cutoff} on loops and cages. Dashed line shows the recommended default value for θ_{cutoff}. (c) Percentage of excluded loops by imposing non-convexity and planarity condition.

To avoid double counting, *GRADE* assigns a unique identifier number to each cage. In addition to water molecules, *GRADE* can also process two types of solute with arbitrary names in the input gro file. Distances between center of mass of each cage and atoms of solutes are calculated and if these distances are smaller than 0.2 nm, the solutes are considered trapped inside the cage. A summary file containing the number of filled and empty cages is provided as output file. Identifying less common types of cages, e.g., $6^3 5^{12}$ cages [113], is planned for the next version of *GRADE*.

B.2.5 Order Parameter F_4

In addition to cages, another quantity that is often used to characterize the state of water molecules is the four-body order parameter $F_4 = \langle cos\ 3\phi \rangle$, where ϕ is the torsion angle of the configuration H-O$\cdots$O-H formed with the outermost hydrogen atoms of two neighboring water molecules. This quantity can be used to distinguish between different tetrahedral networks adopted by water. The flag '`-f4`' can be used to compute the F_4 order parameter averaged over all pairs of water molecules in the simulation box.

B.2.6 Program Structure

Source Files *GRADE* is written in C++ and is made up of a main program file and two supporting resource files:

- `GRADE.cpp` – main program file which reads input files, writes output files, and calls several functions;
- `Functions.hpp` – header file, includes function prototypes;
- `Functions.cpp` – resource file, includes functions needed to find rings, cups and cages.

Files contain original source code written by the authors and they can be compiled using common C++ compilers. A complete list of input and output options of *GRADE* can be printed using '`-h`' flag.

Input and Output File *GRADE* reads the three-dimensional position of atoms in a periodic box provided in *gro* format. Several software packages, including *pdb2gmx* which is part of the GROMACS [3] suite and InterMol [238] can be used to convert *pdb* or trajectory files to *gro* format. The name of the input file including the extension (.gro) should follow the '`-i`' flag. The frequency with which frames in the input file are read can be controlled with the flag '`-fr`'.

Four output files are produced by default. One of these files has extension ".xvg", and it contains eight columns which are the frame number, simulation time (if time is provided in the input file), the total number and the number of filled 5^{12}, $6^2 5^{12}$ and $6^4 5^{12}$ cages, respectively. Filled cages are cages that host a solute molecule, i.e., residue name other than SOL in the input file. Three files with ".gro" extension are also produced by *GRADE*. One file for each type of cages which contain positions of water molecules forming the cages, all solute molecules, followed by solute molecules trapped inside cages. The molecule name of the latter is "CBX". This allows trapped solutes to be visualized independently from other solutes using *VMD*

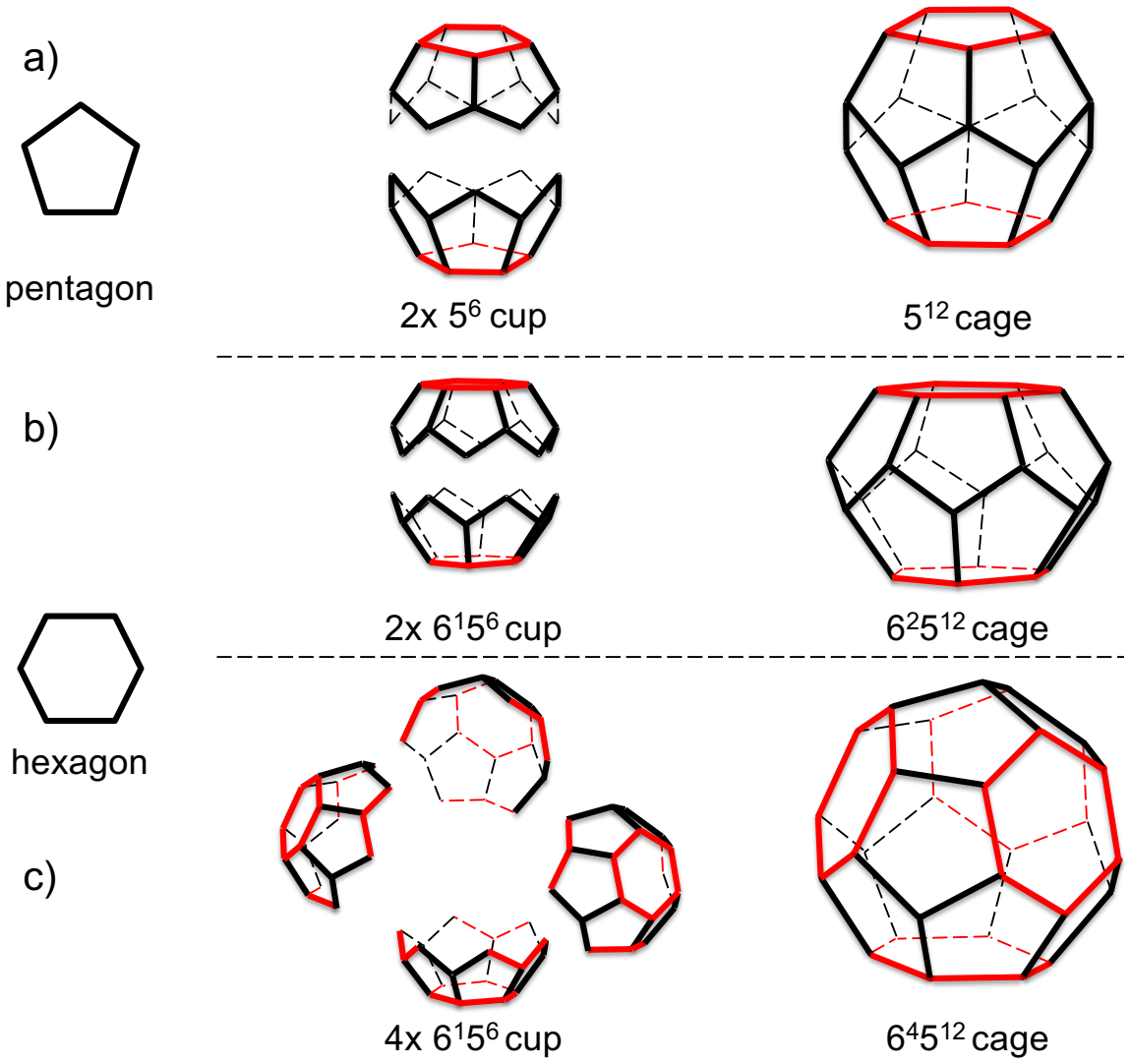


Figure B.6 Structure of cups and cages identified by *GRADE*. a) Five-folded ring (pentagon), two 5^6 cups and 5^{12} cage, b) six-folded ring (hexagon), two $6^1 5^6$ cups and $6^2 5^{12}$ cage, c) four $6^1 5^6$ cups and $6^4 5^{12}$ cage. Fully-coordinated and lateral rings are shown in red and black, respectively.

or other atomic visualization software. By default, names of these output files are formed by concatenating the input name with "_cage512_*i*.gro", "_cage62512_*i*.gro" and "_cage64512_*i*.gro", where i stands for the frame number. Alternative output names can be provided using the '`-o`' flag. The frequency with which output files are written can be controlled with the '`-dt`' flag. By default '`-dt`' and '`-fr`' are set to one.

B.3 Results

As an example of how *GRADE* can be used to study clathrate formation, we now analyze the four simulations B defined in Table B.1. First, *GRADE* needs to be compiled using, e.g., the GNU compiler collection (or any C++ compiler):

```
g++ GRADE.cpp Functions.cpp -o GRADE
```

A Makefile provided with the source files compiles the executable and ensures that the minimum GNU version requirement is met. If trajectory of water and solute molecules are provided in a file named "trajectory.gro", *GRADE* can be executed using the command line:

```
./GRADE -i trajectory.gro -f4 yes
```

This generates five output files: "trajectory.xvg", "trajectory_cage512_*frame*.gro", "trajectory_cage62512_*frame*.gro", "trajectory_cage64512_*frame*.gro" and "F4.gro". A separate gro-file is generated for each analyzed frame or as specified in the command line using the '`-dt`' flag. Figure B.7*a-d* shows the time-evolution of the number of 5^{12} (in red), $6^2 5^{12}$ (blue) and $6^4 5^{12}$ (green) cages of the four simulations in setup B, defined in Table B.1. This time evolution is provided in the output file "trajectory.xvg".

Initially, there are no cages in the simulation box as the system is in the liquid state. After an induction time that varies from $\sim 100 - 250$ ns for the four systems studied, the number of cages increases abruptly to approximately 10-15. This is consistent with theories of blob formation [75] in which the spontaneous nucleation of cages from the liquid state requires the formation of a critical nucleus. After the formation of this nucleus, the number of cages increase in a non-continuous manner whereby short periods of annihilation are followed by regrowth.

Figure B.7e shows snapshots of cages and methane molecules at four points of one of the trajectories (see panel b). Snapshot I corresponds to the first 5^{12} cage that is filled with a methane molecule in the simulation (time = 110 ns). This cage is, however, unstable and it disappears in snapshot II (time =180 ns) which shows the first stable cluster made of eight 5^{12} (in red) and six $6^2 5^{12}$ cages (in blue). This cluster grows through the addition of newly formed cages as depicted in snapshots III (time 250 ns) and IV (time = 580 ns).

The four-body order parameter, $F_4 = \langle cos\ 3\phi \rangle$, is plotted for the four simulations in Figure B.8. This parameter is provided in the output file F4.xvg. F_4 takes values of 0.7, 0 and -0.4 for SI hydrate, liquid water and ice, respectively to distinguish different phases of water [114, 214]. In our simulations, F_4 initially starts from values close to zero, which represents the liquid state, and it increases abruptly as water molecules evolve into clathrate structures.

B.4 Conclusion

In this dissertation, we present a computer code to analyze water structures related to the formation of gas clathrate. The code identifies 5^{12}, $6^2 5^{12}$ and $6^4 5^{12}$ cages as well as the four-body order parameter (F_4) given the atomic coordinates of water molecules. These are main quantities used to quantify clathrate formation in molecular dynamics simulations. We anticipate that this open-source code will serve as a template by

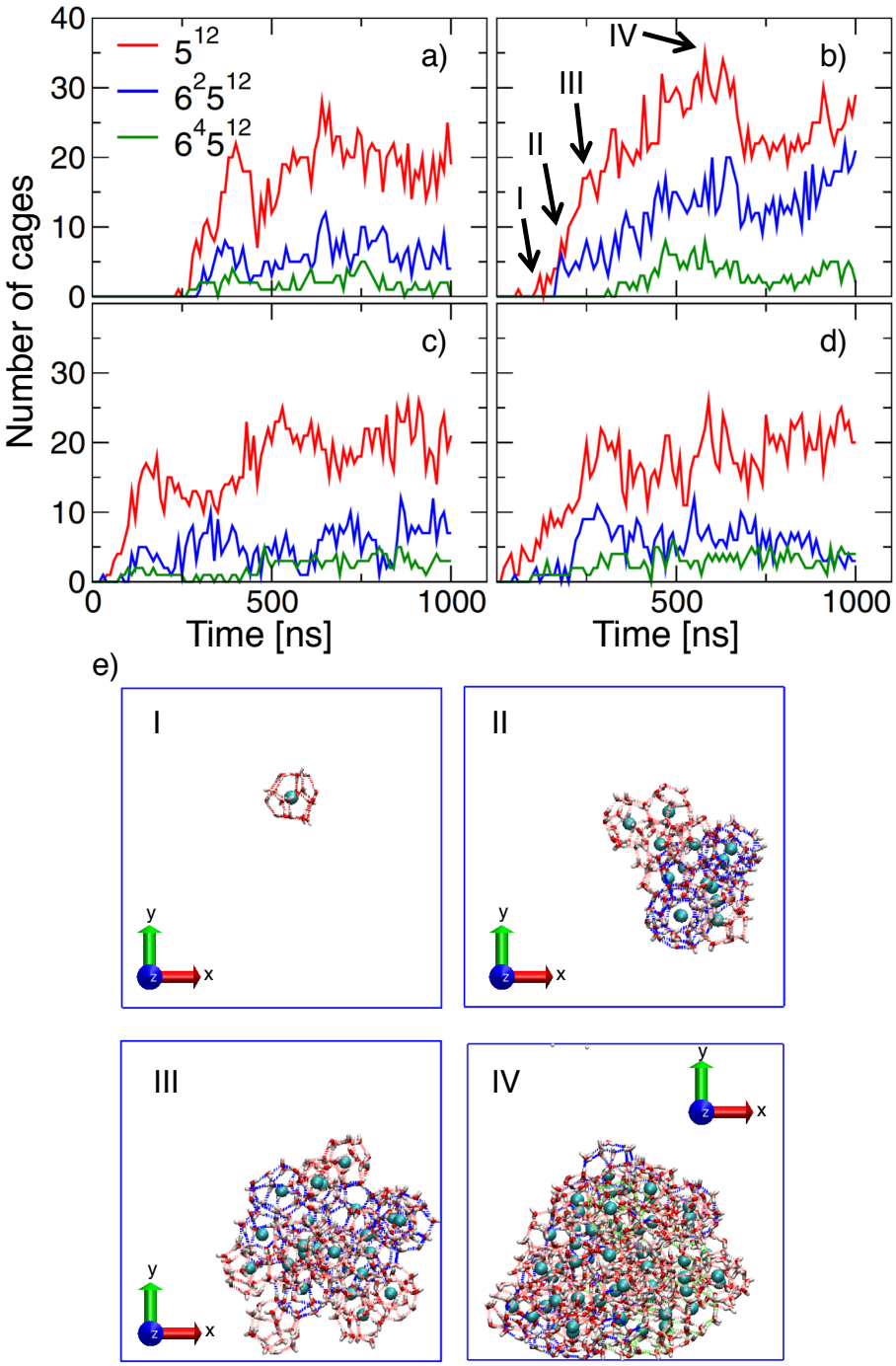


Figure B.7 Cage formation over time. (a-d) Number of 5^{12}, $6^2 5^{12}$ and $6^4 5^{12}$ cages found by *GRADE* in four simulations of 3300 TIP4P/ice water molecules and 200 methane molecules. (e) Visualization of different cages at various points during growth phase specified in panel (b). Water molecules are shown in Licorice representation and Carbon atom of Methane molecules are shown in cyan. Hydrogen bonds are depicted by red, blue and green dashed lines in cages respectively.

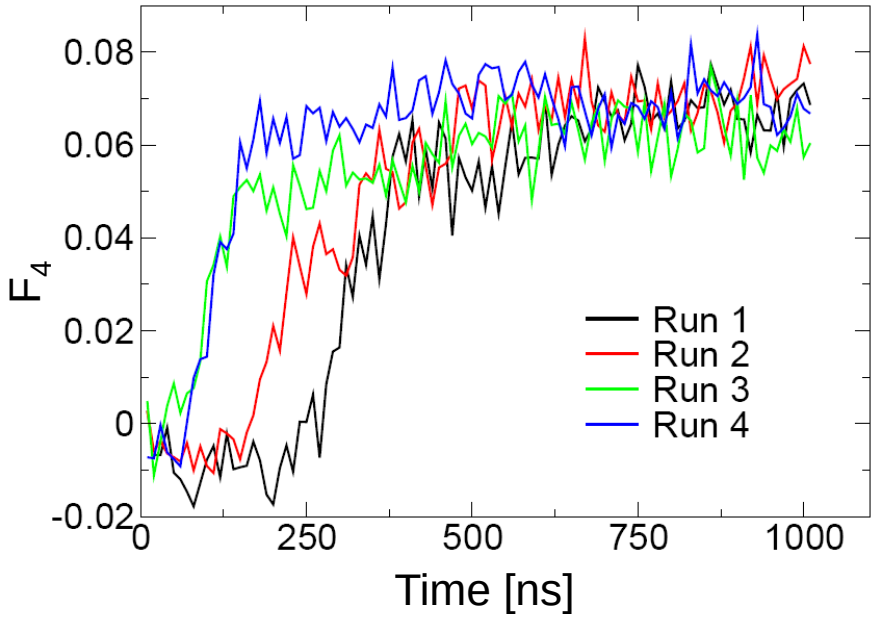


Figure B.8 Time evolution of F_4 order parameter of the four individual simulations at T=270 K and 500 bar.

the community which should be of interest to researchers studying not only natural gases but also antifreezing proteins and hydrophobic interactions. First, the method used to analyze rings in order to compute cages is presented followed by a brief discussion of the computer code. Second, an example of how the code can be used to study the spontaneous nucleation of methane clathrates is presented. We anticipate that the freely available source code will enable research groups to easily analyze their simulation results. Also, the code can be modified to allow the investigation of other order parameters to quantify water structures. It should be mentioned that we have implemented restrictions on the conditions required by water molecules to be considered part of five- and six-folded rings, however, these restrictions have no effect on the number of cages found by the code and they account for a significant speed up of the computer code. Further speed up of *GRADE* can be provided through parallelization which is planned for a future version of the code.

BIBLIOGRAPHY

[1] J. Abascal, E. Sanz, R. G. Fernández, and C. Vega. A potential model for the study of ices and amorphous water: Tip4p/ice. *Journal of Chemical Physics*, 122(23):234511, 2005.

[2] A. Abelein, J. Jarvet, A. Barth, A. Graslund, and J. Danielsson. Ionic strength modulation of the free energy landscape of aβ40 peptide fibril formation. *Journal of the American Chemical Society*, 138(21):6893–6902, 2016.

[3] M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess, and E. Lindahl. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1:19–25, 2015.

[4] L. Adler-Abramovich, L. Vaks, O. Carny, D. Trudler, A. Magno, A. Caflisch, D. Frenkel, and E. Gazit. Phenylalanine assembly into toxic fibrils suggests amyloid etiology in phenylketonuria. *Nature Chemical Biology*, 8(8):701–706, 2012.

[5] B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. *Molecular Biology of the Cell, 4th edition*, volume 91. New York, NY: Garland Science, 2002.

[6] M. P. Allen and D. J. Tildesley. *Computer Simulation of Liquids*. Oxford, UK: Oxford University Press, 2017.

[7] C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181(4096):223–230, 1973.

[8] A. H. Armstrong, J. Chen, A. F. McKoy, and M. H. Hecht. Mutations that replace aromatic side chains promote aggregation of the alzheimer's aβ peptide. *Biochemistry*, 50(19):4058–4067, 2011.

[9] J. Armstrong, M. Boada, M. Rey, N. Vidal, and I. Ferrer. Familial alzheimer disease associated with a713t mutation in app. *Neuroscience Letters*, 370(2-3):241–243, 2004.

[10] R. L. Baldwin. Energetics of protein folding. *Journal of Molecular Biology*, 371(2):283, 2007.

[11] R. L. Baldwin and G. D. Rose. Is protein folding hierarchic? i. local structure and peptide folding. *Trends in Biochemical Sciences*, 24(1):26–33, 1999.

[12] K. A. Ball, A. H. Phillips, D. E. Wemmer, and T. Head-Gordon. Differences in β-strand populations of monomeric aβ40 and aβ42. *Biophysical Journal*, 104(12):2714–2724, 2013.

[13] G. Bates. Huntingtin aggregation and toxicity in huntington's disease. *The Lancet*, 361(9369):1642–1644, 2003.

[14] A. Baumketner and J.-E. Shea. Free energy landscapes for amyloidogenic tetrapeptides dimerization. *Biophysical Journal*, 89(3):1493–1503, 2005.

[15] O. M. Becker, A. D. MacKerell Jr, B. Roux, and M. Watanabe. *Computational Biochemistry and Biophysics*. New York, NY: CRC Press, 2001.

[16] G. Bellesia and J.-E. Shea. Effect of β-sheet propensity on peptide aggregation. *The Journal of Chemical Physics*, 130(14):145103, 2009.

[17] G. Bellesia and J.-E. Shea. What determines the structure and stability of kffe monomers, dimers, and protofibrils? *Biophysical Journal*, 96(3):875–886, 2009.

[18] H. J. Berendsen. *Simulating the physical world: hierarchical modeling from quantum mechanics to fluid dynamics*. Cambridge, UK: Cambridge University Press, 2007.

[19] H. J. Berendsen, J. v. Postma, W. F. van Gunsteren, A. DiNola, and J. Haak. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics*, 81(8):3684–3690, 1984.

[20] J. P. Bernacki and M. M. Regina. Length-dependent aggregation of uninterrupted polyalanine peptides. *Biochemistry*, 50:9200, 2011.

[21] A. Berteotti, A. Barducci, and M. Parrinello. Effect of urea on the β-hairpin conformational ensemble and protein denaturation mechanism. *Journal of the American Chemical Society*, 133:17200, 2011.

[22] Y. Bi and T. Li. Probing methane hydrate nucleation through forward flux sampling method. *The Journal of Physical Chemistry B*, 118:13324, 2014.

[23] Y. Bi, A. Porras, and T. Li. Free energy landscape and molecular pathways of gas hydrate nucleation. *The Journal of Chemical Physics*, 145(21):211909, 2016.

[24] A. Bizjak, T. Urbič, V. Vlachy, and K. A. Dill. The three-dimensional" mercedes benz" model of water. *Acta Chimica Slovenica*, 54(3), 2007.

[25] M. Blaber, X. jun Zhang, J. D. Lindstrom, S. D. Pepiot, W. A. Baase, and B. W. Matthews. Determination of α-Helix Propensity within the Context of a Folded Protein: Sites 44 and 131 in Bacteriophage T4 Lysozyme. *Journal of Molecular Biology*, 235:600–624, 1994.

[26] K. Blennow. Leon mj de, zetterberg h (2006) alzheimer's disease. *The Lancet*, 368:387–403, 2006.

[27] H. R. Bosshard, D. N. Marti, and I. Jelesarov. Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. *Journal of Molecular Recognition*, 17(1):1–16, 2004.

[28] C. J. Bowerman and B. L. Nilsson. Review self-assembly of amphipathic β-sheet peptides: Insights and applications. *Peptide Science*, 98(3):169–184, 2012.

[29] C. Branden and J. Tooze. *Introduction to protein structure*. New York, NY:Garland Publication, 1991.

[30] G. Bussi, D. Donadio, and M. Parrinello. Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1):014101, 2007.

[31] G. Bussi, A. Laio, and M. Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *Physical Review Letters*, 96(9):090601, 2006.

[32] S. Cabani, P. Gianni, V. Mollica, and L. Lepori. Group contributions to the thermodynamic properties of non-ionic organic solutes in dilute aqueous solution. *Journal of Solution Chemistry*, 10(8):563–595, 1981.

[33] M. J. Cannon, A. D. Williams, R. Wetzel, and D. G. Myszka. Kinetic analysis of beta-amyloid fibril elongation. *Analytical Biochemistry*, 328(1):67–75, 2004.

[34] F. Caputo, F. Cascetta, G. Lamanna, G. Rotondo, and A. Soprano. The role of methane hydrates in the failure of a gas pipeline. *Key Engineering Materials*, 577:377–380, 2014.

[35] S. Cavalli, F. Albericio, and A. Kros. Amphiphilic peptides and their cross-disciplinary role as building blocks for nanoscience. *Chemical Society Reviews*, 39(1):241–263, 2010.

[36] J. B. Chaires. Calorimetry and thermodynamics in drug design. *Annual Review of Biophysics*, 37:135–151, 2008.

[37] H. Chan, Z. Zhang, S. Wallin, and Z. Liu. Cooperativity, local-nonlocal coupling, and nonnative interactions: Principles of protein folding from coarse-grained models. *Annual Review of Physical Chemistry*, 62:301, 2011.

[38] W. E. Chang, T. Takeda, E. P. Raman, and D. K. Klimov. Molecular dynamics simulations of anti-aggregation effect of ibuprofen. *Biophysical Journal*, 98(11):2662–2670, 2010.

[39] W.-T. Chen, C.-J. Hong, Y.-T. Lin, W.-H. Chang, H.-T. Huang, J.-Y. Liao, Y.-J. Chang, Y.-F. Hsieh, C.-Y. Cheng, H.-C. Liu, et al. Amyloid-beta (aβ) d7h mutation increases oligomeric aβ42 and alters properties of aβ-zinc/copper assemblies. *PloS One*, 7(4):e35807, 2012.

[40] M. S. Cheung, A. E. García, and J. N. Onuchic. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proceedings of the National Academy of Sciences U.S.A.*, 99(2):685–690, 2002.

[41] F. Chiti and C. M. Dobson. Protein misfolding, functional amyloid, and human disease. *Annual Review of Biochemistry*, 75:333, 2006.

[42] F. Chiti, M. Stefani, N. Taddei, G. Ramponi, and C. M. Dobson. Rationalization of the effects of mutations on peptide andprotein aggregation rates. *Nature*, 424(6950):805–808, 2003.

[43] F. Chiti, N. Taddei, F. Baroni, C. Capanni, M. Stefani, G. Ramponi, and C. M. Dobson. Kinetic partitioning of protein folding and aggregation. *Nature Structural & Molecular Biology*, 9(2):137–143, 2002.

[44] P. Y. Chou and G. D. Fasman. Empirical predictions of protein conformation. *Annual Review of Biochemistry*, 47(1):251–276, 1978.

[45] S. I. Cohen, S. Linse, L. M. Luheshi, E. Hellstrand, D. A. White, L. Rajah, D. E. Otzen, M. Vendruscolo, C. M. Dobson, and T. P. Knowles. Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proceedings of the National Academy of Sciences*, 110(24):9758–9763, 2013.

[46] O. Collet, , and C. Chipot. Non-arrhenius behavior in the unfolding of a short, hydrophobic α-helix. complementarity of molecular dynamics and lattice model simulations. *Journal of American Chemical Society*, 125:6573–6580, 2003.

[47] J. Colletier, A. Laganowsky, M. Landau, M. Zhao, A. Soriaga, L. Goldschmidt, D. Flot, D. Cascio, M. M.R. Sawaya, and D. Eisenberg. Molecular basis for amyloid-β polymorphism. *Proceedings of the National Academy of Sciences USA*, 108:16938, 2011.

[48] T. S. Collett, R. Lewis, W. J. Winters, M. W. Lee, K. K. Rose, and R. M. Boswell. Downhole well log and core montages from the mount elbert gas hydrate stratigraphic test well, alaska north slope. *Marine and Petroleum Geology*, 28(2):561–577, 2011.

[49] P. Cras, F. Van Harskamp, L. Hendriks, C. Ceuterick, C. Van Duijn, S. Stefanko, A. Hofman, J. Kros, C. Van Broeckhoven, and J. Martin. Presenile alzheimer dementia characterized by amyloid angiopathy and large amyloid core type senile plaques in the app 692ala-gly mutation. *Acta Neuropathologica*, 96(3):253–260, 1998.

[50] O. Crescenzi, S. Tomaselli, R. Guerrini, S. Salvadori, A. M. D'Ursi, P. A. Temussi, and D. Picone. Solution structure of the alzheimer amyloid β-peptide (1–42) in an apolar microenvironment: Similarity with a virus fusion domain. *European Journal of Biochemistry*, 269(22):5642–5648, 2002.

[51] R. Cukalevski, B. Boland, B. Frohm, E. Thulin, D. Walsh, and S. Linse. Role of aromatic side chains in amyloid β-protein aggregation. *ACS Chemical Neuroscience*, 3(12):1008–1016, 2012.

[52] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An nlog(n) method for ewald sums in large systems. *The Journal of Chemical Physics*, 98(12):10089–10092, 1993.

[53] P. Das and S. Matysiak. Direct characterization of hydrophobic hydration during cold and pressure denaturation. *The Journal of Physical Chemistry B*, 116(18):5342–5348, 2012.

[54] P. Das, B. Murray, and G. Belfort. Alzheimer's protective a2t mutation changes the conformational landscape of the aβ1–42 monomer differently than does the a2v mutation. *Biophysical Journal*, 108(3):738–747, 2015.

[55] W. Dauer and S. Przedborski. Parkinson's disease: mechanisms and models. *Neuron*, 39(6):889–909, 2003.

[56] F. G. De Felice, M. N. Vieira, L. M. Saraiva, J. D. Figueroa-Villar, J. Garcia-Abreu, R. Liu, L. Chang, W. L. KLEIN, and S. T. Ferreira. Targeting the neurotoxic species in alzheimer's disease: inhibitors of aβ oligomerization. *The FASEB Journal*, 18(12):1366–1372, 2004.

[57] N. S. de Groot, T. Parella, F. X. Aviles, J. Vendrell, and S. Ventura. Ile-phe dipeptide self-assembly: clues to amyloid formation. *Biophysical Journal*, 92(5):1732–1741, 2007.

[58] P. G. Debenedetti and S. Sarupria. Hydrate molecular ballet. *Science*, 326(5956):1070–1071, 2009.

[59] C. L. Dias. Unifying microscopic mechanism for pressure and cold denaturations of proteins. *Physical Review Letters*, 109(4):048104, 2012.

[60] C. L. Dias. Unifying microscopic mechanism for pressure and cold denaturations of proteins. *Physical Review Letters*, 109:048104–048109, 2012.

[61] C. L. Dias, T. Ala-Nissila, M. Grant, and M. Karttunen. Three-dimensional "mercedes-benz" model for water. *The Journal of Chemical Physics*, 131(5):054505, 2009.

[62] C. L. Dias, T. Ala-Nissila, M. Karttunen, I. Vattulainen, and M. Grant. Microscopic mechanism for cold denaturation. *Physical Review Letters*, 100(11):118101, 2008.

[63] C. L. Dias, T. Ala-Nissila, J. Wong-ekkabut, I. Vattulainen, M. Grant, and M. Karttunen. The hydrophobic effect and its role in cold denaturation. *Cryobiology*, 60(1):91–99, 2010.

[64] C. L. Dias, T. Hynninen, T. Ala-Nissila, A. S. Foster, and M. Karttunen. Hydrophobicity within the three-dimensional mercedes-benz model: Potential of mean force. *The Journal of Chemical Physics*, 134(6):02B620, 2011.

[65] C. L. Dias, M. Karttunen, and H. Chan. Hydrophobic interactions in the formation of secondary structures in small peptides. *Physical Review E*, 84(4):041931, 2011.

[66] K. Dill. Dominant forces in protein folding. *Biochemistry*, 29:7133, 1990.

[67] K. Dill and S. Bromberg. *Molecular driving forces: statistical thermodynamics in biology, chemistry, physics, and nanoscience*. New York, NY: Garland Science, 2012.

[68] F. Ding, J. M. Borreguero, S. V. Buldyrey, H. E. Stanley, and N. V. Dokholyan. Mechanism for the α-helix to β-hairpin transition. *Proteins: Structure, Function, and Bioinformatics*, 53:220, 2003.

[69] C. M. Dobson. Protein folding and misfolding. *Nature*, 426:884, 2003.

[70] T. M. Doran, A. J. Kamens, N. K. Byrnes, and B. L. Nilsson. Role of amino acid hydrophobicity, aromaticity, and molecular volume on iapp $(20-29)$ amyloid self-assembly. *Proteins: Structure, Function, and Bioinformatics*, 80(4):1053–1065, 2012.

[71] C. M. Doyle, J. A. Rumfeldt, H. R. Broom, A. Broom, P. B. Stathopulos, K. A. Vassall, J. J. Almey, and E. M. Meiering. Energetics of oligomeric protein folding and association. *Archives of Biochemistry and Biophysics*, 531(1):44–64, 2013.

[72] G. Duarte Ramos Matos, D. Y. Kyu, H. H. Loeffler, J. D. Chodera, M. R. Shirts, and D. L. Mobley. Approaches for calculating solvation free energies and enthalpies demonstrated with an update of the freesolv database. *Journal of Chemical & Engineering Data*, 62(5):1559–1569, 2017.

[73] J. T. Edsall. Apparent molal heat capacities of amino acids and other organic compounds. *Journal of the American Chemical Society*, 54:1506–1507, 1935.

[74] P. Englezos and J. D. Lee. Gas hydrates: A cleaner source of energy and opportunity for innovative technologies. *Korean Journal of Chemical Engineering*, 22(5):671–681, 2005.

[75] N. J. English, M. Lauricella, and S. Meloni. Massively parallel molecular dynamics simulation of formation of clathrate-hydrate precursors at planar water-methane interfaces: Insights into heterogeneous nucleation. *The Journal of Chemical Physics*, 140(20):204714, 2014.

[76] N. J. English and J. MacElroy. Perspectives on molecular simulation of clathrate hydrates: Progress, prospects and challenges. *Chemical Engineering Science*, 121:133–156, 2015.

[77] 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. Alzheimer's disease amyloid propagation by a template-dependent dock-lock mechanism. *Biochemistry*, 39:6288–6295, 2000.

[78] M. Fändrich and C. M. Dobson. The behaviour of polyamino acids reveals an inverse side chain effect in amyloid structure formation. *The EMBO Journal*, 21(21):5682–5690, 2002.

[79] M. Fändrich, M. A. Fletcher, and C. M. Dobson. Amyloid fibrils from muscle myoglobin. *Nature*, 410(6825):165–166, 2001.

[80] M. Fändrich, M. A. Fletcher, and C. M. Dobson. Amyloid fibrils from muscle myoglobin. *Nature*, 410(6825):165, Mar. 2001.

[81] H. S. Frank and M. W. Evans. Free volume and entropy in condensed systems iii. entropy in binary liquid mixtures; partial molal entropy in dilute solutions; structure and thermodynamics in aqueous electrolytes. *Journal of Chemical Physics*, 13:507, 1945.

[82] A. E. Garcia, G. Hummer, and D. M. Soumpasis. Hydration of an α-helical peptide: comparison of theory and molecular dynamics simulation. *Proteins: Structure, Function, and Bioinformatics*, 27(4):471–480, 1997.

[83] C. P. Garnham, R. L. Campbell, and P. L. Davies. Anchored clathrate waters bind antifreeze proteins to ice. *Proceedings of the National Academy of Sciences*, 108(18):7363–7367, 2011.

[84] P. S. Georgoulia and N. M. Glykos. Using j-coupling constants for force field validation: application to hepta-alanine. *The Journal of Physical Chemistry B*, 115(51):15221–15227, 2011.

[85] P. Ghosh, A. Vaidya, A. Kumar, and V. Rangachari. Determination of critical nucleation number for a single nucleation amyloid-β aggregation model. *Mathematical Biosciences*, 273:70–79, 2016.

[86] M. Gobbi, L. Colombo, M. Morbin, G. Mazzoleni, E. Accardo, M. Vanoni, E. Del Favero, L. Cantù, D. A. Kirschner, C. Manzoni, et al. Gerstmann-sträussler-scheinker disease amyloid protein polymerizes according to the "dock-and-lock" model. *Journal of Biological Chemistry*, 281(2):843–849, 2006.

[87] J. Grace, T. Collett, F. Colwell, P. Englezos, E. Jones, R. Mansell, P. Meekison, R. Ommer, M. Pooladi-Darvish, M. Riedel, J. Ripmeester, C. Shipp, and E. Willoughby. Energy from gas hydrates—assessing the opportunities and challenges for canada. *Report of the Expert Panel on Gas Hydrates, Council of Canadian Academies*, 2008.

[88] A. Gringauz. *Introduction to Medicinal Chemistry: How Drugs Act and why.* Weinheim, DE: Wiley-VCH, 1997.

[89] G.-J. Guo and P. M. Rodger. Solubility of aqueous methane under metastable conditions: Implications for gas hydrate nucleation. *The Journal of Physical Chemistry B*, 117:6498–6504, 2013.

[90] G.-J. Guo, Y.-G. Zhang, C.-J. Liu, and K.-H. Li. Using the face-saturated incomplete cage analysis to quantify the cage compositions and cage linking structures of amorphous phase hydrates. *Physical Chemistry Chemical Physics*, 13(25):12048–12057, 2011.

[91] G.-J. Guo, Y.-G. Zhang, and H. Liu. Effect of methane adsorption on the lifetime of a dodecahedral water cluster immersed in liquid water: A molecular dynamics study on the hydrate nucleation mechanisms. *The Journal of Physical Chemistry C*, 111(6):2595–2606, 2007.

[92] G.-J. Guo, Y.-G. Zhang, Y.-J. Zhao, K. Refson, and G.-H. Shan. Lifetimes of cagelike water clusters immersed in bulk liquid water: A molecular dynamics study on gas hydrate nucleation mechanisms. *Journal of Chemical Physics*, 121(3):1542, 2004.

[93] W. Han and K. Schulten. Fibril elongation by aβ17–42: Kinetic network analysis of hybrid-resolution molecular dynamics simulations. *Journal of the American Chemical Society*, 136(35):12450–12460, 2014.

[94] W. Han and Y.-D. Wu. A strand-loop-strand structure is a possible intermediate in fibril elongation: Long time simulations of amyloid-β peptide (10- 35). *Journal of the American Chemical Society*, 127(44):15408–15416, 2005.

[95] J. Hardy and D. J. Selkoe. The amyloid hypothesis of alzheimer's disease: progress and problems on the road to therapeutics. *Science*, 297(5580):353–356, 2002.

[96] J. D. Harper and P. T. Lansbury Jr. Models of amyloid seeding in alzheimer's disease and scrapie: mechanistic truths and physiological consequences of the time-dependent solubility of amyloid proteins. *Annual Review of Biochemistry*, 66(1):385–407, 1997.

[97] S. Hawley. Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry*, 10(13):2436–2442, 1971.

[98] R. W. Hawtin, D. Quigley, and P. M. Rodger. Gas hydrate nucleation and cage formation at a water/methane interface. *Physical Chemistry Chemical Physics*, 10(32):4853–4864, 2008.

[99] L. Hendriks, C. M. Van Duijn, P. Cras, M. Cruts, W. Van Hul, F. Van Harskamp, A. Warren, M. G. McInnis, S. E. Antonarakis, J.-J. Martin, et al. Presenile dementia and cerebral haemorrhage linked to a mutation at codon 692 of the β–amyloid precursor protein gene. *Nature Genetics*, 1(3):218, 1992.

[100] B. Hess, H. Bekker, H. J. Berendsen, and J. G. Fraaije. Lincs: a linear constraint solver for molecular simulations. *Journal of Computational Chemistry*, 18(12):1463–1472, 1997.

[101] B. Hess, C. Kutzner, D. van der Spoel, and E. Lindahl. Gromacs 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *Journal of Chemical Theory and Computation*, 4(3):435–447, 2008.

[102] W. Hosia, N. Bark, E. Liepinsh, A. Tjernberg, B. Persson, D. Hallén, J. Thyberg, J. Johansson, and L. Tjernberg. Folding into a β-hairpin can prevent amyloid fibril formation. *Biochemistry*, 43(16):4655–4661, 2004.

[103] W. Hoyer, C. Grönwall, A. Jonsson, S. Ståhl, and T. Härd. Stabilization of a β-hairpin in monomeric alzheimer's amyloid-β peptide inhibits amyloid formation. *Proceedings of the National Academy of Sciences*, 105(13):5099–5104, 2008.

[104] G. Hummer, S. Garde, A. E. Garcia, and M. E. Paulaitis. The pressure dependence of hydrophobic interactions is consistent with the observed pressure denaturation of proteins. *Proceedings of National Academy of Science*, 95:1552, 1998.

[105] W. Humphrey, A. Dalke, and K. Schulten. VMD – Visual Molecular Dynamics. *Journal of Molecular Graphics*, 14:33–38, 1996.

[106] T. Ikenoue, Y.-H. Lee, J. Kardos, M. Saiki, H. Yagi, Y. Kawata, and Y. Goto. Cold denaturation of alpha-synuclein amyloid fibrils. *Angewandte Chemie International Edition (Online Early view)*, 00:00, 2014.

[107] T. Ikenoue, Y.-H. Lee, J. Kardos, H. Yagi, T. Ikegami, H. Naiki, and Y. Goto. Heat of supersaturation-limited amyloid burst directly monitored by isothermal titration calorimetry. *Proceedings of the National Academy of Sciences*, 111(18):6654–6659, 2014.

[108] H. C. J. L. MacCallum, M. Sabaye Moghaddam and D. Tieleman. Hydrophobic association of α-helices, steric dewetting and enthalpic barriers to protein folding. *Proceedings of the National Academy of Sciences U.S.A.*, 104:6206, 2007.

[109] L. C. Jacobson, W. Hujo, and V. Molinero. Thermodynamic stability and growth of guest-free clathrate hydrates: a low-density crystal phase of water. *The Journal of Physical Chemistry B*, 113(30):10298–10307, 2009.

[110] L. C. Jacobson, W. Hujo, and V. Molinero. Amorphous precursors in the nucleation of clathrate hydrates. *Journal of the American Chemical Society*, 132(33):11806–11811, 2010.

[111] T. R. Jahn, O. S. Makin, K. L. Morris, K. E. Marshall, P. Tian, P. Sikorski, and L. C. Serpell. The common architecture of cross-β amyloid. *Journal of Molecular Biology*, 395(4):717 – 727, 2010.

[112] J. T. Jarrett, E. P. Berger, and P. T. Lansbury. The c-terminus of the β protein is critical in amyloidogenesisa. *Annals of the New York Academy of Sciences*, 695(1):144–148, 1993.

[113] G. A. Jeffrey and R. K. McMullan. *The Clathrate Hydrates*, pages 43–108. Hoboken, NJ: John Wiley and Sons, Ltd, 2007.

[114] F. Jimenez-Angeles and A. Firoozabadi. Nucleation of methane hydrates at moderate subcooling by molecular dynamics simulations. *The Journal of Physical Chemistry C*, 118:11310–11318, 2014.

[115] W. L. Jorgensen. Interactions between amides in solution and the thermodynamics of weak binding. *Journal of American Chemical Society*, 111(10):3770, 1989.

[116] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics*, 79(2):926–935, 1983.

[117] W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[118] I. S. Joung and T. E. Cheatham III. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *The Journal of Physical Chemistry B*, 112(30):9020–9041, 2008.

[119] D. Kaden, A. Harmeier, C. Weise, L. M. Munter, V. Althoff, B. R. Rost, P. W. Hildebrand, D. Schmitz, M. Schaefer, R. Lurz, et al. Novel app/aβ mutation k16n produces highly toxic heteromeric aβ oligomers. *EMBO Molecular Medicine*, 4(7):647–659, 2012.

[120] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen. Evaluation and reparametrization of the opls-aa force field for proteins via comparison with accurate quantum chemical calculations on peptides. *The Journal of Physical Chemistry B*, 105(28):6474–6487, 2001.

[121] D. Kashchiev, R. Cabriolu, and S. Auer. Confounding the paradigm: peculiarities of amyloid fibril nucleation. *Journal of the American Chemical Society*, 135(4):1531–1539, 2013.

[122] J. Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

[123] W. Kauzmann. Some factors in the interpretation of protein denaturation. *Advances in Protein Chemistry*, 14:1, 1959.

[124] W. Kauzmann. Some factors in the interpretation of protein denaturation. *Advances in Protein Chemistry*, 14:1–63, 1959.

[125] J. C. Kendrew, G. Bodo, H. M. Dintzis, T. Parrish, H. Wyckoff, and D. C. Phillips. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature*, 181:662, 1958.

[126] J. C. Kendrew, R. E. Dickerson, B. E. Strandberg, R. G. Hart, D. R. Davies, D. C. Phillips, and V. Shore. Structure of myoglobin: A three-dimensional fourier synthesis at 2 å. resolution. *Nature*, 185(4711):422, 1960.

[127] S. Keten, Z. Xu, B. Ihle, and M. J. Buehler. Nanoconfinement controls stiffness, strength and mechanical toughness of β-sheet crystals in silk. *Nature Materials*, 9:359, 2010.

[128] H.-Y. Kim, M.-K. Cho, D. Riedel, C. O. Fernandez, and M. Zweckstetter. Dissociation of amyloid fibrils of α-synuclein in supercooled water. *Angewandte Chemie International Edition*, 47(27):5046–5048, 2008.

[129] K. Klement, K. Wieligmann, J. Meinhardt, P. Hortschansky, W. Richter, and M. Fandrich. Effect of different salt ions on the propensity of aggregation and on the structure of alzheimers aβ (1-40) amyloid fibrils. *Journal of Molecular Biology*, 373(5):1321–1333, 2007.

[130] B. C. Knott, V. Molinero, M. F. Doherty, and B. Peters. Homogeneous nucleation of methane hydrates: Unrealistic under realistic conditions. *Journal of the American Chemical Society*, 134(48):19544–19547, 2012.

[131] T. P. Knowles, A. W. Fitzpatrick, S. Meehan, H. R. Mott, M. Vendruscolo, C. M. Dobson, and M. E. Welland. Role of intermolecular forces in defining material properties of protein nanofibrils. *Science*, 318(5858):1900–1903, 2007.

[132] C. A. Koh, A. K. Sum, and E. D. Sloan. Gas hydrates: Unlocking the energy from icy cages. *Journal of Applied Physics*, 106:061101–061114, 2009.

[133] M. G. Krone, L. Hua, P. Soto, R. Zhou, B. J. Berne, and J.-E. Shea. Role of water in mediating the assembly of alzheimer amyloid-β aβ16-22 protofilaments. *Journal of the American Chemical Society*, 130(33):11066–11072, 2008.

[134] J. Kühn, L. E. Wong, S. Pirkuliyeva, K. Schulz, C. Schwiegk, K. G. Fünfgeld, S. Keppler, F. D. Batista, H. Urlaub, M. Habeck, et al. The adaptor protein cin85 assembles intracellular signaling clusters for b cell activation. *Science Signaling*, 9(434):ra66, 2016.

[135] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. The weighted histogram analysis method for free-energy calculations on biomolecules. i. the method. *Journal of Computational Chemistry*, 13(8):1011–1021, 1992.

[136] J. E. Ladbury. Calorimetry as a tool for understanding biomolecular interactions and an aid to drug design. *Biochemical Society Transactions*, 38:888–893, 2010.

[137] M. Lauricella, G. Ciccotti, N. J. English, B. Peters, and S. Meloni. Mechanisms and nucleation rate of methane hydrate by dynamical nonequilibrium molecular dynamics. *The Journal of Physical Chemistry C*, 121(43):24223–24234, 2017.

[138] M. Lauricella, S. Meloni, S. Liang, N. J. English, P. G. Kusalik, and G. Ciccotti. Clathrate structure-type recognition: Application to hydrate nucleation and crystallisation. *The Journal of Chemical Physics*, 142(24):244503, 2015.

[139] N. R. Lee, C. J. Bowerman, and B. L. Nilsson. Effects of varied sequence pattern on the self-assembly of amphipathic peptides. *Biomacromolecules*, 14(9):3267–3277, 2013.

[140] N. R. Lee, C. J. Bowerman, and B. L. Nilsson. Sequence length determinants for self-assembly of amphipathic β-sheet peptides. *Peptide Science*, 100(6):738–750, 2013.

[141] J. A. Lemkul and D. R. Bevan. Assessing the stability of alzheimer's amyloid protofibrils using molecular dynamics. *The Journal of Physical Chemistry B*, 114(4):1652–1660, 2010.

[142] J. A. Lemkul and D. R. Bevan. Assessing the stability of alzheimer's amyloid protofibrils using molecular dynamics. *The Journal of Physical Chemistry B*, 114:1652, 2010.

[143] H. Li, C. Tang, and N. S. Wingreen. Nature of driving force for protein folding: A result from analyzing the statistical potential. *Physical Review Letters*, 79:765, 1997.

[144] Y. Liang, D. G. Lynn, and K. M. Berland. Direct observation of nucleation and growth in amyloid self-assembly. *Journal of the American Chemical Society*, 132(18):6306–6308, 2010.

[145] M. M. Lin, O. F. Mohammed, G. S. Jas, and A. H. Zewail. Speed limit of protein folding evidenced in secondary structure dynamics. *Proceedings of the National Academy of Sciences*, 108:16622–16627, 2011.

[146] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved side-chain torsion potentials for the amber ff99sb protein force field. *Proteins: Structure, Function, and Bioinformatics*, 78(8):1950–1958, 2010.

[147] K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror, and D. E. Shaw. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins: Structure, Function and Bioinformatics*, 78(8):1950–1958, jun 2010.

[148] W. Liu, E. Crocker, W. Zhang, J. I. Elliott, B. Luy, H. Li, S. Aimoto, and S. O. Smith. Structural role of glycine in amyloid fibrils formed from transmembrane α-helices. *Biochemistry*, 44(9):3591–3597, 2005.

[149] T. Luhrs, C. Ritter, M. Adrian, D. Riek-Loher, B. Bohrmann, H. Döbeli, D. Schubert, and R. Riek. 3d structure of alzheimer's amyloid-β(1–42) fibrils. *Proceedings of the National Academy of Sciences U.S.A.*, 102:17342, 2005.

[150] P. Luo and R. L. Baldwin. Interaction between water and polar groups of the helix backbone: An important determinant of helix propensities. *Proceedings of the National Academy of Sciences*, 96:4930–4935, 1999.

[151] N. Lupu-Lotan, A. Yaron, A. Berger, and M. Sela. Conformation changes in the nonionizable water-soluble synthetic polypeptide poly-n5-(3-hydroxypropyl)-l-glutamine. *Biopolymers: Original Research on Biomolecules*, 3(6):625–655, 1965.

[152] F. Mahmoudinobar and C. L. Dias. Grade: A code to determine clathrate hydrate structures. *Computer Physics Communications*, 244:385 – 391, 2019.

[153] F. Mahmoudinobar, C. L. Dias, and R. Zangi. Role of side-chain interactions on the formation of α-helices in model peptides. *Physical Review E*, 91(3):032710, 2015.

[154] F. Mahmoudinobar, J. Urban, Z. Su, B. L. Nilsson, and C. L. Dias. Thermodynamic stability of polar and non-polar amyloid fibrils. *Journal of Chemical Theory and Computation*, 2019.

[155] G. Makhatadze, S. Gill, and P. Privalov. Partial molar heat capacities of the side chains of some amino acid residues in aqueous solution: The influence of the neighboring charges. *Biophysical Chemistry*, 38(1-2):33–37, 1990.

[156] G. I. Makhatadze. Thermodynamics of α-Helix Formation. *Advances in Protein Chemistry*, 72:199–226, 2005.

[157] O. S. Makin, E. Atkins, P. Sikorski, J. Johansson, and L. C. Serpell. Molecular basis for amyloid fibril formation and stability. *Proceedings of the National Academy of Sciences of the United States of America*, 102(2):315–320, 2005.

[158] S. Marchesan, A. V. Vargiu, and K. E. Styan. The phe-phe motif for peptide self-assembly in nanomedicine. *Molecules*, 20(11):19775–19788, 2015.

[159] O. Markovitch and N. Agmon. Structure and energetics of the hydronium hydration shells. *The Journal of Physical Chemistry A*, 111(12):2253–2256, 2007.

[160] M. I. Marqués, J. M. Borreguero, H. E. Stanley, and N. V. Dokholyan. Possible mechanism for cold denaturation of proteins at high pressure. *Physical Review Letters*, 91(13):138103, 2003.

[161] S. Marqusee, V. H. Robbins, and R. L. Baldwin. Unusually stable helix formation in short alanine-based peptides. *Proceedings of the National Academy of Sciences U.S.A.*, 86:5286, 1989.

[162] K. E. Marshall, K. L. Morris, D. Charlton, N. O'Reilly, L. Lewis, H. Walden, and L. C. Serpell. Hydrophobic, aromatic, and electrostatic interactions play a central role in amyloid fibril formation and stability. *Biochemistry*, 50(12):2061–2071, 2011.

[163] M. Matsumoto, A. Baba, and I. Ohmine. Topological building blocks of hydrogen bond network in water. *Journal of Chemical Physics*, 127(13):134504–134509, 2007.

[164] M. P. Mattson. Pathways towards and away from alzheimer's disease. *Nature*, 430(7000):631, 2004.

[165] N. A. McDonald and W. L. Jorgensen. Development of an all-atom force field for heterocycles. properties of liquid pyrrole, furan, diazoles, and oxazoles. *The Journal of Physical Chemistry B*, 102(41):8049–8059, 1998.

[166] J. McLaurin, T. Franklin, A. Chakrabartty, and P. Fraser. Phosphatidylinositol and inositol involvement in alzheimer amyloid-β fibril growth and arrest. *Journal of Molecular Biology*, 278(1):183–194, 1998.

[167] J. McLaurin, M. E. Kierstead, M. E. Brown, C. A. Hawkes, M. H. Lambermon, A. L. Phinney, A. A. Darabie, J. E. Cousins, J. E. French, M. F. Lan, et al. Cyclohexanehexol inhibitors of aβ aggregation prevent and reverse alzheimer phenotype in a mouse model. *Nature Medicine*, 12(7):801, 2006.

[168] R. Mishra and R. Winter. Cold- and pressure-induced dissociation of protein aggregation and amyloid fibrils. *Angewandte Chemie International Edition*, 47:6518, 2008.

[169] S. Miyamoto and P. A. Kollman. Settle: An analytical version of the shake and rattle algorithm for rigid water models. *Journal of Computational Chemistry*, 13(8):952–962, 1992.

[170] D. L. Mobley, C. I. Bayly, M. D. Cooper, and K. A. Dill. Predictions of hydration free energies from all-atom molecular dynamics simulations. *The Journal of Physical Chemistry B*, 113(14):4533–4537, 2009.

[171] B. Morel, L. Varela, and F. Conejero-Lara. The thermodynamic stability of amyloid fibrils studied by differential scanning calorimetry. *The Journal of Physical Chemistry B*, 114(11):4010–4019, 2010.

[172] V. Muñoz and L. Serrano. Elucidating the folding problem of helical peptides using empirical parameters. *Nature Structural & Molecular Biology*, 1(6):399–409, 1994.

[173] F. Musiani and A. Giorgetti. Protein aggregation and molecular crowding: Perspectives from multiscale simulations. *International Review of Cell and Molecular Biology*, 329:49–77, 2017.

[174] C. Narayanan and C. L. Dias. Hydrophobic interactions and hydrogen bonds in β-sheet formation. *Journal of Chemical Physics*, 139:115103, 2013.

[175] R. Nelson, M. R. Sawaya, M. Balbirnie, A. O. Madsen, C. Riekel, R. Grothe, and D. Eisenberg. Structure of the cross-β spine of amyloid-like fibrils. *Nature*, 435(7043):773, June 2005.

[176] G. Némethy. Hydrophobic interactions. *Angewandte Chemie International Edition*, 6(3):195–206, 1967.

[177] G. Némethy and H. A. Scheraga. The structure of water and hydrophobic bonding in proteins. iii. the thermodynamic properties of hydrophobic bonds in proteins. *The Journal of Physical Chemistry*, 66(10):1773–1789, 1962.

[178] A. H. Nguyen and V. Molinero. Identification of clathrate hydrates, hexagonal ice, cubic ice, and liquid water in simulations: The chill+ algorithm. *The Journal of Physical Chemistry B*, 119(29):9369–9376, 2014.

[179] H. D. Nguyen and C. K. Hall. Phase diagrams describing fibrillization by polyalanine peptides. *Biophysical Journal*, 87(6):4122, 2004.

[180] H. D. Nguyen, A. J. Marchut, and C. K. Hall. Solvent effects on the conformational transition of a model polyalanine peptide. *Protein Science*, 13:2909, 2004.

[181] P. H. Nguyen, M. S. Li, G. Stock, J. E. Straub, and D. Thirumalai. Monomer adds to preformed structured oligomers of aβ-peptides by a two-stage dock-lock mechanism. *Proceedings of the National Academy of Sciences U.S.A.*, 104:111, 2007.

[182] C. Nilsberth, A. Westlind-Danielsson, C. B. Eckman, M. M. Condron, K. Axelman, C. Forsell, C. Stenh, J. Luthman, D. B. Teplow, S. G. Younkin, et al. The'arctic'app mutation (e693g) causes alzheimer's disease by enhanced aβ protofibril formation. *Nature Neuroscience*, 4(9):887, 2001.

[183] L. Obici, A. Demarchi, G. De Rosa, V. Bellotti, S. Marciano, S. Donadei, E. Arbustini, G. Palladini, M. Diegoli, E. Genovese, et al. A novel aβpp mutation exclusively associated with cerebral amyloid angiopathy. *Annals of Neurology: Official Journal of the American Neurological Association and the Child Neurology Society*, 58(4):639–644, 2005.

[184] E. P. O'Brien, Y. Okamoto, J. E. Straub, B. R. Brooks, and D. Thirumalai. Thermodynamic perspective on the dock-lock growth mechanism of amyloid fibrils. *The Journal of Physical Chemistry B*, 113(43):14421, 2009.

[185] K. T. O'Neil and W. F. DeGrado. A thermodynamic scale for the helix-forming tendencies of the commonly occurring amino acids. *Science*, 250(4981):646–651, 1990.

[186] B. O'Nuallain, S. Shivaprasad, I. Kheterpal, and R. Wetzel. Thermodynamics of aβ (1- 40) amyloid fibril elongation. *Biochemistry*, 44(38):12709–12718, 2005.

[187] J. N. Onuchic, Z. Luthey-Schulten, and P. G. Wolynes. Theory of protein folding: the energy landscape perspective. *Annual Review of Physical Chemistry*, 48(1):545–600, 1997.

[188] S. Padmanabhan, S. Marqusee, T. Ridgeway, T. M. Laue, and R. L. Baldwin. Relative helix-forming tendencies of nonpolar amino acids. *Nature*, 344:268–270, 1990.

[189] D. Paschek. Temperature dependence of the hydrophobic hydration and interaction of simple solutes: An examination of five popular water models. *Journal of Chemical Physics*, 120:6674, 2004.

[190] B. A. Patel, P. G. Debenedetti, F. H. Stillinger, and P. J. Rossky. A water-explicit lattice model of heat-, cold-, and pressure-induced protein unfolding. *Biophysical Journal*, 93(12):4116–4127, 2007.

[191] A. Patriksson and D. van der Spoel. A temperature predictor for parallel tempering simulations. *Physical Chemistry Chemical Physics*, 10(15):2073–2077, 2008.

[192] L. Pauling and R. B. Corey. Two hydrogen-bonded spiral configurations of the polypeptide chain. *Journal of the American Chemical Society*, 72(11):5349–5349, 1950.

[193] L. Pauling and R. B. Corey. The pleated sheet, a new layer configuration of polypeptide chains. *Proceedings of the National Academy of Sciences USA*, 37(5):251–256, 1951.

[194] L. Pauling and R. B. Corey. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proceedings of the National Academy of Sciences U.S.A.*, 37:205, 1951.

[195] A. T. Petkova, Y. Ishii, J. J. Balbach, O. N. Antzutkin, R. D. Leapman, F. Delaglio, and R. Tycko. A structural model for alzheimer's β-amyloid fibrils based on experimental constraints from solid state nmr. *Proceedings of the National Academy of Sciences U.S.A.*, 99(26):16742–16747, 2002.

[196] S. Piana, J. L. Klepeis, and D. E. Shaw. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Current Opinion In Structural Biology*, 24:98–105, 2014.

[197] L. R. Pratt and D. Chandler. Theory of the hydrophobic effect. *Journal of Chemical Physics*, 67:3683, 1977.

[198] P. Privalov, Y. V. Griko, S. Y. Venyaminov, and V. Kutyshenko. Cold denaturation of myoglobin. *Journal of Molecular Biology*, 190(3):487–498, 1986.

[199] P. L. Privalov. Cold denaturation of protein. *Critical Reviews in Biochemistry and Molecular Biology*, 25(4):281, Jan. 1990.

[200] P. L. Privalov. Thermodynamics of protein folding. *Journal of Chemical Thermodynamics*, 29:447–474, 1997.

[201] R. Radhakrishnan and B. L. Trout. A new approach for studying nucleation phenomena using molecular simulations: Application to co2 hydrate clathrates. *Journal of Chemical Physics*, 117(4):1786–1796, 2002.

[202] J. A. Rank and D. Baker. A desolvation barrier to hydrophobic cluster formation may contribute to the rate-limiting step in protein folding. *Protein Science*, 6(2):347, 1997.

[203] S. Rao Jampani, F. Mahmoudinobar, Z. Su, and C. L. Dias. Thermodynamics of aβ16–21 dissociation from a fibril: Enthalpy, entropy, and volumetric properties. *Proteins: Structure, Function, and Bioinformatics*, 83(11):1963–1972, 2015.

[204] N. Rathore, M. Chopra, and J. J. de Pablo. Optimal allocation of replicas in parallel tempering simulations. *The Journal of Chemical Physics*, 122(2):024111, 2005.

[205] M. Reches and E. Gazit. Casting metal nanowires within discrete self-assembled peptide nanotubes. *Science*, 300(5619):625–627, 2003.

[206] G. Reddy, J. E. Straub, and D. Thirumalai. Dynamics of locking of peptides onto growing amyloid fibrils. *Proceedings of the National Academy of Sciences U.S.A.*, 106:11948, 2009.

[207] G. Reddy, J. E. Straub, and D. Thirumalai. Influence of preformed asp23- lys28 salt bridge on the conformational fluctuations of monomers and dimers of aβ peptides with implications for rates of fibril formation. *The Journal of Physical Chemistry B*, 113(4):1162–1172, 2009.

[208] G. Reddy, J. E. Straub, and D. Thirumalai. Dry amyloid fibril assembly in a yeast prion peptide is mediated by long-lived structures containing water wires. *Proceedings of the National Academy of Sciences*, 107(50):21459–21464, 2010.

[209] P. Ricchiuto, A. V. Brukhno, and S. Auer. Protein aggregation: Kinetics versus thermodynamics. *The Journal of Physical Chemistry B*, 116(18):5384–5390, 2012.

[210] J. M. Richardson, M. M. Lopez, and G. I. Makhatadze. Enthalpy of helix-coil transition: Missing link in rationalizing the thermodynamics of helix-forming propensities of the amino acid residues. *Proceedings of the National Academy of Sciences*, 102:1413–1418, 2005.

[211] L. Rizzi and S. Auer. Amyloid fibril solubility. *The Journal of Physical Chemistry B*, 119(46):14631–14636, 2015.

[212] R. C. Rizzo and W. L. Jorgensen. Opls all-atom model for amines: resolution of the amine hydration problem. *Journal of the American Chemical Society*, 121(20):4827–4836, 1999.

[213] A. D. Robertson and K. P. Murphy. Protein structure and the energetics of protein stability. *Chemical Review*, 97:1251, 1997.

[214] P. Rodger, T. Forester, and W. Smith. Simulations of the methane hydrate/methane gas interface near hydrate forming conditions conditions. *Fluid phase equilibria*, 116(1-2):326–332, 1996.

[215] D. J. Rosenman, C. R. Connors, W. Chen, C. Wang, and A. E. García. Aβ monomers transiently sample oligomer and fibril-like configurations: Ensemble characterization using a combined md/nmr approach. *Journal of Molecular Biology*, 425(18):3338–3359, 2013.

[216] D. J. Rosenman, C. Wang, and A. E. García. Characterization of Aβ Monomers through the Convergence of Ensemble Properties among Simulations with Multiple Force Fields. *The Journal of Physical Chemistry B*, 120(2):259–277, 2016.

[217] G. Rossi, G. Giaccone, R. Maletta, M. Morbin, R. Capobianco, M. Mangieri, A. Giovagnoli, A. Bizzi, C. Tomaino, M. Perri, et al. A family with alzheimer disease and strokes associated with a713t mutation of the app gene. *Neurology*, 63(5):910–912, 2004.

[218] G. S. Rule and T. K. Hitchens. *Fundamentals of protein NMR spectroscopy*, volume 5. Springer Science & Business Media, 2006.

[219] C. D. Ruppel. The us geological survey's gas hydrates project: Us geological survey fact sheet 2017-3079, 4p. Technical report, US Geological Survey, 2018.

[220] P. J. Russell. *iGenetics: A Molecular Approach (3rd Edition)*. Pearson, 2009.

[221] A. Šarić, Y. C. Chebaro, T. P. Knowles, and D. Frenkel. Crucial role of nonspecific interactions in amyloid nucleation. *Proceedings of the National Academy of Sciences*, 111(50):17869–17874, 2014.

[222] S. Sarupria and P. G. Debenedetti. Homogeneous nucleation of methane hydrate in microsecond molecular dynamics simulations. *The Journal of Physical Chemistry Letters*, 3(20):2942–2947, 2012.

[223] M. R. Sawaya, S. Sambashivan, R. Nelson, M. I. Ivanova, S. A. Sievers, M. I. Apostol, M. J. Thompson, M. Balbirnie, J. J. Wiltzius, H. T. McFarlane, et al. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature*, 447(7143):453, 2007.

[224] J. A. Schellman. The stability of hydrogen-bonded peptide structures in aqueous solution. *Compt. rend. trav. lab. Carlsberg. Sér. chim.*, 29:230, 1955.

[225] J. P. Schmittschmitt and J. M. Scholtz. The role of protein stability, solubility, and net charge in amyloid fibril formation. *Protein Science*, 12(10):2374–2378, 2003.

[226] M. Schor, J. Vreede, and P. G. Bolhuis. Elucidating the locking mechanism of peptides onto growing amyloid fibrils through transition path sampling. *Biophysical journal*, 103(6):1296–1304, 2012.

[227] N. Schwierz, C. V. Frost, P. L. Geissler, and M. Zacharias. Dynamics of seeded aβ40-fibril growth from atomistic molecular dynamics simulations: Kinetic trapping and reduced water mobility in the locking step. *Journal of the American Chemical Society*, 138(2):527–539, 2016.

[228] F. Sedlmeier and R. R. Netz. Solvation thermodynamics and heat capacity of polar and charged solutes in water. *The Journal of Chemical Physics*, 138(11):03B609, 2013.

[229] M. M. Seibert, A. Patriksson, B. Hess, and D. Van Der Spoel. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *Journal of Molecular Biology*, 354(1):173–183, 2005.

[230] D. J. Selkoe. Alzheimer's disease: genes, proteins, and therapy. *Physiological Reviews*, 81(2):741–766, 2001.

[231] D. J. Selkoe and D. Schenk. Alzheimer's disease: molecular understanding predicts amyloid-based therapeutics. *Annual Review of Pharmacology and Toxicology*, 43(1):545–584, 2003.

[232] F. T. Senguen, T. M. Doran, E. A. Anderson, and B. L. Nilsson. Clarifying the influence of core amino acid hydrophobicity, secondary structure propensity, and molecular volume on amyloid-β 16–22 self-assembly. *Molecular BioSystems*, 7(2):497–510, 2011.

[233] F. T. Senguen, N. R. Lee, X. Gu, D. M. Ryan, T. M. Doran, E. A. Anderson, and B. L. Nilsson. Probing aromatic, hydrophobic, and steric effects on the self-assembly of an amyloid-β fragment peptide. *Molecular BioSystems*, 7(2):486–496, 2011.

[234] N. G. Sgourakis, M. Merced-Serrano, C. Boutsidis, P. Drineas, Z. Du, C. Wang, and A. E. Garcia. Atomic-level characterization of the ensemble of the aβ (1–42) monomer in water using unbiased molecular dynamics simulations and spectral algorithms. *Journal of Molecular Biology*, 405(2):570–583, 2011.

[235] K. Sharp. Entropy-enthalpy compensation: Fact or artifact? *Protein Science*, 10(3):661–667, 2001.

[236] S. Shimizu and H. S. Chan. Temperature dependence of hydrophobic interactions: A mean force perspective, effects of water density, and nonadditivity of thermodynamic signatures. *The Journal of Chemical Physics*, 113:4683, 2000.

[237] S. Shimizu and H. S. Chan. Anti-cooperativity in hydrophobic interactions: A simulation study of spatial dependence of three-body effects and beyond. *Journal of Chemical Physics*, 115:1414, 2001.

[238] M. R. Shirts, C. Klein, J. M. Swails, J. Yin, M. K. Gilson, D. L. Mobley, D. A. Case, and E. D. Zhong. Lessons learned from comparing molecular dynamics engines on the sampl5 dataset. *Journal of Computer-aided Molecular Design*, pages 1–15, 2016.

[239] K. A. T. Silverstein, A. D. J. Haymet, and K. A. Dill. A simple model of water and the hydrophobic effect. *Journal of American Chemical Society*, 120 (13):3166–3175, 1998.

[240] S. Sinha, Z. Du, P. Maiti, F.-G. Klarner, T. Schrader, C. Wang, and G. Bitan. Comparison of three amyloid assembly inhibitors: the sugar scyllo-inositol, the polyphenol epigallocatechin gallate, and the molecular tweezer clr01. *ACS Chemical Neuroscience*, 3(6):451–458, 2012.

[241] N. Sitaram and R. Nagaraj. Host-defense antimicrobial peptides: importance of structure for activity. *Current Pharmaceutical Design*, 8(9):727–742, 2002.

[242] E. Sloan and C. Koh. *Clathrate hydrates of natural gases*. CRC press/Taylor & Francis, 2007.

[243] E. D. Sloan. Fundamental principles and applications of natural gas hydrates. *Nature*, 426:353–363, 2003.

[244] E. D. Sloan. A changing hydrate paradigm–from apprehension to avoidance to risk management. *Fluid Phase Equilibria*, 228:67–74, 2005.

[245] L. Smeller. Pressure-temperature phase diagrams of biomolecules. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*, 1595:11–29, 2002.

[246] P. Soto and R. Zangi. Characterization of the stereochemical selectivity of β-hairpin formation by molecular dynamics simulations. *The Journal of Physical Chemistry B*, 109(3):1281–1288, 2005.

[247] Z. Su and C. L. Dias. Driving β-strands into fibrils. *The Journal of Physical Chemistry B*, 118:10830–10836, 2014.

[248] Z. Su, F. Mahmoudinobar, and C. L. Dias. Effects of trimethylamine-n-oxide on the conformation of peptides and its implications for proteins. *Physical Review Letters*, 119(10):108102, 2017.

[249] Y. Sugita and Y. Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

[250] A. K. Sum, C. A. Koh, and E. D. Sloan. Clathrate hydrates: from laboratory science to engineering practice. *Industrial & Engineering Chemistry Research*, 48(16):7457–7465, 2009.

[251] T. Sun, F.-H. Lin, R. L. Campbell, J. S. Allingham, and P. L. Davies. An antifreeze protein folds with an interior network of more than 400 semi-clathrate waters. *Science*, 343(6172):795–798, 2014.

[252] T. Takeda and D. K. Klimov. Dissociation of aβ_{16-22} amyloid fibrils probed by molecular dynamics. *Journal of Molecular Biology*, 368(4):1202, 2007.

[253] T. Takeda and D. K. Klimov. Replica exchange simulations of the thermodynamics of aβ fibril growth. *Biophysical Journal*, 96(2):442, 2009.

[254] C. Tanford. Protein Denaturation. *Advances in Protein Chemistry*, 23:121–282, 1968.

[255] B. Tarus, J. E. Straub, and D. Thirumalai. Dynamics of asp23-lys28 salt-bridge formation in aβ_{10-35} monomers. *Journal of the American Chemical Society*, 128(50):16159, 2006.

[256] D. Thirumalai, G. Reddy, and J. E. Straub. Role of water in protein aggregation and amyloid polymorphism. *Accounts of Chemical Research*, 45(1):83, 2012.

[257] M. J. Thompson, S. A. Sievers, J. Karanicolas, M. I. Ivanova, D. Baker, and D. Eisenberg. The 3d profile method for identifying fibril-forming segments of proteins. *Proceedings of the National Academy of Sciences*, 103(11):4074–4078, 2006.

[258] Y. F. Tian, J. M. Devgun, and J. H. Collier. Fibrillized peptide microgels for cell encapsulation and 3d cell culture. *Soft Matter*, 7(13):6005–6011, 2011.

[259] L. Tjernberg, W. Hosia, N. Bark, J. Thyberg, and J. Johansson. Charge attraction and β propensity are necessary for amyloid fibril formation from tetrapeptides. *Journal of Biological Chemistry*, 277(45):43243–43246, 2002.

[260] M. Török, S. Milton, R. Kayed, P. Wu, T. McIntire, C. G. Glabe, and R. Langen. Structural and dynamic features of alzheimer's aβ peptide in amyloid fibrils studied by site-directed spin labeling. *Journal of Biological Chemistry*, 277(43):40810–40815, 2002.

[261] G. M. Torrie and J. P. Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

[262] T. Urbic, S. Najem, and C. L. Dias. Thermodynamic properties of amyloid fibrils in equilibrium. *Biophysical Chemistry*, pages –, 2017.

[263] V. N. Uversky, J. Li, and A. L. Fink. Evidence for a partially folded intermediate in α-synuclein fibril formation. *Journal of Biological Chemistry*, 276:10737, 2001.

[264] G. Vriend. What if: a molecular modeling and drug design program. *Journal of Molecular Graphics*, 8(1):52–56, 1990.

[265] A. Vsaric, Y. C. Chebaro, T. P. Knowles, and D. Frenkel. Crucial role of nonspecific interactions in amyloid nucleation. *Proceedings of the National Academy of Sciences*, 111(50):17869–17874, 2014.

[266] G. W. Vuister and A. Bax. Quantitative j correlation: a new approach for measuring homonuclear three-bond j (hnh. alpha.) coupling constants in 15n-enriched proteins. *Journal of the American Chemical Society*, 115(17):7772–7777, 1993.

[267] M. R. Walsh, G. T. Beckham, C. A. Koh, E. D. Sloan, D. T. Wu, and A. K. Sum. Methane hydrate nucleation rates from molecular dynamics simulations: Effects of aqueous methane concentration, interfacial curvature, and system size. *The Journal of Physical Chemistry C*, 115:21241–21248, 2011.

[268] M. R. Walsh, C. A. Koh, E. D. Sloan, A. K. Sum, and D. T. Wu. Microsecond simulations of spontaneous methane hydrate nucleation and growth. *Science*, 326(5956):1095–1098, 2009.

[269] J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, and D. A. Case. Development and testing of a general amber force field. *Journal of Computational Chemistry*, 25(9):1157–1174, 2004.

[270] S. Wärmländer, A. Tiiman, A. Abelein, J. Luo, J. Jarvet, K. L. Söderberg, J. Danielsson, and A. Gräslund. Biophysical studies of the amyloid β-peptide: Interactions with metal ions and small molecules. *ChemBioChem*, 14(14):1692–1704, 2013.

[271] R. Wetzel. Kinetics and thermodynamics of amyloid fibril assembly. *Accounts of Chemical Research*, 39(9):671–679, 2006.

[272] A. D. Williams, S. Shivaprasad, and R. Wetzel. Alanine scanning mutagenesis of aβ(1-40) amyloid fibril stability. *Journal of Molecular Biology*, 357(4):1283 – 1294, 2006.

[273] J. J. Wiltzius, S. A. Sievers, M. R. Sawaya, D. Cascio, D. Popov, C. Riekel, and D. Eisenberg. Atomic structure of the cross-β spine of islet amyloid polypeptide (amylin). *Protein Science*, 17(9):1467–1474, 2008.

[274] www.alz.org. *2016 Alzheimer's Disease Facts and Figurs*. Alzheimer's and Dimentia 2016:12(4), 2016.

[275] T. Yagasaki, M. Matsumoto, and H. Tanaka. Effects of thermodynamic inhibitors on the dissociation of methane hydrate: a molecular dynamics study. *Physical Chemistry Chemical Physics*, 17(48):32347–32357, 2015.

[276] D. Yuhara, B. C. Barnes, D. Suh, B. C. Knott, G. T. Beckham, K. Yasuoka, D. T. Wu, and A. K. Sum. Nucleation rate analysis of methane hydrate from molecular dynamics simulations. *Faraday discussions*, 179:463–474, 2015.

[277] R. Zangi. Driving force for hydrophobic interaction at different length scales. *The Journal of Physical Chemistry B*, 115(10):2303–2311, 2011.

[278] R. Zangi. Side-chain–side-chain interactions and stability of the helical state. *Physical Review E*, 89:012723, 2014.

[279] L. Zhou, N. Brouwers, I. Benilova, A. Vandersteen, M. Mercken, K. Van Laere, P. Van Damme, D. Demedts, F. Van Leuven, K. Sleegers, et al. Amyloid precursor protein mutation e682k at the alternative β-secretase cleavage β-site increases aβ generation. *EMBO Molecular Medicine*, 3(5):291–302, 2011.

[280] A. Zipp and W. Kauzmann. Pressure denaturation of metmyoglobin. *Biochemistry*, 12(21):4217–4228, 1973.

[281] Y. Zou, Z. Qian, Y. Chen, H. Qian, G. Wei, and Q. Zhang. Norepinephrine inhibits alzheimers amyloid-β peptide aggregation and destabilizes amyloid-β protofibrils: A molecular dynamics simulation study. *ACS Chemical Neuroscience*, 10(3):1585–1594, 2019.