The Dual Role of Choline-*O*-sulfate on Chemical Denaturation of Proteins and Amyloid Aggregation: A Computational Study

A Thesis Submitted

in Partial Fulfillment of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Srijita Paul

to the

Department of Chemistry

Indian Institute of Technology Guwahati, India

2020

DEDICATED TO
MY PARENTS

Declaration

I hereby declare that the matter manifested in this thesis entitled "The Dual Role of Choline-*O*-sulfate on Chemical Denaturation of Proteins and Amyloid Aggregation: A Computational Study" is the result of research carried out by me in the Department of Chemistry, Indian Institute of Technology Guwahati, India under the supervision of Prof. Sandip Paul.

In keeping with the general practice of reporting scientific observations, due acknowledgement has been made wherever the work described is based on the findings of other investigators.

Srijita Paul
IIT Guwahati

Certificate

It is certified that the work contained in this thesis entitled, "The Dual Role of Choline-*O*-sulfate on Chemical Denaturation of Proteins and Amyloid Aggregation: A Computational Study" has been carried out by Miss Srijita Paul for the Degree of Doctor of Philosophy under my supervision and the same has not been submitted elsewhere for a degree.

Prof. Sandip Paul
Thesis Supervisor
Department of Chemistry
Indian Institute of Technology Guwahati
Guwahati-781039, India

Preface

On accomplishment of my doctorate studies, it is a genuine pleasure to express here a few words of appreciation to the people who made this journey a reality and an unforgettable experience for me. First and foremost, I would like to express my sincere gratitude to my supervisor, Prof. Sandip Paul, for his excellent guidance, constant encouragement, enormous support, and help during the years in his association. His door was always open for the countless discussion sessions with him which taught me how science and its challenges could be made interesting to ease out a solution, even during the tough times in the Ph.D. pursuit. I am also thankful to him for giving me the freedom to pursue my ideas, and I could not have imagined having a better advisor and mentor for my Ph.D. career.

Besides my supervisor, I am highly obliged to the doctoral committee members, Prof. Gopal Das, Dr. Lal Mohan Kundu, Dr. Kalyan Raidongia for periodically assessing my work and providing valuable suggestions for its improvement. I would like to thank Dr. Bhubaneswar Mandal for being supportive during the difficulties in my research work. My sincere thanks go to all other faculty members in the department for their kind help at various stages of my doctoral work. I gratefully acknowledge the Ministry of Human Resource and Development (MHRD), India for financial support and IIT Guwahati for providing the research facilities to carry out my research work. I would like to appreciate the IIT Guwahati super-computing facility PARAM-ISHAN, without which the completion of my dissertation was not possible within this time period.

No words will be sufficient to thank my lab mates Krishna, Saikat, Rabindranath, Rituparna, Aritra, Madhusmita, and Rimjhim for their unstinting support, helpful discussions, and creating a pleasant working environment in the lab. A heartfelt thank to my seniors Dr. Rahul Sarma, Dr. Subrata Paul, Dr. Bhanita Sharma, Dr. Gargi Borgohain, and Dr. Shubhadip Das for sharing their valuable suggestions and experiences during my Ph.D. days. Also, I would like to acknowledge the project students Vikram, Komal, Himangshu, and Rakesh whose contributions can't be ignored.

I take this opportunity to express my sincere gratitude and respect to all my teachers in school, college, and university days for preparing me academically and non-academically in various aspects of life. All the learning from them will be an asset in every walk of my life. I am grateful to my teachers Dr. Sujoy Roy Chowdhury and Late Dr. Dinabandhu Kundu for their authentic guidance in my tough times.

I extend my sincere thanks and bundles of love to my friends Adrija, Dibyandu, Devki, Puja, Subhadip, Raisa, Somasri, and Payel for listening to me, and being my true friends and mental support in different stages of my life. I am truly grateful to Titas Kumar Mukhopadhyay for being my inspiration and strength. His constant support, encouragement, academic assistance, and scientific discussions enrich me a lot to achieve this goal. I owe a lot to my IITG friends Ruma, Sheuly, Swapna, Priya, Anisha, Eshita, Mood Mohan, Suchetana, Anwesha, Anindya, Anjishnu, and Arghya for making my Ph.D. journey joyful and memorable.

Finally, this dream could not have been fulfilled without the endless love, support, and blessings from my family members. I extend my sincere gratitude to my mother for always being my best friend, philosopher, and guide. I am genuinely grateful to my parents for their great sacrifices and understandings, for helping me to be a better human being, and allowing me to fly high. And last but not least, I am blessed to have my brother as one of the strongest pillars of my life. A famous quote reads- "Every journey begins with a single step"- and today, only because of my family, I move a step uphill towards my zenith.

Srijita Paul

2020

"Creating a new theory is not like destroying an old barn and erecting a sky scraper in its place. It is rather like climbing a mountain, gaining new and wider views, discovering new connections between our starting point and its rich environment."

— Albert Einstein

Outline of the Thesis

Chapter 1

Introduction

"Proteins can do almost everything....Before we can hope to understand how genes work, how muscles contract, how nerves conduct electricity, how embryos develop, or how our bodies function, we must attain a deep understanding of proteins."

– B. Alberts et al., Molecular Biology of the Cell, Fifth Edition, 2008, Garland Science, New York, USA, p125

■ PROTEIN MISFOLDING AND ITS CONSEQUENCES

Proteins are the building blocks of all living organisms. They operate as crucial components in different cellular functions including replication of genetic material, catalysis of metabolic reactions, preservation of the cellular structure, cellular signaling, immune responses, cell adhesion, cell cycle, responding to stimuli, and transporting molecules from one location to another [1]. Humans produce an assortment of around 30,000 different proteins, each with a different role. Each protein consists a unique sequence of its own, leading to a particular native conformation corresponding to a specific physiological function. Proteins exhibit a diverse range of variance, spanning from small peptides to large multimers. They are formed by peptide bonds that link the constituent amino acids. The variation among the peptides depends on their amino acid sequence and side-chain groups of the amino acid residues. The 20 different naturally occurring amino acids in biological species are classified into three different categories depending on their side-chains, i.e. acidic, basic, neutral, or hydrophobic amino acids. The amino acid ordering determines the primary structure of the protein by creating a unique polypeptide chain, which is relatively flexible. These polypeptides can fold into three different secondary structures: α-helix, β-sheet, and random coil. The α-helices and the β-sheets are the ordered structures formed by the strong inter-chain interactions between the amino acid residues, and determine the three-dimensional structure of the protein. At the same time, the random coil is the less ordered structure permitting free rotation around each bond. The tertiary structure refers to the distribution of the α-helices, β-sheets and the random coils in the protein. These elements are folded into a compact conformation stabilized by hydrogen bonds or ionic interactions. The term quaternary structure is used to describe multimeric proteins, in which the different polypeptide chains are connected. To function properly, the protein must achieve its proper conformation and location within the packed environment inside the cell. By the protein folding process, a newly synthesized protein molecule folds into its characteristic 3-D structure [2]. Exposure to environmental changes may disrupt this three dimensional structure resulting in inappropriate shape or misfolding of the protein [3].

Protein unfolding can be defined as a process where the spatial arrangement of the polypeptide chains within the biomolecule changes from the native protein to a highly disordered one i.e., alteration of three-dimensional structure [4]. Some physico-chemical conditions that are known to affect the folding of biomolecules are temperature, pressure, pH , solvent, co-solvent or salt concentration, etc. Based on experimental

and theoretical evidences, it is well known that a high concentration of urea can cause protein denaturation in solution and, hence, inhibit many important biological activities [5, 6, 7, 8]. Urea is found as a waste product in mammalian kidneys and is a strong denaturant at high concentrations[9, 10]. Many experimental and simulations studies have been done for years to understand the urea-conferred denaturation of proteins and polypeptides [11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. Simulations of the protein CI2 in 8 M urea by Bennion and Daggett showed that protein rapidly unfolds in urea solution through expansion of the hydrophobic core [15]. The authors further proposed that urea interact directly with the peptide backbone, and the average residence times for urea around the protein hydrophobic and hydrophilic residues are higher than the corresponding water residence times [15]. Moreover, Berne and co-workers found from the μs MD simulations of hen egg-white lysozyme in 8 M urea solution that the dispersion interactions between urea and the protein fractions (the backbone and side-chains) are stronger than water [11]. It causes the intrusion of urea into the protein interior and facilitates its preferential binding to all regions of the protein. Further, many simulations, focused on small peptides to reduce the complexity in the interactions between solution species and protein residues, provides useful insight [14, 23, 24, 25]. Pettitt and co-workers simulated the deca-alanine peptide with different conformations in binary urea solution and showed that protein denaturation is governed by vdW interaction [26]. The addition of urea to pure water was shown to stabilize the extended states more than the compact helix through enhancement of more favorable protein-urea vdW interactions [26].

Besides traditional denaturants, researchers have promoted the use of micelle such as sodium dodecyl sulfate (SDS) for the denaturation of protein, and formation of higher order structures. These surfactants work at millimolar concentration which makes them extremely transparent towards the circular dichroism spectra measurement, and also, they can be used in gel electrophoresis to unfold a protein [27, 28]. It is well documented in literature that ionic surfactants have a stronger interaction with protein molecules, which unfolds the protein by salting in the peptide backbone through electrostatic interaction as well as solvating hydrophobic moieties by hydrophobic side chains [29]. Generally, neutral surfactants except few [30] are not able to unfold the proteins, whereas ionic surfactants can easily do it at a millimolar concentration, far below its critical micelle concentration (CMC) [27]. For these reasons, the anionic surfactants are thought to be 1000 times stronger denaturants than the traditional chemical denaturants like urea and guanidinium chloride

[27]. Researchers have combined the properties of both the traditional and surfactant denaturants to obtain the best possible protein denaturant n-dodecylguanidinium acetate, commonly named as dodine [31]. It is a n-alkyl derivative of guanidinium and transparent protein denaturant for circular dichroism and infrared studies [32]. Gelman and co-workers found that dodine can effectively unfold the helical domains of the fast-folding lambda repressor fragment (λ_{6-85}) and WW domain at a millimolar concentration, far below its critical micelle concentration (CMC) [31].

The unfolded state of a protein is thermodynamically unstable. To attain minimum energy and more stability, unfolded proteins tend to aggregate [2]. Peptide aggregation, which is sub-sequence of protein unfolding follows two steps; the primary step is the nucleation process in which the proteins reversibly attach to a growing core followed by the development of larger aggregates [33]. The aggregation procedure is initiated by protein segments with hydrophobic amino acid residues and β-sheet predisposition [34]. Protein aggregation can occur via different structural intermediates (oligomers) varying from unstructured amorphous aggregates to highly ordered fibrils called amyloid, generally enriched in cross-β structure (see Figure 1-1) [35, 36, 37]. Protein misfolding and the subsequent aggregation is associated with various, often highly debilitating, diseases for which no sufficient cure is available yet.

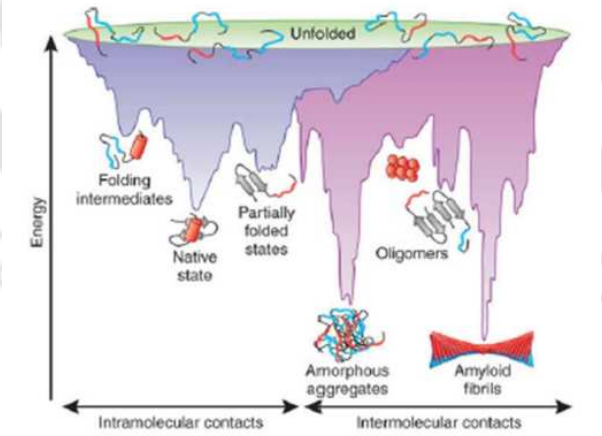

Figure 1-1. *Energy landscape scheme of protein folding and aggregation. The purple regime shows the multitude of conformations funneling to the native state via intramolecular contacts; and the pink area shows the conformations moving toward amorphous aggregates or amyloid fibrils via intermolecular contacts. Aggregate formation can occur from intermediates populated during de novo folding or by destabilization of the native state into partially folded states. Figure adapted from reference [37].*

Protein misfolding diseases are associated with a multitude of disorders, including protein aggregation and plaque formation. Protein misfolding diseases, in which the protein aggregates accumulate either systemically or locally in certain organs or tissues are commonly known as amyloidosis. Amyloidosis can be either hereditary or acquired [38]. The disease typically develops in individuals after the age of 40 and can affect both genders. The deposition of amyloid fibril can occur in all organs of the body, causing various complications such as neuropathy, congestive cardiac and renal failure, hepatosplenomegaly, and skin lesion [39].

The first known protein-misfolding disease was sickle cell anemia in which a single point mutation changes a glutamic acid in the β-globulin chain of hemoglobin into a valine [40]. Neurodegenerative diseases including Alzheimer's disease (AD), Type II diabetes (T2D), Prion protein disease, Creutzfeld-Jacob's disease, Parkinson's disease, Huntingtons disease, Gauchers disease are described as protein misfolding diseases [41, 42, 43, 44, 45, 46, 47, 48].

Alzheimer's disease is one of the most studied conformational diseases. The small amyloid-β peptide aggregates containing 40- to 42-amino acid residues (39- and 43-aa peptides have also been identified) accumulate, and form senile plaques in the brain of Alzheimers patients. Amyloid-β is cleaved off from its amyloid precursor protein (APP) by β- and γ-secretases. When the cleavage takes place in the endoplasmatic reticulum, the γ-secretase generates 42-aa Aβ_{1-42}. However, if it occurs in the trans-Golgi network, a shorter 40-aa Aβ_{1-40} is formed [2].

Human islet amyloid polypeptide (hIAPP) is an intrinsically disordered protein, and its aggregation is associated with another metabolic disease: Type 2 diabetes. hIAPP is a 37-residue peptide co-secreted with insulin by islet β-cell [41, 49, 50, 51]. The hIAPP has a variety of conformational states prone to aggregation through the N-terminus and the central regions [52, 53, 54, 55]. Small hIAPP oligomers with a high surface hydrophobicity, formed during the nucleation conformational phase, are very toxic, and permeate the membrane to produce non-selective pores [52, 55]. These form highly ordered fibrils with cross-β structure. Although some aggregates are amorphous assemblies, the fibers on the surface of the cell membranes are cytotoxic and cause the fragmentation of the membrane by micellization [52, 56].

To fully map the folding/unfolding process, all conformational ensembles (from native state to denatured state) require characterization, and the mechanism of conversion

to be explored. Experimental approaches can provide only limited information regarding these aspects. Molecular dynamics (MD) simulation is the most realistic technique available that allows time-dependent monitoring of the detailed interactions of the protein with its surrounding chemical environments. In present days protein unfolding simulations are carried out equally as they reliably depict possible folding pathways of protein when viewed in reverse [57]. The theory of microscopic reversibility dictates that the mechanism of folding can be probed in both the directions of unfolding and refolding under same conditions [57]. A huge number of research works have been reported in the approach of fully mapping the protein folding pathways over years. Anticipating the process of folding and unfolding of protein will contribute to understanding the biological processes, such as protein degradation and translocation, aging, and human diseases. In view of this, it is of relevance to investigate the unfolding of proteins, since it plays a crucial role in the development of different amyloid diseases and many other cellular processes [47, 58]. This leads us to focus on simulations of proteins in solutions to investigate the detailed processes of protein folding/unfolding equilibrium.

■ COUNTERACTION AND INHIBITION OF PROTEIN MISFOLDING

Protein conformation is sensitive to its surrounding environmental conditions such as temperature, pressure, and the nature of the molecules comprising the volume around it [59]. Some small molecules that accumulate in high concentrations inside the cell in response to osmotic stress are known as osmolytes. Nature exemplifies certain evidences of using protecting and denaturing osmolytes to restore cellular functions. For example, certain marine animals have adapted life at very high pressure and salinity by using osmolytes having denaturing effect such as urea [60] and protecting osmolyte such as betaine, trimethylamine-*N*-oxide (TMAO), trehalose, etc. to counteract the effect of denaturant [23, 24, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70]. Protecting osmolytes are often classified into two classes: 'compatible' or 'counteracting' osmolytes [71]. Compatible osmolytes increase protein stability but do not affect the functional activity of the protein, whereas counteracting osmolytes cause changes in protein function. Some methylamines (e.g. trimethylamine-*N*-oxide (TMAO), betaine, etc.) are considered as counteracting osmolytes. Some amino acids (e.g. proline and glycine) and polyols (e.g. trehalose, sucrose, and sorbitol) are considered as compatible osmolytes.

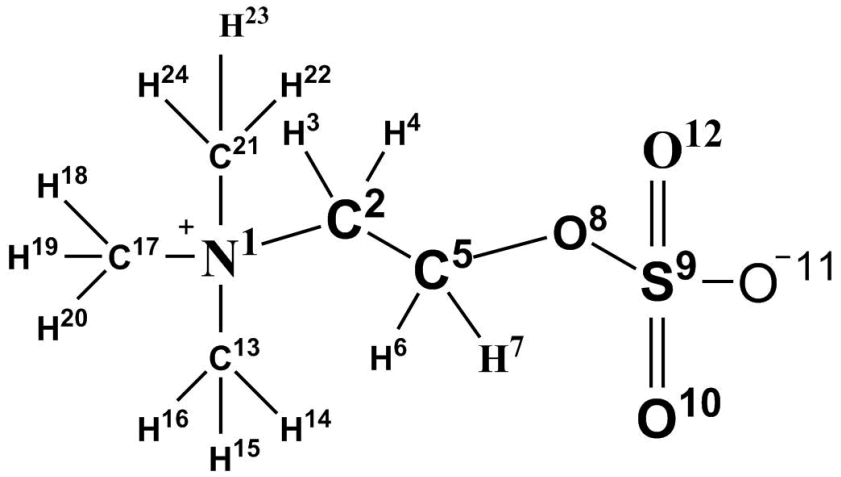

Figure 1-2. *Molecular structure and atom numbering of choline-O-sulfate.*

Choline-*O*-sulfate (2-(trimethyleammonio)ethyl sulfate), abbreviated as COS, is a molecular chaperone widely distributed in nature. It is a naturally occurring ester type molecule with a tertiary amino group and a sulfate group (see Figure 1-2). It is highly available in plants, lichens, red algae, and in spores and mycelia of the higher fungi [72, 73, 74, 75]. The most interesting part of COS is its unique sulfate group, which has distinct structural and functional properties. The sulfate group of COS plays an important role in the microbial transformation of sulfur in soil. The formation of COS also serves in the detoxification of SO_4^{2-}. It can also function as a source of choline and sulfur after hydrolysis by choline sulfutases. It is also well known as an osmoprotectant in several plants, fungal and bacterial species, for example, COS is used as a compatible solute in halophytic *Limonium* species, and *Bacillus subtilis* [72]. It is a compatible osmolyte accumulated under saline conditions by members of the halophytic genus *Limonium* and other *Plumbaginaceae* [76]. Experimental studies showed that β-alanine betaine and choline-*O*-sulfate have osmoprotective properties comparable to glycine betaine and they replace glycine betaine as osmoregulatory solutes [72]. Though the antidoting effect of different osmolyte molecules such as trehalose, TMAO, etc. on urea-induced protein denaturation has been studied rigorously for many years, the effectiveness of choline-*O*-sulfate against extreme chemical and environmental conditions is not much explored.

There is no efficient therapy for inhibiting or reversing protein aggregation and deposition in patients with protein misfolding diseases. Extensive research is going on over the past few decades for designing more efficient medications against these highly devastat-

ing disorders. The main therapeutic approaches to protein misfolding diseases are 1) small compounds that inhibit aggregation, 2) immunotherapy (antibody therapy and vaccination), and 3) compounds that interfere with amyloid-cell or amyloid-protein interactions or responses of similar kind among which inhibition of the aggregation by targeting the aggregation-prone areas of the amyloid are thought to be most efficient approach for the treatment of amyoidosis. Rigorous studies on protein misfolding delineates the inhibition of protein aggregation by small compounds, such as 2,4dinitrophenol [77], di and tri substituted aromatic molecules [78], curcumin, β-cyclodextrin derivatives, hematin, meclocycline, indomethacin, and Congo red [79]. The inhibitory effect involves the stabilization of the native fold of potentially amyloidogenic proteins or inhibition of their oligomerization or fibrillation [79, 80]. Typical inhibitors such as inorganic nanoparticles, sulfonated dyes, etc. are often toxic, and some of them are carcinogenic too [81], whereas a variety of biopolymers and macromolecules that can act as inhibitors have the difficulty of transportation through the blood-brain barrier (BBB). The BBB is the homeostatic defence mechanism against pathogens and toxins. It determines whether or not a given drug (unless lipid-soluble, small ($<$ 600 Da), electrically neutral and weakly basic), can reach the central nervous system (CNS), limiting the cerebral penetration by polymers. Therefore, much effort has been made recently focusing on the development of drugs which are safe, readily available, passes through BBB and at the same time can prevent or delay the progress of the amyloidosis.

Several osmolytes, including trehalose , α-D-mannosylglycerate and ectoine, are effective in preventing amyloid formation of Aβ peptides associated with Alzheimer's disease *in vitro* [82, 83, 84]. Since these osmolytes are non-toxic to the cellular environment, they represent potential inhibitors of neurodegenerative disorders too. Nameki et al. first studied the anti-aggregating property of COS on amylin aggregation *in vitro* and found COS as the best inhibitor in comparison to other inhibitors having similar tertiary amino groups [85]. They used the ThT fluorescence assay to compare the activities of COS and various structural analogs, including glycine betaine, carnitine, acetylcholine and non-detergent sulfobetaines (NDSBs) and found that COS is the most effective inhibitor of hIAPP amyloid formation and the order is as follows: COS $>$ NDSB-195, -201, -211, -221 and -256 $>$ acetylcholine, carnitine and glycine betaine. Moreover, as COS was reportedly a better effective inhibitor than glycine betaine, acetylcholine, or carnitine, all of which have the same quaternary amine group, this suggests that the sulfate group of COS is more important in

suppressing amyloid formation than the other acid groups.

COS can act as a precursor of choline and sulfate after hydrolysis by choline sulfatases [86]. It should be remembered that acetylcholine is a vital neurotransmitter for memory, and people with Alzheimer's disease are deficient of acetylcholine in their brains. Choline is a very important intermediate to produce acetylcholine. An acetyl group is transferred from the coenzyme acetyl-CoA to choline, catalyzed by choline acetyltransferase yielding acetylcholine [87]. So, for the production of acetylcholine, high-affinity uptake of choline is very important, and this choline is produced only from plasma or by the metabolism of choline-containing compounds. So, COS is non-toxic as well as it can be a good supplier of choline to neurons.

Due to its unique characteristic structural and functional features compared to the other solutes available in nature, we have chosen COS as a compatible solute in the protein misfolding pathway. Therefore, in the current dissertation, an effort has been made to understand the molecular mechanism of counteraction of urea- and dodine-induced protein denaturation and inhibition of protein aggregation by choline-*O*-sulfate. The simulation works and concluding remarks are presented in the subsequent six chapters. The next section of the present chapter deals with the basic techniques of MD and REMD simulations that are employed in our works. The detailed analyses of the applications of these techniques for specific systems are given in later chapters. This is followed by a brief description of the work presented in the current thesis.

■ METHODOLOGY

In this thesis, we have used classical MD simulation technique. It has been widely used to investigate the structure and dynamics of biomolecular systems, such as proteins, nucleic acids, and small molecules like amino acids, sugars and drugs. In MD simulation, the potential energy function (U) is described by all interactions between the atoms that are covalently bonded as well as non-bonded interactions between atoms and molecules in the condensed phase. The interactions between particles are governed by the so-called force field parameterization [88].

The potential energy function is written as a sum of bonded and non-bonded interaction terms

$$U = U_{bond} + U_{angle} + U_{dihedral} + U_{vdw} + U_{Coulomb} \tag{1.1}$$

The first three terms (U_{bond}, U_{angle}, $U_{dihedral}$) are the bonded terms, which describe the

bond stretching, angle bending, and torsion rotation, and the last two terms are for the non-bonded potential. In bonded terms, the bond and angle contributions are described by harmonic potentials and all of the interactions between directly bonded atoms (1-2 interactions), angles (1-3 interactions, where two atoms bonded to a common atom), and torsion (interactions between pairs of 1-4 atoms) are defined as:

$$U_{bond} = \sum_{bonds} K_b(b_{ac} - b_{eq})^2 \tag{1.2}$$

$$U_{angle} = \sum_{angles} K_\theta(\theta_{ac} - \theta_{eq})^2 \tag{1.3}$$

$$U_{dihedral} = \sum_{dihedrals} \frac{V_n}{2}(1 + cos(n\phi - \delta)) \tag{1.4}$$

The letters b, θ, ϕ, and δ represent the bond length, bond angle, dihedral angle, and phase angle, respectively. The subscripts ac stands for actual and eq stands for equilibrium. The parameters K_b, K_θ, and V_n are the force constants for bond length, bond angle, and dihedral angle, respectively.

The non-bonded potentials are calculated using two terms, the first one is the Lennard-Jones term (U_{vdw}) [89] describing the van der Waals interaction [90], and the second one is the Coulomb term ($U_{coulomb}$) [91] that deals with the electrostatic interactions between particles having partial charges on them. The non-bonding interaction terms are defined as:

$$U_{vdw} = \sum_i \sum_{i<j} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{1.5}$$

$$U_{coulomb} = \sum_i \sum_{i<j} [\frac{q_i q_j}{4\pi\epsilon_o r_{ij}}] \tag{1.6}$$

where the overall sum is over all the atom pairs i and j. Lennard-Jones parameters σ and ϵ are the diameter of atomic sites and well depth energy, respectively. r_{ij} is the inter-atomic distance. q_i and q_j are the partial charges on interaction sites i and j and ϵ_o is the electrical permittivity.

The aim of the MD simulation is to observe the evolution of atomic coordinates in time. We consider an N-particle system characterized by the following Hamiltonian

$$H = \sum_{i=1}^{N} \frac{p_i^2}{2m} + U(\mathbf{r}^N) \tag{1.7}$$

where m is the mass of each particle, p_i is the momentum of the i-th particle and $U(\mathbf{r}^N)$ is the total potential energy of the system which includes all particle-particle interactions.

The coordinates of the particles are denoted by $\mathbf{r}^N = \{\mathbf{r}_1, \cdots, \mathbf{r}_N\}$. The position and velocity of i-th particle is represented by $\mathbf{r}_i$ and $\mathbf{v}_i$, respectively. The method of molecular dynamics consists of solving the equation

$$a_i = \frac{\mathbf{F_i}}{m_i} \tag{1.8}$$

where $i = 1, 2, ...N$, m_i is the mass of i-th particle and $\mathbf{F}_i$ is the force acting on particle i. This equation is obtained easily from the Lagrangian

$$L = \frac{1}{2}\sum_{i=1}^{N} m_i \mathbf{v_i}.\mathbf{v_i} - \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} u(r_{ij}) \tag{1.9}$$

where the potential U has been assumed to be the sum of pair potentials u_{ij}. The Lagrangian equation of motion is

$$\frac{d}{dt}(\frac{\partial L}{\partial \dot{q}_i}) - \frac{\partial L}{\partial \dot{q}_i} = 0 \tag{1.10}$$

It is clear from eq. 1.10 that the dynamics of particles is described by 3N number of second order differential equations.

It is also possible to write down the Hamiltonian (H) for the system and solve the the Hamiltonian equations of motion

$$\dot{\mathbf{q}}_k = \frac{\partial H}{\partial p_k} \tag{1.11}$$

$$\dot{\mathbf{p}}_k = -\frac{\partial H}{\partial q_k} \tag{1.12}$$

where $\mathbf{q}_k$ and $\mathbf{p}_k$ represent generalized coordinates and momenta. For a system with pairwise interaction potential, the Hamiltonian is

$$H = \frac{1}{2}\sum_{i=1}^{N} m_i \mathbf{v_i}.\mathbf{v_i} + \frac{1}{2}\sum_{i=1}^{N}\sum_{j\neq i}^{N} u(r_{ij}) \tag{1.13}$$

and Eqs. 1.11 and 1.12 yield

$$\frac{dr_i}{dt} = \frac{\mathbf{p_i}}{m_i} \tag{1.14}$$

$$-\dot{\mathbf{p}}_i = -\nabla \mathbf{u} = \mathbf{F_i} \tag{1.15}$$

where i=1,2,....N. There are now $6N$ first order differential equations to be solved.

The equation of motion is solved numerically to yield particle velocities and positions as a function of time. It is usually integrated by using finite difference approach. The Verlet algorithm is one of the most commonly used algorithm for this purpose [92]. The

advantage of using the Verlet algorithm is that its implementation is straightforward and storage requirements are modest. Although, it has the disadvantage of moderate precision during the calculation and velocity does not appear explicitly in the Verlet integration. As an improvement to the Verlet algorithm, the leap-frog algorithm [93] has been developed. But, it has a disadvantage that the positions and velocities are not synchronized. As an alternative of Verlet or the leapfrog algorithm, Velocity Verlet algorithm has been developed and the following relations are used to calculate new position and velocity at the same time:

$$r(t+dt) = r(t) + v(t)dt + \frac{1}{2}a(t)dt^2 \tag{1.16}$$

$$v(t+dt) = v(t) + \frac{1}{2}[a(t) + a(t+dt)]dt \tag{1.17}$$

To calculate the velocities at time t+dt, this method requires acceleration at time t and t+dt. In the present work, we have employed Velocity Verlet algorithm.

REMD algorithm of Sugita and Okamoto [94] has become a widely-used tool for simulation of macromolecules that arises by applying the parallel tempering method to MD simulation. The REMD algorithm runs multiple isothermal MD simulations in parallel at a sequence of increasing temperatures (T_0, T_1,...,T_n) and intermittently attempts to swap simulations between temperatures i and j. The proposed swap accepted with probability P_{acc} is given by the Metropolis criterion below:

$$P_{acc} = min[1, exp[(\beta_i - \beta j)(E_i - E_j)]] \tag{1.18}$$

where $\beta = 1/k_B T$ and E_i, E_j is the configurational energy of the i and j th states respectively. If the exchange is possible, the temperatures of the neighboring replicas will be exchanged, and the velocity will be scaled and reassigned according to the new temperature; otherwise, the two replicas will continue on their trajectories with the same temperature. In this way, replicas of the system which were trapped in local minima after the exchange have gained the kinetic energy to cross higher energy barriers.

Development of Metropolis criterion involves following steps: In standard replica exchange molecular dynamics, the simulated system consists of of M non interacting copies (or replicas) at M different temperatures. The position, momentum and temperature for each replica are denoted by $q^{[i]}$, $p^{[i]}$ and T_m , and here, i=1,....,M and m=1,...,M. The equilibrium probability for the generalized ensemble is given by:

$$W(p^{[i]}, q^{[i]}, T_m) = exp\left[-\sum_{i=1}^{M} \frac{1}{k_B T_m} H(p^{[i]}, q^{[i]})\right] \tag{1.19}$$

In the expression, the Hamiltonian H(p[i], q[i]) is the sum of kinetic energy K(p[i]) and potential energy E(q$^{[i]}$). At temperature T_m , we denote p[i] , q[i] by x_m^i and further we define X = $x_1^{[i(1)]}$,...,$x_M^{[i(M)]}$ as one state of the generalized ensemble. Now we consider a pair of replicas that are exchanged. Suppose we exchange replicas i and j, which are at temperatures T_m and T_n, respectively. The exchange can be represented as:

$$X = ...; x_m^{[i]}; ...; x_n^{[j]}; ... \rightarrow X^{'} = ...; x_m^{[j]}; ...; x_n^{[i]}; ... \tag{1.20}$$

To maintain detailed balance of the generalized system, microscopic reversibility has to be satisfied, thus giving

$$W(X)\rho(X \rightarrow X^{'}) = W(X^{'})\rho(X^{'} \rightarrow X) \tag{1.21}$$

where ρ(X $\rightarrow$ X$^{'}$) is the exchange probability between two states X and X$^{'}$. An important step in the derivation of the exchange criterion is the substitution of the Boltzmann factor for the weight of each conformation into eq. 1.21, yielding eq. 1.22 (given below). Note that, it is not strictly correct until equilibrium has been reached, (this is the point at which the structures are actually considered for exchange with this probability).

$$\begin{aligned} &exp\left[-\frac{1}{k_B T_m} H(p^{[i]}, q^{[i]}) - \frac{1}{k_B T_n} H(p^{[j]}, q^{[j]})\right] \rho(X \rightarrow X^{'}) = \\ &exp\left[-\frac{1}{k_B T_n} H(p^{[j]}, q^{[j]}) - \frac{1}{k_B T_m} H(p^{[i]}, q^{[i]})\right] \rho(X^{'} \rightarrow X) \end{aligned} \tag{1.22}$$

By rearranging eq. 1.22, we obtain the Metropolis exchange probability (see eq. 1.18). It must be reiterated that eq. 1.22 is valid only for equilibrated ensembles that follow Boltzmann distribution. This assumption is followed till the end of the simulation. Use of this exchange probability results into adoption of the correct ensemble by each replica. In the standard REMD, several replicas at different temperatures are simulated simultaneously and independently for a chosen number of MD steps. Then exchange is attempted between a pair of replicas (the probability of success is calculated using eq. 1.18). If the exchange is accepted, the temperatures of the replicas will be swapped, and the velocities will be scaled accordingly. Otherwise, if the exchange is rejected, both of the replicas will continue with their current trajectories with the same thermostat temperature.

For the error estimation we have employed the well-established Block Average method described by Flyvbjerg and Petersen [95] in which the estimation of the error on a time average of correlated data is based on the correlation function for these data.

■ PRESENT WORK

For unveiling the role of Choline-*O*-sulfate, in this thesis, we have examined the counteracting effects of COS against urea (traditional chemical denaturant) as well as dodine (surfactant denaturant) induced denaturation of protein and studied the detailed mechanism of the inhibitory effects of COS on amyloid like aggregation of hIAPP as well as Aβ_{16-22} peptide.

The present chapter (**Chapter 1**) of the thesis includes a review of related experimental and theoretical works that exist in the literature together with the basic techniques of MD simulations. In this chapter, we have introduced COS as a protecting osmolyte against protein denaturation as well as a potent inhibitor in peptide aggregation through literature survey. In **Chapter 2**, we have studied the synergistic behavior of urea-COS mixture through classical molecular dynamics simulation. Here we have studied all possible interactions of urea and COS present in a mixture to find out the mechanism of the counteraction of urea by COS against urea-induced denaturation of the protein. Also, we have validated our results by comparing between two force field parameters of COS, i.e, CHARMM General Force Field parameters (CGenFF) and General AMBER Force Field (GAFF) parameters. **Chapter 3** deals with the direct application of COS as a protecting osmolyte in the protein folding-unfolding process. This chapter has been divided into two parts, **Part A** and **Part B**. **Part A** describes how COS nullifies the deleterious effects of urea on a 15 residue modeled peptide named S-peptide through classical molecular dynamics simulation. **Part B** includes the findings on the counteracting ability of COS against urea on a small globular protein called Trp-cage by an enhanced sampling method called Replica Exchange Molecular Dynamics (REMD) simulation. In this part, by employing REMD method, also we have studied the effects of urea and COS on the mini-protein over a range of temperatures (290 K to 420 K), and get the melting temperature curve of the protein. In the next chapter i.e. **Chapter 4**, we have discussed the unfolding of the terminal helices of the λ-repressor protein by an unconventional denaturant dodine and its stability in presence of COS. In **Chapter 5**, we have reported our research work on the inhibitory effects of COS in the aggregation of human islet amyloid polypeptide,

responsible for Type-II diabetes mellitus (T2Dm). In **Chapter 6**, we have presented COS as a potent inhibitor in the self-association of Aβ_{16-22} peptide, associated with Alzheimer's disease. In the last chapter i.e., **Chapter 7**, we have summarized our overall findings to bring a concrete conclusion portraying COS as an efficient osmolyte as well as an inhibitor.

Chapter 2

Synergistic Behavior of Urea and Choline-*O*-sulfate in Aqueous Solution

"Osmolyte compatibility is proposed to result from the absence of osmolyte interactions with substrates and cofactors, and the nonperturbing or favorable effects of osmolytes on macromolecular-solvent interactions."

– Paul H. Yancey et al. *Science* **1982**, *217*, 1214-1222.

Overview

Urea and COS both are osmolytes, intriguingly, having opposite effects on the protein structure. Urea is well-known for the destabilization of the protein structure. Though the increase in the intracellular pool of betaine in different types of halophytic plants against salt stress reveals COS as an osmoprotective molecule, the mechanism of accumulation is still unexplored. This chapter focuses on the detailed investigation of the interdependent behavior of urea and COS in a mixed osmolyte solution to know how urea becomes a attenuated denaturing agent in presence of COS. We have carefully studied the intra and inter-molecular interactions between the co-solutes and solvent both at room temperature and high temperature to find out the underlying mechanism behind the COS induced protection of biomolecules against urea through classical molecular dynamics simulation. Two different force field parameters i.e., CHARMM General Force Field parameters (CGenFF) and General AMBER Force Field (GAFF) parameters have been employed for this study to show the robustness and force field independency of the simulation results. Different analysis techniques have been used to explore the coherent interactions between COS and urea as well as their solvation properties in presence of each other which shows that in presence of COS, urea becomes a weaker denaturing agent than its individual counterparts. The water-water interaction shows that the mixture effect of urea and COS strengthen the water network in the system leading to the solvent assisted stability of the protein. The enhanced solvation of urea and COS in the urea-COS mixture and their mutual interactions through hydrogen bonding results in the exclusion of urea as well as COS from the protein vicinity, henceforth, provides the required stability. This synergistic behavior of urea and COS is the major reason behind the counteraction by COS against urea-induced denaturation of a protein.

■ INTRODUCTION

Urea and Choline-*O*-sulfate both belong to a very important class of small organic molecules called Smollett [18, 60, 72, 76, 96, 97, 98, 99]. Urea has well-established effects on the solubility of other co-solutes in aqueous solutions which is important for the structural stability of proteins and biomolecules [5, 6, 7, 8]. Urea is found as a waste product in mammalian kidneys and is a strong denaturant at high concentrations [9, 10] whereas osmolytes such as Trimethylamine-*N*-oxide (TMAO), trehalose, glycine betaine, etc are widely employed for the stabilization of the proteins found in the living organisms of deep sea, deserts, and icy lands, living in harsh condition like high and low temperature, high pressure, high salinity, etc [23, 24, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70]. Though the denaturation of protein by urea and counteraction of it by protecting osmolytes is the most common topic in the study of protein folding-unfolding process both experimentally and computationally, the molecular mechanism underlying this compensation is, however, still controversial. Some recent studies show that urea directly interact with protein to destabilize [7, 100, 101, 102, 103, 104, 105] whereas the competing theories agree that urea alters the water structure and dynamics hence unfold the protein through an indirect mechanism [106, 107].

The osmoprotective effects of Choline-*O*-sulfate is well established for several gram-negative bacterial species, osmotolerant and xerotolerant fungus *Penicillium fellutanum* and soil bacterium like *B. subtilis* [72, 73, 74, 75]. In detail investigation on the synergistic behavior of urea and COS in a mixture is required to know how urea becomes a weaker denaturing agent in presence of COS. In this chapter, we have taken care of every possible interaction among the solute and solvent i.e., urea, COS, and water to find out the underlying mechanism behind the counteraction of urea-induced denaturation of proteins by COS through classical molecular dynamics simulation.

In the next section, we have presented a description of the models and simulation method. Results are discussed thereafter and ended with a section whereof we have included our concluding remarks with a brief summary.

■ MODELS AND SIMULATION METHOD

We have carried out a number of classical molecular dynamics simulations with two different types of force field parameters those are CHARMM General Force Field parameters (CGenFF) [108, 109] and General AMBER Force Field (GAFF) parameters [110]. We have

prepared a total number of 18 systems involving both types of force field parameters listed in Table 2-1. S0 to S7 represents the systems including CGenFF parameters and A0 to A7 indicates the systems with GAFF parameters. System S0 contains only water, whereas in system S1 we have added urea to prepare a 8 M urea solution as 8 M urea is well-documented denaturant of protein. Systems S2, S3, and S4 contain COS in water with three different concentrations of COS i.e., 0.3 M, 0.5 M, and 1.0 M respectively. S5 to S7 represent the systems involving both COS and urea where the concentration of urea is fixed to 8 M but the concentration of COS varies from 0.3 to 1.0 M. The compositions of systems A0 to A7 are exactly same as systems S0 to S7 with GAFF parameters. We have chosen TIP3P (Transferable Intermolecular Potential 3P) as the water model for all the simulations [111]. For CHARMM parameter simulations CGenFF model of urea [108, 109] and for GAFF parameter simulations, Smith model of urea was chosen following the earlier works [112, 113, 114, 115]. All the components are packed in a cubic box using the PACKMOL program [116].

It is to be noted here that even after carrying out extensive literature searching we were unable to find the force field parameters for COS. Together with this, since no experimental results are available for counteracting ability of COS against urea-conferred protein denaturation, we could not explore the robustness of our results. So, we have compared the results using two different force field parameters for COS. For CHARMM General force field parameters of COS, CGenFF (version 3.0.1) is used (see Table 2-2) and for GAFF parameters, the RESP suite of AMBER12 program package was employed [117] to obtain its partial charges and the GAFF force field parameters were used with the help of the ANTECHAMBER module built in AMBER12 (see Table 2-3) [117, 118, 119, 120].

Simulations using CHARMM parameters were conducted using NAMD 2.10 package [121], and the other set of simulations using GAFF parameters were carried out using AMBER14 package [122]. For the systems run in NAMD, the first step of the simulation was minimization using conjugate method followed by heating in canonical ensemble (NVT) [123]. Then the systems were equilibrated for 1 ns at NPT (isothermal isobaric ensemble) [124]. Langevin piston method was applied to maintain the system pressure at 1 atm [125]. During this period, the simulation cell volumes were allowed to fluctuate isotropically. Then a final production run was continued up to 200 ns at NPT for each system. For systems S0 to S7, simulations were carried out at 310 K employing Langevin dynamics method with a damping coefficient of 5 ps^{-1} [125]. For AMBER systems (systems

A0 to A7) at first minimization was done for 10000 steps using equally distributed steepest descent method and conjugate gradient method. Then the systems were heated gradually at NVT to maintain an ambient temperature of 310 K similar to that of CHARMM systems [123]. Such heating benefits to an effective unwinding of the systems to overcome all the local minimum boundary heights. This is trailed by extinguishing the systems to the coveted temperature with a uniform interval of 25 K in canonical ensemble (NVT) followed by 1 ns equilibration for all systems at isothermal-isobaric (NPT) ensemble [124]. Subsequently, the simulations were then expanded to 200 ns in NPT ensemble to analyze the diverse contributory properties for all the systems. Langevin dynamics method was employed to maintain the temperature of the systems with a collision frequency of 1 ps^{-1} [126].

Table 2-1. Overview of Systems[a]

System	N_C	N_U	N_W	Box length (Å)	Temp (K)	M_C	M_U
S0	0	0	2000	38.97	310 K	0	0
S1	0	480	2100	45.98	310 K	0	8.20
S2	17	0	3000	45.26	310 K	0.30	0
S3	30	0	3000	45.76	310 K	0.52	0
S4	65	0	3000	46.95	310 K	1.04	0
S5	30	800	3000	53.58	310 K	0.32	8.63
S6	50	800	3000	54.07	310 K	0.53	8.40
S7	110	900	3000	56.20	310 K	1.03	8.42
S7a	110	900	3000	56.78	350 K	1.00	8.16
A0	0	0	2000	39.36	310 K	0	0
A1	0	480	2100	46.11	310 K	0	8.13
A2	17	0	3000	45.65	310 K	0.30	0
A3	30	0	3000	46.14	310 K	0.51	0
A4	65	0	3000	47.06	310 K	1.04	0
A5	30	800	3000	53.65	310 K	0.32	8.60
A6	50	800	3000	54.10	310 K	0.52	8.40
A7	110	900	3000	56.18	310 K	1.03	8.42
A7a	110	900	3000	56.74	350 K	1.00	8.18

[a] N_C, N_U, and N_W represent the number of choline-*O*-sulfate, urea and water molecules respectively. M_C and M_U are the molar concentrations of choline-*O*-sulfate and urea of the respective systems.

In case of AMBER, a constant pressure of 1 atm is maintained for NPT simulations, using Berendsen barostat with a pressure coupling constant of 2 ps [127]. An integration time step of 2 fs was used for both sets of simulations, and the trajectories were saved after every 4 ps of the simulation. A cut-off distance of 12 Å was used for the calculation of short-ranged Lennard-Jones interactions and for the long-ranged electrostatic interactions,

particle mesh Ewald (PME) method was employed [128]. All bonds involving the hydrogen atoms were constrained using the SHAKE algorithm [129] and periodic boundary condition is applied in all three directions. Also, we have simulated the systems S7 and A7 at 350 K temperature namely S7a and A7a to study the effects of higher temperature on these systems.

Table 2-2. CGenFF parameters for COS[a]

Atomic Numbers	Partial Charges (e)	ϵ (kcal/mol)	$R_{min}/2$ (Å)
N1	-0.604	-0.2000	1.8500
C2	-0.099	-0.0550	2.1750
H3,H4	0.250	-0.0460	0.7000
C5	-0.280	-0.0560	2.0100
H6,H7	0.090	-0.0350	1.3400
O8	-0.280	-0.1000	1.6500
S9	1.330	-0.3800	1.9750
O10-12	-0.650	-0.1200	1.7000
C13,C17,C21	-0.349	-0.0770	2.2150
H14-16,H18-20,H22-24	0.250	-0.0460	0.7000

[a] Atomic numbers, partial charges, ϵ and $R_{min}/2$ values for different atomic sites of choline-*O*-sulfate. e is the elementary charge.

Table 2-3. Partial charges of COS for GAFF[a]

Atomic Numbers	Partial Charges (e)
N1	0.099481
C2	-0.037744
H3,H4	0.093643
C5	0.264445
H6,H7	0.017079
O8	-0.450507
S9	1.428239
O10-12	-0.673977
C13,C17,C21	-0.397820
H14-16,H18-20,H22-24	0.187782

[a] Partial charges for different atomic sites of choline-*O*-sulfate used in GAFF force field systems. e is the elementary charge.

For visualization and analysis purposes, Visual Molecular Dynamics (VMD) plugin and CPPTRAJ implemented in AMBER were used whenever required [130, 131].

■ ANALYSIS METHODS

Site-site Radial Distribution Function

The radial distribution function (RDF) provides a relationship between the radial density of co-solvent around a solute species at a particular distance to that of the bulk solution. An excess of co-solvent molecules correlates with a value higher than 1, whereas a deficit of co-solvent yields a value lower than 1. The function can be used to characterize some structural properties of the solution and is expressed by the formula

$$g_{\alpha\beta}(r) = \frac{1}{4\pi r^2 \rho_\beta} \frac{dN(r)}{dr} \tag{2.1}$$

where r is the inter-atomic distance between atoms α and β, ρ_β is the bulk density of β and N(r) is the average number of β at a distance between r and r + dr from α. $g_{\alpha\beta}$(r)dr is the probability of finding β in the range r to r + dr. Pair distributions often display oscillating behavior with the g(r) approaching a value of 1 at the half box length. The first sphere of nearest neighbors is indicated by a peak in g(r) and limited by a minimum in g(r) and is referred to as the first solvation shell.

Hydrogen Bonding

To study the solvation of the co-solutes in different solution mixture, hydrogen bonding interaction is an important property to calculate. Following earlier works, a maximum D-A distance of 3.5 Å and a maximum D-H-A angle of 45° are fixed as the geometric criteria to determine different hydrogen bond numbers where D and A are the donor and acceptor atoms [23, 24, 66, 67, 132, 133, 134, 135, 136, 137].

Coordination Number

The coordination numbers ($n_{\alpha\beta}$) are defined as

$$n_{\alpha\beta} = 4\pi\rho_\beta \int_{r_1}^{r_2} r^2 g_{\alpha\beta}(r) dr \tag{2.2}$$

where $n_{\alpha\beta}$ represents the number of atoms of type β surrounding atom α in a shell extending from r_1 to r_2 and ρ_β is the number density of β in the system. To calculate the first solvation shell coordination number, the values of r_1 and r_2 are set to 0 and 3.5 Å respectively, following the earlier works [69, 138, 139].

Dimer Existence Auto-correlation Function

Dimer Existence Auto-correlation Function (DACF) is defined as the auto-correlation of a function β_{ij}, for a particular set of molecules called i and j and it varies from distance

1 to 0 until the distance criteria considered for the dimer formation between the molecules breaks for the first time according to the equation 2.3:[140, 141, 142].

$$DACF(\tau) = N \cdot \left\langle \sum_{t=0}^{T-\tau} \beta_{ij}(t+\tau) \cdot \beta_{ij}(t) \right\rangle_{i,j} \tag{2.3}$$

An extreme separation of 4.0 Å between the center of masses of a urea and COS molecule is defined as the distance criteria for the dimer formation. In this distance criteria, we have taken into account the nearest neighbor of a molecule to form the dimer and when this next neighbor is no longer the nearest one, the distance criterion breaks.

Potential of Mean Force

Potential of mean force between urea and COS is calculated in terms of free energy, employing two different methods i.e., Adaptive Biasing Force method (ABF) implemented in NAMD in case of CHARMM general force field parameters [143, 144, 145], and Umbrella Sampling method implemented in AMBER when GAFF parameters have been used for the systems [146].

Adaptive Biasing Force Method :

In the amenity of ABF method, an external biasing force is applied to the set of atoms under investigation along a reaction coordinate to overcome the potential energy barrier, which is difficult to achieve by the system within the computational time [143, 144, 145]. The biasing force is determined locally by the rigorous sampling of the system conformations at each step and updated continuously at the particular temperature of the system. To find out the free energy of separation between urea and COS, we have set the distances between two different pairs of atoms namely urea-nitrogen (N_{urea}) and sulfate oxygen of COS (O_{COS}) and urea-carbon (C_{urea}) and COS-nitrogen (N_{COS}) ranging from 3 to 12 Å as the reaction coordinates, and these reaction coordinates are divided into a number of windows each of having an equal length of 2 Å which are further divided into small bins of 0.2 Å . For ABF calculation, we have constructed a system with one COS and one urea in 3000 water molecules using the CHARMM parameters and initially, the system is simulated for 50 ns following the same protocol described for NAMD in the Models and Simulation Method section both at 310 K and 350 K. For the first window, the initial configuration is taken from the last coordinate of the 50 ns MD trajectory and then for the rest of the windows, the initial configurations are generated from the last trajectory of its previous window. Each of the windows are simulated for 10 ns using a harmonic force constant of 10 kcal/mol/Å to constrain the molecules into the given boundary limits.

Umbrella Sampling Method :

In the umbrella sampling protocol, similar to the ABF method, we have chosen the distance between two different pairs of atoms i.e., urea-nitrogen (N_{urea}) and sulfate oxygen of COS (O_{COS}) and urea-carbon (C_{urea}) and COS-nitrogen (N_{COS}) as the reaction coordinates which are increased from 3 to 12 Å and are further segmented into several equally spaced windows each having a width of 0.2 Å which are allowed to overlap with each other for generating a continuous and smooth energy curve. To calculate PMF using this method, we have generated a system similar to that of ABF method containing an urea and a COS molecules in 3000 water and the initial coordinates for the PMF calculation is generated from the 50 ns classical MD simulation following the same method stated above for the AMBER simulation, both at 310 K and 350 K. The starting configuration of each window is taken from the final trajectory of its previous window. To estimate the free energy between two molecules, the atoms within the COS molecule is restrained with a harmonic force constant of 10 kcal/mol/Å and the urea atoms are pulled along the reaction coordinate using an umbrella biasing potential. To obtain the free energy curves Weighted Histogram Analysis Method (WHAM) was used [147, 148].

■ RESULTS AND DISCUSSION

Effect of COS and Urea on Water Structure

For the stability of protein, the arrangement of water network around the protein is important. To check the effect of co-solutes, urea, and COS on the water structure, we have calculated the radial distribution function between the water-oxygen atoms for different systems. From Figure 2-1, we can see, with the addition of 8 M urea into the system, the RDF peak intensity gets enhanced, as urea was found to be a kosmotrope, that strengthen the water structure due to stronger water-water hydrogen bonds and a more rigid occupation of the tetrahedral coordination positions [12, 64, 149]. In presence of 1.0 M COS, the peak intensity increases further which is very much relatable with other osmolytes like Glycine betaine, TMAO, etc [12, 69]. It is noticeable that, in presence of 1.0 M COS and 8 M urea, the peak intensity is the most pronounced indicating that the combined effect of urea and COS on the water structure and hydrogen bonding network is larger than individual effects of urea and COS.

The O_W-O_W hydrogen bond plot (Figure 2-2) shows that the addition of only

urea or only COS has very little effect on the water-water hydrogen bonding but the mixed effect of urea and COS on it is highly pronounced and strengthen the water network into the system which provides the required stability to the protein molecules. It is worth mentioning here that increasing the concentration of COS, does not have much impact on the water structure.

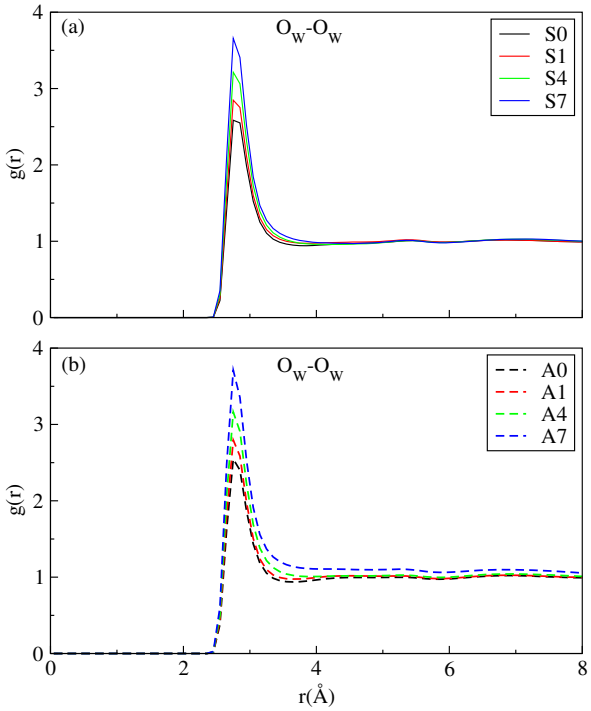

Figure 2-1. *Radial distribution function plot between the water oxygen atoms for (a) pure water (S0), urea-water mixture (S1), 1.0 M COS-water mixture (S4), and 1.0 M COS-urea-water mixture (S7) for CHARMM parameters and (b) pure water (A0), urea-water mixture (A1), 1.0 M COS-water mixture (A4), and 1.0 M COS-urea-water mixture (A7) for GAFF parameters.*

Effect of COS on Urea-Water Interaction

Next we have examined the interactions involving urea and water and the effects of COS on it. We have plotted the pair correlation functions between the urea-nitrogen and water-oxygen, and urea-oxygen and water-oxygen for both CHARMM and GAFF parameters (see Figure 2-3). The first peaks of the RDFs involving urea and water matches

well with the previous studies of urea-osmolytes interactions [12, 69]. It is noticeable that with the addition of 1.0 M COS into the system the urea-water interaction gets enhanced indicating the increased solvation of urea into the system in presence of COS due to which the urea molecules become less available for the protein molecule and hence become a poor denaturant. It is noticeable that in case of CHARMM parameters the enhancement of urea-water interaction is more significant than that of GAFF parameters. Also, the gradual increment of COS concentration into the system has very less effect on the urea-water hydrogen bonding structure.

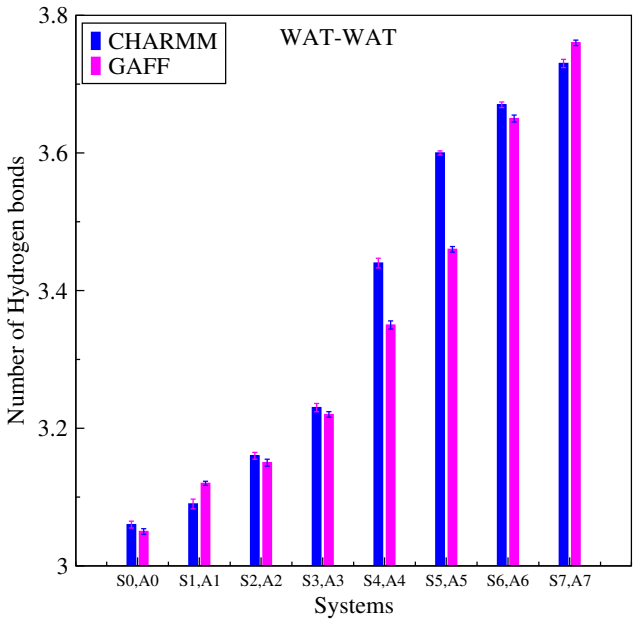

Figure 2-2. *Water-water hydrogen bond numbers (per water) for systems S0 to S7 including CHARMM parameters and systems A0 to A7 including GAFF parameters. Errors are calculated using block average method.*

The first peak of the RDF plot corresponds to the hydrogen bonding interaction, so we have measured the hydrogen bonding interaction between the urea and water to observe the effect of COS on it (see Figure 2-4). It is found that with the addition of COS into the systems the hydrogen bonding interaction between the urea and water becomes stronger.

As shown in the RDF plot, the intensity of urea-water hydrogen bonding interaction is lower in case of GAFF parameters than CHARMM parameters. Also, we have calculated the first shell coordination number of water around urea following the above mentioned criterion and found that with increasing concentration of COS, the number of water molecules gets reduced from the first solvation shell of urea due to the replacement of water by COS as expected.

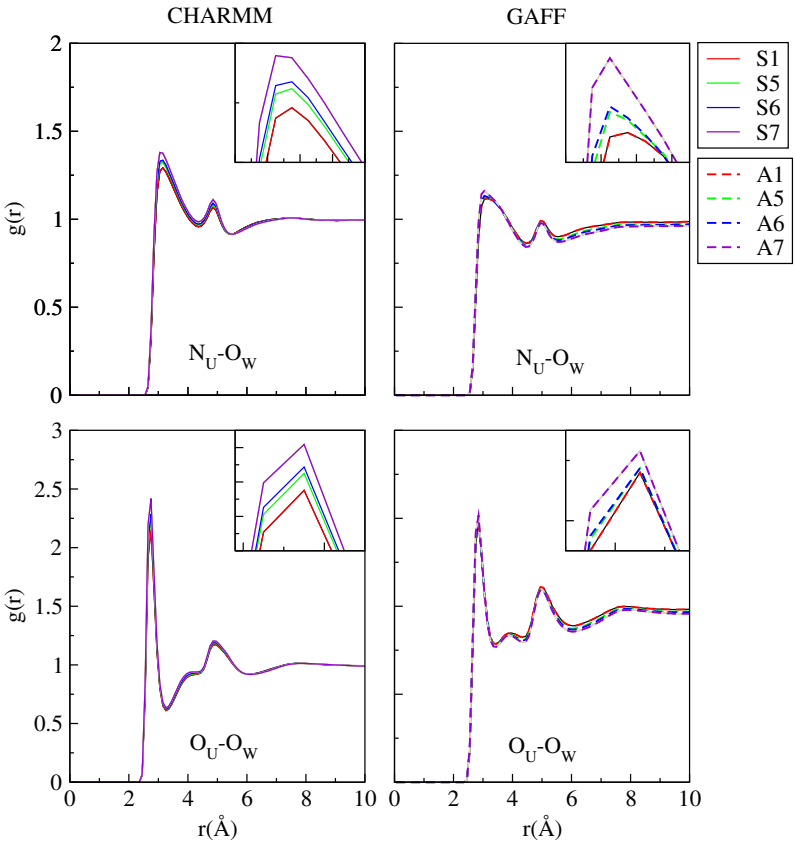

Figure 2-3. *Radial distribution function plot involving N_U-O_W and O_U-O_W for 8.0 M urea systems (S1, A1), 0.3 M COS + urea systems (S5, A5), 0.5 M COS + urea systems (S6, A6), and 1.0 M COS + urea systems (S7, A7). Left hand panel is for CHARMM parameter systems and right hand panel is for GAFF parameter systems. The magnified view of the 1st RDF peaks are provided in the insets.*

Effect of urea on COS-Water Interaction

Investigation on the COS-urea interaction needs to study the solvation of COS into the water and the effect of urea on it. In this regard, we have calculated the radial distribution function involving the nitrogen and sulfate oxygen atom of COS and water-oxygen (see Figure 2-5) and found that after the addition of urea into the 1.0 M COS system, the first RDF peak gets pronounced suggesting the increased hydrogen bonding between the COS and water. This increased solvation of COS into the water in presence of urea indicates better exclusion of COS molecules from the protein surface in the urea-COS mixture. A gradual enhancement of COS-water interaction can be seen with the increasing concentration of COS into the system.

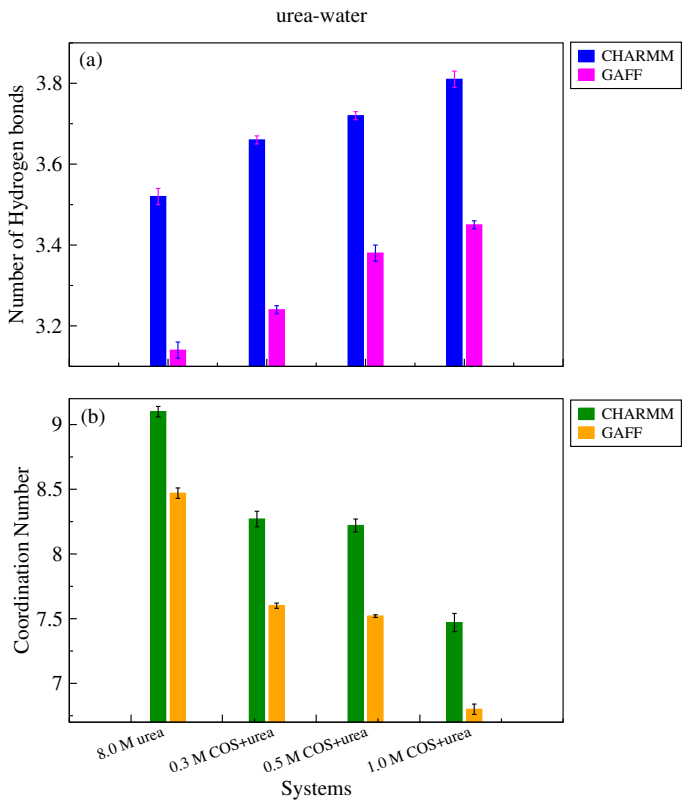

Figure 2-4. *(a) Number of hydrogen bonds between urea and water (per urea) and (b) Coordination numbers of water around urea (per urea) for CHARMM and GAFF parameters for systems containing 8.0 M urea (systems S1, A1), 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6) and 1.0 M COS + urea (systems S7, A7). Errors are calculated using block average method.*

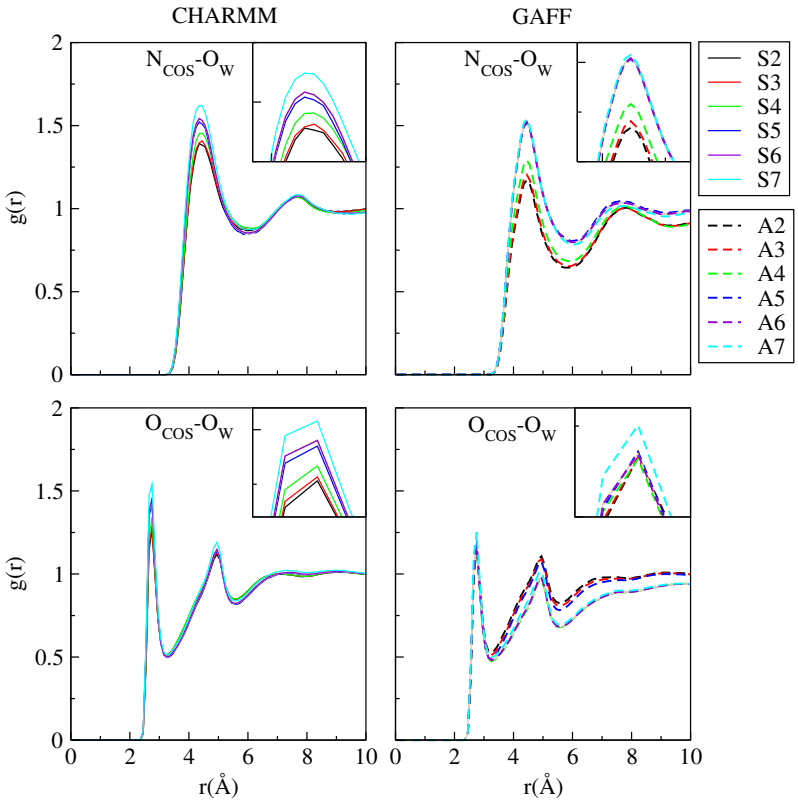

Figure 2-5. *Radial distribution function plot of N_{COS}-O_W and O_{COS}-O_W for 0.3 M COS (systems S2, A2), 0.5 M COS (systems S3, A3), 1.0 M COS (systems S4, A4) and 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), 1.0 M COS + urea (systems S7, A7). Left hand panel is for CHARMM parameter systems and right hand panel is for GAFF parameter systems. The magnified view of the 1st RDF peaks are provided in the insets.*

Considering the hints from the RDF plot, we have determined the hydrogen bond number between COS and water (see Figure 2-6, part a) and observed that with the addition of urea into the COS-water solution, the COS-water hydrogen bonding interaction becomes substantial and with increasing COS concentration this hydrogen bond number increases gradually indicating more solvation of COS molecules into water, hence, more exclusion of osmolyte from the protein backbone. Also, we have estimated the coordination number of water molecules around 3.5 Å of each COS molecules (see Figure 2-6, part b) and found that with the addition of urea into the system, the water molecules get excluded from the first solvation shell of the COS molecules due to the replacement of water molecules by urea which indicates the hydrophobic solvation of COS by urea in the urea-COS mixture.

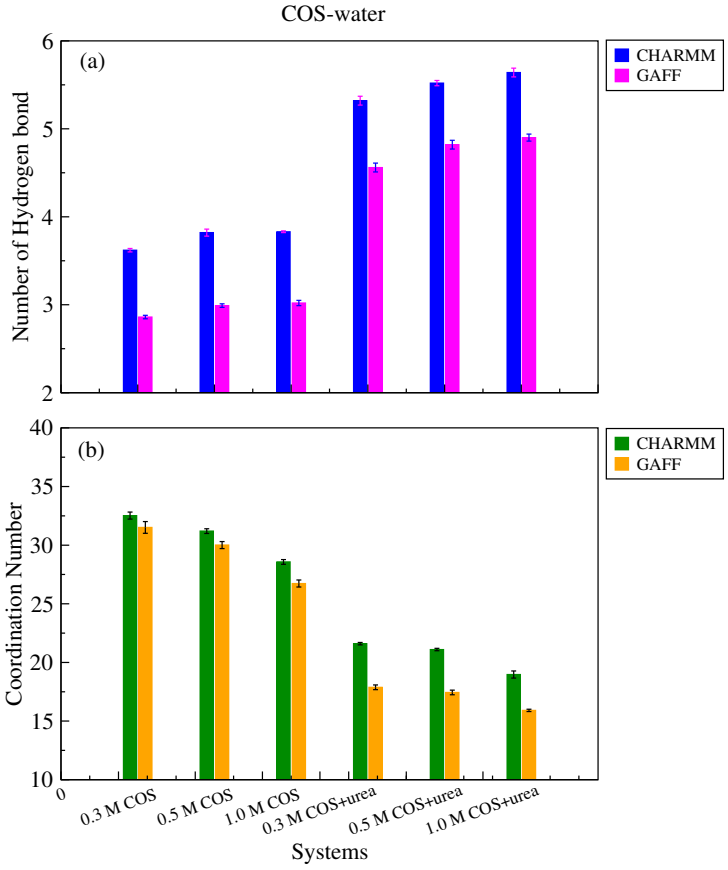

Figure 2-6. *(a) Number of hydrogen bonds between COS and water (per COS) and (b) Coordination numbers of water around COS (per COS) for CHARMM and GAFF parameters for the systems containing 0.3 M COS (systems S2, A2), 0.5 M COS (systems S3, A3), 1.0 M COS (systems S4, A4), 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7). Errors are calculated using block average method.*

Effect of COS on urea structure

The hydrogen bonds between urea and water were found to be significantly weaker than those between water molecules, which drives urea to the self-aggregation due to the hydrophobic interaction between them [149]. So, during the study of COS-urea interaction, it is important to study the effect of COS on the urea structure and hydrogen-bonded network. We have plotted the radial distribution function involving the urea-oxygen and urea-nitrogen (O_U-N_U) and urea-carbon atoms (C_U-C_U) (see Figure 2-7). The first RDF

peak of O_U-N_U indicates the presence of intra-molecular hydrogen bond between the urea molecules which get very less affected by the addition of COS into the systems. From the C_U-C_U RDF plot it can be seen that COS has no effect on the hydrophobic interaction between the urea molecules which means in presence of COS, the urea-urea structure remains preserved.

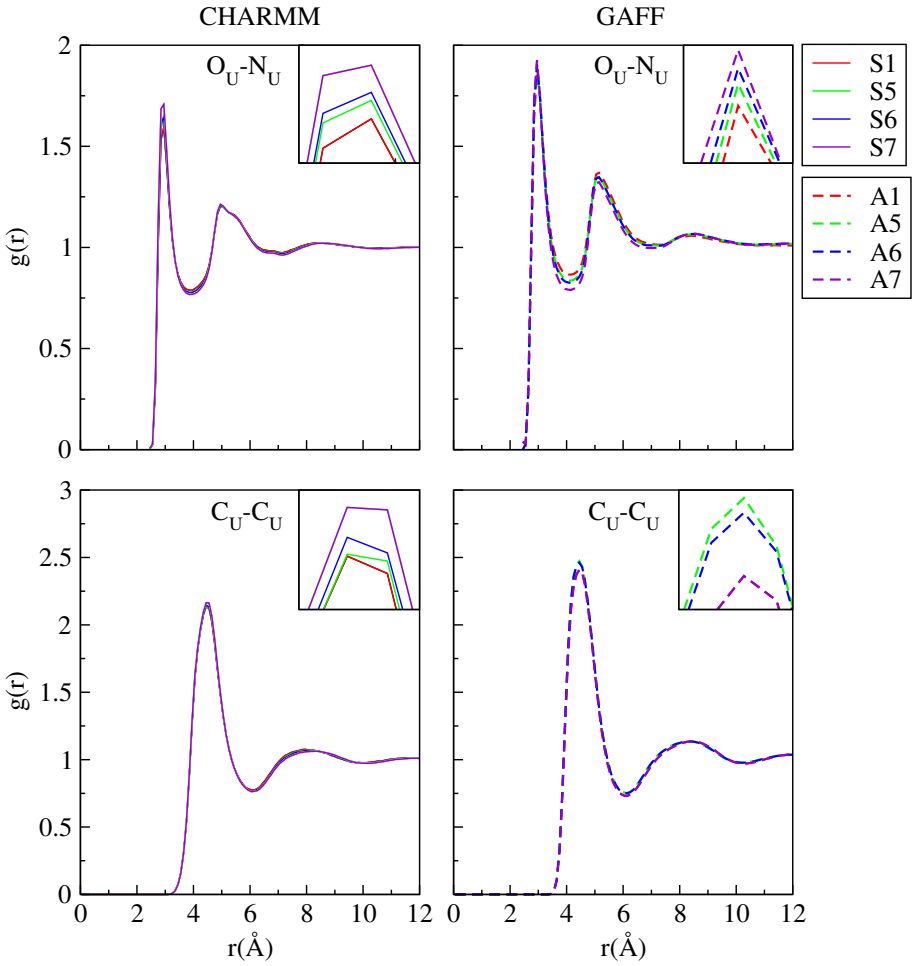

Figure 2-7. *Radial distribution function plot of O_U-N_U and C_U-C_U for 8.0 M urea (systems S1, A1), 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7). Left hand panel is for CHARMM parameter systems and right hand panel is for GAFF parameter systems. The magnified view of the 1st RDF peaks are provided in the insets.*

Also, we have checked the N_U-N_U and O_U-O_U radial distribution plot for all the urea containing systems (see Figure 2-8) and found that with the addition of COS into the

systems, these interactions get diminished slightly that means with the increasing concentration of COS, the decrease in urea structure is not so prominent. Previous studies with other similar types of osmolytes reveal that glycine betaine decreases the self-association of urea while TMAO enhances the urea structure. In our study, we find that COS does not alter the urea structure in the solution rather enhances the solvation of urea into the water to reduce its denaturing power. The hydrogen bond analysis shows that the urea-urea hydrogen bonding interaction almost remains unaffected with the increasing concentration of COS into the systems (Figure 2-9).

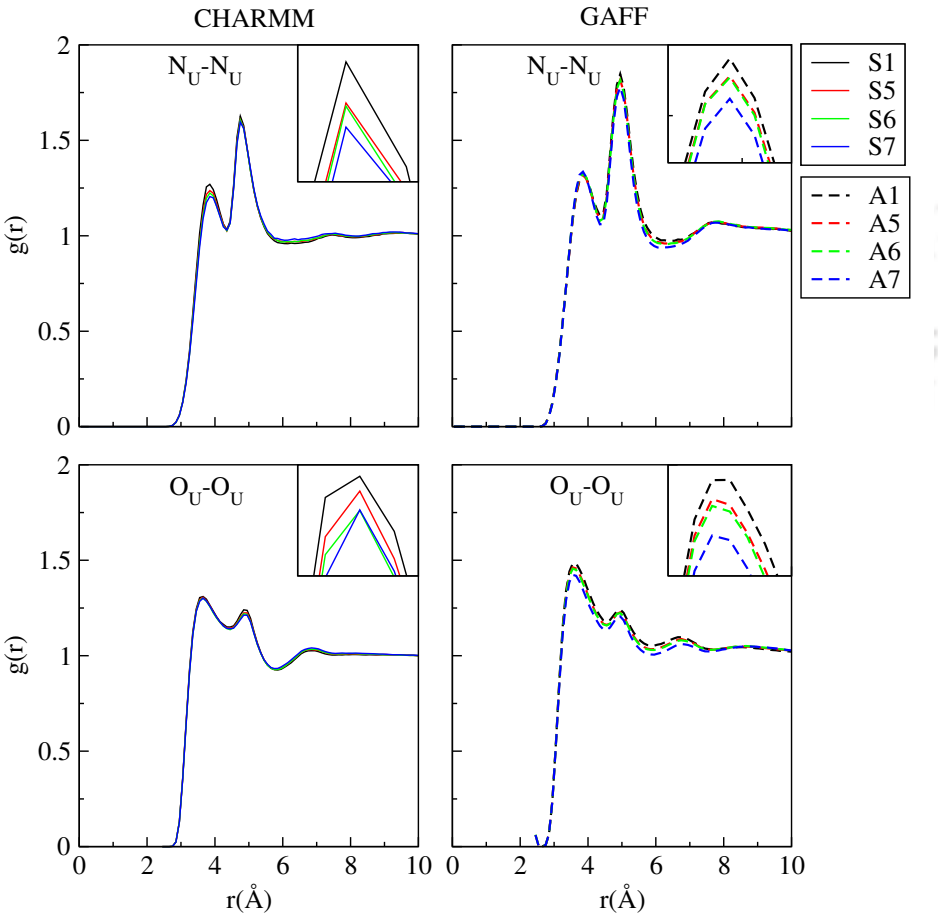

Figure 2-8. *Radial distribution function plot of N_U-N_U and O_U-O_U for 8.0 M urea (systems S1, A1), 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7). Left hand panel is for CHARMM parameter systems and right hand panel is for GAFF parameter systems. The magnified view of the 1st RDF peaks are provided in the insets.*

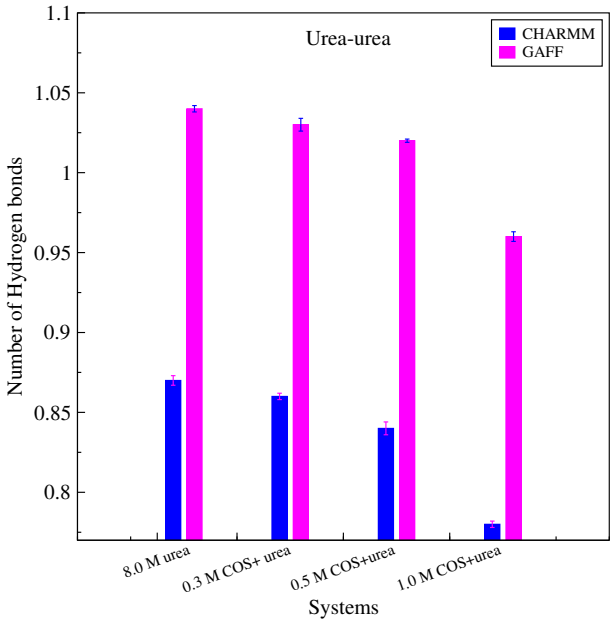

Figure 2-9. *Inter-molecular hydrogen bond numbers between the urea molecules (per urea) for CHARMM and GAFF parameters for systems containing 8.0 M urea (systems S1, A1), 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7). Errors are calculated using block average method.*

Effect of urea on COS structure

The radial pair distribution plot involving nitrogen and oxygen atoms of COS (Figure 2-10) shows a peak around 4 Å which is probably due to the non-bonded partial charge interaction between sulfate oxygen and amide nitrogen of COS and this interaction decreases significantly in presence of urea as the COS molecules get involved with the urea molecules in the mixture. Also, hydrophobic association involving the methyl carbon atoms of COS molecules gets reduced significantly with the addition of urea into the system suggesting the better involvement of COS with urea over COS. In agreement with the RDF plot data of N_{COS}-O_{COS}, the hydrogen bond number estimation between COS molecules is zero, as there is no possibility of hydrogen bond formation between them.

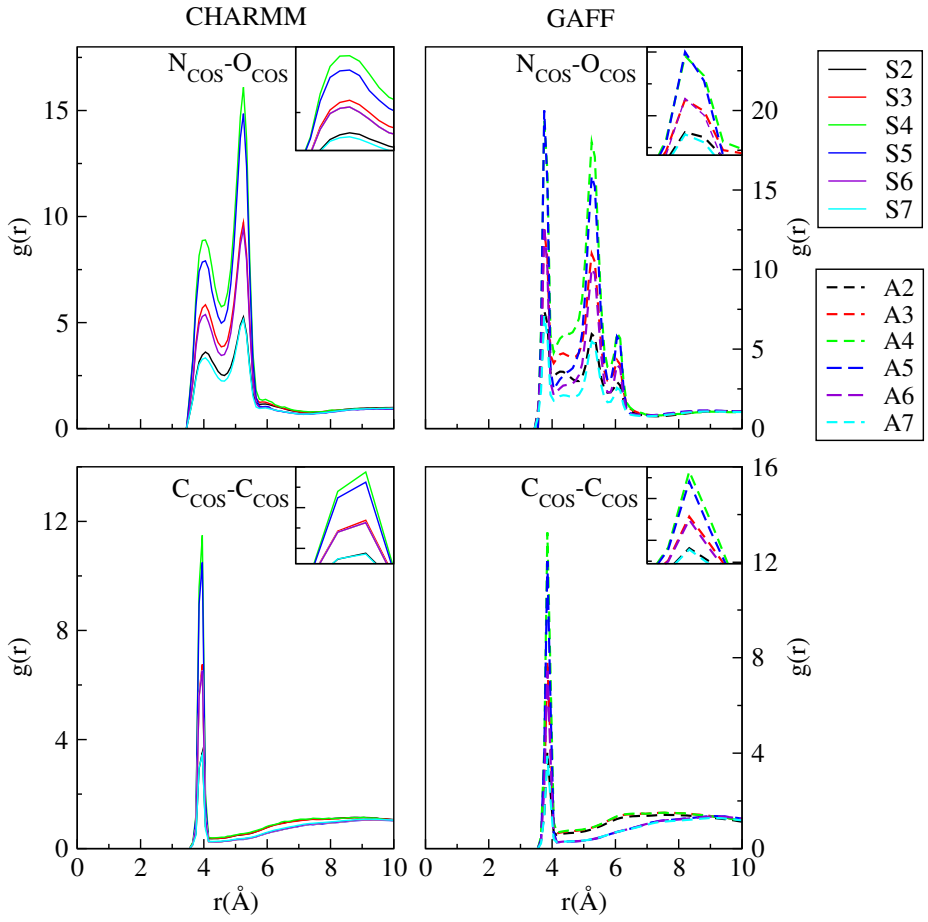

Figure 2-10. *Radial distribution function plot involving N_{COS}-O_{COS} and C_{COS}-C_{COS} for 0.3 M COS (systems S2, A2), 0.5 M COS (systems S3, A3), 1.0 M COS (systems S4, A4) and 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), 1.0 M COS + urea (systems S7, A7). Left hand panel is for CHARMM parameter systems and right hand panel is for GAFF parameter systems. The magnified view of the 1st RDF peaks are provided in the insets.*

Urea-COS interaction

To understand the role of urea-COS interactions in aqueous solution for the stability of a protein, now we have determined the probability of finding a COS molecule around a urea molecule by calculating the pair correlation function involving the COS-nitrogen and urea-oxygen atoms and COS-oxygen and urea-nitrogen atoms (see Figure 2-11). For the N_{COS}-O_U RDF plot, the appearance of a sharp peak at 4.2 Å indicates the presence of bonding interaction between COS and urea which slightly gets enhanced with increasing concentration of COS into the medium. Also, we have calculated the RDF involving O_{COS}-N_U and observed a sharp peak at 2.9 Å suggesting the existence of hydrogen bond between them and the second hump around 5 Å indicates the hydrophobic association of urea facing the methyl groups of COS. It is noticeable that the urea-COS interaction does not alter much with the addition of COS into the systems which denotes that after a certain concentration of COS, the urea-COS interaction becomes independent of the COS concentration. It is worth mentioning here that in case of GAFF parameters though the urea-water interaction was slightly weaker than the CHARMM parameters, the urea-COS interaction is much more pronounced for GAFF parameters which compensate each other and produce COS as a good protecting osmolyte using both of these force field parameters.

As suggested by the RDF plot, we have estimated the hydrogen bond numbers between the urea and COS molecules (Figure 2-12a) and observed that each COS molecules form a significant number of hydrogen bonds with the urea molecules. It is to be noted here that the hydrogen bonding interaction in case of systems built with GAFF parameters is quite stronger than the CGenFF systems. Also, the H-bond numbers do not change significantly depending on the concentration of the COS into the systems. Hence, it can be concluded that the osmolytic effects of COS in presence of urea is concentration independent after a certain ratio of urea/COS. To investigate the hydrophobic interaction between them, we have also measured the coordination number of COS around urea (per urea) (Figure 2-12b) and urea around COS (per COS) (Figure 2-12c) and observed that both of these first shell solvation numbers rise higher with the increasing concentration of COS into the systems due to the replacement of water molecules by COS and urea around each other which leads us to draw the conclusion that both hydrogen bonding interaction between urea

and COS as well as the hydrophobic solvation of urea molecules by COS play an important role to prevent the disruption of protein structure by urea molecules.

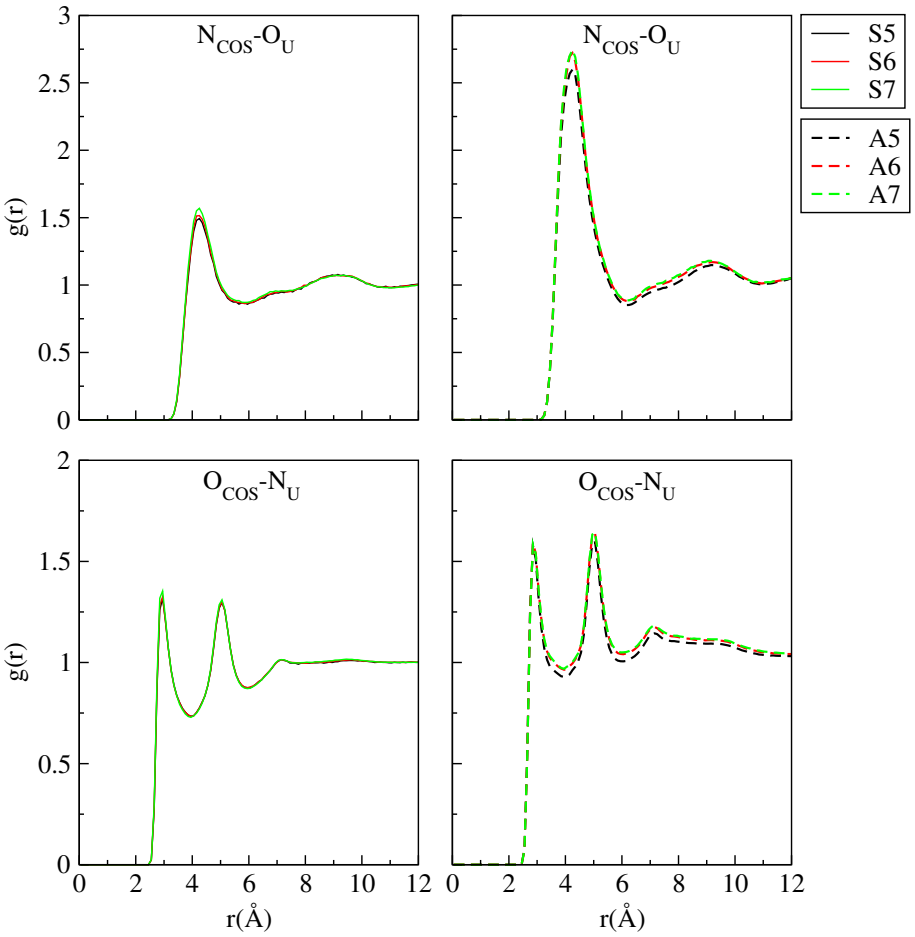

Figure 2-11. *Radial distribution function plot involving N_{COS}-O_U and O_{COS}-N_U for 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7). Left hand panel is for CHARMM parameter systems and right hand panel is for GAFF parameter systems.*

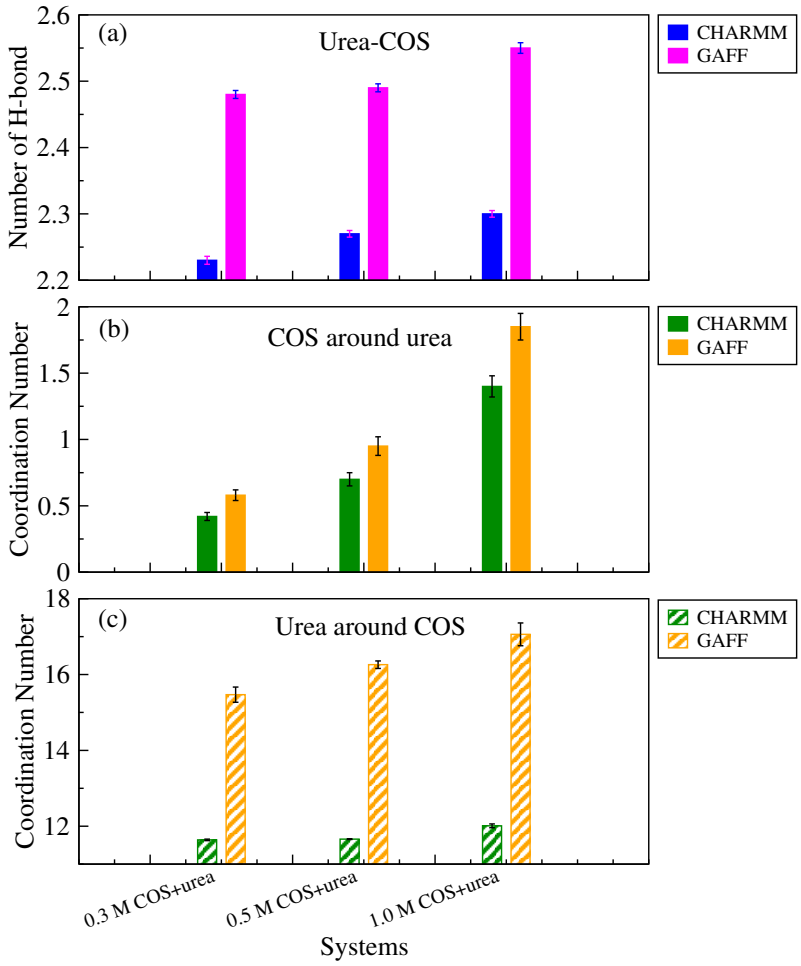

Figure 2-12. *(a) Number of hydrogen bonds between urea and COS (per COS), (b) Coordination numbers of COS around urea (per urea), and (c) Coordination number of urea around COS (per COS) for both CHARMM and GAFF parameters for the systems containing 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7). Errors are calculated using block average method.*

We have also shown the spatial distribution function (SDF) plot of COS around an urea as well as urea around a COS molecule using "TRAVIS"- a trajectory analysis software [140, 150] with an iso-surface value of 0.5. Figure 2-13 shows that with increasing concentration of COS into the solution, the density of both urea and COS increases smoothly around each other suggesting that in the urea-COS mixture, these solute particles interact with each other strongly leading to the exclusion of osmolytes from the protein backbone, henceforth, helps it to remain in its native conformation.

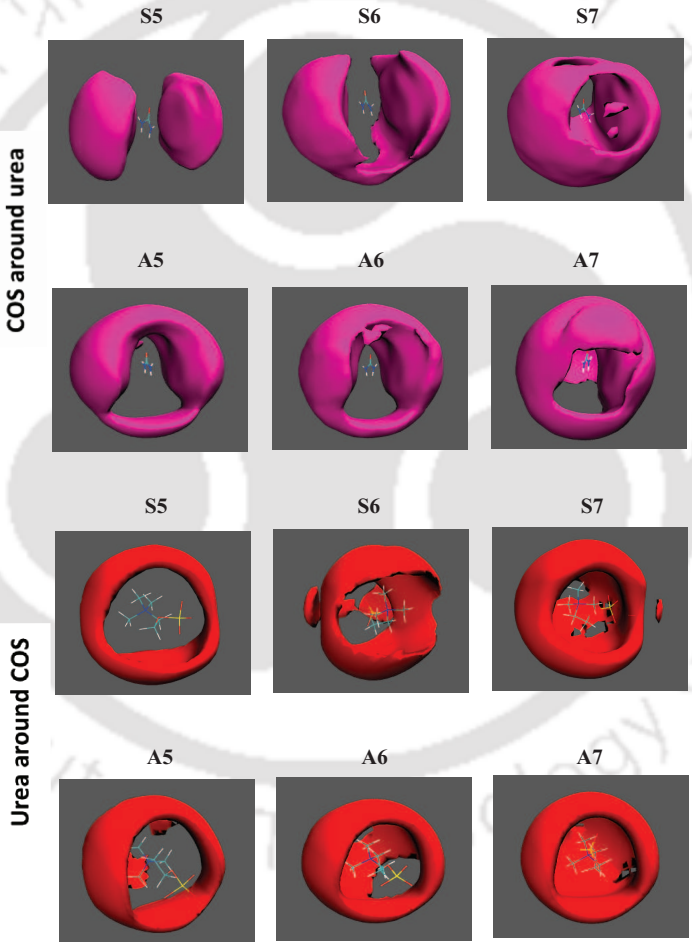

Figure 2-13. *Spatial distribution function plot of COS around urea (pink) as well as urea around COS (red) for both CHARMM and GAFF parameters for the systems containing 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6) and 1.0 M COS + urea (systems S7, A7) using an iso-surface value of 0.5.*

In order to examine the dynamics of urea-COS association in a mixed osmolyte solution, next we have estimated the lifetimes of a dimer (τ) by calculating the dimer existence auto-correlation function (DACF), following the criteria mentioned in the Analysis Method section. The DACF approaches zero at different lifetimes for different systems. From Figure 2-14, it is clearly visible that with increasing concentration of COS the lifetime of dimer increases gradually, suggesting the stronger association of urea and COS in the mixture. It is also observable that the lifetime of a dimer of urea-COS, for a particular concentration of COS, is higher in case of GAFF parameters compared to that of CHARMM parameters.

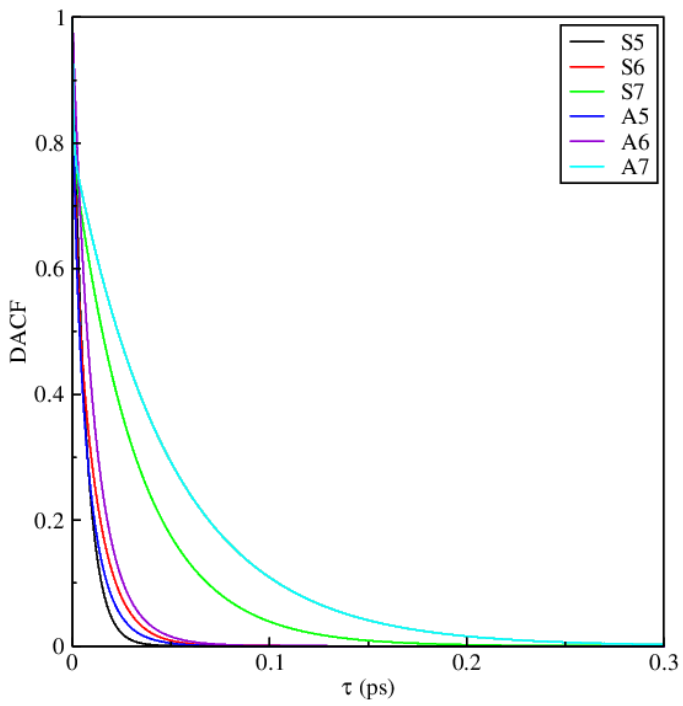

Figure 2-14. *DACF plot for the dimer of COS-urea for the systems containing 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7) for both CHARMM and GAFF parameters.*

Interaction Energy

For the detailed inspection of the interactions between the co-solutes present in the solutions, we have determined the electrostatic as well as van der Waals (vdW) interaction energies involving the COS and urea molecules into the systems (see Figure 2-15). From part a and b of Figure 2-15, it is clearly visible that the urea-urea interaction energies (both vdW and electrostatic) change a little with the addition of COS into the systems which denotes the preservation of urea structure in presence of COS. The interaction energies involving the COS molecules are very much negligible for both CGenFF and GAFF parameters indicating that COS-COS interaction is highly unfavorable in water as well as in presence of urea. Next, we have calculated the interaction energies between the urea and COS molecules present in a system (see Figure 2-15, part e and f) and noted that both the vdW and electrostatic interaction energies are sufficiently favorable between these co-solutes due to which the urea molecules become reluctant to get involved in other interactions.

Effect of Higher Temperature on Urea-COS interaction

Temperature plays a crucial role in the protein folding-unfolding process. So, to study the osmolytic effects of COS in presence of urea, it is important to investigate the behavior of co-solutes at a higher temperature in the mixed urea-COS solution. To venture into the effects of COS as a protective osmolyte at higher temperature, we have simulated the system containing 1.0 M COS in 8 M urea solution at 350 K temperature employing both CGenFF and GAFF parameters namely systems S7a and A7a (see Table 2-1). First, we have looked into the water-water interaction in this mixture at 350 K temperature. The pair correlation function involving water-oxygen atoms (Figure 2-16) shows that at higher temperature the first O_W-O_W RDF peak gets reduced significantly for both force field parameters which clearly says that in presence of 1.0 M COS in urea solution, high temperature weakens the water-water hydrogen bonding network, leading to the loss of structural stability of the protein molecule. From Table 2-4, we can see the water-water H-bond number decreases significantly at 350 K with respect to the lower temperature suggesting that at high temperature COS is less effective to prevent the urea-induced

denaturation of the protein.

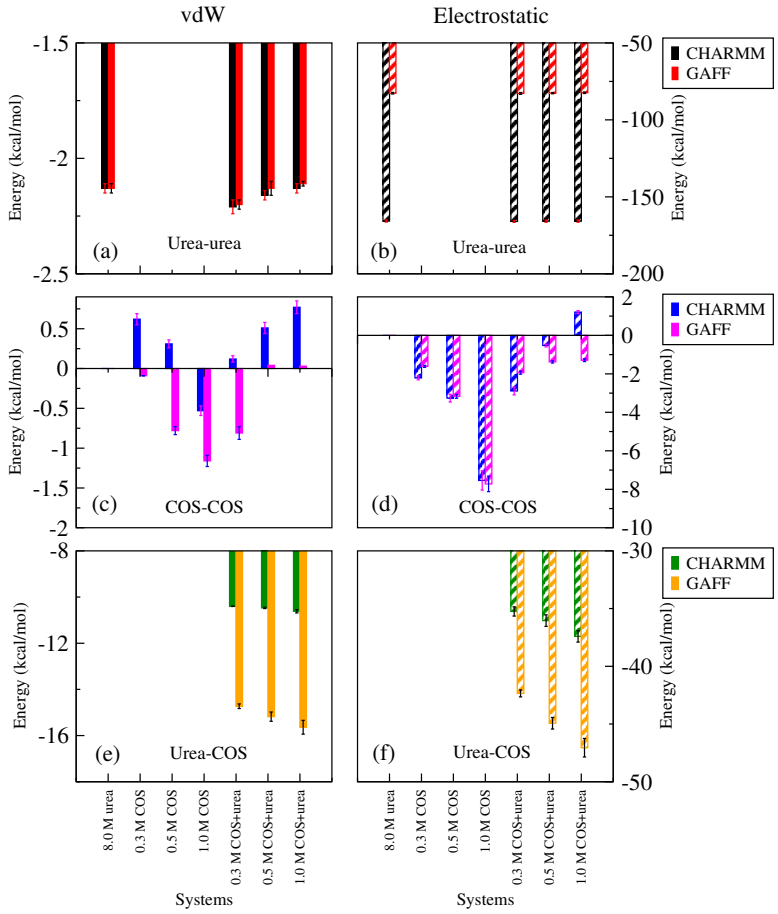

Figure 2-15. *(a) and (b) urea-urea vdW and electrostatic interaction energies (per urea) respectively for the systems containing 8.0 M urea (systems S1, A1), 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7). (c) and (d) COS-COS vdW and electrostatic interaction energies (per COS) respectively for the systems containing 0.3 M COS (systems S2, A2), 0.5 M COS (systems S3, A3), 1.0 M COS (systems S4, A4), 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7). (e) and (f) urea-COS vdW and electrostatic interaction energies (per COS) respectively for the systems containing 0.3 M COS + urea (systems S5, A5), 0.5 M COS + urea (systems S6, A6), and 1.0 M COS + urea (systems S7, A7). All energies are expressed in kcal/mol unit. Errors are calculated using block average method.*

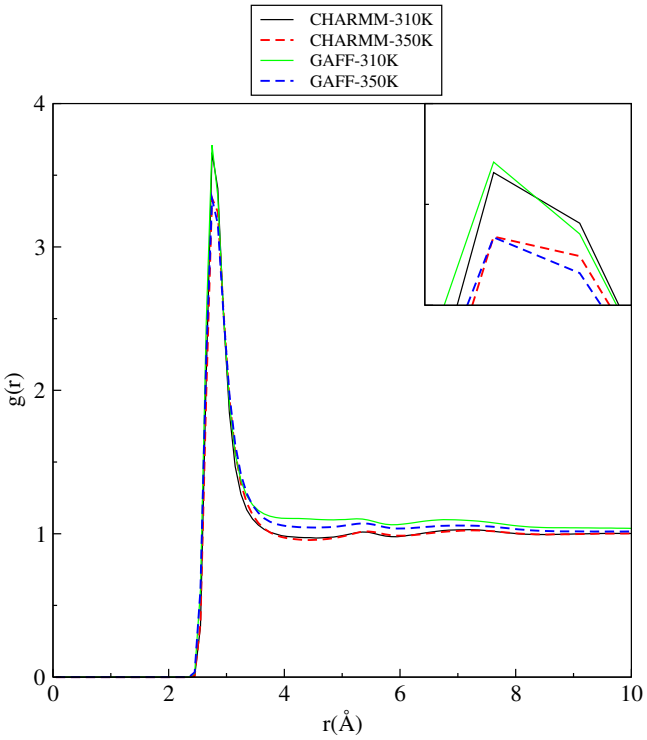

Figure 2-16. *Radial distribution function plot between the water oxygen atoms for the system containing 1.0 M COS + urea for both CGenFF and GAFF parameters at 310 K and 350 K temperatures. The magnified view of the 1st RDF peaks are provided in the inset.*

Table 2-4. Hydrogen bond numbers[a]

System	HB_{WW}	HB_{UW}	HB_{CW}	HB_{UU}	HB_{CC}	HB_{UC}
S7	3.73 (±0.006)	3.81 (±0.02)	5.64 (±0.05)	0.78 (±0.002)	0	2.30 (±0.005)
S7a	3.01 (±0.007)	3.67 (±0.03)	5.52 (±0.03)	0.72 (±0.003)	0	2.16 (±0.007)
A7	3.76 (±0.004)	3.45 (±0.01)	4.90 (±0.04)	0.96 (±0.003)	0	2.55 (±0.008)
A7a	3.09 (±0.006)	3.45 (±0.02)	4.89 (±0.03)	0.93 (±0.001)	0	2.49 (±0.006)

[a] Average hydrogen bond numbers involving water-water (per water), urea-water (per urea), COS-water (per COS), urea-urea (per urea), COS-COS (per COS), and urea-COS (per COS) where W, U, C stand for water, urea, and COS, respectively. Errors are calculated using block average method.

Next, we have focused on the solvation of urea and COS in the water at higher temperature for these systems and determined the radial distribution function involving oxygen and nitrogen atoms of urea and COS with the water-oxygen atom (see Figure 2-17). The probability distribution plot of water around COS shows that by raising the temperature, the solvation of COS molecules by water decreases slightly which denotes the comparatively less exclusion of the osmolyte from the protein surface which can promote the destabilization of the protein molecule. The distribution of water around urea also does not alter prominently with increasing the temperature of the system indicating that temperature has the least effect on the solvation of urea in water in presence of 1.0 M COS. From Table 2-4 it is observed that the decrease in hydrogen bond number between urea-water as well as COS-water is very negligible. Also, we have estimated the first shell hydration number of urea and COS (presented in Table 2-5) and noted that the water molecules around these solute particles get expelled slightly at 350 K temperature indicating comparatively less solvation of these molecules at higher temperature leading to the greater availability of osmolytes for the protein backbone.

Next, we have followed up the self-interaction of urea and COS molecules into the solution by finding the pair correlation function involving their nitrogen, oxygen atoms and carbons atoms (see Figure 2-18, part a, b, c, and d). From the reduction of the first RDF peak involving urea-oxygen and urea-nitrogen atoms, it is clear that the hydrogen bonding interaction between the urea molecules gets disrupted to some extent at high temperature which increases the number of free urea molecules into the system to interact with other solute particles. In accordance with the RDF plot data, the urea-urea hydrogen bond number (Table 2-4) decreases slightly with increasing the temperature of the system. Similarly, the pair correlation function involving different atoms of COS (Figure 2-18, part c and d) shows that raising the temperature of the system keeps the interaction between the COS molecules unchanged. Next, we have examined the involvement of urea and COS with each other at 350 K by finding the probability distribution function of sulfate oxygen of COS around urea-nitrogen and amide-nitrogen of COS around urea-carbon (Figure 2-18, part e and f). It is noticeable that at the higher temperature the first RDF peak of O_{COS}-N_U around 2.8 Å diminishes prominently suggesting the reduction of hydrogen bonding interaction between urea and COS along with the suppression of second peak around 5.1

Å which indicates the depletion of the hydrophobic association between these molecules at 350 K for both CGenFF and GAFF parameters.

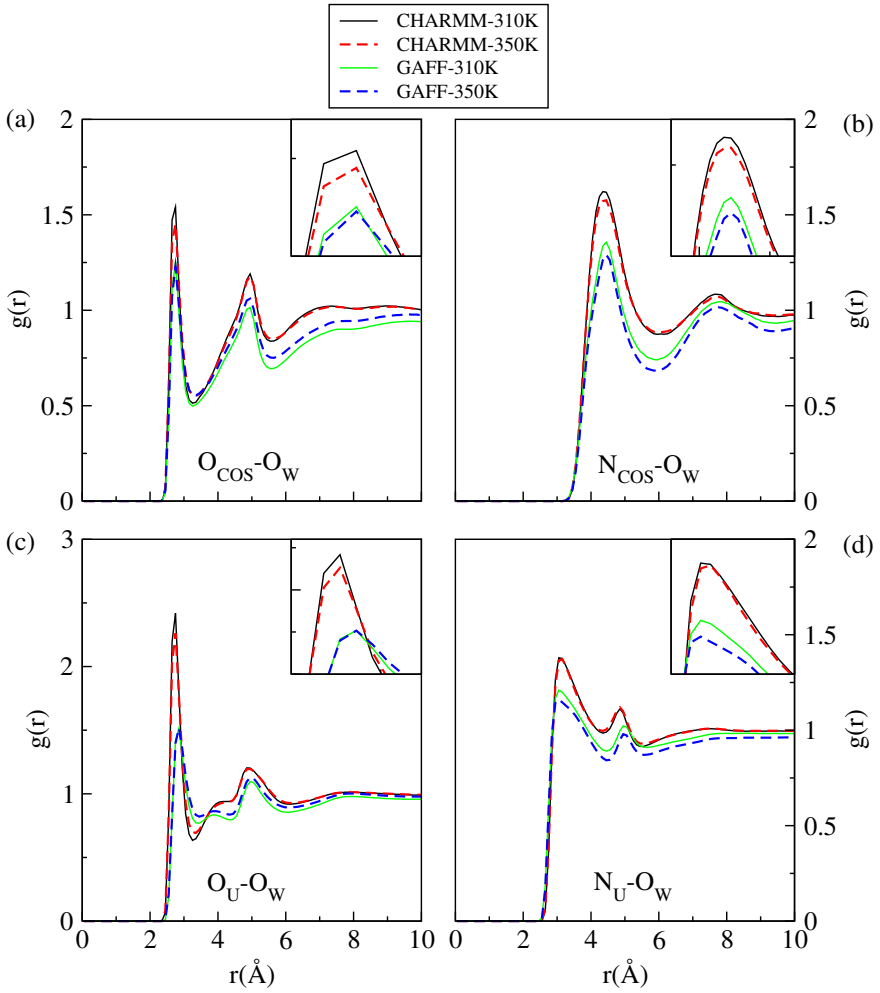

Figure 2-17. *Radial distribution function plot involving (a) sulfate oxygen of COS and water-oxygen, (b) COS-nitrogen and water-oxygen, (c) urea-oxygen and water-oxygen, and (d) urea-nitrogen and water-oxygen for the system containing 1.0 M COS + urea for both CGenFF and GAFF parameters at 310 K and 350 K temperatures. The magnified view of the 1st RDF peaks are provided in the insets.*

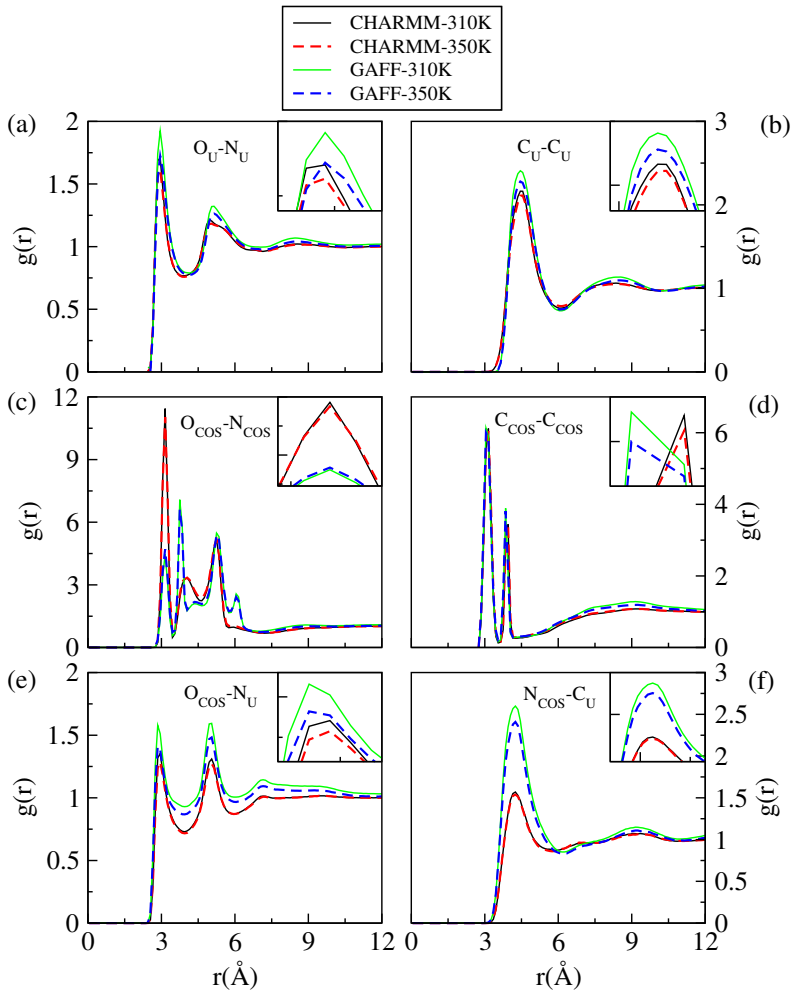

Figure 2-18. *RDF plot involving (a) oxygen and nitrogen atoms of urea, (b) carbon atoms of urea, (c) sulfate oxygen and nitrogen of COS, (d) carbon atoms of COS, (e) sulfate oxygen of COS and urea-nitrogen, and (d) COS-nitrogen and urea-carbon atoms for the system containing 1.0 M COS + urea for both CGenFF and GAFF parameters at 310 K and 350 K temperatures. The magnified view of the 1st RDF peaks are provided in the insets.*

The RDF peaks and the hydrogen bond calculation (Table 2-4) depict that by raising the temperature of the system to 350 K, hydrogen bonding interaction between urea and COS gets reduced due to which the urea molecules becomes more available for interacting with the peptide backbone leading to its denaturation. The coordination numbers

of urea around a COS and COS around a urea molecule (Table 2-5) denote that at higher temperature, the hydrophobic solvation of COS by urea and vice-versa decreases significantly promoting the urea-induced rupture of protein conformation at higher temperature even in the presence of COS.

Table 2-5. 1st shell solvation number[a]

System	N_{WU}	N_{WC}	N_{UC}	N_{CU}
S7	7.47 (±0.07)	18.97 (±0.3)	12.01 (±0.05)	1.40 (±0.08)
S7a	7.36 (±0.03)	18.28 (±0.2)	11.12 (±0.09)	1.35 (±0.03)
A7	6.80 (±0.04)	15.91 (±0.1)	17.06 (±0.3)	1.85 (±0.07)
A7a	6.73 (±0.03)	15.71 (±0.1)	15.99 (±0.2)	1.73 (±0.04)

[a] First shell coordination number of water around urea (per urea), water around COS (per COS), urea around COS (per COS) and COS around urea (per urea) where W, U, and C represent water, urea, and COS respectively. Errors are calculated using block average method.

Table 2-6. Interaction energies[a]

System	vdW_{CC}	$Elec_{CC}$	vdW_{UU}	$Elec_{UU}$	vdW_{UC}	$Elec_{UC}$
S7	0.77 (±0.08)	1.21 (±0.08)	-2.13 (±0.02)	-165.78 (±0.7)	-10.62 (±0.07)	-37.40 (±0.5)
S7a	0.09	-0.54 (±0.01)	-1.86 (±0.02)	-160.58 (±0.9)	-9.93 (±0.05)	-34.22 (±0.4)
A7	-0.03	-1.29 (±0.07)	-2.11 (±0.01)	-82.38 (±0.5)	-15.64 (±0.3)	-47.05 (±0.8)
A7a	-0.60 (±0.008)	-1.27 (±0.03)	-1.96 (±0.02)	-81.02 (±0.3)	-13.32 (±0.2)	-38.90 (±0.6)

[a] Electrostatic and vdW interaction energies involving COS-COS (per COS), urea-urea (per urea), and urea-COS (per COS). U and C represent urea and COS respectively. All energies are expressed in kcal/mol unit. Errors are calculated using block average method.

We have also determined the electrostatic and vdW interaction energies involving urea and COS molecules present in the systems for both low and high temperatures (listed in Table 2-6). It can be noticed that the urea-urea interaction energy (mainly the electrostatic interaction) becomes slightly unfavorable at the higher temperature, indicating the disruption of urea network in the solution and most importantly the urea-COS interaction energy (both vdW and electrostatic) becomes significantly unfavorable with increasing the temperature of the system for both CGenFF and GAFF parameters, indicating that

higher temperature disturbs the urea structure as well as the urea-COS interaction into the system, setting the urea molecules free to participate in other kinds of interactions.

Free Energy of Separation

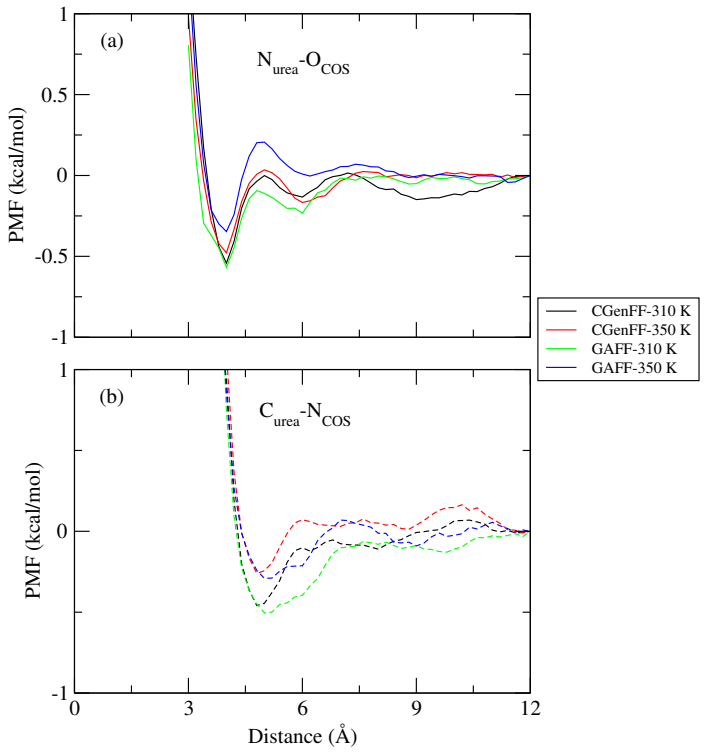

Figure 2-19. *Potentials of mean force between (a) N_{urea}-O_{COS} (b) C_{urea}-N_{COS} in terms of free energy (in kcal/mol) for the systems containing 1.0 M COS + urea for both CGenFF and GAFF parameters at 310 K and 350 K.*

To investigate the stability of the dimer formation between urea and COS in water in terms of free energy, we have computed the potential of mean force between these two molecules as it shows the continuous change in free energy as a function of the reaction coordinate of the system.

From the PMF plot between N_{urea} and O_{COS} (see Figure 2-19a), it can be noticed that the free energy curve contains the first minimum at 4 Å both at 310 K and 350 K, suggesting the formation of a direct hydrogen bond between O_{COS}...H_U, along with a second

hump near 6 Å indicating the presence of hydrophobic interaction between them. On the other hand, free energy plot involving C_{urea} and N_{COS} (see Figure 2-19b) shows a minimum at 4.8 Å indicating that the urea-COS interaction is favored by the hydrophobic interaction which is nothing but the dispersion interaction between the methyl carbons of the tertiary amino group of COS and urea-carbon. Thus, the PMF study using both CGenFF and GAFF parameters reveals the formation of hydrogen-bonded as well as hydrophobically associated urea-COS complex in a mixed osmolyte solution. Also, it is noticeable that by increasing the temperature of the system, the free energy of association between urea and COS molecules becomes less favorable, suggesting weak interaction between them at higher temperature.

■ SUMMARY AND CONCLUSIONS

Urea-COS interaction is the key feature elucidating their biological roles in the protein folding-unfolding process. This report attributes the molecular level insights into the inter-dependent behavior of urea-COS, present in an aqueous mixture, through classical molecular dynamics simulation study using two different types of force field parameters. To establish the mechanism behind the effects of the co-solutes on the peptide backbone, a thorough examination of the urea-COS interaction and their solvation in presence of each other has been done. Detailed analysis shows that in an aqueous solution of urea and COS, the increased solute-solvent interaction along with enhanced water-water hydrogen bond network excludes the osmolytes from the peptide surface owing to the solvent assisted stability of the protein. Also, the involvement of urea and COS with each other makes the former less available for the peptide backbone, hence, prevent the urea-induced unfolding of the protein. The free energy of separation between urea and COS shows that urea-COS interaction is accompanied by the hydrogen bonding interaction between the sulfate oxygen of COS and urea-nitrogen as well as the hydrophobic solvation of the urea-carbon by the methyl groups of COS molecules. Moreover, it is worth mentioning that the comparative study based on force field model dependency draws very similar conclusions on the inter-dependent interactions of urea and COS in mixed osmolyte solution stating the robustness of our simulation study. As per our knowledge, this is the first study on the synergistic behavior of urea-COS, present in an aqueous solution, revealing the indirect mechanism of

the osmolytic effects of COS against urea-induced denaturation of proteins. The effect of COS in urea-conferred protein denaturation will be treated in Chapter 3.

Chapter 3

Osmolytic Effects of Choline-*O*-sulfate Against Urea-Induced Denaturation of Proteins

"Most native proteins manifest in their immunological properties a high degree of specificity, which is diminished by denaturation."

– A. E. Mirsky, Linus Pauling. *Proc. Natl. Acad. Sci. USA* **1936**, *22*, 439-447.

The event of protein folding is associated with essential biological functionalities and unfolding of protein native state can cause intra-cellular toxicity leading to biological dysfunctions and even cell death. Urea has well-established effects on the solubility of other co-solutes in aqueous solutions which is important for the structural stability of proteins and biomolecules. Though the denaturation of protein by urea and counteraction of it by protecting osmolytes is the most common topic in the study of protein folding-unfolding process both experimentally and computationally, the molecular mechanism underlying this compensation is still unexplored [151, 152, 153]. Some recent studies showed that urea directly interacts with protein to destabilize whereas the competing theories agree that urea alters the water structure and dynamics, hence, unfold the protein through an indirect mechanism [5, 6, 7, 8, 12, 13, 57]. Though the counteracting effects of different osmolyte molecules such as trehalose, TMAO, glycine betaine, etc on urea-induced denaturation of protein have been studied rigorously for years [63, 65, 152, 154, 155, 156, 157], the effectiveness of COS against extreme chemical and environmental conditions is not much explored. Owing to the cooperative phenomenon of urea-COS in aqueous solution, discussed in the Chapter-2, in this chapter, we have shown the direct applications of COS as a protecting osmolyte against urea-induced denaturation of two different types of proteins. This chapter has been divided into two parts: Part A and Part B. Part A describes how COS nullifies the deleterious effects of urea on a 15 residue modeled peptide named S-peptide through classical molecular dynamics simulation, while Part B includes the findings on the counteracting ability of COS against urea on Trp-cage mini-protein by an enhanced sampling method called Replica Exchange Molecular Dynamics (REMD).

Part A:

Counteracting Behavior of Choline-*O*-sulfate on Urea-Induced Denaturation of S-peptide

Overview

In this chapter, classical molecular dynamics simulation is performed to explore how COS molecules countervail the denaturing effect of urea on the S-peptide analogue. The calculations of different conformational changes of the peptide heavy-atoms imply that the 4-12 residues of the peptide lose its native conformation in ∼8 M urea solution whereas in the ternary peptide-urea-COS system with 0.3 M COS concentration, the native conformation of the peptide remains preserved. The estimation of the average number of hydrogen bonds between different solution species indicates that it is the preferential urea-COS interactions, which influence peptide-urea interactions significantly. This observation is further confirmed by the calculation of the atomic density map of urea around the peptide heavy atoms and time-averaged relative local distribution functions involving peptide and urea. Moreover, the exclusion of COS molecules from the peptide surface is also confirmed by the determination of the number of COS molecules in the first solvation shell of the peptide as well as from the calculated time-averaged relative local distribution functions involving peptide and COS. A sharp drop in the diffusion coefficient values of all solution species is observed as COS is added. These findings suggest that it is the preferential solvation of urea molecules by COS, which makes the former (urea) less available for the peptide to show its deleterious effect, and hence, the native conformation of the peptide is retained.

■ INTRODUCTION

Osmolytes are small organic molecules that influence the protein stability dramatically. In the protein folding- unfolding equilibrium, unfolded (U) $\rightleftharpoons$ native (N), protecting osmolytes like betaine, glycine, glycine betaine, amino acids, trimethyleamine-*N*-oxide (TMAO), sugar trehalose, etc shift the equilibrium towards the state N, whereas denaturing osmolytes like urea and guanidine hydrochloride force the equilibrium towards the state U [151, 152, 153]. Protecting osmolytes play an important role in the intracellular protein stability and help living organisms to survive in extreme osmotic stress conditions like high pressure, very high and low temperatures, and in presence of salts. All of these compatible solutes are soluble in water and interact with the protein surface favorably or unfavorably during the denaturation and renaturation process.

It is a well-known fact that proteins get destabilized in $\sim$ 8 M aqueous urea solution and loses the ability to perform biological activities due to the change in its structural conformation. Molecular mechanism of urea-induced protein denaturation is being studied for years. Two opposing mechanisms are controversially discussed, which are by no means mutually exclusive. While the first one proposes the direct interaction between urea and protein through the backbone and polar or non-polar side chains, the second one is the indirect mechanism, which states it is the urea-induced changes in the water structure responsible for protein denaturation [5, 6, 7, 8]. The concurrent direct and indirect effects in urea-conferred protein denaturation is also widely available in the literature [12, 13, 57].

In this chapter, by using classical molecular dynamics simulation, we have studied the antidoting effect of COS on urea-induced protein denaturation. To study the urea-conferred protein denaturation and the counteracting effect of COS on it, we have used the S-peptide analog as a model peptide, consisting of 15 residues with sequence ALA-GLU-THR-ALA-ALA-ALA-LYS-PHE-LEU-ARG-GLU-HIS-MET-ASP-SER (see Figure 3A-1). The classical molecular dynamics simulations are performed using aqueous protein solutions in presence and absence of other co-solutes, namely, urea and COS. In the native state, out of 15 residues, 9 residues (residues 4-12) of the peptide remain in α-helical conformation. All the simulations are done at 310 K temperature, which is far below the melting temperature of the peptide i.e., 385 K [67]. We have calculated different structural and dynamical

quantities for arriving to the concluding remarks on the counteraction mechanism of COS on urea-induced protein denaturation.

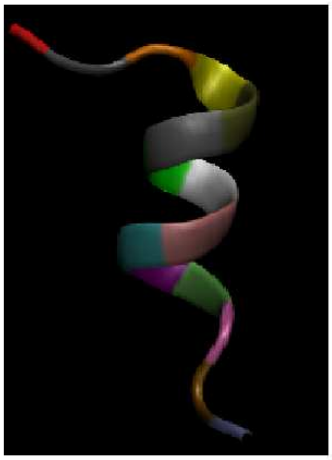

Figure 3A-1. *Initial structure of the helical peptide with residue-wise color identity.*

The remaining of this part consists of three sections including the Models and Simulation Method, Results and Discussion, and Summary and Conclusions.

■ MODELS AND SIMULATION METHOD

A series of classical molecular dynamics (MD) simulations were performed to examine the counteracting effect of COS on urea-conferred protein denaturation. For this, one S-peptide molecule was immersed separately in pure water, in 8 M aqueous urea solution as well as in two different ternary water-urea-COS systems with varying COS concentrations of 0.1 M and 0.3 M. The different systems considered in this study are outlined in Table 3A-1. Note that the first system (system S0) is considered to examine the stability of the peptide in absence of other co-solutes such as urea and COS. The very next system (i.e. system S1) gives information about the denaturing effect of concentrated aqueous urea solution on the peptide conformation. And lastly, two different ternary water-urea-COS mixtures are considered to examine how the deleterious effect of urea on peptide structure is getting abolished by increasing the COS concentration. For water, the mTIP3P (a modified version of TIP3P) model was adopted [111]. CHARMM General force field was used

for urea and COS (see Table 1-1) and CHARMM36 all-atom force field parameters were employed for the S-peptide analog [108, 109]. The same partial charges and parameters for all atomic sites of COS, presented in Table 2-2 of chapter 2, were used here. The starting configuration of each of the systems was generated using PACKMOL program [116]. In order to verify if the results of this study are dependent on the initial configurations of the systems, six more independent simulations (two more for each of the systems) of 250 ns each for systems S0, S1, and S3 were carried out. These "test systems" are termed as S0a, S0b, S1a, S1b, S3a, and S3b (see Table 3A-1). Each of these simulations was started from different initial configurations.

Table 3A-1. Overview of Systems[a]

System	N_P	N_U	N_C	N_W	volume (nm^3)	M_U	M_C
S0	1	0	0	10000	297.81	0	0
S0a	1	0	0	5000	151.42	0	0
S0b	1	0	0	10000		0	0
S1	1	2500	0	10000	477.66	8.68	0
S1a	1	1500	0	5000	260.06	9.57	0
S1b	1	2500	0	10000	481.70	8.61	0
S2	1	2500	30	10000	483.73	8.58	0.10
S3	1	1500	50	5000	270.71	9.23	0.30
S3a	1	1500	50	5000	297.81	9.24	0.30
S3b	1	1500	50	5000	297.81	9.21	0.30

[a] N_P, N_U, N_C, and N_W represent the number of peptide, urea, choline-*O*-sulfate, and water molecules, respectively. M_U and M_C are the molar concentrations of urea and choline-*O*-sulfate respectively for different systems.

The initial structure of the peptide was obtained from the Protein Data Bank (PDB ID : 2RNS) [158]. First, 15 residues of the peptide were extracted from the original PDB and some modifications were done to obtain a stable initial structure, like residues LYS1, GLU9, and GLN11 were replaced by ALA1, LEU9, and GLU11 respectively and the histidine residue was deprotonated [159, 160, 161]. To neutralize a system, one Na^+ ion

was added.

All simulations were carried out using the NAMD 2.10 code [121]. At first all the systems were energy minimized by employing conjugate method. Then the temperature of each of the systems was gradually increased to 310 K by a short MD run of 20 ps using canonical ensemble (NVT) [123]. After that, each of the systems was equilibrated for 1 ns at 310 K and at a constant pressure of 1 atm in isothermal isobaric ensemble (NPT). The temperature of the system was controlled by using the Langevin dynamics with friction constant of 5 ps^{-1}. Langevin piston method with a piston period of 100 fs, damping time constant of 50 fs, and 310 K piston temperature were employed to maintain the desired pressure [124, 125]. After the equilibration step, for all the systems a production run of 250 ns was performed in NPT ensemble. Finally, all the simulations were continued for another 10 ns in microcanonical ensemble (NVE) to calculate the diffusion coefficients of different solution species. For all the simulations a 2 fs time step was used. Particle mesh Ewald (PME) method was employed to calculate all non-bonded long-ranged electrostatic interactions [128]. To calculate all non-bonded short-ranged interactions, a cut-off distance of 12 Å was used and a switch distance of 10 Å was set up. SHAKE algorithm was used to constrain all rigid covalent bonds that involve hydrogen atoms [129]. For analysis purpose the trajectories were saved for every 2 ps.

For visualization and analysis purposes, Visual Molecular Dynamics (VMD) plugin and TCL scripts were used whenever required [130].

■ RESULTS AND DISCUSSION

Convergence

In literature, different parameters such as root mean square deviation (RMSD), intra-protein and protein-solution species average hydrogen bond numbers, helical percentage of the protein, principal components, and diffusivity have been used in order to assess the convergence of the simulations [162, 163, 164, 165]. In this study, we examined the convergence of our simulations by estimating the average number of intra-peptide hydrogen bonds as the simulation progresses for system S3 and the same is shown in Figure 3A-2. In the same figure, we have also shown the average number of hydrogen bonds between

peptide and other co-solutes and solvent with respect to simulation time. The choice of this system S3 over other systems is due to the fact that the counteraction of COS against urea is maximum for this system. Here, it is to be remembered that though we have shown both intra-peptide as well as peptide-different solution species hydrogen bonds, it is the former one in which we are more interested into. As depicted in Figure 3A-2, the fluctuation in the intra-peptide hydrogen bond is not very significant, in particular above 160 ns. This observation together with the calculations of the atom positional RMSD of the backbone C_α atoms of residues 4 to 12 of the peptide vs. simulation time (discussed below) implies that the systems are converged enough.

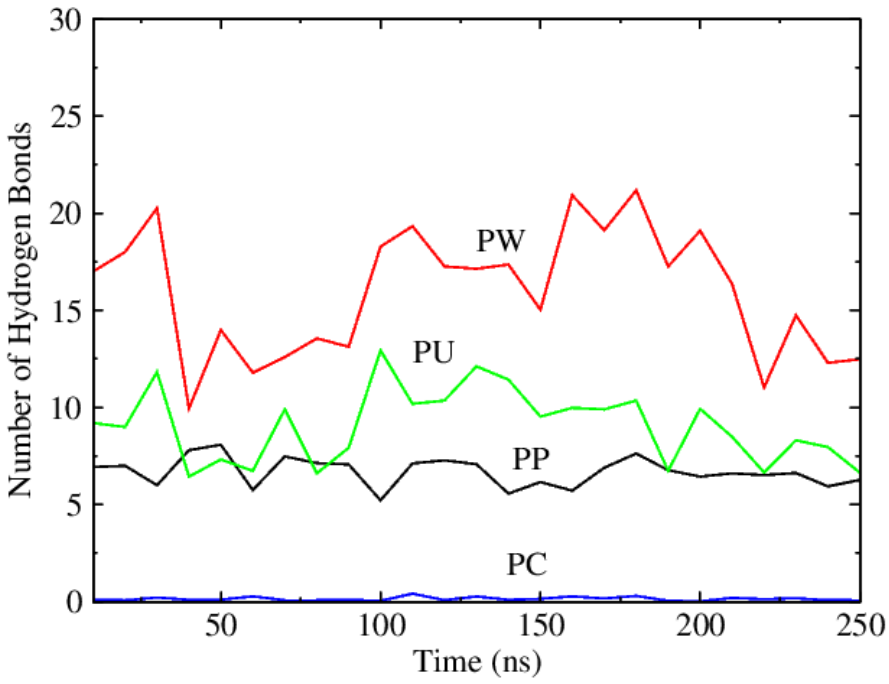

Figure 3A-2. *Average number of intra-peptide as well as peptide-solution species hydrogen bonds vs. simulation time for system S3. P, W, U, and C refer to peptide, water, urea, and choline-O-sulfate, respectively.*

Protein Conformation in Different Solutions

At first, it is important to examine how does the peptide's native conformation change on the addition of different solutes i.e., urea and COS. Towards this, for different systems considered here, we emphasized on the calculations of the atom positional RMSD of the C_α atoms of the peptide backbone of residues 4 to 12 as time progresses and they are shown in Figure 3A-3. Concentrating on the RMSD of the pure water system (system S0), first, we notice that the average RMSD value is around 0.58 Å implying a stable helical conformation of the peptide for the residues 4-12. For binary water-urea system (system S1), the examination of change in RMSD values with simulation time reveals that up to 110 ns, the peptide retains its native conformation, and then it starts to unfold. For this system, the RMSD value reaches 4.5 Å indicating that the helical conformation of residues 4-12 is not maintained. Here we note that these observations are in accordance with the results obtained from the previously reported urea-induced denaturation of S-peptide [16, 68, 166, 167, 168]. Next, to explore the counteracting effect of COS on urea-induced peptide denaturation, we considered the water-urea-COS system with a COS concentration of 0.10 M (system S2). The effect of COS on the retention of peptide conformation in aqueous urea solution is quite apparent as the RMSD decreases to 2.5 Å (from 4.5 Å in binary water-urea solution) at the end of the simulation. Thus, it can be stated that 0.10 M COS concentration protects the helical conformation of the peptide to some extent, if not fully. This finding encourages us to examine how does RMSD value change in another ternary water-urea-COS system with an elevated COS concentration (system S3). It can be seen that when COS concentration is increased to 0.30 M, the average RMSD value of the system is 0.55, which is very close to that of the pure water system (system S0). This observation suggests the counteracting ability of COS acts fully on urea-induced peptide denaturation at a concentration of 0.30 M. Furthermore, in Figure 3A-3, we also compared the change in the RMSD values of C_α atoms of 4-12 residues of the peptide for different systems with respect to the "test systems" (for example system S0 with systems S0a and S0b and so on). Remember that each of these "test simulations" were performed starting from a different initial configuration and we followed the exactly same procedure for carrying out the simulations as that adopted for the main simulations. Nevertheless, these test

simulations produce very similar RMSD values as that of the main simulation.

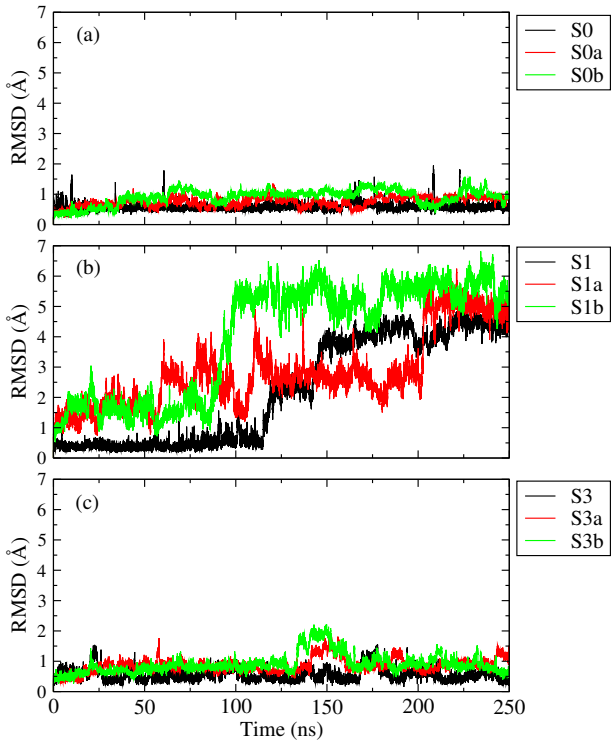

Figure 3A-3. *The atom positional root-mean-square deviation (RMSD) of the heavy-atoms of 4-12 residues of the peptide vs simulation time for systems (a) S0, S0a, and S0b; (b) S1, S1a, and S1b; (c) S3, S3a, and S3b.*

Next, the changes in peptide conformations in different solutions are probed by calculating the radius of gyrations (R_g) of backbone C_α atoms of 4-12 residues. For different systems, the changes in the R_g values vs. simulation time are shown in part a of Figure 3A-4. A much lower R_g value is obtained for systems S0 and S3 whereas, the maximum R_g value is observed for system S1 followed by system S2. Moreover, the R_g values of systems S0 and S3 are very much comparable with each other. In part b of Figure 3A-4, for all systems, we present the change in Solvent Accessible Surface Area (SASA) for the heavy-atoms of residues 4-12 vs. simulation time. SASA is defined as the area over which the contact between the solute protein and the solvent can occur and it can be calculated by the use of Sharke-Rupley algorithm [169]. According to this algorithm, many points on

the surface of each of the heavy atoms of the peptide are created, and then it is examined whether each point is occluded or exposed to the solvent. In this study, a probe radius of 1.4 Å is used to calculate SASA. It is to be remembered here that this probe radius corresponds to the radius of a water molecule. Figure 3A-4,b depicts an elevation in the SASA value for the binary water-urea system, compared to that of the system in pure water. As COS is added to the binary peptide-urea system, the SASA value decreases and the SASA values for systems S0 and S3 behave in a similar fashion. These findings are in line with that of the RMSD and R_g values of different systems discussed above.

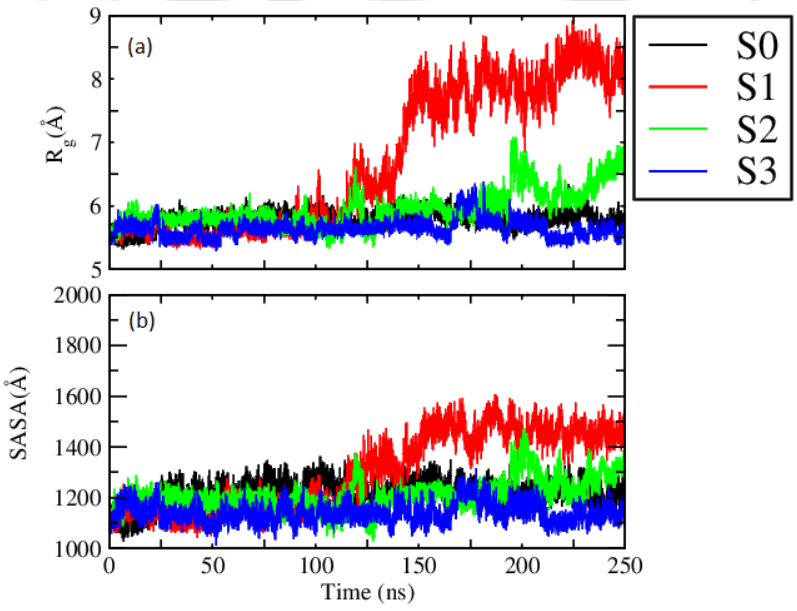

Figure 3A-4. *(a) Radius of gyration (R_g) of the backbone C_α-atoms of residues 4-12 vs simulation time. (b) Solvent accessible surface area (SASA) of the heavy-atoms of 4-12 residues vs simulation time.*

The above discussions, which imply urea-induced denaturation of the peptide in the water-urea binary mixture, a partial and a complete retention of its helical conformation in water-urea-COS ternary mixtures of COS concentrations 0.10 M and 0.30 M respectively, are further confirmed by estimating the percentage of helicity for the full peptide as well as for the residues 4-12 using STRIDE algorithm implemented in VMD [130] (see Table 3A-2). We find that for a given system, the helical percentage for the full peptide is less than that for residues 4-12, as only residues 4-12 falls in the helical region within the full peptide.

Addition of urea destroys the helicity of the peptide completely and with the addition of 0.10 M COS, the helical percentage is retained partially (100% in system S0 to 47.26% in system S2). Remarkably, in presence of 0.30 M COS solution, the helical percentage reaches the value corresponding to that of the pure water system. Furthermore, the percentages of helicity of different "test systems" are compared with respective main simulated systems (S0, S1, and S3). It can be seen that the helical percentages of the peptide in different "test systems" are very much comparable with that of the respective main systems (for example, the helical percentage of the full peptide in system S0 is 77.61% and the same for systems S0a and S0b are 74.42% and 71.42% respectively).

Table 3A-2. Helicity percentage[a]

System	Full Peptide (%)	4 to 12 Residues (%)
S0	77.61	100
S0a	74.42	100
S0b	71.42	100
S1	0	0
S1a	0	0
S1b	0	0
S2	30.30	47.26
S3	73.49	98.84
S3a	73.23	98.30
S3b	61.26	95.37

[a] Helical percentages for full peptide as well as for residues 4-12 in different systems.

The effects of addition of co-solute urea and COS on the peptide conformation are further verified by visualizing the snapshots taken at 50 ns intervals for all the systems (see Figure 3A-5). Note that, in order to have good visual clarity water and other co-solute molecules are left-off. From this figure, the deleterious effect of urea on peptide's native structure and the antidoting effect of COS on urea-induced denaturation of it can be easily identified.

So far the discussions presented above provide an overview of the peptide's conformational stability in pure water, the effects of urea alone and the effects of mixed osmolytes

with varying COS concentrations on it, without giving much detailed information about the underlying mechanism by which urea causes denaturation and the roles played by COS molecules that undo the effect of urea in ternary water-urea-COS mixture. Thus, at this point, it is important to examine at the molecular level to understand urea-induced peptide denaturation and the factors that contribute to COS-induced retention of peptide's native conformation. To answer these, in the following sections we focus our aim towards exploring the molecular level interactions between the peptide and different solution species.

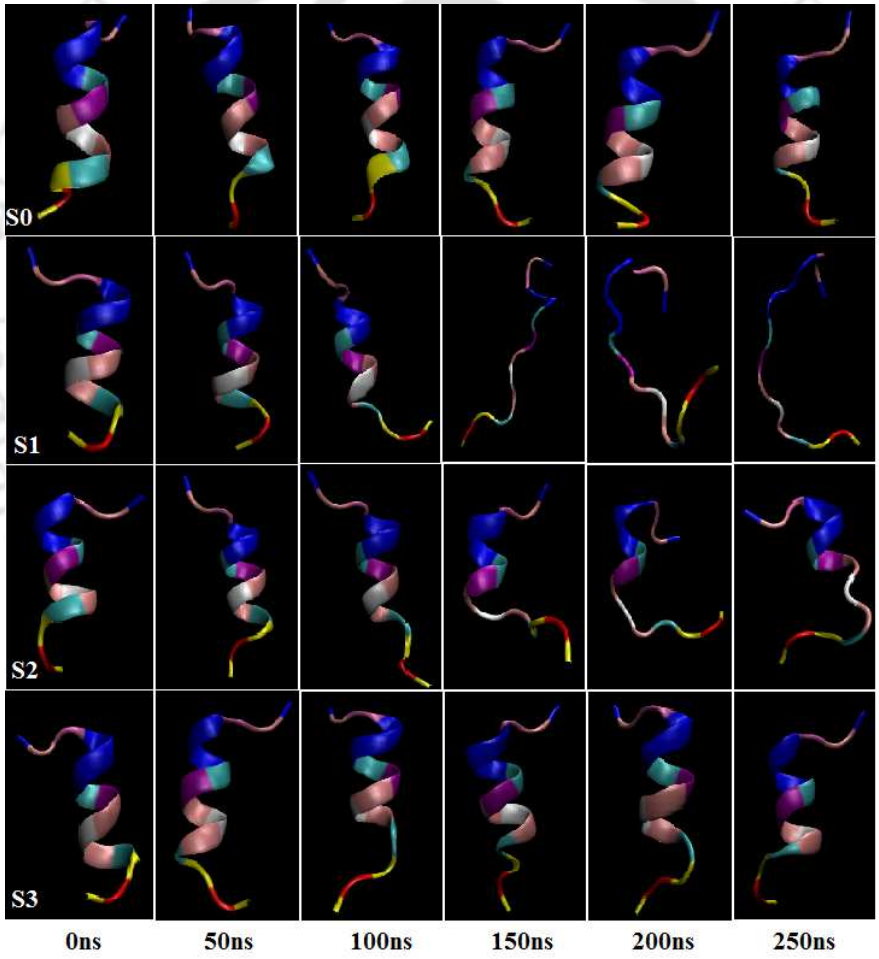

Figure 3A-5. *Snapshots of the peptide for different systems (from top to bottom). From left to right refer to the snapshots at 0, 50, 100, 150, 200, and 250 ns.*

Interaction Energy

As discussed above, the urea-induced protein denaturation follows the direct or indirect or a combination of these two mechanisms simultaneous, we have decomposed the peptide-urea interaction into two components viz. van der Waals (vdW) and electrostatic energies [16, 167]. In Table 3A-3, we present the average peptide-urea vdW and electrostatic energies for systems S1 to S3. In the same table we also show the contributions of these two energy components for interactions between other solution species. It is found that depending upon the number of incorporated COS molecules a significant change in the average peptide-urea vdW energy is observed. For example, the peptide-urea vdW energy for binary water-urea (in absence of COS) system is -96.08 kcal/mol and the same are -80.28 and -47.42 kcal/mol for ternary water-urea-COS systems, S2 and S3 respectively. Thus, addition of COS makes peptide-urea vdW energy unfavorable and this effect is more pronounced as one moves from system S2 to S3. Now, concentrating on the peptide-urea average electrostatic energy, it is observed that the addition of small amount of COS in system S2 does not have much effect on it, while in system S3, a sharp drop in the electrostatic energy of the peptide-urea interaction is observed (-277.24 kcal/ mole in system S1 to -80.40 kcal/mol in system S3). It implies that 0.30 M COS makes this interaction energy about 197 kcal/mol more unfavorable than system S1. Further, it is apparent that in the total peptide-urea interaction energy, though it is the electrostatic energy component which contributes more than that of the vdW component, the contribution of the latter is also significant. This suggests that both the vdW and electrostatic interactions between the peptide and urea contribute to peptide denaturation.

Table 3A-3. Interaction energies[a]

System	S0	S1	S2	S3
PW (vdW)	-39.84	-20.96	-20.06	-18.64
PW (elec)	-389.55	-251.10	-230.54	-149.40
PU (vdW)	—	-96.08	-80.28	-47.42
PU (elec)	—	-277.24	-297.60	-80.40
PC (vdW)	—	—	-1.05	-1.16
PC (elec)	—	—	-8.60	-7.72
UC/C (vdW)	—	—	-10.62	-11.28
UC/C (elec)	—	—	-37.68	-39.97
WC/C (vdW)	—	—	-5.61	-4.96
WC/C (elec)	—	—	-72.98	-66.49
UW/U (vdW)	—	-1.56	-1.53	-1.27
UW/U (elec)	—	-22.93	-22.63	-20.68

[a] The decomposition of total interaction energy into van der Waals (vdW) and electrostatic (elec) energy components between different solution species. P, W, U, and C refer to peptide, water, urea, and choline-*O*-sulfate respectively. The term UC/C, WC/C, and UW/U represent average urea-COS energy per COS molecule, water-COS energy per COS molecule, and urea-water energy per urea molecule. All the energy values are expressed in kcal/mol unit.

Thus, it is of importance to look at the peptide-COS interactions in different systems, as it can provide molecular insights on whether the less favorable peptide-urea interaction in system S3 in comparison to system S1 is the direct result of more favorable peptide-COS interaction or not. As presented in Table 3A-3, it can be noticed that the peptide-COS interaction, which is mostly dominated by electrostatic energy component, is negligible compared to the peptide-urea interactions. This leads us to examine the indirect role of COS molecules to wipe out the effects of urea on peptide conformation. In this context, we have also estimated average urea-COS interaction energies per COS molecule for systems S2 and S3 (see Table 3A-3) and obtained highly favorable vdW and electrostatic urea-COS interaction energies, dominated by electrostatic component. Also, with increasing concentration of COS into the systems, these interaction energies become

more favorable.

Regarding peptide conformation in aqueous medium, since water molecules play a vital role, it is of utmost importance to examine how do the peptide-water interactions get modified in different aqueous solutions of urea and COS. For this, as discussed above, we have also decomposed total peptide-water interaction energy into electrostatic and van der Waals components (Table 3A-3). It is apparent that for all systems, it is the peptide-water electrostatic interaction, which dominates over its vdW component. Concentrating on peptide-water vdW energy first, it is interesting to notice that in binary urea solution (system S1) this energy becomes less favorable compared to that of system in pure water (system S0). Remarkably, addition of COS (systems S2 and S3) does not influence peptide-water vdW energy component much. Now, on examining peptide-water electrostatic energy component for different systems it can be noticed that the addition of urea to peptide water system makes this interaction energy more unfavorable as urea replaces some water molecules from the peptide solvation shell (discussed below). Moreover, the presence of COS causes a further unfavorable peptide-water electrostatic interaction. But, we also find that the number of water molecules in the peptide's solvation shell does not change much as one moves from binary peptide-urea system (system S1) to ternary peptide-urea-COS systems (systems S2 and S3), which implies, albeit indirectly, that the unfavorable peptide-water interactions for systems S2 and S3 (when compared to that for pure water system) are not purely due to less number of available water molecules in the peptide solvation shell (discussed below).

The interesting change in peptide-water interaction energies with the addition of urea and COS for different systems (as discussed above) has encouraged us to investigate the interaction energies between water-urea and water-COS for different systems (see Table 3A-3). It is found that water-urea interaction energy is dominated by electrostatic interactions between them and the contribution of vdW to total water-urea interaction energy is negligible. What is remarkable is that the presence of COS does not influence the interactions between water and urea much on going from system S1 to S3. Interestingly, the highly negative value of water-COS interaction energy makes it the most favorable one among all other interactions. Moreover, as we move from system S2 to system S3, water-COS interaction energy becomes slightly unfavorable, setting the COS molecules free to

interact with the urea molecules more.

Thus, the discussions so far lead us to propose that it is not the favorable peptide-COS interactions that nullify the deleterious effect of urea on the peptide, rather the preferential urea-COS interactions that prevent the peptide molecule to interact with urea molecules, this in turn, helps it to retain its native conformation. These findings also direct us to examine different structural and hydrogen bond properties across all the systems (discussed below), which we believe, would provide a much deeper insights on the interactions between different solution species and would also shed some lights on the underlying molecular mechanism of COS-induced retention of native conformation of the peptide.

Peptide Solvation

In this subsection, we try to examine the average structures of different solution species. Towards this, we first consider the pair-correlation functions involving the peptide-water and peptide-urea. Then, by using VMD, the average mass density maps are determined, which would help us in visualizing the densities of different solution species viz. water, urea, and COS around the peptide surface. In Figure 3A-6, the radial distribution functions (RDFs) involving peptide heavy-atoms and water-oxygen (Figure 3A-6, part a), peptide backbone-oxygen and water-hydrogen (Figure 3A-6, part b), and peptide backbone-hydrogen and water-oxygen (Figure 3A-6, part c) are shown. Focusing on the RDF of heavy atom-water oxygen first (see Figure 3A-6a), we observe the appearance of a small first peak around 2.8 Å and together with this the appearance of a slightly more pronounced second peak around 3.8 Å is also quite noticeable. The first peak can be attributed to the existence of hydrogen bonding interaction between the peptide and water and the appearance of the second peak is due to hydrophobic hydration. The height of the first peak in this RDF is far below the bulk density which suggests the exclusion of water molecules from the vicinity of the peptide surface. Considering the fact that protein folds spontaneously in nature, the exclusion of water molecules from the peptide surface is not surprising. Now, looking into the effect of urea and COS on this RDF, we find that urea alone increases the heights of both these peaks. The addition of COS into the binary urea solution decreases the peak heights, and the effect is more pronounced in the

ternary solution of 0.30 M COS concentration. Furthermore, a comparison of the water density around the peptide backbone oxygen and hydrogen atoms (see Figure 3A-6, part b and c) reveals that the water density is much higher around the former than that for the latter. Moreover, the effect of urea and COS on these RDFs are very similar to that for peptide heavy atom-water oxygen radial distribution function. These findings imply that the peptide backbone oxygen plays the most significant role in participating in the hydrogen bonding interaction.

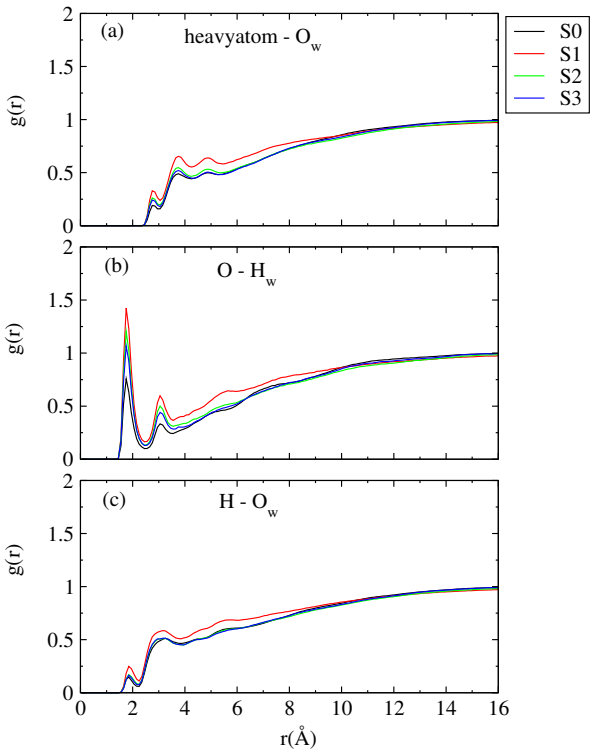

Figure 3A-6. *Pair-correlation functions involving: (a) peptide heavy-atoms and water-oxygen, (b) peptide backbone-oxygen and water-hydrogen, and (c) peptide backbone-hydrogen and water-oxygen.*

The distribution functions involving peptide and the selected atomic sites of urea are displayed in Figure 3A-7. It is noticed that the addition of COS decreases the first peak height of peptide heavy atom-urea carbon RDFs (see Figure 3A-7, part a). Similar trends are also observed for the pair correlation functions involving peptide backbone oxygen-urea hydrogen and peptide backbone hydrogen-urea oxygen (see Figure 3A-7, part b and c).

This indicates that on the addition of COS, more and more urea molecules are getting excluded from the peptide solvation shell and this is further confirmed by the calculations of the contour density maps of urea around the peptide (see Figure 3A-8).

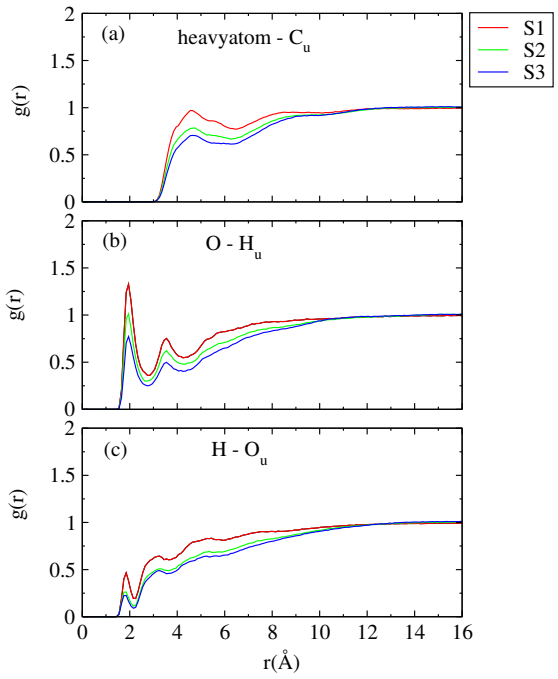

Figure 3A-7. *Site-site radial distribution functions involving: (a) peptide heavy-atoms and urea-carbon, (b) backbone-oxygen and urea-hydrogen, and (c) backbone-hydrogen and urea-oxygen.*

Here it is worth mentioning that for the determination of all the density maps, VolMap plugin implemented in VMD is used with an iso-value of 0.5 and a cell side of 0.5 Å and a 3.5 Å cut off distance from the peptide surface are considered. The VolMap plugin creates volumetric maps (3D grids containing a value at each grid point) based on the atomic coordinates and properties of a specified atom selection. It can compute different types of 3D maps among which we have computed the number density map. As can be seen that the addition of COS to the binary peptide-urea system (system S1) causes a depletion in the number of urea molecules from the close proximity of the peptide surface and this effect is very prominent for system S3. Furthermore, it is quite apparent from part b and c of Figure 3A-8 that the backbone oxygen atom of the peptide is more accessible

for urea molecules than the backbone hydrogen. It is worth mentioning here that these observations do not say anything about the direct roles played by COS molecules to prevent urea-induced protein denaturation.

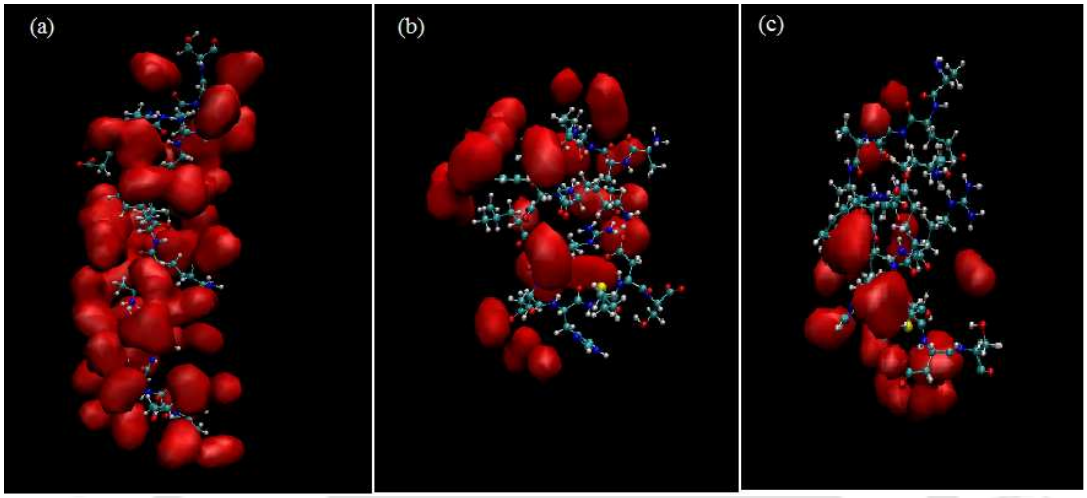

Figure 3A-8. *Contours of urea density within 3.5 Å of the peptide surface for systems (a) S1, (b) S2, and (c) S3 using an iso-surface value of 0.5.*

However, to examine whether the urea molecules are replaced by COS from the vicinity of the peptide surface, we, further, examine the distribution functions involving the peptide and COS as well as the contour density maps of COS around the peptide surface, in the same way described above, for the systems S2 and S3, and are presented in Figure 3A-9 and Figure 3A-10 respectively. It is observed that for both systems S2 and S3, the densities of COS molecules around the peptide surface are quite low compared to the density of urea and this finding excludes the fact of replacement of urea by COS molecules through direct peptide-COS interaction. This is, further, confirmed from the estimations of the number of COS molecules around the peptide surface as well as from the calculations of the average number of peptide-COS hydrogen bonds for these two systems (discussed below).

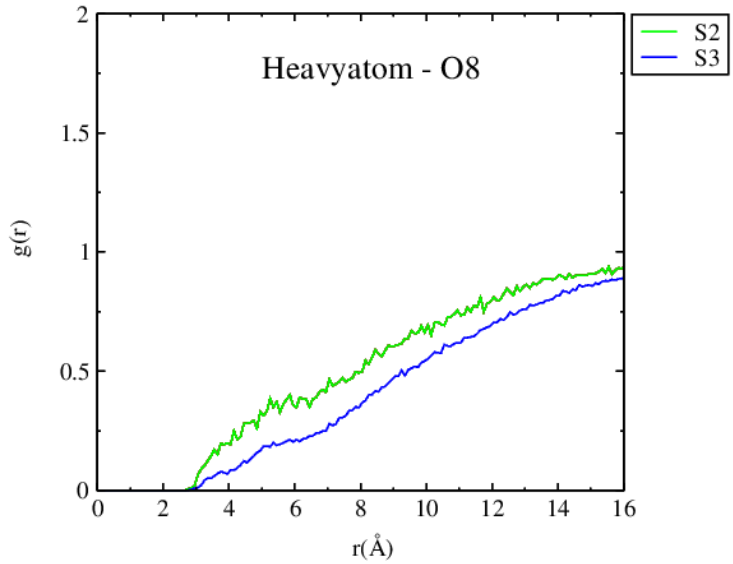

Figure 3A-9. *Site-site radial distribution function involving peptide heavy-atoms and O8 oxygen atom of choline-O-sulfate.*

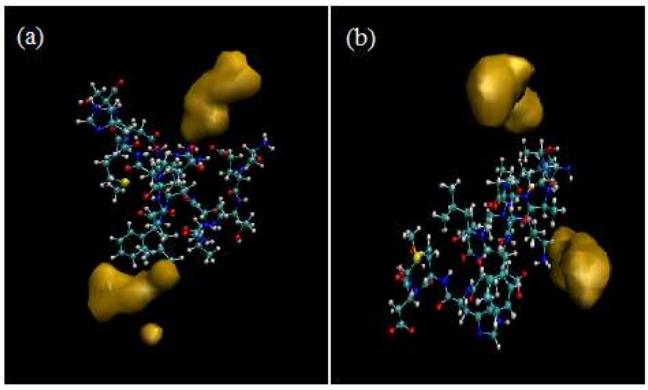

Figure 3A-10. *Contours of choline-O-sulfate density within 3.5 Å of the peptide surface for systems (a) S2 and (b) S3 using an iso-surface value of 0.5.*

Next, in order to probe the solvation pattern of the peptide, for different systems, the average numbers of water, urea and COS molecules around the heavy-atoms of 4-12 residues of the peptide are estimated by using equation 2.2 (see Table 3A-4).

For the calculation of the number of particle in the first solvation shell of a reference particle, a typical value of the position of the first minimum in the corresponding g(r) is considered as the cut off distance. The distribution functions involving the heavy-atoms of 4-12 residues of the peptide, oxygen atom of water, carbon atom of urea and O8 oxygen atom of COS are considered to determine the average numbers of water, urea and COS molecules in the peptide solvation shell. In Table 3A-4, we have also incorporated the normalized first shell coordination number of water and urea and for this, the normalization is carried out with respect to the number densities of water and urea of system S1 respectively. Note that, by doing this we are ruling out the effect of change in number density from system to system on coordination numbers. It can be noticed that the number of water molecules in the peptide solvation shell is 16.13 in pure water system. In binary peptide-urea system, this number decreases and we also find the presence of a large number of urea molecules in the peptide solvation shell. The addition of COS molecules to binary peptide-urea system decreases the number of urea molecules and the effect is very much profound for system S3. In specific, in 0.30 M COS solution there is a sharp drop of number of urea molecules from the first solvation shell of the peptide.

Table 3A-4. First shell solvation number[a]

System	water	urea	COS
S0	16.13	—	—
S1	13.21	33.0	—
S2	14.37 (14.56)	28.80 (29.13)	0.002
S3	10.67 (12.10)	25.09 (23.69)	0.001

[a] The average number of first shell water, urea, and choline-*O*-sulfate molecules around the peptide heavy-atoms of residues 4 to 12. The values in the parentheses are the normalized coordination numbers in which the normalization is carried out with respect to the number density of water and urea in system S1 for first shell coordination number of water and urea.

Now, concentrating on the number of COS molecules in the peptide solvation shell for systems S2 and S3, we observe that the average number of COS molecules in the peptide solvation shell is negligibly small which is already shown in the contour density map of COS around the peptide (see Figure 3A-10).

To examine if the peptide interacts preferentially with urea or COS over water, we have determined preferential interaction parameter. For this, following previously reported works [61, 170, 171], a two domain model is considered namely "local" domain and "bulk" domain. The first domain (i.e. "local" domain) is the region surrounding the peptide and the second domain ("bulk" domain) represents the region far away from the peptide. Then the preferential interaction of the peptide with urea over water ($\Gamma_{urea}(r)$), and the same for peptide-COS over water ($\Gamma_{cos}(r)$) can be calculated as:

$$\Gamma_{urea}(r) = N^{local}_{urea}(r) - \left(\frac{N^{bulk}_{urea}}{N^{bulk}_{water}}\right) \times N^{local}_{water}(r) \tag{3.1}$$

$$\Gamma_{cos}(r) = N^{local}_{cos}(r) - \left(\frac{N^{bulk}_{cos}}{N^{bulk}_{water}}\right) \times N^{local}_{water}(r) \tag{3.2}$$

In the above eq. 3.1, $N^{local}_{urea}(r)$ and N^{bulk}_{urea} represent the average local number of urea carbon present within the distance r from the heavy atoms of the peptide residues 4-12 and total number of urea molecules in a given system respectively. For calculating $N^{local}_{water}(r)$ and $N^{local}_{cos}(r)$ in equation 3.2, the distribution functions involving water oxygen- peptide heavy atoms of residues 4-12 and O8 oxygen atom COS and peptide heavy atoms of residues 4-12 are considered. It is also quite apparent from the above couple of equations that the positive Γ value represents the preferential accumulation of species urea or COS over water in the vicinity of the peptide surface. For different systems, the values of $\Gamma_{urea}(r)$ and $\Gamma_{cos}(r)$ are shown in Figure 3A-11. It is observed that the values of $\Gamma_{urea}(r)$ and $\Gamma_{cos}(r)$ are positive and negative respectively suggesting preferential accumulation of urea molecules and the exclusion of COS molecules from the peptide surface. Now, considering the effect of addition of COS on $\Gamma_{urea}(r)$ (see Figure 3A-11, part a), first, we notice that the value of $\Gamma_{urea}(r)$ decreases as the concentration of COS is increased implying the exclusion of some urea molecules from the peptide surface. On the other hand, a significant decrease in the $\Gamma_{cos}(r)$ value observed as COS concentration is increased from 0.1 to 0.3 M (see

Figure 3A-11, part b). In brief, as more number of COS is introduced in system S3, the interaction between COS and peptide backbone becomes more unfavorable, which causes exclusion of more number of COS molecules in the vicinity of the peptide, suggesting better involvement of COS molecules with urea over peptide backbone.

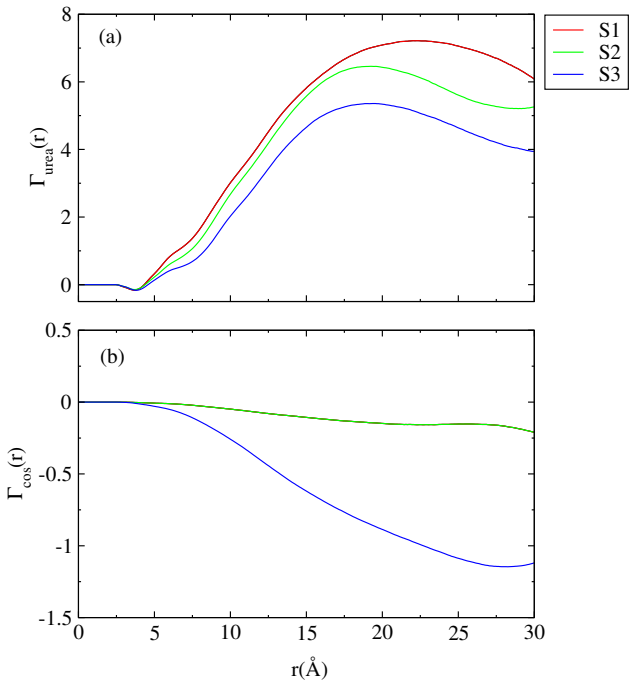

Figure 3A-11. *The preferential interaction parameters involving (a) peptide heavy-atoms and urea ($\Gamma_{urea}(r)$), and (b) peptide heavy-atoms and choline-O-sulfate ($\Gamma_{cos}(r)$) for different systems.*

Hydrogen Bond Properties

It is a well-known fact that the stabilization of the protein's native state is strongly influenced by the hydrogen bonding interactions between the protein and the other species present in the solution. The favorable hydrogen bonding interactions between the protein and other solution species can lead to a reduction in the intra-protein hydrogen bond. Thus, it is important to estimate the average number of peptide-peptide, peptide-water, peptide-urea, and peptide-COS hydrogen bonds in different systems. As used in previously reported studies [132, 133, 134, 135, 136, 137, 139], in this study also maximum D-A distance of 3.5 Å and simultaneously a maximum of D-H-A angle of 45° are used as geometric criteria of

hydrogen bond formation to determine the different hydrogen bond numbers. Here, D and A are the donor and acceptor atoms respectively.

For different systems, the average number of different hydrogen bonds are presented in Table 3A-5. The same table also contains the average number of water-water (per water), urea-urea (per urea), water-urea (per urea), water-COS (per COS), and urea-COS (per COS) hydrogen bonds for all the systems. For different systems, focusing on the average number of intra-peptide hydrogen bonds, first, we find that its value is 6.37 for the peptide in pure water (system S0). In system S1, the presence of urea alone causes a sharp drop in the intra-peptide hydrogen bond number to 4.23, which justifies the urea-conferred denaturation of the peptide. In the ternary peptide-urea-COS system with 0.10 M COS concentration (system S2), there is an increase in its value to 4.87 and interestingly, for 0.30 M COS solution (system S3), the average number of intra-peptide hydrogen bond reaches to 6.04 which is very close to that of the pure water system. The addition of 8 M urea causes a sharp decrease in the peptide-water hydrogen bond number from 22.21 in system S0 to 15.05 in system S1, and the presence of 0.10 M COS has a negligible effect on to it. In brief, in 0.10 M COS solution (system S2) peptide-urea hydrogen bond number remains essentially unchanged when we compare its value for system S1, and, for system S3 a moderate decrease in its value is observed. Further, the low peptide-COS hydrogen bond numbers for both systems S2 and S3 confirm that it does not play any significant role in the stabilization of peptide's native conformation. These observations are also in agreement with the calculations of preferential interaction parameters discussed above. Here, it is interesting to note that the total number of hydrogen bonds formed by the peptide (with itself as well as with other solution species) for systems S0 and S3 are essentially identical. A decrease in the average number of water-water (per water) hydrogen bonds is also observed. This decrease is attributed to the formation of large number of water-urea (per urea) and water-COS (per COS) hydrogen bonds. We also observe the presence of large number of urea-COS (per COS) hydrogen bonds in system S2 and its value increases from 2.69 in system S2 to 3.25 for system S3. Interestingly, the average number of urea-urea hydrogen bonds (per urea) is getting almost unaffected by the inclusion of COS molecules, already discussed in chapter 1 that the addition of COS into the system does not alter the urea structure. Thus, we can safely propose that COS interacts with urea molecules and replaces

some of them from the peptide surface to show its antidoting effect on urea-induced peptide denaturation, and the preferential solvation of urea molecules by COS molecules plays a significant role.

Table 3A-5. Hydrogen bond numbers[a]

System	S0	S1	S2	S3
HB_{PP}	6.37	4.23	4.87	6.04
HB_{PW}	22.21	15.05	15.82	13.79
HB_{PU}	—	15.23	11.95	7.55
HB_{PC}	—	—	0.11	0.17
HB_{total}	28.58	34.51	38.75	27.55
HB_{WW}	3.16	2.40	2.45	2.20
HB_{WU}	—	3.79	4.16	3.44
HB_{WC}	—	—	6.86	5.41
HB_{UC}	—	—	2.69	3.25
HB_{UU}	—	1.08	1.07	1.13

[a] Number of intra-peptide hydrogen bonds (HB_{PP}) and hydrogen bonds between peptide-water (HB_{PW}), peptide-urea (HB_{PU}), and peptide-COS (HB_{PC}). HB_{total} is the sum total of HB_{PP}, HB_{PW}, HB_{PU}, and HB_{PC}. HB_{WW} (per water), HB_{WU} (per urea), HB_{WC} (per COS), HB_{UC} (per COS), and HB_{UU} (per urea) are the average number of hydrogen bonds between water-water, water-urea, water-COS, urea-COS, and urea-urea, respectively.

This fact is further justified by the contour density map of COS around urea for systems S2 and S3 (see Figure 3A-12). It is observed that in system S2, a high density of COS molecules surrounds the oxygen and hydrogen atoms of urea and the density of COS increases significantly in the system containing 0.30 M COS (system S3). In this regard, it is worth mentioning that in the previous chapter (chapter 2) the study of synergistic behavior of urea and COS in aqueous solution shows that due to the enhancement in the hydrogen bonding between urea and COS, the former becomes less available to interact with the peptide backbone, hence, COS can protect the protein structure from the urea-induced denaturation of it [69].

Dynamics of the Solution Species

As hypothesized by the previous studies that some protectant osmolytes such as trehalose and TMAO slow down the translational dynamics of different species present in the solution, which in turn, indirectly helps the protein molecule to retain its native conformation [167, 168, 172], we have also determined the diffusion coefficients (D) of water, urea, and COS for different systems from the long time slope of mean square displacement (MSD) using Einstein's relation:

$$D = \frac{\lim_{t\to\infty} \langle (|r(t) - r(0)|^2 \rangle}{6t} \tag{3.3}$$

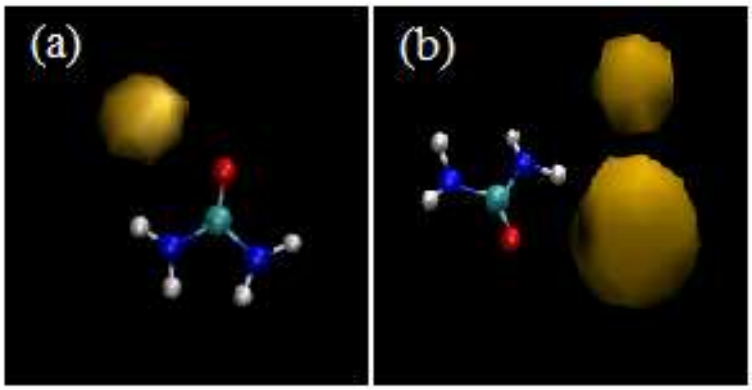

Figure 3A-12. *Contours of choline-O-sulfate density within 3.5 Å of urea for systems (a) S2 and (b) S3 using an iso-surface value of 0.5.*

The diffusion coefficient values of different solution components for different systems are listed in Table 3A-6. The diffusion coefficient value of water in pure water system is 5.27×10^{-5} cm^2 sec^{-1}, which decreases to 3.45 $\times10^{-5}$ cm^2 sec^{-1} in binary urea system. The addition of COS into binary urea system causes a further sharp reduction in its value to 1.02 $\times10^{-5}$ cm^2 sec^{-1} in system S3. A sharp drop in the diffusion coefficient value of urea molecules is also observed in presence of COS molecules. The formation of strong hydrogen bonding interaction between COS-urea and COS-water might be responsible for this

dynamical slow down of different solution species. In the parentheses of the same table (see Table 3A-6) we present the diffusion coefficient values of water, urea, and COS in absence of peptide at 310 K temperature. The diffusion coefficients of water, urea, and COS are calculated by performing a separate set of simulations in which the concentrations of urea and COS are maintained as that of the systems with the peptide molecule. As expected, it is observed that the presence of the peptide molecule slows down the translational motion of all solution species. In this context, it is worth mentioning that Mark et al. reported the diffusion coefficient value of the mTIP3P model of water is $5.93\times10^{-9}m^2s^{-1}$ at 298 K temperature without any co-solute [173].

Table 3A-6. Diffusion Coefficients[a]

System	D_W	D_U	D_{COS}
S0	5.27 (6.10)	—	—
S1	3.45 (4.10)	1.75 (2.62)	—
S2	2.45	1.41	2.00
S3	1.02 (2.61)	0.50 (1.31)	0.95 (1.41)

[a] Diffusion coefficients of water (D_W), urea (D_U), and COS (D_{COS}). In the parentheses the diffusion coefficient values of water, urea, and COS in absence of the peptide are presented. The values are expressed in $\times10^{-5}$cm^{2}sec^{-1} unit.

■ SUMMARY AND CONCLUSIONS

In this study, we have performed classical MD simulation on S-peptide analog to explore the antidoting effect of COS on urea-induced peptide denaturation. From the calculations of C_α-atom RMSD, radius of gyration, and SASA values, the deleterious effect of urea (in absence of COS) and the antidoting effect of 0.30 M COS solution in ternary peptide-urea-COS solution are clearly observed. These observations are further confirmed by the determination of peptide helicity percentage and visualization of the snapshots of the peptide at regular time intervals for different systems.

The total inter-particle energies are decomposed into van der Waals(vdW) and electrostatic energy components. It is found that urea-conferred peptide denaturation is

mainly governed by the direct urea-peptide electrostatic interactions. The contribution of peptide-urea vdW energy to the total peptide-urea interaction is also convincing. Both vdW and electrostatic peptide-COS interactions are negligibly small. We also find a highly favorable urea-COS interaction in 0.3 M COS-urea solution.

In order to gain much deeper insights into the effects of COS on urea-conferred peptide denaturation, at first, the structures of the different solutions are further probed by the calculations of selected site-site pair correlation functions involving peptide, water, urea, and COS. It is observed that in both peptide-water and peptide-urea hydrogen bonding interaction, it is the peptide backbone oxygen atom, which plays a vital role and this interaction decreases as COS is added. The exclusion of COS molecules from the peptide surface is also quite evident from the radial distribution functions involving peptide heavy atoms and O8 atom of COS and negative preferential interaction between peptide-COS over water. The calculations of the average number of water, urea, and COS molecules in the first solvation shell of the peptide as well as the determination of peptide-urea and peptide-COS volume density maps act as corroborative evidence to the findings of the peptide-water, peptide-urea, and peptide-COS distribution functions discussed above. The slowing down of translational motion of water and urea in presence of COS (due to favorable water-COS, water-urea, and urea-COS interactions) is further probed by the calculations of translational diffusion coefficients of these species. The density map of COS around a urea molecule suggests the preferential solvation of urea molecules by COS.

Thus, from the above discussions it can be safely proposed that in the urea-induced denaturation of the peptide, the direct peptide-urea interactions play a major role. The retention of peptide's native conformation in ternary peptide-urea-COS with 0.30 M COS concentration arises mainly due to: (i) exclusion of some urea molecules by COS from the peptide surface and (ii) favorable urea-COS interactions, which make urea molecules less available for the peptide to deliver its action.

Part B:

Choline-*O*-sulfate as a Protecting Osmolyte Against Urea-Induced Denaturation of Trp-cage Mini-protein

Overview

The present chapter discusses the folding-unfolding equilibrium of a small globular protein Trp-cage in presence of denaturing and protecting osmolytes urea and COS respectively, employing Replica Exchange Molecular Dynamics (REMD). Extensive trajectory analysis and free energy calculations show that while 6 M urea quite easily denatures the protein, 0.5 M, and 1 M COS is able to protect the protein from urea induced denaturation at room temperature. However, REMD simulations reveal that while the protein in pure water can withstand a temperature as high as 420 K without melting, the protecting effect of 0.5 M and 1 M COS is operative up to 300 K and 340 K, respectively. This study furnishes appropriate evidence to shed light into the protection mechanism of COS regarding urea induced protein unfolding, thereby putting forward the use of COS as a proper protecting osmolyte towards different genres of proteins.

■ INTRODUCTION

Protein dynamics is a physical process in which the polypeptide chain acquires its native 3-dimensional structure from a random coil conformation for different biological functions [174, 175, 176]. Misfolded protein causes inactive or toxic functionality. So, learning the protein folding-unfolding process has great biological significance to understand different biological activity in a reproducible and diligent manner [176, 177, 178]. The stability of the native state of a globular protein is restricted by different thermodynamic and environmental conditions [179, 180, 181, 182, 183]. The marginal stability between the folded state and the random conformation of a protein depends upon the balance between different driving forces like electrostatic and hydrophobic interaction, polymer entropy, temperature, and pressure, etc [184]. Chemical denaturants like urea, guanidine hydrochloride can perturb the folded structure of the protein at a concentration of 6-8 M [185]. The efficiency of urea as a protein denaturant is well known and has been applied to study the protein folding-unfolding process for years [186, 187, 188, 189] . Urea can directly exert its deleterious effects by binding with the protein backbone or can indirectly unfold the protein by solvation of hydrophobic groups in terms of altering the water structure. Anyways, the unfolding of a protein structure does not follow any particular steps, rather it perceives some random pathways depending upon its amino acid sequences [177, 178].

To study a complex biological system through molecular dynamics simulation, investigation of global conformational transition is more important than the local fluctuations of the molecule. But literature survey reveals that in case of conventional MD simulation, due to the kinetic trapping, the molecule of interest spends maximum time in the lower energy minima producing highly rugged free energy landscapes due to insufficient sampling. Also for protein simulation, due to the lack of convergence of the systems, ordinary MD simulation can not produce the reliable estimates of experimental quantities [190, 191, 192]. So, in order to elucidate the thermodynamics and kinetics of protein, a thorough sampling of the conformational space of the system is required. A plethora of enhanced sampling methods have been evolved to deal with this multi-minima problem among which the replica exchange method is a well known approach [94, 193, 194, 195].

Since from the birth of Replica Exchange Molecular Dynamics (REMD) in 1999 by

Sugita and Okamoto [94], it has been a very popular approach to carry out biomolecular simulations. We have considered REMD simulation technique over the regular canonical molecular dynamics (MD) simulation because of the fact that for the later technique it is usually difficult to obtain accurate canonical distributions both at low- and room temperatures [94], as simulations at such temperatures tend to get trapped in one of the huge numbers of local or minimum energy states. REMD overcomes the energy barriers by attempting state exchange moves. One replica at a higher temperature exchanges its temperature with the immediate lower temperature replica and vice versa. This exchange facilitates barrier crossings of the energy landscape and helps to create the suited trajectories for configurational transitions. This in turn helps to trace out the trajectory of conformational transitions followed by the protein under consideration. REMD ensures efficient sampling by crossing the energetic barriers. This method has numerous other advantages too. It produces information over a range of temperatures (so that we get the melting temperature curve of the protein), it works equally on explicit and implicit solvents. In comparison to constant temperature MD (of protein simulation), it reduces the sampling time by a factor of 20 or so.

Here we have studied the folding-unfolding dynamics of a globular mini-protein named Trp-cage, through replica exchange molecular dynamics simulation. Although Trp-cage is small-sized protein, it acquires tertiary structure like a large globular protein within 4 μs, having a hydrophobic core with a central Trp residue, multiple secondary domains like N-terminal α-helix, 3_{10}-helix, and a C-Terminal polyproline II segment and salt-bridge between residues 9 and 16 which make it popular as a model protein for computational studies (see Figure 3B-1) [196, 197]. Also Trp-cage holds the thermodynamic features like the temperature-dependent unfolding free energy and enthalpy similar to that of a globular protein [198]. Besides the high temperature [198, 199, 200] pressure [201], and cold denaturation [197] of TRP-cage, unfolding of this protein by chemical denaturant urea is also well known both experimentally and theoretically [59, 139, 196, 200, 202]. Denatured proteins can exhibit a wide range of conformational changes leading to functional inactivity which is a consequence of cell death [203]. Protecting osmolytes like trimethylamine N-oxide (TMAO), betaine, glycerol, and sugars (trehalose) play a major role to maintain the function of intracellular proteins by retaining their native structure in extreme envi-

ronmental conditions [67, 139, 204, 205]. In part A of this chapter, we have found COS as a good osmolyte against urea-induced denaturation of a cellular protein called S-peptide at a very low concentration. Now in this part, we tried to make another approach of COS as a protecting osmolyte against urea-induced denaturation of globular protein TRP-cage, employing Replica Exchange Molecular Dynamics simulation.

The rest of this part is divided into three sections describing the Models and Simulation Method, Results and Discussion, and finally the Summary and Conclusions of this work.

■ MODELS AND SIMULATION METHOD

We have chosen the 20 amino acid residue mini-protein Trp-cage (PDB Id : 1L2Y) with amino acid sequence NLYIQWLKDGGPSSGRPPPS for this study (see Figure 3B-1) [206]. To carry out the classical MD simulations, four systems are prepared namely S0, S1, S2, and S3 (see Table 3B-1). In system S0, a single protein molecule was immersed into water to study the behavior of the protein without any co-solute considering it as the reference system to study the behavior of the protein in presence of urea and COS. In system S1, the protein was solvated in $\sim$ 6 M urea to study its conformational changes in urea solution and S2, S3 are the systems with protein in urea-COS mixture where the concentration of COS is varying from 0.5 M to 1.0 M. It is worth mentioning here that, we have started with COS concentration of 0.1 M, following the previous study and keep increasing the concentration till we get a good result. The protein was solvated using TIP3P water in a cubic box [111, 173]. All the systems were prepared using PACKMOL [116]. The protein carries one positive charge, so one chloride ion was added in each system to neutralize.

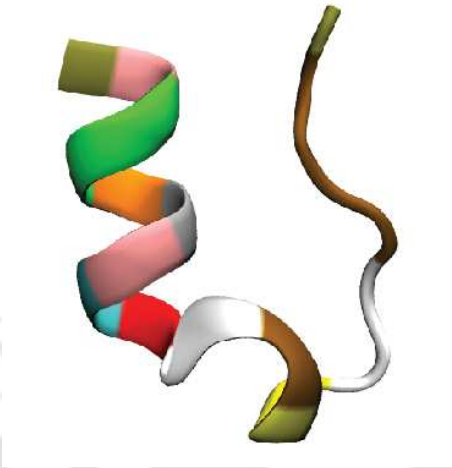

Figure 3B-1. *Initial structure of Trp-cage mini-protein.*

Table 3B-1. Overview of Systems[a]

System	N_P	N_U	N_C	N_W	Box length (Å)	M_U	M_C
S0	1	0	0	2000	39.87	0	0
S1	1	300	0	2000	43.89	5.90	0
S2	1	500	40	3000	51.51	6.07	0.5
S3	1	500	100	3000	52.86	5.62	1.1

[a] N_P, N_U, N_C, and N_W represent the number of peptide, urea, choline-*O*-sulfate, and water molecules respectively. M_U and M_C are the molar concentrations of urea and choline-*O*-sulfate of the respective systems.

Following the earlier works on TRP-cage, we decided to use AMBER force field parameters for our simulation [59, 139]. To model the protein, ff14SB force field parameters were used [207] and Smith model of urea was employed to denature the protein [112, 113, 114, 115]. For COS molecule General AMBER force field (GAFF) parameters were used, presented in Table 2-3 of chapter 2 [120]. In the beginning of the simulation, each of the MD systems was minimized for 10000 steps by equally distributed steepest descent method and conjugate gradient method. Then, the system was heated gradually with a regular interval of 50 K using canonical ensemble (NVT) [123]. After this, each of

the systems was equilibrated for 4 ns in isothermal-isobaric (NPT) ensemble at a constant pressure of 1 atm and temperature of 300 K [124, 126]. This results in a system of adequate density with corresponding edge lengths. Again the systems were subjected to an equilibration at constant volume for 1 ns. The coordinates obtained from this equilibration was used for final replica exchange molecular dynamics (REMD) simulation run [94]. For all REMD simulations, 12 replicas for system S0 and 14 replicas for rest of the systems were distributed exponentially in the temperature range from 290 K to 420 K. The temperatures of the corresponding replicas for systems S0 are 290.0, 300.31, 310.89, 321.73, 332.86, 344.28, 356.01, 368.04, 380.37, 393.04, 407.04 and 420.0 K and for rest of the systems containing 14 replicas, the corresponding temperatures are 290.0, 299.27, 308.75, 318.46, 328.39, 338.56, 348.95, 359.61, 370.51, 381.67, 393.10, 404.80, 416.78 and 420.0 K. Experimental studies reported that the melting temperature of Trp-cage is 315 K [208, 209, 210], whereas in computational studies depending on the force field parameters used, the melting temperature of this mini-protein varies from 360 to 450 K [139, 200, 211, 212]. Initially, all the replicas are equilibrated for 5 ns at their desired temperatures in NVT ensembles as the initial parts of REMD trajectories must be excluded from the thermodynamic analysis. Then each of the simulations was continued up to 200 ns as the previously reported simulations studies on urea-induced denaturation of TRP-cage say that this much long simulation is fair enough for the convergence of the REMD systems containing this protein [59, 139, 200]. Exchanges were attempted within every 4 ps between all neighboring replicas with the average acceptance ratio of 20%. The acceptance ratio of each replica exchange between two adjacent states i and j is governed by the Metropolis equation (eq. 1.18, discussed in the methodology section of Chapter 1).

For REMD simulation, the coordinates were saved after every 2000 integration steps. Langevin thermostat was used for temperature control with a collision frequency of 1 ps^{-1} and all bonds involving hydrogen atoms were constrained using SHAKE algorithm [127, 129]. Particle Mesh Ewald (PME) summation method was employed for long-range electrostatic interactions [128], and a non-bonded cut-off distance of 9 Å was used. Periodic boundary conditions were employed in all three directions to reduce the surface effects. All the simulations were carried out using AMBER14 package [122], and the temperature based trajectories are analyzed using VMD, and CPPTRAJ available in AMBER [130, 131].

■ RESULTS AND DISCUSSION

Structural and Conformational Analysis

First of all, we have tried to analyze different types of structural and conformational changes of the peptide in different solution environments. To quantify the structural deviation of the peptide, we have computed the probability distribution of the radius of gyration (R_g) of the peptide backbone at both low and high temperatures of the corresponding systems (see Figure 3B-2). We can see, in pure water (system S0) the probability distribution of the radius of gyration (R_g) of the C_α-atoms of the peptide backbone is maximum at around 7.0 Å followed by the system containing 1.0 M COS solution (system S3), indicating the most compact structure of the peptide in both of these systems whereas in presence of urea and 0.5 M COS concentration (systems S1 and S2), the peak height around 7 Å is reduced significantly and broadened up to 15 Å due to the loss of compactness of the peptide structure in these systems. It is noticeable that in pure water system at higher temperature the structural rigidity of the peptide is somehow lesser than that of lower temperature and also at higher temperature like 420 K, 1.0 M COS solution is not so effective to preserve the native structure of the peptide in presence of urea.

To depict the structural changes of the peptide backbone, next, we have calculated the percentage of helicity of the protein at different temperatures corresponding to each replica (see Figure 3B-3) using Dictionary of Secondary Structure of Proteins (DSSP) algorithm [213]. The temperature-wise scanning of helicity percentage shows that in pure water system the protein acquires maximum of helical content, and increase in temperature does not show much effect on it. In presence of urea at 290 K, there is a sharp loss in the helicity percentage (from 51.13% to 18.97%). With increasing temperature, this loss is enhanced further, and finally, above 330 K the protein completely loses its helical structure. Various simulation studies reported that the melting temperature of Trp-cage protein varies between 360 to 450 K depending upon the force field parameters used in the simulation [139, 200, 211, 212].

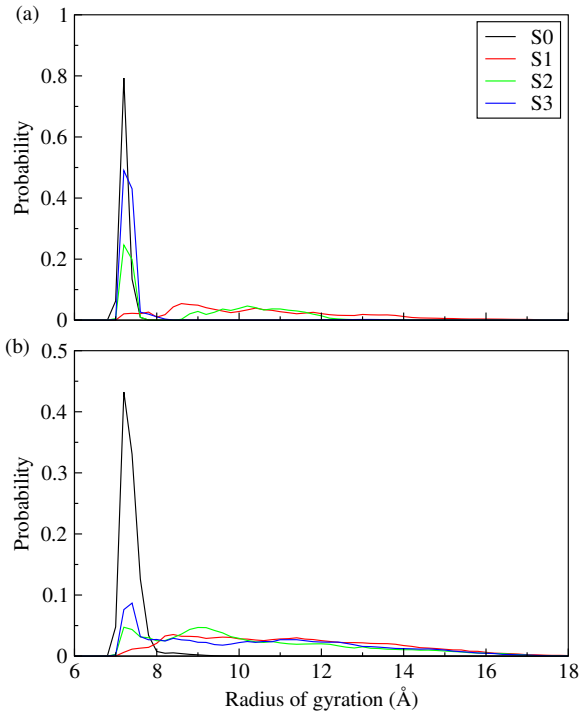

Figure 3B-2. *Probability distribution of the radius of gyration (R_g) of the C_α-atoms of the peptide backbone at (a) 300 K (b) 420 K.*

In this study, we have found that the protein is stable up to 420 K in pure water system. But in presence of co-solutes like urea and COS, the melting temperature of the protein gets altered [139, 212]. From Figure 3B-3, it can be observed that 0.5 M COS solution is able to retain the helical structure of the protein at 290 K. Above this temperature, it starts to lose its helical conformation leading to the complete disruption of the secondary structure around 400 K, whereas, in presence of 1.0 M COS concentration, the native structure is protected up to 340 K. So, it is clear that higher temperature accelerates the urea-induced denaturation process on Trp-cage protein while at low temperature region, COS can diminish this deleterious effect of urea at a minimal concentration.

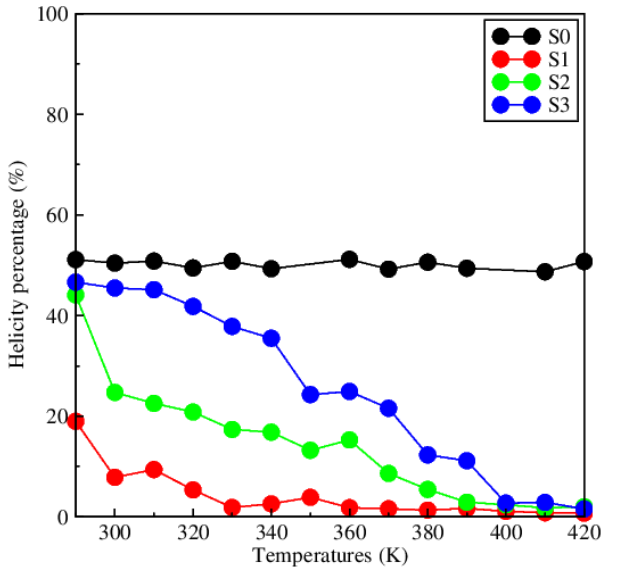

Figure 3B-3. *Temperature-wise helicity percentage for different systems.*

Role of Different Stabilizing Parameters

To find out the effects of different osmolytes on the protein folding-unfolding behavior, now we have focused on the dynamics of the three stabilizing parameters of Trp-cage that are hydrophobic core [214, 215, 216], salt-bridge distance [212, 216], and orientational preferences of the indole planes in different environments [114]. First, we discuss the effect of the hydrophobic core formed by the residues Tyr3, Leu7, Gly11, Pro12, Pro18, and Pro19 surrounding the central residue Trp6, both at low and high temperatures. From Figure 3B-4, it is apparent that for the protein in pure water the distance between the hydrophobic core and the central Trp6 residue of the protein is well maintained within 3 Å for both high and low temperatures but with the addition of urea into the system, the compact arrangement of the hydrophobic core gets disrupted. In presence of 0.5 M COS, at lower temperature, the distance is intact to some extent and in 1.0 M COS solution, the arrangement of the hydrophobic core is similar to that of the pure water system.

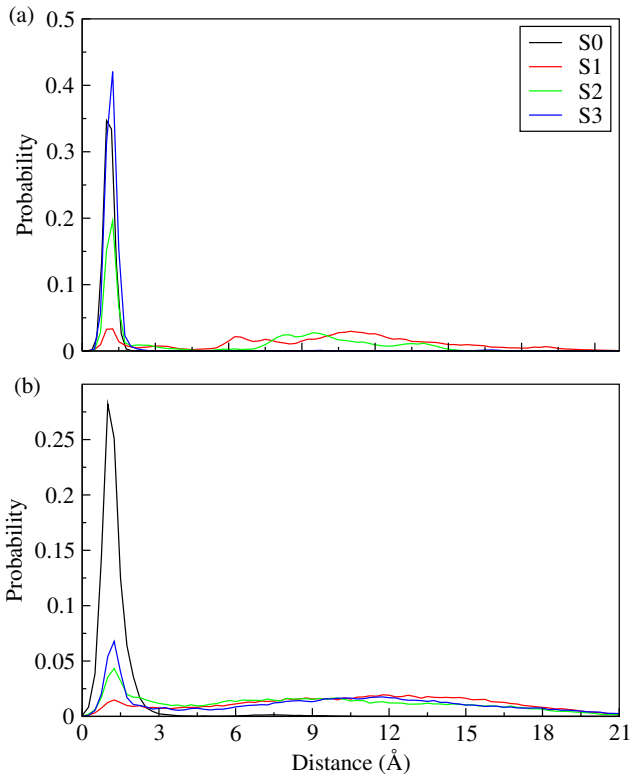

Figure 3B-4. *Probability distribution of the distance between the hydrophobic core and the central Trp6 residue at (a) 300 K and (b) 420 K.*

As mentioned above the presence of salt bridge between residues Asp9 and Arg16 is another stabilizing factor of Trp-cage. For different systems the probability of salt bridge formation between these two residues at low and high temperatures is shown in Figure 3B-5. For the protein in pure water, the salt bridge is beautifully preserved but the addition of 6 M urea into the system completely destroys it which accelerates the unwinding of the folded structure in presence of urea, however, the protein can retain its salt bridge structure again in presence of COS. It is noteworthy that although at high temperature the helical structure of the protein is well maintained in pure water (discussed above), the probability of salt bridge formation decreases (see Figure 3B-5, part b), suggesting that existence of salt bridge between residues Asp9 and Arg16 is not a salient feature for the stability of this protein.

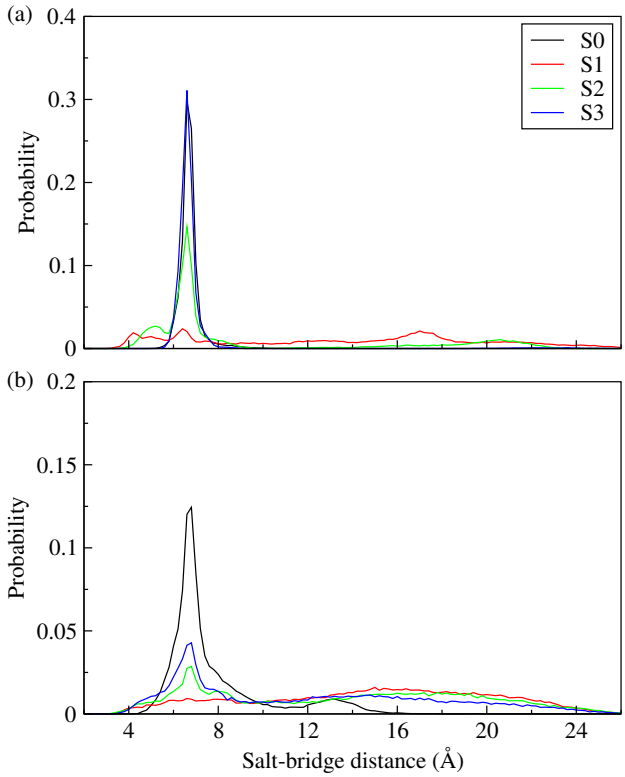

Figure 3B-5. *Probability distributions of the salt bridge distances for different systems at (a) 300 K and (b) 420 K.*

To calculate the orientational preference of the indole plane of residue Trp6 with other ring planes of residues Tyr3, Pro12, Pro17, Pro18, and Pro19 at 300 K temperature, we have considered the angle between two vector normals of the corresponding planes. P(θ) is the probability distribution of the inter-plane orientational angle varying from 0° to 180° (see right-hand panel of Figure 3B-6). An angle 0° or 180° defines parallel orientation whereas angle 90° denotes the perpendicular orientation between two planes. In the same figure, also we have incorporated the probability distribution of the center of mass (COM) distance between two planes of the above mentioned residues (see left-hand panel of Figure 3B-6). From Figure 3B-6, it can be observed that in pure water as well as COS containing system, the COM distance between the aromatic planes of residues Trp6 and Tyr3 is 6 Å but in protein-urea binary solution the distance fluctuates up to 15 Å. Also, the orientational

angle between these two planes is around 30° in pure water and in 1.0 M COS solution, indicating a tilted orientation between them, while in system S1 and S2 the preferential orientational angle varies up to 150° confirming a broad distribution between them. For the protein in pure water, the 5-membered rings of residues Pro12 and Pro18 remain within 5 Å of the indole plane of Trp6, and angles between them is within 15° which clearly states the parallel orientation between these planes. Presence of 6 M urea violates these arrangement denoting the unfolding of the structure, whereas a higher concentration of COS helps to retain the preferred orientation of these residues like the pure water system. The 5-membered rings of both residues Pro17 and Pro19 prefer to be in a perpendicular orientation with respect to the central Trp6 residue in absence of any co-solute, and in presence of 1.0 M COS, but in presence of only urea or low concentration of COS, it goes from perpendicular to parallel arrangement leading to an increased center of mass distance between them. It is worth mentioning that here the plane of a ring is defined as the area connecting the atoms of the ring. It is true that ring of a free proline is puckered but it is also a fact that, in proteins, the ring of a proline residue exists predominantly in two discrete states, up puckered and down puckered which is significantly flexible and coupled with backbone conformation, and this strain is relaxed to allow the proline ring to flatten it in the rare planar conformation [217, 218].

Interaction Energy

In order to decipher the energetics of different types of interactions present, we calculated the interaction energies between the solution species at room temperature for different systems (see Figure 3B-7). All types of interaction energies are decomposed into vdW and electrostatic energy components. It is noticed that the vdW, as well as electrostatic interaction energy between the protein and urea molecules, is the highest in system S1 (see Figure 3B-7, part a). On the other hand, the peptide-COS interaction (see Figure 3B-7, part b) is favorable but weaker than the peptide-urea interaction, and it becomes more favorable with increasing amount of COS into the systems. Also, calculation of the interaction energies between the co-solute particles present in the systems (other than peptide) that is urea and COS (per COS molecules, see Figure 3B-7, part c) reveals that COS molecules interact favorably with the urea molecules. It is important to point out that the

total urea-COS interaction energy in a system is quite higher than the total peptide-COS interaction energy, suggesting that indirect solvation of urea molecules by COS plays the major role to reduce the denaturing effects of urea on Trp-cage mini-protein. It is to be noted that between electrostatic and vdW energy components, the former contributes much higher in total interaction energy and it is true for all interactions considered here.

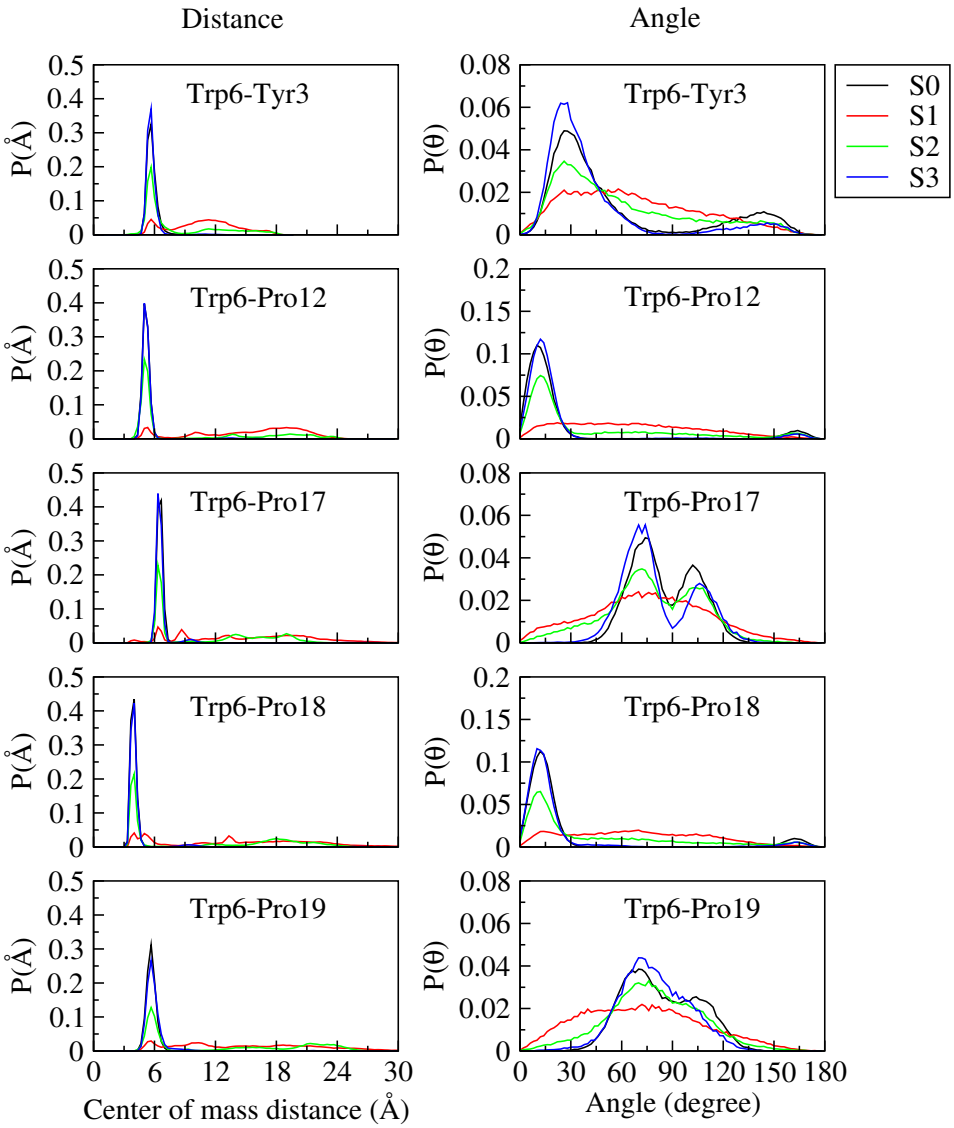

Figure 3B-6. *Probability distributions of COM distances (left panel) and different orientational angles (right panel) of indole plane with other ring planes at 300 K temperature.*

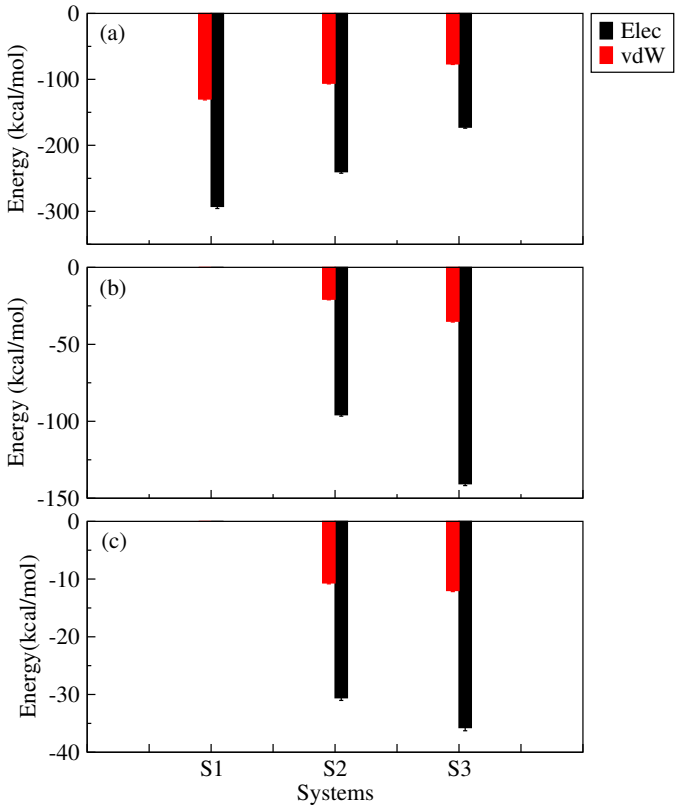

Figure 3B-7. *vdW and electrostatic interaction energies (in kcal/mol) between (a) protein-urea, (b) protein-COS, and (c) urea-COS (per COS molecule).*

Hydrogen Bonding Interaction and Hydrophobic Solvation

Next, we have calculated the radial distribution function (RDF) between different solution species (see Figure 3B-8). The pair correlation function between the peptide heavy-atoms and the urea carbon atoms in system S1 shows the highest peak around 5 Å which decreases with the addition of COS into the system suggesting that presence of COS molecules into the systems helps the peptide to remain less affected by the urea molecules. Part b of Figure 3B-8 presents the probability of finding sulfate oxygen of COS around the peptide heavy-atoms, and it can be observed that the first peak of the plot is not so much pronounced like the RDF of peptide-urea, indicating comparatively weak interaction between peptide and COS molecules into the systems. Also, we have calculated the pair

correlation function between the urea nitrogen and sulfate oxygen of COS, and found a sharp peak around 3 Å suggesting the presence of hydrogen bonding interaction between them due to which the urea molecules get solvated by the COS molecules and become less available for the peptide backbone.

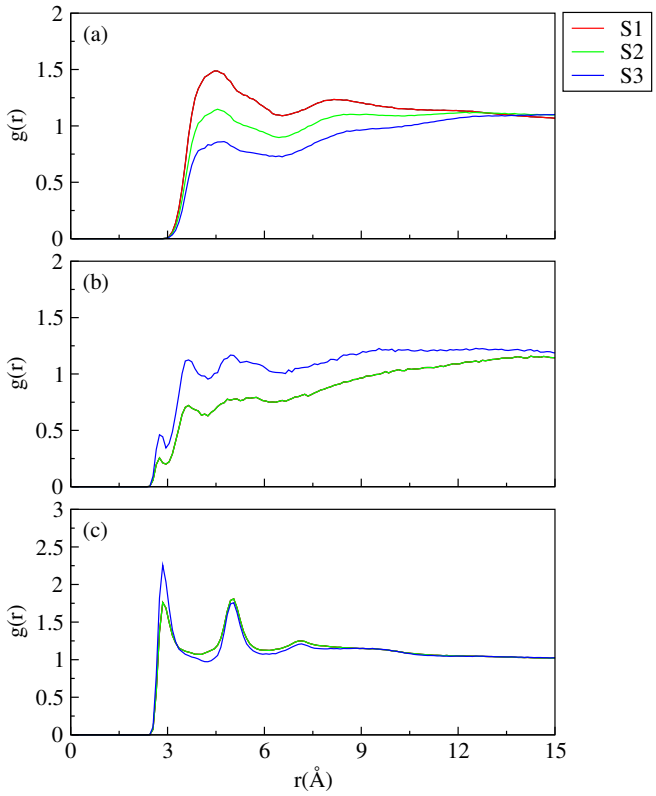

Figure 3B-8. *Pair correlation functions between (a) peptide heavy-atoms and urea-carbon, (b) peptide heavy-atoms and sulfate-oxygen of COS, and (c) urea-nitrogen and sulfate-oxygen of COS at 300 K.*

To study the folding-unfolding behavior of a protein, the study of hydrogen bonding properties is necessary. Considering the presence of highly favorable electrostatic interaction energy between the solution species, we have calculated the average hydrogen bond numbers between different solute particles present in the systems at low (300 K) and high (420 K) temperatures (see Table 3B-2). Following earlier works, a cut-off distance of D-A $\leqslant$ 3.5 Å and $\angle$D-H-A $\leqslant$ 45° is set as the criteria of the H-bond calculation for the present study (where 'D' and 'A' are the donor and acceptor atoms respectively) [139]. From Ta-

ble 3B-2 we observe that for the protein in pure water, the intra-protein hydrogen bond number is maximum and its value is slightly higher at low temperature than that of high temperature. In presence of co-solute urea, the intra-protein hydrogen bond number decreases significantly indicating a loss of helical structure of Trp-cage in this system. The addition of 0.5 M COS can prevent this rupture of secondary structure to some extent at low temperature, but at high temperature, the loss of intra-peptide hydrogen bond is as prominent as that of system S1. Further increase in COS concentration helps the peptide to preserve its native conformation at low temperature by retaining the same number of intra-peptide hydrogen bonds, but at high temperature (420 K), the loss of intra-protein H-bond number indicates the ineffectiveness of COS in the protection of the conformation of the mini protein against simultaneous chemical and temperature-induced denaturation. Next, we consider the hydrogen bonding interaction of the protein with the solvent and other co-solutes. In presence of 6 M urea, the protein-water hydrogen bonding interaction becomes weaker due to the presence of a rich amount of protein-urea hydrogen bonds which suggests the better solvation of protein molecule in urea over water. COS can diminish the protein-urea hydrogen bonding interaction by the formation of hydrogen bonds with protein as well as urea due to which the urea molecules get excluded from the protein surface at a higher concentration of COS. It is noticeable that the protein-COS hydrogen bonding interaction is weaker in comparison to that of urea-COS hydrogen bonding interaction suggesting that COS prevents the protein denaturation mainly through the indirect mechanism by solvating the urea molecules. On contrary, at high temperature, the intra-protein as well as all other types of hydrogen bonding interaction becomes weaker leading to the denaturation of the peptide even in presence of high COS concentration.

Next, we have focused on the solvation of protein in different systems at low and high temperatures to obtain a clear view about the role of different solvent and co-solutes on protein folding-unfolding process (see Table 3B-3). To calculate the coordination number of a particle in the first solvation shell (FSS), we have calculated the number of molecules present within 3.5 Å of the heavy-atoms of another molecule [139]. The hydration of protein is maximum in pure water both at high and low temperatures which strongly supports the solvent-assisted stability of the protein in water. In presence of urea, the first shell hydration number of protein decreases due to the replacement of water molecules by urea

from the vicinity of the protein backbone, leading to the unfolding of the peptide helical structure. If we introduce a small amount of COS (0.5 M) into the system, the number of urea molecules in the first solvation shell slightly goes down due to the involvement of COS with protein as well as urea, but the amount is not enough to help in the retention of protein conformation against the urea induced denaturation. Further increasing the concentration of COS, at low temperature due to the increased solvation of the protein as well as urea by COS molecules, the former becomes less available for the protein backbone, henceforth, the protein remains intact in its native state. At high temperature, the number of solvent molecules around the protein backbone becomes slightly higher due to the more exposure of the unfolded protein towards the solvent molecules. Here we can see, at a lower temperature, COS can prevent the deleterious effect of urea on protein by directly interacting with the protein backbone as well as indirectly solvating the urea molecules by itself among which the indirect mechanism plays the most crucial role here.

Table 3B-2. Hydrogen bond numbers[a]

		S0	S1	S2	S3
HB_{PP}	300 K	13.28	3.74	7.55	11.59
	420 K	11.51	3.30	4.64	5.04
HB_{PW}	300 K	29.70	23.13	23.91	25.44
	420 K	27.31	22.50	22.95	24.82
HB_{PU}	300 K		15.63	14.71	11.37
	420 K		9.55	10.05	10.05
HB_{PC}	300 K			1.45	2.25
	420 K			0.89	2.07
HB_{UC}	300 K			1.98	1.83
	420 K			1.47	1.38
HB_{WW}	300 K	3.16	2.61	2.58	2.53
	420 K	2.69	2.27	2.23	2.17

[a] Average hydrogen bond numbers of intra-protein, protein-water, protein-urea, protein-COS, urea-COS (per COS), and water-water (per water) at 300 K and 420 K. P, W, U, and C represent protein, water, urea, and COS, respectively.

Table 3B-3. First shell solvation number[a]

		S0	S1	S2	S3
$N_{W(O)}$	300 K	43.60	33.64	33.07	34.50
	420 K	46.07	38.04	37.97	40.81
$N_{U(C)}$	300 K	–	3.64	3.40	2.68
	420 K	–	2.63	2.78	2.76
$N_{U(O)}$	300 K	–	9.68	8.90	7.52
	420 K	–	6.95	7.23	7.08
$N_{U(N)}$	300 K	–	17.05	16.40	14.14
	420 K	–	12.25	13.12	13.26
$N_{COS(C)}$	300 K	–	–	1.28	3.05
	420 K	–	–	1.28	3.31
$N_{COS(O)}$	300 K	–	–	2.82	5.20
	420 K	–	–	1.95	4.72
$N_{COS(N)}$	300 K	–	–	0.02	0.04
	420 K	–	–	0.02	0.06
$N_{ure-cos}$	300 K	–	–	4.96	6.98
	420 K	–	–	4.89	6.09

[a] First shell coordination number of water-oxygen ($N_{W(O)}$), urea-carbon ($N_{U(C)}$), urea-oxygen ($N_{U(O)}$), urea-nitrogen ($N_{U(N)}$), COS-carbon ($N_{COS(C)}$), COS-oxygen ($N_{COS(O)}$), COS-nitrogen ($N_{COS(N)}$) around the heavy-atoms of the protein, and $N_{ure-cos}$ is the coordination number of COS around a urea molecule.

Also, we have measured the spatial density function (SDF) of the urea and COS molecules around the heavy-atom of the protein at room temperature by using "VolMap" plugin implemented in VMD, using an iso-surface value of 0.5 for all the systems at 300 K [130]. From Figure 3B-9, it is clearly visible that the spatial density of urea around protein decreases with increasing COS concentration into the system, though the change in density of COS around the protein is not so much prominent. It is worth mentioning here that in system S1, the density of urea molecules surrounding the whole protein are very high. In system S2, though the density of the urea molecules get reduced from the protein backbone, the urea molecules are mostly concentrated at the N-terminal part of the hydrophobic core

of the peptide (mainly residues Trp6 to Tyr3) leading to its disruption. Also, we have analyzed the spatial density of COS around a urea molecule, and noted that the density of COS around urea nitrogen atoms is significantly higher in 1.0 M COS solution due to which the urea molecules get blocked by COS.

Free Energy Landscapes

For detailed analysis, next, we have shown the free energy surfaces of different systems where free energy change (ΔG) is plotted against the root mean square deviation (RMSD) of the C_α-atoms of the peptide backbone at different temperatures (see Figure 3B-10) [114, 115, 139, 212]. The free energies are calculated using the following equation [139]:

$$\Delta G(V) = -k_B T[lnP(V) - lnP_{max}] \tag{3.4}$$

where P(V) is the probability distribution of the RMSD of C_α-atoms of the peptide backbone at coordinate V obtained from each trajectory of the corresponding simulation, and P_{max} is the maximum probability found. From the free energy landscape plot, it is observed that the pure water system contains a continuous folded basin within RMSD $\leqslant$ 2.0 Å which is stable up to 420 K [202]. In system S1, the low energy basin shifted towards higher RMSD value which goes up to 8 Å indicating a complete disruption of the protein folded state. In presence of 0.5 M COS, a small low energy basin is found within the temperature range 290 to 320 K, but with increasing concentration of COS, a significant amount of folded-basin is obtained which is stable up to 340 K, and then it starts to spread towards higher RMSD value revealing the unfolding of the protein at high temperature region even in the presence of COS as a counteracting osmolyte.

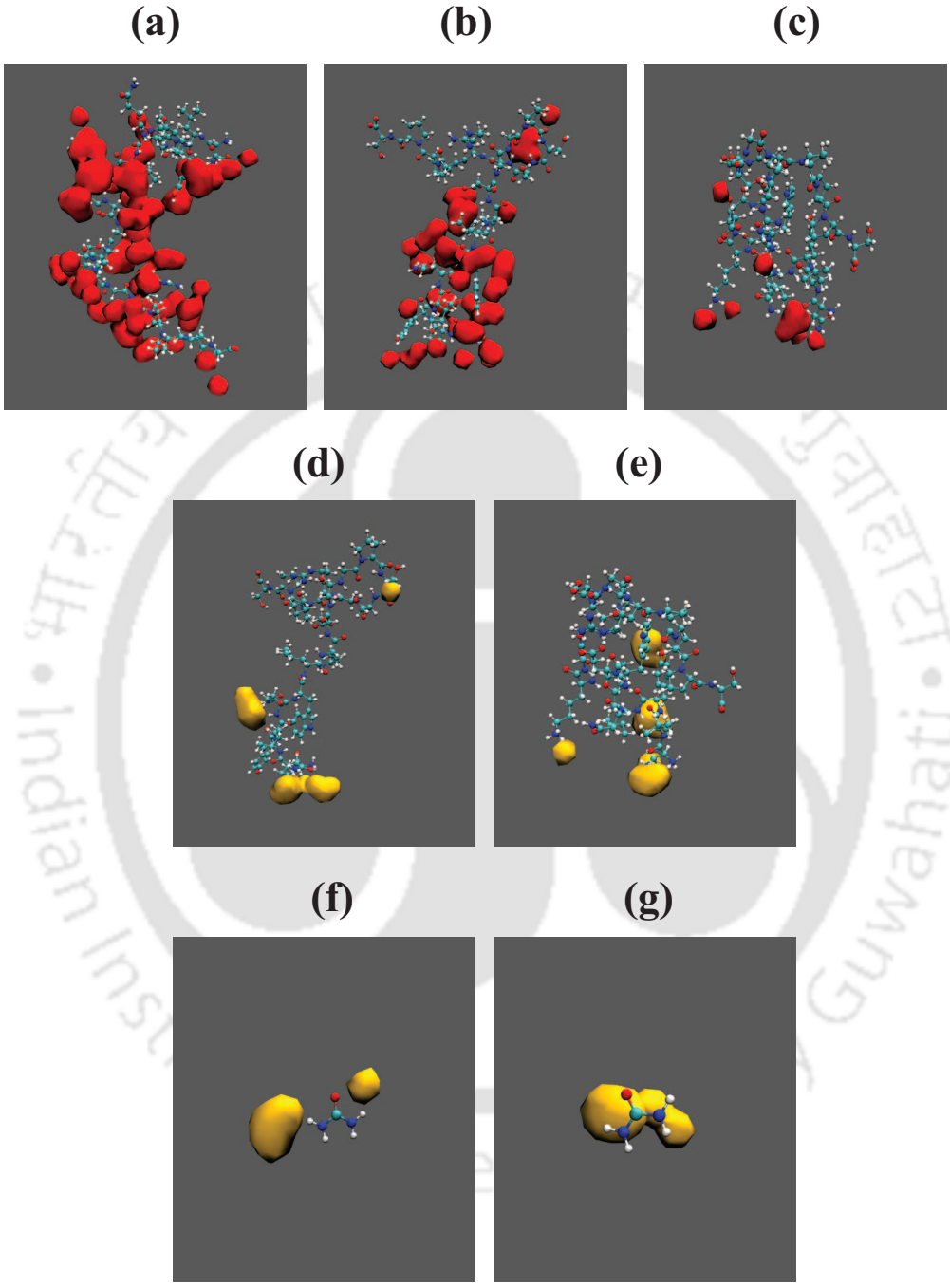

Figure 3B-9. *(a), (b), and (c) are SDF plots of urea (red) around protein for systems S1, S2, and S3, respectively; (d), (e) are SDF plots of COS (yellow) around protein for systems S2, S3, respectively; and (f), (g) are SDF plots of COS (yellow) around a urea molecule for systems S2, S3, respectively at 300 K.*

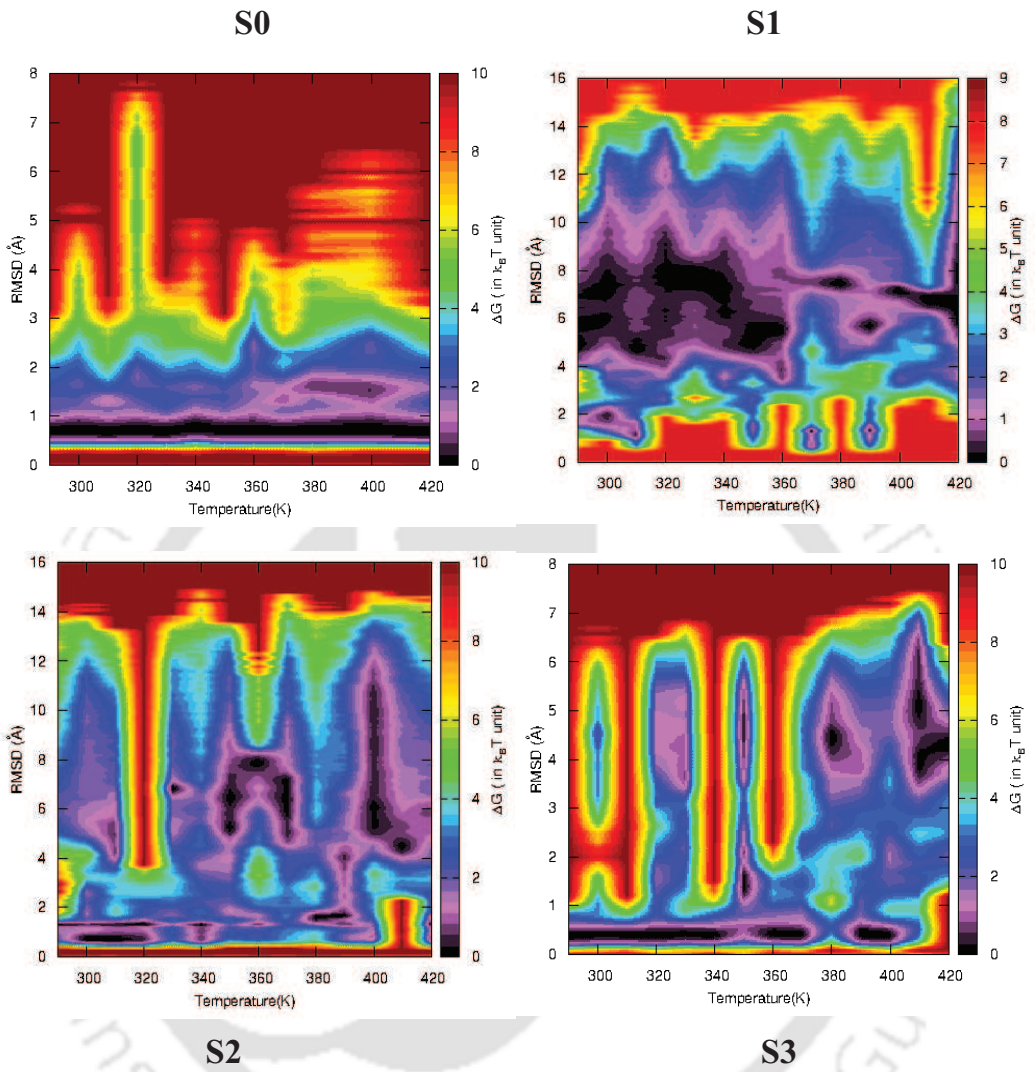

Figure 3B-10. *Free energy landscape plot as a function of RMSD and temperature for different systems.*

Next, we have shown another free energy surface plot obtained from the dihedral principal component analysis (dPCA) (see Figure 3B-11) [114, 115, 139, 165, 219]. For principal component analysis (PCA) of a system containing M atoms, the correlated internal motions are represented by the co-variance matrix [219]:

$$\sigma_{ij} = \langle (x_i - \langle x_i \rangle)(x_j - \langle x_j \rangle) \rangle \tag{3.5}$$

where x_i,....,x_{3M} are the mass-weighted Cartesian coordinates of the N-particle system, and

⟨...⟩ denotes the average over all sampled conformations. The PCA analysis based on the sin and cos transformed components of dihedral angles is performed using the 'Carma: a MD analysis program' [220].

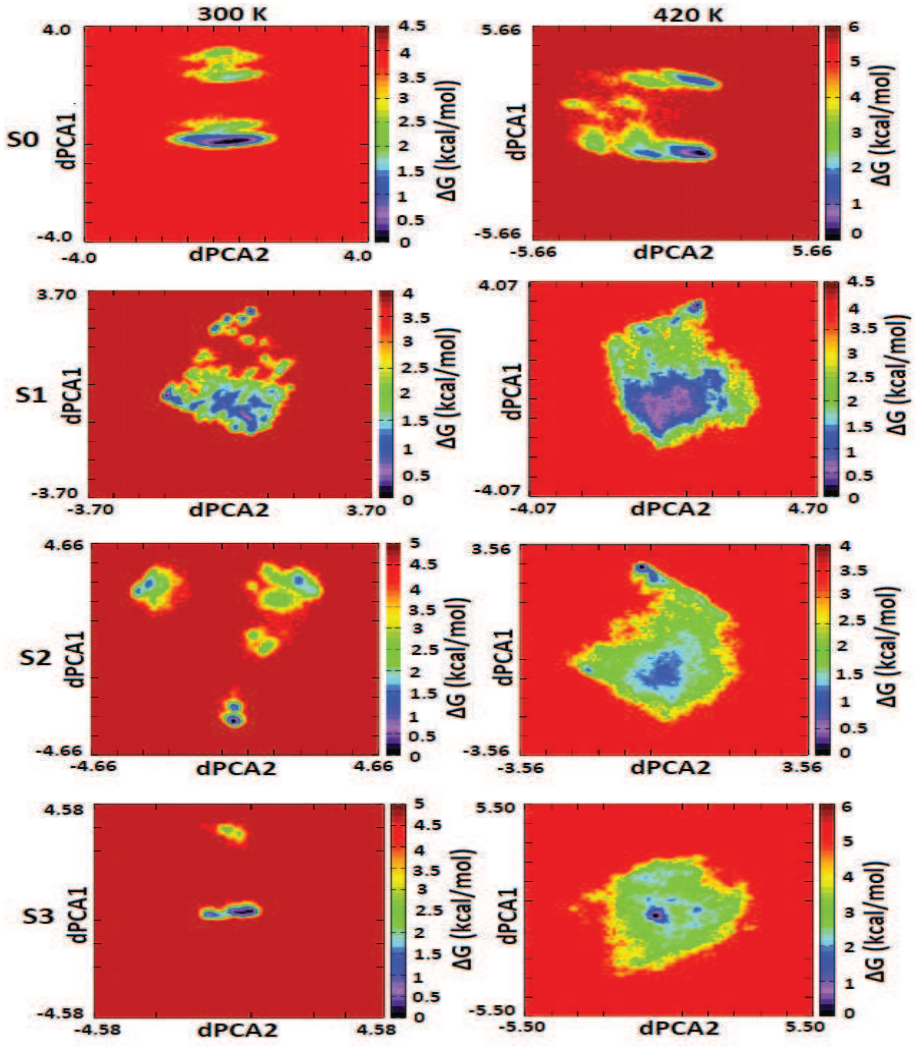

Figure 3B-11. *Free energy surface plot of dPCA1 versus dPCA2 for all the systems at 300 K (left-hand panel) and 420 K (right-hand panel).*

The free energy profile of the pure water system contains a concrete lower energy basin strongly suggesting the presence of a highly energetically stable conformation of the protein in this system. It is worth mentioning that, this low energy profile is observed in both room temperature as well as high temperature like 420 K which makes it very clear

that Trp-cage is stable up to 420 K in absence of any co-solute. Moving from system S0 to S1, a highly rugged energy surface is observed in both high and low temperatures where no trace of energy minimum is found, indicating the presence of completely energetically unstable conformation of the protein in 6 M urea. In presence of 0.5 M COS along with urea at room temperature, several energy basins are found separated with high energy barriers confirming that reversal of the unfolded to folded state is hard to achieve. But, with the addition of more COS into the system, at 300 K the free energy landscape becomes very compact like the system in pure water containing a significant portion of low energy profile, suggesting that the protein retains its most energetically stable conformation in this system. But, at high temperature region, the population of higher ΔG value increases showing the loss of conformational stability of the mini-protein.

Umbrella Sampling

We have calculated the potential of mean forces (PMFs) in terms of free energy by employing an enhanced sampling method called umbrella sampling (see Figure 3B-12) [146]. For PMF calculation, we have additionally constructed three more systems where a single peptide is immersed into 12000 water for pure water system. In the next system, 1600 urea is added to mimic the 6 M urea solution, and then 280 COS molecules are added in the 3rd system along with urea and protein to maintain the COS concentration of 1.0 M like as system S3. For these systems, at first we have carried out 50 ns normal MD simulation at 300 K in NPT ensemble following the same procedure described above in the Model and Simulation section. The coordinates collected from the final trajectory of the MD simulation of each corresponding system are used as the initial coordinates for running umbrella sampling simulations. The distance between C_α(Leu2) to C_α(Pro19) was chosen as the reaction coordinate which is increased from 3 to 42 Å along the X-axis with a harmonic force constant of 1 kcal/mol/Å^2 [216]. The reaction coordinate was chosen following the earlier atomic force microscope experiments in which the stability and folding of a protein was investigated considering the end-to-end distance as the reaction coordinate [206]. Due to the high flexibility of the terminal residues, the distance between the C_α-atoms of the terminal residues of Trp-cage is not considered as the reaction coordinate for PMF study [206]. The reaction coordinate is divided into 157 equally spaced windows each

of having 0.25 Å width, and the initial configuration of each window is generated from the final trajectory of its previous window. Each window is simulated for 10 ns, and thus a total of 1.57 μs simulation run is performed for each system to obtain the free energy of separation between two C_α atoms of residues 2 and 19 of Trp-cage at different solution environments. Weighted Histogram Analysis Method (WHAM) is used to generate the free energy profile of the PMF calculation [147, 148].

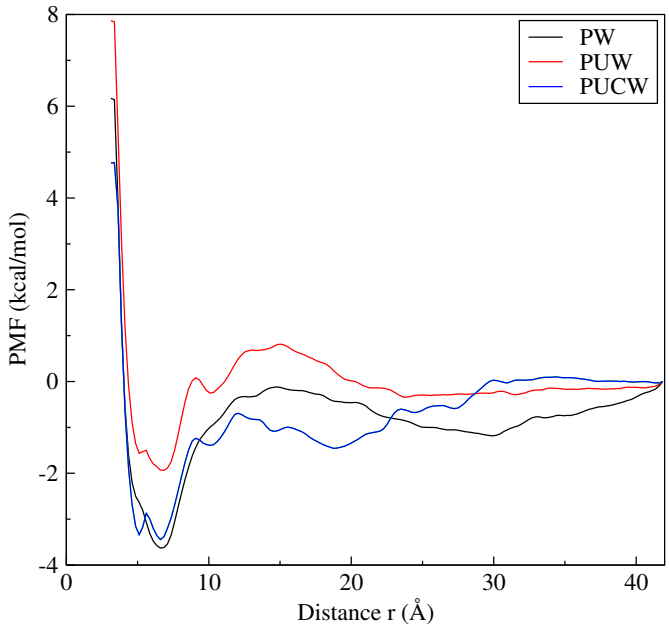

Figure 3B-12. *Potentials of mean force (in kcal/mol) as a function of distance between the backbone C_α-atoms of residues 2 and 19 of Trp-cage at 300 K.*

The unfolding pathway of Trp-cage mini-protein can be divided into three hierarchical steps (i) disruption of the hydrophobic core along with the salt-bridge between Arg16 and Asp9, (ii) unfolding of the α-helix and (iii) partial unwrapping of the helical domain. From Figure 3B-12 it can be noticed that in pure water the free energy profile contains a minimum at around 6 Å which corresponds to the native state of the protein. Then the free energy value increases smoothly up to 13 Å indicating a complete distortion of the

hydrophobic core at this distance. Also, the free energy of separation between two penultimate residues are most negative in this system which clearly indicates that unfolding is difficult to achieve. In case of protein in binary urea-water system, the energy minimum at 6 Å shifts upward (towards the more positive value), disclosing the easier disruption of the hydrophobic core in presence of urea, and at around 10 Å a respectively shallow minimum indicates the solvent assisted complete unreeling of the α-helix. In presence of 1.0 M COS solution, the energy well corresponding to the native structure is as deep as that of pure water system, implying the retention of energetically stable conformation of the protein in presence of 1.0 M COS.

■ SUMMARY AND CONCLUSIONS

In this chapter, we have ventured into the destabilization of the α-helical structure of a mini-protein Trp-cage in presence of 6 M urea, and the mechanism of counteraction by COS, employing an enhanced sampling method called replica exchange molecular dynamics (REMD) simulation. Temperature scanning conformational analysis shows that, Trp-cage is most stabilized in absence of any co-solute up to 420 K. In presence of urea, it loses its compact structure both at low and high temperatures whereas addition of the osmolyte COS into the system helps the protein to keep its secondary structure intact almost similar to the protein in pure water system up to 340 K temperature. It is found that, the distance between the hydrophobic core and the central Trp6 residue, and the orientational preferences of different residue-rings with the central indole plane have significant contributions towards the stabilization of the protein in different environment, whereas the salt-bridge interaction between residues Asp9 and Arg16 does not have significant impact on it. The interaction energy, hydrogen bonding interaction, and the coordination number calculation show that with the increased concentration of COS in the systems, the urea molecules get excluded from the protein backbone due to the comparatively weak protein-COS interaction as well as significant solvation of the urea molecules by COS present in the solution. The free energy landscape plots of ΔG vs RMSD at different temperatures reveal that in pure water as well as in presence of the co-solute COS, the RMSD remains within 2.0 Å while in presence of only urea or even with low concentration of COS, this value goes up to 8 Å even at the low temperature, indicating structural deformation of the protein. The

energy surface of the dPCA plot for pure water system contains a low energy basin in both high and low temperatures denoting the presence of energetically stable conformation of the protein in it. In presence of urea and 0.5 M COS, highly rugged energy surfaces separated by high energy barrier suggest the complete loss of helical structure in these systems. Nevertheless, in 1.0 M COS solution at room temperature, the protein retains its most stable structure, confirmed from the occurrence of low energy basin in the free energy diagram. Lastly, the most negative free energy of separation between two penultimate residues (Leu2 and Pro19) in pure water and 1.0 M COS solution indicates that unwinding of the folded structure of the mini-protein in these solutions is highly difficult to achieve. Therefore, this work clearly demonstrates that COS can effectively prohibit the urea-induced unfolding of the concerned protein (Trp-cage) at higher concentrations ($\geqslant 0.5$ M) delineating the mechanism behind it. Although in the previous part of this chapter (part A), we have established COS as a counteracting osmolyte against urea-induced denaturation of S-peptide, successful application of it on the globular mini-protein Trp-cage would help to draw attention to the wider approach of COS in the field of protein chemistry.

Combining the results of Part A and Part B, it is reasonable to conclude that COS excludes the urea molecules from the protein backbone, thus, preventing it to show its denaturing effects on the protein. Henceforth, COS can be treated as a effective osmolyte against the urea-conferred denaturation of proteins.

Chapter 4

The Conformational Stability of Terminal Helices of λ-repressor Protein in Aqueous Dodine and Choline-*O*-sulfate Solutions

'The growing efforts to understand the unique features of membrane proteins have encouraged the development of mixed micelles to study the equilibria and kinetics of this class of proteins, and traits which unite globular and membrane proteins have also emerged. These issues emphasize the amazing power of surfactants to both extend the protein conformational landscape and at the same time provide convenient and reversible short-cuts between the native and denatured state for otherwise obdurate membrane proteins."

– D. Otzen. *Biochimica et Biophysica Acta* **2011**, *1814*, 562-591.

Overview

In this chapter, we have ventured into the denaturation of fast-folding λ_{6-85}-repressor protein at a millimolar concentration of dodine, and henceforth, evaluated the candidature of COS as a protecting osmolyte against it, employing classical molecular dynamics simulations. Our simulation results show that the terminal helices of λ-repressor protein get unfolded in presence of $\sim$ 15 mM dodine while 0.5 M and higher concentration of COS can prevent this deleterious effect of dodine. Careful analyses of a set of simulations with increasing COS concentration reveals that a higher concentration of COS can provide remarkable stability to the protein, even better than its native state in water. Different interaction parameters show that in aqueous dodine, both the dodinium and acetate ions interact strongly with the terminal helices of the protein to denature whereas in presence of COS, due to the preferential interaction of COS with the protein molecule, dodine molecules get excluded from the protein surface. In addition, the favorable interaction of COS with dodinium head group makes the dodinium ions less available to the vicinity of protein surface which also plays an indirect but decisive role to prevent the unfolding of the terminal helical domains of the λ-repressor protein.

■ INTRODUCTION

In the world of protein science, several techniques are available for denaturing the proteins *in vitro* [221]. There are a number of chemical denaturants that are being used to unfold the thermally stable proteins [222]. Some proteins are there which have fast-folding kinetics because of the small energy barriers between their folded and unfolded states. These fast folding proteins demand chemical denaturants along with other perturbates to acquire the denatured state [223, 224]. Urea and guanidinium chloride are the most commonly used denaturants for *in vitro* chemical disruption of a protein's native structure [221]. These traditional chemical denaturants require very high concentration (molar level concentration) to destroy the secondary structure of a protein. Though these denaturants are good for the fluorescence measurement technique, due to the high concentration and strong absorption in the far UV region, several experimental methods required for the detection of the secondary structure of a protein are hampered.

Years ago, some modified compounds of basic guanidinium hydrochloride salt were prepared such as guanidinium, carbamoylguanidinium, guanylguanidinium salts etc where the guanidinium cation of urea was substituted by other head groups to increase their denaturing effect.[225] In 1983, Walker and his group made a powerful denaturant by the simple addition of a long alkyl tail to the nitrogen-based chaotrope biguanide which can denature the large protein penicillinase at a concentration less than 1.5 mM, but surprisingly no other group used the compound further, and no other application of it was found on smaller proteins [226]. People have also experimented with long alkyl chain detergents to increase the effectiveness of the denaturants [227].

Study of protein-surfactant interaction is a topic of high scientific interest for years due to the wide application of surfactants from the estimation of molecular weights of proteins to efficient washing powder enzymes and products for personal hygiene [228, 229]. Researchers have promoted the use of micelle such as sodium dodecyl sulfate (SDS) for the denaturation of protein, and the formation of higher-order structure. These surfactant works at a millimolar concentration which makes them extremely transparent towards the circular dichroism spectra measurement, and also, they can be used in gel electrophoresis to unfold a protein [27, 28]. Besides the experimental works, there are many simulation

works also reported in the literature on protein detergent interaction. The first molecular dynamics simulation study on peptide and detergent micelle system was done by Wymore and Wong in 1999 [230]. Since then, simulation studies are being performed to investigate the assembly of detergents around proteins by using atomistic as well as coarse-grained model [231, 232]. Böckmann and Caflisch studied the self-assembly of the short-chain lipid dihexanoylphosphatidylcholine (DHPC) in water around the outer membrane protein X (OmpX) by molecular dynamics simulation, and reported that the interactions between zwitterionic lipid head groups of the DHPC and charged residues of the protein molecule promote the formation of micelles around the protein [233]. Chris et al. in their simulation study found that surfactant dodecylphosphocholine forms equatorial micelles by solvating the hydrophobic belt of the protein through its acyl chain [234]. In 2005, high resolution NMR study by Cheryl and co-workers reveals that the compact hydrophobic core of Human Saposin C adopts an open conformation with an exposed hydrophobic pocket in presence of sodium dodecyl sulfate (SDS) [235].

Researchers have combined the properties of both the traditional and surfactant denaturants to obtain the best possible protein denaturant called dodine [31]. n-dodecylguanidinium acetate, commonly named as dodine, is an n-alkyl derivative of guanidinium (see Figure 4-1, part c). Dodine is a transparent protein denaturant for circular dichroism and infrared studies [32]. Gelman and co-workers found that dodine can effectively unfold the helical domains of the fast-folding lambda repressor fragment (λ_{6-85}) and WW domain at a millimolar concentration, far below its critical micelle concentration (CMC). Bacteriophage lambda C1 repressor is important to maintain lysogeny, where the phage replicates non-disruptively along with the host, and it consists of two domains connected by a linker which are N-terminal DNA-binding domain that mediates the interactions with RNA polymerase and C-terminal dimerization domain. λ_{6-85} is a five-helix bundled protein, obtained from the DNA binding site of the Lambda C1 repressor, and its helical domains fold by different mechanisms such as two-state folding [236], helical intermediates [237], downhill folding [238] depending on the sequence and solvent conditions. Recently, Gargi et al. studied the unfolding of the λ_{6-85} repressor mutate by a very low concentration of dodine, and its counteraction by trehalose [165]. They used the fast-folding λ_{YA} mutant of the protein which contains the mutations Y22W/Q33Y/G46A/G48A. The

λ_{YA} mutant is already a very stable incipient downhill folder, and its structure is quite similar to the wild-type protein. The polypeptide chain between mutations Y22W and Q33Y increases the rigidity near helix 2 [239]. Since alanine is a significantly better helix forming residue than glycine, the G46A and G48A mutation helps to increase the melting temperature of the N-terminal domain [240].

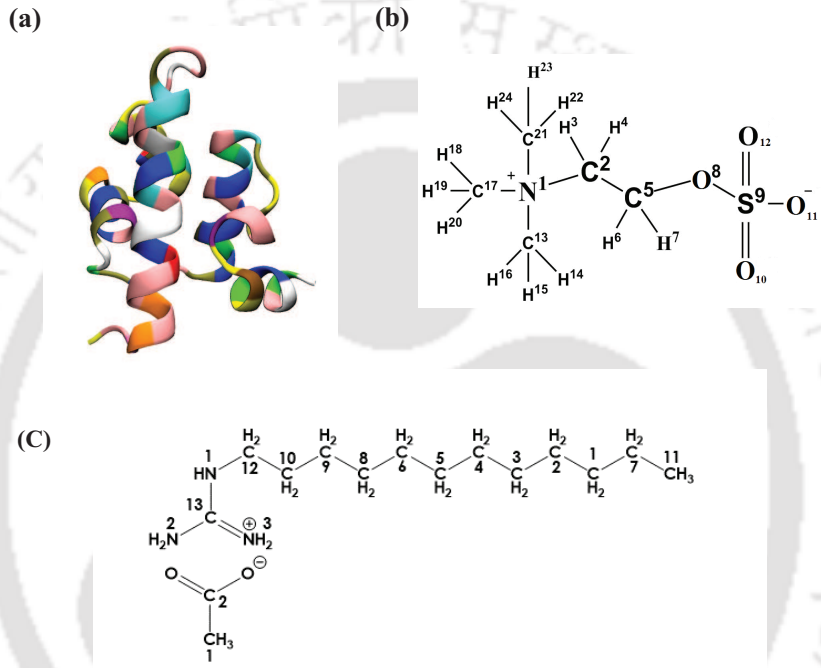

Figure 4-1. *(a) Initial structure of mutated λ_{6-85} repressor, (b) and (c) are molecular structures of COS and dodine respectively with atom numbering.*

Though the denaturation of λ_{6-85} repressor protein by very low concentration of dodine is well established both experimentally and theoretically, very few counteracting osmolytes against it are reported to prevent the structural disruption of such biologically important proteins [165]. In the previous chapters, we have revealed that COS can act as an effective osmolyte against urea-induced denaturation of proteins. In this current chapter, we have studied the counteracting ability of COS against dodine-induced denaturation of the mutated λ_{6-85} repressor protein by molecular dynamics simulation study. Experimental results show that in presence of 1.65 mM dodine concentration, full denaturation of the λ_{6-85} repressor can be achieved at 368 K temperature, and this melting point can be reduced by increasing the concentration of dodine in the solution [31]. MD simulations were done on the temperature and pressure denaturation of this protein where it has been found that refolding of the full protein takes a longer simulation time up to 20 to 50 μs in explicit solvent which is very much computationally expensive [241, 242]. It has been also found both experimentally and theoretically that removal of three helices from the five helices of the λ_{6-85} can reduce it to its smallest fragment which exactly behaves as the full-length protein [243]. To reduce the computational cost and convergence issues, we have particularly focused on the stability of the terminal helices H1 and H5 in ternary aqueous dodine-COS mixture with varying COS concentrations as it is believed that the terminal H1 and H5 of λ-repressor protein get more affected than other helical regions in stress condition [165, 242].

The rest of this chapter is arranged as follows: The simulation protocols are described in the Models and Simulation Method section, results are discussed in the section Results and Discussion, thereafter, a brief summary with concluding remarks are presented in the last section named Summary and Conclusions.

■ MODELS AND SIMULATION METHOD

To study the antidoting effect of COS on the denaturation of λ-repressor protein by dodine, we have carried out a series of classical molecular dynamics simulations. We have constructed five different systems (listed in Table 4-1) among which S0 is the pure water system containing protein and water, S1 is the protein-dodine-water mixture system, and S2, S3, S4 are the ternary systems containing protein, water, dodine along with three

different concentrations of COS i.e., 0.1 M, 0.5 M, and 1.0 M respectively. All solution species were packed in a cubic box using PACKMOL [116]. The protein, containing a total of 80 amino acid residues (PDB ID: 3KZ3) with point mutations at Tyr22Trp, Gln33Tyr, Gly46Ala, and Gly48Ala (see Figure 4-1, part a), was chosen for the study [239]. It has five helical domains and these are H1 (residues 9 to 25), H2 (residues 33 to 40), H3 (residues 44 to 52), H4 (residues 59 to 69), and H5 (residues 73 to 80). Following earlier work, we have used AMBER force field parameters for all the species [165]. For protein, ff14SB force field parameters, and for dodine and COS molecules General AMBER Force Field (GAFF) parameters were employed in this study (see Table 2-3 of chapter 2) [120, 207]. We chose TIP3P water model to solvate the protein as it is found to be the best combination with the AMBER protein force field parameters [111, 173]. The partial charges for all atomic sites of dodine were taken from the previously reported work [165] and for COS, the same partial charges and parameters given in chapter 2 (see Table 2-3) were used. The protein contains an overall charge of +1, so we have neutralized the systems by adding one chloride ion in each.

Table 4-1. Overview of Systems[a]

System	N_P	N_D	N_C	N_W	Box length (Å)	M_D	M_C
S0	1	0	0	12000	72.60	0	0
S1	1	4	0	12000	72.75	1.72×10^{-2}	0
S2	1	4	25	12000	73.05	1.71×10^{-2}	0.10
S3	1	4	125	12000	74.23	1.62×10^{-2}	0.50
S3a	1	4	125	12000	74.20	1.62×10^{-2}	0.50
S4	1	4	300	12000	76.12	1.51×10^{-2}	1.12
S4a	1	4	300	12000	76.10	1.51×10^{-2}	1.13

[a] N_P, N_D, N_C, and N_W represent the number of protein, dodine, choline-*O*-sulfate, and water molecules, respectively. M_D and M_C are the molar concentrations of dodine and choline-*O*-sulfate, respectively.

All the simulations were carried out using AMBER14 package [122]. At the starting of the simulation, 10000 steps minimization was done by equally distributed steepest descent method and conjugate gradient method. Following the previously reported work

[165], we have maintained a temperature of 320 K for all the systems throughout the whole simulation. After minimization, we heated the systems gradually from 0 to 320 K with an interval of 40 K using canonical ensemble (NVT) [123]. Then we equilibrated the systems for 5 ns in isothermal-isobaric (NPT) ensemble, and finally a production run of 1000 ns was continued in NPT ensemble for all the systems [124]. A constant pressure of 1 atm was maintained during the simulation using Berendsen barostat with a pressure coupling constant of 2 ps [126]. The temperature was controlled using Langevin dynamics method with a collision frequency of 1 ps^{-1} [127]. All bonds involving H-atom were restrained by SHAKE algorithm [129]. Particle-mesh Ewald method [128] was employed to calculate the long-range electrostatic interactions, and periodic boundary condition was applied in all the three directions. To check the robustness of simulation results, we have constructed another two systems containing 0.5 M and 1.0 M COS concentrations, namely S3a and S4a, starting from different initial configurations (see Table 4-1).

Before analyzing the trajectories, we have checked the convergence of our systems. To establish the robustness of the results obtained from a MD simulation, proper convergence of the system is very important. To check the convergence of the simulations, we have plotted every 50 ns block averaged helicity percentage of the N-terminal and C-terminal helices i.e., H1 and H5 respectively (see Figure 4-2) using DSSP algorithm [213] implemented in AMBER-CPPTRAJ [131] for pure water system (S0), binary dodine-water system, (S1) and system with the highest concentration of COS i.e., system S4. It is already reported that these two terminal helices are the most flexible helical domains of the λ-repressor protein. So, conformational convergence of these two helical domains should be the most important criteria for the convergence of the whole system, and from Figure 4-2, it is clear that the helicity percentage of H1 and H5 for all three systems are well converged after 500 ns.

■ RESULTS AND DISCUSSION

Structural Properties

In the protein folding-unfolding study, investigation of several structural changes of the protein from its native state is very important. For this purpose, we have calculated

the root mean square deviation (RMSD) of the C_α-atoms of the peptide backbone relative to its crystal structure as the simulation progresses (see Figure 4-3, part a).

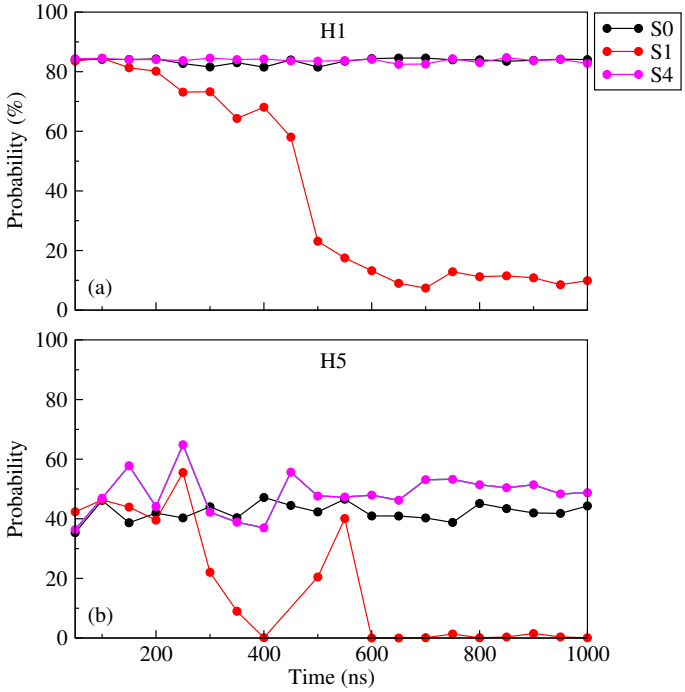

Figure 4-2. *Average helix percentages for helices (a) H1 and (b) H5 versus simulation time.*

As reported in the previous work [165], from Figure 4-3 it can be observed that the C_α-RMSD value for the system in pure water is within 2.0 Å . In system S1, the RMSD value starts to deviate after 250 ns, and reaches up to 5.5 Å , indicating a visible structural change of the protein in presence of dodine. In presence of 0.1 M COS, the RMSD value is very much unstable and higher than the pure water system from the beginning, and after 500 ns it deviates up to 5.5 Å like system S1, indicating that such a low concentration of COS is not sufficient to protect the structural changes of the λ-repressor protein by dodine molecules. In system S3, in presence of 0.5 M COS, the RMSD value is fairly low like the S0 system, and in presence of 1.0 M COS solution RMSD value throughout the simulation is very much stable, and it is slightly lower than the pure water system (within 1.5 Å) which means that 0.5 M and 1.0 M COS can prevent the dodine-induced structural deformation

of the protein. The RMSD values for systems S3a and S4a match well with the RMSD values of systems S3 and S4 respectively, indicating the consistency of results (see Figure 4-3, part c).

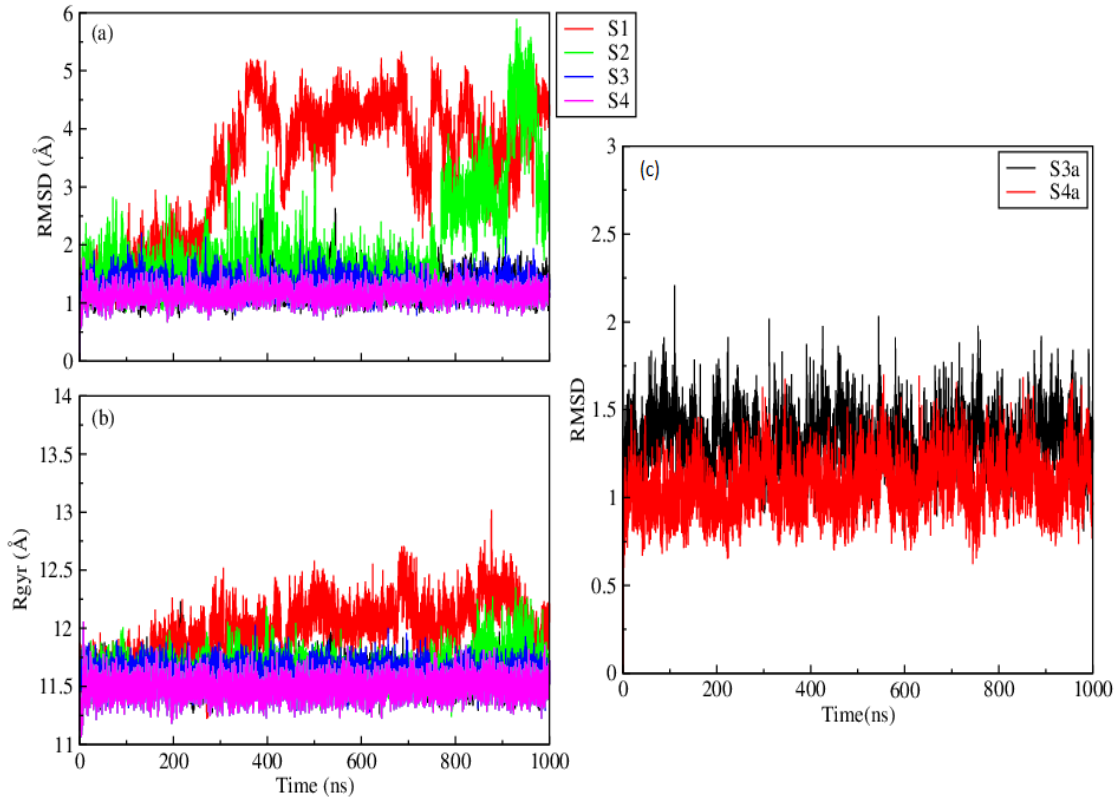

Figure 4-3. *(a) Root Mean Square Deviation (RMSD), (b) Radius of gyration (R_{gyr}) of C_α-atoms of the protein backbone with simulation time for systems S0, S1, S2, S3, and S4. (c) RMSD of C_α-atoms of the protein backbone with simulation time for systems S3a and S4a.*

Next, we have estimated the changes in the radius of gyration (R_{gyr}) of the peptide backbone from its native state with respect to the simulation time (see Figure 4-3, part b). From Figure 4-3b we can see, the R_{gyr} value of the native protein in pure water is within 12 Å which matches the previously reported work by Yanxin et al. [242] where the R_{gyr} value of the native peptide is shown around 12 Å. In the dodine-water system (S1), the R_{gyr} value increases up to 13 Å. In 0.1 M COS solution the R_{gyr} value is higher than the

pure water system, while in presence of 0.5 M and higher concentration of COS, the R_{gyr} values remain similar as that of pure system which clearly indicates that 0.5 M and higher concentration of COS helps the protein to be in its native state, even in the presence of dodine molecules.

In addition, to provide the time courses of RMSD and R_{gyr} changes during simulations, we have shown the snapshots of the λ-repressor protein conformation in different solution environments at different time intervals (see Figure 4-4). It can be observed that in pure water all the helical bundles of the protein remain folded from the beginning to the end of the simulation, whereas with the addition of dodine into the system, the helix H5 starts to unfold in the midway of the simulation, and at the end of the simulation, both the terminal helices (H1 and H5) get unreeled. From the snapshots of the protein structure, it is clear that in presence of a low concentration of COS, the helix H5 gets unwrapped after 500 ns, and with increasing concentration of COS all the helices including the terminal helices remain intact till the end of the simulation.

Conformational Changes of the Protein

In the native state, the λ-repressor protein has a compact structure with five helical domains. To study the folding-unfolding behavior of the protein in presence of dodine and COS, the determination of conformational changes of different helical domains is really important. Following earlier work of Martin and co-workers [242], we have calculated the changes in the secondary structure of the protein with simulation time using STRIDE algorithm implemented in VMD (see Figure 4-5) [130]. In pure water system (S0), we see the presence of five helical domains intact throughout the full course of the simulation, but in dodine-water binary system (S1), after 200 ns the two terminal helices (H1 and H5) lose their helical conformations and convert into turn conformation. In case of system S2, in presence of a very low concentration of COS, the C-terminal helix H5 remains very unstable from the starting of simulation, and after 500 ns it is totally disrupted to the random turn conformation. In presence of 0.5 M (systems S3 and S3a) and 1.0 M COS (systems S4 and S4a), all the five helices remain in their native states till the end of the simulation.

For in-depth study, next, we have calculated the percentage of α-helix contents of

the full-peptide as well as five helical domains using DSSP algorithm [213] implemented in AMBER-CPPTRAJ (see Figure 4-6) [131], considering the last 500 ns of the simulation run.

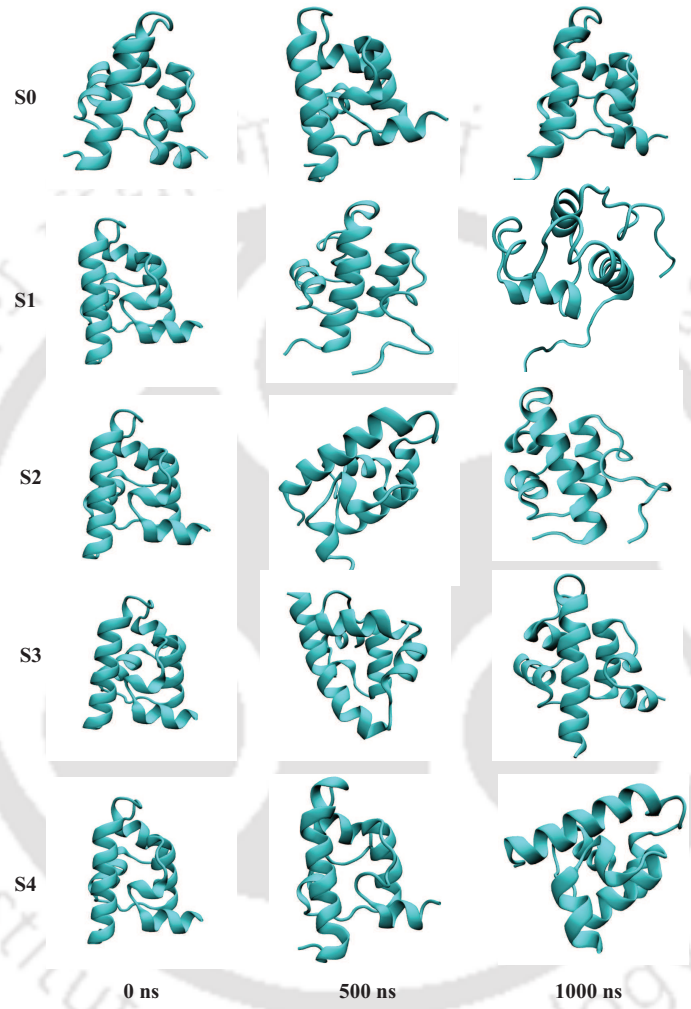

Figure 4-4. *Snapshots of the protein conformation at 0 ns, 500 ns, and 1000 ns for systems S0 to S4.*

In pure water system (S0), the helicity percentage of the full-peptide is 62.76 % which is very much comparable with the previously reported simulation result, where the percentage of α-helix content for the mutated λ_{6-85} repressor protein is reported to be 65.5 % [242]. In the dodine-water binary system, the percentage of α-helix is found to be 37.8 % . This result also shows good agreement with the previous study of high-temperature and high-pressure denaturation of the λ-repressor protein, where the secondary structure,

mainly α-helix content, drops from 60 % to 30 % within 0.15 μs due to the unfolding of different helices [242]. In presence of 0.1 M COS, the helical content is lower than the pure water system (55.54 %), but with increasing concentration of COS, the protein acquires the native state, and in presence of 1.0 M COS, the helicity percentage rises to 65.64 % which is slightly higher than the system in pure water.

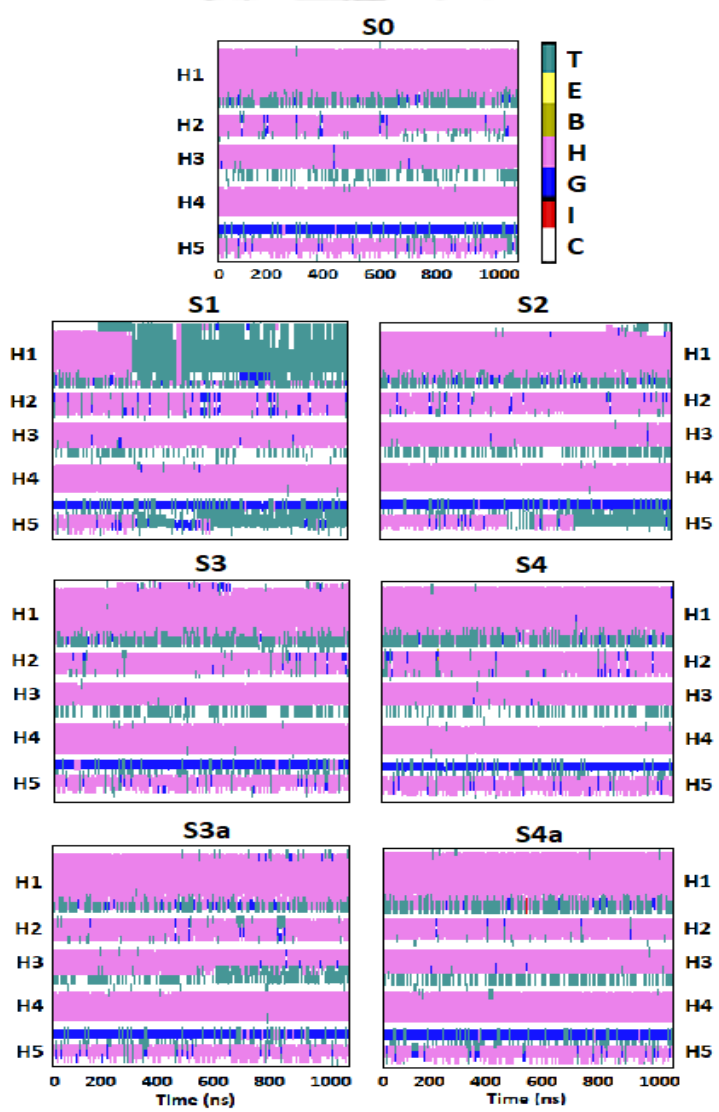

Figure 4-5. *Secondary structure analysis with respect to time for systems S0, S1, S2, S3, S3a, S4, and S4a. In the color bar T=Turn, E=Extended β-sheet, B=Isolated bridge-beta, H=α-helix, G=3_{10}-helix, I=π-helix, C=Coil.*

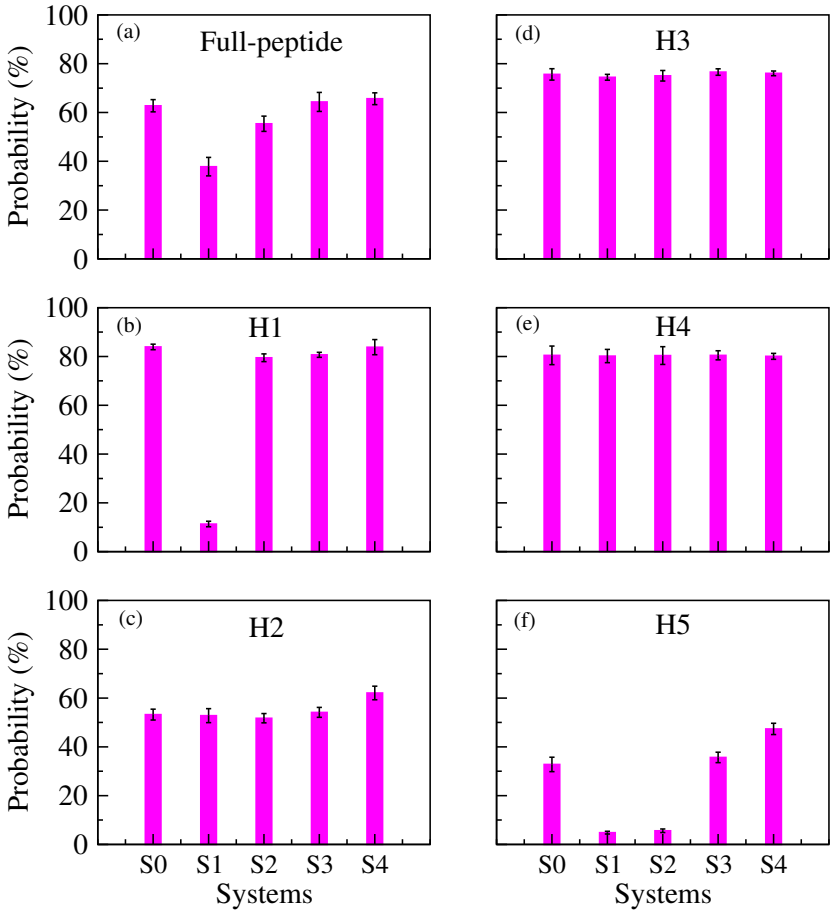

Figure 4-6. *Helicity percentage of (a) full-peptide, (b) helix H1, (c) helix H2, (d) helix H3, (e) helix H4, and (f) helix H5 for all systems considering the last 500 ns simulation trajectory. Standard errors are calculated using block average method.*

To investigate which helical domains are mainly affected by the dodine molecules, we have again calculated the α-helix percentage of all the five helices for all the systems. In part b of Figure 4-6, one can clearly observe a sharp drop in the helical content of the N-terminal helix in system S1 in presence of millimolar concentration of dodine, whereas, in presence of COS, this helical domain remains intact for systems S2, S3, and S4. For helices H2, H3, and H4, no conformational change is observed in any system. In case of

the C-terminal helix, which is reported as the most flexible helical domain of λ-repressor protein, gets fully unfolded in presence of dodine as well as 0.1 M COS (see Figure 4-6, part f). So, it can be said that a very low concentration of COS such as 0.1 M is not sufficient to protect the dodine-induced unfolding of the protein fully. In presence of 0.5 M COS, the helicity percentage of H5 domain is like the system in pure water. It increases in presence of 1.0 M COS indicating that higher concentration of COS can increase the stability and conformational rigidity of the λ-repressor protein than its native state. It is worth to mention here that, the percentage of helicity of full-protein, as well as helices H1 and H5 for systems S3a and S4a, are quite similar to that of systems S3 and S4 respectively, suggesting the results are independent of initial configurations (see Figure 4-7).

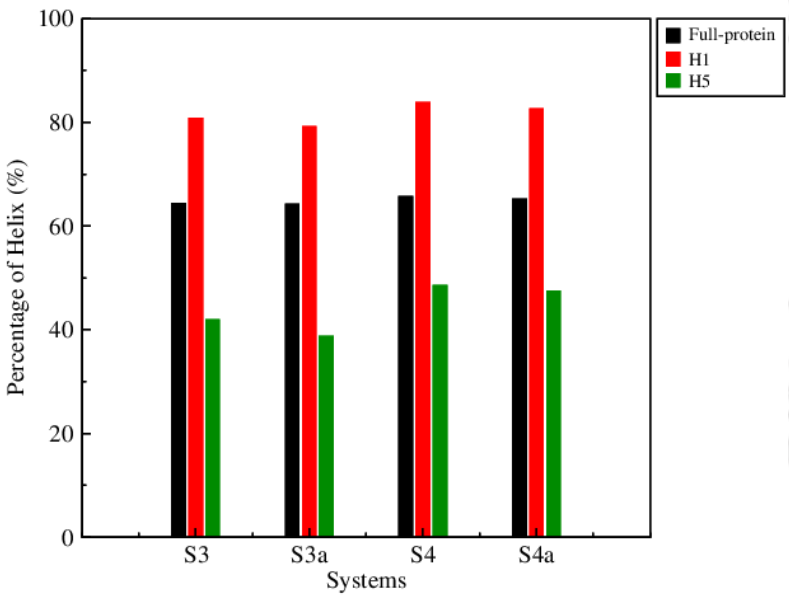

Figure 4-7. *Helicity percentage of full-protein as well as H1 and H5 for systems S3, S3a and S4, S4a.*

Dihedral Angle Principal Component Analysis

We have carried out dihedral angle Principal Component Analysis (dPCA) based on the dihedral angles (ϕ_i ψ_i) of the peptide backbone using equation 3.5, described in part B of chapter 3. The sin and cos transformed dihedral angles are considered as the order parameters for the structural changes of the helical core [244]. We have computed

the dPCA analysis by using Carma [220], and Figure 4-8 represents the multi dimensional free energy surfaces at 320 K resulting from all the structures into the first two principal components.

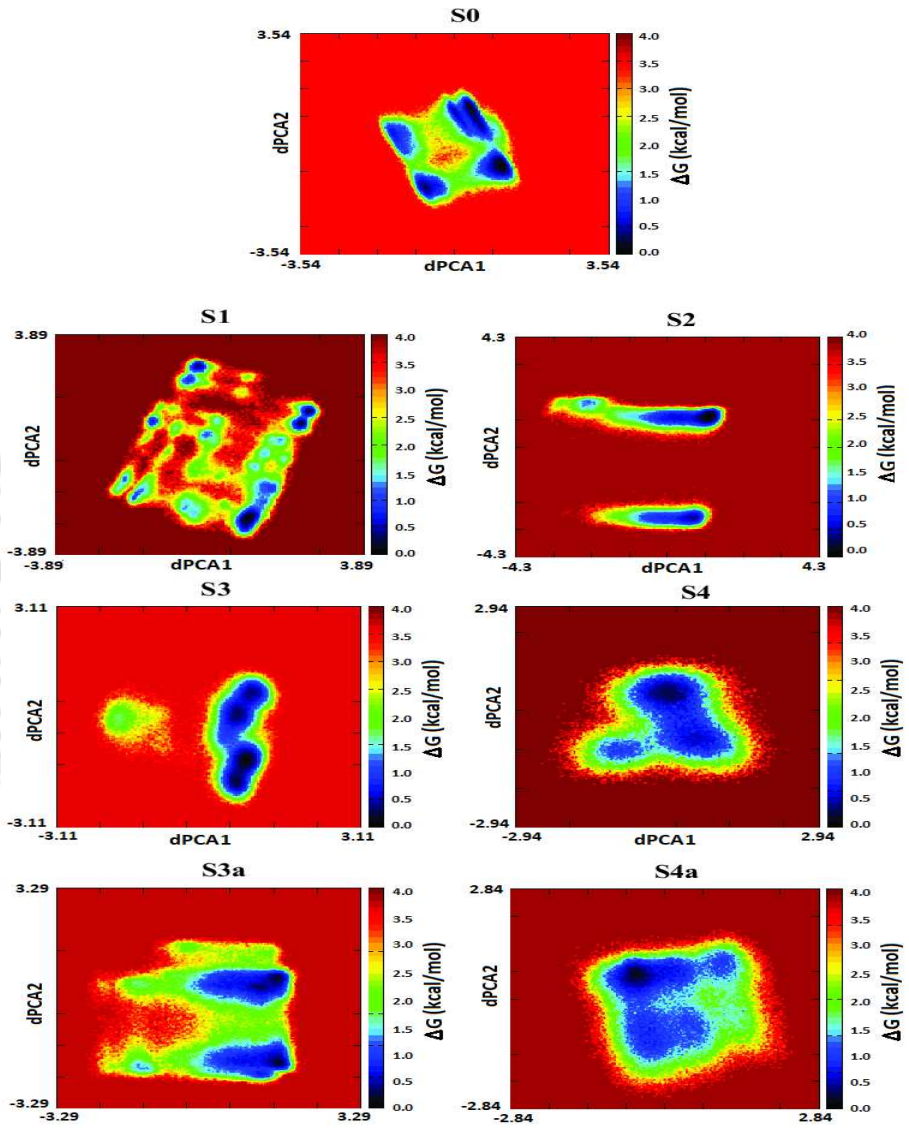

Figure 4-8. *Free energy landscapes of dPCA-1 vs dPCA-2 for systems S0, S1, S2, S3, S3a, and S4, S4a.*

In case of pure water system (S0), the free energy surface contains numerous minima of comparable energy values which clearly indicates that the highly stable conformational states of the peptide are easily convertible to each other. With the addition of the dodine into the system (system S1), the free energy landscape becomes highly rugged

containing very little amount of folded basin which strongly specifies the disruption of the native conformation in this system. In system S2 in presence of 0.1 M COS, the free energy surface contains two clusters with two minima separated by a high energy barrier indicating the partial loss of the native structure, and the intermediate conversion is difficult to achieve. In system S3, the free energy landscape quite resembles that of the system S0 containing two conformationally stable folded basins separated with a low energy barrier. In 1.0 M COS solution, the presence of one compact cluster with the largest folded basin strongly suggest the most conformationally stable structure of the peptide backbone present in this system. The free energy surface for system S3a contains two large folded basins separated by low energy barrier similar to that of system S3, and for system S4a, the free energy landscape contains one large cluster like system S4 (see Figure 4-8).

Interaction Energies between the Solution Species

To understand how COS protects the λ-repressor protein from the denaturing effects of dodine, we have calculated the total interaction energies between different solution species and decomposed them into vdW as well as electrostatic interaction energies (see Figure 4-9). Part a of Figure 4-9 depicts the interaction energies involving protein and dodine (both the dodinium and acetate ions are considered) for systems S1 to S4 from which we observe that the protein-dodine interaction energy (both vdW and electrostatic) decreases gradually with increasing concentration of COS. To know the individual contribution of dodinium and acetate ion, we have also calculated the interaction energies of these two ions with protein separately (see Figure 4-10). It can be seen that in case of dodinium ion due to the presence of positively charged head group as well as long alkyl chain, both the vdW and electrostatic interaction energies contribute significantly to the interaction with the protein, but in case of acetate ion, the contribution from vdW energy is almost zero. Next, we have estimated the protein-COS interaction energies for systems S2 to S4 (see Figure 4-9, part b) and found that in 1.0 M COS system protein has the strongest interaction with COS molecules, and also it is noticeable that the electrostatic interaction dominates over the vdW interaction energy. So, it can be said that, with the addition of COS molecules due to the stronger interaction between protein and COS, the protein-dodine interaction becomes unfavorable which helps to protect the protein from

denaturation.

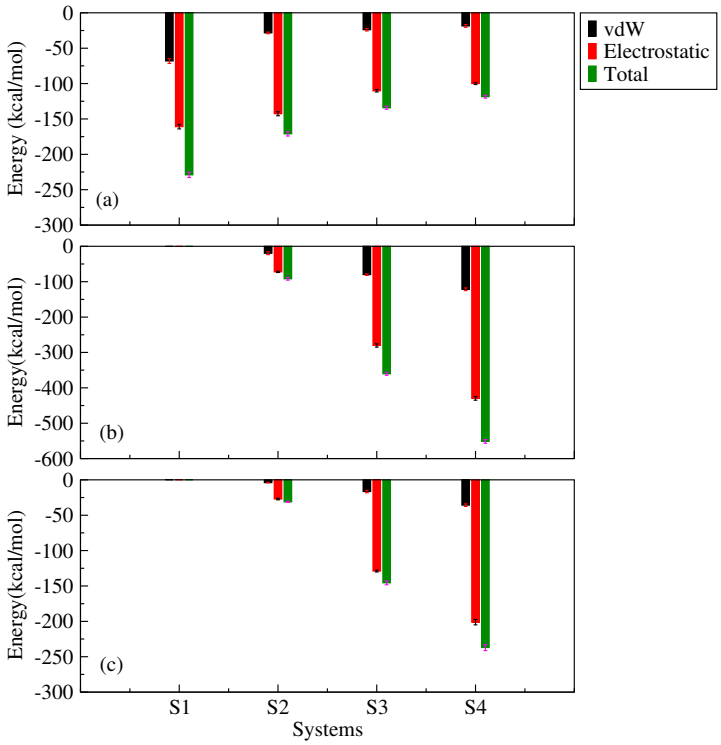

Figure 4-9. *(a) Protein-dodine, (b) protein-COS, and (c) COS-dodine (per dodine) interaction energies (in kcal/mol) for different systems considering the last 500 ns simulation trajectory. Standard errors are calculated using block average method.*

To examine whether any indirect effect of COS is involved or not, we have also calculated the interaction energies between COS and dodine molecules (see Figure 4-9, part c). It is worth mentioning here that, the COS-dodine interaction energies in systems S2, S3, and S4 are more favorable than the protein-dodine interaction energies of the corresponding systems. So, due to this favorable interaction between the dodine and COS molecules, dodine molecules get solvated by COS, hence, become less available for the protein backbone. So, we can conclude that both the direct interaction of COS with protein as well as the indirect exclusion of dodine molecules from the protein backbone by COS helps to keep the protein in its native state. It is to be noted here that, in case of COS-dodine interaction energy, the entire contribution arises from COS-dodinium ion

interaction energy, and no COS-acetate ion interaction is present in the solution.

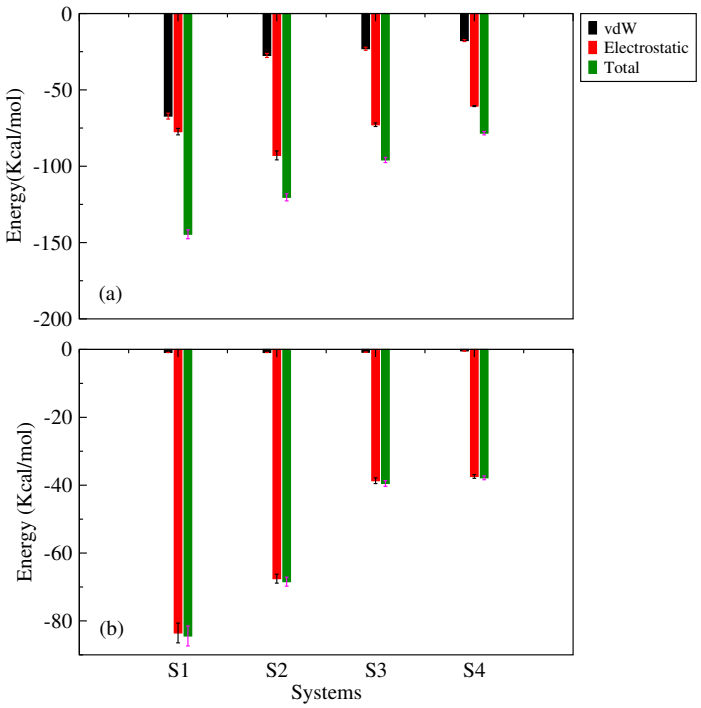

Figure 4-10. *(a) Protein-dodinium ion and (b) protein-acetate ion interaction energies (in kcal/ mol) for different systems considering the last 500 ns simulation trajectory. Standard errors are calculated using block average method.*

Hydrogen Bonding Interaction

An analysis of structures in a 'transition-state ensemble' shows that the helix to coil transitions occur predominantly through the breaking of hydrogen bonds at the helix ends, particularly at the C-terminus. So, to understand the structural properties of a protein, the determination of hydrogen bond number is an obvious choice. Following previous simulation work by Gargi et al., we have calculated different types of average hydrogen bond number using the criteria $\angle$D-A-H $\leqslant 60^{\circ}$ and D-A $\leqslant 3.5$ Å where D and A refer to donor and acceptor atoms [165]. All the hydrogen bonds are calculated by considering the last 500 ns of the simulation for each system. To check the stability of

the protein in different solutions, first, we have calculated the intra-protein hydrogen bond numbers for all the systems (Table 4-2). In pure water, the intra-protein hydrogen bond number is 60.55 which is quite similar to that of the earlier reported result (57.40) [165]. In presence of dodine solution in system S1, a decrease in the intra-protein hydrogen bond number is observed and it drops to 43.09. In 0.1 M COS solution, the H-bond number is still lower than the pure water system which indicates the loss of helical content of the protein in S2 system due to the rupture of intra-protein hydrogen bonds. In systems S3 and S4, the intra-protein hydrogen bond number is fairly high, and in fact, slightly higher than the native state. The intra-protein hydrogen bond number for systems S3a and S4a is also high, like systems S3 and S4 (see Table 4-2), suggesting that at 0.5 and 1.0 M concentrations, COS protects the protein from unfolding and provides extra stability to some extent. To know the interaction between the protein and the solvent, we have calculated average protein-water hydrogen bond numbers and found that, in S1 system due to the loss of compact structure of the protein molecule it becomes more exposed to the water molecules, hence, the protein-water hydrogen bond number is sufficiently higher than the other systems followed by the system S2 where the protein becomes partially unfolded. In the systems S4 and S4a (see Table 4-2), the protein-water hydrogen bond number is lower than system S0 because of the more compact structure of the protein in presence of 1.0 M COS. Next, we have estimated the hydrogen bond numbers between the protein-dodine and protein-COS and found that with the addition of COS molecules protein-dodine hydrogen bond number decreases, and a sharp increase in the protein-COS hydrogen bond number is observed in system S3 and S4. Moreover, in systems S3a and S4a, the protein-dodine hydrogen bond number decreases with increasing protein-COS hydrogen bond number indicating a good consistency of our results (see Table 4-2). Also, we have decomposed the protein-dodine hydrogen bond numbers into protein-dodinium and protein-acetate ion hydrogen bond numbers, and incorporated these in Table 4-2. The protein-dodinium H-bond number decreases with the addition of COS molecules into the systems, but in system S2 the protein-acetate ion hydrogen bond number is slightly higher than that of system S1. As the electrostatic interaction energy between the COS and dodine molecules is very high (see Figure 4-9), we have also determined the COS-dodine hydrogen bond numbers for systems S2 to S4 and found that hydrogen bonding interaction

between COS and dodine is stronger than protein-dodine interaction in systems S3 and S4. For the calculation of hydrogen bond numbers involving dodine, we have considered both the dodinium and acetate ions together. So, we can conclude that in presence of higher concentrations of COS, the COS molecules block the dodine molecules through strong hydrogen bonding interaction, and make them less available to attack the protein.

Table 4-2. Hydrogen bond numbers[a]

	HB_{PP}	HB_{PW}	HB_{PD}	HB_{P-DOD}	HB_{P-ACE}	HB_{PC}	HB_{CD}
S0	60.55(±0.87)	118.72(±1.31)	—	—	—	—	—
S1	43.09(±1.81)	158.86(±2.72)	4.60(±0.32)	3.32(±0.25)	1.30(±0.16)	—	—
S2	52.95(±1.22)	132.43(±1.42)	4.01(±0.29)	2.27(±0.21)	1.74(±0.19)	3.62(±0.09)	0.63(±0.07)
S3	61.11(±0.53)	117.06(±0.94)	1.97(±0.17)	1.44(±0.12)	0.53(±0.09)	14.08(±0.21)	2.77(±0.12)
S3a	60.68(±0.42)	121.69(±1.12)	1.85(±0.25)	1.38(±0.17)	0.47(±0.08)	11.83(±0.19)	2.79(±0.09)
S4	62.82(±0.49)	110.81(±1.01)	1.31(±0.13)	0.98(±0.11)	0.33(±0.03)	20.18(±0.48)	5.53(±0.35)
S4a	61.96(±0.64)	115.51(±1.31)	1.35(±0.13)	0.97(±0.13)	0.38(±0.06)	26.82(±0.61)	6.04(±0.39)

[a] Average hydrogen bond numbers for intra-protein, protein-water, protein-dodine, protein-COS, and COS-dodine for different systems over last 500 ns simulation. P, W, D, C represent protein, water, dodine, and COS respectively. DOD and ACE stand for dodinium and acetate ion respectively. Standard errors are calculated using block average method and provided in the parenthesis.

Protein Solvation

To examine the solvation of the protein in solution, we have calculated some selected pair correlation functions involving protein and different solution species (see Figure 4-11). The intense first peak of the radial distribution function between the protein and the hydrophobic tail of the dodinium ion in system S1 indicates that in binary protein dodine solution, the long carbon chain of the dodine molecules solvates the protein by hydrophobic interaction which disrupts the intra-protein hydrogen bonds leading to the denaturation of the helical domains. The RDF plot involving the protein and the head group of the dodinium ion also has the maximum peak height for system S1 (see Figure 4-11, part b), suggesting the presence of strong interactions between them. These RDF peak heights get

reduced significantly with increasing concentration of COS in the systems which indicates the gradual depletion of the dodinium ion from the protein backbone in presence of COS.

We have also estimated the pair correlation function between the protein and the acetate ion (Figure 4-11, part c) and found that the peaks are not so pronounced like the RDF peaks involving protein and dodinium ion which is probably due to the smaller size of the acetate ion compared to the long alkyl tail dodinium ion. Next, we have determined the radial distribution functions between the protein and the methyl carbon atom and sulfate oxygen atom of the COS molecule (see Figure 4-11, part d and e). In the RDF plot involving protein and methyl carbon atom of COS, the first peak is fairly lower from the bulk density, possibly due to the unfavorable interaction between them. In case of RDF plot involving protein and the sulfate oxygen atom of COS, the first peak appears at 3.0 Å which clearly indicates the presence of hydrogen bonding interaction between them, and with increasing concentration of COS, the peak heights get more pronounced from systems S2 to S4. Due to the presence of a considerable amount of hydrogen bond numbers between COS and dodine, we have also plotted the RDFs between the sulfate oxygen atom of COS and nitrogen atom of the dodinium head for all COS containing systems (see Figure 4-11, part f). The appearance of the first peak around 3.0 Å indicates the hydrogen bonding interaction between the negatively charged sulfate group of COS and positively charged head group of dodinium ion as well as the appearance of the second peak around 5 Å suggests the hydrophobic interaction between the methyl groups of COS with the alkyl tail of dodinium ion.

To get a more clear idea about the solvation of the protein by different solution species, we have calculated the coordination numbers of different molecules around 3.5 Å of the heavy-atoms of the protein (see Table 4-3) considering the last 500 ns simulation of each system. The coordination number of water oxygen around the protein increases sharply from system S0 to S1. This is because of the more exposure of the protein towards water molecules due to its unfolding in presence of dodine. In presence of 0.1 M COS (system S2), as the C-terminal helix gets unfolded, the number of water oxygen around the protein is almost similar to that of the system S1, whereas with the addition of more COS molecules into the systems this number gradually decreases, and at 1.0 M COS concentration the number of water molecules around the protein is slightly lower than the pure water system

indicating the more compactness of the protein structure in presence of 1.0 M COS. The number of dodinium carbon and nitrogen atoms, and the carbon and oxygen atoms of the acetate ion around the protein surface decreases with increasing concentration of COS, and simultaneously the number of carbon and oxygen atoms of the COS molecules increases significantly. From these results, we can conclude that, in presence of higher concentration of COS, due to the favorable interaction with the protein the COS molecules come closer to the vicinity of the protein surface which causes the depletion of dodine from it.

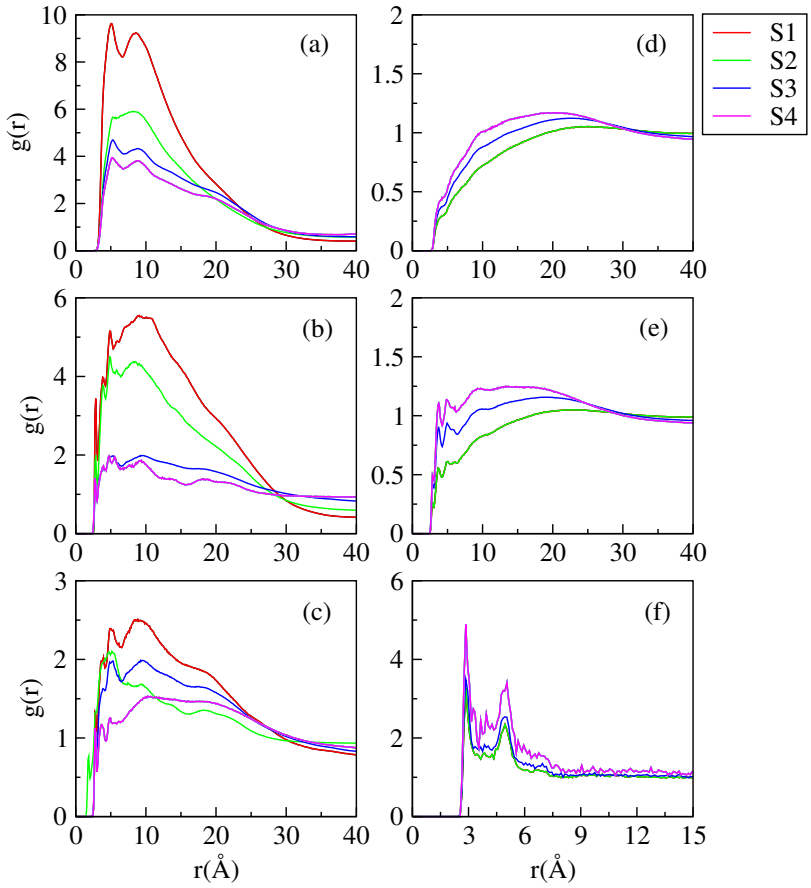

Figure 4-11. *Radial distribution functions involving: (a) Protein-dodinium tail, (b) protein-dodinium head, (c) protein-acetate ion, (d) protein-methyl carbon of COS, (e) protein-sulfate oxygen of COS, and (f) COS-dodinium ion for different systems considering the last 500 ns of the simulation.*

Table 4-3. Coordination numbers[a]

System	S0	S1	S2	S3	S4
WAT(O)	125.22 (± 3.21)	140.29 (± 3.54)	139.67 (± 2.81)	123.35 (± 3.63)	114.85 (± 2.02)
DOD(C)	—	4.42 (± 0.46)	3.10 (± 0.39)	0.83 (± 0.19)	0.82 (± 0.11)
DOD(N)	—	1.71 (± 0.23)	1.33 (± 0.20)	1.26 (± 0.09)	0.97 (± 0.12)
ACE(C)	—	0.42 (± 0.06)	0.39 (± 0.08)	0.25 (± 0.03)	0.17 (± 0.05)
ACE(O)	—	0.77 (± 0.11)	0.76 (± 0.13)	0.50 (± 0.07)	0.32 (± 0.03)
COS(C)	—	—	1.36 (± 0.23)	5.30 (± 0.35)	8.56 (± 0.49)
COS(O)	—	—	2.33 (± 0.11)	9.25 (± 0.33)	13.09 (± 0.27)

[a] Coordination numbers of the heavy-atoms of water, dodinium ion, acetate ion, and COS around protein. DOD and ACE represent dodinium and acetate ions respectively. Standard errors are calculated using block average method and provided in the parenthesis.

Also, we have calculated the average distance between the center of mass of the C_α atoms of the protein backbone and heavy-atoms of dodine and COS, listed in Table 4-4. From these data, we see, the average distance between the protein and dodine increases with increasing concentration of COS in the systems, whereas COS molecules come closer to the protein backbone gradually from system S2 to S4.

Table 4-4. Average distance[a]

System	PD	PC
S1	10.06 (± 0.34)	–
S2	10.75 (± 0.29)	6.03 (± 0.21)
S3	11.94 (± 0.42)	3.46 (± 0.17)
S4	12.88 (± 0.39)	2.44 (± 0.11)

[a] Average distance between the center of mass of the C_α atoms of the protein backbone and heavy-atoms of dodine and COS. P, D, and C represent protein, dodine, and COS respectively. Standard errors are calculated using block average method and provided in the parenthesis.

The spatial distribution function (SDF) of dodine and COS around 5.0 Å of the protein surface is determined using 'TRAVIS' (see Figure 4-12) [140]. For all the SDF plots, we have used the same iso-surface value of 0.5. From the spatial distribution plot of dodine around the protein molecule (see Figure 4-12, part a-d) we can see that the dodinium ions are more localized near the terminal helices in system S1 which is responsible for the unfolding of H1 and H5 of the λ-repressor protein in this system. With the addition of COS molecules into the systems, the density of dodine becomes less, and in presence of 1.0 M COS, it is almost negligible. Next, part e-g of Figure 4-12 depicts the contour density map of COS molecules around the protein surface which clearly shows that the density of COS molecules around the protein increases with increasing concentration of COS into the systems (from S2 to S4) due to which the stability of the protein increases gradually. The next SDF plot of COS around the dodinium ion (see Figure 4-12, part h-j) shows that the density of COS molecules around the dodinium ion is the highest in case of system S4 which also plays a vital role to reduce the deleterious effects of dodine on the protein molecule. Also, it is noticeable that the COS molecules are localized around the head group of the dodinium ion indicating favorable interaction between the positively charged head group of the dodinium and negatively charged sulfate group of COS.

Interaction of Dodine and COS with Terminal helices

From the above discussions, it is clear that the conformations of terminal helices of λ-repressor protein are being affected in presence of the denaturant dodine. Thus, to know the more specific interaction of dodine and COS with the terminal helices i.e., H1 and H5, we have calculated the interaction energies of these helices with dodinium, acetate, and COS (see Figure 4-13). In system S1, dodinium ion has significant vdW as well as electrostatic interaction with both of the terminal helices. In system S2, the interaction energy with helix H1 decreases but with H5 it remains the same as that of system S1. It is noticeable that, in system S2, a decrease in the vdW interaction energy of dodinium ion with H1 is more pronounced than the electrostatic energy. From this result, we propose that the hydrophobic interaction of long alkyl tail of the dodinium ion with helices contributes predominantly in the unfolding. The interaction energy between helices and acetate ion contains only electrostatic interaction energy which decreases with increasing concentration

of COS for H1, but in case of H5, the electrostatic interaction energy with the acetate ion increases slightly in system S2, which is also responsible for the disruption of helix H5 in this system. In case of the COS-helix interaction energy, the electrostatic energy component dominates over the vdW energy component, and it becomes more favorable with increasing concentration of COS into the systems due to which the interaction of dodinium and acetate with terminal helices decreases and the protein attains better stability than its native state.

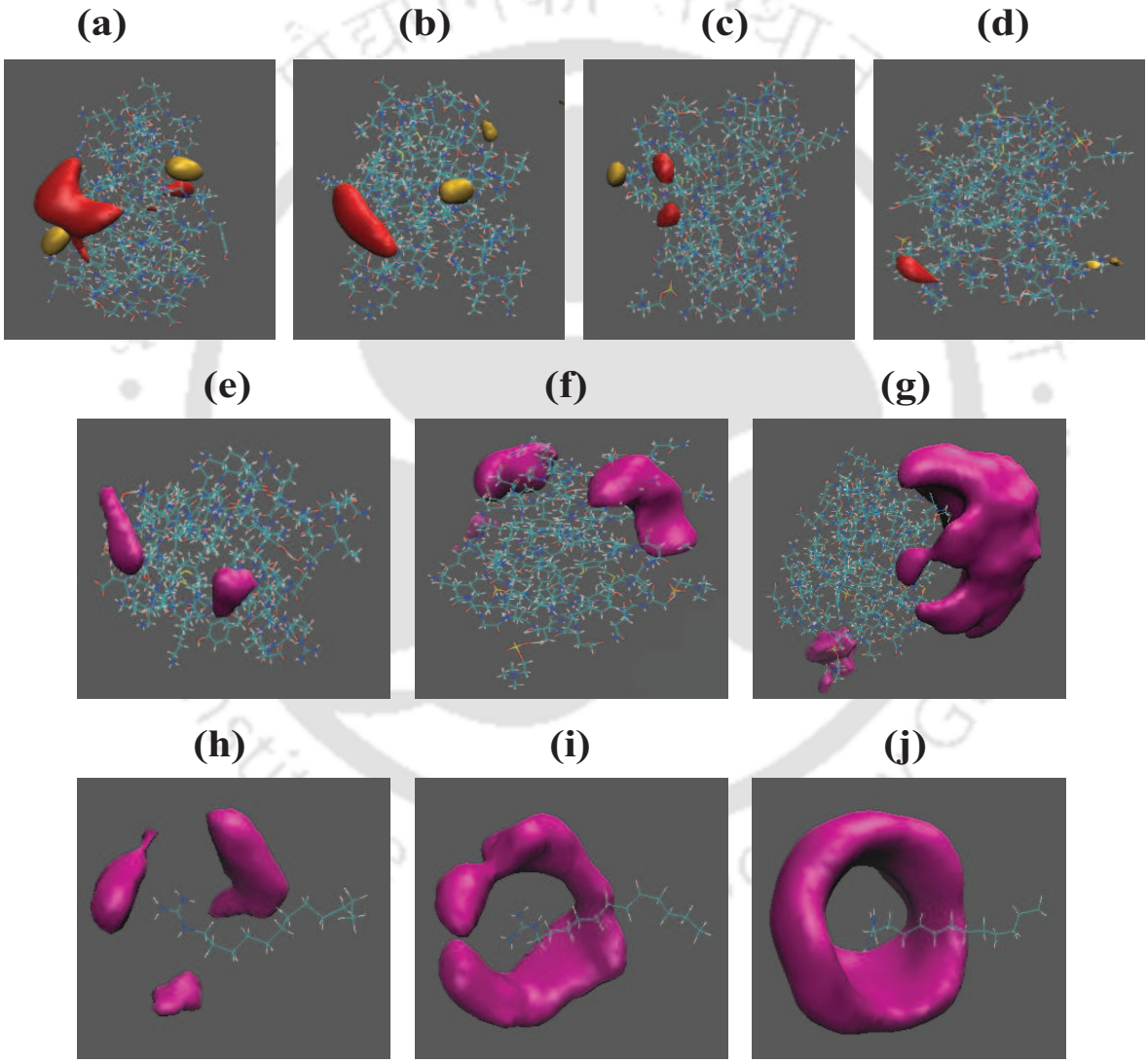

Figure 4-12. *Spatial distribution functions of dodine around protein for systems (a) S1, (b) S2, (c) S3, (d) S4; COS around the protein for systems (e) S2, (f) S3, (g) S4; and dodine around COS for systems (h) S2, (i) S3, and (j) S4.*

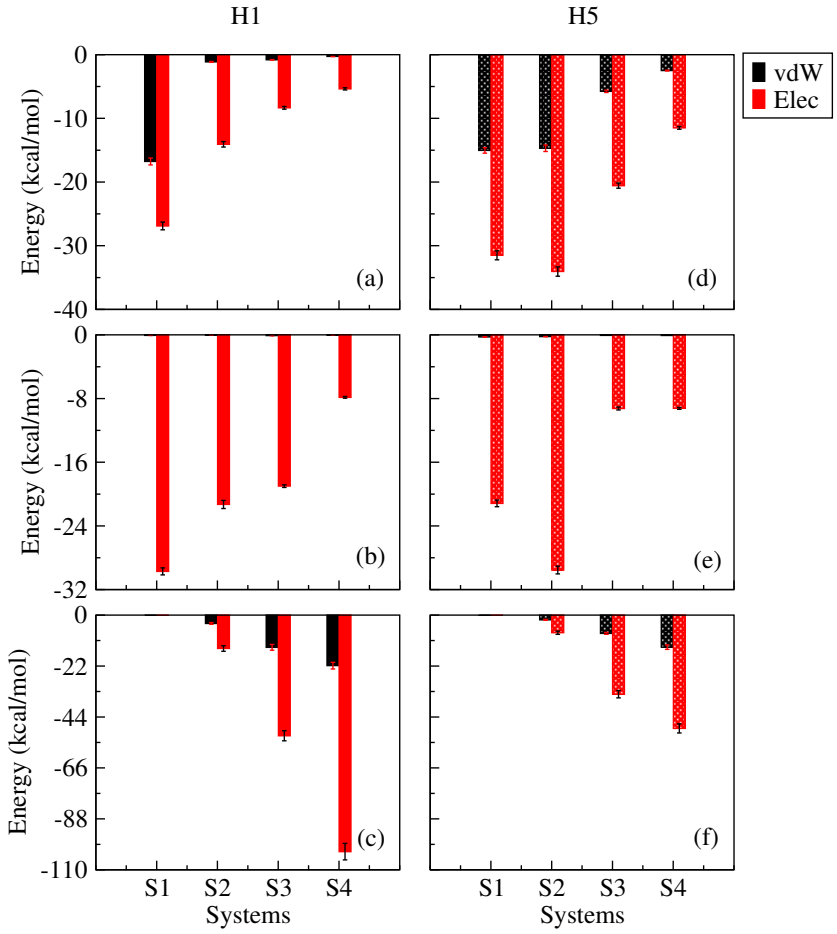

Figure 4-13. *vdW and electrostatic interaction energies between (a) H1-dodinium, (b) H1-acetate, (c) H1-COS, (d) H5-dodinium, (e) H5-acetate, and (f) H5-COS. Energies are expressed in kcal/mol unit.*

Next, we have estimated the average hydrogen bond numbers between the terminal helices and dodinium, acetate, and COS molecules (see Figure 4-14). For helix H1, the hydrogen bond numbers with dodinium ion decrease with the addition of COS into the systems. In case of H5-dodinium ion interaction, like the electrostatic interaction energy, the hydrogen bond number between them also increases slightly in system S2 than that of system S1. From part e of Figure 4-14 we see that, in system S2, helix H5 has the strongest hydrogen bonding interaction with the acetate ions, clearly indicating that both the dodinium and acetate ions conjugally play vital role in the unfolding of helix H5 in this system. From part c of Figure 4-14 we observe that, in system S2, COS has significant

hydrogen bonding interaction with helix H1 due to which the dodinium and acetate ions get excluded from this helical domain, and this helix gets stabilized in presence of 0.1 M COS. On the other hand, there is very poor hydrogen bonding interaction between the helix H5 and COS in system S2 due to which the dodinium and acetate ions can make favorable interaction with this helical domain and denature it. Also due to the exclusion of these ions from helix H1 in presence of 0.1 M COS, more dodinium and acetate ions are available to interact with helix H5 due to which the interaction energy as well as hydrogen bond number between them is higher than system S1. For systems S3 and S4, both the helices have a significant number of hydrogen bonds with COS which is one of the most important factors for the stability of terminal helices in 0.5 and 1.0 M COS solution.

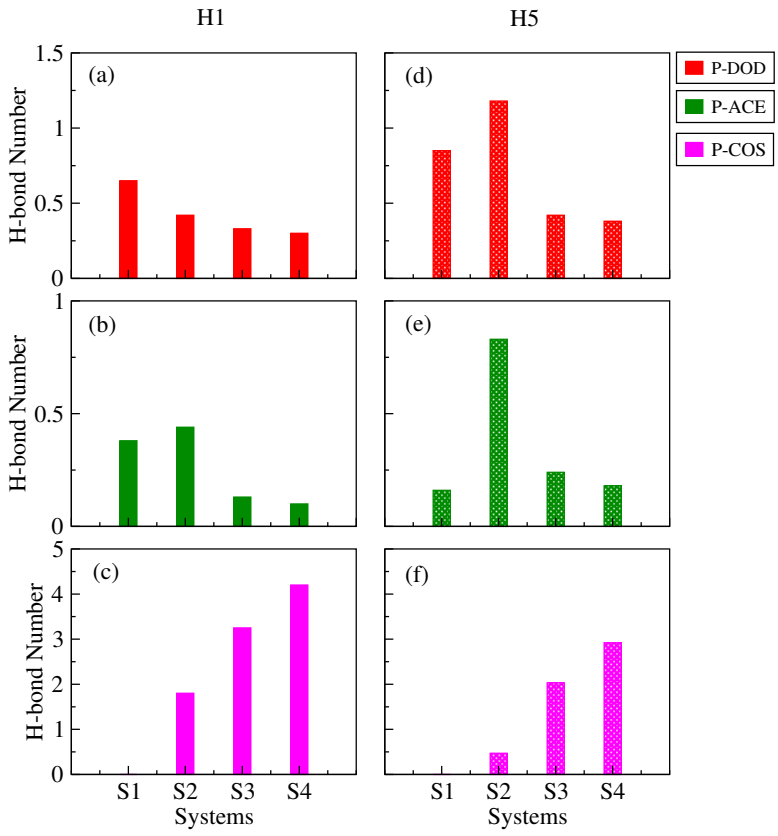

Figure 4-14. *Hydrogen bond numbers between (a) H1-dodinium, (b) H1-acetate, (c) H1-COS, and (d) H5-dodinium, (e) H5-acetate, (f) H5-COS.*

The coordination numbers of the heavy-atoms of dodine and COS around 3.5 Å of the terminal helical domains of the protein are also determined (see Figure 4-15). From part a of Figure 4-15 we see, in system S1, the helix H1 has the highest number of dodine around it which decreases with the incorporation of COS molecules into the systems (from system S2 to S4), but in case of helix H5, system S2 has the maximum coordination number of dodine around it which is already explained in the above sections. Part b of Figure 4-15 shows that in system S2, helix H1 has a higher coordination number of COS around it than helix H5, and with increasing concentration of COS, these numbers increase gradually from system S3 to S4.

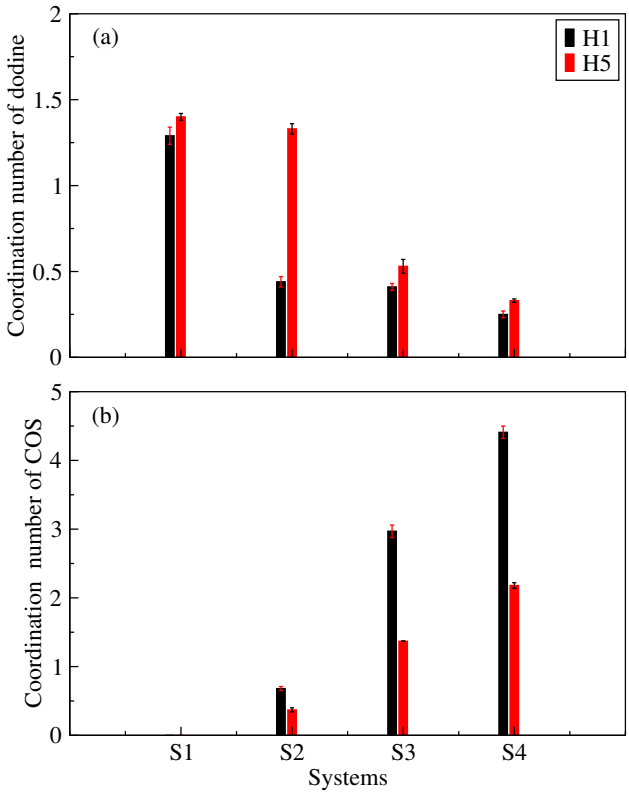

Figure 4-15. *Coordination number of (a) dodine and (b) COS around terminal helices of the protein for different systems.*

■ SUMMARY AND CONCLUSIONS

In summary, herein we have demonstrated the effectiveness of dodine as a protein denaturant, and simultaneously identified the protection mechanism of COS against dodine-induced protein denaturation, using DNA binding bacteriophage λ-repressor protein as a typical entity. Simulation of the protein-dodine system, even at a very low concentration ($\sim$ 15 mM) of dodine, displays significant deviation in the root mean square deviation (RMSD) as well as the radius of gyration (R_{gyr}) of the C_α atoms of the protein backbone after 200 ns, indicating structural deformation of the protein. Calculation of the helicity percentage of the full protein and different helical domains for different systems show that even in presence of millimolar concentration of dodine, the terminal helices (H1 and H5) face a sharp drop in the helical content compared to that observed in pure water. However, our simulations establish that the protein can be well protected from dodine-inflicted denaturation with increasing COS content in the solution. When present in only 0.1 M concentration, COS is unable to protect the protein from denaturation, as evident from the deviation in the RMSD and R_{gyr} value after 500 ns, and decreased helicity percentage for helix H5 in system S2. Nevertheless, our analyses suggest that increasing the concentration of COS triggers a number of changes in the subtle balance of protein-dodine interactions. At higher concentrations of COS, due to the favorable interaction of COS with protein through hydrogen bonding interaction, dodine molecules get excluded from the protein backbone, and also owing to the significant extent of interaction between COS and dodine, the latter does not prefer to interact with the protein. The dihedral angle principal component analysis (dPCA) plot for the pure system contains several folded basins with comparable energies indicating the presence of conformationally stable structures which are easily convertible with each other, whereas, in dodine-water and 0.1 M COS system, presence of small folded basins separated by high energy barriers indicates that intermediate conversion is difficult to achieve. On the other hand, in 0.5 and 1.0 M COS solution, the presence of compact clusters with large folded basins strongly suggests that in these systems the compactness and rigidity of the protein structure is high enough, and the secondary structure is not perturbed.

Identification of specific interactions of terminal helices with different solution

species reveals that, in the dodine-water binary system, the strong hydrophobic as well as electrostatic interaction of dodinium ion causes disruption of these helical domains, especially the strong electrostatic interaction with acetate ions is also playing an indispensable role in the denaturation. In presence of very low concentration of COS, the COS molecules interact with H1 but the interaction with helix H5 is indeed poor. Consequently, dodinium and acetate ions get excluded from the helix H1, leaving it in the native state, which, in turn, makes more dodine molecules available to interact with helix H5 to make it unfolded. At higher concentrations of COS, the interaction of both these helical domains with COS is much pronounced due to which interaction with dodinium and acetate ions becomes progressively feeble. *En bloc*, we conclude that dodine is capable of unfolding the terminal helices of the λ-repressor protein even at a very low concentration while presence of 0.5 M and higher concentration of COS can efficaciously behave as an antidote by eliminating the dodine molecules from the immediate vicinity of the protein surface through direct interaction of COS with protein backbone by hydrogen bonding as well as hydrophobic interactions, and indirectly solvating the dodine molecules by COS. In fact, we find that even a high concentration of COS bestows preferential structural stability to this protein.

We think, the present study provides substantial computational insights into the mechanism of dodine-induced protein denaturation and protection of the same employing COS, which is expected to trigger the experimentalists to judge the biological facets of the protein-COS-dodine interactions both *in vitro* and *in vivo*, thereby paving the way for the unequivocal establishment of COS as a potent novel counteracting osmolyte.

Chapter 5

Inhibitory Effects of Choline-*O*-sulfate on hIAPP Protofibrilation

"Several strategies are being pursued to inhibit and/or reverse protein misfolding and aggregation, with the hope that some of them will result in the generation of drugs that will be useful for the treatment of neurodegenerative diseases."

– C. Soto. *Nat. Rev. Neurosci.* **2003**, *4*, 49-60.

Overview

Type 2 diabetes mellitus (T2Dm) is a neurodegenerative disease, which occurs due to the self-association of human islet amyloid polypeptide (hIAPP), also known as human amylin. It was reported experimentally that choline-*O*-sulfate (COS), a small organic molecule having a tertiary amino group and sulfate group, can prevent the aggregation of human amylin without providing the mechanism of the action of COS in the inhibition process. In this chapter, we investigate the influence of COS on the full-length hIAPP peptide by performing 500 ns classical molecular dynamics simulations. From pure water simulation (without COS), we have identified the residues 11-20 and 23-36 that mainly participate in the fibril formation but in presence of 1.07 M COS, these residues become totally free of β-sheet conformation. Our results also show that the sulfate oxygen of COS directly interacts with the peptide backbone which leads to the local disruption of peptide-peptide interaction. Moreover, the presence of favorable peptide-COS vdW interaction energy and high coordination number of COS molecules in the first solvation shell of the peptide indicates that the hydrophobic solvation of the peptide residues by COS molecules also play a crucial role in the prevention of β-sheet formation. Finally, from the potential of mean force (PMFs) calculations we observe that the free energy between two peptides is more negative in absence of COS, and with increasing concentration of COS, it becomes significantly unfavorable indicating that the peptide dimer formation is most stable in pure water, which becomes less favorable in presence of COS.

■ INTRODUCTION

Neurodegenerative diseases like Alzheimer's disease (AD), type II diabetes (T2D), prion disease, Creutzfeld-Jacob's disease, Parkinson's disease are caused mainly by the formation of amyloid of different peptides [41, 42, 45, 46, 47, 48]. The formation of fibrils is the common characteristic feature of these diseases. Although different peptides are responsible for different amyloid fibrils, all pose similar types of physicochemical and structural properties of amyloidosis [245, 246]. Some research study found that different amyloid peptides including Aβ and tau, hIAPP and rIAPP aggregate to form protofibrils containing the same type of cross-β-structures formed by the pure amyloid peptides [247, 248, 249, 250, 251, 252, 253].

Type 2 diabetes mellitus (T2Dm) is a metabolic disease [254] which occurs due to the deposition of amyloid fibrils, and its advancement is associated with peripheral insulin resistance [41, 49, 50]. It is reported by some groups that human islet amyloid polypeptide (hIAPP), a full-length 37-residue peptide well-known as human amylin, accumulates to form the amyloid fibril of T2Dm [253, 255]. Though it is found that hIAPP oligomers are the major toxic species which cause the β-cell dysfunction of pancreatic islet, molecular structures of these oligomers are still difficult to track down [55]. The source of amyloid peptides is islet β-cells of pancreas, and it is co-secreted with insulin hormone [256]. hIAPP aggregates to form oligomeric structure containing ordered β-sheet conformation from random monomeric structure [257, 258, 259, 260, 261, 262, 263, 264, 265], and it is the most toxic species present in the pancreatic β-cells [266]. Thus, the earliest study of the β-sheet formation by Jaikaran et al. [257] proposes that the residues 24-29 and 32-37 participates in the anti-parallel β-sheet formation, and a β-turn is taken place at the residue 31. The residues 18-29 form extended β-sheet structure enhanced by residue 20. According to this study, the aggregation process is initiated by the hydrophobic interaction of different residues of hIAPP, and mainly the hydrogen bonding interaction makes the ordered secondary structure stable. A further study reveals that the anti-parallel β-sheet formation usually takes place in the regions 12-17, 22-27, and 31-37, while the residues 18-21 and 28-30 are involved in the formation of β-turns [264]. A recent study proposed that the side chain-side chain hydrophobic interactions between the residues 15-17, 32, and

residues 23-27 as well as the inter-peptide hydrophilic interactions of the region 28-31 is mainly responsible for the β-sheet formation [258]. Here it is worth mentioning that all three models on the β-sheet forming regions predicted by different groups have substantial similarities between them. Moreover, the proposed models of β-sheet forming regions includes the area where the aggregation occurs [267], as well as the π-π stacking interactions between different aromatic residues of the peptide, is also mentioned as an important factor behind the oligomerization of amylin [268]. By summarizing all these three models and merging them into one model it is found that the regions 12-17, 23-27, and 32-37 forms the β-sheet structure, and residues 20-21 involves in the formation of β-turns. The other residues like residues 18, 19, 22, and 28-31 have both the possibilities either to get involved in the formation of the β-sheet or β-turn. Therefore, the best possible way to prevent T2Dm is the inhibition of hIAPP aggregation.

Since the aggregation of amylin molecules is the main cause for acquiring cytotoxic properties, a handful number of research works have been performed in the last two decades to examine the molecular mechanism and the region (of amylin) responsible for its aggregation. Besides the study of hIAPP aggregation, nowadays, researchers are highly interested in developing different inhibitors that can stop the nucleation oligomerization of amyloid peptides. There are a numbers of inhibitors reported for the prevention of hIAPP aggregation (all of which are not possible to be mentioned here), still there is lack of rigorous research on it [268, 269, 270, 271, 272, 273, 274, 275, 276]. A recent study reported lysophosphatidylcholine inhibits the aggregation of human amylin through some non-specific interactions [277]. Nameki et al. studied experimentally the effect of choline-*O*-sulfate on the amyloid formation of human islet amyloid polypeptide [85]. Here, we are going to explore COS as an inhibitor on hIAPP fibrillation. They proposed that COS inhibits the aggregation of hIAPP by suppressing the ordered secondary structure formation from a random coil conformation. They assumed that mainly the sulfate group of COS plays the most important role in the prevention of β-sheet formation. The significant interaction between the sulfate oxygen of COS and the amide group of the peptide backbone prevents the interactions between the peptide backbone moieties, which facilitates the formation of β-sheet.

In this chapter, our goal is to provide molecular insights into the mechanism of how

COS inhibits amyloid formation. The rest of this chapter contains three sections including Models and Simulation Method, Results and Discussion, and the last section whereof we have included our concluding remarks with a brief summary.

■ MODELS AND SIMULATION METHOD

To examine the inhibitory effect of COS in amyloid formation, a series of classical MD simulations are performed with varying concentrations of COS. Previously, there are structural models of hIAPP consisting of three β-strands in one monomer [264, 257]. Recently, there are high-resolution amylin fibril structures made of two β-strands which are mainly U-shaped. These types of models of hIAPP are proposed by Luca et al. [258], Wiltzius et al. [259], and Bedrood et al. [260]. Further, Hansmann et al. in their study, explicitly discussed the initial structure of the hIAPP peptide and concluded that the X-ray models of hIAPP are topologically very similar to that of solid-state NMR model proposed by Luca et al [278]. This U-shaped model of hIAPP is found to be quite similar to the fibril models, detected recently from the brain tissues. These reasons lead us to use the U-shaped model of hIAPP as the starting configuration of our study. We have collected the PDB of the decamer amylin from Robert Tycko's research group following several works on amylin fibrils [55, 252, 258, 279]. Then we have extracted a monomer of full-length hIAPP of 37 residues from its decamer. The 37 residue peptide has the amino acid sequence K^{1}-C^{2}-N^{3}-T^{4}-A^{5}-T^{6}-C^{7}-A^{8}-T^{9}-Q^{10}-R^{11}-L^{12}-A^{13}-N^{14}-F^{15}-L^{16}-V^{17}-H^{18}-S^{19}-S^{20}-N^{21}-N^{22}-F^{23}-G^{24}-A^{25}-I^{26}-L^{27}-S^{28}-S^{29}-T^{30}-N^{31}-V^{32}-G^{33}-S^{34}-N^{35}-T^{36}-Y^{37}.

Initially, the monomer is U-shaped having a disulfide bond between two CYS residues (residue 2 and residue 7). hIAPP 1-37 mainly consists of three main regions. The N-terminal region that is region 1-19 is associated with the membrane and insulin binding, while the amyloidogenic region and the C-terminal of the peptide that is regions 20-29 and 30-37 participate in the self-aggregation of the peptide. hIAPP contains an intramolecular disulfide bond between Cys2 and Cys7 which is believed to be responsible for the stabilization of the N-terminus and the C-terminus of the hIAPP 1-37. In the monomeric state, the hIAPP peptides form β-strand-loop-β-strand (U-bend) conformation which is consisted of two anti-parallel β-sheets connected by one turn [i.e., β-strand (Lys1-Val17)-turn (His18-Leu27)-β-strand (Ser28-Tyr37)].

Before starting the simulation, five full-length hIAPP monomers were placed in a cubic box with sufficient distance between them. Using PACKMOL [116] four systems namely S0, S1, S2, and S3 were prepared, which are listed in Table 5-1. In system S0 five peptide monomers of hIAPP were immersed in pure water. For the rest of the systems, COS molecules were added with increasing concentrations of 0.2 M to 1.07 M. For this study, we have kept the number of peptide molecules fixed, and only the concentration of COS was changed. CHARMM General force field was adopted for COS. Here we used the same partial charges and parameters of different atomic sites of COS, mentioned in chapter 2 (see Table 2-2). The all-atom CHARMM36 force field and potential parameters were employed for the peptide, while the mTIP3P model was employed for water due to its good compatibility with CHARMM protein force field [108, 109, 111]. It may be noted that the choice of the protein force field and water model is an issue in simulating Intrinsically Disordered Proteins (IDPs) like hIAPP. In a recent study, Garcia and co-workers showed that the radii of gyration of different IDPs, obtained using the CHARMM force field, agree well with experimental data [280]. This further justifies the choice of the protein force field and water model in the present study. Each monomer of the peptide carries +2 charge due to the presence of two positively charged residues K^1 and R^{11}. So for 5 monomers, each system has a total net charge of +10, and thus, 10 chloride ions were added in each system to make them neutral [55, 279].

All simulations were carried out at ambient temperature (300 K). For simulation purpose NAMD 2.10 package was used [121]. At first, energy minimization was done by using the conjugate method followed by heating in canonical ensemble (NVT) [123]. Then all of the systems were equilibrated for 1 ns using isothermal isobaric (NPT) ensemble [124]. Langevin piston method was used to maintain the 1 atm pressure of the system [125]. Finally, a 500 ns production run was continued with the same NPT ensemble for each system. The temperature of each of the systems was controlled by Langevin dynamics method with a damping coefficient of 5 ps^{-1}. An integration time step of 2 fs was used for all the simulations, and the trajectories were saved after every 4 ps of the simulation. The short-ranged Lennard-Jones interactions were calculated using a cutoff distance of 12 Å while a switch distance of 10 Å was applied. Particle mesh Ewald (PME) method was employed to calculate the long-range electrostatic interactions [128]. All bonds involving

the hydrogen atoms were constrained using the SHAKE algorithm [129]. Visual Molecular Dynamics (VMD) plugin and TCL scripts were used for visualization and analysis purposes [130].

In order to examine if the results presented in this study are dependent on the initial configurations, we have prepared two more pairs of systems namely S0a, S0b, and S3a, S3b (listed in Table 5-1). Note that the systems S0a and S0b resemble system S0 except that these systems are having different initial configurations. The same is true for systems S3, S3a, and S3b.

Table 5-1. Overview of Systems[a]

System	N_P	N_C	N_W	volume (nm^3)	M_C
S0	5	0	15000	469.40	0
S0a	5	0	15000	469.29	0
S0b	5	0	15000	469.37	0
S1	5	60	15000	480.28	0.21
S2	5	150	15000	500.49	0.50
S3	5	350	15000	539.31	1.07
S3a	5	350	15000	539.45	1.07
S3b	5	350	15000	539.33	1.07
P0	2	0	15000	454.22	0
P1	2	60	15000	466.02	0.21
P2	2	150	15000	485.95	0.51
P3	2	350	15000	526.92	1.10

[a] N_P, N_C, and N_W represent the number of peptide, choline-*O*-sulfate, and water molecules, respectively. M_C is the molar concentration of choline-*O*-sulfate for different systems.

Following previous works, the potential of mean force (PMF) between two peptides was also calculated using adaptive biasing force (ABF) method [143, 144, 145]. For this, another set of four systems were prepared, systems P0 to P3 (see Table 5-1). Each of these systems contains two monomers of the peptide, and the concentration of COS is maintained as the previous systems S1 to S3. Then following the same procedure described above, these

four systems were simulated for 100 ns, and the initial configurations for the calculation of free energies were prepared from the final trajectory of this 100 ns production run.

■ RESULTS AND DISCUSSION

Secondary structure analysis in different solution

To see the formation of β-sheet structure of hIAPP in systems S0 to S3, we have represented the snapshots of the initial, intermediate, and final states of the peptides (see Figure 5-1). The formation of β-sheets of the peptides in pure water is observed. For systems containing COS (systems S1 to S3), we find that with increasing concentration of COS, the tendency of peptide-peptide aggregation deceases. It is worth mentioning that only the snapshots do not produce enough information about a simulated system. So, we have further studied in detail the effects of COS in the hIAPP protofibrilation (discussed below).

Next, we estimate the percentage of β-sheet, α-helix, and random coil conformation in different systems using DSSP algorithm [213], and they are presented in Table 5-2a. In system S0, the oligomers contain 24.56% β-sheet conformation followed by 0% α-helix and 50.29% random coil structure. Although initially, we placed the monomers in the simulation box maintaining sufficient distance between them, the formation of oligomeric structure denotes the high aggregation tendency of hIAPP in pure water. Moreover, for systems S0a and S0b, almost similar percentages of secondary structure conformation as that of system S0 are observed (see Table 5-2b). In system S1 with 0.2 M COS concentration, a sharp decrease in β-sheet formation can be observed (11.64%), and the percentage of helix content is also increased to 4.79% . With increasing the concentration of COS in the systems, the probability of finding β-sheet structure gradually becomes more lower. At the highest concentration of COS (system S3), though the total β-sheet percentage is not so much less than that of system S2, the percentage of extended β-sheet formation becomes zero, and also the percentage of random coil formation is the highest in this system due to the conversion of the β-sheet to random coil structure. In the experimental work by Nameki et al., they also found that in absence of COS, the β-sheet percentage increases with decreasing percentage of random coil conformation, but in presence of COS, this

conversion is very less with a slow process [85].

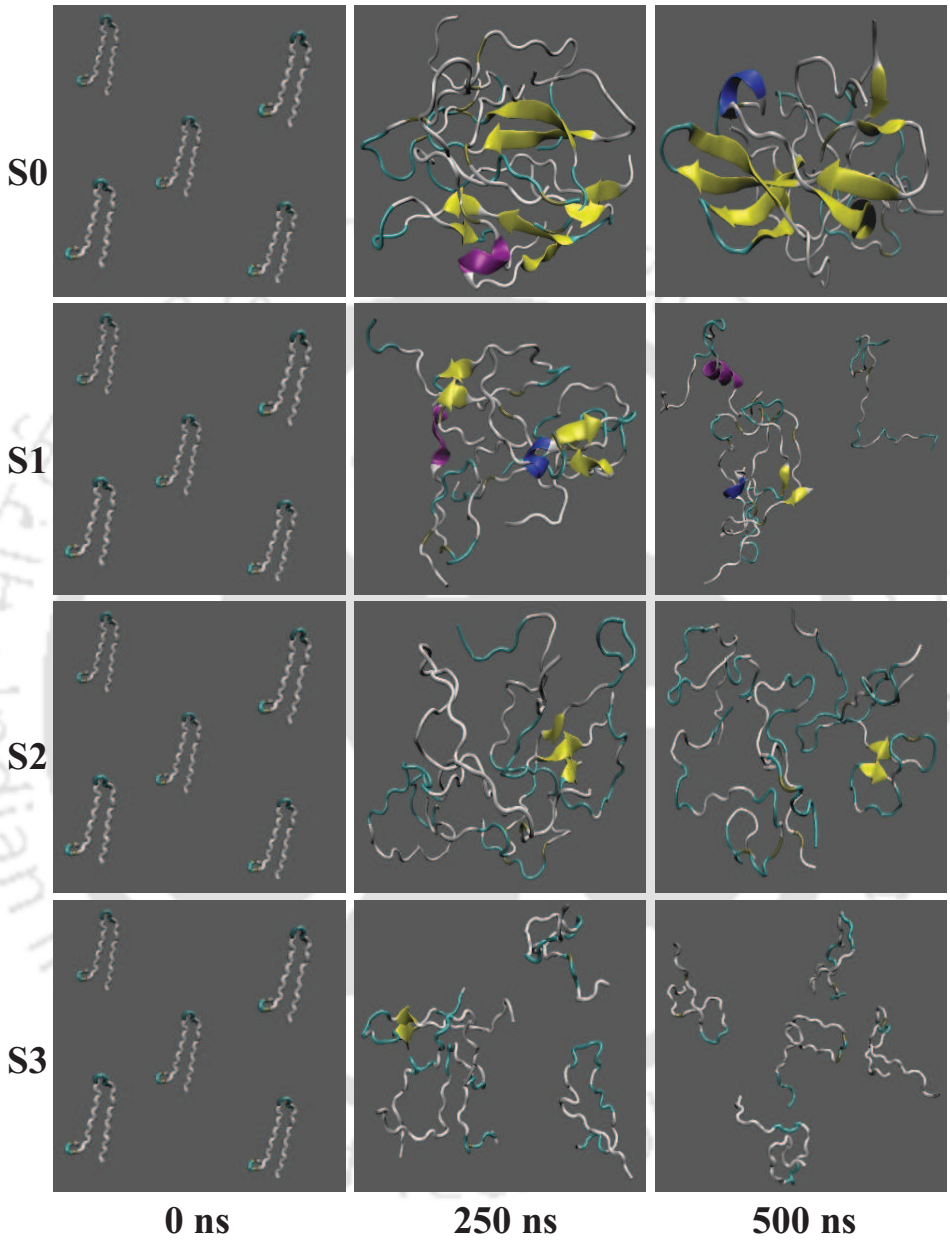

Figure 5-1. *Snapshots of different systems at 0 ns, 250 ns, and 500 ns (from left to right). Water and COS molecules are left-off for better visual clarity.*

Table 5-2a. Percentage of different secondary structure conformations[a]

System	S0	S1	S2	S3
%E (Intra)	12.39 (±0.27)	0	0	0
%E (Inter)	6.71 (±0.12)	4.12 (±0.19)	3.59 (±0.12)	0
%E (Total)	19.10 (±0.36)	4.12 (±0.19)	3.59 (±0.12)	0
%B (Intra)	0	0	0	2.21 (±0.16)
%B (Inter)	5.46 (±0.16)	7.52 (±0.25)	5.31 (±0.11)	5.80 (±0.22)
%B (Total)	5.46 (±0.16)	7.52 (±0.25)	5.31 (±0.11)	8.01 (±0.10)
%Total Intra-β	12.39 (±0.27)	0	0	2.21 (±0.16)
%Total Inter-β	12.17 (±0.18)	11.64 (±0.21)	8.90 (±0.11)	5.80 (±0.22)
%Total β-sheet	24.56 (±0.32)	11.64 (±0.21)	8.90 (±0.11)	8.01 (±0.10)
%H	0	4.79 (±0.17)	0	0
%C	50.29 (±0.11)	49.47 (±0.89)	50.33 (±0.50)	62.28 (±0.51)

[a] percentages of extended β-sheets (E) (both intra and inter peptide), bridge-β (B) (both intra and inter peptide), helix (H), and random coil (C) conformations for system S0 to S3. Standard errors are calculated using block average method and provided in the parenthesis.

Here also we have decomposed the probability of total extended β-sheet and isolated β-bridge formation into intra-peptide and inter-peptide β-sheet conformations. In S0 system, the probability of intra-peptide extended β-sheet formation is very high (12.39%) which becomes totally zero in all the COS containing systems (S1, S2, and S3). Probability of inter-peptide extended β-sheet formation also decreases slowly with increasing concentration of COS. Thus, it is found that in presence of COS, intra-peptide β-sheet formation is more suppressed than inter-peptide one. We have also calculated the percentage of β-sheet, α-helix and random coil conformations for systems S3a and S3b (see Table 5-2b) where we find that the percentages of extended β-sheet formation in these systems are also zero, similar to the system S3. Thus, the closeness of the percentages of different peptide conformations for system S0 with S0a and S0b, and S3 with S3a and S3b confirms the robustness of our simulation results.

Table 5-2b. Percentage of different secondary structure conformations[a]

System	% E	% B	% Total β	% C
S0	19.10	5.46	24.56	50.29
S0a	18.26	2.36	20.62	52.33
S0b	17.77	5.35	23.12	55.78
S3	0	8.01	8.01	62.28
S3a	0	7.45	7.45	66.55
S3b	0	9.06	9.06	61.15

[a] percentage of β-sheet (E), bridge-β (B), random coil (C) conformations for systems S0, S0a, S0b and S3, S3a, S3b.

To obtain the residue-based secondary structure information, we have calculated the probability of formation of secondary structure for different amino acid residues for system S0, and for the system with the highest concentration of COS that is system S3 (see Figure 5-2, part a) [51]. It is apparent that mainly residues 11-20 and 23-36 have a relatively higher propensity to sample β-sheet conformation than other residues [281]. This means the residues 11-20 and 23-36 participate in β-sheet formation while 21-22 is taking part in the formation of β-turn. What is remarkable is that the presence of 1.07 M COS completely diminishes the β-sheet formation for residues 4, 12, 14, 17-18, 20, 24-25, 27, 29-30, and 33-36. Moreover, the presence of COS also decreases the probability of formation of β-sheet for few other residues of the peptides significantly. Without COS molecules, mostly the hydrophobic residues including L12, 16, 27; F15, 23; V17; G24; A25; I26, and some polar residues mainly T4, 30 and 36; H18; S19, 20, 28, 29, and 34; N14, 31, and 35 display higher interaction probability to form oligomers, indicating that the hydrophobic, as well as polar interactions, play important role in the aggregation of hIAPP.

We have also compared the probability of residue-wise β-sheet formation for system S0 with systems S0a and S0b, and system S3 with systems S3a and S3b (see Figure 5-2, part b and c). A very similar probability of residue-wise β-sheet (extended β-sheet + isolated bridge-β) formation is observed for systems S0, S0a, and S0b; and this fact is also true for another set of systems i.e., systems S3, S3a, and S3b.

We have also decomposed the residue-wise total beta-sheet formation into intra-

peptide and inter-peptide total β-sheet formation for systems S0 and S3 (see Figure 5-3). The residue-wise intra-peptide and inter-peptide β-sheet formation both decrease in presence of 1.07 M COS, but the decrease in the intra-peptide β-sheet formation is more significant than inter-peptide one.

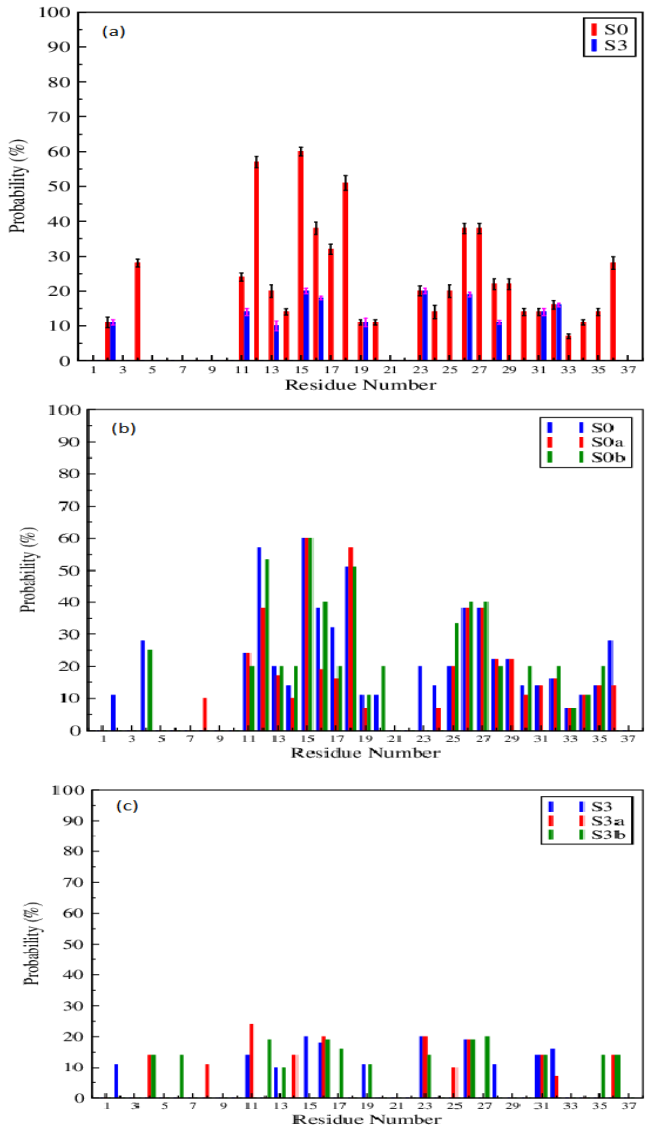

Figure 5-2. *Probability of residue-wise total β-sheet (extended β-sheet + isolated bridge-β) formation in systems (a) S0 and S3; (b) S0, S0a, and S0b; and (C) S3, S3a, and S3b. Standard errors are calculated using block average method.*

The residue-residue contact map between different residues of the peptides considering the C_α-C_α distances is computed taking the last 100 ns trajectory of the simulation (see Figure 5-4). It is noticed that in system S0 the residues of one peptide are in good contact with the residues of other peptides [282]. However, with the increasing concentration of COS, the pair-wise inter-peptide contact between the residues of the peptides decreases, and in presence of 1.07 M COS, all the peptides get separated from each other. The presence of white color denotes the distance between two peptides is more than or equals to 10 Å .

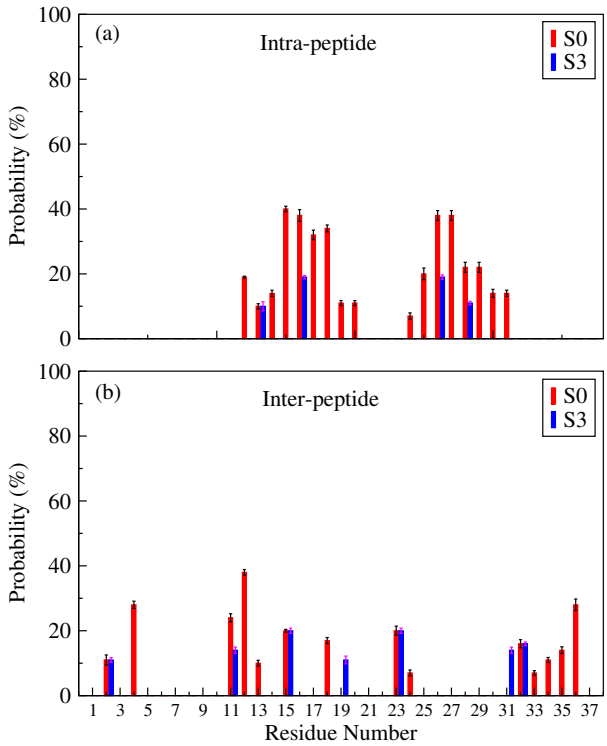

Figure 5-3. *Probability of residue-wise (a) intra-peptide and (b) inter-peptide β-sheet formation in systems S0 and S3. Standard errors are calculated using block average method.*

Also, we have presented the contour density map of peptides around one particular peptide for system S0 and S3 (see Figure 5-5), considering the last 100 ns average pdb of the simulation and a cut-off distance of 5 Å which shows that the density of peptides around one peptide is significantly low in system S3 than that of system S0.

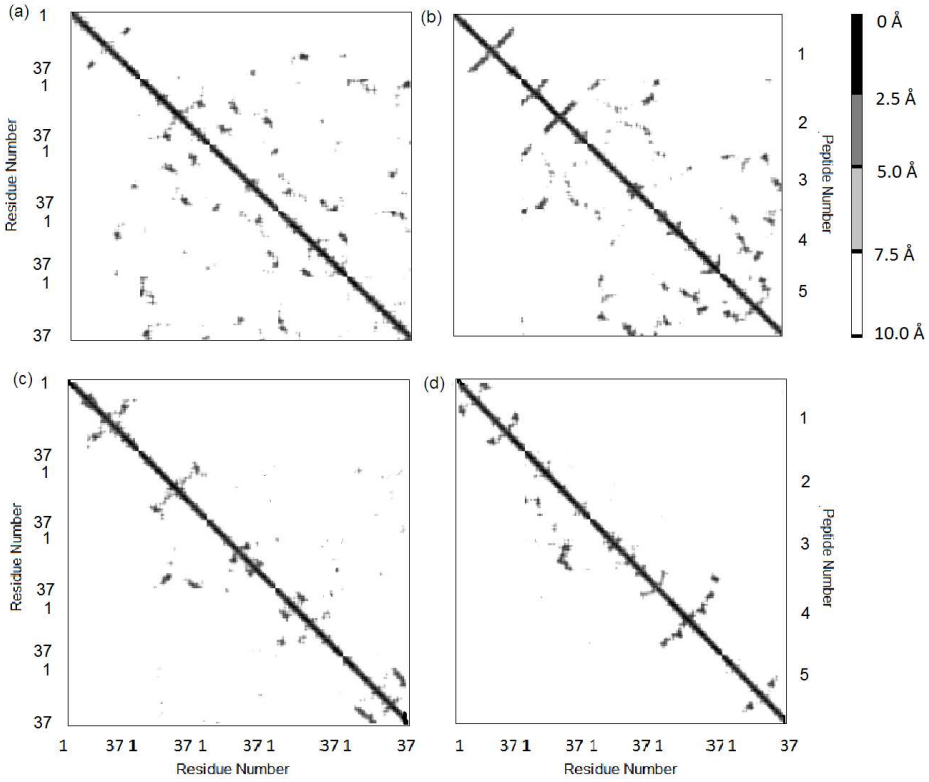

Figure 5-4. *Residue-residue contact map between the peptides. (a), (b), (c), and (d) represent systems S0, S1, S2, and S3 respectively.*

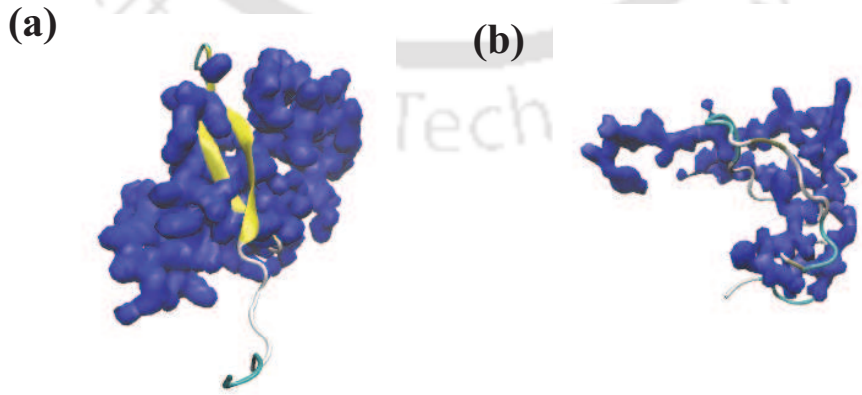

Figure 5-5. *Contour density plot of peptides around a peptide molecule for systems (a) S0, and (b) S3.*

Next, we discuss the secondary structural contents of different monomeric forms of the hIAPP peptide. For that, we have calculated the residue-specific secondary structures of the peptide monomers using the STRIDE algorithm [283] implemented in the VMD software. The changes in the secondary structure of 5 peptides as time progresses throughout the whole trajectory of the simulation (500 ns) are displayed in Figure 5-6. It should be mentioned that the secondary structure of the peptide is force field dependent. Liang et al. showed that the oligomeric conformational study of hIAPP peptide using CHARMM general force field in modified TIP3P water gives reasonably similar results to that of the experimental data, that is hIAPP monomer remains in a disordered state in solution, and oligomers of hIAPP consisting β-sheet structure such as trimer, tetramer, and pentamer are highly stable in water [280, 284]. In part a of Figure 5-6, the significant occurrence of yellow color denotes that system S0 contains a high amount of extended β-sheet conformation, and it remains intact up to the end of the simulation. It implies that in pure water hIAPP monomers aggregate to form ordered β-sheet structure from random coil conformation which is in good accordance with the experimental data. Further, from part b-d of Figure 5-6, we can observe that with the addition of COS, the yellow-colored portion becomes lesser, and in the 1.07 M COS solution the β-sheet formation becomes almost negligible. These results suggest that COS suppresses the formation of ordered secondary structure from random coil conformation of amylin, which is also observed experimentally [85].

Inter Peptide and Peptide-COS Interaction Energy

From the above discussion, it is clear that the increasing concentration of COS makes peptide-peptide interactions unfavorable. To study the effect of COS on the inter-peptide interactions of hIAPP, we have decomposed both the total peptide-peptide and peptide-COS interaction energies into van der Waals (vdW) and electrostatic energy components, and they are incorporated in Table 5-3. It is to be noted here that the electrostatic energy component dominates over the vdW energy values for both peptide-peptide and peptide-COS interactions. With increasing concentration of COS, both the inter-peptide electrostatic and vdW energies become more positive, indicating the less favorable peptide-peptide interaction in presence of COS. On the other hand, the peptide-COS interaction

energies follow the opposite trend to that of peptide-peptide interaction energy. In brief, with increasing concentration of COS, the peptide-COS interaction energies (both electrostatic and vdW) become more and more negative which means with the addition of COS the peptide-COS interaction becomes more favorable. It should be remembered that like inter-peptide interaction, peptide-COS interaction is also dominated by electrostatic interaction energy. Further, we have calculated the COS-COS interaction energies for systems S1, S2, and S3 and found that both the vdW and electrostatic interaction energies between COS molecules are very much unfavorable in comparison to peptide-COS interaction energies, which clearly indicates that COS preferably interacts with peptide molecules rather than self-interaction.

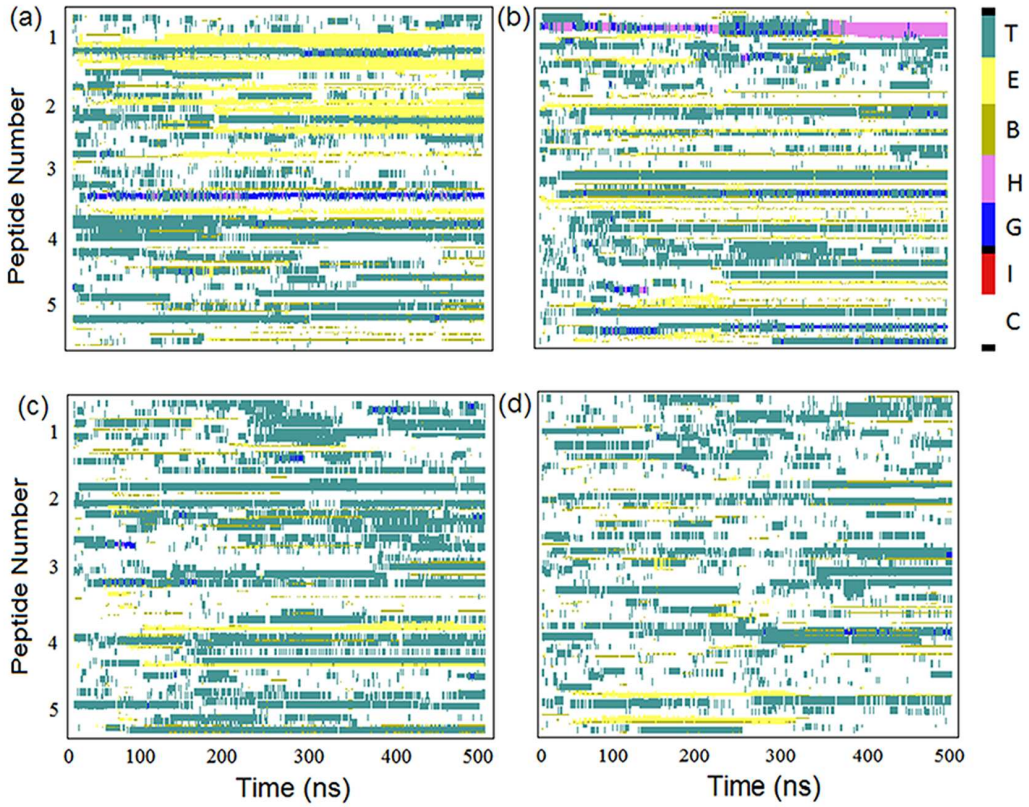

Figure 5-6. *Secondary structure analysis with simulation time. (a), (b), (c), and (d) represent systems S0, S1, S2, and S3, respectively. In the color bar T=Turn, E=Extended β-sheet, B=Isolated bridge-beta, H=α-helix, G=3_{10}-helix, I=π-helix, C=Coil.*

Table 5-3. Interaction energies[a]

System	S0	S1	S2	S3
PP (vdW)	-90.64 (±0.87)	-93.56 (±0.69)	-78.72 (±0.44)	-67.14 (±0.61)
PP (elec)	-456.80 (±1.31)	-435.68 (±1.95)	-421.86 (±1.01)	-393.09 (±1.36)
PC (vdW)	—	-11.74 (±0.17)	-59.62 (±0.19)	-91.84 (±0.16)
PC (elec)	—	-48.29 (±0.54)	-115.59 (±0.28)	-202.25 (±0.91)
CC (vdW)	—	+0.77 (±0.0007)	+0.32 (±0.0002)	-0.59 (±0.00012)
CC (elec)	—	-0.29 (±0.00016)	-2.83 (±0.0011)	-7.47 (±0.023)

[a] Decomposition of total peptide-peptide (per peptide), peptide-COS (per peptide), and COS-COS (per COS) interaction energies into average van der Waals (vdW) and electrostatic (elec) energy components. P and C represent the peptide and COS, respectively. Standard errors are calculated using block average method and provided in the parenthesis. All the energy values are expressed in kcal/mol unit.

Peptide Solvation

From the above discussion, it is observed that hIAPP peptides aggregate in pure water to form ordered β-sheet structure, and here we found residues 11-20 and 23-36 are mainly responsible for this β-sheet formation. The intra-peptide as well as inter-peptide radial distribution functions (see Figure 5-7) show that in pure water system, the inter-peptide and intra-peptide interactions both are very strong. The addition of COS molecules into the systems causes a reduction in these interactions.

To investigate which residues are mainly involved in β-sheet formation, we have calculated the pair correlation functions involving different residues of the peptide. The radial distribution functions between F15 and A25 residues of different systems are shown in the part a of Figure 5-8. The RDF plot in system S0 shows a strong peak at 6.5 Å. The hydrophobic F15 residue strongly interacts with the hydrophobic A25 residue leading to the aggregation. For system S1, in presence of 0.2 M COS, the peak height of the RDF is reduced significantly, and increasing the concentration of COS causes a depletion of the peak height further.

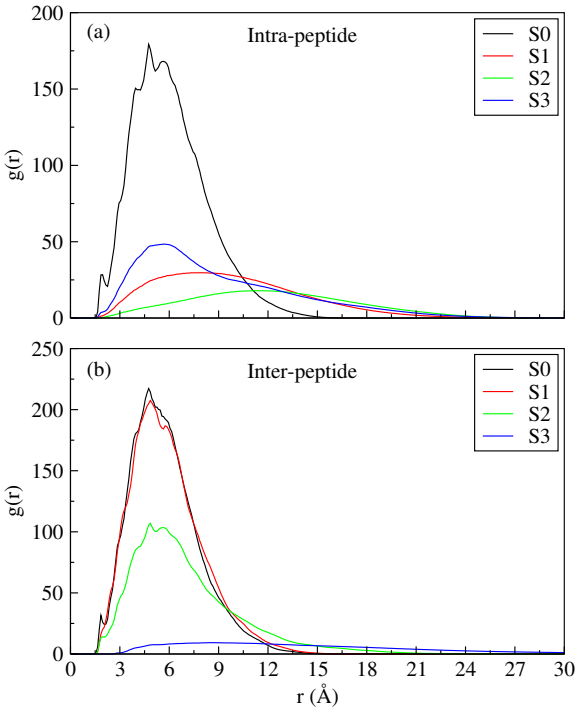

Figure 5-7. *Site-site radial distribution functions between (a) intra-peptide and (b) inter-peptide for systems S0, S1, S2, and S3.*

Nameki et al. in their experimental study suggested that may be COS efficiently interacts with the amide group of the peptide backbone, and prevents interaction between the backbone moieties which reduces the formation of β-sheet and subsequent amyloid formation [85]. They have also observed that COS is a more effective inhibitor than glycine betaine, acetylcholine, carnitine which have the same quaternary amine group, suggesting that the sulfate group of COS plays an important role in the prevention of self-association of hIAPP. So, to study the atomistic interaction of COS with peptide, we have calculated the radial distribution function between the amide nitrogen atom and the sulfate oxygen of COS (atoms O10-12) for different systems, and these are shown in part b of Figure 5-8. The appearance of a sharp first peak at 2.85 Å is quite apparent for all the three systems containing COS (systems S1, S2, and S3), and the height of this peak is the highest for the system containing the highest concentration of COS (system S3). This sharp peak at 2.85 Å indicates the presence of hydrogen bonding interaction between the amide group of

the peptide backbone and the sulfate group of COS, which may play an important role to inhibit the aggregation of amylin.

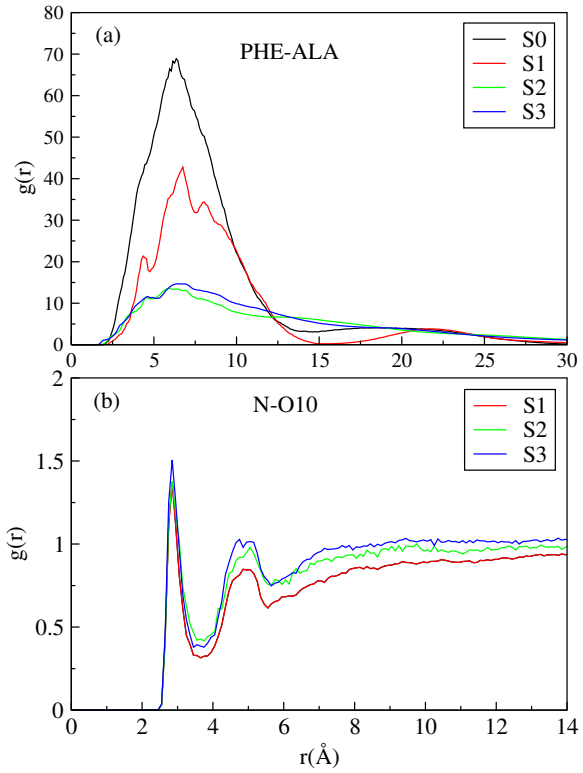

Figure 5-8. *Site-site radial distribution functions between (a) Phe and Ala residues of the hIAPP peptide for systems S0, S1, S2, and S3; (b) amide-nitrogen atom of the peptide backbone and sulfate-oxygen atom of choline-O-sulfate for systems S1, S2, and S3.*

By considering the last 100 ns of the simulation, we calculate the average number of first shell COS molecules around each residue of the peptide. (see Figure 5-9) using eq. 2.2 of chapter 2. A cut-off distance of 3.5 Å is used to calculate the number of COS in the first solvation shell of the peptide. We can see the presence of a higher coordination number of COS molecules around the residues which are prone to β-sheet formation in pure water i.e., residues 2, 4, 11-20, 23-36. The coordination number of COS around the hydrophobic residues like L12, F15, F23, I26, and V32 is very high due to the solvation of hydrophobic moieties by COS molecules. It is also clear that with increasing the concentration of COS,

the number of COS molecules around the residues of the peptides increases, indicating a stronger interaction between peptide and COS at higher COS concentration. From the contour density plot of COS within 3.5 Å of the peptide molecules (see Figure 5-10), we can clearly visualize the increasing density of COS molecules around peptides with increasing concentration of COS.

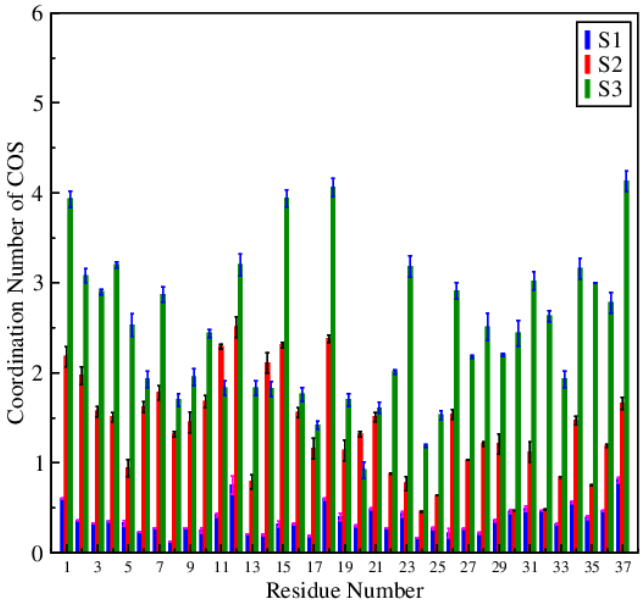

Figure 5-9. *First shell coordination number of COS around each residue of the peptide for systems S1, S2, and S3. Standard errors are calculated using block average method.*

Hydrogen bond properties

Moreover, it is a well-known fact that the inter-peptide hydrogen bonding interaction plays an important role in the formation and stabilization of β-sheet structure of the peptide oligomers. To shed light on the role of hydrogen bonding interaction in the formation of ordered secondary structure in different solutions, we have calculated the average inter-peptide as well as peptide-COS hydrogen bond numbers (per peptide) considering last 100 ns simulation for all the systems, listed in Table 5-4. To calculate the hydrogen bond numbers, a maximum distance of 3.5 Å between the donor (D) and acceptor (A) and, a D-H-A angle of less than or equals to 45° is considered [282]. Here the polar atoms i.e.,

N, O, and S are considered as the donor and acceptor atoms. The number of inter-peptide hydrogen bonds in system S0 is found to be 14.92 (per peptide). As the concentration of COS is increased, a reduction in the peptide-peptide hydrogen bond number is observed, whereas the number of peptide-COS hydrogen bond increases.

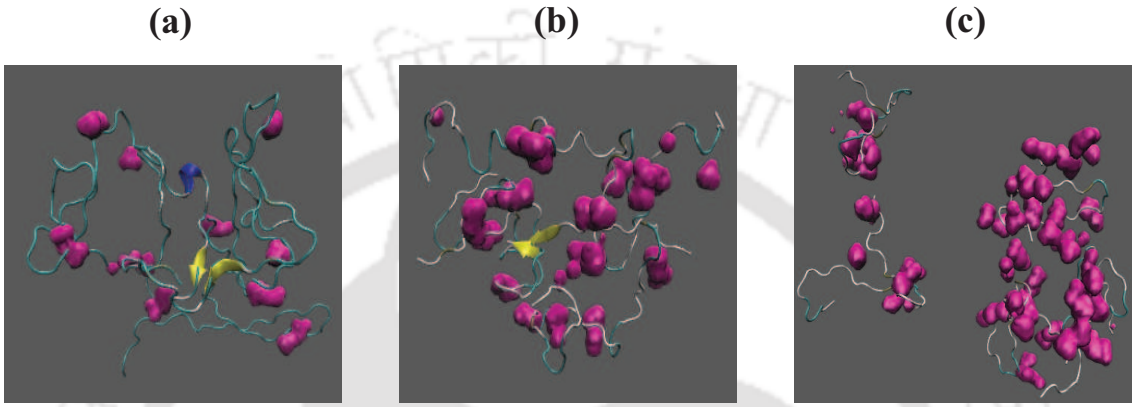

Figure 5-10. *Contour density plot of COS molecules (pink) within 3.5 Å of the peptide surface for systems (a) S1, (b) S2, and (c) S3.*

Table 5-4. Hydrogen bond numbers[a]

System	S0	S1	S2	S3
HB_{PP}(Backbone)	7.35 (±0.017)	4.80 (±0.011)	3.55 (±0.017)	4.36 (±0.011)
HB_{PP}(Side-chain)	7.57 (±0.023)	9.55 (±0.021)	8.77 (±0.027)	6.27 (±0.019)
HB_{PP}(Total)	14.92 (±0.026)	14.35 (±0.037)	12.32 (±0.019)	10.63 (±0.022)
HB_{PC}(Backbone)	—	0.47 (±0.006)	1.72 (±0.009)	3.25 (±0.016)
HB_{PC}(Side-chain)	—	0.46 (±0.006)	1.60 (±0.013)	2.35 (±0.020)
HB_{PC}(Total)	—	0.93 (±0.008)	3.32 (±0.011)	5.60 (±0.031)

[a] Average hydrogen bond numbers between inter-peptide backbones, inter-peptide side-chains, total inter-peptide backbone and side-chains (per peptide), and peptide backbone-COS, peptide side-chain-COS, total peptide backbone and side-chain-COS (per peptide) for systems S0, S1, S2, and S3. Standard errors are calculated using block average method and provided in the parenthesis.

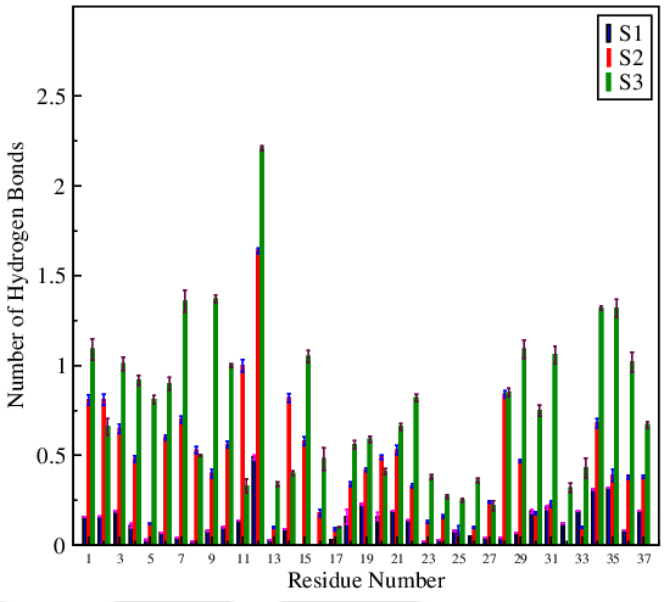

Figure 5-11. *Residue-wise hydrogen bond numbers of the peptide with COS (per peptide) for systems S1, S2, and S3. Standard errors are calculated using block average method.*

To examine the residue-based peptide-COS hydrogen bond interaction, we have extended our study towards the number of peptide-COS hydrogen bond calculation for different amino acid residues of amylin (see Figure 5-11). We find that residue L12 has the highest number of hydrogen bonds with COS in system S3. It is to be noted here that, from Figure 5-2, we can see the residue L12 has the second highest probability of beta-sheet formation in pure water which totally vanishes in presence of 1.07 M COS in system S3. Other residues like residue 1-11, 14, 15, 22, 28-37 also have very high hydrogen bonding interaction with COS which means along with the polar residues, COS gets involved in the strong hydrogen bonding interaction with hydrophobic residues like L12 and F15. Also, It is worth mentioning here that with increasing concentration of COS, the number of hydrogen bonds formed between COS and different residues of hIAPP also increases significantly. By comparing Figures 5-2 and 5-11, we also find that the residues 16-18, 24-27 strongly participate in β-sheet formation in pure water which is significantly reduced in presence of COS molecules, but they do not have stronger hydrogen bond interaction with COS. Thus, it is safe to state that for these residues the hydrogen bonding interactions with

COS do not play any role in the inhibition of fibril formation. Nevertheless, in Figure 5-9, the presence of a significant number of COS molecules in the first solvation shell of these residues indicates COS-induced hydrophobic solvation of them helps to reduce the β-sheet formation of these regions. The presence of a significant amount of peptide-COS vdW interaction energy also supports the presence of hydrophobic interaction between the hIAPP peptide and COS molecules.

Potential of Mean Force

To provide more clear evidence on the inhibitory effect of COS on peptide aggregation, we have calculated the free energies of association between two peptides for different systems (P0, P1, P2, and P3) in terms of their potential of mean forces (see Figure 5-12). To calculate the PMF, we have used the adaptive biasing force (ABF) method implemented in NAMD [143, 144, 145]. In the ABF protocol, an external biasing force is continuously applied to a set of atoms (here the full peptide) along a specific reaction coordinate permitting the system to overcome sufficiently large potential energy barriers, which cannot be passed by the system within computationally affordable time scales. The biasing force is estimated locally at a certain temperature from the extensive sampling of the system conformations at each step and is simultaneously updated [286]. To calculate the free energies to see the stability of the peptide dimer in each system, we have chosen the three-dimensional distances between the center of masses of the two peptides as the reaction coordinate which ranges from 3 to 25 Å. The reaction coordinate is further segmented into several equally spaced windows of 1 Å width, and each of these windows is simulated for 10 ns at 300 K. The initial configurations for all the windows are generated from the 10 ns ABF simulations of the previous window. Each window is further divided into small bins of 0.2 Å width, and the ensemble averages of the adaptive biasing forces are calculated only if 4000 samples are collected in each of the bins. A harmonic force of 100 kcal/mol/Å is applied to the upper and lower boundaries of the reaction coordinate of that window, allowing the solute particles to be bound within given limits. To generate the PMF between two peptides, one peptide is harmonically restrained and the other is set free.

In pure water system (system P0) the PMF is most negative (-40 kcal/mol), and

the global minimum is observed at a distance of 5 Å which means the dimer is most stable in pure water when they are 5 Å apart from each other. Also, it contains a well defined second and third minimum at a distance of 7 and 10 Å . In system P1 where the concentration of COS is $\sim$ 0.2 M, the PMF is less than that of system P0 (-30 kcal/mol), but it is still enough favorable which means 0.2 M COS is not so effective to reduce the stability of the dimer. In systems P2 and P3, the PMFs increase significantly (towards more positive value), and the global minima are also shifted to a larger distance, indicating that in presence of higher concentration of COS the peptides get well separated, and the possibility of dimer formation becomes less.

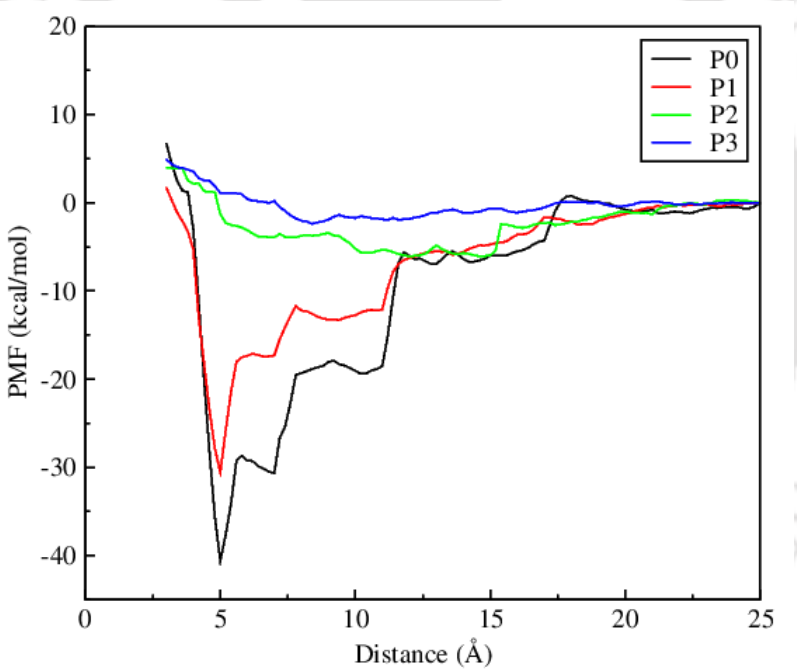

Figure 5-12. *PMF plots (in kcal/mol) between two peptides for systems P0, P1, P2, and P3.*

Destabilization of oligomerized hIAPP by COS

To study the inhibitory effect of COS on preformed protofibril of hIAPP, we have prepared a system taking the oligomerized structure of hIAPP, formed after the 500 ns simulation run of the pure water system (S0), and ∼1 M COS solution was added to it. A 1000 ns simulation run was carried out following the same simulation protocol described in the Models and Simulation method section. The average percentages of total β-sheet and coil conformations of each 200 ns time interval (see Figure 5-13) shows that the probability of β-sheet formation decreases with time, and coil conformation increases gradually. Also, with the progression of simulation time, the inter-peptide H-bond numbers decrease with increasing peptide-COS hydrogen bond numbers (see Figure 5-14). The secondary structure analysis with simulation time (see Figure 5-15) indicates that 1.07 M COS breaks the β-strand of the fibrillated hIAPP.

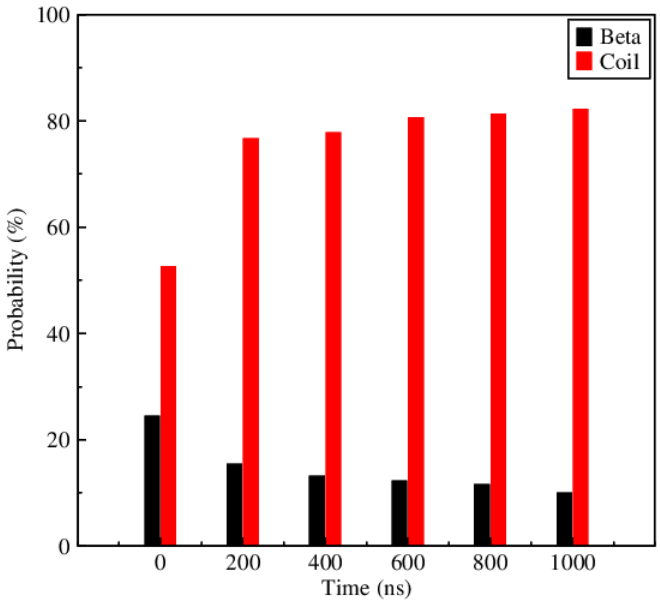

Figure 5-13. *Average percentages of β-sheet and coil conformations at 200 ns time interval. Data presented at 0 ns corresponds to the percentages of β-sheet and coil conformations of the initial fibril structure of the oligomerized peptide.*

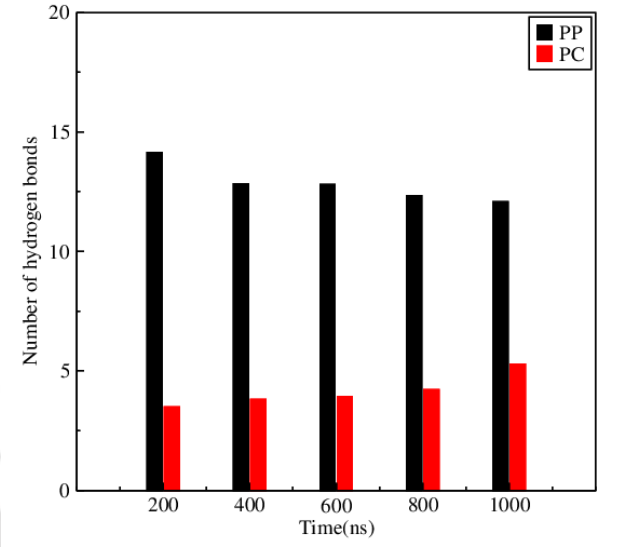

Figure 5-14. *Average inter-peptide and peptide-COS (per peptide) H-bond numbers at 200 ns time interval.*

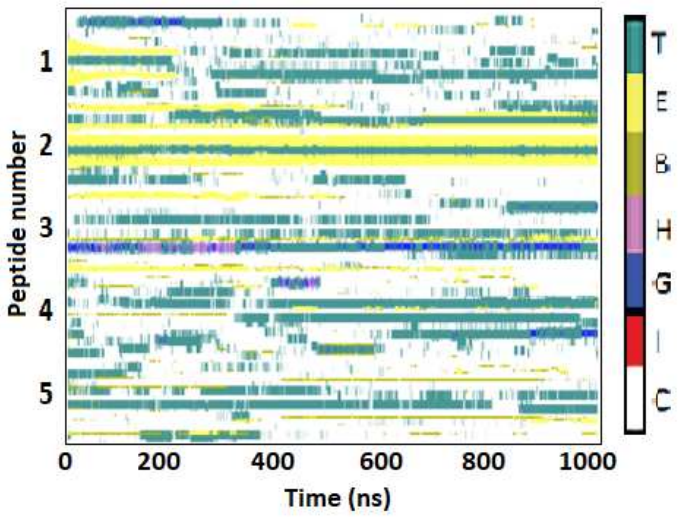

Figure 5-15. *Secondary structure analysis with simulation time for system containing oligomerized hIAPP. In the color bar T=Turn, E=Extended β-sheet, B=Isolated bridge beta, H=α-helix, G=3_{10}-helix, I=π-helix, C=Coil.*

■ SUMMARY AND CONCLUSIONS

The accumulation and deposition of amylin in the pancreas is an established cause of type II diabetes mellitus. A previous experimental study showed that COS, a small ester-type molecule, can prevent the aggregation of this amyloid peptide, hence, inhibits the β-sheet formation [85]. In this study, we have investigated the influence of COS on the full-length hIAPP peptide by performing 500 ns MD simulations. Four systems have been considered here among which one system contains hIAPP in pure water, which shows hIAPP forms ordered secondary structure in pure water. Previous experimental and theoretical studies reported different regions of amylin responsible for β-sheet formation. Here, from the pure water simulation, we have identified that mainly residues 11-20 and 23-36 participate in the formation of β-sheet structure. Further, we have considered three more systems with 0.2 M, 0.5 M, and 1.07 M COS; and noticed that by increasing the concentration of COS the presence of ordered secondary structure is reduced, and at the highest concentration of COS it vanishes totally. To demonstrate the inhibition of β-sheet formation in presence of COS, we have done several secondary structure analysis like percentage of β-sheet calculation, residue-residue contact map, and VMD secondary structure analysis with simulation time. We find that the presence of ordered β-sheet structure decreases gradually with the increasing concentration of COS in the systems.

To investigate in detail, we have plotted residue-wise β-sheet formation for the pure water system and system with the highest concentration of COS. From this plot, we observe that the residues 11-20 and 23-36 are the most amyloidogenic region of the peptide, but in 1.07 M COS, the β-sheet formation in some residues like 12, 14, 17-18, 24-27, 29-30, and 33-36 becomes totally zero. From the intra-peptide and inter-peptide β-sheet percentage calculations, we have found the intra-peptide interactions are more suppressed by the COS molecules than inter-peptide ones. To know how COS interacts with the peptide, we have plotted the site-site radial distribution functions between the backbone amide nitrogen and the sulfate oxygen of COS for systems S1, S2, and S3. This plot contains a sharp peak at 2.85 Å clearly indicating the presence of strong hydrogen bonding interaction between them. Furthermore, we have shown the coordination number of COS around each residue of the peptides within 3.5 Å and observed a higher coordination number of COS molecules

around different residues in system S3.

The inter-peptide and peptide-COS interaction energies (both electrostatic and vdW energy components), hydrogen bond numbers (per peptide), as well as residue-wise hydrogen bond numbers of the peptide with COS, depict that the inter-peptide hydrogen bond number and interaction energy are the highest in the pure water system which is reduced in presence of COS. On the other hand, the peptide-COS hydrogen bond numbers as well as interaction energies increase gradually with increasing the concentration of COS. From the residue-wise hydrogen bond number plot, we find that, residue L12, which has the second highest probability of β-sheet formation has the strongest hydrogen bonding interaction with COS. Other residues which participate in aggregation like residues 1-11, 14-15, 22, 28-37 also have a significant number of hydrogen bonds with COS due to which the probability of β-sheet formation of these residues becomes lower in presence of COS. Interestingly, we observe that the presence of β-sheet structure in residues 16-18 and 24-27 is reduced significantly in system S3, though they do not have strong hydrogen bond interaction with COS. Nevertheless, the coordination number of COS around these residues is very high which leads us to propose that, hydrophobic solvation of these residues by COS inhibits the aggregation of these regions. Lastly, from the PMF calculation, we observe that the free energy of separation between two peptides is most favorable in absence of COS, and with increasing concentration of COS it becomes significantly unfavorable, indicating that the dimer formation is most stable in pure water which becomes less favorable in presence of COS. Moreover, the destabilization of the preformed oligomer of hIAPP in 1.07 M COS solution confirms the inhibition of the amyloid aggregation in presence of COS.

Thus, from this study, we can conclude that hIAPP peptides aggregate in pure water to form ordered secondary structure, and residues 11-20 and 23-36 are the most amyloidogenic region of this peptide. The hydrogen bonding interaction between the sulfate oxygen of COS and the backbone amide nitrogen of the peptide, as well as the hydrophobic solvation of the side-chains of some residues by COS molecules, play the most important role in the inhibition of the amyloid aggregation of hIAPP.

Chapter 6

Inhibitory Effects of Choline-*O*-sulfate on the Aβ_{16-22} Peptide Aggregation

"Amyloidoses thus include some of the most feared and costly diseases in the Western world. Alzheimer's may very soon be the most prevalent and socially disruptive illness in the aging populations of all developed countries."

– C. M. Dobson. *Nature* **2002**, *418*, 729-730.

Overview

The aggregation of Aβ_{16-22} peptide, the smallest fragment of full-length Aβ_{1-42} with seven residues, plays a very crucial role in the Aβ toxicity, hence, causes Alzheimer's disease. Alzheimer's disease (AD) is a progressive neurodegenerative disease associated with brain disorder, and currently, there is no treatment available in the market to cure it permanently. So, drug designing for the curable treatment of AD is a very challenging problem. In this chapter, we have investigated the inhibitory effect of choline-*O*-sulfate (COS) in the oligomerization of Aβ_{16-22} peptide. We have carried out a series of classical simulations using two different types of force field parameters namely CGenFF and GAFF and found that this small seven residue peptide aggregates in pure water to form fibrils, but in presence of the inhibitor COS the formation of β-sheet structure is prevented. Residue-wise secondary structure analysis shows that the hydrophobic core of the peptide LVFFA contains a high percentage of ordered β-sheet conformation in pure water which becomes negligible with the addition of COS. Different types of interaction energies, radial distribution functions, coordination number calculation, and hydrogen bond analysis show that COS interacts with the peptide backbone nitrogen through hydrogen bonding as well as the solvation of peptide hydrophobic residues by it play the major role to prevent the peptide aggregation in water. It is worth to mention that the results obtained from both of the force field parameters draw very similar conclusions. In this chapter, we have also shown that COS can disaggregate the preformed 5-stranded protofibril of Aβ_{16-22} at a concentration of 0.95 M.

■ INTRODUCTION

Alzheimer's disease (AD) is a neurodegenerative disease that initiates slowly and becomes dreadful with time [43, 44, 45, 46, 47]. It is associated with brain disorder and 60-70 % cases of dementia are caused by it. Statistics say that, in 2015, approximately 29.8 million people worldwide were affected by AD among which 1.9 million people died [287]. In current years, there is no curable treatment available for AD, but treatments for its symptoms are available and research is going on in this field. At the molecular level, two types of peptide aggregation in the brain is associated with Alzheimer's disease (i) extracellular amyloid plaques (Aβ peptides), and (ii) intracellular neurofibrillary tangles (tau proteins) [288, 289, 290]. Comparative studies revealed that the aggregation of Aβ peptides is the major cause of this neurodegeneration. So, the inhibition of this amyloid aggregation is marked as the best therapeutic strategy for the treatment of AD [291]. Though all drugs targeting Aβ have failed till now due to several reasons, like high concentration of peptide in *in vitro* compared to *in vivo*, specific interactions of drugs with other molecules, involvement of other proteins in AD apart from Aβ-peptide, etc [292]. Generally Aβ peptide consists of 39 to 43 amino acid residues among which Aβ_{40} and Aβ_{42} are the predominant species *in vivo*. Among the soluble oligomers of amyloid β-peptides, Aβ_{1-42} is considered as a major biological component for Alzheimer's disease in body fluids, such as, serum and cerebrospinal fluid (CSF) [46, 293].

The *in silico* study of full-length Aβ_{1-42} will cost huge computationally. Several short sequences of full-length β-amyloid peptide have been reported by solid-state NMR study which also can form amyloid fibrils in isolation. The solid-state NMR measurements by Robert Tycko and his group showed that the seven-residue Aβ_{16-22} peptide with the sequence N-acetyl-KLVFFAE-NH_2 participates in highly ordered fibril formation, and they suggested that it is the shortest fibril-forming fragment of full-length Aβ peptide which can be used as the model system for physical studies of amyloid fibril formation [294]. We have chosen Aβ_{16-22} as our model peptide due to the following reasons i) Computational study of Aβ monomeric peptide in explicit solvent is feasible but there is convergence issues reported on the full-length peptide [295, 296, 297], ii) It is an ideal model system for the computational study of the amyloid aggregation and fibril formation due to it's

shorter length [17, 282, 298, 299, 300, 301, 302, 303, 304], and iii) It contains the central hydrophobic core of the full-length Aβ peptide that is LVFFA which is identified as the major part of amyloid aggregation and forms anti-parallel β-sheet structure, and also, it is considered as the primary target of the inhibitors used for preventing the aggregation of Aβ-peptide [305, 306].

We have already mentioned that the current treatments of Alzheimer's disease cannot stop its progression, rather, they can temporarily slow down the worsening of dementia symptoms. Today, there is a worldwide effort to find better treatment for this acute disease by preventing its development in the human brain. Researchers found that inhibition of the amyloid aggregation can be the most effective therapeutic strategy for the treatment of Alzheimer's disease [282, 298, 307, 308, 309, 310, 311, 312, 313]. There are several types of inhibitors which are found capable of inhibiting Aβ amyloid aggregation such as, protein like drugs i.e., N-methylated peptides or D-amino acid peptides [314, 315, 316], small molecules, and polyphenols such as BAF31, O4, EGCG, carnosine, Taxifolin, NQTrp, caffeine etc [282, 317, 318, 319, 320, 321, 322, 323]. Recently, one research group has reported norepinephrine as a good inhibitor on amyloid aggregation of Aβ_{1-42} [324].

There are three different ways by which these molecules show their inhibitory effects on amyloid aggregation i) The inhibitors bind to the oligomers to limit the fibril fragmentation, and hence, reduce its toxicity, ii) The inhibitor molecules promote the formation of larger fibrils, and reduce the lifetime of toxic oligomers, and iii) The small molecules interact with the amyloid peptides and prevent aggregation. The third mode of interaction is thought to be the most effective and non-toxic pathway for the inhibition of fibril formation [325].

Though no experimental work is reported proving the inhibitory effect of COS in the fibrillation of Aβ-peptide, it is reported as a good inhibitor in the aggregation of hIAPP peptide (an intrinsically disordered protein like Aβ), which is further supported by our theoretical study, discussed in the previous chapter (chapter 5) [85]. In fact, Hagihara et al. found in their experimental study that COS is a better inhibitor than other inhibitors like glycine betaine, carnitine, acetylcholine, NDSBs, etc, containing the same tertiary amino group. So, in this current chapter, we have shown the mechanism of COS-induced inhibition of the amyloid aggregation of Aβ_{16-22} using classical molecular dynamics simulation. What

follows in the rest of this chapter are a brief description of the models and simulation method, discussion of the results obtained, and finally, our concluding remarks, and a brief summary.

■ MODELS AND SIMULATION METHOD

Classical molecular dynamics (MD) simulations were performed to investigate the inhibitory effects of COS on the aggregation of Aβ_{16-22} peptide. For this purpose, three different systems namely S0, S1, and S2 were considered, and they are outlined in Table 6-1. S0 is the pure water system having five peptide molecules and water, whereas systems S1 and S2 contain five peptides, water, and COS with concentrations of 0.5 M and 0.95 M respectively. Initial configurations of the systems were prepared using PACKMOL in which the peptides were placed randomly in a cubic box [116]. All-atom CHARMM36 force field parameters were considered to model the peptide and CHARMM General force field was used for COS for these systems (see Table 2-2, chapter-2) [108, 109]. It may be noted that the choice of the protein force field and water model is an issue in simulating Intrinsically Disordered Proteins (IDPs) like hIAPP, Aβ. In a recent study, Garcia and co-workers showed that the radii of gyration of different IDPs, as obtained using the CHARMM force field agree well with experimental data [280].

The Aβ_{16-22} with the sequence of KLVFFAE was chosen as the initial structure of the peptide. The C-terminal and N-terminal of the peptide were capped using neutral acetyl and amide groups respectively. The mTIP3P model of water was employed for the simulations [111]. These simulations were carried out using NAMD 2.10 package [121]. At first, energy minimization was done by using the conjugate method followed by heating in canonical ensemble (NVT) [123]. Then 1 ns equilibration was done for all the systems using isothermal-isobaric (NPT) ensemble [124]. The pressure of the systems was fixed to 1 atm by using Langevin piston method [125]. Finally, a 500 ns production run was continued in NPT ensemble for all systems. The temperature of each of the system was controlled using Langevin dynamics method with a damping coefficient of 5 ps^{-1} [125]. An integration time step of 2 fs was used for all simulations, and the trajectories were saved after every 4 ps of the simulation. The short-range Lennard-Jones interactions were calculated using 12 Å cut-off distance, and 10 Å switch distance. Particle Mesh Ewald (PME) method

was employed to calculate the long-range electrostatic interactions [128]. All bonds that involve the hydrogen atoms were constrained using the SHAKE algorithm [129]. In order to examine if the results presented in this study are independent of the initial configurations, we have prepared two more pairs of systems namely S0a, S0b, and S2a, S2b, listed in Table 6-1. Note that the systems S0a and S0b resemble system S0 except that these systems are having different initial configurations. The same is true for systems S2, S2a, and S2b.

As we are not aware of any experimental evidence on the inhibitory effect of COS on Aβ_{16-22} peptide aggregation, and no ideal force field has been identified to study the protein oligomerization till now, to validate our results more accurately, we have checked the force field dependency of our work. We have prepared another two systems namely A0 and A1 (Table 6-1), where A0 is the pure water system, and A1 is the system with 0.94 M concentration of COS as in the system S2. Recently Carballo-Pacheco et al. did a comparative study among different force fields on the aggregation of Aβ_{16-22} peptide and proposed that CHARMM22 and AMBER99SB*ILDN follow almost similar kinetics of oligomer formation, and the most interacting region of the peptides are also same for these two force fields. They also showed that the results obtained from the comparative aggregation study between Aβ_{16-22} and mutant of Aβ_{16-22} (F19V,F20V) using CHARMM22 and AMBER99SB*ILDN are in line with the experimental observations. This group in their previous force field comparison study also showed that both CHARMM22 and AMBER99SB*ILDN force fields model IDPs such as Aβ_{42} accurately, and strongly suggested not to use Gromos54a7 and OPLS-AA force field parameters to study protein oligomer formation. Man et al. in their study on Amyloid-β_{1-42} dimers obtained very similar percentages of secondary structure for both CHARMM22 and AMBER14SB force field parameters [326, 327, 328]. So, following these works, we have considered ff14SB force field parameters for the peptide as our alternative force field model, and the TIP3P model of water was used which is the best combination with the AMBER protein force field [111, 207, 173]. GAFF parameters were employed for COS (see Table 2-3, chapter-2). These simulations were carried out using AMBER14 package [122]. The temperature was fixed at 300 K throughout the simulation for all the systems. The initial configurations of all the systems were generated using PACKMOL program [116]. For all simulations, at first, energy minimization was done for 10000 steps in which the first 5000 steps were carried out applying

the steepest descent method, and for the rest 5000 steps, conjugate gradient method was applied. Then the systems were heated up to 300 K starting from 0 K temperature at an interval of 50 K for 20 ps. After step-wise heating, all the systems were equilibrated for 5 ns at isothermal isobaric ensemble [124], and a final production run of 500 ns was continued to study different molecular properties of the systems. All molecules were placed in a cubic box. Periodic boundary conditions were applied in all three directions. Langevin dynamics method with 1 ps^{-1} collision frequency was applied to control the temperature for all the systems [127]. A constant pressure of 1 atm was maintained by using Berendsen barostat with a relaxation time of 2 ps [126]. Particle Mesh Ewald (PME) method was applied to estimate the long-ranged electrostatic interactions [128]. All covalent bonds containing hydrogen atoms were constrained by SHAKE algorithm [129]. For non-bonding short-ranged interactions, a cut off distance of 10 Å was used. Another two sets of simulations for each of the systems A0 and A1, namely A0a, A0b; and A1a, A1b were conducted with different initial configurations (see Table 6-1) to check the statistical significance of our results.

Visual Molecular Dynamics (VMD) package, TCL scripts, and CPPTRAJ were used for visualization and analysis purpose [130, 131]. To verify the convergence of our simulation, we have plotted the Root Mean Square Deviation (RMSD) of the C_α-carbon atoms of the peptide backbone with time, and 20 ns block average inter-peptide hydrogen bond numbers for systems S0 (for CGenFF) and A0 (for GAFF) and found that both the RMSD and inter-peptide average hydrogen bond numbers get well converged within 500 ns simulation time (see Figure 6-1).

Table 6-1. Overview of Systems[a]

System	N_P	N_C	N_W	Volume (nm^3)	M_C
S0	5	0	5000	154.21	0
S0a	5	0	5000	154.32	0
S0b	5	0	5000	154.42	0
S1	5	50	5000	164.61	0.50
S2	5	100	5000	174.46	0.95
S2a	5	100	5000	174.32	0.95
S2b	5	100	5000	174.87	0.95
A0	5	0	5000	157.90	0
A0a	5	0	5000	158.00	0
A0b	5	0	5000	158.06	0
A1	5	100	5000	177.00	0.94
A1a	5	100	5000	176.92	0.94
A1b	5	100	5000	176.96	0.94
P0	2	0	5000	151.16	0
P1	2	100	5000	172.28	0.96
U0	2	0	3000	93.71	0
U1	2	60	3000	104.80	0.95

[a] N_P, N_C, and N_W represent the number of peptide, choline-*O*-sulfate, and water molecules, respectively. M_C is the molar concentration of choline-*O*-sulfate for different systems.

■ RESULTS AND DISCUSSION

Secondary structure analysis in different solution

Previous theoretical studies reported that Aβ_{16-22} peptides generally acquire a random coil conformation with a very small amount of α helix structure. In pure water, it has a high tendency to aggregate and to form ordered β-sheet structure [282, 329]. To show the different conformational changes of Aβ_{16-22} peptide in different solutions, we have presented the snapshots of different systems at the beginning of the simulations, midway of the simulation, and at the final stage of the simulation for systems S0, S1, and S2 (see Figure

6-2). Initially, for all the systems the peptides were separated well from each other. As the simulation proceeds, in pure water, the peptides start to aggregate and form ordered β-sheet structure which remains intact till the end of the simulation. The snapshots of systems S1 and S2 show no β-sheet conformation at different stages of the simulation.

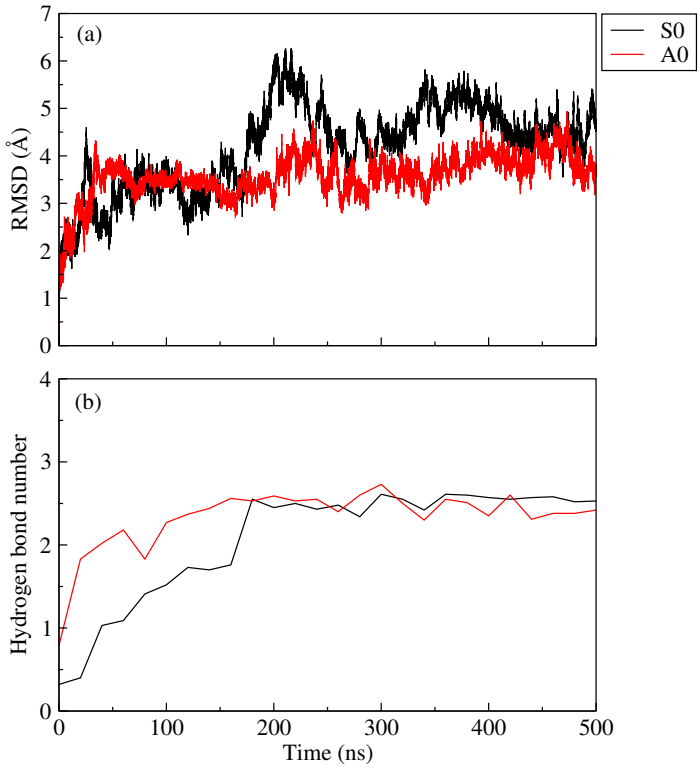

Figure 6-1. *(a) Root Mean Square Deviation (RMSD) of C_α-atoms of the peptide backbone and (b) average inter-peptide hydrogen bond numbers (per peptide) versus simulation time for systems S0 and A0.*

To visualize the conformational changes of the peptides in different solutions with GAFF, we have also considered the snapshots of the peptides at different time intervals for systems A0 and A1 (Figure 6-2) and found that, in system A0, the peptides aggregate with the progression of the simulation and form ordered β-sheet structure from the random coil conformation. In system A1, the peptides are well separated from each other till the end of the simulation keeping their coil conformation intact like that of system S2. It is to be noted here that the presentation of snapshots is a very preliminary analysis to show

the conformational changes. Thus, for strong evidence, we have further carried out several secondary structure analyses, such as percentage of secondary structure calculation, the probability of residue-wise β-sheet formation, secondary structure analysis with time, and residue-residue contact map of the peptides in different systems for both of CGenFF and GAFF parameters.

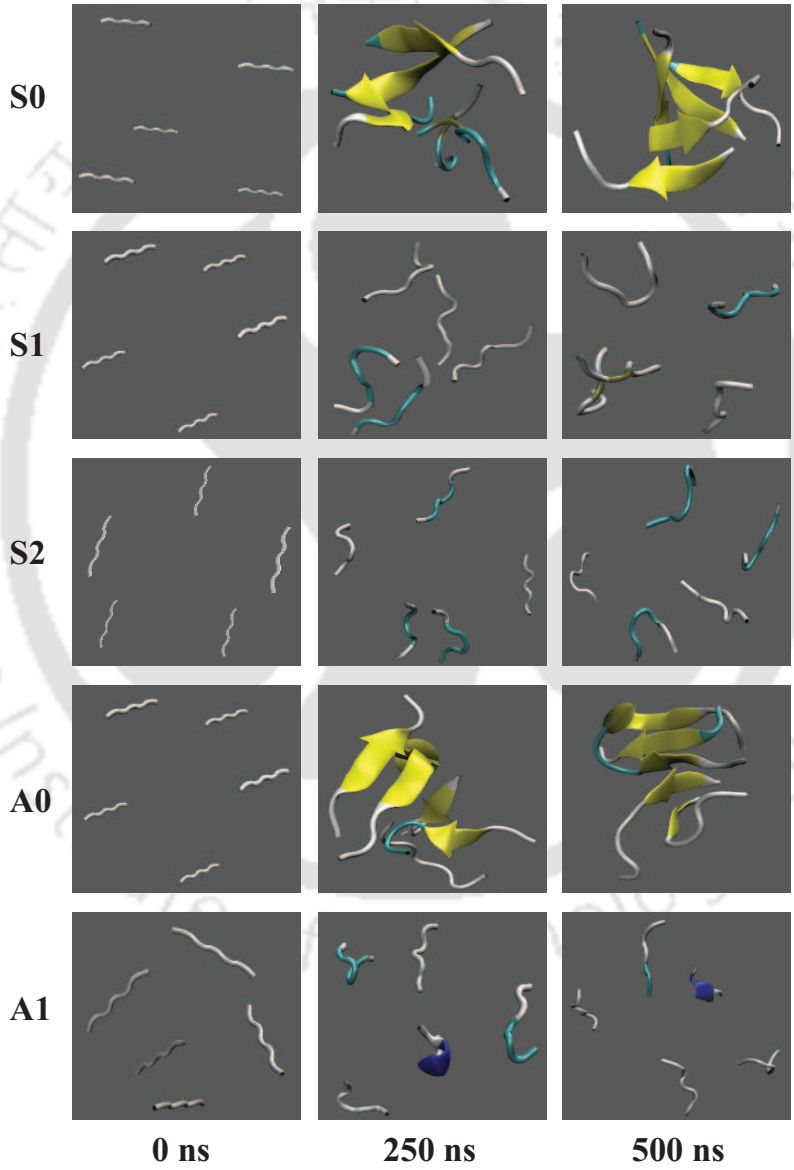

Figure 6-2. *Snapshots of the different systems at 0 ns, 250 ns, and 500 ns (from left to right). Water and COS molecules are left-off for better visual clarity.*

Next, we have calculated the percentages of different secondary structure conformations of the peptide, such as, extended β-sheet, bridge-beta, turn, coil, and helix for systems S0, S1, and S2; and they are presented in Table 6-2a. We can see that in pure water the percentage of β-sheet conformation is 39.09 %, and all these β-sheets are in anti-parallel arrangement. Zheng et al. in their recent experimental study found that in water at pH 7.3 and at room temperature, the full Aβ_{40} peptide contains 57 % of β-sheet conformation [329]. Though there are no experimental findings on β-sheet percentage available for Aβ_{16-22} peptide, as the structural properties of 16-22 residues of the Aβ peptide quite resemble with the Aβ_{40} peptide, so we can say that our simulation result is in agreement with the experimental value. As COS is added, the percentage of extended β-sheet conformation becomes zero, and only 6.46 % isolated beta-bridge conformation is found in system S1, which almost vanishes in system S2. It is also observed that the percentage of coil conformation increases significantly (from 51.93 to 93.86 %) as one moves from system S0 to S2 with decreasing β-sheet conformations of the systems, denoting that COS inhibits the peptides to form ordered secondary structure from random coil conformation.

Table 6-2a. Percentage of secondary structure conformations[a]

System	S0	S1	S2	A0	A1
% E	38.00 (±1.67)	0	0	28.23 (± 1.54)	0
% B	1.09 (± 0.29)	6.46 (±0.21)	0.71 (±0.0004)	2.30 (± 0.47)	2.51 (± 0.02)
% Total β	39.09 (±1.17)	6.46 (±0.21)	0.71 (±0.0004)	30.53 (± 1.29)	2.51 (± 0.02)
% Anti-para β	39.09	1.43	0	0.88	0.28
% parallel β	0	5.04	0.71	29.65	2.23
% H	0	0	0	0.52(± 0.007)	5.75 (± 0.18)
% C	51.93 (±1.11)	89.88 (±1.89)	93.86 (±1.50)	64.44 (± 1.31)	83.04 (± 2.27)
% T	8.98 (±0.23)	3.66 (±0.13)	5.43 (±0.17)	4.51 (± 0.19)	8.70 (± 0.22)

[a] percentages of extended β-sheet (E), isolated bridge-β (B), anti-parallel and parallel β-sheets, random coil (C), and turn (T) conformations for systems S0, S1, S2; and A0, A1. The standard errors are calculated using block average method.

To compare the CGenFF results with GAFF force field results, the percentages of secondary structure conformations of the Aβ_{16-22} peptides for systems A0 and A1 are also determined (Table 6-2a). It is noticed that the total β-sheet percentage in pure water is

30.53 % which is close to that of system S0, but here the difference is that the β-sheets are mostly in parallel arrangement. In 0.95 M COS solution, the percentage of extended β-sheet conformation becomes zero (like that of systems S1 and S2), and only 2.51 % bridge-beta conformation is present. Also, the helical and random coil conformations increase from 0.52 to 5.75 %, and 64.44 to 83.04 % respectively. This implies that both the CHARMM General Force Filed and the General AMBER Force Field produces very comparable results, which is also reported by Matthes et al [330].

In order to examine the robustness of our results, in Table 6-2b we compare the percentages of secondary structure conformations for system S0 with S0a and S0b, and system S2 with S2a and S2b, and similarly, for GAFF force field parameters, the percentages of secondary structure conformations for system A0 is compared with A0a and A0b; and A1 is compared with A1a and A1b. The closeness of these values implies that our results do not depend on the initial configurations of the systems.

Table 6-2b. Percentage of secondary structure conformations[a]

System	% E	% B	% Total β	% H	% C	% T
S0	38.00	1.09	39.09	0	51.93	8.98
S0a	32.11	5.31	37.42	11.04	46.22	5.32
S0b	34.33	0	34.33	0	64.47	1.20
S2	0	0.71	0.71	0	93.86	5.43
S2a	0	1.14	1.14	0	87.95	10.91
S2b	0	1.82	1.82	0	90.53	7.65
A0	28.23	2.30	30.53	0.52	64.44	4.51
A0a	24.62	6.62	31.24	1.34	60.85	6.57
A0b	21.70	4.06	25.76	13.85	49.20	11.19
A1	0	2.51	2.51	5.75	83.04	8.70
A1a	0	2.58	2.58	3.60	85.56	8.26
A1b	0	3.01	3.01	3.25	85.91	7.83

[a] Percentages of extended β-sheet (E), bridge-β (B), total β-sheet, helix (H), random coil (C), and turn (T) conformations for systems S0, S0a, S0b; S2, S2a, S2b; and A0, A0a, A0b; A1, A1a, A1b.

In the part a of Figure 6-3, we show the residue-wise total β-sheet formation of the peptide in different solutions. In system S0, residues 2-5 i.e., LVFF have very high probability of β-sheet formation, and residues 1 and 6 also participate in the β-sheet formation to some extend. In system S1, there is a very low probability of β-sheet formation for residues 3 to 6, whereas in system S2 the probability of β-sheet formation for each and every residue is negligible. Thus, we can conclude that the central hydrophobic core of Aβ_{16-22} that is LVFF has very high tendency to aggregate in pure water to form ordered secondary structure, but in presence of COS, they remain separated from each other due to favorable interaction of COS with the peptide (discussed below).

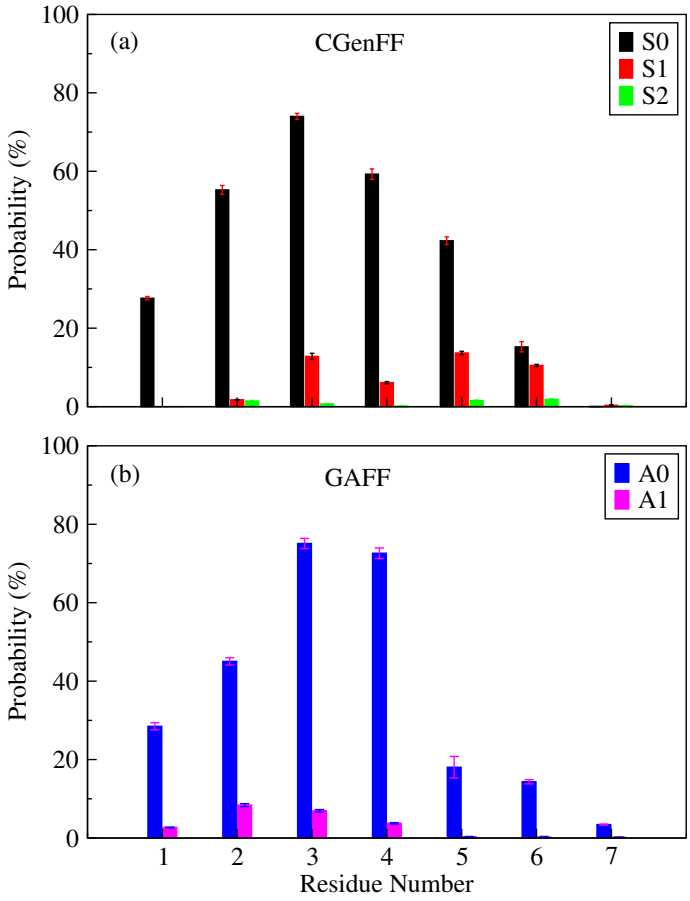

Figure 6-3. *Probability of residue-wise total β-sheet formation (extended β-sheet + bridge-β) for systems (a) S0, S1, and S2; and (b) A0 and A1. Standard errors are calculated using block average method.*

The probability of β-sheet formation for each residue of the Aβ_{16-22} peptide with AMBER force field parameters (see Figure 6-3, part b) reveals that in pure water residues 2 to 4 (LVF) have the highest probability of ordered structure formation as well as other residues like 1, 5, and 6 also participates in the β-sheet formation. On the other hand, in system A1, residues 1 to 4 have very negligible probabilities of β-sheet formation. So, comparing part a and b of Figure 6-3 we can find that, the results obtained from both CHARMM General and AMBER General force field confirm that the region 2 to 5 is the aggregation prone-area of the Aβ_{16-22} peptide and the presence of COS inhibits the aggregation. It was recently reported by Zou et al. that in a full-length Aβ_{1-40} peptide, residues 15-22 and 30-37 contain most populated β-hairpin conformation and residues 17-21 (residues 2-5 in our case) have the highest probability of β-sheet formation. Also, Carballo-Pacheco et al. found in their simulation study that with CHARMM22 and AMBER99SB*ILDN force field, the central PHE residue of the peptide that is F4 and the terminal residues show the strongest interaction probability in the aggregation of Aβ_{16-22} [324, 326].

To study the inhibitory effects of COS on the amyloid aggregation of Aβ_{16-22} peptide, we have further carried out secondary structure analysis of the peptides as the simulation progresses by using STRIDE algorithm implemented in VMD (see Figure 6-4) for both sets of the systems containing CGenFF and GAFF parameters. System S0 contains a rich amount of extended β-sheet as well as bridge-beta conformation, but in systems S1 and S2, there is no trace of extended β-sheet structure, and a little amount of bridge-beta formation is found in presence of COS. Similarly, system A0 also contains a rich amount of extended β-sheet conformation, whereas system A1 has a negligible amount of bridge-beta conformation and no extended β-sheet structure at all.

Next, we introduce the residue-residue contact map of the C_α atoms of the peptides in Figure 6-5, for both the CGenFF and GAFF systems. For both force field parameters, in the pure water system, the occurrence of black to grey color indicates the least distance between the peptides, leading to the aggregated structure. After incorporating the COS molecules into the systems, the contact map holds the maximum of white color, as in these systems the distance between the peptides is greater than 10 Å. Thus, both secondary structure analysis and residue-residue contact maps obtained from General AMBER Force Field are very much similar to that obtained using CHARMM General Force Field.

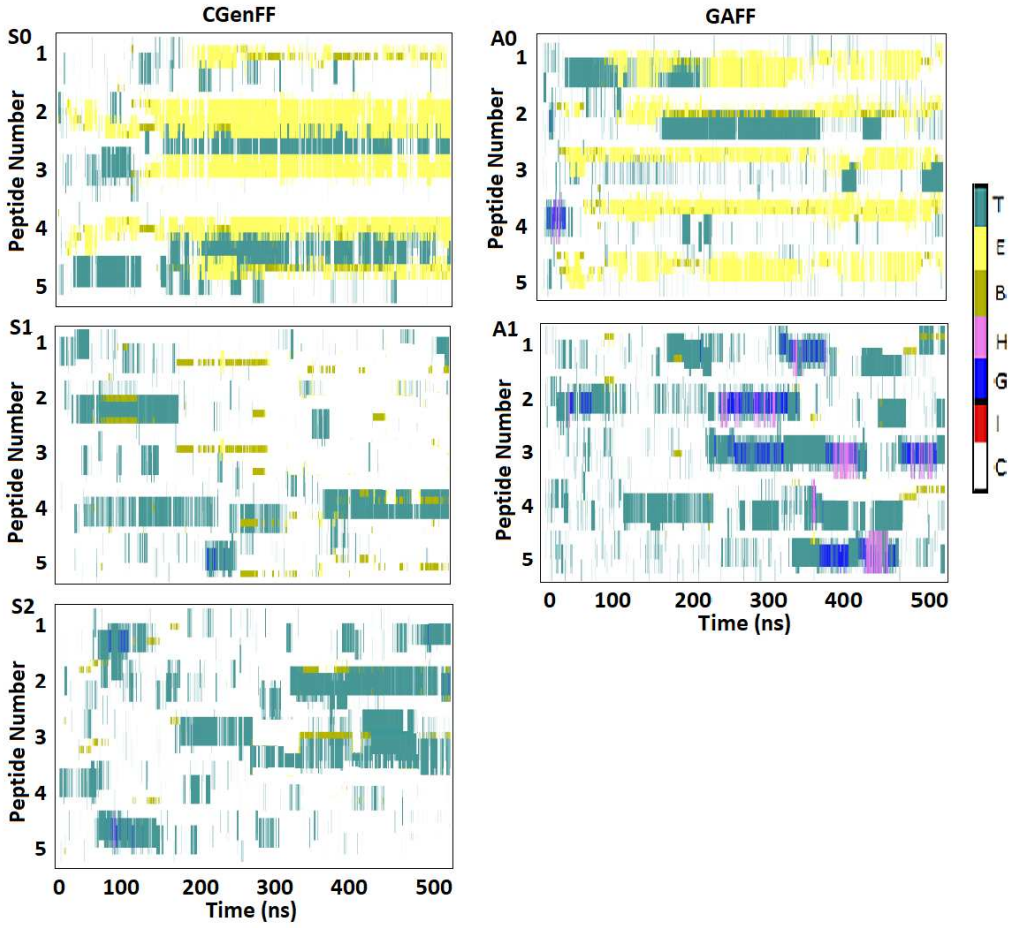

Figure 6-4. *Secondary structure analysis as the simulation progresses. In the color bar: T=Turn, E=Extended β-sheet, B=Isolated bridge-beta, H=α-helix, G=3_{10}-helix, I=π-helix, and C=Coil.*

Inter Peptide and Peptide-COS Interaction Energies

From the secondary structure analysis discussed above, we have come to a decision that in pure water Aβ_{16-22} peptides aggregate to form β-sheet structure, while COS molecules prevent their self-assembly in water. To investigate how does COS inhibit the self-association of Aβ_{16-22} peptides, we have calculated the interaction energies (both van der Waals and electrostatic) between the peptides as well as peptide and COS for systems S0, S1, and S2 (see Table 6-3). The domination of electrostatic energy component over vdW one is quite apparent in both inter-peptide as well as peptide-COS interactions.

Moreover, the addition of COS prevents the inter-peptide interactions. On the other hand, the peptide-COS interactions become more favorable as COS is added.

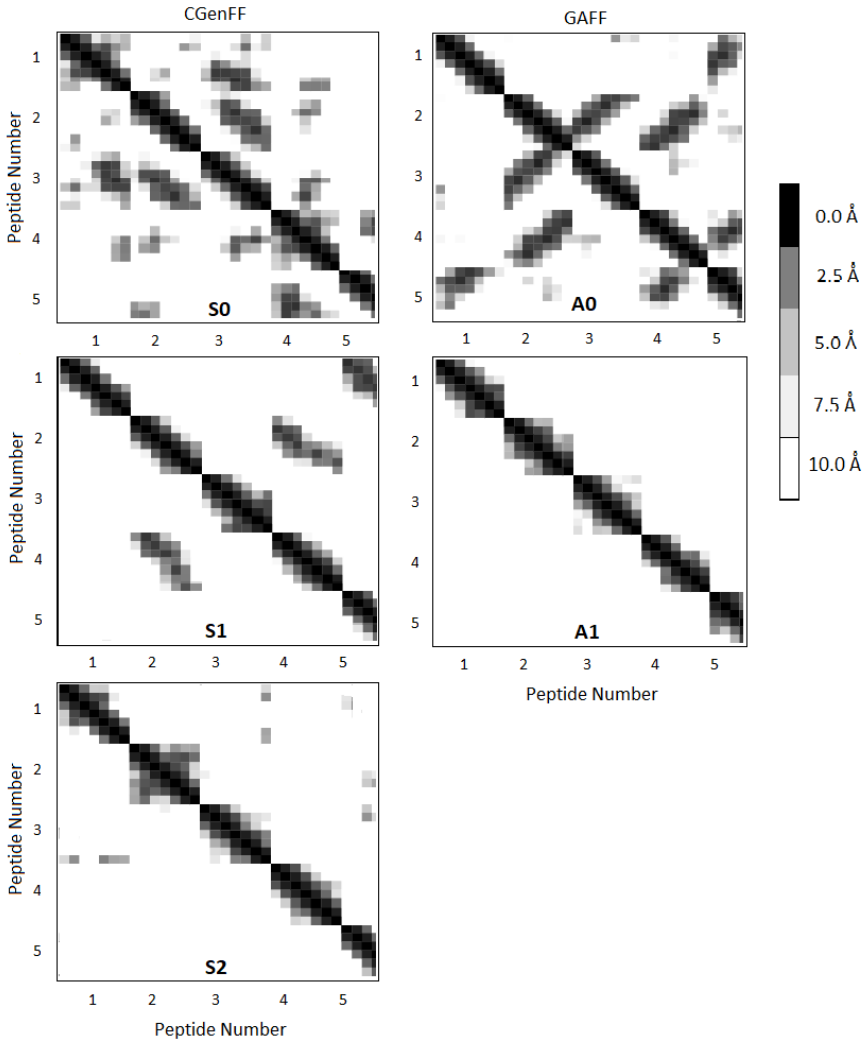

Figure 6-5. *Residue-residue contact maps of the peptides for different systems.*

The interaction energies between the peptides, and peptides and COS molecules for AMBER force field systems are also decomposed into vdW and electrostatic energy components, incorporated in the same Table 6-3. In pure water system A0, the inter-peptide interaction energies (both electrostatic and vdW) are significantly more favorable than system A1, and it is also to be noted that like CGenFF results, the electrostatic energy components dominate over the vdW energy values. In presence of COS, both

the electrostatic and vdW energy components contribute significantly to the peptide-COS interaction energy, which means COS interacts with the peptide both through hydrogen bonding interaction as well as hydrophobic solvation.

Table 6-3. Interaction energies[a]

System	PP (vdW)	PP (elec)	PC (vdW)	PC (elec)
S0	-9.68 (± 0.31)	-48.40 (± 0.129)	–	–
S1	-0.80 (± 0.007)	-7.47 (± 0.081)	-6.44 (± 0.021)	-29.83 (± 0.119)
S2	-0.32 (± 0.016)	-1.07 (± 0.028)	-17.37 (± 0.112)	-48.44 (± 0.216)
A0	-12.37 (± 0.117)	-61.23 (± 0.142)	–	–
A1	+3.52 (± 0.011)	-31.72 (± 0.023)	-29.15 (± 0.109)	-70.34 (± 0.369)

[a] Decomposition of total peptide-peptide and peptide-COS interaction energies into average van der Waals (vdW) and electrostatic (elec) energy components (per peptide). P and C represent the peptide and COS, respectively. Standard error of each calculation is provided in the parenthesis. All the energy values are expressed in kcal/mol unit.

We have also plotted the residue-wise electrostatic and vdW interaction energies between peptide and COS for CGenFF and GAFF systems (see Figure 6-6). From part a of Figure 6-6, we can see that vdW interaction energies between COS and central hydrophobic residues of the peptide are highly favorable which indicates the presence of strong hydrophobic interaction between them. For GAFF force field parameters (see Figure 6-6, part b) it can be noticed that, the residues F-4 and F-5 i.e., the PHE residues have the strongest hydrophobic interaction with COS. So, generalizing these two results, we can conclude that the hydrophobic core residues mainly the two PHE groups play the key role in the oligomerization process, and have strong hydrophobic interactions with COS. The residue-wise electrostatic interaction energies show that in case of CGenFF parameters (see Figure 6-6, part c), the terminal residues K-1 and K-7, and the central hydrophobic core 3VFF5 interact strongly with COS, and similarly, for GAFF parameters, the electrostatic interactions between residues K-1, K-4, K-7, and COS are very much favorable (see Figure 6-6, part d). It is worth mentioning here that, the residues LVFF which have the highest percentage of β-sheet in pure water and the negligible probability of the same in presence of

COS (see Figure 6-3) have significant vdW as well as electrostatic interactions with COS. Also, it can be noticed that the electrostatic energy component dominates over the vdW interaction energy for both of the force field parameters indicating the presence of strong hydrogen bonding interaction between the peptide residues and COS molecules.

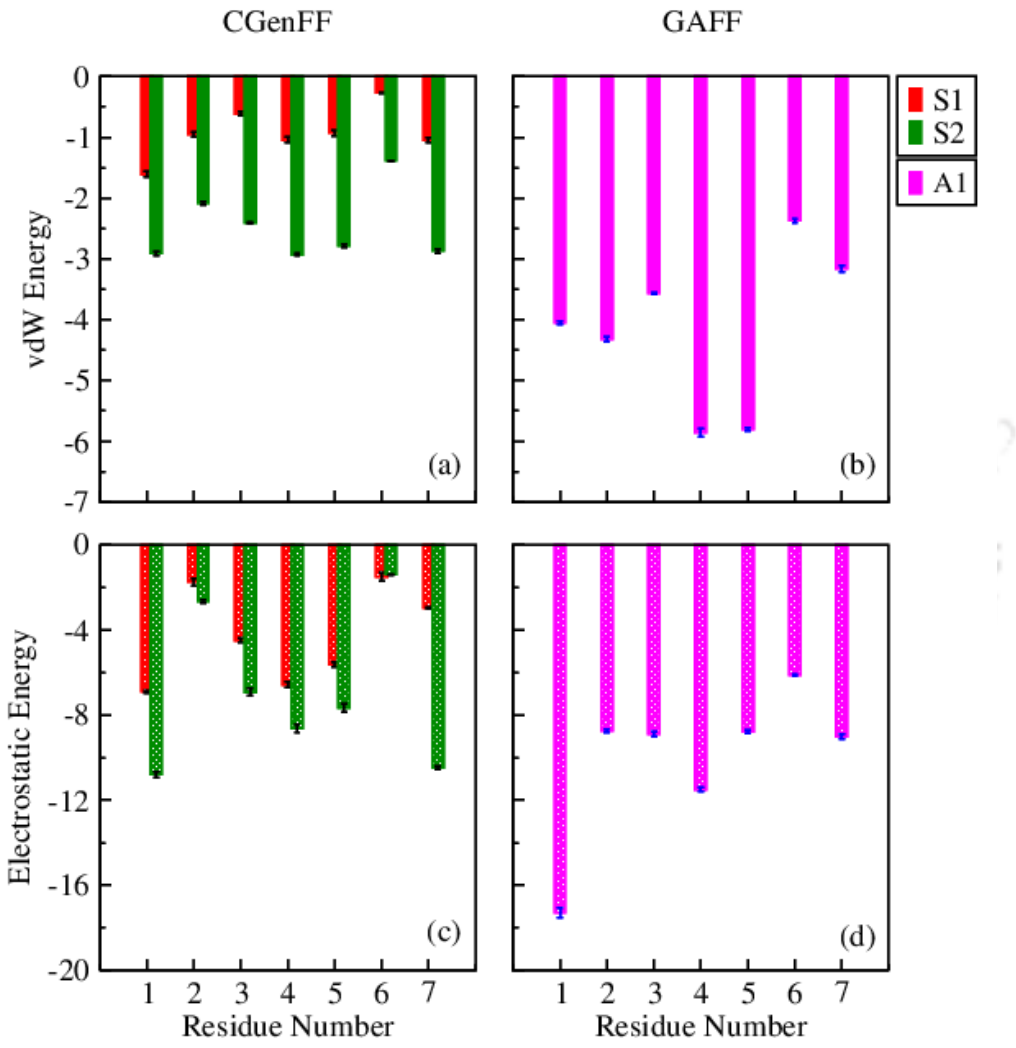

Figure 6-6. *(a) and (b) represents vdW interaction energies between peptide residues and COS (per peptide), (c) and (d) represents residue-wise electrostatic interaction energies between peptide and COS (per peptide) for systems S1, S2, and A1. All energies are expressed in kcal/mol unit. Standard errors are calculated using block average method.*

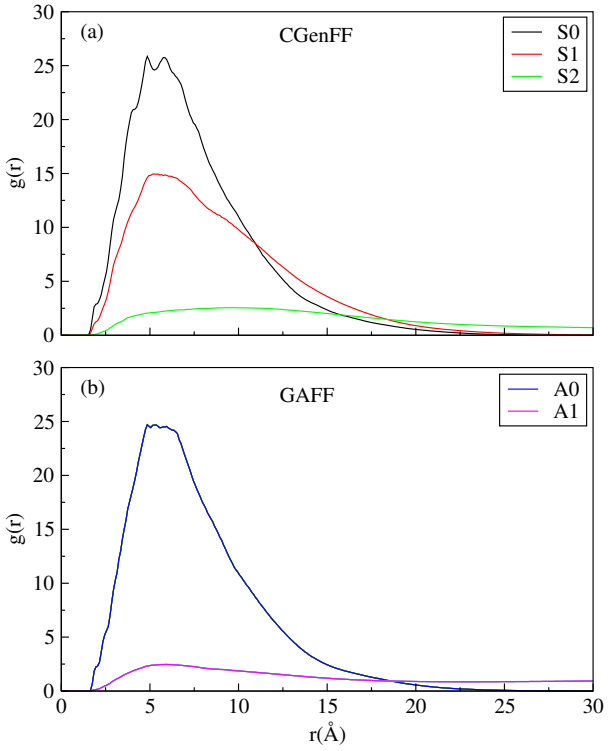

Figure 6-7. *Site-site radial distribution functions between peptides for systems (a) S0, S1, and S2; and (b) A0, A1.*

The results obtained from the interaction energy analysis lead us to check the radial distribution function between peptides as well as peptide and COS. The radial distribution functions (RDFs) between the peptides for different systems are shown in Figure 6-7. For calculation of these RDFs, the full Aβ_{16-22} peptide is considered. In case of CHARMM general force field systems (Figure 6-7, part a), we can clearly notice that for system S0 there is a broad peak that appears at 5.0 Å indicating a strong interaction between the peptides in pure water system. In system S1, in presence of 0.5 M COS, the RDF peak height is reduced significantly indicating that 0.5 M COS solution inhibits the self-association of the peptides to some extent. With increasing the concentration of COS in system S2, the peak height is reduced further denoting that in higher concentrations of COS, due to

stronger interactions between the peptide and COS, inter-peptide interactions are reduced significantly.

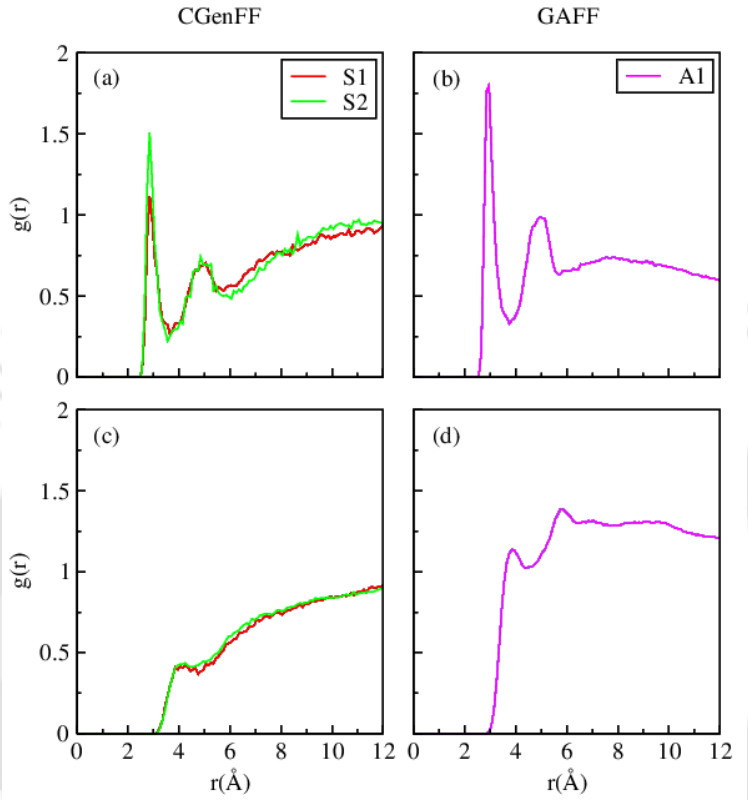

Figure 6-8. *Site-site radial distribution functions (a) and (c) between amide-nitrogen atoms of the peptide backbone and sulfate-oxygen atoms of choline-O-sulfate, (b) and (d) between side-chain carbon atoms of the peptides and methyl-carbon atoms of choline-O-sulfate molecules for both CGenFF and GAFF parameters.*

To investigate how COS helps to reduce the inter-peptide interaction in water, we have plotted site-site pair correlation functions between the backbone nitrogen of the peptide and sulfate oxygen of COS (see Figure 6-8, part a). In system S1, there is an appearance of a sharp peak at 2.8 Å suggesting that the sulfate group of COS is involved in strong hydrogen bonding interaction with the backbone nitrogen of the peptide. With increasing concentration of COS, in the S2 system, the peak height is enhanced further, indicating a better interaction between peptide and COS at higher concentrations of COS. Note that, it has already been reported in an experimental study that, the sulfate group

of COS plays the most important role to inhibit the aggregation of human islet amyloid peptide which is also an intrinsically disordered protein like Aβ peptide [85]. Next, the presence of a significant amount of peptide-COS vdW interaction energy suggests us to calculate the RDF between the side-chain carbon atoms of the peptide and the hydrophobic methyl carbons of COS (see Figure 6-8, part c). The presence of a small peak at 4.0 Å implies the hydrophobic solvation of peptide side-chains by COS molecules.

Now considering the GAFF systems, it can be noticed that the peptide-peptide RDF peak (see Figure 6-7, part b) is very intense in between 5 to 6 Å for system A0, whereas for system A1, the peak height is reduced significantly indicating the reduction in the peptide-peptide interaction in presence of COS. The pair correlation function between the peptide backbone nitrogen and sulfate oxygen of COS for system A1 (see Figure 6-8, part b) shows a sharp peak at 3.0 Å indicating that strong hydrogen bonding interaction is present between these atoms. This is in high consensus with the results obtained using CGenFF parameters which strongly suggest that the sulfate oxygen atoms of the COS participate in the hydrogen bonding interaction with peptide backbone nitrogens. The pair correlation function between the side-chain carbon atoms of the peptide and methyl carbons of COS (see Figure 6-8, part d) shows a relatively low intense peak at 3.9 Å and 5.8 Å which implies the presence of hydrophobic interactions between these moieties.

We have also determined the number of COS molecules present in the first solvation shell of each residue of the peptide and plotted in a bar diagram (see Figure 6-9). With CGenFF parameters (Figure 6-9, part a) it is apparent that, besides the terminal residues, for residues 4 and 5 (the PHE residues), the number of COS molecules in their first solvation shell is very high. From part b of Figure 6-9, it is apparent that the first shell coordination numbers of COS using GAFF parameters are also remarkably higher around the hydrophobic PHE residues (F-4 and F-5) than the other residues. From Figure 6-3 it is barely seen that, these residues have high probability of β-sheet formation in pure water. PHE is a hydrophobic residue, and it has already been reported previously that in the self-assembly of Aβ_{16-22} peptide the two PHE residues play the key role to form β-sheet structure [282]. So, from our results, we can propose that, in the prevention of the self-assembly of Aβ_{16-22} peptides, the solvation of its hydrophobic PHE residues by COS plays an important role.

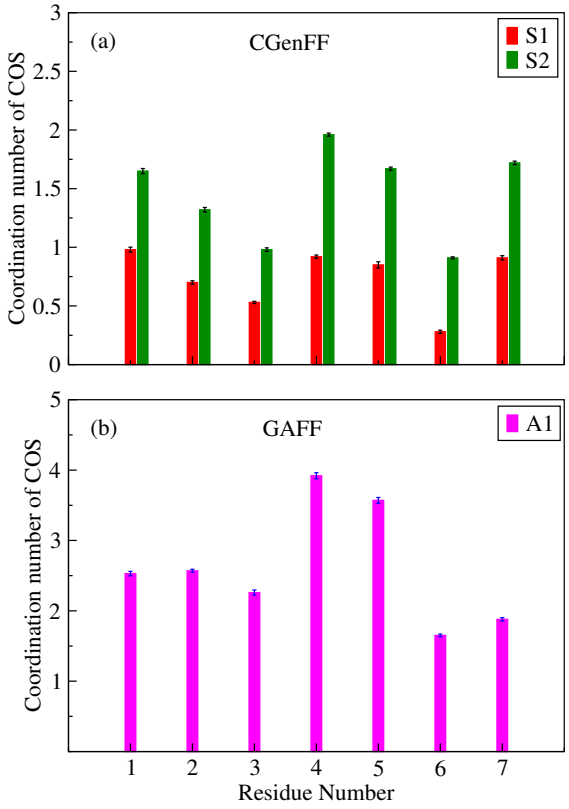

Figure 6-9. *First shell coordination number of COS molecules around each residue of the peptide for systems (a) S1, S2; and (b) A1. Standard errors are calculated using block average method.*

Hydrogen bond properties

In the peptide aggregation and fibril formation, hydrogen bonding interaction plays a crucial role. From the RDF plot between the backbone nitrogen atoms of the peptide and sulfate oxygen atoms of the COS (Figure 6-8, part a and b) we have found that there is a possibility of strong hydrogen bonding interaction between peptide and COS. So, we have estimated the inter-peptide and peptide-COS average hydrogen bond numbers using the geometric criteria D-A $\leqslant$ 3.5 Å and $\angle$D-H-A $\geqslant$ 45° (see Table 6-4) [282]. Here D and A refer to donor and acceptor atoms respectively. The inter-peptide hydrogen bond number using CHARMM force field parameters is very high in system S0, on the other hand, with

the addition of 0.5 M COS solution, this number sharply decreases from 2.55 to 0.83 (per peptide), and in system S2 with the highest concentration of COS, it is reduced further to 0.24 per peptide. The peptide-COS hydrogen bond numbers (per peptide) follow the opposite trend to that of the inter-peptide hydrogen bond number, and it increases with increasing the concentration of COS.

Table 6-4. Hydrogen bond numbers[a]

System	HB_{PP}	HB_{PC}
S0	2.55 (± 0.014)	—
S1	0.83 (± 0.009)	3.09 (± 0.015)
S2	0.24 (± 0.007)	4.12 (± 0.027)
A0	2.50 (± 0.021)	—
A1	0.52 (± 0.006)	3.67 (± 0.020)

[a] Average hydrogen bond numbers between peptide-peptide and peptide-COS (per peptide). Standard error of each calculation is provided in the parenthesis.

We have also estimated different types of intermolecular hydrogen bond numbers for GAFF systems using the same criteria as used in the hydrogen bond analysis for CGenFF systems. The inter-peptide hydrogen bond number for system A0 is 2.50 (per peptide) which is very similar to that of CGenFF data, and it also matches well with the previously reported simulation work on Aβ_{16-22} (they found around 13 hydrogen bonds for 5 peptides that is 2.6 per peptide) [282]. In system A1, the inter-peptide hydrogen bond number is reduced to 0.52 whereas the peptide-COS hydrogen bond number is 3.67 per peptide, which is also quite similar to that of peptide-COS hydrogen bond number with CGenFF parameters.

We have also calculated residue-wise inter-peptide as well as peptide-COS hydrogen bond numbers for all the systems considered here (see Figure 6-10). This is important as it provides information about the specific role of COS in preventing peptide aggregation. It can be noticed that the residues 2LVFFA5 of the peptide which have the highest probability to form β-sheet conformation in pure water (see Figure 6-3), posses a very high number of inter-peptide hydrogen bonds in pure water systems, and it gets reduced significantly

with the addition of COS into the systems. Similarly, the terminal residues and the central hydrophobic residues (VFF) of the peptide have higher hydrogen bonding interaction with COS in systems S1 and S2, and for system A1, residues 1 to 5 form a significant number of hydrogen bonds with COS than the other residues of the Aβ_{16-22} peptide.

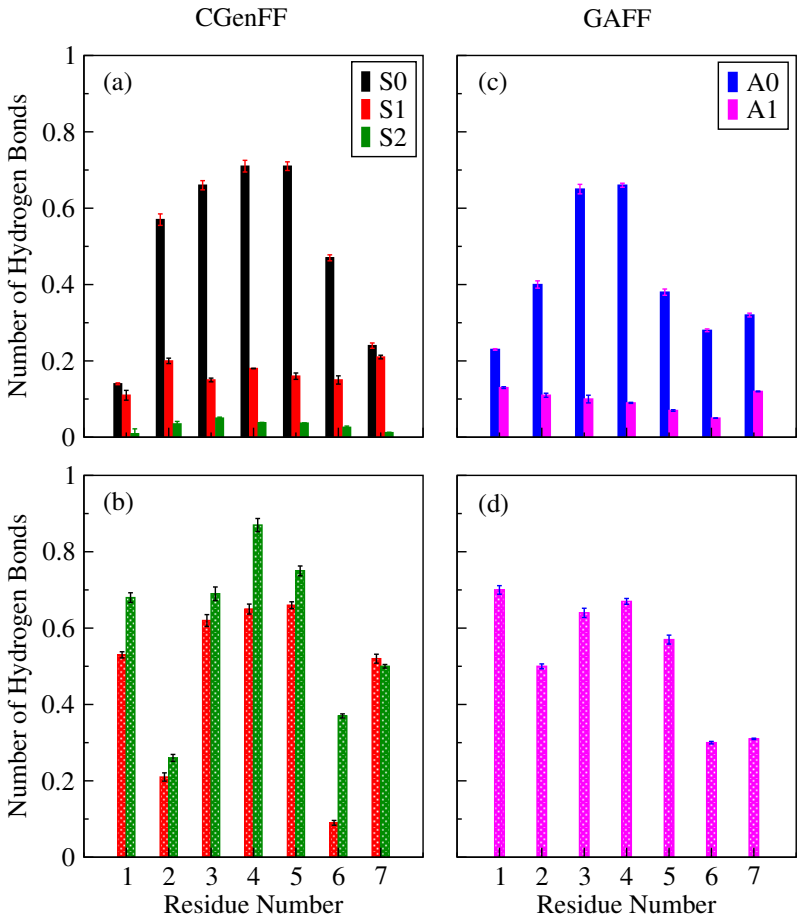

Figure 6-10. *(a) and (c) represent the residue-wise inter-peptide hydrogen bond numbers (per peptide) for systems S0, S1, S2; and A0, A1, respectively. (b) and (d) represent the residue-wise peptide-COS hydrogen bond numbers (per peptide) for systems S1, S2, and A1, respectively. Standard errors are calculated using block average method.*

So, from the above analysis, we can safely propose that the results obtained from both the CHARMM and AMBER force field parameters are in good agreement with each other. Furthermore, we conclude that both hydrogen bonding interaction between peptide

and COS as well as the solvation of the hydrophobic residues by COS molecules play a major role to inhibit the aggregation of Aβ_{16-22} peptides.

Potential of Mean Force

Adaptive Biasing Force Method

To examine the stability of the dimer formed between the peptides in pure water as well as in the COS solution, we have estimated the potential of mean force (PMF) between two peptides (Figure 6-11). The adaptive biasing force (ABF) method is implemented for this purpose [143, 144, 145]. PMF can be used to find out how free energy changes as a function of a coordinate of the system studied. To calculate the PMFs between two peptides, we have chosen the distance between the center of mass of two peptides as the reaction coordinate which ranges from 3 to 15 Å. We have chosen the maximum distance of 15 Å because beyond this distance there is no possibility of dimer formation at all (see Figure 6-5 and 6-7). The reaction coordinate is divided equally into 12 windows of 1 Å width, each of which is simulated for 10 ns at 300 K. The initial configurations for all these windows are generated from the last trajectory of the previous window (described in Section II). Each of these windows is further segmented into small bins of 0.2 Å and the ensemble averages of the adaptive biasing forces are calculated only if 4000 samples are collected in each of these bins. A harmonic force constant of 100 kcal/mol/Å is applied to the upper and lower boundaries of the reaction coordinate of the particular window, allowing the solute particles to be bound within the given limits. To generate the PMFs between two peptides, we restrained one peptide with harmonic constraint by setting the other peptide free.

We have already mentioned in the Models and Simulation Method section that, we have constructed two different systems namely P0 and P1 for the PMF calculation using ABF method (see Table 6-1). These systems contain two peptides, and the system P1 contains the COS concentration of 0.96 M, exactly similar to the system S2. From Figure 6-11, we can see that the binding energy between two peptides in pure water is about -12 kcal/mol, and it has a well-depth that appears at 5.0 Å. This is attributed to the formation of a stable dimer of the peptides. In system P1, the binding energy value is more positive

than system P0, which clearly indicates that the interaction between two peptides in the P1 system is very weak.

Umbrella Sampling Method

The PMFs between two peptides in different solution for GAFF parameters are calculated using Umbrella Sampling method (Figure 6-12) [282]. For this, we have considered two different systems, U0 and U1, described in Table 6-1. Initially, normal MD simulation of 100 ns was performed for both the systems for equilibration. The initial configurations for umbrella sampling are obtained from the final configurations of the normal MD simulations. A series of umbrella sampling simulations were then carried out to calculate the potential of mean forces of different desired systems [146]. Here, we have increased the distance between the center of mass of two peptides from 3 to 15 Å alike the ABF method, and this distance is further segmented into 60 small windows, each of having equal width of 0.2 Å . In this method, at first, we have generated a set of initial configurations for each window from its previous final trajectories. In our case, to calculate the PMFs between two peptides, one peptide is harmonically restrained with a force constant of 10 kcal/mol/Å at the increasing center of mass distance from the reference peptide molecule using an umbrella biasing potential. The windows are allowed to overlap slightly with each other to generate a smooth PMF curve. The PMFs of different systems using GAFF parameters are presented in Figure 6-12.

The potential of mean force between two peptides in system U0 has a sharp minimum at 4.8 Å with a free energy value of -14.2 kcal/mol, suggesting that the dimer is most stable at this distance, and this free energy value is quite comparable with that of CGenFF result. In system U1, there is no sharp change in the binding energy between two peptides, and the PMF value is more positive than that of system U0. The PMFs of different systems, calculated from the umbrella sampling method, follow the same trend as that of the PMFs obtained from ABF method. Nevertheless, from Figure 6-12, it is clear that the highly negative binding energy between the peptides in system U0 is due to the stable dimer formation between the peptides, which becomes unfavorable in presence of COS molecules.

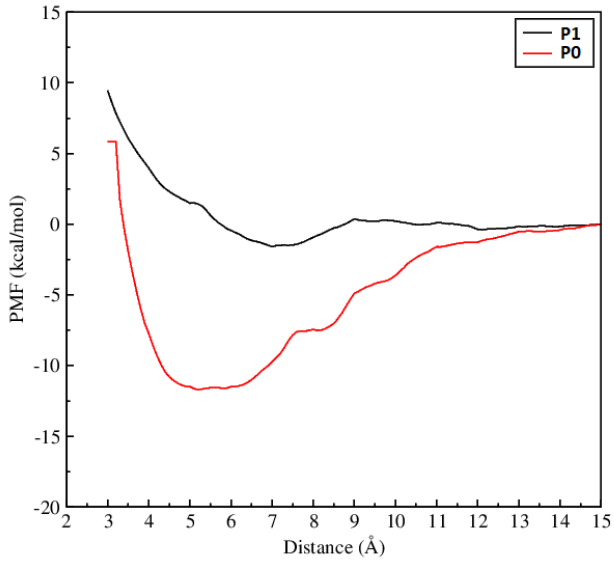

Figure 6-11. *Potential of mean forces (PMF) between the center of mass of two peptides for systems P0 and P1.*

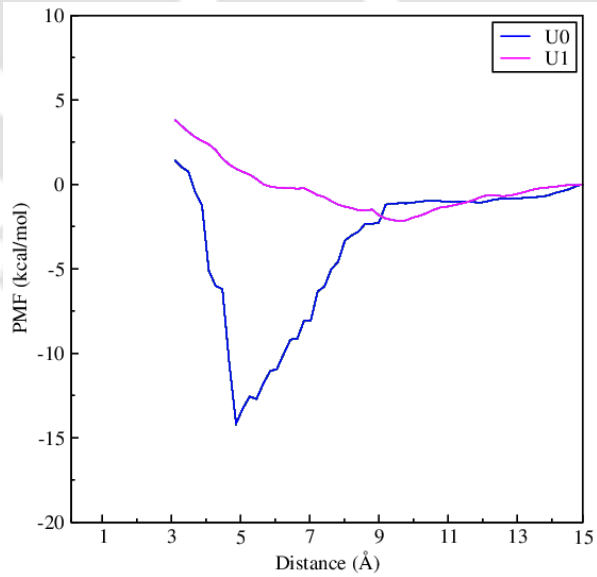

Figure 6-12. *Potential of mean forces (PMF) between the center of mass of two peptides for systems U0 and U1.*

Destabilization of Preformed Protofibril of Aβ_{16-22} by COS

To show the inhibitory effects of COS on amyloid Aβ_{16-22} aggregation, it is necessary to check whether COS can destabilize the preformed protofibril of Aβ_{16-22} or not. To examine the stability of protofibril of Aβ_{16-22} in presence of COS, we have carried out another two sets of simulations starting from the preformed protofibril of Aβ_{16-22}. As there is no crystal or NMR structure of only Aβ_{16-22} protofibril available in the protein data bank, following the work of Li et al. [298], we have considered the 5-stranded β-sheet protofibril of Aβ_{16-22} formed after the 500 ns production run of systems S0 (for CGenFF) and A0 (for GAFF) for these two simulations. Two systems were constructed namely B1 and B2 (see Table 6-5) for CGenFF and GAFF parameters respectively, containing 5-stranded protofibril of Aβ_{16-22} and 0.95 M COS, and 1000 ns NPT production run was carried out for each of these systems in NAMD and AMBER by following the same simulation protocols discussed in the Models and Simulation Method section.

Table 6-5. System overviews[a]

System	N_P	N_C	N_W	Volume (nm^3)	M_C
B1	1	100	5000	174.05	0.95
B2	1	100	5000	177.03	0.94

[a] N_P, N_C, and N_W represent the number of protofibril, choline-*O*-sulfate, and water molecules, respectively. M_C is the molar concentration of choline-*O*-sulfate for different systems.

First, we have represented the snapshots of systems B1 and B2 at different time intervals to show the changes in the conformation of the protofibril at different time steps. From Figure 6-13, we can see that in CGenFF system the protofibril gets destabilized by COS after 400 ns, and after 1000 ns only 2-stranded β-sheet is present in the system. For GAFF system (system B2), the destabilization of protofibril is more rapid than the CGenFF system, and after 1000 ns, no extended β-sheet is present.

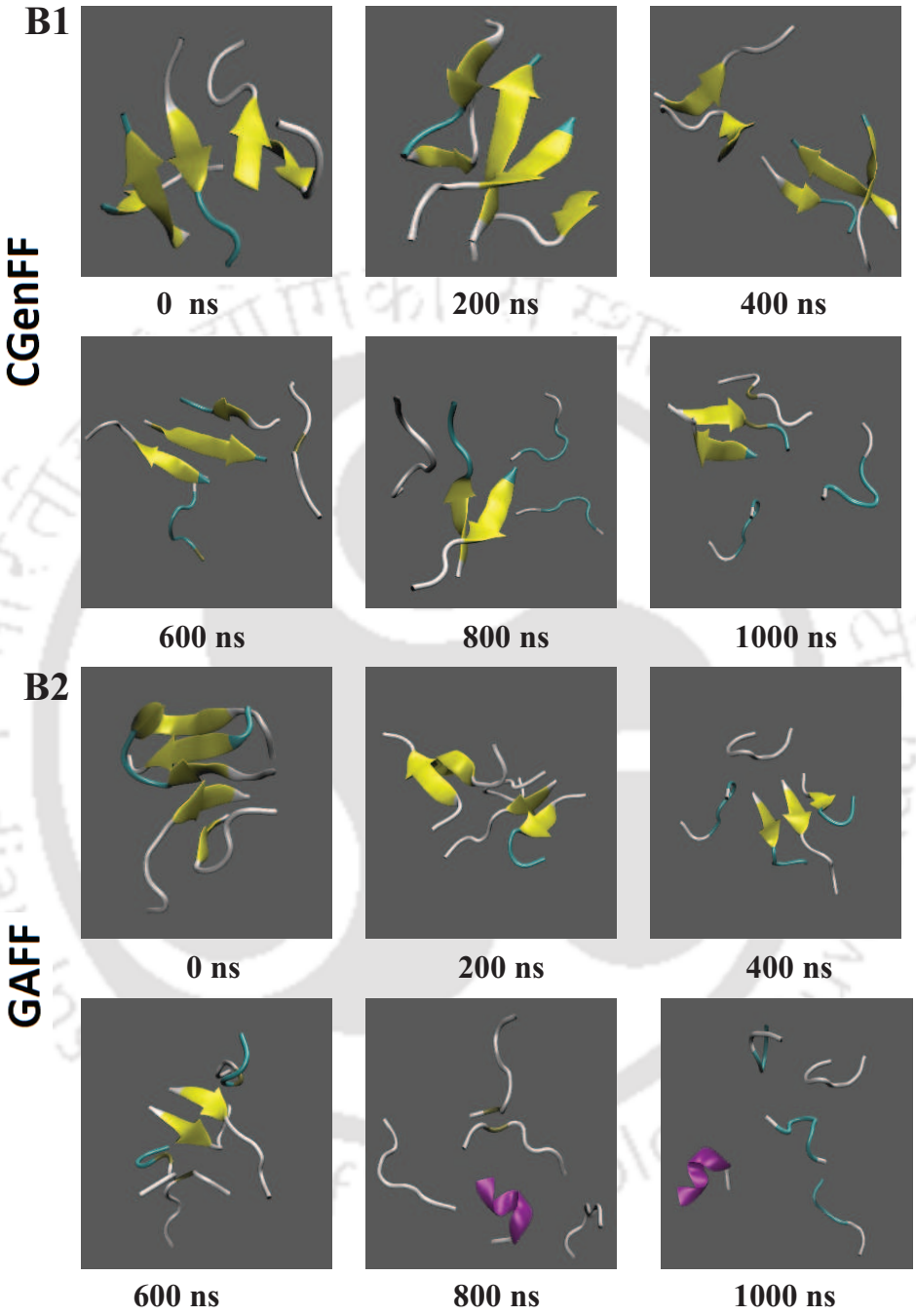

Figure 6-13. *Snapshots of the systems B1 and B2 at 0, 200, 400, 600, 800, and 1000 ns. Water and COS molecules are left-off for better visual clarity.*

Next, we have analyzed the secondary structure of the protofibril with respect to time and found that, as the simulation progresses, in CGenFF system, the extended β-sheet conformation appears partially, and for GAFF system it almost vanishes (see Figure 6-14).

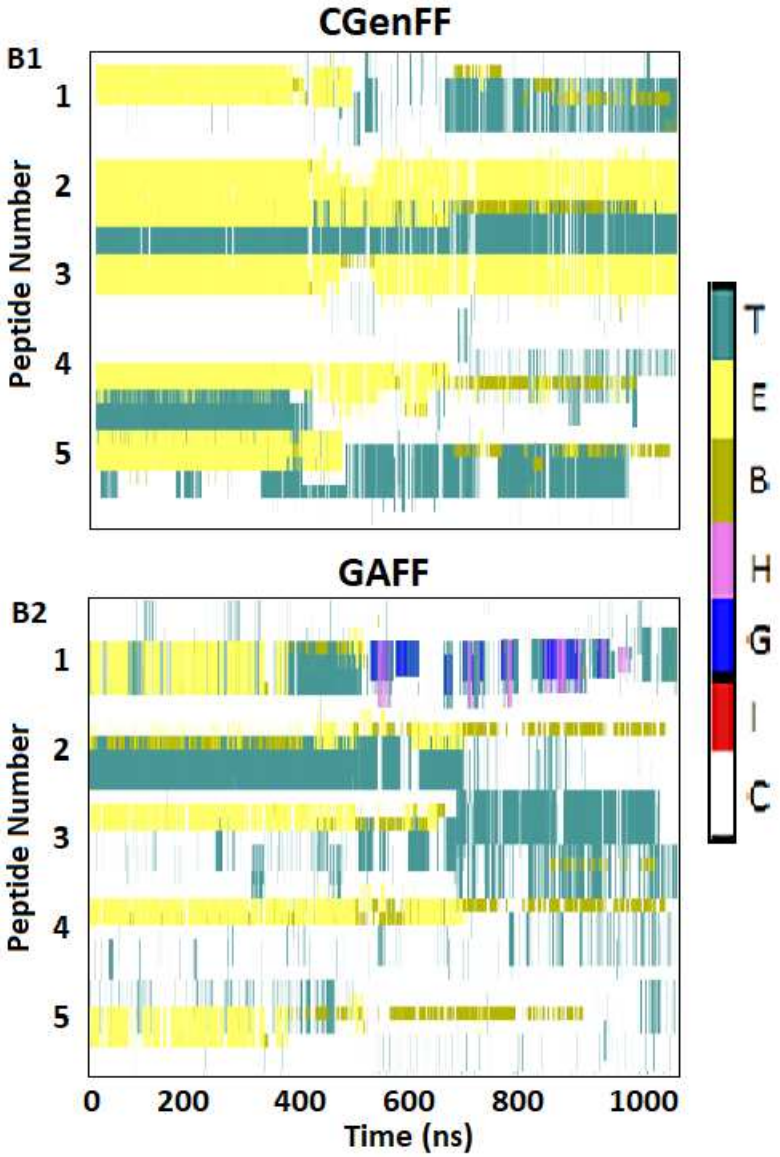

Figure 6-14. *Secondary structure analysis with simulation time for systems B1 and B2. In the color bar: T=Turn, E=Extended β-sheet, B=Isolated bridge-beta, H=α-helix, G=3_{10}-helix, I=π-helix, and C=Coil.*

We have also determined the average percentages of β-sheet, helix, and random coil conformations for these two systems for every 200 ns time interval (see Figure 6-15). For both the systems B1 and B2, the percentages of β-sheet conformation decrease with simulation time, and the formation of random coil structure gets enhanced. It is worth to mention here that in GAFF system, the occurrence of helical conformation also increases significantly as the simulation progresses.

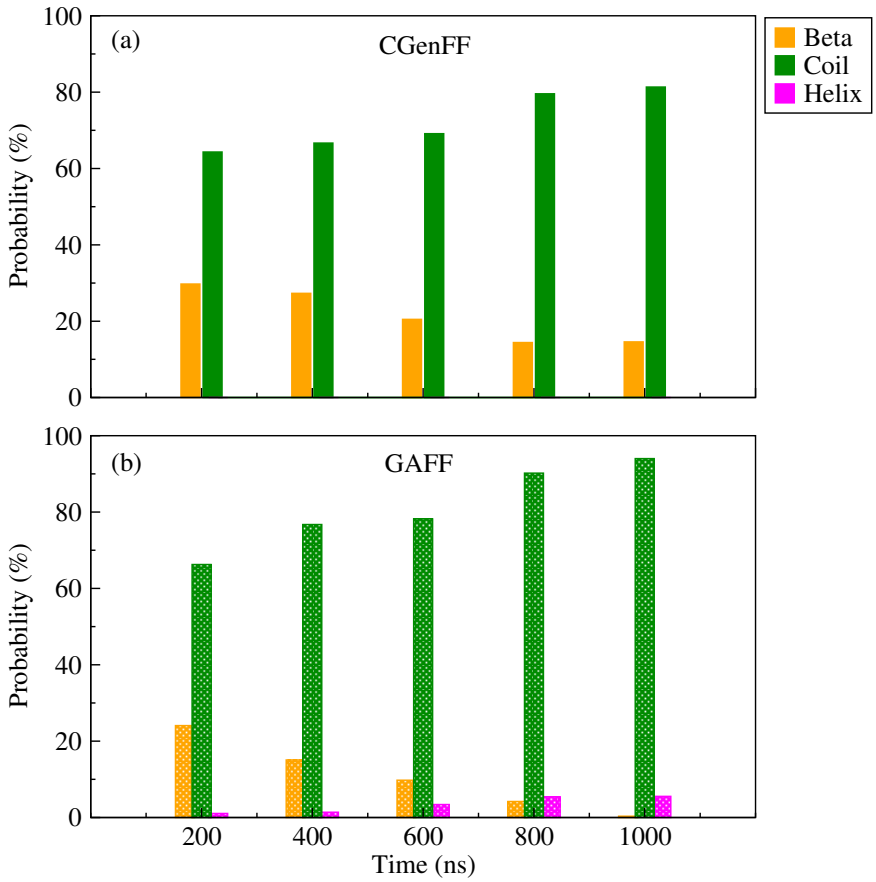

Figure 6-15. *Probability of β-sheet, helix, and random coil conformations at each 200 ns time interval for systems (a) B1, and (b) B2. Percentages are calculated taking the block averages of every 200 ns.*

For in detail analysis, we have also calculated the residue-wise β-sheet formation for both these systems at different time intervals, and presented the data in Figure 6-16. At 0-200 ns time interval, the probability of fibril formation is the highest for almost each and every residue of the protofibril which gets reduced slowly as the simulation progresses.

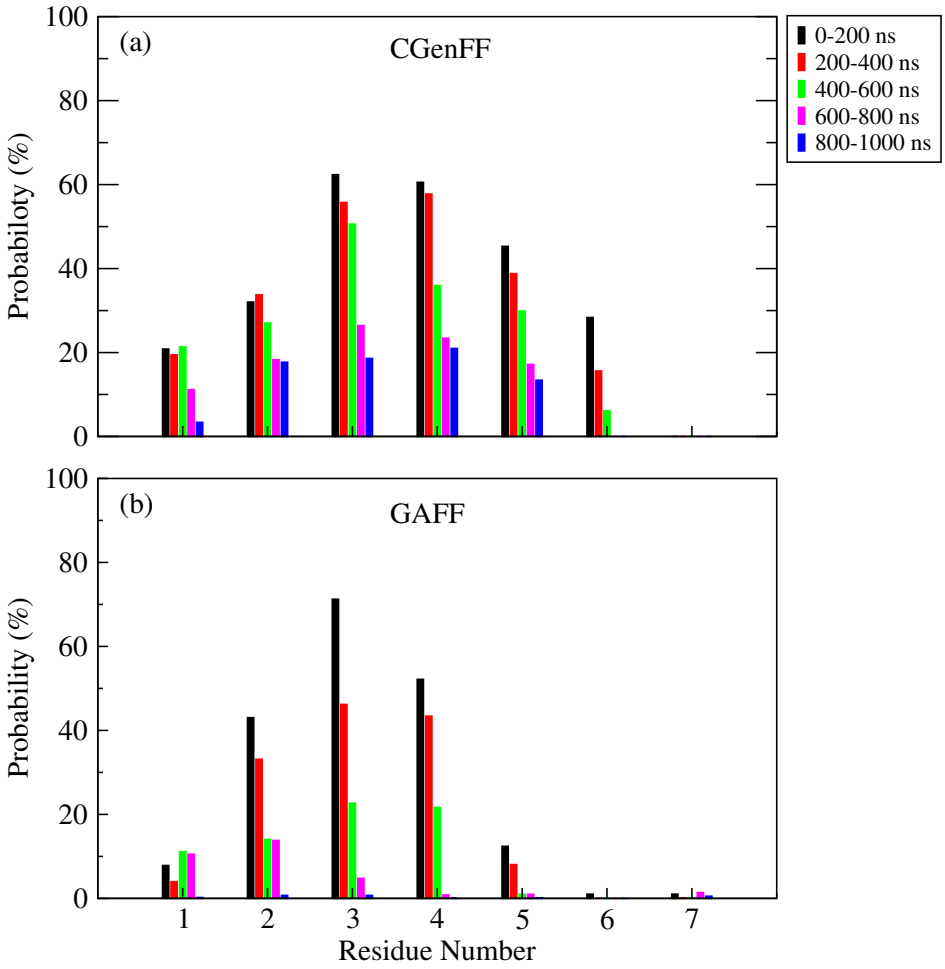

Figure 6-16. *Probability of residue-wise β-sheet formation at each 200 ns time interval for systems (a) B1, and (b) B2. Percentages are calculated taking the block averages of every 200 ns.*

Lastly, we have presented the average inter-peptide and peptide-COS hydrogen bond numbers for every 200 ns time interval in Figure 6-17, and have found that for both CGenFF and GAFF parameters, the inter-peptide hydrogen bond numbers get reduced with simulation time, and peptide-COS hydrogen bond numbers increase simultaneously. Reduction of β-sheet percentage and inter-peptide hydrogen bond number of the protofibril strongly suggest that 0.95 M COS is efficient enough to disaggregate the preformed fibril of Aβ_{16-22}.

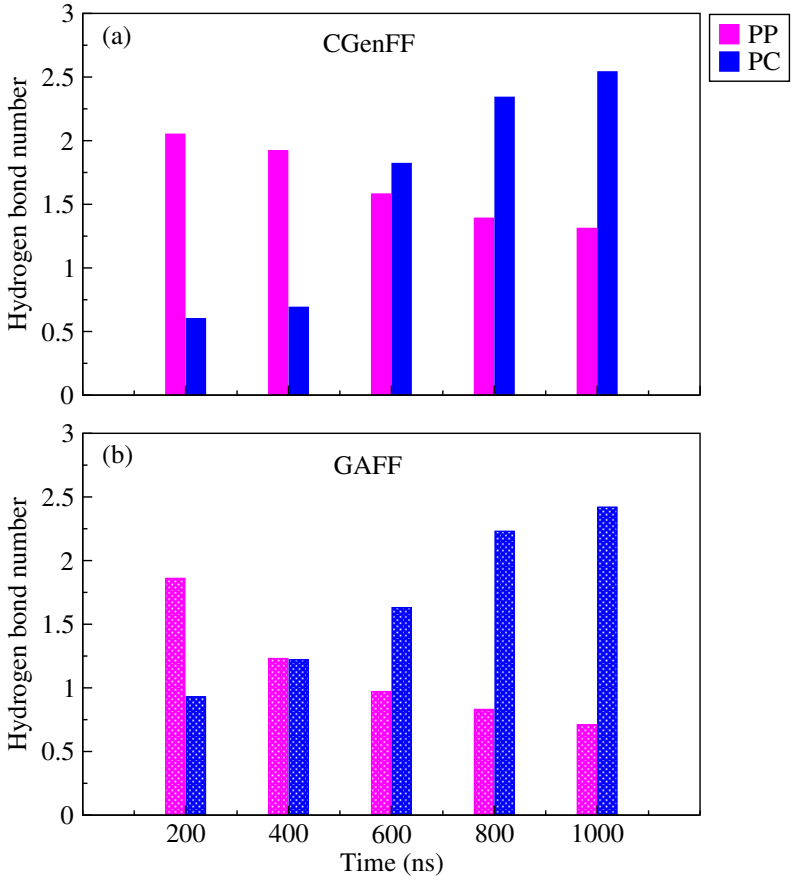

Figure 6-17. *Average inter-peptide and peptide-COS hydrogen bond numbers at each 200 ns time interval for systems (a) B1, and (b) B2.*

■ SUMMARY AND CONCLUSIONS

The aggregation of Aβ-peptide is thought to be the primary cause of Alzheimer's disease. Here, we have performed a theoretical study on the application of COS as an inhibitor in amyloid fibrillation. A comparative study with two different types of force field parameters i.e., CHARMM General Force Field and General AMBER force field parameters, is carried out. It is worth mentioning that, the results obtained from both the CGenFF and GAFF parameters are quite similar to each other which helps us to draw an exact conclusion on our findings. From different secondary structure analysis, we find that, the Aβ_{16-22} peptide aggregates in pure water to form ordered β-sheet structure from random coil conformation leading to the protofibrillation. The central hydrophobic core LVFFA and mainly residues V-3, F-4, and F-5 acquire the maximum β-sheet percentage in pure water, while in presence of 0.5 M and a higher concentration of COS, the β-sheet conformation is almost absent in these residues. The radial distribution functions and interaction energies between the peptide and COS, strongly suggest the presence of hydrogen bonding interaction as well as solvation of the hydrophobic groups of the peptides by COS. The inter-peptide hydrogen bond numbers for residues LVFFA decrease significantly with the addition of COS into the systems, and simultaneously these residues form a significant number of hydrogen bonds with COS. The coordination number of COS around different residues of the peptide, especially, around the residues F-4 and F-5 is significantly higher compared to the other residues, which indicates that the hydrophobic solvation of these residues by COS molecules plays an important role to prevent the peptide aggregation in water. Finally, the destabilization of preformed protofibril of Aβ_{16-22} in presence of 0.95 M COS indicates that COS is quite a good and efficient inhibitor in the prevention of amyloid aggregation of Aβ_{16-22} peptide.

Chapter 7

Summary and Our View on the Dual Role of Choline-*O*-sulfate

"The association of solved problems with those unsolved may throw new light on our difficulties by suggesting new ideas. It is easy to find a superficial analogy which really expresses nothing. But to discover some essential common features, hidden beneath a surface of external differences, to form, on this basis, a new successful theory, is important creative work"

— Albert Einstein

To explore the underlying mechanism of the prevention of protein misfolding by choline-*O*-sulfate at different harsh conditions, we systematically investigated solvation characteristics of various functional groups comprising proteins by employing different MD simulation techniques. First, we started our investigation with the study of the generic behavior of the well-known chemical denaturant urea and COS in an aqueous mixture at room temperature as well as at a higher temperature to explore their physical interactions and thermodynamic properties in presence of each other. Detailed examination of every possible interaction between the solution species shows that urea becomes a weaker denaturing agent than its individual counterparts in presence of COS. The mixed osmolyte solution increases the water-water hydrogen bonding interaction owing to the solvent assisted stability of a poly-peptide and the mutual exclusion of the both osmolytes in presence of each other can reduce the effect of co-solutes on the biomolecule. Further, we found that at higher temperature due to decreased interactions between the solution species, the urea molecules become free, hence, can interact with another solute molecule like protein (if present in the solution) and can knock down its stability. Also, we have validated our results by comparing between two force field parameters of COS, i.e, CHARMM General Force Field parameters (CGenFF) and General AMBER Force Field (GAFF) parameters. The conclusions of this study lead us to further investigate the direct action of COS on the stabilization of proteins against urea-conferred denaturation of proteins.

First, we studied the destabilization of S-peptide in 8 M urea solution and its counteraction by COS at room temperature using classical MD simulation. This study gives us a satisfactory result on the counteracting ability of COS against urea and shows that COS plays an indirect role in the prevention of the urea-conferred denaturation of S-peptide. Like the widely employed osmolytes TMAO and glycine betaine, preferential interaction of COS with the urea molecules over the peptide-backbone excludes the later from the protein surface, hence, cancel its candidature as a protein denaturant. Next, to explore the mechanisms of protein stabilization and destabilization by COS and urea respectively over a wide range of temperatures, we have investigated different structural properties of the Trp-cage mini-protein in pure water, binary urea solution and mixed urea-COS solution by employing Replica Exchange Molecular Dynamics in a temperature range of 290 K to 420 K. This time, we again found COS as an effective osmolyte to counteract

urea up to 340 K. But, as found in the study of synergistic behavior of urea and COS in aqueous solution, at higher temperature, COS is not operative to exclude the urea molecules from the vicinity of the protein surface. Also, we produced the melting curve of the protein at different temperatures from different free energy landscape plots.

Next, we took another move to study the counteracting effect of COS against an unconventional denaturant dodine. For this, we have studied the stability of the terminal helices of the λ-repressor protein in presence of a very low concentration of aqueous dodine solution, and in ternary dodine-COS mixture. Time dependent tracking of different structural changes and free energy landscape plots of the conformational phase-space reveal that at a concentration of $\geqslant$0.5 M, COS provides even a better stability to the protein than that of pure water system. Examination of all possible interactions between the solution species reveals that along with the protein-COS interaction, the sulfate group of COS interacts strongly with the positively charged dodinium head group of dodine and exclude them from the protein backbone, preventing the dodine-induced unfolding of the terminal helices of the λ-repressor protein.

To shed light into the another role of choline-*O*-sulfate as an inhibitor in peptide aggregation, we have studied the self-assembly of hIAPP peptide in pure water as well as in COS solution. Different trajectory analysis shows that the sulfate oxygen of COS strongly interacts with the backbone amide nitrogen of the poly-peptide and prevent them from the fibril formation. Next, we studied the inhibitory effects of COS on another amyloidogenic poly-peptide Aβ_{16-22}, the shortest fibril forming segment of full-length Aβ peptide. In this study, we have employed two different force field parameters, CGenFF and GAFF parameters, to produce more reliable estimation of our results. Simulations using both force field parameters show that the peptides aggregate in pure water to form fibril structures and the central hydrophobic residues play the key role in the formation of extended β-sheet structure. Careful analysis of the simulation trajectories of COS containing systems shows that due to the significant interactions of COS with the central residues, the peptides get separated from each other preventing the self-association of the Aβ_{16-22} peptides.

This thesis represents the molecule choline-*O*-sulfate as a protecting osmolyte in the protein folding-unfolding process as well as a potent inhibitor in the field of peptide aggregation. Successful application of COS on two different kinds of proteins (cellular protein

S-peptide and globular protein Trp-cage) against the chemical denaturant urea would help to draw attention to the wider approach of COS as a counteracting osmolyte in protein misfolding. This naturally occurring molecule is not only able to withstand against the traditional chemical denaturant urea, its counteracting ability is quite appreciable against the surfactant denaturant dodine. Moreover, this non-toxic molecule can inhibit the fibrillation of hIAPP and Aβ_{16-22} peptides, associated with Type-II diabetes mellitus and Alzheimers disease, respectively, which can be attributed to the therapeutic approach to these neurodegenerative diseases and molecular drug discovery. In short, this dissertation can provide a new insight in the field of biochemistry related to protein dysfunction.

Bibliography

1. Cuanalo-Contreras, K.; Mukherjee, A.; Soto, C. *Int. J. Cell Biol.* **2013**, *2013*, 1-10.
2. Herczenik, E.; Gebbink, M. F. B. G. *FASEB J.* **2008**, *22*, 2115-2133.
3. Balch, W. E.; Morimoto, R. I.; Dillin, A.; Kelly, J. W. *Science* **2008**, *319*, 916-919.
4. Kauzmann, W. *Adv. Protein Chem.* **1959**, *14*, 1-63.
5. Finer, E. G.; Franks, F.; Tait, M. J. *J. Am. Chem. Soc.* **1972**, *94*, 4424-4429.
6. Das, A.; Mukhopadhyay, C. *J. Phys. Chem. B* **2009**, *113*, 12816-12824.
7. Almarza, J.; Rincon, L.; Bahsas, A.; Brito, F. *Biochemistry* **2009**, *48*, 7608-7613.
8. Vanzi, F.; Madan, B.; Sharp, K. *J. Am. Chem. Soc.* **1998**, *120*, 10748-10753.
9. Yancey, P. H. *Am. Zool.* **2001**, *41*, 699-709.
10. Yancey, P. H. *J. Exp. Biol.* **2005**, *208*, 2819-2830.
11. Hua, L.; Zhou, R.; Thirumalai, D.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2008** *105*, 16928-16933.
12. Wei, H.; Fan, Y.; Gao, Y. F. *J. Phys. Chem. B* **2010**, *114*, 557-568.
13. Wei, H.; Shao, Q.; Gao, Y. Q. *Phys. Chem. Chem. Phys.* **2010**, *12*, 9292-9299.
14. Wei, H.; Yang, L.; Gao, Q. Y. *J. Phys. Chem. B* **2010**, *114*, 11820-11826.
15. Bennion, B. J.; Daggett, V. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 5142-5147.
16. Zhang, N.; Liu, F.; Dong, X.; Sun, Y. *J. Phys. Chem. B* **2012**, *116*, 7040-7047.
17. Klimov, D. K.; Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 14760-14765.
18. Camilloni, C.; Rocco, A. G.; Eberini, I.; Gianazza, E.; Broglia, R. A.; Tiana, G. *Biophys. J.* **2008**, *94*, 4654-4661.
19. Das, A.; Mukhopadhyay, C. *J. Phys. Chem. B* **2008**, *112*, 7903-7908.
20. O'Brien, E. P.; Ziv, G.; Haren, G.; Brooks, B. R.; Thirumalai, D. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13403-13408.
21. Tirado-Rives, J.; Orozco, M.; Jorgensen, W. L. *Biochemistry* **1997**, *36*, 7313-7329.

22. Gao, M.; She, Z.-S.; Zhou, R. *J. Phys. Chem. B* **2010**, *114*, 15687-15693.

23. Paul, S.; Paul, S. *J. Phys. Chem. B* **2015**, *119*, 9820-9834.

24. Sarma, R.; Paul, S. *J. Phys. Chem. B* **2013**, *117*, 677-689.

25. Caballero-Herrera, A.; Nordstrand, K.; Berndt, K. D.; Nilsson, L. *Biophys. J.* **2005**,*89*, 842-857.

26. Kokubo, H.; Hu, C. Y.; Pettitt, B. M. *J. Am. Chem. Soc.* **2011**, *133*, 1849-1858.

27. Otzen, D. *Biochim. Biophys. Acta* **2011**, *1814*, 562-591.

28. Garavito, R. M.; Ferguson-Miller, S. *J. Biol. Chem.* **2001**, *276*, 32403-32406.

29. Otzen, D. E.; Christiansen, L.; Schülein, M. *Protein Sci.* **1999**, *8*, 1878-1887.

30. Otzen, D. E.; Sehgal, P.; Westh, P. *J. Coll. Int. Sci.* **2009**, *329*, 273-283.

31. Gelman, H.; Perlova, T.; Gruebele, M. *J. Phys. Chem. B* **2013**, *117*, 13090-13097.

32. Guin, D.; Sye, K.; Dave, K.; Gruebele, M. *Protein Sci.* **2016**, *25*, 1061-1068.

33. Nelson, R.; Sawaya, M. R.; Balbirnie, M.; Madsen, A. O.; Riekel, C.; Grothe, R.; Eisenberg, D. *Nature* **2005**, *435*, 773-778.

34. Linding, R.; Schymkowitz, J.; Rousseau, F.; Diella, F.; Serrano, L. *J. Mol. Biol.* **2004**, *342*, 345-353.

35. Rousseau, F.; Schymkowitz, J.; Serrano, L. *Curr. Opin. Struct. Biol.* **2006** *16*, 118-126.

36. Huang, T. H.; Yang, D. S.; Fraser, P. E.; Chakrabartty, A. *J. Biol. Chem.* **2000**, *275*, 36436-36440.

37. Hartl, F. U.; Hayer-Hartl, M. *Nat. Struct. Mol. Biol.* **2009**, *16*, 574-581.

38. Merlini, G.; Bellotti, V. *N. Engl. J. Med.* **2003**, *349*, 583-596.

39. Gertz, M. A.; Rajkumar, S. V. *Curr. Treat. Options Oncol.* **2002**, *3*, 261-271.

40. Hunt, J. A.; Ingram, V. M. *Nature* **1959**, *184*, 640-641.

41. Kayed, R.; Bernhagen, J.; Greenfield, N.; Sweimeh, K.; Brunner, H.; Voelter, W.; Kapurniotu, A. *J. Mol. Biol.* **1999**, *287*, 781-796.

42. Thomas, P. J.; Qu, B. H.; Pedersen, P. L. *Trends Biochem. Sci.* **1995**, *20*, 456-459.

43. Selkoe, D. J. *Nature* **2003**, *426*, 900-904.

44. Soto, C. *FEBS Lett.* **2001**, *498*, 204-207.

45. Soto, C. *Nat. Rev. Neurosci.* **2003**, *4*, 49-60.

46. Matveeva, E. G.; Rudolph, A.; Moll, J. R.; Thompson, R. B. *ACS Chem. Neurosci.* **2012**, *3*, 982-987.

47. Dobson, C. M.; *Nature* **2002**, *418*, 729-730.

48. Forman, M. S.; Trojanowski, J. Q.; Lee, V. W. *Nat. Med.* **2004**, *10*, 1055-1063.

49. Haataja, L.; Gurlo, T.; Huang, C. J.; Butler, P. C. *Endocr. Rev.* **2008**, *29*, 303-316.

50. Abedini, A.; Schmidt, A. M. *Febs. Lett.* **2013**, *587*, 1119-1127.

51. Mo, Y.; Lei, J.; Sun, Y.; Zhang, Q.; Wei, G. *Scientific Reports* **2016**, *6*, 1-8.

52. Brender, J. R.; Krishnamoorthy, J.; Sciacca, M. F. M.; Vivekanandan, S.; D'Urso, S.; Chen, J.; Rosa, C. L.; Ramamoorthy, A. *J. Phys. Chem. B* **2015**, *119*, 2886-2896.

53. Beck Erlach, M.; Kalbitzer, H. R.; Winter, R.; Kremer, W. *Biophys. Chem.* **2019**, *254*, 106239.

54. Patel, H. R.; Pithadia, A. S.; Brender, J. R.; Fierke, C. A.; Ramamoorthy, A. *J. Phys. Chem. Lett.* **2014**, *5*, 1864-1870.

55. Zhao, J.; Yu, X.; Liang, G.; Zheng, J. *Biomacromolecules* **2011** *12*, 210-220.

56. Nanga, R. P. R.; Brender, J. R.; Vivekanandan, S.; Ramamoorthy, A. *Biochim. Biophys. Acta. Biomembr.* **2011**, *1808*, 2337-2342.

57. Daggett, V. *Chem. Rev.* **2006**, *106*, 1898-1916.

58. Fink, A. L. *Fold. Des.* **1998**, em 3, 9-23.

59. Canchi, D. R.; , Paschek, D.; García, A. E. *J. Am. Chem. Soc.* **2010**, *132*, 2338-2344.

60. Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N. *Science* **1982**, *217*, 1214-1222.

61. Canchi, D. R.; Garcia, A. E. *Annu. Rev. Phys. Chem.* **2013**, *64*, 273-293.

62. Ganguly, P.; Hajari, T.; Shea, J. E.; van der Vegt, N. F. A. *J. Phys. Chem. Lett.* **2015**, *6*, 581-585.

63. Zou, Q.; Bennion, B. J.; Daggett, V.; Murphy, K. P. *J. Am. Chem. Soc.* **2002**, *124*, 1192-1202.

64. Paul, S.; Patey, G. N. *J. Am. Chem. Soc.* **2007**, *129*, 4476-4482.

65. Gluick, T. C.; Yadav, S. *J. Am. Chem. Soc.* **2003**, *125*, 4418-4419.

66. Sarma, R.; Paul, S. *J. Chem. Phys.* **2012**, *137*, 94502-1-12.

67. Paul, S.; Paul, S. *J. Phys. Chem. B* **2015**, *119*, 1598-1610.

68. Kumar, N.; Kishore, N. *Biophys. Chem.* **2014**, *189*, 16-24.

69. Kumar, N.; Kishore, N. *J. Chem. Phys.* **2013**, *139*, 115104-1-10.

70. Guinn, E. J.; Pegram, L. M.; Capp, M. W.; Pollock, M. N.; Record, M. T. *Proc. Nat. Aca. Sci. U.S.A* **2011**, *108*, 16932-16937.

71. Bowlus, R. D.; Somero, G. N. *J Exp. Zool.* **1979**, *208*, 137-151.

72. Hanson, A. D.; Rathinasabapathi, B.; Hamberlin, B.; Guge, A. D. *Plant Physiol.* **1991**, *97*, 1199-1205.

73. Cùnovas, D.; Vargas, C.; Csonka, L. N.; Ventosa, A.; Nieto, J. J. *J. Bacteriol.* **1996**, *178*, 7221-7226.

74. Mason, T.; Blunden. G. *Bot. Mar.* **1989**, *32*, 313-316.

75. Park, Y.-I.; Gander, J. E. *Appl. Environ. Microbiol.* **1998**, *64*, 273-278.

76. Nau-Wagner, G.; Boch, J.; Ann Le Good, J.; Bremer, E. *Appl. Environ. Micro.* **1999**, *65*, 560-568.

77. Vieira, M. N.; Figueroa-Villar, J. D.; Meirelles, M. N.; Ferreira, S. T.; De Felice, F. G. *Cell Biochem. Biophys.* **2006**, *44*, 549-553.

78. Morais-de-Sa, E.; Neto-Silva, R. M.; Pereira, P. J.; Saraiva, M. J.; Damas, A. M. *Acta Crystallogr D Biol. Crystallogr.* **2006**, *62*, 512-519.

79. Necula, M.; Kayed, R.; Milton, S.; Glabe, C. G. *J. Biol. Chem.* **2007**, *282*, 10311-10324.

80. Palladini, G.; Perfetti, V.; Merlini, G. *Swiss Med. Wkly.* **2006**, *136*, 715-720.

81. Frid, P.; Anisimov, S. V.; Popovic, N. *Brain Res. Rev.* **2007**, *53*, 135-160.

82. Liu, R.; Barkhordarian, H.; Emadi, S.; Park, C. B.; Sierks, M. R. *Neurobiol. Dis.* **2005**, *20*, 74-81.

83. Liu, F. F.; Ji, L.; Dong, X. Y.; Sun, Y. *J Phys Chem B.* **2009**, *113*, 11320-11329.

84. Kanapathipillai, M.; Lentzen, G.; Sierks, M.; Park, C. B. *FEBS. Lett.* **2005**, *579*, 4775-4780.

85. Hagihara, M.; Takei, A.; Ishii, T.; Hayashi, F.; Kubota, K.; Wakamatsu, K.; Nameki, N. *FEBS Open Bio* **2012**, *2*, 20-25.

86. Markham, P.; Robson, G. D.; Bainbridge, B. W.; Trinci, A. P. *FEMS Microbiol. Rev.* **1993**, *10*, 287-300.

87. Taylor, P.; Brown, J. H. *Synthesis, Storage and Release of Acetylcholine*; Lippincott-Raven: Philadelphia, 1999.

88. Ponder, J. W.; Case, D. A. *Adv. Prot. Chem.* **2003**, *27*, 27-85.

89. Lennard-Jones, J. E. *Proc. R. Soc. London, Ser. A* **1924**, *106*, 463-477.

90. van der Waals, J. D. *Verhandelingen der Koninklijke Akademie der Wetenschappen* **1893**, *1*, 1.

91. Coulomb, C. A. *Collection de ḿemoires relatifs à la physique* **1884**, Gauthier-Villars, 569-638.

92. Verlet, L. *Phys. Rev.* **1967**, *159*, 98-103.

93. Hockney, R. W. *Meth. Comp. Phys.* **1970**, *9*, 136-211.

94. Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141-151.

95. Flyvbjerg, H.; Petersen, H. G. Error Estimates on Averages of Correlated Data. *J. Comp. Phys.* **1989**, *91*, 461-466.

96. Garcia-Perez, A.; Burg, M. B. *Hypertension* **1990**, *16*, 595-602.

97. Yancey, P. H.; Somero, G. N. *Biochem. J.* **1979**, *183*, 317-323.

98. Rossky, P. J. *Proc. Nat. Aca. Sci. U.S.A* **2008**, *105*, 16825-16826.

99. Galvão, T. C.; De. Lorenzo, V.; Cànovas, D. *Mol. Micro.* **2006**, *62*, 1643-1654.

100. Auton, M.; Bolen, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 15065-15068.

101. Auton, M.; Marcelo, L.; Holthauzen, F.; Bolen, D. W. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 15317-15322.

102. Bolen, D. W.; Rose, G. D. *Annu. Rev. Biochem.* **2008**, *77*, 339-362.

103. Stumpe, M. C.; Grubmüller, H. *J. Am. Chem. Soc.* **2007**, *129*, 16126-16131.

104. Lee, S.; Shek, Y. L.; Chalikian, T. V. *Biopolymers* **2010**, *93*, 866-879.

105. Micciulla, S.; Michalowsky, J.; Schroer, M. A.; Holm, C.; von Klitzing, R.; Smiatek, J. *Phys. Chem. Chem. Phys.* **2016**, *18*, 5324-5335.

106. Frank, H. S.; Franks, F. *J. Chem. Phys.* **1968**, *48*, 4746-4757.

107. Dill, K. A. *Biochemistry* **1990**, *29*, 7133-7155.

108. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I. et al. *J. Comput. Chem.* **2010**, *31*, 671-690.

109. Yu, W.; He, X.; Vanommeslaeghe, K.; MacKerell, A. D. Jr. *J. Comput. Chem.* **2012**, *33*, 2451-2468.

110. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

111. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926-935.

112. Smith, L. J.; Berendsen, H. J. C.; van Gunsteren, W. F. *J. Phys. Chem. B.* **2004**, *108*, 1065-1071.

113. Kuffel, A.; Zielkiewicz, J. *J. Chem. Phys.* **2010**, *133*, 035102-1-9.

114. Borgohain, G.; Paul, S. *J. Mol. Liq.* **2017**, *231*, 174-184.

115. Borgohain, G.; Paul, S. *J. Mol. Liq.* **2017**, *233*, 431-441.

116. Martinez, L.; Andrade, R.; Birgin, E. G.; Martinez, J. M. *J Comput Chem.* **2009**, *30*, 2157-2164.

117. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M. et al. *AMBER 12*; University of California: San Francisco, CA, 2012.

118. Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

119. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. *J. Mol. Graph. Modell.* **2006**, *25*, 247-260.

120. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. *J. Comp. Chem.* **2004**, *25*, 1157-1174.

121. Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283-312.

122. Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M. et al. *AMBER 14*; University of California: San Francisco, CA, 2014.

123. Tuckermar, M.; Berne, B. J. *J. Chem. Phys.* **1992**, *97*, 1990-2001.

124. Martyna, G. J.; Tobias, D. J.; Klein, M. L. *J. Chem. Phys.* **1994**, *101*, 4177-4189.

125. Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R. *J. Chem. Phys.* **1995**, *103*, 4613-4621.

126. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684-3690.

127. Hunenberger, P. H. *Adv. Polym. Sci.* **2005**, *173*, 105-149.

128. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, J. L. *J. Chem. Phys.* **1995**, *103*, 8577-8592.

129. Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327-341.

130. Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33-38.

131. Roe, D. R.; Cheatham, T. E. III. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.

132. Luzar, A.; Chandler, D. *Phys. Rev. Lett.* **1996**, *76*, 928-931.

133. Balasubramanian, S.; Pal, S.; Bagchi, B. *Phys. Rev. Lett.* **2002**, *89*, 115505.

134. Chandra, A. *Phys. Rev. Lett.* **2000**, *85*, 768-771.

135. Chandra, A. *J. Phys. Chem. B* **2003**, *107*, 3899-3906.

136. Sarma, R.; Paul, S. *J. Chem. Phys* **2011**, *135*, 174501.

137. Sarma, R.; Paul, S. *J. Phys. Chem. B* **2013**, *117*, 9056-9066.

138. Borgohain, G.; Paul, S. *J. Phys. Chem. B* **2016**, *120*, 2352-2361.

139. Borgohain, G.; Paul, S. *Mol. Simul.* **2017**, *43*, 52-64.

140. Brehm, M.; Kirchner, B. *J. Chem. Inf. Model.* **2011**, *51*, 2007-2023.

141. Chattaraj, K. G.; Paul, S. *J. Chem. Inf. Model.* **2018**, *58*, 1610-1624.

142. Chattaraj, K. G.; Paul, S. *J. Chem. Phys.* **2019**, *150*, 064501-1-17.

143. Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169-9183.

144. Darve, E.; Rodrguez-Gmez, D.; Pohorille, A. *J. Chem. Phys.* **2008**, *128*, 144120-1-13.

145. Darve, E.; Wilson, M. A.; Pohorille, A. *Mol. Simul.* **2002**, *28*, 113-144.

146. Kastner, J. *WIREs Comput. Mol. Sci.* **2011**, *1*, 932-942.

147. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

148. Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40-57.

149. Stumpe, M. C.; Grubmüller, H. *J. Phys. Chem. B* **2007**, *111*, 6220-6228.

150. Svishchev, I. M.; Kusalik, P. G. *J. Chem. Phys.* **1993**, *99*, 3049-3058.

151. Street, T. O.; Bolen, D. W.; Rose, G. D. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 13997-14002.

152. Mirsky, A. E.; Pauling, L. *Proc. Natl. Acad. Sci. U. S. A.* **1936**, *22*, 439-447.

153. Schellman, J. A. *Biophys. Chem.* **2002**, *96*, 91-101.

154. Patist, A.; Zoerb, H. *Colloids Surf. B* **2005**, *40*, 107-113.

155. Watanabe, M. *Appl. Entomol. Zool.* **2006**, *41*, 15-31.

156. Wang, A.; Bolen, D. W. *Biochemistry* **1997**, *36*, 9101-9108.

157. Hu, C. Y.; Lynch, G. C.; Kokubo, H.; Pettitt, B. M. *Proteins: Struct. Funct. Genet.* **2010**, *78*, 695-704.

158. Kim, E. E.; Varadarajan, R.; Wyckoff, H. W.; Richards, F. M. *Biochemistry* **1992**, *31*, 12304-12314.

159. Sarma, R.; Paul, S. *J. Chem. Phys.* **2013**, *139*, 034504-1-10.

160. Tirado-Rives, J.; Jorgensen, W. L. *Biochemistry* **1991**, *30*, 3865-3871.

161. Zhang, Z.; Zhu, Y.; Shi, Y. *Biophys. Chem.* **2001**, *89*, 145-162.

162. Hess, B. *Phys. Rev. E* **2002**, *65*, 031910.

163. Smith, L. J.; Daura, X.; van Gunsteren, W. F. *Proteins: Struct. Funct. Bioinform.* **2002**, *48*, 487-496.

164. Linga, B.; Suna, M.; Bia, S.; Jinga, Z.; Liub, Y. *J. Mol. Graphics Modell.* **2012**, *35*, 1-10.

165. Borgohain, G.; Mandal, B.; Paul, S. *Phys. Chem. Chem. Phys.* **2017**, *19*, 13160-13171.

166. Zhang, Z.; Zhu, Y.; Shi, Y. *Biophys. Chem.* **2001**, *89*, 145-162.

167. Paul, S.; Paul, S. *J. Phys. Chem. B* **2015**, *119*, 10975-10988.

168. Paul, S.; Paul, S. *J. Phys. Chem. B* **2015**, *119*, 9820-9834.

169. Shrake, A.; Rupley, J. A. *J. Mol. Biol.* **1973**, *79*, 351-71.

170. Record, M. T.; Anderson, C. F. *Biophys. J.* **1995**, *68*, 786-794.

171. Parsegian, V. A.; Rand, R. P.; Rau, D. C. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 3987-3992.

172. Lee, S. L.; Debenedetti, P. G.; Errington, J. R. *J. Chem. Phys.* **2005**, *122*, 204511-1-10.

173. Mark, P.; Nilsson, L. *J. Phys. Chem. A* **2001**, *105*, 9954-9960.

174. Anfinsen, C. B. *Biochem. J.* **1972**, *128*, 737-749.

175. Anfinsen, C. B. *Science* **1973** *181*, 223-230.

176. Creighton, T. E. *Curr. Biol.* **1995**, *5*, 353-356.

177. Dobson, C. M.; Šali, A.; Karplus, M. *Angew. Chem. Int. Ed.* **1998**, *37*, 868-893.

178. Dobson, C. M. *Nature* **2003**, *426*, 884-890.

179. Ritchie, K. P.; Keller, B. M.; Syed, K. M.; Lepock, J. R. *Int. J. Hyperthermia* **1994**, *10*, 605-618.

180. Privalov, P. L. *Adv. in Prot. Chem.* **1979**, *33*, 167-241.

181. Privalov, P. L. *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 281-305.

182. Heremans, K. *Ann. Rev. Biphys. Bioeng.* **1982**, *11*, 1-21.

183. Timasheff, S. N. *Annu. Rev. Biophys. Biomol. Struct.* **1993**, *22*, 67-97.

184. Jagannathan, B.; Marqusee, S. *Biopolymers* **2013**, *99*, 860-869.

185. Myers J.K. (2014) Chemical Denaturation. In: Bell E. (eds) Molecular Life Sciences. Springer, New York, NY.

186. O'Brien, E. P.; Dima, R. I.; Brooks, B.; Thirumalai, D. *J. Am. Chem. Soc.*, **2007**, *129*, 7346-7353.

187. Greene, R. F. Jr.; Pace, C. N. *J. Biol. Chem.* **1974**, *249*, 5388-93.

188. Pace, C. N. *Meth. Enzym.* **1968**, *131*, 266-280.

189. Zangi, R,; Zhou, R.; Berne, B. J. *J. Am. Chem. Soc.* **2009**, *131*, 1535-1541.

190. Kubitzki, M. B.; Groot, B. L. De. *Biophys. J.* **2007**, *92*, 4262-4270.

191. Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. *Science* **1991**, *254*, 1598-1603.

192. Frauenfelder, H.; Leeson, D. T. *Nat. Struct. Biol.* **1998**, *5*, 757-759.

193. Okabe, T.; Kawata, M.; Okamoto, Y.; Mikami, M. *Chem. Phys. Lett.* **2001**, *335*, 435-439.

194. Rhee, Y. M.; Pande, V. S. *Biophys. J.* **2003**, *84*, 775-786.

195. Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042-6051.

196. Heyda, J.; Kožíšek, M.; Bednárova, L.; Thompson, G.; Konvalinka, J.; Vondrek, J.; Jungwirth, P. *J. Phys. Chem. B* **2011**, *115*, 8910-8924.

197. Kim, S. B.; Palmer, J. C.; Debenedetti, P. G. *Proc. Nat. Aca. Sci. USA* **2016**, *113*, 8991-8996.

198. Paschek, D.; Hempel, S.; García, A. E. *Proc. Nat. Aca. Sci. USA* **2008**, *105*, 17754-17759.

199. Williams, D. V.; Byrne, A.; Stewart, J.; Andersen, N. H. *Biochemistry* **2011**, *50*, 1143-1152.

200. Gattin, Z.; Riniker, S.; Hore, P. J.; Mok, K. H.; van Gunsteren, W. F. *Prot. Sci.* **2009**, *10*, 2090-2099.

201. Day, R.; Paschek, D.; Garcia, A. E. *Proteins* **2010**, *78*, 1889-1899.

202. Borgohain, G.; Paul, S. *Comput. Theor. Chem.* **2018**, *1131*, 78-89.

203. Samson, A. L.; Knaupp, A. S.; Sashindranath, M.; Borg, R. J.; Au, A. E. -L.; Cops, E. J.; Saunders, H. M.; Cody, S. H.; McLean, C. A.; Nowell, C. J. et al. *Cell Rep.* **2012**, *2*, 889-901.

204. Yancey, P. H.; Clark, M. E.; Hand, S. C.; Bowlus, R. D.; Somero, G. N. *Science* **1982**, *217*, 1214-1222.

205. Guinn, E. J.; Pegram, L. M.; Capp, M. W.; Pollock, M. N.; Record Jr, M. T. *Proc. Nat. Aca. Sci. USA* **2011**, *108*, 16932-16937.

206. Neidigh, J. W.; Fesinmeyer, R. M.; Anderson, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425-430.

207. Maier, J. A.; Martinez, C.; Kasavajhala, K. Wickstrom, L. Hauser, K. E.; Simmerling, C. *J. Chem. Theory Comput.* **2015**, *11*, 3696-3713.

208. Streicher, W. W.; Makhatadze, G. I. *Biochemistry* **2007**, *46*, 2876-2880.

209. Iavarone, A. T.; Parks, J. H. *J. Am. Chem. Soc.* **2005** *127*, 8606-8607.

210. Ahmed, Z.; Beta, I.; Mikhonin, A.; Aleksandr, V.; MikhoninSanford, A. A. *J. Am. Chem. Soc.* **2005**, *127*, 10943-10950.

211. García, A. E.; Sanbonmatsu, K. Y. *Proc. Nat. Acad. Sci. USA*, **2002** *99*, 2782-27870.

212. Tian, J.; García, A. E. *J. Am. Chem. Soc.* **2011**, *133*, 15157-15164.

213. Kabsch, W.; Sander, C. *Biopolymers* **1983**, *22*, 2577-2637.

214. Barua, B.; Lin, J. C.; Williams, V. D.; Kummler, P.; Neidigh, J. W.; Andersen, N. H. *Protein Eng. Des. Sel.* **2008**, *21*, 171-185.

215. Wu, X.; Yang, G.; Zu, Y.; Fu, Y.; Zhou, L.; Yuan, X. *Comput. Theor. Chem.* **2011**, *973*, 1-8.

216. Goyal, S.; Chattopadhyay, A.; Kasavajhala, K.; Priyakumar, U. D. *J. Am. Chem. Soc.* **2017**, *139*, 14931-14946.

217. Ho, B. K.; Coutsias, E. A.; Seok, C.; Dill, K. A. *Protein. Sci.* **2005**, *4*, 1011-1018.

218. DeTar, D. F.; Luthra, N. P. *J. Am. Chem. Soc.* **1977**, **99**, 1232-1244.

219. Mu, Y.; Nguyen, P. H.; Stock, G. *Proteins* **2005**, *58*, 45-52.

220. Glykos, N. M. *J. Comput. Chem.* **2006**, *27*, 1765-1768.

221. England, J. L.; Haran, G. *Annu. Rev. Phys. Chem.* **2011**, *62*, 257-277.

222. Dill, K. A.; Shortle, D. *Annu. Rev. Biochem.* **1991**, *60*, 795-825.

223. Rhoades, E.; Cohen, M.; Schuler, B.; Haran, G. *J. Am. Chem. Soc.* **2004**, *126*, 14686-14687.

224. Huang, F.; Ying, L.; Fersht, A. R. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 123-127.

225. Castellino, F. J.; Barker, R. *Biochemistry* **1968**, *7*, 4135-4138.

226. Mitchinson, C.; Pain, R.; Vinson, J.; Walker, T. *Biochim. Biophys. Acta* **1983**, *743*, 31-36.

227. Decker, R. V.; Foster, J. F. *Biochemistry* **1966**, *5*, 1242-1254.

228. Siezen, R. J.; Leunissen, J. A. *Protein Sci.* **1997**, *6*, 501-523.

229. Hansen, J. H.; Petersen, S. V.; Andersen, K. K.; Enghild, J. J.; Damhus, T.; Otzen, D. *Biopolymers* **2008**, *91*, 221-231.

230. Wymore, T.; Wong, T. C. *Biophys. J.* **1999**, *76*, 1213-1227.

231. Psachoulia, E.; Bond, P. J.; Sansom, M. S. P. *Biochemistry* **2006**, *45*, 9053-9058.

232. Renthal, R.; Brancaleon, L.; Pena, I.; Silva, F.; Chen, L. Y. *Biophys. Chem.* **2011**, *159*, 321-327.

233. Böckmann, R. A.; Caflisch, A. *Biophys. J.* **2005**, *88*, 3191-3204.

234. Neale, C.; Ghanei, H.; Holyoake, J.; Bishop, R. E.; Prive, G. G.; Pomes, R. *Chem. Phys. Lipids* **2013**, *169*, 72-84.

235. Hawkins, C. A.; Alba, E. de.; Tjandra, N. *J. Mol. Biol.* **2005**, *346*, 1381-1392.

236. Huang, G. S.; Oas, T. G. *Proc. Natl Acad. Sci. USA*. **1995**, *92*, 6878-6882.

237. Kim, S. J.; Matsumura, Y.; Dumont, C.; Kihara, H.; Gruebele, M. *Biophys. J.* **2009**, *97*, 295-302.

238. Liu, F.; Gruebele, M. *J. Mol. Biol.* **2007**, *370*, 574-584.

239. Liu, F.; Gao, Y. G.; Gruebele, M. *J Mol Biol.* **2010**, *397*, 789-798.

240. Hecht, M. H.; Sturtevant, J. M.; Sauer, R. T. *Prot. Struc. Func. Bio.* **1986**, *1*, 43-46.

241. Prigozhin, M. B.; Liu b, Y.; Wirth, A. J.; Kapoor, S.; Winter, R.; Schulten, K.; Gruebele, M. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 8087-8092.

242. Liu, Y.; Prigozhin, M. B.; Schulten, K.; Gruebele, M. *J. Am. Chem. Soc.* **2014**, *136*, 4265-4272.

243. Prigozhin, M. B.; Sarkar, K.; Law, D.; Swope,W. C.; Gruebele, M.; Pitera, J. *J. Phys. Chem. B* **2011**, *115*, 2090-2096.

244. Margulis, C. J.; Stern, H. A.; Berne, B. J. *J. Phys. Chem. B* **2002**, *106*, 10748-10752.

245. Pike, C. J.; Walencewicz, A. J.; Glabe, C. G. Cotman, C. W. *Brain Res.* **1991**, *563*, 311-314.

246. Lansbury, Jr. P. T. *Biochemistry* **1992**, *31*, 6865-6870.

247. Yu, X.; Luo, Y.; Dinkel, P.; Zheng, J.; Wei, G.; Margittai, M.; Nussinov, R.; Ma, B. *J. Biol. Chem.* **2012**, *287*, 14950-14959.

248. Siddiqua, A.; Luo, Y.; Meyer, V.; Swanson, M. A.; Yu, X.; Wei, G.; Zheng, J.; Eaton, G. R.; Ma, B.; Nussinov, R.; Eaton, S. S.; Margittai, M. *J. Am. Chem. Soc.* **2012**, *134*, 10271-10278.

249. Meyer, V.; Dinkel, P. D.; Luo, Y.; Yu, X.; Wei, G.; Zheng, J.; Eaton, G. R.; Ma, B.; Nussinov, R.; Eaton, S. S.; Margittai, M. *Angew. Chem., Int. Ed.* **2014**, *53*, 1590-1593.

250. Luo, Y.; Dinkel, P.; Yu, X.; Margittai, M.; Zheng, J.; Nussinov, R.; Wei, G.; Ma, B. *Chem. Commun.* **2013**, *49*, 3582-3584.

251. Miller, Y.; Ma, B.; Nussinov, R. *Biochemistry* **2011**, *50*, 5172-5181.

252. Zhang, M. Z.; Hu, R. D.; Liang, G. Z.; Chang, Y.; Sun, Y.; Peng, Z. M.; Zheng, J. *J. Phys. Chem. B* **2014**, *118*, 7026-7036.

253. Ashburn, T. T.; Auger, M.; Lansbury Jr, P. T. *J. Am. Chem. Soc.* **1992**, *114*, 790-791.

254. National Diabetes Fact Sheet: National Estimates and General Information on Diabetes and Prediabetes in the United States; Department of Health and Human Services, Centers for Disease Control and Prevention: Atlanta, GA, 2011.

255. Barrow, C. J.; Zagorski, M. G. *Science* **1991**, *253*, 179-182.

256. Khemte'mourian, L.; Killian, J. A.; Höppener, J. W. M.; Engel, M. F. M. *Exp. Diabetes Res.* **2008**, *2008* 1-9.

257. Emma, T.; Jaikaran, A. S.; Higham, C. E.; Serpell, L. C.; Zurdo, J.; Gross, M.; Clark, A.; Fraser, P. E. *J. Mol. Biol.* **2001**, *308*, 515-525.

258. Luca, S.; Yau, W.-M.; Leapman, R.; Tycko, R. *Biochemistry* **2007**, *46*, 13505-13522.

259. Wiltzius, J. J. W.; Sievers, S. A.; Sawaya, M. R.; Cascio, D.; Popov, D.; Riekel, C.; Eisenberg, D. *Protein Sci.* **2008**, *17*, 1467-1474.

260. Bedrood, S.; Li, Y.; Isas, J. M.; Hegde, B. G.; Baxa, U.; Haworth, I. S.; Langen, R. *J. Biol. Chem.* **2012**, *287*, 5235-5241.

261. Wiltzius, J. J. W.; Sievers, S. A.; Sawaya, M. R.; Eisenberg, D. *Protein. Sci.* **2009**, *18*, 1521-1530.

262. Goldsbury, C. S.; Cooper, G. J. S.; Goldie, K. N.; Muller, S. A.; Saafi, E. L.; Gruijters, W. T. M.; Misur, M. P.; Engel, A.; Aebi, U.; Kistler, J. *J. Struct. Biol.* **1997**, *119*, 17-27.

263. Goldsbury, C.; Goldie, K.; Pellaud, J.; Seelig, J.; Frey, P.; Muller, S. A.; Kistler, J.; Cooper, G. J. S.; Aebi, U. *J. Struct. Biol.* **2000**, *130*, 352-362.

264. Kajava, A. V.; Aebi, U.; Steven, A. C. *J. Mol. Biol.* **2005**, *348*, 247-252.

265. Xu, W.; Ping, J.; Li, W.; Mu, Y. *J. Chem. Phys* **2009**, *130*, 164709.

266. Susa, A. C.; Wu, C.; Bernstein, S. L.; Dupuis, N. F.; Wang, H.; Raleigh, D. P.; Shea, J. E.; Bowers, M. T. *J. Am. Chem. Soc.* **2014**, *136*, 12912-12919.

267. Shim, S. H.; Gupta, R.; Ling, Y. L.; Strasfeld, D. B.; Raleigh, D. P.; Zanni, M. T. *Proc. Nat. Acad. Sci. USA* **2009**, *106*, 6614-6619.

268. Gazit, E. A. *FASEB J.* **2002**, *16*, 77-83.

269. Porat, Y.; Abramowitz, A.; Gazit, E. *Chem. Biol. Drug Des.* **2016**, *67*, 27-37.

270. Liu, T.; Bitan, G. *Chem. Med. Chem.* **2012**, *7*, 359-374.

271. Bieschke, J. *Neurotherapeutics* **2013**, *10*, 429-439.

272. Ehrnhoefer, D. E.; Bieschke, J.; Boeddrich, A.; Herbst, M.; Masino, L.; Lurz, R.; Engemann, S.; Pastore, A.; Wanker, E. E. *Nat. Struct. Mol. Biol.* **2008**, *15*, 558-566.

273. Meng, F.; Abedini, A.; Plesner, A.; Verchere, C. B.; Raleigh, D. P. *Biochemistry* **2010**, *49*, 8127-8133.

274. Meng, F.; Abedini, A.; Plesner, A.; Middleton, C. T.; Potter, K. J.; Zanni, M. T.; Verchere, C. B.; Raleigh, D. P. *J. Mol. Biol.* **2010**, *400*, 555-566.

275. Noor, H.; Cao, P.; Raleigh, D. P. *Protein Sci.* **2012**, *21*, 373-382.

276. Guo, J.; Sun, W.; Li, L.; Liu, F.; Lu, W. *RSC Adv.* **2017**, *7*, 43491-43501.

277. Xing, Y.; Pilkington, E. H.; Wang, M.; Nowell, C. J.; Kakinen, A.; Sun, Y.; Wang, B.; Davis, T. P.; Ding, F.; Ke, P. C. *Phys. Chem. Chem. Phys* **2017**, *19*, 30627-30635.

278. Berhanu, W. M.; Hansmann, U. H. E. *PLoS One* **2014**, *9*, 97051.

279. Zou, Y.; Qian, Z.; Sun, Y.; Wei, G.; Zhang, Q. *J. Phys. Chem. B* **2017**, *121*, 9203-9212.

280. Rosenman, D. J.; Wang, C. Garciá, A. E. *J. Phys. Chem. B* **2016**, *120*, 259-277.

281. Tenidis, K. Waldner, M.; Bernhagen, J.; Fischle, W.; Bergmann, M.; Weber, M.; Merkle, M. L.; Voelter, W.; Brunner, H.; Kapurniotu, A. *J. Mol. Biol.* **2000**, *295*, 1055-1071.

282. Sharma, B.; Paul, S. *J. Phys. Chem. B* **2016**, *120*, 9019-9033.

283. Frishman, D.; Argos, P. *Proteins: Struct. Funct. Genet.* **1995**, *23*, 566-579.

284. Liang, G.; Zhao, J.; Yu, X.; Zheng, J. *Biochemistry* **2013**, *52*, 1089-1100.

285. Paul, S. and Patey, G. N. *J. Phys. Chem. B* **2008**, *112* 11106-11111.

286. Mukhopadhyay, T. K.; Datta, A. *J. Phys. Chem. B* **2017**, *121*, 811-822.

287. Global, Regional, and National Life Expectancy, All-cause Mortality, and Cause-specific Mortality for 249 Causes of Death, 1980-2015: a Systematic Analysis for the Global Burden of Disease Study 2015. *Lancet.* **2016** *388*, 1459-1544.

288. Vivekanandan, S.; Brender, J. R.; Lee, S. L.; Ramamoorthy, A. *Biochem. Biophys. Res. Commun.* **2011**, *411*, 312-316.

289. Goedert, M.; Spillantini, M. G. *Mol. Brain.* **2017**, *10*, 18-26.

290. Lippens, G.; Sillen, A.; Landrieu, I.; Amniai, L.; Sibille, N.; Barbier, P.; Leroy, A.; Hanoulle, X.; Wieruszeski, J. M. *Prion* **2007**, *10*, 21-25.

291. Sharma, B.; Kalita, S.; Paul, A.; Mandal, B.; Paul, S. *RSC Adv.* **2016**, *6*, 78548-78558.

292. Doig, A. J.; Castillo-Frias, M. P. D.; Berthoumieu, O.; Tarus, B.; Nasica-Labouze, J.; Sterpone, F.; Nguyen, P. H.; Hooper, N. M.; Faller, P.; Derreumaux, P. *ACS Chem. Neurosci.* **2017**, *8*, 1435-1437.

293. Lin, Y. S.; Bowman, G. R.; Beauchamp, K. A.; Pande, V. S. *Biophys. J.* **2012**, *102*, 315-324.

294. Balbach, J. J.; Ishii, Y.; Antzutkin, O. N.; Leapman, R. D.; Rizzo, N. W.; Dyda, F.; Reed, J.; Tycko, R. *Biochemistry* **2000**, *39*, 13748-13759.

295. Sgourakis, N. G.; Merced-Serrano, M.; Boutsidis, C.; Drineas, P.; Du, Z.; Wang, C.; Garcia, A. E. *J. Mol. Biol.* **2011**, *405*, 570-583.

296. Huet, A.; Derreumaux, P. *Biophys. J.* **2006**, *91*, 3829-3840.

297. Davis, C. H.; Berkowitz, M. L. *Proteins* **2010**, *78*, 2533-2545.

298. Li, H.; Luo, Y.; Derreumaux, P.; Wei, G. *Biophys. J.* **2011**, *101*, 2267-2276.

299. Gnanakaran, S.; Nussinov, R.; Garciá, A. E. *J. Am. Chem. Soc.* **2006**, *128*, 2158-2159.

300. Wu, C.; Wang, Z.; Lei, H.; Duan, Y.; Bowers, M. T.; Shea, J. E. *J. Mol. Biol.* **2008**, *384*, 718-729.

301. Nguyen, P. H.; Li, M. S.; Stock, G.; Straub, J. E.; Thirumalai, D. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 111-116.

302. Ma, B.; Nussinov, R. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14126-14131.

303. Soto, P.; Griffin, M. A.; Shea, J. E. *Biophys. J.* **2007**, *93*, 3015-3025.

304. Favrin, G.; Irbäck, A.; Mohanty, S. *Biophys. J.* **2004**, *87*, 3657-3664.

305. Tjernberg, L. O.; Naslund, J.; Lindqvist, F.; Johansson, J.; Karlstrom, A. R.; Thyberg, J.; Terenius, L.; Nordstedt, C. *J. Biol. Chem.* **1996**, *271*, 8545-8548.

306. Nilsberth, C.; Westlind-Danielsson, A.; Eckman, C. B.; Condron, M. M.; Axelman, K.; Forsell, C.; Stenh, C.; Luthman, J.; Teplow, D. B.; Younkin, S. G. et al. *Nat. Neurosci.* **2001**, *4*, 887-893.

307. Kayed, R.; Head, E.; Thompson, J. L.; Mclntire, T. M.; Milton, S. C.; Cotman, C. W.; Glabel, C. G. *Science* **2003**, *300*, 486-489.

308. Nilsson, K. P. *FEBS Lett.* **2009**, *583*, 2593-2599.

309. Kukar, T. L. *Nature* **2008**, *453*, 925-929.

310. Chen, J.; Armstrong, A. H.; Koehler, A. N.; Hecht, M. H. *J. Am. Chem. Soc.* **2010**, *132*, 17015-17022.

311. Hong, Y.; Meng, L.; Chen, S.; Leung, C. W. T.; Da, L.-T.; Faisal, M.; Silva, D.-A.; Liu, J.; Lam, J. W. Y.; Huang, X. et al. *J. Am. Chem. Soc.* **2012**, *134*, 1680-1689.

312. Giger, K.; Vanam, R. P.; Seyrek, E.; Dubin, P. L. *Biomacromolecules* **2008**, *9*, 2338-2344.

313. Cabaleiro-Lago, C.; Quinlan-Pluck, F.; Lynch, I.; Dawson, K. A.; Linse, S. *ACS Chem. Neurosci.* **2010**, *1*, 279-287.

314. Pujadas, L.; Rossi, D.; Andres, R.; Teixeira, C. M.; Serra-Vidal, B.; Parcerisas, A.; Maldonado, R.; Giralt, E.; Crulla, N.; Soriano, E. *Nat. Commun.* **2014**, *5*, 3443-1-11.

315. Bose, P.; Chatterjee, U.; Nerelius, C.; Govender, T.; Norström, T.; Gogoll, A.; Sandegren, A.; Gthelid, E.; Johansson, J.; Arvidsson, P. I. *J. Med. Chem.* **2009** *52*, 8002-8009.

316. Amijee, H.; Bate, C.; Williams, A.; Virdee, J.; Jeggo, R.; Spanswick, D.; Scopes, D. I. C.; Mazzitelli, J.; Doig, A. *Biochemistry* **2012**, *51*, 8338-8352.

317. Jiang, L.; Liu, C.; Leibly, D.; Landau, M.; Zhao, M.; Hughes, M. P.; Eisenberg, D. S. *Elife.* **2013**, *2*, 857-883.

318. Ehrnhoefer, D. E.; Bieschke, J.; Boeddrich, A.; Herbst, M.; Masino, L.; Lurz, R.; Engemann, S.; Pastore, A.; Wanker, E. E. *Nat. Struct. Mol. Biol.* **2008**, *15*, 558-566.

319. Wang, S. H.; Liu, F. F.; Dong, X. Y.; Sun, Y. *J. Phys. Chem. B* **2010**, *114*, 11576-11583.

320. Liu, F. F.; Dong, X. Y.; He, L.; Middelberg, A. P.; Sun, Y. *J. Phys. Chem. B* **2011**, *115*, 11879-11887.

321. Sato, M.; Murakami, K.; Uno, M.; Nakagawa, Y.; Katayama, S.; Akagi, K.; Masuda, Y.; Takegoshi, K.; Irie, K. *J. Biol. Chem.* **2013**, *288*, 23212-23214.

322. Palhano, F. L.; Lee, J.; Gimster, N. P.; Kelly, J. W. *J. Am. Chem. Soc.* **2013**, *135*, 7503-7510.

323. Scherzer-Attali, R.; Pellarin, R.; Convertino, M.; Frydman-Marom, A.; Egoz-Matia, N.; Peled, S.; Levy-Sakin, M.; Shalev, D. E.; Caflisch, A.; Gazit, E.; Segal, D. *PLoS One* **2010**, *5*, e11101.

324. Zou, Y.; Qian, Z.; Chen, Y.; Qian, H.; Wei, G.; Zhang, Q. *ACS Chem. Neurosci.* **2019**, *10*, 1585-1594.

325. Nasica-Labouze, J.; Nguyen, P. H.; Sterpone, F.; Berthoumieu, O.; Buchete, N. V.; Coté, S.; Simone, A. D.; Doig, A. J.; Faller, P.; Garcia, A. et al. *Chem. Rev.* **2015**, *115*, 3518-3563.

326. Carballo-Pacheco, M.; Ismail, A. E.; Strodel, B. *J. Chem. Theory Comput.* **2018**, *14*, 6063-6075.

327. Carballo-Pacheco, M.; Strodel, B. *Protein Sci.* **2017**, *26*, 174-185.

328. Man, V. H.; Nguyen, P. H.; Derreumaux, P. *J. Phys. Chem. B* **2017**, *22*, 5977-5987.

329. Zheng, W.; Tsai, M. Y.; Wolynes, P. G. *J. Am. Chem. Soc.* **2017**, *139*, 16666-16667.

330. Matthes, D.; Gapsys, V.; Brennecke, J. T.; de Groot, B. L. *Sci. Rep.* **2016**, *6*, 33156-1-19.

List of Publications

1. Paul, S.; Paul, S. How Does Aqueous Choline-*O*-Sulfate Solution Nullify the Action of Urea in Protein Denaturation? *J. Chem. Info. and Mod.* **2018**, *58*, 1858-1869.

2. Paul, S.; Paul, S. Inhibitory Effect of Choline-*O*-sulfate on Aβ_{16-22} Peptide Aggregation: A Molecular Dynamics Simulation Study. *J. Phys. Chem. B* **2019**, *123*, 3475-3489.

3. Paul, S.; Paul, S. Molecular Dynamics Simulation Study on the Inhibitory Effects of Choline-*O*-sulfate on hIAPP Protofibrilation. *J. Comput. Chem.* **2019**, *40*, 1957-1968.

4. Paul, S.; Paul, S. The Conformational Stability of Terminal Helices of λ-repressor Protein in Aqueous Dodine and Choline-*O*-sulfate Solutions. *Int. J. Biol. Macromol.* **2020**, *154*, 1332-1346.

5. Paul, S.; Paul, S. In Silico Study of Osmolytic Effects of Choline-*O*-sulfate on Urea Induced Unfolding of Trp-cage Mini-protein: An Atomistic View from Replica Exchange Molecular Dynamics Simulation. *Arch. Biochem. Biophys. doi: 10.1016/j.abb.2020.108484* (Article in press).

6. Paul, S.; Kumari, K.; Paul, S. Molecular Insight into the Effects of Enhanced Hydrophobicity on Amyloid-like Aggregation. *J. Phys. Chem. B.* *doi.org/10.1021/acs.jpcb.0c06000* (Article in press).

7. Paul, S.; Paul, S. Molecular Insights into the Urea-Choline-*O*-sulfate Interaction in Aqueous Solution Elucidating Their Biological Roles. *J. Phys. Chem. B.* (Manuscript under revision).

8. Paul, S.; Paul, S. Self-assembly of Human Calcitonin and Its Inhibition by Its Most Relevant Derivatives. (Manuscript under preparation).

Conferences Attended

1. Presented a poster entitled "Effect of Choline-*O*-sulfate as a Bioprotectant on Urea Induced Protein Denaturation: A Molecular Dynamics Simulation Study" in the conference " Emerging Trends in Chemical Sciences (ETCS-2018)" held at Dibrugarh University, Assam, India.

2. Presented a poster entitled "Effect of Choline-*O*-sulfate as a Bioprotectant on Urea Induced Protein Denaturation: A Molecular Dynamics Simulation Study" in the conference "Frontier in Chemical Sciences (FICS-2018)" held at IIT Guwahati, Assam, India.

2. Presented a poster entitled "Molecular Dynamics Simulation Study on the Inhibitory Effects of Choline-*O*-sulfate on hIAPP Protofibrilation" in the conference "Theoretical Chemistry Symposium (TCS-2019)" held at BITS Pilani, Pilani, Rajasthan, India.

4. Presented a poster entitled "Molecular Dynamics Simulation Study on the Inhibitory Effects of Choline-*O*-sulfate on hIAPP Protofibrilation" in the conference "Research Conclave (RC-2019)" held at IIT Guwahati, Assam, India.