Triggers for Protein Conformational Changes and Self-Assembly Probed with Fully Atomistic Computer Simulations

Thesis submitted to AcSIR for the Award of the Degree of

DOCTOR OF PHILOSOPHY

In

Chemical Sciences

By

Sneha Menon

Registration No.: 10CC13A26019

Under the guidance of

Dr. Nayana Vaval   **Dr. Neelanjana Sengupta**

(Supervisor)   **(Co-supervisor)**

CSIR–National Chemical Laboratory, Pune, India

DECLARATION

I hereby declare that the thesis entitled "**Triggers for protein conformational changes and self-assembly probed with fully atomistic computer simulations**" submitted for the degree of **Doctor of Philosophy in Chemical Sciences** to the Academy of Scientific & Innovative Research (AcSIR), has been carried out by me at the Physical and Materials Chemistry Division of CSIR–National Chemical Laboratory, Pune under the guidance of **Dr. Nayana Vaval** (Supervisor) and **Dr. Neelanjana Sengupta** (Co-supervisor). Such material as has been obtained by other sources has been duly acknowledged in this thesis. The work is original and has not been submitted in part or full by me for any other degree or diploma to any other Institution or University.

Date: **Sneha Menon**

Place: **CSIR–NCL, Pune**   (Ph.D. candidate)

"I am seeking, I am striving, I am in it with all my heart."

– Vincent van Gogh

Acknowledgements

I take this opportunity to express my sincere gratitude and appreciation to all those were supportive of my endeavor to earn a Ph.D.; inspired and helped me in this arduous doctoral journey.

I would first like to thank Dr. Neelanjana Sengupta, under whose supervision I commenced my Ph.D. at CSIR-National Chemical Laboratory (CSIR-NCL). Her advice, encouragement, positive criticism and confidence in me has been instrumental in molding my scientific temperament and helped me accomplish my research goals. She ensured that her relocation to another institute did not hamper my work adversely and continued supervising my doctoral research to completion. I am immensely thankful that she incorporated me in her new lab at the Indian Institute of Science Education and Research Kolkata (IISER Kolkata) and helped me have a productive and comfortable stay at the institute.

I sincerely thank my present and past advisors, Dr. Nayana Vaval and Dr. Debashree Ghosh, for supporting and guiding me through all the official procedures at CSIR-NCL. I am grateful to my Doctoral Advisory Committee members, Dr. Sayan Bagchi, Dr. Chetan Gadgil and Dr. T. G. Ajithkumar for evaluating my progress and providing useful commentary that helped in improving my research outlook. I thank the former and present CSIR-NCL directors, Dr. Sourav Pal and Dr. Ashwini Kumar Nangia; departmental heads of Physical Chemistry Division, Dr. Anil Kumar and Dr. P. A. Joy, for facilitating the necessary research infrastructure. I would like to extend my sincere thanks to the Student Academic Office for their administrative support and the AcSIR office at CSIR-NCL for providing a platform to facilitate the Ph.D. process. I acknowledge DBT-BINC for providing the research fellowship and the Accounts & Bills sections for mediating the fellowship release.

I express my deep appreciation to my senior labmates at CSIR-NCL; Jaya, Asis and Prathit for the constant help and support they provided me. I had a wonderful experience sharing workspace with the DG group members (Mukunda,

Table of Contents

List of figures

x

List of tables

Abstract

A central tenet of biology states that proteins attain three-dimensional structures that are uniquely suited to their associated biological functions. The stability and activity of proteins are modulated by the primary amino acid sequence as well as external environmental factors such as temperature, pH, pressure, molecular crowding, solvent composition, etc. Intrinsically disordered proteins (IDPs) defy the structure-function paradigm and populate thermodynamically equivalent low energy states separated by barriers. Misfolding or dysregulation of IDPs render them exceedingly prone to aggregate into higher order structures including polymorphic soluble oligomers and amyloid fibrils that accumulate in intracellular and extracellular regions. These polymeric assemblies have inappropriate interactions with the cellular components and have been linked to several neurodegenerative and other diseases. The broad conformational heterogeneity and transient nature of IDPs pose manifold challenges to the conventional experimental techniques used in their detailed structural characterization. Computer simulations play essential complementary roles and generate structural ensembles of IDPs that can help in the interpretation of the ensemble-averaged experimental observations. In this thesis, fully atomistic molecular dynamics simulations has been employed to discern the effects of temperature, crowded environment and structural perturbations on the conformational stability, dynamics and early self-assembly of representative IDPs, namely the full-length amyloid beta (Aβ) peptide, cellular prion protein (PrP^C) and an amyloid oligomer of αB-crystallin (ABC). Investigation of the thermal stability of the amyloid

oligomer reveals that the cold thermal response of the amyloid assembly contrasts with the general behavior exhibited by globular proteins. The oligomer is increasingly stabilized as the temperature is lowered below the melting temperature of the water model used in the study. However, the equilibrium free energy profile based on packing of the hydrophobic core is similar to the response of globular proteins. Evaluation of the energetic coupling between the oligomer and the hydration waters highlights the crucial association of the solvation waters with the oligomer structure and stability. In view of the growing evidences linking type-2 diabetes mellitus (T2DM) with Alzheimer's disease (AD), the physical effects of glucose crowding on the early self-assembly (dimerization) of the Aβ peptide is studied. The results suggest that glucose crowding does not promote Aβ self-assembly and the dimers formed in this environment are weakly bound than those formed in pure water. Glucose molecules cluster away from the surface of the peptides and entrap the solvation waters thus shielding the hydrophobic interactions, which drive the process of Aβ self-assembly. The structural and dynamic effects of a minor perturbation in the network of native interactions of PrP^C are studied. Eliminating a single inter-domain hydrogen bond cascades into enhanced conformational flexibility of the protein, overall increased solvent exposure of the hydrophobic core and disruption of the inter-domain salt-bridge network. These changes corroborate with those reported for PrP^C under destabilizing conditions of low pH, high temperature and in disease mutants. The results suggest that perturbing the cooperative interplay of the native network of PrP^C interactions trigger structural and dynamical changes leading to conformations prone to pathogenic transitions. The thesis provides insights into aspects of conformational stability, dynamics and self-assembly of amyloidogenic proteins.

List of publications

1. Menon, S.; Sengupta, N. Perturbations in inter-domain associations may trigger the onset of pathogenic transformations in PrP^C: insights from atomistic simulations. *Mol. Biosyst.* **2015,** *11* (5), 1443–1453.

2. Menon, S.; Sengupta, N. Influence of Hyperglycemic Conditions on Self-Association of the Alzheimer's Amyloid β (Aβ$_{1-42}$) Peptide. *ACS Omega* **2017,** *2* (5), 2134–2147.

3. Menon, S.; Sengupta, N. Computational approaches to understanding the biological behaviour of intrinsically disordered proteins. *Curr. Sci.* (00113891) **2017,** *112* (7), 1444–1454.

4. Menon, S.; Sengupta, N. Influence of crowding and surfaces on protein amyloidogenesis: A thermo-kinetic perspective. *Biochim. Biophys. Acta Proteins Proteom.* **2019**.

5. Menon, S.; Sengupta, N. The Cold Thermal Response of an Amyloid Oligomer Differs from Typical Globular Protein Cold Denaturation. *J. Phys. Chem. Lett.* **2019,** *10* (10), 2453–2457.

Introduction

1.1 Protein Folding

Proteins are complex biological macromolecules performing myriad essential functions that sustain living systems. They are constituted of various combinations of 20 naturally occurring amino acid residues connected by peptide bonds and synthesized by the ribosomal machinery of the cell. The specific order of amino acids in a protein, referred to as its *"primary structure"*, is encoded in the deoxyribonucleic acid (DNA) that is the genetic material of all living organisms. The polypeptide chain upon synthesis undergoes post-translational modifications and folds to attain a unique three-dimensional structure associated with distinct biological function. One of the prime theories and a major milestone in the field of protein science is the *"thermodynamic hypothesis"* put forth by Christian Anfinsen in the early 1960s, based on denaturation and renaturation studies of the protein, ribonuclease A[1-2]. This hypothesis asserts that the native state, i.e. the three-dimensional structure adopted by a protein under normal physiological conditions (pH, temperature, ionic strength,

---

A part of this Chapter is based on the review articles:

1. Menon, S.; Sengupta, N. Influence of crowding and surfaces on protein amyloidogenesis: A thermo-kinetic perspective. *Biochim. Biophys. Acta Proteins Proteom.* **2019**.

2. Menon, S.; Sengupta, N. Computational approaches to understanding the biological behaviour of intrinsically disordered proteins. *Curr. Sci.* (00113891) **2017,** *112* (7), 1444–1454.

presence of metal ions, etc.), is energetically most favorable or at the global minimum of their accessible free energies. The native conformation is dictated by the sum total of interatomic interactions and thus by the primary amino-acid sequence[3]. Thereafter, in the late 1960s, Cyrus Levinthal conjectured the protein folding paradox, also referred to as "*Levinthal's paradox*". According to this conjecture, a protein would require cosmological timescales to attain the native fold if it systematically samples every possible conformation[4]. Thus it was suggested that proteins must fold via specific "*folding pathways*". The paradox is essentially a matter of two mutually exclusive alternatives: (i) thermodynamic control in which a protein folds to the global minimum energy structure in a path independent manner, (ii) kinetic control in which a protein folds quickly or in biological timescales in a pathway dependent manner[5-6]. This pivotal argument led to a major experimental and theoretical quest to characterize the kinetics of protein folding and identify partially structured intermediates along the folding pathway[7-10].

1.1.1   Energy Landscape Perspective of Protein Folding

The classical view of protein folding is based on simple phenomenological models and postulates a unique folding pathway with structured and sequential intermediates[5,8-9]. The new view[5,11-13], derived from theoretical models and computer simulations describe the process in terms of folding funnels or energy landscapes. Protein folding energy landscape is a statistical description of the potential energy surface of a protein as a function of its conformational degrees of freedom[14]. Each conformation represents a point on the multidimensional potential energy surface; the unfolded conformations are the high-energy species whereas folded conformations are energetically downhill[15]. As folding proceeds, a nascent polypeptide chain diffuses

downward in terms of energy owing to the increase in the number of stabilizing interactions[15-17]. The folding process is thus a stochastic search of the multiple conformations accessible to a polypeptide chain such that the protein folds rapidly and efficiently[16].

In recent years, a growing arsenal of advanced experimental techniques[18-28] and substantial enhancement in computing power[29-32] harnessed in computational methods has revolutionized our understanding of the protein folding process[33]. Since the number of conformations that can be sampled by a polypeptide chain grows exponentially with its chain length, folding of small single domain proteins (<100 amino acid residues in length) were mainly studied by experimental and theoretical techniques[34-36]. The folding landscape of small globular proteins is generally smooth, devoid of deep valleys and high barriers; the folding reaction can be modeled by a two-state transition between a disordered unfolded state and the ordered native state[17,37-38]. Most globular proteins fold on rugged energy landscapes and need to cross substantial kinetic energy barriers; consequently populating transient partially folded intermediate species[16-17,39]. The roughness in the folding funnel can be attributed to topological frustration during the formation of native interactions and the extent of non-native interactions[40-42]. A schematic representation of the funnel-shaped protein folding energy landscape is depicted in Figure 1.1. During folding, unfolded polypeptides sample a multitude of conformations as they evolve downhill towards the thermodynamically favorable native state, occupying the narrowest part of the funnel with minimum energy.
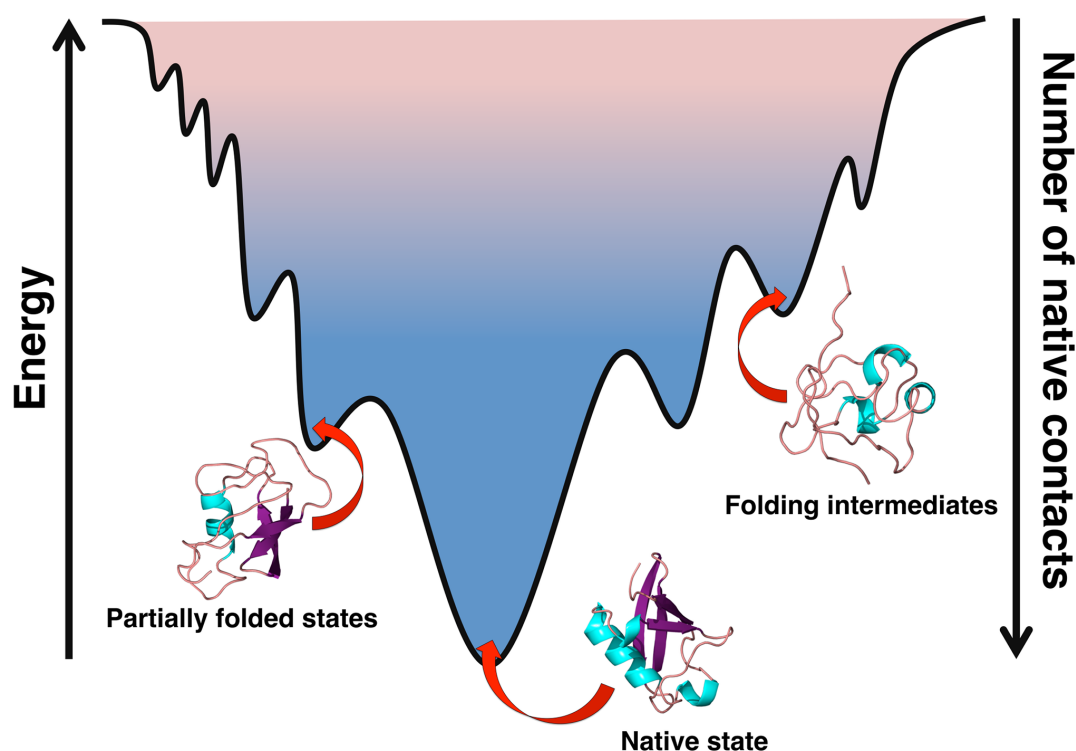
Figure 1.1. Schematic description of a protein folding energy landscape. Folding proceeds with an increment in the number of native interactions and a corresponding decrease in the energy as the polypeptide attains the native conformation.

Protein folding processes, in essence, are driven by random fluctuations in the unfolded conformation that results in the formation of native and non-native interactions; the former assumed to be more stable[17]. Rapid and efficient conformational search during protein folding involves formation of native interactions between key residues that establish the folding nucleus about which the rest of the structure organizes to attain the native topology[17]. These intermediate states have rudimentary native-like architecture that when correctly formed invariably leads to the native state[16]. The role of these heterogeneous species in assisting the folding process or serving as kinetic traps is largely debated[16-17,43]. Characterization of the

species encountered in the process of protein folding by employing advanced biophysical methods of faster timescales and greater sensitivity in combination with computational methods and simulations have provided insights into the underlying principles of protein folding[44-46]. According to one plausible model of protein folding known as the "*hydrophobic collapse*" model, in the early steps of folding, hydrophobic forces lead to the collapse of the unfolded polypeptide chain and the burial of non-polar amino acid residues within the core of the protein[47-48]. Locally folded structures called "*secondary structure elements*" are formed concomitant with the increase in compaction. The free energy preference for sequestering hydrophobic residues coupled with the weak preference for formation of various kinds of secondary structures drives the protein molecule towards the native conformational space, thus restricting the conformational search during folding[15-17,49]. In an alternative model referred to as the "*framework*" model, nascent secondary structures form first which coalesce and are connected by long–range interactions to slowly develop tertiary structure[50-52]. The most commonly occurring secondary structures in proteins are α-helices and β-sheets that are primarily stabilized by hydrogen-bonding interactions between the carbonyl and amide groups of the polypeptide backbone[16-17]. In an α-helix, the carbonyl oxygen of the peptide linkage of an amino acid is hydrogen-bonded to the amide hydrogen of the fourth amino acid down the chain[53]. The helical twist is right-handed in all proteins with 3.6 amino acids in each turn and side chains protruding outward from the helical backbone. In a β-sheet, segments of a polypeptide chain with extended backbone, termed as β-strands, are laterally connected by backbone hydrogen bonds; the side chains of adjacent amino acid residues point in opposite directions, above and below the plane of the sheet. The β-strands can run in the same direction (parallel β-sheet) or in the opposite direction

(antiparallel β-sheet), having the same or opposite amino-to-carboxyl orientations, respectively. Secondary structure elements subsequently form domains stabilized by tertiary interactions. The final stage of folding is highly cooperative in which close packing of the side chains of residues, excluding water from the protein core results in the formation of the native state[16,49].

1.2   Protein Misfolding, Aggregation and Amyloid Formation

The protein folding process is complex and inherently error-prone due to the enormous number of possible conformations that polypeptide chains can adopt[16-17,54-55]. Since the kinetically trapped partially folded species inevitably expose the residues that are otherwise buried in the folded native state, to the solvent; they entail the chances of misfolding and aggregation owing to the formation of non-native contacts with other proteins in the cellular crowded environment[6,16-17]. Such situations are further aggravated by disease-associated mutations, environmental stress and aging[55]. Cells have therefore evolved rescue strategies with molecular chaperones and protein degradation machineries or "*proteostasis network*" that identify and annihilate such misfolded protein products for maintaining a healthy proteome[54]. Molecular chaperones synthesized in all cell-types and cellular compartments are involved in *de novo* protein folding and aid in efficient folding, while preventing aberrant aggregation reactions[56]. Some chaperones are associated with the ribosome and interact with nascent chains emerging from the ribosome by transiently shielding the exposed hydrophobic residues, thus averting formation of unfavorable intra- and intermolecular interactions[56]. On the other hand, some chaperone proteins function downstream of the ribosome and assist the final stages of the folding process[57].

Despite the cellular strategies of protein quality control, proteins often misfold and have a heightened tendency to aggregate, resulting in the malfunctioning of the cellular machinery[58]. Such events are linked to a range of debilitating human diseases, generally referred to as protein misfolding or conformational diseases that can be categorized into loss-of-function and toxic gain-of-function diseases[16,55,58-59]. Loss-of-function pathologies arise from incorrect protein folding mostly due to familial mutations, leading to protein dysfunction and disease[60-61]. In other cases of disorders, misfolding of a particular protein is manifested by the formation of intracellular and (mostly) extracellular aggregates associated with cellular toxicity. These pathologies include neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), spongiform encephalopathies like Creutzfeldt-Jakob diseases (CJD) as well as type-2 diabetes, cardiovascular diseases and cancer, caused by deposition of such aggregates in tissues of the brain, heart and spleen[16,62-63]. Aggregation may also result from genetic mutations in the causative proteins such as in Huntington's disease (HD) and in early-onset AD and PD; the majority, however, are sporadic and age-related, facilitated by a gradual decline in the cellular proteostasis capacity with aging[58,64-65].

1.2.1 Nature of the Amyloid State

Protein aggregation diseases, also called "*amyloidosis*", involve abnormal build-up of a specific protein in the form of elongated and unbranched fibrils known as "*amyloid fibrils*"[16]. Figure 1.2 depicts the folded and amyloid fibrillar form of human α-synuclein protein implicated in Parkinson's disease. Amyloid quantities can be as low as to be practically undetectable, to even kilograms of aggregates found in systemic amyloid diseases[66]. The precursor proteins involved in these diseases are structurally

diverse including small peptides, natively unfolded proteins, natively folded monomeric proteins, and large protein assemblies[17]. A remarkable feature of these diseases is that despite having no sequence similarity and varied native folds, the insoluble, aggregated forms of these proteins have a well-defined fibrillar nature of very similar morphologies and a characteristic cross-β diffraction pattern[67-70].

a. **b.**
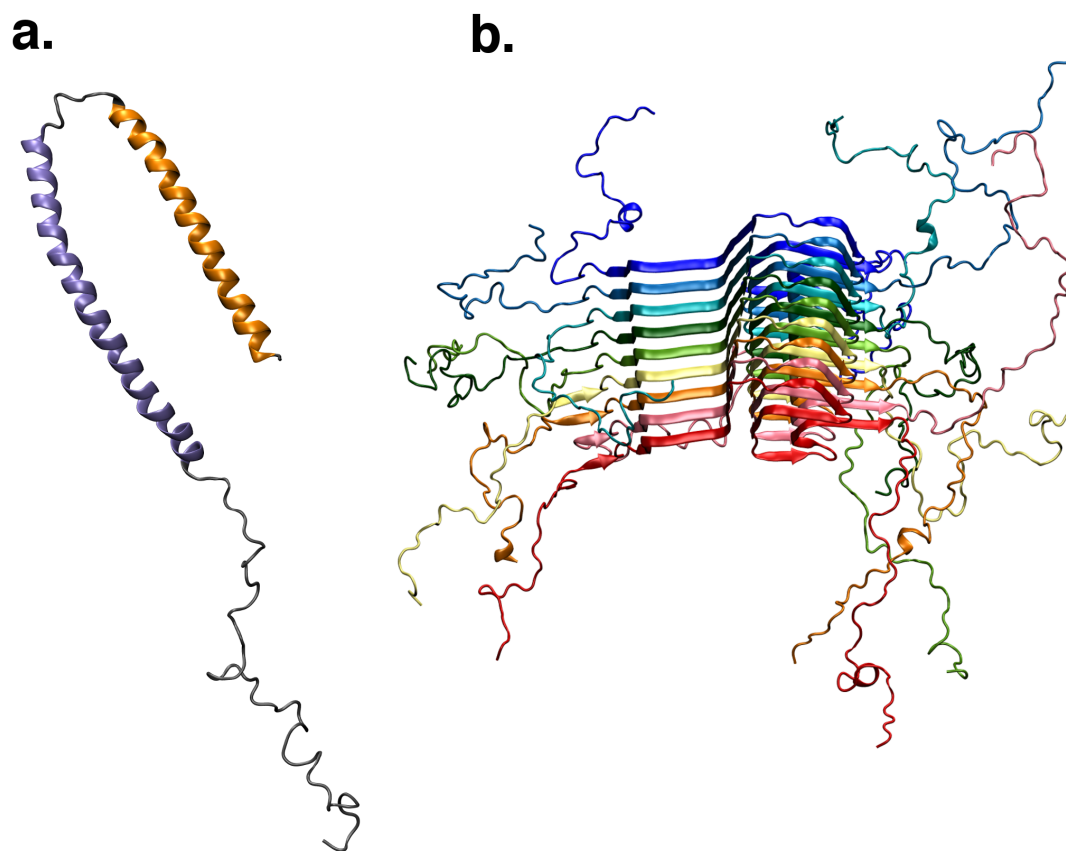
Figure 1.2. (a) Solution nuclear magnetic resonance (NMR) structure of the micelle bound human α-synuclein[71] (PDB ID: 1XQ8) (b) Solid-state NMR structure of α-synuclein amyloid fibril[72] (PDB ID: 2N0A).

The core of amyloid fibrils is composed of β-strands running perpendicular to the long fibril axis giving rise to a cross-β pattern of two diffuse reflections; one at

~4.8 Å along the vertical axis corresponding to the fibril direction arises from the stacked extended β-strands and the second diffuse reflection at a spacing of ~10 Å along the horizontal axis arises from the separation of the adjacent β-sheets[67-70]. Several biophysical methods such as cryo-electron microscopy[73-74], X-ray fiber diffraction[75-80], site-directed spin labeling[81], atomic force microscopy[82], solid-state nuclear magnetic resonance (ssNMR)[72,83-84] have been progressively employed to elucidate the atomic-level structural details of the amyloid fibril. An atomic model of $A\beta_{42}$ amyloid fibril[74] with parallel cross-β structure is shown in Figure 1.3. A pair of β-sheets forms the repeating unit of all amyloid protofilaments, running parallel to the fibril with their β-strands perpendicular to the fibril axis[85]. Each sheet is a standard Pauling-Corey β-sheet model[86], in which each β-strand forms backbone hydrogen bonds with the strand above and below it; the β-strands can be either parallel[74-75,87] or antiparallel[88]. The β-sheets associate in pairs and the side chains emanating from the two sheets are tightly interdigitated[85]. The interface of this motif is devoid of water and has been termed as *"steric zipper"*[78,85]. Two or more protofilaments are tightly twisted around each other and held together by weak forces to form mature amyloid fibrils. They appear as long, unbranched filaments when observed by electron microscopy, typically 7–13 nm in diameter and often several microns in length[89-90]. These fibrillar structures display remarkable mechanical properties such as high tensile strength that can be largely attributed to the array of hydrogen bonds of the cross-β structure[91].
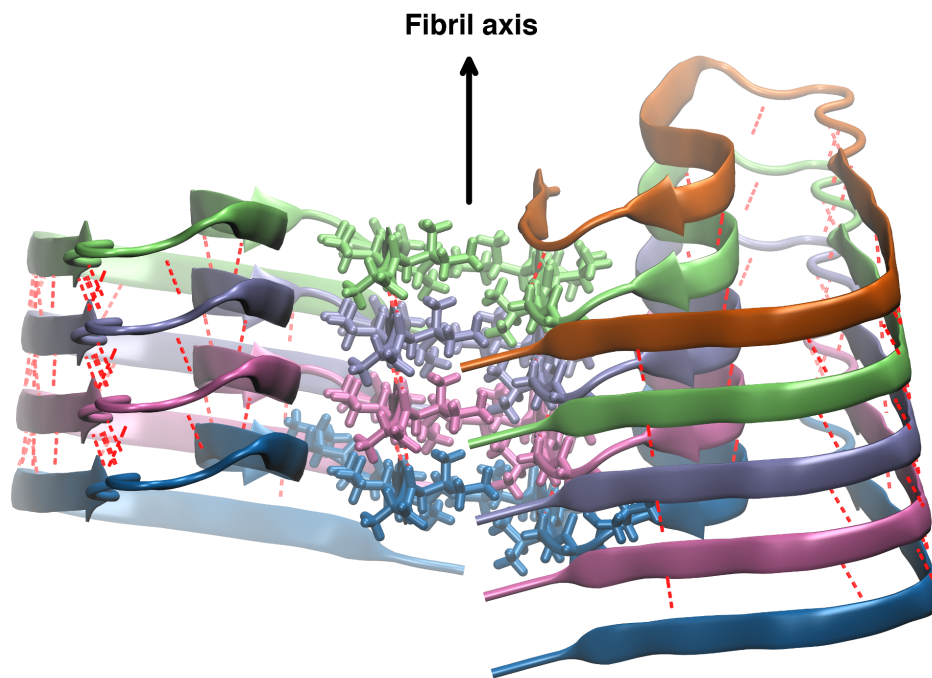
Figure 1.3. Atomic model of Aβ$_{42}$ fibril determined by cryo-electron microscopy[74] (PDB ID: 5OQV). Two intertwined protofilaments with Aβ$_{42}$ molecules are stacked in a parallel in-register manner. The interface consists of interdigitated residue side chains. Interstrand backbone hydrogen bonds in each sheet, shown as red broken lines are parallel to the fibril axis.

1.2.2   Models of Amyloid Formation

Understanding protein aggregation is one of the grand challenges in biophysics with its relevance spanning pathological events to biotechnological and pharmaceutical applications. It is a complex process that has generally been described within the thermo-kinetic framework of nucleated polymerization processes[92-96]. This involves an initially unfavorable process, wherein primarily enthalpic factors cause monomers to coalesce into a critical nucleus, the pathway element with the net highest free energy. The size and structure of the critical nucleus has therefore drawn substantial

attention, particularly for key amyloidogenic proteins[97-101]. Addition of monomers to the critical nucleus results in energetically downhill growth into amyloid protofibrils and then to fibrils accompanied by a disorder to order transformation[102-103]. Indeed, *in vitro* amyloidogenesis depicts many characteristics of the nucleated polymerization process, such as the requirement of a threshold (critical) concentration for amyloid formation, a lag time prior to fibrillogenesis, and a concentration dependence of the rate of fibril formation[104]. Also, monomers and fibrillar species should be detectable during the entire aggregation time course while the concentrations of the nucleus should be very low at all times. However, this model is not consistent with many experimental observations of early non-fibrillar oligomers without detectable amyloid fibril formation[105-107]. This indicates that several proteins exhibit amyloid formation by processes more complex than nucleation-polymerization[105-108]. Another competing hypothesis accounting for the sigmoidal protein aggregation kinetics is the nucleated conformational conversion model. This model suggests that oligomeric intermediates are formed at high monomer concentrations and undergo a slow conformational transition to form protofibrils and eventually fibrils[107-109]. The nature of structural transition as well as the role of monomeric forms in fibril elongation are however elusive. Protein amyloidosis is further complicated by more intricate reactions of aggregation and fragmentation that determine whether fibrils undergo elongation or fragmentation into smaller aggregates[95,110-111]. It is thus a highly complex process that is sensitive to the reaction conditions and accumulates a plethora of soluble and insoluble intermediates over a multitude of kinetically competing pathways operating in a huge range of timescales. Importantly, the observation of self-assembly and fibrillar growth in soluble and functional proteins (such as insulin) has supported the notion that a protein's functionally folded state is merely metastable, and kinetically

separated from deeper conformational minima corresponding to amyloids and their intermediates in an extended free energy landscape[17,56,112-114] depicted in Figure 1.4. In the process of conversion from the unfolded state to the native folded conformation, protein folding and aggregation are competing reactions[56,115]. The downhill progression of folding towards the thermodynamically favorable native state is driven by "*intramolecular*" interactions. During this process, proteins may adopt kinetically trapped, folded or partially folded species. These kinetically trapped intermediates are prone to establish "*intermolecular*" interactions, thereby leading to the assembly of various aggregated species such as amorphous aggregates, β-sheet rich oligomers, protofibrils and amyloid fibrils[17]. The landscape in the aggregation regime is more rugged relative to folding, characterized by deeper kinetic traps, with amyloid fibrils postulated as the thermodynamic ground state[112-114,116]. The emergence of protein self-assembly and amyloid formation as crucial triggers for a range of proteopathies has encouraged extensive efforts into extricating the detailed molecular mechanisms underlying these processes[95,97-98,101,111,117-122].
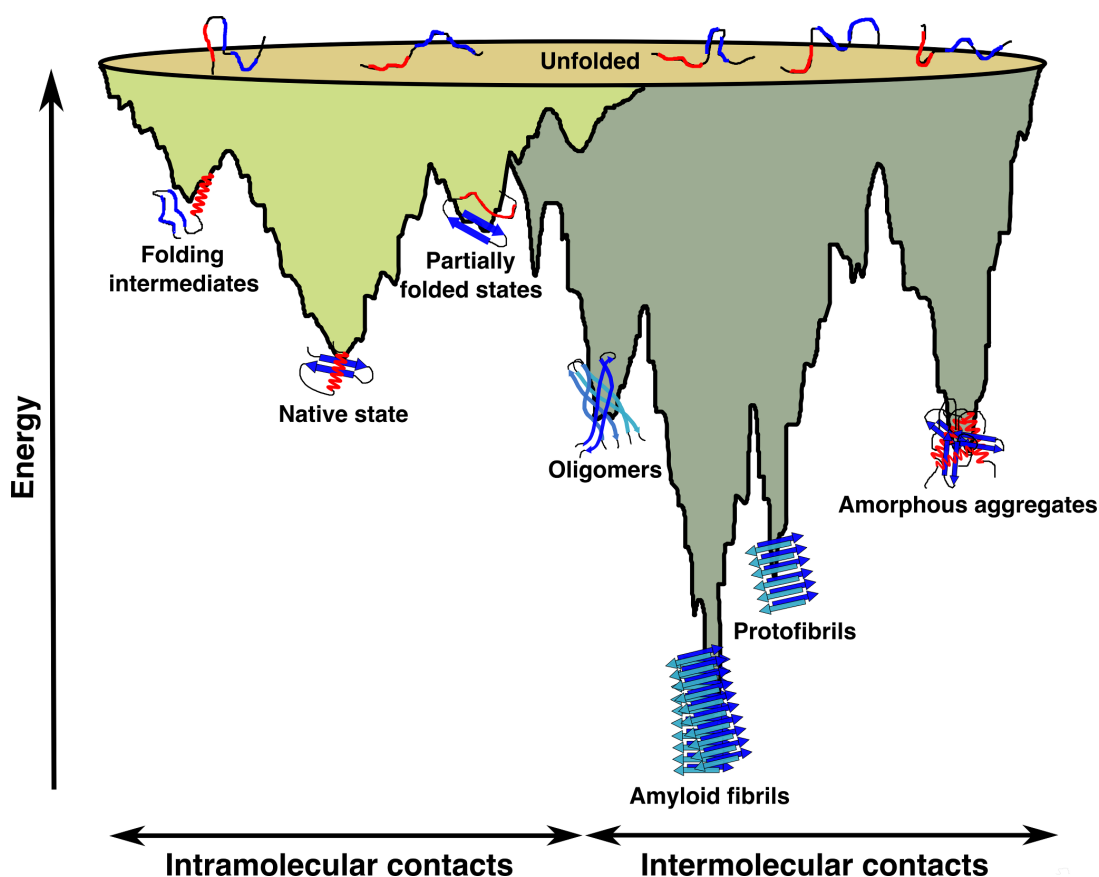
Figure 1.4. Schematic energy landscape for protein folding and aggregation.

1.3 Intrinsically Disordered Proteins

Over the past century, functional integrity of proteins has been intrinsically associated with their ability to fold rapidly into unique, three-dimensional structures. Interestingly, newer perspectives developed over the last few decades establish the existence of a class of polypeptides known as "*intrinsically disordered proteins*" (IDPs) whose biological roles are fundamentally associated to their ability to adopt multiple conformations at physiological conditions[123-125]. In addition, many proteins also have intrinsically disordered segments or regions (IDRs) that are key to their functions. These IDPs or IDRs are highly abundant in nature and involved in a plethora of biological activities. The inherent structural plasticity of IDPs confers

them with unique functional modalities in cellular processes such as cell-cycle regulation, gene expression, protein-protein interactions, etc[126]. Several evidences have recognized structural disorder in proteins that act as chaperones[127-129]. A well-characterized example of such a protein is the heat shock protein (Hsp)-33 that functions as a chaperone under oxidative stress[127]. The transitions of this protein between ordered and disordered conformations are key to its functioning as a chaperone. The promiscuous binding nature of IDPs makes them well suited as hubs in protein interaction networks, especially those involved in cell signaling, as exemplified by the tumor suppressor protein, p53[130]. On the flip side, these characteristics also render IDPs more prone to aggregation and associated diseases[63]. The formation of insoluble and intractable aggregates of IDPs highlights their implications in human diseases (listed in Table 1.1). Examples of such IDPs are amyloid beta (Aβ) and α-synuclein involved in neurodegenerative diseases[62] (Alzheimer's and Parkinson's disease, respectively), p53 and HPV associated with cancer[131] etc.

Owing to their structural heterogeneity, dynamic nature and subtle dependence on solvent and thermodynamic conditions, detailed structural characterization and probing the mechanistic origins of IDP function can pose manifold greater challenges than corresponding studies of well-defined folded proteins. In recent years, however, large strides have been made in biophysical experimental techniques for probing IDP structure and function, most notably in solution[132-133] and solid-state NMR[134], small-angle X-ray scattering[135-136], fluorescence microscopy[137-138], cryo-EM[139], fluorescence correlation spectroscopy[140-141], and some single-molecule techniques[142-143]. It is important to note, however, that experiments are essentially a 'top down' approach to understanding IDP behaviour, and often provide only limited insights on the specific

nature of interactions and the resultant physical forces underlying their observed thermodynamic, kinetic, structural and thereby functional behavior. In this regard, it has been realized that theoretical models and advanced molecular computations can provide an alternative 'bottoms up' view, and thereby can provide a powerful repertoire of complementary methods to probe the physical underpinnings of IDP structure and function[144-146].

Conformational energy landscapes of IDPs describing their thermodynamic free energy as a function of one or more collective variables are typically more rugged in comparison to the 'funnel shaped' landscapes adopted by the landscapes of folded proteins, illustrated in Figure 1.4. Thus, the conformational ensembles of IDPs are characterized by several thermodynamically equivalent low energy states. These states are often separated by relatively small energy barriers and therefore the landscapes have sometimes been referred to as 'glassy'[147]. Infrequently, however, thermodynamically equivalent states may be separated by barriers that exceed thermal levels by more than an order of magnitude; such situations may 'trap' specific states, and may be trigger the onset of systemic malfunction or disease in biological organisms[17,56]. Specially designed computational studies are often necessary to understand the thermodynamic and kinetic origins of such traps and to strategize plausible physico-chemical means to lower the barriers and repopulate the biologically advantageous conformations[148-150].

Disease	Type of disease	Associated IDP	[a]Number of residues	[b]Structural class of the native protein or peptide
Alzheimer's disease	Neurodegenerative	Amyloid-β peptide	40 or 42	Natively unfolded
Parkinson's disease	Neurodegenerative	α-synuclein	140	Natively unfolded
Spongiform encephalopathies	Neurodegenerative	Prion protein	253	Natively unstructured (res 1-120); α -helical (res 121-230)
Huntington's disease	Neurodegenerative	Huntingtin with polyglutamine expansion	3144	Largely natively unfolded
Amyotrophic lateral sclerosis	Neurodegenerative	Superoxide dismutase 1	153	All-β, Ig like
Type-2 Diabetes	Non-neuropathic localized	Amylin	37	Natively unfolded
Insulin amyloidoses	Non-neuropathic localized	Insulin	Two chains of lengths 21 and 30 residues	All-α, insulin like
Cataract	Non-neuropathic localized	γ-crystallins	Variable	All-α, γ-crystallin like

Disease	Type of disease	Associated IDP	[a]Number of residues	[b]Structural class of the native protein or peptide
Haemodialysis-related amyloidosis	Non-neuropathic systemic	β_2-microglobulin	99	All-β, Ig like
Senile systemic amyloidosis	Non-neuropathic systemic	Wild-type transthyretin	127	All-β, prealbumin like

[a]The number of residues correspond to that of the processed polypeptide chains that aggregate and not the precursor protein.

[b]Structural class refers to the fold of the processed protein or peptides prior to aggregation according to the Structural Classification Of Proteins (SCOP) database.

Table 1.1. Some amyloid-related human diseases and the associated protein/peptide in the amyloid state

1.4 Factors Affecting Protein Stability and Self-Assembly

Protein stability is attributed to the net balance of forces that determines the state it attains; natively folded or unfolded. It can be quantified using the difference in Gibbs free energy between the unfolded (U) and folded (N) conformation as,

$$\Delta G_U = G_U - G_N$$

$$= \Delta H - T\Delta S \qquad (1.1)$$

where, ΔH denotes the change in enthalpy, T is the temperature and ΔS is the entropy change or extent of disorder between the two states. The overall free energy change is determined by the effects arising from the exposure of the internal non-polar and polar groups, their solvent interaction as well as the consequential changes in the

solvent-solvent interactions[49]. A negative value of ΔG indicates that the unfolded conformation is favored over the folded conformation. ΔG is positive when the folded state has a lower free energy and hence the folding reaction is favored. An unfolded protein has high entropy and also high enthalpy owing to few stabilizing interactions. The folded state is stabilized by non-bonded interactions such as hydrophobic, electrostatic, hydrogen bonds and van der Waals interactions as well as disulfide bonds and hence has low enthalpy and entropy. The temperature dependent enthalpic and entropic quantities compensate each other to such a large extent that the folded states are only marginally stable relative to the unfolded state[151-156].

The stability of the native state of proteins is a function of the primary amino acid sequence as well as the surrounding environmental variables such as temperature, pressure, pH, ionic strength and solvent composition. Variations in these extrinsic factors from ambient conditions manifest as perturbations in the molecular determinants of protein structural stability and functional integrity. Mutations in protein structure, as small as a point mutation or change of a single amino acid residue may result in change or loss of the native structure, which in turn can alter protein stability and function[157-158]. Destabilization of the native state leads to an increased population of partially folded intermediates, susceptible to amyloid formation[16-17,159-161]. External factors can also modulate amyloid fibrillation kinetics and morphology of the intermediates and the final aggregated species[162-169]. Moreover, biological milieu in which proteins fold and function is a crowded environment that may cause noticeable alterations to the protein structure, dynamics, and folding[170-179]. Therefore, it is increasingly recognized that estimations of protein stability and self-assembly should be studied with careful considerations of the cell

interior[179-187]. This section is intended to provide a brief overview of some of the factors that affect protein stability and self-assembly processes.

1.4.1   Temperature

Proteins *in vivo* adapt to function efficiently within the normal temperature range, which defines cell survival. This generally implies that proteins have a limiting temperature range in which their structural integrity is maintained and undergo denaturation beyond this range[188]. On increasing the temperature above the native temperature of the organism, the enhanced thermal energy can disrupt the intermolecular interactions leading to unfolding or denaturation and have a deleterious effect due to loss of function[189]. However, proteins are also known to display decreased stability at low temperatures, a phenomenon called "*cold denaturation*"[190-194]. Discerning the thermal stability of proteins is crucial to elucidate the adaptability of organisms to extreme environments[195-197]. It also has practical applications in the context of cryopreservation and optimization of protein-based biotechnological and biopharmaceutical processes[198-199].

For small globular proteins, thermal denaturation can be approximated as a cooperative transition between two macroscopic states, native and denatured[200]. The free-energy of thermal unfolding of small proteins obtained from microcalorimetry experiments exhibit an inverse bell-shaped profile and crosses the zero point of free energy difference (ΔG=0) at two points corresponding to the heat and cold denaturation temperatures[190,201]. This indicates that as temperature is increased or decreased from the temperature at which the native state resides, the population of folded conformations gradually decreases with higher population of unfolded species beyond the transition temperatures. The enthalpic and entropic contributions are

strongly temperature dependent and hence have important consequences in aspects of protein stability and self-assembly phenomena[202-203]. Heat denaturation causes an increase in enthalpic and entropic contributions while cold denaturation causes a decrease in both[190,192,204-206]. Furthermore, the change in specific heat upon order-disorder transition of proteins, $\Delta C_p = C_{pD} - C_{pN}$ (where, C_p is the heat capacity at constant pressure; D and N are the denatured and native states, respectively) is thus generally assumed to be positive at the transition or melting temperatures. Therefore, cold and heat denaturation may be driven by different mechanisms that access non-overlapping ensembles[207-208].

Cold unfolding of most proteins occurs below the freezing point of water that precludes experimental observation of this phenomenon due to freezing of the solvent. The most common approach to circumvent the inaccessibility of experimental techniques to extremely low temperatures is to raise the cold-melting temperature of the protein by weakening the folded protein through pressure, pH, chemical denaturants, *ad hoc* mutations[190,201,209-212] or supercooling the sample solutions[207,213-214]. Numerous theoretical and simulation studies have been performed to understand the mechanistic elements of cold unfolding of proteins[49,208,215-218]. The phenomenon of cold denaturation is largely attributed to the weakening of the hydrophobic effect and reduced free energy penalty of the entropically unfavorable interaction between the exposed hydrophobic residues and surrounding water at low temperature[49,207-208,215-216,218-219]. There is substantial ordering of water molecules around the nonpolar residues, characterized by low entropy and low enthalpy owing to the enhanced network of hydrogen bonds of the hydration shell waters. The solvent-induced packing defects at the protein surface and favorable water-protein interactions leads to destabilization of the native protein structure upon cooling[208]. In contrast to globular

proteins, cold thermal response of protein amyloid assemblies is less explored[214,220-221] and lacks general consensus on the detailed molecular mechanism.

1.4.2 Pressure

Understanding the effects of pressure on protein structure and stability is important in the context of extremophiles that have evolved to live under high hydrostatic pressure[222-223]. Increasing pressure, similar to thermal effects, disrupts the thermodynamic equilibrium between the folded native state and the unfolded denatured state[224]. For a two state equilibrium between the native and denatured states in response to thermodynamic variations, the van't Hoff equation can be used to derive the pressure-temperature phase diagram for protein stability[225].

The volume change (ΔV_u) upon protein unfolding under high-pressure conditions is the primary factor determining the effect of hydrostatic pressure on protein stability[224,226]. Le Chatelier's principle states that the system equilibrium shifts toward the states that occupy the least volume in order to minimize the external perturbation of high-pressure conditions[226]. Hence, if ΔV_u is negative, it will lead to decrease in stability since the equilibrium is shifted towards the denatured states while stability increases if ΔV_u is positive[224]. Experimental and simulation studies probing the effects of hydrostatic pressure on protein stability have shown that most proteins denature or unfold upon increase in pressure[227-230]. The volume changes calculated from these studies however are very small, ranging from –4 % to +0.5 %, relative to the total volume of the protein[231-233]. Native folded proteins contain small packing defects (voids and cavities) of varying shapes and sizes and are often devoid of solvent[226,234]. These voids and cavities occupy 20–30 % of the total volume fraction of the native protein[224,234-235]. Using a probe size of 0.04 Å, Chen and Makhatadze

calculated the decrease in void volume between the folded and unfolded states of proteins and found this quantity to be quite large[224]. They asserted that the large decrease in volume is countered by the increased hydration volume consequently leading to only ~7 % decrease in the total molar volume upon unfolding[224]. Studies have shown that change in the void volume of a protein greatly influences ΔV_u value[236-237] while change in the size of the protein, which reflects in its solvent accessibility has no effect[226,238]. Thus, while theoretical studies report density differences between bulk water and hydration waters[239-241], these differences do not contribute significantly to the pressure denaturation of proteins[226].

1.4.3 Macromolecular Crowding

The cellular interior is an extremely crowded milieu where macromolecular concentration can be in the range of 80–400 g/L occupying 5% to 40% of the total volume of the cell[242-243]. This crowded environment differs dramatically from the dilute, idealized conditions usually used in most biophysical studies. It is increasingly clear that it is crucial to consider protein folding and aggregation from the standpoint of the cell since nascent polypeptides are synthesized and released in an environment crowded with biomolecules. Macromolecular crowding, often termed as the "*excluded volume effect*", is known to have significant effects on the stability of biological macromolecules, equilibria of macromolecular associations as well as the rates and extent of supramolecular amyloid assemblies[171,244-245]. The macromolecular crowding effect arises from two phenomena, hare-core repulsions and nonspecific chemical (soft) interactions[246-247]. The hard-core repulsions can be attributed to the excluded volume effect of the crowded medium and are inherently entropic[248]. Most studies on protein folding and stability have mainly focused on steric excluded volume effects of

crowding and found to reduce configurational entropy of extended states and stabilize compact globular conformations[245,249-250]. Soft or chemical interactions can be attractive or repulsive in nature having destabilizing or stabilizing effects, respectively[247].

The complexity of the aggregation process and the partitioning between different pathways makes it difficult to predict the potential altering effects of crowding on different steps and species along the pathways. Several studies designed to probe this aspect indicate that macromolecular crowding may have variable effects on different aggregation pathways; it may be favorable for some aggregating pathways and unfavorable for some others[172,179,181,183,185-187,251-253]. A general physical picture (Figure 1.5) of the complex process of protein aggregation and amyloid formation in crowded environments can be drawn based on several experimental and theoretical studies[244,254]. The acceleratory or hindering effect of crowding on the complex aggregation process can be ascribed to two key competing effects. First, the excluded volume effect in crowded environments promotes protein self-association, leading to increased thermodynamic favorability towards the formation of more compact intermediates, thereby reducing conformations associated with energetic penalties[172,179,181,183]. In contrast, crowding has limiting effects on protein diffusion, and leads to decreased favorability of oligomerization, and thereby to reduced propensities of fibrillation[179,185,187]. The relative extent of these competing effects depends on the nature and concentration of the crowding agent, as well as the nature of the protein and its amyloidogenic propensity[179,181,183,186-187].
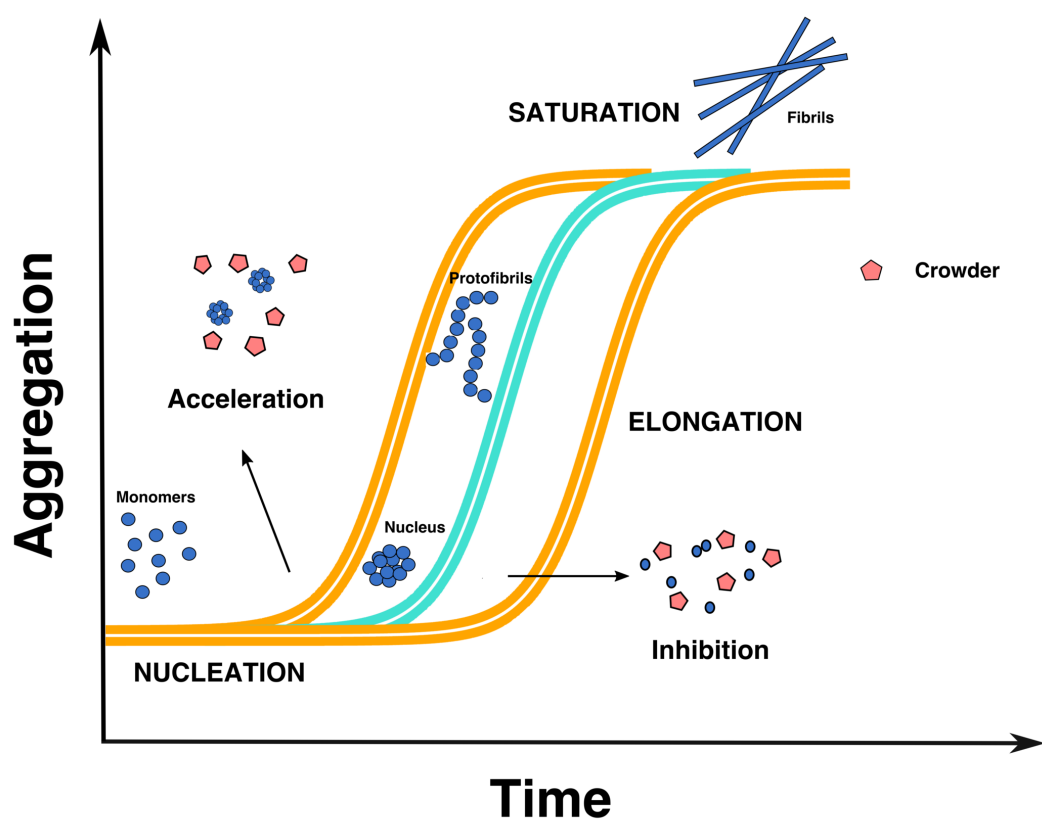
Figure 1.5. Schematic of the modulatory effects of macromolecular crowding on amyloid fibril formation process. The kinetics of fibril formation is shown as a sigmoidal trace that can be attributed to three phases, namely nucleation, elongation and saturation. Amyloid self-assembly can be accelerated by stabilization of the nuclei (shortening of lag phase) or inhibited due to hampered peptide diffusion in crowded environments.

1.4.4   Mutations

Native globular proteins are only marginally stable under physiological conditions, with an overall thermodynamic stability ($\Delta G_{folding}$) in the range of –5 to –15 kcal mol^{-1} [151-156]. The chemical interactions stabilizing the native conformation include disulfide bonds and weak noncovalent interactions like hydrogen bonds, hydrophobic and ionic interactions[49]. Disruption of a single hydrogen bond requires only 2–5 kcal mol^{-1}

implying that even a single amino acid substitution could dramatically alter the stability of a protein and perturb protein-protein interactions due to change in the protein interaction interface. Mutations in proteins occur via molecular mechanisms primarily at the genetic level introducing variability in the population. This variability is important for biodiversity as evolutionary selection pressure retains functionally advantageous mutations[255]. Studies investigating the folding free energies of mutations in proteins have shown that amino acid mutations could be both stabilizing[256-257] and deleterious[258-261], depending upon the protein and nature of the substitutions. For example, most of the familial mutations linked with amyloid diseases lead to destabilization of the native proteins and enhanced propensity to aggregate[16,262-264]. It is crucial to have a comprehensive understanding of how familial and sporadic disease mutations manifest in structural and dynamical changes of the disease-associated protein. These insights can be further leveraged in developing molecular diagnostic strategies to effectively combat diseases.

Systematically mapping the effect of mutations on the kinetics of protein folding and unfolding is a seminal technique used to probe the role played by specific residues in the folding process[10,265-269]. Mutations are strategically engineered into the protein to disrupt stabilizing interactions in the folded structure. These mutations act as specific probes for structural changes during unfolding and kinetic experiments detect transiently formed intermediates during the folding pathway[10,265,270]. Mutagenesis strategies are also utilized in protein engineering to design proteins with enhanced stability, activity and specificity[271-276], predicting evolutionary dynamics of proteins[277] and *de novo* protein design for therapeutic strategies[278-281].

1.5   Thesis Organization

As discussed thus far, proteins fold into unique three-dimensional structures that are intricately associated with their biological functions. Failure to attain the correct fold or loss of the native functional states are known to cause malfunctioning leading to pathological conditions. Misfolded proteins have an enhanced propensity to aggregate into higher order amyloid fibrils implicated in debilitating diseases. It is postulated that proteins, in their physiologically relevant conformations are only metastable with respect to amyloid fibrils.

The stability of the functionally competent states of proteins is crucially linked not only to the intermolecular interactions in the three-dimensional structure but also to the environmental variables such as pH, temperature, pressure, solvent composition etc. Perturbations of the molecular forces in protein structure or deviations in the environmental factors from ambient conditions can have a significant impact on the overall stability of globular proteins. Molecular level understanding of the underlying mechanisms of these perturbations can provide comprehensive insights into the relative stabilities of protein conformational states and how they are modulated by sequence and conditions. These can be effectively harnessed in designing enzymes for industrial usage as well as developing therapeutics. In this context, molecular dynamics (MD) simulations provide atomic-level characterization of the structural and energetics of biomolecular systems like proteins.

The studies encompassing the present thesis have deployed fully atomistic MD simulations and statistical analyses methods to study the potential effects of structural perturbation as well as external factors such as temperature and molecular crowding on the structural integrity and dynamics of proteins. The proteins subjected to these studies are representative IDPs, namely the prion protein, the full-length Aβ peptide,

and an amyloid oligomer of αB-crystallin (ABC). The diverse ensemble of conformational substates that describes the native state of IDPs poses unique challenges in experimentally characterizing the structural features of IDPs. MD simulations are widely used to elucidate the heterogeneous ensembles of IDPs.

The thesis is organized into six chapters as follows:

Chapter 1, the present chapter provides an overview of the protein-folding problem with a brief discussion of the energy landscape theory. Protein misfolding and the consequential aggregation or amyloid formation, its structure and existing models of amyloid formation is discussed. A class of proteins called intrinsically disordered proteins is highlighted and a non-exhaustive list of the associated diseases is provided. Some key factors affecting protein stability and self-assembly process, namely temperature, macromolecular crowding and mutations, explored in this thesis work are discussed.

Chapter 2 provides a theoretical overview of MD simulations method harnessed in this thesis work.

Chapter 3 discusses a molecular investigation of the cold thermal response of an amyloid oligomer. This chapter provides insights into the divergent response exhibited by the amyloid assembly from that of typical globular proteins. It highlights the microscopic influence of the hydration layer on the low-temperature response.

Chapter 4 describes the effect of macromolecular crowding on the early self-assembly or dimerization of the full-length Aβ peptide, implicated in Alzheimer's disease. This chapter provides a comparison of the binding strength and solvation patterns of the assembled dimers in the two environments.

Chapter 5 describes the effects of a minor perturbation of a stabilizing non-bonded interaction of the native prion protein. The pathological relevance of structural and dynamical transitions induced by the perturbation is discussed.

Chapter 6 summarizes the work presented and discusses future directions.

1.6   References

1.   Anfinsen, C. B.; Haber, E.; Sela, M.; White, F. H., Jr. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* **1961,** *47* (9), 1309–1314.

2.   Anfinsen, C. B. The formation and stabilization of protein structure. *Biochem. J.* **1972,** *128* (4), 737–749.

3.   Anfinsen, C. B. Principles that Govern the Folding of Protein Chains. *Science* **1973,** *181* (4096), 223–230.

4.   Levinthal, C. Are there pathways for protein folding? *J. Chim. Phys.* **1968,** *65*, 44–45.

5.   Dill, K. A.; Chan, H. S. From Levinthal to pathways to funnels. *Nat. Struct. Biol.* **1997,** *4* (1), 10–19.

6.   Dobson, C. M.; Šali, A.; Karplus, M. Protein Folding: A Perspective from Theory and Experiment. *Angew. Chem., Int. Ed.* **1998,** *37* (7), 868–893.

7.   Ikai, A.; Tanford, C. Kinetic Evidence for Incorrectly Folded Intermediate States in the Refolding of Denatured Proteins. *Nature* **1971,** *230* (5289), 100–102.

8.   Schmid, F. X. Mechanism of folding of ribonuclease A. Slow refolding is a sequential reaction via structural intermediates. *Biochemistry* **1983,** *22* (20), 4690–4696.

9.   Kim, P. S.; Baldwin, R. L. Specific Intermediates in the Folding Reactions of Small Proteins and the Mechanism of Protein Folding. *Annu. Rev. Biochem.* **1982,** *51* (1), 459–489.

10.   Matouschek, A.; Kellis, J. T.; Serrano, L.; Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **1989,** *340* (6229), 122–126.

11.   Harrison, S. C.; Durbin, R. Is there a single pathway for the folding of a polypeptide chain? *Proc. Natl. Acad. Sci. U.S.A.* **1985,** *82* (12), 4028.

12. Baldwin, R. L. The nature of protein folding pathways: The classical versus the new view. *J. Biomol. NMR* **1995,** *5* (2), 103–109.

13. Baldwin, R. L. Matching speed and stability. *Nature* **1994,** *369* (6477), 183–184.

14. Onuchic, J. N.; Luthey-Schulten, Z.; Wolynes, P. G. Theory of protein folding: The Energy Landscape Perspective. *Annu. Rev. Phys. Chem.* **1997,** *48* (1), 545–600.

15. Onuchic, J. N.; Wolynes, P. G. Theory of protein folding. *Curr. Opin. Struct. Biol.* **2004,** *14* (1), 70–75.

16. Dobson, C. M. Protein folding and misfolding. *Nature* **2003,** *426* (6968), 884–890.

17. Jahn, T. R.; Radford, S. E. The Yin and Yang of protein folding. *FEBS J.* **2005,** *272* (23), 5962–5970.

18. Royer, C. A. Probing Protein Folding and Conformational Transitions with Fluorescence. *Chem. Rev.* **2006,** *106* (5), 1769–1784.

19. Kelly, S. M.; Jess, T. J.; Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta, Proteins Proteomics* **2005,** *1751* (2), 119–139.

20. Balakrishnan, G.; Weeks, C. L.; Ibrahim, M.; Soldatova, A. V.; Spiro, T. G. Protein dynamics from time resolved UV Raman spectroscopy. *Curr. Opin. Struct. Biol.* **2008,** *18* (5), 623–629.

21. Fabian, H.; Naumann, D. Methods to study protein folding by stopped-flow FT-IR. *Methods* **2004,** *34* (1), 28–40.

22. Haustein, E.; Schwille, P. Fluorescence Correlation Spectroscopy: Novel Variations of an Established Technique. *Annu. Rev. Biophys. Biomol. Struct.* **2007,** *36* (1), 151–169.

23. Schuler, B.; Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **2008,** *18* (1), 16–26.

24. Dyson, H. J.; Wright, P. E. Elucidation of the Protein Folding Landscape by NMR. *Meth. Enzymol.* **2005,** *394*, 299–321.

25. Jensen, M. R.; Zweckstetter, M.; Huang, J.-r.; Blackledge, M. Exploring Free-Energy Landscapes of Intrinsically Disordered Proteins at Atomic Resolution Using NMR Spectroscopy. *Chem. Rev.* **2014,** *114* (13), 6632–6660.

26. Englander, S. W. Protein Folding Intermediates and Pathways Studied by Hydrogen Exchange. *Annu. Rev. Biophys. Biomol. Struct.* **2000,** *29* (1), 213–238.

27. Sakakibara, D.; Sasaki, A.; Ikeya, T.; Hamatsu, J.; Hanashima, T.; Mishima, M.; Yoshimasu, M.; Hayashi, N.; Mikawa, T.; Wälchli, M.; Smith, B. O.; Shirakawa, M.; Güntert, P.; Ito, Y. Protein structure determination in living cells by in-cell NMR spectroscopy. *Nature* **2009,** *458*, 102–105.

28. Inomata, K.; Ohno, A.; Tochio, H.; Isogai, S.; Tenno, T.; Nakase, I.; Takeuchi, T.; Futaki, S.; Ito, Y.; Hiroaki, H.; Shirakawa, M. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. *Nature* **2009,** *458*, 106–109.

29. Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002,** *9* (9), 646–652.

30. Hubner, I. A.; Deeds, E. J.; Shakhnovich, E. I. Understanding ensemble protein folding at atomic detail. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103* (47), 17747–17752.

31. Shaw, D. E.; Dror, R. O.; Salmon, J. K.; Grossman, J. P.; Mackenzie, K. M.; Bank, J. A.; Young, C.; Deneroff, M. M.; Batson, B.; Bowers, K. J.; Chow, E.; Eastwood, M. P.; Ierardi, D. J.; Klepeis, J. L.; Kuskin, J. S.; Larson, R. H.; Lindorff-Larsen, K.; Maragakis, P.; Moraes, M. A.; Piana, S.; Shan, Y.; Towles, B.; Millisecond-scale molecular dynamics simulations on Anton. In *Proceedings of the ACM/IEEE Conference on Supercomputing* (SC09) (ACM Press, New York, 2009)

32. Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010,** *330* (6002), 341–346.

33. Bartlett, A. I.; Radford, S. E. An expanding arsenal of experimental methods yields an explosion of insights into protein folding mechanisms. *Nat. Struct. Mol. Biol.* **2009,** *16*, 582–588.

34. Li, A.; Daggett, V. Identification and Characterization of the Unfolding Transition State of Chymotrypsin Inhibitor 2 by Molecular Dynamics Simulations. *J. Mol. Biol.* **1996,** *257* (2), 412–429.

35. Li, A.; Daggett, V. Molecular dynamics simulation of the unfolding of barnase: characterization of the major intermediate. *J. Mol. Biol.* **1998,** *275* (4), 677–694.

36. Li, L.; Shakhnovich, E. I. Constructing, verifying, and dissecting the folding transition state of chymotrypsin inhibitor 2 with all-atom simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98* (23), 13014.

37. Fersht, A. R. Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2000,** *97* (4), 1525–1529.

38. Sánchez, I. E.; Kiefhaber, T. Evidence for Sequential Barriers and Obligatory Intermediates in Apparent Two-state Protein Folding. *J. Mol. Biol.* **2003,** *325* (2), 367–376.

39. Brockwell, D. J.; Radford, S. E. Intermediates: ubiquitous species on folding energy landscapes? *Curr. Opin. Struct. Biol.* **2007,** *17* (1), 30–37.

40. Chavez, L. L.; Onuchic, J. N.; Clementi, C. Quantifying the Roughness on the Free Energy Landscape: Entropic Bottlenecks and Protein Folding Rates. *J. Am. Chem. Soc.* **2004,** *126* (27), 8426–8432.

41. Chavez, L. L.; Gosavi, S.; Jennings, P. A.; Onuchic, J. N. Multiple routes lead to the native state in the energy landscape of the β-trefoil family. *Proc. Natl. Acad. Sci. U.S.A.* **2006,** *103* (27), 10254–10258.

42. Do, T. D.; LaPointe, N. E.; Nelson, R.; Krotee, P.; Hayden, E. Y.; Ulrich, B.; Quan, S.; Feinstein, S. C.; Teplow, D. B.; Eisenberg, D.; Shea, J.-E.; Bowers, M. T. Amyloid β-Protein C-Terminal Fragments: Formation of Cylindrins and β-Barrels. *J. Am. Chem. Soc.* **2016,** *138* (2), 549–557.

43. Roder, H.; Colón, W. Kinetic role of early intermediates in protein folding. *Curr. Opin. Struct. Biol.* **1997,** *7* (1), 15–28.

44. Friel, C. T.; Alastair Smith, D.; Vendruscolo, M.; Gsponer, J.; Radford, S. E. The mechanism of folding of Im7 reveals competition between functional and kinetic evolutionary constraints. *Nat. Struct. Mol. Biol.* **2009,** *16*, 318–324.

45. Korzhnev, D. M.; Salvatella, X.; Vendruscolo, M.; Di Nardo, A. A.; Davidson, A. R.; Dobson, C. M.; Kay, L. E. Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR. *Nature* **2004,** *430* (6999), 586–590.

46. Várnai, P.; Dobson, C. M.; Vendruscolo, M. Determination of the Transition State Ensemble for the Folding of Ubiquitin from a Combination of Φ and Ψ Analyses. *J. Mol. Biol.* **2008,** *377* (2), 575–588.

47. Robson, B.; Pain, R. H. Analysis of the code relating sequence to conformation in proteins: Possible implications for the mechanism of formation of helical regions. *J. Mol. Biol.* **1971,** *58* (1), 237–257.

48. Dill, K. A. Theory for the folding and stability of globular proteins. *Biochemistry* **1985,** *24* (6), 1501–1509.

49. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990,** *29* (31), 7133–7155.

50. Udgaonkar, J. B.; Baldwin, R. L. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. *Nature* **1988,** *335* (6192), 694–699.

51. Jennings, P. A.; Wright, P. E. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science* **1993,** *262* (5135), 892–896.

52. Ptitsyn, O. B. Kinetic and equilibrium intermediates in protein folding. *Protein Eng. Des. Sel.* **1994,** *7* (5), 593–596.

53. Nelson, D. L.; Lehninger, A. L.; Cox, M. M. *Lehninger principles of biochemistry*. Macmillan: 2008.

54. Balchin, D.; Hayer-Hartl, M.; Hartl, F. U. In vivo aspects of protein folding and quality control. *Science* **2016,** *353* (6294), aac4354.

55. Klaips, C. L.; Jayaraj, G. G.; Hartl, F. U. Pathways of cellular proteostasis in aging and disease. *J. Cell Biol.* **2018,** *217* (1), 51–63.

56. Hartl, F. U.; Bracher, A.; Hayer-Hartl, M. Molecular chaperones in protein folding and proteostasis. *Nature* **2011,** *475*, 324–332.

57. Kim, Y. E.; Hipp, M. S.; Bracher, A.; Hayer-Hartl, M.; Ulrich Hartl, F. Molecular Chaperone Functions in Protein Folding and Proteostasis. *Annu. Rev. Biochem.* **2013,** *82* (1), 323–355.

58. Hipp, M. S.; Park, S.-H.; Hartl, F. U. Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* **2014,** *24* (9), 506–514.

59. Dobson, C. M.; Ellis, R. J.; Fersht, A. R.; Dobson Christopher, M. The structural basis of protein folding and its links with human disease. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **2001,** *356* (1406), 133–145.

60. Thomas, P. J.; Qu, B.-H.; Pedersen, P. L. Defective protein folding as a basis of human disease. *Trends Biochem. Sci.* **1995,** *20* (11), 456–459.

61. Sahni, N.; Yi, S.; Taipale, M.; Fuxman Bass, Juan I.; Coulombe-Huntington, J.; Yang, F.; Peng, J.; Weile, J.; Karras, Georgios I.; Wang, Y.; Kovács, István A.; Kamburov, A.; Krykbaeva, I.; Lam, Mandy H.; Tucker, G.; Khurana, V.; Sharma, A.; Liu, Y.-Y.; Yachie, N.; Zhong, Q.; Shen, Y.; Palagi, A.; San-Miguel, A.; Fan, C.; Balcha, D.; Dricot, A.; Jordan, Daniel M.; Walsh, Jennifer M.; Shah, Akash A.; Yang, X.; Stoyanova, Ani K.; Leighton, A.; Calderwood, Michael A.; Jacob, Y.; Cusick, Michael E.; Salehi-Ashtiani, K.; Whitesell, Luke J.; Sunyaev, S.; Berger, B.; Barabási, A.-L.; Charloteaux, B.; Hill, David E.; Hao, T.; Roth, Frederick P.; Xia, Y.; Walhout, Albertha J. M.; Lindquist, S.; Vidal, M. Widespread Macromolecular Interaction Perturbations in Human Genetic Disorders. *Cell* **2015,** *161* (3), 647–660.

62. Knowles, T. P. J.; Vendruscolo, M.; Dobson, C. M. The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* **2014,** *15*, 384–396.

63. Uversky, V. N.; Oldfield, C. J.; Dunker, A. K. Intrinsically Disordered Proteins in Human Diseases: Introducing the D2 Concept. *Annu. Rev. Biophys.* **2008,** *37* (1), 215–246.

64. Douglas, P. M.; Dillin, A. Protein homeostasis and aging in neurodegeneration. *J. Cell Biol.* **2010,** *190* (5), 719–729.

65. Labbadia, J.; Morimoto, R. I. The Biology of Proteostasis in Aging and Disease. *Annu. Rev. Biochem.* **2015,** *84* (1), 435–464.

66. Chiti, F.; Dobson, C. M. Protein Misfolding, Functional Amyloid, and Human Disease. *Annu. Rev. Biochem.* **2006,** *75* (1), 333–366.

67. Astbury, W. T.; Dickinson, S.; Bailey, K. The X-ray interpretation of denaturation and the structure of the seed globulins. *Biochem. J.* **1935,** *29* (10), 2351–2360.1.

68. Sunde, M.; Serpell, L. C.; Bartlam, M.; Fraser, P. E.; Pepys, M. B.; Blake, C. C. F. Common core structure of amyloid fibrils by synchrotron X-ray diffraction. *J. Mol. Biol.* **1997,** *273* (3), 729–739.

69. Sunde, M.; Blake, C. The Structure of Amyloid Fibrils by Electron Microscopy and X-Ray Diffraction. *Adv. Protein Chem.* **1997,** *50*, 123–159.

70. Sipe, J. D.; Cohen, A. S. Review: History of the Amyloid Fibril. *J. Struct. Biol.* **2000,** *130* (2), 88–98.

71. Ulmer, T. S.; Bax, A.; Cole, N. B.; Nussbaum, R. L. Structure and Dynamics of Micelle-bound Human α-Synuclein. *J. Biol. Chem.* **2005,** *280* (10), 9595–9603.

72. Tuttle, M. D.; Comellas, G.; Nieuwkoop, A. J.; Covell, D. J.; Berthold, D. A.; Kloepper, K. D.; Courtney, J. M.; Kim, J. K.; Barclay, A. M.; Kendall, A.; Wan, W.; Stubbs, G.; Schwieters, C. D.; Lee, V. M. Y.; George, J. M.; Rienstra, C. M. Solid-state NMR structure of a pathogenic fibril of full-length human α-synuclein. *Nat. Struct. Mol. Biol.* **2016,** *23*, 409–415.

73. Jiménez, J. L.; Guijarro, J. I.; Orlova, E.; Zurdo, J.; Dobson, C. M.; Sunde, M.; Saibil, H. R. Cryo-electron microscopy structure of an SH3 amyloid fibril and model of the molecular packing. *EMBO J.* **1999,** *18* (4), 815–821.

74. Gremer, L.; Schölzel, D.; Schenk, C.; Reinartz, E.; Labahn, J.; Ravelli, R. B. G.; Tusche, M.; Lopez-Iglesias, C.; Hoyer, W.; Heise, H.; Willbold, D.; Schröder, G. F. Fibril structure of amyloid-β(1–42) by cryo–electron microscopy. *Science* **2017,** *358* (6359), 116–119.

75.   Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. Ø.; Riekel, C.; Grothe, R.; Eisenberg, D. Structure of the cross-β spine of amyloid-like fibrils. *Nature* **2005,** *435* (7043), 773–778.

76.   Makin, O. S.; Atkins, E.; Sikorski, P.; Johansson, J.; Serpell, L. C. Molecular basis for amyloid fibril formation and stability. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (2), 315–320.

77.   Balbirnie, M.; Grothe, R.; Eisenberg, D. S. An amyloid-forming peptide from the yeast prion Sup35 reveals a dehydrated β-sheet structure for amyloid. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98* (5), 2375–2380.

78.   Sawaya, M. R.; Sambashivan, S.; Nelson, R.; Ivanova, M. I.; Sievers, S. A.; Apostol, M. I.; Thompson, M. J.; Balbirnie, M.; Wiltzius, J. J. W.; McFarlane, H. T.; Madsen, A. Ø.; Riekel, C.; Eisenberg, D. Atomic structures of amyloid cross-β spines reveal varied steric zippers. *Nature* **2007,** *447*, 453–457.

79.   Apostol, M. I.; Sawaya, M. R.; Cascio, D.; Eisenberg, D. Crystallographic Studies of Prion Protein (PrP) Segments Suggest How Structural Changes Encoded by Polymorphism at Residue 129 Modulate Susceptibility to Human Prion Disease. *J. Biol. Chem.* **2010,** *285* (39), 29671–29675.

80.   Wiltzius, J. J. W.; Sievers, S. A.; Sawaya, M. R.; Cascio, D.; Popov, D.; Riekel, C.; Eisenberg, D. Atomic structure of the cross-β spine of islet amyloid polypeptide (amylin). *Protein Sci.* **2008,** *17* (9), 1467–1474.

81.   Serag, A. A.; Altenbach, C.; Gingery, M.; Hubbell, W. L.; Yeates, T. O. Identification of a Subunit Interface in Transthyretin Amyloid Fibrils: Evidence for Self-Assembly from Oligomeric Building Blocks. *Biochemistry* **2001,** *40* (31), 9089–9096.

82.   Ionescu-Zanetti, C.; Khurana, R.; Gillespie, J. R.; Petrick, J. S.; Trabachino, L. C.; Minert, L. J.; Carter, S. A.; Fink, A. L. Monitoring the assembly of Ig light-chain amyloid fibrils by atomic force microscopy. *Proc. Natl. Acad. Sci. U.S.A.* **1999,** *96* (23), 13175–13179.

83.   Petkova, A. T.; Ishii, Y.; Balbach, J. J.; Antzutkin, O. N.; Leapman, R. D.; Delaglio, F.; Tycko, R. A structural model for Alzheimer's β-amyloid fibrils based on experimental constraints from solid state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99* (26), 16742–16747.

84.   Tycko, R. Insights into the Amyloid Folding Problem from Solid-State NMR. *Biochemistry* **2003,** *42* (11), 3151–3159.

85.   Eisenberg, D. S.; Sawaya, M. R. Structural Studies of Amyloid Proteins at the Molecular Level. *Annu. Rev. Biochem.* **2017,** *86* (1), 69–95.

86.   Pauling, L.; Corey, R. B. Configurations of Polypeptide Chains With Favored Orientations Around Single Bonds: Two New Pleated Sheets. *Proc. Natl. Acad. Sci. U.S.A.* **1951,** *37* (11), 729–740.

87. Tycko, R. Solid-State NMR Studies of Amyloid Fibril Structure. *Annu. Rev. Phys. Chem.* **2011,** *62* (1), 279–299.

88. Qiang, W.; Yau, W.-M.; Luo, Y.; Mattson, M. P.; Tycko, R. Antiparallel β-sheet architecture in Iowa-mutant β-amyloid fibrils. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (12), 4443–4448.

89. Hartl, F. U. Protein Misfolding Diseases. *Annu. Rev. Biochem.* **2017,** *86* (1), 21–26.

90. Adamcik, J.; Mezzenga, R. Proteins Fibrils from a Polymer Physics Perspective. *Macromolecules* **2012,** *45* (3), 1137–1150.

91. Knowles, T. P.; Fitzpatrick, A. W.; Meehan, S.; Mott, H. R.; Vendruscolo, M.; Dobson, C. M.; Welland, M. E. Role of Intermolecular Forces in Defining Material Properties of Protein Nanofibrils. *Science* **2007,** *318* (5858), 1900–1903.

92. Jarrett, J. T.; Lansbury, P. T. Seeding "one-dimensional crystallization" of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **1993,** *73* (6), 1055–1058.

93. Pallitto, M. M.; Murphy, R. M. A Mathematical Model of the Kinetics of β-Amyloid Fibril Growth from the Denatured State. *Biophys. J.* **2001,** *81* (3), 1805–1822.

94. Wogulis, M.; Wright, S.; Cunningham, D.; Chilcote, T.; Powell, K.; Rydel, R. E. Nucleation-Dependent Polymerization Is an Essential Component of Amyloid-Mediated Neuronal Cell Death. *J. Neurosci.* **2005,** *25* (5), 1071–1080.

95. Xue, W.-F.; Homans, S. W.; Radford, S. E. Systematic analysis of nucleation-dependent polymerization reveals new insights into the mechanism of amyloid self-assembly. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (26), 8926–8931.

96. Morriss-Andrews, A.; Shea, J.-E. Computational Studies of Protein Aggregation: Methods and Applications. *Annu. Rev. Phys. Chem.* **2015,** *66* (1), 643–666.

97. Bitan, G.; Kirkitadze, M. D.; Lomakin, A.; Vollers, S. S.; Benedek, G. B.; Teplow, D. B. Amyloid β-protein (Aβ) assembly: Aβ40 and Aβ42 oligomerize through distinct pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2003,** *100* (1), 330.

98. Urbanc, B.; Cruz, L.; Yun, S.; Buldyrev, S. V.; Bitan, G.; Teplow, D. B.; Stanley, H. E. *In silico* study of amyloid β-protein folding and oligomerization. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101* (50), 17345–17350.

99.   Haass, C.; Selkoe, D. J. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. *Nat. Rev. Mol. Cell Biol.* **2007,** *8*, 101–112.

100.   Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T. Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **2009,** *1*, 326–331.

101.   Ono, K.; Condron, M. M.; Teplow, D. B. Structure–neurotoxicity relationships of amyloid β-protein oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106* (35), 14745–14750.

102.   Hills, R. D.; Brooks, C. L. Hydrophobic Cooperativity as a Mechanism for Amyloid Nucleation. *J. Mol. Biol.* **2007,** *368* (3), 894–901.

103.   Auer, S.; Dobson, C. M.; Vendruscolo, M. Characterization of the nucleation barriers for protein aggregation and amyloid formation. *HFSP J.* **2007,** *1* (2), 137–146.

104.   Powers, E. T.; Powers, D. L. The Kinetics of Nucleated Polymerizations at High Concentrations: Amyloid Fibril Formation Near and Above the "Supercritical Concentration". *Biophys. J.* **2006,** *91* (1), 122–132.

105.   Lomakin, A.; Teplow, D. B.; Kirschner, D. A.; Benedek, G. B. Kinetic theory of fibrillogenesis of amyloid β-protein. *Proc. Natl. Acad. Sci. U.S.A.* **1997,** *94* (15), 7942–7947.

106.   Sabaté, R.; Estelrich, J. Evidence of the Existence of Micelles in the Fibrillogenesis of β-Amyloid Peptide. *J. Phys. Chem. B* **2005,** *109* (21), 11027–11032.

107.   Lee, J.; Culyba, E. K.; Powers, E. T.; Kelly, J. W. Amyloid-β forms fibrils by nucleated conformational conversion of oligomers. *Nat. Chem. Biol.* **2011,** *7*, 602–609.

108.   Collins, S. R.; Douglass, A.; Vale, R. D.; Weissman, J. S. Mechanism of Prion Propagation: Amyloid Growth Occurs by Monomer Addition. *PLoS Biol.* **2004,** *2* (10), e321.

109.   Serio, T. R.; Cashikar, A. G.; Kowal, A. S.; Sawicki, G. J.; Moslehi, J. J.; Serpell, L.; Arnsdorf, M. F.; Lindquist, S. L. Nucleated Conformational Conversion and the Replication of Conformational Information by a Prion Determinant. *Science* **2000,** *289* (5483), 1317–1321.

110.   Xue, W.-F.; Hellewell, A. L.; Gosal, W. S.; Homans, S. W.; Hewitt, E. W.; Radford, S. E. Fibril Fragmentation Enhances Amyloid Cytotoxicity. *J. Biol. Chem.* **2009,** *284* (49), 34272–34282.

111. Xue, W.-F.; Hellewell, A. L.; Hewitt, E. W.; Radford, S. E. Fibril fragmentation in amyloid assembly and cytotoxicity. *Prion* **2010,** *4* (1), 20–25.

112. Jahn, T. R.; Radford, S. E. Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch. Biochem. Biophys.* **2008,** *469* (1), 100–117.

113. Gazit, E. The "Correctly Folded" State of Proteins: Is It a Metastable State? *Angew. Chem., Int. Ed.* **2002,** *41* (2), 257–259.

114. Baldwin, A. J.; Knowles, T. P. J.; Tartaglia, G. G.; Fitzpatrick, A. W.; Devlin, G. L.; Shammas, S. L.; Waudby, C. A.; Mossuto, M. F.; Meehan, S.; Gras, S. L.; Christodoulou, J.; Anthony-Cahill, S. J.; Barker, P. D.; Vendruscolo, M.; Dobson, C. M. Metastability of Native Proteins and the Phenomenon of Amyloid Formation. *J. Am. Chem. Soc.* **2011,** *133* (36), 14160–14163.

115. Joerg, G.; Michele, V. Theoretical Approaches to Protein Aggregation. *Protein Pept. Lett.* **2006,** *13* (3), 287–293.

116. Buell, A. K.; Dhulesia, A.; White, D. A.; Knowles, T. P. J.; Dobson, C. M.; Welland, M. E. Detailed Analysis of the Energy Barriers for Amyloid Fibril Growth. *Angew. Chem., Int. Ed.* **2012,** *51* (21), 5247–5251.

117. Fawzi, N. L.; Okabe, Y.; Yap, E.-H.; Head-Gordon, T. Determining the Critical Nucleus and Mechanism of Fibril Elongation of the Alzheimer's Aβ_{1-40} Peptide. *J. Mol. Biol.* **2007,** *365* (2), 535–550.

118. Cohen, S. I. A.; Linse, S.; Luheshi, L. M.; Hellstrand, E.; White, D. A.; Rajah, L.; Otzen, D. E.; Vendruscolo, M.; Dobson, C. M.; Knowles, T. P. J. Proliferation of amyloid-β42 aggregates occurs through a secondary nucleation mechanism. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110* (24), 9758–9763.

119. Jeong, J. S.; Ansaloni, A.; Mezzenga, R.; Lashuel, H. A.; Dietler, G. Novel Mechanistic Insight into the Molecular Basis of Amyloid Polymorphism and Secondary Nucleation during Amyloid Formation. *J. Mol. Biol.* **2013,** *425* (10), 1765–1781.

120. Meisl, G.; Yang, X.; Hellstrand, E.; Frohm, B.; Kirkegaard, J. B.; Cohen, S. I. A.; Dobson, C. M.; Linse, S.; Knowles, T. P. J. Differences in nucleation behavior underlie the contrasting aggregation kinetics of the Aβ40 and Aβ42 peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2014,** *111* (26), 9384–9389.

121. Fu, Z.; Aucoin, D.; Davis, J.; Van Nostrand, W. E.; Smith, S. O. Mechanism of Nucleated Conformational Conversion of Aβ42. *Biochemistry* **2015,** *54* (27), 4197–4207.

122. Nguyen, H. D.; Hall, C. K. Molecular dynamics simulations of spontaneous fibril formation by random-coil peptides. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101* (46), 16180–16185.

123. Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999,** *293* (2), 321–331.

124. Tompa, P. Intrinsically disordered proteins: a 10-year recap. *Trends Biochem. Sci.* **2012,** *37* (12), 509–516.

125. Dyson, H. J. Making Sense of Intrinsically Disordered Proteins. *Biophys. J.* **2016,** *110* (5), 1013–1016.

126. Uversky, V. N. Intrinsically disordered proteins from A to Z. *Int. J. Biochem. Cell Biol.* **2011,** *43* (8), 1090–1103.

127. Tompa, P.; Csermely, P. The role of structural disorder in the function of RNA and protein chaperones. *FASEB J.* **2004,** *18* (11), 1169–1175.

128. Kovacs, D.; Tompa, P. Diverse functional manifestations of intrinsic structural disorder in molecular chaperones. *Biochem. Soc. Trans.* **2012,** *40* (5), 963.

129. Foit, L.; George, J. S.; Zhang, B. W.; Brooks, C. L.; Bardwell, J. C. A. Chaperone activation by unfolding. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110* (14), E1254–E1262.

130. Oldfield, C. J.; Meng, J.; Yang, J. Y.; Yang, M. Q.; Uversky, V. N.; Dunker, A. K. Flexible nets: disorder and induced fit in the associations of p53 and 14-3-3 with their partners. *BMC Genomics* **2008,** *9* (1), S1.

131. Iakoucheva, L. M.; Brown, C. J.; Lawson, J. D.; Obradović, Z.; Dunker, A. K. Intrinsic Disorder in Cell-signaling and Cancer-associated Proteins. *J. Mol. Biol.* **2002,** *323* (3), 573–584.

132. Bussell, R.; Eliezer, D. Residual Structure and Dynamics in Parkinson's Disease-associated Mutants of α-Synuclein. *J. Biol. Chem.* **2001,** *276* (49), 45996–46003.

133. Eliezer, D.; Barré, P.; Kobaslija, M.; Chan, D.; Li, X.; Heend, L. Residual Structure in the Repeat Domain of Tau: Echoes of Microtubule Binding and Paired Helical Filament Formation. *Biochemistry* **2005,** *44* (3), 1026–1036.

134. Havlin, R. H.; Tycko, R. Probing site-specific conformational distributions in protein folding with solid-state NMR. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (9), 3284–3289.

135. Mylonas, E.; Hascher, A.; Bernadó, P.; Blackledge, M.; Mandelkow, E.; Svergun, D. I. Domain Conformation of Tau Protein Studied by Solution Small-Angle X-ray Scattering. *Biochemistry* **2008,** *47* (39), 10345–10353.

136. Bernadó, P.; Svergun, D. I. Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. Biosyst.* **2012,** *8* (1), 151–167.

137. Maji, S. K.; Perrin, M. H.; Sawaya, M. R.; Jessberger, S.; Vadodaria, K.; Rissman, R. A.; Singru, P. S.; Nilsson, K. P. R.; Simon, R.; Schubert, D.; Eisenberg, D.; Rivier, J.; Sawchenko, P.; Vale, W.; Riek, R. Functional Amyloids As Natural Storage of Peptide Hormones in Pituitary Secretory Granules. *Science* **2009,** *325* (5938), 328–332.

138. Singh, P. K.; Maji, S. K. Amyloid-Like Fibril Formation by Tachykinin Neuropeptides and Its Relevance to Amyloid β-Protein Aggregation and Toxicity. *Cell Biochem. Biophys.* **2012,** *64* (1), 29–44.

139. Schmidt, M.; Rohou, A.; Lasker, K.; Yadav, J. K.; Schiene-Fischer, C.; Fändrich, M.; Grigorieff, N. Peptide dimer structure in an Aβ(1–42) fibril visualized with cryo-EM. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (38), 11858–11863.

140. Bhowmik, D.; MacLaughlin, C. M.; Chandrakesan, M.; Ramesh, P.; Venkatramani, R.; Walker, G. C.; Maiti, S. pH changes the aggregation propensity of amyloid-β without altering the monomer conformation. *Phys. Chem. Chem. Phys.* **2014,** *16* (3), 885–889.

141. Chandrakesan, M.; Bhowmik, D.; Sarkar, B.; Abhyankar, R.; Singh, H.; Kallianpur, M.; Dandekar, S. P.; Madhu, P. K.; Maiti, S.; Mithu, V. S. Steric Crowding of the Turn Region Alters the Tertiary Fold of Amyloid-$\beta_{18\text{-}35}$ and Makes It Soluble. *J. Biol. Chem.* **2015,** *290* (50), 30099–30107.

142. Elbaum-Garfinkle, S.; Rhoades, E. Identification of an Aggregation-Prone Structure of Tau. *J. Am. Chem. Soc.* **2012,** *134* (40), 16607–16613.

143. Arya, S.; Mukhopadhyay, S. Ordered Water within the Collapsed Globules of an Amyloidogenic Intrinsically Disordered Protein. *J. Phys. Chem. B* **2014,** *118* (31), 9191–9198.

144. Zerze, G. H.; Miller, C. M.; Granata, D.; Mittal, J. Free Energy Surface of an Intrinsically Disordered Protein: Comparison between Temperature Replica Exchange Molecular Dynamics and Bias-Exchange Metadynamics. *J. Chem. Theory Comput.* **2015,** *11* (6), 2776–2782.

145. Ithuralde, R. E.; Roitberg, A. E.; Turjanski, A. G. Structured and Unstructured Binding of an Intrinsically Disordered Protein as Revealed by Atomistic Simulations. *J. Am. Chem. Soc.* **2016,** *138* (28), 8742–8751.

146. Das, P.; Matysiak, S.; Mittal, J. Looking at the Disordered Proteins through the Computational Microscope. *ACS Cent. Sci.* **2018,** *4* (5), 534–542.

147. Vitalis, A.; Wang, X.; Pappu, R. V. Quantitative Characterization of Intrinsic Disorder in Polyglutamine: Insights from Analysis Based on Polymer Theories. *Biophys. J.* **2007,** *93* (6), 1923–1937.

148. Wales, D. J.; Bogdan, T. V. Potential Energy and Free Energy Landscapes. *J. Phys. Chem. B* **2006,** *110* (42), 20765–20776.

149.   Raval, A.; Piana, S.; Eastwood, M. P.; Shaw, D. E. Assessment of the utility of contact-based restraints in accelerating the prediction of protein structure using molecular dynamics simulations. *Protein Sci.* **2016,** *25* (1), 19–29.

150.   Jana, Asis K.; Sengupta, N. Adsorption Mechanism and Collapse Propensities of the Full-Length, Monomeric $A\beta_{1-42}$ on the Surface of a Single-Walled Carbon Nanotube: A Molecular Dynamics Simulation Study. *Biophys. J.* **2012,** *102* (8), 1889–1896.

151.   Brandts, J. F.; Rose, A. H. Thermobiology. Academic Press: New York, **1967**.

152.   Privalov, P. L.; Khechinashvili, N. N. A thermodynamic approach to the problem of stabilization of globular protein structure: A calorimetric study. *J. Mol. Biol.* **1974,** *86* (3), 665–684.

153.   Savage, H. J.; Elliott, C. J.; Freeman, C. M.; Finney, J. L. Lost hydrogen bonds and buried surface area: rationalising stability in globular proteins. *J. Chem. Soc. Faraday Trans.* **1993,** *89* (15), 2609–2617.

154.   Ruvinov, S.; Wang, L.; Ruan, B.; Almog, O.; Gilliland, G. L.; Eisenstein, E.; Bryan, P. N. Engineering the Independent Folding of the Subtilisin BPN' Prodomain: Analysis of Two-State Folding versus Protein Stability. *Biochemistry* **1997,** *36* (34), 10414–10421.

155.   Vogl, T.; Jatzke, C.; Hinz, H.-J.; Benz, J.; Huber, R. Thermodynamic Stability of Annexin V E17G: Equilibrium Parameters from an Irreversible Unfolding Reaction. *Biochemistry* **1997,** *36* (7), 1657–1668.

156.   Giver, L.; Gershenson, A.; Freskgard, P.-O.; Arnold, F. H. Directed evolution of a thermostable esterase. *Proc. Natl. Acad. Sci. U.S.A.* **1998,** *95* (22), 12809–12813.

157.   Shoichet, B. K.; Baase, W. A.; Kuroki, R.; Matthews, B. W. A relationship between protein stability and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **1995,** *92* (2), 452–456.

158.   Tokuriki, N.; Tawfik, D. S. Stability effects of mutations and protein evolvability. *Curr. Opin. Struct. Biol.* **2009,** *19* (5), 596–604.

159.   Greenbaum, E. A.; Graves, C. L.; Mishizen-Eberz, A. J.; Lupoli, M. A.; Lynch, D. R.; Englander, S. W.; Axelsen, P. H.; Giasson, B. I. The E46K Mutation in α-Synuclein Increases Amyloid Fibril Formation. *J. Biol. Chem.* **2005,** *280* (9), 7800–7807.

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

161. Jones, E. M.; Surewicz, K.; Surewicz, W. K. Role of N-terminal Familial Mutations in Prion Protein Fibrillization and Prion Amyloid Propagation in Vitro. *J. Biol. Chem.* **2006,** *281* (12), 8190–8196.

162. Morel, B.; Varela, L.; Azuaga, A. I.; Conejero-Lara, F. Environmental Conditions Affect the Kinetics of Nucleation of Amyloid Fibrils and Determine Their Morphology. *Biophys. J.* **2010,** *99* (11), 3801–3810.

163. Fraser, P. E.; Nguyen, J. T.; Surewicz, W. K.; Kirschner, D. A. pH-dependent structural transitions of Alzheimer amyloid peptides. *Biophys. J.* **1991,** *60* (5), 1190–1201.

164. Wood, S. J.; Maleeff, B.; Hart, T.; Wetzel, R. Physical, Morphological and Functional Differences between pH 5.8 and 7.4 Aggregates of the Alzheimer's Amyloid Peptide Aβ. *J. Mol. Biol.* **1996,** *256* (5), 870–877.

165. Gursky, O.; Aleshkov, S. Temperature-dependent β-sheet formation in β-amyloid $Aβ_{1-40}$ peptide in water: uncoupling β-structure folding from aggregation. *Biochim. Biophys. Acta Prot. Struct. Mol. Enzym.* **2000,** *1476* (1), 93–102.

166. Cruz, L.; Urbanc, B.; Borreguero, J. M.; Lazo, N. D.; Teplow, D. B.; Stanley, H. E. Solvent and mutation effects on the nucleation of amyloid β-protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (51), 18258–18263.

167. Kalhor, H. R.; Kamizi, M.; Akbari, J.; Heydari, A. Inhibition of Amyloid Formation by Ionic Liquids: Ionic Liquids Affecting Intermediate Oligomers. *Biomacromolecules* **2009,** *10* (9), 2468–2475.

168. De Felice, F. G.; Vieira, M. N. N.; Meirelles, M. N. L.; Morozova-Roche, L. A.; Dobson, C. M.; Ferreira, S. T. Formation of amyloid aggregates from human lysozyme and its disease-associated variants using hydrostatic pressure. *FASEB J.* **2004,** *18* (10), 1099–1101.

169. Torrent, J.; Balny, C.; Lange, R. High Pressure Modulates Amyloid Formation. *Protein Pept. Lett.* **2006,** *13* (3), 271–277.

170. van den Berg, B.; Ellis, R. J.; Dobson, C. M. Effects of macromolecular crowding on protein folding and aggregation. *EMBO J.* **1999,** *18* (24), 6927–6933.

171. Ellis, R. J. Macromolecular crowding: an important but neglected aspect of the intracellular environment. *Curr. Opin. Struct. Biol.* **2001,** *11* (1), 114–119.

172. Hatters, D. M.; Minton, A. P.; Howlett, G. J. Macromolecular Crowding Accelerates Amyloid Formation by Human Apolipoprotein C-II. *J. Biol. Chem.* **2002,** *277* (10), 7824–7830.

173. Cheung, M. S.; Klimov, D.; Thirumalai, D. Molecular crowding enhances native state stability and refolding rates of globular proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (13), 4753–4758.

174. Stagg, L.; Zhang, S.-Q.; Cheung, M. S.; Wittung-Stafshede, P. Molecular crowding enhances native structure and stability of α/β protein flavodoxin. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (48), 18976–18981.

175. Li, C.; Charlton, L. M.; Lakkavaram, A.; Seagle, C.; Wang, G.; Young, G. B.; Macdonald, J. M.; Pielak, G. J. Differential Dynamical Effects of Macromolecular Crowding on an Intrinsically Disordered Protein and a Globular Protein: Implications for In-Cell NMR Spectroscopy. *J. Am. Chem. Soc.* **2008,** *130* (20), 6310–6311.

176. Mukherjee, S.; Waegele, M. M.; Chowdhury, P.; Guo, L.; Gai, F. Effect of Macromolecular Crowding on Protein Folding Dynamics at the Secondary Structure Level. *J. Mol. Biol.* **2009,** *393* (1), 227–236.

177. Cino, E. A.; Karttunen, M.; Choy, W.-Y. Effects of Molecular Crowding on the Dynamics of Intrinsically Disordered Proteins. *PLoS ONE* **2012,** *7* (11), e49876.

178. Mikaelsson, T.; Ådén, J.; Johansson, Lennart B. Å.; Wittung-Stafshede, P. Direct Observation of Protein Unfolded State Compaction in the Presence of Macromolecular Crowding. *Biophys. J.* **2013,** *104* (3), 694–704.

179. Breydo, L.; Reddy, K. D.; Piai, A.; Felli, I. C.; Pierattelli, R.; Uversky, V. N. The crowd you're in with: Effects of different types of crowding agents on protein aggregation. *Biochim. Biophys. Acta Proteins Proteom.* **2014,** *1844* (2), 346–357.

180. Shtilerman, M. D.; Ding, T. T.; Lansbury, P. T. Molecular Crowding Accelerates Fibrillization of α-Synuclein: Could an Increase in the Cytoplasmic Protein Concentration Induce Parkinson's Disease? *Biochemistry* **2002,** *41* (12), 3855–3860.

181. Uversky, V. N.; M. Cooper, E.; Bower, K. S.; Li, J.; Fink, A. L. Accelerated α-synuclein fibrillation in crowded milieu. *FEBS Lett.* **2002,** *515* (1-3), 99–103.

182. Munishkina, L. A.; Cooper, E. M.; Uversky, V. N.; Fink, A. L. The effect of macromolecular crowding on protein aggregation and amyloid fibril formation. *J. Mol. Recognit.* **2004,** *17* (5), 456–464.

183. Munishkina, L. A.; Ahmad, A.; Fink, A. L.; Uversky, V. N. Guiding Protein Aggregation with Macromolecular Crowding. *Biochemistry* **2008,** *47* (34), 8993–9006.

184. White, D. A.; Buell, A. K.; Knowles, T. P. J.; Welland, M. E.; Dobson, C. M. Protein Aggregation in Crowded Environments. *J. Am. Chem. Soc.* **2010,** *132* (14), 5170–5175.

185. Magno, A.; Caflisch, A.; Pellarin, R. Crowding Effects on Amyloid Aggregation Kinetics. *J. Phys. Chem. Lett.* **2010,** *1* (20), 3027–3032.

186. Latshaw, D. C.; Cheon, M.; Hall, C. K. Effects of Macromolecular Crowding on Amyloid Beta (16–22) Aggregation Using Coarse-Grained Simulations. *J. Phys. Chem. B* **2014,** *118* (47), 13513–13526.

187. Latshaw, David C.; Hall, Carol K. Effects of Hydrophobic Macromolecular Crowders on Amyloid β (16–22) Aggregation. *Biophys. J.* **2015,** *109* (1), 124–134.

188. Argos, P.; Rossmann, M. G.; Grau, U. M.; Zuber, H.; Frank, G.; Tratschin, J. D. Thermal stability and protein structure. *Biochemistry* **1979,** *18* (25), 5698–5703.

189. Privalov, P. L. Stability of Proteins Small Globular Proteins. *Adv. Protein Chem.* **1979,** *33*, 167–241.

190. Privalov, P. L.; Griko, Y. V.; Venyaminov, S. Y.; Kutyshenko, V. P. Cold denaturation of myoglobin. *J. Mol. Biol.* **1986,** *190* (3), 487–498.

191. Privalov, P. L. Cold Denaturation of Protein. *Crit. Rev. Biochem. Mol. Biol.* **1990,** *25* (4), 281–306.

192. Griko, Y. V.; Privalov, P. L. Calorimetric study of the heat and cold denaturation of beta-lactoglobulin. *Biochemistry* **1992,** *31* (37), 8810–8815.

193. Damaschun, G.; Damaschun, H.; Gast, K.; Misselwitz, R.; Müller, J. J.; Pfeil, W.; Zirwer, D. Cold denaturation-induced conformational changes in phosphoglycerate kinase from yeast. *Biochemistry* **1993,** *32* (30), 7739–7746.

194. Gast, K.; Damaschun, G.; Damaschun, H.; Misselwitz, R.; Zirwer, D. Cold denaturation of yeast phosphoglycerate kinase: kinetics of changes in secondary structure and compactness on unfolding and refolding. *Biochemistry* **1993,** *32* (30), 7747–7752.

195. Franks, F.; Hatley, R. H. M. Stability of proteins at subzero temperatures: thermodynamics and some ecological consequences. *Pure Appl. Chem.* **1991,** *63*(10), 1367–1380.

196. Russell, N. J. Toward a molecular understanding of cold activity of enzymes from psychrophiles. *Extremophiles* **2000,** *4* (2), 83–90.

197. Rothschild, L. J.; Mancinelli, R. L. Life in extreme environments. *Nature* **2001,** *409* (6823), 1092–1101.

198. Hubel, A.; Skubitz, A. P. N. Principles of Cryopreservation. In *Biobanking of Human Biospecimens: Principles and Practice*; Springer International Publishing, **2017,** 1–21.

199. Huang, P.-S.; Boyken, S. E.; Baker, D. The coming of age of de novo protein design. *Nature* **2016,** *537*, 320–327.

200. Franks, F.; Hatley, R. H. M.; Friedman, H. L. The thermodynamics of protein stability: Cold destabilization as a general phenomenon. *Biophys. Chem.* **1988,** *31* (3), 307–315.

201. Griko, Y. V.; Privalov, P. L.; Sturtevant, J. M.; Venyaminov, S. Cold denaturation of staphylococcal nuclease. *Proc. Natl. Acad. Sci. U.S.A.* **1988,** *85* (10), 3343–3347.

202. Baldwin, R. L. Temperature dependence of the hydrophobic interaction in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1986,** *83* (21), 8069–8072.

203. Fu, L.; Freire, E. On the origin of the enthalpy and entropy convergence temperatures in protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **1992,** *89* (19), 9335–9338.

204. Agashe, V. R.; Udgaonkar, J. B. Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry* **1995,** *34* (10), 3286–3299.

205. Pastore, A.; Martin, S. R.; Politou, A.; Kondapalli, K. C.; Stemmler, T.; Temussi, P. A. Unbiased Cold Denaturation: Low- and High-Temperature Unfolding of Yeast Frataxin under Physiological Conditions. *J. Am. Chem. Soc.* **2007,** *129* (17), 5374–5375.

206. Graziano, G. On the mechanism of cold denaturation. *Phys. Chem. Chem. Phys.* **2014,** *16* (39), 21755–21767.

207. Sabelko, J.; Ervin, J.; Gruebele, M. Cold-Denatured Ensemble of Apomyoglobin: Implications for the Early Steps of Folding. *J. Phys. Chem. B* **1998,** *102* (10), 1806–1819.

208. Lopez, C. F.; Darst, R. K.; Rossky, P. J. Mechanistic Elements of Protein Cold Denaturation. *J. Phys. Chem. B* **2008,** *112* (19), 5961–5967.

209. Kunugi, S.; Tanaka, N. Cold denaturation of proteins under high pressure. *Biochim. Biophys. Acta Prot. Struct. Mol. Enzym.* **2002,** *1595* (1), 329–344.

210. Meersman, F.; Dobson, C. M.; Heremans, K. Protein unfolding, amyloid fibril formation and configurational energy landscapes under high pressure conditions. *Chem. Soc. Rev.* **2006,** *35* (10), 908–917.

211. Ravindra, R.; Winter, R. On the Temperature–Pressure Free-Energy Landscape of Proteins. *ChemPhysChem* **2003,** *4* (4), 359–365.

212. Buchner, G. S.; Shih, N.; Reece, A. E.; Niebling, S.; Kubelka, J. Unusual Cold Denaturation of a Small Protein Domain. *Biochemistry* **2012,** *51* (33), 6496–6498.

213. Szyperski, T.; Mills, J. L.; Perl, D.; Balbach, J. Combined NMR-observation of cold denaturation in supercooled water and heat denaturation enables accurate measurement of ΔC_p of protein unfolding. *Eur. Biophys. J.* **2006,** *35* (4), 363–366.

214. Kim, H.-Y.; Cho, M.-K.; Riedel, D.; Fernandez, C. O.; Zweckstetter, M. Dissociation of Amyloid Fibrils of α-Synuclein in Supercooled Water. *Angew. Chem., Int. Ed.* **2008,** *47* (27), 5046–5048.

215. Dias, C. L.; Ala-Nissila, T.; Karttunen, M.; Vattulainen, I.; Grant, M. Microscopic Mechanism for Cold Denaturation. *Phys. Rev. Lett.* **2008,** *100* (11), 118101.

216. Matysiak, S.; Debenedetti, P. G.; Rossky, P. J. Role of Hydrophobic Hydration in Protein Stability: A 3D Water-Explicit Protein Model Exhibiting Cold and Heat Denaturation. *J. Phys. Chem. B* **2012,** *116* (28), 8095–8104.

217. Bianco, V.; Franzese, G. Contribution of Water to Pressure and Cold Denaturation of Proteins. *Phys. Rev. Lett.* **2015,** *115* (10), 108101.

218. Kim, S. B.; Palmer, J. C.; Debenedetti, P. G. Computational investigation of cold denaturation in the Trp-cage miniprotein. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113* (32), 8991–8996.

219. Yang, C.; Jang, S.; Pak, Y. A fully atomistic computer simulation study of cold denaturation of a β-hairpin. *Nat. Commun.* **2014,** *5*, 5773.

220. Mishra, R.; Winter, R. Cold- and Pressure-Induced Dissociation of Protein Aggregates and Amyloid Fibrils. *Angew. Chem., Int. Ed.* **2008,** *47* (35), 6518–6521.

221. Ikenoue, T.; Lee, Y.-H.; Kardos, J.; Saiki, M.; Yagi, H.; Kawata, Y.; Goto, Y. Cold Denaturation of α-Synuclein Amyloid Fibrils. *Angew. Chem., Int. Ed.* **2014,** *53* (30), 7799–7804.

222. Horikoshi, K. Barophiles: deep-sea microorganisms adapted to an extreme environment. *Curr. Opin. Microbiol.* **1998,** *1* (3), 291–295.

223. Lauro, F. M.; Bartlett, D. H. Prokaryotic lifestyles in deep sea habitats. *Extremophiles* **2008,** *12* (1), 15–25.

224. Chen, C. R.; Makhatadze, G. I. Molecular determinant of the effects of hydrostatic pressure on protein folding stability. *Nat. Commun.* **2017,** *8*, 14561.

225. Hawley, S. A. Reversible pressure-temperature denaturation of chymotrypsinogen. *Biochemistry* **1971,** *10* (13), 2436–2442.

226. Roche, J.; Royer Catherine, A. Lessons from pressure denaturation of proteins. *J. Royal Soc. Interface* **2018,** *15* (147), 20180244.

227. Zipp, A.; Kauzmann, W. Pressure denaturation of metmyoglobin. *Biochemistry* **1973,** *12* (21), 4217–4228.

228. Brandts, J. F.; Oliveira, R. J.; Westort, C. Thermodynamics of protein denaturation. Effect of pressure on the denaturation on ribonuclease A. *Biochemistry* **1970,** *9* (4), 1038–1047.

229. Seemann, H.; Winter, R.; Royer, C. A. Volume, expansivity and isothermal compressibility changes associated with temperature and pressure unfolding of staphylococcal nuclease. *J. Mol. Biol.* **2001,** *307* (4), 1091-1102.

230. Paliwal, A.; Asthagiri, D.; Bossev, D. P.; Paulaitis, M. E. Pressure Denaturation of Staphylococcal Nuclease Studied by Neutron Small-Angle Scattering and Molecular Simulation. *Biophys. J.* **2004,** *87* (5), 3479–3492.

231. Schweiker, K. L.; Fitz, V. W.; Makhatadze, G. I. Universal Convergence of the Specific Volume Changes of Globular Proteins upon Unfolding. *Biochemistry* **2009,** *48* (46), 10846–10851.

232. Chalikian, T. V.; Breslauer, K. J. On volume changes accompanying conformational transitions of biopolymers. *Biopolymers* **1996,** *39* (5), 619–626.

233. Lin, L.-N.; Brandts, J. F.; Brandts, J. M.; Plotnikov, V. Determination of the Volumetric Properties of Proteins and Other Solutes Using Pressure Perturbation Calorimetry. *Anal. Biochem.* **2002,** *302* (1), 144–160.

234. Richards, F. M. Areas, Volumes, Packing, and Protein Structure. *Anal. Biochem. Annu. Rev. Biophys. Bioeng.* **1977,** *6* (1), 151–176.

235. Liang, J.; Dill, K. A. Are Proteins Well-Packed? *Biophys. J.* **2001,** *81* (2), 751–766.

236. Ando, N.; Barstow, B.; Baase, W. A.; Fields, A.; Matthews, B. W.; Gruner, S. M. Structural and Thermodynamic Characterization of T4 Lysozyme Mutants and the Contribution of Internal Cavities to Pressure Denaturation. *Biochemistry* **2008,** *47* (42), 11097–11109.

237. Roche, J.; Caro, J. A.; Norberto, D. R.; Barthe, P.; Roumestand, C.; Schlessman, J. L.; Garcia, A. E.; García-Moreno E, B.; Royer, C. A. Cavities determine the pressure unfolding of proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (18), 6945–6950.

238. Rouget, J.-B.; Aksel, T.; Roche, J.; Saldana, J.-L.; Garcia, A. E.; Barrick, D.; Royer, C. A. Size and Sequence and the Volume Change of Protein Folding. *J. Am. Chem. Soc.* **2011,** *133* (15), 6020–6027.

239. Bianco, V.; Franzese, G.; Dellago, C.; Coluzza, I. Role of Water in the Selection of Stable Proteins at Ambient and Extreme Thermodynamic Conditions. *Phys. Rev. X* **2017,** *7* (2), 021047.

240. Ghosh, T.; García, A. E.; Garde, S. Molecular Dynamics Simulations of Pressure Effects on Hydrophobic Interactions. *J. Am. Chem. Soc.* **2001,** *123* (44), 10997–11003.

241. Sarupria, S.; Garde, S. Quantifying Water Density Fluctuations and Compressibility of Hydration Shells of Hydrophobic Solutes and Proteins. *Phys. Rev. Lett.* **2009,** *103* (3), 037803.

242. Fulton, A. B. How crowded is the cytoplasm? *Cell* **1982,** *30* (2), 345-347.

243. Zimmerman, S. B.; Minton, A. P. Macromolecular crowding: biochemical, biophysical, and physiological consequences. *Annu. Rev. Biophys. Biomol. Struct.* **1993,** *22*, 27–65.

244. Ellis, R. J.; Minton Allen, P. Protein aggregation in crowded environments. *Biol. Chem.* **2006,** *387*(5), 485–497.

245. Minton, A. P. Models for Excluded Volume Interaction between an Unfolded Protein and Rigid Macromolecular Cosolutes: Macromolecular Crowding and Protein Stability Revisited. *Biophys. J.* **2005,** *88* (2), 971–985.

246. Elcock, A. H. Models of macromolecular crowding effects and the need for quantitative comparisons with experiment. *Curr. Opin. Struct. Biol.* **2010,** *20* (2), 196–206.

247. Wang, Y.; Li, C.; Pielak, G. J. Effects of Proteins on Protein Diffusion. *J. Am. Chem. Soc.* **2010,** *132* (27), 9392–9397.

248. Minton, A. P. Excluded volume as a determinant of macromolecular structure and reactivity. *Biopolymers* **1981,** *20* (10), 2093–2120.

249. Minton, A. P. Implications of macromolecular crowding for protein assembly. *Curr. Opin. Struct. Biol.* **2000,** *10* (1), 34–39.

250. Sasahara, K.; McPhie, P.; Minton, A. P. Effect of Dextran on Protein Stability and Conformation Attributed to Macromolecular Crowding. *J. Mol. Biol.* **2003,** *326* (4), 1227–1237.

251. Fung, J.; Darabie, A. A.; McLaurin, J. Contribution of simple saccharides to the stabilization of amyloid structure. *Biochem. Biophys. Res. Commun.* **2005,** *328* (4), 1067–1072.

252. Menon, S.; Sengupta, N. Influence of Hyperglycemic Conditions on Self-Association of the Alzheimer's Amyloid β (Aβ$_{1-42}$) Peptide. *ACS Omega* **2017,** *2* (5), 2134–2147.

253. O'Brien, E. P.; Straub, J. E.; Brooks, B. R.; Thirumalai, D. Influence of Nanoparticle Size and Shape on Oligomer Formation of an Amyloidogenic Peptide. *J. Phys. Chem. Lett.* **2011,** *2* (10), 1171–1177.

254. Zhou, H.-X.; Rivas, G.; Minton, A. P. Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences. *Annu. Rev. Biophys.* **2008,** *37* (1), 375–397.

255. Bastolla, U.; Dehouck, Y.; Echave, J. What evolution tells us about protein physics, and protein physics tells us about evolution. *Curr. Opin. Struct. Biol.* **2017,** *42*, 59–66.

256. Taverna, D. M.; Goldstein, R. A. Why are proteins so robust to site mutations? *J. Mol. Biol.* **2002,** *315* (3), 479–484.

257. Reddy, B. V.; Datta, S.; Tiwari, S. Use of propensities of amino acids to the local structural environments to understand effect of substitution mutations on protein stability. *Protein Eng.* **1998,** *11* (12), 1137–1145.

258. Shortle, D.; Stites, W. E.; Meeker, A. K. Contributions of the large hydrophobic amino acids to the stability of staphylococcal nuclease. *Biochemistry* **1990,** *29* (35), 8033–8041.

259. Matthews, B. W. Structural and genetic analysis of protein stability. *Annu. Rev. Biochem.* **1993,** *62*, 139–160.

260. Liu, R.; Baase, W. A.; Matthews, B. W. The introduction of strain and its effects on the structure and stability of T4 lysozyme. *J. Mol. Biol.* **2000,** *295* (1), 127–145.

261. van der Kamp, M. W.; Daggett, V. Pathogenic Mutations in the Hydrophobic Core of the Human Prion Protein Can Promote Structural Instability and Misfolding. *J. Mol. Biol.* **2010,** *404* (4), 732–748.

262. Singh, J.; Udgaonkar, J. B. The Pathogenic Mutation T182A Converts the Prion Protein into a Molten Globule-like Conformation Whose Misfolding to Oligomers but Not to Fibrils Is Drastically Accelerated. *Biochemistry* **2016,** *55* (3), 459–469.

263. Yin, S.; Pham, N.; Yu, S.; Li, C.; Wong, P.; Chang, B.; Kang, S.-C.; Biasini, E.; Tien, P.; Harris, D. A.; Sy, M.-S. Human prion proteins with pathogenic mutations share common conformational changes resulting in enhanced binding to glycosaminoglycans. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (18), 7546–7551.

264. Grant, M. A.; Lazo, N. D.; Lomakin, A.; Condron, M. M.; Arai, H.; Yamin, G.; Rigby, A. C.; Teplow, D. B. Familial Alzheimer's disease mutations alter the stability of the amyloid β-protein monomer folding nucleus. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (42), 16522–16527.

265. Matouschek, A.; Kellis, J. T.; Serrano, L.; Bycroft, M.; Fersht, A. R. Transient folding intermediates characterized by protein engineering. *Nature* **1990,** *346* (6283), 440–445.

266. Fersht, A. R. Protein folding and stability: the pathway of folding of barnase. *FEBS Lett.* **1993,** *325* (1), 5–16.

267. Jäger, M.; Nguyen, H.; Crane, J. C.; Kelly, J. W.; Gruebele, M. The folding mechanism of a β-sheet: the WW domain. *J. Mol. Biol.* **2001,** *311* (2), 373–393.

268. Hart, T.; Hosszu, L. L. P.; Trevitt, C. R.; Jackson, G. S.; Waltho, J. P.; Collinge, J.; Clarke, A. R. Folding kinetics of the human prion protein probed by temperature jump. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106* (14), 5651–5656.

269. Hosszu, L. L. P.; Tattum, M. H.; Jones, S.; Trevitt, C. R.; Wells, M. A.; Waltho, J. P.; Collinge, J.; Jackson, G. S.; Clarke, A. R. The H187R Mutation of the Human Prion Protein Induces Conversion of Recombinant Prion Protein to the PrPSc-like Form. *Biochemistry* **2010,** *49* (40), 8729–8738.

270. Merlo, C.; Dill, K. A.; Weikl, T. R. Φ values in protein-folding kinetics have energetic and structural components. *Proc. Natl. Acad. Sci. U.S.A.* **2005,** *102* (29), 10171–10175.

271. Socha, R. D.; Tokuriki, N. Modulating protein stability – directed evolution strategies for improved protein function. *FEBS J.* **2013,** *280* (22), 5582-5595.

272. Scott, D. J.; Kummer, L.; Tremmel, D.; Plückthun, A. Stabilizing membrane proteins through protein engineering. *Curr. Opin. Chem. Biol.* **2013,** *17* (3), 427–435.

273. Moroz, Y. S.; Dunston, T. T.; Makhlynets, O. V.; Moroz, O. V.; Wu, Y.; Yoon, J. H.; Olsen, A. B.; McLaughlin, J. M.; Mack, K. L.; Gosavi, P. M.; van Nuland, N. A. J.; Korendovych, I. V. New Tricks for Old Proteins: Single Mutations in a Nonenzymatic Protein Give Rise to Various Enzymatic Activities. *J. Am. Chem. Soc.* **2015,** *137* (47), 14905–14911.

274. Heydenreich, F. M.; Vuckovic, Z.; Matkovic, M.; Veprintsev, D. B. Stabilization of G protein-coupled receptors by point mutations. *Front Pharmacol.* **2015,** *6*, 82.

275. Wilding, M.; Hong, N.; Spence, M.; Buckle, A. M.; Jackson, C. J. Protein engineering: the potential of remote mutations. *Biochem. Soc. Trans.* **2019,** *47* (2), 701–711.

276. Gobeil, S. M. C.; Ebert, M. C. C. J. C.; Park, J.; Gagné, D.; Doucet, N.; Berghuis, A. M.; Pleiss, J.; Pelletier, J. N. The Structural Dynamics of Engineered β-Lactamases Vary Broadly on Three Timescales yet Sustain Native Function. *Sci. Rep.* **2019,** *9* (1), 6656.

277. DePristo, M. A.; Weinreich, D. M.; Hartl, D. L. Missense meanderings in sequence space: a biophysical view of protein evolution. *Nat. Rev. Genet.* **2005,** *6* (9), 678–687.

278. Chevalier, A.; Silva, D.-A.; Rocklin, G. J.; Hicks, D. R.; Vergara, R.; Murapa, P.; Bernard, S. M.; Zhang, L.; Lam, K.-H.; Yao, G.; Bahl, C. D.; Miyashita, S.-I.; Goreshnik, I.; Fuller, J. T.; Koday, M. T.; Jenkins, C. M.; Colvin, T.; Carter, L.; Bohn, A.; Bryan, C. M.; Fernández-Velasco, D. A.; Stewart, L.; Dong, M.; Huang, X.; Jin, R.; Wilson, I. A.; Fuller, D. H.; Baker, D. Massively parallel de novo protein design for targeted therapeutics. *Nature* **2017,** *550,* 74–79.

279. Gallardo, R.; Ramakers, M.; De Smet, F.; Claes, F.; Khodaparast, L.; Khodaparast, L.; Couceiro, J. R.; Langenberg, T.; Siemons, M.; Nyström, S.; Young, L. J.; Laine, R. F.; Young, L.; Radaelli, E.; Benilova, I.; Kumar, M.; Staes, A.; Desager, M.; Beerens, M.; Vandervoort, P.; Luttun, A.; Gevaert, K.; Bormans, G.; Dewerchin, M.; Van Eldere, J.; Carmeliet, P.; Vande Velde, G.; Verfaillie, C.; Kaminski, C. F.; De Strooper, B.; Hammarström, P.; Nilsson, K. P. R.; Serpell, L.; Schymkowitz, J.; Rousseau, F. De novo design of a biologically active amyloid. *Science* **2016,** *354* (6313), aah4949.

280. Paul, K. R.; Hendrich, C. G.; Waechter, A.; Harman, M. R.; Ross, E. D. Generating new prions by targeted mutation or segment duplication. *Proc. Natl. Acad. Sci. U.S.A.* **2015,** *112* (28), 8584–8589.

281. Childers, M. C.; Daggett, V. Insights from molecular dynamics simulations for computational protein design. *Mol. Syst. Des. Eng.* **2017,** *2* (1), 9–33.

Methodology

2.1   Abstract

Computational and theoretical modeling approaches, over the years, have garnered significance in complementing experimental techniques to understand the physics and chemistry of biomolecular systems. Biomolecules like proteins are inherently flexible at ambient conditions and these dynamical properties reflect the mechanistic aspects of their biological functions. This necessitates molecular level understanding of the structure and dynamics of these macromolecules thereby providing insights in the structure-dynamics-function linkage. In this regard, molecular dynamics (MD) simulations have proven valuable as they capture the behavior of proteins and other biomolecules in atomic detail and at very fine temporal resolution. MD simulations provide insights in a wide variety of biomolecular processes such as protein folding, conformational changes, ligand binding, the structural basis of diseases and designing small molecules, peptides and proteins in therapeutics. The key motivation of this thesis is to probe the structural and dynamical response of proteins and protein assemblies subject to alterations of thermodynamic variables such as temperature, structural perturbations such as mutations and in crowded environments. The proteins investigated here belong to a class of proteins called *"intrinsically disordered proteins"* (IDPs) that attribute their biological function to the intrinsic structural disorder and are also implicated in several aggregation diseases. Experimental

limitations in understanding the complex nature of conformational fluctuations of this class of proteins have motivated the rigorous implementation of computer modeling and simulations in capturing the molecular origin of their unique structure and function. We have extensively used atomistic MD simulations for the work presented in this thesis. This chapter gives an overview of the basic tenets of classical atomistic MD simulations and some of the associated statistical analyses performed routinely to analyze MD data.

2.2   Molecular Dynamics Simulations

The underlying principle of classical atomistic MD simulations (in contrast with quantum or *ab initio* simulations, which are not discussed here) is that classical Newtonian mechanics can be applied to molecular systems to accurately describe the dynamics of the system. Electrons are considered implicitly through the Born-Oppenheimer approximation, which states that timescales of nuclear motion are much slower than electronic motions. Thus, nuclear motions, rotations and vibrations can be studied independently from electronic fluctuations since electrons are assumed to find their equilibrium distribution once the positions of the nuclei are known.

In the classical atomistic MD model, the nuclei and electrons together are treated as spherical particles or hard spheres, possessing a net point charge. The radii as well as net charges of the atoms are determined experimentally or from high-level quantum calculations. Atomic interactions are calculated based on harmonic approximations or classical potentials that determine the spatial distributions of atoms and their corresponding energies. In MD simulations, the time evolution of a system of interacting particles is determined by solving Newton's equations of motion thus sampling the phase space of the system yielding a trajectory defined by the positions

and velocity vectors[1]. Application of force on each atom causes it to accelerate that is dependent upon the atomic mass. This fundamental relationship, described by Newton's second law of motion, can be used to calculate the motion of a particle under an applied force as:

$$\vec{F} = m\vec{a} = m\frac{d\vec{v}}{dt} = m\frac{d^2\vec{r}}{dt^2} \qquad (2.1)$$

The current position ($\vec{r}$), velocity ($\vec{v}$) and instantaneous forces ($\vec{F}$) acting upon a particle of mass m are used to numerically integrate the second-order differential equation (Equation 2.1) to yield the velocities and change in positions at the next time step. The net force experienced by a particle due to interactions from the rest of the system is derived from the negative gradient or first derivative of the potential energy (V) with respect to the atom positions as,

$$\vec{F} = -\frac{dV}{dr} = -\nabla V \qquad (2.2)$$

The equations of motion are deterministic in nature, i.e. the positions and velocities at time zero determine the position and velocities at any time, t. The calculation of a trajectory therefore, requires an initial set of positions and an initial distribution of velocities. The initial coordinates may be obtained from experimental methods such as X-ray crystallography and NMR spectroscopy or theoretical homology models or a combination of the two. The initial velocity vector is typically chosen randomly such that the total kinetic energy of the system corresponds to the value expected at the target temperature T and adjusted such that the total momentum of the system is zero. This is usually accomplished by assigning velocity components from a Maxwell-Boltzmann or Gaussian distribution, which gives the probability that an atom i of mass m_i has a velocity v_{ix} in the x direction at a temperature T.

$$p(v_{ix}) = \left(\frac{m_i}{2\pi k_B T}\right)^{\frac{1}{2}} \exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{k_B T}\right]$$ (2.3)

The instantaneous temperature can be calculated from the particle's momentum, p given by the relation,

$$T = \frac{1}{(3N)}\sum_{i=1}^{N}\frac{|p|}{2m_i}$$ (2.4)

2.2.1 Integration of Equations of Motion

The integration of the equations of motion is a many-body problem that cannot be solved analytically and hence they are discretized and solved numerically. The *"finite difference method"* is used to propagate the system in time with continuous potentials that are assumed to be pairwise additive. The basic idea is that the integration is discretized into small time steps of δt. The total force on each particle in the system at time t is calculated as the vector sum of interaction with the other particles in the configuration. The accelerations of the particles can be computed from the forces and then used along with the positions and velocities at time t to obtain the new positions and velocities at a time $t+\delta t$. In the same way, the forces on the particles at the new positions are calculated and used to determine the positions and velocities at the next time step $t+2\delta t$.

Some of the commonly used algorithms for integrating Newton's equations of motion in MD simulations are described in this section.

2.2.1.1 Verlet Algorithm

The Verlet algorithm[2] is the most widely used method employed in MD simulations. It uses the positions and accelerations at time t and the positions from the previous

step, $r(t-\delta t)$ where δt is a small timestep, to calculate the new positions, $r(t+\delta t)$, at the next time step, $t+\delta t$. The basic formula can be derived from a Taylor series expansion and is read as,

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \qquad (2.5)$$

$$r(t-\delta t) = r(t) - v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \qquad (2.6)$$

The summation of the above two equations yields,

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + a(t)\delta t^2 \qquad (2.7)$$

It must be noted that the velocities are not required for the calculation of the new positions. The velocities can be derived as:

$$v(t) = \frac{[r(t+\delta t) - r(t-\delta t)]}{2\delta t} \qquad (2.8)$$

The Verlet algorithm is straightforward and of modest precision. Several flavors of the Verlet algorithm have been developed to propagate velocities explicitly, such as the velocity Verlet method[3], the leap-frog algorithm[4] and Beeman's algorithm[5].

2.2.1.2   Velocity Verlet Algorithm

The widely used velocity Verlet algorithm utilizes a computer's finite precision efficiently and yields positions, velocities and accelerations at the same time. The equations implemented in the velocity Verlet method are as follows:

$$r(t+\delta t) = r(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \qquad (2.9)$$

$$v(t+\delta t) = v(t) + \frac{1}{2}[a(t) + a(t+\delta t)]\delta t \qquad (2.10)$$

The standard implementation scheme of the algorithm is as follows:

(1) From the initial configuration at time t, the new positions are calculated using equation 2.9

(2) The velocities at an intermediate time step $t + \frac{1}{2}\delta t$ are determined as:

$$v\left(t + \frac{1}{2}\delta t\right) = v(t) + \frac{1}{2}\delta t a(t) \tag{2.11}$$

(3) The new forces are calculated using the current positions thereby giving $a(t + \delta t)$. The velocities at time $t + \delta t$ are determined using:

$$v(t + \delta t) = v\left(t + \frac{1}{2}\delta t\right) + \frac{1}{2}\delta t a(a + \delta t) \tag{2.12}$$

This method is simple and easy to implement, with accuracy reaching up to the order of $O(\Delta t^4)$.

The choice of the timestep δt is an important consideration in a simulation protocol; it should be smaller that the fastest motion in the system. The size of the timestep is particularly relevant to the computational demands since a simulation with a longer timestep will rapidly sample the phase space in less iterations. The highest frequency vibrations, typically bonds containing hydrogen atoms, limit the magnitude of the simulation timestep. It is therefore of considerable interest to increase the timestep by imposing constraints on the internal coordinates thus freezing the high frequency motions without affecting the other internal degrees of freedom. Some commonly used methods for implementing constrained dynamics include the SHAKE[6], RATTLE[7] and LINCS[8] algorithms.

2.3   Force Fields

The potential energy function used to calculate the energy and geometry of a molecular system is called the "*force field*". This section provides an overview of atomistic force fields employed in classical MD simulations that is used in the studies pertinent to this thesis. An atomistic force field comprises of a functional form describing the multidimensional potential energy surface of a system as well as a parameter set of atom types and bonded parameters contained in the system. A typical molecular mechanics force field consists of components describing intra and intermolecular forces. Among the most commonly used force fields are the CHARMM[9-10], AMBER[11-12], GROMOS[13] and OPLS[14-15] force fields. The energies of the biomolecules studied in this thesis were calculated using the CHARMM force field and hence explained below in terms of this force field. The CHARMM potential energy function is based on intrinsic coordinates on bonded and pairwise non-bonded energy terms. The bonded potentials, E_{bonded} calculate the energy corresponding to deviations of bond lengths and angles away from their equilibrium values as well as bond rotations in a molecule, as depicted in Figure 2.1. The non-bonded energy terms, $E_{non\text{-}bonded}$ account for pairwise interactions between non-bonded atoms *i.e.*, atoms separated by three or more covalent bonds. The total potential energy, E_{total} of an N atom system as defined by the functional form of the CHARMM force field is given by:

$$E_{total} = E_{bonded} + E_{nonbonded}$$

$$= \left(E_{bonds} + E_{angles} + E_{dihedrals} + E_{impropers} + E_{UB} \right) + \left(E_{vdW} + E_{coulomb} \right) \qquad (2.13)$$

2.3.1 Bonded Potentials
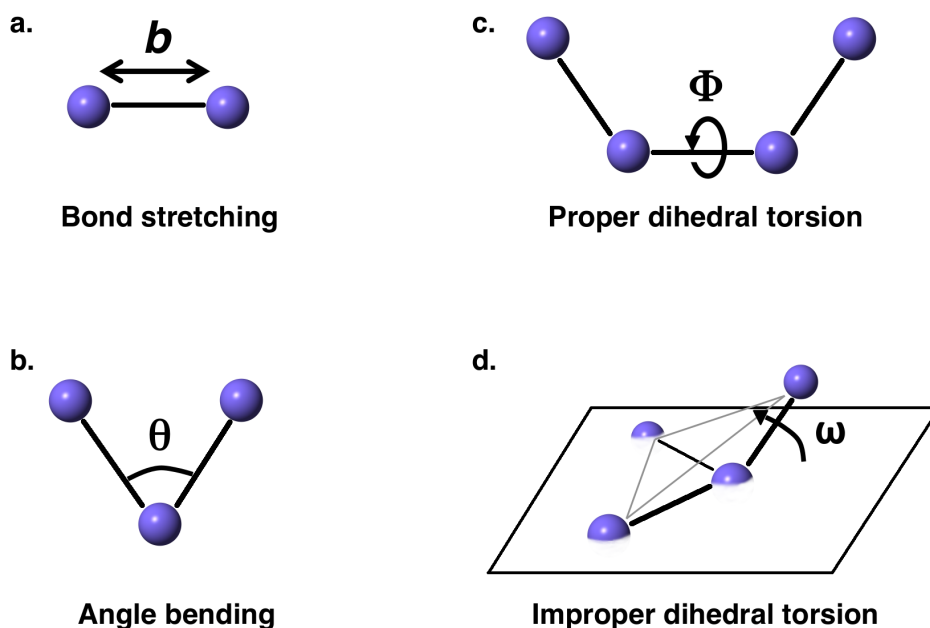
Figure 2.1. Schematic representation of the bonded interactions terms, (a) bond stretching, (b) angle bending, (c) proper and (d) improper dihedrals, contributing to the bonded potential in the force field.

E_{bonds} is defined by a harmonic potential representing the interactions between two atoms separated by one covalent bond and is given by:

$$E_{bonds} = \sum_{bonds} k_b (b - b_0)^2 \qquad (2.14)$$

This term accounts for the energy of a bond of bond length, b as it deviates from the ideal or equilibrium bond length, b_0 and k_b is the force constant that determines the strength of the bond. The ideal bond length, b_0 and the force constant k_b depend upon the chemical nature of the atom constituents of a bond and are specific for each atom pair.

E_{angles} corresponds to the angular potential upon alteration of bond angles, θ from their equilibrium values, θ_0. This term is also represented as a harmonic potential of force constant, k_θ.

$$E_{angles} = \sum_{angles} k_\theta \left(\theta - \theta_0\right)^2 \tag{2.15}$$

$E_{dihedrals}$ represents the torsional angle potential which models the torsional barriers between four atoms separated by three covalent bonds (*i.e.*1,4 pairs). It describes the rotation of a dihedral angle, ϕ around a bond with a force constant, k_ϕ, phase γ and multiplicity, *n*. This potential is expressed as:

$$E_{dihedrals} = \sum_{dihedrals} k_\phi \left(1 + \cos\left(n\phi - \gamma\right)\right) \tag{2.16}$$

$E_{impropers}$ is the energy term that accounts for improper dihedral angles in which the four atoms are not bonded in a sequential manner *i.e.* three atoms centered around a fourth atom. This potential has been designed to maintain chirality and planarity about certain groups in a molecular structure and is defined as:

$$E_{impropers} = \sum_{impropers} k_\omega \left(\omega - \omega_0\right)^2 \tag{2.17}$$

where k_ω is the force constant around the *'out-of-plane'* torsional angle.

E_{UB} is the Urey-Bradley energy term and accounts for the stretch-bend cross term by using 1,3-non-bonded interactions *i.e.* atoms bound to a common atom. This potential is modeled by a harmonic function of the distance between the 1,3 atoms and has the form:

$$E_{UB} = \sum k_u \left(u - u_0 \right)^2 \tag{2.18}$$

where k_u is the Urey-Bradley force constant and u is the distance between the 1,3 atom pairs under consideration.

2.3.2   Non-bonded Potentials

The non-bonded potential, $E_{non\text{-}bonded}$ has two components arising from the van der Waals interaction energy, E_{vdW} and electrostatic interaction energy, $E_{coulomb}$.

Van der Waals Potential

Van der Waals interactions comprises of attractive and repulsive forces between two atoms that are not bonded to each other. These interactions can be associated with dipole-dipole or dipole-induced dipole forces that arise due to fluctuations in the electron clouds. The energy of these interactions varies as a function of the interatomic distance. It is zero at infinite distance and as the distance between the atoms decreases, the energy decreases, reaches a minimum and then rapidly increases as the distance further decreases.

The van der Waals potential for a non-bonded distance, r_{ij} for an atom pair (indexed i and j) is commonly described by a Lennard-Jones 12-6 potential as:

$$E_{vdW} = \sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \tag{2.19}$$

where r_{ij} is the interatomic distance between the two atoms, ε_{ij} is the depth of the Lennard-Jones potential and σ_{ij} is the location where the potential is zero, also known as the collision diameter. This potential mimics the balance of two forces, a short-ranged repulsive force that varies as r^{-12}, and a slower decaying attractive or

dispersive force that varies as r^{-6}. In the CHARMM force field, each atom type is assigned the characteristic Lennard-Jones parameters *i.e.* σ and ε rather than specific atom pairs. To calculate the van der Waals interactions between different types of atom pairs, the Lorentz-Berthelot combining rules[16] are applied to the corresponding one-body parameters as:

$$\sigma_{ij} = \frac{1}{2}\left(\sigma_i + \sigma_j\right) \qquad (2.20)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \qquad (2.21)$$

Since van der Waals forces are short-ranged, for computational convenience, these interactions can be computed for atom pairs only within a pre-defined cutoff radius.

Coulomb Potential

The ionic interactions between charged groups can be approximated by Coulomb's law for each atom pair as:

$$E_{coulomb} = \sum_{i,j} \frac{q_i q_j}{\varepsilon r_{ij}} \qquad (2.22)$$

where ε is the effective dielectric constant of the medium, r_{ij} is the distance between a pair of atoms with partial charges q_i and q_j. Owing to the long-ranged nature of electrostatic interactions, simple truncation methods based on cutoff distances are poor approximations to the energy functions. Hence, more sophisticated algorithms such as the Ewald summation and other lattice techniques[17-20] and fast multipole methods[21-22] have been developed for the correct treatment of long range forces. The simulations performed for the work presented in the following chapters were done using the *particle mesh Ewald* (PME) method[19] as implemented in the CHARMM force field.

2.4   Periodic Boundary Conditions

In MD simulations, it is crucial to treat boundaries and boundary effects appropriately since it allows the calculation of macroscopic properties from a relatively small number of particles. Boundary or surface effects dominate in such small systems since the ratio of number of surface atoms to the total number of atoms is large. The application of "*periodic boundary conditions*" (PBC) in simulations eliminates artifacts arising due to surface effects and enables relatively small number of atoms to experience forces similar to a bulk system. Essentially, the modeled system of finite size is placed in a *unit cell* and considered to have replicas in three-dimensional space forming an infinite lattice[1]. Figure 2.2 depicts a standard square lattice in two-dimensions, with the unit cell (center) surrounded by its images or replicas. Non-cubic periodic cell shapes such as the hexagonal prism, truncated octahedron, rhombic dodecahedron and elongated dodecahedron can also be modeled.

In a simulation with PBC, only the coordinates of the unit cell are recorded and propagated. As an atom crosses the boundary and leaves (enters) the unit cell, an image enters (leaves) to replace it with the same velocity, from the opposite direction such that the total number of atoms in the simulation cell is conserved.
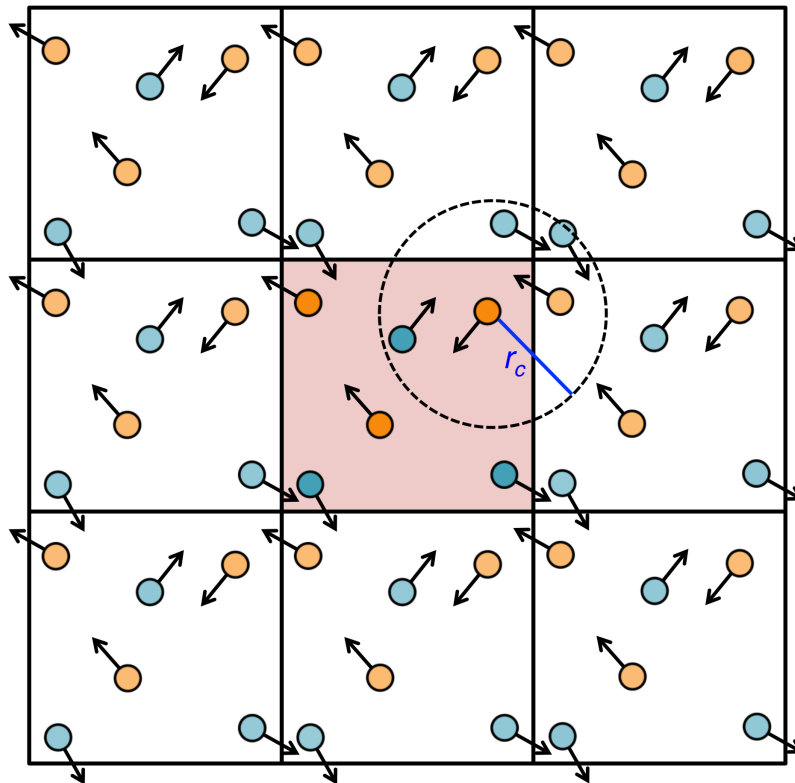
Figure 2.2. Schematic representation of periodic boundary conditions in two dimensions applied to a cubic unit cell (shaded in red). A vector depicts the velocity of each constituent atom. Short-range interactions are considered only within the spherical cutoff.

With the incorporation of periodicity, the number of interacting atom pairs increases enormously because each atom not only interacts with the other atoms in the box but also with its images in the neighboring replicas. In practical terms, this problem is countered by using PBC in combination with the "*minimum image convention*". In the minimum image convention, an atom *'i'* interacts only with the closest image of the other $(N-1)$ atoms in the periodic array. In addition, a spherical cutoff (r_c) is applied to restrict the potential in a finite range. A consistent combination of spherical cutoffs and minimum image convention is obtained when the cutoff is at most half the length of the simulation cell.

2.5   Non-bonded Neighbor Lists

The time required to calculate the distance between every pair of atoms within the system is proportional to the number of distinct pair separations, $N(N-1)/2$, in an N atom system. To reduce the cost of performing the non-bonded interaction calculations, a *non-bonded pair list* is made which contains all the atom pairs for which non-bonded interactions should be calculated (as implemented in the NAMD simulation package used in two chapters of this thesis). A distance greater than the non-bonded cutoff is specified by the parameter *pairlistdist* to construct a list of all the neighbors of each atom with a pair separation within this distance. Over the next few MD steps, only the atom pairs present in the neighbor list are considered for the force calculation routine. The neighbor list is reconstructed at regular intervals during the course of the simulation. The value of *pairlistdist* should be chosen such that no atom, initially outside this distance moves more than the non-bonded cutoff distance before the neighbor list is updated. It is also important to update the neighbor list at the correct frequency such that it ensures energy conservation and efficiency.

2.6   Thermodynamic Ensembles

MD simulations generates microscopic information of a system in the form of trajectories that can be analyzed using the principles of statistical mechanics to yield macroscopic properties such as pressure, energy, heat capacity etc. The atomic positions 'r' and momenta 'p' forming the coordinates in a multidimensional space called "*phase space*" describe the microscopic state of a system. An "*ensemble*" is defined as a collection of all possible microscopic states of a system that possess the same thermodynamic state. A simulation trajectory consists of a sequence of points in

phase space, corresponding to different conformations of the system and their momenta, generated as a function of time in the same ensemble. A fundamental postulate of statistical mechanics, the "*ergodic hypothesis*" states that the time average is equal to the ensemble average[1]. This implies that if a system is allowed to evolve in time indefinitely, it will eventually sample all possible states in the phase space. MD simulations can generate several different ensembles corresponding to different thermodynamic constraints.

a. Microcanonical ensemble (NVE): The thermodynamic state characterized by a fixed number of atoms, N, a fixed volume, V, and a fixed energy, E; corresponding to an isolated system.

b. Canonical ensemble (NVT): An ensemble of systems whose thermodynamic state is defined by a fixed number of atoms, N, a fixed volume, V, and a fixed temperature, T.

c. Isobaric-Isothermal ensemble (NPT): The thermodynamic state of this ensemble has fixed number of atoms, N, a constant pressure, P, and a constant temperature, T.

d. Grand canonical ensemble (mVT): The thermodynamic state for this ensemble is characterized by a fixed chemical potential, m, a fixed volume, V, and a constant temperature, T.

The simulation algorithmic framework discussed thus far conserves the energy of the system, corresponding to the microcanonical ensemble. However, studies of molecular properties of systems under thermal or mechanical equilibrium with the surroundings require simulations to be performed with constant temperature or pressure or both. The description in the following subsections gives a flavor of the methods for temperature and pressure control, enabling simulations of the NVT and NPT ensemble.

2.6.1   Temperature Control

The *"equipartition theorem"* of statistical mechanics states that each particle in an equilibrium system has an average kinetic energy of $k_B T/2$, where k_B is the Boltzmann constant and T is the temperature of the system. The temperature T of a system with N_f degrees of freedom is related to the mean kinetic energy as:

$$\langle K.E \rangle = \frac{1}{2} N_f k_B T$$

(2.23)

The total kinetic energy of the system is the sum of the individual kinetic energies of the particles in the system. The temperature of the system can be therefore expressed as:

$$T = \sum_{i=1}^{N} \frac{m_i v_i^2}{k_B N_f}$$

(2.24)

In the above expression, m_i and v_i represent the mass and velocity of the *i'th* particle in the system. The above relation suggests that the temperature of the system can be regulated based on the velocity phase space variable of the particles. In the following subsections, some of the methods designed to generate thermodynamic ensembles at constant temperature, referred to as *'thermostat'* are discussed.

2.6.1.1   Velocity Rescaling

A simple approach to control temperature is to scale the velocity vector at each timestep such that the system kinetic temperature, T reaches the target value, T_0. This can be achieved by scaling the velocity of each particle as:

$$v_{new} = c_t v_{old}$$

(2.25)

The scaling factor, c_t is calculated as:

$$c_t = \sqrt{\frac{T_0}{T}}$$

(2.26)

This method however, leads to rapid pumping of energy among the various degrees of freedom and does not reproduce the fluctuations expected from the canonical ensemble.

2.6.1.2 Berendsen Thermostat

Another approach to isothermal molecular dynamics is the weak coupling method introduced by Berendsen *et al.*[23]. This method couples the system to an external heat bath fixed at the target temperature. The velocities of the particles are scaled such that rate of change of temperature is proportional to the temperature difference between the bath and the system.

$$\frac{dT(t)}{dt} = \frac{1}{\tau}\left(T_{bath} - T(t)\right)$$

(2.27)

Here, $T(t)$ is the instantaneous kinetic temperature, T_{bath} is the temperature of the heat bath and τ is the time constant of the coupling to the heat bath.

The change in temperature between subsequent timesteps is given by:

$$\Delta T = \frac{\delta t}{\tau}\left(\frac{T_{bath}}{T(t)} - 1\right)$$

(2.28)

This scales the velocity vector as:

$$\lambda^2 = 1 + \frac{\delta t}{T}\left(\frac{T_{bath}}{T(t)} - 1\right)$$

(2.29)

This scaling of the velocities exponentially decays the system temperature to the desired value. The factor τ is the coupling parameter that controls the characteristic decay time and its magnitude determines the strength of the coupling to the heat bath.

When τ is large, the coupling will be weak. When τ is small, the coupling to the heat bath will be strong such that the energy exchange between the system and the thermal reservoir will be significant. When $\tau = \delta t$, the algorithm reduces to the simple velocity rescaling method for constant temperature simulations. This method is deterministic, stable and easy to implement, but it does not generate rigorous canonical ensemble.

2.6.1.3 Andersen Thermostat

The Andersen thermostat[24] implements stochastic molecular dynamics for canonical sampling. In this method, a particle is chosen at random intervals and its velocity is reassigned a random value from the Maxwell Boltzmann distribution. This is equivalent to coupling the system to a heat bath and bombarding randomly chosen particles with a thermal particle at the desired temperature T. The strength of the coupling to the heat reservoir is specified by a collision frequency, υ. This parameter signifies the mean rate at which each particle suffers stochastic collisions. The probability that any random particle suffers a stochastic collision in any small time interval is $\upsilon \Delta t$. The time intervals between two successive collisions is distributed according to the relation:

$$P(t) = \upsilon e^{-\upsilon t} \tag{2.30}$$

where $P(t)\Delta t$ is the probability that the collision interval is between t and $t + \Delta t$. For each particle i of the N particle system, a random number r_i is generated. If $r_i < \upsilon \Delta t$, then the velocity of the $i'th$ particle is rescaled with a value from the Maxwell Boltzmann distribution.

2.6.1.4 Langevin Thermostat

Another stochastic thermostat employs Langevin dynamics and random damping force to the momenta of the particles. In this formalism, for each particle i of the system, Newton's equations of motion are modified as:

$$f_i - \gamma_i \frac{dr_i}{dt} + R_i = m_i \frac{d^2 r_i}{dt^2} \tag{2.31}$$

In the above expression, γ_i represents the frictional drag on the particle, R_i is the random force acting on particle i due to the interaction potential. The random fluctuating force represents the thermal kicks to the particles from the surrounding particles. The damping factor and the random force combine to simulate the correct canonical ensemble.

2.6.1.5 Nosé-Hoover Thermostat

Extended system methods of temperature control, originally proposed by Nosé[25] and subsequently developed by Hoover[26], incorporate an extended Lagrangian containing additional, artificial coordinates and velocities. The system is coupled to a thermal reservoir at the target temperature, thereby adding an additional degree of freedom to the physical system, and assigned a fictitious mass. The augmented equations of motion of the extended system represent the evolution of both the internal and external variables in the desired ensemble. The magnitude of the fictitious mass ensures that the thermalization process is efficient. The Nosé-Hoover thermostat is an efficient method to simulate a real canonical ensemble with correct fluctuations.

2.6.2   Pressure Control

MD simulations in the isobaric ensemble are more relevant for validation of experimental measurements that are made under constant pressure conditions. In the same spirit as temperature control by weak coupling to a heat bath, for isobaric simulations, the system is coupled to a pressure bath[23]. The atomic coordinates and volume of the simulation box are scaled as the instantaneous pressure approaches the external applied pressure. However, a major drawback of this approach is that it does not produce trajectories from any known ensemble that makes it difficult to relate fluctuations in any observable to thermodynamic properties.

In the extended pressure coupling method[24], an additional degree of freedom, corresponding to the volume of the simulation cell is included, that fluctuates to balance the internal and external applied pressures. This extra degree of freedom effectively acts as a piston and is assigned a mass of W. Solving the extended Lagrangian produces trajectories consistent with the isobaric ensemble.

The Langevin piston method[27] allows partial damping of the piston by describing the piston degrees of freedom with a Langevin equation as:

$$\frac{dr_i}{dt} = \frac{p_i}{m_i} + \frac{1}{3V}\frac{dV}{dt}r_i \tag{2.32}$$

$$\frac{dp_i}{dt} = f_i - \frac{1}{3V}\frac{dV}{dt}p \tag{2.33}$$

$$\frac{d^2V}{dt^2} = \frac{1}{W}\left[P(t) - P_{ext}\right] - \gamma\frac{dV}{dt} + R(t) \tag{2.34}$$

where r_i, p_i, m_i and f_i denote the position, momentum, mass and force acting on the $i'th$ particle in the system, respectively. $P(t)$ is the instantaneous pressure, V is the

volume of the system, P_{ext} is the imposed pressure, W is the piston mass, γ is the collision frequency and $R(t)$ is a random force obtained from a Gaussian distribution. $R(t)$ has zero mean and a variance given by:

$$\langle R(0)R(t)\rangle = \frac{2\gamma k_B T \delta(t)}{W} \tag{2.35}$$

When $\gamma=0$, this method is same as the extended system method in which the motion of the piston is completely undamped.

2.7   Analysis of MD Simulation Trajectories

MD simulation yields a trajectory that describes the structural ensemble explored by the system in terms of temporal evolution of the positions and velocities of the atoms in the system. Equilibrated MD trajectories can be further processed by statistical measures to obtain structural, dynamical and functional insights about the biomolecular system studied. Various observables are calculated from the trajectories based on the ergodic hypothesis that equates equilibrium time averaged properties to ensemble averages. Some of the routinely used measures of biomolecular MD data analysis to extract equilibrium properties are briefly described below.

(i) Root mean squared deviation (RMSD): It is a measure of the mean distances between atom positions of two superimposed structures. The smaller the deviation, the more spatially equivalent are the two conformations. RMSD analysis of the trajectory can be used to understand the global conformational changes of the biomolecules (for example, protein) as a function of simulation time. RMSD is mathematically represented as:

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(r_i^{\alpha} - r_i^{\beta}\right)^2} \tag{2.36}$$

Here, r_i^α and r_i^β are the atomic positions of the *i'th* atom in the two aligned structures.

(ii) Root mean squared fluctuation (RMSF): It is a measure of the extent of deviation or fluctuation in the mean position of an atom i with respect to the corresponding position in the reference structure. It is calculated using the formula:

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}\left(x_i(t)-\overline{x}_i\right)^2} \tag{2.37}$$

Here, T is the time period over which the averaging is done for every atom i, $x_i(t)$ is the atomic position at time t and $\overline{x}(t)$ is the time-averaged position of atom i.

(iii) Radius of gyration (R_g): It is an indicator of the overall compactness of a protein conformation. R_g can be calculated from the root mean square distance of each atom from the center of mass of the molecule. R_g can be mathematically expressed as:

$$R_g = \sqrt{\frac{1}{N_i}\sum_i\left(r_i-r_{cm}\right)^2} \tag{2.38}$$

where $(r_i - r_{cm})$ is the distance between the atom i and the center of mass of the molecule.

(iii) Constant volume specific heat capacity (C_V): The isochoric specific heat capacity at temperature T is related to the energetic fluctuations as:

$$C_V = \frac{\left\langle \delta E^2 \right\rangle}{k_B T^2} \tag{2.39}$$

In the above expression, the numerator is the variance of energy and k_B is the Boltzmann constant. C_V is calculated to determine the size of thermodynamic

fluctuations in the system. It is a major thermodynamic quantity measured in biophysical studies of proteins to obtain insights in protein folding and its stability. Protein unfolding is associated with a positive heat capacity difference between the unfolded and folded states at the transition temperature, both in heat and cold denaturation events.

(iv) Principal Component Analysis (PCA)

Protein dynamics is manifested as a temporal change in its molecular structure or conformation. The broad range of time and spatial scales of these accessible motions can be appropriately represented as a vector space of a large number of dimensions equal to the number of degrees of freedom describing these motions. MD simulations generate trajectories sampling the accessible conformational ensemble characterized by these degrees of freedom. The configurational space can be separated into two subspaces:

(a) An "*essential subspace*" comprising only a few degrees of freedom in which anharmonic motion occurs. Most of the positional fluctuations are concentrated in this subspace and are essential for function.

(b) The remaining space with all other degrees of freedom represent much less important motion and has a narrow Gaussian distribution that can be considered as physically constrained.

Interestingly, it has been found that the vast array of protein dynamics can be characterized by a surprisingly small number of collective degrees of freedom[28]. Owing to the complexity of biomolecular systems, however, simulation trajectories can be challenging to analyze and decipher the functional essential motions among the noise of protein dynamics.

MD simulation trajectory is multivariate data in which a large number of variables (atomic positions) are typically interrelated, correlated or dependent on each other. These large data sets are most suited for statistical multivariate analysis to extract meaningful information. In particular, principal component analysis (PCA) is the simplest multivariate statistical technique that is used to reduce or simplify the original large number of dependent variables to a smaller and independent set to explain the phenomena of interest. In the context of biomolecular simulations, PCA is based on the assumption that the major collective modes of fluctuation dominate functional dynamics of the biomolecule. This approach has the advantage that the individual modes can be examined and visualized separately and hence enables the major modes of collective motion to be separated for more local fluctuations.

The theory underlying the PCA technique is briefly described in this section. PCA reduces the dimensionality of a multivariate data set by taking the p interrelated variables, $x_1, x_2, ..., x_p$, and based on variances, finding combinations of these variables to generate a transformed set of variables, $z_1, z_2, ... , z_p$ that are uncorrelated. These transformed independent variables (indices z_i) are called the "*principal components*" (PCs). The steps involved in the analysis are listed below:

(i) *Construction of the covariance matrix:* The process of determining the principal components starts with the construction of the $p \times p$ covariance or correlation matrix from a set of configurations from MD trajectories. It is necessary to first transform the trajectory to eliminate the overall rotational and translational motions that do not contribute to the real dynamics of the system. This is accomplished by aligning the structures in the trajectory to a reference structure, which is typically an experimental structure or an average structure from the trajectory. The alignment is commonly done

using a subset of atoms, α-carbons or the heavy atoms in the structure. A matrix $x(t)$ describes the aligned trajectory, where x is a *3N*-dimensional vector of all atomic coordinates represented by a column vector and t is the timestamp that defines each of the rows. A variance-covariance matrix of positional fluctuations is constructed as:

$$C = \left\langle \left(x(t) - \langle x \rangle \right) \left(x(t) - \langle x \rangle \right)^T \right\rangle \tag{2.40}$$

where $\langle \ \rangle$ denotes an ensemble average.

(ii) The covariance matrix C is a symmetric matrix that can be diagonalized by an orthogonal coordinate transformation T as:

$$C = T \Lambda T^T \tag{2.41}$$

Here, Λ is the diagonal (eigenvalue) matrix and T contains the eigenvectors of C as columns. The eigenvectors (μ_i) and eigenvalues (λ_i) correspond to the modes and amplitude of collective motion.

(iii) The eigenvectors are sorted such that their eigenvalues are in a decreasing order.

(iv) If μ_i is the *i'th* eigenvector of C, then the original configuration can be projected onto each of the principal components to yield the principal coordinates $p_i(t)$ as:

$$p_i(t) = \mu_i \cdot \left(x(t) - \langle x \rangle \right) \tag{2.42}$$

The variance $\left\langle p_i^2 \right\rangle$ is equal to the corresponding eigenvalue λ_i

Although PCA technique can be a very powerful tool in MD data analysis, it has some limitations that can introduce bias and lead to misinterpretation of data. An important source of error is the sampling limitation of MD simulations. If the conformational sampling is insufficient, PCA will provide an incomplete representation of the phase space of the molecule. The correct choice of the set of atoms depending upon the desired molecular motion is crucial to a successful PCA.

v) Free energy

The probability of finding a molecular system in one microstate or another depends on the free-energy difference between the two states. Free-energy calculations are useful for quantifying biophysical properties of biomolecules such as protein–protein and protein-ligand binding affinities, partition coefficients, protein stability etc. According to the principles of statistical mechanics, free energy difference between two states can be calculated from the ensemble averages of atomic-level conformations of the system, which can be generated from MD or Monte Carlo (MC) simulation methods. The free energy of a system can be expressed as Helmholtz free energy, for the NVT ensemble or the Gibbs free energy relevant to NPT ensemble. In MD, the free energies are calculated as a function of specific order parameters also referred to as "*collective variables*" or "*reaction coordinates*" that describe transformations as the system evolves from the initial, reference state to the final or target state. The choice of the collective variables may significantly influence the accuracy and efficiency of the free energy calculations. In this section, the formalism of two of the widely employed simulation methods for calculating free energy changes is briefly discussed.

Thermodynamic Integration

This method is specifically utilized to obtain Helmholtz free energy differences of a canonical ensemble. The evolution of the system is described in terms of a continuous coupling parameter λ varying from 0 to 1. The free energy as a function of λ can be expressed as:

$$\Delta A = \int_0^1 \frac{\partial A(\lambda)}{\partial \lambda} d\lambda \qquad (2.43)$$

The partition function of a canonical ensemble is given by:

$$Q_{NVT} = \frac{1}{N!}\frac{1}{h^{3N}}\int\int dp^N\, dr^N \exp\left[-\frac{H\left(p^N,r^N\right)}{k_B T}\right] \tag{2.44}$$

where $H(p^N, r^N)$ is the Hamiltonian of the system.

The Helmholtz free energy of the system can be obtained from the partition function as:

$$A = -k_B T \ln Q_{NVT} \tag{2.45}$$

With the partition function expressed in terms of λ, the above free energy expression can be written as:

$$\Delta A = -k_B T \int_0^1 \frac{1}{Q_{NVT}(\lambda)}\frac{\partial Q_{NVT}(\lambda)}{\partial \lambda} d\lambda \tag{2.46}$$

From equations (2.43) and (2.44), the following relation can be obtained:

$$\frac{\partial A(\lambda)}{\partial \lambda} = \int\int dp^N\, dr^N \frac{\partial H\left(p^N,r^N\right)}{\partial \lambda}\left\{\frac{\exp\left[-H\left(p^N,r^N\right)/k_B T\right]}{Q_{NVT}(\lambda)}\right\} \tag{2.47}$$

$$= \left\langle \frac{\partial H\left(p^N,r^{N,\lambda}\right)}{\partial \lambda}\right\rangle_\lambda \tag{2.48}$$

Thus, the free energy change is given by the formula:

$$\Delta A = \int_0^1 \left\langle \frac{\partial H\left(p^N,r^N,\lambda\right)}{\partial \lambda}\right\rangle_\lambda d\lambda \tag{2.49}$$

While applying the thermodynamic integration method, a number of simulations are run sampling the ensemble of equilibrium conformations at a series of λ values that are discrete but closely spaced.

The average quantity:

$$\left\langle \frac{\partial H\left(p^{N}, r^{N}, \lambda\right)}{\partial \lambda} \right\rangle_{\lambda} \tag{2.50}$$

is calculated for each simulation run. This quantity can then be plotted as a function of

λ and the area under the curve gives the free energy difference ΔA.

Umbrella Sampling

When large energetic barriers separate the two thermodynamic states of interest,

regions in configurational space are insufficiently sampled. Several techniques have

been developed to resolve the sampling issue. In the umbrella sampling method, the

reaction coordinate is restrained and driven to the desired final value by using a bias

potential[29]. The sampling is usually done in a series of windows and are finally

combined using with the weighted histogram analysis method (WHAM)[30-31] or

umbrella integration[32]. This method drives a system over an energy barrier by

augmenting the potential energy function of the system with a bias potential. The

modified potential function to connect energetically separated regions in phase space

is also referred to as the "*umbrella potential*".

The bias potential w_i corresponding to a window *i* is an additional energy function that

is a function of the reaction coordinate ξ. The net potential is given by:

$$E^{b}(r) = E^{u}(r) + w_{i}(\xi) \tag{2.51}$$

Here, the superscript *b* denotes biased potential and superscript *u* refers unbiased

quantities. A commonly used bias function is a quadratic, harmonic potential:

$$w(\xi) = K/2 \left(\xi - \xi^{ref} \right)^{2} \tag{2.52}$$

The unbiased probability distribution of the system along ξ can be expressed as:

$$\rho(\xi) = \frac{\int \exp\left[-(1/k_BT)E(r)\right]\delta\left[\xi'(r)-\xi\right]d^N r}{\int \exp\left[-(1/k_BT)E(r)\right]d^N r}$$

(2.53)

The biased simulation generates a non-Boltzmann distribution along the reaction coordinate. Assuming an ergodic system, this biased probability can be expressed as:

$$\rho'(\xi) = \frac{\int \exp\left\{-(1/k_BT)\left[E(r)+w_i\left(\xi'(r)\right)\right]\right\}\delta\left[\xi'(r)-\xi\right]d^N r}{\int \exp\left\{-(1/k_BT)\left[E(r)+w_i\left(\xi'(r)\right)\right]\right\}d^N r}$$

(2.54)

The integration is performed over all degrees of freedom expect ξ, which is the only factor on which the bias depends,

$$\rho'(\xi) = \exp\left[-(1/k_BT)w_i(\xi)\right]$$

$$\times \frac{\int \exp\left[-(1/k_BT)E(r)\right]\delta\left[\xi'(r)-\xi\right]d^N r}{\int \exp\left\{-(1/k_BT)\left[E(r)+w_i\left(\xi'(r)\right)\right]\right\}d^N r}$$

(2.55)

Using equation (2.53),

$$\rho(\xi) = \rho'(\xi)\exp\left[(1/k_BT)w_i(\xi)\right]$$

$$\times \frac{\int \exp\left\{-(1/k_BT)\left[E(r)+w_i\left(\xi(r)\right)\right]\right\}d^N r}{\int \exp\left[-(1/k_BT)E(r)\right]d^N r}$$

$$= \rho'(\xi)\exp\left[(1/k_BT)w_i(\xi)\right]$$

$$\times \frac{\int \exp\left[-(1/k_BT)E(r)\right]\exp\left\{-(1/k_BT)w_i\left[\xi(r)\right]\right\}d^N r}{\int \exp\left[-(1/k_BT)E(r)\right]d^N r}$$

$$= \rho'(\xi)\exp\left[(1/k_BT)w_i(\xi)\right]\left\langle \exp\left[-(1/k_BT)w_i(\xi)\right]\right\rangle$$

(2.56)

From the above equation, $A_i(\xi)$ can be evaluated as:

$$A_i(\xi) = -k_B T \ln \rho'(\xi) - w_i(\xi) + F_i \qquad (2.57)$$

In the above expression, $\rho'(\xi)$ is obtained from the biased simulation, $w_i(\xi)$ is provided analytically and $F_i = -k_B T \ln \langle \exp[-(1/k_B T) w_i(\xi)] \rangle$

The free energy difference, $\Delta A(\xi)$, between the states described by ξ and ξ_0 is expressed as:

$$\Delta A(\xi) = A(\xi) - A(\xi_0) = -k_B T \ln \frac{\rho(\xi)}{\rho(\xi_0)} \qquad (2.58)$$

Combining equations (2.53), (2.54) and (2.58):

$$\Delta A(\xi) = -k_B T \ln \frac{w(\xi_0) \rho'(\xi)}{w(\xi) \rho'(\xi_0)}$$

$$= -k_B \left[\ln \frac{\rho'(\xi)}{\rho'(\xi_0)} + \ln w(\xi_0) - \ln w(\xi) \right] \qquad (2.59)$$

The free energy difference, $\Delta A'(\xi)$, calculated by sampling the system with the non-Boltzmann probability density $\rho'(\xi)$ can be written as:

$$\Delta A'(\xi) = -k_B T \ln \frac{\rho'(\xi)}{\rho'(\xi_0)} \qquad (2.60)$$

Substituting the above expression in equation (2.59), one can obtain:

$$\Delta A(\xi) = \Delta A'(\xi) + k_B T \left[\ln w(\xi) - \ln w(\xi_0) \right] \qquad (2.61)$$

The above formula is used to obtain free energy difference of interest from a simulation performed using non-Boltzmann sampling.

Umbrella sampling method and its various modified versions are extensively for free

energy calculation in biomolecular systems[30,32-34].

2.8   References

1.   Leach, A. R. Molecular modelling: principles and applications. Harlow, England; New York: Prentice Hall, **2001**.

2.   Verlet, L. Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.* **1967,** *159* (1), 98-103.

3.   Swope, W. C.; Andersen, H. C.; Berens, P. H.; Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *J. Chem. Phys.***1982,** *76* (1), 637-649.

4.   Hockney, R. W. Methods in Computational Physics. *Alder, B* **1970**, 136-211.

5.   Beeman, D. Some multistep methods for use in molecular dynamics calculations. *J. Comput. Phys.* **1976,** *20* (2), 130-139.

6.   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23* (3), 327-341.

7.   Andersen, H. C. Rattle: A "velocity" version of the shake algorithm for molecular dynamics calculations. *J. Comput. Phys.* **1983,** *52* (1), 24-34.

8.   Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997,** *18* (12), 1463-1472.

9.   Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983,** *4* (2), 187-217.

10.   MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998,** *102* (18), 3586-3616.

11.   Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.* **1995,** *91* (1), 1-41.

12.   Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules J. Am. Chem. Soc. 1995, 117, 5179−5197. *J. Am. Chem. Soc.* **1996,** *118* (9), 2309-2309.

13.   Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004,** *25* (13), 1656-1676.

14.   Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996,** *118* (45), 11225-11236.

15.   Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B* **2001,** *105* (28), 6474-6487.

16.   Stone, A. J. The theory of intermolecular forces. Clarendon. Oxford: **1996**.

17.   Ewald, P. P. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.* **1921,** *369* (3), 253-287.

18.   Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.

19.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103* (19), 8577-8593.

20.   Hockney, R. W.; Eastwood, J. W. Computer simulation using particles. crc Press: **1988**.

21.   Greengard, L.; Rokhlin, V. A fast algorithm for particle simulations. *J. Comput. Phys.* **1987,** *73* (2), 325-348.

22.   Board, J. A.; Causey, J. W.; Leathrum, J. F.; Windemuth, A.; Schulten, K. Accelerated molecular dynamics simulation with the parallel fast multipole algorithm. *Chem. Phys. Lett.* **1992,** *198* (1), 89-94.

23.   Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984,** *81* (8), 3684-3690.

24.   Andersen, H. C. Molecular dynamics simulations at constant pressure and/or t emperature. *J. Chem. Phys.* **1980,** *72* (4), 2384-2393.

25.   Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **1984,** *52* (2), 255-268.

26.   Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **1985,** *31* (3), 1695-1697.

27.   Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.* **1995,** *103* (11), 4613-4621.

28.   Amadei, A.; Linssen, A. B. M.; Berendsen, H. J. C. Essential dynamics of proteins. *Proteins* **1993,** *17* (4), 412-425.

29.   Kästner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011,** *1* (6), 932-942.

30.   Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* **1992,** *13* (8), 1011-1021.

31.   Souaille, M.; Roux, B. t. Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations. *Comput. Phys. Commun.* **2001,** *135* (1), 40-57.

32.   Kästner, J.; Thiel, W. Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: "Umbrella integration". *J. Chem. Phys.* **2005,** *123* (14), 144104.

33.   Wang, J.; Gu, Y.; Liu, H. Determination of conformational free energies of peptides by multidimensional adaptive umbrella sampling. *J. Chem. Phys.* **2006,** *125* (9), 094907.

34.   Bursulaya, B. D.; Brooks, C. L. Folding Free Energy Surface of a Three-Stranded β-Sheet Protein. *J. Am. Chem. Soc.* **1999,** *121* (43), 9947-9951.

The Cold Thermal Response of an Amyloid Oligomer Differs from Typical Globular Protein Cold Denaturation
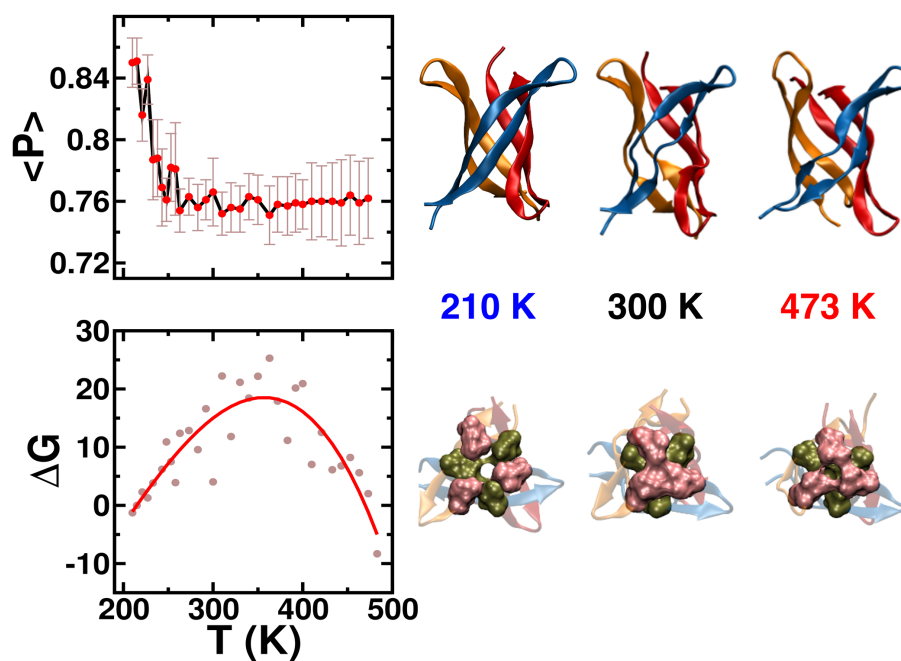
Based on:

Menon, S.; Sengupta, N. The Cold Thermal Response of an Amyloid Oligomer Differs from Typical Globular Protein Cold Denaturation. *J. Phys. Chem. Lett.* **2019,** *10* (10), 2453–2457.

3.1 Abstract

In contrast with the general behavior of folded proteins, the cold thermal response of amyloid assemblies is difficult to elicit with simple models. We exploit exhaustive simulations to evaluate the thermal response of a barrel-shaped model amyloid oligomer, with a distinct hydrophobic core akin to that of folded proteins. Cumulative thermal data over the range of 210–483 K indicate a sharp inflection and rise in structural stability as the temperature is lowered below the melting temperature of the water model. This is not commensurate with the equilibrium free energy profile obtained with core packing as the order parameter. However, energetic analyses and the size of their fluctuations indicate the crucial role of hydration in mediating structural transitions, beyond the expected temperature-dependent hydrophobic effect. Structural ordering of the hydration layer over bulk water is maximized at the transition and vanishes at high temperatures. This is a first direct demonstration of the microscopic influence of hydration water on the low-temperature response of an amyloid assembly close to the cryo-regime.

3.2 Introduction

The conformational and functional response of proteins to intensive thermodynamic variations, particularly temperature, is of great interest owing to their physiological and biophysical ramifications. Few examples of such ramifications include role of proteins in the viability of robust life forms such as tardigrades, adaption of life forms to extreme conditions[1-3], and even the plausibility of extra-terrestrial life[4]. In addition, the cold response of proteins and protein assemblies is crucially important from the standpoint of cryogenic preservation of biospecimens and recovery of their functionality with temperature reversal[5]. Understanding and effectively leveraging such phenomena necessitate molecular-level demonstrations of the interplay of key thermodynamic factors underlying thermostability of proteins and protein assemblies.

Studies of thermal stability of globular proteins now surpass three decades[6-11]. Temperature induced unfolding in small globular proteins is generally well approximated as a cooperative, two-state transition[6-8,12]. Indirect observation via external manipulation of the thermodynamic conditions may alter underlying energetics[13-14]. However, despite experimental limitations, free energy decompositions using microcalorimetric methods indicate that the overall temperature trends of enthalpy and entropy effectively counter and cancel each other at the 'cold' and 'heat' denaturation points[1,6-7]. Increasing the temperature reduces the fractional population with native contacts, with the population balance reversed beyond a melting temperature that coincides with a sharp peak in the protein heat capacity[15-16]. The counterintuitive favorability of the unfolded state increases at temperatures below physiological conditions and is variously referred to as protein *"cold denaturation"* or *"cold melting"*[16-17]. Cold denaturation has been directly observed in just a handful of globular proteins[11,18-20]. In general, it is often achieved by supercooling small sample

volumes, lowering the freezing point of water, or by enhancing denaturation by various means (*pH, mutation, denaturation*)[6,13-14,21]. The cold melting temperature of globular proteins is frequently estimated by extrapolation of free energy difference plots to the low temperature regime to the point where the difference reduces to zero[16].

Studies of the cold thermal response of amyloid assemblies are more recent, and till date has been reported for only a handful of self-assembled aggregates[22-25]. Folding and amyloidogenesis are competing processes that result in sharply divergent physiological consequences[26-27]. The molecular-level interactions that lead to folding and self-assembly are not different but superposed differently in the two processes. Unlike the downhill folding transition in a rough energy landscape, amyloidogenesis begins with slow nucleation followed by favorable growth into ordered fibrils[26]. Hydration plays subtle roles during amyloidogenesis; backbone hydrogen bonds offer substantial stability to the cross β-sheet architecture of the assemblies[28]. A recent study of dissociation of α-synuclein amyloid fibrils in supercooled water demonstrates that the fibrils have lower stability than globular proteins at low temperatures[22]. Ikenoue *et al.*[24] have demonstrated that while the fibrils of α-synuclein dissociate to monomers within 253 K, the fibrils of other key amyloids only show small decrease in stability in this temperature range. Underlying microscopic mechanisms leading to their dissimilar cold response over folded proteins remain unresolved.

The differential response of amyloid fibrils and globular proteins to cold thermal conditions must largely arise from the dissimilarities in the relative contributions of the non-bonded interactions. The hydrophobic effect contributes predominantly to the folding free energy of single domain globular proteins, but its

effects are relatively subtle in amyloid self-assembly[29-31]. Backbone hydrogen bonds are more crucial for the cross β-sheet architecture in amyloids, whereas side chain hydrogen bonds are relatively more important in folded proteins[31]. In both cases, however, specific side chain packing is a requisite feature that, for example, preserves the dry hydrophobic core in folded proteins or provides extra stability to amyloid fibrils[25]. Both protein folding and amyloidogenesis take place in predominantly aqueous milieu. In folding, initial collapse of the polypeptide chain led by the expulsion of solvating water molecules results in the formation of a hydrophobic core[25,32]. In the amyloid pathways of several polypeptides, initial self-assembly is facilitated by subtle yet significant dewetting transitions[33]. High temperature protein unfolding is attributed to increased configurational entropy. In contrast, weakening of the hydrophobic effect and a sharp reduction in the entropic cost of apolar solvation are considered as primary driving forces of protein cold denaturation, supported by a number of theoretical models and computer simulations[17,34-40].

The present report is an investigation of the thermal response of a peptide oligomer that has the distinguishing features of an amyloid assembly, but has an interior stabilized by the packing of hydrophobic side chains akin to folded globular proteins. The amyloidogenic segment was derived from the eye-lens protein, αB crystallin (ABC) and the oligomeric crystal structure reported by Laganowsky *et al.*[41]. The oligomer is a three double-stranded antiparallel barrel of cylindrical shape (termed "*cylindrin*") formed from a tandem repeat of the 11-residue segment, KVKVLGDVIEV, with a sequence variation of Leu replacing Val at position 2 and covalently linked by a double glycine linker and an additional N-terminal glycine[41]. The cylindrin assembly could serve as a model for the elusive toxic oligomers implicated in various amyloid diseases[41]; this hypothesis is supported by recent

reports of β-barrel arrangement along amyloid pathways[42-43]. The cylindrical β-barrel is consolidated with strategically located intra and inter-peptide hydrogen bonding interfaces. The strong interface is formed by a network of inter-hairpin backbone hydrogen bonds between the tightly twisted β-strands, while the weak interface is stabilized predominantly by inter-peptide backbone, side chain and water-mediated hydrogen bonds[41]. In addition, the interior of the β-barrel is characterized by tight packing of the inward facing hydrophobic side chains of valine and leucine residues. Schematic representations of the stabilizing non-bonded interactions of the oligomer are depicted in Figure 3.1. Herein, we have leveraged fully atomistic simulations of the oligomer over a temperature span exceeding 230 K to delineate the cold response of amyloid oligomers from that observed for globular proteins.
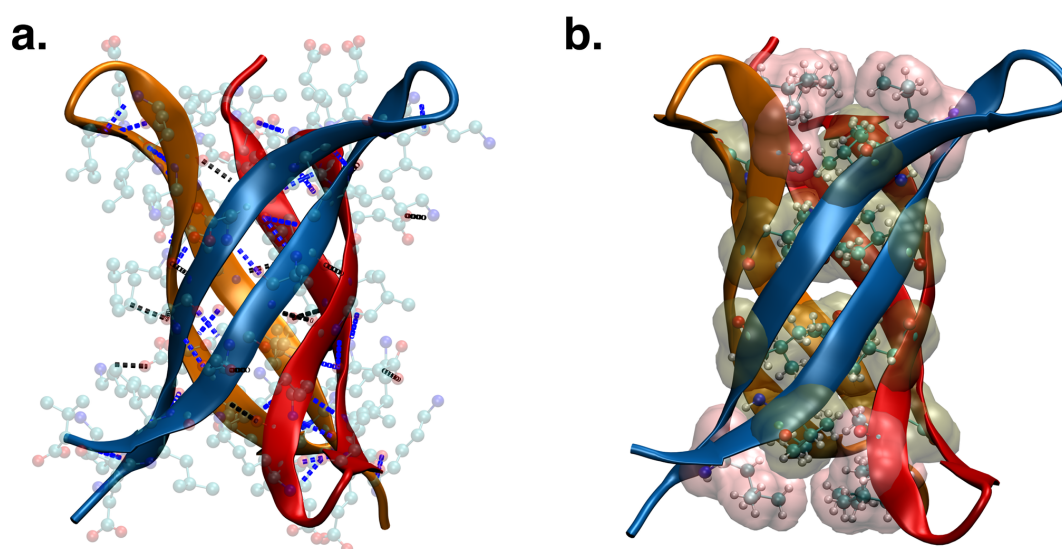
Figure 3.1   Crystal structure of the cylindrin oligomer composed of 3 peptides, colored in blue, orange and red. (a) Hydrogen bonds at the strong and weak interface are depicted as broken lines in blue and black, respectively, along with the side chain atoms of outward facing residues (b) Surface representation of the hydrophobic core (leucine in pink; valine in green) are depicted and overlaid with side chain atoms.

3.3   Methods

3.3.1   System Setup and MD Simulations

The X-ray crystal structure of the oligomer, resolved by Eisenberg *et al*. was obtained from the Protein Data Bank (PDB ID: 3SGR)[41] and solvated with water molecules in a cubic simulation box such that the protein assembly is at least 15 Å from the edges of the box. All-atom molecular dynamics (MD) simulations were performed using the GROMACS package (version 5.0.7)[44] at 33 temperatures ranging from 210 to 483 K (list of temperatures in Table 3.1). The CHARMM36 force field[45] was employed for the protein and TIP4P-2005 water model[46] was used due to its ability to reproduce temperature-dependent thermodynamic properties[47-48]. The particle mesh Ewald (PME) method[49] was used for computing long-range electrostatic interactions with a cut-off of 12 Å and the covalent bonds involving hydrogen atoms were constrained using the LINCS algorithm[50]. Temperature and pressure coupling (1 bar) of the system was done using a v-rescale thermostat[51] and Parrinello-Rahman barostat[52], respectively. Energy minimization of the system was first performed using the steepest descent method. The leap-frog integrator algorithm with a time step of 2 fs was used to integrate Newton's equations of motion in the isothermal-isobaric (NPT) ensemble. Trajectories of 1 μs were generated at each temperature, amounting to a cumulative simulation time of 33 μs. We note that the volume of the box remains stable over the last 0.95 μs of each simulation trajectory.

3.3.2 Trajectory Analysis

3.3.2.1 Structural Persistence Parameter

The structural persistence parameter, P, characterizing the secondary structure content of a protein, relative to a reference structure, is defined as:

$$P = \frac{1}{N_{res}} \sum_{j=1}^{N_{res}} e^{-\left(\frac{\Delta\phi_j}{\Delta\phi_{max}}\right)} \cdot e^{-\left(\frac{\Delta\psi_j}{\Delta\psi_{max}}\right)} \tag{3.1}$$

Here $(\Delta\phi_j, \Delta\psi_j)$ represent the magnitude of the changes in backbone torsion angles over the crystal structure; $(\Delta\phi_{max}, \Delta\psi_{max})$ represent the magnitudes of the maximum changes allowed in the dihedral angles in the Ramachandran space ignoring the direction of rotation; N_{res} represents number of residues. P represents the extent of secondary structural persistence of the original conformation; a value of 1.0 is indicative of a fully unchanged secondary structure, while lower values denote decreased overall secondary structural persistence of the reference conformation.

3.3.2.2 Contact Area

The contact area of the inner hydrophobic core of the oligomer is calculated as:

$$CA = \sum_{i=1}^{n} SASA_{hphob-res} - SASA_{core} \tag{3.2}$$

Here, $SASA_{hphob-res}$ represents the solvent accessible surface area (SASA) of the each of the individual 'n' hydrophobic residues of the core and $SASA_{core}$ is the SASA of the entire hydrophobic core. The SASA is calculated within the VMD package[53] with a spherical probe of diameter 1.4 Å.

3.3.2.3   Radius of Gyration

The radius of gyration of the oligomer was calculated using the *gmx gyrate* tool of Gromacs.

3.3.2.4   Free Energy

The free energy change ($\Delta G^{CA}(T)$) due to distortion of core packing of the oligomer interior at a particular temperature T is calculated as:

$$\Delta G^{CA}(T) = G_D(T) - G_N(T) = -RT \ln\left(\frac{1 - x(T)}{x(T)}\right) \tag{3.3}$$

where $x(T)$ is the fraction of the oligomer in the native conformational state (N) at temperature T. A minimum change of 10% over crystal structure values represents the distorted (D) states at T. This cutoff was chosen based on analysis of the probability distribution of the hydrophobic contact area at 300 K (see Figure 3.2).
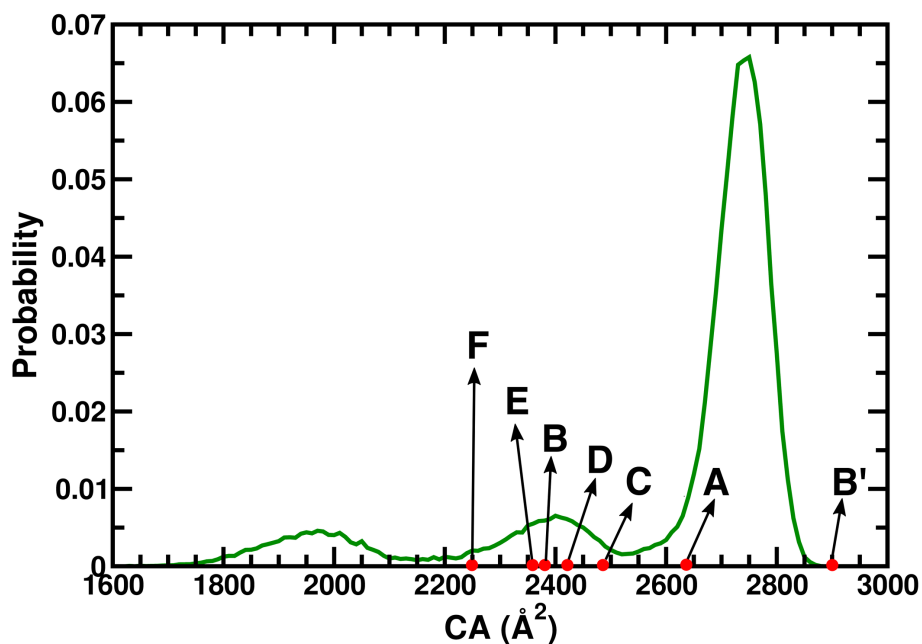
Figure 3.2 Probability distribution of the contact area (CA) obtained from the simulation trajectory at 300 K. Point A represents the crystal structure contact area (2648.0 Å^2). Points 'B' and 'B''' correspond to ±10% change in CA from the crystal structure value. A 10% cutoff covers the entire dominant mode in the distribution. Points 'C' (6% change) and 'D' (8.5% change) represent positions of the first minimum and the first maximum, respectively, towards the lower edge of the dominant mode. Points 'E' and 'F' represent 11% and 15% change over the crystal structure value, respectively.

3.3.2.5 Tetrahedral Order Parameter

The structural order of water molecules can be quantified using the tetrahedral order parameter[54] (q) that is given by:

$$q = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos\theta_{jk} + \frac{1}{3} \right)^2 \tag{3.4}$$

Here, θ_{jk} represents the angle formed between the oxygen atom of a given water molecule and the O–O bond vectors r_{ij} and r_{ik}, of its nearest neighbors, j and k

(j, $k \leq 4$). The average value of q varies between 0 and 1; the latter corresponding to perfectly tetrahedral ice.

3.3.2.6 Water-mediated Inter-residue Contacts

A water-mediated inter-residue contact is defined as a contact formed between a pair of residues that are hydrogen bonded to a common water molecule. The donor-acceptor distance threshold for a hydrogen bond was considered as 3.5 Å and the angle connecting the donor, hydrogen and the acceptor atoms did not exceed 30°.

3.4 Results and Discussion

3.4.1 Structural Response to Temperature Changes

We begin by evaluating conformational transitions from the initial state at varying thermal regimes. Figure 3.3 depicts the backbone root-mean-square-deviation (RMSD) from the crystal structure at representative temperatures spanning the low to high regime. At the lowest temperature of 210 K, the RMSD undergoes only a slow increase over the simulation period and saturates to about 1.4 Å at 0.8 μs, suggesting that the crystal structure-like conformation is largely preserved at these conditions. The conformational fluctuations undergo an overall increase with rising temperature. In the physiological regime, the RMSD fluctuates in the range of 1.8–1.9 Å throughout the simulation timescale. The high temperature regime is marked by large periodic deviations from the crystal structure over time spans of 100–200 ns. The highest simulation temperature of 483 K brings about dissociation in the oligomeric assembly at 0.3 μs; we have not considered the dissociated state in the analyses below.
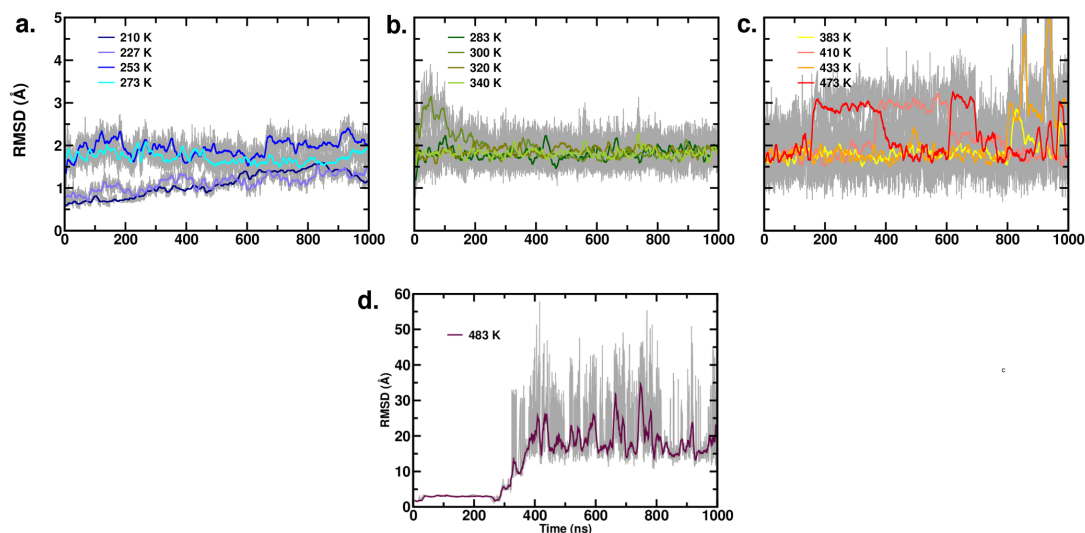
Figure 3.3 Time evolution of backbone RMSD for representative temperatures in the (a) low, (b) medium and (c and d) high temperature regime shown as running averages, from the starting structure.
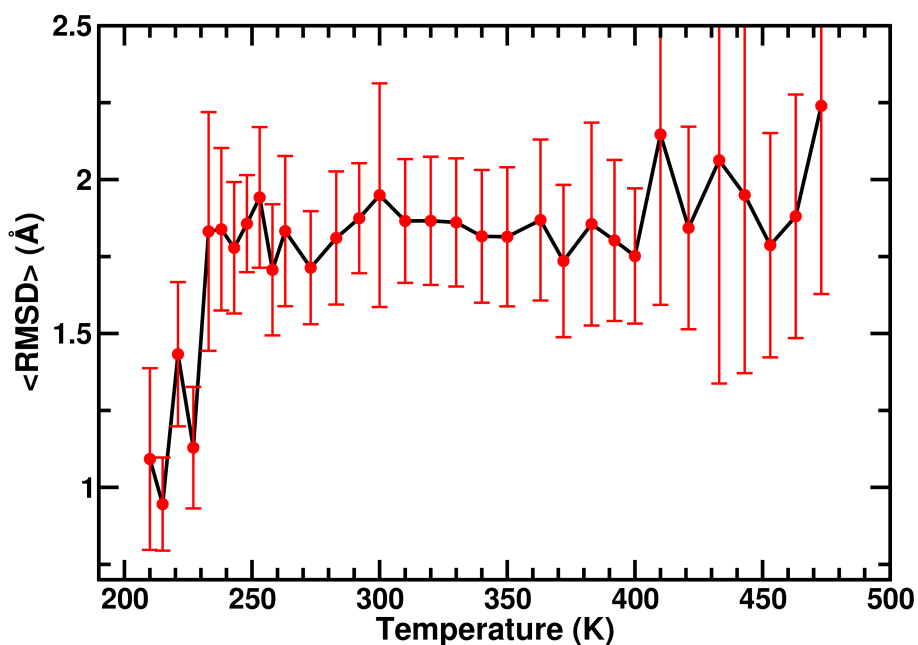
Figure 3.4 Thermal trend of the mean values of the backbone RMSD of the oligomer calculated over simulation trajectories at individual temperatures. Standard deviations are shown as error bars.

No.	Temperature (K)	<RMSD> (Å)
1	210	1.3
2	215	1.06
3	221	1.66
4	227	1.36
5	233	1.96
6	238	1.89
7	243	1.94
8	248	1.89
9	253	2.17
10	258	1.82
11	263	1.76
12	273	1.80
13	283	1.86
14	292	1.89
15	300	1.92
16	310	1.86
17	320	1.85
18	330	1.83
19	340	1.81
20	350	1.77
21	363	1.77

No.	Temperature (K)	<RMSD> (Å)
22	372	1.74
23	383	1.97
24	392	1.81
25	400	1.82
26	410	1.74
27	421	1.86
28	433	3.26
29	443	1.70
30	453	1.68
31	463	1.79
32	473	2.20
33	483	17.89

Table 3.1 List of simulation temperatures and the mean RMSD for the last 100 ns of the simulation trajectories

Our preliminary observations contrast the usual cold response of folded globular proteins, where an increase in RMSD at low temperatures signals cold denaturation and loss of native conformational states[38-39]. We characterized the extent of secondary structure persistence of the oligomer at varying temperatures from the structural persistence parameter defined by Equation 3.1. The $<P>$ vs. T plot, along with representative snapshots spanning the thermal regime, are depicted in Figure 3.5. Consistent with the trends of RMSD, structural persistence is highest at the lowest temperatures of 210 and 215 K. With increase in temperature, a sharp drop in $<P>$ to

a value of 0.76 is observed up to 263 K; it thereafter remains stable till 473 K. The trend in the <*P*> value indicates that the regime commencing with cooling below ~250 K marks important structural transitions in the structural evolution of the system. In this structural "*transition regime*", the oligomeric stability rises almost linearly with lowering of temperature. This transition regime coincides with the transition observed in the thermal trend of the backbone RMSD, as reported in Figure 3.4. Collectively, these observations indicate that the global structural integrity of this amyloid system is susceptible to increasing distortion with heat, but relatively resistant to denaturation in the low temperature regime; the latter appears to contrast the cold denaturation phenomenon of globular proteins[38-39].
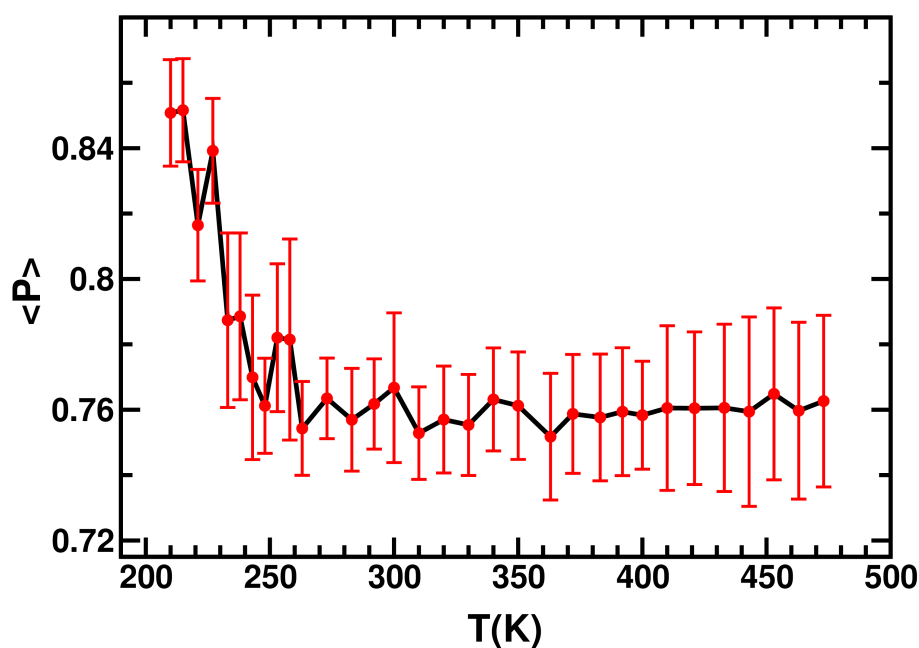
Figure 3.5 Thermal trend of <*P*> with standard deviations shown as error bars.

We further assessed the finer architectural properties to identify the effect of temperature on two key aspects of the system's stability, namely hydrophobic

associations and hydrogen-bonding interactions. The β-barrel oligomeric stability may therefore be attributed to a combination of thermodynamic factors associated with both these features. The mid-axis region, situated ~11 Å from the axis ends, is characterized by the presence of 6 glycine residues. At the ends of the barrel, two sets of leucine and valine triplets, with inward pointing side chains are located ~10 Å from the mid-axis region; leucine triplet is present on the outer side. Hydrophobic interactions between the side chains of the valine triplets close the axial channel and keep it dry. In the text below, we refer to the valine triplets on either ends of the barrel as the "*gate*", and the mid-axial region as the "*waist*" of the barrel (Figure 3.6).
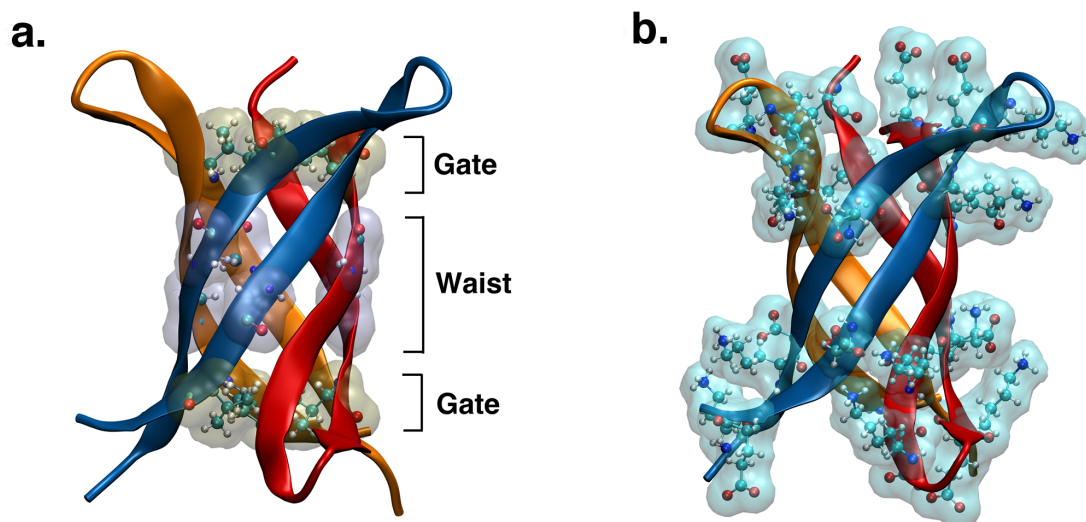
Figure 3.6. (a) The residues in the gate and waist region are shown as atoms in CPK representation with a transparent surface (valine in green, glycine in blue). (b) Solvent facing hydrophilic residues are shown as atoms with a cyan colored transparent surface representation.
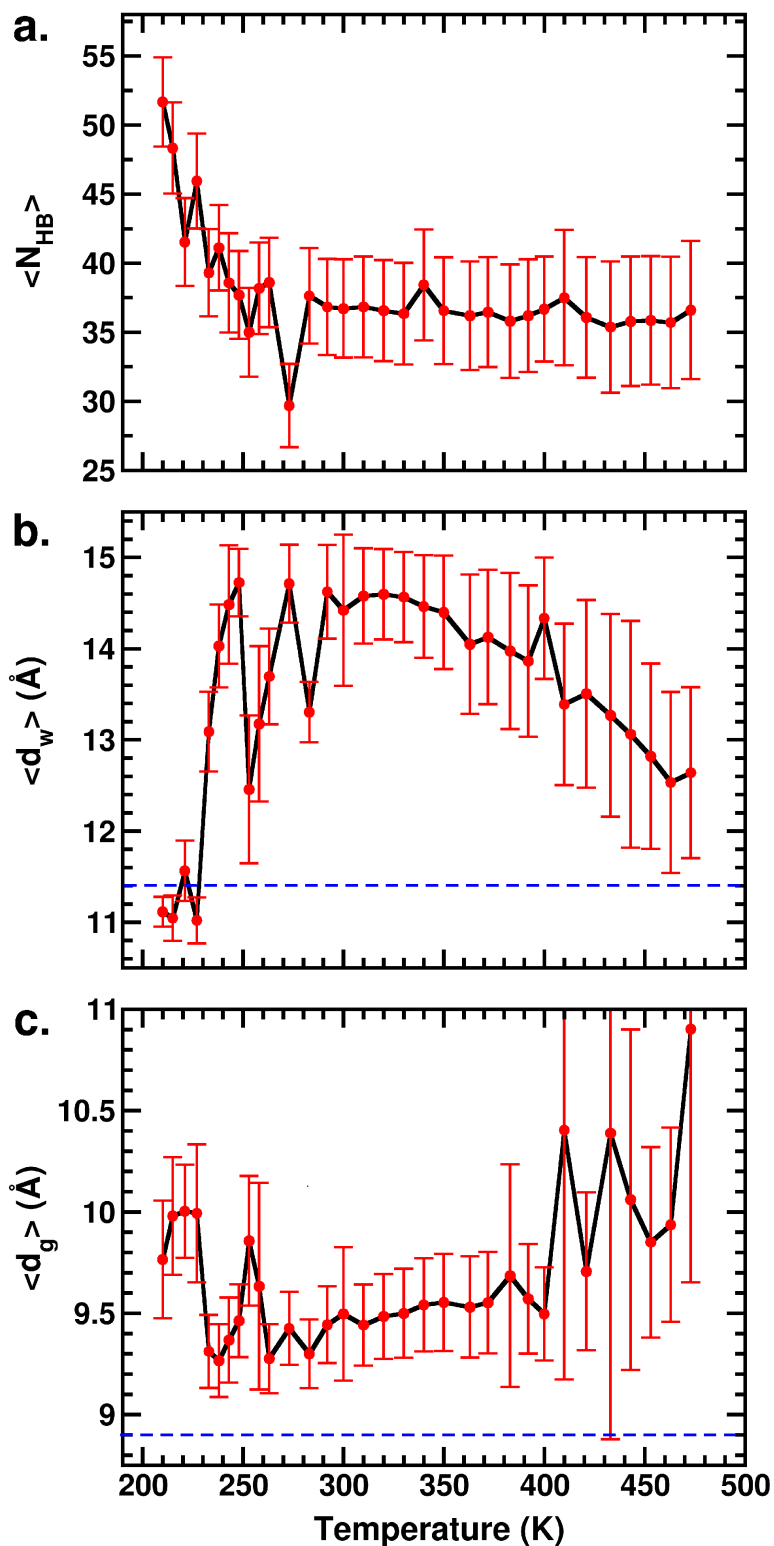
Figure 3.7 Thermal trend of the average value of (a) number of inter-residue hydrogen bonds (b) barrel waist diameter (d_w) and (c) gate diameter (d_g) for the last 800 ns of the trajectories. Standard deviations are shown as error bars.

As depicted in Figure 3.7, the total number of interfacial residue-residue hydrogen bonds within the oligomer shows an overall decline with increasing temperature in the transition regime and then attains saturation at higher temperatures. In Figure 3.7b and c, we have compared temperature dependence of the average pore waist ($<d_w>$) and the average gate (d_g) diameters. The $<d_w>$ is lowest and close to the crystal structure value in the low temperature regime up till ~227 K, followed by a rapid increase by ~2–2.5 Å with marked fluctuations in the transition regime, and is stable over physiological temperatures. As the high temperature regime is reached, a nearly smooth decrease with temperature is observed. It is crucial to note that the general thermal trend of the mean gate diameter, $<d_g>$, is opposite to that of $<d_w>$. While the lowest value of $<d_g>$ is achieved below 273 K, there is an overall rise in this value with the lowering of temperature till 210 K. The low temperature trend is marked by strong fluctuations, which commences as the transition regime of the P-value plot. The $<d_g>$ remains stable in the physiological regime and increases rapidly, accompanied by large fluctuations, in the high temperature regime. The pore dimension at the gate ends increase by ~1 Å at low temperatures with respect to the crystal structure, while it increases by ~0.5 Å at intermediate temperatures.

Lack of convergence in the trends of structural parameters suggests an intriguing interplay of factors towards the overall structural robustness of the amyloid-like β-barrel oligomer over a large temperature range exceeding 230 K that borders on the cryo as well as the high temperature melting regime. Interestingly, despite indications of structural robustness, the transition regime is commensurate with the beginning of high fluctuations in the low temperature regimes. Since the hydrophobic interactions of residues at the gate ends keep the inner core dry, change in dimensions of the gate end at low temperature signals that cooling may affect the

packing and hydration of the barrel interior. The gate widening at high temperatures, on the other hand, could possibly arise from increasing thermal disorder. Thus, while the gross structural changes indicate that withdrawal of thermal energy preserves the structural integrity of the oligomer, subtle yet key structural deviations hint possible cold denaturation events. These observations led us to further investigations as described in subsequent sections.

3.4.2   Packing of the Hydrophobic Core and Hydration of the Barrel

A key driving force underpinning globular protein stability is the sequestration of the hydrophobic core from water. The hydrophobic effect is also important in protein self-assembly and amyloid fibrillogenesis; however, other non-bonded interactions, including hydrogen bonding and electrostatic interactions, alter the overall balance. Hansmann *et al.*[55] recently highlighted the role of hydrophobic interactions within the barrel interior on the stability of the oligomer's cylindrin geometry. Herein, we proceed to investigate the extent to which temperature affects hydrophobic packing in the oligomer. We quantified the packing of the inner core by calculating the contact area (Equation 3.2) of the hydrophobic residues whose side chains are packed to form the core of the barrel, depicted in Figure 3.8. The mean core contact area drops sharply below 260 K, by about 200 $Å^2$ from that in the crystal structure (~2647 $Å^2$). This indicates loosening of the sterically packed side chains of hydrophobic residues in the low temperature range studied. In the physiological regime of temperatures, the mean contact area remains stable at ~50 $Å^2$ higher than the crystal structure value. At high temperatures, the contact area falls smoothly below the reference value, also indicating a perturbed packing of the core. Interestingly, we observe that the thermal

trend of the contact area reflects that of the mean pore waist diameter (Figure 3.7b). This suggests that the relatively tighter packing of the hydrophobic core at physiological temperatures causes an expansion of the barrel in the waist region. In the waist region, glycine residues (G7 and G20) are involved in main chain hydrogen bonds with core forming valines (V5 and V18), located in the adjacent β-hairpin i.e. at the weak interface of interactions. We note that the tighter packing of the core residues at physiological temperatures reduces the propensity of these glycine residues to form hydrogen bonds with the neighboring valines at the weak interface, which renders them more flexible. The increased flexibility at the center of the β-strands thus introduces a bulge at the waist of the barrel that leads to an increase in central diameter of the pore observed at these temperatures.
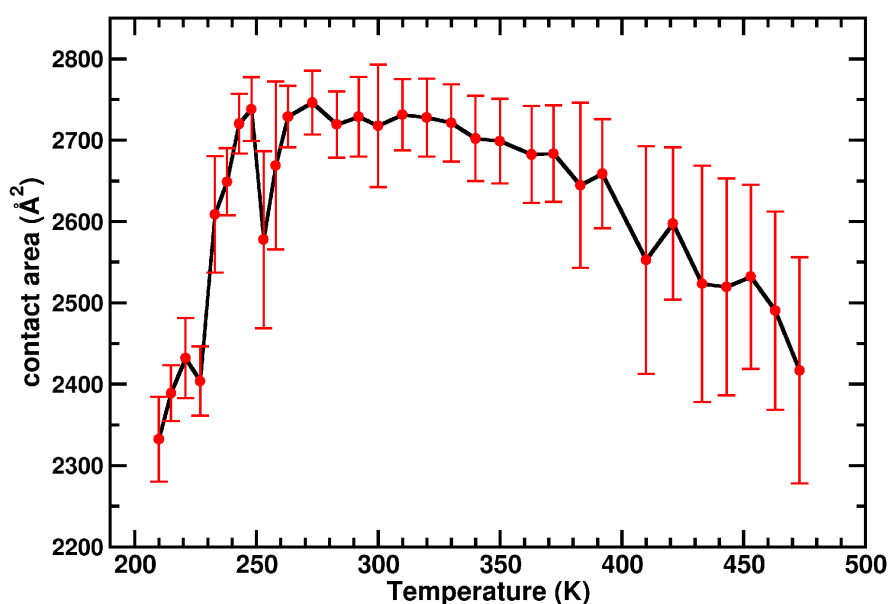
Figure 3.8 Thermal trend of the total contact area of the hydrophobic core residues for the last 800 ns of the trajectories. Standard deviations are shown as error bars.

Figure 3.9 depicts the distributions of solvent accessible surface area (SASA) of the conformations explored at representative temperatures from the low, intermediate and high temperature range; the thermal trend of average value of SASA is provided in Figure 3.10. Upon cooling, the oligomer exhibits a distinct shift in SASA towards higher values indicating enhanced hydration. The low temperature regime below the transition regime is characterized by sharp rises and fall in solvation of the protein. The spike in solvation appears at 227 K with a mean value of $\sim$5600 Å^2 which then plummets upon cooling or heating. A second small peak in <SASA> is observed at 253 and 258 K with mean values of $\sim$5400 Å^2. It is rather interesting to note that this trend of SASA at low temperatures is consistent with the trend of $<d_g>$ (Figure 3.7c) indicating that the increased hydration of the barrel may be partly attributed to the perturbed arrangement of residues at the gate ends. The <SASA> fluctuates closely, around 5300 Å^2 at physiological and high temperatures; the high temperature ensembles however, exhibit greater deviations in solvation as indicated by their broader distribution of SASA (Figure 3.9).
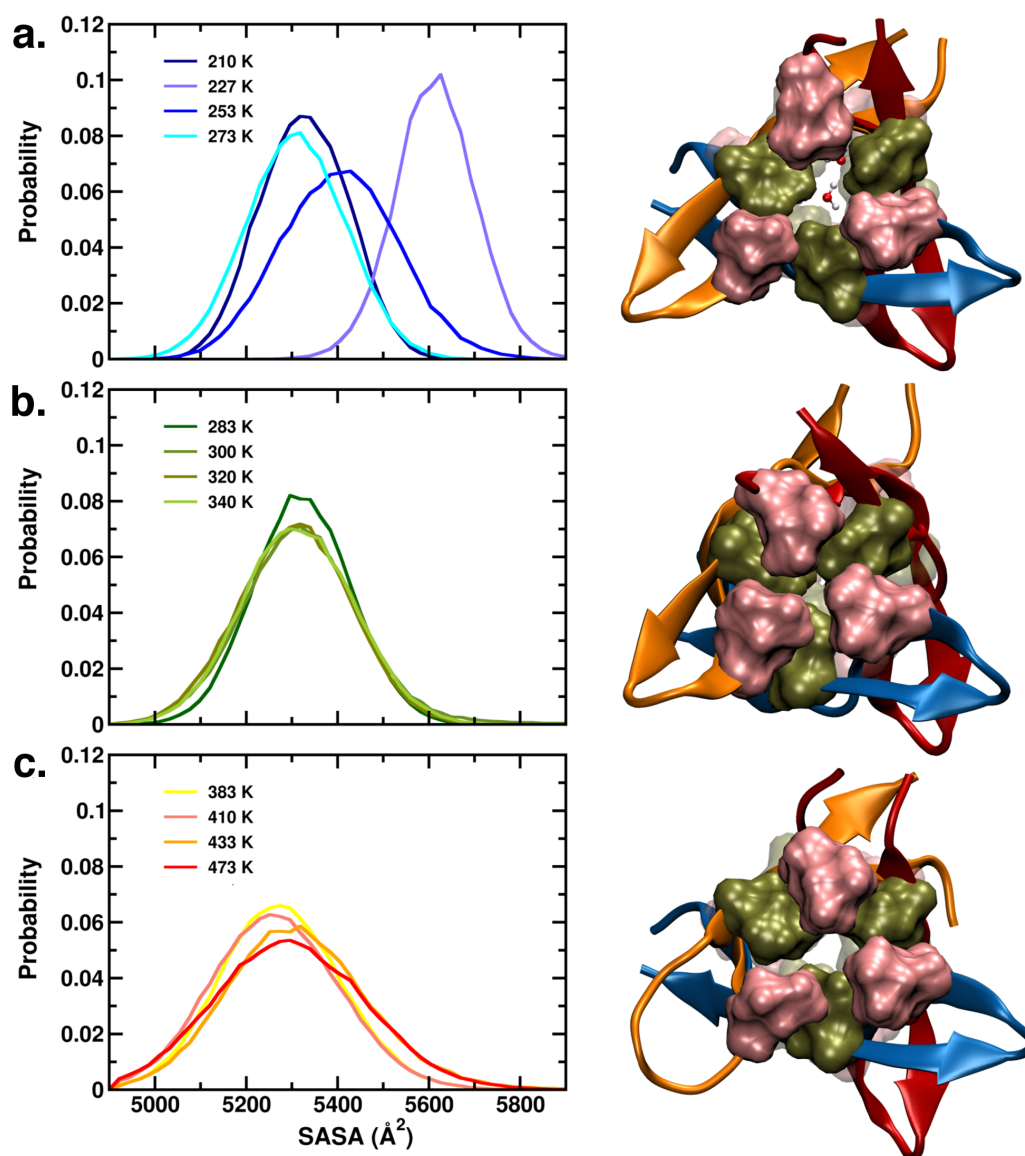
Figure 3.9 Probability distributions of SASA of the oligomer at representative temperatures in the (a) low, (b) physiological and (c) high temperature regime. Snapshots of the oligomer at one representative temperature in each regime (low: 210 K, physiological: 300 K and high: 473 K) are provided with surface representations of leucine and valine residues at the gate ends of the oligomer shown in pink and green, respectively. Waters inside the core are shown in CPK representation.
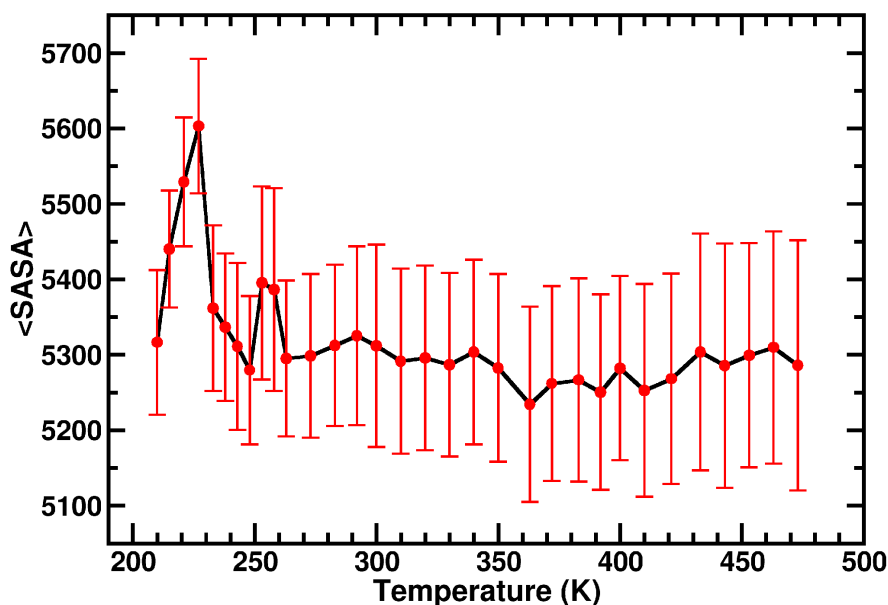
Figure 3.10 Thermal trend of average SASA of the oligomer for the last 800 ns of the trajectories. Standard deviations are shown as error bars.

3.4.3 Thermal Response of the Hydrophobic Core

External perturbations that shift the balance of non-bonded forces (primarily hydrophobic, electrostatic and van der Waals) in proteins and protein assemblies can alter their energy landscapes. In folded proteins, however, the hydrophobic effect emerges as a predominant factor that is attributed to a large fraction of the overall free energy change of folding; indeed, this appears consistent with the temperature dependent unfolding curve of globular proteins (discussed previously). Amyloid aggregates, on the other hand, are kinetically stable entities characterized by ordered hydrogen bonded networks, though hydrophobic forces are understood to play crucial roles, especially during the early events of aggregation[31].

We have leveraged the unique β-barrel architecture of ABC with a distinct hydrophobic interior and our large simulation ensembles to delineate and compare the

key non-bonded effects, and their relative contributions to the net thermostability of the system. Temperature induced unfolding in small globular proteins has been well approximated as a two-state transition, with the folded and the denatured states equally populated at the transition temperatures[6-8,56]. In case of the oligomer studied here, we designated the two states as "*native*" (*N*) and partial "*distorted*" (*D*); classification of the '*D*' states relative to the crystal structure is described in Methods. We used Boltzmann weighting of the equilibrated trajectory data to obtain temperature dependence of free energy of distortion of core packing of the oligomer interior, quantified from the contact area (CA) of the hydrophobic residues.
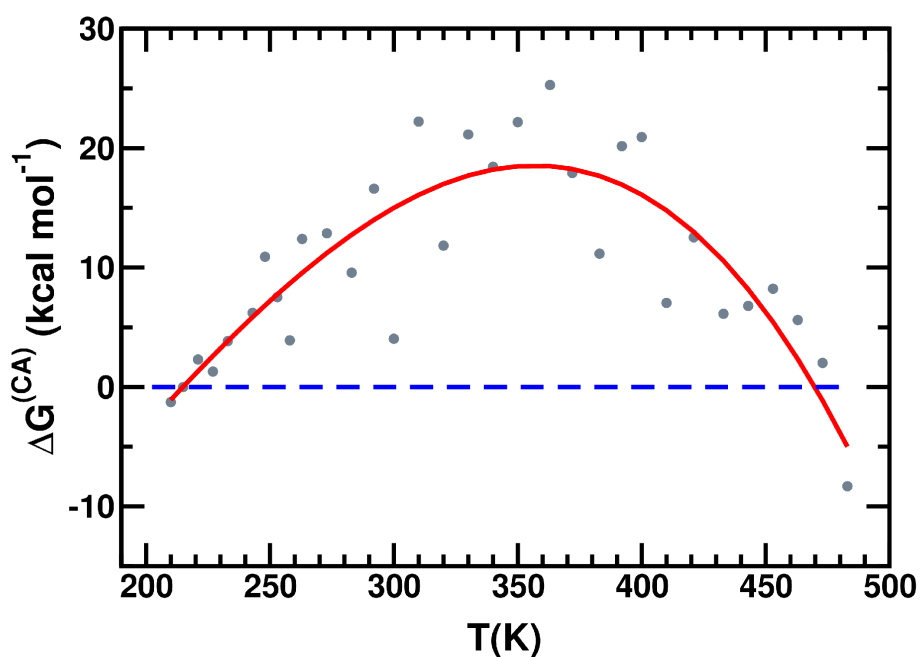
Figure 3.11 $\Delta G^{(CA)}$ profile generated as a function of temperature

Figure 3.11 depicts the free energy change ($\Delta G^{(CA)}$) for distortion of the native like-oligomer with temperature. A polynomial fit yields a bell-shaped stability trend reminiscent of the overall free energy response of globular proteins. For several

globular proteins, the folded state is usually stabilized over unfolded state under physiological conditions by 10–30 kcal mol^{-1} [12,39,57]. In the case of the amyloid system studied here, the largest value of $\Delta G^{(CA)}$ (~25.3 kcal mol^{-1}) is obtained at ~363 K, corroborating the exceptional high temperature structural robustness of the oligomer. The $\Delta G^{CA}(T)$ profile depicts a growing distortion of the hydrophobic core with lowering of temperature. $\Delta G^{CA}(T)$ is 0 at a low temperature of ~215 K (T_{low}), and a high temperature of ~463 K (T_{high}). However, as discussed earlier, the distortion of the core is not reflected significantly in the overall structural changes (see Figure 3.3), plausibly owing to the compensating effects of the intra- and inter-peptide hydrogen bonds. We note here the markedly dissimilar trends in $<P(T)>$ and $\Delta G^{CA}(T)$; this indicates that unlike globular proteins, the weakened hydrophobic packing at very low temperatures is inadequate for destabilizing amyloid oligomeric structure. It also suggests that the overall free energy profile of an amyloid oligomer should be dissimilar to that of globular proteins. In particular, the amalgamation of different non-bonded effects in the amyloid oligomer gives rise to marked thermostability, and may be expected to confer complex topologies to the multi-dimensional free energy landscapes.

3.4.4 Amyloid Hydration Modulates Low Temperature Fluctuations

Prima-facie, the previous results suggest no important associations between the local solvent environment and the system's thermal stability. However, we note that the sharp inflexion in $P(T)$ occurs close to the melting temperature of the water model at 0.1 MPa pressure[46]. To unravel underlying hydration effects, we evaluated the temperature dependence of the energetic coupling between the oligomer's

stabilization and its first solvation layer (defined as water molecules within 4.5 Å of the heavy atoms). In panels a and b of Figures 3.12, we report the mean protein internal energy difference and mean protein–water interaction energy difference over the corresponding value at 300 K as:

$$\Delta E_a(T) = \left\langle E_a(T) - E_a(300) \right\rangle \tag{3.5}$$

$$\Delta E_{aw}(T) = \left\langle E_{aw}(T) - E_{aw}(300) \right\rangle \tag{3.6}$$

Here, $E_a(T)$ and $E_{aw}(T)$ are the oligomer internal energy and its interaction with the hydration layer at temperature T, respectively.

A decrease (increase) in energy is indicative of enhanced (reduced) interaction strength. We note that up to physiological temperatures, both $\Delta E_a(T)$ and $\Delta E_{aw}(T)$ decrease smoothly with the lowering of temperature. This indicates progressive strengthening of the oligomer, and reduced interaction of the protein with hydration waters with reduction in temperature up to this regime. Interestingly, both trends are disrupted and characterized by greater fluctuations as the system shifts to the cold structural transition regime. While the oligomer is further strengthened upon cooling, its interaction with the hydration waters shows a sharp reversal in trend. The oligomer exhibits strongest interaction with hydration waters at ~253K; the interaction trails off as the temperature is decreased further. These observations strongly suggest that stability of the amyloid oligomer at low temperatures is crucially linked with interactions associated with the hydration layer.
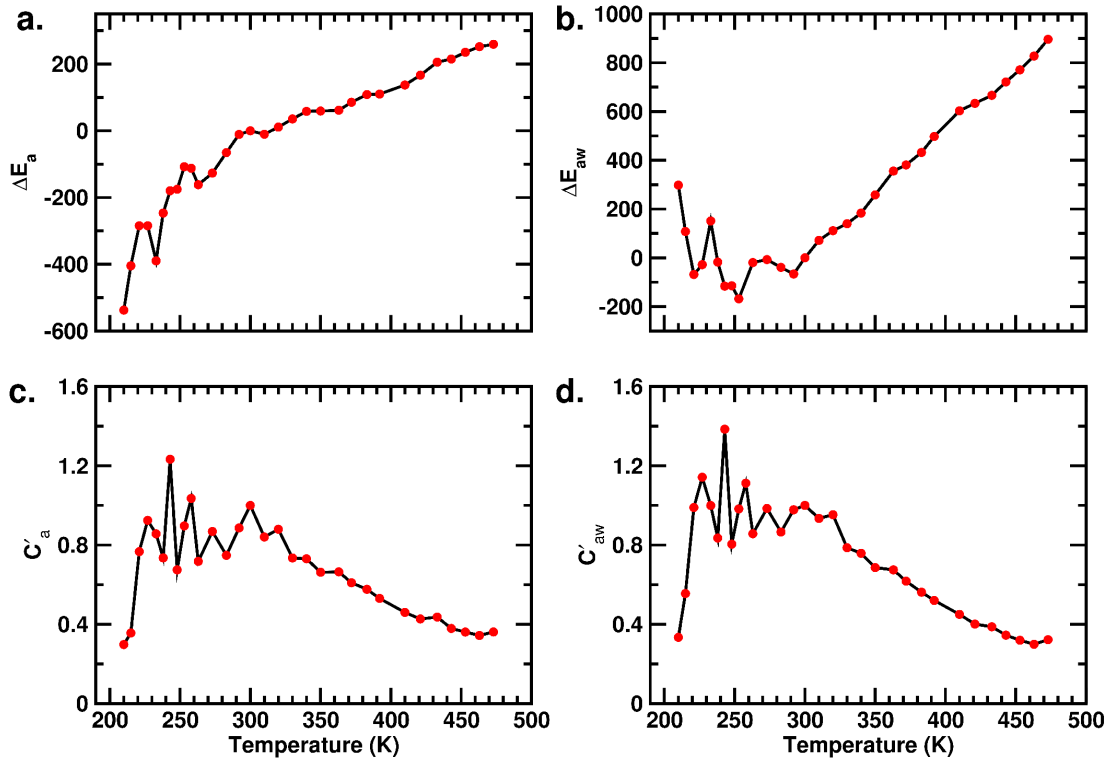
Figure 3.12 Thermal trends of (a) ΔE_a , (b) ΔE_{aw}, (c) C'_a and (d) C'_{aw} . See text for details.

The heat capacity of globular proteins is characterized by a sharp peak at the high temperature melting transition, accompanied by a net increase[15-16]. Experimental studies as well as more recent theoretical models suggest a peak in partial heat capacity of globular proteins at the cold denaturation event as well[6-7,35].

The isochoric heat capacity of an equilibrated system at temperature T is related to the energetic fluctuation (variance) as:

$$C = \frac{\left\langle \delta E^2 \right\rangle}{k_B T^2} \tag{3.7}$$

We calculated the following ratios over 300 K values using 0.8 μs of equilibrated data:

$$C'_a = \frac{\left\langle \left[\delta E_a(T)\right]^2 \right\rangle}{\left\langle \left[\delta E_a(300)\right]^2 \right\rangle} \left(\frac{T_{300}}{T}\right)^2 \tag{3.8}$$

$$C'_{aw} = \frac{\left\langle \left[\delta E_{aw}(T)\right]^2 \right\rangle}{\left\langle \left[\delta E_{aw}(300)\right]^2 \right\rangle} \left(\frac{T_{300}}{T}\right)^2 \tag{3.9}$$

These quantities are plotted as a function of temperature in panels c and d of Figure 3.12. It is revealed that $C'_a(T)$ and $C'_{aw}(T)$ are nearly identical in their thermal trends despite the apparently dissimilar temperature dependence of the corresponding energetics. Importantly, both $C'_a(T)$ and $C'_{aw}(T)$ display a rise with lowering of temperature, before a sharp fall commencing at ~227 K. This surprising result indicates an intimate, temperature dependent coupling between the hydration layer and the overall amyloid energetics with clear low temperature manifestations. We note that the broad, hydration influenced peaks in $C'_a(T)$ and $C'_{aw}(T)$ occur at temperatures close to the heat capacity peaks reported for bulk TIP4P/2005 water[58].

It is well known that water diverges from its bulk characteristics in the vicinity of biomolecular solutes[59-61]. Globular protein unfolding is known to be accompanied by structural reordering of the hydration layer[62-63]. Our results thus far clearly implicate the hydration layer in the complex thermal response of the barrel oligomer. We therefore investigated the oligomer hydration layer's temperature dependent structural response by evaluating the tetrahedral order parameter (q) introduced by Errington and Debenedetti[54].
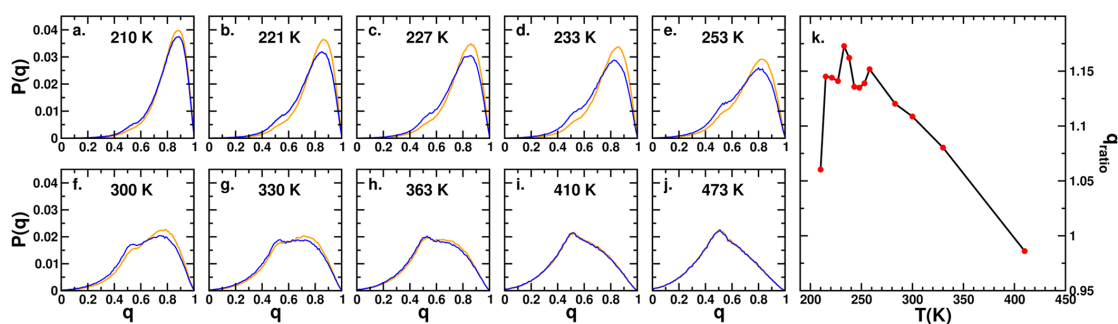
Figure 3.13 (a-j) Normalized probability density distributions of the tetrahedral order parameter (q) of hydration water (orange) and bulk water (blue) for 10 representative temperatures; (k) population ratio at q_{high} for hydration layer to bulk waters. See text for details.

Normalized probability distributions of the q-value calculated with 0.8 µs of equilibrated data are compared for hydration layer and bulk water at various temperatures in Figure 3.13a–j. We note the presence of two peaks that occur at q_{high} of ~0.8 and q_{low} of ~0.55 for bulk water at physiological temperatures (300 K). For the oligomer's hydration layer at 300 K, there is a noticeable decrease (increase) in the peak height at q_{low} (q_{high}) indicating its enhanced structural ordering over bulk, consistent with earlier reports[64]. In the high temperature regime, relative populations at q_{high} and q_{low} decrease and increase, respectively. Structural differences between bulk and hydration waters diminish with increasing temperature, and vanish from 363 K onwards. Using the normalized distributions, we estimated the hydration layer vs. bulk population ratios (q_{ratio}) of the peak heights obtained at q_{high}. An interesting water structural ordering over bulk, distinct from the high temperature regime, emerges upon cooling. With temperatures decreased below 300 K, q_{ratio} consistently increases over 1.0 within the temperature range that corresponds to peak positions of $C'_a(T)$ and $C'_{aw}(T)$ (see Figure 3.13k). As the temperature is lowered further, there is a

113

sharp fall in q_{ratio}. Thus, the maximum structural ordering of the hydration layer over corresponding bulk water is directly associated with fluctuations in the assembly-water interactions. These analyses provide corroborating evidence of solvent coupling to the thermal response of protein oligomers.

3.5   Conclusions

This exposition reveals molecular phenomena underlying differential cold response of amyloid oligomers and folded globular proteins. The enhanced structural integrity with lowering of temperature supports the experimental observation of greater resistance of amyloid cold degeneration. Unlike globular proteins, the presence of a hydrophobic core in the amyloid assembly is insufficient for triggering low temperature destabilization. However, temperature dependent energy fluctuations within the hydration layer are expected to have strong modulatory effects on the thermodynamic response. We note here previous work establishing the intimate thermal association between temperature driven protein conformational fluctuations and hydration dynamics[65]. In this light, we note that cooling results in slight increase in the number of water-mediated inter-residue contacts (Figure 3.14). Further detailed studies are required to unravel their nature and lifetimes, and possible associations with the structural transitions.
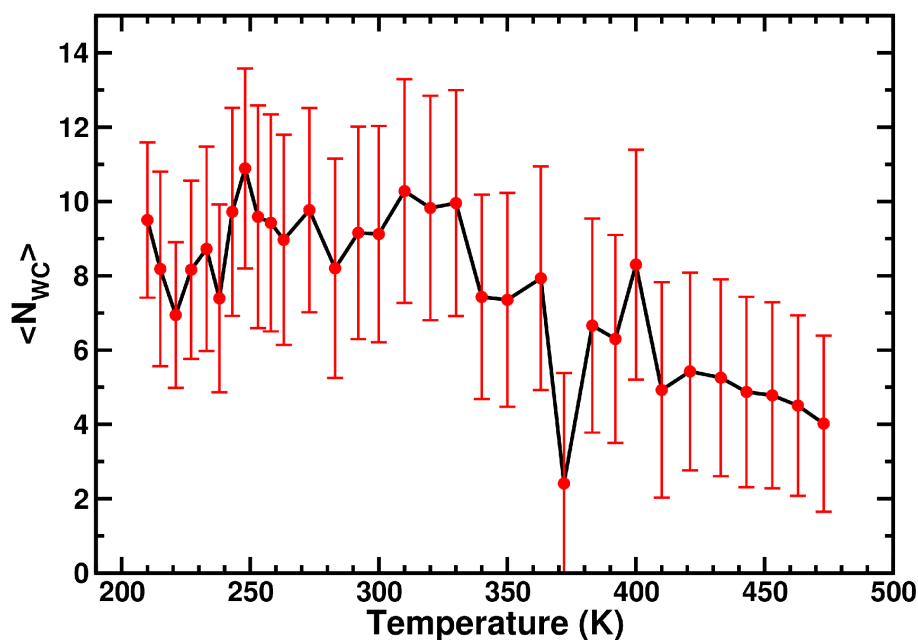
Figure 3.14 Thermal trend of the mean number of water-mediated inter-residue contacts, $\langle N_{WC} \rangle$, over the trajectories simulated at the individual temperatures. Standard deviations are shown as error bars.

Our study is a first direct demonstration of the critical importance of temperature dependent microscopic influence of water on amyloidogenic conformation and stability that goes beyond its roles in sequestering hydrophobic contacts. The association of hydration water with the divergent behavior of folded proteins, strands and fibrillar assemblies to low temperature has important bearings. It is currently believed that water undergoes a liquid-liquid phase transition (LLPT) between high- and a low-density form at extreme cooling and at high pressure[58,66]. This phenomenon onsets below 180 K and at pressures orders of magnitude higher than atmospheric pressure for the TIP4P/2005 water model[58]. The large structural and thermodynamic changes thus brought about can be expected to sharply influence water-protein associations. We note that recent investigations implicate the barrel-shaped cylindrin motif as an on-pathway, toxic motif encountered by a number of

disease-related amyloidogenic proteins[41-42,67]. In this light, the observations reported in this study can be expected to represent general features of the cold response of crucial, on-pathway toxic amyloid oligomers. However, further detailed studies are necessary to delineate any sequence-specific differences. Investing efforts on establishing collective variables that distinguish the complex energy landscapes of amyloid assemblies vis-à-vis protein thermal response should be useful. This could potentially arise from the powerful energy landscape formalisms currently employed for amyloid assembly pathways[68]. Further studies of the interplay of thermodynamic conditions, solvent state, hydrophobicity and protein conformations are necessary for improved understanding and also for applications ranging from cryo-preservation to studying the plausibility of extreme life forms.

3.6 References

1. Franks, F.; Hatley, R. H. Stability of proteins at subzero temperatures: thermodynamics and some ecological consequences. *Pure Appl. Chem.* **1991,** *63* (10), 1367–1380.

2. Jaenicke, R. Protein stability and molecular adaptation to extreme conditons. *Eur. J. Biochem.* **1991,** *202* (3), 715–728.

3. Boothby, T. C.; Tapia, H.; Brozena, A. H.; Piszkiewicz, S.; Smith, A. E.; Giovannini, I.; Rebecchi, L.; Pielak, G. J.; Koshland, D.; Goldstein, B. Tardigrades Use Intrinsically Disordered Proteins to Survive Desiccation. *Mol. Cell* **2017,** *65* (6), 975–984.e5.

4. Rothschild, L. J.; Mancinelli, R. L. Life in extreme environments. *Nature* **2001,** *409*, 1092–1101.

5. Hubel, A.; Skubitz, A. P. N. Principles of Cryopreservation. In *Biobanking of Human Biospecimens*, Springer: **2017**, 1–21.

6. Privalov, P. L.; Griko, Y. V.; Venyaminov, S. Y.; Kutyshenko, V. P. Cold denaturation of myoglobin. *J. Mol. Biol.* **1986,** *190* (3), 487–498.

7. Griko, Y. V.; Privalov, P. L.; Sturtevant, J. M.; Syu, V. Cold denaturation of staphylococcal nuclease. *Proc. Natl. Acad. Sci. U.S.A.* **1988,** *85* (10), 3343–3347.

8. Franks, F.; Hatley, R. H. M.; Friedman, H. L. The thermodynamics of protein stability: Cold destabilization as a general phenomenon. *Biophys. Chem.* **1988,** *31* (3), 307–315.

9. Agashe, V. R.; Udgaonkar, J. B. Thermodynamics of denaturation of barstar: evidence for cold denaturation and evaluation of the interaction with guanidine hydrochloride. *Biochemistry* **1995,** *34* (10), 3286–3299.

10. Sanfelice, D.; Morandi, E.; Pastore, A.; Niccolai, N.; Temussi, P. A. Cold Denaturation Unveiled: Molecular Mechanism of the Asymmetric Unfolding of Yeast Frataxin. *ChemPhysChem* **2015,** *16* (17), 3599–3602.

11. Rösner, H. I.; Caldarini, M.; Prestel, A.; Vanoni, M. A.; Broglia, R. A.; Aliverti, A.; Tiana, G.; Kragelund, B. B. Cold Denaturation of the HIV-1 Protease Monomer. *Biochemistry* **2017,** *56* (8), 1029–1032.

12. Privalov, P. L. Stability of Proteins Small Globular Proteins. In *Advances in Protein Chemistry* Academic Press: **1979,** *33*, 167–241.

13. Adrover, M.; Esposito, V.; Martorell, G.; Pastore, A.; Temussi, P. A. Understanding Cold Denaturation: The Case Study of Yfh1. *J. Am. Chem. Soc.* **2010,** *132* (45), 16240–16246.

14. Buchner, G. S.; Shih, N.; Reece, A. E.; Niebling, S.; Kubelka, J. Unusual Cold Denaturation of a Small Protein Domain. *Biochemistry* **2012,** *51* (33), 6496–6498.

15. Prabhu, N. V.; Sharp, K. A. Heat capacity in proteins. *Annu. Rev. Phys. Chem.* **2005,** *56*, 521–548.

16. Cooper, A. Protein Heat Capacity: An Anomaly that Maybe Never Was. *J. Phys. Chem. Lett.* **2010,** *1* (22), 3298–3304.

17. Privalov, P. L. Cold denaturation of protein. *Crit. Rev. Biochem. Mol. Biol.* **1990,** *25* (4), 281–306.

18. Pastore, A.; Martin, S. R.; Politou, A.; Kondapalli, K. C.; Stemmler, T.; Temussi, P. A. Unbiased Cold Denaturation: Low- and High-Temperature Unfolding of Yeast Frataxin under Physiological Conditions. *J. Am. Chem. Soc.* **2007,** *129* (17), 5374–5375.

19. Li, Y.; Shan, B.; Raleigh, D. P. The Cold Denatured State Is Compact but Expands at Low Temperatures: Hydrodynamic Properties of the Cold Denatured State of the C-terminal Domain of L9. *J. Mol. Biol.* **2007,** *368* (1), 256–262.

20. Bothe, Jameson R.; Tonelli, M.; Ali, Ibrahim K.; Dai, Z.; Frederick, Ronnie O.; Westler, William M.; Markley, John L. The Complex Energy Landscape of the Protein IscU. *Biophys. J.* **2015,** *109* (5), 1019–1025.

21. Sabelko, J.; Ervin, J.; Gruebele, M. Cold-Denatured Ensemble of Apomyoglobin: Implications for the Early Steps of Folding. *J. Phys. Chem. B* **1998,** *102* (10), 1806–1819.

22. Kim, H.-Y.; Cho, M.-K.; Riedel, D.; Fernandez, C. O.; Zweckstetter, M. Dissociation of Amyloid Fibrils of α-Synuclein in Supercooled Water. *Angew. Chem.* **2008,** *120* (27), 5124–5126.

23. Mishra, R.; Winter, R. Cold- and Pressure-Induced Dissociation of Protein Aggregates and Amyloid Fibrils. *Angew. Chem. Int. Ed.* **2008,** *47* (35), 6518–6521.

24. Ikenoue, T.; Lee, Y.-H.; Kardos, J.; Saiki, M.; Yagi, H.; Kawata, Y.; Goto, Y. Cold Denaturation of α-Synuclein Amyloid Fibrils. *Angew. Chem.* **2014,** *126* (30), 7933–7938.

25. Adamcik, J.; Mezzenga, R. Amyloid Polymorphism in the Protein Folding and Aggregation Energy Landscape. *Angew. Chem. Int. Ed.* **2018,** *57* (28), 8370–8382.

26. Jahn, T. R.; Radford, S. E. The Yin and Yang of protein folding. *FEBS J.* **2005,** *272* (23), 5962–5970.

27.   Routledge, K. E.; Tartaglia, G. G.; Platt, G. W.; Vendruscolo, M.; Radford, S. E. Competition between intramolecular and intermolecular interactions in an amyloid-forming protein. *J. Mol. Biol.* **2009,** *389* (4), 776–786.

28.   Fitzpatrick, A. W.; Knowles, T. P. J.; Waudby, C. A.; Vendruscolo, M.; Dobson, C. M. Inversion of the Balance between Hydrophobic and Hydrogen Bonding Interactions in Protein Folding and Aggregation. *PLoS Comput. Biol.* **2011,** *7* (10), e1002169.

29.   Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990,** *29* (31), 7133–7155.

30.   Spolar, R. S.; Ha, J. H.; Record, M. T. Hydrophobic effect in protein folding and other noncovalent processes involving proteins. *Proc. Natl. Acad. Sci. U.S.A.* **1989,** *86* (21), 8382–8385.

31.   Dobson, C. M. Protein folding and misfolding. *Nature* **2003,** *426*, 884–890.

32.   Cheung, M. S.; García, A. E.; Onuchic, J. N. Protein folding mediated by solvation: Water expulsion and formation of the hydrophobic core occur after the structural collapse. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99* (2), 685.

33.   Berne, B. J.; Weeks, J. D.; Zhou, R. Dewetting and Hydrophobic Interaction in Physical and Biological Systems. *Annu. Rev. Phys. Chem.* **2009,** *60* (1), 85–103.

34.   Lopez, C. F.; Darst, R. K.; Rossky, P. J. Mechanistic elements of protein cold denaturation. *J. Phys. Chem. B* **2008,** *112* (19), 5961–5967.

35.   Bakk, A.; Høye, J. S. One-dimensional Ising model applied to protein folding. *Physica A* **2003,** *323*, 504–518.

36.   Graziano, G. On the molecular origin of cold denaturation of globular proteins. *Phys. Chem. Chem. Phys.* **2010,** *12* (42), 14245–14252.

37.   Matysiak, S.; Debenedetti, P. G.; Rossky, P. J. Role of Hydrophobic Hydration in Protein Stability: A 3D Water-Explicit Protein Model Exhibiting Cold and Heat Denaturation. *J. Phys. Chem. B* **2012,** *116* (28), 8095–8104.

38.   Yang, C.; Jang, S.; Pak, Y. A fully atomistic computer simulation study of cold denaturation of a β-hairpin. *Nat. Commun.* **2014,** *5*, 5773.

39.   Kim, S. B.; Palmer, J. C.; Debenedetti, P. G. Computational investigation of cold denaturation in the Trp-cage miniprotein. *Proc. Natl. Acad. Sci. U.S.A.* **2016,** *113* (32), 8991.

40.   Chatterjee, P.; Bagchi, S.; Sengupta, N. The non-uniform early structural response of globular proteins to cold denaturing conditions: A case study with Yfh1. *J. Chem. Phys.* **2014,** *141* (20), 205103.

41.   Laganowsky, A.; Liu, C.; Sawaya, M. R.; Whitelegge, J. P.; Park, J.; Zhao, M.; Pensalfini, A.; Soriaga, A. B.; Landau, M.; Teng, P. K. Atomic view of a toxic amyloid small oligomer. *Science* **2012,** *335* (6073), 1228–1231.

42.   Pan, J.; Han, J.; Borchers, C. H.; Konermann, L. Structure and Dynamics of Small Soluble Aβ(1–40) Oligomers Studied by Top-Down Hydrogen Exchange Mass Spectrometry. *Biochemistry* **2012,** *51* (17), 3694–3703.

43.   Bernstein, S. L.; Dupuis, N. F.; Lazo, N. D.; Wyttenbach, T.; Condron, M. M.; Bitan, G.; Teplow, D. B.; Shea, J.-E.; Ruotolo, B. T.; Robinson, C. V.; Bowers, M. T. Amyloid-β protein oligomerization and the importance of tetramers and dodecamers in the aetiology of Alzheimer's disease. *Nat. Chem.* **2009,** *1*, 326–331.

44.   Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19–25.

45.   Huang, J.; MacKerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013,** *34* (25), 2135–2145.

46.   Abascal, J. L. F.; Vega, C. A general purpose model for the condensed phases of water: TIP4P/2005. *J. Chem. Phys.* **2005,** *123* (23), 234505.

47.   Ashbaugh, H. S.; Collett, N. J.; Hatch, H. W.; Staton, J. A. Assessing the thermodynamic signatures of hydrophobic hydration for several common water models. *J. Chem. Phys.* **2010,** *132* (12), 124504.

48.   Vega, C.; Abascal, J. L. F. Simulating water with rigid non-polarizable models: a general perspective. *Phys. Chem. Chem. Phys.* **2011,** *13* (44), 19663–19688.

49.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103* (19), 8577–8593.

50.   Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1998,** *18* (12), 1463–1472.

51.   Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007,** *126* (1), 014101.

52.   Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981,** *52* (12), 7182–7190.

53.   Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33–38.

54.   Errington, J. R.; Debenedetti, P. G. Relationship between structural order and the anomalies of liquid water. *Nature* **2001,** *409* (6818), 318–321.

55.   Berhanu Workalemahu, M.; Hansmann Ulrich, H. E. The stability of cylindrin β-barrel amyloid oligomer models–A molecular dynamics study. *Proteins* **2013,** *81* (9), 1542–1555.

56.   Privalov, P. L.; Khechinashvili, N. N. A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J. Mol. Biol.* **1974,** *86* (3), 665–684.

57.   Dill, K. A.; Shortle, D. Denatured states of proteins. *Annu. Rev. Biochem.* **1991,** *60* (1), 795–825.

58.   Singh, R. S.; Biddle, J. W.; Debenedetti, P. G.; Anisimov, M. A. Two-state thermodynamics and the possibility of a liquid-liquid phase transition in supercooled TIP4P/2005 water. *J. Chem. Phys.* **2016,** *144* (14), 144504.

59.   Bagchi, B. *Water in Biological and Chemical Processes: From Structure and Dynamics to Function*. Cambridge University Press: Cambridge, **2013**.

60.   Laage, D.; Elsaesser, T.; Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017,** *117* (16), 10694–10725.

61.   Pal, S.; Maiti, P. K.; Bagchi, B.; Hynes, J. T. Multiple Time Scales in Solvation Dynamics of DNA in Aqueous Solution: The Role of Water, Counterions, and Cross-Correlations. *J. Phys. Chem. B* **2006,** *110* (51), 26396–26402.

62.   Pal, S.; Bandyopadhyay, S. Thermal unfolding of barstar and the properties of interfacial water around the unfolded forms. *J. Chem. Phys.* **2013,** *139* (23), 235101.

63.   Capponi, S.; White, S. H.; Tobias, D. J.; Heyden, M. Structural Relaxation Processes and Collective Dynamics of Water in Biomolecular Environments. *J. Phys. Chem. B* **2019,** *123* (2), 480–486.

64.   Das, P.; Matysiak, S. Direct Characterization of Hydrophobic Hydration during Cold and Pressure Denaturation. *J. Phys. Chem. B* **2012,** *116* (18), 5342–5348.

65.   Mukherjee, S.; Mondal, S.; Bagchi, B. Mechanism of Solvent Control of Protein Dynamics. *Phys. Rev. Lett.* **2019,** *122* (5), 058101.

66.   Palmer, J. C.; Martelli, F.; Liu, Y.; Car, R.; Panagiotopoulos, A. Z.; Debenedetti, P. G. Metastable liquid–liquid transition in a molecular model of water. *Nature* **2014,** *510,* 385–388.

67.   Liu, C.; Zhao, M.; Jiang, L.; Cheng, P.-N.; Park, J.; Sawaya, M. R.; Pensalfini, A.; Gou, D.; Berk, A. J.; Glabe, C. G.; Nowick, J.; Eisenberg, D. Out-of-

register β-sheets suggest a pathway to toxic amyloid aggregates. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (51), 20913.

68.   Röder, K.; Wales, D. J. Energy Landscapes for the Aggregation of Aβ$_{17-42}$. *J. Am. Chem. Soc.* **2018,** *140* (11), 4018–4027.

Influence of Hyperglycemic Conditions on Self-Association of the Alzheimer's Amyloid beta (Aβ$_{1\text{-}42}$) peptide
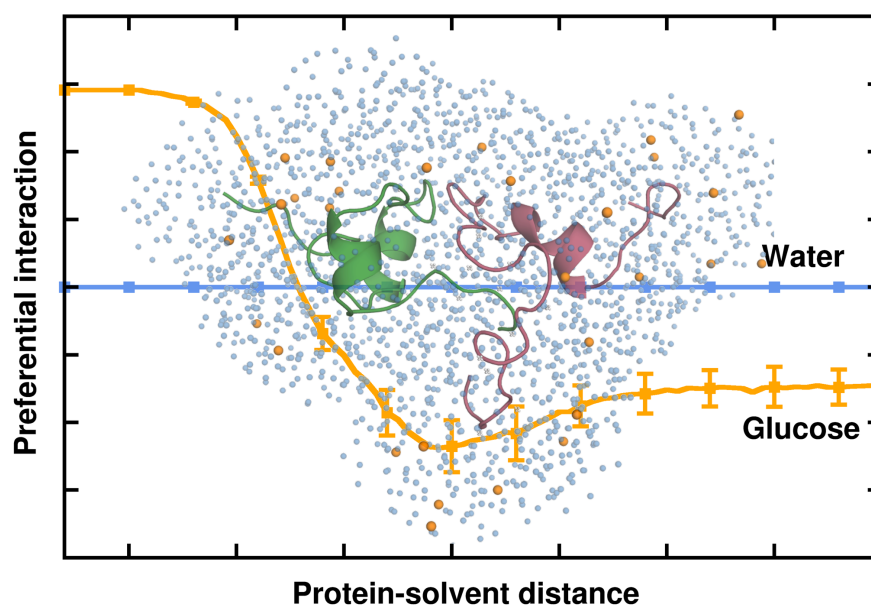
Based on:

4.1 Abstract

Clinical studies have identified a correlation between type-2 diabetes mellitus and cognitive decrements en route to the onset of Alzheimer's disease (AD). Recent studies have established that post-translational modifications of the amyloid beta (Aβ) peptide under hyperglycemic conditions; particularly, the process of glycation exacerbates its neurotoxicity and accelerates AD progression. In view of the assertion that macromolecular crowding has an altering effect on protein self-assembly, it is crucial to characterize the effects of hyperglycemic conditions via crowding on Aβ self-assembly. Toward this purpose, fully atomistic molecular dynamics simulations were performed to study the effects of glucose crowding on Aβ dimerization, which is the smallest known neurotoxic species. The dimers formed in the glucose-crowded environment were found to have weaker associations as compared to those formed in water. Binding free energy calculations show that the reduced binding strength of the dimers can be mainly attributed to the overall weakening of the dispersion interactions correlated with substantial loss of interpeptide contacts in the hydrophobic patches of the Aβ units. Analysis to discern the differential solvation pattern in the glucose-crowded and pure water systems revealed that glucose molecules cluster around the protein, at a distance of 5–7 Å, which traps the water molecules in close association with the protein surface. This preferential exclusion of glucose molecules and resulting hydration of the Aβ peptides has a screening effect on the hydrophobic interactions, which in turn diminishes the binding strength of the resulting dimers. Our results imply that physical effects attributed to crowded hyperglycemic environments are incapable of solely promoting Aβ self-assembly, indicating that further mechanistic studies are required to provide insights into the

self-assembly of post-translationally modified Aβ peptides, known to possess aggravated toxicity, under these conditions.

4.2   Introduction

Alzheimer's disease (AD) is the most common form of senile dementia and currently affects nearly 45 million people worldwide. It is a progressive, multifactorial, and irreversible disorder characterized by various pathological markers in the brain, particularly fibrillar deposits of the 4 kDa Amyloid beta (Aβ) peptide in the neuronal synapses[1-3]. Aβ is an intrinsically disordered protein (IDP) and hence defies the long-standing protein structure-function paradigm[4-7]. The lack of a single, well-defined equilibrium structure usually makes IDPs highly prone to self-assembly and aggregation, and in several cases such as Aβ, the insoluble aggregates are associated with the onset of debilitating neurodegenerative and other diseases[8-10]. The amyloid hypothesis postulates that the aggregation of Aβ into insoluble, fibrillar aggregates marks the onset of AD[11]. However, in recent years, accumulating evidence substantiates the hypothesis that small, soluble Aβ oligomers, rather than mature fibrils formed subsequently, may be the critical players in the pathology of AD[12-16]. Hence, uncovering the mechanisms of early self-assembly and oligomeric interactions, as well as factors that can potentially accelerate or slow down the rate of self-assembly of Aβ are among the key prerequisites for developing effective therapies against AD onset and progression.

Over the last few decades, increasing clinical evidence has shown a correlation of AD onset and cognitive decline with the occurrence of hyperglycemia and type-2 diabetes mellitus (T2DM) in elderly individuals[17-21]. It has been observed that

elevated blood glucose levels caused by several factors including insulin dysfunction and resistance may increase the chances of AD pathogenesis[21-22]. However, conclusive evidence demonstrating mechanistic linkages between excess glucose in the bloodstream and the onset of AD is still lacking. The search for the underlying causative factors is further complicated due to seemingly contradictory evidences. For example, it has been suggested that glucose may have some beneficial effects on the cognitive abilities of healthy individuals, whereas hyperglycemia may trigger neuronal death by excessive amyloid deposition in those already predisposed towards AD[23]. An emerging consensus appears to associate the post-translational chemical modifications of Aβ in hyperglycemic environments to its rate of self-assembly[23-24]. Particularly, Aβ modified as an advanced glycation end product (AGE) is thought to possess aggravated toxicity compared to that of the unmodified Aβ[25-26]. Using extensive atomistic computer simulations, we have recently demonstrated that Aβ with AGE modified lysines possesses greater β-sheet propensity and is thermodynamically predisposed to stronger self-association[27].

It is noteworthy here that although AGE modifications of Aβ are found to enhance the peptide's self-assembly, there exist no studies to date on how hyperglycemic conditions may directly influence the protein's self-assembly thermodynamics and thereby modulate the process in an alternative manner independent of plausible chemical modifications. This aspect becomes particularly important when one notes that macromolecular crowding of the aqueous environment can play a major role in altering the physical characteristics of a protein and its rate of self-association[28-34]. Particularly, simple sugars such as glucose, trehalose, sucrose, and polysaccharides such as dextran and Ficoll, frequently used as molecular crowding agents and as components for mimicking cytoplasmic crowding

environments, can have profound effects on protein self-assembly[35-39]. Various experimental and theoretical studies have demonstrated modest to drastic effects of macromolecular crowding on protein self-assembly and aggregation[29,40-47]. It is further interesting to note recent works demonstrating that mixtures of solvents may influence protein conformation and solubilities in a manner distinct from those brought about by pure solvents[48-49].

In light of the crucial influence of glucose in the aggregation propensities of Aβ, and thereby in the onset of AD, it is imperative to decouple its potential physical effects vis-à-vis crowding and its possible roles via chemical modifications in the modulation of Aβ self-assembly. Molecular dynamics (MD) simulations have been widely used to provide molecular insights into the structure, dynamics and self-assembly of amyloidogenic proteins such as Aβ, islet amyloid polypeptide, α-synuclein, and prion[50-69]. Herein, we present a systematic study based on fully atomistic computer simulations of the role of glucose molecules in modulating the dimerization of independent, full-length Aβ units. This earliest step in Aβ self-assembly is a crucial component in the nucleation-polymerization growth of Aβ oligomers, protofibrils and fibrils[50,70]. Our studies reveal that high glucose concentrations have a small effect on the overall properties of the Aβ monomer including conformational fluctuations, structure compactness, intrapeptide contacts and secondary structure propensities. Upon analyzing the early dimerization of Aβ peptides in glucose solution, we observed that there is a small but appreciable weakening in the binding strength of the dimers, concurrent with an observable loss in contacts between the hydrophobic domains of the peptide units. A component-wise analysis of the binding free energy reveals that the loss arises primarily from the weakening of the van der Waals (vdW) interaction energies accompanying the loss in

residue contacts. Considering the potential crowding effects brought about by glucose molecules, we further investigated solvent distributions in the vicinity of the Aβ dimeric complexes and found important effects that arise due to the presence of glucose. Our calculated preferential interaction parameters indicate that glucose molecules form a "cage" within about 5–7 Å of the protein heavy atoms that trap water molecules in the vicinity of the protein and create a distinctive increase in side chain hydration. This excess hydration reduces the efficacy of the hydrophobic effect and weakens the interactions between the hydrophobic domains of the Aβ units, accounting for an approximately 50% reduction in the strength of the binding free energy. Our results offer strong credence to the hypothesis that "standalone" physical effects of hyperglycemic conditions are incapable of consolidating Aβ self-assembly and enhancing its aggregation. Therefore, the observed clinical effects of hyperglycemic conditions on AD should be primarily via chemical modifications of Aβ, and plausibly through AGE modifications.

4.3   Methods

4.3.1   System Setup and MD Simulations

All MD simulations reported in this study were performed using the NAMD simulation package[71]. The details of the Aβ conformations generated and used are described below. The simulations were performed under periodic boundary conditions using the NAMD2.9 simulation package[71]. The CHARMM22 force field with CMAP correction[72] was used to simulate the peptides, the CHARMM36 all-atom carbohydrate force field[72] was used for glucose parameters, and the TIP3P[73] water model was used for the solvent. We point out that the CHARMM force field has been

noted to sample Aβ conformations with high levels of accuracy[74]. A time step of 2 fs was used. A constant temperature of 310 K was maintained with Langevin dynamics at a collision frequency of 1 ps^{-1}, and a pressure of 1 atm was maintained with the Nosé–Hoover method[75]. Long-range electrostatic interactions were computed using the particle-mesh Ewald (PME) method[76] and covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm[77]. The systems were first energy minimized for 15000 steps using the conjugate gradient method followed by simulations in the isothermal–isobaric (NPT) ensemble. Three independent MD simulations were performed for full-length Aβ units in pure water (PW) and glucose–water mixture (PG) solvents. Each trajectory was of 200 ns duration, amounting to a cumulative simulation time of 0.6 μs for each system.

4.3.1.1 Aβ Simulations

The details of the generation of the Aβ monomeric conformation can be found in previous work conducted in our group[27,78]. Briefly, the solution state NMR structure of the full-length Aβ$_{1-42}$ peptide obtained in a 3:7 mixture of hexafluoro-2-propanol and water (PDB ID: 1Z0Q)[79] was heated in the gas phase at 373 K. From the ensemble of the random coil configurations, ten structures were independently simulated at 310 K in explicit water for a minimum of 150 ns, generating a cumulative simulation data set of over 1.6 μs. Principal Component Analysis was then performed on the ensemble of the Aβ conformations and a conformation of the peptide representing one of the most populated clusters was chosen as the starting monomeric structure for our studies. The representative structure was benchmarked with experimental data by simulating it for 6 ns and comparing the ^{15}N and ^{13}C$_{\alpha}$

chemical shifts, which were calculated using the SHIFTS program[80]. We point out here that the structural propensities of the chosen initial conformation are remarkably similar to those of the full-length Aβ conformation reported to populate the peptide's ensemble in water[51,53].

The dimeric simulations were initiated by placing the two Aβ monomers at a center of mass distance of 33 Å and solvating in pure water (PW) and glucose solution to obtain the PW-D and PG-D systems, respectively. The PW-D system was solvated explicitly in a cubic box containing 22284 TIP3P[73] water molecules. The glucose simulation box of the PG-D system was built by placing the two monomeric structures at a center of mass distance of 33 Å with a random distribution of glucose molecules using the Packmol[81] program, followed by solvation with TIP3P[73] water molecules. The glucose concentration chosen was 108 g L^{-1}, which amounted to 218 glucose molecules in a cubic box containing 14968 water molecules. The minimum distance from the extremities of the protein to the edge of the simulation box was at least 15 Å. Similarly, for the monomeric systems, the Aβ peptide conformer described above was solvated in a cubic box of water and glucose solution at a concentration of 108 g L^{-1} to obtain the PW-M and PG-M systems, respectively. For each protein–solvent system, three independent trajectories of 200 ns each were generated. The system and simulation details are summarized in Table 4.1. All of the trajectories were equilibrated for 50 ns and the nontemporal analyses were done for the equilibrated part (last 150 ns) of the trajectories.

Type	Name	Solvent	d_0	S_W	S_G	Box dimensions (Å^3)	T_{total}
Monomer	PW-M	water	--	12905	--	74 x 79 x 69	0.6
Monomer	PG-M	water-glucose	--	9136	135	71 x 76 x 67	0.6
Dimer	PW-D	water	33	22284	--	107 x 91 x 71	0.6
Dimer	PG-D	water-glucose	33	14968	218	103 x 87 x 67	0.6

Table 4.1 Details of the simulated Aβ systems. The initial intermonomer center of distance (d_0 in Å), number of solvent water (S_W) and solvent glucose (S_G) molecules, initial dimensions of the simulation box, and the cumulative simulation times (T_{total}, in μs) are specified.

4.3.2 Trajectory Analysis

4.3.2.1 Secondary Structure

Secondary structural propensities for individual residues were obtained using the STRIDE algorithm[82], as implemented in VMD[83].

4.3.2.2 Protein-Protein Interaction Energy

The nonbonded interaction energies (electrostatic and vdW) between the peptide units were calculated using the NAMD Energy plugin available in the NAMD package[71]. The interaction energy, E, composed of electrostatic and vdW interactions, for a pair of atoms of charges q_i and q_j, separated by a distance r_{ij}, is given by:

$$E = 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{Dr_{ij}} \tag{4.1}$$

where the parameters σ_{ij}, ε_{ij} and D are obtained from the force field.

4.3.2.3   Solvent Accessible Surface Area

Solvent accessible surface area (SASA) values were calculated using the VMD package[83] by rolling a spherical probe of radius 1.4 Å over the protein residue surface.

4.3.2.4   Binding Free Energy

The absolute binding free energies between the two Aβ monomers were obtained using the Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) method, as implemented in the NAMD package[84-85]. The calculation was performed using the single trajectory method on the dimer as well as each of the monomeric subunits constituting the complex. The total free energy of each of the components is defined as:

$$G_{total} = H_{MM} + G_{solv-pol} + G_{solv-np} - TS_{config} \tag{4.2}$$

where H_{MM}, $G_{solv-pol}$, $G_{solv-np}$ and S_{config} represent the total internal energy, the polar solvation free energy, the nonpolar solvation free energy, and the configurational entropy, respectively. The internal energy, H_{MM}, constitutes the bond, angle, dihedral, improper, electrostatic, and vdW energies. The solvent dielectric constant of water at 310 K was used to compute the polar solvation free energy[86]. The nonpolar solvation free energy, $G_{solv-np}$, is quantified as the product of the surface tension of water ($\gamma = 0.0072$) and the SASA of the solute.

The binding free energy is estimated as the difference:

$$\Delta G_{binding} = G_{total(complex)} - G_{total(monomer1)} - G_{total(monomer2)} \qquad (4.3)$$

The entropic changes are ignored as in previous recent studies[56,84-85,87]. The binding free energy of the dimer complex is thus obtained as:

$$\Delta G_{binding} = \Delta H_{MM} + \Delta G_{solv-pol} + \Delta G_{solv-np}$$

$$= \left(\Delta E_{electrostatic} + \Delta E_{vdW} + \Delta E_{bonded}\right) + \Delta G_{solv-pol} + \Delta G_{solv-np} \qquad (4.4)$$

where, $\Delta E_{electrostatic}$, ΔE_{vdW} and $\Delta E_{internal}$ are the changes in the electrostatic, vdW, and bonded energies, respectively.

4.3.2.5 Preferential Interaction Parameters

To obtain information about the enrichment or exclusion of solution species at the protein surface, we estimated the solute–solvent preferential interaction parameters. These parameters have provided profound insights on protein solvation in several earlier works[36,88-91]. Accordingly, the time-averaged normalized preferential interaction parameters of solution species, water (P_{Ow}) and glucose (P_{glc}) can be defined as:

$$P_{Ow}(r) = \frac{n_{Ow}(r)\left[N_{Ow} + N_{glc}\right]}{N_{Ow}\left[n_{Ow}(r) + n_{glc}(r)\right]} \qquad (4.5)$$

$$P_{glc}(r) = \frac{n_{glc}(r)\left[N_{glc} + N_{Ow}\right]}{N_{glc}\left[n_{glc}(r) + n_{Ow}(r)\right]} \qquad (4.6)$$

where n_{Ow} and n_{glc} refer to the local number of water oxygen atoms and glucose molecules, respectively, located at a distance r from the heavy atoms of the protein. N_{Ow} and N_{glc} correspond to the total number of water oxygen atoms and glucose molecules in the simulation box, respectively. If the ratio $P_x(r)$ is greater than 1 in close proximity of the peptide, then the respective solvent species preferentially

interacts with the peptide. Conversely, if the ratio is lower than 1, the solvent molecules are preferentially excluded from the surface of the peptide.

4.3.2.6 Tetrahedral Order Parameter

To quantify the local degree of tetrahedrality, we calculated the tetrahedral order parameter[92-93] (q) of water molecules (by considering only the water oxygen atoms as potential hydrogen bonding sites) around the heavy atoms of high probability interpeptide contact residues in the dimeric (PW-D and PG-D) systems as:

$$q = 1 - \frac{3}{8} \sum_{j=1}^{3} \sum_{k=j+1}^{4} \left(\cos\theta_{jk} + \frac{1}{3} \right)^2 \tag{4.7}$$

θ_{jk} represents the angle between the O–O vectors i,j and i,k from the oxygen atom, i with its nearest neighbors (j and k; ≤ 4). The average values of q range from to 0 to 1; the order parameter value of 1 resembling the perfect tetrahedral structure of water in its ice-state. We have considered the water molecules, which lie within 5 Å of the residue heavy atoms.

4.4 Results and Discussion

4.4.1 Effects on Monomeric Conformation

We begin first by investigating the effect of glucose molecules on the conformational dynamics of the Aβ monomeric conformation. Representative snapshots from the PW-M and PG-M ensembles are illustrated in panels a and b of Figure 4.1. In Figure 4.1c, we present the root-mean-square deviation (RMSD) of the monomer relative to the initial structure in the PW-M and PG-M systems, averaged over multiple simulation trajectories. We find that amongst the two systems, the average RMSD of

the monomer is only slightly lower in glucose solution than that in PW, indicating a very marginal difference in fluctuations of the monomer in the presence of glucose. The mean RMSD values for the PG-M and PW-M systems, averaged over the last 150 ns, are 12.2 (±0.8) Å and 13.7 (±0.6) Å, respectively.
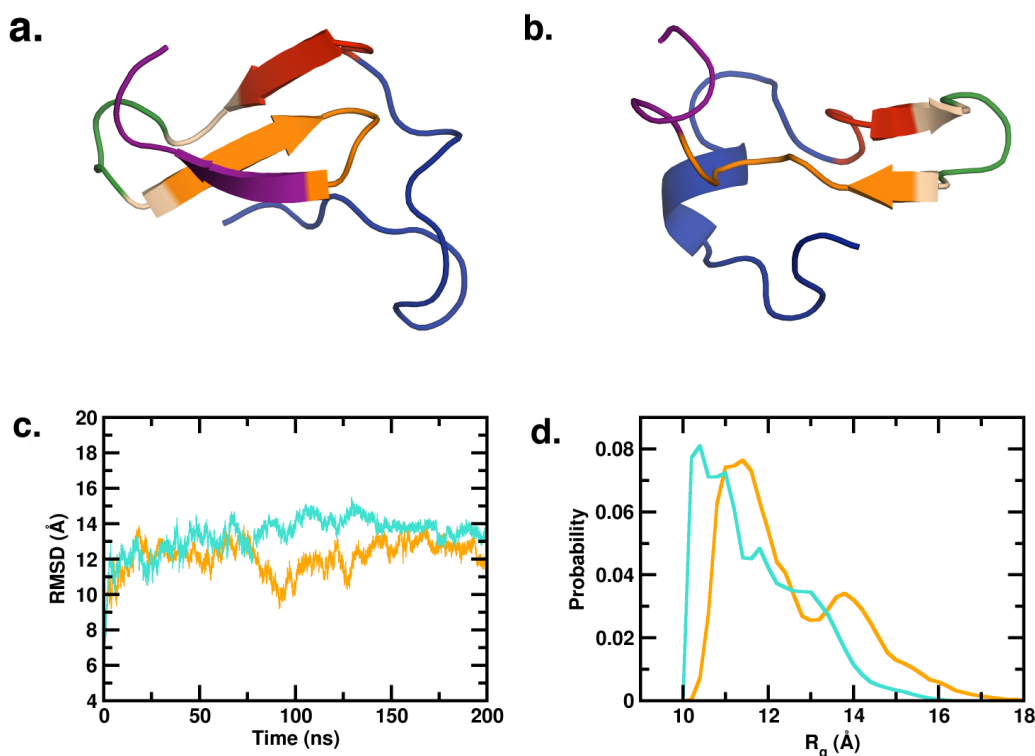
Figure 4.1 Representative structure from the monomeric ensembles (a) PW-M and (b) PG-M. The peptides are colored segment-wise. (N-terminal region (NTR)-blue, central hydrophobic core (CHC)-red, turn region (TR)-green, second hydrophobic region (SHR)-orange, C-terminal region (CTR)-magenta) (c) Time evolution of the backbone RMS deviations from the starting structure, averaged over multiple trajectories. (d) Distributions of the radius of gyration (R_g) of the PW-M and PG-M ensembles. Data for the PW-M and PG-M systems are shown in turquoise and orange, respectively.

Several studies have reported that in water, Aβ peptide adopts collapsed conformations that are attributable largely to the strong hydrophobic interactions of the CHC, and the hydrophobic regions at the C-terminus of the peptide[94-101]. We analyzed the radius of gyration (R_g) of the monomer of the two systems as a measure of the peptide's overall compactness. The mean R_g value of the Aβ monomer is 12.9 (±0.8) Å in the PG-M system and 11.5 (±0.5) Å in the PW-M system over the last 150 ns of the trajectories, indicating a small decrease of the mean R_g by 1.4 Å in the latter system. In Figure 4.1d, we compare the distributions of R_g values, where the overall shift to slightly higher values accompanied by a slight narrowing of the distribution is evident in the PG-M system. We note here that the probability distribution of R_g of the PW-M system peaks at 10.4 Å, which is in close agreement with a previous report[53]. The peak positions in the R_g distribution in PG-M can be observed at 11.4 Å and 13.8 Å, further underscoring the changes in compactness in the presence of glucose within the solvent environment.
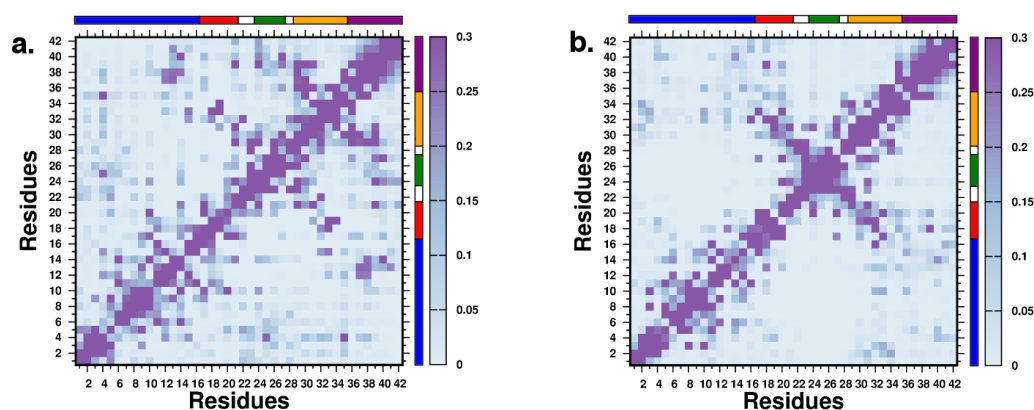
Figure 4.2 Intramonomer residue-residue contact probabilities for the (a) PW-M and (b) PG-M systems. Axes denote the residue numbers. The color scale for the contact probability is shown at the extreme right of each plot. The color bar at the top and right of each plot represents the segments in the Aβ peptide.

As reported extensively in previous studies[27,52], the collapse of the Aβ peptide in aqueous environment arises due to dewetting transitions and the resulting favorable interactions between distal hydrophobic residues located within the peptide sequence. To understand the key inter-residue interactions that are altered in the presence of glucose, we have compared the intramonomer contact maps, which are presented in Figure 4.2. For our analyses, we considered five segments of the Aβ peptide: NTR (D_1AEFRHDSGYEVHHQK$_{16}$), CHC (L_{17}VFFA$_{21}$), TR (V_{24}GSN$_{27}$), SHR (G_{29}AIIGLM$_{35}$) and CTR (V_{36}GGVVIA$_{42}$). The NTR and TR segments are mostly hydrophilic, whereas the CHC, SHR, and CTR segments are mainly hydrophobic. The contacts obtained in PW-M are similar to those reported in previous studies. The strongest inter-residue contacts are observed in the CHC/SHR, TR/CTR and SHR/CTR regions. Importantly, a large number of the strongest contacts observed in PW-M are lost in the PG-M system; upon inspection, it was revealed that the contacts lost in PG-M are predominantly hydrophobic contacts. The number of high probability, non-nearest neighbor contacts, defined as contacts between residues spaced by three or more units in the sequence, is 41 in the PW-M system and 28 in the PG-M system. Interestingly, we find that in addition to the loss in key hydrophobic contacts, a few strong contacts involving nonhydrophobic residues emerge in the PG-M system, suggesting a subtle alteration in the role of the solvent environment in contact formation upon the addition of glucose. This aspect is further corroborated when we compare the SASA of the high probability contact residues in both systems. The SASA per contact residue side chain is 69.2 Å^2 in PW-M and 81.0 Å^2 in PG-M, indicating that the contacts formed are relatively more exposed to the solvent in the latter system.

We finally examined the residue-wise secondary structural propensities of the monomers in the two systems; the comparisons are depicted in Figure 4.3. In PW-M, the β-strands are mainly located in the NTR, CHC, SHR, and CTR regions; it is noteworthy that these regions participate maximally in the intramonomer contacts and in the overall compactification of the peptide monomer in water. We point out that similar conformational contacts in these regions have been reported previously in experimental and computational studies[51,78,98,101-103]. The comparison between the PW-M and PG-M systems indicates that the presence of glucose is correlated with an overall decrease in β-strand propensities within the monomer.
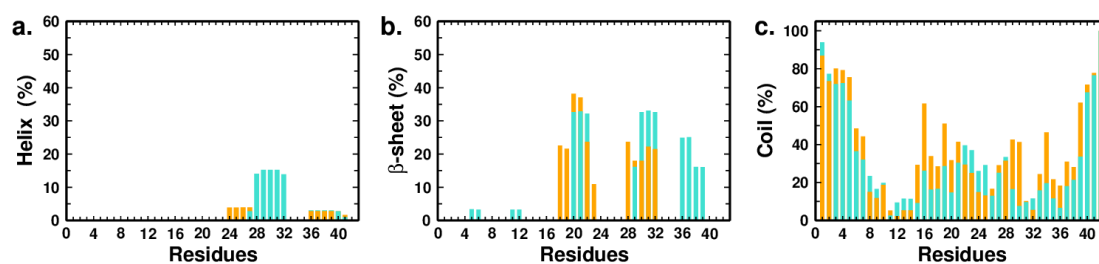
Figure 4.3 Residue-wise percentage secondary structure content of the (a) helix, (b) β-sheet, and (c) coil for the PW-M (in turquoise) and PG-M (in orange) ensembles.

These preliminary analyses indicate that the presence of glucose in the aqueous environment triggers small changes in the peptide's conformational fluctuations, compactness, intrapeptide hydrophobic contacts, solvent accessibility, and in the overall secondary structural propensities. In the following sections, we investigate in detail the ramifications of these changes on the peptide's self-association into the dimeric structure, and the associated differential role of solvation attributed to the presence of glucose.

4.4.2   Intermonomer Association and Structural Propensities

It is well-known that the Aβ peptide can self-associate to form several different assembly forms ranging from dimers to higher oligomers and aggregates of amyloid fibrils[13,16,101,104-105]. The Aβ dimer is of particular interest as it is the smallest neurotoxic species that impairs synaptic plasticity and memory, and further it is a key component in the nucleation mechanism[104,106-109]. Experimental and computational studies showed that Aβ$_{1-42}$ forms stable dimers in solution[110-112]. Herein, we have characterized the physical effect of glucose crowding on the spontaneous Aβ dimeric assembly process from the initial monomeric state. The dimerization event was first monitored via the center of mass distance between the two monomers in PW and in the glucose solution. Figure 4.4a shows the time evolution of the center of mass distance between the two monomeric units. The distributions of the distances obtained from the first 25 ns, as well as from the last 150 ns of the independent trajectories of the PW-D and PG-D systems are compared in Figure 4.4b,c. We observe that within the first 25 ns, the intermonomer distance in the PW-D system decreases dramatically from 33 Å to 10 Å. A similar phenomenon is observed in the PG-D system in which the intermonomer peptide distance, on an average, decreases to 11 Å within the initial 25 ns. The mean intermonomer center of mass distances in the PW-D and PG-D systems are similar within the first 25 ns of simulations, being 21.5 (±7.1) Å and 25.3 (±5.5) Å, respectively. In the PW-D system, after the initial 25 ns, the intermonomer distance fluctuates around 10–11 Å for the remaining part of the trajectory. However, in the PG-D system, beyond 25 ns, the distance increases to about 20 Å in the next 10 ns and then fluctuates between 18 and 20 Å subsequently. Within the last 15 ns, the interpeptide center of mass distance in this system again decreases to 16 Å. This is reflected in the bimodal distance distribution in the PG-D system, with a smaller

population peaking at 16 Å and a relatively larger population peaking around 18–20 Å. The mean intermonomer center of mass distances in the PW-D and PG-D systems over the last 150 ns are 10.5 (±1.2) Å and 17.8 (±1.9) Å, respectively. It is worth noting here that although both systems exhibit a drop in the interpeptide distance within the first 25 ns and a relative stability in this value over the last 150 ns, at any point of time, the mean distance is always greater in the PG-D system, suggestive of a discernable effect of glucose crowding on the spontaneous dimerization ability of Aβ in water.
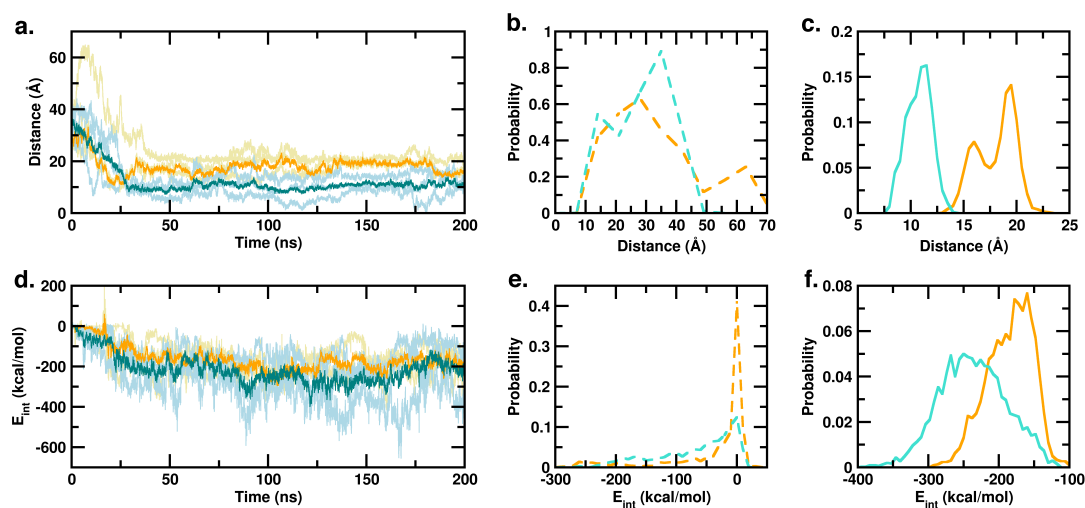
Figure 4.4 Evolution of the (a) interpeptide center of mass distance and (d) interpeptide interaction strength over the simulation timescale. Data for the PW-D and PG-D trajectories are shown in light blue and gold, respectively, and the averages corresponding to them are shown in teal and orange, respectively. (b and e) Probability distributions for the first 25 ns and (c and f) last 150 ns, corresponding to the data in (a) and (d). PW-D and PG-D are shown in turquoise and orange, respectively.

We further evaluated the intermonomer interaction energies as a function of simulation time in the PW-D and the PG-D systems; this quantity has previously been used as an (preliminary) indicator of the interpeptide interactions (binding strengths)[27,78,110,113-114]. Figure 4.4d depicts the time evolution of the interactions as observed in the trajectories of the PW-D and PG-D systems. Spontaneous dimerization of the Aβ peptide in both systems is marked by a lowering of the interaction energies within the first 50 ns, followed by a stability in the interaction energies over the latter part of the trajectories. Figure 4.4e,f depicts the probability distributions of the interpeptide interactions over the first 25 ns and the last 150 ns. The distribution over the first 25 ns peaks at 0.0 kcal mol^{-1} in both the PW-D and PG-D systems, indicating a lack of any significant initiation of dimerizing interactions in the earliest part of the trajectories. We mention here that the lack of significant interactions in the earliest parts of the simulations, noted previously in other reports, corresponds to the diffusive part of the Aβ dimerization process[110,115]. Unlike the early distributions, the distributions over the latter parts of the trajectory peak at −268.0 kcal mol^{-1} and −160.0 kcal mol^{-1}, respectively, in the PW-D and PG-D systems, indicating the presence of strong interpeptide dimerizing interactions. It is noteworthy that although the peaks corresponding to the two systems are well separated, there is a distinct degree of overlap between the two distributions. However, the distribution for the PG-D system is markedly narrower than that of the PW-D system, indicating a lower extent of fluctuations in the interactions in the presence of glucose in the solvent environment. The mean and standard deviation values of the interaction energies over the final 25 ns of the PG-D and PW-D simulation trajectories are −175.0 (±23.7) and −192.2 (±29.8) kcal mol^{-1}, respectively.

We next proceeded to analyze the structural features of the dimers formed in the two systems by comparing the intermonomer residue-residue contact probability map, illustrated in Figure 4.5a,b respectively. We computed the contact probability for the last 150 ns of each trajectory, leading to a cumulative simulation time of 450 ns for each system. As in other recent studies, a pair of residues forms a contact if the center of mass distance between their side-chains does not exceed 7 Å[27,78,98,114]. As apparent from the comparison of contact maps, there is a significant reduction in the total number of intermonomer contacts in PG-D as compared to that of the PW-D system. We note that in the PW-D ensemble, the region of high contact density involves hydrophobic interactions between the CHC and CTR segments, which is in agreement with the previously reported studies[27,116]. The SHR/CHC, SHR/TR, SHR/SHR, and SHR/CTR regions also display high contact density. Interestingly, there is a high density of contacts between the hydrophilic NTR and hydrophobic CTR regions. In addition, a moderate density of contacts is localized in the NTR/CHC and NTR/TR regions. On the other hand, in the PG-D system, a reduction in the intermonomer contacts is evident amongst the hydrophobic segments CHC, SHR, and CTR, whereas there is a modest density of contacts between the hydrophilic NTR with NTR as well as the hydrophobic CHC and SHR segments. Contacts are also formed between the hydrophilic TR and hydrophobic CTR region.
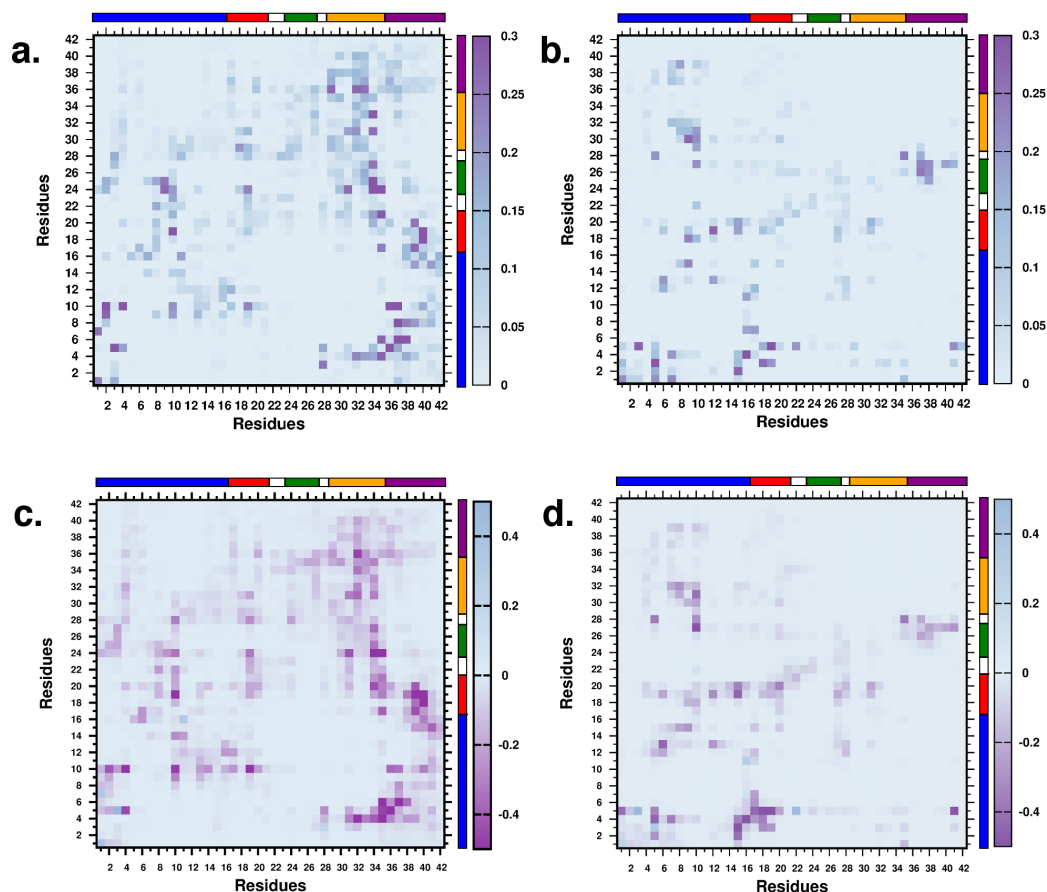
Figure 4.5 Interpeptide residue-wise contact probability maps for the (a) PW-D and (b) PG-D systems. Interpeptide residue-wise average vdW interaction energies (in kcal mol^{-1}) for the (c) PW-D and (d) PG-D ensembles.

We have further provided the intermonomer interaction energy maps corresponding to the average vdW interaction among the residues in Figure 4.5c,d for the PW-D and PG-D systems, respectively. We observe that the inter-residue contact probabilities among the monomers are vividly reflected in the average vdW interaction energies. Especially noticeable is the substantial weakening of vdW interaction energies in the hydrophobic patches of the peptides in the PG-D system, which corroborates with a loss of contacts in these regions. We further analyzed the secondary structural propensities of the Aβ peptides of the dimeric systems in the two solvent

environments. Comparing the overall secondary structure content in the two systems, depicted in Figure 4.6a-c, we observe a subtle β-sheet propensity of the residues in the hydrophobic CHC, TR and CTR segments in PW-D. It is interesting to note that β-sheet propensity in these regions is absent in the PG-D system. The discussion above demonstrates that the spontaneous dimerization of Aβ peptide is discernably compromised in the presence of glucose molecules. Importantly, the dimerization process is primarily affected by the distinct loss in key hydrophobic contacts that are known to play important roles in Aβ assembly.
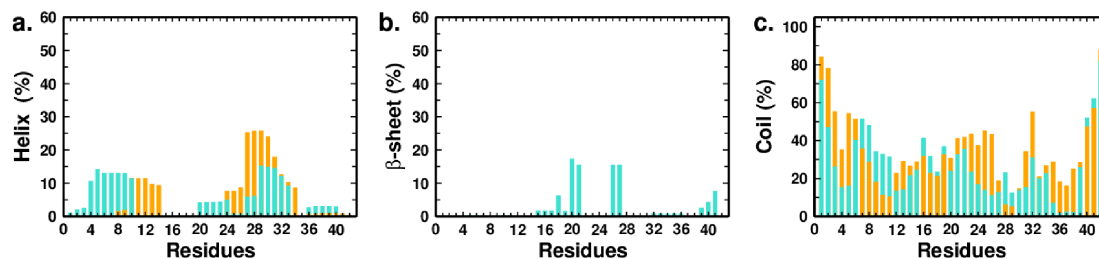
Figure 4.6 Residue-wise percentage secondary structure content of the (a) helix, (b) β-sheet, and (c) coil for the PW-D (in turquoise) and PG-D (in orange) ensembles.

4.4.3 Thermodynamics of Aβ Binding

The analyses presented thus far demonstrate that hyperglycemic conditions within the aqueous environment of the full-length Aβ peptide alter its conformational fluctuations, secondary structure, compactness, solvent exposure, and propensity for self-association in a noticeable manner. For deeper insights into the origins of the distinct weakening of Aβ dimerization observed in the presence of glucose, we have calculated and compared the intermonomer binding free energy of the peptide units, along with the individual contributing components. This was done with the MM-

GBSA protocol as described in the Methods section. We point out that this method has been routinely used to obtain binding affinities of biomolecules[84-85,117]. The mean and standard deviations of the various contributions to the total binding free energy are presented in Table 4.2. From these data sets, it can be observed that the mean value of the binding free energy, ΔG_{bind}, is lower in the PW-D system than that in the PG-D system by a value ranging from 25.2 to 27.2 kcal mol^{-1}, reflecting the relatively stronger dimerizing interactions in the former system. It is observed in both systems that the favorable binding free energy of dimerization originates predominantly from the nonpolar terms, namely ΔE_{vdW} and $\Delta G_{solv-np}$. Interestingly, the fluctuations observed in each component of the binding free energy are consistently higher in the PW-D system. The contribution of the electrostatic interactions between the two monomers (ΔE_{elec}) is offset by the contribution arising due to the polar solvation free energy ($\Delta G_{solv-pol}$). It is important to note here that a critical component of the binding ΔG_{bind} is obtained from the solvation free energy of the nonpolar moieties of the dimerizing units, $\Delta G_{solv-np}$; this quantity is markedly lower in the PW-D system. This observation suggests a relatively greater thermodynamic favorability of sequestering hydrophobic contacts in the PW-D system; this is corroborated by results discussed later. The magnitude of the differences in $\Delta G_{solv-np}$ between the two systems varies between 6.6 and 7.5 kcal mol^{-1}. Overall, these results suggest that in the crowded environment of glucose solution, there is a distinct weakening of the binding free energies that contributes to the dimerization of the full-length Aβ units. Such behavior qualitatively agrees with experimental and computational findings that sugars, which function as osmolytes within cells, hamper the aggregation of amyloidogenic peptides or globular proteins[36,118-121]. We point out that previous reports of Aβ dimerization have ignored the contributions arising due to loss in configurational entropy[115,122]. We

estimated the cumulative configurational entropy of the protein backbone atoms of the Aβ peptides for the PW-D and PG-D systems, calculated for the initial 10 ns of the dimerizing trajectories when the proteins exist as individual units as well as for the last 10 ns when the proteins have dimerized, using Schlitter's method[123]. The entropy change of Aβ peptides upon dimerization *(ΔS*$_{\text{monomer-dimer}}$*)* is small and comparable for the PW-D and PG-D systems, the values being 1.1 x 10^{-3} and 0.9 x 10^{-3} kcal mol^{-1} K^{-1}, respectively. Therefore, they are not considered in the calculation of ΔG_{bind}. The results presented thus far establish that the self-association of full-length Aβ units is thermodynamically weakened when an excess of glucose molecules is present.

Contribution	PW-D	PG-D
ΔG_{bind}	−54.739 (±18.685)	−27.584 (±11.847)
ΔH_{MM}	−241.996 (±105.642)	−185.145 (±68.752)
ΔE_{elec}	−169.391 (±95.823)	−150.682 (±71.374)
ΔE_{vdW}	−72.605 (±18.753)	−34.463 (±19.895)
ΔG_{solv}	187.257 (±90.258)	157.561 (±61.259)
$\Delta G_{\text{solv-np}}$	−15.660 (±3.712)	−8.202 (±3.090)
$\Delta G_{\text{solv-pol}}$	202.917 (±93.074)	165.763 (±61.767)

Table 4.2 Individual contributions of the interpeptide binding free energies calculated for the last 150 ns of each trajectory in the PW-D and PG-D systems (in kcal mol^{-1}). See the text for details. Standard deviations are provided in brackets.

4.4.4 Glucose Caging Modulates Aβ Hydration and Interactions

The solvent environment has a profound influence on the self-assembly behavior of IDPs such as Aβ[52,115,124-128]. Importantly, dewetting transitions and hydrophobic associations between Aβ monomers play a crucial role in a peptide's self-association[52,99-100]. In Figure 4.7, we depict representative snapshots of the dimeric state of the protein in the absence and in the presence of glucose in the solvent environment. The weakened association between the Aβ units and the resultant loss in the compactification of the dimeric state in the presence of glucose in the solvent is evident upon comparison.
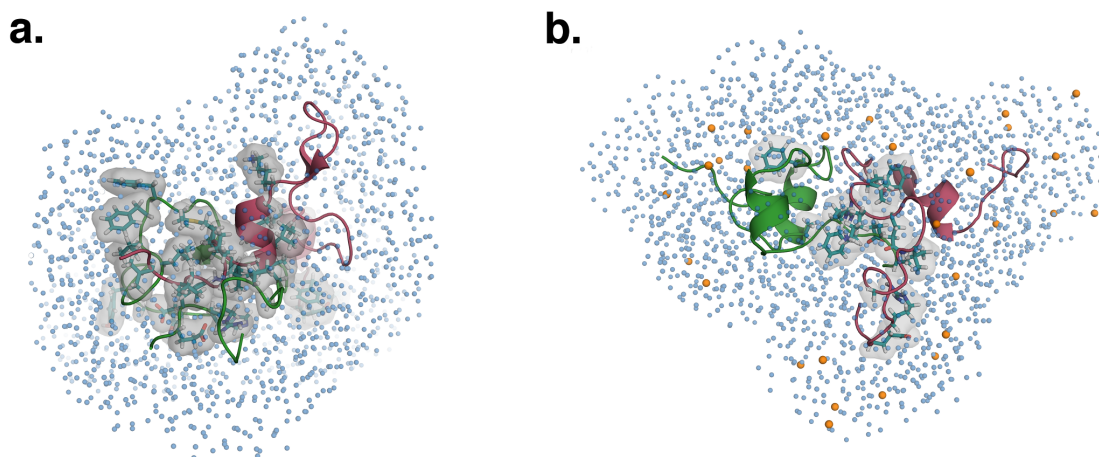
Figure 4.7 Representative structures from the dimeric ensembles (a) PW-D and (b) PG-D; the two peptide units are colored in pink and green. The residues involved in interpeptide high probability contacts are depicted by a translucent gray surface, with side chains represented as sticks and colored teal. Glucose molecules around the PG-D dimer within a distance of 7 Å from the protein units are shown as orange colored spheres and the water oxygens around the dimers in both the systems are shown as spheres colored skyblue.

In light of our findings, we further investigated how the nature of surface hydration could be altered in the presence of glucose. We first computed the selected site-site radial distribution functions, $g(r)$, involving the protein and the different solution species in PW and glucose solution, for both the monomeric and dimeric systems. The $g(r)$ between the protein heavy atoms and center of mass of glucose molecules in the PG-M and PG-D systems is shown in Figure 4.8a,b, respectively. We note that the protein-glucose $g(r)$ begins to gradually increase from 3.5 Å and forms a broad peak centered at 6 Å. This indicates that there is a high density of glucose molecules in the shell between 5 and 7 Å around the protein. Figure 4.9a,b display the $g(r)$ calculated between the protein heavy atoms and water oxygen atoms in the monomeric (PG-M and PW-M) and dimeric (PG-D and PW-D) systems, respectively. Interestingly, we observe that in the first and second peaks of the protein-water $g(r)$, positioned at 2.8 and 3.8 Å, respectively, there is a marginal but noticeable enhancement in the hydration (see the figure inset) in the glucose system over that in PW; this enhancement is consistently observed in both the monomeric and dimeric systems. For insights into the origin of this marginal difference, we calculated the radial distribution functions between the water oxygens and the full side chains of the residues that participate in internal contacts (monomeric systems), and in the interpeptide contacts (dimeric systems) with high probability. This comparison, presented in Figure 4.8c,d for the monomeric and dimeric systems, respectively, clearly shows marked enhancements in the first and second solvation peaks in the presence of glucose. This indicates that the differences observed earlier can be largely attributed to the enhanced hydration of the internal contacts formed within the monomer in the PG-M compared to that in the PW-M system, and to the interpeptide contacts formed within the dimer in the PG-D compared to that in the PW-D system.

In Table 4.3, we have tabulated the number of high probability, nonlocal internal contacts (in monomer) and interpeptide (in dimer) contacts observed in absence and in the presence of glucose. It is noted that while some of the contacts are common, there is an overall decrease in the number of hydrophobic contacts in the presence of glucose. Interestingly, in the dimeric PG-D system, most of the nonlocal contacts formed involve the participation of polar residues (see Table 4.4), reflecting the relatively greater difference between the $g(r)$ observed in Figure 4.8d. We further note here that the average SASA of the side chains that contributes to the nonlocal interpeptide contacts increase about three-fold in the PG-D system in comparison to the PW-D system. (see Table 4.5) Furthermore, we evaluated the mean tetrahedral order parameter (q) of the water molecules that lie within 5 Å of the interpeptide contact residues in the dimeric systems (see Table 4.5). The tetrahedral order parameter is an indicator of the structural ordering of the local hydration waters[93,129-130]. A marginal decrease in the average q of the waters in the vicinity of the contacts in the PG-D relative to that in the PW-D system reflects a small decrease in the overall local ordering of hydration waters around the contact residues.
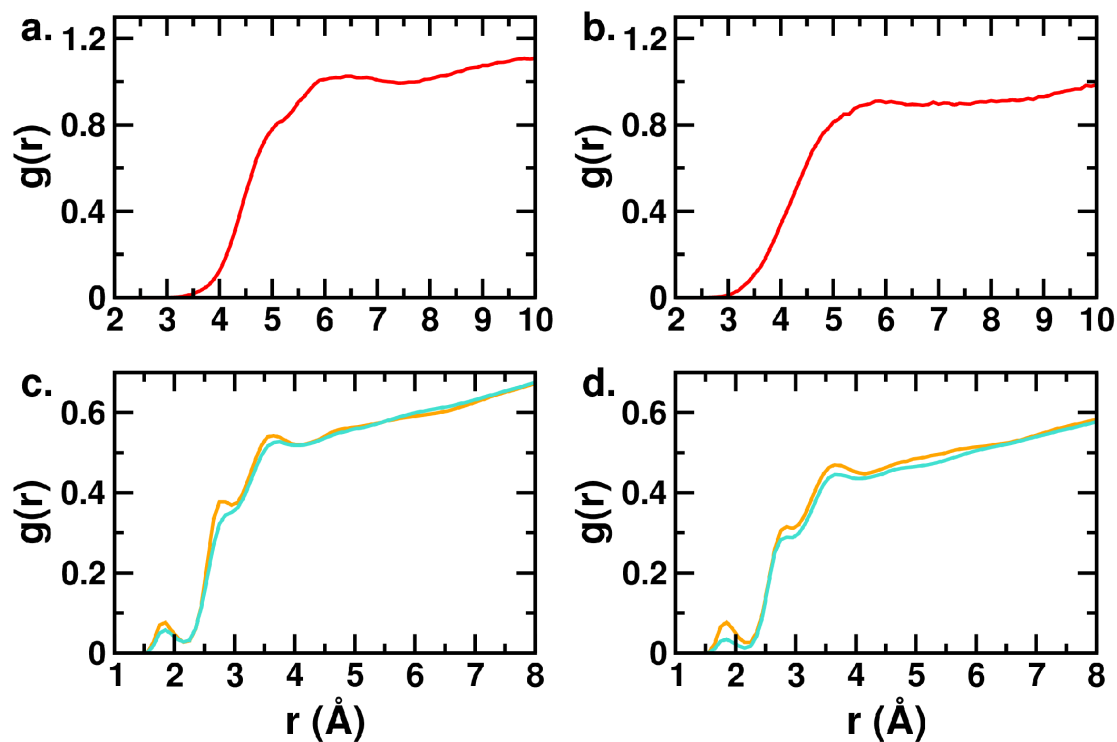
Figure 4.8 Site-site radial distribution function, $g(r)$, between protein heavy atoms and glucose center of mass of the (a) PG-M and (b) PG-D systems is represented in red. The $g(r)$ between the water oxygen atoms and full side chains of residues involved in high probability contacts of the (c) monomeric (PW-M and PG-M) and (d) dimeric (PW-D and PG-D) systems. Data corresponding to the PW and glucose solution (PG) are shown in turquoise and orange, respectively.
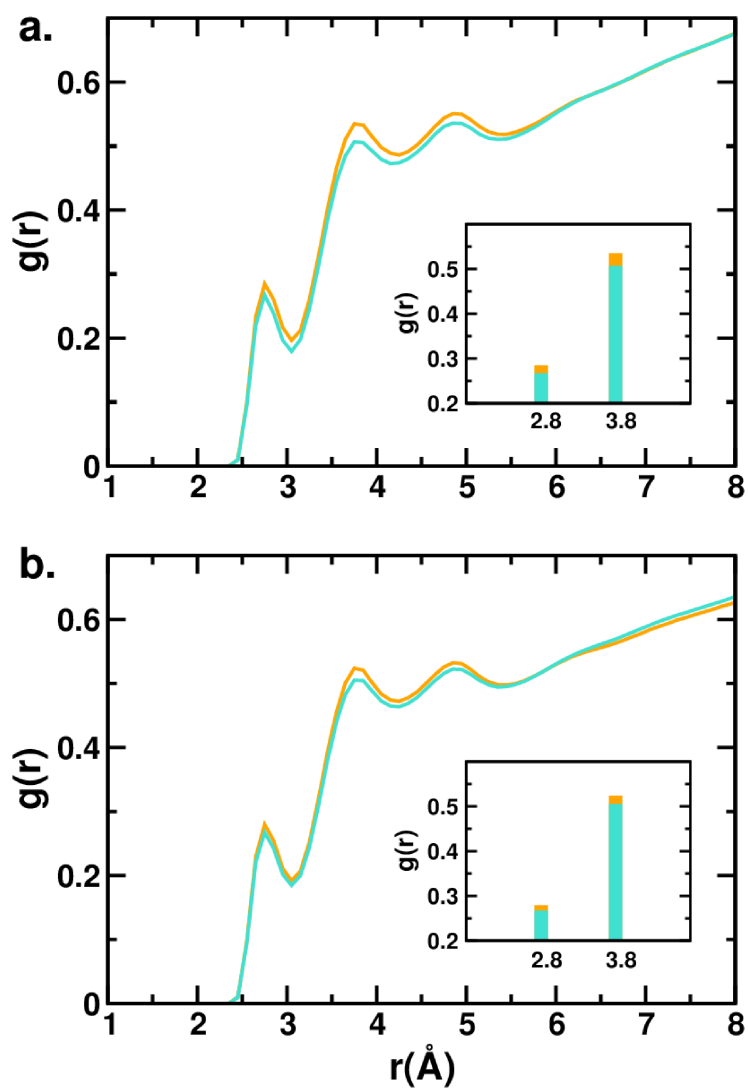
Figure 4.9 Site-site radial distribution function, *g*(*r*), of the protein heavy atoms and water oxygen atoms of the (a) monomeric (PW-M and PG-M) and (b) dimeric (PW-D and PG-D) systems. The inset bar plots show the first and second *g(r)* peaks. Data for PW and glucose solution are shown in turquoise and orange, respectively.

	PG-M	PW-M	PG-D	PW-D
N_{tot}	28	41	8	31
N_P	15	21	6	14
N_H	13	20	2	17

Table 4.3 Summary of number of nonlocal contacts in the monomeric systems (PG-M and PW-M), and the dimeric systems (PG-D and PW-D). N_{tot} represents the total number of contacts, N_P represents the number of contacts involving at least one non-hydrophobic residue, and N_H represents the number of contacts involving only hydrophobic residues

Contact pair	System		Type
Ala 2-Gln 15		PG-D	P
Glu 3-Arg 5		PG-D	P
Glu 3-Val 18		PG-D	P
Glu 3-Lys 28	PW-D		P
Phe 4-Lys 16		PG-D	P
Arg 5-Glu 3	PW-D	PG-D	P
Arg 5-Phe 19		PG-D	P
Arg 5-Val 36	PW-D		P
His 6-Met 35	PW-D		P
His 6-Gly 37	PW-D		P
His 6-Gly 38	PW-D		P
Asp 7-Asp 1	PW-D		P
Ser 8-Gly 37	PW-D		P
Tyr 10-Ala 2	PW-D		P
Tyr 10-Phe 4	PW-D		P

Contact pair	System	Type
Tyr 10-Val 36	PW-D	P
Tyr 10-Asn 27	PW-D	P
Phe 19-Tyr 10	PW-D	P
Asn 27-Leu 34	PW-D	P
Phe 4-Met 35	PW-D	H
Gly 9-Ala 2	PW-D	H
Leu 17-Val 39	PW-D	H
Val 18-Val 40	PW-D	H
Phe 19-Val 12	PG-D	H
Phe 19-Val 40	PW-D	H
Phe 20-Val 39	PW-D	H
Ala 21-Met 35	PW-D	H
Val 24-Gly 9	PW-D	H
Val 24-Ile 31	PW-D	H
Val 24-Leu 34	PW-D	H
Val 24-Met 35	PW-D	H
Gly 25-Leu 34	PW-D	H
Ala 30-Gly 9	PG-D	H
Ile 31-Leu 34	PW-D	H
Gly 33-Leu 34	PW-D	H
Val 36-Gly 29	PW-D	H
Val 36-Ile 32	PW-D	H
Val 36-Gly 33	PW-D	H

Table 4.4 Interpeptide residue contacts found in the dimeric Aβ in PW-D and in the system with glucose (PG-D). Each contact is classified as purely hydrophobic (H) when both the participating residues are hydrophobic and polar with the participation of at least one polar or charged residue (P).

Contribution	PW-D	PG-D
$<q>$	0.399	0.306
$<SASA>/N_{con-res}$	50.922	155.277

Table 4.5 Average values of the tetrahedral order parameter ($<q>$) and SASA/contact residue ($<SASA>$ / $N_{con-res}$) of the dimeric systems (PW-D and PG-D).

To further ascertain the preferential hydration of Aβ peptides in glucose solution, we characterized the relative local distribution of water and glucose molecules around the Aβ dimers, as described in the Methods section. Figure 4.10 depicts the normalized fraction of glucose molecules (P_{glc}) and water oxygen (P_{Ow}), as a function of the distance from the protein heavy atoms in PW and glucose solutions. Commensurate with the $g(r)$ trends, it can be observed that up to a distance of 4.5 Å from the protein heavy atoms, P_{Ow} is greater than 1, which signifies that the Aβ peptides are more preferentially hydrated in the glucose solution. In addition, P_{glc} is lower than 1 at a distance below 4.5 Å with a peak greater than 1 in the distance range 5–7 Å. This indicates that the glucose molecules are excluded from the surface of the dimer and form a dense space-filling network surrounding the dimer at a distance of 5–7 Å, which causes a depletion of water in this region. This cage-like network traps the water molecules at the surface of the protein resulting in enhanced hydration of the dimer. Considering the relevance of a dewetting-induced hydrophobic collapse to Aβ self-assembly, we further investigated if the phenomenon of glucose caging and enhanced protein surface hydration occurs at the dimer interface. Analyses of the preferential interaction parameters for the dimer interface region, depicted in Figure 4.9, reveal that concurrent to the whole dimer, the

154

interfacial region is characterized by a water-enriched hydration shell resulting from the caging effect of glucose. We remark here that the presence of the water molecules caged at the protein surface by the glucose clusters reduces the overall interactions between hydrophobic residues that provide a major driving force for Aβ self-assembly, and account for the reduced binding strength of the resulting dimers.
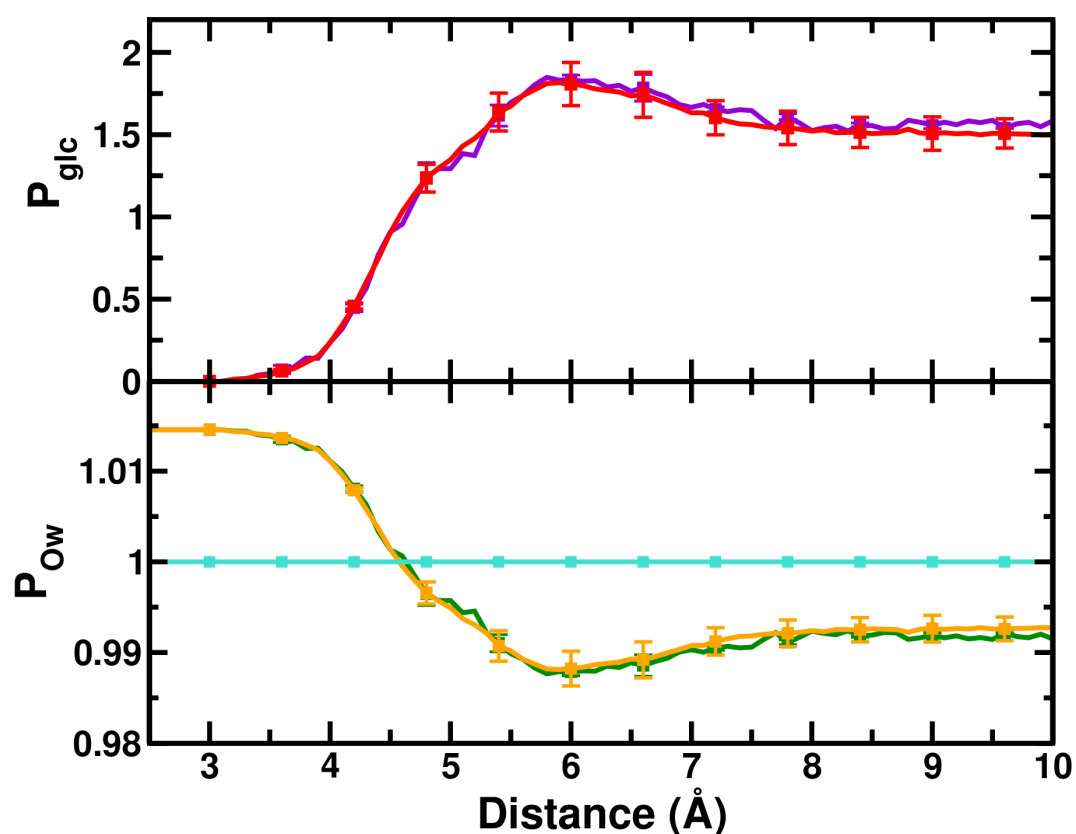
Figure 4.10   Time-averaged normalized preferential interaction parameters of the relative local distribution of glucose (P_{glc}) and water (P_{Ow}) in the dimeric trajectories. The upper panel shows P_{glc} for the PG-D dimer (in red) and the dimer interface (in violet) while the lower panel represents P_{Ow} for PW-D and PG-D (whole dimer), in turquoise and orange, respectively while the ratio for PG-D dimer interface is shown in green.

4.5   Conclusions

Numerous epidemiological studies over the last few decades have linked T2DM to an increased risk of AD[17-21]. Notably, evidence has suggested that hyperglycemia-mediated glycated Aβ (Aβ-AGE) is significantly more pathogenic than the unglycated one and augments AD progression both in vitro and in vivo[25-26]. However, with accumulating knowledge on the implications of macromolecular crowding on protein self-assembly, there are currently no studies probing the physical aspects of crowded hyperglycemic conditions on the thermodynamics of Aβ self-assembly. Further, it is important to note complex effects that may be brought about on biomolecular conformations by the crowding interactions of solvent mixtures[48-49]. In light of such observations, we have, in the present study, used classical MD simulations to delineate the physical effects of the crowded environment of aqueous glucose solution on the conformational stability and the self-assembly characteristics of full-length Aβ peptide. We find that the glucose-crowded environment has a narrow but discernable impact on the Aβ monomer with respect to its conformational fluctuations, compactness, internal contacts, solvent exposure, and in its overall secondary structure propensities. Our simulations of the early self-assembly of Aβ monomers reveal that the resultant dimers in glucose solution exhibited weakened peptide-peptide binding free energies and a substantial loss in the number of intermonomer contacts. It is noteworthy that the reduced binding strength of the dimers mainly arises from overall weakening of the dispersion interactions that is commensurate with the loss of inter-residue contacts in the hydrophobic segments of the peptides. Considering the critical role that hydration water plays in protein aggregation as well as the excluded volume effect owing to the presence of glucose molecules, we evaluated the local hydration pattern of the dimers to elucidate if

crowding modulates Aβ hydration. Our analysis of the preferential interactions of Aβ with the solvent species indicates that glucose molecules cluster around the peptides at a distance of 5–7 Å and enriches the shell in the vicinity of the protein surface with water molecules. This preferential hydration of the Aβ peptides and the caging effect of glucose molecules screen the hydrophobic interactions between the peptides and weaken the binding strength of the resulting dimers. Our results demonstrate the physical effects of hyperglycemic conditions and the resultant crowding effects on the conformational properties and early self-assembly of Aβ.

Further studies in our laboratory are underway to dissect the effects of crowding on the microscopic details of the structural and dynamical properties of the hydration layer of the Aβ dimers in glucose solution. In view of the enhanced Aβ neurotoxicity upon hyperglycemia induced chemical modifications, it is further important to gain a molecular level understanding of the self-assembly of these chemically modified Aβ peptides under hyperglycemic conditions. These studies will aid in gaining molecular insights into the copathogenesis of T2DM and AD as well as provide incentives to design effective therapeutic strategies to counteract the harmful effects of these debilitating diseases.

4.6   References

1.   Masters, C. L.; Simms, G.; Weinman, N. A.; Multhaup, G.; McDonald, B. L.; Beyreuther, K. Amyloid plaque core protein in Alzheimer disease and Down syndrome. *Proc. Natl. Acad. Sci. U.S.A.***1985,** *82* (12), 4245–4249.

2.   LeBlanc, A. Increased production of 4 kDa amyloid beta peptide in serum deprived human primary neuron cultures: possible involvement of apoptosis. *J. Neurosci.* **1995,** *15* (12), 7837–7846.

3.   Busciglio, J.; Gabuzda, D. H.; Matsudaira, P.; Yankner, B. A. Generation of beta-amyloid in the secretory pathway in neuronal and nonneuronal cells. *Proc. Natl. Acad. Sci. U.S.A.***1993,** *90* (5), 2092–2096.

4.   Wright, P. E.; Dyson, H. J. Intrinsically unstructured proteins: re-assessing the protein structure-function paradigm. *J. Mol. Biol.* **1999,** *293* (2), 321–331.

5.   Hegyi, H.; Tompa, P. Intrinsically Disordered Proteins Display No Preference for Chaperone Binding In Vivo. *PLoS Comput. Biol.* **2008,** *4* (3), No. e1000017.

6.   Chebaro, Y.; Ballard, A. J.; Chakraborty, D.; Wales, D. J. Intrinsically Disordered Energy Landscapes. *Sci. Rep.* **2015,** *5*, No. 10386.

7.   Dyson, H. J. Making Sense of Intrinsically Disordered Proteins. *Biophys. J.* **2016,** *110* (5), 1013–1016.

8.   Ross, C. A.; Poirier, M. A. Protein aggregation and neurodegenerative disease. *Nat. Med.* **2004,** *10*, S10−S17.

9.   Uversky, V. N. Intrinsically disordered proteins from A to Z. *The International J. Biochem. Cell Biol.* **2011,** *43* (8), 1090–1103.

10.   Uversky, V. N. Introduction to Intrinsically Disordered Proteins (IDPs). *Chem. Rev.* **2014,** *114* (13), 6557–6560.

11.   Hardy, J.; Selkoe, D. J. The Amyloid Hypothesis of Alzheimer's Disease: Progress and Problems on the Road to Therapeutics. *Science* **2002,** *297* (5580), 353–356.

12.   Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis. *Science* **2003,** *300* (5618), 486–489.

13.   Walsh, D. M.; Selkoe, D. J. Aβ Oligomers–a decade of discovery. *J. Neurochem.* **2007,** *101* (5), 1172–1184.

14.	Liao, M. Q.; Tzeng, Y. J.; Chang, L. Y. X.; Huang, H. B.; Lin, T. H.; Chyan, C. L.; Chen, Y. C. The correlation between neurotoxicity, aggregative ability and secondary structure studied by sequence truncated Aβ peptides. *FEBS Lett.* **2007,** *581* (6), 1161–1165.

15.	Haass, C.; Selkoe, D. J. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid β-peptide. *Nat. Rev. Mol. Cell Biol.* **2007,** *8* (2), 101–112.

16.	Ahmed, M.; Davis, J.; Aucoin, D.; Sato, T.; Ahuja, S.; Aimoto, S.; Elliott, J. I.; Van Nostrand, W. E.; Smith, S. O. Structural conversion of neurotoxic amyloid $\beta_{1\text{-}42}$ oligomers to fibrils. *Nat. Struct. Mol. Biol.* **2010,** *17* (5), 561–567.

17.	Leibson, C. L.; Rocca, W. A.; Hanson, V. A.; Cha, R.; Kokmen, E.; O'Brien, P. C.; Palumbo, P. J. Risk of Dementia among Persons with Diabetes Mellitus: A Population-based Cohort Study. *Am. J. Epidemiol.* **1997,** *145* (4), 301–308.

18.	Haan, M. N. Therapy Insight: type 2 diabetes mellitus and the risk of late-onset Alzheimer's disease. *Nat. Clin. Pract. Neuro.* **2006,** *2* (3), 159–166.

19.	Takeda, S.; Sato, N.; Rakugi, H.; Morishita, R. Molecular mechanisms linking diabetes mellitus and Alzheimer disease: beta-amyloid peptide, insulin signaling, and neuronal function. *Mol. BioSyst.* **2011,** *7* (6), 1822–1827.

20.	Candore, G. Can Alzheimer disease be a form of type 3 diabetes. *Rejuvenation Res.* **2012,** *15* (2), 217–221.

21.	Biessels, G. J.; Strachan, M. W. J.; Visseren, F. L. J.; Kappelle, L. J.; Whitmer, R. A. Dementia and cognitive decline in type 2 diabetes and prediabetic stages: towards targeted interventions. *Lancet Diabetes Endocrinol.* **2014,** *2* (3), 246–255.

22.	Gasparini, L.; Netzer, W. J.; Greengard, P.; Xu, H. Does insulin dysfunction play a role in Alzheimer's disease? *Trends Pharmacol. Sci.* **2002,** *23* (6), 288–293.

23.	Messier, C.; Gagnon, M. Glucose regulation and cognitive functions: relation to Alzheimer's disease and diabetes. *Behav. Brain Res.* **1996,** *75* (1–2), 1–11.

24.	Takeuchi, M.; Yamagishi, S.-i. Involvement of toxic AGEs (TAGE) in the pathogenesis of diabetic vascular complications and Alzheimer's disease. *J. Alzheimer's Dis.* **2009,** *16* (4), 845–858.

25.	Li, X. H.; Du, L. L.; Cheng, X. S.; Jiang, X.; Zhang, Y.; Lv, B. L.; Liu, R.; Wang, J. Z.; Zhou, X. W. Glycation exacerbates the neuronal toxicity of [beta]-amyloid. *Cell Death Dis.* **2013,** *4*, No. e673.

26. Chen, C.; Li, X. H.; Tu, Y.; Sun, H. T.; Liang, H. Q.; Cheng, S. X.; Zhang, S. Aβ-AGE aggravates cognitive deficit in rats via RAGE pathway. *Neuroscience* **2014,** *257*, 1–10.

27. Jana, A. K.; Batkulwar, K. B.; Kulkarni, M. J.; Sengupta, N. Glycation induces conformational changes in the amyloid-β peptide and enhances its aggregation propensity: molecular insights. *Phys. Chem. Chem. Phys.* **2016,** *18* (46), 31446–31458.

28. Zhou, H.-X.; Rivas, G.; Minton, A. P. Macromolecular crowding and confinement: biochemical, biophysical, and potential physiological consequences. *Annu. Rev. Biophys.* **2008,** *37*, 375–397.

29. van den Berg, B.; Ellis, R. J.; Dobson, C. M. Effects of macromolecular crowding on protein folding and aggregation. *EMBO J.* **1999,** *18* (24), 6927–6933.

30. Roosen-Runge, F.; Hennig, M.; Zhang, F.; Jacobs, R. M. J.; Sztucki, M.; Schober, H.; Seydel, T.; Schreiber, F. Protein self-diffusion in crowded solutions. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108* (29), 11815–11820.

31. Arrio-Dupont, M.; Foucault, G.; Vacher, M.; Devaux, P. F.; Cribier, S. Translational Diffusion of Globular Proteins in the Cytoplasm of Cultured Muscle Cells. *Biophys J.* **2000,** *78* (2), 901–907.

32. Wang, Y.; Li, C.; Pielak, G. J. Effects of Proteins on Protein Diffusion. *J. Am. Chem. Soc.* **2010,** *132* (27), 9392–9397.

33. Dix, J. A.; Verkman, A. S. Crowding Effects on Diffusion in Solutions and Cells. *Annu. Rev. Biophys.* **2008,** *37* (1), 247–263.

34. Minton, A. P. Implications of macromolecular crowding for protein assembly. *Curr. Opin. Struct. Biol.* **2000,** *10* (1), 34–39.

35. Zhou, B.-R.; Zhou, Z.; Hu, Q.-L.; Chen, J.; Liang, Y. Mixed macromolecular crowding inhibits amyloid formation of hen egg white lysozyme. *Biochim. Biophys. Acta, Proteins Proteomics* **2008,** *1784* (3), 472–480.

36. Liu, F.-F.; Ji, L.; Dong, X.-Y.; Sun, Y. Molecular Insight into the Inhibition Effect of Trehalose on the Nucleation and Elongation of Amyloid β-Peptide Oligomers. *J. Phys. Chem. B* **2009,** *113* (32), 11320–11329.

37. Zhou, H.-X. Influence of crowded cellular environments on protein folding, binding, and oligomerization: Biological consequences and potentials of atomistic modeling. *FEBS Lett.* **2013,** *587* (8), 1053–1061.

38. Breydo, L.; Reddy, K. D.; Piai, A.; Felli, I. C.; Pierattelli, R.; Uversky, V. N. The crowd you're in with: Effects of different types of crowding agents on protein aggregation. *Biochim. Biophys. Acta, Proteins Proteomics* **2014,** *1844* (2), 346–357.

39. Spiga, E.; Abriata, L. A.; Piazza, F.; Dal Peraro, M. Dissecting the Effects of Concentrated Carbohydrate Solutions on Protein Diffusion, Hydration, and Internal Dynamics. *J. Phys. Chem. B* **2014,** *118* (20), 5310–5321.

40. Uversky, V. N.; M. Cooper, E.; Bower, K. S.; Li, J.; Fink, A. L. Accelerated α-synuclein fibrillation in crowded milieu. *FEBS Lett.* **2002,** *515* (1-3), 99–103.

41. Hatters, D. M.; Minton, A. P.; Howlett, G. J. Macromolecular Crowding Accelerates Amyloid Formation by Human Apolipoprotein C-II. *J. Biol. Chem.* **2002,** *277* (10), 7824–7830.

42. Munishkina, L. A.; Ahmad, A.; Fink, A. L.; Uversky, V. N. Guiding Protein Aggregation with Macromolecular Crowding. *Biochemistry* **2008,** *47* (34), 8993–9006.

43. Munishkina, L. A.; Cooper, E. M.; Uversky, V. N.; Fink, A. L. The effect of macromolecular crowding on protein aggregation and amyloid fibril formation. *J. Mol. Recognit.* **2004,** *17* (5), 456–464.

44. Magno, A.; Caflisch, A.; Pellarin, R. Crowding Effects on Amyloid Aggregation Kinetics. *J. Phys. Chem. Lett.* **2010,** *1* (20), 3027–3032.

45. Ma, Q.; Fan, J.-B.; Zhou, Z.; Zhou, B.-R.; Meng, S.-R.; Hu, J.-Y.; Chen, J.; Liang, Y. The Contrasting Effect of Macromolecular Crowding on Amyloid Fibril Formation. *PLoS One* **2012,** *7* (4), No. e36288.

46. Ma, B.; Xie, J.; Wei, L.; Li, W. Macromolecular crowding modulates the kinetics and morphology of amyloid self-assembly by β-lactoglobulin. *Int. J. Biol. Macromol.* **2013,** *53*, 82–87.

47. Latshaw, D. C.; Cheon, M.; Hall, C. K. Effects of Macromolecular Crowding on Amyloid Beta (16–22) Aggregation Using Coarse-Grained Simulations. *J. Phys. Chem. B* **2014,** *118* (47), 13513–13526.

48. Xia, Z.; Das, P.; Shakhnovich, E. I.; Zhou, R. Collapse of Unfolded Proteins in a Mixture of Denaturants. *J. Am. Chem. Soc.* **2012,** *134* (44), 18266–18274.

49. Das, P.; Xia, Z.; Zhou, R. Collapse of a Hydrophobic Polymer in a Mixture of Denaturants. *Langmuir* **2013,** *29* (15), 4877–4882.

50. Gnanakaran, S.; Nussinov, R.; García, A. E. Atomic-Level Description of Amyloid β-Dimer Formation. *J. Am. Chem. Soc.* **2006,** *128* (7), 2158–2159.

51. Sgourakis, N. G.; Yan, Y.; McCallum, S. A.; Wang, C.; Garcia, A. E. The Alzheimer's Peptides Aβ40 and 42 Adopt Distinct Conformations in Water: A Combined MD/NMR Study. *J. Mol. Biol.* **2007,** *368* (5), 1448–1457.

52.   Krone, M. G.; Hua, L.; Soto, P.; Zhou, R.; Berne, B. J.; Shea, J.-E. Role of Water in Mediating the Assembly of Alzheimer Amyloid-β Aβ16−22 Protofilaments. *J. Am. Chem. Soc.* **2008,** *130* (33), 11066–11072.

53.   Lin, Y.-S.; Bowman, Gregory R.; Beauchamp, Kyle A.; Pande, Vijay S. Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophys. J.* **2012,** *102* (2), 315–324.

54.   Viet, M. H.; Nguyen, P. H.; Ngo, S. T.; Li, M. S.; Derreumaux, P. Effect of the Tottori Familial Disease Mutation (D7N) on the Monomers and Dimers of Aβ40 and Aβ42. *ACS Chem. Neurosci.* **2013,** *4* (11), 1446–1457.

55.   Atsmon-Raz, Y.; Miller, Y. Non-Amyloid-β Component of Human α-Synuclein Oligomers Induces Formation of New Aβ Oligomers: Insight into the Mechanisms That Link Parkinson's and Alzheimer's Diseases. *ACS Chem. Neurosci.* **2016,** *7* (1), 46–55.

56.   Das, P.; Chacko, A. R.; Belfort, G. Alzheimer's Protective Cross-Interaction between Wild-Type and A2T Variants Alters Aβ42 Dimer Structure. *ACS Chem. Neurosci.* **2017,** *8*, 606–618.

57.   Fusco, G.; De Simone, A.; Arosio, P.; Vendruscolo, M.; Veglia, G.; Dobson, C. M. Structural Ensembles of Membrane-bound α-Synuclein Reveal the Molecular Determinants of Synaptic Vesicle Affinity. *Sci. Rep.* **2016,** *6*, No. 27125.

58.   Zheng, J.; Jang, H.; Ma, B.; Tsai, C.-J.; Nussinov, R. Modeling the Alzheimer Aβ17-42 Fibril Architecture: Tight Intermolecular Sheet-Sheet Association and Intramolecular Hydrated Cavities. *Biophys. J.* **2007,** *93* (9), 3046–3057.

59.   Zheng, J.; Jang, H.; Ma, B.; Nussinov, R. Annular Structures as Intermediates in Fibril Formation of Alzheimer Aβ$_{17-42}$. *J. Phys. Chem. B* **2008,** *112* (22), 6856–6865.

60.   Zhang, M.; Hu, R.; Liang, G.; Chang, Y.; Sun, Y.; Peng, Z.; Zheng, J. Structural and Energetic Insight into the Cross-Seeding Amyloid Assemblies of Human IAPP and Rat IAPP. *J. Phys. Chem. B* **2014,** *118* (25), 7026–7036.

61.   Zhang, M.; Hu, R.; Chen, H.; Chang, Y.; Gong, X.; Liu, F.; Zheng, J. Interfacial interaction and lateral association of cross-seeding assemblies between hIAPP and rIAPP oligomers. *Phys. Chem. Chem. Phys.* **2015,** *17* (16), 10373–10382.

62.   Zhang, M.; Hu, R.; Chen, H.; Chang, Y.; Ma, J.; Liang, G.; Mi, J.; Wang, Y.; Zheng, J. Polymorphic cross-seeding amyloid assemblies of amyloid-[small beta] and human islet amyloid polypeptide. *Phys. Chem. Chem. Phys.* **2015,** *17* (35), 23245–23256.

63. Anand, P.; Nandel, F. S.; Hansmann, U. H. E. The Alzheimer β-amyloid (A β 1–39) dimer in an implicit solvent. *J. Chem. Phys.* **2008,** *129* (19), No. 195102.

64. Han, M.; Hansmann, U. H. E. Replica exchange molecular dynamics of the thermodynamics of fibril growth of Alzheimer's Aβ42 peptide. *J. Chem. Phys.* **2011,** *135* (6), No. 065101.

65. Xi, W.; Wang, W.; Abbott, G.; Hansmann, U. H. E. Stability of a Recently Found Triple-β-Stranded Aβ1–42 Fibril Motif. *J. Phys. Chem. B* **2016,** *120* (20), 4548–4557.

66. Miller, Y.; Ma, B.; Nussinov, R. Polymorphism of Alzheimer's Aβ$_{17-42}$ (p3) Oligomers: The Importance of the Turn Location and Its Conformation. *Biophys. J.* **2009,** *97* (4), 1168–1177.

67. Miller, Y.; Ma, B.; Nussinov, R. The Unique Alzheimer's β-Amyloid Triangular Fibril Has a Cavity along the Fibril Axis under Physiological Conditions. *J. Am. Chem. Soc.* **2011,** *133* (8), 2742–2748.

68. Baram, M.; Atsmon-Raz, Y.; Ma, B.; Nussinov, R.; Miller, Y. Amylin-Aβ oligomers at atomic resolution using molecular dynamics simulations: a link between Type 2 diabetes and Alzheimer's disease. *Phys. Chem. Chem. Phys.* **2016,** *18* (4), 2330–2338.

69. DeMarco, M. L.; Daggett, V. From conversion to aggregation: Protofibril formation of the prion protein. *Proc. Natl. Acad. Sci. U.S.A.* **2004,** *101* (8), 2293–2298.

70. Wei, G.; Mousseau, N.; Derreumaux, P. Computational Simulations of the Early Steps of Protein Aggregation. *Prion* **2007,** *1* (1), 3–8.

71. Kalé, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *J. Comput. Phys.* **1999,** *151* (1), 283–312.

72. Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Chem.* **2004,** *25* (13), 1584–1604.

73. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The J. Chem. Phys.* **1983,** *79* (2), 926–935.

74. Carballo-Pacheco, M.; Strodel, B. Comparison of force fields for Alzheimer's A β42: A case study for intrinsically disordered proteins. *Protein Sci.* **2016**, *26*, 174–185.

75.   Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics* **1995,** *103* (11), 4613–4621.

76.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103* (19), 8577–8593.

77.   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23* (3), 327–341.

78.   Jana, A. K.; Sengupta, N. Aβ self-association and adsorption on a hydrophobic nanosurface: competitive effects and the detection of small oligomers via electrical response. *Soft Matter* **2015,** *11* (2), 269–279.

79.   Tomaselli, S.; Esposito, V.; Vangone, P.; van Nuland, N. A. J.; Bonvin, A. M. J. J.; Guerrini, R.; Tancredi, T.; Temussi, P. A.; Picone, D. The α-to-β Conformational Transition of Alzheimer's Aβ-(1–42) Peptide in Aqueous Media is Reversible: A Step by Step Conformational Analysis Suggests the Location of β Conformation Seeding. *ChemBioChem* **2006,** *7* (2), 257–267.

80.   Osapay, K.; Case, D. A. A new analysis of proton chemical shifts in proteins. *J. Am. Chem. Soc.* **1991,** *113* (25), 9436–9444.

81.   Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009,** *30* (13), 2157–2164.

82.   Heinig, M.; Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **2004,** *32* (2), W500–W502.

83.   Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996,** *14* (1), 33–38.

84.   Vergara-Jaque, A.; Comer, J.; Monsalve, L.; González-Nilo, F. D.; Sandoval, C. Computationally Efficient Methodology for Atomic-Level Characterization of Dendrimer–Drug Complexes: A Comparison of Amine- and Acetyl-Terminated PAMAM. *J. Phys. Chem. B* **2013,** *117* (22), 6801–6813.

85.   Gorham Jr, Ronald D.; Rodriguez, W.; Morikis, D. Molecular Analysis of the Interaction between Staphylococcal Virulence Factor Sbi-IV and Complement C3d. *Biophys. J.* **2014,** *106* (5), 1164–1173.

86.   Fernández, D. P.; Mulev, Y.; Goodwin, A. R. H.; Sengers, J. M. H. L. A Database for the Static Dielectric Constant of Water and Steam. *J. Phys. Chem. Ref. Data* **1995,** *24* (1), 33–70.

87. Zhang, L.; Xiao, X.; Yuan, Y.; Guo, Y.; Li, M.; Pu, X. Probing Immobilization Mechanism of alpha-chymotrypsin onto Carbon Nanotube in Organic Media by Molecular Dynamics Simulation. *Sci. Rep.* **2015,** *5*, No. 9297.

88. Lerbret, A.; Bordat, P.; Affouard, F.; Hédoux, A.; Guinet, Y.; Descamps, M. How Do Trehalose, Maltose, and Sucrose Influence Some Structural and Dynamical Properties of Lysozyme? Insight from Molecular Dynamics Simulations. *J. Phys. Chem. B* **2007,** *111* (31), 9410–9420.

89. Lerbret, A.; Affouard, F.; Bordat, P.; Hédoux, A.; Guinet, Y.; Descamps, M. Molecular dynamics simulations of lysozyme in water/sugar solutions. *Chem. Phys.* **2008,** *345* (2–3), 267–274.

90. Zhang, N.; Liu, F.-F.; Dong, X.-Y.; Sun, Y. Molecular Insight into the Counteraction of Trehalose on Urea-Induced Protein Denaturation Using Molecular Dynamics Simulation. *J. Phys. Chem. B* **2012,** *116* (24), 7040–7047.

91. Paul, S.; Paul, S. Molecular Insights into the Role of Aqueous Trehalose Solution on Temperature-Induced Protein Denaturation. *J. Phys. Chem. B* **2015,** *119* (4), 1598–1610.

92. Chau, P. L.; Hardwick, A. J. A new order parameter for tetrahedral configurations. *Mol. Phys.* **1998,** *93* (3), 511–518.

93. Errington, J. R.; Debenedetti, P. G. Relationship between structural order and the anomalies of liquid water. *Nature* **2001,** *409* (6818), 318–321.

94. Zhang, S.; Iwata, K.; Lachenmann, M. J.; Peng, J. W.; Li, S.; Stimson, E. R.; Lu, Y. a.; Felix, A. M.; Maggio, J. E.; Lee, J. P. The Alzheimer's Peptide Aβ Adopts a Collapsed Coil Structure in Water. *J. Struct. Biol.* **2000,** *130* (2), 130–141.

95. Lim, K. H.; Collver, H. H.; Le, Y. T. H.; Nagchowdhuri, P.; Kenney, J. M. Characterizations of distinct amyloidogenic conformations of the Aβ (1–40) and (1–42) peptides. *Biochem. Biophys. Res. Commun.* **2007,** *353* (2), 443–449.

96. Bernstein, S. L.; Wyttenbach, T.; Baumketner, A.; Shea, J.-E.; Bitan, G.; Teplow, D. B.; Bowers, M. T. Amyloid β-Protein: Monomer Structure and Early Aggregation States of Aβ42 and Its Pro19 Alloform. *J. Am. Chem. Soc.* **2005,** *127* (7), 2075–2084.

97. Baumketner, A.; Bernstein, S. L.; Wyttenbach, T.; Bitan, G.; Teplow, D. B.; Bowers, M. T.; Shea, J.-E. Amyloid β-protein monomer structure: A computational and experimental study. *Protein Sci.* **2006,** *15* (3), 420–428.

98. Lee, C.; Ham, S. Characterizing amyloid-beta protein misfolding from molecular dynamics simulations with explicit water. *J. Comput. Chem.* **2011,** *32* (2), 349–355.

99. Jana, Asis K.; Sengupta, N. Adsorption Mechanism and Collapse Propensities of the Full-Length, Monomeric Aβ1-42 on the Surface of a Single-Walled Carbon Nanotube: A Molecular Dynamics Simulation Study. *Biophys. J.* **2012,** *102* (8), 1889–1896.

100. Jana, A. K.; Jose, J. C.; Sengupta, N. Critical roles of key domains in complete adsorption of A[small beta] peptide on single-walled carbon nanotubes: insights with point mutations and MD simulations. *Phys. Chem. Chem. Phys.* **2013,** *15* (3), 837–844.

101. Nasica-Labouze, J.; Nguyen, P. H.; Sterpone, F.; Berthoumieu, O.; Buchete, N.-V.; Coté, S.; De Simone, A.; Doig, A. J.; Faller, P.; Garcia, A.; Laio, A.; Li, M. S.; Melchionna, S.; Mousseau, N.; Mu, Y.; Paravastu, A.; Pasquali, S.; Rosenman, D. J.; Strodel, B.; Tarus, B.; Viles, J. H.; Zhang, T.; Wang, C.; Derreumaux, P. Amyloid β Protein and Alzheimer's Disease: When Computer Simulations Complement Experimental Studies. *Chem. Rev.* **2015,** *115* (9), 3518–3563.

102. Barz, B.; Olubiyi, O. O.; Strodel, B. Early amyloid β-protein aggregation precedes conformational change. *Chem. Commun.* **2014,** *50* (40), 5373–5375.

103. Nagel-Steger, L.; Owen, M. C.; Strodel, B. An Account of Amyloid Oligomers: Facts and Figures Obtained from Experiments and Simulations. *Chembiochem* **2016,** *17* (8), 657–676.

104. Ono, K.; Condron, M. M.; Teplow, D. B. Structure–neurotoxicity relationships of amyloid β-protein oligomers. *Proc. Natl. Acad. Sci. U.S.A.* **2009,** *106* (35), 14745–14750.

105. Miller, Y.; Ma, B.; Nussinov, R. Polymorphism in Alzheimer Aβ Amyloid Organization Reflects Conformational Selection in a Rugged Energy Landscape. *Chem. Rev.* **2010,** *110* (8), 4820–4838.

106. Shankar, G. M.; Li, S.; Mehta, T. H.; Garcia-Munoz, A.; Shepardson, N. E.; Smith, I.; Brett, F. M.; Farrell, M. A.; Rowan, M. J.; Lemere, C. A.; Regan, C. M.; Walsh, D. M.; Sabatini, B. L.; Selkoe, D. J. Amyloid-[beta] protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nat. Med.* **2008,** *14* (8), 837–842.

107. Walsh, D. M.; Hartley, D. M.; Kusumoto, Y.; Fezoui, Y.; Condron, M. M.; Lomakin, A.; Benedek, G. B.; Selkoe, D. J.; Teplow, D. B. Amyloid β-Protein Fibrillogenesis: structure and biological activity of protofibrillar intermediates. *J. Biol. Chem.* **1999,** *274* (36), 25945–25952.

108. Nuallain, B.; Freir, D. B.; Nicoll, A. J.; Risse, E.; Ferguson, N.; Herron, C. E.; Collinge, J.; Walsh, D. M. Amyloid β-Protein Dimers Rapidly Form Stable Synaptotoxic Protofibrils. *J. Neurosci.* **2010,** *30* (43), 14411–14419.

109. Villemagne, V. L.; Perez, K. A.; Pike, K. E.; Kok, W. M.; Rowe, C. C.; White, A. R.; Bourgeat, P.; Salvado, O.; Bedo, J.; Hutton, C. A. Blood-borne amyloid-β dimer correlates with clinical markers of Alzheimer's disease. *J. Neurosci.* **2010,** *30* (18), 6315–6322.

110. Chong, S.-H.; Ham, S. Atomic-level investigations on the amyloid-β dimerization process and its driving forces in water. *Phys. Chem. Chem. Phys.* **2012,** *14* (5), 1573–1575.

111. Zhu, X.; Bora, R. P.; Barman, A.; Singh, R.; Prabhakar, R. Dimerization of the Full-Length Alzheimer Amyloid β-Peptide (Aβ42) in Explicit Aqueous Solution: A Molecular Dynamics Study. *J. Phys. Chem. B* **2012,** *116* (15), 4405–4416.

112. Lv, Z.; Roychaudhuri, R.; Condron, M. M.; Teplow, D. B.; Lyubchenko, Y. L. Mechanism of amyloid β−protein dimerization determined using single−molecule AFM force spectroscopy. *Sci. Rep.* **2013,** *3*, No. 2880.

113. Lin, J.; Buettner, R.; Yuan, Y.-C.; Yip, R.; Horne, D.; Jove, R.; Vaidehi, N. Molecular dynamics simulations of the conformational changes in signal transducers and activators of transcription, Stat1 and Stat3. *J. Mol. Graphics Modell.* **2009,** *28* (4), 347–356.

114. Jose, J. C.; Chatterjee, P.; Sengupta, N. Cross Dimerization of Amyloid-β and αSynuclein Proteins in Aqueous Environment: A Molecular Dynamics Simulations Study. *PLoS One* **2014,** *9* (9), No. e106883.

115. Chong, S.-H.; Ham, S. Impact of chemical heterogeneity on protein self-assembly in water. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (20), 7636–7641.

116. Côté, S.; Laghaei, R.; Derreumaux, P.; Mousseau, N. Distinct Dimerization for Various Alloforms of the Amyloid-Beta Protein: Aβ1–40, Aβ1–42, and Aβ1–40(D23N). *J. Phys. Chem. B* **2012,** *116* (13), 4043–4055.

117. Gohlke, H.; Kiel, C.; Case, D. A. Insights into Protein–Protein Binding by Binding Free Energy Calculation and Free Energy Decomposition for the Ras–Raf and Ras–RalGDS Complexes. *J. Mol. Biol.* **2003,** *330* (4), 891–913.

118. Singer, M. A.; Lindquist, S. Multiple Effects of Trehalose on Protein Folding In Vitro and In Vivo. *Mol. Cell* **1998,** *1* (5), 639–648.

119. Liu, R.; Barkhordarian, H.; Emadi, S.; Park, C. B.; Sierks, M. R. Trehalose differentially inhibits aggregation and neurotoxicity of beta-amyloid 40 and 42. *Neurobiol. Dis.* **2005,** *20* (1), 74–81.

120. Ignatova, Z.; Gierasch, L. M. Chapter Twenty-One - Effects of Osmolytes on Protein Folding and Aggregation in Cells. In *Methods in Enzymology*, Dieter, H.; Helmut, S., Eds. Academic Press: 2007; Vol. Volume 428, pp 355–372.

121. Estrela, N.; Franquelim, H. G.; Lopes, C.; Tavares, E.; Macedo, J. A.; Christiansen, G.; Otzen, D. E.; Melo, E. P. Sucrose prevents protein fibrillation through compaction of the tertiary structure but hardly affects the secondary structure. *Proteins* **2015,** *83* (11), 2039–2051.

122. Urbanc, B.; Cruz, L.; Ding, F.; Sammond, D.; Khare, S.; Buldyrev, S. V.; Stanley, H. E.; Dokholyan, N. V. Molecular Dynamics Simulation of Amyloid β Dimer Formation. *Biophys. J.* **2004,** *87* (4), 2310–2321.

123. Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **1993,** *215* (6), 617–621.

124. Chong, S. H.; Ham, S. Interaction with the Surrounding Water Plays a Key Role in Determining the Aggregation Propensity of Proteins. *Angew. Chem., Int. Ed.* **2014,** *53* (15), 3961–3964.

125. Ma, J.; Komatsu, H.; Kim, Y. S.; Liu, L.; Hochstrasser, R. M.; Axelsen, P. H. Intrinsic Structural Heterogeneity and Long-Term Maturation of Amyloid β Peptide Fibrils. *ACS Chem. Neurosci.* **2013,** *4* (8), 1236–1243.

126. Arya, S.; Mukhopadhyay, S. Ordered Water within the Collapsed Globules of an Amyloidogenic Intrinsically Disordered Protein. *J. Phys. Chem. B* **2014,** *118* (31), 9191–9198.

127. Chong, S.-H.; Ham, S. Distinct Role of Hydration Water in Protein Misfolding and Aggregation Revealed by Fluctuating Thermodynamics Analysis. *Acc. Chem. Res.* **2015,** *48* (4), 956–965.

128. Dalal, V.; Arya, S.; Mukhopadhyay, S. Confined Water in Amyloid-Competent Oligomers of the Prion Protein. *ChemPhysChem* **2016,** *17* (18), 2804–2807.

129. Gupta, M.; Nayar, D.; Chakravarty, C.; Bandyopadhyay, S. Comparison of hydration behavior and conformational preferences of the Trp-cage mini-protein in different rigid-body water models. *Phys. Chem. Chem. Phys.* **2016,** *18* (48), 32796–32813.

130. Khatua, P.; Jose, J. C.; Sengupta, N.; Bandyopadhyay, S. Conformational features of the Aβ42 peptide monomer and its interaction with the surrounding solvent. *Phys. Chem. Chem. Phys.* **2016,** *18* (43), 30144–30159.

Structural Perturbations in the Prion Protein (PrP^C) Trigger Pathogenic Transformations
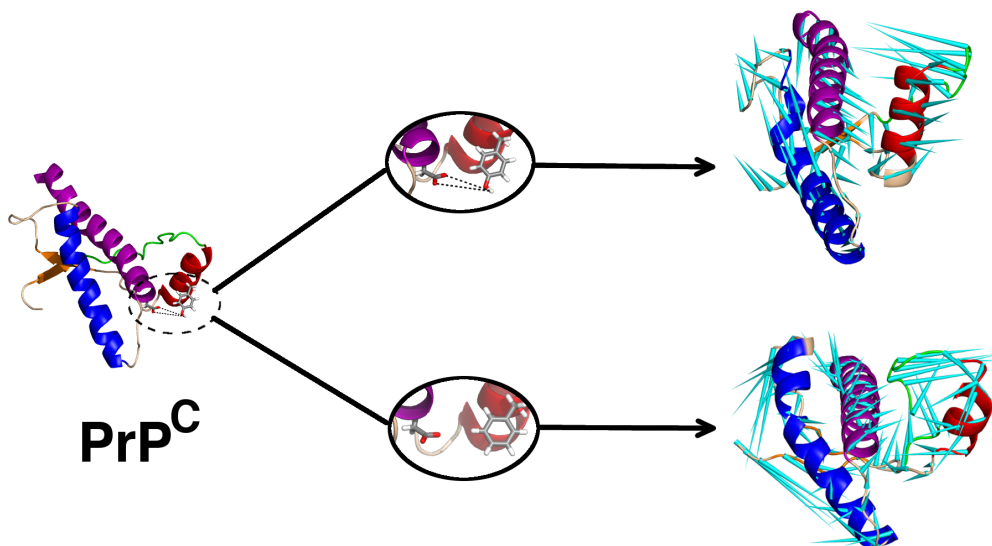
Based on:

5.1 Abstract

Conversion of the predominantly α-helical cellular prion protein (PrP^C) to the misfolded β-sheet enriched Scrapie form (PrP^{Sc}) is a critical event in prion disease pathogenesis. However, the conformational triggers that lead to the isoform conversion (PrP^C to PrP^{Sc}) remain obscure, and conjectures about the role of unusually hydrophilic, short helix H1 of the C-terminal globular domain in the transition are varied. Helix H1 is anchored to helix H3 via a few stabilizing polar interactions. We have employed fully atomistic molecular dynamics simulations to study the effects triggered by a minor perturbation in the network of these non-bonded interactions in PrP^C. The elimination of just one of the key H1-H3 hydrogen bonds led to a cascade of conformational changes that are consistent with those observed in partially unfolded intermediates of PrP^C, with pathogenic mutations and at low pH environments. Our analyses reveal that the perturbation results in enhanced conformational flexibility of the protein. The resultant enhancement in the dynamics leads to overall increased solvent exposure of the hydrophobic core residues and concomitant disruption of H1-H3 inter-domain salt bridge network. This study lends credence to the hypothesis that perturbing the cooperativity of the stabilizing interactions in the PrP^C globular domain can critically affect its dynamics and may lead to structural transitions of pathological relevance.

5.2   Introduction

Prions are proteinaceous infectious agents that cause fatal neurodegenerative diseases. These diseases, collectively termed as transmissible spongiform encephalopathies (or TSEs), include Creutzfeldt–Jakob disease, fatal familial insomnia, Gerstmann–Sträussler–Scheinker disease, Kuru in humans, scrapie in sheep, and bovine spongiform encephalopathy in cattle[1-2]. Such diseases occur as sporadic, acquired or inherited disorders hallmarked by the accumulation of an abnormal isoform of the prion protein[1,3-4]. According to the "*protein-only*" hypothesis, prion diseases arise from conformational changes in the normal isoform of cellular prion protein (PrP^C) to a protease-resistant, pathogenic form called Scrapie (PrP^{Sc})[3,5]. The PrP^{Sc} template has the ability to further recruit cellular isoforms of PrP^C and subsequently induce them to alter their conformation to the misfolded form, thus ensuring self-propagation[1,3,6].

The cellular prion protein, PrP^C, is a highly conserved cell surface glycoprotein of ~210 amino acids that is bound to the cell membrane by a glycosylphosphatidylinositol (GPI) anchor at its C-terminal residue[7]. The N-terminal domain of PrP^C is largely disordered, while the C-terminal region consists of a structured, globular domain. The globular domain, depicted in Figure 5.1b, is composed of two short β-strands, S1 (Y$_{128}$MLG$_{131}$) and S2 (V$_{161}$YYR$_{164}$), forming an anti-parallel β-sheet, and three α-helices, H1 (D$_{144}$YEDRYYRENM$_{154}$), H2 (N$_{173}$NFVHDCVNITIKQHTVTTTK$_{194}$) and H3 (E$_{200}$TDVKMMERVVEQMCITQ-YERESQAYYQR$_{228}$), with H2 and H3 covalently bridged by a disulfide-bond between Cys179 and Cys214[8]. The structure of the scrapie form, PrP^{Sc}, is poorly defined as its insolubility and high conformational heterogeneity eludes high-resolution structural analysis techniques[9]. However, it is known that PrP^{Sc} has substantial β-sheet content and reduced α-helical content compared to the native

form, which implies that the structural transition of PrP^C to PrP^{Sc} involves large conformational rearrangements[10].

The conformational changes that drive PrP^C to its pathogenic PrP^{Sc} counterpart, as well as the mechanistic details of PrP^{Sc} self-propagation have not yet been established[9]. Several models have been proposed for the structure of PrP^{Sc} that provide clues on the fibrillation and oligomerization processes[11-14]. However, a point of contention concerns the regions in PrP^C that harbor residue sequences vulnerable to conformational changes thereby facilitating initial seeding and further promoting fibrillar growth. Previous studies have suggested that the S1H1S2 region in PrP^C plays a crucial role in the structural transition[11-12,15-17]. Interestingly, based on experimental and computational studies, views have also emerged about the potential role of H2 and H3 in the conversion of PrP^C to PrP^{Sc} [18-23]. The PrP^C structure is stabilized by a core of hydrophobic residues, together with many salt-bridge and hydrogen-bonding interactions between the secondary structure elements[8,24-26]. A number of pathogenic mutations reside in the hydrophobic core of PrP^C that display varied effects on the mature protein[27]. Numerous studies indicate that perturbations to the structure, brought about by amino acid point mutations and changes in environmental conditions such as pH, temperature, and presence of denaturants can substantially destabilize the PrP^C conformer, affect its thermodynamic stability and induce structural transformations of pathological significance[28-31].

In the globular domain of PrP^C, helix H1 has certain intriguing properties[32] and its plausible roles in triggering the onset of TSEs has been the central topic in a series of experimental and computational studies[33-41]. In contrast with other helices usually found in globular proteins, H1 is characterized by a large abundance of hydrophilic, solvent exposed residues, making it the most soluble of all the protein α-

helices found in the PDB[32]. Circular dichroism (CD) and Nuclear magnetic resonance (NMR) studies of the isolated forms of H1 report its extreme stability, and thereby suggest the lack of its role in PrP[Sc] generation[34]. Conversely, molecular modeling studies have proposed a model of PrP[Sc] aggregate, namely the β-nucleation model in which helix H1 unravels and favorably adopts β-sheet conformations[32]. These aggregates whose core consists of hydrophilic H1 components are mainly stabilized by the formation of inter-molecular salt-bridges between H1 residues. Many experimental and computational studies have postulated that a key molecular event in PrP[C] misfolding is the detachment of the S1H1S2 region from the H2H3 subdomain[17,42-45]. Further, evidence from computer simulations suggests that a partially unfolded intermediate of the globular prion domain that is a possible candidate for PrP aggregation, displays the detachment and high mobility of the H1 from the relatively stable core formed by helices H2 and H3[17]. Such an analogous behavior has also been observed in simulation studies of PrP[C] in response to low pH as well as in some disease-associated mutants[15-16,46-48]. These studies suggest that interactions between the helices H1 and H3 could be playing significant roles in the conformational propensities of PrP[C].

We point out that all mammalian species of the prion protein are highly conserved in sequence and architecture[8]. This is evident from a superimposition of PDB structures corresponding to different species[8,49-51] as shown in Figure 5.1a, as well as the RMSD and percent identity values reported in Table 5.1. Local differences in the backbone conformations are manifested in the C-terminal of helix H3, the loop between S2 and H2 and the orientation of H1. Helix H1, owing to its charged nature, does not contribute to the hydrophobic core and is engaged in a cluster of electrostatic interactions, stabilizing the tertiary structure. Three non-local salt bridges E146–

K204, E146–R208 and R156–D202 connect the stable H2H3 core with the more flexible domain consisting of H1 and strands S1 and S2. The interface of H1 (C-terminal) and H3 consists of three polar residues Y149, Y150 and N153. Of these residues, residue Y149 is involved in an inter-helix side chain (H1H3) hydrogen-bonding interaction with residue D202. Residue Y150 forms a hydrogen bond with P137 that is located in the S1H1 loop region, while residue N153 forms an intra-helical backbone hydrogen bond with Y149. These interactions involving H1 residues are depicted in Figure 5.1b. We note here that although the hydrogen bond Y149–D202 is not reported in the NMR structure of human prion protein used in this work, it is highly likely to form the bond during the course of an unbiased MD simulation.

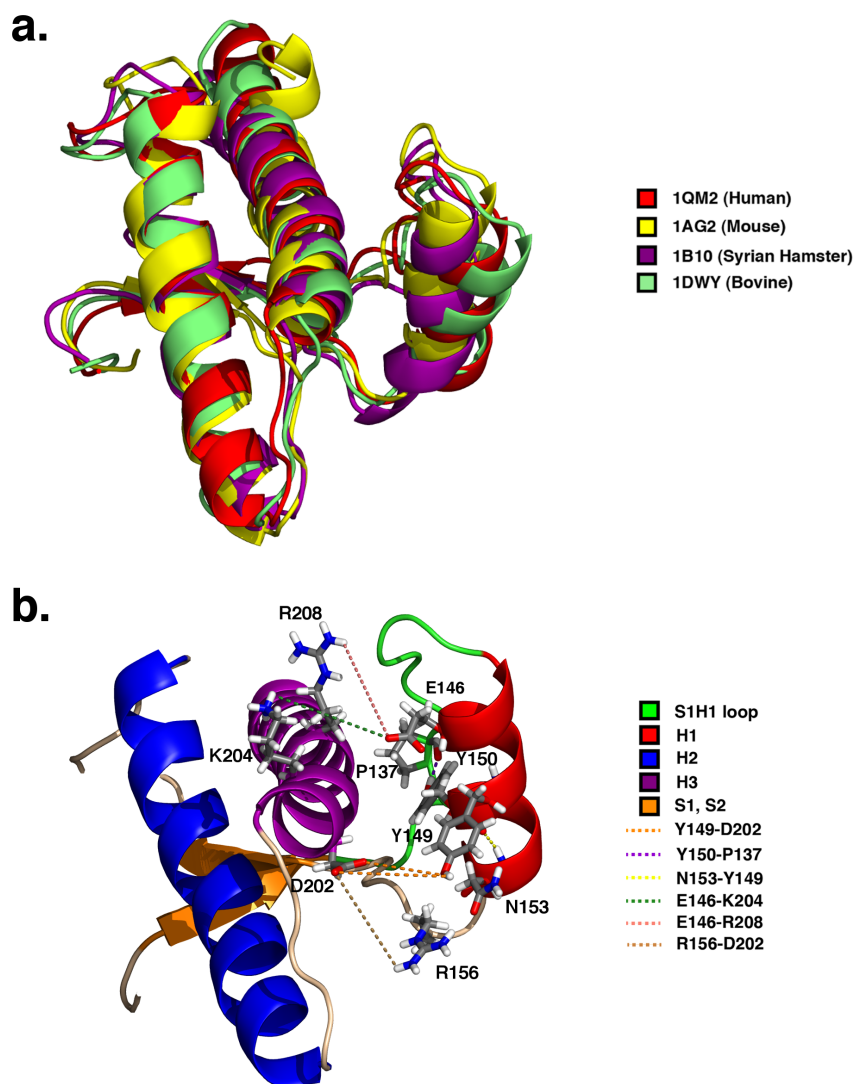**Figure 5.1** (a) Superimposed structures of prion protein of four different species: Human, Mouse, Syrian Hamster and Bovine PrP. (b) Charged interactions of helix H1 residues with H3 and the S1H1 loop. These include hydrogen bond interactions: Y149–D202, Y150–P137 and N153–Y149; and salt bridge interactions: E146–K204, E146–R208 and R156–D202. The residues are represented as sticks and the interactions as dashed lines.

PDB ID	1QM2	1AG2	1B10	1DWY
1QM2	–	2.612	2.156	1.371
1AG2	80.73	–	2.065	2.007
1B10	85.58	86.11	–	1.614
1DWY	91.43	83.33	86.67	–

Table 5.1 Structural superimposition of mammalian prion proteins of four species 1QM2: Human, 1AG2: Mouse, 1B10: Syrian Hamster, 1DWY: Bovine PrP. The upper triangular half of the matrix is RMSD (Å) and the lower triangular half is percent identity among pairwise PrP structures.

Molecular dynamics (MD) simulations have been extensively used to provide an atomistic understanding of the structural, dynamical and self-assembly propensities of amyloidogenic proteins[13,52-57]. In this study, using fully atomistic MD simulations, we have studied the conformational and dynamical implications of a minor perturbation of H1 interactions, namely the loss of a single side chain-side chain hydrogen bond Y149–D202, that links H1 with the rest of the globular domain of PrP^C. This perturbation was brought about by the replacement of residue Y149 with phenylalanine. Phenylalanine is identical to tyrosine except for the absence of the –OH group, which thus effectively obliterates the possibility of forming the Y149–D202 hydrogen bond in the artificially mutated system Y149F. Moreover, this replacement largely bypasses volume constraints without drastically affecting the structure of the protein. Such a minor perturbation could also be induced by stochastic, thermal or dynamical fluctuations or alterations to the solvent conditions. With 0.5 μs of simulated data in each case and the dominant clusters elicited from

Principal Component Analysis (PCA), we have evaluated and compared the conformational and dynamical behavior of the globular C-terminal domain of PrP^C in its native form (Figure 5.2a) with that of the singly mutated form with the Y149F mutation in H1 (Figure 5.2b).
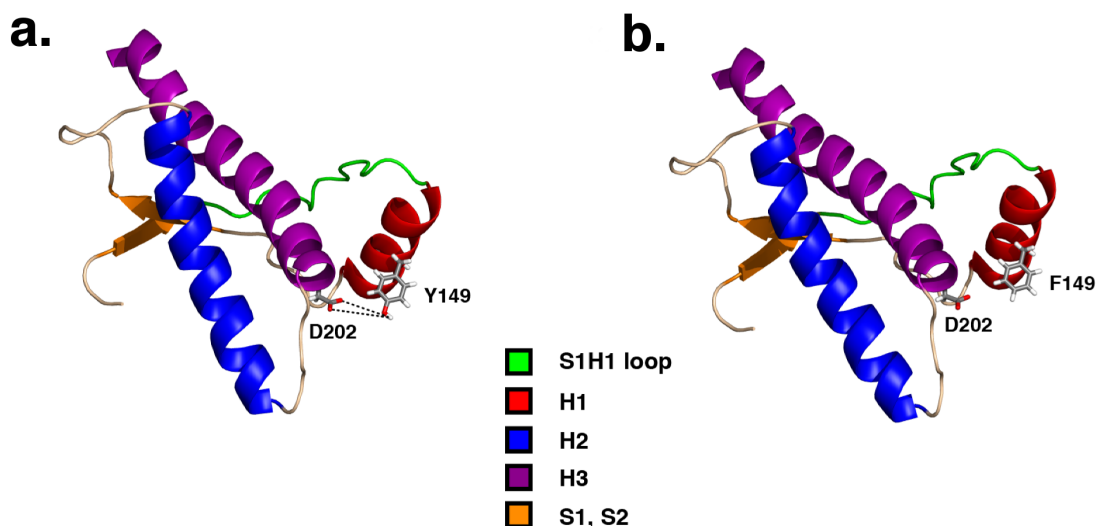
Figure 5.2 (a) Structure of the C-terminal globular domain (residues 125–228) of Wild type human PrP^C indicating the Y149–D202 hydrogen bond (b) Mutant Y149F, in which residue Y149 in H1 is replaced by a hydrophobic residue phenylalanine.

Our analyses show that this apparently small perturbation in H1-H3 interaction can result in conformational changes causing an overall weakening of the structural stability of PrP^C. The hydration of the hydrophobic core and weakening of H1-H3 salt-bridge propensities due to the perturbation suggests a cooperative interplay in the network of these associations in maintaining structural stability of the protein. Interestingly, the decay of the internal associations resulted in marginal decrease in the degree of helicity, and a small overall increase in the length of the β-sheet domains. These changes in structure and dynamics are in agreement with the

conformational changes observed at low pH, high temperature environments and in response to pathogenic point mutations[15-16,30-31,46-48]. Our results suggest that the origins of PrP^C to PrP^{Sc} transition may be present in stochastic perturbations to the native network of interactions that stabilize the PrP form.

5.3   Methods

5.3.1   System Setup and MD Simulations

The NMR structure of the C-terminal domain (residues 125–228) of human PrP^C (PDB ID: 1QM2) determined by *Zahn et al.*[8] were chosen as the initial structure for wild type prion protein (hereafter WT). A variant of the protein was obtained by replacing the Tyr residue at position 149 of the WT protein with a residue of Phe, to yield the Y149F system. The NH^{3+} and COO^- groups were added to the N and C-termini of the protein and three Na^+ counterions were used to neutralize the systems. Each system was solvated explicitly in a cubic box containing approximately 9600 TIP3P water molecules[58] and simulated under periodic boundary conditions. The systems were simulated using the NAMD2.8 simulation package[59] with the CHARMM22 all-atom force field with CMAP correction[60]. A time step of 2 fs was used. A constant temperature of 310 K was maintained using Langevin dynamics with a collision frequency of $1ps^{-1}$. A pressure of 1 atmosphere was maintained using the Nosé-Hoover algorithm[61]. Covalent bonds involving hydrogen atoms were constrained using the SHAKE algorithm[62]. Long-range electrostatic interactions were computed by particle-mesh Ewald (PME) method[63]. The cutoff for non-bonded interactions was 12 Å, with smooth truncation starting from 10 Å. The systems were initially energy minimized for 15000 steps on the basis of the conjugate gradient method followed by simulations in the isothermal–isobaric ensemble for 100 ns each.

Five independent MD simulations were performed for WT and Y149F to generate trajectories of 100 ns each, amounting to a total of 0.5 μs of simulation time for each system. One control simulation trajectory was generated for each of the systems with the same simulation protocols described above, using the AMBER force field[64].

5.3.2 Trajectory Analysis

5.3.2.1 Principal Component Analysis

Principal component analysis (PCA) of the combined trajectories of WT and Y149F was performed to capture the motional complexity of the systems using the Carma program[65]. PCA is a widely used technique to obtain functionally relevant collective motions from MD simulation ensembles[66-70]. The analysis is based on the diagonalization of the variance-covariance matrix of atomic fluctuations along the simulation trajectory. The eigenvectors and eigenvalues obtained from the diagonalization represent the various modes of motions and can be used to cluster the ensemble of structures. The eigenvectors corresponding to the largest eigenvalues are the 'Principal Components' (PC) that describe the largest amplitude collective motions of the protein. The probability distributions of the first two principal components (PC1 and PC2) corresponding to the C_α atomic fluctuations are used to calculate the free energy using the following equation:

$$\Delta G = -k_B T \ln \left[\frac{p}{p_{\max}} \right] \qquad (5.1)$$

where k_B is the Boltzmann's constant; T is the absolute temperature; p is the probability distribution of the first two principal components; and $p_{\max}$ is the maximum probability.

The two extreme projections sampled along the trajectory on PC1 were further used to build porcupine plots that depict a graphical view of the dominant motion represented by the most populated cluster. In a porcupine plot, each C_α atom has a cone pointing in the direction of its motion along the trajectory; the length of the cone represents the amplitude of the motion. The plots were generated using PyMOL[71].

5.3.2.2   Dynamic Cross Correlation Analysis

To identify the protein regions whose motions are correlated with one another, the cross-correlation of the atomic fluctuations obtained for the most populated cluster was obtained from PCA. The pairwise cross-correlation coefficient, C_{ij}, for the displacement of all C_α atom pairs, i and j, is given by:

$$C_{ij} = \frac{\langle \Delta r_i \cdot \Delta r_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle \langle \Delta r_j^2 \rangle}} \tag{5.2}$$

where Δr_i and Δr_j are the displacement vectors of the atoms i and j, from their mean positions respectively, which is determined from all configurations in the cluster. All frames of the cluster are superimposed on the initial structure and a matrix of all atom-wise cross-correlations is generated. The elements of the matrix are displayed in a graphical representation termed as a "*dynamic cross-correlation map*" (DCCM). The value of $C_{ij} = 1$ for completely positively correlated motion while $C_{ij} = -1$ for completely negatively correlated motion. Positively correlated residues move in the same direction, whereas negatively correlated residues move in the opposite direction. The magnitude of the cross-correlations of fluctuations of backbone C_α atoms was calculated and plotted using the Bio3D Package[72].

5.3.2.3 Configurational Entropy

We have calculated the configurational entropy per C_α atom of the WT and Y149F systems using Schlitter's method[73] as implemented in the Carma[65] program. According to Schlitter's formula, the absolute entropy S is approximated as:

$$S_{abs} < S = \frac{1}{2} k_B \ln \det \left[1 + \frac{k_B T e^2}{\hbar^2} M\sigma \right]$$
(5.3)

Here, k_B is the Boltzmann's constant, $\hbar$ is Planck's constant reduced by 2π, T is the absolute temperature, e is the Euler value, M is the diagonal mass matrix of rank $3N$ and σ is the covariance matrix of the atomic positional fluctuations.

5.4   Results and Discussion

5.4.1   Conformational Integrity

We have studied conformational and dynamical changes in the PrP^C globular domain brought about by an elimination of an inter-domain hydrogen bond, namely Y149–D202, via a tyrosine to phenylalanine mutation at the residue position 149. In Figure 5.3a, we compare the probability distributions of the side chain interaction energies of residue pairs Y149–D202 and F149–D202, belonging to the native and the mutated systems, respectively. Mean non-bonded interaction energies between these residues in the WT and Y149F systems are −11.46 (±7.66) and −0.87 (±0.94) kcal mol^{-1}, respectively. Overall, the loss of the inter-domain H-bond between helices H1 and H3 causes a significant weakening of the side chain interactions. To compare the conformational stabilities, the root-mean-square-deviation (RMSD) of the backbone atoms was calculated for the two systems. In Figure 5.3b, we have plotted the RMSD of the backbone atoms of the WT and Y149F proteins relative to the initial structure

over the simulation timescale, averaged over the five independent runs. The average RMSD of the mutant Y149F is markedly higher than that of WT system over the entire timescale, indicating that compared to the WT, marked conformational changes occur in the mutant system. The mean RMSD values for the WT and the Y149F systems, averaged over the last 80 ns, are 3.22 ($\pm$0.10) Å and 3.97 ($\pm$0.18) Å, respectively. We further compared the cumulative configurational entropy per C_α atom as a measure of the extent of disorder of the two systems. As depicted in Figure 5.3c, the net configurational entropy of Y149F is larger than that of WT. The mean configurational entropy per C_α atom for the WT and the Y149F systems, averaged over the last 2 ns, is 32.94 ($\pm$0.12) and 34.13 ($\pm$0.13) $JK^{-1}mol^{-1}$, respectively. Thus, the loss of the single H1-H3 hydrogen bond induces fluctuations and conformational disorder in PrP^C.

We also evaluated the above characteristics for the control simulation generated with the AMBER force field, which is presented in Figure 5.4. The probability distribution of the side chain interaction energies, depicted in Figure 5.4a, indicates that the interaction strength of residue pair F149–D202 of the mutant is significantly reduced as compared to the WT residue pair Y149–D202. The mean interaction strengths in the WT and the Y149F systems are -13.765 ($\pm$6.96) and -0.92 ($\pm$0.84) kcal mol^{-1}, respectively, and therefore comparable to the values obtained with CHARMM. We note here that the RMSD trends obtained for AMBER trajectories show a higher deviation for the WT. However, similar to the corresponding data obtained with CHARMM, the configurational entropy per C_α atom of the mutant is higher than that of the WT system. The mean configurational entropy averaged over

the last 2 ns of the WT and Y149F systems simulated with AMBER are 36.55 ($\pm$0.09) and 38.84 ($\pm$0.12) $JK^{-1}mol^{-1}$ respectively.

Figure 5.3 (a) Probability distributions of interaction energy strength between residues Y149–D202 in WT and F149–D202 in Y149F system. (b) Backbone RMS deviations from the starting structure as a function of simulation time (c) Cumulative configurational entropy per C_α atom of the WT and Y149F systems.
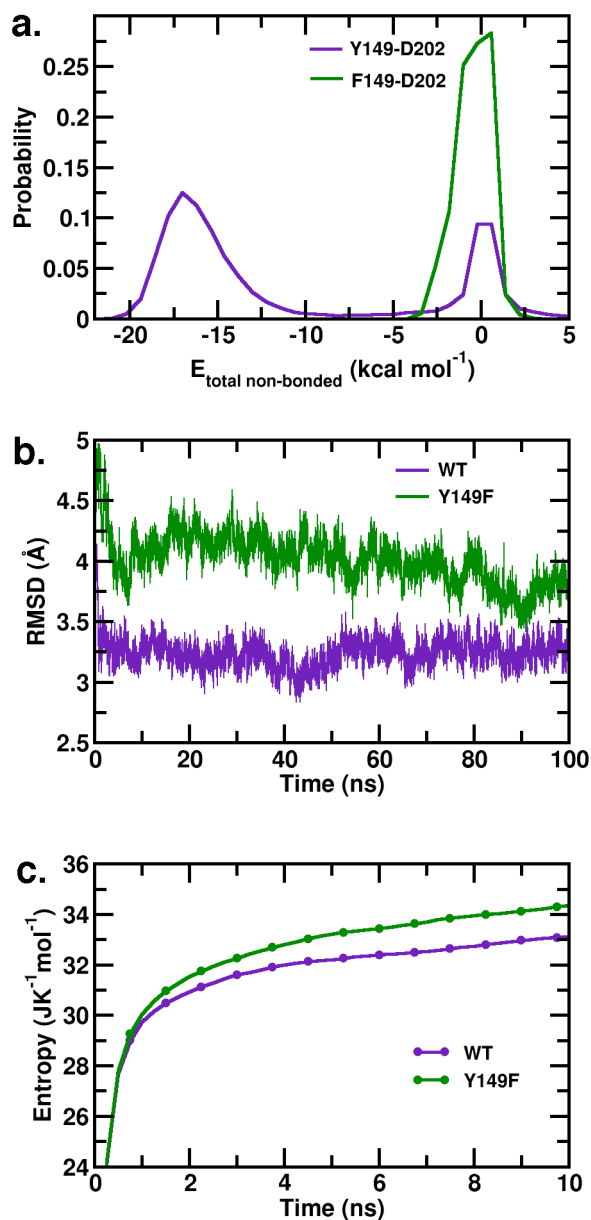
Figure 5.4 Control simulation, using the AMBER force field. (a) Probability distributions of interaction energy strength between residues Y149–D202 in WT and F149–D202 in Y149F system. (b) Backbone RMS deviations from the starting structure as a function of simulation time (c) Cumulative configurational entropy per Cα atom of the WT and Y149F systems.
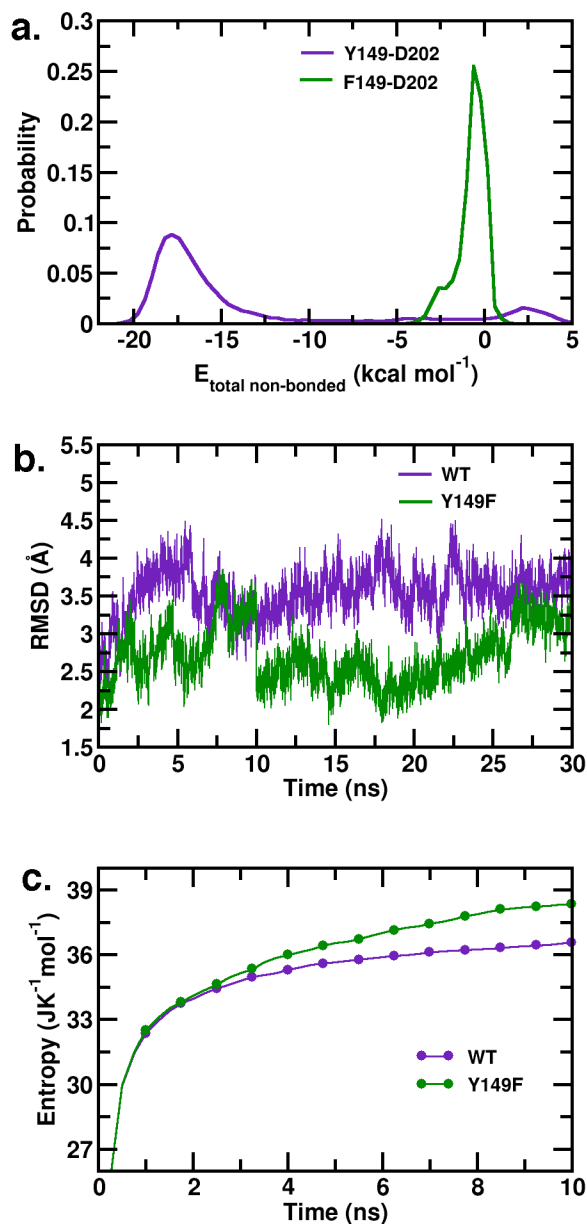
5.4.2 Principal Component Analysis (PCA) and Essential Dynamics

PCA, as described in *Methods*, was performed on the simulation ensembles of the WT and Y149F systems. The overall flexibility of the two systems was calculated by the trace of the diagonalized covariance matrix of the C_α atomic fluctuations. The trace values for the WT and Y149F systems were found to be 701.82 and 1042.92 Å^2, respectively, suggesting that there is higher flexibility in the collective motion of the protein in Y149F as compared to that of WT. In Figure 5.5a and b, we present the free energy landscape projected on the first (PC1) and second (PC2) principal components, and provide snapshots of representative conformations of the most populated cluster of the two systems. In all further analyses, we considered the most populated clusters obtained from the PCA of the simulated systems for comparing the structural and dynamical characteristics of PrP^C of the WT system and the system with induced perturbations in inter-domain interactions.
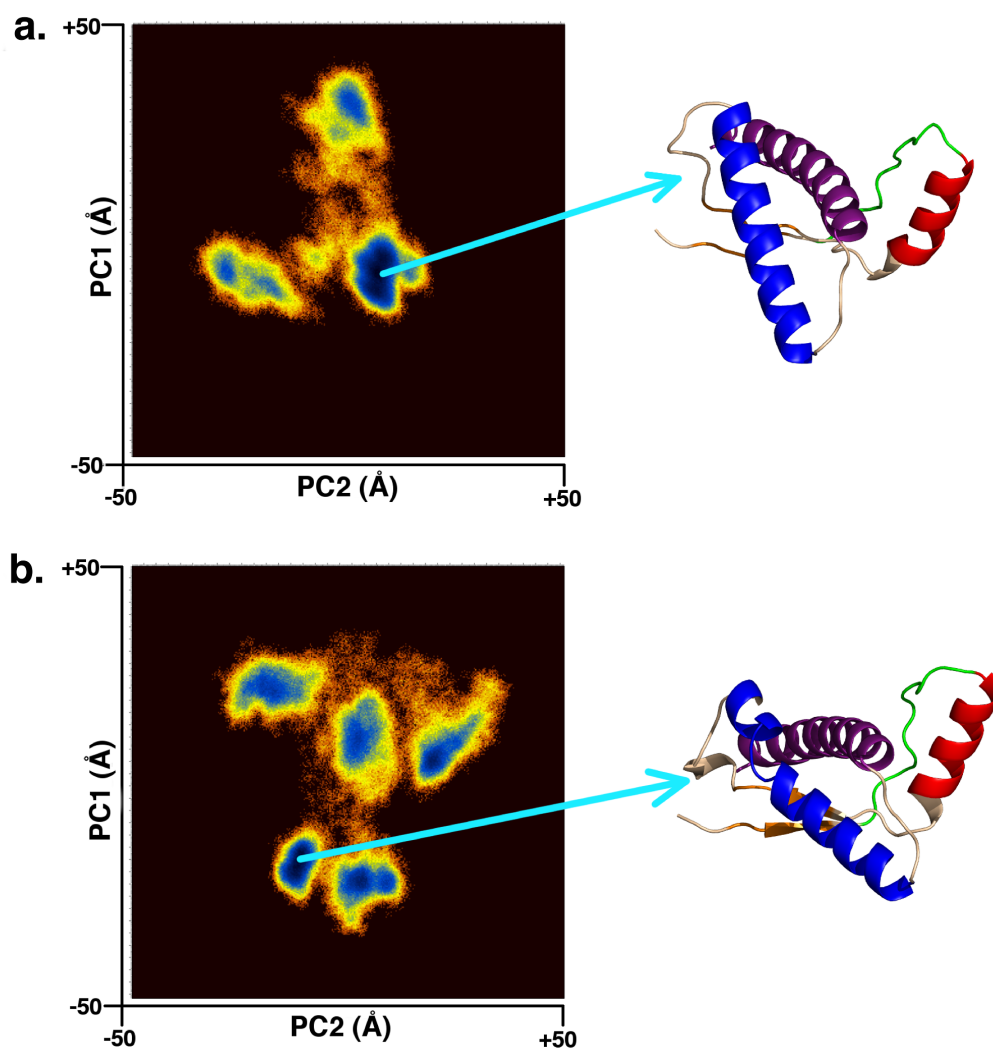
Figure 5.5 Free energy landscape on the plane defined by the first (PC1) and second (PC2) principal components for the a) WT ensemble, and b) the Y149F ensemble. Representative structures for the most populated cluster are depicted.

To identify key differences in the modes of motion in the WT and Y149F systems, we generated porcupine plots from the extreme projections on PC1 generated from the simulation ensembles; these are presented in Figures 5.6a and b. The unidirectional motion of the helices H1 and H3 indicates dynamical correlation between them in the WT system. However, this correlation is largely lost in the perturbed Y149F system, and the two helices are observed to move in opposite

directions. The S1H1 loop, preceding helix H1, exhibits large amplitude displacements in both the systems, albeit in opposite directions. In WT, the S1H1 loop moves in a direction towards the H2H3 sub domain with which it forms the hydrophobic core in PrP^C. In Y149F, on the other hand, it moves in a direction away from H2H3 subdomain. Moreover, the short, anti-parallel β-sheet linked by a hydrogen bond network shows concerted motion. We note here that earlier studies have reported the movement of H1 and the S1H1 loop away from H3[15-16,47-48]. However, we observed that S1H1 loop movement is opposite to that of H1. The helices H2 and H3 that together form bulk of the hydrophobic core do not move unidirectionally, but the magnitude of this displacement is low in both the systems. We remark here that H2 and H3 are connected by a disulfide bridge between residues Cys179 and Cys214. In Figure 5.7, we have plotted the distribution of the distance between the sulfur atoms of these residues. The mean distance between the two atoms increases from 4.99 (±0.36) Å in the WT, to 7.40 (±0.68) Å in the Y149F system. While the possibility of bond breakage cannot be assessed with classical simulations, this analysis indicates that the structural changes in the Y149F system markedly affects the stability attributed to the disulfide bridge.
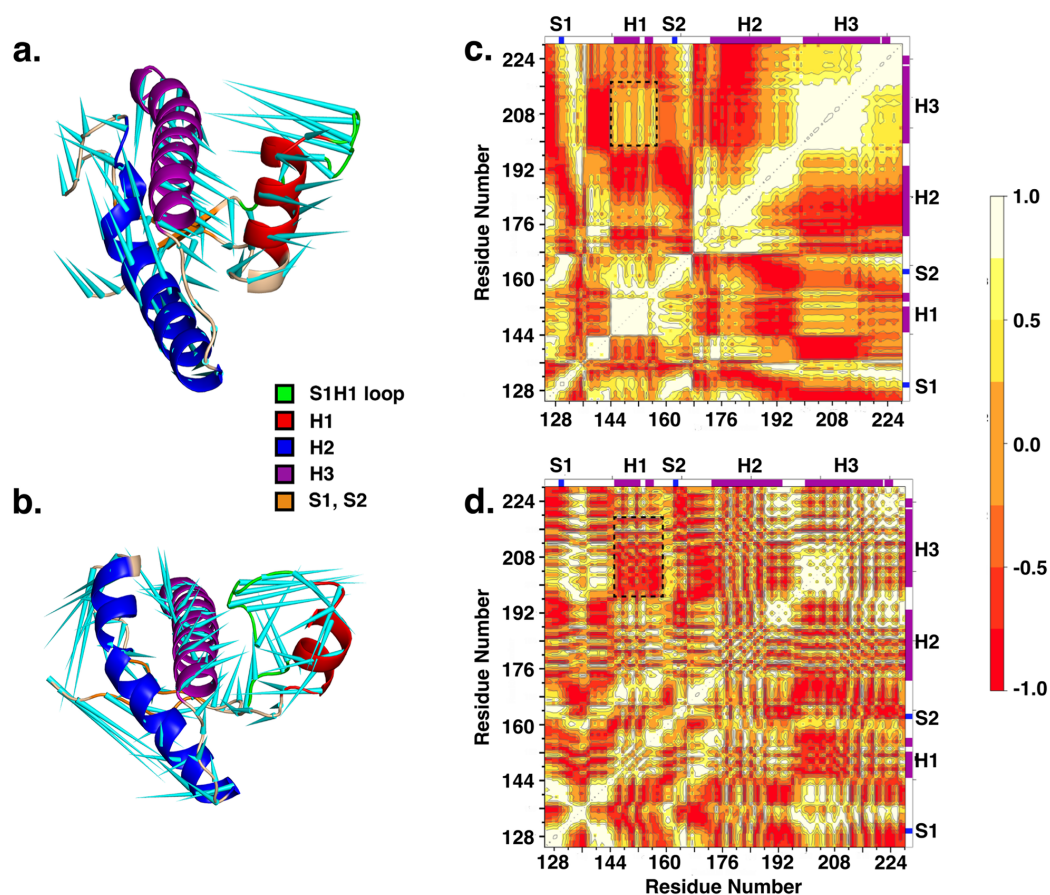
Figure 5.6. Porcupine plots of the first principal component (PC1) obtained from PCA depicting dominant motions of residues in (a) WT and (b) Y149F of the most populated cluster. The dominant motions in PC1 are illustrated as cones colored in cyan. The length of the arrows represents the amplitude of the cones while the direction indicates the direction of motion. Dynamic Cross-Correlation Map computed for the most populated cluster of c) WT and d) Y149F systems. Axes denote the residue numbers. The color scale for correlation and anti-correlation are shown at the right of each plot.
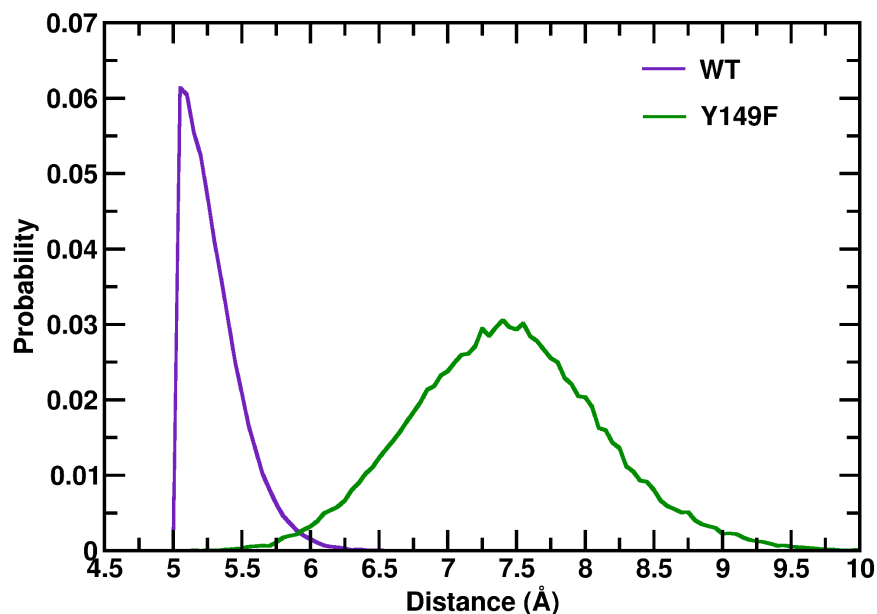
Figure 5.7 Distance between the sulfur atoms of the disulfide bond Cys179-Cys214 between helices H2-H3 in the most populated cluster for the WT and Y149F systems.

To further corroborate the observed dynamics, we quantified the interatomic cross correlations of the fluctuations in C_α atomic positions of the two systems. In Figure 5.6c and d, we have illustrated the Dynamic Cross-Correlation Map (DCCM) representing the correlated motions of WT and Y149F ensembles projected along PC1. In the WT, H1 is positively correlated with N-terminal of H3, in agreement with their concerted motion as a unit observed in the porcupine plot. On the other hand, in Y149F, H1 has a strong negative correlation with H3; interestingly, this correlation is of a higher degree with the N-terminal end of H3, which is the region of contact of H1 and H3. Further, helices H2 and H3 are anticorrelated with each other in WT and have less pronounced negative correlation in the mutant. The short, anti-parallel β-strands are positively correlated with each other in both the systems as they are connected by a hydrogen bond network and move in phase. These analyses show that small

perturbation in the network of interactions in the PrP^C domain can critically affect its flexibility and modes of internal motions.

5.4.3 Hydration of the Hydrophobic Core

In PrP^C, a tightly packed hydrophobic core consists of the twenty residues Met134, Pro137, Ile139, Phe141, Pro158, Val161, Phe175, Val176, Cys179, Val180, Ile184, Phe198, Val203, Met205, Met206, Val209, Met210, Cys213, Cys214 and Ile215.[24] These residues provide hydrophobic contacts between helices H2 and H3; between the S1H1 loop and H3, and between the anti-parallel β-sheet and H2 or H3. The hydrophobic core residues between helices H2–H3 and at the S1H1 loop–H3 interface are depicted in Figures 5.8a and b respectively. We note here that several pathogenic mutations are located in the hydrophobic core, and importantly, a number of studies relate the instability of the hydrophobic core to the early steps in the misfolding process[16,27,48]. Herein, we investigated how the dynamical changes induced by the minor perturbation in the H1–H3 interactions could affect the stability of the hydrophobic core. We first assessed the solvent-accessible surface area of the hydrophobic core, $SASA_{HC}$, calculated within the VMD package[74] by running a spherical probe of 1.8 Å radius over the protein surface, as a measure of its compactness and solvent exposure. In Figure 5.8c, we depict the probability distributions of $SASA_{HC}$ computed for the WT and Y149F. In Y149F, there is a clear shift of the $SASA_{HC}$ of the hydrophobic core towards higher values; the peak positions in the WT and Y149F systems are centered at ~425 and ~550 $Å^2$, respectively. We have plotted, in Figure 5.9, the SASA of the hydrophobic core residues, averaged over the independent trajectories in each system as a function of simulation time, along with corresponding standard deviations. Thus, the perturbation

caused by loss of a single H1–H3 hydrogen bond in Y149F effectively leads to about 30 % increase in the solvent exposure of the hydrophobic core. We also analyzed the probability distributions of SASA$_{HC}$ of the AMBER simulation trajectories. There is enhanced hydration of the Y149F system as compared to the WT, with the peak positions at ~325 and ~510 Å^2, respectively, which is comparable to the observation with CHARMM. We note here that the SASA$_{HC}$ increase is observed in Y149F despite the slightly higher hydrophobicity of phenylalanine compared to tyrosine. The increase, therefore, should be attributable to the increased conformational flexibility and the loss of structural coherence in the perturbed system.

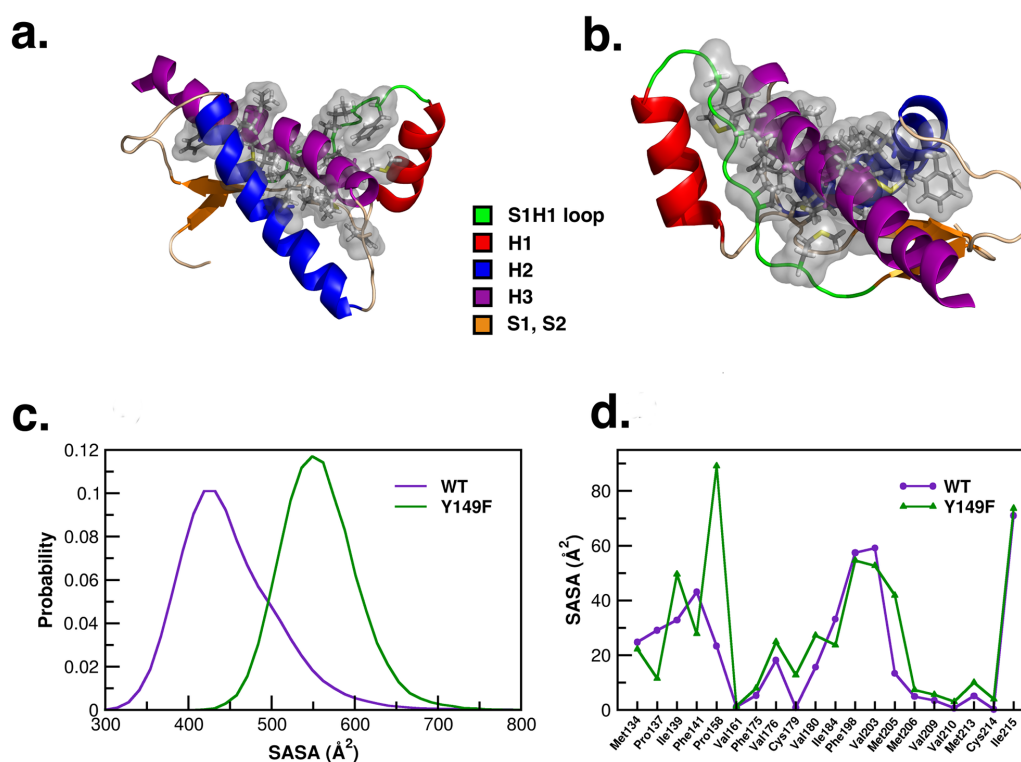**Figure 5.8** The hydrophobic core residues of PrP C-terminal globular domain at (a) H2-H3 interface (b) S1H1 loop-H3 interface. The residues are shown as translucent grey surface, with side chains represented as sticks. (c) Probability distributions of SASA of hydrophobic core residues calculated for cluster 1 obtained from PCA. (d) Per-residue side chain SASA of the hydrophobic core residues.
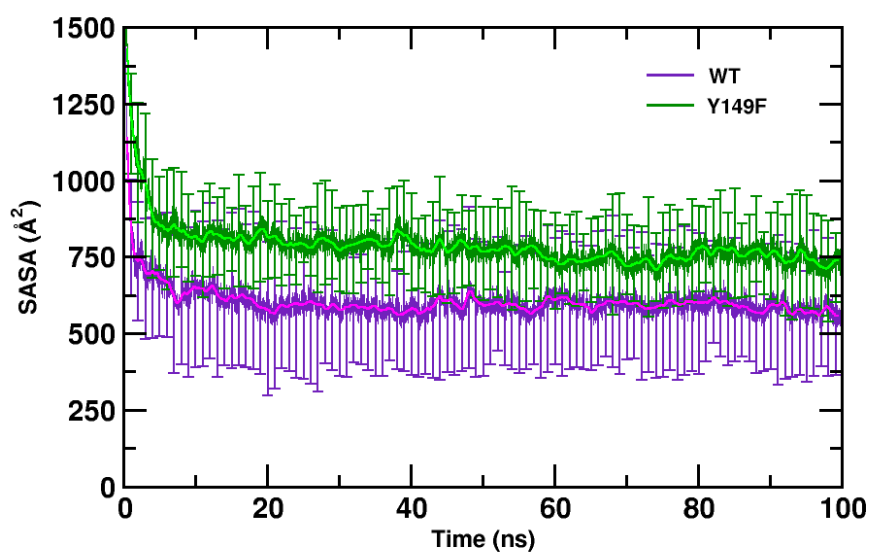
Figure 5.9 Mean SASA of the hydrophobic core residues, calculated over simulation time. The standard deviations show the spread of the values among multiple trajectories.

To elucidate the residue-level changes, we present in Figure 5.8d, the residue-wise mean solvent accessible surface area, or $SASA_{res}$, of the individual hydrophobic core residues analyzed for the WT and Y149F systems. The comparison reveals that the mean $SASA_{res}$ of residues Ile139, Pro158, Phe175, Val176, Cys179, Val180, Met205, Met206, Val209, Val210, Met213, Cys214 and Ile215 is higher in Y149F. Of these residues, Met205, Val209 and Met213 in H3 have hydrophobic interactions with residues of the S1H1 loop. This is commensurate with the observed decorrelated motion of H1 and H3 in Y149F that causes the exposure of the H3 residues linked by hydrophobic contacts with the S1H1 loop. Residues Phe175, Val176, Val180 and Ile184 of H2 have hydrophobic interactions with residues Val203, Met206, Val210 and Ile215 of H3. The increased $SASA_{res}$ of these residues shows that the stability of hydrophobic interactions in the H2H3 subdomain is affected due to increased solvent

192

exposure. Helices H2 and H3 are connected by a disulfide bond formed between residues Cys179–Cys214 that also shows enhanced SASA$_{res}$. This suggests that the packing of helices H2 and H3 in the hydrophobic core is perturbed significantly in the Y149F system. Residue P158, located in the loop between H1 and S2, shows remarkable increase in SASA$_{res}$ in Y149F as compared to the WT system. In the WT system, the side chain of P158 is oriented towards H3, which is part of the hydrophobic core, and thus shielded from the surrounding water. However, in Y149F, the enhanced dynamics and flexibility of the S1H1S2 region in a direction away from the hydrophobic core, results in exposure of the P158 side chain to solvent molecules. The analyses above shows that the dynamics of PrP^C resulting from the perturbation of H1–H3 interaction and the resulting mobilities of the structural elements causes marked decrease in the stability of the hydrophobic core.

In addition, we analyzed two hydration sites previously identified by De Simone *et al.*[75] with tightly bound structurally conserved waters that are necessary in maintaining local elements of the PrP^C fold. The water at site 1 mediates the interaction between the carbonyl of S132, the amide of V161 and the Q217 side chain oxygen; thus connecting three protein regions that belong to different secondary structure elements S1, S2 and H3, respectively. In Figure 5.10a, we have plotted the probability distributions of the distance between S132:O–V161:N in the dominant cluster of the two systems. The peak positions of the distribution are ~5.2 Å and ~7 Å for WT and Y149F systems, respectively. The increase in the distance between the backbones of S132 and V161 of ~2Å in Y149F indicates that the increased fluctuations and dynamics in this system destabilize the water-mediated interaction at site 1. Another identified hydration site, with long residence time (exceeding 1ns) is located at the end of strand S2 and H2. This site bridges the backbone oxygen of

residue F175 and the amide of R164; it also interacts with the side chain of D178. We analyzed the distance between the atoms of the F175:O–R164:N interaction; the probability distance distributions of this interaction are depicted in Figure 5.10b. Interestingly, the mutant exhibits a bimodal distribution with the higher peak position at ~5.5 Å, while the WT system has a peak position of ~7.5 Å. The reduced distance between the interacting atoms indicates that the water-mediated interaction at site 2 is strengthened in the mutant Y149F. We mention here that there is very little overlap of the hydrophobic core residues with the residues participating in these tightly bound hydration sites.
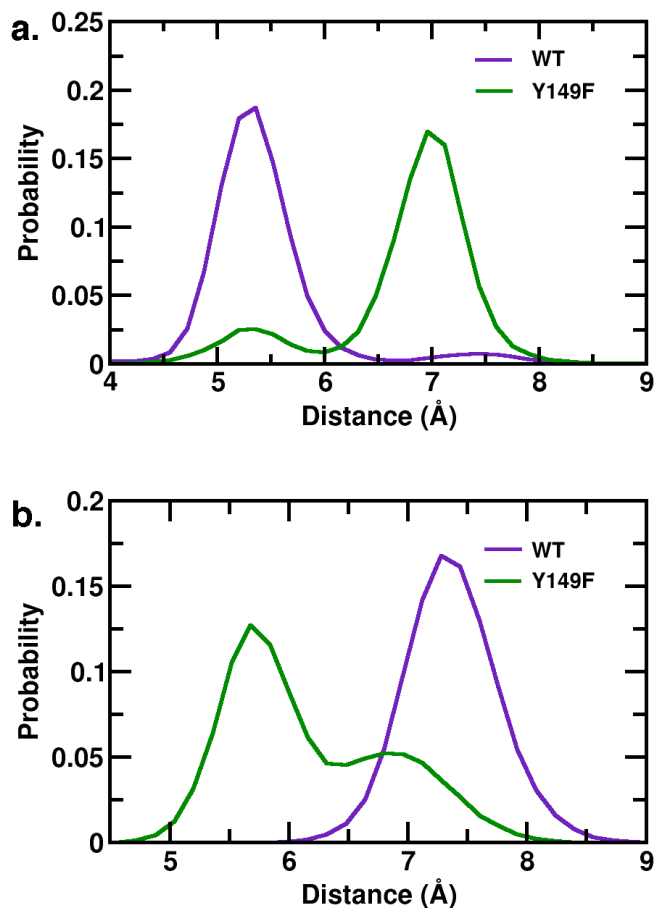
Figure 5.10 Probability distance distributions of atoms (a) S132:O–V161:N and (b) F175:O–R164:N belonging to hydration Site 1 and Site 2, respectively.

5.4.4   Inter-domain Salt Bridge Stability

The PrP^C structure contains a diverse set of salt bridges that have been proposed to play an important role in stabilizing secondary structural elements and maintaining the overall tertiary fold[26]. Hence, disruption of the salt bridge network may substantially destabilize the folded conformation. To understand how a minor perturbation of the interactions that tether H1 to the rest of the protein may affect the stability of the salt bridge network, we examined the three non-local salt-bridges that anchor H1 to H3: E146–K204, E146–R208 and R156–D202. Previous studies have asserted that these salt-bridges contribute to the stabilization of PrP^C and that their

abolition, either due to protonation of the acidic amino acids in response to low pH or by relevant genetic mutations, decreases the stability of PrP^C and favors misfolding[16,26-27,76-77]. In Figure 5.11a-c, we depict distributions of the salt bridge distances (d_{SB}) between each pair in the WT and the perturbed Y149F system. In Y149F, the d_{SB} distributions shift to higher values compared to the unmutated system for all the three salt-bridge pairs. In Table 5.2, we have reported the mean values of the inter-residue distances and interaction energies of these salt-bridge forming residue pairs in the most populated cluster of the WT and Y149F systems. The interactions for all the three salt-bridges are weakened in Y149F as compared to the WT system.

Salt-bridge	$< d >$ (Å)		$< E >$ (kcal mol^{-1})	
	WT	Y149F	WT	Y149F
E146-K204	10.40 (±4.95)	13.47 (±2.58)	-21.21 (±32.38)	-1.45 (±2.74)s
E146-R208	8.30 (±4.26)	16.16 (±1.43)	-31.11 (±37.33)	-0.045 (±0.08)
R156-D202	5.15 (±2.10)	8.47 (±1.31)	-59.61 (±30.79)	-7.59 (±6.46)

Table 5.2 Mean inter-residue distances, d and interaction energies, E of the salt-bridge forming residues of the most populated cluster of the WT and Y149F systems. Standard deviations are provided within braces
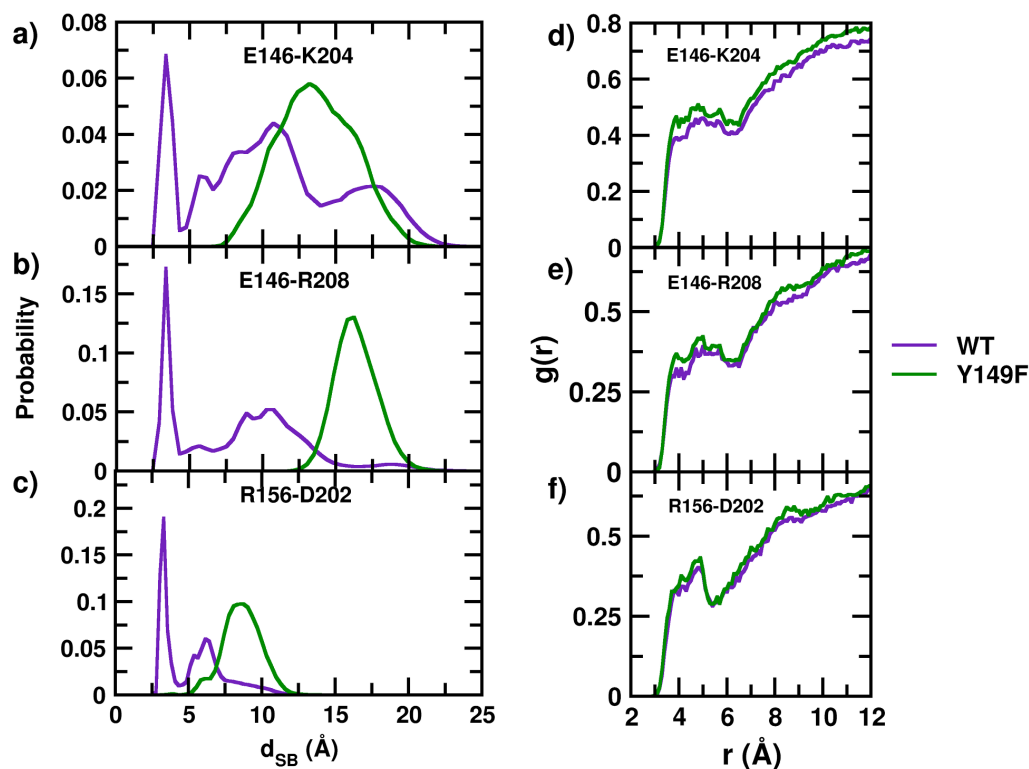
Figure 5.11 Salt-bridge analyses of the most populated cluster obtained from PCA. Distribution of the H1-H3 inter-domain salt-bridge distance, d_{SB}, in the most populated cluster obtained from PCA for the residue pairs (a) E146–K204, (b) E146–R208 and (c) E156–D202. Radial distribution functions (RDFs) calculated between the oxygens of the solvent water molecules, and the C_β of the salt-bridge forming residues (d) E146–K204, (e) E146–R208 and (f) R156–D202

We note here that the formation of stable salt-bridge networks is generally associated with a desolvation barrier[78-81]. Thus, the overall destabilization of the salt bridge network in Y149F may be commensurate with a higher local desolvation barrier resulting from the increased hydration accompanying the dynamical instability in this system. We probe this effect by comparing the protein-water radial distribution functions, $g(r)$, calculated between the oxygen atoms of solvent water molecules and the C_β atom of the salt-bridge forming residues for the WT and Y149F systems; these results are shown in Figures 5.11d-f. Commensurate with stronger salt-bridge

interaction in the E146–K204, E146–R208 and R156–D202 pairs in the WT, the first and second solvation peaks in $g(r)$ are marginally weaker in this system compared to Y149F. These analyses illustrate how a minor loss in the H1-H3 association results in major perturbations within the salt-bridge network of the folded domain of PrP^C.

5.4.5   Secondary Structural Propensity

We finally attempt to understand if the disruption of H1-H3 interactions and the resulting conformational dynamics may cascade into secondary structural effects relevant for prion propagation. In Figure 5.12, we compare the residue-wise helical and β-sheet propensities of dominant clusters in the WT and Y149F systems, obtained with the STRIDE algorithm[82] in VMD[74].
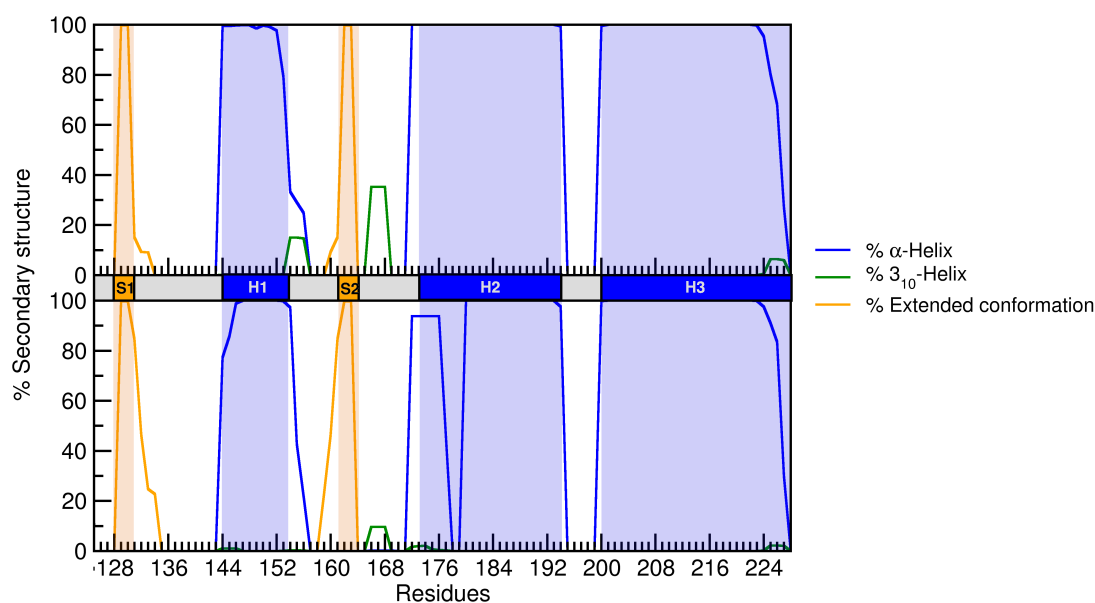

Figure 5.12 Percentage of secondary structure content per residue in the WT (upper panel) and Y149F (lower panel) systems. The secondary structure assignments were done by STRIDE algorithm in VMD.

The structure of helix H1 is found to be stable in both the WT and Y149F systems. However, the region near the C-terminal end of H1, consisting of residues $M_{154}YR_{156}$, is labile. In the WT system, these residues formed a 3_{10}-helix in about 20% of the conformations within the most populated cluster. On the other hand, in the Y149F system, these residues adopt α-helical conformations. We note here that increased disorder is known to exist in residues $D_{167}EYSN_{171}$ in the S2H2 loop[8]. Our analysis shows that these residues exhibit transient 3_{10} helicity in both systems, but with a distinctly lower propensity in Y149F. We further note that experimental and computational studies have shown that residues at the C-terminal end of H2 are frustrated in their helical state[8]. However, surprisingly, we observe here that the N-terminal end of H2 up to the disulfide-bond forming residue, Cys179, undergoes structural loss in Y149F system, while the C-terminal region remains fairly stable. Furthermore, we observed that the lengths of the β-strands in Y149F were enhanced relative to the WT system. This was corroborated with an analysis of inter-strand backbone hydrogen bonds in S1 and S2; the donor-acceptor distance threshold for a hydrogen bond was 4.0 Å and the angle connecting the donor, hydrogen and the acceptor atoms was more than 135º. In the starting structure, the short anti-parallel β-sheet consists of four inter-strand backbone hydrogen bonds, namely, M129:N–Y163:O, M129:O–Y163:N, G131:N–V161:O and M134:N–N159:O. These hydrogen bonds are fairly stable in the WT system. In addition, two new hydrogen bonds, V161:N–G131:O and A133:N–N159:O, are observed in the Y149F system, resulting in a slight elongation of the native β-sheet. The glycine at the end of strand S1, which promotes a conserved β-bulge, was proposed to be a "negatively designed element"[83] to prevent edge-to-edge intermolecular β-sheet aggregation in PrP^C [84]. Furthermore, residue R220 interacts with S132, thus sustaining the β-bulge in the edge strand S1.

We remark here that the enhanced fluctuation of the S1H1 loop in Y149F causes the displacement of S132, which destabilizes the β-bulge. Interestingly, the disruption of the β-bulge exposes the unsaturated amides and carbonyls on the edge strand S1, which results in the formation of the V161:N–G131:O bond. A fifth hydrogen bond, A133:N–N159:O is formed in Y149F. These additional hydrogen bonds in Y149F have the effect of elongating and stabilizing the native β-sheet. This observation is consistent with previous simulation studies on partially unfolded states of PrP^C, effects of pathogenic mutations, and influence of low pH environment[17,46]. The effect of perturbing inter-domain associations is thus manifested in notable secondary structural changes of the globular domain. However, it is appropriate to keep in mind that secondary structure propensities could be sensitive to force field effects. Therefore, the correlation between the conformational dynamics and secondary structural propensities yielded by MD trajectories generated with other atomistic force fields should also be compared.

5.5   Conclusions

In this work, we have examined the effects of the disruption of a single hydrogen bond between the H1 and H3 helical domains on the overall behavior of the folded, C-terminal globular domain of PrP^C. We observed that this relatively minor perturbation in the inter-domain association cascaded into several key conformational and dynamical effects. The effect is manifested in the dynamics of the various secondary structural domains, which in turn resulted in increased solvent exposure of residues of the hydrophobic core and overall reduced stability of the inter-domain salt bridges. The changes were further accompanied by an overall helical destabilization and subtle elongation of the β-strands, which could potentially mark the onset of

conformational transitions to the PrP^{Sc} forms. We note that the observed changes are consistent with the conformational changes observed in misfolding studies, at low pH or in studies with pathogenic mutations[15-17,27,48].

As mentioned previously, earlier studies have highlighted the role of several inter-domain non-bonded interactions in the early structural determinants of PrP^C to PrP^{Sc} transition[17,42-45,64,77]. Our study strongly underscores that the cooperative interplay of various non-bonded associations, particularly inter-helical hydrogen bonding between H1 and H3, hydrophobic contacts and salt bridge network, are necessary to maintain the structural integrity of the folded domain of PrP^C. Small, transient perturbations in the network of these interactions may critically disrupt this cooperativity and thereby induce dynamical and conformational instabilities potentially leading to structures that are prone to pathogenic transitions. In the light of these findings, it could be useful to design anti-prion strategies that bias the folded state by reducing dynamical fluctuations and stabilizing key inter-domain associations. Further investigations on longer timescales, coupled with precise experimental measurements, are required to characterize the transient nature of the observed fluctuations, and the extent to which they are capable of modulating the heterogeneous pathways inherent to the PrP^C to PrP^{Sc} transition.

5.6 References

1. Liemann, S.; Glockshuber, R. Transmissible spongiform encephalopathies. *Biochem. Biophys. Res. Commun.* **1998,** *250* (2), 187–193.

2. Aguzzi, A.; Calella, A. M. Prions: protein aggregation and infectious diseases. *Physiol. Rev.* **2009,** *89* (4), 1105–1152.

3. Prusiner, S. B. Scrapie Prions. *Annu. Rev. Microbiol.* **1989,** *43* (1), 345–374.

4. Prusiner, S. B. Molecular biology of prion diseases. *Science* **1991,** *252* (5012), 1515–1522.

5. Weissmann, C. Molecular genetics of transmissible spongiform encephalopathies. *J. Biol. Chem.* **1999,** *274* (1), 3–6.

6. Prusiner, S. B. Prions. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (23), 13363–13383.

7. Stahl, N.; Borchelt, D. R.; Hsiao, K.; Prusiner, S. B. Scrapie prion protein contains a phosphatidylinositol glycolipid. *Cell* **1987,** *51* (2), 229–240.

8. Zahn, R.; Liu, A.; Lührs, T.; Riek, R.; von Schroetter, C.; López García, F.; Billeter, M.; Calzolai, L.; Wider, G.; Wüthrich, K. NMR solution structure of the human prion protein. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (1), 145–150.

9. Diaz-Espinoza, R.; Soto, C. High-resolution structure of infectious prion protein: the final frontier. *Nat. Struct. Mol. Biol.* **2012,** *19* (4), 370–377.

10. Pan, K. M.; Baldwin, M.; Nguyen, J.; Gasset, M.; Serban, A.; Groth, D.; Mehlhorn, I.; Huang, Z.; Fletterick, R. J.; Cohen, F. E. Conversion of alpha-helices into beta-sheets features in the formation of the scrapie prion proteins. *Proc. Natl. Acad. Sci. U. S. A.* **1993,** *90* (23), 10962–10966.

11. Wille, H.; Michelitsch, M. D.; Guénebaut, V.; Supattapone, S.; Serban, A.; Cohen, F. E.; Agard, D. A.; Prusiner, S. B. Structural studies of the scrapie prion protein by electron crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (6), 3563–3568.

12. Govaerts, C.; Wille, H.; Prusiner, S. B.; Cohen, F. E. Evidence for assembly of prions with left-handed β-helices into trimers. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (22), 8342–8347.

13. DeMarco, M. L.; Daggett, V. From conversion to aggregation: Protofibril formation of the prion protein. *Proc. Natl. Acad. Sci. U. S. A.* **2004,** *101* (8), 2293–2298.

14. Cobb, N. J.; Sönnichsen, F. D.; McHaourab, H.; Surewicz, W. K. Molecular architecture of human prion protein amyloid: A parallel, in-register β-structure. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (48), 18946–18951.

15. DeMarco, M. L.; Daggett, V. Molecular Mechanism for Low pH Triggered Misfolding of the Human Prion Protein. *Biochemistry* **2007,** *46* (11), 3045–3054.

16. van der Kamp, M. W.; Daggett, V. Influence of pH on the Human Prion Protein: Insights into the Early Steps of Misfolding. *Biophys. J.* **2010,** *99* (7), 2289–2298.

17. De Simone, A.; Zagari, A.; Derreumaux, P. Structural and Hydration Properties of the Partially Unfolded States of the Prion Protein. *Biophys. J.* **2007,** *93* (4), 1284–1292.

18. Dima, R. I.; Thirumalai, D. Exploring the Propensies of Helices in PrP[C] to Form β Sheet Using NMR Structures and Sequence Alignments. *Biophys. J.* **2002,** *83* (3), 1268–1280.

19. Adrover, M.; Pauwels, K.; Prigent, S.; de Chiara, C.; Xu, Z.; Chapuis, C.; Pastore, A.; Rezaei, H. Prion fibrillization is mediated by a native structural element that comprises helices H2 and H3. *J. Biol. Chem.* **2010,** *285* (27), 21004–21012.

20. Chakroun, N.; Prigent, S.; Dreiss, C. A.; Noinville, S.; Chapuis, C.; Fraternali, F.; Rezaei, H. The oligomerization properties of prion protein are restricted to the H2H3 domain. *FASEB J.* **2010,** *24* (9), 3222–3231.

21. Tycko, R.; Savtchenko, R.; Ostapchenko, V. G.; Makarava, N.; Baskakov, I. V. The α-Helical C-Terminal Domain of Full-Length Recombinant PrP Converts to an In-Register Parallel β-Sheet Structure in PrP Fibrils: Evidence from Solid State Nuclear Magnetic Resonance. *Biochemistry* **2010,** *49* (44), 9488–9497.

22. Chen, J.; Thirumalai, D. Helices 2 and 3 Are the Initiation Sites in the PrP[C] -> PrP[SC] Transition. *Biochemistry* **2013,** *52* (2), 310–319.

23. Chakroun, N.; Fornili, A.; Prigent, S. h.; Kleinjung, J.; Dreiss, C. i. A.; Rezaei, H.; Fraternali, F. Decrypting Prion Protein Conversion into a β-Rich Conformer by Molecular Dynamics. *J. Chem. Theory Comput.* **2013,** *9* (5), 2455–2465.

24. Riek, R.; Wider, G.; Billeter, M.; Hornemann, S.; Glockshuber, R.; Wüthrich, K. Prion protein NMR structure and familial human spongiform encephalopathies. *Proc. Natl. Acad. Sci. U. S. A.* **1998,** *95* (20), 11667–11672.

25. Zuegg, J.; Gready, J. E. Molecular Dynamics Simulations of Human Prion Protein: Importance of Correct Treatment of Electrostatic Interactions. *Biochemistry* **1999,** *38* (42), 13862–13876.

26. Guest, W. C.; Cashman, N. R.; Plotkin, S. S. Electrostatics in the stability and misfolding of the prion protein: salt bridges, self energy, and solvation. *Biochem. Cell Biol.* **2010,** *88* (2), 371–381.

27. van der Kamp, M. W.; Daggett, V. Pathogenic Mutations in the Hydrophobic Core of the Human Prion Protein Can Promote Structural Instability and Misfolding. *J. Mol. Biol.* **2010,** *404* (4), 732–748.

28. Liemann, S.; Glockshuber, R. Influence of Amino Acid Substitutions Related to Inherited Human Prion Diseases on the Thermodynamic Stability of the Cellular Prion Protein. *Biochemistry* **1999,** *38* (11), 3258–3267.

29. Calzolai, L.; Zahn, R. Influence of pH on NMR Structure and Stability of the Human Prion Protein Globular Domain. *J. Biol. Chem.* **2003,** *278* (37), 35592–35596.

30. El-Bastawissy, E.; Knaggs, M. H.; Gilbert, I. H. Molecular dynamics simulations of wild-type and point mutation human prion protein at normal and elevated temperature. *J. Mol. Graphics Modell.* **2001,** *20* (2), 145–154.

31. Santini, S.; Claude, J.-B.; Audic, S.; Derreumaux, P. Impact of the tail and mutations G131V and M129V on prion protein flexibility. *Proteins* **2003,** *51* (2), 258–265.

32. Morrissey, M. P.; Shakhnovich, E. I. Evidence for the role of PrP^C helix 1 in the hydrophilic seeding of prion aggregates. *Proc. Natl. Acad. Sci. U. S. A.* **1999,** *96* (20), 11293–11298.

33. Speare, J. O.; Rush, T. S.; Bloom, M. E.; Caughey, B. The Role of Helix 1 Aspartates and Salt Bridges in the Stability and Conversion of Prion Protein. *J. Biol. Chem.* **2003,** *278* (14), 12522–12529.

34. Ziegler, J.; Sticht, H.; Marx, U. C.; Müller, W.; Rösch, P.; Schwarzinger, S. CD and NMR Studies of Prion Protein (PrP) Helix 1: novel implications for its role in the PrP^C → PrP^{Sc} conversion process. *J. Biol. Chem.* **2003,** *278* (50), 50175–50181.

35. Sharman, G. J.; Kenward, N.; Williams, H. E.; Landon, M.; Mayer, R. J.; Searle, M. S. Prion protein fragments spanning helix 1 and both strands of β sheet (residues 125-170) show evidence for predominantly helical propensity by CD and NMR. *Folding Des.* **1998,** *3* (5), 313–320.

36. Liu, A.; Riek, R.; Zahn, R.; Hornemann, S.; Glockshuber, R.; Wüthrich, K. Peptides and proteins in neurodegenerative disease: Helix propensity of a polypeptide containing helix 1 of the mouse prion protein studied by NMR and CD spectroscopy. *Pept. Sci.* **1999,** *51* (2), 145–152.

37. Norstrom, E. M.; Mastrianni, J. A. The Charge Structure of Helix 1 in the Prion Protein Regulates Conversion to Pathogenic PrP^{Sc}. *J. Virol.* **2006,** *80* (17), 8521–8529.

38. Santini, S.; Derreumaux, P. Helix H1 of the prion protein is rather stable against environmental perturbations: molecular dynamics of mutation and deletion variants of PrP(90-231). *Cell. Mol. Life Sci.* **2004,** *61* (7-8), 951–960.

39. Watzlawik, J.; Skora, L.; Frense, D.; Griesinger, C.; Zweckstetter, M.; Schulz-Schaeffer, W. J.; Kramer, M. L. Prion Protein Helix1 Promotes Aggregation but Is Not Converted into beta-Sheet. *J. Biol. Chem.* **2006,** *281* (40), 30242–30250.

40. Ji, H.-F.; Zhang, H.-Y.; Shen, L. The Role of Electrostatic Interaction in Triggering the Unraveling of Stable Helix 1 in Normal Prion Protein. A Molecular Dynamics Simulation Investigation. *J. Biomol. Struct. Dyn.* **2005,** *22* (5), 563–570.

41. Camilloni, C.; Schaal, D.; Schweimer, K.; Schwarzinger, S.; De Simone, A. Energy Landscape of the Prion Protein Helix 1 Probed by Metadynamics and NMR. *Biophys. J.* **2012,** *102* (1), 158–167.

42. Schwarzinger, S.; Horn, A. H. C.; Ziegler, J.; Sticht, H. Rare Large Scale Subdomain Motions in Prion Protein can Initiate Aggregation. *J. Biomol. Struct. Dyn.* **2006,** *23* (6), 581–590.

43. Eghiaian, F.; Daubenfeld, T.; Quenet, Y.; van Audenhaege, M.; Bouin, A.-P.; van der Rest, G.; Grosclaude, J.; Rezaei, H. Diversity in prion protein oligomerization pathways results from domain expansion as revealed by hydrogen/deuterium exchange and disulfide linkage. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (18), 7414–7419.

44. Viles, J. H.; Donne, D.; Kroon, G.; Prusiner, S. B.; Cohen, F. E.; Dyson, H. J.; Wright, P. E. Local Structural Plasticity of the Prion Protein. Analysis of NMR Relaxation Dynamics. *Biochemistry* **2001,** *40* (9), 2743–2753.

45. Kachel, N.; Kremer, W.; Zahn, R.; Kalbitzer, H. R. Observation of intermediate states of the human prion protein by high pressure NMR spectroscopy. *BMC Struct. Biol.* **2006,** *6* (1), 16.

46. Chen, W.; van der Kamp, M. W.; Daggett, V. Diverse effects on the native β-sheet of the human prion protein due to disease-associated mutations. *Biochemistry* **2010,** *49* (45), 9874–9881.

47. Cheng, C.; Daggett, V. Molecular Dynamics Simulations Capture the Misfolding of the Bovine Prion Protein at Acidic pH. *Biomolecules* **2014,** *4* (1), 181–201.

48. Guo, J.; Ning, L.; Ren, H.; Liu, H.; Yao, X. Influence of the pathogenic mutations T188K/R/A on the structural stability and misfolding of human prion protein: Insight from molecular dynamics simulations. *Biochim. Biophys.Acta, Gen. Subj.* **2012,** *1820* (2), 116–123.

49.   Riek, R.; Hornemann, S.; Wider, G.; Billeter, M.; Glockshuber, R.; Wuthrich, K. NMR structure of the mouse prion protein domain PrP(121-231). *Nature* **1996,** *382* (6587), 180–182.

50.   James, T. L.; Liu, H.; Ulyanov, N. B.; Farr-Jones, S.; Zhang, H.; Donne, D. G.; Kaneko, K.; Groth, D.; Mehlhorn, I.; Prusiner, S. B.; Cohen, F. E. Solution structure of a 142-residue recombinant prion protein corresponding to the infectious fragment of the scrapie isoform. *Proc. Natl. Acad. Sci. U. S. A.* **1997,** *94* (19), 10086–10091.

51.   López García, F.; Zahn, R.; Riek, R.; Wüthrich, K. NMR structure of the bovine prion protein. *Proc. Natl. Acad. Sci. U. S. A.* **2000,** *97* (15), 8334–8339.

52.   Autiero, I.; Saviano, M.; Langella, E. In silico investigation and targeting of amyloid β oligomers of different size. *Mol. BioSyst* **2013,** *9* (8), 2118–2124.

53.   Chong, S.-H.; Yim, J.; Ham, S. Structural heterogeneity in familial Alzheimer's disease mutants of amyloid-beta peptides. *Mol. BioSyst* **2013,** *9* (5), 997–1003.

54.   Autiero, I.; Langella, E.; Saviano, M. Insights into the mechanism of interaction between trehalose-conjugated beta-sheet breaker peptides and Aβ (1–42) fibrils by molecular dynamics simulations. *Mol. BioSyst.* **2013,** *9* (11), 2835–2841.

55.   Buchete, N.-V.; Tycko, R.; Hummer, G. Molecular Dynamics Simulations of Alzheimer's β-Amyloid Protofilaments. *J. Mol. Biol.* **2005,** *353* (4), 804–821.

56.   Jana, A. K.; Jose, J. C.; Sengupta, N. Critical roles of key domains in complete adsorption of Aβ peptide on single-walled carbon nanotubes: insights with point mutations and MD simulations. *Phys. Chem. Chem. Phys.* **2013,** *15* (3), 837–844.

57.   Jana, A. K.; Sengupta, N. Adsorption Mechanism and Collapse Propensities of the Full-Length, Monomeric A**β**1-42 on the Surface of a Single-Walled Carbon Nanotube: A Molecular Dynamics Simulation Study. *Biophys. J.* **2012,** *102* (8), 1889–1896.

58.   Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *The J. Chem. Phys.* **1983,** *79* (2), 926–935.

59.   Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. NAMD2: Greater Scalability for Parallel Molecular Dynamics. *Journal of Computational Physics* **1999,** *151* (1), 283–312.

60.   Mackerell, A. D. Empirical force fields for biological macromolecules: Overview and issues. *J. Comput. Phys.* **2004,** *25* (13), 1584–1604.

61.   Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. *J. Chem. Phys.,* **1995,** *103* (11), 4613–4621.

62.   Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977,** *23* (3), 327–341.

63.   Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995,** *103*, 8577.

64.   Alonso, D. O. V.; An, C.; Daggett, V. Simulations of biomolecules: characterization of the early steps in the pH-induced conformational conversion of the hamster, bovine and human forms of the prion protein. *Philos. Trans. A Math. Phys. Eng. Sci.* **2002,** *360* (1795), 1165–1178.

65.   Glykos, N. M. Software news and updates. Carma: a molecular dynamics analysis program. *J. Comput. Chem.* **2006,** *27* (14), 1765–1768.

66.   Lin, Y.-S.; Bowman, G.; Beauchamp, K.; Pande, V. Investigating How Peptide Length and a Pathogenic Mutation Modify the Structural Ensemble of Amyloid Beta Monomer. *Biophys. J.* **2012,** *102* (2), 315–324.

67.   Haider, S.; Parkinson, G. N.; Neidle, S. Molecular Dynamics and Principal Components Analysis of Human Telomeric Quadruplex Multimers. *Biophys. J.* **2008,** *95* (1), 296–311.

68.   Fadouloglou, V. E.; Stavrakoudis, A.; Bouriotis, V.; Kokkinidis, M.; Glykos, N. M. Molecular Dynamics Simulations of BcZBP, A Deacetylase from Bacillus cereus: Active Site Loops Determine Substrate Accessibility and Specificity. *J. Chem. Theory Comput.* **2009,** *5* (12), 3299–3311.

69.   Michel Espinoza-Fonseca, L.; Ilizaliturri-Flores, I.; Correa-Basurto, J. Backbone conformational preferences of an intrinsically disordered protein in solution. *Mol. BioSyst.* **2012,** *8* (6), 1798–1805.

70.   Bhakat, S.; Martin, A. J. M.; Soliman, M. E. S. An integrated molecular dynamics, principal component analysis and residue interaction network approach reveals the impact of M184V mutation on HIV reverse transcriptase resistance to lamivudine. *Mol. BioSyst.* **2014,** *10* (8), 2215–2228.

71.   Delano, W. L. The PyMOL Molecular Graphics System. 2002.

72.   Grant, B. J.; Rodrigues, A. P. C.; ElSawy, K. M.; McCammon, J. A.; Caves, L. S. D. Bio3d: an R package for the comparative analysis of protein structures. *Bioinformatics* **2006,** *22* (21), 2695–2696.

73.   Schlitter, J. Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chem. Phys. Lett.* **1993,** *215* (6), 617–621.

74.   Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996,** *14* (1), 33–38.

75.   De Simone, A.; Dodson, G. G.; Verma, C. S.; Zagari, A.; Fraternali, F. Prion and water: Tight and dynamical hydration sites have a key role in structural stability. *Proc. Natl. Acad. Sci. U. S. A.* **2005,** *102* (21), 7535–7540.

76.   van der Kamp, M. W.; Daggett, V. The consequences of pathogenic mutations to the human prion protein. *Protein Eng., Des. Sel.* **2009,** *22* (8), 461–468.

77.   Bamdad, K.; Naderi-Manesh, H. Contribution of a putative salt bridge and backbone dynamics in the structural instability of human prion protein upon R208H mutation. *Biochem. Biophys. Res. Commun.* **2007,** *364* (4), 719–724.

78.   Salari, R.; Chong, L. T. Desolvation costs of salt bridges across protein binding interfaces: Similarities and differences between implicit and explicit solvent models. *J. Phys. Chem. Lett.,* **2010,** *1* (19), 2844–2848.

79.   Meuzelaar, H.; Tros, M.; Huerta-Viga, A.; van Dijk, C. N.; Vreede, J.; Woutersen, S. Solvent-Exposed Salt Bridges Influence the Kinetics of α-Helix Folding and Unfolding. *J. Phys. Chem. Lett.,* **2014,** *5* (5), 900–904.

80.   Salari, R.; Chong, L. T. Effects of High Temperature on Desolvation Costs of Salt Bridges Across Protein Binding Interfaces: Similarities and Differences between Implicit and Explicit Solvent Models. *J. Phys. Chem. B* **2012,** *116* (8), 2561–2567.

81.   Jose, J. C.; Chatterjee, P.; Sengupta, N. Cross Dimerization of Amyloid-β and αSynuclein Proteins in Aqueous Environment: A Molecular Dynamics Simulations Study. *PLoS One* **2014,** *9* (9), No. e106883.

82.   Heinig, M.; Frishman, D. STRIDE: a web server for secondary structure assignment from known atomic coordinates of proteins. *Nucleic Acids Res.* **2004,** *32* (2), W500-W502.

83.   Richardson, J. S.; Richardson, D. C. Natural β-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl. Acad. Sci. U. S. A.* **2002,** *99* (5), 2754–2759.

84.   De Simone, A.; Dodson, G. G.; Fraternali, F.; Zagari, A. Water molecules as structural determinants among prions of low sequence identity. *FEBS Lett.* **2006,** *580* (10), 2488–2494.

Summary and Future Outlook

6.1   Abstract

The body of work assembled in this thesis sought to investigate the effects of structural perturbation and environmental factors on the conformational dynamics and self-assembly of proteins classified as intrinsically disordered proteins (IDPs). Towards this purpose, MD simulations technique was employed as it can provide a molecular-level understanding of the physical underpinnings of IDP structure and dynamics. The principal goals of these studies were, (1) to evaluate the thermal response of an amyloid oligomer, (2) to delineate the physical effects of glucose crowding on the early self-assembly of the Aβ peptides, and (3) to probe the structural and dynamical implications of a minor perturbation in one of the native intermolecular interactions of the cellular prion protein. In this chapter, the key results of the studies performed with the above stated objectives are summarized and the future perspectives emerging from this work are briefly discussed.

6.2   Summary

This thesis presents an exploration of the conformational and dynamical response of amyloidogenic proteins to extrinsic (temperature, molecular crowding) and intrinsic (structural) factors. The results discussed here are a step towards better understanding

of the molecular determinants of the stability of this class of proteins implicated in diseases. The conclusions and future perspectives pertaining to these studies are discussed in this section.

The cold thermal response of an amyloid oligomer was found to be divergent from that typically observed in globular proteins[1]. The model amyloid oligomer studied has characteristic features of an amyloid assembly and a hydrophobic core similar to that of globular proteins. Exhaustive conformational sampling in a broad temperature range revealed a sharp inflection indicative of structural transition and growth in overall structural stabilization upon cooling, in contrast to that exhibited by folded proteins. However, the equilibrium free energy profile of hydrophobic core packing reflects the temperature-dependent hydrophobic effect upon globular protein unfolding. A significant role of solvation waters in mediating the structural response of the oligomer emerges from the energetic coupling between the oligomer stabilization and the hydration layer as well as the extent of fluctuations therein. Furthermore, the structural ordering of the solvation shell exhibits maximal difference over bulk at the transition and become negligible at high temperatures. The differential cold thermal response and the critical role of the hydration waters on the stability of amyloid assemblies and modulating its thermal behavior emerge from this exposition.

Evidences from several epidemiological studies linking type-2 diabetes mellitus (T2DM) with AD pathogenesis[2-4] have encouraged elucidation of the molecular basis of hyperglycemic conditions in T2DM in the pathophysiology of AD. An investigation in this direction was performed to discern the physical effects of glucose crowding in hyperglycemic conditions on the early self-assembly or dimerization of Aβ peptides[5]. The results imply that crowding in excess glucose

conditions do not promote Aβ self-assembly. The weaker dimers formed in glucose solutions relative to pure water can be attributed to greater hydration of the peptides owing to clustering of glucose molecules away from the peptide surface screening the hydrophobic interactions that drive the self-assembly process. Recent evidences indicate that post-translationally modified glycated Aβ peptides are more toxic than the unglycated peptide[6-7] and have enhanced propensity to self-assemble[8]. Hence, it would be beneficial to understand the self-assembly of these modified peptides in crowded hyperglycemic solvent conditions.

Native interactions are key determinants of protein stability and play important roles in mediating protein-folding mechanisms[9-10]. Disruption of these interactions under non-ideal conditions compromise protein stability leading to misfolding and aggregation associated with diseases[9,11]. In this thesis, the significance of native non-bonded interactions in maintaining the structural integrity of the cellular prion protein (PrP^C) is underscored[12] by studying the conformational and dynamical effects of a relatively minor perturbation of a native non-bonded interaction. Elimination of a single hydrogen bond led to a cascade of structural and dynamical transformations consistent with those observed in destabilizing conditions of low pH and in pathogenic mutants of PrP^C. Enhanced flexibility of the perturbed protein exposed the sequestered hydrophobic core residues with a concomitant loss of the inter-domain salt-bridge network. Since the small perturbation introduced is not a familial disease-causing mutation, this study indicates that stochastic, transient perturbations in the native network of interactions have the potential to cause conformational and dynamical changes inducing pathogenic transformations in PrP^C. This strategy can be effectively employed in verifying the contributions of the other native interactions to the overall stability of PrP^C. This information will be useful in designing therapeutic

approaches directed at preventing or reducing PrP^C transition to the pathogenic scrapie form.

6.3 Future Outlook

The work presented in this thesis is a step towards understanding aspects of conformational stability, dynamics and self-assembly of amyloidogenic proteins. Understanding the stability of both globular proteins and IDPs as a function of key thermodynamic variables, sequence and other influencing external factors is of fundamental importance and can be explored using interdisciplinary approaches combining insights from experimental biophysical techniques as well as theoretical and computational studies. Characterization of the protein conformational energy landscapes by using various perturbation variables such as temperature, pressure, pH, chemical denaturants etc. can provide comprehensive insights into the relative stabilities of the various conformational states populating the energy landscape of proteins and how they are modulated by amino-acid sequence and environmental conditions. The overall knowledge base thus developed may be leveraged for deeper understanding of the molecular basis of protein stability. Before concluding, some directions of future research that lead from the work presented in this thesis are discussed below.

i) Order parameters for protein thermodynamic response

Development of order parameters encompassing the multifactorial determinants of oligomer stability would be useful in describing the overall free energy profile of the amyloid self-assembly. In addition to the core packing parameter studied in this thesis, these may incorporate other stabilizing factors such as number of native

residue-residue and water-mediated inter-residue hydrogen bonds, fraction of native contacts and other geometric parameters that best characterize the thermally induced conformational transitions.

ii) The work presented here indicates the temperature-dependent microscopic influence of hydration waters on the stability of the self-assembled amyloid motif is greater than its primary role in sequestering the hydrophobic residues as seen in globular proteins. Further detailed investigations are needed to delineate the temperature-dependent role of water in mediating inter-residue stabilizing hydrogen bonding interactions, the lifetimes of these hydrogen bonds and ensuing structural transitions.

iii) Evidences indicate that amyloidogenic proteins share common structural motifs regardless of the sequence[13-17]. However, further detailed studies carried out over the various polymorphic forms if naturally occurring, artificial and functional amyloids, and also heterogeneous assemblies with other biomolecules, would be helpful in consolidating these findings.

iv) Hydration dynamics

Studies of the role of solvent in conformational and self-assembly behavior of proteins have showed that the coupling of water molecules with the protein surface causes distinct deviations from bulk behavior[18-20]. These aspects are extensively studied for globular proteins but largely unexplored for IDPs and amyloidogenic proteins[21-26]. The dynamical retardation induced in the hydration waters in the vicinity of the protein is characterized by sublinear diffusion, slower tumbling time and higher

residence times in the solvation shell[22-26]. In view of the role of solvent in modulating the dimerization of Aβ peptides in glucose-crowded environment studied in this thesis, further systematic investigation of the translational and rotational dynamics of the water molecules entrapped near the surface by glucose crowding studied in this thesis would provide additional mechanistic insights. Particularly, local hydration dynamics around the interpeptide contacts of the dimers formed in pure water and glucose solution can be compared and correlated depending upon the nature of the contacts.

6.4   References

1.   Menon, S.; Sengupta, N. The Cold Thermal Response of an Amyloid Oligomer Differs from Typical Globular Protein Cold Denaturation. *J. Phys. Chem. Lett.* **2019,** *10* (10), 2453–2457.

2.   Haan, M. N. Therapy Insight: type 2 diabetes mellitus and the risk of late-onset Alzheimer's disease. *Nat. Clin. Pract. Neurol.* **2006,** *2* (3), 159–166.

3.   Leibson, C. L.; Rocca, W. A.; Hanson, V. A.; Cha, R.; Kokmen, E.; O'Brien, P. C.; Palumbo, P. J. Risk of Dementia among Persons with Diabetes Mellitus: A Population-based Cohort Study. *Am. J. Epidemiol.* **1997,** *145* (4), 301–308.

4.   Takeda, S.; Sato, N.; Rakugi, H.; Morishita, R. Molecular mechanisms linking diabetes mellitus and Alzheimer disease: beta-amyloid peptide, insulin signaling, and neuronal function. *Mol. Biosyst.* **2011,** *7* (6), 1822–1827.

5.   Menon, S.; Sengupta, N. Influence of Hyperglycemic Conditions on Self-Association of the Alzheimer's Amyloid β (Aβ$_{1-42}$) Peptide. *ACS Omega* **2017,** *2* (5), 2134–2147.

6.   Li, X. H.; Du, L. L.; Cheng, X. S.; Jiang, X.; Zhang, Y.; Lv, B. L.; Liu, R.; Wang, J. Z.; Zhou, X. W. Glycation exacerbates the neuronal toxicity of β-amyloid. *Cell Death Dis.* **2013,** *4*, No. e673.

7.   Chen, C.; Li, X. H.; Tu, Y.; Sun, H. T.; Liang, H. Q.; Cheng, S. X.; Zhang, S. Aβ-AGE aggravates cognitive deficit in rats via RAGE pathway. *Neuroscience* **2014,** *257*, 1–10.

8.   Jana, A. K.; Batkulwar, K. B.; Kulkarni, M. J.; Sengupta, N. Glycation induces conformational changes in the amyloid-β peptide and enhances its aggregation

propensity: molecular insights. *Phys. Chem. Chem. Phys.* **2016,** *18* (46), 31446–31458.

9.   Jahn, T. R.; Radford, S. E. The Yin and Yang of protein folding. *FEBS J.* **2005,** *272* (23), 5962–5970.

10.   Best, R. B.; Hummer, G.; Eaton, W. A. Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110* (44), 17874–17879.

11.   Dobson, C. M. Protein folding and misfolding. *Nature* **2003,** *426* (6968), 884–890.

12.   Menon, S.; Sengupta, N. Perturbations in inter-domain associations may trigger the onset of pathogenic transformations in PrP^C: insights from atomistic simulations. *Mol. Biosyst.* **2015,** *11* (5), 1443–1453.

13.   Kayed, R.; Head, E.; Thompson, J. L.; McIntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabe, C. G. Common Structure of Soluble Amyloid Oligomers Implies Common Mechanism of Pathogenesis. *Science* **2003,** *300* (5618), 486–489.

14.   Bucciantini, M.; Calloni, G.; Chiti, F.; Formigli, L.; Nosi, D.; Dobson, C. M.; Stefani, M. Prefibrillar Amyloid Protein Aggregates Share Common Features of Cytotoxicity. *J. Biol. Chem.* **2004,** *279* (30), 31374–31382.

15.   Jahn, T. R.; Makin, O. S.; Morris, K. L.; Marshall, K. E.; Tian, P.; Sikorski, P.; Serpell, L. C. The Common Architecture of Cross-β Amyloid. *J. Mol. Biol.* **2010,** *395* (4), 717–727.

16.   Laganowsky, A.; Liu, C.; Sawaya, M. R.; Whitelegge, J. P.; Park, J.; Zhao, M.; Pensalfini, A.; Soriaga, A. B.; Landau, M.; Teng, P. K.; Cascio, D.; Glabe, C.; Eisenberg, D. Atomic View of a Toxic Amyloid Small Oligomer. *Science* **2012,** *335* (6073), 1228–1231.

17.   Liu, C.; Zhao, M.; Jiang, L.; Cheng, P.-N.; Park, J.; Sawaya, M. R.; Pensalfini, A.; Gou, D.; Berk, A. J.; Glabe, C. G.; Nowick, J.; Eisenberg, D. Out-of-register β-sheets suggest a pathway to toxic amyloid aggregates. *Proc. Natl. Acad. Sci. U.S.A.* **2012,** *109* (51), 20913–20918.

18.   Pal, S. K.; Peon, J.; Bagchi, B.; Zewail, A. H. Biological Water: Femtosecond Dynamics of Macromolecular Hydration. *J. Phys. Chem. B* **2002,** *106* (48), 12376–12395.

19.   Ebbinghaus, S.; Kim, S. J.; Heyden, M.; Yu, X.; Heugen, U.; Gruebele, M.; Leitner, D. M.; Havenith, M. An extended dynamical hydration shell around proteins. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (52), 20749–20752.

20.   Laage, D.; Elsaesser, T.; Hynes, J. T. Water Dynamics in the Hydration Shells of Biomolecules. *Chem. Rev.* **2017,** *117* (16), 10694–10725.

21. Li, T.; Hassanali, A. A.; Kao, Y.-T.; Zhong, D.; Singer, S. J. Hydration Dynamics and Time Scales of Coupled Water−Protein Fluctuations. *J. Am. Chem. Soc.* **2007,** *129* (11), 3376–3382.

22. Bandyopadhyay, S.; Chakraborty, S.; Bagchi, B. Exploration of the Secondary Structure Specific Differential Solvation Dynamics between the Native and Molten Globule States of the Protein HP-36. *J. Phys. Chem. B* **2006,** *110* (41), 20629–20634.

23. Sengupta, N.; Jaud, S.; Tobias, D. J. Hydration Dynamics in a Partially Denatured Ensemble of the Globular Protein Human α-Lactalbumin Investigated with Molecular Dynamics Simulations. *Biophys. J.* **2008,** *95* (11), 5257–5267.

24. Tarek, M.; Tobias, D. J. The Dynamics of Protein Hydration Water: A Quantitative Comparison of Molecular Dynamics Simulations and Neutron-scattering Experiments. *Biophys. J.* **2000,** *79* (6), 3244–3257.

25. Rani, P.; Biswas, P. Diffusion of Hydration Water around Intrinsically Disordered Proteins. *J. Phys. Chem. B* **2015,** *119* (42), 13262–13270.

26. Jose, J. C.; Khatua, P.; Bansal, N.; Sengupta, N.; Bandyopadhyay, S. Microscopic Hydration Properties of the $A\beta_{1-42}$ Peptide Monomer and the Globular Protein Ubiquitin: A Comparative Molecular Dynamics Study. *J. Phys. Chem. B* **2014,** *118* (40), 11591–11604.