Scuola Internazionale Superiore di Studi Avanzati

Doctoral Thesis

Atomistic Simulations of Model Amyloid Beta Aggregates, Water Networks and their Optical Properties

---

Author:
Kwang Hyok Jong

Supervisors:
Professor Ali A. Hassanali
Professor Giuseppe Legname

A thesis submitted in fulfillment of the requirements
for the degree of Doctor of Philosophy

in the

Laboratory of Prion Biology
Neuroscience

SCUOLA INTERNAZIONALE SUPERIORE DI STUDI AVANZATI

Abstract

Functional and Structural Genomics
Neuroscience

Doctor of Philosophy

Atomistic Simulations of Model Amyloid Beta Aggregates, Water Networks and their Optical Properties

by Kwang Hyok Jong

Based on the amyloid hypothesis, amyloid oligomers and the fibrils that they aggregate into, have been implicated in neurodegenerative diseases. Most of these amyloid proteins live in a solvent environment. The role of solvent in modulating the structural and dynamical properties of amyloid proteins remains poorly understood. In this thesis, computer simulations are used to reveal the structural properties of the amyloid protein and the coupling between protein and water using model systems.

After assessing the validity of the force fields by comparison with high-level quantum chemistry calculations, we examine further the conformational free energy landscape of an amyloid protein. Different conformations characterized in the free energy surface are driven by internal protein interactions as well as interactions between protein and water, resulting in the collective reorganization of protein and water hydrogen bond networks. We show that these proteins are surrounded by water wires that add a roughness to the free energy surface.

To better understand the water hydrogen bond network and particularly the water wires around protein, we used data-science algorithms allowing for the dimensionality and free energy landscape of different water coordinates to be determined. These results confirm that using water wire coordinates encodes more information on the underlying secondary structure of the protein.

Finally, ab initio calculations are used to investigate the optical properties of amyloid proteins to help rationalize recent experiments suggesting the intrinsic fluorescence in fibrils that can occur without aromatic residues.

Acknowledgements

First of all, I would like to express my deepest gratitude to my Diploma and PhD supervisor Professor Ali A. Hassanali for his full support, expert guidance, and generosity throughout my research and work. It was my great honor and a real privilege to share to his exceptional knowledge, crazy passion for science, and extraordinary humanity. I am also extremely thankful to my co-supervisor Professor Giuseppe Legname for his constant support and constructive suggestions.

I extend my special thanks to Professor Alessandro Laio and Professor Luca Grisanti for their generous help and great advice. I am also very thankful to Dr. Yavar A. Azar, Dr. Narjes Ansari and Dr. Emiliano Polli for their great help and collaboration.

I would like to express my sincere thanks to Professor JinU Kang for giving me an opportunity to start my scientific life at ICTP and SISSA. I am very grateful to Professor Fernando Quevedo, Professor Stefano Ruffo, Professor Sandro Scandolo, Professor Ralph Gebauer, Professor Nicola Seriani for their strong support and help. I am very grateful to Professor CholSu Kim, Professor HakChol Pak, Professor CholSun Kim for their continuous help and mental support.

I would like to thank UiRi Mun, CholJun Kang, OkSong An for their support and cooperation. I also thank my friends Kaaya Ismail, Erick Buko, Nawaz Qaisrani, Giulia Sormani, Francesca Cuturello, Andrea Papale, and Lucia Coronel for their continuous encouragement and friendship.

Finally, I'd like to acknowledge ICTP and SISSA for the financial and computational support.

Contents

List of Figures

List of Tables

Dedicated to my beloved family

Chapter 1

Introduction

1.1 Brief introduction of Alzheimer Disease and Amyloid Hypothesis

Alzheimer disease (AD) is a neurodegenerative disease that slowly starts and progressively destroys the brain cells [Burns and Iliffe, 2009]. AD is known to be one of the dominant causes of around 60 % of dementia. The most common symptoms are short-term memory loss, language problems, and behavioral issues. This disease was first described by a German psychiatrist and pathologist, Dr. Alois Alzheimer. He carried out an autopsy on the brain of a patient who had died after several years of a progressive mental disorder characterized by confusion and memory loss. In the patient's cortex responsible for memory, he found the strange bundles of nerves, which he termed neurofibrillary tangles and amyloid plaques. The neurofibrillary tangles and amyloid plaques in AD patient's brain are shown in the right panel of Figure 1.1, compared to the normal brain in the left panel. In late 1906, he presented his findings, speculating that these abnormal tangles and plagues would be responsible for the patient's mental disorder.

For more than a century after the first description of AD, researchers have tried to understand the cause of this disease and search for drugs to cure it. Despite decades of research, the cause of AD is still unknown except for a few cases with identified abnormal genes. There exist several competing hypotheses trying to explain the cause of AD including cholinergic hypothesis, amyloid hypothesis, tau hypothesis and so on. The amyloid hypothesis, for example, has dominated research on AD for last decades after its postulation in 1991 [Goate et al., 1991] in which amyloid plaques outside the neuron cell are the fundamental cause of AD. The establishment of this assumption is firstly inspired by the success of isolation of amyloid β from the amyloid precursor protein (APP) in the cell membrane and subsequent studies on its procedure, revealing that β-secretase and γ-secretase cleave APP to create amyloid peptide [Allsop et al., 1988]. The amyloid hypothesis has taken off by the discovery of the dominantly inherited mutations which are responsible for the familial forms of AD [Goate et al., 1991; Scheuner et al., 1996]. The individual amyloid β proteins are accumulated into larger structures called "oligomers" and further accumulation produces insoluble fibrils, which then aggregate into the plaques. An illustration of

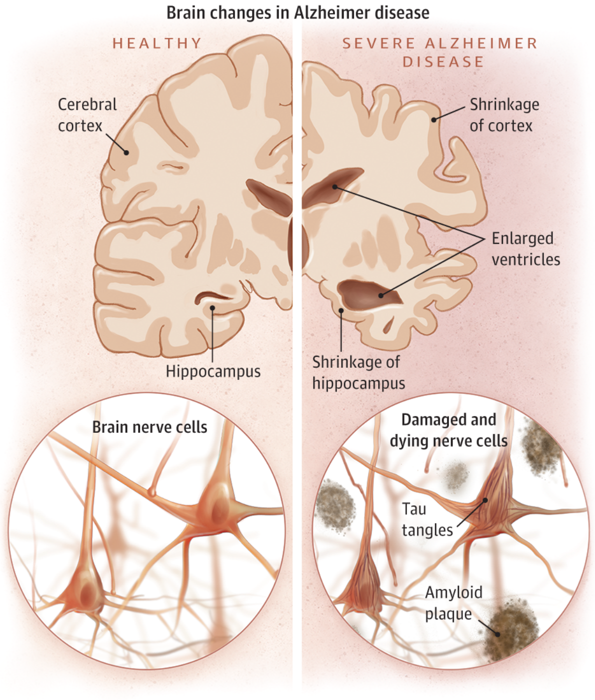
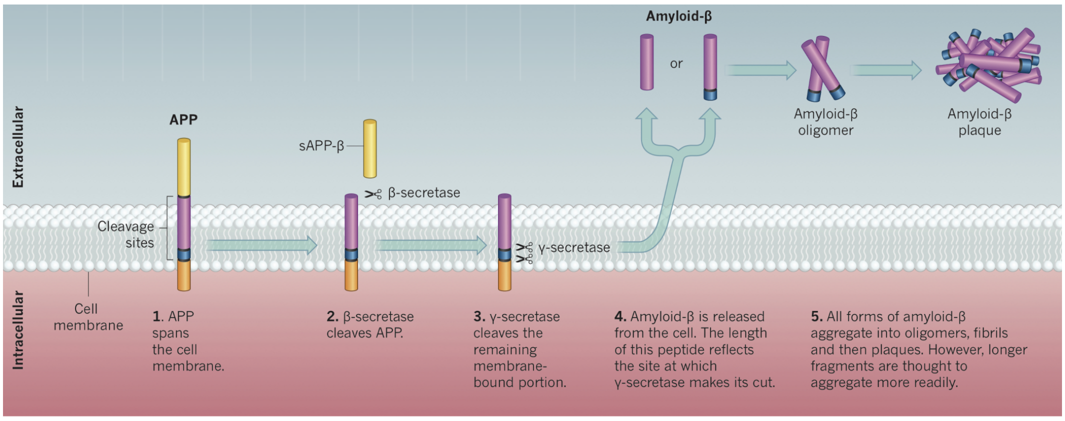


Figure 1.1: Comparison of AD patient's brain with the neurofibrillary tangles and the amyloid plaques to the normal brain. This figure is from [Jin, 2015].

the mechanism from APP to amyloid plaques is shown in Figure 1.2. In the amyloid hypothesis, the rest of AD procedure such as the formation of tau-tangle protein and synapse dysfunction is considered to be triggered by the amyloid β aggregation [Hardy and Selkoe, 2002]. A recent change in amyloid hypothesis [Selkoe and Hardy, 2016], considering soluble small oligomers as one of the dominant causes of AD, directs the studies on both fibrils and oligomers which will be introduced in the following section.



1. APP spans the cell membrane.

2. β-secretase cleaves APP.

3. γ-secretase cleaves the remaining membrane-bound portion.

4. Amyloid-β is released from the cell. The length of this peptide reflects the site at which γ-secretase makes its cut.

5. All forms of amyloid-β aggregate into oligomers, fibrils and then plaques. However, longer fragments are thought to aggregate more readily.

Figure 1.2: Diagram of amyloid β production. This figure is reproduced from [Makin, 2018]

1.2 Structure of Amyloid Fibrils and Oligomers

1.2.1 Amyloid Fibrils

As we discussed earlier in Section 1.1, in the last several decades the conversion of amyloid β proteins from their soluble forms into the insoluble fibrils has been discovered as a key reason for AD [Hardy and Selkoe, 2002; Aguzzi, Sigurdson, and Heikenwaelder, 2008; Collinge, 2001; Hebda and Miranker, 2009]. Therefore, there have been a lot of experimental and theoretical efforts to understand the physical and chemical processes underlying its aggregation [Nasica Labouze et al., 2015; Paravastu et al., 2008; Kirkitadze, Condron, and Teplow, 2001; Hou et al., 2004; Zhang et al., 2000; Baumketner and Shea, 2007; Petkova et al., 2002; Khemtémourian et al., 2011; Xu et al., 2013; Shanmugam and Polavarapu, 2004; Wei and Shea, 2006; Rosenman et al., 2013; Bitan et al., 2003; Riccardi, Nguyen, and Stock, 2012; Carballo Pacheco and Strodel, 2016; Makin et al., 2005; Pellarin and Caflisch, 2006; Fawzi et al., 2007; Auer, Dobson, and Vendruscolo, 2007; Baftizadeh et al., 2012; Barducci et al., 2013; Sanfelice et al., 2014; Qiang et al., 2012; Lührs et al., 2005].

With the development of experimental techniques in molecular biology and crystallography, the structure of the amyloid fibrils has been clarified, providing deep insights into exploring the pathogenesis of this disease. These fibrils are three-dimensional architectures resulting from the aggregation of misfolded amyloid proteins with a sequence made up of 39-43 amino acids. These fibrils were found to be composed of cross-β strands aligned perpendicular to the fibril axis and exhibit extensive hydrogen bonding along the length of fibrils [Nelson and Eisenberg, 2006], as revealed by the X-ray diffraction. The structure of fibrils reveals a wide range of polymorphism from thin and straight fibrils to wide and twisted ribbons [Paravastu et al., 2008; Schmidt et al., 2016; Wälti et al., 2016]. Recent solid-state NMR and cryo-EM experiments show the existence of the parallel in-register β sheet in the amyloid fibrils which potentially maximizes favorable interactions between hydrophobic as well as polar side chains [Schmidt et al., 2016; Wälti et al., 2016].

In Figure 1.3a, we showed a close-up view of the atomic resolution of cross-β amyloid fibrils of transthyretin protein. The left panel of Figure 1.3a shows the TEM image of the fibrils where the white scale bar is 50 nm, and the right panel is MAS NMR atomic resolution structure of the fibrils fitted into cryo-EM. The cross-section of the fibrils is shown in the right panel. [Fitzpatrick et al., 2013]. The atomic resolution structure of this cross β fibrils shows that extensive hydrogen bonding interactions between backbone groups and hydrophobic interactions between sidechain groups have a large contribution to the stability of fibrils, but the exact details of how these interactions couple with each other still remains an open question. A hierarchal organization of cross β amyloid fibrils has been shown in Figure 1.3b, showing how the fibril is formed from the monomer β strands collecting into the β sheet, protofilament and further fibrils.

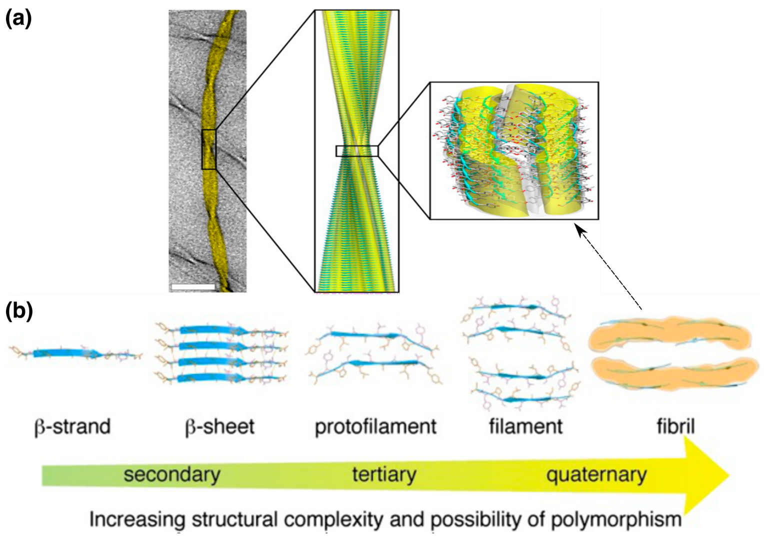


Figure 1.3: (a) TEM image of fibrils with the white scale bar of 50 nm (left panel), close-up view of the atomic-resolution structure of the fibril fitted into the cryo-EM reconstruction (center panel) and fibril surface (right panel); oxygen, carbon, and nitrogen atoms are shown in red, gray, and blue, respectively. (b) hierarchy of atomic-resolution motifs involved in the self-assembly from β strand to fibrils [Fitzpatrick et al., 2013].

1.2.2 Oligomers

While insoluble amyloid fibrils are believed to be the main toxic species, it has been reported that soluble oligomers such as the dimer, trimer and high-order ones also have toxicity [Kayed et al., 2003; Shankar et al., 2008; Cappai and Barnham, 2008]. No effect of removing amyloid plaques on the inhibition of AD are observed in several clinical trials [Hardy, 2009], meaning that the small amyloid oligomers might be thought of as primary toxic species of AD [Baglioni et al., 2006]. Although both oligomers and fibrils are present in the brain of an AD patient, soluble oligomers could be better correlated with disease than are the classic amyloid fibrils [CA et al., 1999; Lue et al., 1999; Gaspar et al., 2010; Takahashi et al., 2004]. It has been shown that the oligomers hold a variety of exchangeable and distinct structures which may contribute to the pathology of AD via different mechanisms [Lee et al., 2017]. Therefore, there is a lot of interest in understanding the conformational and dynamical heterogeneity of oligomers and the exact mechanism of oligomeric-induced neurotoxicity. In order to understand how the oligomerization occurs from monomer to high-order oligomers, it is important to provide a comprehensive microscopic description of both the structural landscape and molecular interactions for the monomer as a fundamental step.

1.3 Amyloid β Protein in Aqueous Solution

All the processes of amyloid oligomerization and fibrilization described earlier occur in a water medium. It is well known that the water is fundamentally related to the structure, stability, dynamics and function of the biomolecules like proteins [Halle, 2004; Nakasako, 2004]. For example, water dictates the hydrophobic force, which is considered to be the driving force of protein folding, to make the hydrophobic sidechains collapse and also mediates binding in the complex formation through the hydrogen bonding [Huggins, Marsh, and Payne, 2011; Papoian, Ulander, and Wolynes, 2003; Privalov and Makhatadze, 1993]. Protein-protein and protein-ligand interactions [Huggins, Marsh, and Payne, 2011; Papoian, Ulander, and Wolynes, 2003; Dubins et al., 2000], both hydrophobic [Russo, Teixeira, and Ollivier, 2009; Russo, Ollivier, and Teixeira, 2008] and electrostatic forces [Foderà et al., 2013; Zhang, Witham, and Alexov, 2011] are mediated by the solvent. Despite the extensive experimental and theoretical studies, the role of the solvent is relatively poorly understood in the amyloid oligomerization and fibril formation.

Due to the heterogeneity and high propensity to aggregate in amyloid proteins, the atomistic resolution 3D structures of monomer, as well as small oligomers in the solvent, are not manageable by traditional NMR and X-ray crystallography. Only low resolution atomistic structures are available by a combination of several techniques such as TEM, CD, AFM and so on [Hou et al., 2004; Bernstein et al., 2009; Maity and Lyubchenko, 2016]. Recent solid-state NMR spectroscopy experiment for amyloid fibrils show the existence of the different dynamic water layers around the fibrils [Wang et al., 2017]. Hierarchy of water layers in fibrils are shown in Figure 1.4; 1) three rigid water layers tightly bound to individual protofilament, 2) the dynamic water layer between the protofilaments, and 3) the bulk-like water layer. The existence of these distinct water layers strongly suggest a coupling between the protein and water even in the aggregation process, but microscopic detailed understanding is still missing.

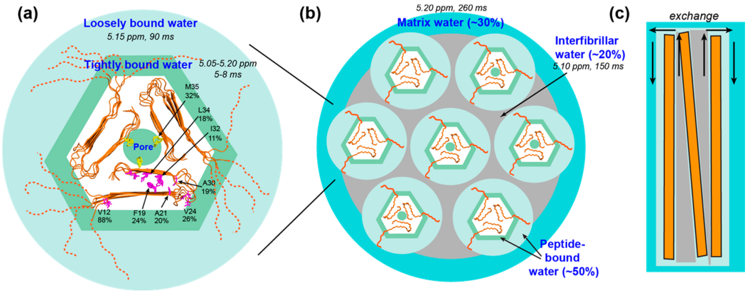


Figure 1.4: Schematic of the different water layers around the wild-type $A\beta_{40}$ fibrils. (a) Cross-section of the protofilament where N-terminus is indicated as a dashed line and C-terminus residues are shown in sticks. Water pore inside of the protofilament and two distinct hydration water layers are shown with different colors of hydration levels, (b) top view of fibrils with water layer between protofilaments and (c) side view of fibrils (orange) within water solvent. This figure is reproduced from [Wang et al., 2017].

There have been several theoretical attempts to understand the coupling between protein and water degrees of freedom in amyloid systems. Using molecular dynamics simulations, D. Thirumalai et. al. [Thirumalai, Reddy, and Straub, 2012] tried to provide a perspective on how interaction between the protein and water affect protein folding as well as its fibrilization. Through discussions for different systems of hydrophobic Aβ_{16-22} and hydrophilic yeast prion Sup35, they suggested aggregation from monomers to fibrils, in general, occurs by a two-step scenario: 1) aggregation-prone monomers aggregate into disordered oligomers and 2) rearrangement of the structures within the oligomers produces the ordered fibrils. From their perspective, interestingly, the two steps of the aggregation can be driven by release of water in the hydration shell into the bulk which would be entropically favorable. Different role of water is suggested for different hydrophilic and hydrophobic surfaces. For hydrophobic systems, water in the hydration shell is favorably expelled into bulk creating a dry interface, but for hydrophilic systems water molecules at the interface form very stable water wire resulting in significant slowdown of aggregation.

1.4 Intrinsic Fluorescence of Aβ fibrils

Monitoring the aggregation process in-vivo and in-vitro experiments is very important in understanding the aggregation mechanism of amyloid protein and developing the drugs to inhibit AD. Fluorescence, as one of the most powerful experimental tools, has been widely used in the protein studies such as protein folding and aggregation [Lakowicz, 2006; Lakowicz, 2006; Munishkina and Fink, 2007]. Since the proteins with the aromatic residues such as tryptophan, tyrosine and phenylalanine can absorb in the ultraviolet light of around 280nm and emit light at around 350 nm, it is usually referred to as “protein fluorescence” [Lakowicz, 2006]. Recently, there have been reports of a different type of intrinsic protein fluorescence in the amyloid β fibrils, which occurs in the visible range and appears to be independent of the presence of aromatic residues [Shukla et al., 2004; Sharpe et al., 2011; Chan et al., 2013; Pinotsi et al., 2016; Grisanti et al., 2017a]. As shown in Figure 1.5, the fluorescence is observed only when amyloid proteins aggregate into fibrils [Chan et al., 2013].

Recent experimental observations and theoretical calculations show that the intrinsic fluorescence is not exclusive to the amyloid aggregates and can originate from both monomeric and oligomeric proteins [Bhattacharya et al., 2017; Ye et al., 2017; Prasad et al., 2017; Chen et al., 2018].

Several studies show that intrinsic fluorescence for amyloid fibrils can occur by the reorganization of the electronic structure via hydrogen bonding between the amid groups as well as termini in close proximity [Bhattacharya et al., 2017; Ye et al., 2017; Chen et al., 2018]. The fluorescence experimental studies for a series of non-aromatic biogenic and synthetic peptides based on alanine, valine and isoleucine also show that the intrinsic fluorescence in the aggregated state is associated with abundant hydrogen bonds between amide groups [Ye et al., 2017]. Besides the amide groups,

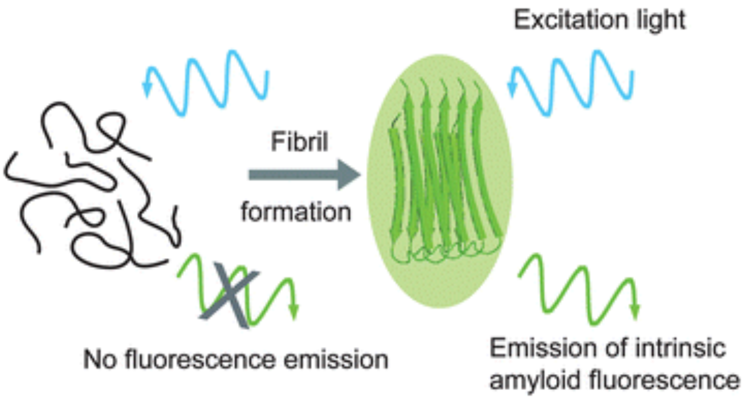


Figure 1.5: Diagram of monitoring the formation of amyoid fibrils by the intrinsic fluorescence. This figure is reproduced from [Chan et al., 2013].

the charge transfer between charged residue groups for charged amino acid monomers and poly-peptides is considered as one of dominant origins of the intrinsic fluorescence [Ansari et al., 2018; Prasad et al., 2017; Johansson and Koelsch, 2017].

Several works have suggested fluorescence might be utilized as a label-free diagnostic tool to probe the structural and dynamical transition of amyloid protein aggregates. For example, Ansari et al. [Ansari et al., 2018] showed that the absorption intensities of human c-Myc protein and its mutant at wavelengths of 250-800 nm are dependent on 3D proximity of charged functional groups across the protein. Moreover, significant changes in the spectra by changing pH in range of 3-11 as well as by the application of different temperatures and salts show a strong correlation between their secondary structure and fluorescence. Increase of absorption spectra with time in hen lysozyme at pH=2 directly correlated with the growth of aggregates, as confirmed by the increasing thioflavin T fluorescence. Meanwhile, sensitivity of absorption spectra by charge transfer transition to a proximity of different charged groups is very supportive to a feasibility to employ protein with non-aromatic residues as a new luminescent material [Mandal, Paul, and Venkatramani, 2018; Prasad et al., 2017; Chen et al., 2018].

1.5 Aims and Organization of the thesis

The present thesis is concerned with the structural complexity of amyloid proteins, the coupling of amyloid proteins and the surrounding water, and understanding the origin of intrinsic fluorescence of the amyloid proteins. Atomistic molecular dynamics (MD) simulations, data-science algorithms, and time-dependent density functional theories have been employed. For many of our studies, we use the hexapeptide 2Y3J (PDB code) model system with six amino acids (AIIGLM), $A\beta_{30-35}$, which resides on the C-terminus of full-length amyloid β protein and is shown to aggregate into fibrils [luca_2017]. Liu et al. [Liu et al., 2004] have shown that $A\beta_{30-35}$ plays a critical role in the aggregation process by including short anti-parallel strands in the surrounding residues, which in turn could promote the fibril formation of full-length amyloid β protein.

The thesis is organized in the following way:

A review of the theory of molecular dynamics and time-dependent density functional theory are discussed in Chapter 2. In addition, a brief introduction to the software I developed to study hydrogen bond network in complex system is discussed.

In Chapter 3, we deal with the validation of the force fields for small protein/water clusters using high-level quantum chemistry calculations. Despite small differences between force fields, all of them predict the potential energy surfaces in reasonably good agreement with the quantum chemistry calculations. We employ an energy decomposition analysis to show that most of discrepancies in the force fields originate from not including many-body polarization.

In Chapter 4, we use an enhanced sampling technique, well-tempered metadynamics, to explore the conformational landscape of model Aβ protein. We characterize distinct conformations on the free energy landscape, driven by a combination of several interactions between polar and apolar groups of protein. Furthermore, we show that conformational fluctuations can also be affected by the reorganization of the water wires threading the protein. The directed water wires provide an additional roughness to the conformational free energy landscape.

In order to better understand the coupling between protein and water, in Chapter 5 we examine using data science algorithms the complexity of different water coordinates around well-studied small peptide tri-alanine. Because of the complexity of water coordinates, we utilize the sophisticated data-science oriented algorithms which are able to provide their dimensionality together with underlying free energy landscape. The results reveal that water wires are very good coordinates for identifying the roughness in the free energy landscape and also encoding information on the underlying secondary structure.

In chapter 6, we carry out the ab initio calculations using time-dependent density functional theory to elucidate ground and excited state properties of different Aβ_{30-35} conformations sampled from classical molecular dynamics in chapter 4. We show that low energy excitations are commonly characterized by the charge transfer between polar groups including termini. Intra hydrogen bonds show strong correlation with low energy excitation, indicating that low energy excitations might originate from the reorganization of the electronic structure via internal hydrogen bonds including backbone and termini.

A comprehensive overview of the main conclusions of the thesis and future perspectives are summarized in chapter 7.

Chapter 2

Methodological Background

In this chapter, I review basic theory of molecular dynamics, data-science algorithms, and time-dependent density functional theory used in this thesis. I also briefly discuss software development to study hydrogen bond network of protein and water system.

2.1 Classical Molecular Dynamics (MD)

2.1.1 Overview of classical MD

Molecular dynamics is a computer simulation to understand movements of molecules and provide static and dynamical properties of the system. Note that classical terminology implies neglection of quantum behavior of the system, meaning that motions of electrons are not considered and all electrons are considered instantly follow positions of nuclei. Movements of system composed of N particles are determined by Newton's equation of motion:

$$\vec{a}_i = \frac{d\vec{v}_i}{dt} = \frac{\vec{F}_i}{m_i}, i = \{1 ... N\} \tag{2.1}$$

where $\vec{r}_i$, $\vec{v}_i$, $\vec{a}_i$, m_i and $\vec{F}_i$ are position, velocity, acceleration, mass and force of i-th particle, respectively. Force on the i-th particle ($\vec{F}_i$) in Equation 2.1 is obtained by derivative of interaction energy (U) at its position

$$\vec{F}_i = -\frac{\partial U}{\partial \vec{r}}\Big|_{\vec{r}=\vec{r}_i}, i = \{1 ... N\} \tag{2.2}$$

With the current computational resources, it is not yet feasible to calculate accurate potential energy of large biomolecules by means of quantum chemistry electronic structure, so that we use simple classical function, called Force Field, with a number of parameters to give a good agreement with experimental data and/or quantum chemistry calculations [Ponder and Case, 2003]. Several families of force fields are developed by different groups such as AMBER [Lindorff Larsen et al., 2010], CHARMM [MacKerell, Banavali, and Foloppe, 2000] and OPLS [Kaminski et al., 2001; Jacobson et al., 2002]. As you can see in chapter 3, we examined the qualities of these force field for hybrid protein water clusters. For example, functional form of AMBER force field is defined as following:

$$
\begin{aligned}
U(\vec{r}^N) = \sum_{bonds} k_b(\ell - \ell_0)^2 + \sum_{angles} k_a(\theta - \theta_0)^2 + \sum_{dihedrals} \sum_n \frac{1}{2} V_n [1 + cos(n\phi - \gamma)] \\
+ \sum_{j=1}^{N-1} \sum_{i=j+1}^{N} f_{ij} \Big\{ \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} + \epsilon_{ij} \left[\left(\frac{r_{0ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{r_{0ij}}{r_{ij}} \right)^{6} \right] \Big\}
\end{aligned}
\tag{2.3}
$$

where each term is an approximation for different interaction energy. The first three terms are bonded terms for interactions of atoms linked by covalent bonds with bonds (ℓ), angles (θ), and dihedrals (ϕ) and the last term is a non-bonded term for long-range electrostatic and van der Waals energies (Lennard-Jones 6-12 potential). A large number of different parameter sets exist even for the same family of force field. In case of AMBER force field in Equation2.3, there are AMBER94, AMBER96, AMBER99, AMBER99SB, AMBER03, and so on [Duan et al., 2003; Best, Zheng, and Mittal, 2014; Lindorff Larsen et al., 2010] provided for the different types biological systems.

After force on each atom is obtained by force field, its position and velocity are determined by the integration of Equation 2.1. There exist several algorithms to integrate Newton's equation of motion such as Verlet algorithm [Verlet, 1967], the velocity Verlet algorithm [Swope et al., 1982] and so on. The Verlet algorithm, one of frequently used algorithms, is derived by the Taylor expansion of coordinates of a particle at time $t \pm \Delta t$ where Δt is time step,

$$
\vec{r}_i(t \pm \Delta t) = \vec{r}_i(t) \pm \vec{v}_i(t)\Delta t + \frac{\vec{a}_i(t)}{2}\Delta t^2 \pm \frac{\dddot{\vec{r}}_i}{3!}\Delta t^3 + O(\Delta t^4) \tag{2.4}
$$

addition of which gives

$$
\vec{r}_i(t + \Delta t) \approx \vec{r}_i(t) - \vec{r}_i(t - \Delta t) + \vec{a}_i(t)\Delta t^2 + O(\Delta t^4) \tag{2.5}
$$

Equation 2.5 can be used in classical MD to advance positions of the atoms over time, but this explicitly does not include the terms of the velocities. The explicit cooperation with velocities is apparently solved in the velocity Verlet algorithm which allows us to calculate the atomic positions and velocities at time $t + \Delta t$ by the positions and velocities at time t:

$$
\begin{aligned}
\vec{r}(t + \Delta t) &= \quad \vec{r}(t) + \vec{v}(t)\Delta t + \frac{1}{2}\vec{a}(t)\Delta t^2 \\
\vec{v}(t + \Delta t) &= \quad \vec{v}(t) + \frac{\vec{a}(t) + \vec{a}(t + \Delta t)}{2}\Delta t
\end{aligned}
\tag{2.6}
$$

2.1.2 Metadynamics as an Enhanced Sampling Technique

Although MD has been widely used in different aspects, its result is meaningful only if its run is long enough for the system to visit all energetically relevant configurations. In other words, ergodicity of the system might be guaranteed by the long timescale of the simulation in case of the existence of high free energy barriers between configurations. In such cases, change from one relevant configuration to another can take place only by rare fluctuations over free energy barrier and obtaining sufficient statistics requires an enormous amount of simulation time.

In last years, several methods have been proposed to address this timescale problem of MD, often offered to as enhanced sampling techniques. Metadynamics [Laio and Parrinello, 2002] belongs to a class of methods in which sampling is enhanced by the introduction of an additional bias potential acting on a few collective variables (CVs) whose fluctuations are very important for rare event of interest to take place. Besides metadynamics, a number of methods belong to this class, such as umbrella sampling method [Torrie and Valleau, 1977], local elevation [Huber, Torda, and Gunsteren, 1994], conformational flooding [Grubmüller, 1995; Müller, Meijere, and Grubmüller, 2002], adaptive biasing force [Darve and Pohorille, 2001], Wang-Landau algorithm [Wang and Landau, 2001], energy landscape paving [Hansmann and Wille, 2002], steered MD [Park and Schulten, 2004], self-healing umbrella sampling [Marsili et al., 2006] and Gaussian-mixture umbrella sampling [Maragakis, Vaart, and Karplus, 2009].

Let's assume that the system is composed of N particles with positions at $\vec{r}$ ($\{\vec{r}_i\}$, $i \in 1...N$) and $\vec{s}$ is a set of D collective variables of particle positions:

$$\vec{s}(\vec{r}) = (s_1(\vec{r}), s_2(\vec{r}), ..., s_D(\vec{r})) \tag{2.7}$$

In metadynamics, an external history-dependent bias potential which is a function of CVs is added to Hamiltonian (H) of the system:

$$H = T + U + V_G \tag{2.8}$$

where T is total kinetic energy, U is potential energy to describe the system, and V_G is bias potential. The bias potential is usually presented as a sum of Gaussian hills in CV space to inhibit the system from revisiting configurations that have already been sampled. At time t, bias potential deposited along the trajectory can be written as

$$V_G(\vec{s}, t) = \int_0^t dt' \omega exp\left(-\sum_{i=1}^{D} \frac{(s_i - s_i(t')^2}{2\sigma_i^2}\right) \tag{2.9}$$

where ω is energy rate, a constant obtained by ratio of Gaussian height W to deposition stride τ_G ($\frac{W}{\tau_G}$), and σ_i is the width of the Gaussian for the ith CV.

In Figure 2.1, we show how metadynamics is working for simple 1D potential with three local minima A, B, and C. The left panel of Figure 2.1 shows the underlying original potential (black line) and the evolution of total potential by deposition of Gaussian hills along the trajectory (color lines). In the right panel of Figure 2.1, we show time series of variable. At t=0, the system was set in local minimum B. In a standard MD simulation, the system would remain stuck in this basin for a while because energy barriers are larger than thermal fluctuations. Instead, in metadynamics simulation, underlying bias potential grows by the deposition of the Gaussian hills, pushing the system out of basin B into a new local minimum at around t=180. Since basin B has a lower barrier with basin A than basin C, the system naturally falls into basin A and is trapped in this basin, until it is completely filled (t≈600). After this time, the system starts to diffuse in the region of both basin A and B. Continuous filling on these basins A and B allow the system to access region of basin C. Starting from t≈1500 when all basins are compensated by bias potential, the system evolves like a random walk over time. The bias potential V_G provides an unbiased estimate of underlying free energy

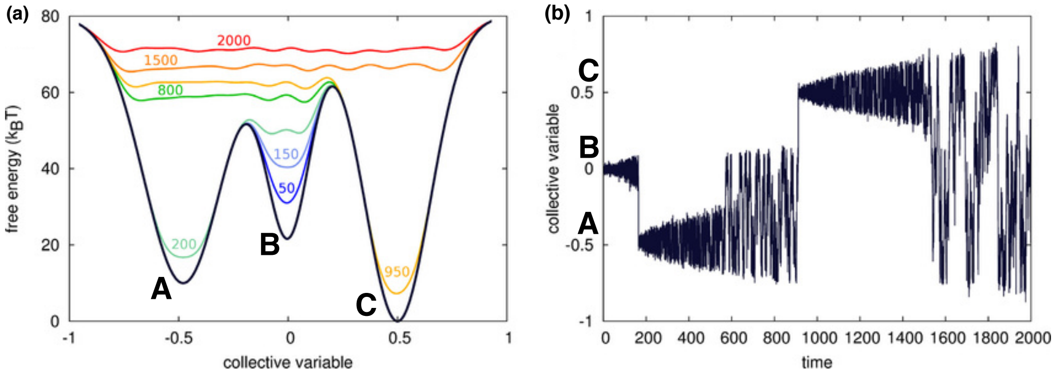


Figure 2.1: Schematic of metadynamics for one dimensional model potential. a) Underlying potential (black line) and total potential filled by Gaussian deposits along the trajectory (colored lines). b) Time evolution of collective variable during the simulation. This figure is reproduced from [Pietrucci, 2017]

$$V_G(\vec{s}, t \to \infty) = -F(\vec{s}) + C \tag{2.10}$$

where C is an irrelevant additive constant and the free energy F($\vec{s}$) is defined as

$$F(\vec{s}) = -\frac{1}{\beta} log P(\vec{s}) = -\frac{1}{\beta} log \left(\int d\vec{r} \delta(\vec{s} - \vec{s}(\vec{r})) e^{-\beta U(\vec{r})} \right) \tag{2.11}$$

where $P(\vec{s})$ is probability distribution, $\beta = (k_B T)^{-1}$, k_B is the Boltzmann constant, T is the temperature of the system.

Despite several benefits of the metadynamics, this method has two major disadvantages: 1) Choice of CVs appropriate for describing the complicated process is nontrivial. 2) Constant height of Gaussian hills makes it difficult for bias potential to converge to the underlying potential energy modulo constant. The bias potential overfills the underlying free energy, oscillating around it and pushing the system towards high energy region.

There is no systematic way to choose appropriate CVs, but it can be usually proceeded by short standard MD simulations which give a general overview of a physicochemical process of the system. A solution to the convergence of metadynamics is provided by well-tempered metadynamics [Barducci, Bussi, and Parrinello, 2008]. In well-tempered metadynamics, the bias deposition rate decreases along trajectory by using different expression of bias potential:

$$V_G(\vec{s},t) = k_B \Delta T ln \left(1 + \frac{\omega N(\vec{s},t)}{k_B \Delta T}\right), \tag{2.12}$$

$$\dot{V}_G(\vec{s},t) = \frac{\omega \delta_{\vec{s},\vec{s}(t)}}{1 + \frac{\omega N(\vec{s},t)}{k_B \Delta T}} = \omega e^{-\frac{V_G(\vec{s},t)}{k_B \Delta T}} \delta_{\vec{s},\vec{s}(t)} \tag{2.13}$$

where $\dot{V}_G(\vec{s},t)$ is time derivative of bias potential, $N(\vec{s},t) = \int_0^t \delta_{\vec{s},\vec{s}(t')dt'}$ is histogram of CVs during the simulation, and ΔT is an input parameter with the dimension of a temperature. In pratice, Equations 2.12 and 2.13 are implemented by rescaling the Gaussian height W according to

$$W = \omega \tau_G e^{-\frac{V_G(\vec{s},t)}{k_B \Delta T}} \tag{2.14}$$

Unlike standard metadynamics, bias potential does not fully compensate the underlying free energy, but it converges to

$$V_G(\vec{s}, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(\vec{s}) + C \tag{2.15}$$

where C is a irrelevant constant. In long time limit, probability distribution of CVs can converge to

$$P(\vec{s}) \propto e^{-\frac{F(\vec{s})}{k_B (T + \Delta T)}} \tag{2.16}$$

As you can see in Equation 2.16, for $\Delta T \to 0$, ordinary MD is recovered, whereas $\Delta T \to \infty$ limit corresponds to standard metadynamics.

The probability distribution of CVs can be reconstructed by Equation 2.16, but for the other degrees of freedom is distorted in a nontrivial way. Different techniques have been proposed to reweight a metadynaimcs run and recover unbiased distribution of degrees of freedom other than CVs [Tiana, G., 2008; Marinelli et al., 2009; Bonomi, Barducci, and Parrinello, 2009]. Recently, P. Tiwary et.al. introduced a very simple reweighting scheme to recover unbiased distributions of generic operators [Tiwary and Parrinello, 2015]. Under the action of internal potential $U(\vec{r},t)$ and bias potential $V(\vec{s},t)$, instantaneous distribution $P(\vec{r},t)$ can be written

$$P(\vec{r},t) = \frac{e^{-\beta(U(\vec{r}) + V(s(\vec{r}),t))}}{\int d\vec{r} e^{-\beta(U(\vec{r}) + V(s(\vec{r}),t))}} \tag{2.17}$$

which can be rewritten in the form

$$P(\vec{r},t) = e^{-\beta(V(s(\vec{r}),t) - c(t))} \cdot P_0(\vec{r}) \tag{2.18}$$

where $P_0(\vec{r})$ is the unbiased probability distribution and $c(t)$ is defined as

$$c(t) = \frac{1}{\beta} log \frac{\int d\vec{s} e^{-\beta F(\vec{s})}}{\int d\vec{s} e^{-\beta(F(\vec{s})+V(\vec{s},t))}} \tag{2.19}$$

They show how to calculate time dependent function $c(t)$ in Equation 2.18 which provides important information about the system.

In well-tempered metadynamics, $c(t)$ can be obtained by

$$e^{\beta c(t)} \approx \frac{k_B \Delta T}{\gamma \omega \Delta t (2\pi)^{D/2} det\sigma} = \int d\vec{s} \left[e^{\gamma V(\vec{s},t+\Delta t)/k_B \Delta T} - e^{\gamma V(\vec{s},t)/k_B \Delta T} \right] \tag{2.20}$$

where $det\sigma$ is the determinant of the CV space, Δt is time interval between successive Gaussian depositions. The equilibrium average of an arbitrary operator $\langle O(\vec{r}) \rangle_0$ can be computed by an average for metadynamics run qauntity

$$\langle O(\vec{r}) \rangle_0 = \langle O(\vec{r}) e^{\beta(V(\vec{s},t)-c(t))} \rangle \tag{2.21}$$

2.2 Development of Software for Hydrogen Bond Network Analysis

As described in Chapter 1, complex processes of biomolecules like protein folding and aggregation usually occur in an aqueous solution and the solvent is considered to play a fundamental role in these processes. To investigate the role of water, we examine directed hydrogen bond water wires connecting donor and acceptor groups of the protein. The water wires help us to understand the properties of the hydrogen bond network around the protein. Donor groups include the N-terminus and amide N−H bonds of the backbone, while the C-terminus and the carbonyl C=O groups are acceptor groups. Oxygen atoms of water molecules (O_w) act as both donors and acceptors. Among many possible water wires connecting same donor and acceptor groups, we look at the shortest water wires in which the smallest number of water molecules are involved. In order to extract the shortest water wires, we use graph theory by treating the donor and acceptor groups as vertices on graph and hydrogen bonds between vertices with edges. If two vertices v_i and v_j are connected by an edge e_{ij}, they are said to be adjacent. The edges e_{ij} encode information about the hydrogen bonds. We assign the weights, corresponding to strengths of hydrogen bonds, to edges leading to a weighted directed adjacency matrix A=A(G), which is a $M \times M$ matrix with M being the number of vertices. There are broadly two criteria to define a hydrogen bond, one based on geometry and another based on energetics. In this work, we use the former and define the weight (A_{ij}) of edge (e_{ij}) based on a combination of distance and angle of hydrogen bond [Luzar and Chandler, 1996]

$$A_{ij} = \frac{1}{f(d_{DA}) \sum_H f(d_{HA}) f(\angle HDA)} \tag{2.22}$$

where H is the index of hydrogen atom, d_{DA} is distances between donor and acceptor atoms, d_{HA} is distance between hydrogen and acceptor atoms and, $\angle HDA$ is an angle between hydrogen and acceptor atoms at donor atom. Switching function ($f(\cdot)$) measures how much values differ from certain quantities defined as

$$f(s) = \frac{1 - (\frac{s-d_0}{s_0})^m}{1 - (\frac{s-d_0}{s_0})^n} \tag{2.23}$$

where d_0 is minimum criteria of s value, and the exponents (m,n) and s_0 determine smoothness of switching function. The parameters m and n for all of d_{DA}, d_{HA} and $\angle HDA$ are set to 12 and 24, respectively. d_0 is set to zero for all variables, while values of s_0 are set to 0.35 nm, 0.25 nm, and 30°, respectively.

After constructing the directed adjacency matrix A, we extract the shortest water wires from the adjacency matrix using Dijkstra's algorithm [DIJKSTRA, 1959]. Dijkstra's algorithm is known to be one of the most efficient algorithms to find shortest paths from "source" node to other "target" nodes. The donor and acceptor groups of protein are considered as source and target groups in the graph, respectively. The weight of each node is initially assigned with distance from the source, where the adjacent nodes of the source have weights of edges and the other nodes have infinity values. The weight of the source itself is set to zero. The weights of nodes are gradually upgraded with smaller values of weight until all target nodes are reached. In this work, I have developed C++ program to construct the adjacency matrix and extract the water wires connecting donor and acceptor groups.

2.3 Data Science Algorithm for Computing Dimensionality and Free Energy Surface in Very High-Dimensional Space

Estimating underlying free energy is one of the key goals to understand the complex process which occurs in a biological system Several approaches [Kumar et al., 1995; Kumar, Payne, and Vásquez, 1996; Ceriotti, Tribello, and Parrinello, 2011] are available to perform this task.

As you can in Equation 2.11, free energy is defined by probability density and therefore, free energy estimation is in principle density estimation. There are two different categories of approaches, parametric and nonparametric ones, available for density estimation [Scott, 2015; Läuter, 1986; Dempster, Laird, and Rubin, 1977; Izenman, 1991; Sheather, 2004]. While in parametric approaches density function is assumed to has a specific form like Gaussian [Läuter, 1986; Dempster, Laird, and Rubin, 1977], no strong assumption is made in nonparametric methods [Izenman, 1991; Sheather, 2004]. For example, in the k-nearest neighbor estimator (k-NN) [Mack and Rosenblatt, 1979], density is measured as the ratio of the number of nearest neighbors k to the volume they occupy without any assumption. It is known that the choice of the value of k could induce systematic error.

In this work, We have used the density estimator proposed by A. Rodriguez and A. Laio [Rodriguez and Laio, 2014]. Let us $X_1, ..., X_N$ a set of N-independent and identically distributed random vectors in $\Re^D$ and assume that the X_i lie on a manifold of dimension d($\leq$D), constant in the data set. Note that the dimension d is often called intrinsic dimension of the data set [Facco et al., 2017]. Let $r_{i,\ell_{\ell \leq k}}$ be the sequence of ordered distanced from ith point to its k-nearest neighbors; $r_{i,1}$ is the distance between i and its nearest neighbor, $r_{i,2}$ is the distance with its second nearest neighbor, and so on. In a sufficiently small neighborhood of each point, the distance can be approximated by the Euclidean distance in $\Re^D$. The volume of the hyperspherical shell enclosed between neighbors $\ell - 1$ and ℓ is given by

$$\Delta v_{i,\ell} = \omega_d (r^d_{i,\ell} - r^d_{i,\ell-1}) \tag{2.24}$$

where ω_d is the volume of the $d-$ sphere with unitary radius. If the density is constant around the point (i), all the $\Delta v_{i,\ell}$ are independently drawn from an exponential distribution with rate equal to the density ρ:

$$P(\Delta v_{i,\ell} \in [v, v + dv]) = \rho e^{-\rho v} dv \tag{2.25}$$

Consider two shells $\Delta v_{i,\ell}$ and $\Delta v_{i,m}$, and let R be the quantity $\frac{\Delta v_{i,\ell}}{\Delta v_{i,m}}$. In the case of constant density, the probability distribution of R can be exactly obtained from Equation 2.25

$$P(R \in [\bar{R}, \bar{R} + d\bar{R}]) = \int_0^\infty dv_{i,\ell} \int_0^\infty dv_{i,m} \rho^2 e^{-\rho(v_{i,\ell}+v_{i,m})} 1_{\{\frac{v_{i,m}}{v_{i,\ell}} \in [\bar{R}, \bar{R}+d\bar{R}]\}} = d\bar{R} \frac{1}{(1+\bar{R})^2} \tag{2.26}$$

where 1 represents the indicator function. Dividing both sides of Equation 2.26 by $d\bar{R}$, we obtain probability distribution for R, g(R)

$$g(R) = \frac{1}{(1+R)^2} \tag{2.27}$$

$g(R)$ does not depend explicitly on the dimension d, which is included in the definition of R. If we defined new quantity $\mu \doteq \frac{r_{i,2}}{r_{i,1}} \in [1, +\infty)]$, R and μ are related by the equality

$$R = \mu^d - 1 \tag{2.28}$$

From this equation, we can easily derive an formula for the distribution of μ

$$f(\mu) = d\mu^{-d-1} 1_{[1,+\infty)}(\mu) \tag{2.29}$$

while the cumulative distribution of μ is obtained by integration

$$F(\mu) = (1 - \mu^{-d}) 1_{[1,+\infty)}(\mu) \tag{2.30}$$

As you can see in Equations 2.29 and 2.30, f and F are independent on the local density, but depend explicitly on the dimension d. The dimension (intrinsic dimension) d can be estimated through the following equation (TWO-NN [Facco et al., 2017])

$$\frac{log(1 - F(\mu))}{log(\mu)} = d \tag{2.31}$$

In practice, the dimension d is estimated form the slope of the line fitting the value of $-log\left(1 - F^{emp}\left(\frac{r_{i,2}}{r_{i,1}}\right)\right)$ as a function of $log\left(\frac{r_{i,2}}{r_{i,1}}\right)$, where F^{emp} is for the empirical cumulative distribution.

Since all the volumes $v_{i,\ell}$ are independently drawn from an exponential distribution with rate equal to density ρ, the log-likelihood function of the parameter ρ given the observation of the k-nearest neighbor distances from point i is

$$L(\rho|\{v_{i,\ell}\}_{\ell\leq k}) \doteq L_{i,k}(\rho) = klog\rho - \rho\sum_{\ell=1} kv_{i,\ell} = klog\rho - \rho V_{i,k} \tag{2.32}$$

where $V_{i,k} = \sum_{\ell=1}^{k} v_{i,\ell}$ is the volume of the hypershpere with center at i containing k data points.

By maximizing L with respect to ρ, it is found that $\rho = k/V_{i,k}$, the standard k-NN estimator of the local density. The estimated error on ρ is given by the asymptotic standard deviation of the parameter estimate: $\epsilon_\rho = \frac{\rho}{\sqrt{k}} = \frac{\sqrt{k}}{V_{i,k}}$. While the error ϵ_ρ gets smaller as k increases, the density in the neighborhood of the point i within the distance $r_{i,k}$ might become inconstant. The value of k can be chosen as the largest possible value for which the condition of constant density holds within a given level of confidence [Rodriguez et al., 2018].

Once the optimal k for point i ($\hat{k}_i$) is found, free energy of point i from density estimator can be computed as $F_i = -k_B Tlog(\rho_i) = -log\left(\frac{\hat{k}_i}{V_{i,\hat{k}_i}}\right)$.

2.4 Time-dependent Density Functional Theory

2.4.1 Ground State Density Functional Theory

Ground state density functional theory is a computational method to investigate the electronic structure of many-body systems in the ground state, using functionals of electron density. This theory is put on a firm footing by the theoretical works of Walter Kohn and Pierre Hohenberg [Hohenberg and Kohn, 1964]. First Hohenberg-Kohn (HK) theorem states that for non-degenerate ground states, all ground state properties of the many-electron system such as external potential and total energy are uniquely determined by electron density (up to an additive constant). Second HK theorem proves that exact electron density minimizes total energy. HK theorems are shown to hold even for degenerate ground-states [Levy, 1982] and an analogous theorem was applied to the spin densities, $n_\alpha(\vec{r}), n_\beta(\vec{r})$, where $\alpha, \beta = \pm\frac{1}{2}$.

The ground state electronic state of the N electron system is described by a wavefunction $\Psi(\vec{r}_1, ..., \vec{r}_N)$ which satisfies time independent Schrödinger equation

$$H\Psi = [T + V_{ee} + V_{ext}]\,\Psi = E\Psi \tag{2.33}$$

where H is Hamiltonian, E is total energy, T is kinetic energy, V_{ee} is electron-electron interaction, and V_{ext} is external potential energy, each of which is defined below

$$\begin{aligned} T &= -\frac{1}{2}\sum_{i=1}^{N}\nabla_i^2 && (2.34)\\ V_{ee} &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{1}{|\vec{r}_i - \vec{r}_j|} && (2.35)\\ V_{ext} &= \sum_{i=1}^{N} v_{ext}(\vec{r}_i) && (2.36)\end{aligned}$$

The electron density, $n(\vec{r})$, is given by normalized wavefunction Ψ

$$n(\vec{r}) = N\int d\vec{r}_2 \cdots \int d\vec{r}_N \Psi^*(\vec{r}, \vec{r}_2, ..., \vec{r}_N)\Psi(\vec{r}, \vec{r}_2, ..., \vec{r}_N) \tag{2.37}$$

At ground states, Equation 2.37 can be reversed and wavefunction Ψ_0 can be determined as a unique function of density, i.e., $\Psi_0 = \Psi[n_0]$. Ground state energy can also be defined as a function of n_0

$$E_0 = E[n_0] = \langle\Psi[n_0]|T + V_{ee} + V_{ext}|\Psi[n_0]\rangle \tag{2.38}$$

From Equation 2.38, ground state density n_0 and energy $E[n_0]$ could be obtained by minimization of energy functional $E[n]$

$$E[n] = \langle\Psi[n]|T + V_{ee} + V_{ext}|\Psi[n]\rangle = T[n] + V_{ee}[n] + \int d\vec{r} v_{ext}(\vec{r})n(\vec{r}) \tag{2.39}$$

with respect to $n(\vec{r})$. Note that $T[n]$ and $V_{ee}[n]$ are called universal functionals because they are the same for any N electron system regardless of external potential. This minimization can be done using the Euler equation

$$\begin{aligned} \frac{\delta}{\delta n(\vec{r})}\left[E[n] - \mu\int d\vec{r}' n(\vec{r}')\right] &= 0 && (2.40)\\ \frac{\delta T[n] + V_{ee}[n]}{\delta n(\vec{r})} + v_{ext}(\vec{r}) &= \mu && (2.41)\end{aligned}$$

where μ is a Lagrange multiplier to put constraint on the total number of electrons. Solving Equation 2.41 will provide ground state density n_0 and all ground state properties.

Let us consider a non-interacting system of N electrons where electron-electron interaction energy term V_{ee} is absent. We denote the Hamiltonian of this system by

H_s

$$H_s = T + V_s = \sum_{i=1}^{N}\left(-\frac{\nabla_i^2}{2} + v_s(\vec{r}_i)\right) \tag{2.42}$$

Application of HK theorems to this system ensures a one-to-one correspondence between external potential ($v_s(\vec{r})$) and ground state density ($n_0 s(\vec{r})$). The total energy functional can be written as

$$E_s[n] = T_s[n] + \int d\vec{r} n(\vec{r}) v_s(\vec{r}) \tag{2.43}$$

Euler equation 2.40 becomes

$$\frac{\delta T_s[n]}{\delta n(\vec{r})} + v_s(\vec{r}) = \mu \tag{2.44}$$

Solving Equation 2.44 will give the exact ground state density $n_0 s(\vec{r})$ of non-interaction system. For non-interacting systems, many-body ground state wavefunction reduces to a single Slater determinant [Kohn and Sham, 1965]

$$\Psi_s(\vec{r}_1, \cdots, \vec{r}_N) = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(\vec{r}_1) & \phi_2(\vec{r}_1) & \cdots & \phi_N(\vec{r}_1) \\ \phi_1(\vec{r}_2) & \phi_2(\vec{r}_2) & \cdots & \phi_N(\vec{r}_2) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \phi_1(\vec{r}_N) & \phi_2(\vec{r}_N) & \cdots & \phi_N(\vec{r}_N) \end{vmatrix} \tag{2.45}$$

where sigle particle orbitals $\phi_j(\vec{r})$ satisfy the Schrödinger equation

$$\left(-\frac{1}{2}\nabla^2 + v_s(\vec{r})\right)\phi_j(\vec{r}) = \epsilon_j \phi_j(\vec{r}) \tag{2.46}$$

and ground state density is obtained from N lowest single particle orbitals

$$n_s(\vec{r}) = \sum_{j=1}^{N} |\phi_j(\vec{r})|^2 \tag{2.47}$$

Equations 2.46 and 2.47 also provide exact ground state density of non-interacting system.

To use the single-particle picture, a fully interacting system can be mapped to a non-interacting one which gives the exact same density. Total energy functional in Equation 2.39 is rewritten in the following way

$$E[n] = T[n] + V_{ee}[n] + \int d\vec{r} v_{ext}(\vec{r}) n(\vec{r}) = T_s[n] + \int d\vec{r} v_{ext}(\vec{r}) n(\vec{r}) + E_H[n] + E_{XC}[n] \tag{2.48}$$

where $E_H[n]$ is Coulomb energy or Hartree energy given by

$$E_H[n] = \frac{1}{2}\int d\vec{r}d\vec{r}'\frac{n_(\vec{r})n_(\vec{r}')}{|\vec{r}-\vec{r}'|} \tag{2.49}$$

and $E_{XC}[n]$ is exchange-correlation energy

$$E_{XC}[n] = T[n] - T_s[n] + V_{ee}[n] - \frac{1}{2}\int d\vec{r}d\vec{r}'\frac{n_(\vec{r})n_(\vec{r}')}{|\vec{r}-\vec{r}'|} \tag{2.50}$$

Using the Euler equation 2.40 for total energy functional, we obtain

$$\frac{\delta T_s[n]}{\delta n(\vec{r})} + v_{ext}(\vec{r}) + \int d\vec{r}'\frac{n_(\vec{r}')}{|\vec{r}-\vec{r}'|} + \frac{\delta E_{XC}[n]}{\delta n(\vec{r})} = \mu \tag{2.51}$$

Compared to Equations 2.44, 2.46 and 2.47, the following equations, called Kohn-Sham (KS) equations are derived

$$v_s[n](\vec{r}) = v_{ext}(\vec{r}) + \int d\vec{r}'\frac{n_(\vec{r}')}{|\vec{r}-\vec{r}'|} + v_{XC}[n](\vec{r}) \tag{2.52}$$

$$v_{XC}[n](\vec{r}) = \frac{\delta E_{XC}[n]}{\delta n(\vec{r})} \tag{2.53}$$

$$\left(-\frac{1}{2}\nabla^2 + v_s(\vec{r})\right)\phi_j(\vec{r}) = \epsilon_j\phi_j(\vec{r}) \tag{2.54}$$

$$n_0(\vec{r}) = \sum_{j=1}^{N}|\phi_j(\vec{r})|^2 \tag{2.55}$$

Once $E_{XC}[n]$ is known exactly or approximated, $v_{XC}[n](\vec{r})$ is determined by differentiation. KS equations can be solved self-consistently for electron density and orbitals, and the total energy can be determined by inserting these into the total energy functional $E = T_s + E_H + E_{XC} + V_{ext}$.

KS equations can be easily generalized into their spin resolved forms

$$v_{s\sigma}[n_\alpha, n_\beta](\vec{r}) = v_{ext\sigma}(\vec{r}) + \int d\vec{r}'\frac{n_(\vec{r}')}{|\vec{r}-\vec{r}'|} + v_{XC\sigma}[n_\alpha, n_\beta](\vec{r}) \tag{2.56}$$

$$v_{XC}[n_\alpha, n_\beta](\vec{r}) = \frac{\delta E_{XC}[n_\alpha, n_\beta]}{\delta n_\sigma(\vec{r})} \tag{2.57}$$

$$\left(-\frac{1}{2}\nabla^2 + v_{s\sigma}[n_\alpha, n_\beta](\vec{r})\right)\phi_{j\sigma}(\vec{r}) = \epsilon_{j\sigma}\phi_{j\sigma}(\vec{r}) \tag{2.58}$$

$$n_0(\vec{r}) = n_{0\alpha}(\vec{r}) + n_{0\beta}(\vec{r}) = \sum_{\sigma=\alpha,\beta}\sum_{j=1}^{N_\sigma}|\phi_{j\sigma}(\vec{r})|^2 \tag{2.59}$$

Please note that $E_{XC}[n_\alpha, n_\beta]$ is not known exactly and must be approximated. There exist many different approximations of varying accuracy and computational cost. In any ground state DFT calculation, we must use approximations for the functional dependence of the XC energy on the spin densities. There exists a hierarchy of such approximations such as local density approximation (LDA), generalized gradient approximation (GGA) and hybrid functionals.

Please note that we have used a more accurate level theory of MP2 [Møller and Plesset, 1934; Binkley and Pople, 1975] than ground-state DFT for the properties of ground state electronic structures in Chapter 3.

2.4.2 Time-dependent Density Functional Theory

Time-dependent density functional theory, as an extension of ground state density functional theory, is a method to investigate the properties of the many-body system in the presence of time-dependent external potential such as an electric and magnetic field. Time evolution of an electronic many-body system is governed by time-dependent many-body Schrödinger equation under a initial condition

$$i\frac{\partial}{\partial t}\Psi(\vec{r}_1,\cdots,\vec{r}_N,t) = H(t)\Psi(\vec{r}_1,\cdots,\vec{r}_N,t) \tag{2.60}$$

$$H(t) = T + V_{ee} + V_{ext}(t) \tag{2.61}$$

$$\Psi(t=0) = \Psi_0 \tag{2.62}$$

where $V_{ext}(t)$ is time-dependent external potential.

Runge-Gross (RG) theorem [Runge and Gross, 1984a], as formal foundation of TDDFT, states that the densities $n(\vec{r},t)$ and $n'(\vec{r},t)$ evolving from a common initial state Ψ_0 under influence of two external potentials $v_{ext}(\vec{r},t)$ and $v'_{ext}(\vec{r},t) \neq v_{ext}(\vec{r},t) + c(t)$ (both Taylor expandable about the initial time 0) are always different. Therefore, there is a one-to-one correspondence between densities and potentials and the time-dependent potential is unique functional of time-dependent density, vice versa (for a given fixed initial state). Since potentials are functionals of densities, Hamiltonian $H(t)$ and thus wavefunction$\Psi(t)$ are functionals of electron density. All physical observables become functionals of density

$$O(t) = \langle \Psi[n,\Psi_0]|O|\Psi[n,\Psi_0]\rangle = O[n,\Psi_0](t) \tag{2.63}$$

The van Leeuwen theorem [Leeuwen, 1999] shows that the time dependent density $n(\vec{r},t)$ of an interacting system associated with external potential $v_{ext}(\vec{r},t)$ and initial state Ψ_0 can be reproduced by a noninteracting system with effective potential $v_s[n](\vec{r},t)$. Therefore, we can easily apply KS approaches into TDDFT to drive TD Kohn-Sham (TDKS) equations

$$\left[-\frac{\nabla^2}{2} + v_s[n](\vec{r},t)\right]\phi_j(\vec{r},t) = i\frac{\partial}{\partial t}\phi_j(\vec{r},t) \tag{2.64}$$

with initial condition

$$\phi_j(\vec{r},0) = \phi_j^0(\vec{r}) \tag{2.65}$$

whose potential should reproduce the exact spin densities of interacting system

$$n(\vec{r},t) = \sum_{j=1}^{N} |\phi_j(\vec{r},t)|^2 \tag{2.66}$$

The effective potential $v_s[n](\vec{r},t)$ is given by

$$v_s[n](\vec{r},t) = v_{ext}(\vec{r},t) + \int d\vec{r}'\, \frac{n(\vec{r}',t)}{|\vec{r}-\vec{r}'|} + v_{XC}[n](\vec{r},t) \tag{2.67}$$

This defines the exchange-correlation potential $v_{XC\sigma}(\vec{r},t)$.

Just as in ground state DFT, TDKS equations are often written in a form with spin indices

$$\left[-\frac{\nabla^2}{2} + v_{ext\sigma}(\vec{r},t) + \int d\vec{r}'\, \frac{n(\vec{r}',t)}{|\vec{r}-\vec{r}'|} + v_{XC\sigma}[n](\vec{r},t)\right] \phi_{j\sigma}(\vec{r},t) = i\frac{\partial}{\partial t}\phi_{j\sigma}(\vec{r},t) \tag{2.68}$$

where total density is given by sum of spin up and down densities

$$n(\vec{r},t) = \sum_{\sigma=\alpha,\beta} n_\sigma(\vec{r},t) = \sum_{\sigma=\alpha,\beta} \sum_{j=1}^{N_\sigma} |\phi_{j\sigma}(\vec{r},t)|^2 \tag{2.69}$$

In case that an external potential is very small, the system does not deviate much from the initial ground state. In such cases, it is often not necessary to obtain a full solution of time-dependent Schrödinger equation or TDKS equation for extracting small deviations from the ground state. This task can be accomplished using response theory. Linear-response theory is applicable for a weak perturbation such as long-wavevlength optical field

$$\delta v_{ext}(\vec{r},t) = -\xi \exp(i\omega t)\mathbf{z} \tag{2.70}$$

In the general case of the response of the ground-state to an arbitrary weak external field, the system's first-order response is characterized by the non-local susceptibility

$$\delta n_\sigma(\vec{r},t) = \sum_{\sigma'} \int dt' \int d\vec{r}' \chi_{\sigma\sigma'}[n_0](\vec{r},\vec{r}';t-t')\delta v_{ext\sigma'}(\vec{r}',t'). \tag{2.71}$$

This susceptibility χ is a functional of the ground-state density, $n_0(\vec{r})$. A similar equation describes the density response in the KS system:

$$\delta n_\sigma(\vec{r},t) = \sum_{\sigma'} \int dt' \int d\vec{r}' \chi_{s\sigma\sigma'}[n_0](\vec{r},\vec{r}';t-t')\delta v_{s\sigma'}(\vec{r}',t') \tag{2.72}$$

where χ_s is the KS response function, constructed from KS energies and orbitals:

$$\chi_{s\sigma\sigma'}(\vec{r},\vec{r}',\omega) = lim_{\eta\to 0_+}\delta_{\sigma\sigma'} \sum_q \left\{ \frac{\Phi_{q\sigma}(\vec{r})\ \Phi^*_{q\sigma'}(\vec{r}')}{\omega-\omega_q+i\eta} - \frac{\Phi^*_{q\sigma}(\vec{r})\ \Phi_{q\sigma'}(\vec{r}')}{\omega+\omega_q-i\eta} \right\} \tag{2.73}$$

where q is a double index, representing a transition from occupied KS orbital i to unoccupied KS orbital a, $\omega_{q\sigma} = \epsilon_{a\sigma} - \epsilon_{i\sigma}$, and $\Phi_{q\sigma}(\vec{r}) = \phi^*_{i\sigma}(\vec{r})\phi_{a\sigma}(\vec{r})$. Thus χ_s is completely determined by the ground state KS potential. It is the susceptibility of the non-interacting electrons sitting in the KS ground state potential.

The response χ of the system, as central equation of TDDFT linear response, is obtained as a Dyson-like equation [Petersilka, Gossmann, and Gross, 1996]

$$\begin{aligned}\chi_{\sigma\sigma'}(\vec{r},\vec{r}',\omega) &= \chi_{s\sigma\sigma'}(\vec{r},\vec{r}',\omega) + \sum_{\sigma_1\sigma_2}\int d\vec{r}_1 \int d\vec{r}_2 \chi_{s\sigma\sigma_1}(\vec{r},\vec{r}_1,\omega) \\ &\times \left(\frac{1}{|\vec{r}_1-\vec{r}_2|} + f_{XC\sigma_1\sigma_2}(\vec{r}_1,\vec{r}_2,\omega)\right)\chi_{\sigma_2\sigma'}(\vec{r}_2,\vec{r}',\omega) \qquad (2.74)\end{aligned}$$

where $f_{XC\sigma_1\sigma_2}(\vec{r}_1,\vec{r}_2,\omega)$ is Fourier transformation of XC kernerl function $f_{XC\sigma\sigma'}[n_0](\vec{r},\vec{r}',t-t')$

$$f_{XC\sigma\sigma'}[n_0](\vec{r},\vec{r}',t-t') = \left.\frac{\delta v_{XC\sigma}(\vec{r},t)}{\delta n_{\sigma'}(\vec{r}',t')}\right|_{n_0} \qquad (2.75)$$

The poles of linear susceptibility $\chi(\vec{r},\vec{r}',\omega)$ are excitation frequencies of the system. Casida [Casida, 1996] used well-known random-phase approximation (RPA) approach, to produce equations in which these poles of χ are found as the solution to an eigenvalue problem. The density change $\delta n_\sigma(\vec{r},t)$ can be expanded on the basis of KS transitions,

$$\delta n_\sigma(\vec{r},\omega) = \sum_q \left(P_{q\sigma}(\omega)\Phi^*_{q\sigma}(\vec{r}) + P_{\bar{q}\sigma}(\omega)\Phi_{q\sigma}(\vec{r})\right) \qquad (2.76)$$

where $q = (i,a)$ and $\bar{q} = (a,i)$. This representation yields two coupled matrix equations [Bauernschmitt and Ahlrichs, 1996]:

$$\left[\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}^* & \mathbf{A}^* \end{pmatrix} - \omega \begin{pmatrix} -\mathbb{1} & 0 \\ 0 & \mathbb{1} \end{pmatrix}\right]\begin{pmatrix} \mathbf{X} \\ \mathbf{Y} \end{pmatrix} = -\begin{pmatrix} \delta\mathbf{v} \\ \delta\mathbf{v}^* \end{pmatrix} \qquad (2.77)$$

where $A_{q\sigma q'\sigma'} = \delta_{qq'}\delta_{\sigma\sigma'}\omega_{q\sigma} + K_{q\sigma q'\sigma'}$, $B_{q\sigma q'\sigma'} = K_{q\sigma q'\sigma'}$, $X_{q\sigma} = P_{q\sigma}$, $Y_{q\sigma} = P_{\bar{q}\sigma}$ and

$$K_{q\sigma q'\sigma'}(\omega) = \int d\vec{r} \int d\vec{r}' \; \Phi_{q\sigma}(\vec{r}) f_{HXC\sigma\sigma'}(\vec{r},\vec{r}',\omega)\Phi^*_{q'\sigma'}(\vec{r}') \qquad (2.78)$$

with

$$\delta v_{q\sigma}(\omega) = \int d\vec{r} \; \Phi_{q\sigma}(\vec{r})\delta v_{ext}(\vec{r},\omega) \; . \qquad (2.79)$$

At an excitation energy, $\delta v = 0$ and choosing real KS orbitals and since $(\mathbf{A}-\mathbf{B})$ is positive definite, we get:

$$\sum_{q'\sigma'} \tilde{\Omega}_{q\sigma q'\sigma'}(\omega) \; \vec{a}_{q'\sigma'} = \omega^2 \; \vec{a}_{q\sigma}, \qquad (2.80)$$

where

$$\tilde{\Omega} = (\mathbf{A} - \mathbf{B})^{1/2}(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B})^{1/2} \ ,$$

or

$$\tilde{\Omega}_{q\sigma q'\sigma'}(\omega) = \omega_{q\sigma}^2 \delta_{qq'}\delta_{\sigma\sigma'} + 2\sqrt{\omega_{q\sigma}\omega_{q'\sigma'}} K_{q\sigma q'\sigma'} \ . \tag{2.81}$$

Oscillator strengths f_q may be calculated [Casida, 1996] from the normalized eigenvectors using

$$f_{q\sigma} = \frac{2}{3}\left(|\vec{x}^T \mathbf{S}^{-1/2}\vec{a}_q|^2 + |\vec{y}^T \mathbf{S}^{-1/2}\vec{a}_q|^2 + |\vec{z}^T \mathbf{S}^{-1/2}\vec{a}_q|^2 \right) \ , \tag{2.82}$$

where

$$S_{qq'} = \delta_{qq'}\delta_{\sigma\sigma'}/w_{q'} \tag{2.83}$$

Chapter 3

Accuracy of Force Fields for Energy Prediction in Solvated Small Peptides

A version of this chapter has been published as:
Jong, K.; Ansari N.; Grisanti, L.; Hassanali, A. Understanding the Quantum Mechanical Properties of Hydrogen Bonds in Solvated Biomolecules from Cluster Calculations. J. Mol. Liq. 2018, 263, 501 - 509.

In this chapter, I access the validation of the classical force fields using high-level quantum chemistry calculations and address the origin of their discrepancies.

3.1 Introduction

Hydrogen bonds (HBs) form one of the most important and studied interactions in the physical chemistry and chemical physics of biological systems [Jeffrey and Saenger, 1991; Grabowski, 2006; Pauling, 1967]. HBs are critical for stabilizing the secondary structure of proteins such as alpha helices and beta-sheets [Hubbard and Kamran Haider, 2001; Pace et al., 2014; Bordo and Argos, 1994], as well as in base-pairing of DNA and RNA Fonseca Guerra et al., 2000; Khakshoor et al., 2012; Nakashima, Fukuoka, and Saitou, 2013. HBs are also the most important interaction in liquid water [Stillinger, 1980; Henchman and Irudayam, 2010] and in its interactions with biological matter which is known to affect their structure, dynamics and function [Zheng and Merz, 1992; Sass, Schmid, and Grzesiek, 2007; Rawat and Biswas, 2014].

A particular class of biological systems that has received a lot of attention in the last decade are aggregates of proteins or polypeptides [Ross and Poirier, 2004; Horwich, 2002]. An important feature that characterizes some of these aggregates is the formation of a dense network of hydrogen bonds [Fitzpatrick et al., 2011; Pinotsi et al., 2016; Jong, Grisanti, and Hassanali, 2017]. The aggregation process typically happens in a condensed liquid phase and involves the coupling of numerous degrees of freedom involving both protein and water hydrogen bonds [Jong, Grisanti, and Hassanali, 2017]. Besides being implicated in neurodegenerative diseases such as Alzheimers and Parkinsons [Hardy and Selkoe, 2002; Rubinsztein, 2006], aggregates featuring hydrogen bond networks have also been found to have rather intruiging optical properties [Pinotsi et al., 2016; Grisanti et al., 2017a; Bhattacharya et al., 2017].

There have been numerous theoretical and computational studies examining the structural properties of protein aggregates [Carballo Pacheco and Strodel, 2016; Baftizadeh et al., 2012; Yang and Gao, 2015; Morriss-Andrews and Shea, 2015; Röhrig et al., 2006]. The vast majority of these studies, use classical empirical potentials where effects such as polarization and charge transfer are not taken into account. These effects have been shown to be important in several types of hydrogen bonding interactions [Tan et al., 2005; Chen and Dannenberg, 2011; Dang, 1998]. There have been a few studies using density-functional theory (DFT) based calculations of peptide aggregates showing the importance of quantum-mechanical effects along hydrogen bonds [Wang, Forsman, and Woodward, 2016]. These effects can play an important role in both the ground and excited electronic states.

In this chapter, we conduct a theoretical study on hydrogen bonding interactions of hybrid peptide-water clusters. These systems serve as model systems for peptide molecules and proteins in the condensed liquid phase since it is computationally prohibitive to examine quantum mechanical effects of solvated biomolecules consisting of 1000s of atoms. We use high-level quantum chemistry calculations to compare energetics obtained from several biological and water force fields for backbone-backbone and backbone-water hydrogen bonds. Interestingly, we find that for many of the biomolecular forcefields examined, pairwise interactions between molecular groups are in reasonable agreement with those from ab initio. However, three body interactions involving peptide-peptide or peptide-water hydrogen bonds are also significant and we show that this is rooted in mostly effects of polarization as shown by an energy decomposition analysis (EDA). The extent of the importance of three body interactions is sensitive to the type of hydrogen bond network either of the peptide or water.

This chapter is structured as follows. In Section 3.2, we begin by discussing the computational methods employed in this study. We then move on in Section 3.3 to show the comparison between the empirical potential energy surfaces (PES) to those obtained from ab initio. Section 3.4 explores the importance of many body interactions in peptide-peptide and peptide-water hydrogen bonds and in Section 3.6 we show EDA for the many body interactions. Finally, we end in Section 3.7 with our discussion and conclusions.

3.2 Computational Methods and Model Systems

We use a combination of empirical force fields and ab initio calculations in order to examine both the quality of the potentials as well as to understand the role of many body interactions. Most biological forcefields do not explicitly account for many-body electronic effects. For those like GROMOS and OPLS that are designed to reproduce experimental quantities, one could possibly argue that many-body effects are included in an effective way even though the functional form of the potentials is exclusively given by pairwise interactions. It is important however, to understand the physical origins of the limitations of current potentials and in particular, to disentangle the missing physics associated with the underlying interactions. We take first steps in this direction by studying the nature of the interaction potential for small model systems. We first describe the six different systems that were used to construct the potential energy surfaces. System **BB** (panel a) as shown in Figure 3.1) probes backbone-backbone interactions involving two amide-hydrogen bonds, while **NC** probes the strong $NH_3{}^+$-COO^- termini salt bridge. **BW1** and **BW2** interrogate backbone-water hydrogen bonds including N−H and C=O. **NW** and **CW** examine the termini-water interactions.

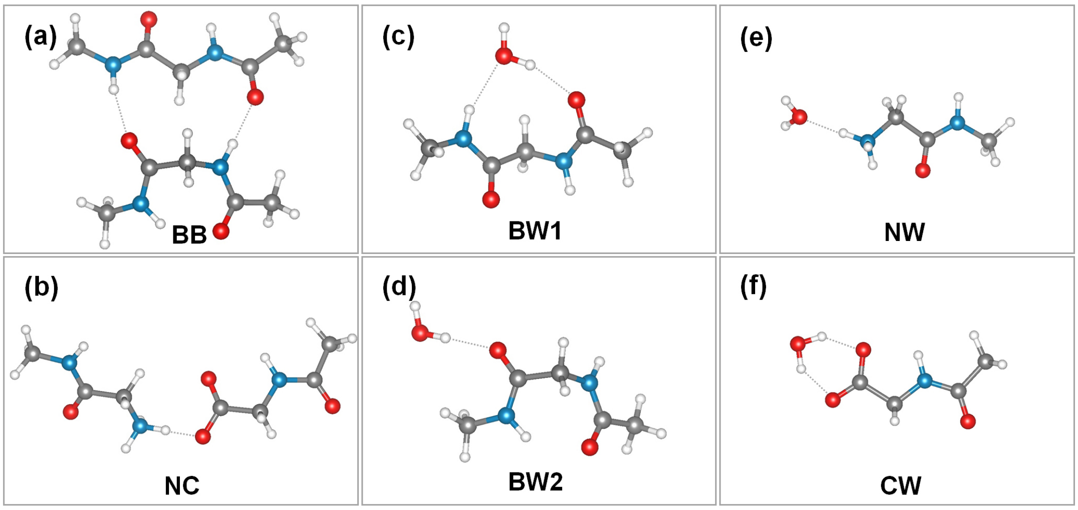


Figure 3.1: Benchmark structures of (a) backbone and backbone **BB**, (b) $NH_3{}^+$-terminus and COO^--terminus **NC**, (c) backbone (N−H and C=O) and water **BW1**, (d) backbone (C=O) and water **BW2**, (e) $NH_3{}^+$-terminus and water **NW**, and (f) COO^--terminus and water **CW**. O:red, N:blue, H:white, C:gray

All the quantum chemistry calculations were done using the Gaussian 09 code [Gaussian09 Revision E.01]. All the six structures in Figure 3.1 described earlier were first optimized at the Hartree-Fock level of theory using a 6-311++G** basis set. After the optimization, PES paths were constructed with GaussView [Dennington, Keith, and Millam, 2016] using a grid spacing of 0.2-2.0 Å which allowed for a part of the repulsive and attractive region to be probed. Once the optimized configurations of the monomers are fixed, we changed the distance between them to generate the grid of the potential energy surfaces. For **BB**, **BW1** and **CW** in the panels (a), (c) and (f) with two hydrogen bonds, we moved one monomer along the direction between C_α atom and C_α atom, C_α atom and the water oxygen atom, and the carbon atom of C-terminus and the water oxygen atom, respectively. In the cases of **NC**, **BW2** and **CW** with one hydrogen bond, the grid is generated along the direction between the donor and acceptor atoms of the hydrogen bond.

The ab initio potential energy constructed on configurations along the grid were then determined using the MP2 level of theory with aug-cc-pVDZ basis set. All the energies were corrected for basis set superposition error (BSSE) [Boys and Bernardi, 1970].

In order to examine how the most popular biological forcefields fare against ab initio calculations, the potential energy surfaces were also constructed for the six systems using several empirical potentials. In particular, PES were constructed using AMBER99SB-ILDN [Lindorff Larsen et al., 2010], CHARMM27 [MacKerell, Banavali, and Foloppe, 2000] and OPLSAA [Kaminski et al., 2001; Jacobson et al., 2002]. As indicated earlier, some of the panels studied involve peptide-water interactions: AMBER99SB-ILDN, was combined with TIP4P-D [Piana et al., 2015], CHARMM27 with TIP3P [Jorgensen et al., 1983], and finally OPLSAA with TIP4P [Jorgensen et al., 1983]. Besides the fixed-point charge models, we also conducted benchmarks with the polarizable AMOEBA forcefield with the AMOEBA14 polarizable atomic mutlipole water model [Shi et al., 2013; Ren and Ponder, 2004; Ren and Ponder, 2003].

The parameters for the empirical potential include the following. A cutoff of 16 Å was used the van-der-Waals interactions. Electrostatics were treated with particle mesh Ewald-Switch (PME-switch) method [Darden, York, and Pedersen, 1993; Essmann et al., 1995] with a Coulomb switching cutoff at 16 Å. The potential energy surfaces extracted from the empirical force fields were conducted in a unit cell with a large cubic box of length 300 Å. For the AMOEBA calculations, we set a cutoff of 16 Å for the van-der-Waals interactions and PME was employed with a real-space cutoff of 16 Å. The convergence threshold for the induced dipole moment was set to 10^{-10} debye per atom.

3.3 Hydrogen Bond Potential Energy Surfaces

We begin our discussion with the PES involving the backbone-backbone HB interactions (**BB**) and N- and C- termini HB interactions (**NC**). For clarity, we note that the term binding energy will be used in the ensuing discussion to be that inferred from the difference in energy between the minimum and the point where the two molecules are far away from each other. Overall, we will see that the binding energies inferred from the PES are reasonably well predicted from the empirical force fields when compared with the ab initio calculations. Figure 3.2a shows the PES involving the amide bond hydrogen bond interactions of two peptide chains. The MP2 binding energy for this panel is approximately 75 kJ/mol. It is rather encouraging to see that all the forcefields including AMOEBA are within 10% of the MP2 prediction. All the forcefields underestimate the binding energy by about 7 kJ/mol. This, however, falls within the variations in energies that are observed across different empirical force fields. Figure 3.2b shows the NH_3^+-COO^- termini PES which is essentially dominated by a Coulomb interaction and all the forcefields except AMOEBA and AMBER99SB-ILDN reproduce the energetics within 3%.

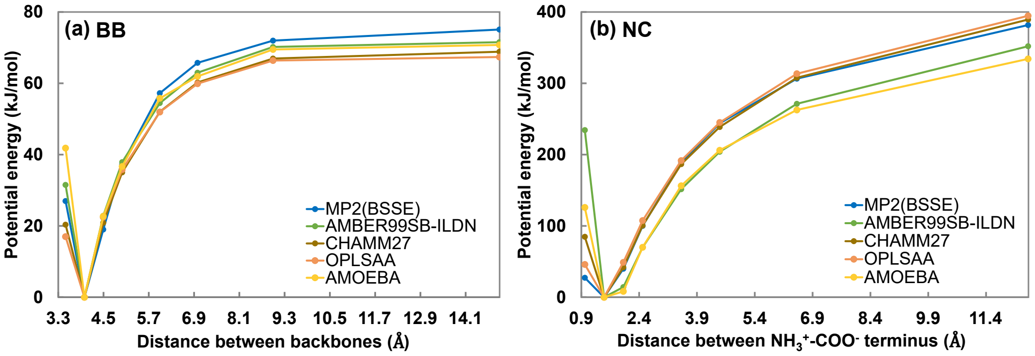


Figure 3.2: Benchmark of empirical force fields for energies between (a) two backbones and (b) $NH_3{}^+$-terminus and COO^--terminus

Figure 3.3 compares the PES for a single water binding to the peptide fragment (**BW1** and **BW2**) shown in Figure 3.1 panels c) and d). In **BW1** we see that the water molecule accepts one hydrogen bond from the N−H and donates one hydrogen bond to the carbonyl oxygen, resulting in a larger binding energy compared to **BW2**. In **BW1** (Figure 3.3a) we see that AMOEBA and CHARMM27/TIP3P perform the best and are within 2% of the MP2 energies. The combination of AMBER99SB-ILDN with TIP4P-D overbinds water in this configuration by about 10%, while OPLSAA/TIP4P underbinds water by about 6%. In the case of water interacting with the carbonyl group **BW2** (Figure 3.3b), AMBER99SB-ILDN with TIP4P-D and AMOEBA overbind water by respectively by 20% and 10%, while the CHARMM27 with TIP3P and OPLSAA with TIP4P appear to work better.

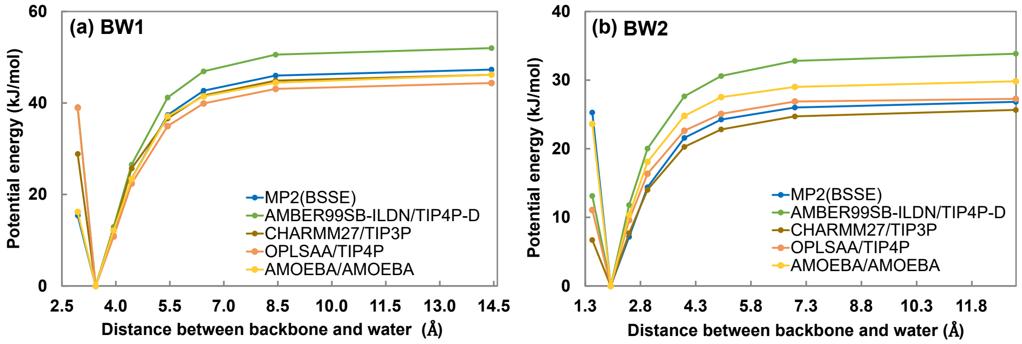


Figure 3.3: Benchmark of empirical force fields for energies in HB interactions between (a) backbone (N−H and C=O) and water and (b) backbone (C=O) and water.

Figure 3.4 illustrate the PES obtained from the $NH_3{}^+$-COO^- termini and water hydrogen bonding interactions **NW** and **CW** (see Figure 3.1). For configuration **NW** involving NH_3^+-terminus hydrogen bonding with water, the AMBER99SB-ILDN, CHARMM27, and OPLSAA forcefields overbind the water by 15%, while the AMOEBA forcefield underbind the water by 5%. On the other hand, in configuration **CW** involving the COO^--terminus - water HBs, all the forcefields overbind the water by about 7-36%. The force fields are shown to work better for the NH_3^+-terminus with water molecules rather than the COO^--terminus with water molecules.

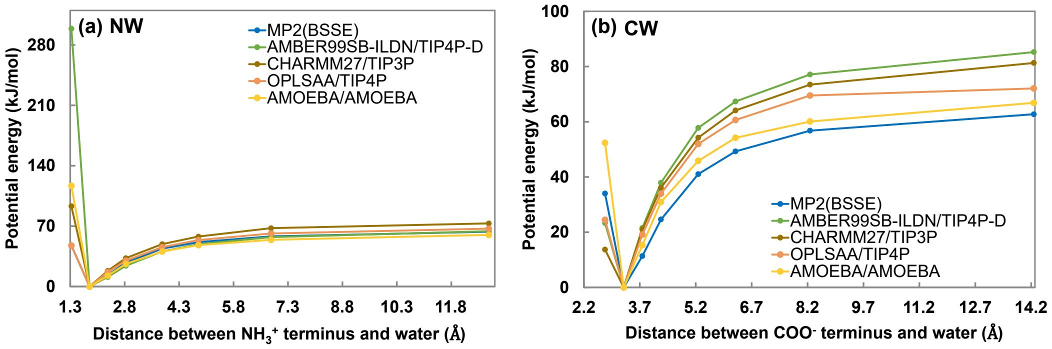


Figure 3.4: Benchmark of empirical force fields for energies between (a) NH_3^+-terminus and water and (b) COO^--terminus and water.

The preceding results indicate that there is no particular force field and water model combination that outperforms the rest. In order to see this more clearly and to also obtain a global assessment of the force fields, we define C^{FF} as the square of the percentage relative error in the binding energy of the force field with respect to the MP2 binding energy :

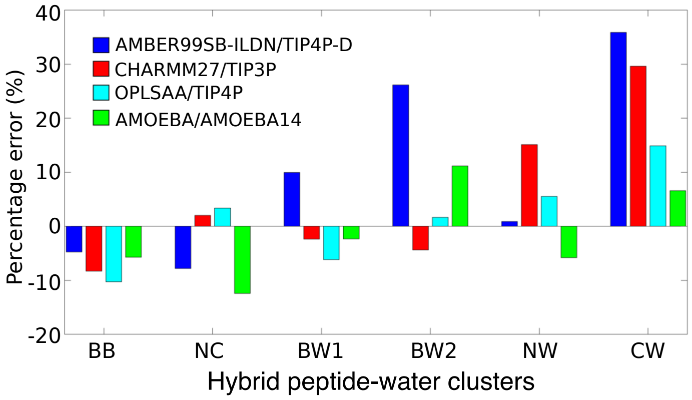
$$C^{FF} = \sum_i (\delta \Delta E_i)^2 = \sum_i \left(\frac{\Delta E_i^{FF} - \Delta E_i^{MP2}}{\Delta E_i^{MP2}} \times 100 \right)^2 \tag{3.1}$$

where the index FF indicates the force field, the index i belongs to the six benchmark structures in Figure 3.1 and ΔE is the binding energy. The global behaviour of the different force fields expressed by δE_i for each system i is summarized in the left panel of Figure 3.5, collecting the deviation from the reference results (MP2) as a percentage, while the total errors C^{FF} are collected in the right panel of Figure 3.5. There is no particular FF that can be identified to work across all types of interactions examined here. On the other hand, we can also identify specific interactions that have a systematic error across all force fields. This is observed in **BB**, where energies are underestimated and **CW** where the binding energies are overestimated. Interestingly, configuration **CW**, including water and COO^--terminus, also turns out to give largest relative errors. Based on this small subset of interactions examined here, we see that the AMOEBA and OPLSAA force fields appear to perform the best.

All the binding energies for the structures studied are shown in Table 3.1.

3.4 Many-Body Interactions in Hydrogen Bonds

In the previous section, we have shown that the empirical forcefields generally account for the correct evaluation of hydrogen bonding energetics in comparison to results from the quantum chemistry calculations. On the other hand, the agreement is not



	C^{FF}
AMBER99SB-ILDN	2152.3
CHARMM27	1202.9
OPLSAA	409.9
AMOEBA	394.4

Figure 3.5: (left) Percentage error bars of binding energies estimated by AMBER99SB-ILDN/TIP4P-D, CHARMM27/TIP3P, OPLSAA/TIP4P and AMOEBA/AMOEBA14 relative to MP2 binding energy for the benchmark systems in Figure 3.1 and (right) Total sum of the percentage error square of the binding energies of the force fields of AMBER99SB-ILDN with TIP4P-D, CHARMM27 with TIP3P, OPLSAA with TIP4P and AMOEBA with AMOEBA14 with respect to the MP2 binding energy.

Table 3.1: Binding energies of the benchmark structures calculated by the MP2(BSSE) and the force fields of AMBER99SB-ILDN with TIP4P-D, CHARMM27 with TIP3P, OPLSAA with TIP4P and AMOEBA with AMOEBA14 (kJ/mol)

	MP2(BSSE)	AMBER99SB-ILDN	CHARMM27	OPLSAA	AMOEBA
BB	75.09	71.53	68.85	67.38	70.77
NC	381.82	351.97	389.51	394.73	334.30
BW1	47.29	51.99	46.17	44.36	46.19
BW2	26.83	33.84	25.65	27.27	29.82
NW	63.57	64.13	73.18	67.07	59.88
CW	62.73	85.23	81.30	72.07	66.85

quantitative and perhaps more importantly, relies on an examination of pairwise interactions only. One of the missing underlying physics in empirical potentials is the lack of many body electronic effects. There have been several theoretical studies showing the importance of three-body interactions in the hydrogen bonding of water as well as the hydrogen bonding between the protein and water [Xantheas, 2000; Cobar et al., 2012; Best et al., 2012; Qi, Leverentz, and Truhlar, 2013; Bianco et al., 2017a; Bianco et al., 2017b; Davtyan et al., 2012]. In this next section, we examine the many-body interaction contributions coming from peptide-peptide and peptide-water interactions.

For a system of n particles, which could include molecules, ions or peptide chains, the n-particle interaction energy can be separated out into various contributions coming from a relaxation term, two particle, three particle and so-forth. Typically the two-particle interaction dominates the total interaction energy. The total interaction energy as a difference between total energy ($E(1,2,3,...,n)$) and the sum of the energy of the individual isolated monomer ($E^0(i)$; i=1,2,...,n) can be expressed in the following manner:

$$\Delta E = E(1,2,3,...,n) - \sum_{i=1}^{n} E^0(i) \tag{3.2}$$

$$\left.\begin{aligned}
&\equiv \sum_{i=1}^{n} E(i) - \sum_{i=1}^{n} E^0(i) \quad \text{Relaxation} \\
&+ \sum_{i=1}^{n-1}\sum_{j>i}^{n} \Delta^2 E(ij) \quad \text{two-body} \\
&+ \sum_{i=1}^{n-2}\sum_{j>i}^{n-1}\sum_{k>j}^{n} \Delta^3 E(ijk) \quad \text{three-body} \\
&+ \sum_{i=1}^{n-3}\sum_{j>i}^{n-2}\sum_{k>j}^{n-1}\sum_{\ell>k}^{n} \Delta^4 E(ijk\ell) \quad \text{four-body} \\
&+ \cdots + \Delta^n E(1,2,3,4,...,n) \quad n\text{-body}
\end{aligned}\right\} \tag{3.3}$$

where $\Delta^2 E(ij)$ is two-body term, $\Delta^3 E(ijk)$ is three-body term, ..., and $\Delta^n E(1,2,3,4,...,n)$ is n-body term.

In this work, we don't consider the relaxation term in Equation 3.3. The two, three and four-body contributions can be computed using the following equations:

$$\Delta^2 E(ij) = E(ij) - \{E(i) + E(j)\} \tag{3.4}$$

$$\begin{aligned}\Delta^3 E(ijk) = E(ijk) &- \{E(i) + E(j) + E(k)\} \\ &- \{\Delta^2 E(ij) + \Delta^2 E(ik) + \Delta^2 E(jk)\}\end{aligned} \tag{3.5}$$

$$\begin{aligned}\Delta^4 E(ijk\ell) = E(ijk\ell) &- \{E(i) + E(j) + E(k) + E(\ell)\} \\ &- \{\Delta^2 E(ij) + \Delta^2 E(ik) + \Delta^2 E(i\ell) + \Delta^2 E(jk) + \Delta^2 E(j\ell) + \Delta^2 E(k\ell)\} \\ &- \{\Delta^3 E(ijk) + \Delta^3 E(ij\ell) + \Delta^3 E(ik\ell) + \Delta^3 E(jk\ell)\}\end{aligned} \tag{3.6}$$

To understand the contribution of many-body interactions, a series of different benchmark systems were used. Figure 3.6 shows the eight configurations employed to investigate this part. These include interactions between 3 peptide fragments (involving backbone-backbone hydrogen bonds **BBB** or $NH_3{}^+ - COO^-$ termini salt-bridges **NNC**); a single peptide fragment with 2 water molecules interacting with backbone groups in different arrangements (**BWW1**, **BWW2**, **BWW3**); a single peptide fragment interacting with two water molecules at the NH_3^+-terminus **NWW** or at the COO^--terminus **CWW**; a single water molecule bridging two peptide ends **NCW**.

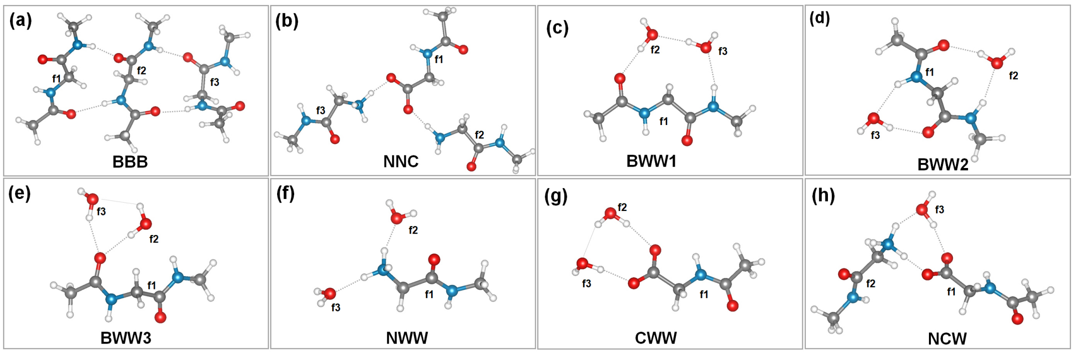


Figure 3.6: Structures for the calculations: a) 3 peptide fragments involving backbone-backbone hydrogen bonds **BBB**, b) 3 fragments involving $NH_3{}^+{-}COO^-$ termini salt-bridges **NNC**, c) a single peptide fragment with 2 water molecules forming a water chain linking the N−H to the C=O groups **BWW1**, d) a single peptide fragment interacting with 2 water molecules but on opposite sides of the peptide fragment **BWW2**, e) a single fragment with 2 water molecules donating two hydrogen bonds to the same C=O group, of which one also donating to a N−H **BWW3**, f) a single peptide fragment where two water molecules accept hydrogen bonds from the NH_3^+-terminus **NWW**, g) a single peptide chain where two water molecules donate hydrogen bonds to the COO^--terminus **CWW**, h) a single water molecule forming a water bridge between two peptide fragments **NCW**

The structures illustrated in Figure 3.6, while not including a realistic environment that would be experienced by a protein or peptide chain during protein folding or protein aggregation [Hartl, Bracher, and Hayer-Hartl, 2011; Ross and Poirier, 2004], provides a first step in trying to understand the importance of many-body interactions in the solvation of biomolecules and the interactions between peptide fragments. Using Equations 3.4 and 3.5, the pairwise and three-body interactions were determined for panels a)-h).

3.5 Many-Body Interactions for Peptide-Water Clusters

We first examine the many body interactions for the three fragments involving the backbone-backbone and the termini-termini hydrogen bonds **BBB** and **NNC** shown in Figure 3.6. In Table 3.2, we show the relative pairwise and three-body interactions. We see for hydrogen bonds involving the peptide backbones, the three-body interactions contributes 4.98 kJ/mol which is about 4% of the total interaction energy. **NNC**, involving the $NH_3{}^+$-COO^- termini has a larger contribution to the three-body interaction with respect to **BBB**, and perhaps more interestingly, is also repulsive. Repulsive three-body interactions have been already observed in the hydrogen bonding network of the small water clusters and has been attributed to the presence of water molecules playing the role as double donors or acceptors [Xantheas, 2000; Cobar et al., 2012; Guevara-Vela et al., 2016]. In the case of **NNC**, the carboxylate group plays the role of double acceptor. Later in the manuscript, the energetic origins of this will be determined using an energy decomposition analysis.

Table 3.2: Two-body and three-body interaction energies of **BBB** and **NNC** benchmark model systems calculated by MP2 (BSSE) (kJ/mol)

	$\Delta^2 E$	$\Delta^3 E$
BBB	-119.09	-4.98
NNC	-716.78	64.22

Fragments **BWW1**, **BWW2** and **BWW3** were chosen to understand the importance of three body interactions between peptide-water hydrogen bonds. In Table 3.3, we show the three-body interactions of configuration **BWW1**, **BWW2**, **BWW3**. A comparison of the 3-body interactions from such structures illustrates the importance of water-networks around the peptide fragment. Specifically, we see that in **BWW1** consisting of a directed water wire of length 3 linking the N−H to the C=O group on the same side, features the three-body interaction that is roughly about a factor of 4 larger than the **BWW2** and **BWW3**. As we will see later, this originates from the fact that the orientation of the hydrogen bonds allows for larger mutual polarization.

The strength of the hydrogen bonds involving the backbone and water are weaker than those between the termini and water. This effect however, originates from essentially Coulomb interactions which is seen by the fact that the three-body interactions of all of **NWW**–**NCW** (panels f) to h)) are repulsive (Table 3.3). The three-body interaction of **NWW**, involving the NH_3^+-terminus and water, is slightly larger by about 2 kJ/mol compared to **CWW** of the COO^--terminus and water. **NCW** which consists of a water wire hydrogen bonding to the $NH_3{}^+$ and COO^- termini of two different fragments incurs a repulsive 3-body interaction that is roughly 18 kJ/mol. Again, each of these fragments involve subtle changes in the network topology which affect the many body polarization. In **CWW** for example, water f1 accepts and donates a single hydrogen bond while f2 is a double donor. On the other hand, f2 and f3 in **NWW** are both single acceptors. In the case of **NCW**, f3 is similar to **CWW** but now involves the $NH_3{}^+$-terminus playing the role of a double donor.

Table 3.3: Two-body and three-body interaction energies of backbone with two waters in three different positions calculated by MP2 (BSSE) (kJ/mol)

	$\Delta^2 E$	$\Delta^3 E$
BWW1	-68.74	-12.81
BWW2	-79.69	-3.02
BWW3	-58.53	-2.14

Table 3.4: Two-body and three-body interaction energies of NH_3^+ and COO^- termini backbones with two water and NC termini backbone with one water calculated by MP2 (BSSE) (kJ/mol)

	$\Delta^2 E$	$\Delta^3 E$
NWW	-146.25	5.30
CWW	-109.30	2.98
NCW	-584.44	18.80

3.6 Energy Decomposition Analysis

In this section, we are to investigate the origin of the contribution of the many body interactions to the force field using the energy decomposition analysis. It is clear that 3-body interactions are significant in the context of biological matter interactions involving hydrogen bonds. The origins of many body electronic effects involves quantum mechanical phenomena such as polarization, electron correlation, exchange and charge transfer which has been studied extensively in numerous other systems [Phipps et al., 2016; Phipps et al., 2015; Su and Li, 2009].

In order to understand the physical origin of the interactions in the fragments studied here, we employed the localized molecular orbital energy decomposition analysis implemented in GAMESS [Su and Li, 2009; Schmidt et al., 1993] package for electronic structure with the MP2 method and cc-pVDZ, basis set. Within the framework of EDA, the interaction energy, ΔE, between the monomers A forming the supermolecule X, is separated into several components based on various different contributions:

$$\Delta E = E_X - \sum_A E_A = \Delta E_{ele} + \Delta E_{ex} + \Delta E_{rep} + \Delta E_{pol} + \Delta E_{disp} \tag{3.7}$$

where E_X is the total energy of the supermolecule X, E_A is the energy of the monomer A, E_{ele} is electrostatic energy, E_{ex} is exchange energy, E_{rep} is repulsion energy, E_{pol} is polarization energy and E_{disp} is dispersion energy.

In order to obtain all the energy components of the three-body interaction ($\Delta^3 E_{ele}$, $\Delta^3 E_{ex}$, etc.), we first applied the EDA method to trimer and all dimers, and then subtracted all the dimer decomposed energy terms from trimer separately using Equation 3.5.

We start by illustrating the EDA results of the backbone-backbone and termini-termini hydrogen bonding interactions (Figure 3.6 a) and b)). In Table 3.5, we show all the decomposed energy terms of three-body interactions. The EDA analysis clearly demonstrates that the three-body interactions of system **BBB** and **NNC** essentially come from the polarization interaction terms. In both cases, the polarization accounts for large contributions of 79% and 81% of the total three-body interactions respectively, indicating that the orbitals undergo significant changes when three monomers form the trimer by hydrogen bonding interactions. For **NNC**, which consists of positive (NH_3^+) and negative (COO^-) monomers, electrostatic and dispersion energies also contribute about 14.2% and 10.9% respectively to the total 3-body interaction.

Table 3.5: EDA of three-body interaction of panel a) and b) (kJ/mol)

	$\Delta^3 E_{ele}$	$\Delta^3 E_{ex}$	$\Delta^3 E_{rep}$	$\Delta^3 E_{pol}$	$\Delta^3 E_{disp}$	$\Delta^3 E$
BBB	-0.67	0.21	-0.38	-3.97	-0.21	-4.98
NNC	9.12	2.01	-6.23	52.30	7.03	64.22

Next, we consider the EDA results for the peptide-water hydrogen bonding interactions of Figure 3.6 panels c)-e), **BWW1–3** which are listed in Table 3.6. It is clear that polarization contribution dominates and that this is the origin of the enhanced effect for fragment **BWW1**. In this fragment, the water molecules are oriented in a fashion forming a directed loop of hydrogen bonds which faciliates mutual polarization.

Table 3.6: EDA of three-body interaction of panel c)-e) (kJ/mol)

	$\Delta^3 E_{ele}$	$\Delta^3 E_{ex}$	$\Delta^3 E_{rep}$	$\Delta^3 E_{pol}$	$\Delta^3 E_{disp}$	$\Delta^3 E$
BWW1	-1.46	0.63	-1.17	-9.46	-1.34	-12.80
BWW2	-0.38	0.04	-0.04	-2.34	-0.33	-3.01
BWW3	-0.12	0.17	-0.42	-1.21	-0.542	-2.09

Finally, EDA results for interactions between N- and C- termini and water molecules (Figure 3.6 panels f)-h), **NWW**, **CWW** and **CNW**) are shown in Table 3.7. The polarization contributions are 4.06 kJ/mol, 3.01 kJ/mol and 15.90 kJ/mol, respectively. This corresponds to about 77%, 100% and 84% of the total three-body interactions. Therefore, we could easily conclude that the physical origin of the repulsive three-body interactions term of the hydrogen bonding interactions between N- and C- termini and water molecules originates from polarization.

Table 3.7: EDA of three-body interaction of panel f)-h) (kJ/mol)

	$\Delta^3 E_{ele}$	$\Delta^3 E_{ex}$	$\Delta^3 E_{rep}$	$\Delta^3 E_{pol}$	$\Delta^3 E_{disp}$	$\Delta^3 E$
NWW	0.54	-0.04	0.38	4.06	0.38	5.27
CWW	0.33	0.79	-1.05	3.01	-0.12	3.01
CNW	1.97	1.46	-1.59	15.90	1.05	18.83

3.7 Conclusion

Hydrogen bond interactions are ubiquitous in biological systems and continue to pose interesting challenges from both experimental and theoretical fronts. A particular class that has attracted our attention recently is protein aggregates relevant for neurodegenerative diseases. There are numerous theoretical studies using classical empirical potentials, studying the mechanisms associated with the formation of these aggregates as well as their structural and dynamical properties. The importance of quantum mechanical effects in these systems still remains poorly undertood.

In this chapter, we take first steps in examining the importance of quantum mechanical effects in peptide-water and peptide-peptide hydrogen bonding. We do this on small hybrid peptide-water clusters that serve as model systems for important interactions in soft matter systems in the condensed phase. We first compare potential energy surfaces obtained from high level ab initio calculations to several classical biological forcefields and respective water models. For the most part, pairwise interactions seem to be reasonably captured by the force field, at least for the fragments examined.

We focused on extracting the many body electronic contributions from cluster calculations optimized at T=0K. In principle, one could also examine how these contributions change as a function of finite temperature. This poses additional challenges in that the sampling of the thermodynamically stable states would need to be adequately sampled which involves both the molecule as well as its liquid environment. In addition, doing high level quantum chemistry calculations for molecules with a large solvation environment is also computationally prohibitive. This is beyond the scope of the current study but would be an interesting topic to explore in the future.

Most of the biological force fields do not explicitly include three body interactions and beyond. We examined the role of these effects and found that three body interactions involving peptide-peptide and peptide-water can contribute about 4-19% of the total interaction energy. Although this is smaller than the effects in bulk water, it is not insignificant. We show further, using energy decomposition analysis that the origin of these effects is mostly dominated by many-body polarization.

Chapter 4

Hydrogen Bond Networks and Hydrophobic Effects in the Amyloid β_{30-35} Chain in Water

A version of this chapter has been published as:
Jong, K.; Grisanti, L.; Hassanali, A. Hydrogen Bond Networks and Hydrophobic Effects in the Amyloid β_{30-35} Chain in Water: A Molecular Dynamics Study. J. Chem. Inf. Model. 2017, 57, 15481562.

In this chapter, we explore the conformational landscape of Amyloid β_{30-35} using well-tempered metadynamics and provide an understanding of the coupling between protein and surrounding water.

4.1 Introduction

There is currently an active effort from both experimental and theoretical fronts to understand the physical and chemical processes underlying protein fibril formation as well as of the early stages of aggregation [Nasica Labouze et al., 2015; Paravastu et al., 2008; Kirkitadze, Condron, and Teplow, 2001; Hou et al., 2004; Zhang et al., 2000; Baumketner and Shea, 2007; Petkova et al., 2002; Khemtémourian et al., 2011; Xu et al., 2013; Shanmugam and Polavarapu, 2004; Wei and Shea, 2006; Rosenman et al., 2013; Bitan et al., 2003; Riccardi, Nguyen, and Stock, 2012; Carballo Pacheco and Strodel, 2016; Makin et al., 2005; Pellarin and Caflisch, 2006; Fawzi et al., 2007; Auer, Dobson, and Vendruscolo, 2007; Baftizadeh et al., 2012; Barducci et al., 2013; Sanfelice et al., 2014; Qiang et al., 2012; Lührs et al., 2005]. These fibrils are three-dimensional architectures resulting from the aggregation of misfolded proteins. One of the most studied fibrils in this regard, are those that develop from the Amyloid β(Aβ) protein, a sequence made up of 39-43 amino acids. Besides serving as excellent model systems to understand physical and chemical processes in a biological context, the Amyloid fibrils have been implicated in neurodegenerative diseases such as Alzheimers and Parkinsons [Hardy and Selkoe, 2002]. Apart from the eventual plaques that form, there is a growing appreciation that oligomers such as dimers and trimers can also play a critical role in the pathology of Alzheimers disease [Shankar et al., 2008]. There is thus a lot of interest to understand the conformational heterogeneity of Aβ since the structural disorder that features it during the early stages of aggregation may have important implications on its subsequent dynamical evolution [Uversky, Oldfield, and Dunker, 2008]. Besides the Aβ protein fibrils, there have been numerous experimental studies showing that fibril-like structures can result from the aggregation of different types of poly-peptide chains typically made up of hydrophobic amino-acids [Gething and Sambrook, 1992; Gsponer and Vendruscolo, 2006; Sugita and Okamoto, 1999; Nguyen and Hall, 2005; Roberts, 2007]. Similar to Aβ, these fibrils are characterized by a dense network of hydrogen bonds between the polar backbone making beta-sheet secondary structures. The exact details of how hydrogen bonding and hydrophobic interactions couple with each other remains an open question. A very recent experimental study for example, used solid-state NMR to show that the abundance of hydrophobic amino acids in Aβ1-42 results in a dense packing of alkylic side chains in the plain perpendicular to the fibril axis [Wälti et al., 2016]. Much less is known however, about how polar and non-polar interactions within the protein and the surrounding solvent couple with each other before aggregation has even started.

Recently, some of us have been involved in trying to understand a rather peculiar and anomalous experimental observation, namely that amyloid fibrils are capable of fluorescing in the absence of aromatic residues [Pinotsi et al., 2016; Grisanti et al., 2017a]. Using state-of-the-art first principles simulations on small model amyloid crystals, we find that salt-bridges between the N and C termini are characterized by strong hydrogen bonds where proton transfer leads to the formation of both zwitterionic and methyl-capped states in the fibril. This feature is tuned by the surrounding hydrogen bond network involving the proximity of water and hydrophobic amino acids [Grisanti et al., 2017a]. An obvious limitation of the tools deployed in this earlier study is that both the model systems and time-scales did not allow us to explore the larger scale conformational fluctuations associated with the hydrogen bond network.

Providing a detailed microscopic description of both the structural landscape and molecular interactions for the monomer with the surrounding solvent bath is key to understanding how subsequent aggregation proceeds. There have thus been numerous theoretical studies examining the conformational fluctuations of both the Aβ monomer as well as smaller segments of it [Lapidus, 2013; Xi, Li, and Wang, 2012; Granata et al., 2015; Straub and Thirumalai, 2011; Liu et al., 2004; Chebaro, Mousseau, and Derreumaux, 2009; Gnanakaran, Nussinov, and García, 2006; Yan and Wang, 2006]. In many of these studies, the model monomer peptides are terminated with two methyl groups and this by construction, hinders the possible formation of strong hydrogen bonds between the N-C termini and possibly between the termini and the backbone.

In this chapter, we use classical molecular dynamics simulations to explore the protein and water networks of the hydrated C-terminal hydrophobic peptide of Aβ_{30-35}. This chain formed a segment of one of the model crystals [Pinotsi et al., 2016; Grisanti et al., 2017a]. Furthermore, Liu et al. [Liu et al., 2004] have shown that Aβ_{30-35} plays a critical role in the aggregation process by including short anti-parallel strands in the surrounding residues, which in turn could promote the fibril formation of full-length Aβ. Using well-tempered metadynamics, we explore the free energy landscape of this chain in its zwitterionic, $NH_3{}^+ \cdot AIIGLM \cdot COO^-$, form. We also perform simulations of a variant of the zwitterionic system where the termini ends are instead capped with methyl groups $CH_3CONH \cdot$ AIIGLM$\cdot CONHCH_3$(methyl-capped system) to examine the importance of the termini. The change in the termini interactions leads to significant changes in the conformational landscape of zwitterionic and methyl-capped systems. We find that the disorder in the conformational landscape of zwitterionic system is driven by a diversity of different interactions such as electrostatic interaction between termini, polar interactions of the backbone, van-der-Waals interactions of the hydrophobic side chains and also hydrogen bonds between the termini and the backbone.

We also investigate the topological properties of the hydrogen bond network surrounding the peptide as probed through the reorganization of water wires connecting N-donor (N$-$H and $NH_3{}^+$) and carbonyl (C=O) groups. Similar types of analysis have been conducted to understand the network structures of bulk water, aqueous solutions and also in understanding the mechanisms of proton transfer in water [Lee et al., 2015; Kim et al., 2014; Giberti et al., 2014; Hassanali et al., 2013; Agmon et al., 2016]. In addition, Thirumalai and co-workers have also shown the formation of single-file water wires between two sheets of the yeast prion protein although these have been investigated mostly in a qualitative manner [Thirumalai, Reddy, and Straub, 2012; Reddy, Straub, and Thirumalai, 2010]. Our analysis reveal subtle differences such as the shortening or lengthening of the wires connecting different parts of the peptide during the conformational fluctuations, providing new insights into the coupling of protein and water motions for this system. The changes in these wires often include those involving the N and C termini in the zwitterionic system and provide a rationale for understanding the changes in the structural disorder when one moves to the methyl-capped system.

This chapter is organized as follows. We begin in Section 4.2 with details of the computational protocols employed in this work. We then move on in Section 4.3.1 to discuss our results on the free energy landscapes of the zwitterionic and methyl-capped systems as revealed by our metadynamics simulations. Here, we also discuss the coupling of polar and non-polar interactions in stabilizing the disordered structures observed in the zwitterionic system. In Section 4.3.2, we discuss our analysis of the reorganization of hydrogen bond networks around the peptide. Finally, we end in Section 4.4 with some conclusions and possible future directions.

4.2 Methods

Below we highlight the molecular dynamics simulation details including a brief overview of the theory behind metadynamics and also how our collective variables were chosen for the metadynamics simulations.

4.2.1 Simulation Details

The zwitterionic system consists of the 6 amino acid sequence AIIGLM. The starting structure for this hexapeptide was extracted from one of the 8 chains forming the model amyloid crystal structure with PDB code 2Y3J [Colletier et al., 2011]. The zwitterionic system has termini with NH_3^+ and COO^- groups, as found at pH=7 solvent conditions. The methyl-capped structure is obtained by modifying the termini of the zwitterionic one by capping the two ends with methyl-amide and acetyl groups. Zwitterionic and methyl-capped systems are solvated with 7052 and 7098 water molecules respectively. The two systems were subsequently run for 20 ns in the NPT ensemble using the Berendsen barostat [Berendsen et al., 1984] at 1 atm and using a time constant of 0.5 ps. The final cubic box lengths used for zwitterionic and methyl-capped systems were then 6.0 nm. All simulations were thermostated at 300K with a Nosé-Hoover thermostat [Nosé, 1984; Hoover, 1985] using a time constant of 0.1 ps in the NVT ensemble. A non-bonded pair list was produced using a cut-off radius of 1.4 nm. The short-range non-bonded pairwise interactions were evaluated by using a shifted Lennard-Jones potential with a cut-off length at 0.9 nm while the long-range electrostatic interactions were handled using the Particle Mesh Ewald-Switch (PME-switch) method [Darden, York, and Pedersen, 1993; Essmann et al., 1995] with a Coulomb switching cut-off at 1.2 nm. A long range dispersion correction was applied to both energy and pressure for the van-der-Waals cutoff. A time step of 0.5fs was used. In our simulations, the water molecules were kept rigid while no bond or angle constraints were put on the protein. This choice was motivated by two aspects. Firstly, previous studies have suggested that not having any constraints on the angles is important for ensuring protein flexibility [Van Gunsteren and Karplus, 1982; Toxvaerd, 1987; Pronk et al., 2013]. Secondly, our earlier ab initio studies have pointed to the role of proton transfer along the N and C termini [Pinotsi et al., 2016; Grisanti et al., 2017a]. It is clearly beyond the scope of the current potential to capture these types of effects, but allowing for some bond flexibility goes along the lines of having a more realistic hydrogen bond between the N and C termini. Bond and angle flexibility of the protein also means that in order to obtain better conservation of energy a smaller timestep is needed. In previous ab-initio simulations of liquid water some of us have shown that one can probably get away with a larger timestep especially for structural properties [Gasparotto, Hassanali, and Ceriotti, 2016]. However, in this work, we decided to stay on the more conservative side and use a smaller timestep of 0.5fs.

For all our simulations, the OPLS-AA force field and TIP4P water model were employed. In a recent study, Smith and co-workers did a systematic comparison of a wide variety of forcefields including OPLS-AA, AMBER, CHARMM and GROMOS on the conformational landscape of Aβ_{21-30} in water compared to available experimental data [Smith et al., 2015]. Their conclusion was that the OPLS-AA forcefield suppressed the formation of helical structures consistent with the experiments and hence recommended the use of either OPLS-AA or GROMOS for their system. Since we are also doing simulations on a segment of the amyloid protein, we decided to conduct our simulations with OPLS-AA and TIP4P.

4.2.2 Free Energy Calculations

Free energy surfaces were explored using well-tempered metadynamics [Barducci, Bussi, and Parrinello, 2008] simulations. Here, we briefly summarize the key theoretical concepts behind the methodology. A history-dependent bias potential, $V_b(\vec{s},t)$(where the quantity $\vec{s}$ is the vector of collective variables), is introduced to enhance the sampling of the free energy surface in the basis of a predefined set of collective variables. $V_b(\vec{s},t)$ is a sum of Gaussian hills with the width $\delta\vec{s}$ and the height w, centered at the values of the collective variables that have already been visited and deposited at the time interval τ_G,

$$V_b(\vec{s},t) = \mathrm{w} \sum_{t=\tau_G,2\tau_G,\cdots} exp\left(-\frac{|\vec{s}-\vec{s}(t)|^2}{2\delta\vec{s}^2}\right) \tag{4.1}$$

Unlike standard metadynamics, the height of Gaussian hills added is modified according to the relationship

$$\mathrm{w} = \omega e^{-[V_b(\vec{s},t)/\Delta T]}\tau_G \tag{4.2}$$

where ω is the initial bias deposition rate with dimension of energy rate and ΔT is a tunable temperature-like parameter that controls how quickly w reduces as the wells are filled. The parameters, ω and ΔT, are chosen to achieve the best efficiency. The free energy can then be computed with the following expression

$$F(\vec{s}) = -\frac{T+\Delta T}{\Delta T}V_b(\vec{s},t\to\infty) \tag{4.3}$$

In this work, we have biased two collective variables based on our analysis of some relatively short unbiased simulations of the zwitterionic system. Over the course of a 80ns MD simulation, we observed a single event where N and C termini came into close proximity forming a strong salt-bridge for a couple of nanoseconds before dissociating away from each other. Therefore we have chosen the end-to-end distance between the N and C termini as one of the collective variables for the well-tempered metadynamics. As we will see later in the manuscript, backbone hydrogen bonding and side-chain hydrophobic packing also plays an important role in the structural disorder we observe in the peptide conformations. In order to enhance the fluctuations in these coordinates, we also biased the radius of gyration as another collective variable as has been done in previous studies [Barducci et al., 2013; Sanfelice et al., 2014].

As we will see later, while we have only biased two collective variables namely, the end-to-end distance and the radius of gyration, we also construct free energy surfaces in other variables that are not biased. In order to do this, re-weighted free energy profiles along unbiased coordinates were constructed with a recent re-weighting algorithm [Tiwary and Parrinello, 2015]. Some examples of collective variables that we found to be important to understand the structural disorder in zwitterionic and methyl-capped systems include, the extent of hydrogen bonding between the backbone amide groups, van-der-Waals packing of the hydrophobic side-chains and the exposure of both the backbone and the side-chains to the surrounding water. In all these cases, contact maps were defined by a switching function

$$C = \sum_{i \in A} \sum_{j \in B} s_{ij} = \sum_{i \in A} \sum_{j \in B} \frac{1 - (\frac{r_{ij}}{r_0})^n}{1 - (\frac{r_{ij}}{r_0})^m} \tag{4.4}$$

where r_{ij} is the distance between atoms i of group A and j of group B, r_0 is the cutoff distance to consider two atoms in contact, and the parameters n and m are exponents of the switching function, s_{ij}. We used $n = 6$, $m = 12$, and $r_0 = 0.35$nm in this work for all quantities evaluated in contact maps. The $r_0 = 0.35$nm corresponds to the first minimum position of the radial distribution functions (RDFs) for the distance between heavy atoms of different groups such as backbone-backbone, sidechain-sidechain, backbone-water, and sidchain-water.

Besides the quantities involving the reorganization of the protein backbone and side chain, We also examined various topological properties like the water networks connecting different parts of the protein. In particular, we examined directed water wires connecting donor and acceptor groups of the peptide. Donor groups include the N-terminus and amide N−H bonds of the backbone, while the C-terminus and the carbonyl C=O groups are acceptor groups. Oxygens of water molecules ($\mathrm{O_w}$) act as both donors and acceptors. All the donors and acceptors are treated as vertices on a graph with edges between them. The vertices include the nitrogen and oxygen atoms of the protein and the oxygen atoms of water molecules. If two vertices v_i and v_j are connected by an edge e_{ij}, they are said to be adjacent. The edges e_{ij} encode information about the hydrogen bonds. We assign a weight to edges leading to a weighted directed adjacency matrix A=A(G), which is an $M \times M$ matrix with M being the number of vertices. There are broadly two criteria to define a hydrogen bond, one based on geometry and another based on energetics. We use the former and define the weights of the edges based on a combination of the distance and angle of the hydrogen bond [Luzar and Chandler, 1996]. More specifically, the elements of adjacency matrix, A_{ij}, for the weighted graph are defined as

$$A_{ij} = \left(\frac{1 - \left(\frac{r_{ij}}{r_0}\right)^n}{1 - \left(\frac{r_{ij}}{r_0}\right)^m} \right) \left(\frac{1 - \left(\frac{\alpha_{ij}}{\alpha_0}\right)^n}{1 - \left(\frac{\alpha_{ij}}{\alpha_0}\right)^m} \right) \tag{4.5}$$

where r_{ij} is the distance between donor and acceptor of H-bond and α_{ij} is the angle H−D−A, r_0, α_0, m, and n are parameters of the edge weight. Note that in the expression above for A_{ij}, $n > m$. The directed water wires are determined by Djisktra's algorithm [DIJKSTRA, 1959] which determines the shortest path between any pair of nodes on a graph. In order to take the advantage of this algorithm, we give small weights to strong hbond wires and therefore set the exponent n with a value lager than m. The parameters r_0 and α_0 for different hydrogen bonds, were set by the values corresponding to the first minima position of the distribution function of both the distance between donor and acceptor and the angle H−D−A. The parameters of Equation 4.5 are set to $r_0 = 0.34$nm, $\alpha_0 = 30°$, $m = 6$, $n = 12$.

All the molecular dynamics simulations were performed using the GROMACS 4.6.7 [Pronk et al., 2013] package and the metadynamics calculations were conducted using the PLUMED2.1 [Tribello et al., 2014] plugin. For both well-tempered metadynamics simulations of zwitterionic and methyl-capped systems, the bias factors, $\gamma = (T + \Delta T)/\Delta T$, were set to 10 and Gaussian functions were deposited every 1ps with an initial height of 0.5kJ/mol, whereas the widths in the radius of gyration and the end-to-end distance of Gaussian functions for zwitterionic were 0.01nm and 0.01nm and widths in the radius of gyration and the end-to-end distance for methyl-capped systems were 0.02nm and 0.02nm respectively. These widths were determined from the fluctuations observed during the unbiased simulations of both systems. The simulations of zwitterionic and methyl-capped systems were run for a total simulation time of 1.5μs and 1.1μs. The visualization of the structures was done using VMD [Humphrey, Dalke, and Schulten, 1996].

In most classical empirical potentials, both the protein and water molecules are non-polarizable. In liquid water for example, the dipole moment of a single water molecule in gas phase is actually smaller than that in the condensed phase due to polarization. In classical potentials this is effectively included by increasing the atomic charges so that the dipole moment is larger. Within the context of our simulations, both the N and C terminus do not polarize the water molecules and hence the dipole moments of water molecules near the protein will be exactly the same as those that are far away. In addition, the water molecules may also polarize parts of the protein such as the amide backbone groups although the extent of this is not entirely clear. As a possible direction of investigation in the future, it is possible to examine the effect of using polarizable forcefields for proteins on this system [Shi et al., 2013; Ponder et al., 2010]. However, besides the protein, one would also have to include polarizable water molecules making the simulations much more computationally expensive.

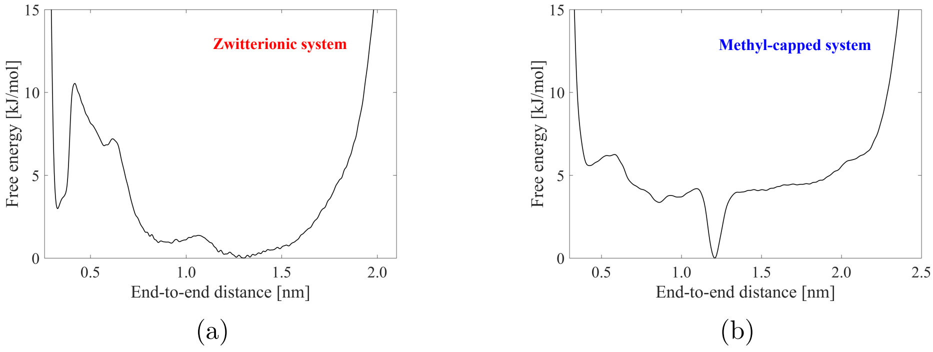
4.3 Results

In this section, we move on to characterizing the free energy surfaces obtained for the zwitterionic and methyl-capped systems. As we will shortly see, changing the termini has quite a drastic effect on the underlying free energy surface and more particularly, the structural disorder that is observed in the polypeptide chain. We rationalize the origin of this by elucidating the re-organization of both the protein and water hydrogen bond networks.

4.3.1 Free Energy Surfaces

4.3.1.1 1D-Free Energy Surfaces

As described earlier, we biased two collective variables, the end-to-end distance and the radius of gyration, in our well-tempered metadynamics simulations. We begin by illustrating in Figure 4.1 the 1D free energy surface along the end-to-end distance obtained for the zwitterionic and methyl-capped systems. This comparison also helps us build our intuition on the underlying interactions that are important for stabilizing different structures we observe. It is clear that just by a cursory visual inspection of the 1d-free energy surface, the two zwitterionic and methyl-capped systems are very different. In the case of zwitterionic, there is a pronounced and narrow minimum at 0.4nm corresponding to the formation of a salt bridge between the N and C terminus, and a much broader basin between 0.8-1.8nm which involve situations where the N and C terminus are far away from each other and hence more solvent exposed. Based on the 1D free energy surface, there are barriers between 7-10kJ/mol to move between these two basins. On the other hand, for the methyl-capped system, the narrow minimum occurs at much larger end-to-end distance distances of about 1.2nm. Approximately 5 kJ/mol above this minimum are two broad and flat basins analagous to what one would see in a typical funnel-like picture of protein folding [Finkelstein and Galzitskaya, 2004].



(a) (b)

Figure 4.1: (a) 1D free energy surface along the end-to-end distance for zwitterionic system and (b) methyl-capped system

The differences observed between zwitterionic and methyl-capped systems in Figure 4.1 are not too surprising given that the fluctuations along this coordinate involve very different driving forces. In the case of zwitterionic system, there is a strong electrostatic interaction that essentially gets completely unscreened by the solvent in the narrow basin below 0.5nm. For methyl-capped system, the termini are made up of methyl groups and hence at short end-to-end distance involve much weaker van-der-Waals interactions which simply creates a shallow minimum at around 0.4nm as seen in Figure 4.1b. This minimum is reminiscent of the contact minimum that occurs in the potential of mean force between two methane molecules in water [Hummer et al., 1996]. This analysis for both zwitterionic and methyl-capped systems integrates out the the radius of gyration and hence coordinates that would be involved in characterizing the degree of change in the compactness of the polypeptide chains. We now move on to discuss the higher dimensional free energy surface obtained along both biased and unbiased collective variables in our metadynamics runs.

4.3.1.2 2D Free Energy Surfaces

Protein Conformation

We begin by illustrating the 2D free energy surface that is obtained by biasing both the end-to-end distance and the radius of gyration collective variables for the zwitterionic system in Figure 4.2. This free energy surface reveals a much richer underlying landscape characterized by structural heterogeneity. The free energy surface features four distinct minima along the end-to-end distance and the radius of gyration at roughly the following locations: (0.35,0.45), (0.35,0.53), (0.75,0.45) and finally (1.3,0.53). As one might expect based on our earlier observations, the former two at shorter end-to-end distance involve narrower basins due to strong electrostatic forces, while the latter two at larger the end-to-end distance are broader owing to the enhanced conformational flexibility. Interestingly, for both situations where there are strong and weak termini interactions in the zwitterionic system, there are two basins with smaller and larger radii of gyration which essentially quantifies the extent of compactness or folded character of the system.

In order to aid future discussions in the paper, we label the four basins in Figure 4.2 A, B, C and D. Also shown in the Figure are representative snapshots as a visual guide to the conformational changes that are involved. Within the statistical errors associated with the convergence of our free energiesall the four states are essentially equally populated. While the NH_3^+ and COO^- groups in both A and B reside very close to each other, the snapshots suggest that both the packing of the backbone and side-chains is quite different. In order to understand these features associated with the states A, B, C and D in more detail, we turn to examining how changes in the radius of gyration involve the reorganization of both backbone and side-chain packing.

Polar and Non-Polar Protein Interactions

To probe deeper into the underlying origins associated with fluctuations along the radius of gyration coordinate, we examined the reweighted free energy surface obtained using two additional unbiased collective variables, namely the degree of

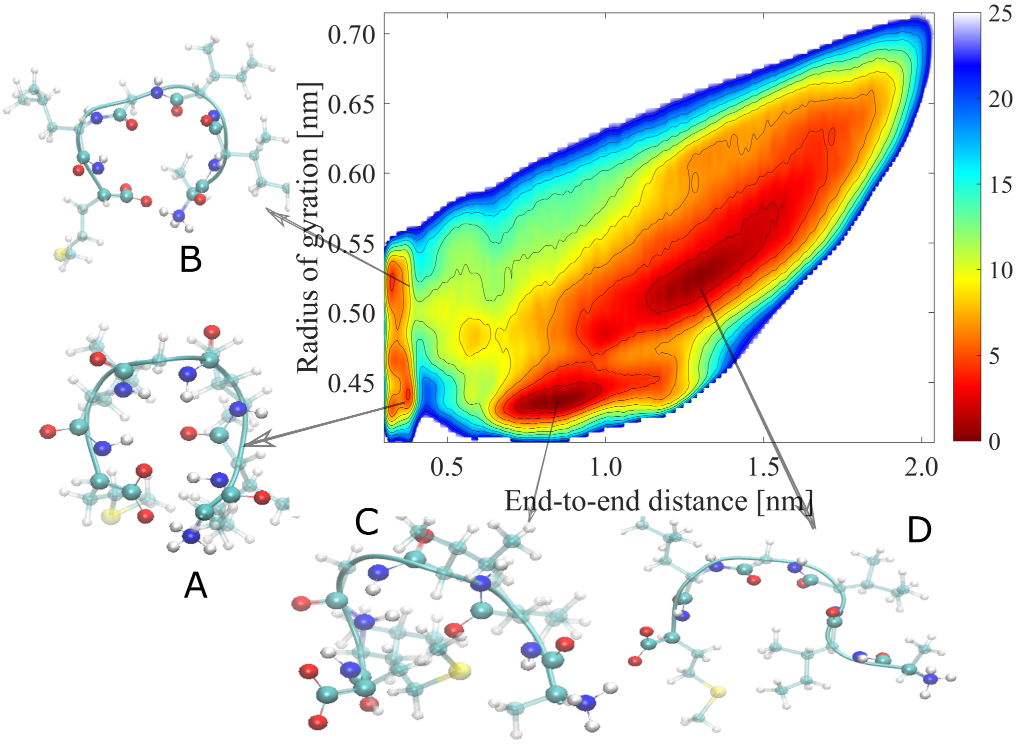
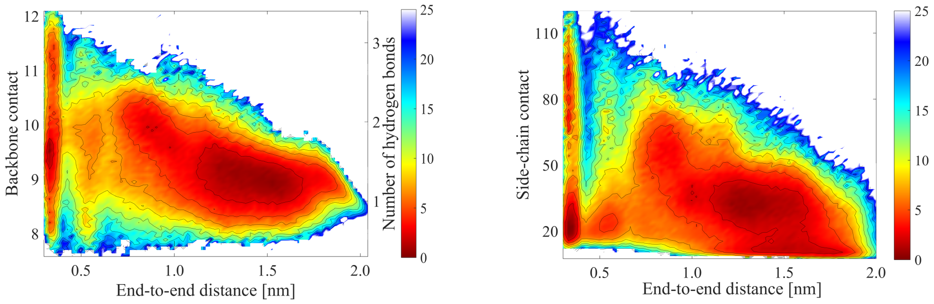


Figure 4.2: 2D free energy surface obtained from 1.5-μs-long well-tempered metadynamics for zwitterionic chain. The free energy surface is contoured by 2.5kJ/mol up to 25kJ/mol. The structures shown around the free energy surface correspond to representative snapshots in the different basins.

backbone and side-chain contacts as measured by contact maps described earlier in the Methods section. The former quantifies the extent of the hydrogen bonding interactions between the amide bonds as illustrated in the changes observed going from snapshots B to A of Figure 4.2. The latter reveals changes in the extent of hydrophobic interactions between the side-chain groups in the peptide.

The left and right panels of Figure 4.3 show the 2D free energy surface obtained from the simulations along the end-to-end distance and backbone contact and side-chain contact respectively. Overall, we observe a complex coupling of polar and non-polar forces acting simultaneously to stabilize the different conformations. For states A and B which occur for short distances when the end-to-end distance is less than 0.5nm, there are three regions along the backbone contact coordinate. We also notice, that the states where the N and C termini are further away from each other (states C and D) are characterized by slightly smaller backbone contact values although the differences are not so drastic indicating that even in the so-called less compact (or extended) states, there are significant hydrogen bonding interactions between the N−H and C=O groups of the backbone. To give a clear description of the hydrogen bonding interactions between the N−H and C=O groups of the backbone, we used another collective variable, the total number of backbone-backbone hydrogen bonds which is shown in the right vertical y-axis of Figure 4.3a. Figure 4.3a shows that both the compact and extended states have more than one intramolecular hydrogen bond. We also examined the individual contributions coming from different backbone-backbone hydrogen bonds. Note that that hydrogen bonds between N−H of GLY and C=O of ILE$_2$, N−H of LEU and C=O of ILE$_2$, and N−H of MET and C=O of ILE$_2$ contribute the most to the total number of hydrogen bonds. The right panel of Figure 4.3 illustrates how the hydrophobic side chains contribute to the conformational fluctuations. For short N-C termini distances, there are conformations consisting of high and low side-chain contact. The changes in the side-chain contact parameter in this regime, are quite significant - the value of side-chain contact undergoes broad fluctuations and can increase by a factor of 4 between states A and B.

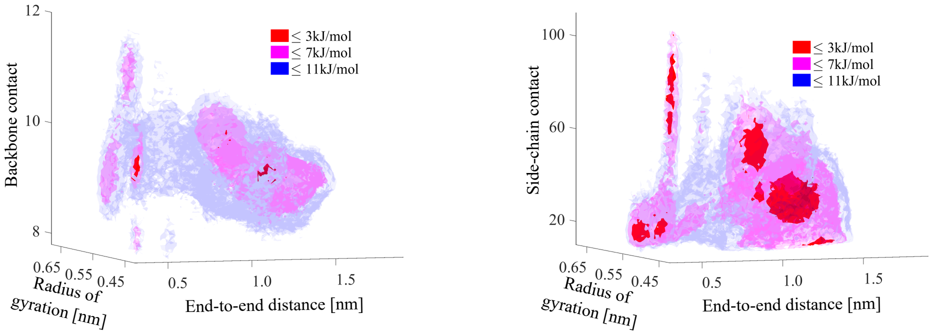


(a) (b)

Figure 4.3: (a) Reweighted 2D free energy surface along end-to-end distance, backbone contact(left y-axis), and total number of backbone-backbone hydrogen bonds(right y-axis) and (b) along end-to-end distance and side-chain contact

The preceding analysis shows the importance of both hydrogen bonding backbone interactions as well as the packing of hydrophobic side-chains once the N and C termini are in close proximity. To elucidate how changes in the backbone contact and side-chain contact result in changes in the radius of gyration and hence also to resolve better the differences in the origin of states A, B, C, and D in Figure 4.2, we examined 3D free energy surface along the end-to-end distance, the radius of gyration, the backbone contact and the side-chain contact as shown in Figure 4.4. In order to ease the visualization of the 3D free energy surface, we identify three regions of the free energy surface, with different colors: less than 3, 7 and 11 kJ/mol shown in red, yellow and blue respectively. A is characterized by a lower radius of gyration compared to B and based on the 3D free energy surface, the former features slightly larger backbone contacts originating from interactions of the amide dipoles (see Figure 4.5). In addition, the backbone contacts for state A are quite similar to the extended state C. Interestingly, the 3D free energy surface shows that the largest backbone contact corresponds to a possible intermediate state that is sampled during the transition between between basins A and B.

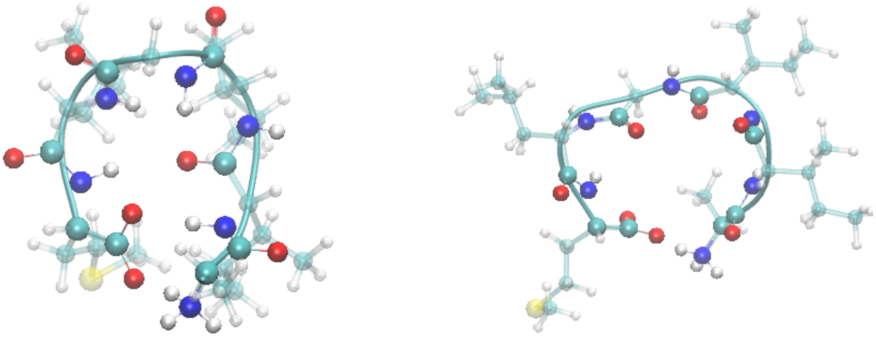
The right panel of Figure 4.4 shows the coupling of the end-to-end distance, the radius of gyration and side-chain contact and provides better separation of all the



(a) (b)

Figure 4.4: (a) Reweighted 3D free energy surface along end-to-end distance, radius of gyration and backbone contact and (b) along end-to-end distance, radius of gyration and side-chain contact.

various states and also into the origins of the changes of the radius of gyration. The side chains are all hydrophobic amino acids and hence an increase in their contacts corresponds to the formation of favorable van-der-Waal interactions between alkylic groups. We clearly see here that moving from lower to higher radius of gyration when the N and C termini are in close contact, involves a rather drastic decrease in the extent of side-chain interactions. States C and D on the other hand, are characterized by side-chain contacts sandwiched between A and B. In C, the side chains of amino acids isoleucine and leucine pack closely with each other while in D which has the lowest side-chain packing, involves isoleucine and methionine. Interestingly, moving to state B results in a conformation where all the side chains are solvent exposed. Finally, the transition from B to A involves a collective collapse of all but one of the hydrophobic side chains. This creates a conformation resembling a Janus-like particle with the polar hydrogen bonds of the backbone on one side and the non-polar side chains on the other (see left and right panels of Figure 4.5).



(a) (b)

Figure 4.5: (a) Representative snapshot of basin A and (b) basin B

From an energetic standpoint, the creation of this compact Janus-like structure involves a marked decrease of side-chain interaction energies. We showed in Figure 4.6 the 3D surface involving end-to-end distance, radius of gyration, and side-chain interaction enthalpy.

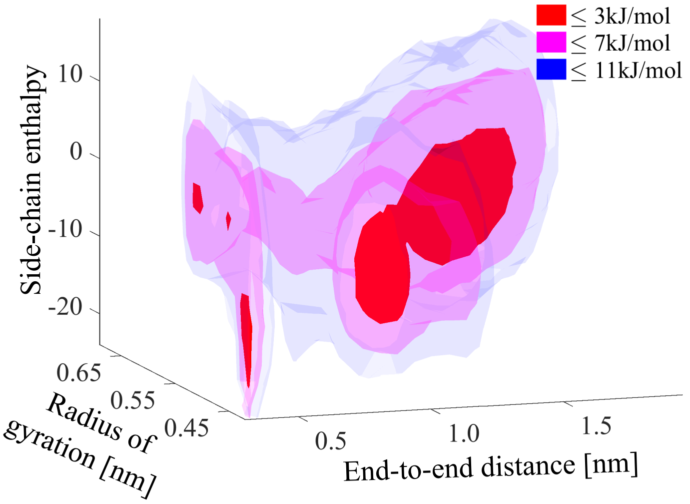


Figure 4.6: Reweighted 3D free energy surface along end-to-end distance, radius of gyration and enthalpy of sidechain

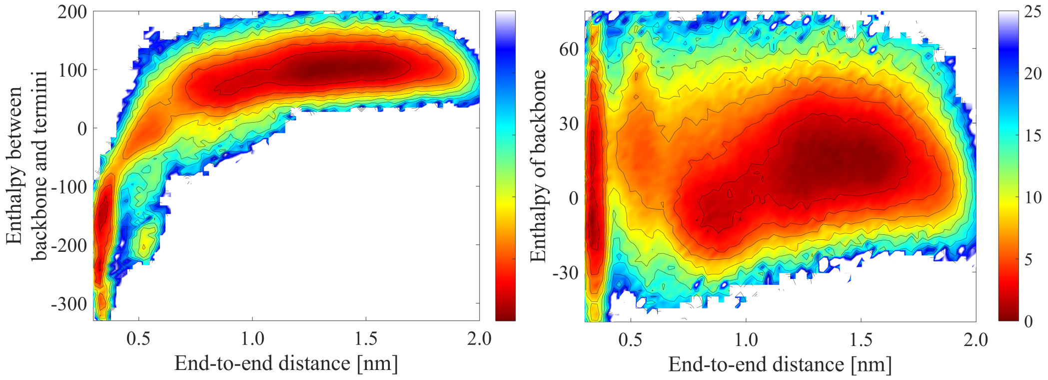
The transition from B to A results in a decrease of sidechain interaction energy by ≈ 15kJ/mol. It is clear from the figure that sidechain interaction energy for state C is also larger than state A by ≈10kJ/mol even though states A and C have similar radius of gyration.

The picture developed up to this point, reveals a rich variety of molecular interactions involving both polar hydrogen bonding and hydrophobic interactions for the zwitterionic chain. Fluctuations in the backbone and side-chain contacts lead to changes in the extent of compactness and hence the radius of gyration. Although we have focused on the hydrogen bonding interactions of just the amide bonds of the backbone, due to the highly flexible character of the termini, polar interactions between the termini and the backbone also play an important role in stabilizing all the various conformations. In Figure 4.7, we showed free energy surfaces along with the end-to-end distance and the enthalpies between termini and backbone and between backbone groups. As you can see in the left panel of Figure 4.7, the transition from the extended state to compact state results in the decrease of termini-backbone interaction energy by 300 kJ/mol. In the right panel of Figure 4.7, the backbone interaction energy are shown where there are less pronounced changes in the backbone interactions consistent with Figure 4.3.

These results are also interesting within the context of understanding the aggregation of amyloid fibrils. As indicated in the introduction, the underlying driving forces associated with the formation of the fibrils include both hydrogen bonding of the backbone and burying of hydrophobic side chains [Wälti et al., 2016]. Our simulations show that these driving forces form an integral part of the conformational fluctuations of our small model peptide.

Protein-Water Interactions

It goes without saying that these fluctuations involving the close packing of the

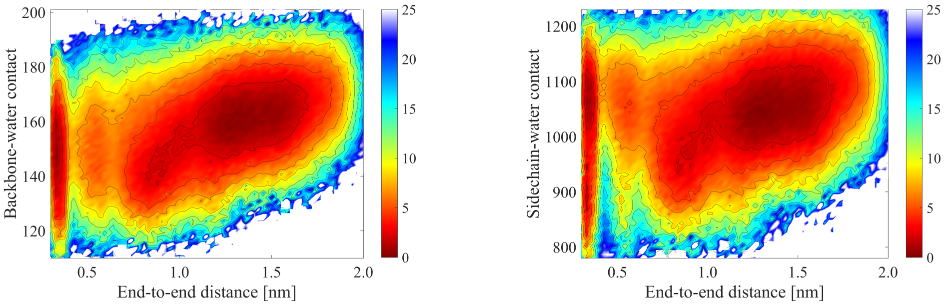


(a) (b)

Figure 4.7: (a) Reweighted 2D free energy surface along end-to-end distance and enthalpy between backbone and termini (BT Enthalpy), (b) along end-to-end distance and enthalpy of backbone (BB Enthalpy)

hydrogen bond network to stabilize the backbone interactions as well as the packing of side chains must be intimately coupled to the reorganization of the surrounding solvent. There have been numerous studies in the literature discussing the importance of this coupling [Bellissent-Funel et al., 2016; Fischer and Verma, 1999; Best, Zheng, and Mittal, 2014; Mattos, 2002; Phillips and Pettitt, 1995; Bizzarri and Cannistraro, 2002; Tarus, Straub, and Thirumalai, 2006; Baumketner and Shea, 2007]. One obvious question that emerges from the preceding analysis regards the changes in the exposure of both the backbone and side-chains to the solvent during the structural transformations along the free energy surface. The left and right panels of Figure 4.8 show the free energy surface along the end-to-end distance and the contact between the backbone/side-chain and water. In constructing this contact map, we focused on water molecules within 3.5 Angstrom of the backbone/side-chain. Although the peptide chain becomes more compact at short N-C termini distances, this structural transformation is manifested much more in the solvent exposure of the side-chain than in the backbone.

In the following section, we will tackle the exposure of the peptide to the solvent through an examination of some topological properties involving the hydrogen bond network. The snapshots shown earlier in Figure 4.5 depicts the physical origins of Figure 4.8. In one case (panel b) the side chains are mostly separated away from each other leaving a lot of room for water molecules to be interspersed between the side chains. On the other hand, for state A which is the most compact state, involves the collective collapse of several of these side chains leading to the expulsion of solvent. The backbone on the other hand, as we will see in more detail in the next section, still remains solvent exposed - in fact, in both state A and B we see that the amide dipoles orient in such a fashion so that several N$-$H groups of the backbone essentially form strong hydrogen bonds with a single C$=$O, while most of the other carbonyl oxygens remain solvent exposed. The fact that the anionic carbonyl groups are exposed to water is consistent with an idea proposed by Collins [Collins and Washabaugh, 1985] who suggested that the structuring of water induced by ions, is rooted in specific hydrogen bonding interactions. He proposed that anions are better at ordering water molecules than cations due to the asymmetry of charge in a water molecule. In the next section we will see how the ordering of water is reflected in the formation of water wires around the peptide chain.



(a)

(b)

Figure 4.8: (a) Reweighted 2D free energy surface along the end-to-end distance and backbone-water contact and (b) along the end-to-end distance and sidechain-water contact

4.3.2 Water Wires Around Amyloid β_{30-35} Chain

Earlier we showed that the backbone-water contact map does not feature any significant changes during the structural conformational changes. On the other hand, there are much larger changes in the solvent density around the hydrophobic side-chains. The backbone-water contact map quantifies the polar interactions formed between the amide groups and the surrounding water but is blind to the orientational correlations originating directly from the hydrogen bond network.

Understanding the role of hydrogen bond networks in shaping both the structural and dynamical properties of biological systems continues to be a topic of active investigation [Smolin et al., 2005; Cui, Ou, and Patel, 2014; Brovchenko et al., 2005; Ahmad et al., 2011]. Thirumalai and co-workers have also shown that long-lived water wires between two beta sheets from the polar fragment of the yeast prion protein, result in long-lived metastable structures [Thirumalai, Reddy, and Straub, 2012; Reddy, Straub, and Thirumalai, 2010]. The analysis of the water wires in these studies is done in a qualitative manner without quantitatively examining the hydrogen bond network. There have been several theoretical studies examining the importance of water networks in maintaining the structural integrity of proteins as well as in the interaction of proteins. In particular, Mazen A. and co-workers [Ahmad et al., 2011] showed that during the encounter of two hydrophilic proteins, adhesive water networks form between them stabilizing intermediate states before native contacts form. In all these studies, the connections between the nodes in the network neglect the directionality of the interactions. It is clear however, that if one is interested in networks associated with water hydrogen bonds as we are here, then one must consider the evolution of directed networks.

In this section, we examine the cooperative and collective behavior associated with the reorganization of the directed adhesive networks enveloping the hexapeptide chain in water. As described in the Methods section, we examined the statistics associated with the shortest directed path connecting candidate donor and acceptor groups. Before we showcase individual water wires forming an important and integral part of the protein structure and how they change during the conformational fluctuations, we begin by quantifying how the collective directed network from all possible donor to acceptor groups evolves during conformation fluctuations of protein. In order to quantify this, we examined the global efficiency of the protein-water network which is defined as [Latora and Marchiori, 2003; Latora and Marchiori, 2001; Stam and Reijneveld, 2007]

$$E_g = \frac{2}{N(N-1)} \sum_{i,j \in N, i \neq j} \frac{1}{\ell_{ij}} \tag{4.6}$$

where N is the number of vertices (N$-$H and C$=$O groups) and ℓ_{ij} is the shortest path length between i and j vertices. If the two vertices i an j are not connected, $1/\ell_{ij}$ is set to zero so that these pairs don't contribute to global efficiency.

Figure 4.9 shows the free energy surface obtained along the end-to-end distance and E_g. Since the global efficiency averages over many directed paths from different parts of the peptide, it is useful to calibrate ourselves first on the typical shortest path lengths (ℓ_{ij}) that the E_g maps onto. An average directed path of length 1 would correspond to a global efficiency of 1. In the limit that there are no directed paths between the nodes, E_g would tend to 0. When the N and C termini are separated far away from each other, Figure 4.9 shows that E_g is peaked at 0.065 corresponding to an average path length of about 16. In this extended state, there are large fluctuations in the directed network connecting the donor N$-$H and C$=$O groups in the peptide as seen in the broad distribution of E_g extending from 0.04 to 0.125 implying the formation of very long wires of length 25 to rather short wires of length 8. Moving from states C and D into the regime where the N and C termini are in close contact, leads to an overall increase in E_g. At short end-to-end distance distances of 0.35 nm, there is a minimum corresponding to a typical path length of 11 which results from the formation of shorter directed paths between the donor and acceptor groups compared to the extended state although there are still rather large fluctuations in the network.

The global efficiency of the network averages over connections involving all terminus-terminus, termini-backbone and backbone-backbone and hence we get a global measure of the changes in the connectivity. Furthermore, the contributions to global efficiency between candidate donor and acceptor groups involve the possibility of many long wires or a few short wires. To examine the contribution of wires from specific donor and acceptor groups of the peptide, to the global efficiency, we also

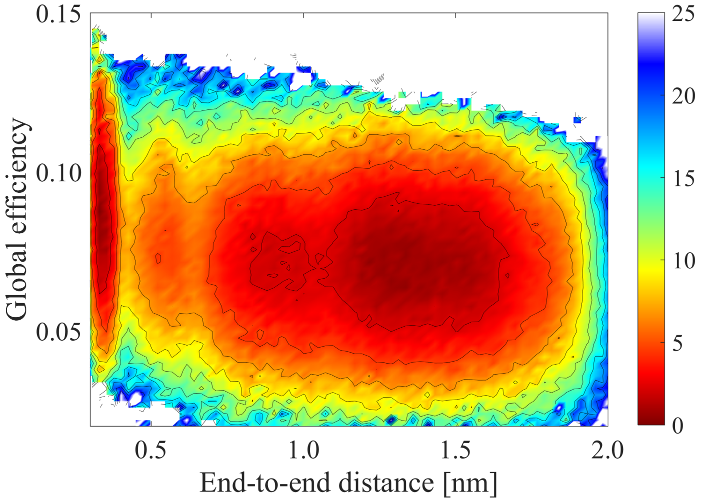


Figure 4.9: Reweighted 2D free energy surface along the end-to-end distance and global efficiency

determined a local efficiency in a similar way to the global efficiency. The difference being that in the local efficiency, the summation goes over the number of nodes specific between a particular donor and acceptor. To understand better the changes of the water networks involving the termini and backbone separately, we show in Figure 4.10 the free energy surface obtained by examining the components of local network. The top left panel elucidates the changes involving the specific connections between the N and C terminus. When the N and C termini are far away from each other the local efficiency is small corresponding to very long water wires. More interesting features emerge when the distance is short - there are several sub-basins along the local efficiency corresponding to different types of salt-bridges which consist of both direct contact between the N and C termini as well as by water mediated interactions due to the formation of short water wires ranging from length 2 to 5. The evolution of the specific water wire between the N and C termini is shown in the bottom left panel of Figure 4.10. Visual examples of the different types of water wires are shown in the bottom left panel of Figure 4.10.

The top right panel of Figure 4.10 elucidates the water networks formed between all the backbone amide groups and carbonyl oxygens. In the extended state, the backbone is stabilized by water wires of varying lengths ranging from 5 to 30 with a dominant peak at 15. Transitioning into the regime where the N and C termini are in close contact results in a more structured landscape implying that in both states A and B, there are different types of sub-states stabilized by unique water-networks. A specific example of these water networks are illustrated in the bottom right panel of Figure 4.10 formed between the isoleucine and leucine residues. We observe here as well that there are specific water wires between the amide groups of these amino acids as a function of the N-C termini distance - see labels 1, 2 and 3.

A similar type of analysis focusing on the networks connecting the N-terminus to the backbone and the backbone to the C-terminus shows distinct patterns in the water

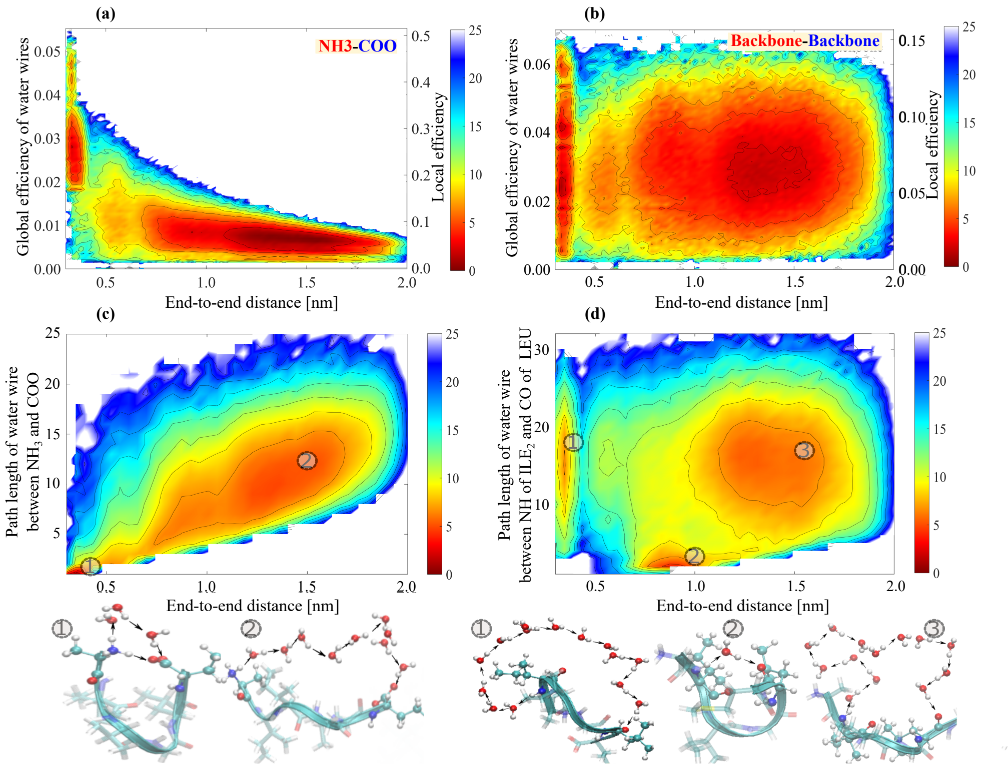


Figure 4.10: Top: (a) reweighted 2D free energy surface along the end-to-end distance global efficiency of water wires between N and C termini(left y-axis) and it's local efficiency(right y-axis), (b) along the end-to-end distance and global and local efficiencies of water wires between N$-$H and C=O of backbone. Bottom: (c) along the end-to-end distance and the shortest path length between NH_3^+ and COO^- and (d) along the end-to-end distance and the shortest path length between NH(ILE$_2$) and C=O(LEU).

wires between the two respective situations. To illustrate these features we show in the left and right panels of Figure 4.11 a comparison of the distribution of water wires from the N-terminus to the C=O group of Glycine and from the N$-$H group of the Glycine to C-terminus respectively. The free energy landscapes associated with these coordinates are distinct, reinforcing the notion that the interactions of the N and C termini with the backbone is mediated by specific water-wire motifs.

The global and local-efficiency network measures associated with the NH_3^+(terminus)-backbone and backbone-COO^-(terminus) are shown in Figure 4.12. The water wires connecting NH_3^+-backbone show that the compact and states have slightly smaller wires but that in both states the fluctuations are quite similar ranging from 10 to 30. The pathway going from extended state to the compact state involves the increase of the global and local connectivities between N-terminus and backbone. On the other hand, free energy surface involving the water wires connecting backbone-COO^- has more structure and it features large fluctuations in the compact state than the wires connecting N-terminus to the backbone. For example, the local minima of average path length of ≈3 appeared in the compact state, where there are different water wires stabilizing the protein structure. The pathway form the extended state

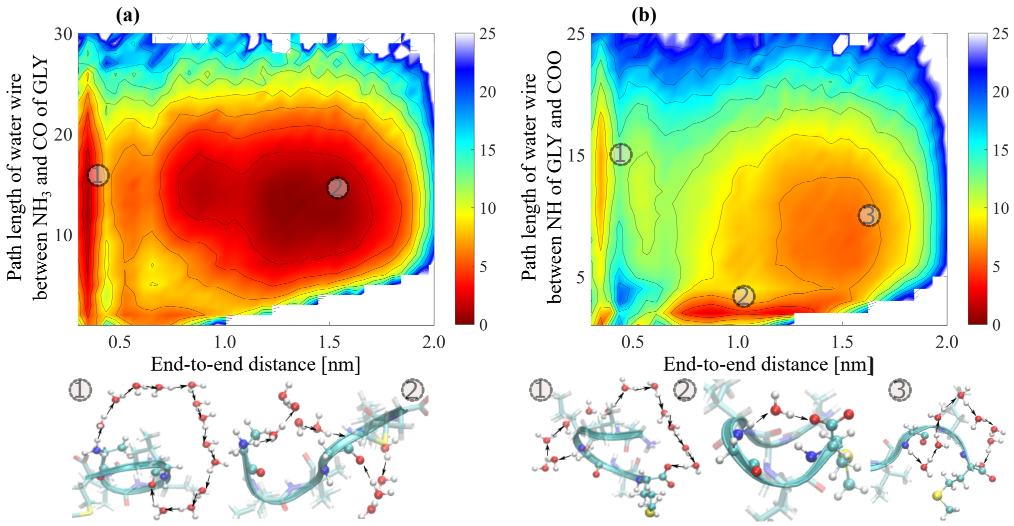
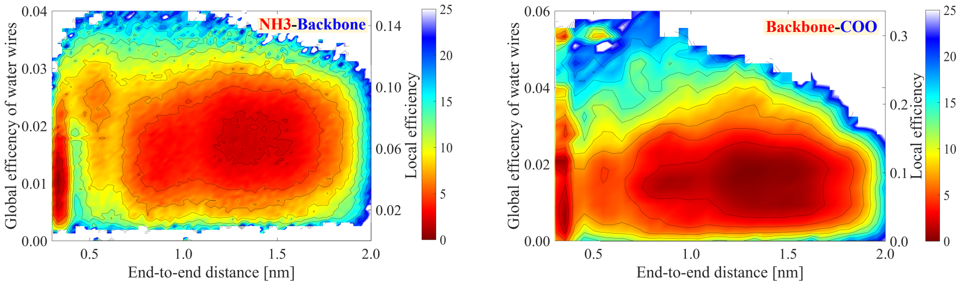


Figure 4.11: (a) Reweighted 2D free energy surface along the end-to-end distance and the shortest path length between NH_3^+ and C=O of GLY and (b) along the end-to-end distance and the shortest path length between N−H of GLY and COO^-

to the compact state doesn't involve any significant change along the global and local connectivity of the backbone-COO^- water wires.

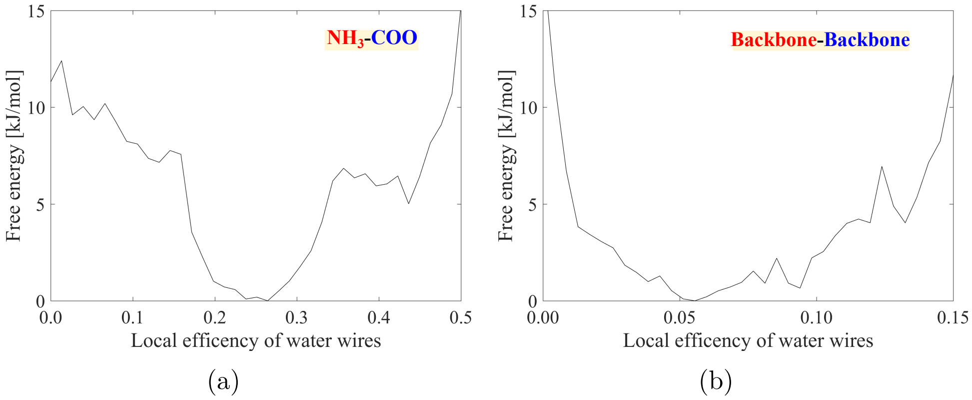


(a) (b)

Figure 4.12: (a) Reweighted 2D free energy surface along end-to-end distance and protein-water H-bond (PW-HB) network connectivity between NH_3^+ termini and backbone, (b) along end-to-end distance and PW-HB network between backbone and COO^- termini.

To better see how the water wires couple with the protein coordinates, we constructed one-dimensional free energy distributions along the various network coordinates when the distance between the N and C terminus was less than 4 Å. Take for example, the distribution along the water-wire network coordinate from the N to C terminus in the left panel of Figure 4.13 which shows a deeper minimum at around 0.25 (corresponding to an average path length of 4). The distribution is however broad, rough and characterized by several local minima that are thermally accessible. Similarly, the free energy distribution along the backbone network coordinate is also broad and rough in the right panel of Figure 4.13.

Another way to interpret this is that the solvent contributes to the configurational entropy of the peptide. This is an issue we are currently exploring. Since we are



(a) (b)

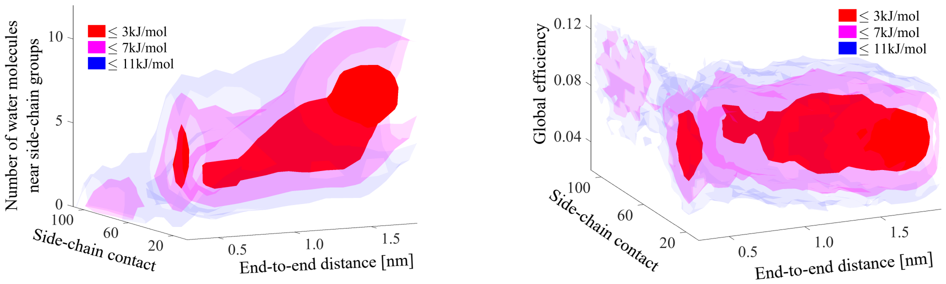
Figure 4.13: (a) 1D free energy along local efficiency of water wires connecting NH3 and COO and (b) along local efficiency of water wires connecting NH and CO groups of backbone

dealing with a short chain made up of 6 amino acids, it is clear that the water wires linking the N and C termini will be more important compared to longer chains. However, the preceding analysis also show the existence of water wires involving the termini and backbone as well as within the backbone itself, which is expected to be more important for longer chain systems. In the future, it will be interesting to examine how the properties of the water wires change as a function of the length of the chain as well as the sequence of amino acids forming the chain.

4.3.3 Coupling of Protein and Water Fluctuations

In this work we have placed particular emphasis on the coupled reorganization of the hydrogen bond networks involving intraprotein contacts as well as that of the protein and surrounding water. The electrostatic interaction between the N and C termini results in the formation of strong salt-bridges that can be both in direct contact or mediated through the solvent by water wires. Besides the salt bridges, there are a myriad of network connections that can form between different donor and acceptor groups of the peptide through the solvent. Thus conformational fluctuations are best seen as a collective reorganization of both protein and water networks. At the same time, the side chains of the amino acids are hydrophobic in nature and the extent to which they pack closely with each other, also plays an important role in stabilizing different types of conformations. The collapse of the side-chains and explusion of water results in the formation of a protein conformation resembling a Janus-like particle.

The coupling of the reorganization of the water networks as well as the collapse of the hydrophobic side-chains, as pointed out earlier, is shown more clearly in the left and right panels of Figure 4.14. The left panel shows the coupling between the squeeze out of water molecules between the side-chains and the increase in their van-der-Waals packing. The transition from states B to A involves a coupling of these two events occurring simultaneously similar to what was observed in earlier simulations



(a) (b)

Figure 4.14: (a) Reweighted 3D free energy surface along the end-to-end distance, the radius of gyration and number of water molecules within 0.5nm of the side-chain groups and (b) along the end-to-end distance, side-chain contact and global efficiency

of a hydrophobic sub-sequence of the amyloid fibril [Reddy, Straub, and Thirumalai, 2010]. Besides the water expulsion of water, there is also a collective reorganization of the hydrogen bond network. This feature is seen in the right panel of Figure 4.14 where we see that during the collapse of the hydrophobic side-chains there is an overall shortening of water wires enveloping the hexapeptide. It would be interesting in future studies to understand how these features change as one moves to larger polypeptides as well as how the mechanisms are altered when the chemistry of the side-chain groups are altered.

The 3D-Surfaces that we have shown in Figure 4.14 involving coupled protein and water reorganization, as well as those shown of the coupled side-chain and backbone fluctuations have important implications on the underlying dynamics of the peptide fragment. While it is beyond the scope of the current study to characterize the dynamical properties, we were interested in getting some theoretical error bars on the quality of the reaction coordinates we have used here. One of the important coordinates in our work is the fluctuation of the N-C termini distance and hence we extracted the transition state theory (TST) estimate of the rate constant of escape from the contact state with d_{ee} ∼0.35nm in Figure 4.1a and compared it to the mean first passage time (MFPT) obtained from ∼60 unbiased trajectories initialized from this region. The TST estimate of the rate constant is ∼10ps^{-1}. Our unbiased trajectories were each run for a simulation time of 500ps - about 80% of the trajectories escaped within this time and the MFPT is ∼145 ps. There is, however, a broad distribution of escape times ranging from about 10-200ps. The difference between the TST estimate and our unbiased trajectories is not surprising since the former always overestimates the rate constant [Hänggi, Talkner, and Borkovec, 1990].

We show below two unbiased trajectories illustrating the evolution of various collective variables and the importance of the collective reorganization in the fluctuations on the free energy surface. In Figure 4.15 we show the evolution of end-to-end distance, radius of gyration, side-chain contact, global efficiency and finally the shortest path between the N and C termini. Between 250 and 270 ps, the distance between the N and C termini increases and the system transits out of state A. In this case, there is no big change in radius of gyration (see the violet curve in the top left panel of Figure 4.15). However, we see that there is a decrease in the global efficiency parameter around the same timescale. Part of the decrease in the global efficiency comes from the formation of a longer wire forming between the N and C termini as seen the bottom right panel of Figure 4.15. Finally, we also see that the transition out of A occurs when the side-chain contact parameter is at a lower value than that associated with the maximum side-chain packing.

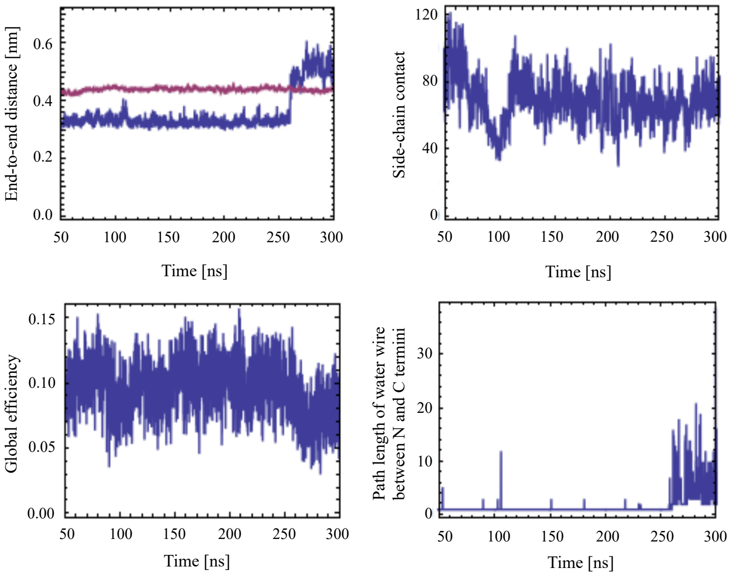


Figure 4.15: collective variables: end-to-end distance, radius of gyration(red line), side-chain contact, global efficiency and the shortest path length between N and C termini as a function of time.

The importance of the fluctuations in the side-chain contact parameter is reflected in another trajectory shown in Figure 4.16. In this case we see that the distance between the N-C termini increases transiently at around 270ps but then decreases again. radius of gyration again remains constant during this transition. Looking at the side-chain contact parameter we see that it has a larger value than that in Figure 4.16. In addition, we also see that a longer water wire between the N and C termini forms transiently but then the disappears when a strong salt bridge is reformed.

4.3.4 Implications on Amyloid Aggregation Mechanisms

In a recent experimental study, it was shown that the Aβ_{1-42} fibrils are characterized by a hydrophobic core involving the dense packing of alkylic side chains. Besides this, there are of course hydrogen bonds between the amide groups of the backbone. In order for all these conformational changes to occur, there must be some changes in the exposure of the side chains and backbone to the surrounding water. There are several simulation studies that have proposed the important role of water wires in controlling the pathways of aggregation and the underlying kinetics although this

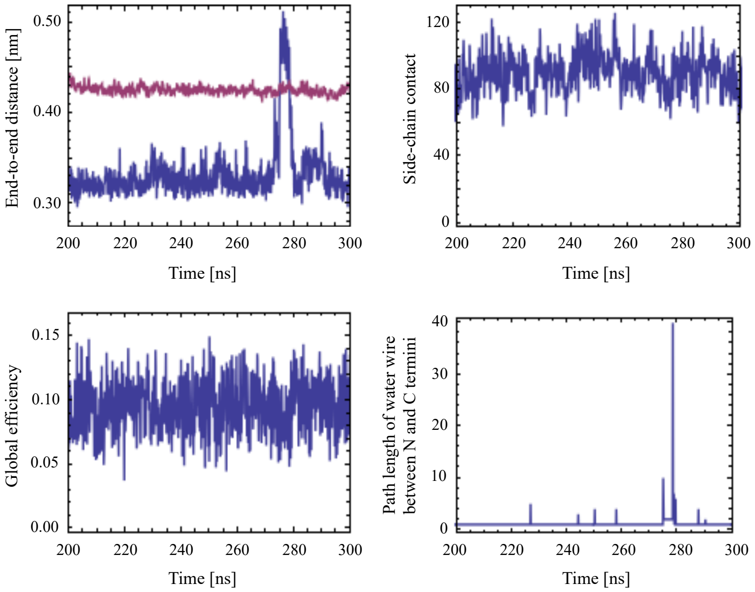


Figure 4.16: collective variables: end-to-end distance, radius of gyration(red line), side-chain contact, global efficiency and the shortest path length between N and C termini as a function of time.

still remains an open question [Thirumalai, Reddy, and Straub, 2012; Reddy, Straub, and Thirumalai, 2010].

We cannot make any quantitative claims about the implications of our results on the aggregation mechanism, but we can deduce some qualitative suggestions from the behavior of the monomer chain. As pointed out earlier, in some recent studies, Thirumalai and co-workers suggested that chronological order in which peptide-peptide contacts versus water expulsion occurs can affect the kinetics of aggregation as well as the amount of water wires that eventually end up in the fibril [Thirumalai, Reddy, and Straub, 2012; Reddy, Straub, and Thirumalai, 2010]. Firstly, it is clear from the analysis of the hydrogen bond networks that the aggregation process will involve the reorganization of many different water wires and that not all of them will be the same. As aggregation proceeds, there will likely be water wires that get trapped in the nucleating core as well as those that get expelled. Since the Aβ_{1-42} is made up different types of side-chains, the underlying dynamics of the water wires will vary depending on the particular polar and non-polar groups in the local environment.

Besides the water wires, there is also the question of the length scales of hydrophobicity and hydrophobic collapse involved in the aggregation mechanism. At the early stages, like in our simulations of the monomer state, we observe fluctuations involving the collective collapse of the hydrophobic side chains and simultaneous reduction of water density around those groups. Clearly for a much larger chain this will be more complicated. However, as aggregation proceeds, there will likely be a competition between hydrophobic collapse involving side-chains within the same monomer and as well as from different monomers or semi-nucleated chains. We can only speculate at this point, that the presence of long living water wires may slow down the kinetics of hydrophobic collapse.

4.3.5 NMR and CD Spectra Predictions

There are currently no experiments examining the structure of the zwitterionic and methyl-capped system in water. Here at this point, we made some theoretical predictions of the NMR chemical shift and CD spectra of both systems. To do this, we used an open source empirical predictor of the NMR chemical shifts namely ShiftX2 [Han et al., 2011] which has been used in numerous other applications [Kragelj et al., 2013; Frank et al., 2015; Zhu, Zhang, and He, 2013]. For the CD spectra, we used the empirical predictor DichroCalc [Bulheller and Hirst, 2009]. Comparing the NMR chemical shift and CD spectra, the latter appears to be more revealing in distinguishing zwitterionic and methyl-capped systems.

Out of all the chemical shifts examined, we found that the proton attached to the nitrogen atoms of the amide groups provided the strongest signal in the chemical shift to possibly distinguish between the different conformations explored by the zwitterionic and methyl-capped systems. In the left panel of Figure 4.17, histograms for the two systems are shown. A total of 300 configurations were used for both the zwitterionic and methyl-capped systems respectively. For zwitterionic and methyl-capped, the configurations were randomly chosen from those within 2 k_BT units of the minima. Overall, the methyl-capped system shows a more structured chemical shift distribution compared to zwitterionic. The fact that the zwitterionic proton chemical shifts have a pronounced population between 8-8.5ppm compared to the methyl-capped may originate from the fact that this region has been shown to be correlated with beta-sheet character [Sibley, Cosman, and Krishnan, 2003]. Although zwitterionic is a single chain, we have found that the conformational fluctuations involving the creation of states A,B and D present some beta-sheet like character.

In the right panel of Figure 4.17, we show the CD spectra for zwitterionic and methyl-capped systems averaged over 100 configurations each. Comparing the two, we see that zwitterionic system has a stronger intensity of absorption at about 195nm while, methyl-capped system has a stronger intensity of absorption at about 203nm. These differences likely correspond to the subtle changes in the extent of beta-sheet like local structures that are more pronounced for the zwitterionic system. It should be stressed that both zwitterionic and methyl-capped systems are short polypeptides and hence cannot really form any stable secondary structures.

4.4 Conclusion

In this chapter, we have explored the conformational landscape of a hexapeptide of the C-terminal fragment of Amyloid β_{30-35} in liquid water using well-tempered metadynamics simulations. The free energy landscape of this system is very rich and characterized by an underlying structural disorder where the N and C termini plays a critical role in stabilizing different conformations. The conformational fluctuations

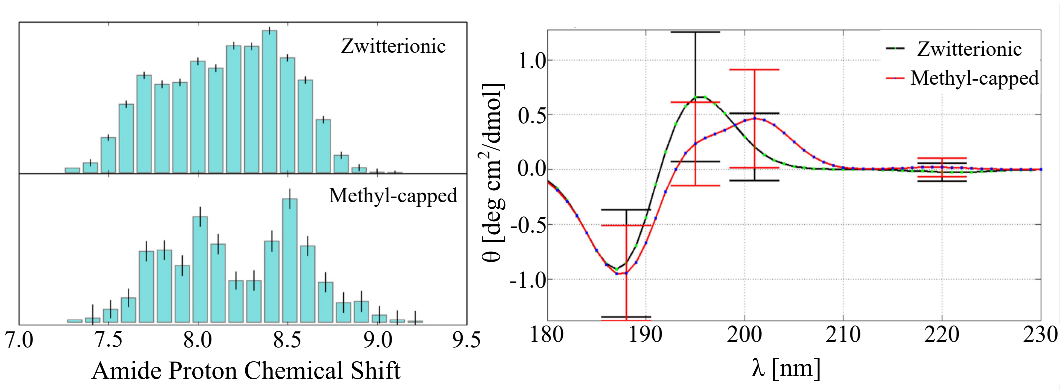


Figure 4.17: Chemical shifts of the proton attached to the nitrogen atoms of the amid groups (left panel) and CD spectra for zwitterionic and methyl-capped systems (right panel). Error bars are shown in both figures.

on the free energy landscape are driven by an intimate coupling of electrostatic interactions, hydrogen bond networks and hydrophobic forces. Although the importance of these interactions are known to be important in the aggregation of biological systems, a molecular characterization and understanding of the underlying processes still remains an active area of research. It is clear that the N and C termini play an important role in stabilizing different conformations of the hexapeptide in water. These conformations involve explicit salt-bridges between the termini and strong hydrogen bonds between the termini and the backbone amide groups. As indicated earlier, in order to assess the sensitivity and to understand how the free energy landscape changes in the absence of standard N and C terminal groups, we also performed simulations of a hexapeptide that is capped with methyl groups. The absence of the standard termini groups has a rather drastic effect on the conformational landscape. In Figure 4.18, the free energy surface along both the end-to-end distance and the radius of gyration confirms the presence of essentially only one minima. Interestingly, if we compare this to the free energy surface of the zwitterionic system, they share similar features of having analagous states A through D. However, the absence of the termini interactions de-stabilizes many of the conformations and hence is likely to affect the mechanisms associated with amyloid aggregation inferred in previous simulations where the termini are capped with methyl groups [Baftizadeh et al., 2012; Urbanc et al., 2004; Plumley et al., 2014; Hughes, Burke, and Doig, 2000; Soto, Griffin, and Shea, 2007].

Although our simulations only examine the importance of the evolution and fluctuation of directed water wires for the monomer chain, it would be interesting to explore in the future, how these properties change during the aggregation process and how this is altered by the number of monomers in the nucleation center [Thirumalai, Reddy, and Straub, 2012]. Although our metadynamics simulations currently do not allow us to extract dynamical properties, an obvious extension of the current work is to understand the dynamics associated with the conformational transitions and the reorganization of water networks around the protein. The tools developed here would be interesting to apply to these problems in the future.

Our studies should also motivate new experiments examining the effect of mutating the termini on spectroscopic observables probed through methods such as circular dichroism [Rybicka et al., 2016] and nuclear magnetic resonance [Karamanos et al., 2015] and also on optical properties such as fluorescence [Pinotsi et al., 2016].

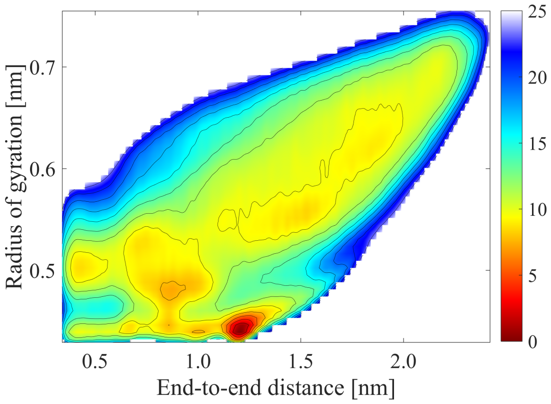


Figure 4.18: 2D free energy surface obtained from 1.1-μs-long well-tempered metadynamics for methyl-capped chain. The free energy surface is contoured by 2.5kJ/mol up to 25kJ/mol.

Chapter 5

A Data Science Approach to Understanding Water Networks Around Tri-Alanine

A version of this chapter has been published as:
Jong, K.; Hassanali, A. A Data Science Approach to Understanding Water Networks Around Biomolecules: The Case of Tri-Alanine in Liquid Water. J. Phys. Chem. B. 122 (32), 7895 (2018)

In this chapter, I use data-science algorithms to examine the complexity of water around tri-alanine protein.

5.1 Introduction

Water is one of the most ubiquitous solvents forming the essential lubricant of life [Franks, 2000; Ball, 2008; Robinson et al., 1996]. Both experimental and theoretical studies in the last century have elucidated with profound detail, the role of water in maintaining the structural and dynamical integrity of biomolecules [Ansari et al., 1992; Fenimore et al., 2002; Zhao, Li, and Tian, 2013; Bellissent-Funel et al., 2016]. In fundamental biophysical processes such as protein folding [Levy and Onuchic, 2006; Makhatadze and Privalov, 1993; Privalov and Makhatadze, 1993], protein-protein and protein-ligand interactions [Huggins, Marsh, and Payne, 2011; Papoian, Ulander, and Wolynes, 2003; Dubins et al., 2000], both hydrophobic [Russo, Teixeira, and Ollivier, 2009; Russo, Ollivier, and Teixeira, 2008] and electrostatic forces [Foderà et al., 2013; Zhang, Witham, and Alexov, 2011] are mediated by the solvent medium.

Despite long study, the microscopic origins and mechanisms associated with the coupling of protein and solvent degrees of freedom still remain an active area of investigation. Numerous experimental and theoretical investigations have shown that proteins perturb both the structure and dynamics of water around them [Fogarty and Laage, 2014; Tarek and Tobias, 2000; Bagchi, 2005; Laage, Elsaesser, and Hynes, 2017]. The extent of this perturbation and how it changes as a function of temperature, has been a subject of lively discussions in the literature. It is well known, for example, that at around 200-220K proteins undergo a dynamical glass-like transition resulting in an enhancement of their intrinsic fluctuations [Vitkup et al., 2000; Gupta, Chakravarty, and Bandyopadhyay, 2016; Jansson, Bergman, and Swenson, 2011].

Protein dynamics is thought to be strongly modulated by the surrounding solvent, and water has often been described as slaving the protein motions [Fenimore et al., 2002; Zhao, Li, and Tian, 2013]. The thermodynamics and dynamics of a protein is ultimately dictated by the underlying free energy surface (FES) in which the fluctuations occur. Due to the sheer complexity of these systems, it is typical practice to focus on a few important protein (or more generally solute) degrees of freedom and to neglect those of the surrounding solvent [Pietrucci and Laio, 2009; Fiorin, Klein, and Henin, 2013]. On the other hand, if the solvent does in fact tune the fluctuations of the protein one wonders how exactly this is manifested in the FES and how sensitive the FES is to the underlying secondary structure of the protein. Answering these questions form the subject of this work with specific focus on identifying the important solvent variables around biomolecules.

Hydrogen bond network analysis has been used in several previous studies to understand the structural and dynamical properties of hydration water around biomolecules as well as in bulk water. For example, examining the spanning hydrogen-bond network of water molecules in the hydration shell around some proteins has been shown to break at the transition temperature of around 307K, which is apparently insensitive to protein structure and average interaction energy between the protein and water molecules [Oleinikova and Brovchenko, 2011]. A more recent study by Sterpone and co-workers used a clustering of the connectivity states involving the hydrogen bond network to demonstrate a direct correlation between protein dynamical transition and the disorder of the extended water-water hydrogen bond network [Rahaman et al., 2017]. The hydrogen bond network in bulk water has also been investigated by a network analysis showing the structural inhomogeneity of water [Rao, Garrett-Roe, and Hamm, 2010]. The morphological structures of ion aggregates and hydrogen bond networks in high salt solutions have also been studied by spectral graph theory giving insights into the connectivity of network upon introduction of ions [Choi and Cho, 2014]

Here, we use a combination of data-science oriented algorithms [Rodriguez and Laio, 2014; Facco et al., 2017; Rodriguez et al., 2018] to investigate the structure and the topography of the free energy landscape of water around a peptide chain, aqueous tri-alanine (Al_3) from molecular dynamics simulations. Understanding the free energy landscape of this system in terms of its secondary structure has been the subject of several previous studies [Marinelli et al., 2009]. Here we concentrate on the importance of the surrounding water. In particular, we estimate the intrinsic dimension and we reconstruct the topography of the landscape by using various water coordinates. We find that the complex structure of the landscape can only be highlighted by using variables capturing the structure of the water network, namely the water wire variables [Jong, Grisanti, and Hassanali, 2017]. More trivial water coordinates, such as the number of hydrogen bonds formed by the solute with the water molecules, wash out the relevant information. We also discover that the free energy landscape of the water molecules is qualitatively different in different conformations of Tri-alanine. In particular, if the molecule is in an α conformation the landscape is extremely rough, while it is more smooth if the molecule is in a β conformation.

This chapter is organized as following. We begin by introducing the details of computational methods in Section 5.2. We then move on in Section 5.3 to discuss our results on the hydrogen bond network around tri-alanine peptide, comparing to the hydrogen bond network in bulk water system. Here we look at the other solvent coordinates like solvent coordinates and discuss electric field on the water wires. Finally, we end in Section 5.4 with some conclusions and possible future directions.

5.2 Computational Methods

5.2.1 MD simulations

MD simulations were performed using GROMACS 2016.4 [Pronk et al., 2013] package with AMBER03 [Duan et al., 2003] force field and the TIP3P [Jorgensen et al., 1983] water model. Al_3 is solvated in a cubic box of length 32 Åwith 1052 water molecules. The time step was set to 2 fs. The real space cut-off the electrostatic interaction and the cutoff of the van der Waals interaction were set to 10 Å. The van der Waals interaction is switched off smoothly from 9 Åto 10 Å. A neighbor list is updated every 10 steps and the cut- off distance of the short-range neighbor list was set to 11 Å. Particle Mesh Ewald method(PME) [Darden, York, and Pedersen, 1993; Essmann et al., 1995] was used to treat the long-range electrostatic interactions with the maximum grid spacing for the Fourier transformation of 1.2 Åand an interpolation oder of 4. The system was simulated within the NPT ensemble using the Berensden barostat [Berendsen et al., 1984] at 1bar. The characteristic time of the thermostat is 0.1 ps and that of the barostat is 2.5 ps. A long-range dispersion correction is applied for both energy and pressure for the van der Waals cutoff. All bond lengths of the alanine system are fixed by the LINCS algorithm [Hess et al., 1997]. The SETTLE [Miyamoto and Kollman, 1992] algorithm was used to fix the angle and bond length of the water molecule.

For some of our analysis we examined water wire distributions in bulk water. For these simulations, a periodic box of 512 water molecules in the cubic box of length 24.81 Å was used. The TIP3P water model was used for these calculations and the parameters for the molecular dynamics are the same as those used with tri-alanine.

5.2.2 Collective Variables for Clustering

In order to understand the coupling of protein and water, we need to establish a series of coordinates involving both the solute and water that will be used for extracting the intrinsic dimensionality as well as for the eventual clustering. These coordinates are described in the following.

5.2.2.1 Solute Coordinates

The conformation of an amino acid can be defined by a series of dihedral angles of the backbone. According to the Ramachandran plot, a pair of the dihedral angles, (ϕ, ψ), of each amino acid would characterize the conformational state of each residue in the peptide chain. In this work, we used the six dihedral angles of Al_3 for the dimensionality and clustering analysis. Besides the dihedral angles, there have been numerous studies that have shown that conformational changes in proteins is coupled to changes in the extent of backbone contacts involving amide hydrogen bonds as well as side chain packing of both polar and non-polar amino acids. These interactions are quantified by the backbone contact (BC) and sidechain contact (SC) details of which can be found in section 4.2.

5.2.2.2 Solvent Coordinates

We consider two types of solvent coordinates which have different degree of complexity. The first is the canonical water coordination number around the N-H and C=O groups. In the case of Al_3, there are a total of eight coordination numbers. To determine the coordination number, we use a smooth switching function like we did for the BC and SC. Details of the functional form can also be found in section 4.2).

The water coordination number, integrates out the orientational and topological properties of the hydrogen bond network that surrounds a biomolecule. There have been several studies pointing to the role of directed hydrogen bond networks around proteins. In a recent work [Jong, Grisanti, and Hassanali, 2017], we showed that for a 6-chain amino acid forming a sub-segment of the amyloid-beta 42 protein, there are numerous water wires threading the peptide chain and that link donor and acceptor groups. For the system examined in this work, the donor groups of the protein are the N-H groups of the backbone, while the acceptor groups are the C=O groups of the backbone and the C-terminus. The water molecules act as both the donor and acceptor groups.

To obtain the water wires, we construct an adjacency matrix with edges between donor and acceptor groups. The vertices(v) of adjacency matrix includes the all nitrogen and oxygen atoms of the termini and amide groups of the backbone as well as the all oxygen atoms of the water molecules. The weight of the edge(e_{ij}) between the vertices v_i and v_j is determined by the strength of a hydrogen bonding interaction. The weight is determined using a combination of several geometrical quantities such as the distance between the donor and acceptor atoms(d_{DA}), the distance between the hydrogen and the acceptor atoms(d_{OH}) and the angle between the hydrogen and the acceptor atoms at the donor atom($\angle$HDA) (see section 2.2 for more details). Since the Dijkstra [DIJKSTRA, 1959] algorithm is used identifying the shortest path, edges with stronger hydrogen bonds have a smaller weight.

5.2.3 Estimating the Intrinsic Dimension

Due to the presence of steric and chemical constraints, the configurations generated by molecular dynamics are approximately contained into a manifold of relatively small dimension. The dimension of this manifold, which is also the minimum number of the variables needed to accurately describe the system, is called intrinsic dimension(ID). Here, we compute the ID by TWO-NN, a recently introduce estimator [Facco et al., 2017]. In this approach, the value of the intrinsic dimension, d, is estimated from the empirical probability distribution of $\mu_i = \frac{r_{2,i}}{r_{1,i}}$ where $r_{1,i}$ is the distance of the first neighbor of point i and $r_{1,i}$ is the distance of the second neighbor.

It can be proved that the cumulative probability distribution of μ is given by $F(\mu) = 1 - \frac{1}{\mu^d}$ independently on the density of the point. This allows inferring the value of d also in data samples characterized by large variations of the density [Facco et al., 2017]. The same approach allows also quantifying the reliability of the estimate of d by block analysis. A meaningful estimate of the ID should indeed be robust with respect to the decimation of the data set.

5.2.4 Reconstructing the free energy landscape

The free energy of each data point is estimated by a parameter-free extension of the k-NN density estimator [Mack and Rosenblatt, 1979] introduced in Ref. [Rodriguez et al., 2018]. In the kNN method the free energy is estimated by $F_i = -\log\left(\frac{k}{r_{k,i}^d}\right)$, where $r_{k,i}$ is the distance of the k-th neighbor of point i and k is a parameter that should be fixed. In the approach of ref. [Rodriguez et al., 2018] the value of k is found independently for each data point by searching, by a statistical test, the largest neighborhood in which the free energy in the embedding manifold can be considered constant. Therefore, in regions in which the free energy varies quickly k will be small, while in regions in which the free energy is smooth k will be large. The approach also allows estimating the error on the free energy of each point. Prior knowledge of the intrinsic dimension d of the embedding manifold is essential for performing the statistical test.

The free energy minima and the saddle points separating them are then found by the extension of the Density Peak clustering approach [Rodriguez and Laio, 2014] introduced in Ref. [d'Errico et al., 2018]. In this approach, the free energy minima are found fully automatically, without choosing a priori their number. The approach provides an estimate of the statistical reliability of each free energy minimum, quantified by its Z-score

$$Z_c = \min_{c'} \frac{(F_{cc'} - F_c)}{(\varepsilon_{F_c} + \varepsilon_{F_{cc'}})}. \tag{5.1}$$

Here c and c' label two different free energy minima; F_c is the free energy of the minimum c; $F_{cc'}$ is the free energy of the saddle point between c and c'; ε_{F_c} and $\varepsilon_{F_{cc'}}$ are the corresponding errors.

5.3 Results and Discussion

5.3.1 Water Networks: Tri-Alanine and Bulk Water

Molecular dynamics simulations of ∼700ns at ambient temperature and pressure were used to examine the free energy landscape of the solvent around tri-alanine in its different configurations. We first determine the intrinsic dimension (ID) of the manifold containing the configurations by the approach recently proposed by Facco and coworkers [Facco et al., 2017]. This algorithm is specifically designed to estimate the ID of data harvested from highly non-uniform probability distributions. Indeed, in this approach the ID measure depends only on the distance of the first and of the second neighbor of each data point. This makes the ID estimate largely independent on variations of the density of the data points, and on the curvature of the manifold containing the data. This procedure also provides an estimate of the length scale over which the ID value is meaningful. Note that the ID is well defined only on a length scale on which the curvature of the manifold containing the data can be neglected.

Following the determination of the ID, a clustering analysis is performed which allows for an accurate determination of the topography of the free energy landscape. In this procedure, the free energy is determined directly on the manifold containing the data point by a procedure based only on the definition of a measure of similarity between the configurations [Rodriguez et al., 2018]. The minima of the free energy landscapes are then found by a non-parametric generalization of Density Peak clustering [Rodriguez et al., 2018]. Technical details can be found in the Methods section. The conformations that tri-alanine populates have previously been analyzed using the three pairs of Ramachandran dihedral angles (ϕ and ψ) [Marinelli et al., 2009]. A total of eight statistically meaningful conformations (or clusters) are present, consistent with previous work [Marinelli et al., 2009].

For simplicity, we focus our discussion in the first part of the manuscript on examining the behavior of water around two limiting conformations, namely (α, α, α) and (β, β, β) where the three indices correspond to the three pairs of Ramachandran angles. Physically, these two conformations correspond to the most alpha-helical and beta-sheet like configurations. We begin by examining the FES along two types of solvent coordinates. The first solvent coordinate is a standard coordination number (CN) of each amide and carboxyl groups of the peptide. These coordinates do not take into account the long range correlations in the hydrogen bond network. The second solvent coordinate we considered is the length of the so-called, water-wire (LW) which is made up of directed pathways of hydrogen bonds from every N-H group to carboxyl group of the amide bonds of the peptide. Water wires have been shown in previous studies to be useful in elucidating directional correlations in liquid water [Collins and

Washabaugh, 1985] and around proteins [Jong, Grisanti, and Hassanali, 2017]. See Methods for technical details on how CN and LW are constructed.

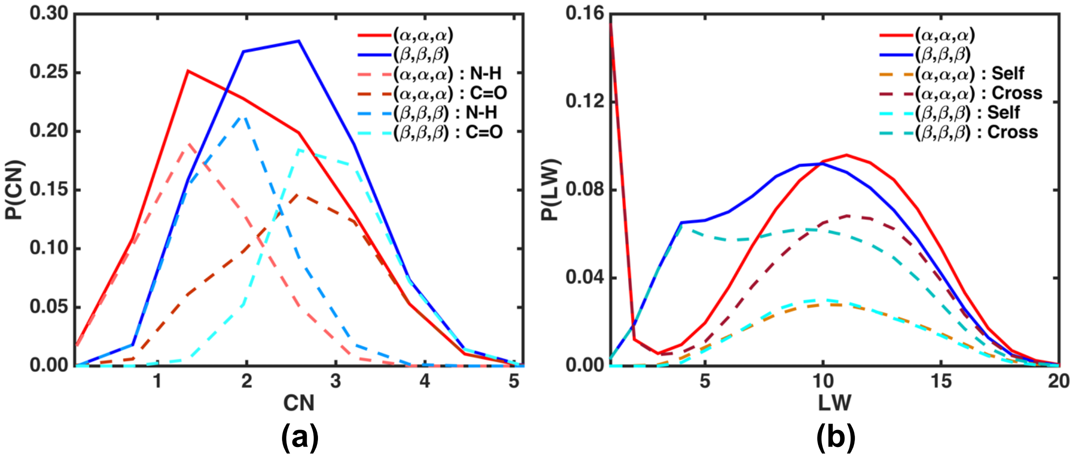


Figure 5.1: Probability distribution of the coordination number (a) and probability distribution of the path length for alpha and beta configurations (b)

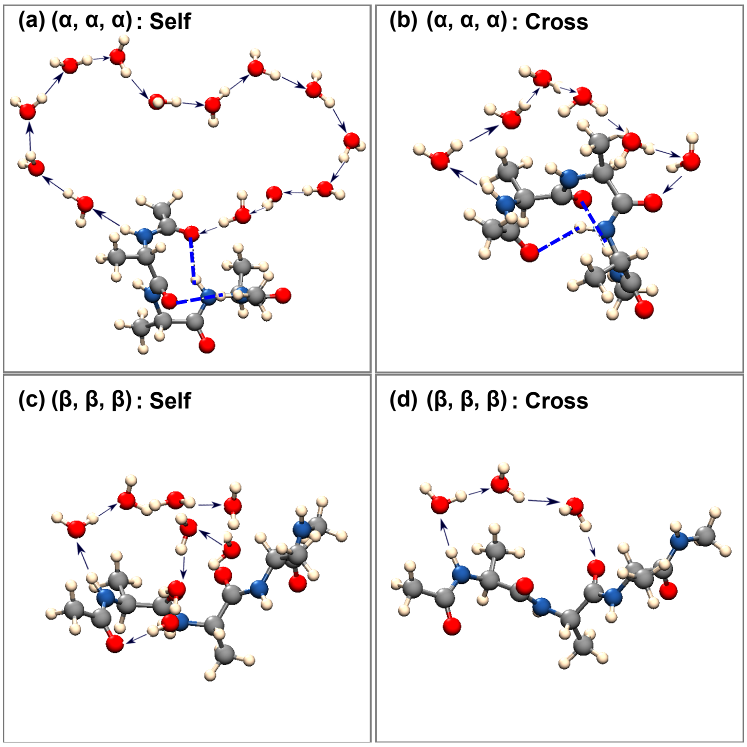


Figure 5.2: Hydrogen bond water wires : self hydrogen bond water wire connecting first N−H group and first C=O group of all-alpha structure (a) , cross water wire connecting first N−H group and third C=O group of alpha wire (b) , self water wire between first N−H group (c) and first C=O group of beta structure and cross wire between first N−H group and third C=O group of beta structure (d) . The direct Hydrogen bond is plotted with blue dashed line. O:red, N:blue, H:white, C:gray

The left and right panels of Figure 5.1 illustrate the probability distributions of

CN and LW respectively, for the all-alpha (α, α, α) and all-beta (β, β, β) structures. It is clear that all-beta is more solvent exposed given that the average coordination number around each group is larger by about 1-2 water molecules compared to all-alpha. Also shown in those distributions, are the separate contributions coming from the water around the N-H and C=O groups where we see clearly that the carbonyl groups tend to exhibit larger coordination numbers on average.

The right panel of Figure 5.1 shows the distribution of LW averaged over all the 16 possible paths formed between any N-H and C=O group of the peptide. Also shown are the distributions formed between the same amide group (denoted self) and across different amide groups (denoted cross). The distribution of the water wires length reveal a remarkable difference between the all-alpha and all-beta conformations. This is a first indication that the structure of the solute leaves an imprint on the water network around it. The all-alpha cluster features a sharp peak at LW=1 and a second broader peak around 11. The LW distribution for all-beta exhibits a peak at LW=9 and a small but significant shoulder at ∼4 that is clearly missing from all-alpha. The peak at LW=1 for all-alpha corresponds to intra-hydrogen bond interactions between the N-H and C=O groups of the peptide (see dashed blue lines shown snapshots of Figure 5.2). These paths do not involve any water molecules. On the other hand, the longer paths explicitly involve water wires passing through the solvent and connecting different amide groups. Another interesting feature of LW for all-alpha is the apparent gap between path lengths of 1 and ∼5 which originates from the fact that the alpha secondary structure imposes steric constraints on the formation of short paths between different donor N-H and acceptor C=O.

Moving on the all-beta secondary structure, since it is completely solvent exposed, it yields qualitatively shorter wires on average compared to the all-alpha structures. The major peak for all-beta is characterized by shorter paths compared to all-alpha since the entire peptide is somewhat uniformly wetted. Interestingly, we see that the shoulder at ∼5 in the beta structures, mostly originates from paths connecting different amide groups on the peptide chain which tend to be shorter (cross) than those formed within the N-H and C=O groups that are co-planar and oriented in opposite directions leading to longer directed path (self). The self-wires for all-alpha and all-beta are characterized by a more subtle difference with the former having a peak at larger values. The differences may become more pronounced for larger secondary structures and also depending on the chemistry of the side-chains involved. Figure 5.2 shows some snapshots from the simulations illustrating the directed water wires around all-alpha (panels a and b) and all-beta (panels c and d) structures where these features for the two limiting secondary structures can be visually appreciated. These results form one of the essential findings of this work namely that different conformations of the solute leave a strong fingerprint on the water network. This feature is captured much less clearly by the ordinary coordination number.

In order to appreciate the water wire distributions around the small peptide better, it is instructive to study the water wire distributions in bulk water without the peptide to compare and contrast the similarities and differences. To this end we examined three types of directed water wires in the bulk: water wires from each water molecule to the nearest neighbor shell defined as waters within 3.5 Å of that water, water wires from each water molecule to the second neighbor shell defined as those waters between 3.5 - 5.7 Å of that water and finally self-wires, which are water wires that connect a water to itself. The top left panel of Figure 5.3 shows the probability distributions of these three types of water wires. Both the first and second neighbor distributions feature a peak at wire length 1 and 2 respectively, followed by a broad shoulder between 3-15. The self-loops are peaked at about 6 consistent with previous studies [Rahman and Stillinger, 1971; Davis and Litovitz, 1965]. Panels b)-d) of Figure 5.3 show examples of these types of wires that exist in bulk water.

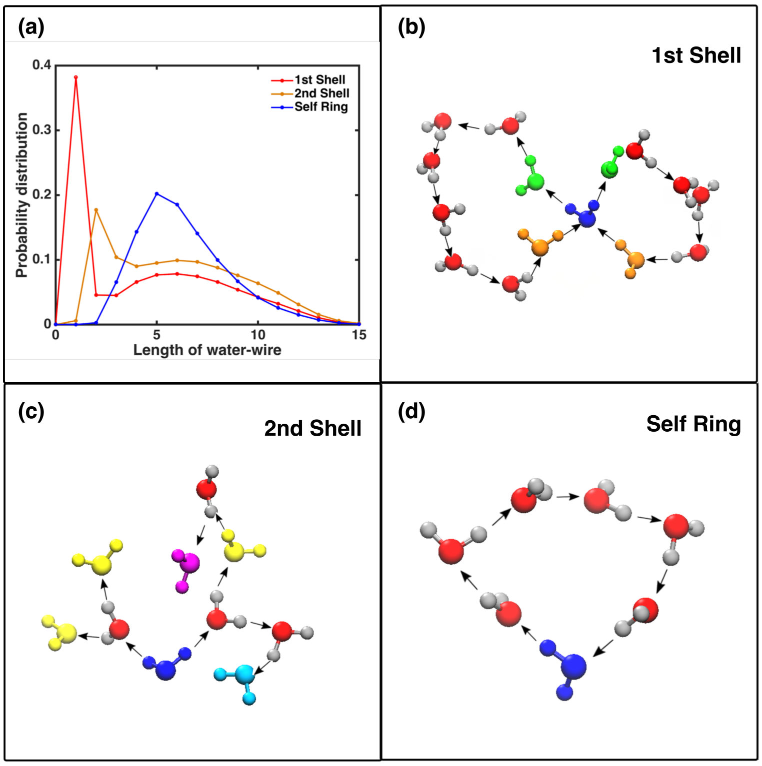


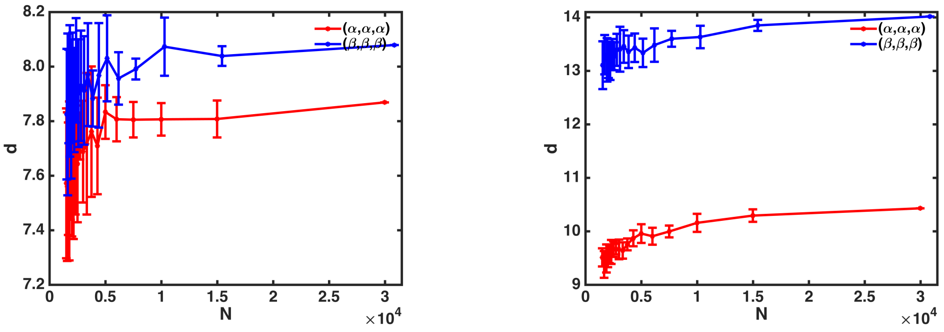
Figure 5.3: Probability distribution of the path length of water wire for the water molecules on the first shell and the second shell as well as the path length of the self ring (a), the water wire connecting the original water (blue color) molecules to the two donor water molecules (orange color) and two acceptor water molecules (green color) in the first shell (b), the water wires for the water molecules in the second shell: water molecules of length of 2 (yellow color), water molecules of length of 3 (cyan color) and water molecule of length of 4 (magenta color) (c) and water wire of self ring (d).

There are some interesting features of the water wire distributions in the bulk that bear some similarities to what we observe in Figure 5.1. The path length distributions for the alpha-structure has similar features to the water wire lengths in the bulk for first and second shell. In the case of the alpha-structure, the wires of length one involve direct hydrogen bonds between the N-H and C=O groups rather than hydrogen bonds within the first solvation shell. Nonetheless, one can see that donor and acceptor groups in the bulk play a similar role to those in the peptide. On the other hand, the self loops in the peptide tend to be longer than in the case of water due to the larger effective length of the amide-peptide bond compared to a water molecule. The conclusion we draw here is that given the flexibility of the water wires in the bulk, they tend to adapt to the conformational restraints of the peptide currently studied. How general these features are will require a more exhaustive investigation of other proteins with varying degrees of polar and non-polar groups. This will be the subject of a forthcoming study.

5.3.2 Dimensionality and Clustering

5.3.2.1 Dimensionality

Given the stark differences in water coordinates around the two secondary structures it is interesting to examine how this is reflected in the free energy landscape of the solvent around the molecule, considering the CN and LW coordinates. Initially, the landscape of CN and LW was determined for the all-alpha and all-beta structures separately. To obtain an accurate description of the FES, an estimate of the ID was determined using the method described earlier. In Figure 5.4, we show block analysis of IDs estimated for the CN and LW. For CN, we find that the ID for the alpha structure is 7.81±0.08, while for the beta structure it is 8.04±0.04. With LW, the corresponding ID values for the secondary structures are 10.30±0.12 and 13.85±0.10



(a) (b)

Figure 5.4: Block analysis of the dimensionality of the coordination number (a) and the path length (b) for the (α,α,α) and (β,β,β) clusters along the block size.

Regardless of the choice of the coordinate, it is clear that the ID along the solvent coordinate is surprisingly large. Physically what this means is that there are a large number of independent directions in which the water network around the peptide chain can change. Interestingly, for both solvent coordinates, the ID of alpha is smaller than that of beta although this difference is much more pronounced when the water wire coordinate is used. The higher ID around the beta-sheet reflects the fact that since it is more exposed to the solvent, the number of directions in which the water network can rearrange is larger than in the alpha-helix.

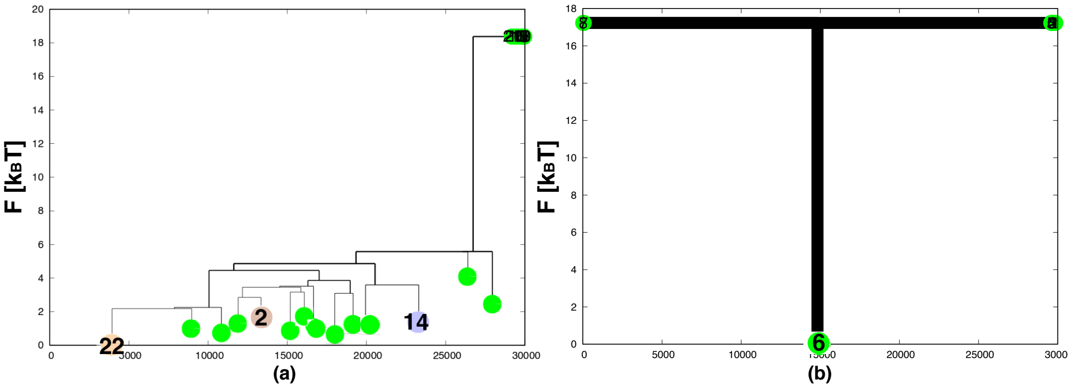


Figure 5.5: Free energy dendrogram of the clusters calculated by the clustering with the path length coordinates (a) and the coordination number coordinates (b) for (α, α, α) at the temperature of 300K (Parameters of clustering: confidence Z is 1.0, smallest allowed barrier is zero and the population threshold to be shown is 300). The free energy minima on both left and right panels are set to zero. The three clusters of 22, 14 and 2 with large population in the path length clustering (left panel) are depicted with the colors of orange, blue and brown.

5.3.2.2 Clustering

With a reliable estimate of the ID in hand, we then estimate the free energy landscape of the solvent. Once again, we consider separately the α and the β structure. A convenient way of representing the statistically relevant free energy minima is to construct a dendrogram [Facco et al., 2017] Each node of the dendrogram in Figure 5.5 implies a minimum of the free energy landscape and the horizontal connection between the nodes means that they are neighboring basins. Moreover the crossing point between the horizontal and vertical lines indicates a saddle point between basins. The dendrogram on the left in Figure 5.5 represents the landscape of the alpha-structures with LW at a statistical confidence level $Z = 1$ (see Methods). It is characterized by ∼15 different local minima and is thus very rough. Repeating the same clustering procedure for these structures with the CN coordinate leads to a rather featureless dendrogram with only a single cluster as seen in the right panel of Figure 5.5.

In Figure 5.6, we showed the dendrograms of LW and CN coordinates for the beta structures at a confidence $Z = 1$. Both LW and CN dendrograms are rather featureless showing only a single cluster.

The free energy minima using the LW coordinates are characterized by a confidence value of 1.0 or less, and of 1.0 or less using the CN coordinates. To assess more quantitatively the roughness of the free energy landscapes and the quality of LW and CN to discriminate differences in secondary structure, we constructed a measure of the roughness, called granularity, as $-\sum_i \frac{N_i}{N} log \frac{N_i}{N}$ where N_i is the population

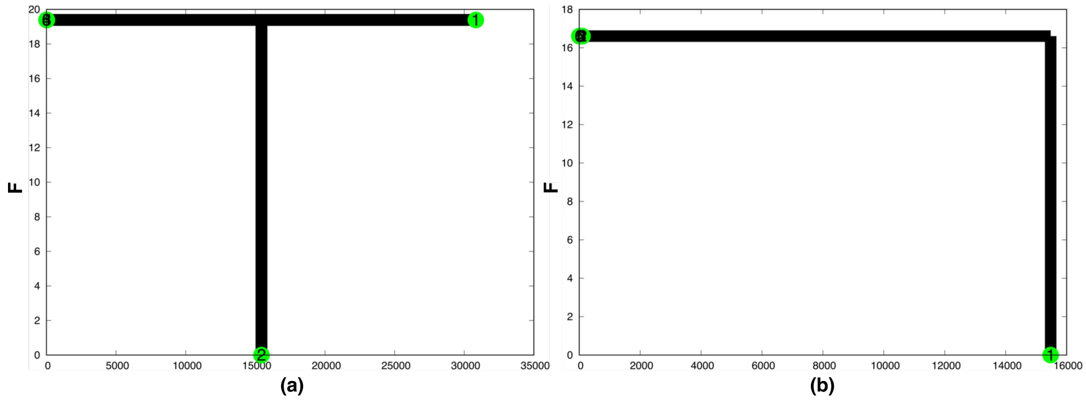


Figure 5.6: (a) Free energy dendrogram of the clusters calculated by the clustering with the path length coordinates and (b) the coordination number coordinates for (β, β, β) at the temperature of 300K (Parameters of clustering: confidence Z is 1.0, smallest allowed barrier is zero and the population threshold to be shown is 300). The free energy minima on both left and right panels are set to zero.

of the cluster i and N is the total number of the population. The granularity for all-alpha and all-beta for water wire coordinates and coordination number is shown in the left and right panels of Figure 5.7, respectively as function of Z-score. The granularity measure confirms that the all-alpha free energy landscape is consistently rougher along the water-wire space compared to all-beta. Furthermore, regardless of the choice of the Z-score, LW consistently leads to a much rougher free energy landscape compared to the CN coordinate. Overall our results show that for the all-alpha case, the different states along water wire space involve rather faint minima on the order to thermal energy. Hence the timescales associated with the fluctuations between these minima will occur on the sub-picosecond regime.

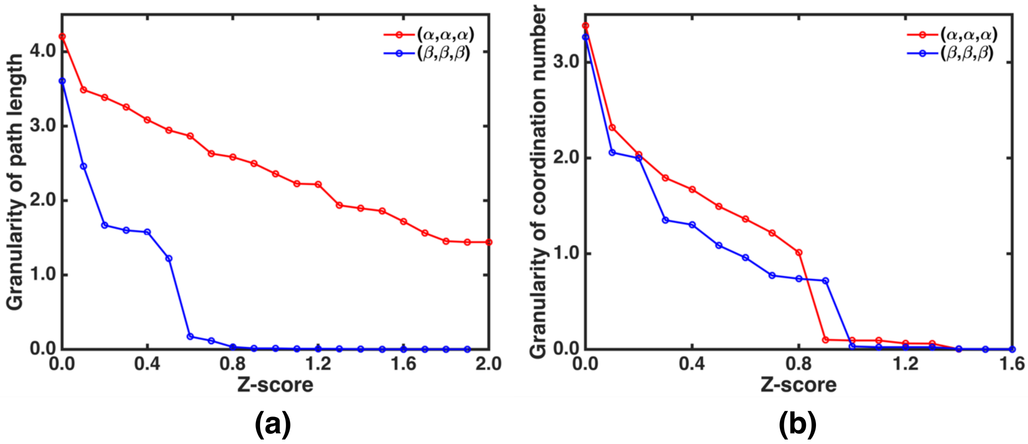


Figure 5.7: Granularity values along the Z-score for the path length coordinates (a) and for the coordination number coordinates (b)

Our results up to this point, show that the ID of the water around tri-alanine is rather unexpectedly large. In the case of LW, the FES for the alpha-structure reveals an underlying roughness implying that there are recurring patterns of water wires at fixed secondary structure. The fluctuations of the protein coordinates associated with tri-alanine can be generically mapped to changes in the extent of polar and non-polar interactions [Jong, Grisanti, and Hassanali, 2017] of the backbone and side-chains respectively, which can be probed using backbone or side-chain contact maps [Dyson, Wright, and Scheraga, 2006; Jong, Grisanti, and Hassanali, 2017]. To disentangle the molecular origins of these features, we examined the statistics of the LW coordinate for the five most populated clusters and also determined the extent of their BC and SC interactions.

Table 5.1: Population, LW, BC, SC and CN for (α, α, α) clusters (first 5 rows). The last row shows the same statistics for the single most populated (β, β, β) cluster.

	Cluster	Population	$\overline{\text{LW}}$	ΔLW	$\overline{\text{BC}}$	ΔBC	$\overline{\text{SC}}$	ΔSC
(α, α, α)	22	7872	9.4600	0.0130	1.3956	0.0062	1.4276	0.0034
	14	5522	8.9529	0.0167	1.5626	0.0076	1.3707	0.0042
	4	2433	9.5194	0.0236	1.4039	0.0105	1.1725	0.0053
	2	2353	9.8481	0.0220	0.9535	0.0077	1.5568	0.0070
	3	2187	9.2663	0.0279	1.5616	0.0099	1.3051	0.0057
(β, β, β)	2	30790	9.2243	0.0054	0.6242	0.0023	1.0834	0.0009

Table 5.1 shows that the different clusters found in Figure 5.5 are characterized by subtle changes in their LW, BC and SC values. Specifically, we observe an anti-correlation between the extent of backbone contacts and the length of short wires. This feature is seen more clearly in the distributions of these parameters shown in Figure 5.8 for a couple of the clusters.

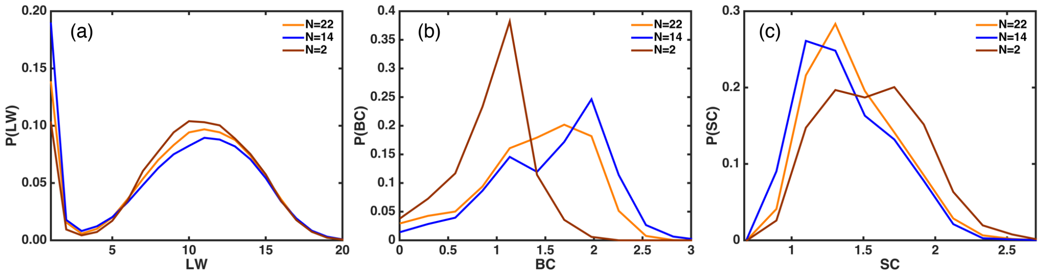


Figure 5.8: Probability distribution of LW (a), probability distribution of BC (b) and probability distribution of SC (c) for the clusters of 22, 14 and 2

Earlier we saw that for the all-alpha structures, there was a characteristic peak at LW=1 involving single hydrogen bonds between amide groups of the peptide. Panels (a) and (b) of Figure 5.8 show that those clusters with a higher likelihood of having these direct contacts and have higher backbone contacts, as expected. Besides BC, in the panel (c) of Figure 5.8 the SC parameter also appears to reflect some changes for the different clusters suggesting that the clustering along LW reveals changes in hydrophobic packing within the peptide. The bottommost row of Table 5.1 shows the same parameters for the only cluster in all-beta which clearly has the smallest BC and SC values compared to all the clusters.

A variable that collects the information of all the wires into a single quantity is the global efficiency defined as $E_g = \frac{2}{N(N-1)} \sum_{i,j \in N, i \neq j} \frac{1}{\ell_{ij}}$, where N is the number of vertices (N−H and C=O groups) and ℓ_{ij} is the shortest path length between i and j vertices. The global efficiency can be thought of as a variable that quantifies the connectivity of all the wires forming the 3D network around the peptide. We determined the global efficiency for the three clusters that formed the focus of this work namely, 22, 14 and 2. The global efficiency distributions are shown in panel a of Figure 5.9.

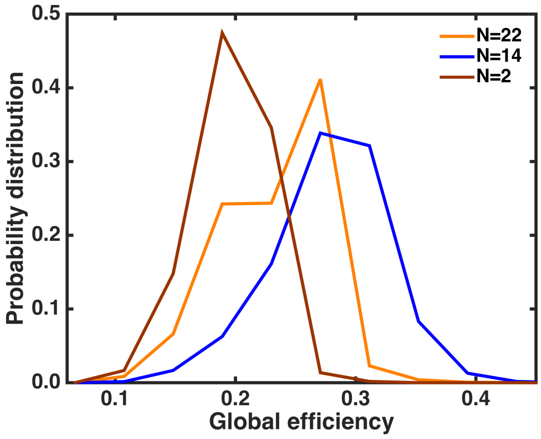
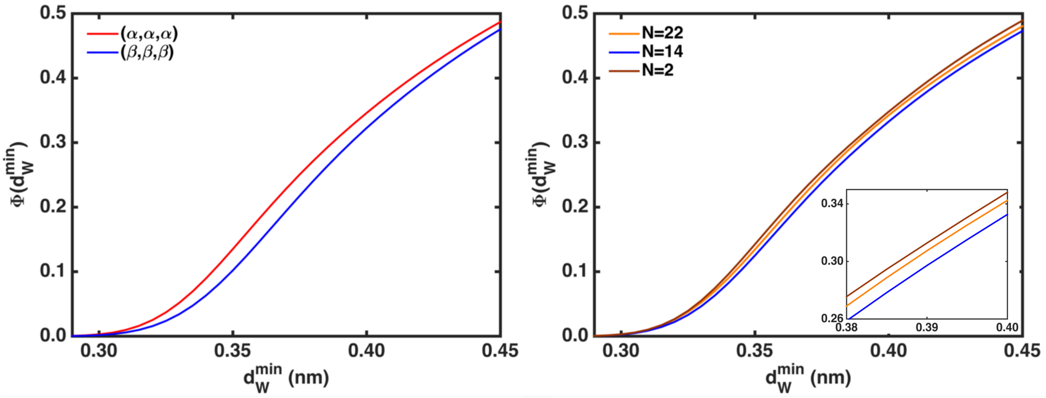


Figure 5.9: Probability distribution of the global efficiency of the clusters 22, 14 and 2 of the structures with the all-alpha constraints .

The FES as a function of the LW coordinates also provides insight into the coupling between the water network and hydrophobic packing involving different secondary structures. In the case of tri-alanine, the all-alpha structure is characterized by an enhanced packing of methyl groups of the side chains. To determine the spatial extent of the water wires relative to the hydrophobic groups, we determined the distribution of the shortest distance between water molecules in the wires to any hydrophobic C_β atom. The left panel of Figure 5.10 shows this distribution for the all-alpha and all-beta secondary structures where it is clear that the water wires are on average closer to the hydrophobic groups for the latter compared to the former. In the right panel of Figure 5.10 similar distributions are shown for the three clusters in panels (a)-(c) of Figure 5.8 where it is seen that the configurations with larger hydrophobic packing (enhanced SC), tend to form water wires where water molecules are slightly closer to the hydrophobic groups.

While we have focused all of our analysis of the FES of the solvent around a subset of secondary structure ensembles namely all-alpha and all-beta, this constraint can be relaxed. To provide a global landscape of the coupling between the water network and the protein conformation, a clustering was performed using LW on all the protein configurations from the trajectory. The key findings are summarized in



(a) **(b)**

Figure 5.10: Cumulative distribution (Φ) of the shortest distance between water molecules in the water wire and the any hydrophobic C_β atoms (a) for the dihedral clusters of (α,α,α) and (β,β,β), and (b) for LW clusters of 22, 14 and 2 in (α,α,α) dihedral cluster.

Figure 5.11 where the correspondence between dihedral clusters and LW clusters are shown. Similar to our earlier results, we see that the LW clusters discriminate between the alpha and beta structures quite well and that the alpha basins is characterized by more roughness compared to that of beta.

A more quantitative measure of the extent of entanglement between the protein conformation and water network as measured by the dihedral and water wire coordinates respectively, can be obtained by computing a normalized mutual information (NMI) [Pfitzner, Leibbrandt, and Powers, 2009]. An NMI value of 0 would imply that a partitioning of clusters obtained using water wires and dihedrals are independent, while a value of 1 would imply that the two partitions are completely correlated and that each cluster from the water wire partitioning would yield a unique dihedral cluster. We find that the NMI between clustering performed based on dihedral and LW yields a value of 0.32 implying that the two partitions are significantly correlated. This feature is essentially what we observe in Figure 5.11 involving the mapping between dihedral and path clusters. As expected, the NMI between dihedral and CN partitions is 0.0002, showing again that this solvent coordinate washes out important information about the coupling between peptide and water degrees of freedom.

5.3.3 Electric Fields and Water Wires

The presence of the water wires connecting donor and acceptor groups in the peptide is an indirect indication of the connection of oriented water molecules. Several previous studies have pointed to the importance of electric field fluctuations along hydrogen bonds in bulk water as well as around proteins. It is thus interesting to explore if there exists a correlation between the presence of the water wires and the electric field experienced along water molecules in the water wires.

We adopt a strategy for determining the electric field proposed by Dellago and coworkers [Reischl, Köfinger, and Dellago, 2009]. Specifically, we estimated the electric

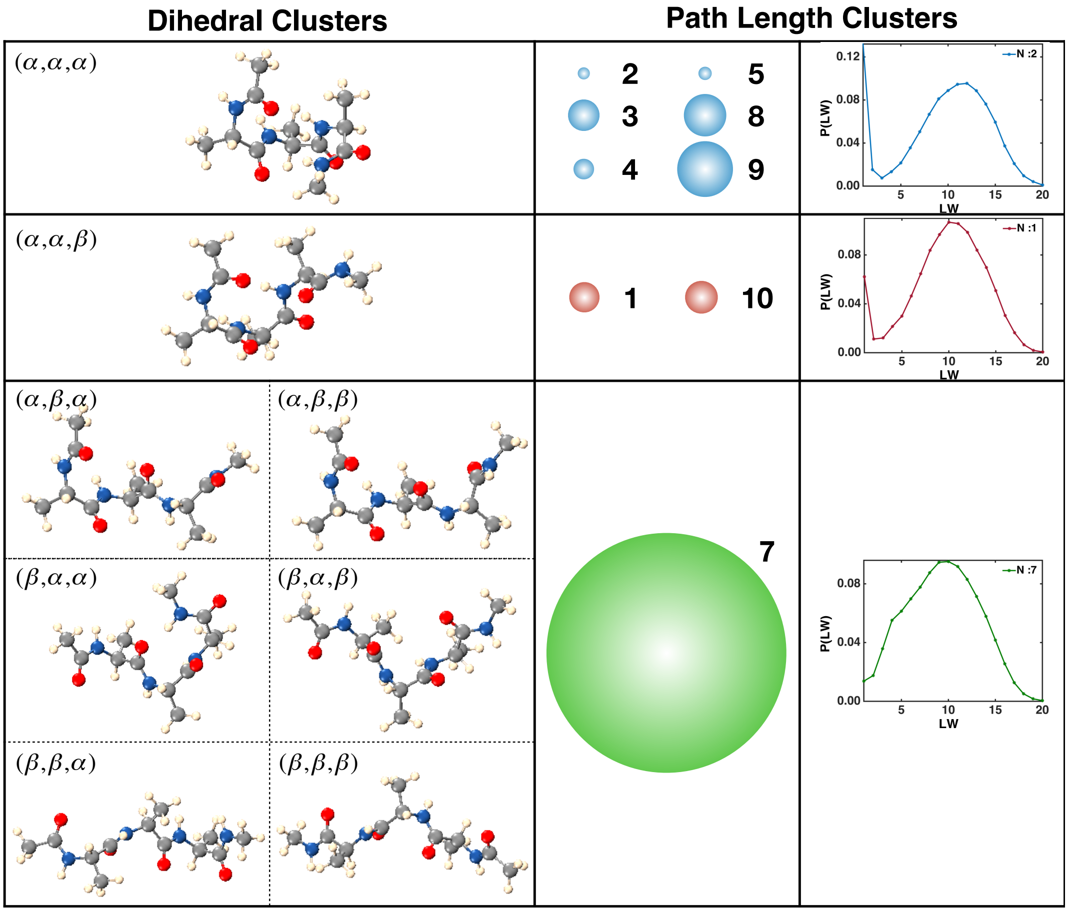


Figure 5.11: Outline of dihedral angle and path length clusters : dihedral clusters (left panel) and corresponding path length clusters with their path length distributions (right panel). The size of the circle of the path length clusters shows the relative population with respect to the population of the cluster 2 (=1750).

field at the center positions of OH bonds of the water molecule by calculating the electrostatic force acting on the test charges located at the center of OH bonds. The electrostatic force was obtained by the Ewald summation projecting on the local Cartesian coordinates of each OH bond where the origin lies on the center of OH bond, unit vector $\vec{n}_x$ lies in the direction of OH bond, pointing toward the hydrogen atom, unit vector $\vec{n}_y$ is perpendicular to $\vec{n}_x$ in the corresponding water molecule plane and unit vector $\vec{n}_z$ is perpendicular to the water molecule plane. We begin by showing in Figure 5.12 the probability distributions of x, y and z components of the electric field (E_x, E_y and E_z) for water molecules in all the water wires around the tri-alanine peptide and compare them to water molecules in bulk water.

Similar to previous studies, the distributions of the electric field are essentially Gaussian due to the central limit theorem. Table 5.2 shows the mean and variance of the distributions obtained from Gaussian fits. The E_x component is on average larger owing to the fact that this represents strong forces along the hydrogen bond direction. Interestingly, as seen from both the distributions and in Table 5.2, the E_x component for water molecules in the wires is larger than than in the bulk by about

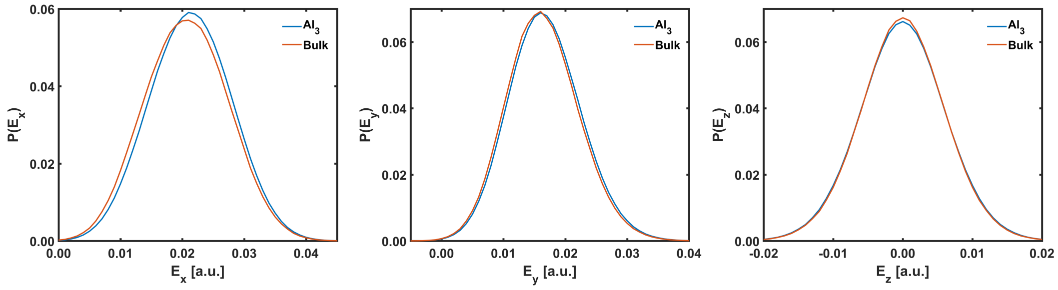


Figure 5.12: Probability distribution of the electric field on the water molecules on the water wires around the tri-alanine peptide and in bulk water in x-direction (a), in y-direction (b) and in z-direction (c) . 1 [a.u.] ≈ 51.422 [V/Å].

0.007 [a.u.].

Table 5.2: The values of mean and variance of the distribution of the electric fields on water molecules on the water wires around the peptide and in the bulk water, obtained from Gaussian fitting.

	Al_3		Bulk water	
	μ [a.u.]	σ [a.u.]	μ [a.u.]	σ [a.u.]
E_x	0.0213	0.0068	0.0206	0.0069
E_y	0.0164	0.0058	0.0161	0.0058
E_z	0.0000	0.0060	0.0000	0.0060

The electric field distributions shown in Figure 5.12 involve the contributions of both the peptide and all the other water molecules in the system, except of course, the water molecule that is tagged. We did not find any correlations between the length of particular wires and the total electric field from the system. However, when we separate out the contribution of the field coming from the peptide alone, we observe some interesting trends which are shown in Figure 5.13.

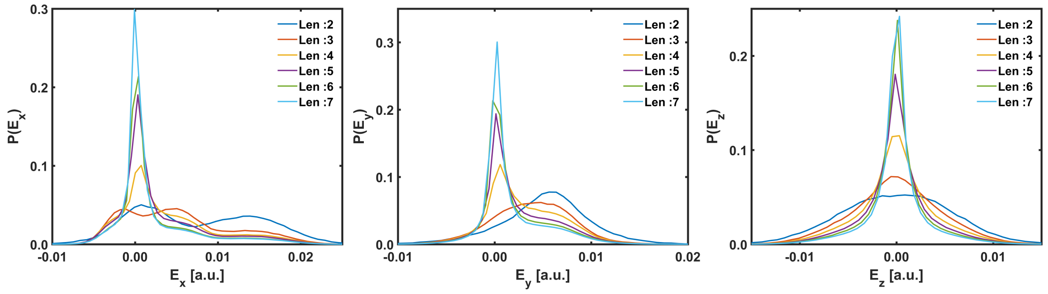


Figure 5.13: Probability distribution of the electric field contributed only by the peptide for the water molecules on the different lengths of water wires in x-direction (a), in y-direction (b) and in z-direction (c) .

Intriguingly, we see that E_x and E_y is characterized by larger fluctuations for shorter wires than longer wires. Furthermore, we also observe that the field contributions from only the peptide are characterized by a peaks centered close to zero due to the opposite orienting effects of N-H and C=O groups on the hydrogen bonds.

5.4 Conclusion

In conclusion, we have taken first steps in this work to elucidate the coupling between peptide structure and the surrounding water medium using a combination of data-science oriented algorithms. We show that these newly developed methods provide rich insights into the physical chemistry of hydrated biomolecules. Specifically, water around the small peptide tri-alanine, resides on a high dimensional manifold which appears to be sensitive to the type of water coordinate as well as the underlying secondary structure of the peptide. We show that directed water wires enveloping the peptide carry more information about the free energy landscape than the canonical radially averaged coordination number.

Besides the water wires from donor to acceptor groups, one can also include other topological properties of the water network such as wires forming between two donor N-H groups or two acceptor C=O groups. In the case of tri-alanine studied here, performing the clustering including these other variables increases the NMI marginally (0.4). It would be interesting in the future to explore the dimensionality of water using different water network coordinates and subsequent clusters emerging from larger and more chemically complex biomolecules.

We also examined the correlation between water molecules along water wires and the electric field they experience. Overall, we find that the electric fields along the direction of the hydrogen bonds for water molecules in the water wires is slightly more enhanced than in bulk water. By examining the electric field contributions coming from the peptide alone, we are able to identify stronger correlations between the electric field and the wire lengths. Again, it would be interesting in future work to examine the correlations between the electric field fluctuations and water wires for more complex biological systems.

Chapter 6

Low Energy Optical Excitations as an Indicator of Intra Hydrogen Bond Interactions

A version of this chapter is to be submitted:
K. Jong, Y. T. Azar, A. Stephens, L. Grisanti, G. Kaminski and A. Hassanali. Low Energy Optical Excitations as an Indicator of Hydrogen Bond Interactions in Amyloid Proteins

In this chapter, we examine the origin of low energy excitation of Amyloid β_{30-35} fibrils without aromatic residues.

6.1 Introduction

Amyloid aggregation is a process by which proteins assemble into fibrils characterized by a high content of β-sheets [Chiti and Dobson, 2006; Greenwald and Riek, 2010; Morriss-Andrews and Shea, 2015]. The formation of fibrils is associated with neurodegenerative diseases such Alzheimers and Parkinsons [Ross and Poirier, 2004; Horwich, 2002]. In the past decades, there has been a lot of experimental evidence showing that small oligomers and protofibrils in the early stages of fibril aggregation, are in fact, more toxic [Horwich, 2002; Rochet and Lansbury, 2000; Mannini et al., 2014; Ladiwala et al., 2012]. An understanding of the physical and chemical processes associated with the initial stages of amyloid aggregation may help with coming up with medical therapies that can intervene at earlier stages of the disease development.

The stability of amyloid fibrils has typically been rationalized by the presence of a dense network of hydrogen bonds forming β-sheet structures [Greenwald and Riek, 2010; Jong, Grisanti, and Hassanali, 2017; "Cooperative hydrogen bonding in amyloid formation"]. Recent experimental using solid-state NMR spectroscopy however, show that 3D structure of Aβ_{1-42} fibrils is also characterized by a hydrophobic core made up of maximally buried hydrophobic side chains [Wälti et al., 2016]. Moreover, it has also been shown that the toxicity of the oligomers increases with increasing hydrophobicity [Rochet and Lansbury, 2000; Mannini et al., 2014; Ladiwala et al., 2012]. The interplay between the hydrophobic interactions between the side chains and the hydrophilic interactions through the dense hydrogen bond network and its role in amyloid aggregation still remains an open question.

One of the obvious experimental challenges in characterizing the mechanisms associated with aggregation like those discussed earlier, is being able to monitor the aggregation process. In this regard, there has been an increasing effort to understand the optical properties of amyloid proteins. Recent experimental studies have shown that amyloid fibrils develop an intrinsic fluorescence during aggregation [Maji et al., 2005; Chan et al., 2013; Pinotsi et al., 2016; Grisanti et al., 2017b]. It has also been observed that these protein aggregates can absorb low energy photons in the energy range of 240-500 nm [Johansson and Koelsch, 2017] with the measured excitation peak at around 365 nm [Pinotsi et al., 2016]. Remarkably, these features do not require the presence of aromatic amino acids.

These anomalous optical properties may not be exclusive to amyloid aggregates [Bhattacharya et al., 2017; Ye et al., 2017; Prasad et al., 2017; Chen et al., 2018]. The fluorescent experimental studies for a series of non-aromatic biogenic and synthetic peptides based on alanine, valine and isoleucine also shows that the intrinsic fluorescence in the aggregated state/condensed phase is associated with the abundant existence of hydrogen bonding between amide groups [Ye et al., 2017]. Recently, Prasad et al. [Prasad et al., 2017] showed that a monomeric protein devoid of aromatic residues features significant absorption between 250-300nm and a long tail in the absorption up to about 800nm. Using electronic structure calculations, they show that the charged amino (NH_3^+) and carboxylate (COO^-) of the spatially proximal lysine and glutamic acid side chains, act as electronic charge acceptors or donors for photoinduced electron transfer either from or to the polypeptide backbone or to each other. They also demonstrate that charge-transfer (CT) excitations involving these charged groups and the peptide backbone appear to be the main source for the optical activity of these proteins in the range between 220-380 nm. In specific cases, the increase of the intrinsic fluorescence during the aggregation for the model system of insulin and lysozyme suggest this phenomena are at least partially caused by the chemical process like oxidation [Tikhonova et al., 2018].

It is shown that this intrinsic fluorescence might be utilized as a label-free diagnostic tool to probe the structural and dynamical transition of amyloid-like aggregates. For example, Ansari et al. [Ansari et al., 2018] showed that the absorption intensities of the PEST fragment of human c-Myc and its mutant at wavelengths of 250-800 nm are dependent on the 3D proximity of the charged functional groups across the protein. Moreover, the significant changes in the spectra by changing pH in the range of 3-11 as well as by the application of different temperatures and salts show a strong correlation between their secondary structure and fluorescence. The increase of the absorbance spectra with time in Hen Egg-White Lysozyme at pH of 2 directly correlated with the growth of aggregates, as confirmed by the increasing thioflavin T fluorescence. Meanwhile, the sensitivity of absorption spectra by the charge transfer transition to the proximity of the different charged groups is also very supportive to the feasibility to employ the non-aromatic residues as a new luminescent biomolecules [Mandal, Paul, and Venkatramani, 2018; Prasad et al., 2017; Chen et al., 2018].

In this chapter, we combine fluorscence spectroscopy experiments and theoretical modeling to specifically examine the role of termini interactions on the optical properties. Optical absorption and fluorescence is measured for a six-chain amino acid 2Y3J (AIIGLM) which forms a segment of the full amyloid beta 1-40. In order to explore the sensitivity of the optical properties to the termini interactions, the experiments were repeated by acetylating the N-terminus. Although atomic force microscopy experiments indicate the formation of some form of fibrilar or crystal aggregates in both systems, the optical properties are strikingly different - acetylation significantly reduces optical activity between 280-350 nm.

In earlier studies, we have shown using density functional theory (DFT) based calculations, that strong hydrogen bonding interactions between the N and C termini of the amyloid proteins, play an important role in tuning their optical properties [Pinotsi et al., 2016]. Due the complexity of the amyloid fibril morphology, these calculations were performed using standard generalized gradient approximation (GGA) functionals on small model crystal structures. When used to determine optical properties, GGA functionals can often suffer from inadequately capturing the physics of CT excitations. Here, we use range corrected hybrid functionals to examine the electronic character of the optical excitations. We suggest that the low energy excitations observed in the experiment appear to be modulated by the extent of hydrogen bonding interactions which is affected by the specific conformations of the peptide. The low energy absorption of photons involves a mixture of CT excitation from C-to-N terminus and C-terminus to backbone groups. This region of the optical spectrum may thus provide a way to examine the extent of packing and specific hydrogen bond interactions in peptide aggregates.

The remainder of the chapter is organized as follows: Section 6.2 devoted to a brief explanation of experimental methods and employed computational methods for the structural preparations and ground (excited) states calculations. In section 6.3, we present our results for both ground and excited states as a function of structural parameters, emphasizing the role of all possible types of intramolecular interaction. Finally, we end in Section 6.4 with some conclusions.

6.2 Computational Methods

In the following section, we describe the computational methods that were performed to help interpret the observations made in the experiments. An important aspect of this work was to be able to use more accurate time-dependent density functional theory (TDDFT) functionals in order to characterize the nature of the optical excitations and the differences observed in the experiments. In order to achieve this, the model systems used for the theory need to be carefully identified which we will now describe.

6.2.1 Construction of Model Systems

The experimental systems studied for both NC and N-Acetylated involve very large-scale molecular aggregates that are currently computationally prohibitive to model. Furthermore, the morphology of both the fibril or crystal structures in the case of NC and N-Acetylated respectively, is currently not fully characterized. In the case of NC, there is a crystal structure available [Colletier et al., 2011] which formed the study of a previous work in group using GGA functionals [Pinotsi et al., 2016; Grisanti et al., 2017b]. However, the unit cell is made up of 752 atoms and thus one cannot use more accurate range-corrected hybrid functionals to study optical properties on this system. We thus focus our efforts on investigating the electronic structure and optical properties of a monomer unit of this structure where the role of the conformation of the peptide can be examined. There are two important motivations for this: firstly, as pointed out in the introduction, very recent experiments have shown that the fold of Aβ_{1-42} is stabilized not only by the well known polar interactions of the beta-sheets, but also the hydrophobic packing of the side chains [Wälti et al., 2016]. In addition, there have also been some studies showing that monomeric proteins lacking aromatic amino acids can also exhibit surprising optical properties [Prasad et al., 2017].

We computed the optical spectra using TDDFT for configurations of NC that were obtained from an earlier work in our group [Jong, Grisanti, and Hassanali, 2017]. We briefly describe some of the details here. We performed microsecond simulations of the NC chain in a box of water using well-tempered metadynamics simulations which allowed for a thorough sampling of the conformations of the chain. These simulations showed that there were at least four important collective variables involved in characterizing the various structures: end-to-end distance (d_{ee}), radius of gyration (Rg), backbone contacts (BC) which quantifies polar interactions within the chain and finally the sidechain contacts (SC) which measures the extent of hydrophobic packing involving the sidechain groups of the amino acids (see SI for the details of the definition of BC and SC). The free energy landscape along d_{ee}, BC and SC is shown in Figure 6.1. A total of 40 configurations were chosen in order to determine the optical spectra. These different configurations were chosen from different regions of the free energy surface (FES). To help with the discussion later on, we define four different basins (A, B, C and D) which are shown in Figure 6.1.

In order to examine the sensitivity of the optical properties to the acetylation of the N- and C- termini, we performed some calculations on two other model systems. N-Acetylated has a negatively charged C-terminus and is thus likely complexed with a positively charged counterion from the buffer solution. Two extended conformations obtained from basins C of NC were used to acetylate the N-terminus afterwhich a sodium ion was placed near the negatively charged C-terminus. These configurations were then optimized using force field AMBER99SB-ILDN [Lindorff Larsen et al., 2010]

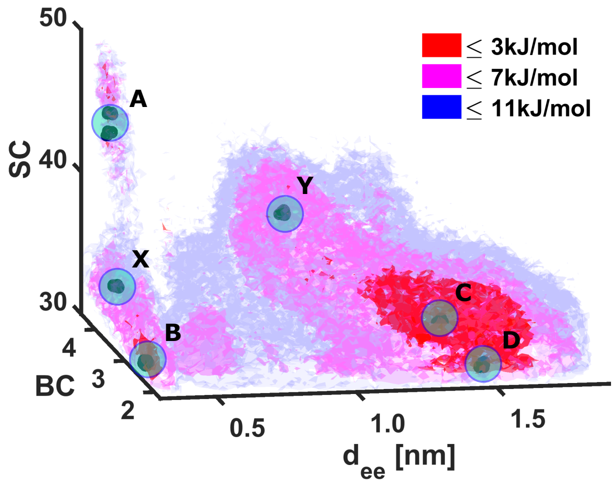


Figure 6.1: 3D free energy surface along d_{ee}, BC and SC for NC system.

keeping the backbone heavy atoms frozen. The second model system is inspired from an earlier work in our group [Jong, Grisanti, and Hassanali, 2017] where we conducted simulations of the NC system modified so that both the N and C termini were capped with methyl groups. This system is dubbed as MET. As expected the FES of MET along the coordinates illustrated in Figure 6.1 is qualitatively different (see SI Figure S2) due to the absence of the strong salt-bridge interactions between the termini. While we currently have no experiments on this particular system, it provides an instructive example of the importance of the N-C termini hydrogen bond interactions on the spectra.

6.2.2 Electronic Structure Calculations

All the optical properties are performed using TDDFT [Marques and Gross, 2004; Runge and Gross, 1984b] implementation in the Gaussian 09 code [Gaussian09 Revision E.01]. In all the ground and excited states calculations, we used a range corrected exchange-correlation functional (CAM-B3LYP) [Yanai, Tew, and Handy, 2004] with the 6-311++G** [McLean and Chandler, 1980; Krishnan et al., 1980] basis set. There have been numerous previous studies showing that CAM-B3LYP performs quite well in characterizing charge transfer excitations as well as Rydberg states [Peach et al., 2008; Brkljača et al., 2014].

Here, we briefly review the formalism behind computing the optical properties. Excitation energies and oscillator strengths were calculated solving the Casida equations,

$$\sum_{jb} \Omega_{ia,jb} F_{jb,I} = \omega_I^2 F_{ia,I} \tag{6.1}$$

where ω_I is the excitation energy of I-th excited state and $F_{ia,I}$ correspond to the contribution of the transition from an occupied KS orbital (ϕ_i), to a virtual orbital (ϕ_a) [Rüger et al., 2016], where i,j,... and a,b,... indices used for occupied and virtual orbitals, respectively. The solution to the equation yields the corresponding oscillator strengths for each excitation through the following:

$$f_I = \frac{2}{3}\,\omega_I |d_I|^2, \quad d_I = \sum_{ia} \sqrt{\frac{2\omega_{ia}}{\omega_I}} F_{ia,I} \vec{d}_{ia} \\ \vec{d}_{ia} = \langle \phi_i | \vec{r} | \phi_a \rangle \tag{6.2}$$

The transition dipole moment for the I-th excitation, d_I, can be expressed as a linear combination of single orbital transitions, d_{ia}. In order to obtain the spectra, these resulting oscillator strengths, f_I, are convoluted with Gaussian functions to obtain the molar absorptivity (ϵ) as a function of frequency (ω) in the following manner:

$$\epsilon(\omega) = 2.174 \times 10^8 \sum_I \frac{f_I}{\Delta_{1/2}} exp[-\,2.773 \frac{(\omega - \omega_I)^2}{\Delta_{1/2}^2}\,] \tag{6.3}$$

The full width at half maximum ($\Delta_{1/2}$) is 0.3 eV and f_I and ω_I are the Ith oscillator strength and excitation frequency, respectively.

In order to characterize the physical origin of the excitations, charge density differences, $\Delta\rho(\mathbf{r})$, are determined for some specific excitations. The spatial distance between the barycenter of positive ($\Delta\rho_+(\mathbf{r})$) and negative ($\Delta\rho_-(\mathbf{r})$) regions can be used to quantify the charge transfer character of each excitation [Savarese et al., 2017]. In addition, we also studied the dipole moment difference between the ground and excited states [Azar and Payami, 2015]:

$$\Delta\mu = \int \mathbf{r}\rho_{exc} d\mathbf{r} - \int \mathbf{r}\rho_{gs} d\mathbf{r} \tag{6.4}$$

where ρ_{gs} and ρ_{exc} are the electronic density of the ground and excited states.

As indicated earlier, we cannot currently perform TDDFT calculations on molecular aggregates of system sizes that are relevant to the experiment. However, the effect of the environment can be included in a mean field manner using polarizable continuum solvent models [Cramer and Truhlar, 1999; Tomasi, Mennucci, and Cammi, 2005] (PCM). For our calculations, we use a conductor-like polarizable continuum model (C-PCM) [Cossi and Barone, 2001] implemented in Gaussian. The static and optical dielectric constants for the protein were chosen to be 10 and 2.5 respectively, as used in several previous studies [Li et al., 2013; Pitera, Falta, and Gunsteren, 2001].

6.3 Results and Discussion

We begin by first showing the experimental results showing the effect of altering the chemistry of the termini group on the excitation and emission spectra, followed by a detailed analysis of our theoretical findings.

6.3.1 Experimental Results

The left and right panels of Figure 6.2 illustrate the AFM images obtained for the NC versus the N-Aceylated systems. While it is clear that NC forms fibrils, N-Acetylated on the other hand appears to form crystal-like structures. The exact morphology of these systems cannot be inferred from the AFM images. Specfically, it is also known that amyloid aggregates are highly polymorphic given the wide variety of structures that have been found using NMR and X-Ray crystallography [Wälti et al., 2016].

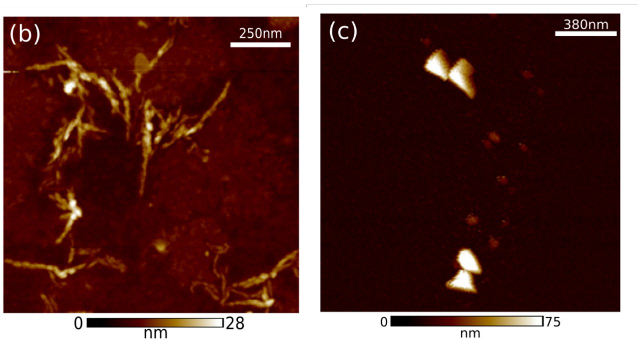


Figure 6.2: Atomic Force images of amyloid proteins in PeakForce tapping mode. Samples grown at pH7.4: (a) Aβ_{30-35}, (b) acetylated-Aβ_{30-35}.

Figure 6.3 shows the absorption spectra for the NC and N-Acetylated systems. Also shown, is the spectra obtained for the full Amyloid Aβ_{1-42}. Interestingly, the spectra for the NC and the full amyloid essentially overlap and thus NC serves as a good model system to understand the photophysics of the amyloid aggregates. Acetylating the N-terminus results in a drastic change in the optical absorption. Although measuring pure absorption spectra is hard for such samples and scattering is surely contributing to the background, the absorption intensity of the entire low-energy band between 250 and 450nm is significantly reduced in going from NC to N-Acetylated and the optical density of the latter overlaps with the noise from the background.

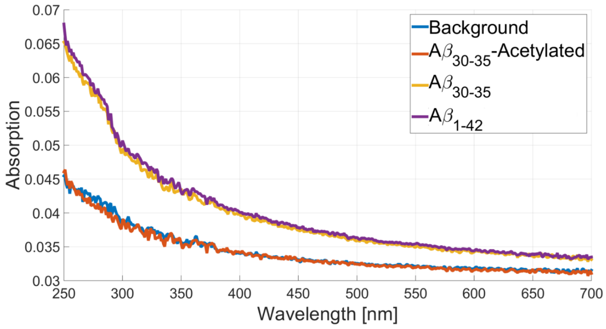


Figure 6.3: Absorption spectra of Aβ_{30-35} and acetylated Aβ_{30-35} compared to pH 7.4 buffer background

Fluorescence excitation and emission spectra were determined for both NC and N-Acetylated structures. The excitation was centered at 290 nm and the emission at

350 nm for both systems. Figure 6.4 shows these spectra for NC and N-Acetylated in the left and right panels respectively. Interestingly, for NC both the excitation and emission spectra are very similar to that of the full amyloid [Pinotsi et al., 2016]. Again, acetylation results in a relative reduction in the low energy component of the fluorescence (near-UV region).

The experimental findings show a striking sensitivity to the presence of having the zwitterionic state of the termini ends and the optical absorption/emission. There are several possible scenarios that could result in this behavior. If the low energy excitation band between 250 and 450nm involves charge reorganization near the N-terminus, it is clear that acetylating could quench this excitation. On the other hand, Figure 6.2 shows, that the morphology of the amyloid aggregate changes upon N-terminal acetylation. The exact details of how this change is manifested in the aggregate remains unknown. However, the change in optical properties could also be the result of differences in polar and non-polar interactions in the aggregate which ultimately change the environmental conditions and nature of the excitations. In the next section, we will explore using our simulations, the molecular origins of these trends.

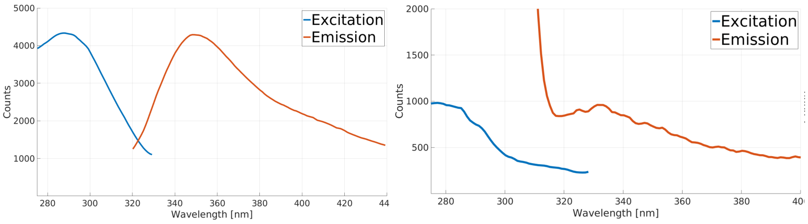


Figure 6.4: Emission and excitation spectra of the intrinsic fluorescence of amyloid protein solutions: (a)Aβ_{30-35}, (b) acetylated Aβ_{30-35}. For the excitation spectrum the emission was collected at 350 nm, and for the emission spectra the excitation was set at 290 nm. The extremely steep peaks in (b) are the artefacts of the emission and excitation windows overlapping.

6.3.2 Theoretical Results

We begin with our analysis of the electronic structure from the ground-state DFT calculations followed by the optical properties obtained with TDDFT. In particular, we will examine how the the optical properties are affected by the peptide conformation and specific environmental interactions.

6.3.2.1 Ground State Electronic Structure

In order to get a qualitative sense of the role of conformation and environmental effects on the electronic structure, we begin by showing in Figure 6.5 the KS energy levels for the occupied and virtual orbitals for all the forty configurations sampled from the FES. The KS energy diagrams in panel (a) of Figure 6.5 shows a significant difference in the HOMO-LUMO gaps between the compact states of A and B and the extended states of C and D in the NC system. While the values of HOMO-LUMO gaps for the states A and B are larger than 4 eV, those of states C and D are less than 2 eV. Panel (b) of Figure 6.5 shows the KS energy diagrams when the protein environment is included using the PCM model. The HOMO-LUMO gaps for the compact states change by a very small amount of ≈0.1 eV, for the extended states the gap increases by ≈ 4 eV.

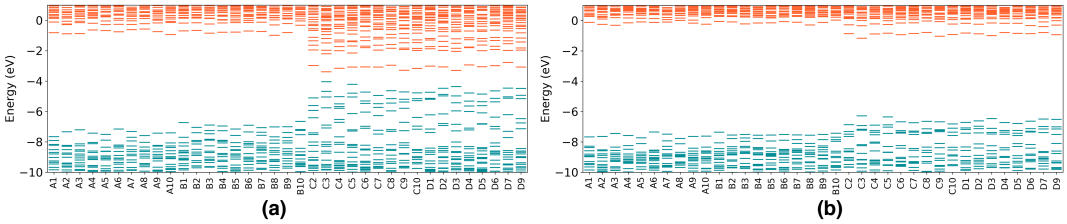


Figure 6.5: KS energy diagram for the configurations of the states A, B, C and D of NC system (a) in gas phase, (b) in the protein environment. Cyan and orange bars represent HOMO(s) and LUMO(s) respectively.

Lida et al. [Iida et al., 2009] have discussed a systematic framework for understanding the orbital energy shifts that arises due to the effect of solvation. The orbital energy shift of ith level, $\Delta\epsilon_i$, is proportional to the dot product of the change in the dipole moment associated with ith molecular orbital ($\vec{\Delta\mu_i}$) and total dipole moment of the molecule ($\vec{\mu}_0$), i.e., $\Delta\epsilon_i \propto -\vec{\Delta\mu_i} \cdot \vec{\mu}_0$. Figure 6.6 shows the spatial distributions and dipole moments for HOMO and LUMO as well as the total dipole moment of the representative compact and extended configurations of NC system. The dipole moments of HOMO and LUMO and the total dipole moment are shown in magenta and gray colors, respectively. As seen in the left and right panels of Figure 6.6, the magnitudes of the dipole moments of the HOMO and LUMO and the total dipole moment of the extended configuration are larger as expected, than ones of the compact configurations. Thus the effect of the solvent environment on the HOMO-LUMO gap is quite sensitive to the specific details of the conformation of the peptide.

6.3.2.2 Absorption Spectra

We move on now to examining the effect of the protein conformation on the absorption spectra. Figure 6.1 shows the free energy surface of the monomer chain from which a series of configurations were sampled to compute the spectra. The left and right panels of Figure 6.7, show the averaged absorption spectra over the different configurations for calculations performed in the gas phase and using the PCM model, respectively. In order to assess the role of protein conformation on the spectra, we separate the contributions coming from the compact (A and B) and extended states (C and D).

The averaged spectra in the gas phase for all NC states in panel (a) show tails beyond 250 nm and especially the extended states C and D have longer tails beyond 300nm. As we can see in the panel (b), there still remain tails in the spectra at wavelengths larger than 250 nm for the states C and D in the protein environment. However, the tails beyond 250 nm vanish for states A and B. Both in the gas phase and

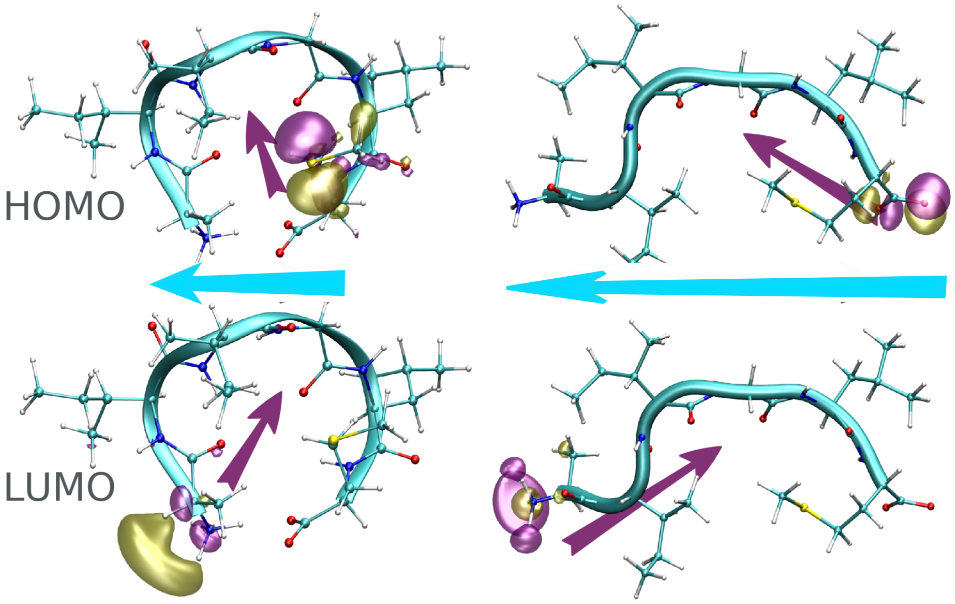


Figure 6.6: HOMO and LUMO isodensity plots of NC compact configuration (left) and the NC extended configuration (right). The cyan arrows represent molecule dipole moments and magenta arrows represent the dipole moments associated with HOMO and LUMO.

PCM environment, the averaged spectra of states C and D exhibit relatively larger values of the molar absorptivity than the ones of states A and B. Interestingly, among the different states of NC system in the PCM model, only state C displays a clear peak at a large wavelength of about 260 nm which is very close to the experimental absorption peak suggesting the dominance of more extended like conformations in the fibril. This feature is also consistent with our previous studies where the optical spectra were determined for the fully periodic 2Y3J crystal, albeit with a lower level of theory [Pinotsi et al., 2016].

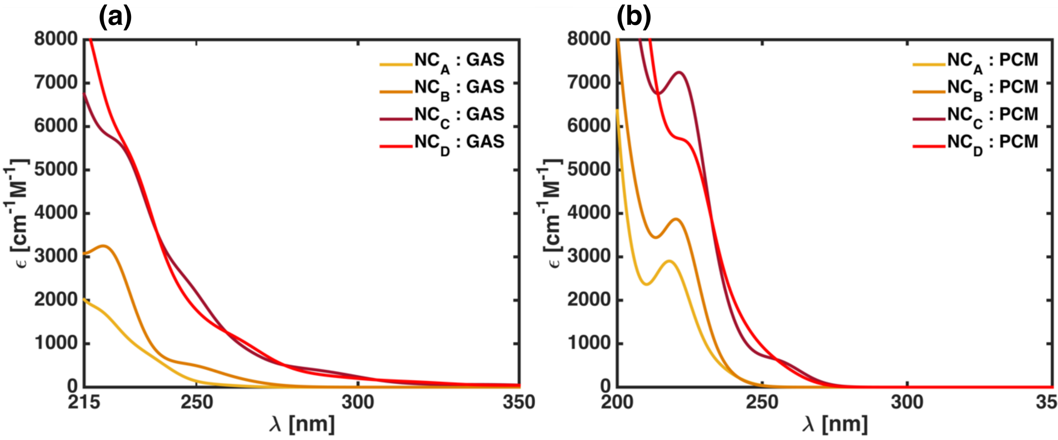


Figure 6.7: averaged absorption spectra on different states of NC system in gas phase (a) and NC system in the protein environment (b).

Since our absorption spectra are constructed from configurations sampled from a molecular dynamics simulation, we are also in a position to investigate the coupling between the optical properties and the conformation of the peptide. Since the conformational fluctuation of the protein can be generally mapped to the changes in the extent of the polar and non-polar interactions of the backbones and side-chains, we estimated the backbone and side-chain contacts (BC and SC) of the peptide. Note that the BC refers to all the intra hydrogen bonds including the interactions between the N and C termini.

For all the configurations of NC, we explored the correlation between the positions of the first peak in the absorption spectra, E_1 (λ), with BC and SC. In Figure 6.8, we show the 2D scatter plots between E_1 (λ) and the structural parameters in the gas phase. The size of the scatter points is representative of the relative oscillator strength of the first peak in the absorption spectra. Both panels a) and b) in Figure 6.8 show a sizable correlation between structural parameters and the positions of the first peak. Lower BC and SC which correspond to more open or extended conformations are characterized by lower excitation energies.

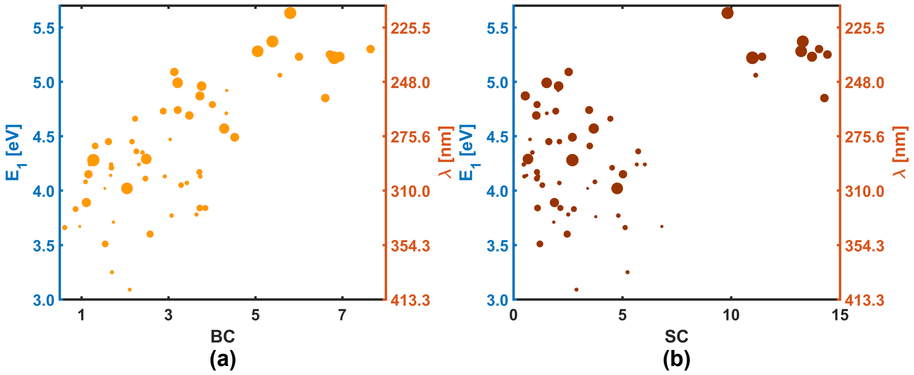


Figure 6.8: Scatter plot between (a) BC and E_1 (λ) and (b) SC and E_1 (λ) for NC system. The size of scatter points are representative of the oscillator strength of the corresponding first peak. The values of Pearson's correlation coefficient for two panels in Figure 6.8 are 0.74 and 0.54 respectively.

From absorption spectra in Figure 6.7, we have already noticed that an inclusion of the protein environment with the PCM changes the optical nature of the excitations quite significantly. In the panels of Figure 6.9, the coupling between the BC and SC parameters and the optical properties in the presence of the PCM is illustrated. Interestingly, the correlation with BC is still quite present (with a correlation coefficient of 0.50) and confirms that the low energy excitations in the spectra are associated with conformations that have the least intra-protein hydrogen bond interactions. On the other hand, for the SC coordinate, correlation is weakened in the PCM environment.

6.3.2.3 Characterization of Excited States

For polypeptides and proteins without any aromatic residues, it is well known that a sequence of peptide bonds exhibits strong absorptions around 190 nm and weak

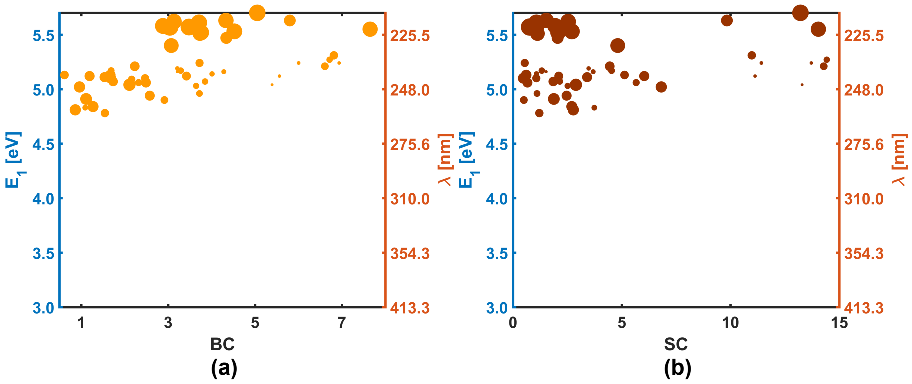


Figure 6.9: Scatter plot between (a) BC and E_1 (λ) and (b) SC and E_1 (λ) for NC system in PCM model. The size of scatter points are representative of the oscillator strength of the corresponding first peak.

absorption around 220 nm [Hunt and Simpson, 1953]. Furthermore, optical activity at wavelengths beyond 250 nm is rarely expected. Absorption between 250 and 290nm is typically associated with the presence of aromatic amino acids such as tryptophan. The origins of low energy excitations we observe may be more complicated than expected. Earlier studies in our group have shown that amyloid fibrils devoid of any conjugated systems can absorb and emit low energy photons in the energy range beyond 350nm. Prasad and co-workers have also recently shown [Prasad et al., 2017] the monomeric proteins lacking aromatic residues show absorption at wavelengths between 250-300 nm. They ascribe this to charge transfer excitations involving charged amino acids.

As discussed earlier, the configurations of NC show a broad absorption spectra with a tail stretching between 250 to 300nm. In order to understand the nature of these lowest energy excitations, we chose some representative configurations for both compact (A3) and extended states (C3 and C7). The integer values correspond to different configurations that were sampled from the FES. Note that C3 makes a dominant contribution to the first peak in the averaged absorption spectra of state C beyond 250 nm using PCM.

We begin by summarizing in Table 6.1 the characteristics of the three low-lying excited states for the chosen three configurations. Table 6.1 present several information about the nature of the low-energy transitions, including the magnitude of the dipole moment difference between the ground and excited states ($\Delta\boldsymbol{\mu}$), and the distance between electron center (ρ_c^-) and hole center (ρ_c^+) ($\Delta\mathbf{r}$) respectively (see Methods Section earlier for more details).

As we can see in Table 6.1, the lowest excited states S1 of A3 in the gas phase have absorption frequencies of less than 250 nm. In both cases, the magnitude of the dipole moment difference, $|\Delta\mu|$, is larger than 2.5 (eÅ) so that this excitation can be considered as a short range CT excitation. Concurrently, the distance between the electron and hole centers is ≈2 Å. Note that a carbon-carbon covalent bond length is 1.2 Å and hence this excitation really involves a very local shifting of charge.

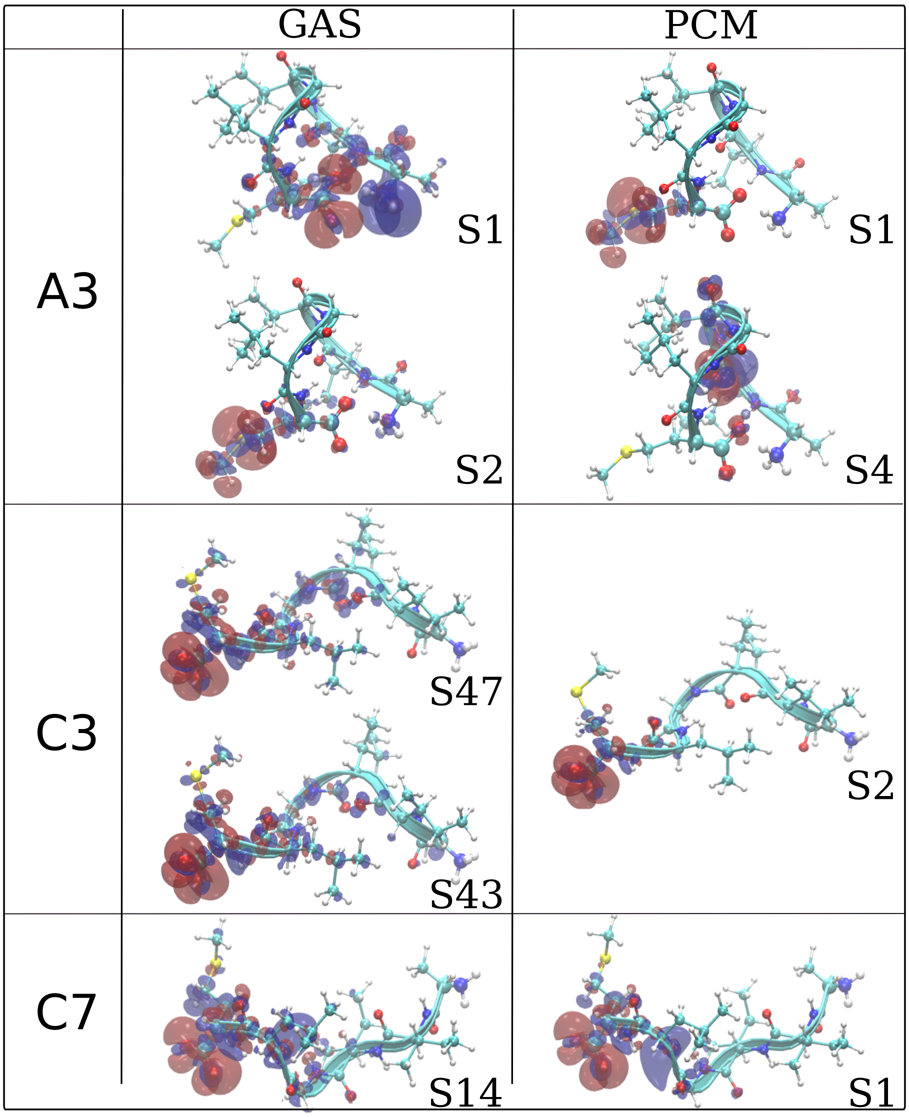


Figure 6.10: Iso-surface plot of charge difference between ground and excited states. The red and blue colors represent positive (electron) and negative (hole) values of the iso-surface which are ±0.002 a.u.

A visualization of the difference between ground and excited state densities ($\Delta\rho = \rho_{exc} - \rho_{gs}$) as seen in Figure 6.10 (A3) confirms that the excitation has a CT character involving the N and C termini (including the methionine side chain) in the gas phase. Note that the red and blue colors represent depletion and increment of charge density, respectively.

Upon the inclusion of the PCM dielectric environment, we observe distinct changes in the excitation character. For the compact conformation, $|\Delta\mu|$ is now less than 1.5 (eÅ) implying the the environment quenches the spatial extent of the excitation. Unlike in the gas phase where the excitation involves a CT from the C to N terminus, with the PCM, the excitation is now a charge reorganization involving the methionine side chain as well as the peptide backbone, in agreement to some of our previous findings [Pinotsi et al., 2016].

Similar features involving the CT character of the lowest energy excitation holds

Table 6.1: The characterization of the excited states for the configurations of A3, C3 and C7 in the gas phase and the protein environment. The first column is the type of the environment, the second column is the configuration name, third column is the postion of the first peak, the forth and fifth columns are the index of the excited states and their contribution percentage on the first peak, the sixth column is the oscillator strength, seventh column is the magnitude of the dipole moment difference between the excited state and the ground state, and the last column is the distance between the electron center and hole center through the excitation (see Methods Section 6.2 for more details).

Env.	Conf.	λ(nm)	state	Perc.(%)	f_I	$\lvert\Delta\mu\rvert$ (eÅ)	$\Delta\mathbf{r}$ (Å)
GAS	A3	237.1	S1	52.4	0.008	2.64	1.99
			S2	19.6	0.003	1.62	1.42
			S4	11.4	0.002	0.30	0.39
			S6	11.3	0.002	0.42	0.52
	C3	324.6	S47	67.0	0.002	7.36	4.88
			S43	16.3	0.001	9.34	5.82
	C7	345.4	S14	90.3	0.004	5.69	3.62
PCM	A3	233.1	S1	37.9	0.007	1.45	1.40
			S4	23.1	0.004	0.46	0.58
			S3	19.1	0.003	0.30	0.40
			S2	17.8	0.003	0.42	0.53
	C3	257.8	S2	99.9	0.026	1.05	0.99
	C7	251.0	S1	94.2	0.017	3.06	2.49

for the extended structures of C3 and C7. In the gas phase, all of the excitations lie beyond 270 nm especially that of C7 which lies above 340 nm. The values of $|\Delta\mu|$ which are larger than 5.69 (eÅ) for the low-lying states, show that these excitations have a CT character from the C-terminus to the backbone. Again, this is reflected in the isosurface plots shown in Figure 6.10. The possibility of charge transfer from a charged terminus to the backbone at wavelengths beyond 250 nm has also been reported previously in small dipeptides [Prasad et al., 2017]. The inclusion of the PCM environment has a significant impact on the character of the excitations. For C3 and C7, we observe that the values of $|\Delta\mu|$ decrease by 6 (eÅ) and subsequently involve a local CT from the C terminus to the backbone.

6.3.2.4 Role of NC-Termini and Inter Hydrogen Bonding Interactions

The experiments show that upon acetylation of the N terminus, there is a significant decrease in the optical density at low energies. The AFM images also indicate that the aggregated state upon acetylation appears to be quite different from the zwitterionic form, although the exact details of the molecular structure remain unknown at this point. In order to assess the importance of the termini groups on the optical properties, we examined the spectra for two different systems. Firstly, we took advantage of an earlier work in our group where we had conducted long simulations of 1.1 μs, where both termini ends are acetylated (dubbed as MET), to extract several configurations in order to perform a thermal averaging of the spectra. Secondly, using two extended configurations of the zwitterionic state, we acetylated the N terminus and introduced a positive counterion (Na^+) near the C-terminus to maintain charge neutrality. These configurations were then optimized keeping the heavy atoms of the peptide besides N- and C-termini at the fixed positions.

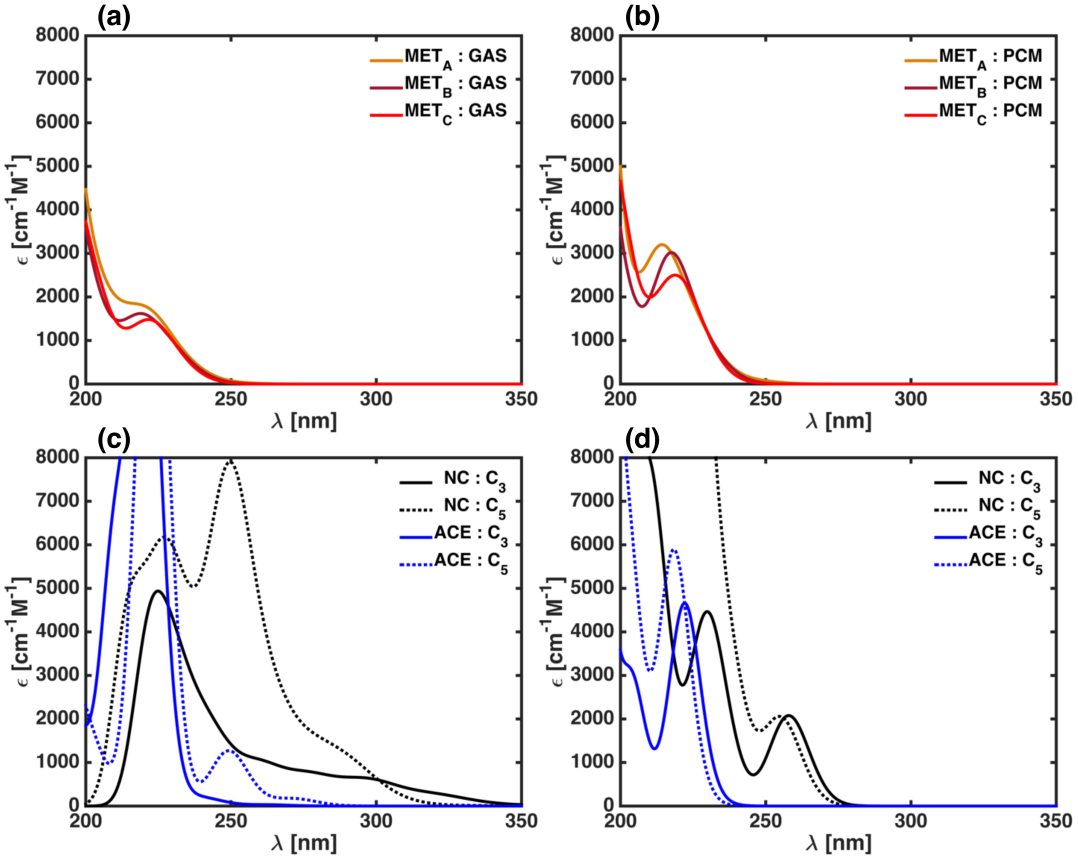


Figure 6.11: averaged absorption spectra on different states of MET in the gas phase (a), MET in the protein environment (b), ACE with a sodium counterion in the gas phase (c) and ACE with a sodium counterion in the protein environment(d). Here "MET" means a configuration capped with two methyl groups on N and C termini and "ACE" means a N-acetylated configuration.

The top left and right panels of Figure 6.11 show the spectra for MET in the gas phase and in PCM. The results are quite striking when compared to that obtained in Figure 6.7 for NC - in the case of MET, the region beyond 250nm is dark and furthermore, the spectra in the low-lying energy region are independent of the environment of the peptide. The double-methylation is of course a drastic perturbation that changes both the N and C termini chemistry and does not tell us about the individual contributions of the N and C termini to the excitation.

The bottom two panels of Figure 6.11 show the spectra obtained for two configurations where the N terminus was acetylated, as done in the experiment. Also shown in the figure, are the spectra that are obtained with the unperturbed NC system in order to see the effect of N-acetylation more clearly. It is clear that acetylation of

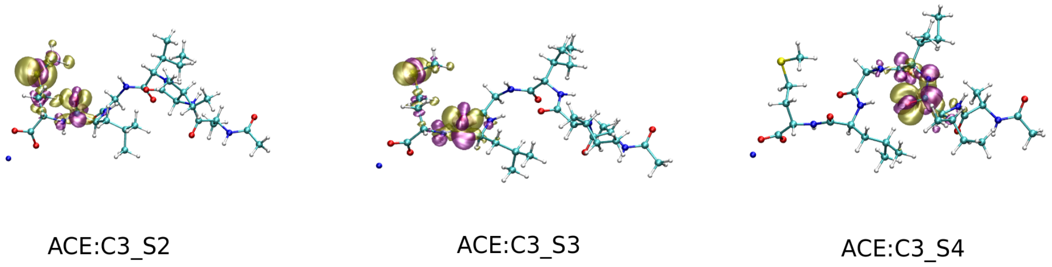


Figure 6.12: Iso-surface plot of charge difference between ground and excited states for acetylated systems. The violet and yellow colors represent charge increment and depletion. Iso-values are set to ±0.002 a.u.

the N terminus results in an overall blue shift of the spectra in both gas phase and PCM model and that this feature is much more pronounced when the PCM model is applied. The iso-surfaces of charge density difference between the ground states and excited states for the N-Acetylated configurations show the dominant contribution of the local excitations near the sulfur atom and the peptide bonds in the PCM model in Figure ??.

The preceding analysis examines the effect of altering the termini of a single chain. We are obviously aware that what we infer about the nature of the optical properties from this system, even with a continuum dielectric medium, is likely to be more complicated in the realistic fibrils or crystal structures.

In order to assess the sensitivity of our interpretations to the chosen model systems, three dimer configurations were generated using the extended conformation of 2Y3J: a) is a head to tail in one direction where the two beta strands only interact via the NC termini hydrogen bonds, b) involves two beta sheets in an anti-parallel configuration and c) involves two beta sheets oriented in a parallel configuration. Due to the size of these systems, we could not compute the full absorption spectra and instead we examined only the first few bright excited states. The excitation energies of and the corresponding oscillator strengths for these first bright low-energy excited state are shown in the Figure 6.13. Interestingly, the first excitation states for all dimer configurations can be characterized by the charge transfer between the termini and between C-terminus and backbone. In panels a) and b) of Figure 6.13, the charge transfer between the N and C termini through the hydrogen bond characterize the lowest excited state for both termini contacted dimer and anti-parallel beta sheet. Comparing to S1 state of monomer compact configuration, although the N and C termini reside on different monomers in dimer system, lowest excited state is characterized by the charge transfer via hydrogen bond with similar strength of monomer compact configuration. The panel c) of Figure 6.13 shows the charge transfer from C-terminus to backbone is the origin of the lowest excitation state of parallel dimer configuration. We might extend this analysis of the absorption spectra of dimer system into ones of the fibrils and conclude that the possible charge transfer between the termini and backbone characterize the low energy excitations of the fibrils.

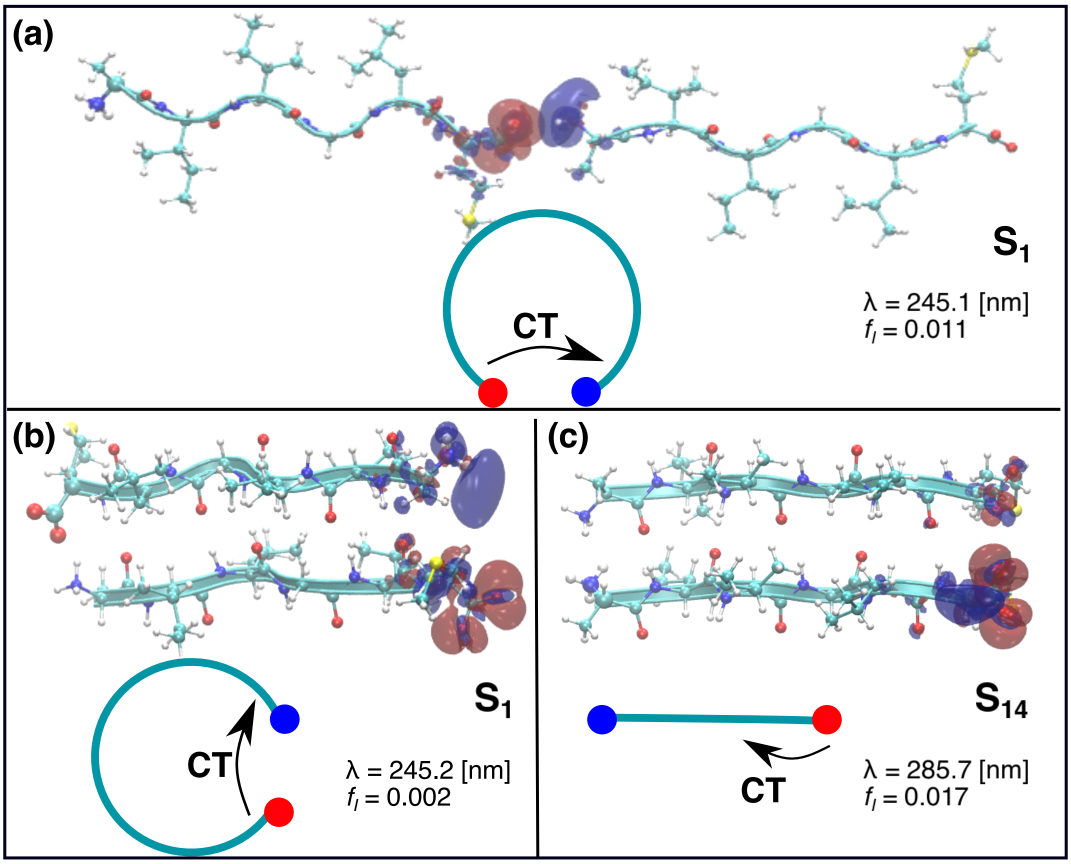


Figure 6.13: Iso-surface plot of charge difference between ground and first excited states for the different dimer configurations : Linear (a), Anti-parallel (b) and Prallel (c). The red and blue colors represent positive (electron) and negative (hole) values of the iso-surface which are ±0.002 a.u.. The values of the wavelength(λ) and the oscillator strength (f_I) of the lowest energy excitation for the different dimer configurations are shown.

6.4 Conclusion

The question of the origin of low energy excitations and subsequent intrinsic fluorescence in biological systems has been the subject of numerous recent experiments and theoretical studies [Pinotsi et al., 2016; Grisanti et al., 2017b; Prasad et al., 2017]. In this work, we have tried to focus our efforts on elucidating the role of intra-protein hydrogen bond contacts on optical properties between 250-300nm beyond the presence of aromatic residues for a some simple model systems.

The fluorescence experiments show a strong sensitivity to the chemical details of the termini. Specifically, acetylation of the N terminus makes the low energy optical activity above 250nm essentially dark. The AFM experiments indicate that both the original zwitterionic chain as well as the N-acetylated chain, appear to form some type of molecular aggregate although the exact details of the structural architecture remain unknown. Using time-dependent density functional theory calculations, we have examined the electronic and optical properties of some model systems of the aggregates both in the gas phase as well as immersed in a continuum dielectric solvent.

Our calculations show that the chemical nature of the termini groups has quite a significant effect on the low energy excitations. In the zwitterionic system, these excitations have a charge transfer character that often involves the termini groups. The nature of these excitations are very sensitive to the environment as well as the conformation of the protein. Our results suggest that the extended conformations are most likely to be the dominant structures in the aggregates for the zwitterionic system although the broad and long nature of the tail in the excitation energies may also be an indication of the heterogeneity in the packing originating from different conformations due to different extents of backbone and side-chain packing. This is somewhat anticipated given recent experimental results of amyloid beta 1-42 showing that an important component of the amyloid structure is the hydrophobic core.

Recent studies by Prasad and co-workers combining theory and experiments have suggested that the low energy excitation regime between 240-500 nm maybe a way to interrogate the secondary structure of a protein. Our results complement this study and reinforce the notion that there is some very rich optical activity in the far UV/visible range that maybe used as an indirect probe into the conformational signature of the protein. In addition, there is the possibility that this region of the spectrum can be used to study molecular aggregates and in the interactions between them in a rather unique way.

Chapter 7

Summary

Using molecular dynamics simulations, D. Thirumalai et al. [Thirumalai, Reddy, and Straub, 2012] studied the role of water in the aggregation process of different hydrophobic and hydrophilic amyloid proteins. While in hydrophobic residues water between peptides is favorably released into the bulk water to accelerate aggregation, in hydrophilic residues water forms a very stable hydrogen bond wire in the pore between peptides which dramatically slows down fibril formation. The existence of different water layers in amyloid fibrils including water pore between peptides is observed in very recent solid-state NMR spectroscopy [Wang et al., 2017], suggesting the important role of water in the aggregation. As we had mentioned in Chapter 1, soluble small oligomers, as primary toxic species, are shown to be more toxic than insoluble fibrils [Sengupta, Nilson, and Kayed, 2016]. Like in the fibrils, water might play an important role in the oligomers, but the importance of water in amyloid oligomers have not been studied. In my thesis, we made the first step in understanding the role of water in the oligomers in the early stage of aggregation.

In Chapter 4, we studied the role of water in the structural fluctuations of a model amyloid protein by reconstructing the conformational free energy landscape projected onto the water wire coordinates connecting donor and acceptor groups in protein. We have shown that water wires add a roughness to the free energy landscape of the protein.

To dig deeper into understanding the origin of this roughness, in Chapter 5 we employed data-science algorithms to reconstruct the free energy landscape in water wire coordinates around tri-alanine protein. Similar to the case of model amyloid protein, we have found a significant roughness in the free energy surface which is sensitive to the underlying secondary structure. In addition, comparing the coordination number to the water wire coordinate shows that the latter is a better solvent coordinate.

The roughness in the free energy landscape along solvent coordinates for both amyloid protein and tri-alanine systems shows the role of water is not exclusive to amyloid proteins. Many studies have shown that water plays an essential role in stability and function of different biomolecules [Nakasako, 2004; Levy and Onuchic, 2006; Makhatadze and Privalov, 1993; Privalov and Makhatadze, 1993; Halle, 2004].

These findings which water adds an additional roughness to the free energy landscape of biomolecules are likely to be a generic phenomenon. These ideas resonate with the slaving role of water in the diffusion of protein on the underlying rough free energy landscape [Zwanzig, 1988; Fenimore et al., 2002].

Since both model amyloid protein and tri-alanine in my thesis are mostly composed of hydrophobic residues, it would be interesting in the future to examine how the roughness changes as full adding more complicated chemistry such as different amino acids large aggregates and temperature and pressure effects.

One of the other focus of my thesis was investigating the optical properties of the amyloid protein. Recently several experiments have shown that protein aggregates including amyloids feature interesting optical properties since they are able to absorb low energy photons at the wavelength beyond 350 nm and emit the visible light centered at around 450 nm even in the absence of aromatic residues. This is currently a very active area of research and perhaps rather controversial.

Inspired by experimental studies on our model systems, we also examined how protein conformation, termini interactions, and the environment tune the optical absorption. This may provide a possibility of using the optical transition to study hydrogen bonding in proteins and maybe even aggregation, without the need of using an external probe.

Bibliography

Agmon, Noam et al. (2016). “Protons and Hydroxide Ions in Aqueous Systems”. In: Chem. Rev. 116.13, pp. 7642–7672. doi: 10.1021/acs.chemrev.5b00736. eprint: http://dx.doi.org/10.1021/acs.chemrev.5b00736. url: http://dx.doi.org/10.1021/acs.chemrev.5b00736.

Aguzzi, Adriano, Christina Sigurdson, and Mathias Heikenwaelder (2008). “Molecular Mechanisms of Prion Pathogenesis”. In: Annual Review of Pathology: Mechanisms of Disease 3.1. PMID: 18233951, pp. 11–40. doi: 10.1146/annurev.pathmechdis.3.121806.154326. eprint: https://doi.org/10.1146/annurev.pathmechdis.3.121806.154326. url: https://doi.org/10.1146/annurev.pathmechdis.3.121806.154326.

Ahmad, Mazen et al. (2011). “Adhesive Water Networks Facilitate Binding of Protein Interfaces”. In: Nat. Commun. 2, p. 261. url: http://dx.doi.org/10.1038/ncomms1258.

Allsop, D et al. (1988). “Immunohistochemical evidence for the derivation of a peptide ligand from the amyloid beta-protein precursor of Alzheimer disease”. In: Proceedings of the National Academy of Sciences 85.8, pp. 2790–2794. issn: 0027-8424. doi: 10.1073/pnas.85.8.2790. eprint: http://www.pnas.org/content/85/8/2790.full.pdf. url: http://www.pnas.org/content/85/8/2790.

Ansari, A et al. (1992). “The role of solvent viscosity in the dynamics of protein conformational changes”. In: Science 256.5065, pp. 1796–1798. issn: 0036-8075. doi: 10.1126/science.1615323. eprint: http://science.sciencemag.org/content/256/5065/1796.full.pdf. url: http://science.sciencemag.org/content/256/5065/1796.

Ansari, Mohd. Ziauddin et al. (2018). “Protein charge transfer absorption spectra: an intrinsic probe to monitor structural and oligomeric transitions in proteins”. In: Faraday Discuss. 207 (0), pp. 91–113. doi: 10.1039/C7FD00194K. url: http://dx.doi.org/10.1039/C7FD00194K.

Auer, Stefan, Christopher M. Dobson, and Michele Vendruscolo (2007). “Characterization of the Nucleation Barriers for Protein Aggregation and Amyloid Formation”. In: HFSP J. 1.2, pp. 137–146. doi: 10.2976/1.2760023. eprint: http://www.tandfonline.com/doi/pdf/10.2976/1.2760023. url: http://www.tandfonline.com/doi/abs/10.2976/1.2760023.

Azar, Yavar T and Mahmoud Payami (2015). “Theoretical description of efficiency enhancement in DSSCs sensitized by newly synthesized heteroleptic Ru complexes”. In: Physical Chemistry Chemical Physics 17.44, pp. 29574–29585.

Baftizadeh, Fahimeh et al. (2012). “Multidimensional View of Amyloid Fibril Nucleation in Atomistic Detail”. In: J. Am. Chem. Soc. 134.8, pp. 3886–3894. doi: 10.1021/ja210826a. eprint: http://dx.doi.org/10.1021/ja210826a. url: http://dx.doi.org/10.1021/ja210826a.

Bagchi, Biman (2005). “Water Dynamics in the Hydration Layer around Proteins and Micelles”. In: Chem. Rev. 105.9, pp. 3197–3219. doi: 10.1021/cr020661+. eprint: https://doi.org/10.1021/cr020661+. url: https://doi.org/10.1021/cr020661+.

Baglioni, Serena et al. (2006). “Prefibrillar Amyloid Aggregates Could Be Generic Toxins in Higher Organisms”. In: Journal of Neuroscience 26.31, pp. 8160–8167. issn: 0270-6474. doi: 10.1523/JNEUROSCI.4809-05.2006. eprint: http://www.jneurosci.org/content/26/31/8160.full.pdf. url: http://www.jneurosci.org/content/26/31/8160.

Ball, Philip (2008). “Water as an Active Constituent in Cell Biology”. In: Chem. Rev. 108.1, pp. 74–108. doi: 10.1021/cr068037a. eprint: https://doi.org/10.1021/cr068037a. url: https://doi.org/10.1021/cr068037a.

Barducci, Alessandro, Giovanni Bussi, and Michele Parrinello (2008). “Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method”. In: Phys. Rev. Lett. 100 (2), pp. 020603–020606. doi: 10.1103/PhysRevLett.100.020603. url: http://link.aps.org/doi/10.1103/PhysRevLett.100.020603.

Barducci, Alessandro et al. (2013). “Free-energy Landscape of Protein Oligomerization From Atomistic Simulations”. In: Proc. Natl. Acad. Sci. U. S. A. 110.49, pp. 4708–4713. doi: 10.1073/pnas.1320077110. eprint: http://www.pnas.org/content/110/49/E4708.full.pdf. url: http://www.pnas.org/content/110/49/E4708.abstract.

Bauernschmitt, Rüdiger and Reinhart Ahlrichs (1996). “Treatment of electronic excitations within the adiabatic approximation of time dependent density functional theory”. In: Chemical Physics Letters 256.4, pp. 454 –464. issn: 0009-2614. doi: https://doi.org/10.1016/0009-2614(96)00440-X. url: http://www.sciencedirect.com/science/article/pii/000926149600440X.

Baumketner, Andrij and Joan Emma Shea (2007). “The Structure of the Alzheimer Amyloid β 10-35 Peptide Probed through Replica-Exchange Molecular Dynamics Simulations in Explicit Solvent”. In: J. Mol. Biol. 366.1, pp. 275–285. issn: 0022-2836. doi: http://dx.doi.org/10.1016/j.jmb.2006.11.015. url: http://www.sciencedirect.com/science/article/pii/S0022283606015440.

Bellissent-Funel, Marie-Claire et al. (2016). “Water Determines the Structure and Dynamics of Proteins”. In: Chem. Rev. 116.13, pp. 7673–7697. doi: 10.1021/acs.chemrev.5b00664. eprint: https://doi.org/10.1021/acs.chemrev.5b00664. url: https://doi.org/10.1021/acs.chemrev.5b00664.

Berendsen, H. J. C. et al. (1984). “Molecular Dynamics with Coupling to an External Bath”. In: J. Chem. Phys. 81.8, pp. 3684–3690. doi: http://dx.doi.org/10.

1063/1.448118. url: http://scitation.aip.org/content/aip/journal/jcp/81/8/10.1063/1.448118.

Bernstein, Summer L. et al. (2009). “Amyloid-b protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer’s disease”. In: Nature Chemistry 1. article, 326 EP –. url: http://dx.doi.org/10.1038/nchem.247.

Best, Robert B., Wenwei Zheng, and Jeetain Mittal (2014). “Balanced Protein–Water Interactions Improve Properties of Disordered Proteins and Non-Specific Protein Association”. In: J. Chem. Theory Comput. 10.11, pp. 5113–5124. doi: 10.1021/ct500569b. url: http://dx.doi.org/10.1021/ct500569b.

Best, Robert B. et al. (2012). “Inclusion of Many-Body Effects in the Additive CHARMM Protein CMAP Potential Results in Enhanced Cooperativity of α-Helix and β-Hairpin Formation”. In: Biophysical Journal 103.5, pp. 1045–1051. issn: 0006-3495. doi: 10.1016/j.bpj.2012.07.042. url: http://dx.doi.org/10.1016/j.bpj.2012.07.042.

Bhattacharya, Arpan et al. (2017). “Direct Evidence of Intrinsic Blue Fluorescence from Oligomeric Interfaces of Human Serum Albumin”. In: Langmuir 33.40. PMID: 28930631, pp. 10606–10615. doi: 10.1021/acs.langmuir.7b02463. eprint: https://doi.org/10.1021/acs.langmuir.7b02463. url: https://doi.org/10.1021/acs.langmuir.7b02463.

Bianco, Valentino et al. (2017a). “How the stability of a folded protein depends on interfacial water properties and residue-residue interactions”. In: Journal of Molecular Liquids 245. Recent Progresses on the Experimental & Theoretical-Computational Techniques for the Study of Liquids and Supercritical Fluids. From Simple to Complex Systems, pp. 129 –139. issn: 0167-7322. doi: https://doi.org/10.1016/j.molliq.2017.08.026. url: http://www.sciencedirect.com/science/article/pii/S0167732217315416.

Bianco, Valentino et al. (2017b). “Role of Water in the Selection of Stable Proteins at Ambient and Extreme Thermodynamic Conditions”. In: Phys. Rev. X 7 (2), p. 021047. doi: 10.1103/PhysRevX.7.021047. url: https://link.aps.org/doi/10.1103/PhysRevX.7.021047.

Binkley, J. S. and J. A. Pople (1975). “Møller-Plesset Theory for Atomic Ground State Energies”. In: Int. J. Quantum Chem. 9.2, pp. 229–236. issn: 1097-461X. doi: 10.1002/qua.560090204. url: http://dx.doi.org/10.1002/qua.560090204.

Bitan, Gal et al. (2003). “Amyloid β-Protein (Aβ) Assembly: Aβ40 and Aβ42 Oligomerize through Distinct Pathways”. In: Proc. Natl. Acad. Sci. U. S. A. 100.1, pp. 330–335. doi: 10.1073/pnas.222681699. eprint: http://www.pnas.org/content/100/1/330.full.pdf. url: http://www.pnas.org/content/100/1/330.abstract.

Bizzarri, Anna Rita and Salvatore Cannistraro (2002). “Molecular Dynamics of Water at the Protein-Solvent Interface”. In: J. Phys. Chem. B 106.26, pp. 6617–6633.

doi: 10.1021/jp020100m. eprint: http://dx.doi.org/10.1021/jp020100m. url: http://dx.doi.org/10.1021/jp020100m.

Bonomi, M., A. Barducci, and M. Parrinello (2009). “Reconstructing the equilibrium Boltzmann distribution from well-tempered metadynamics”. In: Journal of Computational Chemistry 30.11, pp. 1615–1621. doi: 10.1002/jcc.21305. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.21305. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.21305.

Bordo, Domenico and Patrick Argos (1994). “The Role of Side-chain Hydrogen Bonds in the Formation and Stabilization of Secondary Structure in Soluble Proteins”. In: Journal of Molecular Biology 243.3, pp. 504 –519. issn: 0022-2836. doi: https://doi.org/10.1006/jmbi.1994.1676. url: http://www.sciencedirect.com/science/article/pii/S0022283684716767.

Boys, S.F. and F. Bernardi (1970). “The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors”. In: Molecular Physics 19.4, pp. 553–566. doi: 10.1080/00268977000101561. eprint: http://dx.doi.org/10.1080/00268977000101561. url: http://dx.doi.org/10.1080/00268977000101561.

Brkljača, Zlatko et al. (2014). “Calculating CD Spectra of Flexible Peptides: An Assessment of TD-DFT Functionals”. In: Journal of chemical theory and computation 10.8, pp. 3270–3279.

Brovchenko, I. et al. (2005). “Thermal Breaking of Spanning Water Networks in the Hydration Shell of Proteins”. In: J. Chem. Phys. 123.22, p. 224905. doi: 10.1063/1.2121708. url: http://dx.doi.org/10.1063/1.2121708.

Bulheller, Benjamin M. and Jonathan D. Hirst (2009). “DichroCalc—circular and Linear Dichroism Online”. In: Bioinformatics 25.4, p. 539. doi: 10.1093/bioinformatics/btp016. eprint: /oup/backfile/Content_public/Journal/bioinformatics/25/4/10.1093/bioinformatics/btp016/2/btp016.pdf. url: http://dx.doi.org/10.1093/bioinformatics/btp016.

Burns, Alistair and Steve Iliffe (2009). “Alzheimer’s disease”. In: BMJ 338. issn: 0959-8138. doi: 10.1136/bmj.b158. eprint: https://www.bmj.com/content. url: https://www.bmj.com/content/338/bmj.b158.

CA, McLean et al. (1999). “Soluble pool of Abeta amyloid as a determinant of severity of neurodegeneration in Alzheimer’s disease”. In: Annals of Neurology 46.6, pp. 860–866. doi: 10.1002/1531-8249(199912)46:6<860::AID-ANA8>3.0.CO;2-M. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/1531-8249%28199912%2946%3A6%3C860%3A%3AAID-ANA8%3E3.0.CO%3B2-M. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/1531-8249%28199912%2946%3A6%3C860%3A%3AAID-ANA8%3E3.0.CO%3B2-M.

Cappai, Roberto and Kevin J. Barnham (2008). “Delineating the Mechanism of Alzheimer’s Disease Aβ Peptide Neurotoxicity”. In: Neurochemical Research 33.3, pp. 526–532. issn: 1573-6903. doi: 10.1007/s11064-007-9469-8. url: https://doi.org/10.1007/s11064-007-9469-8.

Carballo Pacheco, Martín and Birgit Strodel (2016). “Advances in the Simulation of Protein Aggregation at the Atomistic Scale”. In: J. Phys. Chem. B 120.12, pp. 2991–2999. doi: `10.1021/acs.jpcb.6b00059`. eprint: `http://dx.doi.org/10.1021/acs.jpcb.6b00059`. url: `http://dx.doi.org/10.1021/acs.jpcb.6b00059`.

Casida, Mark E. (1996). “Time-Dependent Density Functional Response Theory of Molecular Systems: Theory, Computational Methods, and Functionals”. In: Recent Developments and Applications of Modern Density Functional Theory. Ed. by J.M. Seminario. Vol. 4. Theoretical and Computational Chemistry. Elsevier, pp. 391 –439. doi: `https://doi.org/10.1016/S1380-7323(96)80093-8`. url: `http://www.sciencedirect.com/science/article/pii/S1380732396800938`.

Ceriotti, Michele, Gareth A. Tribello, and Michele Parrinello (2011). “Simplifying the representation of complex free-energy landscapes using sketch-map”. In: Proceedings of the National Academy of Sciences 108.32, pp. 13023–13028. issn: 0027-8424. doi: `10.1073/pnas.1108486108`. eprint: `http://www.pnas.org/content/108/32/13023.full.pdf`. url: `http://www.pnas.org/content/108/32/13023`.

Chan, Fiona T. S. et al. (2013). “Protein amyloids develop an intrinsic fluorescence signature during aggregation”. In: Analyst 138 (7), pp. 2156–2162. doi: `10.1039/C3AN36798C`. url: `http://dx.doi.org/10.1039/C3AN36798C`.

Chebaro, Yassmine, Normand Mousseau, and Philippe Derreumaux (2009). “Structures and Thermodynamics of Alzheimer’s Amyloid-β Aβ(16-35) Monomer and Dimer by Replica Exchange Molecular Dynamics Simulations: Implication for Full-Length Aβ Fibrillation”. In: J. Phys. Chem. B 113.21, pp. 7668–7675. doi: `10.1021/jp900425e`. eprint: `http://dx.doi.org/10.1021/jp900425e`. url: `http://dx.doi.org/10.1021/jp900425e`.

Chen, Xiaohong et al. (2018). “Prevalent intrinsic emission from nonaromatic amino acids and poly(amino acids)”. In: Science China Chemistry 61.3, pp. 351–359. issn: 1869-1870. doi: `10.1007/s11426-017-9114-4`. url: `https://doi.org/10.1007/s11426-017-9114-4`.

Chen, Yung-Fou and J. J. Dannenberg (2011). “The effect of polarization on multiple hydrogen-bond formation in models of self-assembling materials”. In: Journal of Computational Chemistry 32.13, pp. 2890–2895. issn: 1096-987X. doi: `10.1002/jcc.21870`. url: `http://dx.doi.org/10.1002/jcc.21870`.

Chiti, Fabrizio and Christopher M. Dobson (2006). “Protein Misfolding, Functional Amyloid, and Human Disease”. In: Annual Review of Biochemistry 75.1. PMID: 16756495, pp. 333–366. doi: `10.1146/annurev.biochem.75.101304.123901`. eprint: `https://doi.org/10.1146/annurev.biochem.75.101304.123901`. url: `https://doi.org/10.1146/annurev.biochem.75.101304.123901`.

Choi, Jun-Ho and Minhaeng Cho (2014). “Ion aggregation in high salt solutions. II. Spectral graph analysis of water hydrogen-bonding network and ion aggregate structures”. In: J. Chem. Phys. 141.15, p. 154502. doi: `10.1063/1.4897638`.

eprint: https://doi.org/10.1063/1.4897638. url: https://doi.org/10.1063/1.4897638.

Cobar, Erika A. et al. (2012). "Examination of the hydrogen-bonding networks in small water clusters (n = 2-5, 13, 17) using absolutely localized molecular orbital energy decomposition analysis". In: Phys. Chem. Chem. Phys. 14 (44), pp. 15328–15339. doi: 10.1039/C2CP42522J. url: http://dx.doi.org/10.1039/C2CP42522J.

Colletier, Jacques-Philippe et al. (2011). "Molecular basis for amyloid-β polymorphism". In: Proceedings of the National Academy of Sciences 108.41, pp. 16938–16943. issn: 0027-8424. doi: 10.1073/pnas.1112600108. eprint: http://www.pnas.org/content/108/41/16938.full.pdf. url: http://www.pnas.org/content/108/41/16938.

Collinge, John (2001). "Prion Diseases of Humans and Animals: Their Causes and Molecular Basis". In: Annual Review of Neuroscience 24.1. PMID: 11283320, pp. 519–550. doi: 10.1146/annurev.neuro.24.1.519. eprint: https://doi.org/10.1146/annurev.neuro.24.1.519. url: https://doi.org/10.1146/annurev.neuro.24.1.519.

Collins, Kim D. and Michael W. Washabaugh (1985). "The Hofmeister effect and the behaviour of water at interfaces". In: Q. Rev. Biophys. 18.4, pp. 323–422. doi: 10.1017/S0033583500005369.

Cossi, Maurizio and Vincenzo Barone (2001). "Time-dependent density functional theory for molecules in liquid solutions". In: The Journal of chemical physics 115, p. 4708.

Cramer, Christopher J and Donald G Truhlar (1999). "Implicit solvation models: equilibria, structure, spectra, and dynamics". In: Chemical Reviews 99.8, pp. 2161–2200.

Cui, Di, Shuching Ou, and Sandeep Patel (2014). "Protein-spanning Water Networks and Implications for Prediction of Protein-protein Interactions Mediated through Hydrophobic Effects". In: Proteins: Struct. Funct. Bioinf. 82.12, pp. 3312–3326. issn: 1097-0134. doi: 10.1002/prot.24683. url: http://dx.doi.org/10.1002/prot.24683.

Dang, Liem X. (1998). "Importance of Polarization Effects in Modeling the Hydrogen Bond in Water Using Classical Molecular Dynamics Techniques". In: The Journal of Physical Chemistry B 102.3, pp. 620–624. doi: 10.1021/jp9731258. eprint: http://dx.doi.org/10.1021/jp9731258. url: http://dx.doi.org/10.1021/jp9731258.

Darden, Tom, Darrin York, and Lee Pedersen (1993). "Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems". In: J. Chem. Phys. 98.12, pp. 10089–10092. doi: http://dx.doi.org/10.1063/1.464397. url: http://scitation.aip.org/content/aip/journal/jcp/98/12/10.1063/1.464397.

Darve, Eric and Andrew Pohorille (2001). "Calculating free energies using average force". In: The Journal of Chemical Physics 115.20, pp. 9169–9183. doi: 10.1063/

1.1410978. eprint: https://doi.org/10.1063/1.1410978. url: https://doi.org/10.1063/1.1410978.

Davis, C. M. and T. A. Litovitz (1965). “Two-State Theory of the Structure of Water”. In: J. Chem. Phys. 42.7, pp. 2563–2576. doi: 10.1063/1.1696333. eprint: https://doi.org/10.1063/1.1696333. url: https://doi.org/10.1063/1.1696333.

Davtyan, Aram et al. (2012). “AWSEM-MD: Protein Structure Prediction Using Coarse-Grained Physical Potentials and Bioinformatically Based Local Structure Biasing”. In: The Journal of Physical Chemistry B 116.29, pp. 8494–8503. doi: 10.1021/jp212541y. eprint: https://doi.org/10.1021/jp212541y. url: https://doi.org/10.1021/jp212541y.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). “Maximum Likelihood from Incomplete Data via the EM Algorithm”. In: Journal of the Royal Statistical Society. Series B (Methodological) 39.1, pp. 1–38. issn: 00359246. url: http://www.jstor.org/stable/2984875.

Dennington, Roy, Todd A. Keith, and John M. Millam (2016). GaussView Version 6. Semichem Inc. Shawnee Mission KS.

d’Errico, M. et al. (2018). “Automatic topography of high-dimensional data sets by non-parametric Density Peak clustering”. In: ArXiv e-prints. arXiv: 1802.10549 [stat.ML].

DIJKSTRA, E.W. (1959). “A Note on Two Problems in Connexion with Graphs.” In: Numer. Math. 1, pp. 269–271. url: http://eudml.org/doc/131436.

Duan, Yong et al. (2003). “A Point-charge Force Field for Molecular Mechanics Simulations of Proteins based on Condensed-phase Quantum Mechanical Calculations”. In: J. Comput. Chem. 24.16, pp. 1999–2012. issn: 1096-987X. doi: 10.1002/jcc.10349. url: http://dx.doi.org/10.1002/jcc.10349.

Dubins, David N. et al. (2000). “Role of Water in Protein-Ligand Interactions: Volumetric Characterization of the Binding of 2’-CMol. Phys. and 3’-CMol. Phys. to Ribonuclease A”. In: J. Phys. Chem. B 104.2, pp. 390–401. doi: 10.1021/jp992138d. eprint: https://doi.org/10.1021/jp992138d. url: https://doi.org/10.1021/jp992138d.

Dyson, H. Jane, Peter E. Wright, and Harold A. Scheraga (2006). “The role of hydrophobic interactions in initiation and propagation of protein folding”. In: Proc. Natl. Acad. Sci. U. S. A. 103.35, pp. 13057–13061. issn: 0027-8424. doi: 10.1073/pnas.0605504103. eprint: http://www.pnas.org/content/103/35/13057.full.pdf. url: http://www.pnas.org/content/103/35/13057.

Essmann, Ulrich et al. (1995). “A Smooth Particle Mesh Ewald Method”. In: J. Chem. Phys. 103.19, pp. 8577–8593. doi: http://dx.doi.org/10.1063/1.470117. url: http://scitation.aip.org/content/aip/journal/jcp/103/19/10.1063/1.470117.

Facco, Elena et al. (2017). “Estimating the intrinsic dimension of datasets by a minimal neighborhood information”. In: Scientific Reports 7.1, p. 12140. issn: 2045-2322. doi: `10.1038/s41598-017-11873-y`. url: `https://doi.org/10.1038/s41598-017-11873-y`.

Fawzi, Nicolas Lux et al. (2007). “Determining the Critical Nucleus and Mechanism of Fibril Elongation of the Alzheimer’s Aβ1-40 Peptide”. In: J. Mol. Biol. 365.2, pp. 535–550. issn: 0022-2836. doi: `http://dx.doi.org/10.1016/j.jmb.2006.10.011`. url: `http://www.sciencedirect.com/science/article/pii/S0022283606013507`.

Fenimore, P. W. et al. (2002). “Slaving: Solvent fluctuations dominate protein dynamics and functions”. In: Proc. Natl. Acad. Sci. U. S. A. 99.25, pp. 16047–16051. doi: `10.1073/pnas.212637899`. eprint: `http://www.pnas.org/content/99/25/16047.full.pdf`. url: `http://www.pnas.org/content/99/25/16047.abstract`.

Finkelstein, A.V. and O.V. Galzitskaya (2004). “Physics of Protein Folding”. In: Phys. Life. Rev. 1.1, pp. 23–56. issn: 1571-0645. doi: `http://dx.doi.org/10.1016/j.plrev.2004.03.001`. url: `http://www.sciencedirect.com/science/article/pii/S1571064504000053`.

Fiorin, Giacomo, Michael L. Klein, and Jerome Henin (2013). “Using collective variables to drive molecular dynamics simulations”. In: Mol. Phys. 111.22-23, pp. 3345–3362. doi: `10.1080/00268976.2013.813594`. eprint: `https://doi.org/10.1080/00268976.2013.813594`. url: `https://doi.org/10.1080/00268976.2013.813594`.

Fischer, Stefan and Chandra S. Verma (1999). “Binding of Buried Structural Water Increases the Flexibility of Proteins”. In: Proc. Natl. Acad. Sci. U. S. A. 96.17, pp. 9613–9615. doi: `10.1073/pnas.96.17.9613`. eprint: `http://www.pnas.org/content/96/17/9613.full.pdf`. url: `http://www.pnas.org/content/96/17/9613.abstract`.

Fitzpatrick, Anthony W. et al. (2011). “Inversion of the Balance between Hydrophobic and Hydrogen Bonding Interactions in Protein Folding and Aggregation”. In: PLoS Comput Biol 7.10. Ed. by Vijay S. Pande, e1002169. issn: 1553-734X. doi: `10.1371/journal.pcbi.1002169`. url: `http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3192805/`.

Fitzpatrick, Anthony W. P. et al. (2013). “Atomic structure and hierarchical assembly of a cross-β amyloid fibril”. In: Proceedings of the National Academy of Sciences 110.14, pp. 5468–5473. issn: 0027-8424. doi: `10.1073/pnas.1219476110`. eprint: `http://www.pnas.org/content/110/14/5468.full.pdf`. url: `http://www.pnas.org/content/110/14/5468`.

Foderà, Vito et al. (2013). “Electrostatics Controls the Formation of Amyloid Superstructures in Protein Aggregation”. In: Phys. Rev. Lett. 111 (10), p. 108105. doi: `10.1103/PhysRevLett.111.108105`. url: `https://link.aps.org/doi/10.1103/PhysRevLett.111.108105`.

Fogarty, Aoife C. and Damien Laage (2014). “Water Dynamics in Protein Hydration Shells: The Molecular Origins of the Dynamical Perturbation”. In: J. Phys. Chem. B 118.28, pp. 7715–7729. doi: `10.1021/jp409805p`. eprint: `https://doi.org/10.1021/jp409805p`. url: `https://doi.org/10.1021/jp409805p`.

Fonseca Guerra, Célia et al. (2000). “Hydrogen Bonding in DNA Base Pairs: Reconciliation of Theory and Experiment”. In: Journal of the American Chemical Society 122.17, pp. 4117–4128. doi: `10.1021/ja993262d`. eprint: `http://dx.doi.org/10.1021/ja993262d`. url: `http://dx.doi.org/10.1021/ja993262d`.

Frank, Aaron T. et al. (2015). “Predicting Protein Backbone Chemical Shifts From C_alpha Coordinates: Extracting High Resolution Experimental Observables from Low Resolution Models”. In: J. Chem. Theory Comput. 11.1, pp. 325–331. doi: `10.1021/ct5009125`. url: `http://dx.doi.org/10.1021/ct5009125`.

Franks, Felix (2000). Water. A Matrix of Life. RSC Paperbacks. The Royal Society of Chemistry, pp. 001–004. isbn: 978-0-85404-583-9. doi: `10.1039/9781847552341`. url: `http://dx.doi.org/10.1039/9781847552341`.

Frisch, M. J. et al. Gaussian09 Revision E.01. Gaussian Inc. Wallingford CT 2009.

Gaspar, Renee C. et al. (2010). “Oligomers of β-amyloid are sequestered into and seed new plaques in the brains of an AD mouse model”. In: Experimental Neurology 223.2. Beta-amyloid and tau protein abnormalities in Alzheimer’s disease, pp. 394 –400. issn: 0014-4886. doi: `https://doi.org/10.1016/j.expneurol.2009.09.001`. url: `http://www.sciencedirect.com/science/article/pii/S0014488609003720`.

Gasparotto, Piero, Ali A. Hassanali, and Michele Ceriotti (2016). “Probing Defects and Correlations in the Hydrogen-Bond Network of ab Initio Water”. In: J. Chem. Theory Comput. 12.4, pp. 1953–1964. doi: `10.1021/acs.jctc.5b01138`. eprint: `http://dx.doi.org/10.1021/acs.jctc.5b01138`. url: `http://dx.doi.org/10.1021/acs.jctc.5b01138`.

Gething, Mary Jane and Joseph Sambrook (1992). “Protein Folding in the Cell”. In: Nature 355.6355, pp. 33–45. doi: `10.1038/355033a0`. url: `http://dx.doi.org/10.1038/355033a0`.

Giberti, Federico et al. (2014). “The Role of Quantum Effects on Structural and Electronic Fluctuations in Neat and Charged Water”. In: J. Phys. Chem. B 118.46, pp. 13226–13235. doi: `10.1021/jp507752e`. eprint: `http://dx.doi.org/10.1021/jp507752e`. url: `http://dx.doi.org/10.1021/jp507752e`.

Gnanakaran, S., Ruth Nussinov, and Angel E. García (2006). “Atomic-Level Description of Amyloid β-Dimer Formation”. In: J. Am. Chem. Soc. 128.7, pp. 2158–2159. doi: `10.1021/ja0548337`. eprint: `http://dx.doi.org/10.1021/ja0548337`. url: `http://dx.doi.org/10.1021/ja0548337`.

Goate, Alison et al. (1991). “Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer’s disease”. In: Nature 349, 704 EP –. url: `http://dx.doi.org/10.1038/349704a0`.

Grabowski, Slawomir J. (2006). Hydrogen Bonding - New Insights. Vol. 3. Challenges and Advances in Computational Chemistry and Physics. New York: Springer.

Granata, Daniele et al. (2015). “The Inverted Free Energy Landscape of an Intrinsically Disordered Peptide by Simulations and Experiments”. In: Sci. Rep. 5, p. 15449. url: http://dx.doi.org/10.1038/srep15449.

Greenwald, Jason and Roland Riek (2010). “Biology of Amyloid: Structure, Function, and Regulation”. In: Structure 18.10, pp. 1244–1260. issn: 0969-2126. doi: 10.1016/j.str.2010.08.009. url: https://doi.org/10.1016/j.str.2010.08.009.

Grisanti, Luca et al. (2017a). “A computational study on how structure influences the optical properties in model crystal structures of amyloid fibrils”. In: Phys. Chem. Chem. Phys. 19 (5), pp. 4030–4040. doi: 10.1039/C6CP07564A. url: http://dx.doi.org/10.1039/C6CP07564A.

Grisanti, Luca et al. (2017b). “A computational study on how structure influences the optical properties in model crystal structures of amyloid fibrils”. In: Physical Chemistry Chemical Physics 19.5, pp. 4030–4040.

Grubmüller, Helmut (1995). “Predicting slow structural transitions in macromolecular systems: Conformational flooding”. In: Phys. Rev. E 52 (3), pp. 2893–2906. doi: 10.1103/PhysRevE.52.2893. url: https://link.aps.org/doi/10.1103/PhysRevE.52.2893.

Gsponer, Joerg and Michele Vendruscolo (2006). “Theoretical Approaches to Protein Aggregation”. In: Protein Pept. Lett. 13.3, pp. 287–293. doi: 10.2174/092986606775338407. url: http://www.eurekaselect.com/node/56863/article.

Guevara-Vela, Jose Manuel et al. (2016). “Hydrogen bond cooperativity and anticooperativity within the water hexamer”. In: Phys. Chem. Chem. Phys. 18 (29), pp. 19557–19566. doi: 10.1039/C6CP00763E. url: http://dx.doi.org/10.1039/C6CP00763E.

Gupta, Madhulika, Charusita Chakravarty, and Sanjoy Bandyopadhyay (2016). “Sensitivity of Protein Glass Transition to the Choice of Water Model”. In: J. Chem. Theory Comput. 12.11, pp. 5643–5655. doi: 10.1021/acs.jctc.6b00825. eprint: https://doi.org/10.1021/acs.jctc.6b00825. url: https://doi.org/10.1021/acs.jctc.6b00825.

Halle, Bertil (2004). “Protein hydration dynamics in solution: a critical survey”. In: Philosophical Transactions of the Royal Society of London B: Biological Sciences 359.1448, pp. 1207–1224. issn: 0962-8436. doi: 10.1098/rstb.2004.1499. eprint: http://rstb.royalsocietypublishing.org/content/359/1448/1207.full.pdf. url: http://rstb.royalsocietypublishing.org/content/359/1448/1207.

Han, Beomsoo et al. (2011). “SHIFTX2: Significantly Improved Protein Chemical Shift Prediction”. In: J. Biomol. NMR. 50.1, pp. 43–57. issn: 1573-5001. doi: 10.

1007/s10858-011-9478-4. url: http://dx.doi.org/10.1007/s10858-011-9478-4.

Hänggi, Peter, Peter Talkner, and Michal Borkovec (1990). “Reaction-rate Theory: Fifty Years after Kramers”. In: Rev. Mod. Phys. 62 (2), pp. 251–341. doi: 10.1103/RevModPhys.62.251. url: http://link.aps.org/doi/10.1103/RevModPhys.62.251.

Hansmann, Ulrich H. E. and Luc T. Wille (2002). “Global Optimization by Energy Landscape Paving”. In: Phys. Rev. Lett. 88 (6), p. 068105. doi: 10.1103/PhysRevLett.88.068105. url: https://link.aps.org/doi/10.1103/PhysRevLett.88.068105.

Hardy, John (2009). “The amyloid hypothesis for Alzheimer’s disease: a critical reappraisal”. In: Journal of Neurochemistry 110.4, pp. 1129–1134. doi: 10.1111/j.1471-4159.2009.06181.x. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1471-4159.2009.06181.x. url: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1471-4159.2009.06181.x.

Hardy, John and Dennis J. Selkoe (2002). “The Amyloid Hypothesis of Alzheimer’s Disease: Progress and Problems on the Road to Therapeutics”. In: Science 297.5580, pp. 353–356. issn: 0036-8075. doi: 10.1126/science.1072994. eprint: http://science.sciencemag.org/content/297/5580/353.full.pdf. url: http://science.sciencemag.org/content/297/5580/353.

Hartl, F. Ulrich, Andreas Bracher, and Manajit Hayer-Hartl (2011). “Molecular chaperones in protein folding and proteostasis”. In: Nature 475.7356, pp. 324–332. issn: 0028-0836. doi: 10.1038/nature10317. url: http://dx.doi.org/10.1038/nature10317.

Hassanali, Ali et al. (2013). “Proton Transfer Through the Water Gossamer”. In: Proc. Natl. Acad. Sci. U. S. A. 110.34, pp. 13723–13728. doi: 10.1073/pnas.1306642110. eprint: http://www.pnas.org/content/110/34/13723.full.pdf. url: http://www.pnas.org/content/110/34/13723.abstract.

Hebda, James A. and Andrew D. Miranker (2009). “The Interplay of Catalysis and Toxicity by Amyloid Intermediates on Lipid Bilayers: Insights from Type II Diabetes”. In: Annual Review of Biophysics 38.1. PMID: 19416063, pp. 125–152. doi: 10.1146/annurev.biophys.050708.133622. eprint: https://doi.org/10.1146/annurev.biophys.050708.133622. url: https://doi.org/10.1146/annurev.biophys.050708.133622.

Henchman, Richard H. and Sheeba Jem Irudayam (2010). “Topological Hydrogen-Bond Definition to Characterize the Structure and Dynamics of Liquid Water”. In: The Journal of Physical Chemistry B 114.50, pp. 16792–16810. doi: 10.1021/jp105381s. eprint: http://dx.doi.org/10.1021/jp105381s. url: http://dx.doi.org/10.1021/jp105381s.

Hess, Berk et al. (1997). “LINCS: A linear constraint solver for molecular simulations”. In: J. Comput. Chem. 18.12, pp. 1463–1472. issn: 1096-987X. doi: 10.1002/(SICI)1096-987X(199709)18:12<1463::AID-{J.Comput.Chem.}4>3.0.CO;2-H. url:

http://dx.doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-{J.Comput.Chem.}4>3.0.CO;2-H.

Hohenberg, P. and W. Kohn (1964). "Inhomogeneous Electron Gas". In: Phys. Rev. 136 (3B), B864–B871. doi: 10.1103/PhysRev.136.B864. url: https://link.aps.org/doi/10.1103/PhysRev.136.B864.

Hoover, William G. (1985). "Canonical Dynamics: Equilibrium Phase-space Distributions". In: Phys. Rev. A 31 (3), pp. 1695–1697. doi: 10.1103/PhysRevA.31.1695. url: http://link.aps.org/doi/10.1103/PhysRevA.31.1695.

Horwich, Arthur (2002). "Protein aggregation in disease: a role for folding intermediates forming specific multimeric interactions". In: J Clin Invest 110.9, pp. 1221–1232. issn: 0021-9738. doi: 10.1172/JCI16781. url: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC151620/.

Hou, Liming et al. (2004). "Solution NMR Studies of the Aβ(1—40) and Aβ(1—42) Peptides Establish that the Met35 Oxidation State Affects the Mechanism of Amyloid Formation". In: Journal of the American Chemical Society 126.7. PMID: 14971932, pp. 1992–2005. doi: 10.1021/ja036813f. eprint: https://doi.org/10.1021/ja036813f. url: https://doi.org/10.1021/ja036813f.

Hubbard, Roderick E and Muhammad Kamran Haider (2001). "Hydrogen Bonds in Proteins: Role and Strength". In: John Wiley & Sons, Ltd.

Huber, Thomas, Andrew E. Torda, and Wilfred F. van Gunsteren (1994). "Local elevation: A method for improving the searching properties of molecular dynamics simulation". In: Journal of computer-aided molecular design 8 6, pp. 695–708.

Huggins, David J., May Marsh, and Mike C. Payne (2011). "Thermodynamic Properties of Water Molecules at a Protein-Protein Interaction Surface". In: J. Chem. Theory Comput. 7.11, pp. 3514–3522. doi: 10.1021/ct200465z. eprint: https://doi.org/10.1021/ct200465z. url: https://doi.org/10.1021/ct200465z.

Hughes, Eleri, Ron M. Burke, and Andrew J. Doig (2000). "Inhibition of Toxicity in the β-Amyloid Peptide Fragment β-(25-35) Using N-Methylated Derivatives". In: J. Biol. Chem. 275.33, pp. 25109–25115. doi: 10.1074/jbc.M003554200. eprint: http://www.jbc.org/content/275/33/25109.full.pdf+html. url: http://www.jbc.org/content/275/33/25109.abstract.

Hummer, G et al. (1996). "An Information Theory Model of Hydrophobic Interactions". In: Proc. Natl. Acad. Sci. U. S. A. 93.17, pp. 8951–8955. eprint: http://www.pnas.org/content/93/17/8951.full.pdf. url: http://www.pnas.org/content/93/17/8951.abstract.

Humphrey, William, Andrew Dalke, and Klaus Schulten (1996). "VMD: Visual Molecular Dynamics". In: J. Mol. Graphics 14.1, pp. 33–38. issn: 0263-7855. doi: http://dx.doi.org/10.1016/0263-7855(96)00018-5. url: http://www.sciencedirect.com/science/article/pii/0263785596000185.

Hunt, HD and WT Simpson (1953). "Spectra of Simple Amides in the Vacuum Ultraviolet1". In: Journal of the American Chemical Society 75.18, pp. 4540–4543.

Iida, Kenji et al. (2009). “A systematic understanding of orbital energy shift in polar solvent”. In: The Journal of chemical physics 130.4, p. 044107.

Izenman, Alan Julian (1991). “Review Papers: Recent Developments in Nonparametric Density Estimation”. In: Journal of the American Statistical Association 86.413, pp. 205–224. doi: 10.1080/01621459.1991.10475021. eprint: https://doi.org/10.1080/01621459.1991.10475021. url: https://doi.org/10.1080/01621459.1991.10475021.

Jacobson, Matthew P. et al. (2002). “Force Field Validation Using Protein Side Chain Prediction”. In: The Journal of Physical Chemistry B 106.44, pp. 11673–11680. doi: 10.1021/jp021564n. eprint: http://dx.doi.org/10.1021/jp021564n. url: http://dx.doi.org/10.1021/jp021564n.

Jansson, HelÃ©n, Rikard Bergman, and Jan Swenson (2011). “Role of Solvent for the Dynamics and the Glass Transition of Proteins”. In: J. Phys. Chem. B 115.14, pp. 4099–4109. doi: 10.1021/jp1089867. eprint: https://doi.org/10.1021/jp1089867. url: https://doi.org/10.1021/jp1089867.

Jeffrey, George A. and Wolfram Saenger (1991). Hydrogen Bonding in Biological Structures. Berlin: Springer-Verlag.

Jin, J. (2015). “Alzheimer disease”. In: JAMA 313.14, p. 1488. doi: 10.1001/jama.2015.2852. eprint: /data/journals/jama/933743/jpg150011.pdf. url: +http://dx.doi.org/10.1001/jama.2015.2852.

Johansson, Patrik K. and Patrick Koelsch (2017). “Label-free imaging of amyloids using their intrinsic linear and nonlinear optical properties”. In: Biomed. Opt. Express 8.2, pp. 743–756. doi: 10.1364/BOE.8.000743. url: http://www.osapublishing.org/boe/abstract.cfm?URI=boe-8-2-743.

Jong, KwangHyok, Luca Grisanti, and Ali Hassanali (2017). “Hydrogen Bond Networks and Hydrophobic Effects in the Amyloid β_{30-35} Chain in Water: A Molecular Dynamics Study”. In: Journal of Chemical Information and Modeling 57.7. PMID: 28603985, pp. 1548–1562. doi: 10.1021/acs.jcim.7b00085. eprint: http://dx.doi.org/10.1021/acs.jcim.7b00085. url: http://dx.doi.org/10.1021/acs.jcim.7b00085.

Jorgensen, William L. et al. (1983). “Comparison of simple potential functions for simulating liquid water”. In: J. Chem. Phys. 79.2, pp. 926–935. doi: 10.1063/1.445869. eprint: https://doi.org/10.1063/1.445869. url: https://doi.org/10.1063/1.445869.

Kaminski, George A. et al. (2001). “Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides”. In: The Journal of Physical Chemistry B 105.28, pp. 6474–6487. doi: 10.1021/jp003919d. eprint: http://dx.doi.org/10.1021/jp003919d. url: http://dx.doi.org/10.1021/jp003919d.

Karamanos, Theodoros K. et al. (2015). “Mechanisms of Amyloid Formation Revealed by Solution NMR”. In: Prog. Nucl. Magn. Reson. Spectrosc. 88-89, pp. 86–104.

issn: 0079-6565. doi: http://dx.doi.org/10.1016/j.pnmrs.2015.05.002. url: http://www.sciencedirect.com/science/article/pii/S007965651500028X.

Kayed, Rakez et al. (2003). "Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis". In: Science 300.5618, pp. 486–489. issn: 0036-8075. doi: 10.1126/science.1079469. eprint: http://science.sciencemag.org/content/300/5618/486.full.pdf. url: http://science.sciencemag.org/content/300/5618/486.

Khakshoor, Omid et al. (2012). "Measurement and Theory of Hydrogen Bonding Contribution to Isosteric DNA Base Pairs". In: Journal of the American Chemical Society 134.6, pp. 3154–3163. doi: 10.1021/ja210475a. eprint: http://dx.doi.org/10.1021/ja210475a. url: http://dx.doi.org/10.1021/ja210475a.

Khemtémourian, Lucie et al. (2011). "Low pH Acts as Inhibitor of Membrane Damage Induced by Human Islet Amyloid Polypeptide". In: J. Am. Chem. Soc. 133.39, pp. 15598–15604. doi: 10.1021/ja205007j. eprint: http://dx.doi.org/10.1021/ja205007j. url: http://dx.doi.org/10.1021/ja205007j.

Kim, Seongheun et al. (2014). "Ion Aggregation in High Aalt Solutions: Ion Network Versus Ion Cluster". In: J. Chem. Phys. 141.12, p. 124510. doi: 10.1063/1.4896227. url: http://dx.doi.org/10.1063/1.4896227.

Kiril, Tsemekhman et al. "Cooperative hydrogen bonding in amyloid formation". In: Protein Science 16.4 (), pp. 761–764. doi: 10.1110/ps.062609607. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1110/ps.062609607. url: https://onlinelibrary.wiley.com/doi/abs/10.1110/ps.062609607.

Kirkitadze, Marina D, Margaret M Condron, and David B Teplow (2001). "Identification and Characterization of Key Kinetic Intermediates in Amyloid β-Protein Fibrillogenesis". In: J. Mol. Biol. 312.5, pp. 1103–1119. issn: 0022-2836. doi: http://dx.doi.org/10.1006/jmbi.2001.4970. url: http://www.sciencedirect.com/science/article/pii/S0022283601949708.

Kohn, W. and L. J. Sham (1965). "Self-Consistent Equations Including Exchange and Correlation Effects". In: Phys. Rev. 140 (4A), A1133–A1138. doi: 10.1103/PhysRev.140.A1133. url: https://link.aps.org/doi/10.1103/PhysRev.140.A1133.

Kragelj, Jaka et al. (2013). "Conformational Propensities of Intrinsically Disordered Proteins from NMR Chemical Shifts". In: ChemPhysChem 14.13, pp. 3034–3045. issn: 1439-7641. doi: 10.1002/cphc.201300387. url: http://dx.doi.org/10.1002/cphc.201300387.

Krishnan, R. et al. (1980). "Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions". In: The Journal of Chemical Physics 72.1, pp. 650–654. doi: 10.1063/1.438955. eprint: https://doi.org/10.1063/1.438955. url: https://doi.org/10.1063/1.438955.

Kumar, Shankar, Philip W. Payne, and Maximiliano Vásquez (1996). "Method for free-energy calculations using iterative techniques". In: Journal of Computational Chemistry 17.10, pp. 1269–1275. doi: 10.1002/(SICI)1096-987X(19960730)17:

10<1269::AID-JCC7>3.0.CO;2-M. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/%28SICI%291096-987X%2819960730%2917%3A10%3C1269%3A%3AAID-JCC7%3E3.0.CO%3B2-M. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/%28SICI%291096-987X%2819960730%2917%3A10%3C1269%3A%3AAID-JCC7%3E3.0.CO%3B2-M.

Kumar, Shankar et al. (1995). “Multidimensional free-energy calculations using the weighted histogram analysis method”. In: Journal of Computational Chemistry 16.11, pp. 1339–1350. doi: 10.1002/jcc.540161104. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/jcc.540161104. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.540161104.

Laage, Damien, Thomas Elsaesser, and James T. Hynes (2017). “Water Dynamics in the Hydration Shells of Biomolecules”. In: Chem. Rev. 117.16, pp. 10694–10725. doi: 10.1021/acs.chemrev.6b00765. eprint: https://doi.org/10.1021/acs.chemrev.6b00765. url: https://doi.org/10.1021/acs.chemrev.6b00765.

Ladiwala, Ali Reza A. et al. (2012). “Conformational Differences between Two Amyloid β Oligomers of Similar Size and Dissimilar Toxicity”. In: Journal of Biological Chemistry 287.29, pp. 24765–24773. doi: 10.1074/jbc.M111.329763. eprint: http://www.jbc.org/content/287/29/24765.full.pdf+html. url: http://www.jbc.org/content/287/29/24765.abstract.

Laio, Alessandro and Michele Parrinello (2002). “Escaping free-energy minima”. In: Proceedings of the National Academy of Sciences 99.20, pp. 12562–12566. issn: 0027-8424. doi: 10.1073/pnas.202427399. eprint: http://www.pnas.org/content/99/20/12562.full.pdf. url: http://www.pnas.org/content/99/20/12562.

Lakowicz, Joseph R. (2006). Principles of Fluorescence Spectroscopy. Springer Science and Business Media, pp. 529–606.

Lapidus, Lisa J. (2013). “Understanding Protein Aggregation from the View of Monomer Dynamics”. In: Mol. BioSyst. 9 (1), pp. 29–35. doi: 10.1039/C2MB25334H. url: http://dx.doi.org/10.1039/C2MB25334H.

Latora, V. and M. Marchiori (2003). “Economic Small-world Behavior in Weighted Networks”. In: Eur. Phys. J. B 32.2, pp. 249–263. doi: 10.1140/epjb/e2003-00095-5. url: http://dx.doi.org/10.1140/epjb/e2003-00095-5.

Latora, Vito and Massimo Marchiori (2001). “Efficient Behavior of Small-World Networks”. In: Phys. Rev. Lett. 87 (19), pp. 198701–198704. doi: 10.1103/PhysRevLett.87.198701. url: http://link.aps.org/doi/10.1103/PhysRevLett.87.198701.

Läuter, H. (1986). “Silverman, B. W.: Density Estimation for Statistics and Data Analysis. Chapman & Hall, London – New York 1986, 175 pp., £12.—”. In: Biometrical Journal 30.7, pp. 876–877. doi: 10.1002/bimj.4710300745. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/bimj.4710300745. url: https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.4710300745.

Lee, Hochan et al. (2015). “Spectral Graph Analyses of Water Hydrogen-Bonding Network and Osmolyte Aggregate Structures in Osmolyte-Water Solutions”. In: J. Phys. Chem. B 119.45, pp. 14402–14412. doi: 10.1021/acs.jpcb.5b08029. eprint: http://dx.doi.org/10.1021/acs.jpcb.5b08029. url: http://dx.doi.org/10.1021/acs.jpcb.5b08029.

Lee, Shin Jung C. et al. (2017). “Towards an understanding of amyloid-β oligomers: characterization, toxicity mechanisms, and inhibitors”. In: Chem. Soc. Rev. 46 (2), pp. 310–323. doi: 10.1039/C6CS00731G. url: http://dx.doi.org/10.1039/C6CS00731G.

Leeuwen, Robert van (1999). “Mapping from Densities to Potentials in Time-Dependent Density-Functional Theory”. In: Phys. Rev. Lett. 82 (19), pp. 3863–3866. doi: 10.1103/PhysRevLett.82.3863. url: https://link.aps.org/doi/10.1103/PhysRevLett.82.3863.

Levy, Mel (1982). “Electron densities in search of Hamiltonians”. In: Phys. Rev. A 26 (3), pp. 1200–1208. doi: 10.1103/PhysRevA.26.1200. url: https://link.aps.org/doi/10.1103/PhysRevA.26.1200.

Levy, Yaakov and José N. Onuchic (2006). “Water Mediation in Protein Folding and Molecular Recognition”. In: Annu. Rev. Biophys. Biomol. Struct. 35.1, pp. 389–415. doi: 10.1146/annurev.biophys.35.040405.102134. eprint: https://doi.org/10.1146/annurev.biophys.35.040405.102134. url: https://doi.org/10.1146/annurev.biophys.35.040405.102134.

Li, Lin et al. (2013). “On the Dielectric “Constant” of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi”. In: Journal of Chemical Theory and Computation 9.4. PMID: 23585741, pp. 2126–2136. doi: 10.1021/ct400065j. eprint: http://dx.doi.org/10.1021/ct400065j. url: http://dx.doi.org/10.1021/ct400065j.

Lindorff Larsen, Kresten et al. (2010). “Improved side-chain torsion potentials for the Amber ff99SB protein force field”. In: Proteins: Struct., Funct., Bioinf. 78.8, pp. 1950–1958. issn: 1097-0134. doi: 10.1002/prot.22711. url: http://dx.doi.org/10.1002/prot.22711.

Liu, Ruitian et al. (2004). “Residues 17-20 and 30-35 of Beta-Amyloid Play Critical Roles in Aggregation”. In: J. Neurosci. Res. 75.2, pp. 162–171. issn: 1097-4547. doi: 10.1002/jnr.10859. url: http://dx.doi.org/10.1002/jnr.10859.

Lue, Lih-Fen et al. (1999). “Soluble Amyloid β Peptide Concentration as a Predictor of Synaptic Change in Alzheimer’s Disease”. In: The American Journal of Pathology 155.3, pp. 853–862. issn: 0002-9440. doi: 10.1016/S0002-9440(10)65184-X. url: https://doi.org/10.1016/S0002-9440(10)65184-X.

Luzar, Alenka and David Chandler (1996). “Hydrogen-bond kinetics in liquid water”. In: Nature 379.6560, pp. 55–57. doi: 10.1038/379055a0. url: http://dx.doi.org/10.1038/379055a0.

Lührs, Thorsten et al. (2005). “3D Structure of Alzheimer’s Amyloid-β(1-42) Fibrils”. In: Proc. Natl. Acad. Sci. U. S. A. 102.48, pp. 17342–17347. doi: 10.1073/pnas.

0506723102. eprint: http://www.pnas.org/content/102/48/17342.full.pdf. url: http://www.pnas.org/content/102/48/17342.abstract.

Mack, Y.P and M Rosenblatt (1979). “Multivariate k-nearest neighbor density estimates”. In: J. Multivar. Anal. 9.1, pp. 1–15. issn: 0047-259X. doi: https://doi.org/10.1016/0047-259X(79)90065-4. url: http://www.sciencedirect.com/science/article/pii/0047259X79900654.

MacKerell, Alexander D., Nilesh Banavali, and Nicolas Foloppe (2000). “Development and current status of the CHARMM force field for nucleic acids”. In: Biopolymers 56.4, pp. 257–265. issn: 1097-0282. doi: 10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W. url: http://dx.doi.org/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W.

Maity, Sibaprasad and Yuri L. Lyubchenko (2016). “Probing of Amyloid Ab (14-23) Trimers by Single-Molecule Force Spectroscopy”. In: Jacobs J Mol Transl Med 1.1. 28239686[pmid], p. 004. url: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5321571/.

Maji, Samir et al. (Nov. 2005). “Conformational Dynamics of Amyloid β-Protein Assembly Probed Using Intrinsic Fluorescence †”. In: 44, pp. 13365–76.

Makhatadze, George I. and Peter L. Privalov (1993). “Contribution of Hydration to Protein Folding Thermodynamics: I. The Enthalpy of Hydration”. In: J. Mol. Biol. 232.2, pp. 639–659. issn: 0022-2836. doi: https://doi.org/10.1006/jmbi.1993.1416. url: http://www.sciencedirect.com/science/article/pii/S0022283683714166.

Makin, O. Sumner et al. (2005). “Molecular Basis for Amyloid Fibril Formation and Stability”. In: Proc. Natl. Acad. Sci. U. S. A. 102.2, pp. 315–320. doi: 10.1073/pnas.0406847102. eprint: http://www.pnas.org/content/102/2/315.full.pdf. url: http://www.pnas.org/content/102/2/315.abstract.

Makin, Simon (July 2018). “The amyloid hypothesis on trial”. In: 559, S4–S7.

Mandal, I., S. Paul, and R. Venkatramani (2018). “Optical backbone-sidechain charge transfer transitions in proteins sensitive to secondary structure and modifications”. In: Faraday Discuss. 207 (0), pp. 115–135. doi: 10.1039/C7FD00203C. url: http://dx.doi.org/10.1039/C7FD00203C.

Mannini, Benedetta et al. (2014). “Toxicity of Protein Oligomers Is Rationalized by a Function Combining Size and Surface Hydrophobicity”. In: ACS Chemical Biology 9.10, pp. 2309–2317. doi: 10.1021/cb500505m. eprint: https://doi.org/10.1021/cb500505m. url: https://doi.org/10.1021/cb500505m.

Maragakis, Paul, Arjan van der Vaart, and Martin Karplus (2009). “Gaussian-Mixture Umbrella Sampling”. In: The Journal of Physical Chemistry B 113.14. PMID: 19284746, pp. 4664–4673. doi: 10.1021/jp808381s. eprint: https://doi.org/10.1021/jp808381s. url: https://doi.org/10.1021/jp808381s.

Marinelli, Fabrizio et al. (Aug. 2009). “A Kinetic Model of Trp-Cage Folding from Multiple Biased Molecular Dynamics Simulations”. In: PLoS Comput. Biol 5.8, pp. 1–18. doi: 10.1371/journal.pcbi.1000452. url: https://doi.org/10.1371/journal.pcbi.1000452.

Marques, Miguel AL and Eberhard KU Gross (2004). “Time-dependent density functional theory”. In: Annu. Rev. Phys. Chem. 55, pp. 427–455.

Marsili, Simone et al. (2006). “Self-healing Umbrella Sampling: A Non-equilibrium Approach for Quantitative Free Energy Calculations”. In: The Journal of Physical Chemistry B 110.29. PMID: 16854090, pp. 14011–14013. doi: 10.1021/jp062755j. eprint: https://doi.org/10.1021/jp062755j. url: https://doi.org/10.1021/jp062755j.

Mattos, Carla (2002). “Protein-water Interactions in a Dynamic World”. In: Trends Biochem. Sci. 27.4, pp. 203–208. issn: 0968-0004. doi: http://dx.doi.org/10.1016/S0968-0004(02)02067-4. url: http://www.sciencedirect.com/science/article/pii/S0968000402020674.

McLean, A. D. and G. S. Chandler (1980). “Contracted Gaussian basis sets for molecular calculations. I. Second row atoms, Z=11–18”. In: The Journal of Chemical Physics 72.10, pp. 5639–5648. doi: 10.1063/1.438980. eprint: https://doi.org/10.1063/1.438980. url: https://doi.org/10.1063/1.438980.

Miyamoto, Shuichi and Peter A. Kollman (1992). “Settle: An analytical version of the SHAKE and RATTLE algorithm for rigid water models”. In: J. Comput. Chem. 13.8, pp. 952–962. issn: 1096-987X. doi: 10.1002/jcc.540130805. url: http://dx.doi.org/10.1002/jcc.540130805.

Møller, Chr. and M. S. Plesset (1934). “Note on an Approximation Treatment for Many-Electron Systems”. In: Phys. Rev. 46 (7), pp. 618–622. doi: 10.1103/PhysRev.46.618. url: http://link.aps.org/doi/10.1103/PhysRev.46.618.

Morriss-Andrews, Alex and Joan-Emma Shea (2015). “Computational Studies of Protein Aggregation: Methods and Applications”. In: Annual Review of Physical Chemistry 66.1. PMID: 25648485, pp. 643–666. doi: 10.1146/annurev-physchem-040513-103738. eprint: https://doi.org/10.1146/annurev-physchem-040513-103738. url: https://doi.org/10.1146/annurev-physchem-040513-103738.

Munishkina, Larissa A. and Anthony L. Fink (2007). “Fluorescence as a method to reveal structures and membrane-interactions of amyloidogenic proteins”. In: Biochimica et Biophysica Acta (BBA) - Biomembranes 1768.8. Amyloidogenic Protein–Membrane Interaction, pp. 1862 –1885. issn: 0005-2736. doi: https://doi.org/10.1016/j.bbamem.2007.03.015. url: http://www.sciencedirect.com/science/article/pii/S0005273607000806.

Müller, E. Matthias, Armin de Meijere, and Helmut Grubmüller (2002). “Predicting unimolecular chemical reactions: Chemical flooding”. In: The Journal of Chemical Physics 116.3, pp. 897–905. doi: 10.1063/1.1427722. eprint: https://doi.org/10.1063/1.1427722. url: https://doi.org/10.1063/1.1427722.

Nakasako, Masayoshi (2004). “Water–protein interactions from high–resolution protein crystallography”. In: Philosophical Transactions of the Royal Society of London B: Biological Sciences 359.1448, pp. 1191–1206. issn: 0962-8436. doi: `10.1098/rstb.2004.1498`. eprint: `http://rstb.royalsocietypublishing.org/content/359/1448/1191.full.pdf`. url: `http://rstb.royalsocietypublishing.org/content/359/1448/1191`.

Nakashima, Hiroshi, Ai Fukuoka, and Yuka Saitou (2013). “Hydrogen bonds are related to the thermal stability of 16S rRNA”. In: Journal of Biomedical Science and Engineering 06.01, pp. 19–24. doi: `10.4236/jbise.2013.61003`.

Nasica Labouze, Jessica et al. (2015). “Amyloid β Protein and Alzheimer’s Disease: When Computer Simulations Complement Experimental Studies”. In: Chem. Rev. 115.9, pp. 3518–3563. doi: `10.1021/cr500638n`. eprint: `http://dx.doi.org/10.1021/cr500638n`. url: `http://dx.doi.org/10.1021/cr500638n`.

Nelson, Rebecca and David Eisenberg (2006). “Recent atomic models of amyloid fibril structure”. In: Current Opinion in Structural Biology 16.2. Theory and simulation/Macromolecular assemblages, pp. 260 –265. issn: 0959-440X. doi: `https://doi.org/10.1016/j.sbi.2006.03.007`. url: `http://www.sciencedirect.com/science/article/pii/S0959440X06000455`.

Nguyen, Hung D. and Carol K. Hall (2005). “Kinetics of Fibril Formation by Polyalanine Peptides”. In: J. Biol. Chem. 280.10, pp. 9074–9082. doi: `10.1074/jbc.M407338200`. eprint: `http://www.jbc.org/content/280/10/9074.full.pdf+html`.

Nosé, Shuichi (1984). “A Molecular Dynamics Method for Simulations in the Canonical Ensemble”. In: Mol. Phys. 52.2, pp. 255–268. doi: `10.1080/00268978400101201`. eprint: `http://dx.doi.org/10.1080/00268978400101201`. url: `http://dx.doi.org/10.1080/00268978400101201`.

Oleinikova, Alla and Ivan Brovchenko (2011). “What Determines the Thermal Stability of the Hydrogen-Bonded Water Network Enveloping Peptides?” In: J. Phys. Chem. Lett. 2.7, pp. 765–769. doi: `10.1021/jz200181w`. eprint: `https://doi.org/10.1021/jz200181w`. url: `https://doi.org/10.1021/jz200181w`.

Pace, C. Nick et al. (2014). “Contribution of hydrogen bonds to protein stability”. In: Protein Science 23.5, pp. 652–661. issn: 1469-896X. doi: `10.1002/pro.2449`. url: `http://dx.doi.org/10.1002/pro.2449`.

Papoian, Garegin A., Johan Ulander, and Peter G. Wolynes (2003). “Role of Water Mediated Interactions in Protein-Protein Recognition Landscapes”. In: J. Am. Chem. Soc. 125.30, pp. 9170–9178. doi: `10.1021/ja034729u`. eprint: `https://doi.org/10.1021/ja034729u`. url: `https://doi.org/10.1021/ja034729u`.

Paravastu, Anant K. et al. (2008). “Molecular structural basis for polymorphism in Alzheimer’s β-amyloid fibrils”. In: Proceedings of the National Academy of Sciences 105.47, pp. 18349–18354. issn: 0027-8424. doi: `10.1073/pnas.0806270105`. eprint: `http://www.pnas.org/content/105/47/18349.full.pdf`. url: `http://www.pnas.org/content/105/47/18349`.

Park, Sanghyun and Klaus Schulten (2004). “Calculating potentials of mean force from steered molecular dynamics simulations”. In: The Journal of Chemical Physics 120.13, pp. 5946–5961. doi: 10.1063/1.1651473. eprint: https://doi.org/10.1063/1.1651473. url: https://doi.org/10.1063/1.1651473.

Pauling, Linus (1967). Chemical Bond: Brief Introduction to Modern Structural Chemistry. New York: Cornell University Press.

Peach, Michael JG et al. (2008). “Excitation energies in density functional theory: An evaluation and a diagnostic test”. In: The Journal of chemical physics 128, p. 044118.

Pellarin, Riccardo and Amedeo Caflisch (2006). “Interpreting the Aggregation Kinetics of Amyloid Peptides”. In: J. Mol. Biol. 360.4, pp. 882–892. issn: 0022-2836. doi: http://dx.doi.org/10.1016/j.jmb.2006.05.033. url: http://www.sciencedirect.com/science/article/pii/S0022283606006152.

Petersilka, M., U. J. Gossmann, and E. K. U. Gross (1996). “Excitation Energies from Time-Dependent Density-Functional Theory”. In: Phys. Rev. Lett. 76 (8), pp. 1212–1215. doi: 10.1103/PhysRevLett.76.1212. url: https://link.aps.org/doi/10.1103/PhysRevLett.76.1212.

Petkova, Aneta T. et al. (2002). “A Structural Model for Alzheimer’s β-Amyloid Fibrils based on Experimental Constraints from Solid State NMR”. In: Proc. Natl. Acad. Sci. U. S. A. 99.26, pp. 16742–16747. doi: 10.1073/pnas.262663499. eprint: http://www.pnas.org/content/99/26/16742.full.pdf. url: http://www.pnas.org/content/99/26/16742.abstract.

Pfitzner, Darius, Richard Leibbrandt, and David Powers (2009). “Characterization and evaluation of similarity measures of pairs of clusterings”. In: Knowl. Inf. Syst. 19, pp. 361–394. doi: https://doi.org/10.1007/s10115-008-0150-6.

Phillips, G. N. and B. M. Pettitt (1995). “Structure and Dynamics of the Water around Myoglobin.” In: Protein Sci. 4.2, pp. 149–158. issn: 0961-8368. url: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2143067/.

Phipps, M. J. S. et al. (2016). “Energy Decomposition Analysis Based on Absolutely Localized Molecular Orbitals for Large-Scale Density Functional Theory Calculations in Drug Design”. In: Journal of Chemical Theory and Computation 12.7, pp. 3135–3148. doi: 10.1021/acs.jctc.6b00272. eprint: http://dx.doi.org/10.1021/acs.jctc.6b00272. url: http://dx.doi.org/10.1021/acs.jctc.6b00272.

Phipps, Maximillian J. S. et al. (2015). “Energy decomposition analysis approaches and their evaluation on prototypical protein-drug interaction patterns”. In: Chem. Soc. Rev. 44 (10), pp. 3177–3211. doi: 10.1039/C4CS00375F. url: http://dx.doi.org/10.1039/C4CS00375F.

Piana, Stefano et al. (2015). “Water Dispersion Interactions Strongly Influence Simulated Structural Properties of Disordered Protein States”. In: The Journal of Physical Chemistry B 119.16, pp. 5113–5123. doi: 10.1021/jp508971m. eprint:

http://dx.doi.org/10.1021/jp508971m. url: http://dx.doi.org/10.1021/jp508971m.

Pietrucci, Fabio (2017). “Strategies for the exploration of free energy landscapes: Unity in diversity and challenges ahead”. In: Reviews in Physics 2, pp. 32 –45. issn: 2405-4283. doi: https://doi.org/10.1016/j.revip.2017.05.001. url: http://www.sciencedirect.com/science/article/pii/S2405428317300059.

Pietrucci, Fabio and Alessandro Laio (2009). “A Collective Variable for the Efficient Exploration of Protein Beta-Sheet Structures: Application to SH3 and GB1”. In: J. Chem. Theory Comput. 5.9, pp. 2197–2201. doi: 10.1021/ct900202f. eprint: https://doi.org/10.1021/ct900202f. url: https://doi.org/10.1021/ct900202f.

Pinotsi, Dorothea et al. (2016). “Proton Transfer and Structure Specific Fluorescence in Hydrogen Bond Rich Protein Structures”. In: J. Am. Chem. Soc. 138.9, pp. 3046–3057. doi: 10.1021/jacs.5b11012. eprint: http://dx.doi.org/10.1021/jacs.5b11012. url: http://dx.doi.org/10.1021/jacs.5b11012.

Pitera, Jed W, Michael Falta, and Wilfred F van Gunsteren (2001). “Dielectric properties of proteins from simulation: the effects of solvent, ligands, pH, and temperature”. In: Biophysical journal 80.6, pp. 2546–2555.

Plumley, Joshua A. et al. (2014). “Capping Amyloid β-Sheets of the Tau-Amyloid Structure VQIVYK with Hexapeptides Designed To Arrest Growth. An ONIOM and Density Functional Theory Study”. In: J. Phys. Chem. B 118.12, pp. 3326–3334. doi: 10.1021/jp501890p. eprint: http://dx.doi.org/10.1021/jp501890p. url: http://dx.doi.org/10.1021/jp501890p.

Ponder, Jay W. and David A. Case (2003). Force Fields for Protein Simulations. Vol. 66. Advances in Protein Chemistry. Academic Press, pp. 27 –85. doi: https://doi.org/10.1016/S0065-3233(03)66002-X. url: http://www.sciencedirect.com/science/article/pii/S006532330366002X.

Ponder, Jay W. et al. (2010). “Current Status of the AMOEBA Polarizable Force Field”. In: J. Phys. Chem. B 114.8. PMID: 20136072, pp. 2549–2564.

Prasad, Saumya et al. (2017). “Near UV-Visible electronic absorption originating from charged amino acids in a monomeric protein”. In: Chemical science 8.8, pp. 5416–5433.

Privalov, Peter L. and George I. Makhatadze (1993). “Contribution of Hydration to Protein Folding Thermodynamics: II. The Entropy and Gibbs Energy of Hydration”. In: J. Mol. Biol. 232.2, pp. 660 –679. issn: 0022-2836. doi: https://doi.org/10.1006/jmbi.1993.1417. url: http://www.sciencedirect.com/science/article/pii/S0022283683714178.

Pronk, Sander et al. (2013). “GROMACS 4.5: a High-throughput and Highly Parallel Open Source Molecular Simulation Toolkit”. In: Bioinformatics 29.7, pp. 845–854. doi: 10.1093/bioinformatics/btt055. eprint: http://bioinformatics.oxfordjournals.org/content/29/7/845.full.pdf+html. url: http://bioinformatics.oxfordjournals.org/content/29/7/845.abstract.

Qi, Helena W., Hannah R. Leverentz, and Donald G. Truhlar (2013). “Water 16-mers and Hexamers: Assessment of the Three-Body and Electrostatically Embedded Many-Body Approximations of the Correlation Energy or the Nonlocal Energy As Ways to Include Cooperative Effects”. In: The Journal of Physical Chemistry A 117.21, pp. 4486–4499. doi: `10.1021/jp401463f`. eprint: `https://doi.org/10.1021/jp401463f`. url: `https://doi.org/10.1021/jp401463f`.

Qiang, Wei et al. (2012). “Antiparallel β-sheet Architecture in Iowa-mutant β-amyloid Fibrils”. In: Proc. Natl. Acad. Sci. U. S. A. 109.12, pp. 4443–4448. doi: `10.1073/pnas.1111305109`. eprint: `http://www.pnas.org/content/109/12/4443.full.pdf`. url: `http://www.pnas.org/content/109/12/4443.abstract`.

Rahaman, Obaidur et al. (2017). “Configurational Disorder of Water Hydrogen-Bond Network at the Protein Dynamical Transition”. In: J. Phys. Chem. B 121.28, pp. 6792–6798. doi: `10.1021/acs.jpcb.7b03888`. eprint: `https://doi.org/10.1021/acs.jpcb.7b03888`. url: `https://doi.org/10.1021/acs.jpcb.7b03888`.

Rahman, Aneesur and Frank H. Stillinger (1971). “Molecular Dynamics Study of Liquid Water”. In: J. Chem. Phys. 55.7, pp. 3336–3359. doi: `10.1063/1.1676585`. eprint: `https://doi.org/10.1063/1.1676585`. url: `https://doi.org/10.1063/1.1676585`.

Rao, Francesco, Sean Garrett-Roe, and Peter Hamm (2010). “Structural Inhomogeneity of Water by Complex Network Analysis”. In: J. Phys. Chem. B 114.47, pp. 15598–15604. doi: `10.1021/jp1060792`. eprint: `https://doi.org/10.1021/jp1060792`. url: `https://doi.org/10.1021/jp1060792`.

Rawat, Nidhi and Parbati Biswas (2014). “Hydrogen Bond Dynamics in Intrinsically Disordered Proteins”. In: The Journal of Physical Chemistry B 118.11. PMID: 24571104, pp. 3018–3025. doi: `10.1021/jp5013544`. eprint: `http://dx.doi.org/10.1021/jp5013544`. url: `http://dx.doi.org/10.1021/jp5013544`.

Reddy, Govardhan, John E. Straub, and D. Thirumalai (2010). “Dry Amyloid Fibril Assembly in a Yeast Prion Peptide is Mediated by Long-lived Structures Containing Water Wires”. In: Proc. Natl. Acad. Sci. U. S. A. 107.50, pp. 21459–21464. doi: `10.1073/pnas.1008616107`. eprint: `http://www.pnas.org/content/107/50/21459.full.pdf`. url: `http://www.pnas.org/content/107/50/21459.abstract`.

Reischl, Bernhard, Jürgen Köfinger, and Christoph Dellago (2009). “The statistics of electric field fluctuations in liquid water”. In: Mol. Phys. 107.4-6, pp. 495–502. doi: `10.1080/00268970902865493`. eprint: `https://doi.org/10.1080/00268970902865493`. url: `https://doi.org/10.1080/00268970902865493`.

Ren, Pengyu and Jay W. Ponder (2003). “Polarizable Atomic Multipole Water Model for Molecular Mechanics Simulation”. In: The Journal of Physical Chemistry B 107.24, pp. 5933–5947. doi: `10.1021/jp027815+`. eprint: `http://dx.doi.org/10.1021/jp027815+`. url: `http://dx.doi.org/10.1021/jp027815+`.

— (2004). “Temperature and Pressure Dependence of the AMOEBA Water Model”. In: The Journal of Physical Chemistry B 108.35, pp. 13427–13437. doi: `10.1021/`

jp0484332. eprint: http://dx.doi.org/10.1021/jp0484332. url: http://dx.doi.org/10.1021/jp0484332.

Riccardi, Laura, Phuong H. Nguyen, and Gerhard Stock (2012). “Construction of the Free Energy Landscape of Peptide Aggregation from Molecular Dynamics Simulations”. In: J. Chem. Theory Comput. 8.4, pp. 1471–1479. doi: 10.1021/ct200911w. eprint: http://dx.doi.org/10.1021/ct200911w. url: http://dx.doi.org/10.1021/ct200911w.

Roberts, Christopher J. (2007). “Non-native Protein Aggregation Kinetics”. In: Biotechnol. Bioeng. 98.5, pp. 927–938. issn: 1097-0290. doi: 10.1002/bit.21627. url: http://dx.doi.org/10.1002/bit.21627.

Robinson, G. W. et al. (1996). Water in Biology, Chemistry and Physics : Experimental Overviews and Computational Methodologies. World Scientific, pp. 001–004.

Rochet, Jean-Christophe and Peter T Lansbury (2000). “Amyloid fibrillogenesis: themes and variations”. In: Current Opinion in Structural Biology 10.1, pp. 60 –68. issn: 0959-440X. doi: https://doi.org/10.1016/S0959-440X(99)00049-4. url: http://www.sciencedirect.com/science/article/pii/S0959440X99000494.

Rodriguez, Alex and Alessandro Laio (2014). “Clustering by fast search and find of density peaks”. In: Science 344.6191, pp. 1492–1496. issn: 0036-8075. doi: 10.1126/science.1242072. eprint: http://science.sciencemag.org/content/344/6191/1492.full.pdf. url: http://science.sciencemag.org/content/344/6191/1492.

Rodriguez, Alex et al. (2018). “Computing the Free Energy without Collective Variables”. In: J. Chem. Theory Comput. 14.3, pp. 1206–1215. doi: 10.1021/acs.jctc.7b00916. eprint: https://doi.org/10.1021/acs.jctc.7b00916. url: https://doi.org/10.1021/acs.jctc.7b00916.

Rosenman, David J. et al. (2013). “Aβ Monomers Transiently Sample Oligomer and Fibril-Like Configurations: Ensemble Characterization Using a Combined MD/NMR Approach”. In: J. Mol. Biol. 425.18, pp. 3338–3359. issn: 0022-2836. doi: http://dx.doi.org/10.1016/j.jmb.2013.06.021. url: http://www.sciencedirect.com/science/article/pii/S0022283613004014.

Ross, Christopher A. and Michelle A. Poirier (2004). “Protein aggregation and neurodegenerative disease”. In: Nature Medicine 10, S10. url: http://dx.doi.org/10.1038/nm1066.

Rubinsztein, David C. (2006). “The roles of intracellular protein-degradation pathways in neurodegeneration”. In: Nature 443.7113, pp. 780–786. issn: 0028-0836. doi: 10.1038/nature05291. url: http://dx.doi.org/10.1038/nature05291.

Rüger, Robert et al. (2016). “Tight-binding approximations to time-dependent density functional theory—A fast approach for the calculation of electronically excited states”. In: The Journal of chemical physics 144.18, p. 184103.

Runge, Erich and E. K. U. Gross (1984a). “Density-Functional Theory for Time-Dependent Systems”. In: Phys. Rev. Lett. 52 (12), pp. 997–1000. doi: 10.1103/

PhysRevLett.52.997. url: https://link.aps.org/doi/10.1103/PhysRevLett.52.997.

Runge, Erich and Eberhard KU Gross (1984b). "Density-functional theory for time-dependent systems". In: Physical Review Letters 52.12, p. 997.

Russo, Daniela, Jacques Ollivier, and Jose Teixeira (2008). "Water hydrogen bond analysis on hydrophilic and hydrophobic biomolecule sites". In: Phys. Chem. Chem. Phys. 10 (32), pp. 4968–4974. doi: 10.1039/B807551B. url: http://dx.doi.org/10.1039/B807551B.

Russo, Daniela, JosÃ© Teixeira, and Jacques Ollivier (2009). "The impact of hydration water on the dynamics of side chains of hydrophobic peptides: From dry powder to highly concentrated solutions". In: J. Chem. Phys. 130.23, p. 235101. doi: 10.1063/1.3154383. eprint: https://doi.org/10.1063/1.3154383. url: https://doi.org/10.1063/1.3154383.

Rybicka, Anna et al. (2016). "Thioflavin T: Electronic Circular Dichroism and Circularly Polarized Luminescence Induced by Amyloid Fibrils". In: ChemPhysChem 17.18, pp. 2931–2937. issn: 1439-7641. doi: 10.1002/cphc.201600235. url: http://dx.doi.org/10.1002/cphc.201600235.

Röhrig, Ute F. et al. (2006). "Stability and Structure of Oligomers of the Alzheimer Peptide Aβ_{16-22}: From the Dimer to the 32-Mer". In: Biophysical Journal 91.9, pp. 3217 –3229. issn: 0006-3495. doi: https://doi.org/10.1529/biophysj.106.088542. url: http://www.sciencedirect.com/science/article/pii/S0006349506720343.

Sanfelice, Domenico et al. (2014). "Characterization of the Conformational Fluctuations in the Josephin Domain of Ataxin-3". In: Biophys. J. 107.12, pp. 2932–2940. issn: 0006-3495. doi: http://dx.doi.org/10.1016/j.bpj.2014.10.008. url: http://www.sciencedirect.com/science/article/pii/S0006349514010595.

Sass, Hans-Jürgen, Franziska Fang-Fang Schmid, and Stephan Grzesiek (2007). "Correlation of Protein Structure and Dynamics to Scalar Couplings across Hydrogen Bonds". In: Journal of the American Chemical Society 129.18. PMID: 17429967, pp. 5898–5903. doi: 10.1021/ja068336h. eprint: http://dx.doi.org/10.1021/ja068336h. url: http://dx.doi.org/10.1021/ja068336h.

Savarese, Marika et al. (2017). "Metrics for Molecular Electronic Excitations: A Comparison between Orbital-and Density-Based Descriptors". In: The Journal of Physical Chemistry A 121.40, pp. 7543–7549.

Scheuner, D. et al. (1996). "Secreted amyloid b-protein similar to that in the senile plaques of Alzheimer's disease is increased in vivo by the presenilin 1 and 2 and APP mutations linked to familial Alzheimer's disease". In: Nature Medicine 2, 864 EP –. url: http://dx.doi.org/10.1038/nm0896-864.

Schmidt, Andreas et al. (2016). "Cryo-EM reveals the steric zipper structure of a light chain-derived amyloid fibril". In: Proceedings of the National Academy of Sciences 113.22, pp. 6200–6205. issn: 0027-8424. doi: 10.1073/pnas.1522282113.

eprint: `http://www.pnas.org/content/113/22/6200.full.pdf`. url: `http://www.pnas.org/content/113/22/6200`.

Schmidt, Michael W. et al. (1993). “General atomic and molecular electronic structure system”. In: Journal of Computational Chemistry 14.11, pp. 1347–1363. issn: 1096-987X. doi: `10.1002/jcc.540141112`. url: `http://dx.doi.org/10.1002/jcc.540141112`.

Scott, D. W. (2015). Multivariate density estimation: theory, practice, and visualization. Wiley.

Selkoe, Dennis J and John Hardy (2016). “The amyloid hypothesis of Alzheimer’s disease at 25 years”. In: EMBO Molecular Medicine 8.6, pp. 595–608. issn: 1757-4676. doi: `10.15252/emmm.201606210`. eprint: `http://embomolmed.embopress.org/content/8/6/595.full.pdf`. url: `http://embomolmed.embopress.org/content/8/6/595`.

Sengupta, Urmi, Ashley N. Nilson, and Rakez Kayed (2016). “The Role of Amyloid-β Oligomers in Toxicity, Propagation, and Immunotherapy”. In: EBioMedicine 6, pp. 42 –49. issn: 2352-3964. doi: `https://doi.org/10.1016/j.ebiom.2016.03.035`. url: `http://www.sciencedirect.com/science/article/pii/S2352396416301219`.

Shankar, Ganesh M. et al. (2008). “Amyloid-[beta] Protein Dimers Isolated Directly from Alzheimer’s Brains Impair Synaptic Plasticity and Memory”. In: Nat. Med. 14.8, pp. 837–842. issn: 1078-8956. doi: `10.1038/nm1782`. url: `http://dx.doi.org/10.1038/nm1782`.

Shanmugam, Ganesh and Prasad L. Polavarapu (2004). “Structure of Aβ(25-35) Peptide in Different Environments”. In: Biophys. J. 87.1, pp. 622–630. issn: 0006-3495. doi: `http://dx.doi.org/10.1529/biophysj.104.040907`. url: `http://www.sciencedirect.com/science/article/pii/S000634950473547X`.

Sharpe, Simon et al. (2011). “Solid-State NMR Characterization of Autofluorescent Fibrils Formed by the Elastin-Derived Peptide GVGVAGVG”. In: Biomacromolecules 12.5. PMID: 21456595, pp. 1546–1555. doi: `10.1021/bm101486s`. eprint: `https://doi.org/10.1021/bm101486s`. url: `https://doi.org/10.1021/bm101486s`.

Sheather, Simon J. (2004). “Density Estimation”. In: Statistical Science 19.4, pp. 588–597. issn: 08834237. url: `http://www.jstor.org/stable/4144429`.

Shi, Yue et al. (2013). “Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins”. In: Journal of Chemical Theory and Computation 9.9, pp. 4046–4063. doi: `10.1021/ct4003702`. eprint: `http://dx.doi.org/10.1021/ct4003702`. url: `http://dx.doi.org/10.1021/ct4003702`.

Shukla, Anshuman et al. (2004). “A novel UV laser-induced visible blue radiation from protein crystals and aggregates: scattering artifacts or fluorescence transitions of peptide electrons delocalized through hydrogen bonding?” In: Archives of Biochemistry and Biophysics 428.2, pp. 144 –153. issn: 0003-9861. doi: `https:`

//doi.org/10.1016/j.abb.2004.05.007. url: http://www.sciencedirect.com/science/article/pii/S0003986104002693.

Sibley, Anaika B., Monique Cosman, and V.V. Krishnan (2003). "An Empirical Correlation between Secondary Structure Content and Averaged Chemical Shifts in Proteins". In: Biophys. J. 84.2, pp. 1223–1227. issn: 0006-3495. doi: http://dx.doi.org/10.1016/S0006-3495(03)74937-6. url: //www.sciencedirect.com/science/article/pii/S0006349503749376.

Smith, Micholas Dean et al. (2015). "Force-Field Induced Bias in the Structure of Aβ21-30: A Comparison of , AMBER, CHARMM, and GROMOS Force Fields". In: J. Chem. Inf. Model. 55.12, pp. 2587–2595. doi: 10.1021/acs.jcim.5b00308. eprint: http://dx.doi.org/10.1021/acs.jcim.5b00308. url: http://dx.doi.org/10.1021/acs.jcim.5b00308.

Smolin, Nikolai et al. (2005). "Properties of Spanning Water Networks at Protein Surfaces". In: J. Phys. Chem. B 109.21, pp. 10995–11005. doi: 10.1021/jp050153e. eprint: http://dx.doi.org/10.1021/jp050153e. url: http://dx.doi.org/10.1021/jp050153e.

Soto, Patricia, Mary A. Griffin, and Joan Emma Shea (2007). "New Insights into the Mechanism of Alzheimer Amyloid-β Fibrillogenesis Inhibition by N-Methylated Peptides". In: Biophys. J. 93.9, pp. 3015–3025. issn: 0006-3495. doi: http://dx.doi.org/10.1529/biophysj.107.112086. url: http://www.sciencedirect.com/science/article/pii/S0006349507715564.

Stam, Cornelis J. and Jaap C. Reijneveld (2007). "Graph Theoretical Analysis of Complex Networks in the Brain". In: Nonlinear Biomed. Phys. 1.1, p. 3. issn: 1753-4631. doi: 10.1186/1753-4631-1-3. url: http://dx.doi.org/10.1186/1753-4631-1-3.

Stillinger, Frank H. (1980). "Water Revisited". In: Science 209.4455, pp. 451–457. issn: 0036-8075. doi: 10.1126/science.209.4455.451. eprint: http://science.sciencemag.org/content/209/4455/451.full.pdf. url: http://science.sciencemag.org/content/209/4455/451.

Straub, John E. and D. Thirumalai (2011). "Toward a Molecular Theory of Early and Late Events in Monomer to Amyloid Fibril Formation". In: Annu. Rev. Phys. Chem. 62.1, pp. 437–463. doi: 10.1146/annurev-physchem-032210-103526. eprint: http://dx.doi.org/10.1146/annurev-physchem-032210-103526. url: http://dx.doi.org/10.1146/annurev-physchem-032210-103526.

Su, Peifeng and Hui Li (2009). "Energy decomposition analysis of covalent bonds and intermolecular interactions". In: The Journal of Chemical Physics 131.1, p. 014102. doi: 10.1063/1.3159673. eprint: http://dx.doi.org/10.1063/1.3159673. url: http://dx.doi.org/10.1063/1.3159673.

Sugita, Yuji and Yuko Okamoto (1999). "Replica-exchange Molecular Dynamics Method for Protein Folding". In: Chem. Phys. Lett. 314.1-2, pp. 141–151. issn: 0009-2614. doi: http://dx.doi.org/10.1016/S0009-2614(99)01123-9. url: http://www.sciencedirect.com/science/article/pii/S0009261499011239.

Swope, William C. et al. (1982). “A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters”. In: The Journal of Chemical Physics 76.1, pp. 637–649. doi: 10.1063/1.442716. eprint: https://doi.org/10.1063/1.442716. url: https://doi.org/10.1063/1.442716.

Takahashi, Reisuke H. et al. (2004). “Oligomerization of Alzheimer’s β-Amyloid within Processes and Synapses of Cultured Neurons and Brain”. In: Journal of Neuroscience 24.14, pp. 3592–3599. issn: 0270-6474. doi: 10.1523/JNEUROSCI.5167-03.2004. eprint: http://www.jneurosci.org/content/24/14/3592.full.pdf. url: http://www.jneurosci.org/content/24/14/3592.

Tan, Hongwei et al. (2005). “The Role of Charge Transfer in the Hydrogen Bond Cooperative Effect of cis-N-Methylformamide Oligomers”. In: The Journal of Physical Chemistry A 109.28, pp. 6303–6308. doi: 10.1021/jp051444q. eprint: http://dx.doi.org/10.1021/jp051444q. url: http://dx.doi.org/10.1021/jp051444q.

Tarek, Mounir and Douglas J. Tobias (2000). “The Dynamics of Protein Hydration Water: A Quantitative Comparison of Molecular Dynamics Simulations and Neutron-scattering Experiments”. In: Biophys. J. 79.6, pp. 3244–3257. issn: 0006-3495. doi: https://doi.org/10.1016/S0006-3495(00)76557-X. url: http://www.sciencedirect.com/science/article/pii/S000634950076557X.

Tarus, Bogdan, John E. Straub, and D. Thirumalai (2006). “Dynamics of Asp23-Lys28 Salt-Bridge Formation in Aβ10-35 Monomers”. In: J. Am. Chem. Soc. 128.50, pp. 16159–16168. doi: 10.1021/ja064872y. eprint: http://dx.doi.org/10.1021/ja064872y. url: http://dx.doi.org/10.1021/ja064872y.

Thirumalai, D., Govardhan Reddy, and John E. Straub (2012). “Role of Water in Protein Aggregation and Amyloid Polymorphism”. In: Acc. Chem. Res. 45.1, pp. 83–92. doi: 10.1021/ar2000869. eprint: http://dx.doi.org/10.1021/ar2000869. url: http://dx.doi.org/10.1021/ar2000869.

Tiana, G. (2008). “Estimation of microscopic averages from metadynamics”. In: Eur. Phys. J. B 63.2, pp. 235–238. doi: 10.1140/epjb/e2008-00232-8. url: https://doi.org/10.1140/epjb/e2008-00232-8.

Tikhonova, Tatiana N. et al. (2018). “Dissection of the deep-blue autofluorescence changes accompanying amyloid fibrillation”. In: Archives of Biochemistry and Biophysics 651, pp. 13 –20. issn: 0003-9861. doi: https://doi.org/10.1016/j.abb.2018.05.019. url: http://www.sciencedirect.com/science/article/pii/S0003986117308664.

Tiwary, Pratyush and Michele Parrinello (2015). “A Time-Independent Free Energy Estimator for Metadynamics”. In: J. Phys. Chem. B 119.3, pp. 736–742. doi: 10.1021/jp504920s. eprint: http://dx.doi.org/10.1021/jp504920s. url: http://dx.doi.org/10.1021/jp504920s.

Tomasi, Jacopo, Benedetta Mennucci, and Roberto Cammi (2005). “Quantum mechanical continuum solvation models”. In: Chemical Reviews-Columbus 105.8, pp. 2999–3094.

Torrie, G.M. and J.P. Valleau (1977). “Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling”. In: Journal of Computational Physics 23.2, pp. 187 –199. issn: 0021-9991. doi: https://doi.org/10.1016/0021-9991(77)90121-8. url: http://www.sciencedirect.com/science/article/pii/0021999177901218.

Toxvaerd, So/ren (1987). “Comment on constrained molecular dynamics of macromolecules”. In: J. Chem. Phys. 87.10, pp. 6140–6143. doi: 10.1063/1.453488. eprint: http://dx.doi.org/10.1063/1.453488. url: http://dx.doi.org/10.1063/1.453488.

Tribello, Gareth A. et al. (2014). “PLUMED 2: New Feathers for an Old Bird”. In: Comput. Phys. Commun. 185.2, pp. 604–613. issn: 0010-4655. doi: http://dx.doi.org/10.1016/j.cpc.2013.09.018. url: http://www.sciencedirect.com/science/article/pii/S0010465513003196.

Urbanc, B. et al. (2004). “Molecular Dynamics Simulation of Amyloid β Dimer Formation”. In: Biophys. J. 87.4, pp. 2310–2321. issn: 0006-3495. doi: http://dx.doi.org/10.1529/biophysj.104.040980. url: http://www.sciencedirect.com/science/article/pii/S0006349504737078.

Uversky, Vladimir N., Christopher J. Oldfield, and A. Keith Dunker (2008). “Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept”. In: Annu. Rev. Biophys. 37.1, pp. 215–246. doi: 10.1146/annurev.biophys.37.032807.125924. eprint: http://dx.doi.org/10.1146/annurev.biophys.37.032807.125924. url: http://dx.doi.org/10.1146/annurev.biophys.37.032807.125924.

Van Gunsteren, W. F. and Martin Karplus (1982). “Effect of constraints on the dynamics of macromolecules”. In: Macromolecules 15.6, pp. 1528–1544. doi: 10.1021/ma00234a015. eprint: http://dx.doi.org/10.1021/ma00234a015. url: http://dx.doi.org/10.1021/ma00234a015.

Verlet, Loup (1967). “Computer ”Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules”. In: Phys. Rev. 159 (1), pp. 98–103. doi: 10.1103/PhysRev.159.98. url: https://link.aps.org/doi/10.1103/PhysRev.159.98.

Vitkup, Dennis et al. (2000). “Solvent mobility and the protein ’glass’ transition”. In: Nat. Struct. Biol. 7, pp. 34–38. url: http://dx.doi.org/10.1038/71231.

Wälti, Marielle Aulikki et al. (2016). “Atomic-resolution structure of a disease-relevant Aβ(1–42) amyloid fibril”. In: Proceedings of the National Academy of Sciences 113.34, E4976–E4984. issn: 0027-8424. doi: 10.1073/pnas.1600749113. eprint: http://www.pnas.org/content/113/34/E4976.full.pdf. url: http://www.pnas.org/content/113/34/E4976.

Wang, Fugao and D. P. Landau (2001). “Efficient, Multiple-Range Random Walk Algorithm to Calculate the Density of States”. In: Phys. Rev. Lett. 86 (10), pp. 2050–2053. doi: 10.1103/PhysRevLett.86.2050. url: https://link.aps.org/doi/10.1103/PhysRevLett.86.2050.

Wang, Haiqiang, Jan Forsman, and Clifford E Woodward (2016). “Density functional theory of equilibrium random copolymers: application to surface adsorption of aggregating peptides”. In: Journal of Physics: Condensed Matter 28.24, p. 244011. url: `http://stacks.iop.org/0953-8984/28/i=24/a=244011`.

Wang, Tuo et al. (2017). “Water Distribution, Dynamics, and Interactions with Alzheimer’s β-Amyloid Fibrils Investigated by Solid-State NMR”. In: Journal of the American Chemical Society 139.17. PMID: 28406028, pp. 6242–6252. doi: `10.1021/jacs.7b02089`. eprint: `https://doi.org/10.1021/jacs.7b02089`. url: `https://doi.org/10.1021/jacs.7b02089`.

Wei, Guanghong and Joan Emma Shea (2006). “Effects of Solvent on the Structure of the Alzheimer Amyloid-β(25-35) Peptide”. In: Biophys. J. 91.5, pp. 1638–1647. issn: 0006-3495. doi: `http://dx.doi.org/10.1529/biophysj.105.079186`. url: `http://www.sciencedirect.com/science/article/pii/S0006349506718781`.

Xantheas, Sotiris S (2000). “Cooperativity and hydrogen bonding network in water clusters”. In: Chemical Physics 258.2, pp. 225 –231. issn: 0301-0104. doi: `https://doi.org/10.1016/S0301-0104(00)00189-0`. url: `http://www.sciencedirect.com/science/article/pii/S0301010400001890`.

Xi, Wenhui, Wenfei Li, and Wei Wang (2012). “Template Induced Conformational Change of Amyloid-β Monomer”. In: J. Phys. Chem. B 116.25, pp. 7398–7405. doi: `10.1021/jp300389g`. eprint: `http://dx.doi.org/10.1021/jp300389g`. url: `http://dx.doi.org/10.1021/jp300389g`.

Xu, Weixin et al. (2013). “pH-Dependent Conformational Ensemble and Polymorphism of Amyloid-β Core Fragment”. In: J. Phys. Chem. B 117.28, pp. 8392–8399. doi: `10.1021/jp404034x`. eprint: `http://dx.doi.org/10.1021/jp404034x`. url: `http://dx.doi.org/10.1021/jp404034x`.

Yan, Yilin and Chunyu Wang (2006). “Aβ42 is More Rigid than Aβ40 at the C Terminus: Implications for Aβ Aggregation and Toxicity”. In: J. Mol. Biol. 364.5, pp. 853–862. issn: 0022-2836. doi: `http://dx.doi.org/10.1016/j.jmb.2006.09.046`. url: `http://www.sciencedirect.com/science/article/pii/S0022283606012551`.

Yanai, Takeshi, David P Tew, and Nicholas C Handy (2004). “A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP)”. In: Chemical Physics Letters 393.1-3, pp. 51–57.

Yang, Y. Isaac and Yi Qin Gao (2015). “Computer Simulation Studies of Aβ_{37-42} Aggregation Thermodynamics and Kinetics in Water and Salt Solution”. In: The Journal of Physical Chemistry B 119.3, pp. 662–670. doi: `10.1021/jp502169b`. eprint: `http://dx.doi.org/10.1021/jp502169b`. url: `http://dx.doi.org/10.1021/jp502169b`.

Ye, Ruquan et al. (2017). “Non-conventional fluorescent biogenic and synthetic polymers without aromatic rings”. In: Polym. Chem. (10), pp. 1722–1727. doi: `10.1039/C7PY00154A`. url: `http://dx.doi.org/10.1039/C7PY00154A`.

Zhang, S. et al. (2000). “The Alzheimer’s Peptide Aβ Adopts a Collapsed Coil Structure in Water”. In: J. Struct. Biol. 130.2, pp. 130–141. issn: 1047-8477. doi: http://dx.doi.org/10.1006/jsbi.2000.4288. url: http://www.sciencedirect.com/science/article/pii/S1047847700942886.

Zhang, Zhe, Shawn Witham, and Emil Alexov (2011). “On the role of electrostatics on protein-protein interactions”. In: Phys. Biol. 8.3, pp. 035001–035001. issn: 1478-3967. doi: 10.1088/1478-3975/8/3/035001. url: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3137121/.

Zhao, Li, Wenzhao Li, and Pu Tian (Apr. 2013). “Reconciling Mediating and Slaving Roles of Water in Protein Conformational Dynamics”. In: PLoS One 8.4, pp. 1–7. doi: 10.1371/journal.pone.0060553. url: https://doi.org/10.1371/journal.pone.0060553.

Zheng, Ya-Jun and Kenneth M. Merz (1992). “Study of hydrogen bonding interactions relevant to biomolecular structure and function”. In: Journal of Computational Chemistry 13.9, pp. 1151–1169. issn: 1096-987X. doi: 10.1002/jcc.540130916. url: http://dx.doi.org/10.1002/jcc.540130916.

Zhu, Tong, John Z. H. Zhang, and Xiao He (2013). “Automated Fragmentation QM/MM Calculation of Amide Proton Chemical Shifts in Proteins with Explicit Solvent Model”. In: J. Chem. Theory Comput. 9.4, pp. 2104–2114. doi: 10.1021/ct300999w. url: http://dx.doi.org/10.1021/ct300999w.

Zwanzig, R (1988). “Diffusion in a rough potential”. In: Proceedings of the National Academy of Sciences 85.7, pp. 2029–2030. issn: 0027-8424. doi: 10.1073/pnas.85.7.2029. eprint: http://www.pnas.org/content/85/7/2029.full.pdf. url: http://www.pnas.org/content/85/7/2029.