EFFECT OF IONS AND TEMPERATURE ON THE STRUCTURAL AND DYNAMIC ASPECTS OF BIOMOLECULES

Thesis

Submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

by

OMKAR SINGH

(Reg. No: 187081CY005)

DEPARTMENT OF CHEMISTRY

NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA

SURATHKAL, MANGALURU- 575025

MAY 2024

DECLARATION

By the Ph.D Research Scholar

I hereby declare that the Research Thesis entitled **"Effect of Ions and Temperature on the Structural and Dynamic Aspects of Biomolecules"** which is being submitted to the National Institute of Technology Karnataka, Surathkal in partial fulfillment of the requirements for the award of the degree of **Doctor of Philosophy** in Chemistry is a *bonafide report of the research work carried out by me*. The material contained in this Research Thesis has not been submitted to any University or Institution for the award of any degree.

OMKAR SINGH
Reg.No. 187081CY005
Department of Chemistry

Place: NITK, Surathkal
Date: 02/05/2024

CERTIFICATE

This is to *certify* that the Research Thesis entitled **"Effect of Ions and Temperature on the Structural and Dynamic Aspects of Biomolecules"** submitted by **Omkar Singh**, (Register Number: 187081CY005) as the record of the research work carried out by him, is *accepted as the Research Thesis submission* in partial fulfilment of the requirements for the award of degree of Doctor of Philosophy.

Debashree Chakraborty 03/05/2024

Dr Debashree Chakraborty
Research Guide
(Name and Signature with Date)

06/05/2024

Chairman – DRPC
(Signature with Date and Seal)

H.O.D.
Chemistry Dept.
NITK SURATHKAL
MANGALURU-575 025, KARNATAKA

ACKNOWLEDGEMENTS

There are no proper words to convey my deep gratitude and respect for my research advisor, Dr. Debashree Chakraborty, Associate Professor, Department of Chemistry, NITK. She has inspired me to become an independent researcher and helped me realize the power of critical reasoning. She also demonstrated what a brilliant and hard-working scientist can accomplish. I vow my immense gratitude to her for giving me an opportunity to work under her supervision and for providing the necessary facilities to work on this research project.

My sincere gratitude is due towards my RPAC members, Dr. D Krishna Bhat, Professor, Department of Chemistry, and Dr. Jagannathan T Kalathi, Asst. Professor, Department of Chemical Engineering for their timely assessment and evaluation of my research progress. Their valuable inputs at various stages of my work have contributed greatly in giving the final shape to my research work.

I'm grateful to Dr. Darshak R Trivedi, Head of the Department & Professor, and Dr. Uday Kumar D, Dr. Arun M Isloor, and Dr. D Krishna Bhat, former Heads of the Department of Chemistry for providing the administrative facilities and infrastructure. I also thank all the faculty members of the Department of Chemistry along with the faculty of the Computer Science and Engineering Department Dr. Basavaraj Talawar, NITK Surathkal for their support.

I acknowledge NITK, Surathkal for providing the research fellowship, laboratory facilities, and financial support necessary for the completion of my research work. Also, I'm thankful to the Department of Physics and Department of Computer Science & Engineering, NITK Surathkal for additional computational facilities and PARAM Utkarsh.

I sincerely thank my senior, Dr. Dilip. H. N, Dr. Pushyaraga PV, and all other lab members Dr. Bratin Kumar Das, Apoorva Mathur, Anjana VM, Divya P, and Samrat Sarkar for their constant support at every stage of my research. I am thankful to

all the research scholars and non-teaching staff within and outside the department for their valuable assistance. I owe a deep gratitude to the professionals and the well-wishers within the country and abroad who have guided and helped me during my research work.

I would like to express my gratitude to my in-laws Mr. C K Venugopalan and Mrs. P S Pushpakala, for their unfailing emotional support.

I deeply thank my beloved parents, Mr. Charan Singh and Mrs. Kastoori Devi for their unconditional trust, timely encouragement, and endless patience. It was their love that raised me again when I got weary. The family of my brothers Mr. Kapil Kumar, and Satendra Singh, and my nephews Shivakant Singh, and Priyanshi Singh have also been generous with their love and encouragement despite the long distance between us.

I thank with love from the bottom of my heart to Dr. Pushyaraga P V and Adhyansh Singh, my wife and son. Understanding me best, as a Ph.D. herself, Dr. P V has been my best friend and great companion, loved, supported, encouraged, entertained, and helped me get through this agonizing period in the most positive way.

Finally, I thank all those who have helped me directly or indirectly and extended their support in the completion of my doctoral research work.

Thanking you.

OMKAR SINGH

ABSTRACT

The atomic-level structural and dynamical study of biomacromolecules to understand the biomolecular mechanisms poses a major challenge in molecular biophysics. The biological activity of many macromolecules in the cellular environment depends on the individual or combined effect of external factors such as the presence of mono/di-valent cations, water, and temperature. These factors can trigger the constant motion of the atoms in biomolecules due to inter and intra-molecular interactions, which in turn changes the dynamic aspects. The interactions of biomolecules and biomolecular complexes such as protein-protein, protein-DNA, and protein-RNA with water and solvent molecules have a vital role in the biological system. Water plays an important role in biological processes as a primary solvent. The structure of liquid water can be altered due to its interaction with biomolecules, metal ions, cosolvents, and under thermodynamic conditions. In principle, Molecular Dynamics simulations typically generate substantial insight into biomolecular functions by providing the dynamics of water/hydrated ions and their interactions with the biomolecules. Also, the formation of a strong interaction between the protein-protein, protein-RNA, protein-ions, and protein-water complex can be easily predicted using computer simulations.

To understand the various interactions, several protein-protein, protein-water, protein-ions, and RNA-water systems were analyzed in this research work. This includes the study of hydration shell properties near antimicrobial peptide in the presence of metal ions, the influence of the ion specificity and temperature dependence on the structure of intrinsically disordered sheep prion peptide, conformational evolution, and the effect of the ion-counter ion based on the different model potential on the stability of Frameshift Stimulation Element (FSE) of SARS CoV-2 RNA genome. The results show that the water molecules in a hydrophilic environment are more disruptive, containing broken hydrogen bonds compared to the hydrophobic environment. The hydrophilic and hydrophobic amino acids attribute to the formation of different density regions of water molecules near the peptide surface in the presence of ions. The changes in the secondary structure elements of intrinsically disordered peptides (IDP) are found to be more sensitive in the presence of ions. The transition

from the disordered-to-ordered structure with temperature was thoroughly studied based on the interaction of the hydrophobic and hydrophilic residues with water. Insights into the structural changes of IDP with ions and temperature possess great clinical significance because of its pathogenicity. The temperature dependence of the FSE of the SARS CoV-2 RNA genome shows interchangeable conformations with more stability at lower temperatures compared to higher temperatures. The RNA structure was found to be more dynamic and transient in the mTIP3P water model. The formation of the ion-contact pair near the negatively charged phosphate group (Na^+-PO_4^-) leads to strong RNA-ion interaction due to the lower dielectric constant of the SPC/E model helps to make strong hydration shells with higher hydrogen bond lifetime. The survival probability of ions near the RNA surface and the strength of RNA-Water interactions tend to decrease with the temperature rise. Thus the individual/combined effect of temperature, ions, and water decides the biological activity of the RNA genome. In conclusion, the research work successfully addressed the structural and dynamic behavior of biomolecules, peptides, and RNA, in an aqueous medium with an ionic effect, temperature effect, or both. The outcomes extend the current knowledge about the structure-function relationship and can assist in various biological applications.

Keywords: Antimicrobial Peptide, Intrinsically Disordered Sheep Prion Peptide, SARS CoV-2 RNA Genome, Metal Ions, Molecular Dynamics Simulation, Replica Exchange Molecular Dynamics, Metadynamics, Free Energy Surface, Principle Component Analysis, Network Analysis, Radial Distribution Function, Tetrahedral Order Parameter, H-bond Lifetime, Diffusion Coefficient

INDEX

CHAPTER 3

PREFERENTIAL BINDING AFFINITY OF IONS AND THEIR EFFECT ON STRUCTURE AND DYNAMICS OF WATER NEAR ANTIMICROBIAL PEPTIDE

CHAPTER 4

INFLUENCE OF ION SPECIFICITY AND CONCENTRATION ON THE CONFORMATIONAL TRANSITION OF INTRINSICALLY DISORDERED SHEEP PRION PEPTIDE

CHAPTER 7

SENSITIVITY OF THE RNA CONFORMATIONS ON THE WATER MODELS: EFFECT OF THE HYDRATION SHELLS

List of Figures

List of Tables

List of Appendix

CHAPTER 1

INTRODUCTION

In this section, an incisive introduction of water-biomolecules, protein-ion interactions, temperature effect on the biomolecules, the functional role of biomolecules in biological processes have been briefly discussed. Also, a brief description of the peptides, and nucleic acids, followed by a concise literature review, biomolecular systems of interest, scope, and objectives of the present research work have been explored.

1.1 BACKGROUND

Biomolecules are essential for living organisms as it assists the organisms to grow, sustain and reproduce, regardless of the life forms. The biological function of biomolecules in their cellular environment is greatly affected by various external factors such as water, ions, cosolvents, temperature, pH, etc (Rozhkov and Goryunov 2014). Biomolecules, such as proteins, nucleic acids, and lipids, are essential components of living systems, and their structure and function are intimately linked to their environment. The variations in the concentration of naturally occurring ions in the cell affect biomolecules in several ways by inactivating a process or inducing pathogenicity. The interaction of ions with the different hydrophobic and hydrophilic residues of the protein perturbs the structure of the water molecules near their protein surface. The distortions in the arrangement of water molecules and loss of tetrahedrality of the interfacial water molecules can influence the dynamic properties of biomolecules. Similarly, it is found that biomolecules are highly sensitive to changes in temperature and even small fluctuations can disrupt their structure and function (Wang 1969). Interestingly, the thermal fluctuations in protein and nucleic acid residues can affect the native structure which results in loss of biological activity.

The structure-to-function relationship of biomolecules is determined by a delicate balance of forces, including hydrogen bonding, hydrophobic interactions,

electrostatic interactions, and van der Waals forces. The external factors influence the structure of biomolecules or biomolecular complexes, changing their biological functions. It is known that 70% of biological cells in the human body consist of water; the rest constitutes ions and other biomolecules. Thus the structure and dynamics of water around the biomolecules can be perturbed due to the presence of ions and cosolvents. The effect of ions and temperature on biomolecules and surrounding water molecules are complex and the specific characteristics exhibited by the biomolecules under above-mentioned conditions are still in question.

1.2 EFFECTS OF WATER ON BIOMOLECULES AND THEIR SIGNIFICANCE IN BIOLOGICAL PROCESS

Biomolecules and water are two indispensable parts. The stability of the biomolecules depends on various factors such as the effect of inter-peptide hydrogen bonds (Pace et al. 2014), side-chains, charge (Nick Pace et al. 2014) etc. Among all of these, one very important factor is stability due to the presence of the water molecules nearby it. Water is known as a universal solvent due to its unique and abnormal properties. Its uniqueness is due to the presence of four-fold hydrogen bonds (Lu et al. 2008; da Silva et al. 2011). The hydrogen bonding nature of the water gets modulated by the presence of different solutes molecules such as ions, biomolecules and also by different thermodynamic conditions such as temperature, pressure (Biswas and Mallik 2014). The water molecules found to be on the surface of a protein or nucleic acids are known as hydration water or biological water (Denisov and Halle 1996). The stability of the 3D structure of the biomolecules such as protein and nucleic acids depends on the intramolecular interactions and intermolecular interactions with water, ions etc. Biomolecules, on the other hand, can perturb the structure of the water molecules near their surface due to the presence of the different hydrophobic and hydrophilic residues (Chaudhary and Kaphle 2023; Persson and Halle 2018). It is hypothesized that water has two polymorphic forms which coexist and can be distinguished based on the difference in the hydrogen bond network. The two distinct liquid phases of water are Low-density water (LDW, density 0.91 g/ml) which consists of mainly tetrahedral structure, and high-density water (HDW, density 1.2 g/ml) with distorted tetrahedral structure (Galamba 2013; Mähler and Persson 2012;

Mancinelli et al. 2007; Marcus 2009; Nilsson and Pettersson 2015; Poole et al. 1992; Vedamuthu et al. 1994). In contrast, a continuous distribution model for water is also proposed (Niskanen et al. 2019). These water molecules are important for the structure, stability, dynamics, and function of biological macromolecules (Jungwirth 2015; Levy and Onuchic 2006).

Most of the water molecules participate in hydrogen bond networks in which the polar group of the proteins makes the hydrogen bond with water molecules which are responsible for electrostatic interactions (Curtis et al. 1998). Moreover, the protein chain can make the water-mediated hydrogen bond and stabilize the hydrophilic and hydrophobic surface of the proteins. The hydrophobic patches are more important to direct the protein-protein association compared to protein-water. The enthalpy gain from the water-mediated hydrogen bond is greater than the entropic amount that must be disabled for interfacial water. Therefore, an equilibrium can be observed between the "bound" water and "free" water (Nandi and Bagchi 1998).

$$\text{Bound water } (K_1) \leftrightarrow \text{Free water } (K_2)$$

The equilibrium constant is related to the free energy difference between the molecules, with the usual thermodynamic relation, $\Delta G = -RT\ ln(K)$.

The cooperative binding of the water molecules with biomolecules leads to the conformational transitions of biomolecules to low and high binding affinity and helps to regulate the biomolecular process. The interactions of biomolecules and biomolecular complexes such as protein-protein, protein-DNA, and protein-RNA with water and other solvent molecules have a vital role in the biological system. The formation of a strong interaction between the protein-protein, protein-DNA/RNA, protein-ions, and protein-water complex follows a thermodynamic balance. The change in free energy (ΔG) during the interaction and forming of the complex depends upon the change in entropy (ΔS) and enthalpy (ΔH).

$$\Delta G = \Delta H - T\Delta S \tag{1.1}$$

where T is temperature. The enthalpy term is modulated with the non-covalent interactions (hydrogens bonds, electrostatic interactions), whereas the entropy term is changed with the variations in biomolecular conformations, perturbation in water structure, counter-ions, and other physical quantity. Water molecules near the apolar groups are found to have more crystallinity (Starr and Sciortino 2014) than bulk

(known as the iceberg model (Galamba 2013; Huraskin and Horn 2019; Laage et al. 2009). They are also found to behave differently in presence of various cosolvents and ions (Dilip and Chakraborty 2019; Huraskin and Horn 2019). The effect of ions on water structure and biological process is briefly discussed in next section.

1.3 IMPORTANCE OF METAL IONS IN BIOLOGY

In plants and animals, inorganic ions are necessary for the vital cellular activities such as muscle contraction, regulation of osmotic pressure and water content, neuron activation etc. Ions, such as lithium, sodium, potassium, calcium, and magnesium, are present in biological systems and can affect biomolecules in several ways. For example, ions can interact with charged amino acids in proteins and affect their conformation and stability (Collins 2004). Ions bind to nucleic acids can influence their structure and function. Additionally, ions can affect the structure and properties of lipid membranes. With a significantly higher ionic concentration, the solubility of the peptide/protein decreases known as the "salting out" effect whereas the decrease in ionic strength led to "salting in" (Dill et al. 2005; Shinto et al.2005). Salts affect the electrostatic interactions in solution and influence the structure evolution of the biomolecules by forming the ion-ion contact pair.

It is often seen that the water contains dissolved salts which play major roles in many chemical and biological processes (Jungwirth 2015; Levy and Onuchic 2006) for example, alkali metal ions are known to play an important role in stabilizing proteins by changing their hydration environment (Huraskin and Horn 2019). The most abundant metal cations such as Na^+, K^+, $Ca^{2+,}$ and Mg^{2+} have a significant effect on the structure, stability, and function of nucleic acids and proteins (Friedman et al. 2005). They also help in the regulation of important biomolecular processes (Anderson and Record 1995; Hribar et al. 2002). Salts affect the electrostatic interactions in a solution (Martins de Oliveira et al. 2018; Thomas and Elcock 2006) and thereby influences the solubility of the amino acids and proteins by altering the liquid water structure. Experiments have shown that water molecules are ordered by small or multivalent ions and disordered by large monovalent ions (Chen et al. 2017; Hribar et al. 2002). The binding affinity of ions towards weak carboxylic acids

increases due to the size of ions. The smaller cation ($Li^+ > Na^+ > K^+$) shows strong attraction between the cations and acetate in solvent-shared configurations (Hess and van der Vegt 2009). The hydration radius of the cation is on contrary, proportional to the original size of the ion because the small ions have more hydration energy and can accept more water molecules. Water ordering is in fact, related to the ion charge densities by the Hofmeister effect (Cacace et al. 1997; Collins 2004, 1997; Collins and Washabaugh 1985). Further, ions are classified as either kosmotropes (structure makers) or chaotropes (structure breakers) according to their relative abilities to induce the structuring of water (Dill et al. 2005; Mancinelli et al. 2007). The ion-protein interaction is more attractive for chaotropic ions whereas the ion-protein interaction is more repulsive for kosmotropic ions than uncomplexed proteins. Ion pairs are temporary associations of ions held together by charge attraction. Ion-pair formation and disassociation are fast reactions in dynamic equilibrium. The formation of ion pairs increases the solubility of an electrolyte. The low salt concentrations, the Debye–Hückel term dominates, and salts thus shift the equilibrium toward the solution phase, i.e., an increase in solubility. On the other hand, at high salt concentrations, the contribution from hydration of the ions dominates and salts then shift the equilibrium toward the solid phase, i.e., a decrease in solubility.

The importance of metal ions in life cannot be overlooked. The interaction between metal ions and biomolecules is an essential for various biological processes (Ussing 1959). The solubility of the biomolecules can be influenced due to the ionic strength, polarity of the solvent, temperature and acidity or alkalinity of the system (Lu et al. 2008). The strength of ionic salt solutions in aqueous media and thermodynamic conditions strongly changes the microscopic properties such as diffusion, entropy, hydrogen bond dynamics and macroscopic properties such as density, volume, structural parameters of the solution. The temperature induced conformational changes of biomolecules and the associated water dynamics are discussed in the next section.

1.4 INFLUENCE OF TEMPERATURE ON BIOMOLECULES

Temperature is a significant variable parameter that shows different behaviour at extreme cases and significantly affects the structure and function of biomolecules at

molecular level. Therefore, it is necessary to understand the structural conformation of biomolecules projected under thermal conditions. The temperature effect can either alter or break the non-covalent interactions thereby causing changes in the three-dimensional folding of biomolecules. The mechanism behind the thermal stability of natural biomolecules such as proteins and nucleic acids is still particularly challenging, for a wide range of temperatures (Julió Plana et al. 2019). The alterations in protein conformations, stability, flexibility, and protein unfolding can occur with an increase as well as decrease in temperature. At low temperatures, they are less flexible, and their structure is more rigid, which can be termed as cold denaturation. In case of proteins, at low temperature, the more hydrophobic nature of water solvates the hydrophobic regions resulting the unfolding of 3D structure. At higher temperatures, flexibility increases and they can undergo conformational changes that can affect their function. Experimentally, the thermal folding/unfolding of the protein or nucleic acids is commonly related to the melting temperature (T_m). Similarly, the theoretical prediction of T_m of the biomolecules is calculated with the help of kinetic of folding/unfolding from secondary structure elements.

The dynamic nature of water molecules at the interface of proteins or nucleic acids is crucial for maintaining the biological activity. The water molecules near the surface of biomolecules mainly get perturbed compared to the bulk due to changes in the arrangement of the hydrogen bond network. Hydration water plays two important roles in biological processes by stabilizing the native states of the biomolecules and participating in biological reactions. Hydration shell is intimate to biomolecules that dominates the interactions at short range with the metal ions and water molecules. Also, at different temperatures, the exchange rate of interfacial water molecules, are rare and transient event on the molecular time scale. The interfacial water molecules may favour or disfavour the energetics of biological process. Generally, this fact is overlooked while studying the mechanism of temperature induced conformational changes. Theoretical studies proved to be an efficient way to interpret the minute details of temperature induced changes on water dynamics.

The research work mainly focus on the ionic and temperature dependence on the structural and the dynamic aspects of biomolecules such as proteins and nucleic

acids. These are called the biological building blocks which constitute the basis of living organisms. All cells consist of four fundamental biomacromolecules: nucleic acids, proteins, lipids and carbohydrates. Out of these, nucleic acids and proteins are essential to carry forward the genetic information (central dogma): DNA → RNA → Protein or RNA → Protein. The basic concepts of the protein/peptides and the nucleic acids have been discussed in the following section.

1.5 BIOLOGICAL BUILDING BLOCKS

The biomolecules such as Proteins, DNA, RNA, etc are organic macromolecules that have a significant role in living things including viruses and bacteria. They are essential for normal cell functions. A major aim of biophysics is to understand the biomolecular processes of living organisms and how the uniqueness of protein/nucleic acid structures (secondary, tertiary structures) perform a specific function. Proteins are biopolymers of amino acid residues linked by peptide (amide) bonds where as RNA, DNA are biopolymers composed of nucleotides.

1.5.1 Proteins

Proteins or Peptides are one of the important and decisive biological building blocks. It is considered a macromolecule due to its significantly large size. The secondary or tertiary structure of proteins mainly contain multiple units of amino acids. Therefore, it can be said that proteins are the polymer of structural units of amino acids.

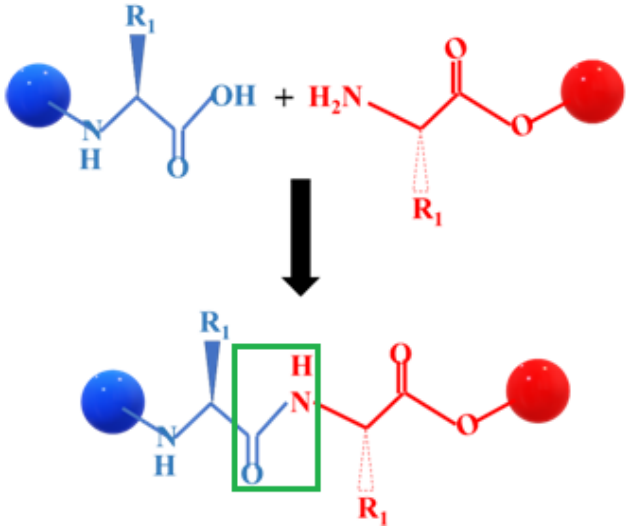


Figure 1.1: Amino acids forming a dipeptide through amide linkage.

The subunits of the protein are made up of amino acids called peptides. The deep understanding on chemical composition in peptides or protein showed that each amino acid consists of an acidic carboxyl group, an amino group, and an organic R group (**Figure 1.1**). There are 20 various types of amino acids present in nature with varying organic group. These amino acids can be joined together in various patterns and orders through amide linkage to form different types of proteins.

For more than five-decades, molecular biochemistry, structural biology, and biophysics have been influenced by the rigid or semi-rigid view of a functional protein or protein domain. These functional entities were used to propose unique and stable 3D structures of proteins or peptides (Uversky 2014). The protein structures can be determined using X-ray crystallography (Smyth 2000) and NMR spectroscopy (Cavalli et al. 2007). The functions of proteins or peptides depend on its energetically stable 3D fold conformation. Some fluctuations that allow for biological function also results from the changes in 3D structures. Conversely, there are disordered proteins and protein domains, such as biologically active intrinsically disordered proteins (IDPs) and intrinsically disordered protein domains (IDPs) that do not contain a properly defined equilibrated structure and mainly exist as highly dynamic, heterogeneous conformers (Tompa 2002), which are crucial for biological processes.

The biological functions of proteins and protein domains include energetical responses, catalyze enzymatic reactions, mechanical structural support, cell growth control, transportation, and storage of molecules. The specificity of protein function depends on the three-dimensional structure of the protein or peptide formed by the polymeric chain of amino acids. In one way, the fluctuations, denaturation, or misfolding of proteins results in the formation of aggregates which affects the native structure of proteins. Such changes in the molecular level result in system malfunction and disease in the organism as seen in Alzheimer’s disease, Parkinson’s disease, diabetes, etc.

1.5.2 Nucleic Acids

Proteins are not only the biomolecules that catalyze chemical reactions. There are some other biomolecules which have the ability to catalyze biochemical reactions.

The RNA molecules, found more than 30 years ago, can participate in the growth of t-RNA precursors (Jaeger 1997) which decodes mRNA and transfers genetic information. Nucleic acids are complex biomolecules mainly present in cells and viruses. They are accountable for storing genetic information and expressing it through protein synthesis. They are made up of small units called nucleotides. Nucleotides consist of pentose sugar attached to nitrogen-containing aromatic base units and phosphate groups (**Figure 1.2)**. Each nucleic acid has four nitrogen-containing base units namely adenine (A), thymine (T)/ uracil (U) guanine (G), and cytosine (C). All nucleic acids contain A, G, and C base units. Presence of T and U determines the type of nucleic acid; DNA and RNA (Silva 2004). RNA can function as a reservoir of genetic information and a catalyst for biological reactions. The ability of RNA to initiate biochemical reactions lies in the folded 3D structures by forming Watson-Crick base pairs (Xia et al. 1998). The ability of certain regulatory proteins and nucleic acids to recognize and bind specific nucleotide sequences has led to their widespread applications in research. These nucleic acids act as key players in forming proteins (Chen et al. 2008). The ribozymes enzyme can catalyze the enzymatic reactions to understand the functional roles of RNA molecules in the cell such as catalysis, replication, transcriptional and translational regulation, and ligand binding (Tanner 1999). The folding of RNA from single-stranded regions to loops or helices can create binding sites for ligand and metal ions (Pan 1995). These mechanisms force infected cells to produce only viral proteins instead of the normal host cell.

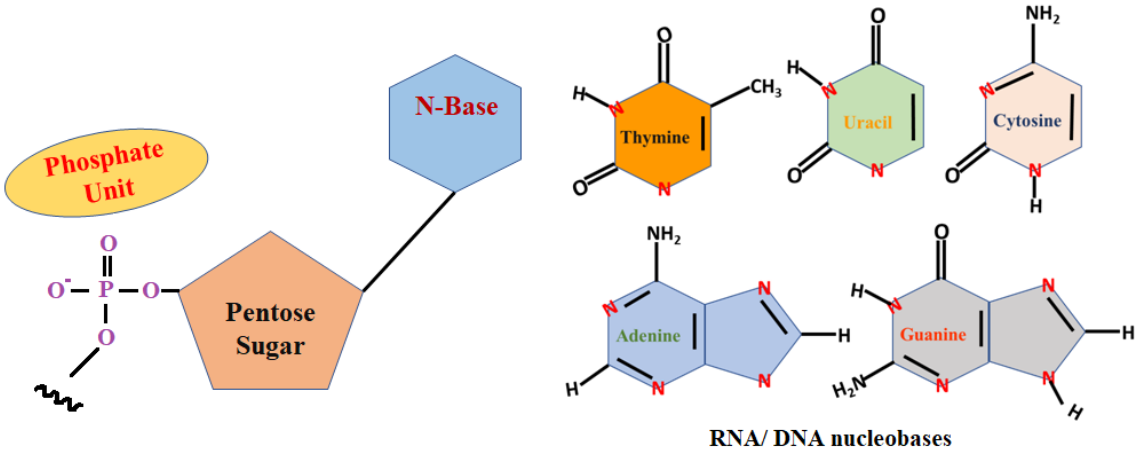


Figure 1.2: Representation of nucleotide containing a pentose sugar attached to a nitrogenous base and phosphate group (left). Chemical structures of DNA and RNA nucleobases (right).

In summary, it can be concluded that the structure-to-function relationship of biomolecules greatly depends on the surrounding environment. Water also assists an important role in biological phenomena. The presence of ions and cosolvents can activate or inactivate the function of proteins/nucleic acids as well as alter the properties of water molecules around them. Similarly, the changes in temperature from normal conditions can affect both the biomolecules and the water structure. Literatures supporting the structure-to-function relationship of biomolecules and the effect of water on biomolecules under various conditions are discussed in the following section.

1.6 LITERATURE SURVEY

Life originated with water and it is very difficult to exist in its absence. The solvation of biomolecules in water has importance in various disciplines within chemistry, and biology to understand the structural, dynamics, and thermodynamic properties. The dynamics of water around biomolecules are complex and non-uniform (Frank and Evans 1945). Soper and Phillips (1986) defined hydrogen bond between two water molecules only if their inter-oxygen distance is less than 3.5 Å and the angle O--H--O is less than 30˚. The importance of water depends on its solvation power for any nonpolar substance (Ben-Naim 1987) and the dynamic of the water gets affected due to the nature of the solute. Poole et al. (1992) explained the phase diagram of liquid water that exhibits rich features that are visible in the equilibrium properties of water. The water molecules near the surface of biomolecules mainly get perturbed compared to the bulk due to changes in the arrangement of the hydrogen bond network. Water plays an important role in protein folding due to the presence of hydrophobic and hydrophilic moieties (Rahman and Stillinger 1971) Various experimental and computational techniques help to determine the 3-D structure of nucleic acids, and protein/peptides which are responsible for their function. The computer simulation method is a better choice to study processes that can be harmful, costly, or even impossible experimentally. Computer simulations can be performed at a very high (extreme) temperature and pressure which are impossible in the laboratory. The goal of MD simulation is to characterize and predict the behavior of real systems at the atomic level. Molecular dynamics simulation can investigate the structure and

dynamic properties of the materials and biological molecules, with better accuracy and experimental correlation. Below we are providing the literature survey of some key articles along with their significant contribution towards the structural and dynamic aspects of water and biomolecules under varying ionic concentration and thermodynamic conditions.

- Ibragimova and Wade (1998) showed that explicit ions are required to stabilize the YAP WW domain (β sheet domain) in addition to the normal charge-balancing counter ions using MD simulations. The stability of a protein is calculated in terms of RMSD, RMSF, interaction energy, and electrostatic interaction.
- The heterogeneous nature of local hydrogen bond patterns were studied by constructing Voronoi polyhedra ($\eta = A3/36\pi V2$) for the oxygen atom distributions. The slow H-bond relaxation is found to be temperature sensitive while the fast relaxation is temperature independent and arise from interstitial molecules coupling with a center water molecule (Yeh and Mou 1999).
- MD simulations performed with various concentrations of NaOH (0.5 M to 19 M) suggests the existence of ordered ion structures in the solution. The structure of hydration shells of ions-water was analyzed using the radial distribution functions, Voronoi tessellations, and the running coordination numbers. The structural and dynamic properties obtained from theoretical methods agreed well with the available experimental data (Zapałowski and Bartczak 2000).
- The relationship between structural order and anomalies of liquid water were studied based on the order parameters of the water structure: translational order parameter (t) and orientational order parameter (q). These parameters have explored the structure of water in the form of tetrahedral and both parameters are strongly coupled to each other (Errington and Debenedetti 2001).
- The thermodynamics of interaction between TMAO and protein functional group (-CONH-) were determined using calorimetric and MD simulation methods. It is found that transfer free energy of –CONH- from water to TMAO is positive, indicating an unfavourable TMAO/-CONH- interaction whereas slightly favourable interaction between TMAO and apolar groups from water.

MD simulations supported these findings and suggest that TMAO stabilizes protein conformations via water structure enhancement through hydrogen bonds (Zou et al. 2002).

- Zhang et al. (2003) calculated the structural and dynamic properties of protein (bacteriophage T4 lysozyme) and S-peptide analog in an aqueous medium via amplified collective motion method. The motions of molecules along the modes are coupled at higher temperatures such as the weak coupling method to amplify collective motion. This methodology guides atomic level molecular simulations and can be applied to protein folding/unfolding, structure prediction.
- MD simulations performed on HP-36 studied the sensitivity of protein-water hydrogen bond lifetime in the protein hydration layer based on the secondary structure elements. The dynamics of interfacial water molecules and their structural properties are correlated with the network of hydrogen bonds between water and the amino acid residues of the protein. The relaxation behaviour of protein-water H-bonds depends on the secondary structure elements (Bandyopadhyay et al. 2005; Garcia and Hummer 2000; Pavlek et al. 2024).
- The absorption of ions on trapped water within the tiny cavities inside the silica gel follow the order $Mg^{2+} < Ca^{2+} < Li^{+} < Na^{+} < K^{+} < Rb^{+} < Cs^{+}$ and proposed that the structure of water inside these pores are different from the bulk water (Remko and Rode 2006; Romero et al. 2018). It can change the physical and chemical properties of the system. Water can change its density so that hydrogen bonds become straighter and stronger or bent and weaker. Water can also change its properties at a protein surface. Many enzymes discriminate between sodium and potassium: carboxyl groups present in enzymes show a preference for K^{+} over Na^{+} ions at a ratio of 2:1. These two ions show the effect on the water: Na^{+} ions slightly increase the water structure whereas K^{+} ions decrease the water structure (Wiggins 2008). It is found that lithium shows exception to follow the Hofmeister series and shows more water affinity than K^{+} and Na^{+} ions (Aziz et al. 2008; Jakobsson et al. 2017). The experimental results (X-ray absorptions) are supported by combined *ab initio* and MD simulations. Another study suggested increasing binding affinity of ions towards weak carboxylic acids, $K^{+} < Na^{+} < Li^{+}$ because of the strong attraction

between the smaller cation and acetate in solvent-shared configurations (Hess and van der Vegt 2009). In RNA, binding affinity follows direct Hofmeister series at backbone but reversed at nucleobases leading to high selectivity of ion binding sites (Cruz-León and Schwierz 2020). In less polarised water, DNA experiences B→A conformational transition for lighter alkali metal counterions (Li^+, Na^+ and K^+) while it keeps B form for heavier ions (Rb^+ and Cs^+) (Wen et al. 2014).

- Yan et al. (2008) studied the relation between excess entropy and its local order parameter via MD simulations using canonical ensemble and TIP5P water model. In conclusion, excess entropy talks about dynamic and thermodynamic properties of water and also the exact location of Low/High density of water.
- Chakraborty and Chandra (2011) performed ab initio molecular dynamics simulations of liquid-vapor interfacial system consists of water-ammonia mixture. The structural and dynamical properties are calculated for the mixture revealed that the relaxation of the interfacial molecules is faster than those of the bulk. However, longer hydrogen bonds lifetimes are observed at interface.
- Ion-protein interactions were studied using polarizable force-field with 30 PDB structures at 0.15 M concentration of Na^+, K^+, Ca^{2+}, Cl^- in presence of NMA, ETOH, PA (1,2,3,4 molecules) as the solvent for each system. The results were analyzed by the structural properties like angular geometry, binding distances between ions and NMA, the coordination number of ions, and thermodynamics properties such as solvent-solvent transfer free energy of salts. The accumulated force field models show the interactions of Na^+, K^+, Ca^{2+}, and Cl^- with the polar group of protein, and amino acids including the functional group like carboxylates, hydroxyl groups, and side chain moieties (Li et al. 2015).
- The entropy concept for water structure and dynamics in a protein hydration layer is analyzed via MD simulations of lipase B. The relationship between the structure and diffusive dynamics of water was studied. The different measurement of hydration dynamics was calculated by selecting every region's (α-helix, β-sheet) water molecule, and the RDFs, excess entropy also showed that protein structure affects the water structure and dynamics through topological and chemical effects. The heterogeneous hydration dynamics

around lipase surface was found to correlate with spectroscopic data of other proteins (Dahanayake and Mitchell-Koch 2018).

- ❖ Mixed solvent (Water-MeOH) promotes protein RNA association, by enhancing salt–linkages between basic amino acid side-chains and acidic phosphates of the sugar–phosphate backbone (Porcelli et al. 2024). Simulations showed that the A-RNA shape was most sensitive to the RNA force field, with some force fields leading to a reduced inclination of the A-RNA duplexes (Kührová et al. 2023).
- ❖ Temperature-dependent simulation studies of amyloid peptides for the study of aggregation and deposition of amyloid fibrils: necessary to support the development of therapeutics (Hosseini and Van Der Spoel 2023; Okumura 2023; Tachi et al. 2022; Zhang et al. 2024). Observed conformational changes upon variations in temperature were not at all driven by electrostatic interactions. The final conclusion was that the simulations were not able to accurately capture the temperature induced conformational changes of IDP (Jephthah et al. 2019; Zerze et al. 2015).
- ❖ All-atom replica exchanges molecular dynamics (REMD) simulations have been done of PrP^{C}–PrP^{C} to understand the molecular mechanism, modes of interactions, and altered stability pattern in the PrP^{C}–PrP^{C} term of monomer and dimer. The conformational changes mainly in terms of the secondary structure, hydrophobic surface exposure, and peptide-peptide intermolecular interactions. The monomers exhibit spontaneous dimerization by a decrease in the solvent-exposed surface area along with an increase in intermolecular contacts involving salt bridges, hydrogen bonds, and hydrophobic interactions. The water molecules can wrap around the chain to almost maintain the hydrogen-bonded network (Chamachi and Chakrabarty 2016; Hande and Chakrabarty 2022; Singh et al. 2017). At different temperatures, the exchange rate of interfacial water molecules, are rare and transient event on the molecular time scale (Strandberg et al. 2023).
- ❖ In amyloid formation, the aberrant β-rich isoform of PrP is more efficient in binding and stimulating the conformational switch of the normal α-helical form *via* a unique binding-induced misfolding and aggregation (Mukhopadhyay 2020). The study of the multidimensional free energy landscapes of diversely

complex proteins were done with REHT and REST2 methods. During the course of simulations, both the REHT and REST2 methods drive the transition of the unfolded starting conformation to a native folded conformation with almost a perfect match with the experimental. The secondary structure changes observed during simulations are found to be similar to those observed in NMR result (Appadurai et al. 2021).

- The effect of ions is significant to IDPs and local ion–residue interactions play an important role in the structures and behaviors of IDPs. Recently, an open-source Python package SPEADI, was developed to analyze the dynamic distribution of ions around IDPs from MD trajectories (Bruyn et al. 2023).
- The charged phosphate groups of the sugar-phosphate backbone are the primary hydration sites for the water molecules and their negative charge attracts the counter ions for establishing the charge neutrality in the aqueous solution. The polyanionic nature of nucleic acids induces stronger interactions with water and solvent molecules than the proteins (Fingerhut 2021).
- In the last three years, Scientists have done a lot of clinical trials and computational studies to find out the antiviral drug for COVID-19 and its various mutants. The interaction of antiviral drug (AD) with WT-A97V and P323L-RdRp was alanyzed using molecular dynamic (MD) simulations and the binding free energies were calculated. The AD bound to WT- and A97V-RdRp had a similar dynamic motion and residue fluctuations, whereas AD interaction with P323L-RdRp exhibited a tighter molecular conformation, with a high internal motion near the active site (Mohammad et al. 2021). Molecular docking and molecular dynamics simulation were performed to study the binding potency of novel guanosine derivatives against RNA-dependent RNA polymerase of SARS-CoV-2 (Elfiky et al. 2022).
- It is reported using MD simulations that higher temperatures have little or no influence on the stability and folding of the SARS-CoV-2 Spike protein (Khan et al. 2022). Computer simulations studies showed that the G15U mutation in the Delta variant s2m leads to significant structural, dynamical, and entropic differences from the parent SARS CoV-2 variant. This provides the foundation

on the mechanism of homodimerization in the Delta variant, which helps to establish structure–function relationship in virus (Makowski et al. 2023).

- A combined NMR/MD study, to assess the temperature-dependent conformational dynamic properties of RNA secondary structure, which is required for the microscopic understanding of the various functions of RNA systems (Ferner et al. 2008). Using a combination of careful experimental and simulation studies to probe the temperature-dependent twist of the RNA double helix: the magnitude of the thermally induced twist decrease observed in the simulations is much weaker than that observed in the experiment (Dohnalová et al. 2024).
- The different water model potentials can affect the structure and dynamics of simulated biomolecules such as proteins (Venugopal et al. 2022) and RNA. The water of different model potentials have different polarizability that can affect the structure and stability of the biomolecules.

The above literatures support the fact that computational studies can be important to propose many novel intuitions/postulate in the research of biological process. Taking into consideration, the two ribosomal peptides (Antimicrobial, Intrinsically Disorder Prion Peptide) and one ribonucleic acid (SARS-CoV-2 RNA-genome) were selected along with the metal ions and thermodynamic conditions for the present research work. A brief discussion about the peptides and SARS-CoV-2 RNA-genome considered in this research work is provided below.

1.6.1 Antimicrobial Peptide and Importance

Antimicrobial peptides (AMPs) are a class of small peptides that widely exist in animals, plants, and even microorganisms. The emergence of antibiotic-resistant microorganisms and the innate immune system of the different organisms paved the way for research on antimicrobial peptides. AMPs have a wide range of inhibitory effects against bacteria, fungi, parasites, and viruses (Lei et al. 2019). AMPs tend to help in metabolic processes in the plasma membrane with extra and intracellular sites. It can frequently change the body's immune system; help kill cancer cells and promote faster wound healing. they are considered to be a crucial part of immunity for all the

eukaryotic cells towards the microbial influx (Ageitos et al. 2017; Das and Chakraborty 2020; Hancock and Sahl 2006; Mondal 2019).

Generally, AMPs consist of 6–100 amino acids units with a diverse group of amino acids, displaying various types of structures, including α-helices, β-sheets, or cyclic structures. There have been many experimental and theoretical studies that suggest that AMPs can exist partially or fully disordered inside the solution and acquire their functional secondary structure upon interaction with the amphiphilic membrane and solvent (Kang et al. 2017; Preußke et al. 2023; Sato and Feix 2006). AMPs are polymers that are composed of hydrophilic, hydrophobic amino acids and therefore, they are amphiphilic in nature. AMPs generally have helical and beta-sheet structures with short amino acid sequences (Huan et al. 2020). The AMPs (positively charged) interact with the negatively charged cell membrane protein through electrostatic interactions. AMPs can be adsorbed on membrane proteins by a conformational change. The selected antimicrobial peptide, PDB ID: 5Z32 (**Figure 1.3**), helps to make conjugate with tungsten disulfide quantum dots which boost the antimicrobial activity of peptide (Mohid et al. 2019). Understanding the structure-function relationships of peptide-water, peptide-peptide, and peptide-salts interactions are important in molecular life and also in biological processes.

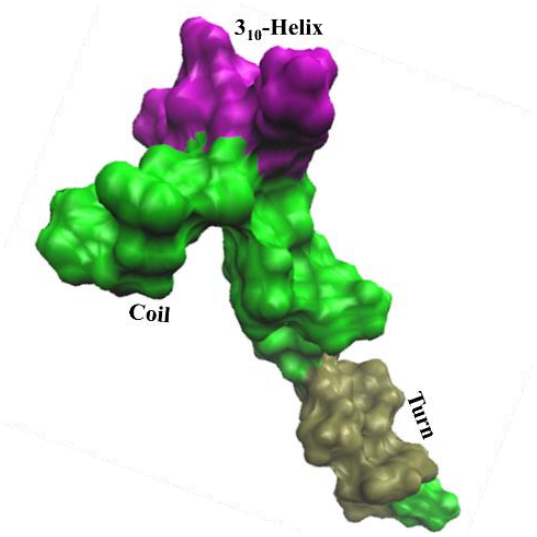


Figure 1.3: Representation of antimicrobial peptide surface (PDB ID: 5Z32).

1.6.2 Intrinsically Disordered Sheep Prion Peptide

The intrinsically disordered peptides/proteins (IDPs) do not contain well-defined three-dimensional structures, but their intrinsic conformational changes propensity has a wide range of biological functions such as signal transduction, transcription, and cell cycle regulation (Wright and Dyson 1999). Prion is a cell surface protein (**Figure 1.4**) and is dominantly expressed in the peripheral and central nervous systems (Bendheim et al. 1992). The primary functions of prion protein include cell adhesion, synapse formation (Santuccione et al. 2005), regulation of circadian rhythm (Tobler et al. 1996), controlling ion homeostasis, cell signaling (Lewis and Hooper 2011), etc. It is known that the transformation of the normal cellular prion isoform (Prp^{C}) to protease-resistant scrapie isoform (Prp^{Sc}) is the initial point of prion disease. Following the misfolding, an exponential increase in the infected prion protein (PrP^{Sc}) population occurs in an autocatalytic manner. The global disorder-to-order transitions can occur leading to the interaction of IDPs with partner proteins (Friedman et al. 2005). The intrinsic disorder is particularly abundant in proteins implicated in human diseases such as cancer, neurodegenerative disease, and diabetes (Jucker and Walker 2013, 2015). The role of metal ions in the context of IDP pathogenesis remains highly debated due to their ability to modify the electrostatic interaction (Huat et al. 2019). A recent study of metal ions such as Li^{+}, Na^{+}, K^{+}, $Mg^{2+,}$ Ca^{2+}, $Zn^{2+,}$ etc, have a significant effect on the prion peptide by shielding the unfavorable charge repulsion (Luczkowski et al. 2015). The Cu^{2+} and Zn^{2+} are also reported to exhibit a significant effect on prion misfolding and the formation of toxic oligomeric species. The water molecules can wrap around the chain to maintain the hydrogen-bonded network (Chamachi and Chakrabarty 2016; Hande and Chakrabarty 2022; Singh et al. 2017). The thermodynamic stability of the PrP^{C} form concerning the possible misfolded conformations, and the associated barriers of their interconversion with the underlying conformational free energy landscape studied by the metadynamics simulation (Chamachi and Chakrabarty 2017). The various mammalian prion peptide revealed that it consists of a reserved monomeric protein fold with a highly flexible N-terminus and a globular C-terminal domain, which contains a small two-stranded, anti-parallel β-sheet and α-helices

(Megy et al. 2009). Sheep prion peptides are generally found in β-hairpin in solution state and it is found to exist in α-helix and β-strand experimentally. The structural changes in sheep prion peptide which are trapped in free energy space cannot be determined by experimental techniques, and therefore, computational study is used to get information about these conformations. Sheep prion peptide is chosen because it is known that prion proteins are highly infectious and have the ability to transmit from animal to human. Among all the mammalian prion proteins reported to date, sheep is highly consumed by humans. Therefore, there is a high chance of transmission of infected prion protein from sheep to human through food items. Further, it is known that all the mammalian prion proteins have a high similarity index. It is also evident from the literature that instead of C-terminal and N-terminal segments the rest of the prion protein (23-231) is almost similar to all mammalian prion proteins.

There are various enhancement sampling method such as all-atom replica exchanges molecular dynamics (REMD) simulations, umbrella sampling, and metadynamics which can efficiently sample the hidden conformations. Therefore, with the help of these techniques, it will be interesting to understand the folding mechanism of the sheep prion intrinsically disordered peptide in the presence of metal ions and under thermodynamic conditions to understand the normal cellular prion isoform (Prp^{C}) changes to protease-resistant scrapie isoform (Prp^{Sc}).

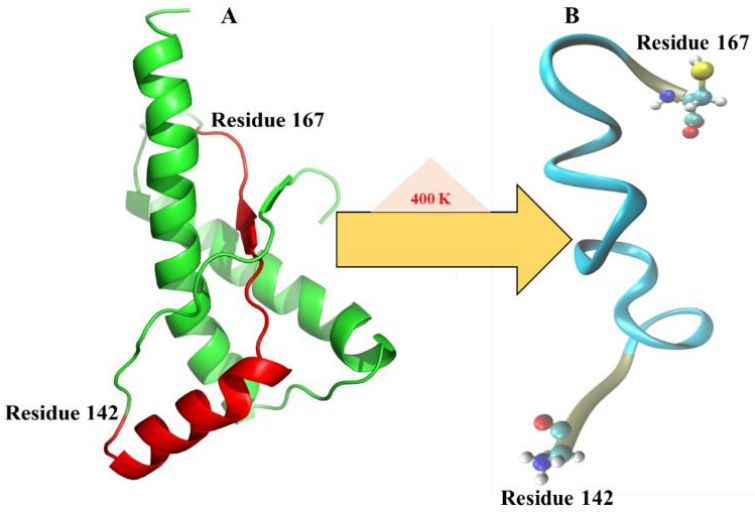


Figure 1.4: (A) Representation of the cellular prion peptide and red shows the sheep prion peptide (intrinsically disordered region). (B) The unstructured arrangement at 400 K of the intrinsically disordered region (142-167).

1.6.3 Severe Acute Respiratory Syndrome Coronavirus2 RNA Genome

Coronavirus has the largest positive-stranded RNA genome of 26 to 32 kb among the known RNA viruses. The sequence analysis of SARS CoV-2 isolates reveals that the genome encodes for 16 non-structural proteins (Nsp 1-16) which forms replicase/transcriptase complex (RTC), 4 structural proteins (spike, envelope, membrane, nucleocapsid) and 9 putative accessory factors (Wu et al. 2020). SARS-CoV-2 carries the largest ssRNA genome and is the causal pathogen of the ongoing COVID-19 pandemic.

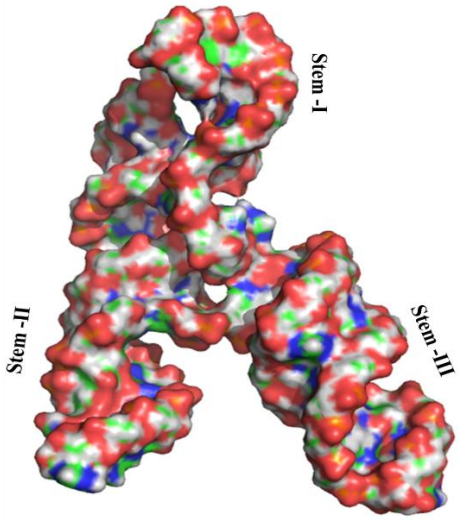


Figure 1.5: Schematic representation of FSE pseudoknot structure from SARS-CoV-2 RNA genome. Different nucleotides are shown by different colours.

The frameshift stimulation element (FSE) from the SARS CoV-2 RNA genome is an ideal drug target for designing and developing antiviral drugs against the COVID-19 pandemic (Schlick et al. 2021) (**Figure 1.5**). It is believed that FSE contains a major ring binding site and other two alternative binding sites: the slippery hairpin site and the $J_{3/2}$ site, where small molecules can bind to perform anti-frameshifting (Zhang et al. 2020). The frameshifting process helps the virus to compact larger genetic material into short genetic elements (Atkins et al. 2016; Niknamian 2021). Small drug molecules which bind to FSE binding site can cause anti-frameshifting, which blocks the production of RNA enzymes important for viral replication (Neupane et al. 2020). The role the alternative conformations found in FSE plays an important role in the frameshifting process and the quantification of this

structural abundance is still a major challenge (Dey et al. 2024; Yan et al. 2023). The binding of the drug molecules depends on the conformation of the pseudoknot structure in SARS coronavirus (Park et al. 2011), which is identical to SARS CoV-2 FSE up to single nucleotide substitution (Yan et al. 2023; Zhang et al. 2020). Therefore, studying the conformation evolution of FSE from the SARS CoV-2 RNA genome with temperature change can help to understand the finer details of the major forces responsible for maintaining the biological activity of RNA.

1.7 SCOPE AND OBJECTIVE OF THE WORK

From the literature survey, it is found that computational techniques can be used effectively to predict the interactions and free energy of protein folding/unfolding. Molecular dynamics (MD) simulations and high-order sampling methods such as REMD, umbrella sampling, metadynamics can capture protein dynamics and associated conformational changes at the atomic level. Since the metal ions can affect the water structure as well as the biomolecules, it is important to understand how the metal ions bind to proteins: before, during, or after protein folding, to perform a biological function. Also, it is equally important to understand the role of water potential induced stability and instability of biomolecules. In the research work, the effects of metal ions and temperature on the solvation structure of the water model as well as the secondary structure of biomolecules have been investigated.

Based on the literature review and above-mentioned scope, the following objectives have been designed.

- To calculate the preferential binding affinity of metal ions and their effect on the structure and dynamics of water near the antimicrobial peptide.
- To determine the influence of ion specificity and concentration on the conformational transition of intrinsically disordered peptides (IDP).
- To obtain insights into disorder-to-order transition in sheep prion peptide.
- To study the temperature-dependent conformational evolution of the SARS CoV-2 RNA genome.

- To explore the multiple and prominent conformational forms of RNA genome at different temperatures through interhelical dynamics and network analysis.
- To determine the sensitivity of the RNA Conformations in SPC/E and mTIP3P Water Models focusing on the effect of the hydration shells.

CHAPTER 2

MATERIALS AND METHODS

In this section, a brief discussion regarding the methodology used in biomolecular simulations is provided. Also, a brief background of various structural analyses and water-structure analyses around the peptide and the RNA molecule is included.

Molecular Dynamic simulation has become an important and very useful tool to answer many questions in the field of biochemistry and biophysics. GROMACS software was used to perform unbiased MD simulations in this research work. The initial coordinates of three biological systems for MD simulations; antimicrobial peptide (5Z32), intrinsically disordered sheep prion peptide (2RMW), and FSE of SARS CoV-2 RNA genome (6XRZ) were retrieved from RCSB Protein Databank. Before simulations, the systems were subjected to energy minimization by Steepest Descent algorithm followed by NVT and NPT equilibration. The peptide and nucleic acids parameters were taken from CHARMM, and AMBER force fields due to their suitability to reproduce experimental observations. The various water model SPC/E, mTIP3P, and TIP4P have been used to solvate the system. The simulations were repeated at least thrice to check the convergence and reliability of the results. The structural properties of biomolecules were investigated in terms of the root-means-squared deviation, radius of gyration, root-mean-square fluctuation, solvent accessible surface area and number of hydrogen bonds. The solvation structure has been studied in terms of radial distribution function, orientation profile, the fraction of hydrogen-bonded water molecules, tetrahedral order parameter, and void distribution. The dynamic properties have been calculated in terms of mean square displacement, hydrogen bond lifetime and survival probability. The free energy, excess entropy, binding affinity, porcupine plots, and native contact map have been calculated to understand the stability of the system. The thermodynamic stability of the peptide and RNA molecules is determined by the free energy of folding and unfolding of the system based on the native and non-native base pairs. Further, we introduced network

analysis method that can sample big data sets effectively to identify the most prominent conformations of the macromolecules which existed during the process of evolution/denaturation and can also show how these conformations are connected to each other. A concise discussion on the theory of various simulation techniques and analyses used in this research work is given in following sections.

2.1 MOLECULAR DYNAMICS SIMULATION

One of the key assumptions in biophysics is that biomolecules are more flexible and dynamic and the atoms which constitute the biomolecules are having a constant motion. Therefore, it is more difficult to understand the interactions at the atomic level during the regulation of their structures, dynamics of the biomolecules, and functions. The structural and dynamical properties of the biomolecular system require a deep knowledge of energetics and basic interaction mechanism at the microscopic level. Computer simulations have played an important role to probe the dynamics of different many-body systems, by capturing the motions of atoms and their interactions under a wide range of biophysical, physio-chemical processes which are more far-away from the scope of the experimental procedure. However, the detailed analysis of the degree of freedom at a microscopic level is an important aspect to probe the physical phenomena of theory. Therefore, the resolution of the primary structure calculated from the experimental tools is the basic and important requirement to perform MD simulations. In general, the trajectories of the molecules at an atomistic level in a molecular system are predicted by the numerical solution of Newton's equation of motion. The forces acting on the particles are to calculate the interatomic potentials during the particle motions. The MD simulations study was first applied to the material systems, afterward, MD simulation became an important and very useful tool to answer many questions in the field of biochemistry and biophysics. Steps to perform a typical MD simulation are shown in **Figure 2.1**. To simulate a wide range of molecular systems, there are different types of force fields developed and is re-parametrized to reproduce the experimental data. Therefore, the output obtained from the MD simulation trajectory can be compared with experiments and considered a useful tool to validate experimental data.

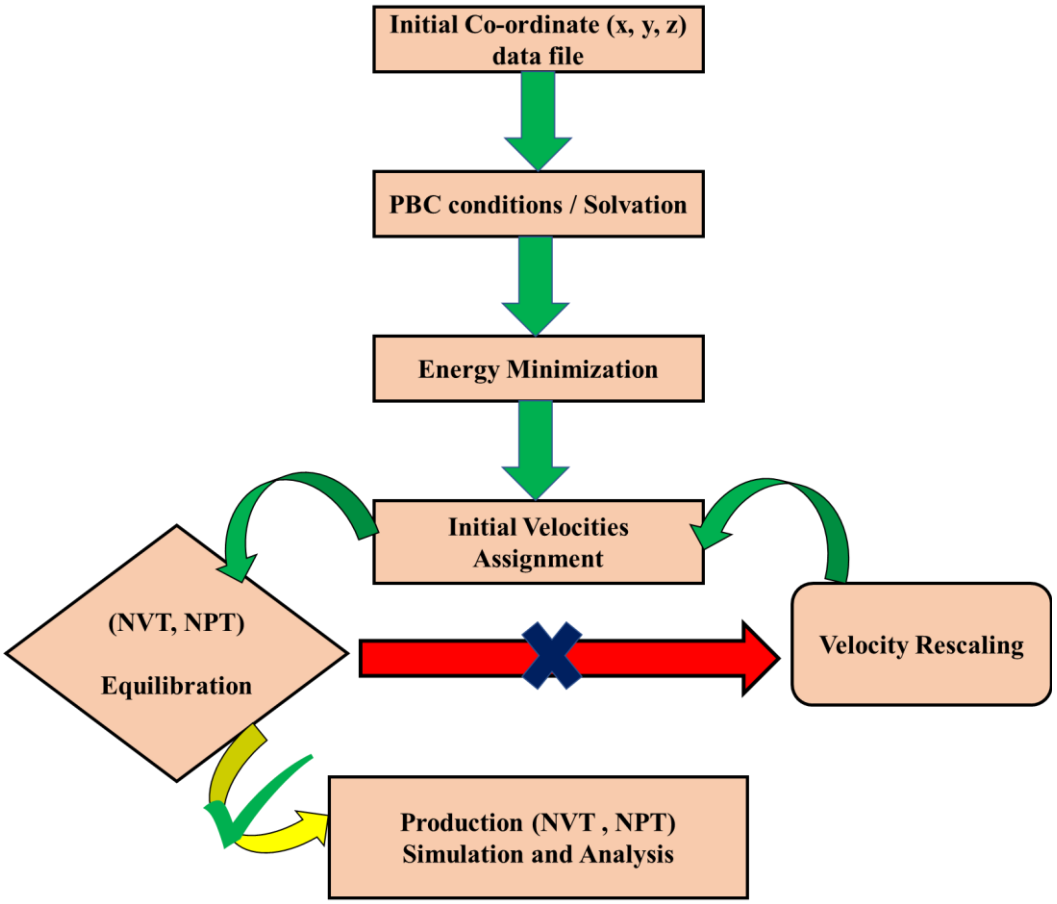


Figure 2.1: Flow chart to perform the Molecular Dynamics simulation.

The major drawback of MD simulation is the high computational cost. In general, many biomolecular events can occur in a nanosecond (ns)/microsecond (μs) to millisecond (ms) time scale. To detect the biological phenomena, one thing needs to extend the MD simulation up to the ms time scale, which requires huge storage space and computation time. There are many higher sampling techniques has been developed, such as replica exchange molecular dynamics, umbrella sampling, free energy perturbation, meta-dynamics, adaptive biasing force, etc to overcome limitation such as computation time and computational cost. Such methods can efficiently detect various biological events by increasing the sampling of the biomolecule in a defined phase space. However, those methods can be efficiently applicable in biomolecular systems containing comparatively a moderate number of atoms. Understanding the structure and dynamics mechanism of very large biomolecules requires some advanced sampling techniques.

2.1.1 Basic Equations of Molecular Dynamics Simulations and Essential Algorithms

Molecular dynamics simulations are performed by solving Newton's equations of motion

$$\boldsymbol{F_i} = m_i \boldsymbol{a_i} \tag{2.1}$$

where m_i is the mass of the atom, $\boldsymbol{a_i} = \frac{d^2 r_i}{dt^2}$ is the acceleration and F_i is the force acting on atom i due to the interaction with other atoms. The force acting on each atom can be expressed in terms of the gradient of potential energy (V) of that particular atom,

$$F = -\nabla_i V = -\frac{dV}{dr_i} \tag{2.2}$$

$$-\nabla_i V = m_i \frac{d^2 r}{dt^2} \tag{2.3}$$

$$-\frac{dV}{dr_i} = m_i \frac{d^2 r}{dt^2} \tag{2.4}$$

The position and velocity of each atom at every time interval can be calculated by solving equation 2.3. To solve the differential equation, the initial position, and velocities of each atom which is generally mentioned in the coordinate file is considered. The collection of the position and velocity of all the atoms of interest is known as trajectory, the final output of MD simulation. To integrate the Newtonian law of motion for the huge collection of particles, different algorithms were developed.

***Verlet Algorithm*:** The basic equation of verlet algorithm is given below

$$r(t + \delta t) = 2r(t) - r(t - \delta t) + a(t)\delta t^2 \tag{2.5}$$

In verlet algorithm, the new position and acceleration are calculated from the position at time t and (t-δt) without using explicit velocities (Verlet 1967).

***Leapfrog Algorithm*:** This algorithm was used to fulfill the drawbacks of verlet algorithm, especially the velocity term (Hockney and Eastwood 1988). By using some approximation for derivative, once it has considered the velocity at the midpoint between time t and (t+δt) express as:

$$v_i\left(t+\frac{1}{2}\delta t\right) = r_i(t+\delta t) - \frac{\mathrm{r_i(t)}}{\delta \mathrm{t}} \tag{2.6}$$

***Velocity verlet Algorithm*:** Unlike the Leapfrog algorithm, the velocity-verlet algorithm calculates the position, and velocity at the same value as the time variable. This approach explicitly incorporates velocity to solve the problem of the first-time step in the verlet algorithm (Swope et al. 1982).

$$r_i(t+\delta t) = r_i(t) + v(t)\delta t + \frac{1}{2}a(t)\delta t^2 \tag{2.7}$$

$$v(t+\delta t) = v(t) + \frac{1}{2}[a(t) + a(t+\delta t)]\delta t \tag{2.8}$$

***Beeman's Algorithm*:** This method provides the numerical integration of the ordinary differential equation of order two. The algorithm is designed in such a way that it allows a huge number of particles in MD simulation Beeman (Beeman 1976). It computes the position r_i of the particle at the time (t + δt) from the data (t - δt),

$$r_i(t+\delta t) = r_i(t) + v(t)\delta t + \frac{1}{6}[a_i(t+\delta t) - 2a_i(t)]\delta t^2 + O\delta t^4 \tag{2.9}$$

$$v_i(t+\delta t)\delta t = r_i(t+\delta t) - r_i(t) + \frac{1}{6}[2a_i(t+\delta t) - a_i(t)]\delta t^2 + O\delta t^4 \tag{2.10}$$

***RESPA Algorithm*:** The reversible reference system propagator algorithm (RESPA) is a newly introduced integrator algorithm suitable for systems with a separation of time scales or with long-range forces. This algorithm is generally used to integrate fast motions with smaller time steps.

2.1.2 Force Fields and Interaction Parameters

Force field is the collection of mathematical functions and parameters which is used to calculate the potential energy of the atoms in the molecular system. There are two

basic terms, bonded and nonbonded interaction between the atoms. There are two kinds of force fields (a) All-atom force and (b) Coarse-grained force fields.

The mathematical function for the force field can be expressed by the following equations:

$$E_{total} = E_{bonded} + E_{non-bonded} \tag{2.11}$$

where, $E_{bonded} = E_{bond} + E_{angle} + E_{dihedral}$; $E_{non-bonded} = E_{electrostatic} + E_{van\ der\ Waal}$

There are many force fields available in the literature, which have different degrees of complexity, and were developed to work on different kinds of systems. Some examples of the force field are: AMBER (Assisted Model Building and Energy Refinement, (Weiner et al. 1984)), CHARMM (Chemistry at HARvard Molecular Mechanics, (Brooks et al. 1983)), GROMOS (GROningen Molecular Simulation package, (van Gunsteren and Berendsen 1987)), OPLS (Optimized Potential for Liquid Simulation, (Jorgensen and Tirado-Rives 1988)), MARTINI etc.

The potential energy of the system can be related by a simple equation of internal coordinates known as a force field. In CHARMM force field, potential energy (V) has the form:

$$\begin{aligned} V = & \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} k_\emptyset[1 + \cos(n\emptyset - \delta)] \\ & + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{Urey-Bradley} k_{UB}(r_{1,3} - r_{1,3;0})^2 \\ & + \sum_{\emptyset,\varphi} V_{CMAP} + \sum_{non-bonded} \varepsilon_{ij}[(\frac{R_{min,ij}}{r_{ij}})^{12} - 2(\frac{R_{min,ij}}{r_{ij}})^6 \\ & + \sum_{non-bonded} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \end{aligned} \tag{2.12}$$

where, the terms corresponding to harmonic bond stretching potential (b is the bond length), harmonic bond angle potential (θ is the bond angle), proper dihedral potential (ϕ is the dihedral angle and δ is phase shift), improper dihedral potential, Urey-

Bradley potential ($r_{1,3}$ is the distance between atoms 1 and 3), grid-based energy correction map, van der Waals potential (r_{ij} is the inter-particle distance, ε_{ij} is the well depth and σ is the distance where potential is zero) and electrostatic potential (q is the charge of particle and ε_0 is the permittivity of free space) respectively. *k* is the force constant for each interaction terms.

For AMBER force field, potential energy (V) has the form:

$$V = \sum_{bonds} k_b(b - b_0)^2 + \sum_{angles} k_\theta(\theta - \theta_0)^2 + \sum_{dihedrals} \frac{1}{2} V_n[1 + \cos(n\emptyset - \delta)] + \sum_{impropers} k_\omega(\omega - \omega_0)^2 + \sum_{non-bonded} \varepsilon_{ij}[(\frac{R_{min,ij}}{r_{ij}})^{12} - 2(\frac{R_{min,ij}}{r_{ij}})^6 + \sum_{non-bonded} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \quad (2.13)$$

The potential energy in AMBER force field is similar to CHARMM force field, where the former lacks Urey-Bradley potential and CMAP correction terms.

2.1.3 Non-Bonded interactions

The interactions between atoms or molecules are governed by a potential that maintains the integrity of the matter and prevents the atoms from collapsing. The non-bonded interactions consistes of a Coulomb term, a repulsion term and a dispersion term. The Coulomb term covers all the charged particles in a system. The most commonly used potential, used first for liquid argon is the Lennard-Jones (L-J) potential, (6-12 interaction) which is a mathematically simple model that approximates the interaction between a pair of neutral atoms or molecules (**Figure 2.2**). A form of this inter-atomic potential was proposed by John Lennard-Jones in 1924 (Bell et al. 2019). The equation of L-J potential can be described as

$$U_{L-J}(r_{ij}) = 4\varepsilon(\left(\frac{\sigma}{r_{ij}}\right)^{12} - \left(\frac{\sigma}{r_{ij}}\right)^6) \quad (2.14)$$

where, $r_{ij} = r_i - r_j$ is the distance between a pair of atoms i and j and ε is the depth of the potential well, σ is the finite distance where the potential is zero.

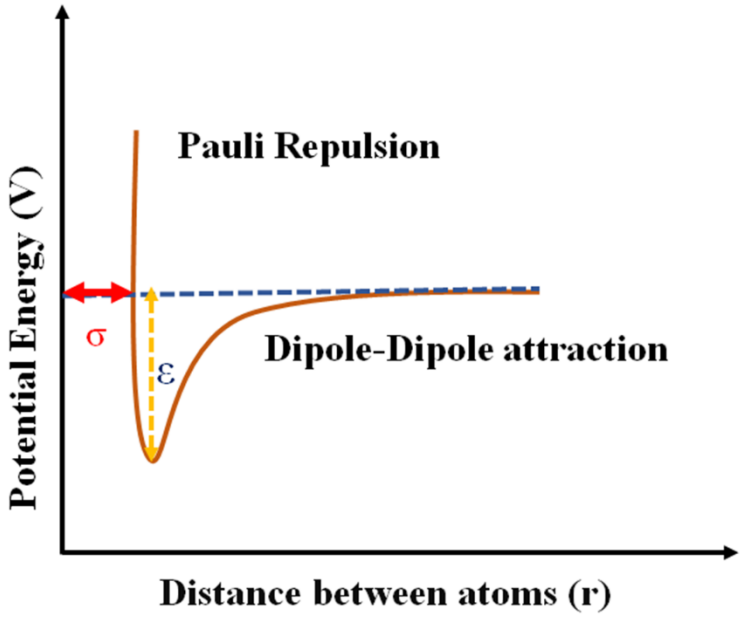


Figure 2.2: A graph V/r for Lennard – Jones potential.

***Electrostatic interactions*:** The Coulomb interaction between two charge particles is given by:

$$V_c(r_{ij}) = f\frac{q_i q_j}{\varepsilon_r r_{ij}} \tag{2.15}$$

where $f = \frac{1}{4\pi\varepsilon_0} = 138.935\ \text{kJ mol}^{-1}\text{nme}^{-2}$Coulomb interaction decays slowly with distance, considered long-range interaction. r_{ij} represents the distance between two atoms having charges q_i and q_j and ε_0 represent the dielectric constant, a number relating the ability of a material to carry the alternating current to the ability of the vacuum to carry alternating current.

2.1.4 Bonded Interactions

Bonded interactions are not exclusively pair interactions, but include 3 and 4-body interactions as well. There is bond stretching (2-body), bond angle (3-body), and

dihedral angle (4-body) interactions. A special type of dihedral interaction (called improper dihedral) is used to force atoms to remain in a plane or to prevent transition to a configuration of opposite chirality.

2.1.5 Important Files for Performing MD Simulation

***Coordinate file*:** This kind of file contains all the atomic positions and initial velocities in the 3D space (XYZ coordinates). Such coordinates files are often retrieved from XRD, and NMR studies of biomolecules. However, the coordinate files can be constructed using drawing software for small organic molecules. The coordinates files of unknown biomolecules can be predicted by homology modeling keeping related molecules as a template. The most widely used coordinate file is the protein Data Bank file (PDB) which can be retrieved from the RCSB-PDB website for known biomolecules.

***Topology files*:** The topology files primarily contain information on the mass and charge of all the atoms subjected to MD simulation. The addition of the charges corresponding to each atom represents the total charge of the entire system. Furthermore, the covalent bonds between the atoms, angels, and dihedrals are defined in the topology file. The topology files are dependent on the chosen force field for MD simulation.

***Parameter files*:** This file contains all the crucial parameters for MD simulation such as the integrator algorithm, time steps for integration (generally in the femtosecond range), bond parameters, nonbonded interaction cut-offs, virtual temperature, pressure controller, etc. The reference temperature and pressure for the simulation are also mentioned in this file.

2.1.6 Ensembles

An ensemble is a large group of microscopic states of a system with certain constant properties. It consists of moveable parameters at an atomistic level for molecular sub-units. Generally, there are different types of ensembles such as the Micro-canonical ensemble (NVE), Isothermal-isobaric ensemble (NPT), and Canonical ensemble (NVT). Some other ensembles such as (μVT) are also used in MD simulation.

***Microcanonical ensemble (NVE)*:** In the case of the microcanonical ensemble, A system (solid, liquid, or gas) has constant volume and energy with a constant number of particles. This ensemble was obtained by solving Newton's equations without any temperature and pressure control. At constant energy, simulations are not recommended for equilibration because, without the energy flow facilitated by the temperature control method, the optimum temperature cannot be achieved.

***Canonical ensemble (NVT)*:** A system has a constant temperature and volume ensemble with a constant number of particles referred to as the Canonical ensemble. The ensemble is obtained by controlling the temperature through direct temperature scaling during the initialization stage and by temperature-bath coupling during the data collection phase.

***Isothermal-isobaric ensemble (NPT)*:** A system having (N) particles at constant temperature and pressure is referred to as the NPT ensemble. This ensemble allows for overcoming both temperature and pressure. The ensemble can also be used during equilibration to achieve the optimum temperature and pressure before changing the constant volume.

2.1.7 Periodic Boundary Conditions

There are 3 types of boundary conditions (BC) that can be specified during the setup of MD simulations: 1) vacuum 2) reflecting wall and 3) periodic boundary conditions (PBC). The vacuum is the simplest BC which mimics a gas-phase environment but the dynamics of the global system properties will not reproduce the condensed phase (van Gunsteren et al. 1995). The reflecting wall BC indicates that the particle might be immediately reflected when it tries to cross the wall. PBC means that the system is placed in a simulation box and it is considered to have infinitely many images in space. Each simulation domain has 26 nearest neighbors in the 3D boundary. When the particle tries to cross the boundary of the simulation box on one side, immediately an image of the particle comes toward the simulation box from the opposite side, therefore, the overall number of particles in the system is conserved. However, to improve the speed and accuracy to compute electrostatic interaction in large-scale MD simulations Particle Mesh Ewald (PME) method has been developed.

2.1.8 Minimization

Even when initial structures are obtained from the experimental data such as PDB of protein or nucleic acids, there may be missing hydrogen atoms and unusual bonds or other atoms in residues. Therefore, it is necessary to do energy minimization of the initial coordinates to remove any potential steric clashes between atoms if the missing atoms are added during the initial structure calculation. In my work, I used either the steepest descent algorithm (Martyna et al. 1992) or conjugate gradient schemes to conduct energy minimization.

2.1.10 Holding on Temperature and Pressure

In all MD simulations reported in this work, I used the velocity-rescale temperature coupling method to additional damping and random forces introduced in the system. The temperature control is implemented through frequent adjustment of the momenta of all atoms in the system. Next, a 20 ns NPT equilibration was carried out with a time step of 2 fs at 1 atm using the respectively The pressure is controlled using the Nose-Hoover thermostat (Martyna et al. 1992) and Parrinello-Rahman pressure coupling scheme (Parrinello and Rahman 1981) algorithm in all MD simulations.

2.1.11 Software Packages for MD Simulation

GROMACS (GROningen MAchine for Chemical Simulations) (Abraham et al. 2015) is a software to perform molecular dynamics simulations of biological macromolecules such as proteins, nucleic acids, and lipids, and also for non-biological molecules such as polymers. GROMACS is a very efficient and fast calculation tool, the behavior of systems (for example, protein-water, protein-protein, nucleic acids in water) under some specific conditions such as temperature, pressure, and pH can be calculated very smoothly. It has also consisted of methods for preparation and analysis including free-energy calculation of the systems. It helps in monitoring the physical movements of atoms and molecules at specific time intervals. GROMACS has different force fields that can be used for energy minimization and different kinds of systems. GROMACS software package were used to conduct MD simulations for studies presented in working chapters. GROMACS is a suite of biomolecular simulation programs that can be used to setup, perform, and analyze MD simulations.

2.2 METADYNAMICS SIMULATION

Metadynamics is an enhanced sampling method and is generally described as "packing of the free energy wells with bias-free energy. In metadynamics, the free-energy surface is a function of some selected degrees of freedom, by choosing the collective variables (CVs). The bias potential which is a function of the collective variables (CVs) is added to the Hamiltonian of the system. This potential can be written as a sum of Gaussians deposited along the system trajectory in the CVs space to discourage the system from revisiting configurations already sampled. First, the bias potential will exert influence on coarse-grained dynamics in the CVs space based on a series of constrained molecular dynamics simulations. Later, the bias was given continuously during the MD simulation either by acting directly on the microscopic coordinates of the system or via an extended Lagrangian formalism. The modified Gaussian potential (V_G),

$$H = T + V + \sum V_G \tag{2.16}$$

where H is total energy or Hamiltonian of the system, T is the kinetic energy, V is the potential energy and V_G is the bias potential or Gaussian potential.

Let F be defined as the set of d functions (collective variables) of the microscopic coordinates (R) of the system: $F(R) = (F_1(R), F_2(R), \dots, F_d(R))$. At time t, the Gaussian potential (bias potential) may be given as,

$$V_G(F,t) = \int_0^t dt' \omega exp\left(-\sum_{i=1}^{d} \frac{(F_i(R) - F_i(R(t')))^2}{2\sigma_i^2}\right) \tag{2.17}$$

where ω is the energy rate and σ_i is the width of the Gaussian for i^{th} collective variable. The energy rate is constant and generally expressed in terms of Gaussian height (W) and a deposition stride (τ_G)

$$\omega = \frac{W}{\tau_G} \tag{2.18}$$

As it is common among CV-based methods, metadynamics is limited in the number of CVs it can handle as the computational cost scales exponentially with the number of CVs. This generally limits metadynamics simulation to three to four CVs. In standard MD simulation, the probability of system to remain in this minimum is more because the energy barriers are higher than thermal fluctuations. Meanwhile, in

metadynamics, as time goes by, Gaussians are deposited causing the underlying bias potential to grow, until the system is pushed out of another basin into a new local minimum. It accelerates the sampling of rare events by forcing the system to be away from local free-energy minima. Metadynamics allows us to explore new reaction pathways as the system tends to escape the minima passing through lowest free-energy saddle point.

2.3 REPLICA EXCHANGE MOLECULAR DYNAMICS (REMD)

The most challenging computational problems in the field of statistical mechanics is the development of methods which are capable of sampling a canonical distribution. Replica Exchange Molecular Dynamics (REMD) is a technique which can quiken the conformational sampling by allowing the exchange of configurations between *n* number of replicas (Nishimoto et al. 2019; Sugita and Okamoto 1999). REMD method was first introduced by Okamoto and co-workers (Sugita and Okamoto 1999) in the study of biomolecules. REMD is a hybrid method with a combination of Molecular Dynamics simulations and Monte Carlo algorithm.

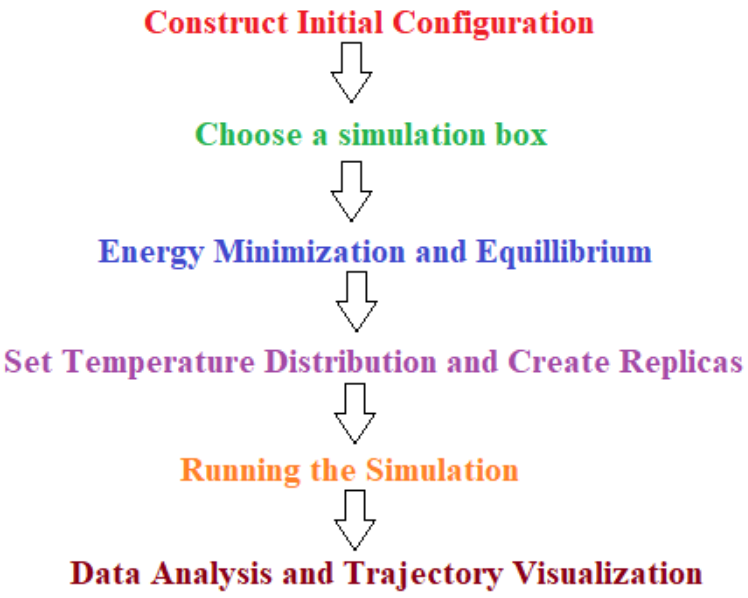


Figure 2.3: Major Steps in performing REMD simulation.

Let us consider a system of N atoms of mass m_k (k=1,…, N) with their coordinate vectors and momentum vectors denoted by $q \equiv \{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_N\}$ and $p \equiv \{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_N\}$, respectively. The Hamiltonian $H(q,p)$ of the system is the sum of the kinetic energy $K(p)$ and the potential energy $U(q)$

$$H(q,p) = K(p) + U(q) \tag{2.19}$$

The REMD method combines the replica exchange algorithm with multiple MD simulations. Thus, the initial step of the REMD simulation is identical to that of the conventional MD simulation. In the final step, a series of parallel MD replicas at different temperatures are created (**Figure 2.3**). The potential energy U (R_1, R_2, . . . , R_N) of N number of replicas (R) is characterized by a large number of local minima separated by high barriers. REMD can overcome these high energy barriers by swapping between neighbouring replicas periodically using Metropolis criteria. Such potential energy functions describe various bio-physical systems including proteins, DNA, RNA, polymer membranes, and polymer blends, to name just a few. In REMD, the exchange of adjacent replicas will be attempted based on the metropolis criterion with the following acceptance probability:

$$p(i \to j) = \min\,(1, e^{-\Delta}) \tag{2.20}$$

$$\Delta = \left(\beta_i - \beta_j\right)\left(U(R_i)\right) - U(R_j)) \tag{2.21}$$

where, $U(R_i)$ and $U(R_j)$ are the potential energy of i^{th} and j^{th} replicas at temperatures β_i and β_j respectively, where $\beta = \frac{1}{k_B T}$.

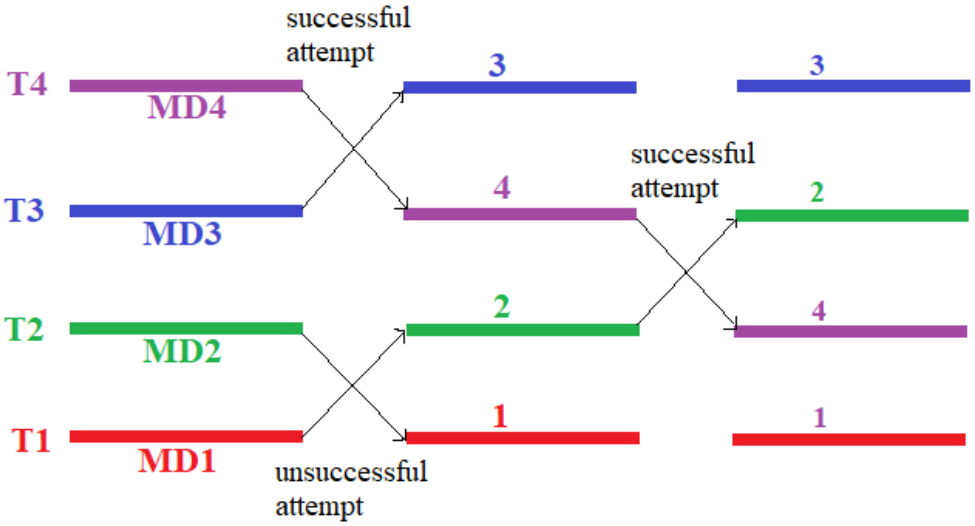


Figure 2.4: Simple illustration of REMD method using different temperature MD simulations at T1, T2, T3 and T4.

During the simulation, the temperatures will be distributed across all the replicas in a geometric progression to achieve a constant exchange rate across the

temperature range under consideration (**Figure 2.4**). The range of temperatures to sample the system has been identified based on the following exponential distribution:

$$T_i = T_0 * e^{ci} \tag{2.22}$$

where T_i and T_0 are i^{th} replica and starting temperatures respectively, c is the desired acceptance ratio and the exponential allows the increase in temperature intervals. REMD helps to overcome energy barriers and enhance the sampling with a better conformational space. Also, it allows to explore the free energy landscape of the secondary structure of biomolecules (proteins and peptides).

2.4 DATA ANALYSES FROM SIMULATION TRAJECTORY

VMD is used to visualize trajectories generated by MD simulations. MDAnalysis and an in-house Python program have been used to analyze the structural and dynamic properties of obtained trajectories. The root-means-squared deviation (RMSD), Radius of gyration (R_g), root mean squared fluctuation (RMSF), solvent accessible surface area (SASA), the distance between atom pairs, number of hydrogen bonds, interaction energy (IE), mean square displacement (MSD), etc are calculated with the GROMACS analysis tool. The free energy, entropy, enthalpy, and free energy landscape (FEL), binding affinity, native contact map (NCM), hydrogen bond lifetime, the fraction of hydrogen-bonded water molecules (f_n), void distribution, and survival probability (SP) has been calculated with a programming language such as python, Fortran, and C.

2.4.1 Structural Parameters

(i) ***Root mean square deviation:*** The root-mean-square deviation (RMSD) of protein, peptide, and nucleic acids was calculated as a function of time in order to evaluate the degree of conformational drift from the initial structure. RMSD can be calculated by least-square fitting the desired conformation to reference conformation ($t_2 = 0$) and is given by

$$RMSD(t_1, t_2) = \left[\frac{1}{M}\sum_{i=1}^{N} m_i \left\|r_i(t_1) - r_i(t_2)\right\|^2\right]^{\frac{1}{2}} \tag{2.23}$$

where r_i is the coordinate vector corresponding to time t_1, t_2. If the calculated RMSD is not mass-weighted, then we consider all $m_i = 1$ and $M = N$.

The energy change of the evolved conformations during a simulation as a function of RMSD can be calculated by

$$\Delta G = -k_B lnP(rmsd) \tag{2.24}$$

where k_B is the Boltzmann constant (8.3145×10^{-3} kJ/mol) and P(rmsd) is the probability of conformations with RMSD (Anandakrishnan et al. 2019). The free energy profiles of the protein provide the information about 1-D free energy landscape which depicts the most stable conformation at a particular RMSD value.

(i) ***The radius of gyration:*** Radius of gyration (Rg) measures the compactness of Biomolecules such as Peptide, proteins, DNA, RNA, and polymers during the simulation. Rg can be calculated using the equation,

$$rGyr(Rg) = \left(\frac{\sum \|r_i\|^2 m_i}{\sum i\, m_i}\right)^{\frac{1}{2}} \tag{2.25}$$

(iii) Root mean square fluctuation: The root-mean-square-fluctuation (RMSF) calculation showed the flexibility of the biomolecules. If we have considered that the x_i is the coordinates of particle i and $\langle x_i \rangle$ is the ensemble average position of i, then the RMSF can be calculated from the given equation:

$$\rho = \sqrt{\langle (x_i - \langle x_i \rangle)^2 \rangle} \tag{2.26}$$

(iv) Hydrogen bond occupancy: In the structural analysis, the hydrogen bond interactions exist between two non-overlapping groups of atoms, if the distance between the donor heavy atom and the acceptor (HD---A), r < 3.5 Å (Tan et al. 2021) and the angle (HD-A), α < 30˚ (Petukhov et al. 2004). The hydrogen bond occupancy is defined as the fraction of the time the molecule is hydrogen-bonded in the total simulation trajectory based on the geometric criteria.

(v) Solvent-accessible surface area (SASA): The Solvent accessible surface area (SASA) is defined as the surface area of a biomolecule or polymer that is accessible to a solvent. The measurement of SASA is usually described in units of $Å^2$.

(vi) Free energy of protein folding/unfolding: The free energy of folding for protein was determined by the changes in α-helix and β-sheet elements.

$$\Delta G = -K_B T \ln\left(\frac{1-f_N}{f_N}\right) \tag{2.27}$$

where f_N is the fraction of the native structure (Miner and García 2017).

(vii) ***Radial Distribution Function:*** Radial Distribution function (RDF) in a system of particles such as atoms, molecules, ions, solvents, ionic liquids, etc, describes the variation in density as a function of distance. It can also be defined as the probability distribution of finding a particle at a distance 'r' from the reference point. The equation of RDF g(r) is given as,

$$g(r) = 4\pi r^2 \rho dr \tag{2.28}$$

In MD simulation, the number of atom pairs in a provided range of separation, *r* is considered to determine the RDF. The general expression to determine the RDF is given by

$$rG(r) = 4\pi r^2(\rho(r) - \rho(0)) \tag{2.29}$$

where *ρ(r)* is the atomic density at distance *r* and *ρ(0)* is the bulk density of atoms (Sha et al. 2011).

(viii) ***Orientational order parameter:*** The orientational order parameter (S_g) is defined as the normalized sum of the squares of the differences between the cosines of the inter-bond angles and the cosine of the angles that would have been made if the bonds were tetrahedrally arranged. In a tetrahedral arrangement, all the angles between bonds will be the same, and therefore,

$$S_g = \frac{3}{32}\sum_{j=1}^{3}\sum_{k=j+1}^{4}\left(cos\psi_{jk} + \frac{1}{3}\right)^2 \tag{2.30}$$

where $\psi_{j,k}$ is the angle between the central oxygen atom and the j^{th} and k^{th} bonds. The factor 3/32 normalizes the S_g values to the range $0 < S_g < 1$. The squaring ensures the contribution from each interbond angle is always greater than or equal to zero. If all the angles are arranged in perfectly tetrahedral order, the value of S_g will be the approach to zero. If the value of S_g is approaching 1 that shows distortion from tetrahedrality.

(ix) Preferential binding affinity: The pair correlation functions can be used to calculate the KB integral value. The physical significance of the KB integral can be understood in terms of mutual affinities between the interacting molecular species in the system. The KB integrals between solution components can be expressed using the equation

$$G_{\alpha\beta} = 4\pi \int_0^\infty r^2 \left[g_{\alpha\beta}(r) - 1\right] dr \tag{2.31}$$

where $g_{\alpha\beta}$ is the radial distribution function, and r is the interatomic separations. A higher value of $G_{\alpha\beta}$ indicates an overall stronger inter-atomic attraction between the species α and β (either direct or mediated by other components). If solvent water is denoted by subscript α, protein by β, and salts by γ, then $v_{\alpha\beta}$preferential binding (Lin et al. 2013; Pierce et al. 2008) of water to protein is given by:

$$v_{\alpha\beta} = \rho_\alpha \left(G_{\alpha\beta} - G_{\beta\gamma}\right) \tag{2.32}$$

(x) Entropy: The excess entropy S_2 of the biomolecular system can be calculated with the help of the pair correlation function. For a system, consisting of a total number of N particles, the pair entropy is given by,

$$\frac{S_2}{Nk_B} = 2\pi\rho \int_0^\infty g\,(r)[lng(r)] - [g(r) - 1]r^2 dr \tag{2.33}$$

Most simple liquids obey Dzugutov scaling law (Errington et al. 2006), which uses microscopic reduction parameters, and is stated as,

$$D^* = 0.049e^{S_2/Nk_B} \tag{2.34}$$

where g(r) is the radial distribution function and ρ is the number density, D* is diffusion, k_B is the Boltzmann constant.

(xi) Secondary Structure Analysis

The evolved secondary structure such as alpha helix, extended β-sheet, turns and coils of the peptide/protein during the course of the simulation was analyzed *via* the DSSP program. The gmx/gmx_mpi_d do_dssp command was used to generate the matrix

file and the gmx/gmx_mpi_d xpm2ps utility was used to generate the time evolution of the secondary structure of the peptides.

2.4.2 Dynamical Parameters

(i) ***Mean square displacement:*** Self-Diffusion coefficients can be calculated as a function of the velocity autocorrelation function or mean square displacement (MSD) method. In the present study, we used the mean square displacement method to calculate the diffusion coefficient. The self-diffusion coefficient is given by

$$D = lim_{t\to\infty} \frac{1}{6N_M t} \left\langle \sum_{j=1}^{N_M} \left[r_j(t) - r_j(0) \right]^2 \right\rangle \tag{2.35}$$

Where D is the diffusion coefficient, N_M is the total number of atoms, t is the time, and $r_j(t)$ is the true displacement vector of the j^{th} atom at time t.

(ii) ***Hydrogen Bond lifetime:*** The hydrogen bond lifetime between the donor and acceptor of the intra-molecule and inter-molecules can be calculated. We define two molecules to be hydrogen-bonded if the interatomic distance between hydrogen bond donor and acceptor pairs is < 2.5 Å.

$$S_{HB}(t) = \frac{\langle h(0).H(t) \rangle}{\langle h(0)^2 \rangle} \tag{2.36}$$

where $S_{HB}(t)$ describes the probability of hydrogen bond formed from t=0 and remains bonded up to time *t*. The *h (0) and H(t)* are the variables based on the hydrogen bond population. The *h(0)* is unity when the particular pair (protein-water, protein-protein, and water-water) hydrogen-bonded at time *t* according to the definition zero otherwise (Kumar et al. 2007; Luzar 2000)

(iii) ***Survival Probability:*** Survival probability (SP) is an important quantity for characterizing the relaxation dynamics of metal cations and cosolvents on the surface of the biomolecules and near the water. It gives information about the probability of spices staying on the surface of a biomolecule or inside the biomolecule. SP is more important and useful when working with complex local environments such as peptides proteins, nucleic acid, other biomolecules, and some multiple functional groups. The SP is given by:

$$SP(t') = \frac{1}{T}\sum_{t=1}^{T}\frac{N(t,t+t')}{N(t)} \quad (2.37)$$

where T is the maximum time of the simulation, t' is the time between the analyzed configuration, N(t) number of particles in a region of interest at time t, and N (t, $t + t'$) is the number of particles in the same region (Araya-Secchi et al. 2014; Debnath et al. 2010). The relaxation trend of the $SP(t')$ provides information about the local dynamics of hydration water, and ions. We have fitted the relaxation curve with the KWW equation defined as $\varphi(t,\tau) = exp^{-(t/\tau)^{\beta}}$; where β is stretched exponential parameter that lies $0 < \beta \leq 1$ and τ is the relaxation time.

To quantify the characteristic relaxation times and for better fit, the normalized survival probability, i.e., S(t) obtained by using a single exponential function as given below is calculated for two different regions such as interface and bulk:

$$s(t) = a_0 exp\left(\frac{-t}{\tau_1}\right) \quad (2.38)$$

where a_0 is the pre-exponential constant, τ_1 is the decay constant corresponding to the time-span of the metal ions that are staying near the carbonyl group of protein and in the bulk water.

2.4.3 Free Energy Landscape

A free energy landscape is a map graph of possible states of a system. The method is frequently used in science to describe all possible conformations occurred in a molecular system. In this generally, the spatial positions of interacting molecules in a system, or parameters can be captured in their corresponding energy levels, mainly Gibbs free energy. Geometrically, the free energy landscape is the graph of the energy function across the configuration space of the system. Boltzmann inversion of conformational distribution (Mu et al. 2004) was employed to compute the FEL based on the following equation

$$\Delta G_i = -k_B T[lnP_i - lnP_{max}] \quad (2.39)$$

where ΔG_i is the free energy of i^{th} coordinate corresponds to the probability distribution of P_i. P_{max} is indicating the maximum probability. T is the absolute temperature, and k_B is the Boltzmann constant.

2.4.4 Principal Component Analysis

The principal component analysis (PCA) is an important statistical calculation that can reduce the multidimensional data into 2 or 3 dimensions. From the MD simulation trajectory, it is necessary to extract the major motion or the essential dynamics of the system that are directly related to the biological function such as protein or nucleic acids folding/unfolding and fluctuation in molecules. The principal component modes of biomolecular dynamics can be calculated by diagonalizing the covariance matrix defined as,

$$C_{ij} = \frac{1}{N}\sum_{k=1}^{N}(r_i^k - \langle r_i \rangle)(r_j^k - \langle r_i \rangle \quad (2.40)$$

In equation (2.40) r_i^k and r_j^k are the pair of elements that correspond to vector r^k which depict the configuration of the system at time step k. $\langle r_i \rangle$ and $\langle r_j \rangle$ are the time-averaged value of the elements for the N structures sampled in the MD simulation. In general, r is the vector that contains all the coordinates of the C_α atom of a protein. The eigenvector corresponds to the largest eigenvalue and provides the low dimensional subspace that can describe the maximum behaviours of the biomolecules.

2.4.5 Clustering Algorithm

Clustering is a type of machine learning (ML) in which the aim is to make a set of quantities into groups called clusters. Ostensibly, the entities within a cluster are more similar to each other than the entities from other clusters. There are many algorithms have been developed to study biomolecular systems with clustering of MD trajectory.

The behaviour of biomolecules such as DNA, RNA, and Protein, which are having a lot of confirmations, can be quantified with a clustering progress as given

$$Cp = 1 - \frac{N}{E} \quad (2.41)$$

where *N* is the actual number of times points such that frames *n* and *n*+1 is in a different cluster and *E* is the expected number of switches or cluster's size.

2.4.6 Network Analysis

Network analysis and clustering have been carried out to group the prominent structures where both orientation and conformation are equally essential to understand mechanism of protein function (Abramyan et al. 2016). The conformations for network analysis were obtained by in-built clustering algorithm in GROMACS software using RMSD as criterion. The selection of a proper cut-off is crucial as a lower cut off value gives too many conformations which may lead to irrelevant results; while a larger cut-off value neglects the finer details of the conformations. An open software, Gephi (Bastian et al. 2009), is used for constructing graphs for the network analysis. Edge and Node files were created by executing the following equation:

$$d(t,t') = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\delta_i(t)\right)^2} \tag{2.42}$$

where δ_i is the Euclidean distance between the coordinates of atom *i* at time *t'* with the same atom in the reference conformation (at *t*) of system containing N number of heavy atoms. If the RMSD is not mass-weighted, then $\delta_i = (r_i(t) - r_i(t'))^2$ where r_i is the coordinate vector of desired atom (*Section 2.4.1 (i)*).

CHAPTER 3

PREFERENTIAL BINDING AFFINITY OF IONS AND THEIR EFFECT ON STRUCTURE AND DYNAMICS OF WATER NEAR ANTIMICROBIAL PEPTIDE

Dissolved ions affect the stability of amino acids and proteins by altering the liquid water structure. In the present study, the behaviour of water at the interface and the bulk region of the biomolecules around hydrophobic and hydrophilic residues under the influence of different alkali metal ions, such as Li^+, Na^+, and K^+ is explored. This is an important unexplored question in understanding many biomolecular processes.

3.1 BACKGROUND

Ions are known to be classified as either kosmotropes (structure makers) or chaotropes (structure breakers) according to their relative abilities to induce the structuring of water (Dill et al. 2005; Mancinelli et al. 2007). Biomolecules, on the other hand, can perturb the structure of water molecules near their surface due to the presence of different hydrophobic and hydrophilic residues (Persson and Halle 2018). The associated distortions and loss of tetrahedral symmetry of the interfacial water molecules near the protein surface, in turn, affect their dynamic properties (Galamba 2013; Poole et al. 1992; Vedamuthu et al. 1994). “Biological water” are the water molecules around a biomolecule (protein, DNA, RNA, or a cellular membrane) that has separate and different properties from the aqueous bulk (Ageitos et al. 2017; Campo 2006; Fogarty et al. 2013; Nayar and Chakravarty 2013; Persson and Halle 2008; Sinha and Bandyopadhyay 2011). These water molecules are important for the structure, stability, dynamics, and function of biomacromolecules (Jungwirth 2015; Levy and Onuchic 2006). Water molecules near the apolar groups are found to have more crystallinity (Starr and Sciortino 2014) than bulk (known as the iceberg model) (Galamba 2013; Huraskin and Horn 2019; Laage et al. 2009). They are also found to behave differently in presence of various cosolvents and ions (Dilip and Chakraborty 2019; Huraskin and Horn 2019).

It is an already well-known fact that the stabilization and destabilization of the biomolecules in presence of TMAO and urea respectively occurs due to the direct and indirect mechanism (Lee et al. 2010). In view of this, it will be enthralling to see whether such mechanism exist when we compare the different size of ions. Since water behaves differently at different interfaces and in presence of different solutes, its structural and dynamical inhomogeneity at the interface of hydrophilic and hydrophobic residues of the same protein molecule in the presence of different ions will be a fascinating thing to study. The presence of different types of water near different residues can create different pressure zones in the solution leading to the difference in the diffusion of water molecules, ions across the different pressure zones which can lead to many biological functions such as folding-unfolding of proteins, transport of nutrients across the cells. This kind of study will be helpful to understand the bigger biological processes such as nutrient uptake and protein denaturation. Studies regarding different kinds of water have been limited only to the aqueous solutions of amino acids and the effect of model potential on the water molecules (Campo 2006; Nayar and Chakravarty 2013).

To address this, in the present study, the water structure near an antimicrobial protein were analyzed and tried to correlate it with the dynamical properties of the solution. The antimicrobial protein (PDB ID: 5Z32) (Mohid et al. 2019) considered here have different subunits such as hydrophilic, hydrophobic, and mixed zone to see the contrasting difference in the water structure. Different concentrations of LiCl, NaCl and KCl salt solutions were considered as the size of the ions increases from $Li^+ > Na^+ > K^+$ ions. The effect of different model potentials is also compared.

3.2 SYSTEM SETUP

The starting coordinates for the simulation of AMP were taken from the NMR structure (PDB ID: 5Z32) (**Figure 3.1**). Molecular dynamic simulation were carried out in a cubic box comprising of the antimicrobial peptide, alkali metal ions, and water molecules. The compositions of each solvent mixture in three different concentrations are given in **Table 3.1**. The systems were neutralized anions. All atomistic simulations were performed with the GROMACS (version 2016.5) software

package (Berendsen et al. 1995; Van Der Spoel et al. 2005). Two different sets of simulations having different force fields were performed to compare the effect of the model potentials. A combination of CHARMM (Best et al. 2012) and SPC/E (Berendsen et al. 1981, 1987; Khalak et al. 2018; Oliveira and Colherinhas 2020) water systems, and a combination of AMBER (Duan et al. 2003) and TIP4P (Mahoney and Jorgensen 2000) were chosen to explore the effects of the force field parameters. The ion potentials used correspond to the CHARMM and AMBER force fields in the respective simulations. The details of the ion potentials are given in **Appendix I**. The different combination of the force field and the water model agrees qualitatively. The SPC/E water model is chosen since it represents the water properties more correctly and there have been reports which show that this combination works well (Průša and Cifra 2020; Tarek and Tobias 2000; Wolf and Groenhof 2012).

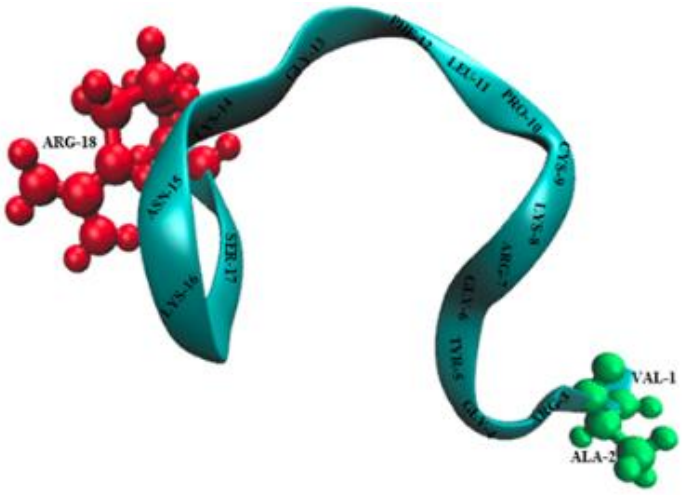


Figure 3.1: Structure of the selected antimicrobial peptide.

The equations of motion were integrated using the leapfrog algorithm with a time step of 10^{-15} s (1 fs) along with minimum image conventions and periodic boundary conditions in all three directions (Allen and Tildesley 2017). The minimum image conventions for calculation of the short-range Lennard-Jones interactions were employed and Particle Mesh Ewald (PME) (Darden et al. 1993; Essmann et al. 1995) sum was used to treat the long-range electrostatic interactions with nonbonded interaction space cut-off of 1.2 nm. Energy minimization was done using the steepest descent algorithm and the systems were equilibrated in NVT ensemble for 20 ns at

300 K. All simulations were performed in NPT ensemble for 100 ns at 300 K and 1 bar using Parrinello-Rahman barostat (τ_p= 2.0 ps) (Parrinello and Rahman 1980) and Velocity-rescale (τ_T = 0.1 ps) (Bussi et al. 2007) with modified Berendsen thermostat (τ_T = 0.1 ps) (Berendsen et al. 1984) were employed to keep the pressure and temperature constant respectively. LINCS algorithm was applied to keep bond lengths constrained. The simulations were well equilibrated and the RMSD values converged in 100 ns simulation (**Appendix II**).

Table 3.1: N_{AMP}, N_{KCl}, N_{NaCl}, N_{LiCl}, N_{water} represent the number of antimicrobial protein units, KCl, NaCl, LiCl, and water molecules in the simulation box.

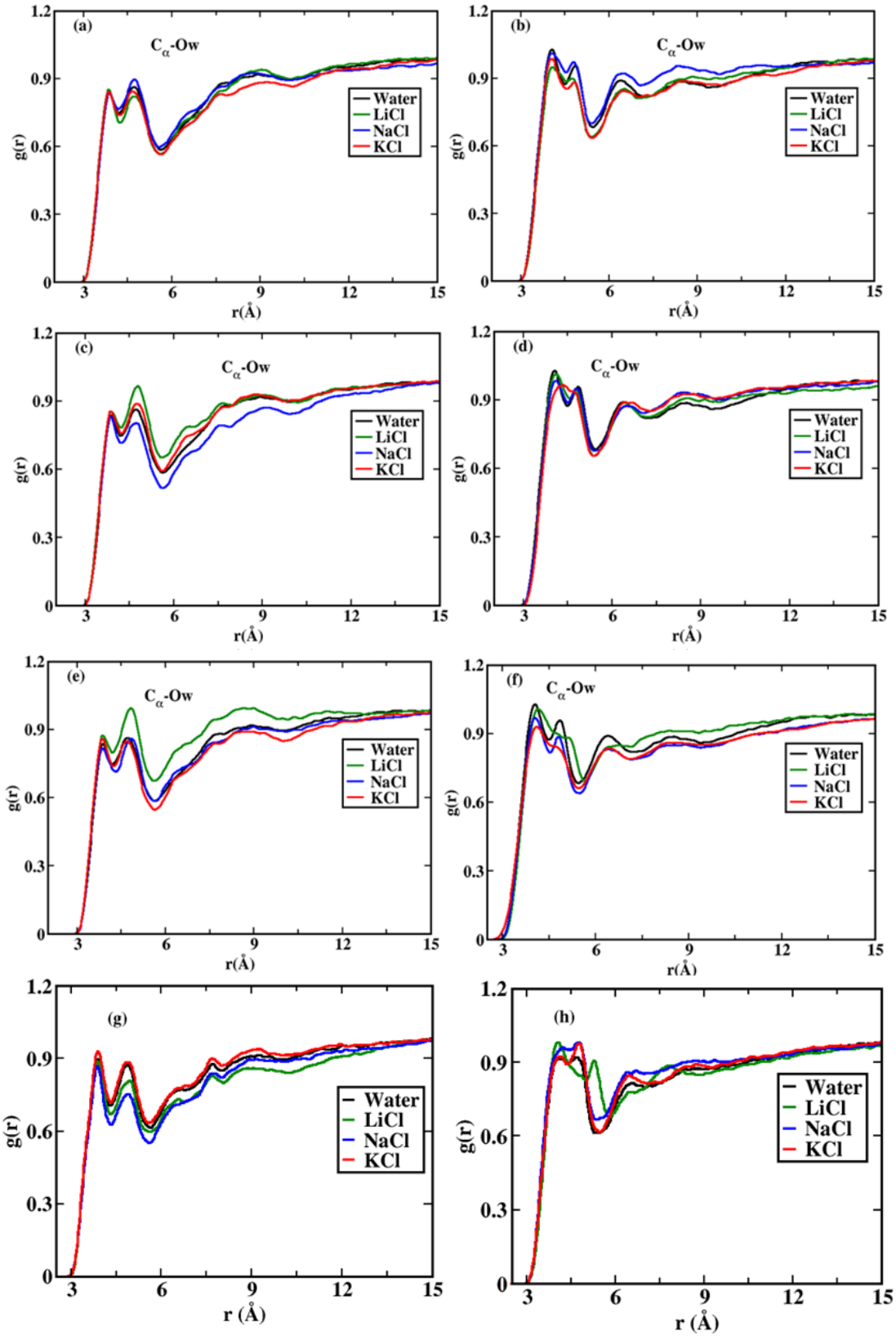
System	N_{AMP}	N_{LiCl}	N_{NaCl}	N_{KCl}	N_{Water}
AMP (0.0 M)	1	1	1	1	3264
AMP (0.4 M)	1	10	10	10	3255
AMP (1.0 M)	1	30	30	30	3235

3.3 RADIAL DISTRIBUTION FUNCTIONS

Salt ions are widely present in chemical and biological systems and are of vital importance to biochemical activities. The two terminals of this peptide have two different environments (**Figure 3.1**). One corresponding to around Residue 2, ALA, and the other corresponding to around residue 18, ARG. The hydrophobic residue ALA is surrounded by VAL and the hydrophilic residue is connected to SER amino acid which shows that they have different environments. These two residues were selected to check the water structure and later, its dynamics in the presence of ions.

The C_α-O_w pair correlation functions of Alanine (hydrophobic) and Arginine (hydrophilic) amino acids were plotted (**Figure 3.2**). The C_α-O_w pair correlation function is chosen since the water structure near C_α is found to be more altered and informative (Dilip and Chakraborty 2019) especially in the presence of the ions. The other RDFs (C_β-Ow, N_H-Ow C_γ-Ow, of ARG and C_β-Ow, N_H-Ow of ALA) (**Appendix III, IV**) are also found to be in the same trend when compared to C_α of ARG and ALA. The first solvation shell of ALA at different concentrations of salt is

found to have little differences (**Figure 3.2 (a)**, **(c)**, and **(e)**) which increases as we go to the higher solvation shells.



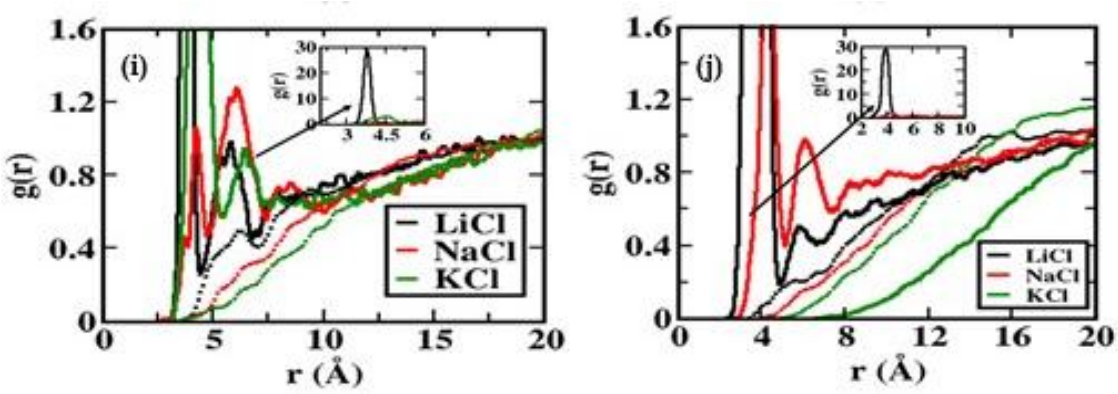


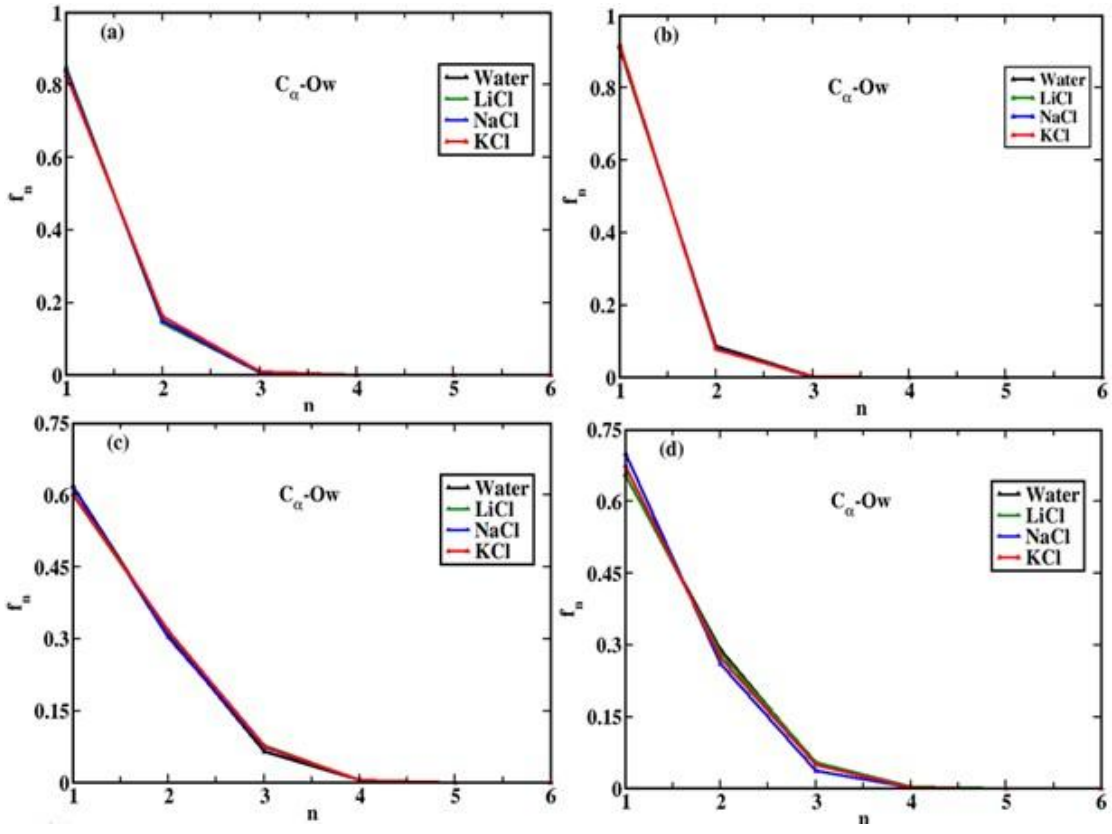
Figure 3.2: Radial distribution functions g(r) of (a-h) C_α-O_w of ALA and ARG amino acid residues of protein in the presence of various salt solutions. (a), (b) are the RDFs of 0 M; (c), (d) is the RDFs of 0.4 M and (e), (f) are the RDFs of 1.0 M of ALA and ARG respectively for the CHARMM-SPC/E water model system. (g) and (h) shows the RDFs of 1.0 M of ALA and ARG for the AMBER-TIP4P model and (i) and (j) show g(r) of C_α-ion of ALA (dash line) and ARG (solid line) amino acids at 1.0 M for CHARMM-SPC/E and AMBER-TIP4P system respectively.

It can be seen that in the presence of salts the peak height of the second and third solvation shells rises with the increase in the concentration and is maximum for LiCl in the case of the CHARMM-SPC/E model concerning the other two types of salt at higher concentrations; whereas, in case of AMBER-TIP4P model, it is KCl. The sharp and well-defined peaks in **Figure 3.2(e)** suggest that the water structure is changed and it becomes more ordered in the presence of Lithium salt. The structural arrangement of water molecules near C_α of ARG is shown in **Figure 3.2 (b)**, **(d)**, and **(f)**. In this case, it is found that all the solvation peaks are changing with the salt concentrations near the ARG residue which shows that the solvation structure of the water molecules is very much different for different salts near the ARG residue. When applied AMBER-TIP4P, force field the differences in the RDF are shown in **Figure 3.2 (g)** and **(h)**. The solvation structure of water molecules is altered in this case too. **Figures 3.2 (i)** and **(j)** show the RDF of C_α-ion near ARG and ALA. As it is evident from the curves that the lithium-ion is found to be present near the protein surface in comparison to the other ions and more in the presence of ARG residue. The presence of the ions near the protein surface will affect the solvation structure around these

residues. For more detailed observation, the number of hydrogen-bonded water molecules near these residues was calculated.

3.4 NUMBER OF HYDROGEN-BONDED WATER MOLECULES

It is noted from the RDF that the solvation structures near the C_α of alanine and arginine are very different in the presence of different salt solutions. For a better understanding of the structure of water molecules near the interface of alanine and arginine, we plotted the fraction (f_n) of oxygen atoms of water molecules that contain *n* number of water-water hydrogen bonds for CHARMM-SPC/E systems in **Figure 3.3**. We have considered 3.25 Å as the distance criteria between two oxygen-oxygen atoms to be hydrogen-bonded (Dilip and Chakraborty 2019). We selected the water molecules which are present within 4.3 Å, 5.6 Å from C_α of alanine and arginine as interfacial water molecules and outside 9.0 Å from C_α of alanine and arginine as bulk water molecules based on the placement of the solvation shells in the RDF for calculations (Dilip and Chakraborty 2019).



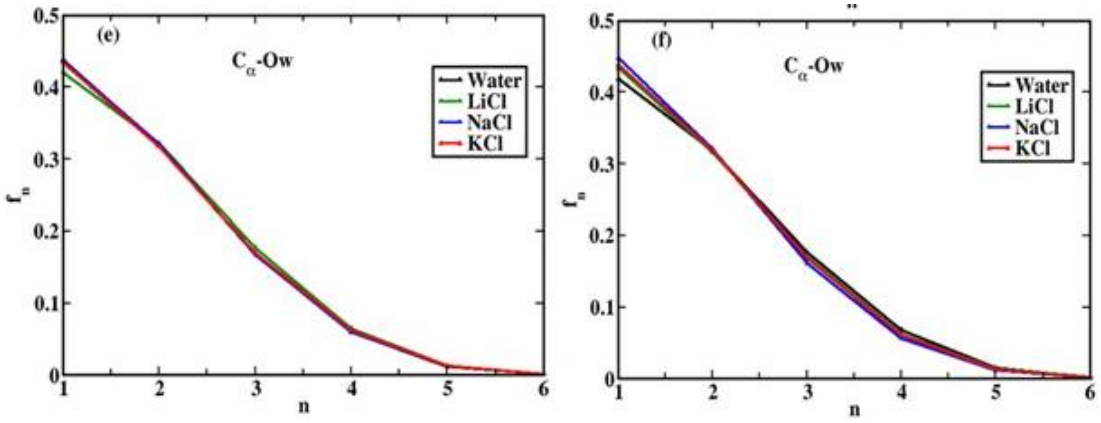


Figure 3.3: The fraction of water molecules having *n* number of hydrogen bonds within the distances [(a), (b) 4.3 Å] [(c), (d) 5.6 Å] and [(e), (f) > 9.0 Å] from C_α of ALA and ARG amino acid residues of protein in presence of 1.0 M salt solutions for CHARMM-SPC/E water model system. Graphs (a), (c), (e) and (b), (d), (f) are around C_α of alanine and arginine respectively.

In all the cases, it is observed that the probability of the existence of the lower coordinated (i.e., one or two coordinated) water molecules is found to be more, and higher coordinated is less. The lower coordinated water molecules exist due to the presence of the broken hydrogen bonds present at the interface of the biomolecules. The percentage of this lower coordinated hydrogen bond decreases as we go towards the higher solvation shells. The water molecules near ARG have more percentage of lower coordinated hydrogen bonds.

It is also observed that the number of hydrogen bonds between the water molecules present near the ALA residue has an insignificant effect due to the presence of different salt whereas; in the case of ARG, the number of hydrogen bonds between the water molecules present at the second **Figure 3.3(d)** and third **Figure 3.3(f)** solvation shells is affected sightly in presence of different salt in case of the SPC/E water model. In all the cases, lithium salts showed fewer broken hydrogen bonds compared to the other two salt solutions. At a higher solvation shell around 9.0 Å, the probability of finding water molecules having four and five coordinated hydrogen bonds increases as shown in **Figure 3.3(e)** and **(f)**. The fraction of water molecules having five coordinated water molecules is very low than four coordinated.

These water molecules arise due to the presence of interstitial water molecules between the solvation shells (Biswas et al. 2018). The order of salt solution which is showing higher coordinated water molecules in the bulk is LiCl > KCl > NaCl. This observation corelates with the coordination number results obtained from *ab initio* molecular dynamics (Bankura et al. 2013) and X-ray and neutron scattering studies (Mancinelli et al. 2007; Megyes et al. 2008; Soper and Weckström 2006). The trend is found to be similar near the C_α of alanine and arginine. To have a better overview of the tetrahedral structure of the water molecules we calculated the orientation profile of the bulk and interfacial water molecules which is discussed in the next section. The AMBER-TIP4P model systems were found to show similar distributions and are shown in **Figure 3.4**.

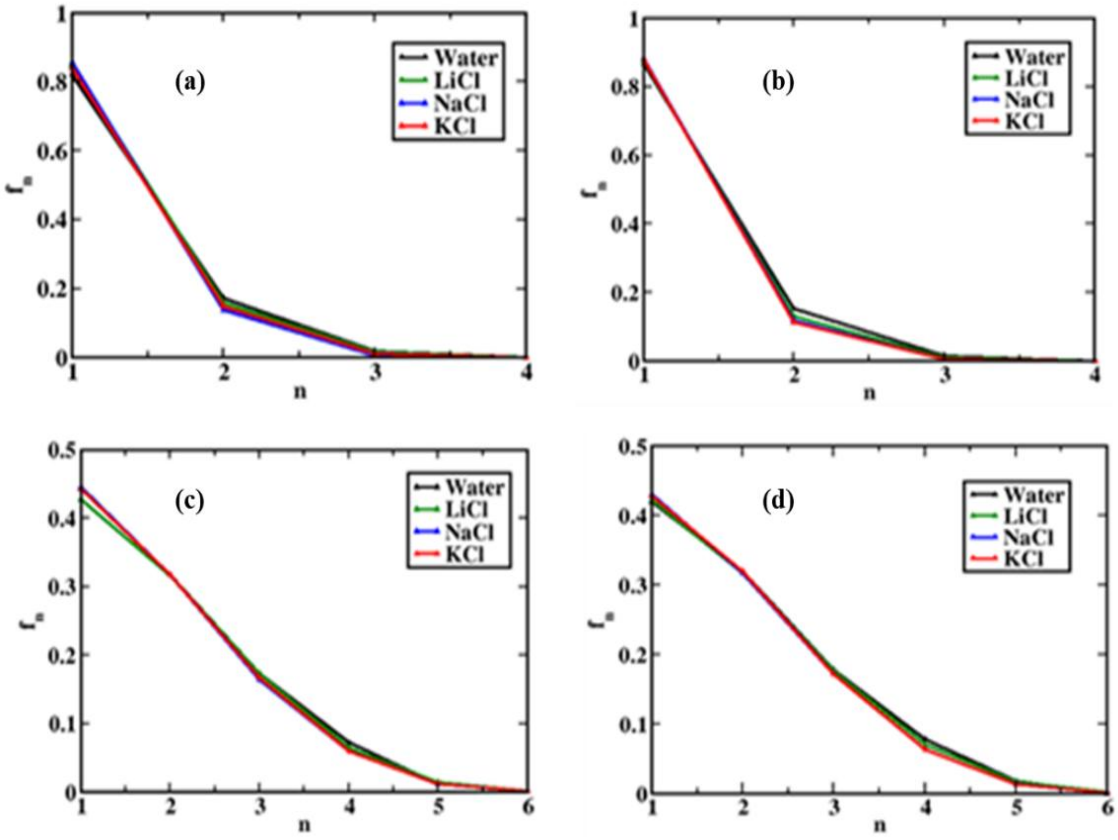
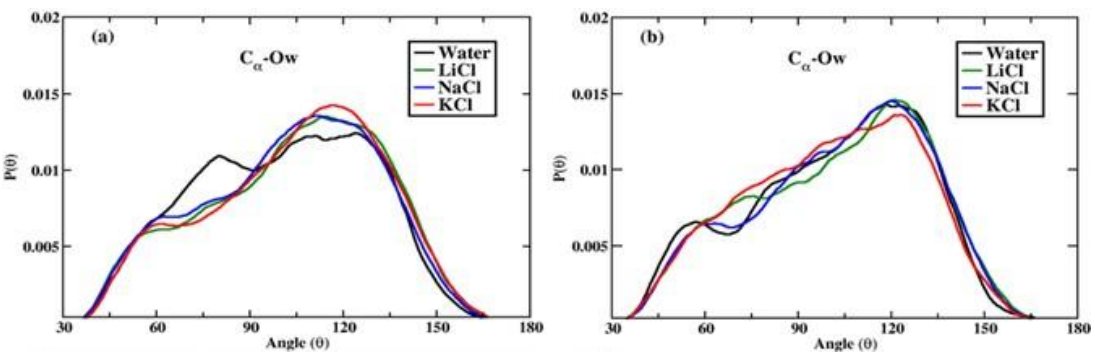


Figure 3.4: The fraction of water molecules having *n* number of hydrogen bonds within the distances [(a), (b) 4.3 Å] [(c), (d) > 9.0 Å] from C_α of ALA and ARG amino acid residues of protein in presence of 1.0 M salt solutions for AMBER-TIP4P model systems. The graphs (a), (c) and (b), (d) are around C_α of alanine and arginine respectively.

3.5 ORIENTATION PROFILE

To unravel the tetrahedral structure of water molecules, we calculated the angular distribution function of the water molecule present near the interface and the bulk region of the C_α of alanine and arginine residues in **Figure 3.5** for CHARMM-SPC/E water model system. A small peak around 50º and a broad distribution around 104.5º are signatures of bulk water molecules as seen in **Figure 3.5 (e)** and **(f)**. The smaller peak can be attributed to the broken or interstitial water molecules and the broad distribution around 104.5º is due to tetrahedral water molecules. In **Figure 3.5 (a)** and **(b)**, we have plotted the probability distribution $P(\theta_{O\text{-}O\text{-}O})$ of the interfacial water molecules within a cut-off 4.3 Å near the C_α of alanine and arginine respectively. The maximum distortion in the angle is found for the water molecules present near the first solvation shell of C_α of the ALA and ARG amino acid residues of the protein. It is interesting to note that the water molecules near the ALA residue in the aqueous solution have only one broad distribution ranging from 80º-130º (**Figure 3.5 (a)**). In the presence of ions, the pronouncement of the two peaks is observed and the placement of the second peak is found to shift towards higher angles indicating deviation from the tetrahedral angle. In the case of ARG, the scenario is quite different in the first solvation shell (4.3 Å) when compared to the ALA residue. The water molecules present in the aqueous solution are found to be more ordered compared to the presence of the ions (**Figure 3.5 (b)**). The small hump seen in the aqueous solution is found to merge onto a broader distribution having a maximum of around 125º in the presence of ions. This suggests that the water molecules present in this region are highly distorted from the tetrahedral structure in the presence of the ions. This can be explained due to the presence of the solvation shell of the ions.



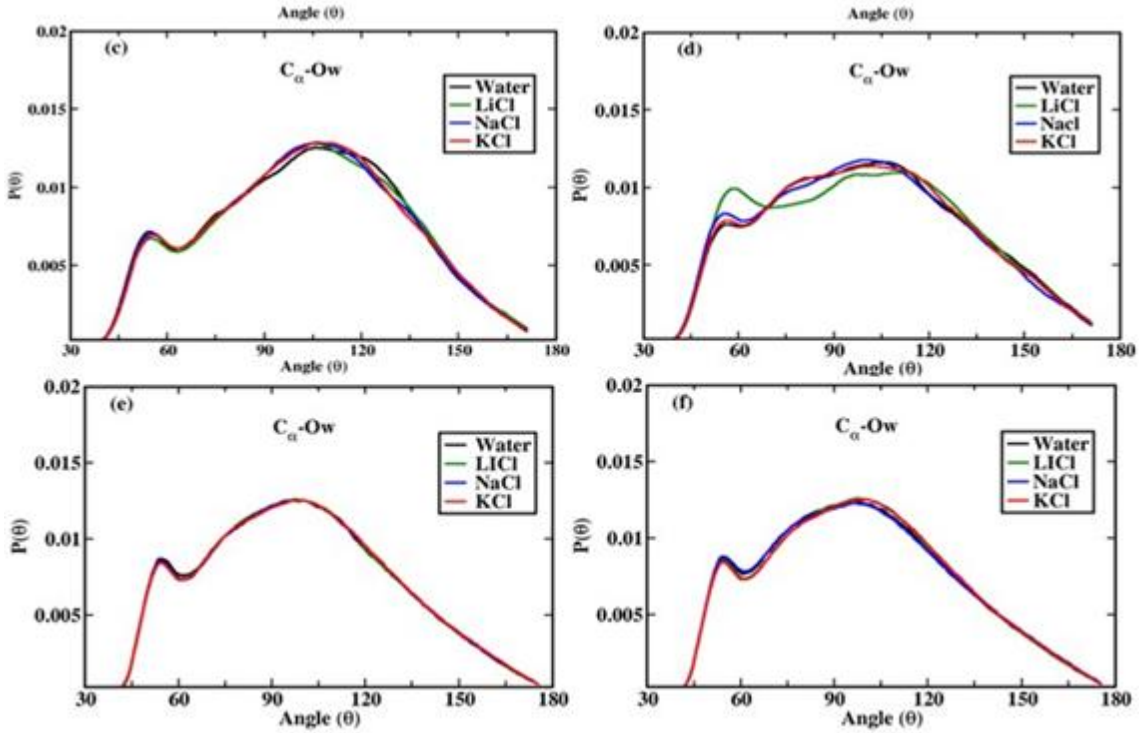


Figure 3.5: The normalized probability distribution of <O-O-O angle of oxygen atoms of water molecules within the distances [(a), (b) 4.3 Å] [(c), (d) 5.6 Å] and [(e), (f) > 9.0 Å] from C_α of ALA and ARG amino acid residues of protein in presence of various salt solutions of 1.0 M for CHARMM-SPC/E water model. Graphs (a), (c), (e) and (b), (d), (f) are around C_α of alanine and arginine respectively.

The <O-O-O angle probability distribution of water molecules in the second solvation shell at a cut-off of 5.6 Å near the C_α of alanine and arginine are shown in **Figure 3.5 (c)** and **(d)** respectively. In both cases, the broader <O-O-O angle distribution shifts towards 104.5º which indicates the water molecules present in the second solvation shell are more tetrahedral. In the case of LiCl, the height of the small peak is more pronounced compared to the other salts near the C_α of arginine. This can be explained as due to the presence of the higher charge density in Lithium ions, the water structure in the second solvation shell of Li is still disrupted. As we are going towards the higher solvation shells, the probability distribution P ($\theta_{O\text{-}O\text{-}O}$) of <O-O-O angle is found to be around 104.5º (**Figure 3.5 (e)** and **(f)**) showing more bulk-like character. For the TIP4P water model (**Figure 3.6**), the change in the structure of the water molecules is found to be less pronounced compared to the SPC/E water model

around the interface of the protein. Here also, it is noticed that the interfacial water molecules are more disrupted than the bulk water. Next, it will be informative if we look into the tetrahedral parameters of the water molecules in different solvation shells.

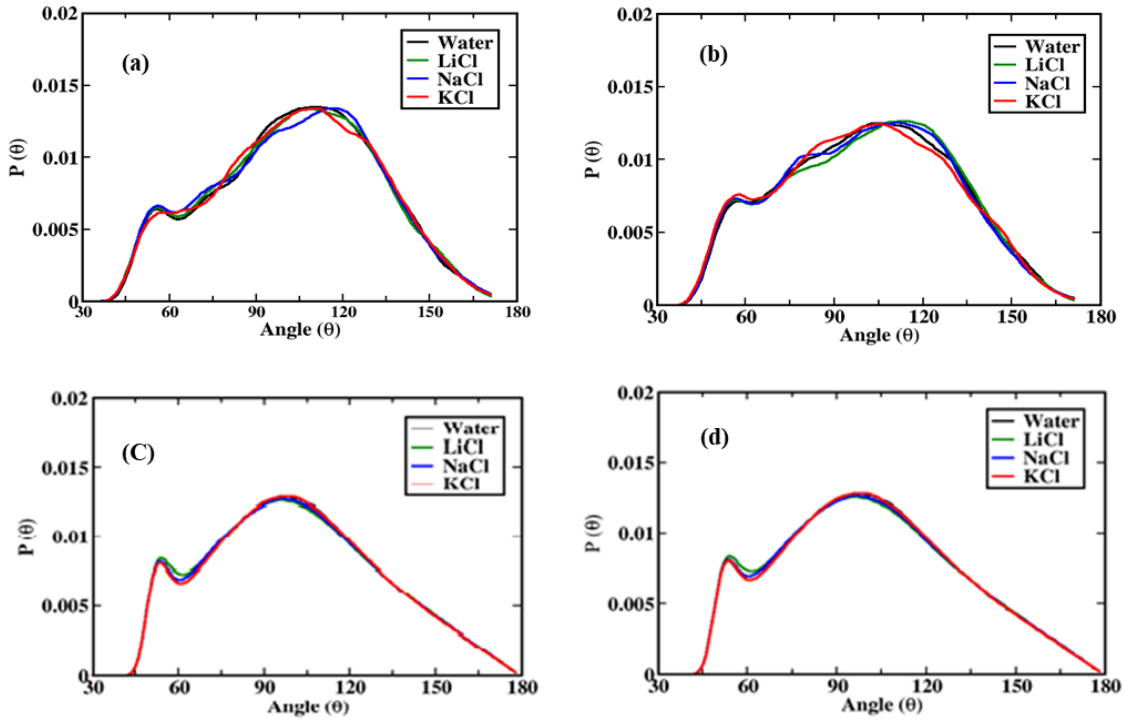


Figure 3.6: The normalized probability distribution of <O-O-O angle of oxygen atoms of water molecules within the distances [(a), (b) 4.3 Å] [(c), (d) > 9.0 Å] from C_α of ALA and ARG amino acid residues of protein in presence of various salt solutions of 1.0 M for AMBER-TIP4P model systems. Graphs (a), (c), (e) and (b), (d), (f) are around C_α of alanine and arginine respectively.

3.6 ORIENTATION TETRAHEDRAL ORDER S_g

The structure of water molecules near the hydrophilic and hydrophobic residues of peptide can be characterized based on the tetrahedral order parameter (Section 2.4.1 (viii)). The curve for the probability distribution of S_g is plotted in **Figure 3.7** for different cut-offs for ARG and ALA residues. It can be seen that the distribution of the first solvation (**Figure 3.7 (a)** and **(b)**) near the amino acid residues are much broader compared to the higher solvation shells (**Figure 3.7 (c)** and **(d)**). This is due to the presence of heterogeneity near the protein surface, which reduces as we go to

the higher solvation shells. The distribution is also found to be broader in the case of ALA residue compared to ARG. It is interesting to note that for the ALA case, initially in the first solvation shell, the S_g values are broader, and have both lower and higher values which suggest that the water structure is heterogeneous. In the presence of ions, the S_g values are a little less indicating more ordered and tetrahedral compared to the aqueous system which reverses in the case of ARG residue. For ARG, we find water molecules in the aqueous system, and the presence of KCl is slightly shifted towards the lower values (**Figure 3.7 (b)**). In the case of the AMBER/TIP4P case, it is found that the water molecules get more ordered in the presence of Li^+ ions near ALA and ARG. As we move towards the higher solvation shell, water molecules in the aqueous system and the presence of LiCl are found to be more ordered compared to in presence of K^+ and Na^+ ions (**Figure 3.7 (c)**, **(d)**, **Appendix V**).

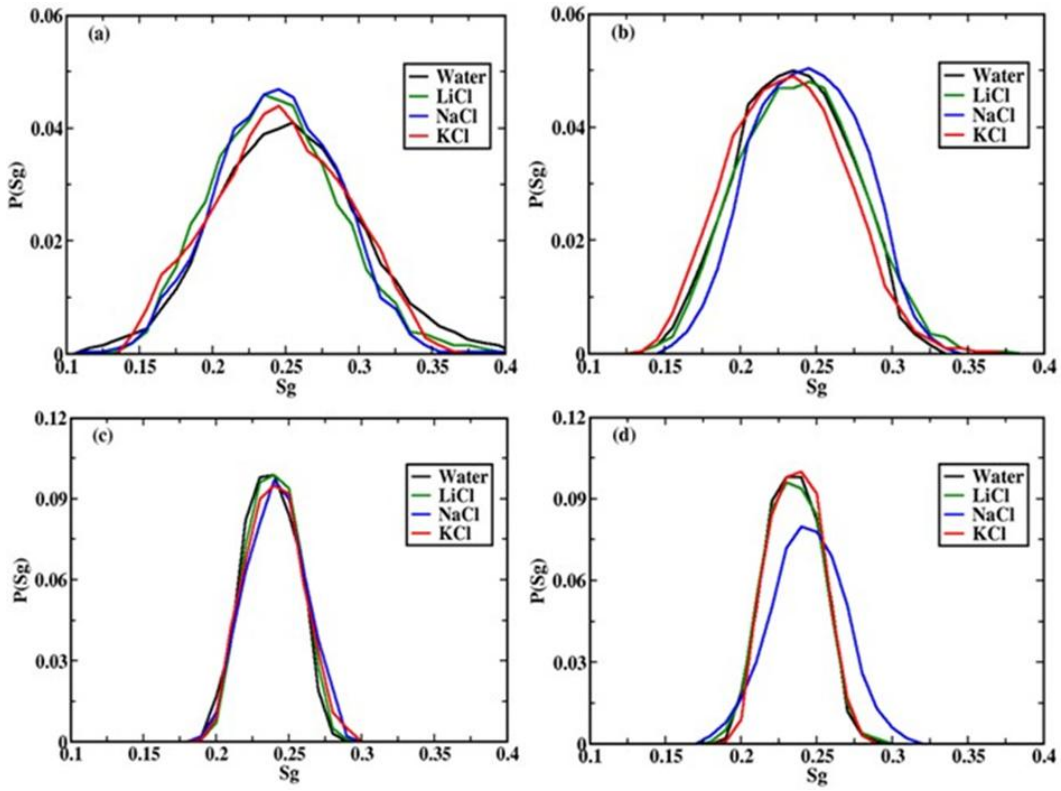


Figure 3.7: The orientation order parameter of water molecules within the distances [(a), (b) 4.3 Å] and [(c), (d) > 9 Å] from C_α of ALA and ARG amino acid residues of protein in presence of various salt solutions 1.0 M for CHARMM-SPC/E water model.

3.7 VORONOI ANALYSIS

As already observed that the presence of ions, affects the structure of the water molecules near the protein surface, it will be interesting to see their changes in the void analysis. We analyzed the voids in water molecules through the Voronoi polyhedra (VP) method, which is one of the efficient ways by which we can analyze the structural details of a system. Especially, it will be important for the current study where the arrangement of the water molecules in the coordination sphere of ions differs a lot from the bulk in the presence of salt solution. From a set of configurations of atoms, we can generate a tessellation in 3D space, known as Voronoi tessellation (Voronoi 1908). The tessellation consists of repetitive units of the VP region of the atoms. The units are defined as the regions consisting of all points that are closer to the reference atom than to any other atom. The void radius is defined as the distance between the vertex of VP and the atom minus the atom radius. The coordinates of the oxygen atoms were considered to generate the tessellation, and the Voronoi analysis was executed by the algorithm described in refs (Chakraborty and Chandra 2011a; b, 2012; Tanemura et al. 1983). The plot of the probability distribution of the voids present in the LiCl and KCl systems is presented in **Figure 3.8**. Void distribution for the second solvation shell (< 5.6 Å for C_α) and bulk water (the water outside 12 Å from C_α) were calculated for comparison.

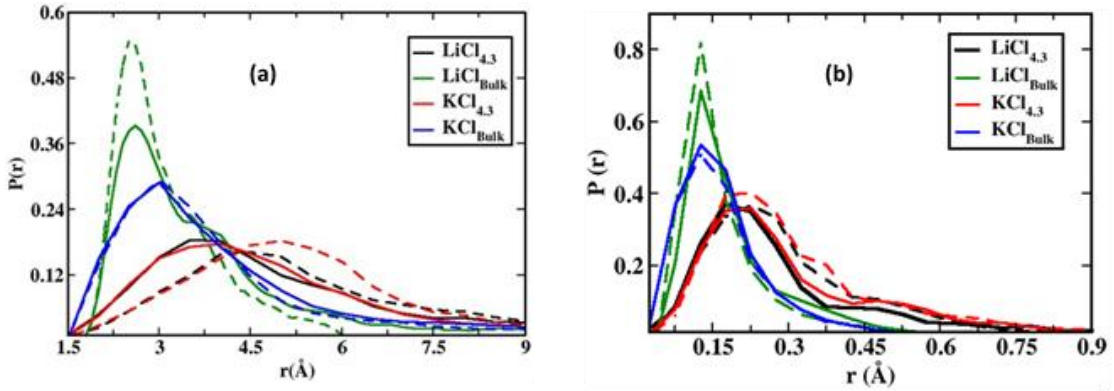


Figure 3.8: Probability distribution of radius of vacancies of water molecules in different salts solution for (a) CHARMM-SPC/E and (b) AMBER-TIP4P water model system. The solid and dashed lines represent ALA and ARG respectively.

It can be seen that the systems have quite a large impact on the void distribution of the system. The water molecules near the first solvation shell of the ions have broader distribution and higher void radius indicating less dense water molecules and more inhomogeneity. It can be also seen that the distribution of water molecules near ARG residue is shifted towards the higher void radius compared to water molecules near ALA residue which may be due to more fraction of broken hydrogen bonds near the ARG residue (**Figure 3.8 (b)**, **(d)**). In bulk water, however, the main peak of the distributions is found to be around an almost similar void radius for ARG and ALA; it is observed more in the case of KCl. It is found that the bulk water has a lower void radius and especially in the case of lithium implying more dense water molecules. A narrower distribution in this region suggests more homogeneity. It is interesting to notice here that the bulk water distribution corresponding to the KCl system is broader and at a higher void radius concerning the LiCl system suggesting more inhomogeneity and less dense water molecules than the lithium case. correlates with the The effects are found to be more pronounced in the case of SPC/E water compared to TIP4P (**Figure 3.8 (b)**). Next, it will be interesting to relate the structural changes as observed in the previous sections with the dynamical properties such as mean square displacement, entropy, and hydrogen bond dynamics.

3.8 MEAN SQUARE DISPLACEMENT

In the present study, we used the mean square displacement method to calculate the diffusion coefficient (Section 2.4.2 (i)). The diffusion coefficients were obtained from the slope of MSD in the long-time region for bulk and interfacial water molecules (Liu et al. 2023) (**Figure 3.9**). For the calculation of MSD for interfacial molecules, we considered water molecules that are present within 4.3 Å from the ALA and ARG residues. When they diffuse in the higher solvation shells, they are not considered interfacial water molecules. Similarly, for the calculation of bulk water molecules, we considered the water molecules which are present > 9 Å from the ALA and ARG residues. The bulk diffusion coefficient is calculated to be 2.53×10^{-5} cm^2 /s, which is slightly higher than the reported 2.40×10^{-5} cm^2/s diffusion coefficient for the SPC/E

water model (Glättli et al. 2003). This may be due to more inhomogeneous water molecules in the current system due to the presence of the protein.

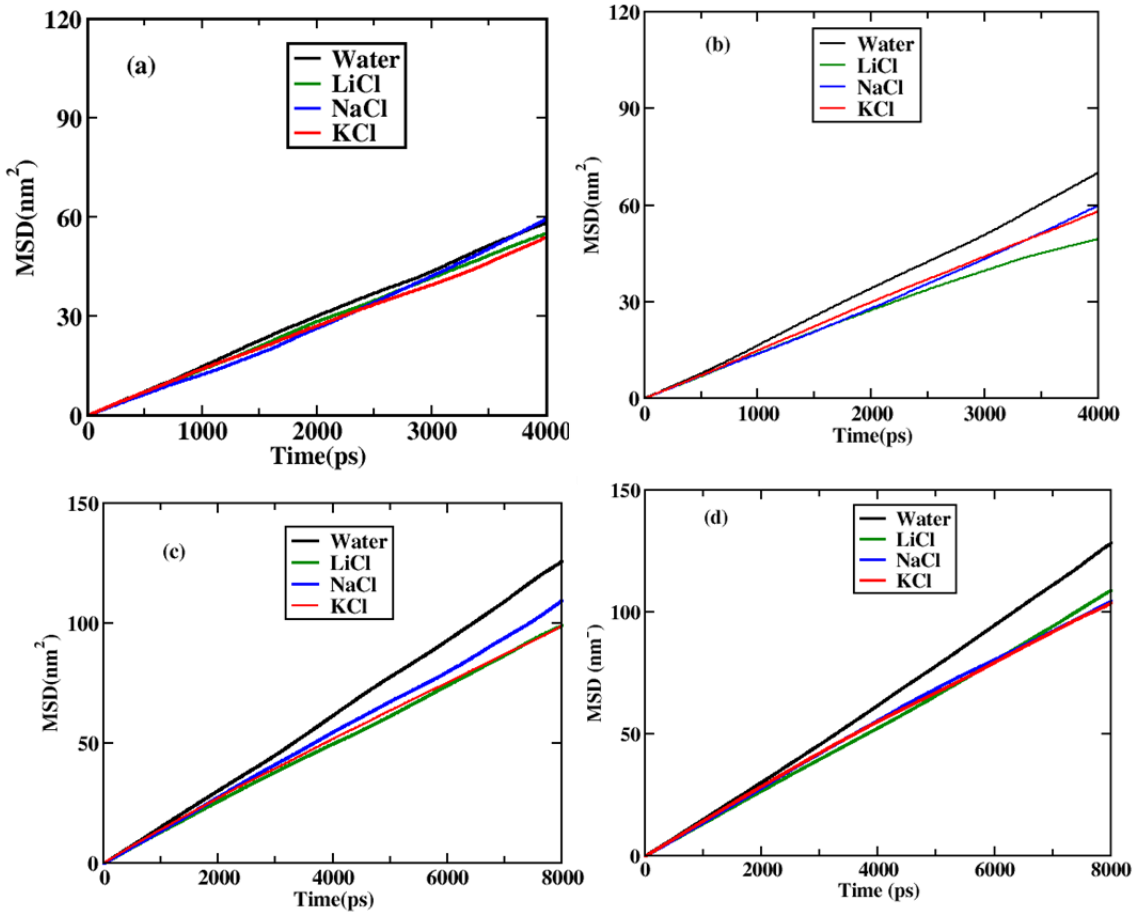


Figure 3.9: Diffusion coefficient of water molecules within the distances [(a), (b) 4.3 Å] and [(c), (d) > 9 Å] from C_α of ALA and ARG amino acid residues of protein in presence of various salt solutions 1.0 M for CHARMM-SPC/E model systems.

The diffusion values of water molecules around hydrophobic and hydrophilic amino acids of the protein are given in **Table 3.2.** The standard deviation reported in the table is calculated through the block average of our simulation trajectory. The details are given in **Appendix VI, VII**. The standard deviation data are found to be mainly between the ±0.02-0.04 range. From **Table 3.2**, it can be seen that the self-diffusion coefficients of the water molecules are found to be lower in presence of the ions. Interestingly, it can be seen that in presence of bigger ions (Na^+ and K^+), the first solvation shell water molecules are found to be more diffusive compared to the smaller ion Li^+. This can be explained because the water molecules are more tightly

bound to the protein surface in case smaller ions like Li^+ which holds the first solvation shell more tightly compared to the other bigger ions.

Table 3.2: Diffusion coefficient of water molecules around amino acids of antimicrobial protein in water, LiCl, NaCl, and KCl in D = (value) x $10^{-5}cm^2/sec$.

System	First Solvation Shell (< 4.3 Å)	Bulk (> 9 Å)
CHARMM-SPC/E model		
Water (ALA, ARG)	2.45 ±0.04, 2.48 ±0.03	2.51 ±0.03, 2.53 ±0.03
LiCl (ALA, ARG)	2.22 ±0.08, 2.14 ±0.07	2.39 ±0.03, 2.41 ±0.03
NaCl (ALA, ARG)	2.36 ±0.04, 2.30 ±0.07	2.43 ±0.03, 2.45 ±0.04
KCl (ALA, ARG)	2.41 ±0.05, 2.38 ±0.05	2.36 ±0.05, 2.46 ±0.04
AMBER-TIP4P model		
Water (ALA, ARG)	3.29 ±0.04, 3.17 ±0.07	3.53 ±0.05, 3.54 ±0.03
LiCl (ALA, ARG)	2.82 ±0.08, 2.56 ±0.08	3.28 ±0.03, 3.21 ±0.04
NaCl (ALA, ARG)	2.69 ±0.06, 2.52 ±0.07	3.32 ±0.03, 3.39 ±0.08
KCl (ALA, ARG)	2.65 ±0.02, 2.51 ±0.05	3.56 ±0.05, 3.50 ±0.03

As a result, the diffusion of the water molecules is less in the presence of Li^+. The value of the self-diffusion coefficient increases towards the higher solvation shell because the influence of the ions diminishes. Further, preferential binding coefficients of the cations were calculated (Section 2.4.1 (ix)).The preferential binding coefficient of the ions up to 5.6 Å from the protein surface was calculated (**Table 3.3**). The preferential binding coefficient of Li^+ ions with the protein is more which implies Li^+ remains with the protein surface more. But the bigger cations like K^+ prefer to remain with the bulk water. As a result, in the case of bigger ions, the self-diffusion value of water is more at the first solvation shell of the protein and does not differ much from the bulk value. Our result matches the experimental trend where the binding coefficient of Li^+ ions is more than Na^+ and K^+ ions (Aziz et al. 2008; Dong et al. 2009).

The results from orientation tetrahedral order, Voronoi analysis and diffusion coefficient agrees well with the dynamics of water in highly concentrated salt solutions studied using NMR approach (Rezaei-Ghaleh 2022).

Table 3.3: The preferential binding coefficient of the metal cation with the protein surface for the CHARMM-SPC/E model system.

System	0.4M	1.0M
LiCl (ALA, ARG)	2.39, 2.56	4.62, 10.23
NaCl (ALA, ARG)	2.13, 1.85	3.02, 2.58
KCl (ALA, ARG)	2.23, 2.27	0.98, 3.85

3.9 ENTROPY

As evident from the diffusion data, the water molecules at different regions behave differently, it will be interesting to relate the excess entropy of this different region with the diffusion. The excess entropy was calculated by pair correlation entropy S_2. RDF of O_w-O_w of water around the C_α of residue ALA and ARG of the protein were considered for the entropy calculation. The provincial values of S_2 can be used to evaluate the relation between the hydration layer of the water structure and the diffusion (transport properties) of water (Dzugutov 1996).

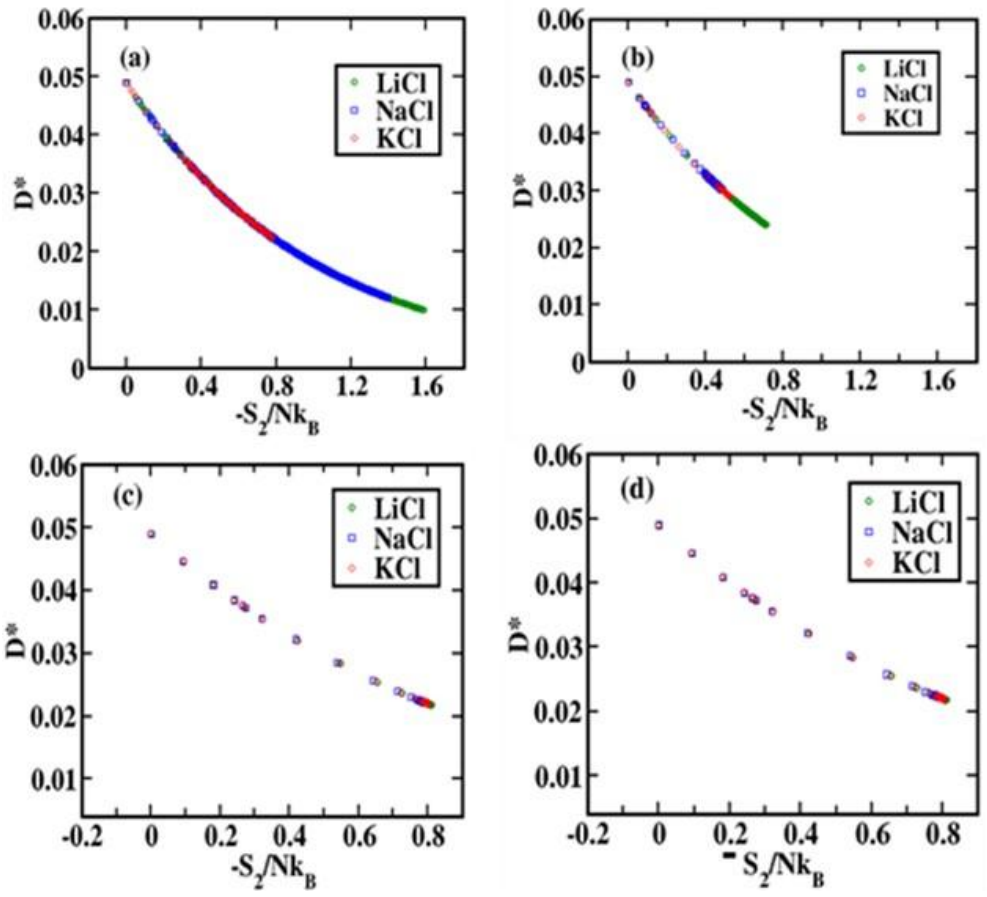


Figure 3.10: The relation between excess entropy and diffusion of water molecules within the distances [(a), (b) 4.3 Å] and [(c), (d) > 9.0 Å] from C_α of ALA and ARG amino acid residues of protein in presence of various salt solutions 1.0 M for CHARMM-SPC/E water model system.

It can be seen, that the diffusion and the pair correlation entropy (excess entropy) of water molecules present near the first solvation shell in the presence of LiCl salt solution are low when compared to NaCl and KCl (**Figure 3.10**). This indicates that the water molecules around both the hydrophobic and hydrophilic residues are open tetrahedral and low dense in the presence of LiCl salt solution (Chopra et al. 2010; Nayar and Chakravarty 2013). At higher solvation shells (outside 9.0 Å), it is seen that all the solution shows a similar type of behavior. A similar trend was found for the AMBER/TIP4P model (**Figure 3.11**).

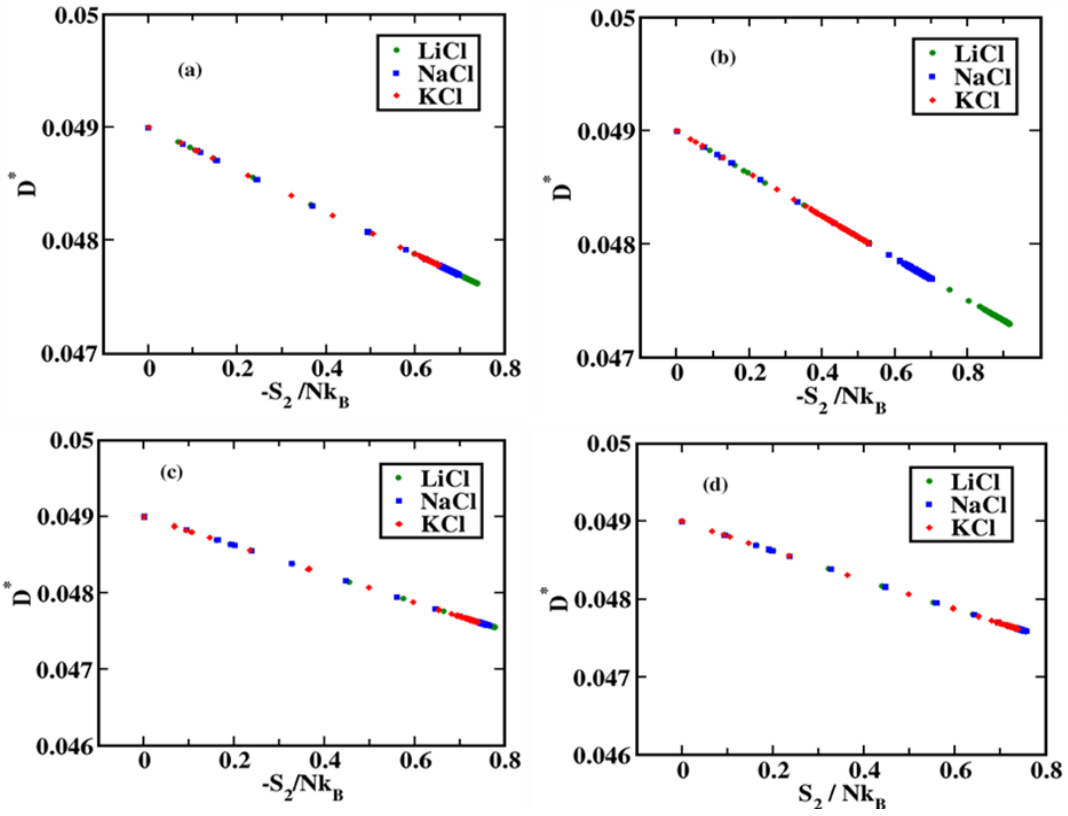


Figure 3.11: The relation between excess entropy and diffusion of water molecules within the distances [(a), (b) 4.3 Å] and [(c), (d) > 9.0 Å] from C_α of ALA and ARG amino acid residues of protein in presence of various salt solutions 1.0 M for AMBER-TIP4P model systems.

3.10 HYDROGEN BOND DYNAMICS

In aqueous media, peptide residues are capable of forming hydrogen bonds with water and they can also modify the regular water-water hydrogen bond network (Campo a* and Raul Grigera 2004; Kumar et al. 2007; Luzar 2000). The presence of different types of water in different regions will also affect the hydrogen-bond dynamics of the system. Therefore, the continuous hydrogen bond autocorrelation functions were calculated as described in Section 2.4.2 (ii).

Hydrogen bond lifetimes formed between water molecules around the peptide for CHARMM-SPC/E and AMBER-TIP4P are given in **Table 3.4**. For aqueous solutions near the protein, the hydrogen bond lifetime is more near the protein surface because of the non-availability of the neighboring water molecules to make new hydrogen bonds. Therefore, once the hydrogen bond is made, it is preserved for a longer time. The hydrogen bond lifetime of the solution increases in the presence of the ions. It can be seen that in the first solvation shell near the ALA residue of aqueous protein solution and in the presence of smaller ions like Li^+ and Na^+, the hydrogen bond lifetime is more; which decreases towards the bulk region. This can be explained due to the formation of strong hydrogen bonds between water molecules around the peptide residue in the presence of smaller ions. The smaller ions prefer to live near the surface of the protein as evident from the preferential binding affinity discussed in the diffusion section, and make strong hydrogen bonds with the water molecules.

Table 3.4: Lifetime (τ_{HB}) of continuous hydrogen bonds (in ps) formed by water-water around amino acids of antimicrobial protein in water, LiCl, NaCl, and KCl.

System	First Solvation Shell (< 4.3 Å)	Bulk (> 9.0 Å)
CHARMM-SPC/E water model		
Water (ALA, ARG)	0.854, 0.842	0.841, 0.837
LiCl (ALA, ARG)	0.883, 0.853	0.869, 0.855
NaCl (ALA, ARG)	0.880, 0.858	0.856, 0.846
KCl (ALA, ARG)	0.864, 0.837	0.855, 0.841

AMBER/TIP4P water model		
Water (ALA, ARG)	0.668, 0.658	0.661, 0.656
LiCl (ALA, ARG)	0.673, 0.662	0.667, 0.658
NaCl (ALA, ARG)	0.670, 0.658	0.665, 0.654
KCl (ALA, ARG)	0.667, 0.657	0.662, 0.653

For bigger ions such as K^+, this effect is less due to the size of the ions, resulting in less increment in hydrogen bond lifetime in the first solvation shell. A similar effect is observed for ARG. The same trend is followed in the AMBER/TIP4P water model system but is less pronounced compared to the CHARMM-SPC/E water model. These results are well correlated with the diffusion of the water molecules.

Therefore, it can be concluded that the ions involved with the protein surface give rise to low-density water at the interface. The water molecules are strongly bonded to these ions and also with themselves due to the electrostatic force of interactions of the ions. These make the water molecules less mobile. As a result, the entropy of water molecules decreases, and the hydrogen bond lifetime between water-water increases in presence of the ions.

3.11 CONCLUSION

In summary, the structural and dynamical properties of water molecules near the hydrophobic and hydrophilic unit of an antimicrobial peptide in presence of LiCl, NaCl, and KCl solutions at various concentrations we investigated using MD simulation. Two different model systems, namely the CHARMM-SPC/E water model system, and the AMBER-TIP4P water model system were studied. The structural properties were analyzed in terms of radial distribution functions, the number of hydrogen bonds, <O-O-O angle distributions, orientational tetrahedral order parameter S_g values. These characteristics were related to dynamical properties like diffusion coefficients, excess entropy, and hydrogen bond dynamics. The water structure in the first solvation shell is found to be maximumly disrupted and lower in density compared to the other solvation shells. The water molecules present hydrophilic environment is found to be more disruptive, less dense containing broken hydrogen bonds compared to the hydrophobic environment. The probability of broken

hydrogen bonds is more in the case of first and second solation shells. The tetrahedrality of the water molecules near the protein surface is found to be more in presence of Li^+ ions. Li^+ being the smallest ion, have more charge density which preferentially binds with the protein surface. As a result, the structure of the water in the first solvation shell is more tetrahedral due to the strong solvation shell of the Li^+ ion. This results to slow diffusion and lower entropy of water molecules which increases in the bulk. The bigger ions like K^+ ions have a lower preferential binding affinity towards the protein surface resulting in higher diffusion of the water molecules in the first solvation shell. The pairwise entropy, S_2, which is the first term in the excess entropy, shows a high correlation with the diffusion coefficients of the water molecules. We found low excess entropy for the first solvation shell water molecules containing Lithium salt and maximum for KCl solutions. This is further confirmed by the calculation of the hydrogen bond dynamics of the water molecules in the first solvation and bulk.

The addition of the salt solutions was found to increase the hydrogen bond strength of water compared to that of pure protein water. An increase in the hydrogen bond strength in the presence of salt solutions was observed with increasing the salt concentration. In the case of Lithium ions, we found the maximum hydrogen bond lifetime in the first solvation shell. The results suggest that the hydrophilic and hydrophobic amino acids attribute to the formation of different density regions of water molecules in the salt solutions which determines the dynamic property of the water molecules. The results of the two different force field combinations showed a similar trend; however, the SPC/E water model showed a more pronounced effect compared to the TIP4P water model.

CHAPTER 4

INFLUENCE OF ION SPECIFICITY AND CONCENTRATION ON THE CONFORMATIONAL TRANSITION OF INTRINSICALLY DISORDERED SHEEP PRION PEPTIDE

The structural sensitivity of IDPs with the ions has been observed experimentally; however, it is still unclear how the presence of different metal ions affects the structural stability. The present study aims to study the effect of size, charge, and ionic concentration on the structure of sheep prion peptide using monovalent and divalent cations.

4.1 BACKGROUND

Metal ions are known to have a significant effect on the various type of biomolecular events by shielding the unfavorable charge repulsion, thus the interaction forces come to play in stabilizing the biomolecular processes like protein folding, and protein-protein associations (Maclean et al. 2002; Song et al. 2007; Tobias and Hemminger 2008). However, the role of metal ions in the context of IDP pathogenesis remains highly debated due to their ability to modify the electrostatic interaction (Huat et al. 2019). This attracted both experimental and theoretical attention. Much of the early works support that multivalent cations such as Cu^{2+}, Zn^{2+}, Ca^{2+}, Mn^{2+}, and Fe^{3+} promote the aggregation-prone molten globule state of IDPs by binding with the charged amino acids at the protein surface (Davies et al. 2011; Fink 1998; Tamamizu-Kato et al. 2006; Tõugu et al. 2011; Uversky and Fink 2004) or increasing the mean residence time of hydration-shell waters (Smith and Cruz 2013). Additionally, the transitional metal ions such as Hg^{2+} are reported to favor the transition of the unstructured coil to β-sheet conformation for IDPs like tau protein (Luczkowski et al. 2015). Moreover, Cu^{2+} and Zn^{2+} exhibit a significant effect on prion misfolding and the formation of toxic oligomeric species (Jones et al. 2004; Wadsworth et al. 1999). Thus, patients with neurodegenerative disorders are found to have elevated concentrations of divalent metal ions (metal dyshomeostasis) in the brain and body

fluids (Szabo et al. 2016). In contrast, Mg^{2+} has a comparatively lesser propensity to increase the polymorphic behavior of IDPs, thereby reducing their self-aggregation (Wicky et al. 2017). Unlike bivalent cations, alkali metal ions such as Na^+ and K^+ are reported to show a less pronounced effect on the stability of monomeric IDPs or their lower-order aggregates. Recent works of Huraskin and co-workers showed that Na^+ makes bridging interaction with the carboxylate groups of amyloid fibrils while K^+ interacts transiently (Abelein et al. 2016; Huraskin and Horn 2019). However, there remains a large knowledge gap in understanding the role of Lithium ions on the conformation of IDPs. Lithium salts are known to be used as psychiatric medication (bipolar disorder) and many cells or animal studies proved that Li^+ downregulates the abnormal protein aggregation by controlling the off and on pathways of amyloidosis (Berntsson et al. 2021). For instance, recent mice studies pointed out that Li^+ plays important role in clearing the lower order and toxic Aβ fibrils from the brain, thereby attenuating the neuronal damage (Pan et al. 2018). Moreover, as a congener of Na^+ ion, Li^+ can easily exploit the various sodium transporters such as $Na^+/K^+/Cl^-$ cotransporters, Na^+/H^+ exchangers, sodium/bicarbonate symporters (NBC), Na^+/H^+ antiporters expressed at the surface of human brain microvascular endothelial cells to cross the blood-brain barrier (BBB) (Luo et al. 2018). This further indicates the utilization of lithium as a promising therapeutic agent for neurodegenerative disorders (Machado-Vieira et al. 2009). Although the application of various lithium salts against prion misfolding and propagation is becoming progressively evident, the underline molecular mechanisms of Li^+ ions to inhibit prion misfolding are far from reaching consensus.

From the theoretical point of view, the fraction of charged amino acids and the net charge of residues are important metrics for understanding the salt-dependent conformational transitions of IDPs. The IDPs with a net charge close to zero are often found to show salt-dependent expansion or unfolding due to the change in the hydrophobic environment (Mao et al. 2010). Additionally, the solubility of the amino acids present in the IDP sequence is reported as the key to understanding their conformational change which is highly salt-type and salt-concentration dependent (Baldwin 1996; Zhou 2005). At low salt concentration, functional folding of IDPs is

readily observed due to the dominance of electrostatic screening (Avbelj and Moult 1995). Whereas, at a high salt regime, the IDPs generally go to a misfolded state and prefer self-assembly by salting-out effect (Goto et al. 2017). In order to investigate the effect of electrostatic screening or salting-out effect on IDP confirmation at the molecular level, previous all-atom implicit solvent and coarse-grain MD simulations are available (Dignon et al. 2018; Vancraenenbroeck et al. 2019; Vitalis and Pappu 2009; Wohl et al. 2021). However, the elucidation of salt-specific interaction along with the salt-induced folding/misfolding at atomic resolution has been missing. Therefore, there is an urgent requirement to introduce a computational model of salt-IDP interaction at atomic resolution.

4.2 SYSTEM SETUP

The 3D structure of the prion peptide segment 142-166 (PDB ID: 2RMW) was retrieved from the solution state NMR model of 26 residues peptide (Bertho et al. 2008). The amino acid sequence of the peptide is shown in **Figure 4.1**. A 1 ns MD simulation of the peptide at 400 K was conducted for obtaining the unstructured conformation as the initial coordinate for our simulations.

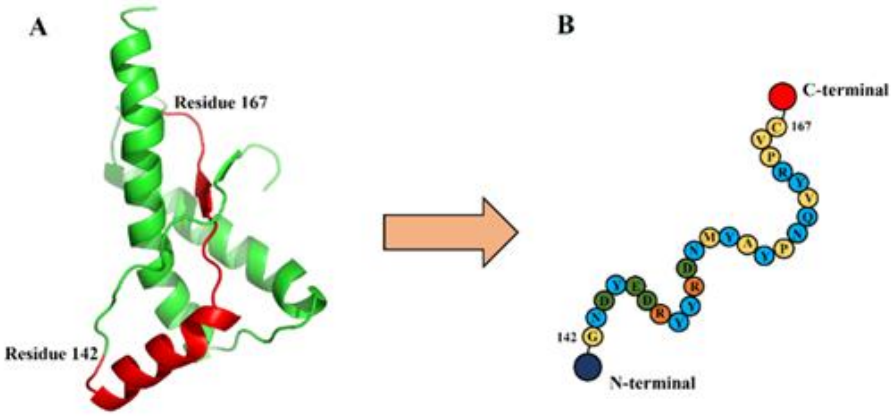


Figure 4.1: System of interest for the present study: (A) The secondary structure arrangement of cellular prion protein (PrP^{C}). The red-colored region is indicating the intrinsically disordered region, susceptible to misfolding. (B) The schematic view of the amino acid sequence of the intrinsically disordered region. Hydrophobic, polar, acidic, and basic residues are colored yellow, sky, green, and brown respectively.

All atomistic simulations were performed employing the GROMACS (Berendsen et al. 1995; Van Der Spoel et al. 2005) software package and CHARMM force field (Best et al. 2012). The N, C-terminal segment of the peptide was kept in a zwitterionic state and the peptide was solvated by TIP3P (Jorgensen et al. 1983) water molecules in a cubic box of 4.9 nm^3. The protonation states of the amino acids were assigned to the pH of 7. At this pH, three arginine residues were protonated and four acidic residues (two Aspartic acids and two Glutamic acids) were deprotonated which makes the overall charge of the protein -1. The overall charge of the system was neutralized by adding one Li^+ ion. In order to study the effect of cations on the folding of prion peptides, the chloride salts such as LiCl, NaCl, KCl, and $MgCl_2$ were added to the system at different concentrations. The details of the simulated salt-peptide systems are shown in **Table 4.1**. In the present work LiCl, NaCl, KCl, and $MgCl_2$ were considered single point charges, and water molecules were characterized by three-point interaction models. The ion parameters (Oliveira and Colherinhas 2020) used in the current study are shown in **Appendix I**.

Table 4.1: Summary of simulated systems‡‡ C_{ion}, N_{LiCl}, N_{NaCl}, N_{KCl}, N_{MgCl2}, $N_{protein\text{-}atom}$, N_W represents ionic concentration, number of LiCl, NaCl, KCl, $MgCl_2$, molecules, number of atoms in prion peptide 142-167, number of water molecules respectively. *3 is showing number of simulation replicas for each system.

System	C_{ion} (M)	N_{LiCl}	N_{NaCl}	N_{KCl}	N_{MgCl2}	$N_{protein\text{-}atom}$	N_W	Simulation time (ns) and sampling interval (ps, subscript)
S1	0	0	-	0	0	440	3710	(300_{10}) *3
S2	0.3	10	-	-	-	440	3691	(300_{10}) *3
S3	0.6	20	-	-	-	440	3671	(300_{10}) *3
S4	0.9	30	-	-	-	440	3648	(300_{10}) *3
S5	0.3	-	10	-	-	440	3691	(300_{10}) *3
S6	0.6	-	20	-	-	440	3671	(300_{10}) *3
S7	0.9	-	30	-	-	440	3651	(300_{10}) *3
S8	0.3	-	-	10	-	440	3691	(300_{10}) *3

S9	0.9	-	-	30	-	440	3671	(300_{10}) *3
S10	0.1	-	-	-	3	440	3710	(300_{10}) *3
S11	0.15	-	-	-	5	440	3698	(300_{10}) *3
S12	0.3	-	-	-	10	440	3691	(300_{10}) *3

The equations of motions were integrated using the leapfrog algorithm with a time step of 2 fs along with periodic boundary conditions in all three directions (Allen and Tildesley 2017). The minimum image conventions were used for calculating the short-range Lennard-Jones interaction terms and Particle Mesh Ewald (PME) (Darden et al. 1993; Essmann et al. 1995) was employed to treat the long-range electrostatic interactions with a nonbonded interaction cut-off of 1.2 nm. The van der Waals interaction cutoff was set to 1.2 nm. LINCS (Hess et al. 1997) algorithm was applied to constrain the bonds attached to hydrogen atoms of the peptide molecule and the SETTLE algorithm (Miyamoto and Kollman 1992) was employed to constrain the geometry of water molecules. The systems were equilibrated in canonical ensemble (NVT) followed by the isothermal-isobaric ensemble (NPT) for 20 ns each respectively by restraining the solute heavy atoms. The thermodynamic equilibrium of the system was ensured by plotting the time evolution of temperature, pressure and potential energy as depicted in **Figure 4.2**. Next, the restrain was removed and the peptide molecule was allowed to move freely during the production run of 300 ns. Later for hydrogen bond dynamic calculations, we have run the simulation with 1 fs timestep for 1 ns from the final trajectory. In order to assess the robustness of our simulation data, three sets of MD simulations were carried out for each system. The temperature and pressure of the system were maintained employing Velocity rescale (Bussi et al. 2007) (τ_t=0.1ps) and Parrinello-Rahman coupling algorithm (Parrinello and Rahman 1981) (τ_p=0.2 ps). The VMD (Humphrey et al. 1996) tool was used to visualize the folding process of the peptide in different salt.

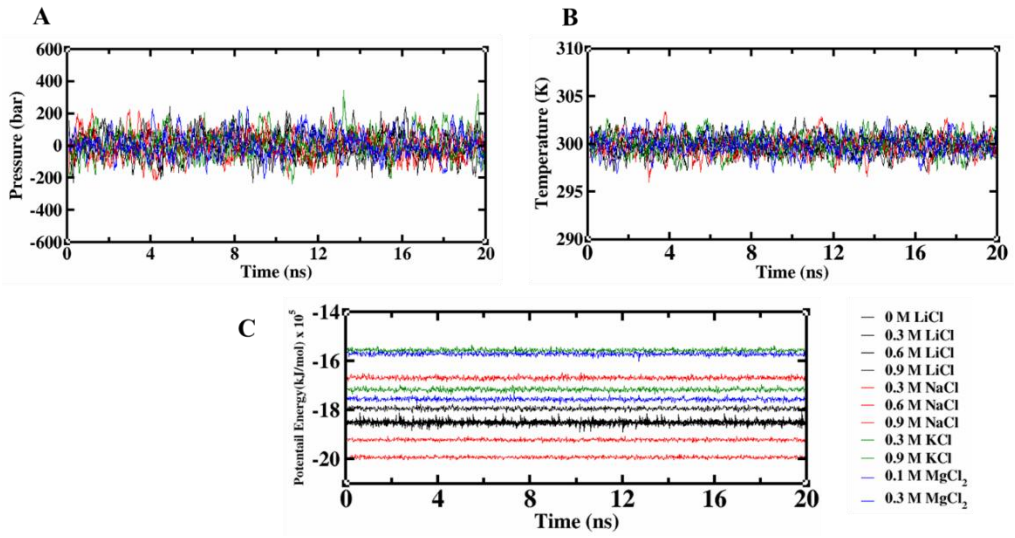


Figure 4.2: Thermodynamic equilibrium graph (A) Time evolution of pressure (B) Temperature and (C) Potential energy for 20 ns in all the cases.

4.3 CONFORMATIONAL TRANSITION OF PRION PEPTIDE IN DIFFERENT SALT SOLUTIONS

To understand the dependency of ionic concentrations and ion specificity on the prion peptide folding/misfolding mechanism, the changes in the conformation of the prion peptide and the amino-acid interactions responsible for such changes are studied during simulation. Although there is a difference in the time evolution of the secondary structure in three different replicas for each salt concentration, the final structures are found to be similar. The first set of data is discussed here and the rest of the additional data is given in Appendix.

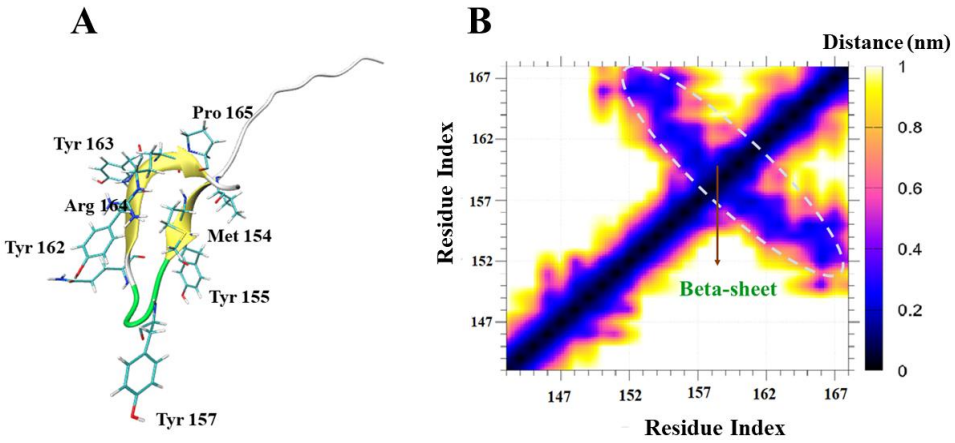
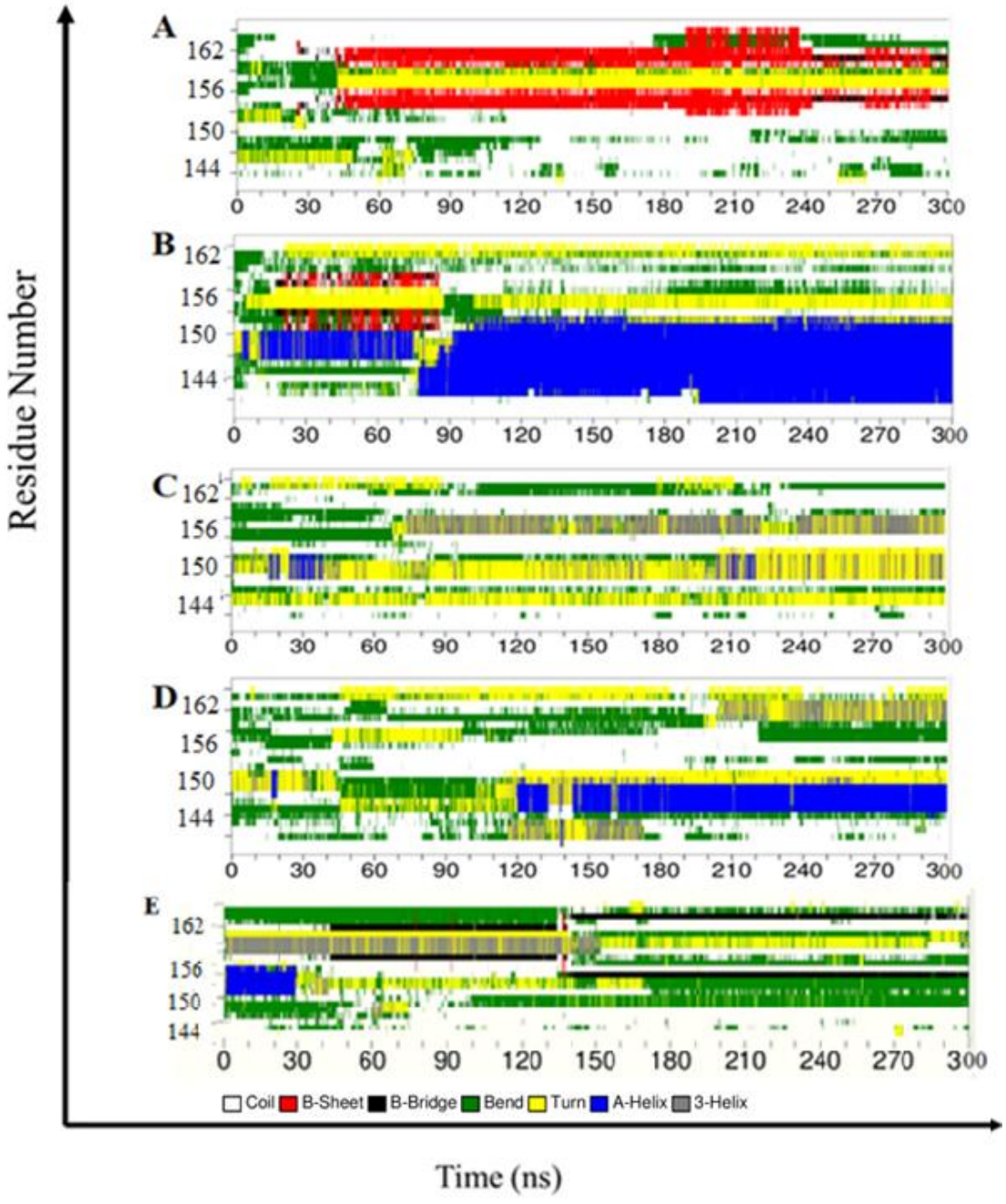
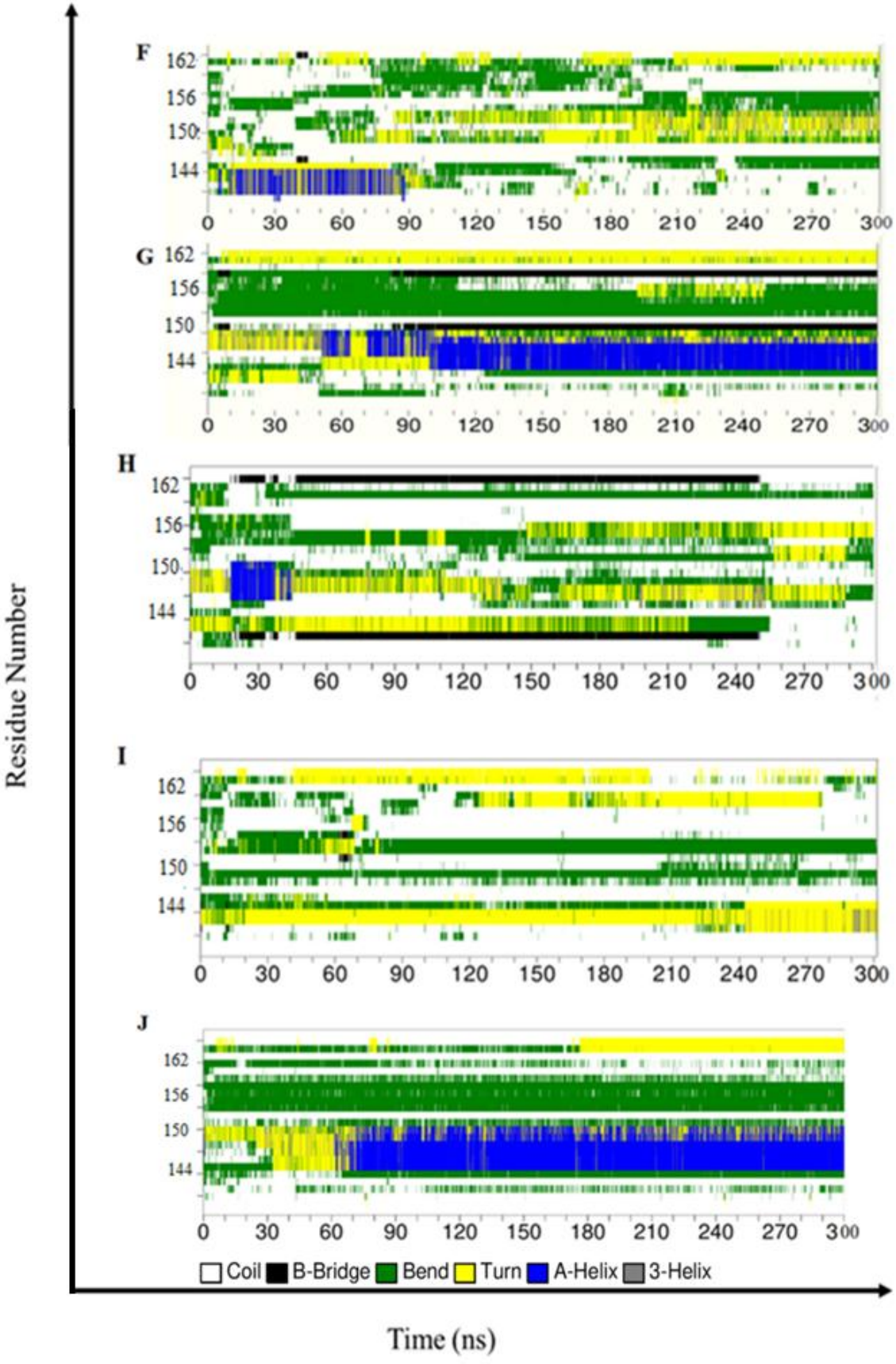


Figure 4.3: (A) The secondary structural arrangement and (B) residue-wise contact map of the sheep prion peptide 142-167 in pure aqueous solution.

Figure 4.3 shows the maximum populated secondary structures of the peptide and contact map in pure aqueous solution. The occupancy of the intra-peptide hydrogen bond during the course of the simulation is provided in **Appendix VIII**. It is defined as the fraction of the time the molecules are hydrogen-bonded in the total simulation trajectory. It can be seen from **Figure 4.3** that in the absence of salts solution, the peptide is found to form a beta-hairpin structure at the C-terminal end of the peptide. The blue color region perpendicular to the diagonal line further validates our observation.




F
G
H
I
J
162
156
150
144
0 30 60 90 120 150 180 210 240 270 300
Residue Number
Time (ns)
Coil
B-Bridge
Bend
Turn
A-Helix
3-Helix

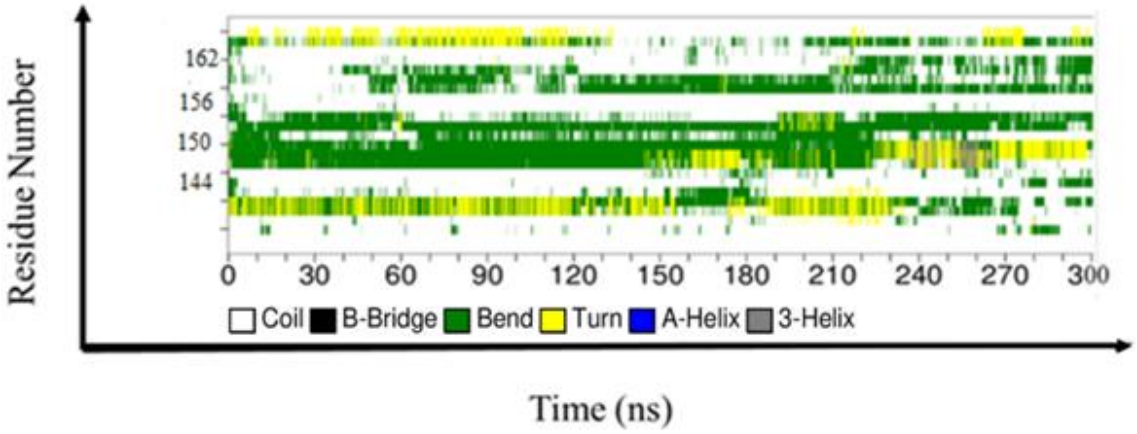


Figure 4.4: Time evaluation of Residue wise secondary structure composition throughout the 300 ns simulation trajectory. (A) 0 M, (B) 0.3 M LiCl, (C) 0.6 M LiCl, (D) 0.9 M LiCl, (E) 0.3 M NaCl, (F) 0.6 M NaCl, (G) 0.9 M NaCl, (H) 0.3 M KCl, (I) 0.9 M KCl, and (J) 0.1 M $MgCl_2$, (K) 0.3 M $MgCl_2$.

The time evolution of the secondary structure element in prion peptide is shown in **Figure 4.4**. It is found that the formation of the beta-sheet initiated from ~30 ns and remains stable up to the end of the simulation. In the other two sets of simulations beta-bridge structure with unstable small α-helix is found which also indicates misfolded conformation of peptides prone to toxicity **(Appendix IX, X)**. This misfolded structure is majorly stabilized by the hydrogen bond between Tyr162-Met154 and Ala156-Gln160 with an occupancy value of 82.6% and 79.6% respectively. In addition, the hydrogen bond between Asn159 and Ala 156 (45%) also stabilizes the hairpin structure during the course of the simulation. The N-terminal end of the peptide is mostly unstructured and remains highly extended without the presence of salt. **Figure 4.5** shows the maximum populated secondary structure in various salt types and concentrations. At 0.3 M LiCl solution, the peptide is found to fold in the N-terminal end which is the opposite direction compared to its folding in pure aqueous solution (**Figure 3-A**). This is further validated by the time evolution of the secondary structure depicted in **Figure 4.4, Appendix IX and X** where the appearance of α-helix/3-10 helix is found. This observation supports the experimental observation of the dual form of the prion peptide at different pH and buffers conditions (Bertho et al. 2008). The conformational evolution of the peptide structure in presence of 0.3 M LiCl is shown in **Figure 4.**

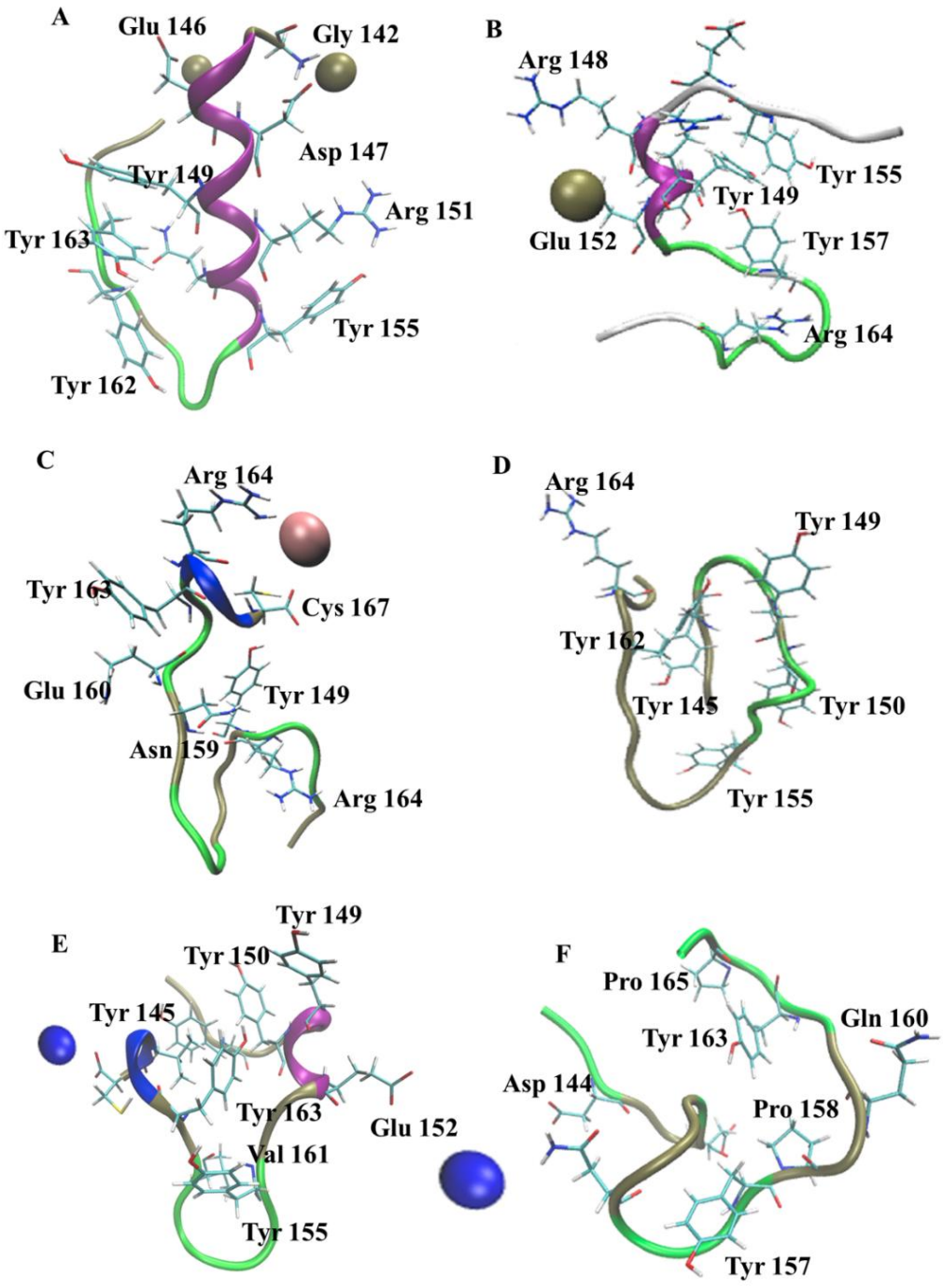


Figure 4.5: The maximum populated secondary structural arrangement of the prion peptide 142-167 correspond to (B) 0.3 M LiCl, (C) 0.9 M LiCl solution, (D) 0.3 M KCl solution, (E) 0.9 M KCl solution, (F) 0.1 M $MgCl_2$ solution.

It can be further noted that the formation of α-helix is initiated from the beginning of the simulation due to the electrostatic interaction between the side chain of Arg 151, Tyr 150, Tyr 149 and hydrophobic interaction between the side chain of

Tyr 149, Met154. This conformation is further stabilized by the π-stacking interaction between Tyr 150 and Tyr 149 (**Figure 4.6-A, B**). The appearance of the beta-sheet structure is also found to be initiated at the same time (~30 ns) as seen in pure aqueous solution **(Figure 4.6-C, D, Figure 4.4-B)**. However, the beta-sheet is found to be interrupted at ~ 85 ns due to the greater interaction of the Li^+ ion with the N-terminal end (**Figure 4.6-E**). It can be seen from **Figure 4.4-B** that the initially formed alpha-helix is interrupted at 73 ns and then the re-folding is propagated from the N-terminal end to the C-terminal end. This is due to the binding of Li^+ between Asp 144 and Glu 146 which essentially screens the electrostatic repulsion between the two acidic residues and initiates the folding by dragging the acidic side chains at a closer distance. This interaction mainly induces the alpha-helical coiling of the peptide at the N-terminal end (**Figure 4.6-F**) and eventually breaks the beta-sheet hydrogen bonds formed at the C-terminal end.

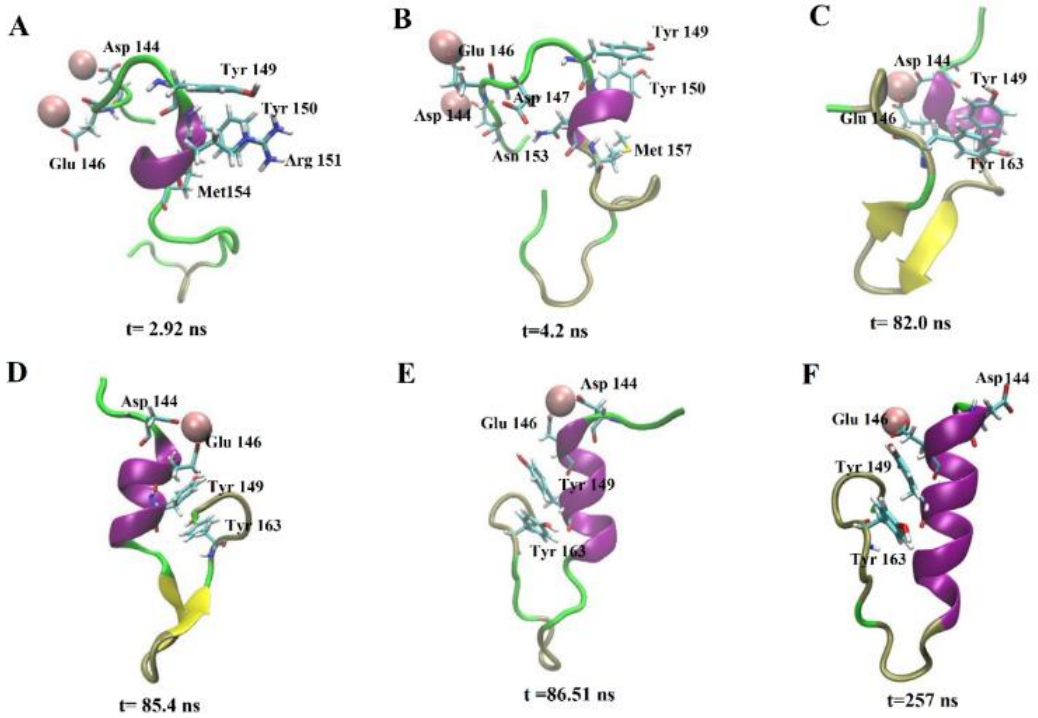


Figure 4.6: The change in the secondary structure arrangement of prion peptide in 0.3 M LiCl at (A) 2.92 ns, (B) 4.2 ns, (C) 82.4 ns, (D) 85.4 ns, (E) 85.4 ns, (F) 257 ns. The Lithium-ion is shown in brown color.

After the formation of a stable alpha helix, the Li^+ ion is found to bind with only Glu 146, and the association of Asp 144, Glu 146 is interrupted. The C-terminal

end is found to be highly unstructured and it formed a loop-like structure that is mostly stabilized by the π-π stacking interaction between Tyr163 and Tyr149. The helical stability of the peptide is attributed to the hydrogen bonding between Met154:Tyr150 and Asn153:Tyr149 residues with a minimum hydrogen bond occupancy of ~78.2 % and 78.1% respectively (**Appendix VIII**). Apart from these hydrogen bonds, additional hydrogen bonds such as Tyr149:Tyr145 (72%), Tyr150:Glu146 (67%), Arg151:Asp147 (64.3%), Glu152:Arg148 (66.4%) at the N-terminal end and Asn159:Ala156 (60.4%), Tyr163:Asn153 (66.0%) at C-terminal end of the peptide are also found. In **Figure 4.7**, we have plotted the pair-wise residual contact maps of the peptides in various salt solutions which highlight the differences in the region-wise contact formed due to the change in the saline environment. The cutoff distance of 0.45 nm is chosen to capture the short-range non-covalent intra-peptide interactions including van der Waals, and hydrogen bonding. The anti-diagonal blue-colored region shows the residual contacts. It is clear from the contact map that the type of salt solution and the salt concentration has a significant effect on the residue-residue interaction profile of the prion peptide. The blue region parallel to the diagonal line is known to be the signature of the helical conformation of the peptide at the N-terminal end residue 144 to 157. This is observed for 0.3 M LiCl, 0.9 M LiCl, and also slightly in the case of 0.1 M $MgCl_2$. The off-diagonal blue-colored region further shows the interaction between the intra-peptide side chains. At a higher LiCl concentration (0.9M), a small fraction of the peptide (Tyr150-Tyr155) is found to be folded in an alpha-helical state at the N-terminal end. The lesser folding in the presence of 0.9 M LiCl concentration is supported by the secondary structure evolution shown in **Figure 4.4-D**. In the other two simulation replicas at the same concentration (0.9 M LiCl) transient β-sheet/β-bridge structure (**Appendix IX-D**) and random coil structure were found (**Appendix X-D**). This result indicates that the refolding of prion peptides in cellular conformation is disturbed in high LiCl concentrations. The intra-peptide hydrogen bond seen in 0.9 M LiCl solution, has comparatively lower hydrogen bond occupancy which indicates the instability of the hydrogen bond to maintain the helical structure of the peptide.

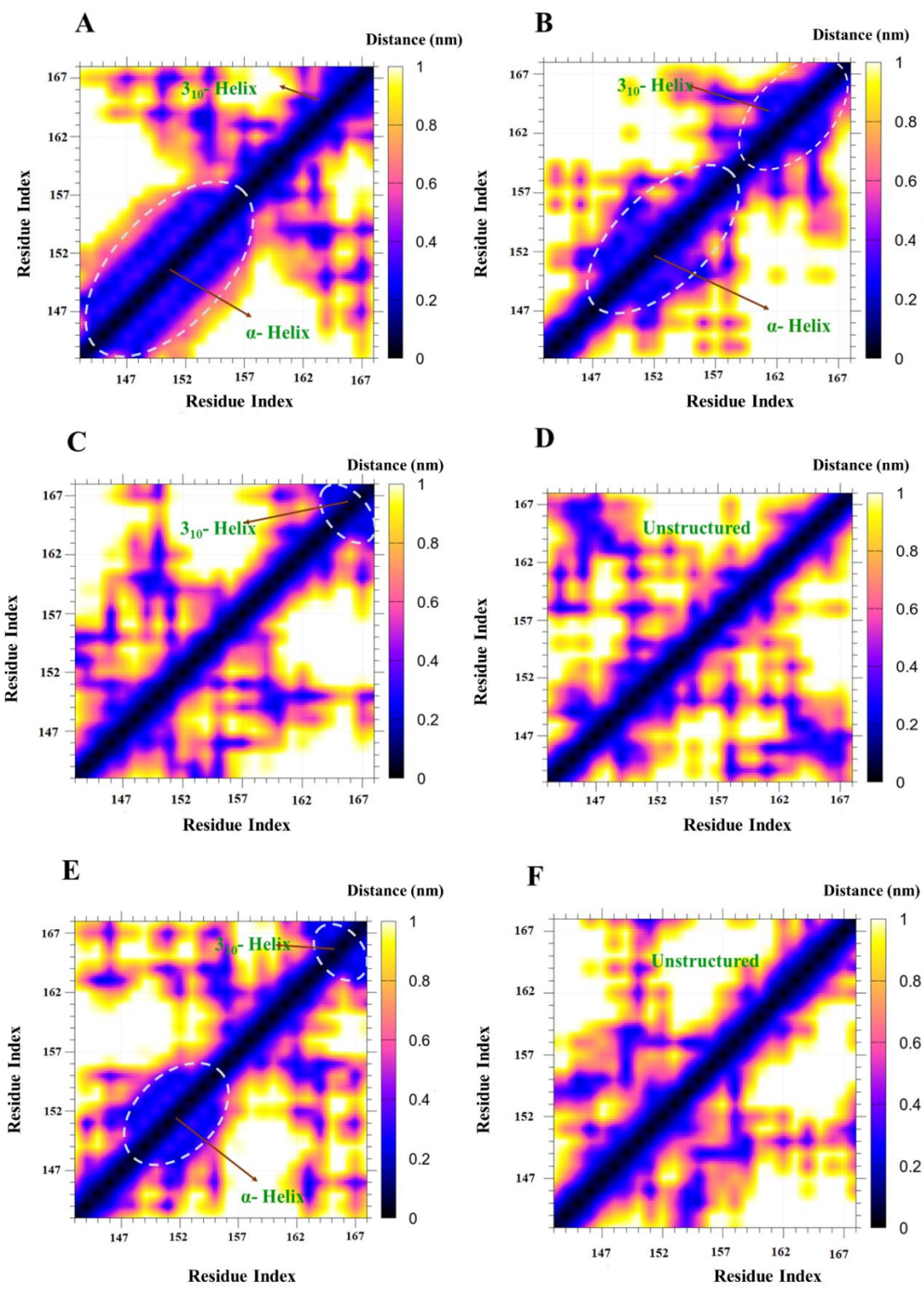


Figure 4.7: Visualization of residue-residue contact map (A) 0.3 M LiCl solution, (B) 0.9 M LiCl solution, (C) 0.3 M KCl solution, (D) 0.9 M KCl solution and (E) 0.1 M $MgCl_2$ (F) 0.3 M $MgCl_2$.

It is found from **Figure 4.5-A** that Li^+ ions (0.3 M concentration) are mostly interacting with the N-terminal segment of the peptide, especially with Glu152. These results indicate the specificity of Li^+ concentration that shifts the dynamic equilibrium of the peptide from the misfolded to the folded state. In contrast, it is evident from **Figure 4.8** that the peptide segment (142-167) is mostly unstructured or polymorphic in the presence of Na^+ ions at 0.3 M (**Figure 4.8-A**) and 0.9 M (**Figure 4.8-B**) concentration. Nevertheless, a small fraction of alpha-helix, 3_{10}-helix, and β-bridge conformation is found to be formed at the various location of the peptide in all the simulation replicas that contain Na^+ (**Figure 4.4-E-G, Appendix IX, X-E-G**). It can be found that Na^+ ions mostly interact with acidic residues such as aspartic acid or glutamic acid like Li^+ ion.

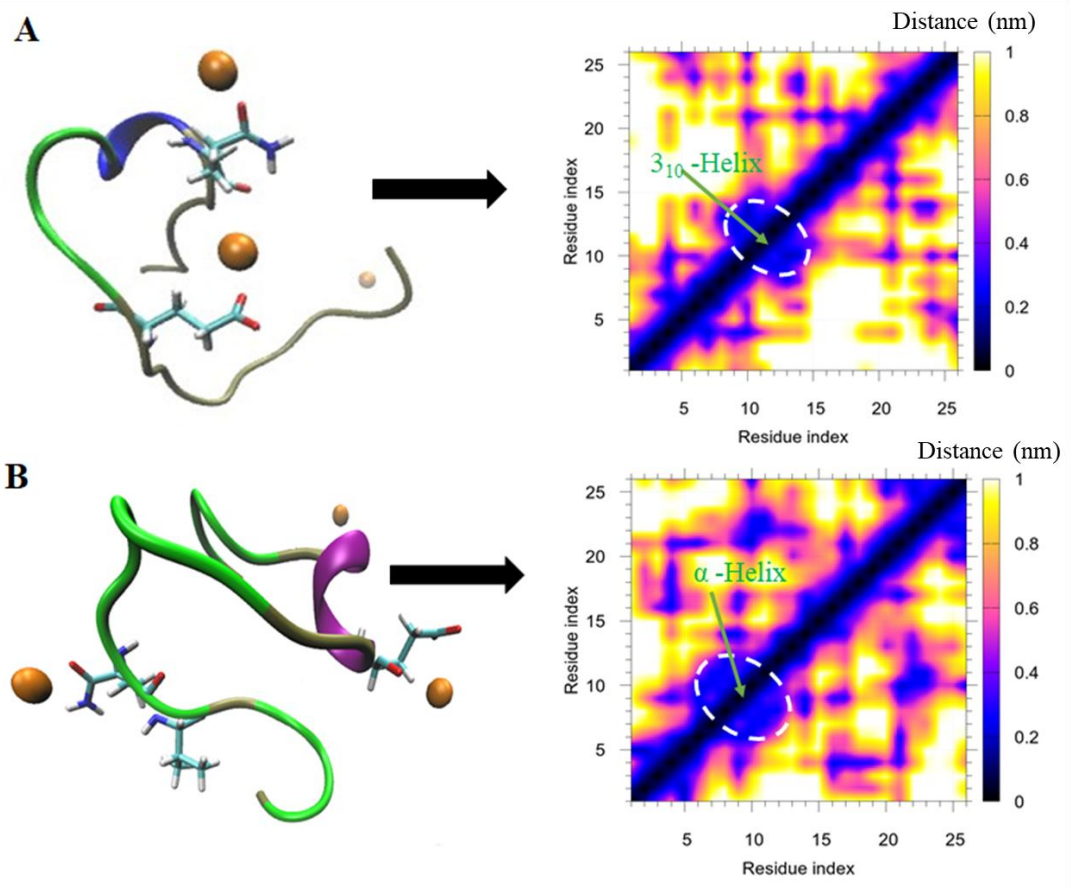


Figure 4.8: The secondary structural arrangement and visualization of the contact map of the prion peptide 142-166 correspond to the global minima at (A) 0.3 M NaCl, and (B) 0.9 M NaCl solution.

However, the interaction of Na^+ ion is comparatively transient which is not efficient in screening the charge repulsion necessary for the alpha-helical folding. In the case of a comparatively bigger cation such as K^+, the peptide is found to form hairpin-like loops and beta bridges favorable to forming extended beta-sheet i.e., the misfolded state (**Figure 4.4-H, I, Appendix IX, X-H, I**). The hydrogen bonds found in the LiCl solution are missing in the case of KCl solutions. At 0.3 M KCl solution, residue pairs Gln160:Pro158 and Arg148:Tyr145 are found to show H-bond occupancy of 62.2% and 60% respectively. It can be seen from **Figure 4.5-D** that the K^+ ions are mostly interacting with the C-terminal Cys167 residue of the peptide. This can be explained due to the higher affinity of the K^+ ion due to its high affinity toward the S atom (Duka and Ahearn 2013). Similarly, at 0.9 M KCl solution, the hydrogen bonds between Val166: Asp144 and Tyr155:Tyr150 stabilize the hair-pin-like structures due to their higher occupancy value (**Appendix VIII**). The contacts are found to be highly disoriented and scattered sparse **(Figures 4.5 C and D)** which indicates the unfolded state of the peptide. The T-shaped π-stacking interaction between Tyr-162 and Tyr-145 stabilizes the compact state of the peptide at 0.9 M KCl concentration. This result indicates that K^+ ions have less propensity to fold the peptide in its natural form. In the presence of Mg^{2+} ions, it is found that the peptide is folded in the alpha-helical state near two terminal ends which is confirmed by the secondary structure profile **(Figure 4.4-J, Appendix IX, X-J)**. The hydrogen bond between Tyr155:Val161 (88.3%) stabilizes the C-terminal helix whereas the N-terminal helix is maintained by the hydrogen bond between Asn153:Tyr149(30.4%), Glu152:Arg148 (63.5%), and Arg151:Asp147 (64.9%). The simultaneous π-stacking interaction between Tyr145, Tyr149, Tyr150, and Tyr162, Tyr163 increases the compactness of the peptide. It can be noted here that the helical structure is quite small compared to what is observed in the Lithium chloride salt case. Further, we calculated the residue-wise coulombic interaction of the ions in all concentrations and depicted in **Figure 4.9** to validate our results discussed above. It can be found that at 0.3 M LiCl concentration the N-terminal residues such as Asp 144, Glu 146, Asp 147, Asp 152, and C-terminal residue Cys 167. In higher LiCl concentration the interaction of Li^+ ion is comparatively less (**Figure 4.9-A**). The interaction of Na^+ and K^+ ions has significantly less interaction with the peptide which causes an imbalance in charge

neutrality of the peptide to restrict the polymorphic nature. It can be noted here that Na^+ has interacted with a greater number of residues where the K^+ ion interacts selectively with the peptide residues (**Figure 4.9-B, C**). Furthermore, the Mg^{2+} ion is found to have intermediate interaction as seen in **Figure 4.9-D**.

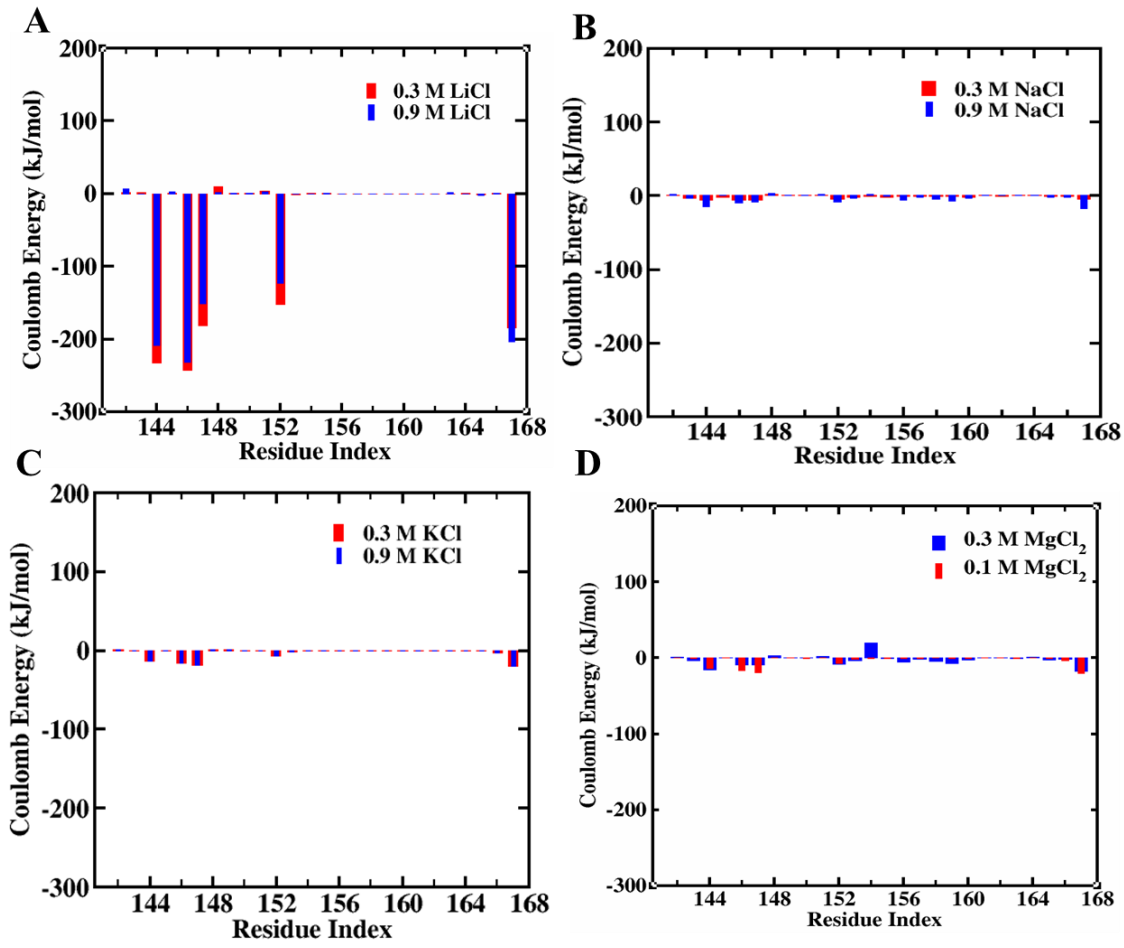


Figure 4.9: Bar plot of the Coulombic interaction energy (A) Li^+-amino acid residue (B) Na^+- amino acid residue (C) K^+- amino acid residue (D) Mg^{2+}-amino acid residue in different salt solutions.

The compactness of the structure can be observed in the probability distribution of the RMSD and R_g of the prion peptide is plotted in **Figure 4.10**. The time evolution and the convergence of the above-mentioned properties for all the simulation sets are shown in **Appendix XI**. In the absence of salts, the distribution of peptide-backbone RMSD and R_g are found to cover a broad region and is more populated towards the higher values (**Figure 4.10- A, B**).

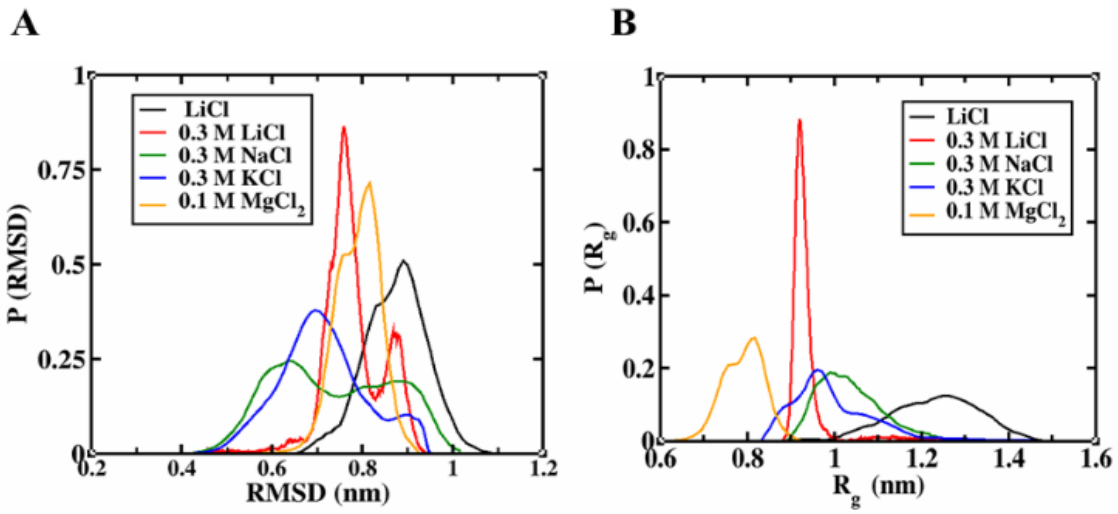


Figure 4.10: Probability distributions of structural parameters depiction on salt concentrations dependence: (A) backbone-RMSD (B) radius of gyration.

The broader distribution of RMSD and R_g indicates a large collection of the peptide conformations or the polymorphic nature which is more prone to toxicity in the absence of an ionic environment. In the presence of LiCl, the peptide is found to deviate increasingly from the initial unstructured conformation resulting in two peaks due to the $\beta \leftrightarrows \alpha$ transition in the peptide conformation (**Figure 4.10-A**). The shift in the probability distribution of peptide-backbone R_g values at lower values indicates the folding of the prion peptide in a very compact state. The most compact state of the prion peptide is found in 0.1 M $MgCl_2$ concentration **Figure 4.11 (C, D)** and the most stable structure is formed at 0.3 m LiCl (narrow peak). At higher LiCl concentration the compactness of the peptide is comparatively lesser as shown in **Figure 4.11**. This is due to the lower extent of secondary structure modification in higher LiCl concentrations as discussed in the earlier sections. The presence of KCl is found to have a minor effect on the conformational transition from the unstructured state resulting in a higher population of the peptide-backbone RMSD towards the lower value which signifies the lesser evolution from the unstructured state. However, the broad probability distribution of R_g indicates a less compact state in the presence of KCl. The effect of salt on the protein structure can be better viewed by the Free energy landscape (FEL) plots which we will be discussing in the next section.

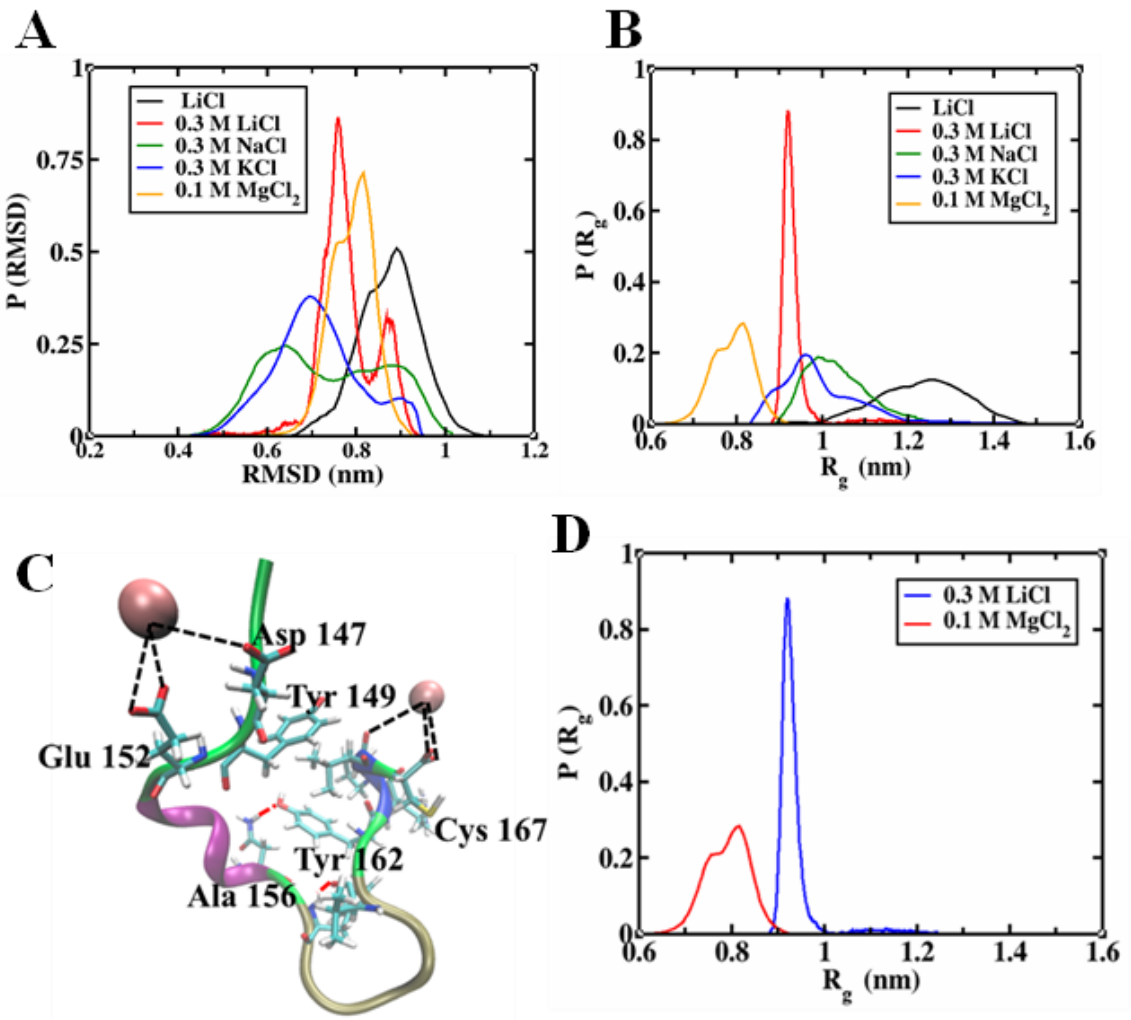


Figure 4.11: Probability distributions of structural parameters depiction on salt concentrations dependence: (A) Backbone-RMSD and (B) Radius of gyration (C) secondary structure arrangement of 0.1 M $MgCl_2$ (D) Distribution of R_g.

4.4 FREE ENERGY LANDSCAPE (FEL)

The potential energy surface or free energy landscape of intrinsically disordered proteins is believed to be highly multi-dimensional. Due to the lack of reference structural conformation, it is notoriously difficult to construct the energy landscape of IDPs in terms of reaction coordinates such as RMSD or R_g. Therefore, it is necessary to take a different approach to analyze the complex multidimensional data generated from the MD simulation of IDPs. Recent reports on the dynamics of IDPs suggest that the principal component analysis (Venugopal et al. 2021) (PCA) or multidimensional

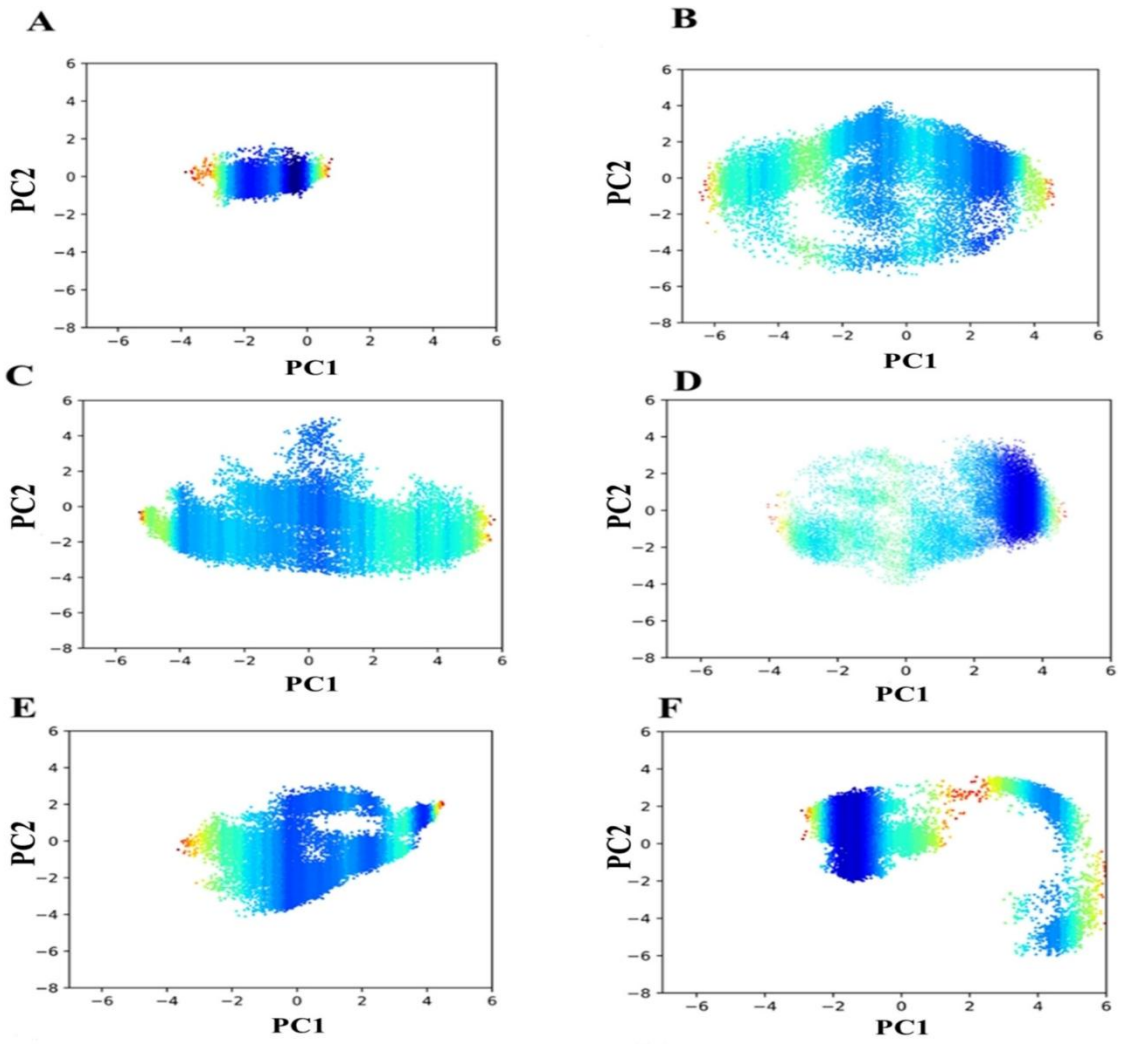
scaling (MDS) (Sanches et al. 2022) method is comparatively efficient to explain the statistical sampling of IDPs in the phase space. First two principal components i.e., PC1 and PC2 were considered as the reaction coordinates to generate the free energy surface of the intrinsically disordered prion peptide in the presence of various salt solutions (**Figure 4.12**). The convergence of the two RCs is measured in terms of cosine content (Khan et al. 2017) and is given in **Table 4.2**. It is believed that the cosine content below 0.5 shows the convergence of the trajectories. It can be found from Table S2 that values are below 0.5 indicating the convergence of the trajectories. The last 200 ns of the simulation trajectory were considered for the calculations. The principal components are found in the covariance matrix. The free energy data can be obtained using the Boltzmann inversion of the 2D-probability distribution of chosen reaction coordinates defined as $\Delta G_i = -k_B T[lnP_i - lnP_{max}]$, where k_B is the Boltzmann constant, T is the temperature and P_{max} is the maximum probability.

Table 4.2: The cosine value of PCs

System	Cosine of PC1	Cosine of PC2
0.3 M LiCl	0.274	0.00048
0.6 M LiCl	0.281	0.0043
0.9 M LiCl	0.361	0.0032
0.3 M NaCl	0.412	0.0081
0.9 M NaCl	0.481	0.0612
0.3 M KCl	0.341	0.0014
0.9 M KCl	0.432	0.0175
0.1 M $MgCl_2$	0.237	0.0023
0.3 M $MgCl_2$	0.462	0.0062

It can be seen from **Figure 4.12** that the projections or sampling points of the prion peptide are widely spaced in the most saline environment except for some specific salt concentration. This indicates the role of specific ionic strength to restrict the conformational sampling of the IDPs. In a purely aqueous environment (**Figure 4.12-I**) the projection is found to span from -6 to 6 along the PC1. The lower free

energy density has corresponded to the structure is found to be the extended beta-sheet conformation. Moreover, the free energy value is dynamic which indicates a more spread-out energy surface where the peptide conformations can float the entire phase space obtained from our MD simulation. In contrast, at 0.3 M LiCl solution, the sampling projection is found to be restricted mostly to the negative value of PC1 and PC2 (**Figure 4.12-A**). This observation is according to the funnel-like energy surface as seen in globular proteins and the structure corresponding to the lower free energy value is found to be an alpha helix, as it exists in its cellular conformation. Further, the increment in LiCl concentration (0.9 M) is found to increase the sampling of the peptide significantly which ultimately generates a more spread-out energy surface (**Figure 4.12-B**).



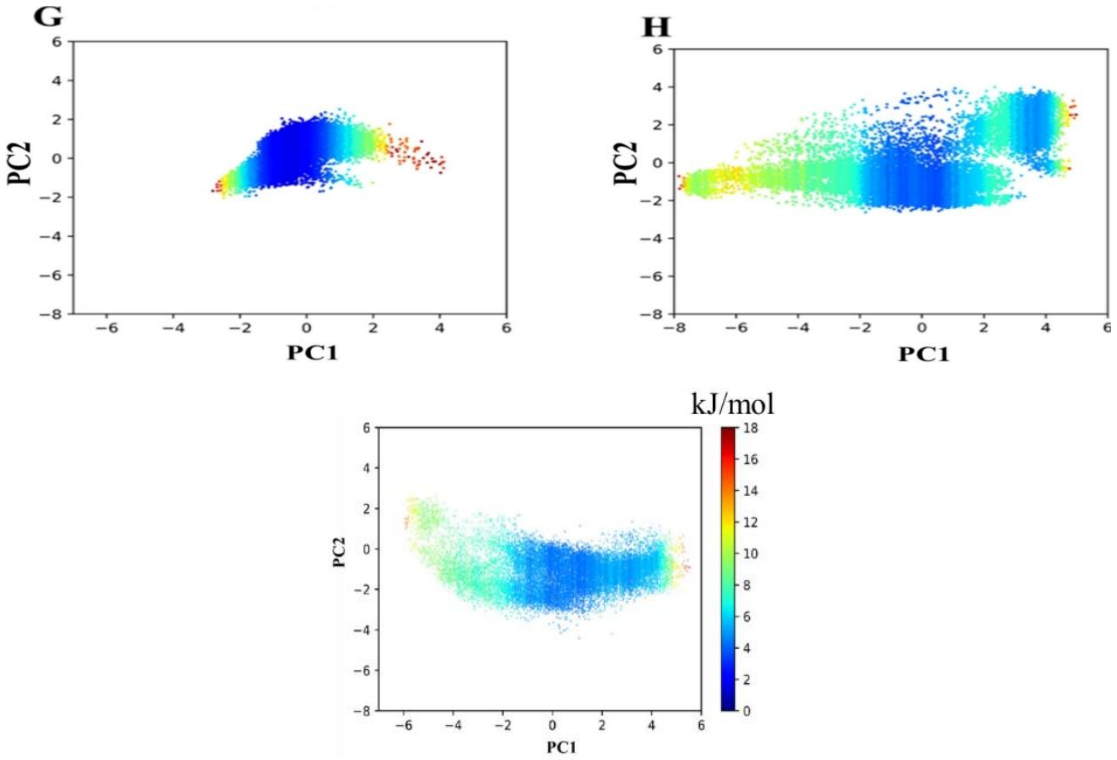


Figure 4.12: The free energy landscape of prion peptide (142-167) at (A) 0.3 M LiCl, (B) 0.9 M LiCl, (C) 0.3 M NaCl, (D) 0.6M NaCl, (E) 0.3 M KCl, (F) 0.9 M KCl, (G) 0.1 M $MgCl_2$, (H) 0.3 M $MgCl_2$ and (I) aqueous solutions.

It is evident from **Figures 4.12-C** and **D** that Na^+ ions are not successful in restricting the polymorphic nature of the peptide. In the case of Na^+ ion, the fraction of helical content is comparatively less as seen in the case of Li^+ ion. The lower free energy region corresponds to the structure shown in **Figure 4.8-A, B**. Similarly in the case of KCl, a border distribution is found in both the concentration, indicating the structural heterogeneity of the peptide that is prone to aggregation (**Figure 4.12-E, F**). Further, we found the restriction of the conformational ensemble in the case of 0.1 M $MgCl_2$ solution. The low free-energy region is indicating a compact helix-loop-helix structure (**Figure 4.12-G**). Next, the increment in Mg^{2+} concentration is found to increase the structural randomness of the peptide as seen in **Figure 4.12-H**.

Thus, it can be summarized that Li^+ ions at some specific concentration have the ability to shift the dynamic equilibrium towards the natural alpha-helical state of the peptide. The higher charge density of Mg^{2+} ions shows a lesser extent of alpha-helical compact structure at a lower concentration. The peptide showed polymorphic

behavior in Na^+ and K^+ ions in a lower concentration. At a higher concentration of all types of ions, the disorderness of the conformation of the peptide increases. To understand the process in further detail, we calculated the effect of ion concentrations on the protein-protein hydrogen bonds and protein-water hydrogen bonds.

4.5 NUMBER OF HYDROGEN BONDS AND TRANSITIONAL ORDER PARAMETER

The effect on the structure of the protein can be better understood if we explore the environment in which the peptide is residing. This can be done if we explore the extent of intrapeptide, water-peptide hydrogen bonding, and heterogeneity of interfacial water molecules in various salt concentrations to assess the stability of the peptide. In **Figure 4.13-A**, we have plotted the number of water-peptide hydrogen bonds (with error bar) and the number of intra-peptide hydrogen bonds for interfacial water molecules with variations in the concentration for the different salt solutions. The water molecules that are present within 5 Å from the carbonyl groups of the protein backbone were considered as interfacial water molecules. It is decided from the RDF of protein-water shown in **Appendix XII-E**. There is not much change in the number of hydrogen bonds except at 0.3 M LiCl and 0.15 M $MgCl_2$. The protein water hydrogen bond is decreased at 0.3 M LiCl salt. The protein is found to be maximumly dehydrated at 0.3 M LiCl when it forms the alpha helix and consequently, the number of intra-peptide hydrogen bonds increased at this concentration (**Figure 4.13-A**). At a higher concentration of LiCl, the number of protein-protein hydrogen bonds and the protein-water hydrogen bond remain constant as there are no major structural changes observed at this concentration. For a similar reason, no major change is observed for KCl at all concentrations. K^+ ion being a bigger cation than Li^+ prefers to live more inside water and interact less with the protein molecule resulting in no substantial change in the protein structure. The behavior of the Na^+ ion is similar to the potassium ion. Mg^{2+} on the other hand, affects the structure of the protein at a much lower concentration, up to 0.15 M. The polyvalent ions induce more pronounced effects than monovalent ions at the same concentration. The structure of the IDP is found to be sensitive toward the type of ion at lower concentrations showing different secondary structures. At higher

concentrations, the structure of the protein is unaltered for the different salt concentrations of similar activity. Therefore, in the case of $MgCl_2$, we see a change in the number of protein-protein and protein-water hydrogen bonds up to 0.15 M.

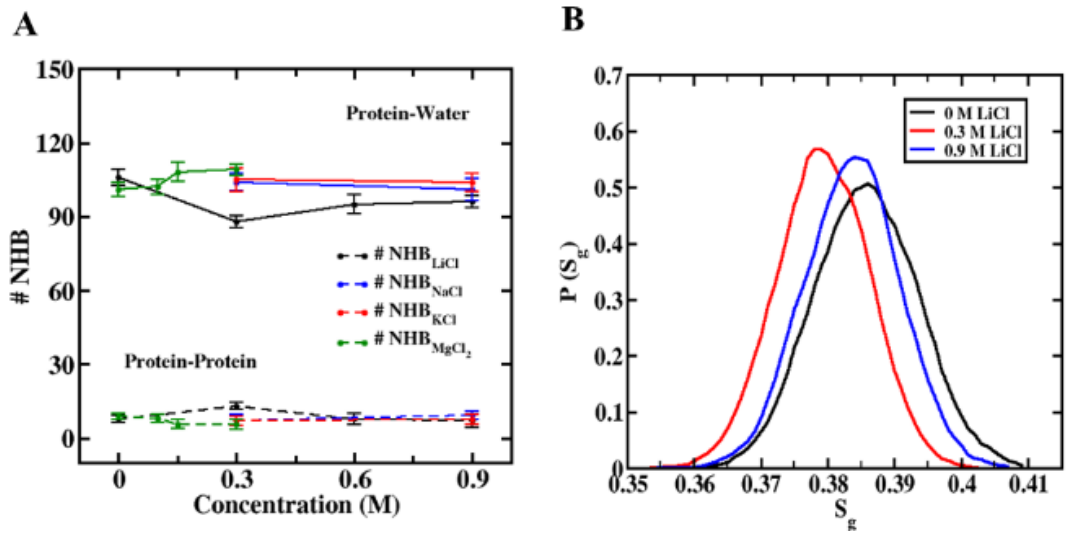
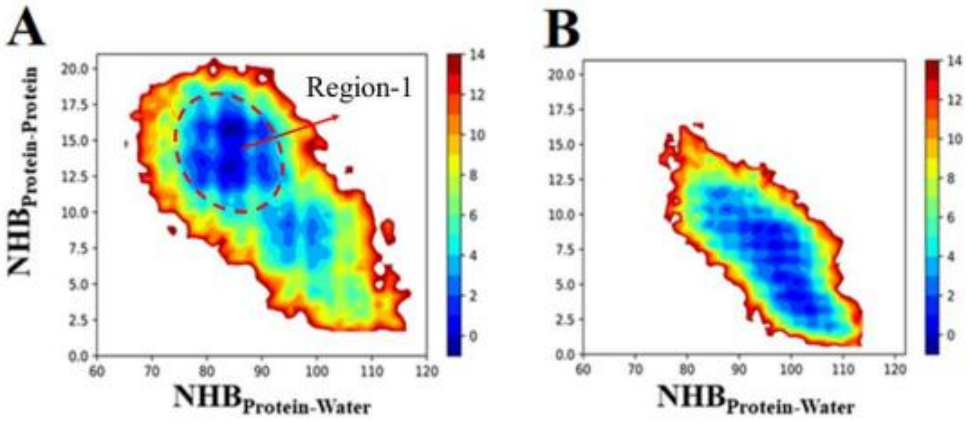


Figure 4.13: (A) Scatter plot number of Hydrogen bond (NHB) between the water-peptide and intra-peptide with salts solution concentrations with error bars. (B) Probability distribution of tetrahedral parameter S_g for different LiCl concentrations.

A similar result is found in the case of the free energy surface plotted between protein-protein and water-protein hydrogen bonds. It is evident that when the structure of the protein is alpha-helix the number of protein-protein hydrogen bonds increases and the number of protein-water hydrogen bonds decreases **(Figure 4.14-A, B, G, H)**. In the case of Na^+ and K^+, there is not much difference as there is not much change in the secondary structure of the protein.



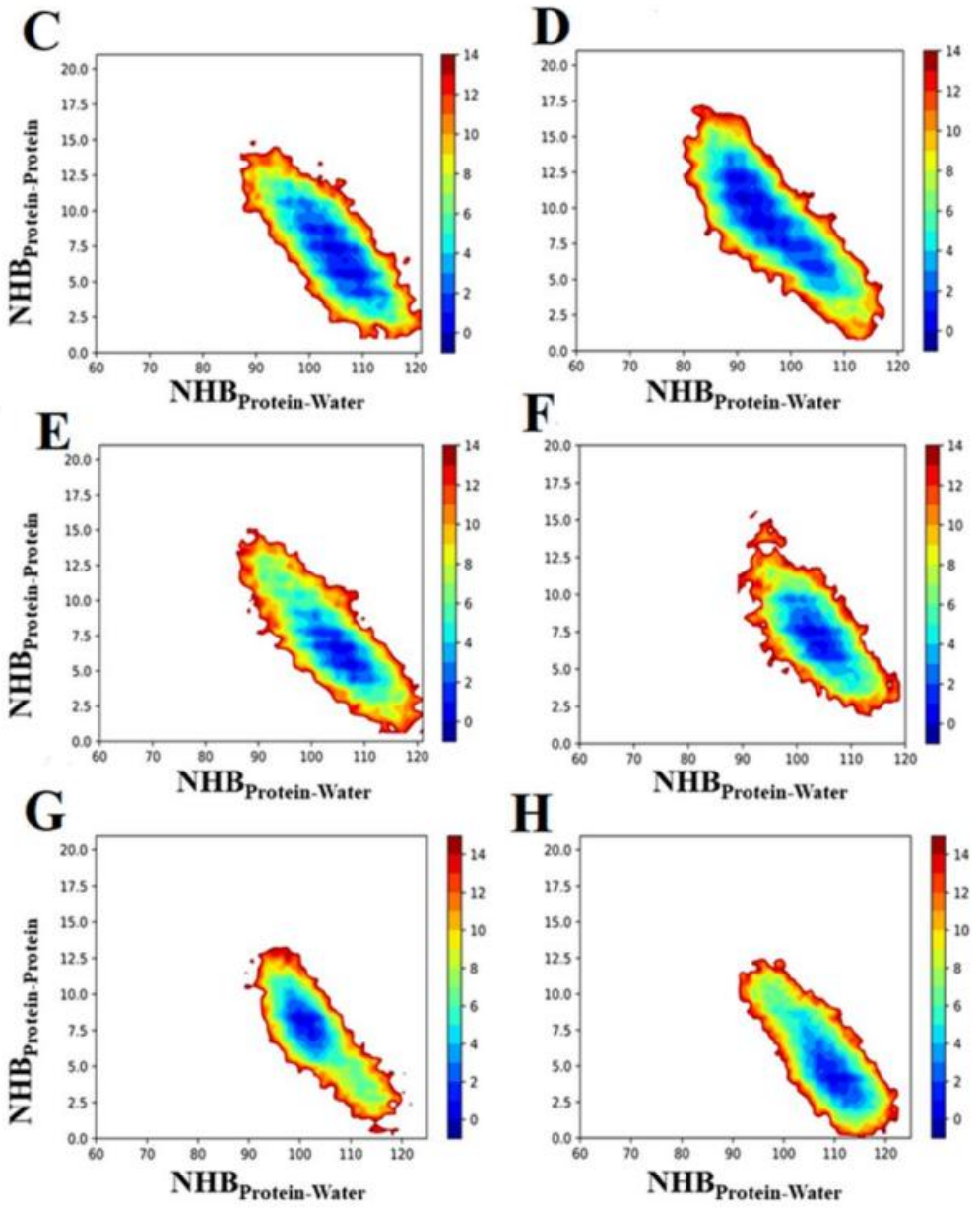


Figure 4.14: The free energy landscape of prion peptide (142-167) at (A) 0.3 M LiCl, (B) 0.9 M LiCl, (C) 0.3 M NaCl, (D) 0.6 M NaCl, (E) 0.3 M KCl, (F) 0.9 M KCl, (G) 0.1 M $MgCl_2$ and (H) 0.3 M $MgCl_2$ solutions.

At 0.3 M LiCl concentration, we found that water around the protein surface is more tetrahedral compared to other cases as shown in **Figure 4.13-B**. The lesser water content on the protein surface makes the water structure more tetrahedral, and less imperfect. The tetrahedrality of the water molecules can be measured in terms of S_g values (Section 2.4.1 (viii)).

It is evident from the above discussion that the size of the ion and the charge density on the ion surface has a major role in the structural modification of intrinsically disordered prion peptide. Thus, it can be summarized that the cations with optimum charge density help in the natural folding of the peptide by increasing the hydrogen bond and stacking interaction between the residues at lower concentrations. Next, it will be interesting to calculate the binding preference of the ions with the protein and the effect of the solvent molecules at these concentrations.

4.6 PREFERENTIAL BINDING OF THE METAL ION WITH PROTEIN SURFACE

The preferential binding coefficient is known to be an important quantity to measure the specificity of the ionic interaction with the peptide molecule. As it is already observed that lithium ions show a different effect on the prion peptide structure and that too at lower concentrations in comparison to the other three ion types (Dilip.H.N. and Chakraborty 2020; Singh and Chakraborty 2021). Li^+ ions have different affinities on the protein surfaces which is due to their smaller size and unit positive charge. To quantify this effect, we calculated the Preferential binding coefficient (Newman 1994) between the ion-protein, ion-water, and protein-water. The structure of the protein and solution can be related to the thermodynamic properties of the KB integral (Section 2.4.1 (ix)).

In **Figure 4.15**, we have shown the preferential binding coefficient of ion-protein, protein-water, ion-water, and ion-counter ion at the interface of the protein surface (5 Å from the carbonyl group of the backbone of the protein) with respect to the concentration of the salt solution. Li^+ has a more preferential binding affinity towards the protein surface compared to the other salts. For Li^+, the preferential binding affinity of ion-protein decreases as the concentration increases initially, then it becomes almost constant at higher concentrations **(Figure 4.15-A)**. The preferential binding affinity of the protein-water increases in the same region and ultimately it becomes almost constant at higher concentrations. Therefore, it shows that at 0.3 M, when Li^+ ion forms alpha-helix with the prion peptide, the protein-ion interaction is maximum and protein-water interaction is minimum. It can be explained that, as the

concentration of the salt increases, the concentration of the counter ion Cl^- also increases, which increases the interactions of the Li^+ ions with the Cl^- ions resulting in to decrease in the ion-protein interactions. As a result, the preferential binding coefficient of the Li^+ towards the Cl^- increases drastically (at 0.9 M) when the ion stops interacting with the protein (**Figure 4.15-D**). This increases the protein-water interactions initially and later becomes constant.

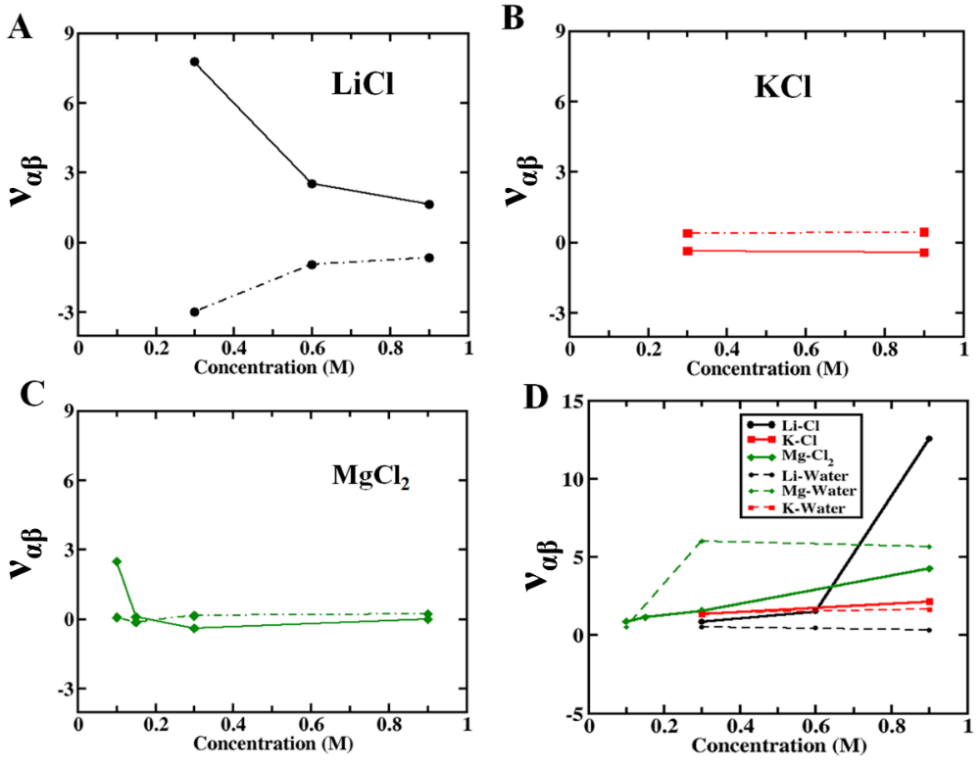
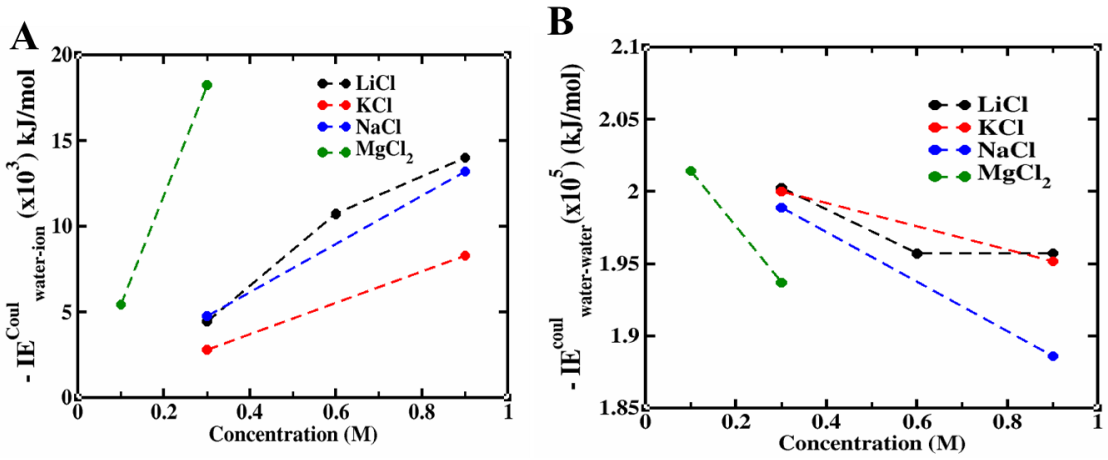


Figure 4.15: The Preferential binding coefficients (PBC) $v_{\alpha\beta}$ within a cutoff of 5.0 Å in (A) LiCl (B) KCl (C) $MgCl_2$ for Ion-protein (Solid line), protein-water (Dash line), and (D) Ion-chlorine and Ion-water in different salts solutions.

For the K^+ ion, all the three preferential binding coefficients are almost constant since it interacts with the protein very little. The K^+ and Cl^- interactions, however, increase with the concentration. It can be noted here that in the case of K^+ ion, the preferential binding coefficient of the protein and water is more compared to K^+ ion-protein which implies that because of its bigger size, K^+ ion is more associated with the water compare to the protein surface in both lower and higher concentrations. It likes to lie outside of the protein surface and so in the presence of K^+ ions no major

change in the secondary structure of the IDP is observed. In the case of divalent cation (Mg^{2+}), more preferential binding affinity with the protein surface is found at a low concentration, at 0.15 M which decreases as the concentration increases. The maximum observed here in the $v_{\alpha\beta}$ is at a lower concentration compared to Li^{+} due to the high charge density of Mg^{2+} ions. One of the major reasons behind less interaction of the Mg^{2+} ions with the protein surface considered in this study with respect to the Li^{+} ions is due to its +2 charge, which is hard to get stabilized with the single valency of the O^{-} atoms on the surface of the carboxylic groups of the prion protein. Moreover, Mg^{+2} is reported to have a high affinity toward water molecules forming $Mg(OH)_6$ complex (Shanker and Bandyopadhyay 2017). If the charge on an IDP protein surface is more and can accommodate Mg^{2+} ions by the adjustment of the amino acids, the preferentially binding affinity of the Mg^{2+} ions towards the protein can be more, which is observed in the case of bigger biomolecular systems. In **Appendix XII**, the radial distribution function and preferential binding graph for protein-ions, ions-water, and protein-water by taking the center of mass as a function of interatomic distance for systems are shown. Further, the water-ion, water-water, and ion-ion interaction energy were calculated to show the efficiency of the chosen force-field (Gee et al. 2011; Saravi and Panagiotopoulos 2021) to capture the interactions at a high salt-concentration regime as depicted in **Figure 4.16**. At higher salt concentrations the increment in water-ion and ion-ion interaction restricts their interaction with a protein molecules. Next, we calculated the survival probability of the ions in the next section.



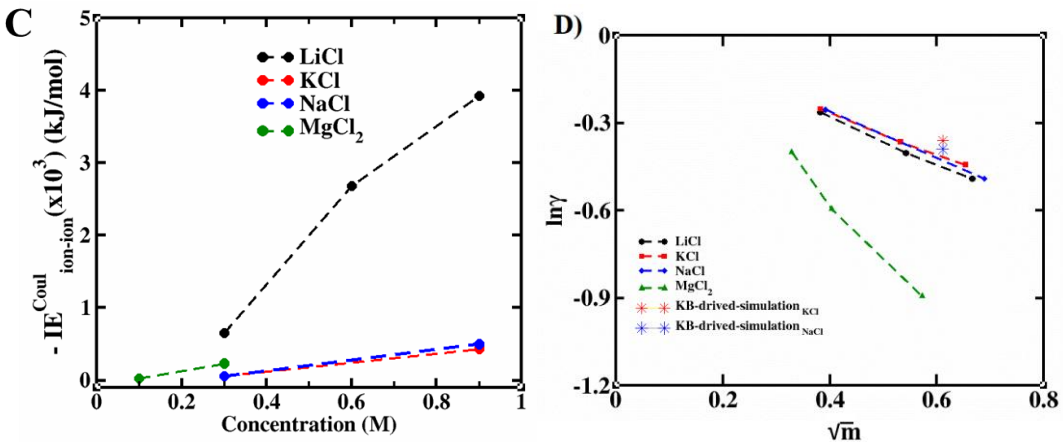


Figure 4.16: Graphs showing the Coulombic interaction energy of (A) Ion-water (B) Water-Water (C) Ion-Ion in different peptide salt solutions and (D) activity coefficients of the salts at various concentrations.

We have also calculated the activity coefficients of the salts as suggested and plotted it against the concentration of the salt solution. **Figure 4.16 (D)** contains the ionic activity and concentration of cations in each simulation. It is found that at higher salt concentrations the activity coefficient is decreasing. Our plots agree well qualitatively with the reported values with other simulation results and we have compared two data from the KB derived ion parameter shown in the star point. The other data cannot be compared due to the different concentration ranges chosen. The comparison of the data and the trend of the plots shows that the parameters used for the ions are reasonable.

4.7 SURVIVAL PROBABILITY AND HYDROGEN BOND LIFETIME

Survival probability (SP) is an important quantity for characterizing the relaxation dynamics of metal cations on the surface of the protein (Section 2.4.2 (iii)). It gives information about the probability of an ion staying on the surface of a protein molecule or inside the protein pocket (Debnath et al. 2010). The calculated beta values are given in **Table 4.3**.

Table 4.3: The relaxation time (τ) and stretching exponential coefficient (β) after fitting by KWW equation given in table.

System	τ (ps)	β
0.3 M LiCl	47.08	0.828
0.9 M LiCl	44.81	0.768
0.3 M NaCl	42.52	0.772
0.9 M NaCl	41.26	0.847
0.3 M KCl	38.82	0.755
0.9 M KCl	41.16	0.817
0.1 M $MgCl_2$	18.62	0.781
0.3 M $MgCl_2$	28.91	0.720

It can be seen that beta values range around 0.7-0.847 for most of the systems which indicates the relaxation process is heterogeneous in nature therefore, we calculated the survival probabilities of the ions both at the interface of the protein and the bulk water. The metal ions within a cut-off of 5.0 Å from the carbonyl group of the protein backbone were considered at the interface (decided from RDF **Appendix XII-A**) and if they are present at > 1 nm away from the protein backbone, they are considered in bulk. The survival probabilities for a long-time window duration of 200 ps were evaluated. The relaxation trend of survival time is given in **Table 4.4**. As already evident from **Table 4.4** the survival probability of the Li^+ ion at the interface of the protein is maximum. It is clear from the survival probability that the Li^+ ions stay 80.98 ps near the protein surface, which stabilizes the alpha-helix structure of the protein. As the ion concentration increases, the lifetime of the ion on the protein surface decreases. At high concentrations, all the ions show a similar survival probability of around 11-7 ps which can be attributed to the high salt effect where the ions no longer interact with the protein surface. Their electrostatic effect is screened by the presence of the solvent molecules and the counterions. The survival probability of the ions at the bulk however increases, with the increase in the concentration of the ions which shows that the interaction between the ions and the bulk water increases as the concentration increases.

Table 4.4: The relaxation trend of survival time of cations during the simulation with standard error.

System	Interface (ps)	Bulk (ps)
0.3 M LiCl	80.37 ± 0.52	42.78 ± 0.60
0.6 M LiCl	22.36 ± 0.22	40.87 ± 0.30
0.9 M LiCl	9.08 ± 0.19	43.32 ± 0.51
0.3 M NaCl	11.75 ± 0.40	32.98 ± 0.08
0.9 M NaCl	10.09 ± 0.23	38.35 ± 0.41
0.3 M KCl	14.08 ± 0.18	26.41 ± 0.72
0.9 M KCl	7.50 ± 0.53	35.76 ± 0.48
0.1 M $MgCl_2$	13.42 ± 0.64	6.795 ± 0.33
0.3 M $MgCl_2$	7.02 ± 0.77	17.82 ± 0.28

To find out the solvation structure of the water molecules near the protein, the hydrogen bond lifetime of the water molecules near the protein surface and the bulk water were calculated (Section 2.4.2 (ii)). In **Table 4.5** we calculated the hydrogen bond between protein nitrogen and oxygen of water, protein oxygen and hydrogen of water, and the water-water hydrogen bond at the interfacial water around the protein. We have considered interfacial molecules up to a cut-off value of 0.5 nm for the calculation of protein-water hydrogen bond lifetime.

Table 4.5: The lifetime (τ_{HB}) of continuous hydrogen bonds (ps) formed between protein-water and water-water in the presence of all salt solutions including the standard errors.

System	Lifetime of hydrogen bond between donor of protein and acceptor of water (ps)	Lifetime of hydrogen bond between acceptor of protein and donor of water (ps)	Lifetime of hydrogen bond between Ow-Hw1/Hw2 (ps)
0 M	0.973 ± 0.013	1.145 ±0.009	0.860 ±0.006
0.3 M LiCl	1.223 ± 0.018	0.934 ± 0.001	0.866 ± 0.002

0.6 M LiCl	1.017 ± 0.004	0.921 ± 0.012	0.874 ± 0.001
0.9 M LiCl	0.957 ± 0.004	0.915 ± 0.030	0.877 ± 0.001
0.3 M NaCl	1.059 ± 0.014	1.053 ± 0.004	0.868 ± 0.003
0.9 M NaCl	1.002 ± 0.018	1.036 ± 0.027	0.872 ± 0.001
0.3 M KCl	1.047 ± 0.012	1.040 ± 0.012	0.865 ± 0.007
0.9 M KCl	0.912 ±0.011	0.938 ± 0.010	0.873 ± 0.006
0.1 M $MgCl_2$	0.957 ± 0.027	1.056 ± 0.015	0.877 ± 0.005
0.3 M $MgCl_2$	0.882 ± 0.003	1.003 ± 0.021	0.884 ± 0.007

In the interfacial region, we have more error bars due to the presence of a lesser number of molecules. It can be found that the protein water hydrogen bond lifetime decreases with the increment of the ionic concentration whereas the water-water hydrogen bond lifetime increases with the concentration of the ions. It is found that the lifetime of protein water hydrogen bond is highest in the case of 0.3 M LiCl solution. Due to the higher survival probability of the ions at the protein surface, the hydration water carried by ions makes a stronger hydrogen bond with the peptide which increases the lifetime of this non-covalent interaction. This also indicates a well-defined hydration layer around the protein required for a stable folded state. In contrast, the lesser lifetime of protein water hydrogen bond indicates a less stable hydration layer around the protein in the presence of Na^+, K^+ ions. At higher salt concentrations the decrease of the protein-water hydrogen bond indicates the disruption of the hydration layer which ultimately causes the polymorphic protein structure.

4.8 CONCLUSION

Lastly, to conclude the ion-specific differences in the structural changes of the IDP can be linked to the preferential binding affinity of the ions and the solvation of the protein with the water. The IDP is found to be maximumly solvated at 0 M concentration; where it forms the β-sheet structure. At lower ionic concentrations, due to the favorable interaction of the Li^+ ion with the protein residues, it prefers to stay at the surface of the protein, thus stabilizing the folding of the protein to the α-helical

state. At this stage, the number of the protein-water hydrogen bond is less which increases the protein-water hydrogen bond lifetime due to the loss of hydrogen bond cooperativity. The strong electrostatic force of interaction of the ions also adds up to increasing the protein-water hydrogen bond lifetime. This lifetime decreases with the increasing ionic concentration as the water molecules get more involved with the ions which increases the water-water hydrogen bond lifetime. As the concentration of the ion increases, the ion-water and ion-counter ion interactions increase that shielding the ion-protein interactions, and hence ion-specific affinity of the IDP protein is lost. Bigger cations like K^+ ions prefer to live inside the water more and hence have a lesser effect on the protein conformation. The residue-wise interaction plots validate this hypothesis. Further, Li^+ ions were found to interact more with Asp 144, Glu 146, Asp 147, Asp 152, Cys 167 residue whereas K^+ was found to interact with Cys 167. With the increase in the concentration of the ions, this interaction decreases. As a result, the survival probability of the ions also decreases. Higher charge density ions like Mg^{2+} find it hard to get stabilized on the surface of the protein for a lower charge protein and hence shows little effect on the conformation at a very lower concentration of the salt. The changes observed here are at lower concentrations. The strength of the specific interactions of ions with the protein surface is found to depend on the concentration of the salts. The free energy profile of the protein shows that in presence of 0.3 M LiCl and 0.1 M $MgCl_2$, the IDP acquires a compact landscape. This study shows the importance of alkali metal ions on the structure of the proteins. The binding of the ions to residues at the protein surface at a physiological salt concentration is important for protein folding and hence, they can be proved to be beneficial for the treatment of neurodegenerative diseases.

CHAPTER 5

INSIGHT INTO DISORDER-TO-ORDER TRANSITION IN SHEEP PRION PEPTIDE

Understanding the factors that govern the conformational ensemble of IDPs and the mechanisms of disorder-to-order transitions gained a great interest in the field of protein science. Here, AMD and REMD simulations were used to obtain insights into different conformational ensembles and the transition from disorder to order state.

5.1 BACKGROUND

Intrinsically disordered proteins or protein regions (IDPs, IDPRs) are the proteins that do not contain singularly equilibrated 3-dimensional (Dyson and Wright 2005). The changes in their conformational structure are known to have significant effects on biological functions such as transcription, cell growth, and gene expression (Wright and Dyson 2015). Unlike globular proteins, IDPs consist of higher charge residues in the sequence which favors disordered conformations due to local or long-range charge repulsion. Hydrophobic amino acids are fewer to stabilize the hydrophobic cores with larger side chains of IDPs/IDPRs. IDPs/IDPRs are capable of undergoing structural transitions to achieve more ordered conformation while binding with other cellular partners. The degree of folding/disorder-to-order conformations depends on the binding moiety such as other proteins or supramolecular assemblies (Uversky 2013). The perturbation in the conformation of IDPs depends on various factors like the alteration of solvent composition, temperature, pH, post-translational modification, and mutations (Uversky 2009). The IDPs and IDPRs are very much accessible for posttranslational modifications, by peptide-peptide interaction within IDPs/IDPRs inside the human proteome (Darling and Uversky 2018). The conformational disorder-to-order transition of IDPs can have important functional consequences, as it allows them to interact with specific partners and carry out their biological roles.

Understanding the factors that govern the conformational ensemble of IDPs and the mechanisms of disorder-to-order transitions gained a great interest in the field

of protein science. The vast conformational space of IDPs votes a great challenge as the structural transitions are triggered by external physical factors and the binding partners. Experimental techniques have limitation in probing some aspects of the conformational space. For example, small-angle X-ray scattering (SAXS) provides information on the overall shapes and sizes of IDPs (Kachala et al. 2015), but not at the atomic level. Similarly, the NMR technique shows the secondary chemical shifts which carry residue-specific information but are still unsuccessful to assign the degrees of freedom of IDPs (Jensen et al. 2014). The experimental technique is still unsuccessful to define the conformational transition from disordered to ordered and vice versa. Molecular dynamics (MD) simulations offer an attractive approach for IDPs, with an atomic representation for each conformation. The large simulation time is able to determine the vastness of conformation sampling of the IDPs but the higher computational cost of such long trajectories restricts the simulation in nanoseconds to a few microseconds time scale. Therefore, the computation of IDPs or large peptides are still computationally challenging at physiochemical conditions by MD simulations. The MD simulations are not always consistent with the experiment, either because of insufficient sampling or some drawbacks to the force fields (Bhowmick et al. 2016). The use of biasing potential or reweighting the ensemble of the MD simulation helps to correlate the results with experimental observation (Crehuet et al. 2019; Fisher et al. 2010). IDP undergoes a multi-funnel-free energy landscape which reveals more stable conformations. In atomistic molecular dynamics (AMD) it is difficult to sample all possible states of the conformations which are trapped in a deep well of the free energy space. Therefore, these conformations can be efficiently sampled by enhanced sampling methods such as REMD (Replica exchange molecular dynamic). The configurational states sampled by using replica-exchange molecular dynamics (REMD) have been used as a useful tool for the assessment of protein folding (Espinoza-Fonseca 2009). An important outcome of REMD simulations is that it allows extended sampling of the free energy space to obtain the IDP conformational changes extensively. It helps the peptides to overcome the free energy barrier. However, Molecular dynamics (MD) simulations generate continuous, atomic-level trajectory data for studying the structural and dynamic properties of disordered regions of the protein (Castro et al. 2019). The study aims to gain insights

on the conformational evolution of sheep prion with the change in temperature and to find the driving force behind these structural changes.

In this chapter, the temperature-induced structural changes in prion peptide IDP were examined using individual molecular dynamic simulation and also validated it with REMD calculation. The structures of the peptide sampled by MD simulation showed inter and intra-molecular stabilizing forces in conformations with higher mobility of the residues. The conformational changes have been investigated mainly in terms of the changes in secondary structure, contact map analysis, surface exposure, hydrogen bonds, and intermolecular interactions. Results obtained from the present investigation provide a deeper understanding of residue mobility and factors affecting the conversion of IDP from a disordered to an ordered state and vice versa.

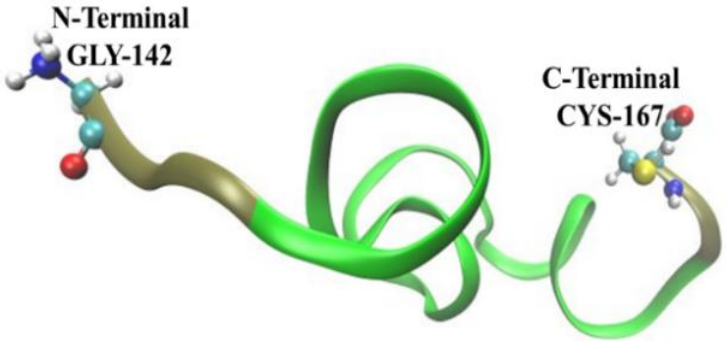


Figure 5.1: 3-D structure of prion peptide segment (PDB ID: 2RMW) showing N- and C- terminal residues.

5.2 SYSTEM SETUP

The initial coordinates of IDP (prion peptide segment 142-166, PDB ID: 2RMW) (Bertho et al. 2008) were obtained as mentioned in Section 4.2 (**Figure 5.1**). Then, the peptide was solvated by TIP3P (Jorgensen et al. 1983) water molecules in a cubic box of length 4.76 nm. The overall charge of the system was neutralized by adding counter ions. The effect of temperature on the folding-unfolding nature of prion peptides was studied using MD simulations at a temperature range of 290 K- 330 K. The details of the simulated peptide systems are shown in **Table 5.1**. Further, REMD simulations (290-330 K) are performed for 600 ns.

Table 5.1: Details of the simulations N_{IDP} and N_{water} represent the corresponding number of prion peptide units and water molecules in the simulation box.

System /Temp (K)	N_{IDP}	N_{Water}	Simulation Time (ns)
1/ 290	1	3513	1000
2/ 300	1	3513	1000
3/ 308	1	3510	1000
4/ 320	1	3510	1000
5/ 330	1	3510	1000

Atomistic Molecular Dynamics Simulations: All simulations were performed in the GROMACS (Berendsen et al. 1995; Van Der Spoel et al. 2005) software package with CHARMM-36 force field (Gil Pineda et al. 2020; Oliveira and Colherinhas 2020). The leapfrog algorithm with a time step of 2 fs along with periodic boundary conditions was used in all three directions (Allen and Tildesley 2017). The short-range Lennard-Jones interaction and Particle Mesh Ewald (PME) (Darden et al. 1993; Essmann et al. 1995) were employed to treat the long-range electrostatic interactions with nonbonded interaction within a cut-off of 1.2 nm. The LINCS (Hess et al. 1997) algorithm was applied to constrain the hydrogen bond atoms of the peptide molecule and the SETTLE (Miyamoto and Kollman 1992) algorithm was employed to constrain the geometry of water molecules. The systems were equilibrated in the isothermal-isochoric ensemble (NVT) followed by the isothermal-isobaric ensemble (NPT) for 20 ns each respectively. The 1 μs production run has been done for every temperature. All the simulations were performed thrice to check the reproducibility of the results. A total of 15 μs production has been done. The temperature and pressure of the system were maintained employing Velocity rescale (Bussi et al. 2007) (τ_t=0.1ps) and Parrinello-Rahman coupling algorithm (Silvestrelli and Parrinello 1999) (τ_p=0.2 ps) respectively. The time average thermodynamic properties have given in **Appendix XIII**. The VMD (Humphrey et al. 1996) tool was used to visualize the folding process of the peptide at different temperatures.

Replica Exchange Molecular Dynamics (REMD): The peptide systems have been solvated with 3513 water molecules of the TIP3P model in a cubic box. The energy minimization of the system has been carried out using a steepest-descent algorithm before molecular dynamics simulations. REMD is used to enhance conformational sampling by allowing the crossing of energy barriers between the conformations from lower temperatures to higher temperatures. 10 replicas within the temperature range of 290−330 K (**Table 5.2**) were used and attempted an exchange at every 10 ps in order to achieve an average exchange rate of 15%. For all these simulations, the position-restrained NVT equilibration has been run for 5 ns using a V-rescale thermostat. The NVT production has been carried out for 600 ns for each replica. In all the simulations, the particle mesh Ewald method has been used to describe the long-range electrostatic interactions with 0.16 nm grid spacing. Leapfrog integrator with the integration time step of 2 fs has been used to solve Newton's equation of motion and trajectories have been saved at every 10 ps.

Table 5.2: The temperature obtained from the exchange probability for replicas.

S.No.	Temperature (K)
1	290.0
2	294.3
3	298.7
4	303.1
5	307.5
6	312.0
7	316.6
8	320.9
9	325.7
10	330.0

5.3 STRUCTURAL CHANGES AND SECONDARY STRUCTURE ANALYSIS

The secondary structure analysis has been done to understand the temperature dependency of the conformational evaluations of the sheep prion peptide during the

simulations. The residue-residue (amino-acids) and solvent-residues interactions are responsible to change the conformation from disordered to ordered conformation of the IDPs. The time average secondary structure for the intrinsically disordered peptide was calculated by DSSP analysis at different temperatures **Figure 5.2**. The other two sets are given in **Appendix XIV**.

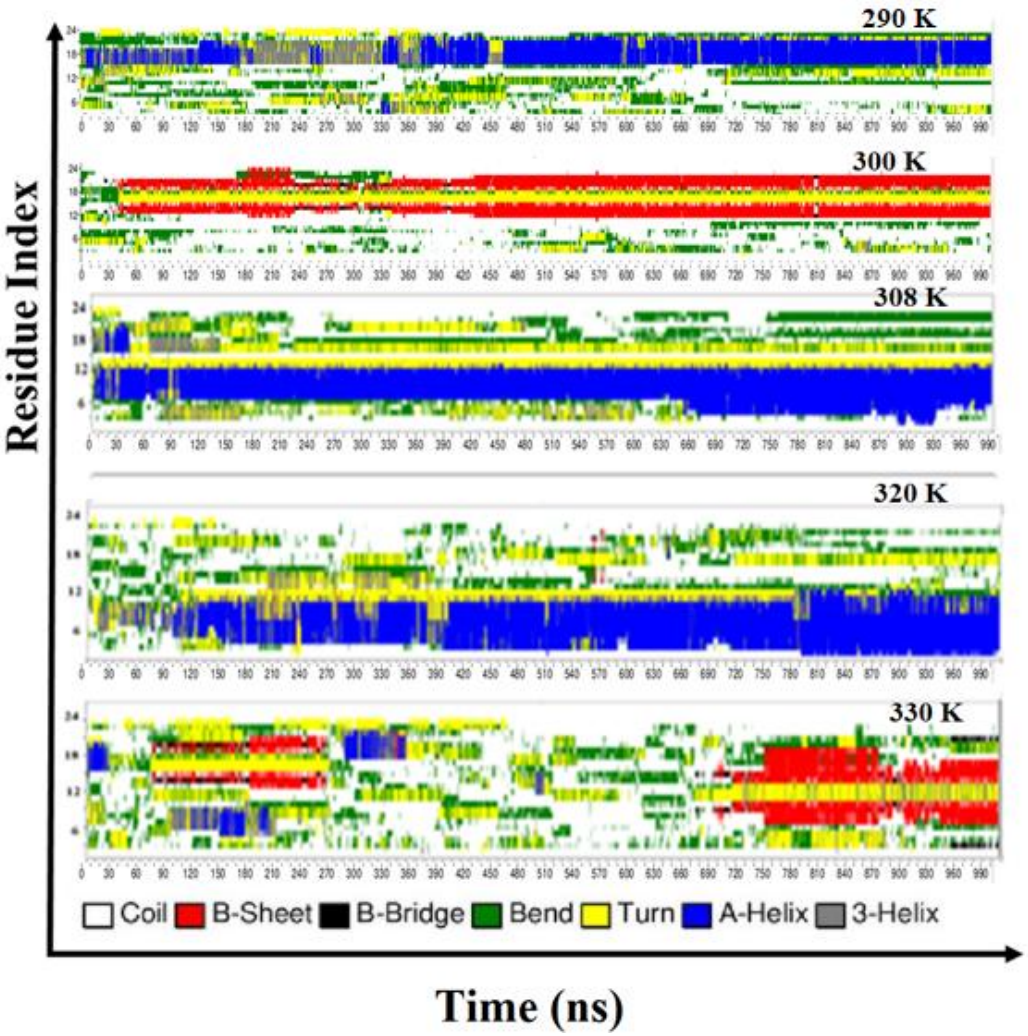


Figure 5.2: Time evaluation of residue-wise secondary structure composition throughout the simulation trajectory at all 5 temperatures.

It can be seen that the secondary structure is dominated by random coil conformation during the initial phase of the simulation. At lower temperatures (290 K), the IDP is found to be in α-helix form towards the C-terminal end whereas the N-terminal is a random coil (**Figure 5.2(a)**), and after 450 ns it was continuously found to be in alpha-helix conformation. At the 300K temperature, the IDP formed β-sheet

(**Figure 5.2(b)**) mainly near the C-terminal, but at 308 K and 320 K α-folded helix formed near the N-terminal residues (**Figure 5.2 (c-d)**). In the case of 330 K, the IDP was again found to be β-sheet and this β-sheet is found to be shifted towards the N-terminal from C-terminal after 700 ns (**Figure 5.2 (e)**).

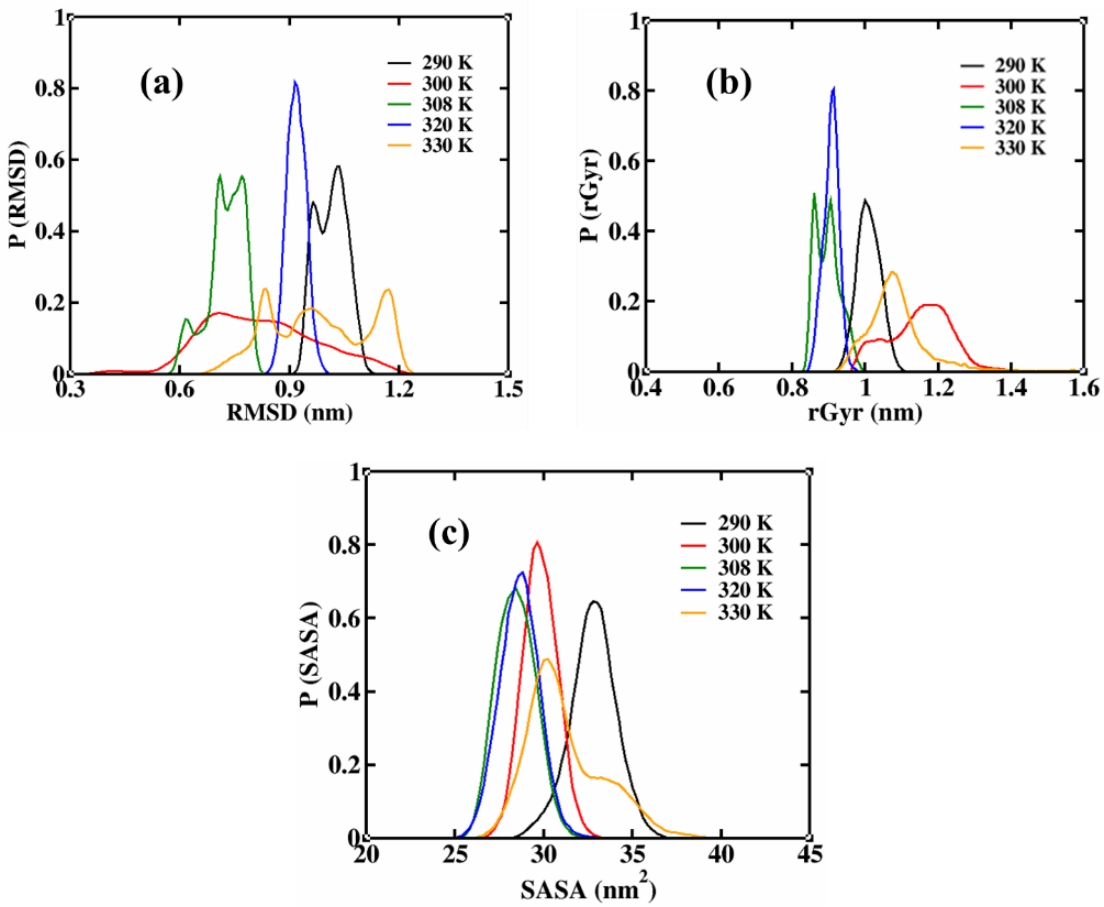


Figure 5.3: The probability distribution of the secondary structural parameters (a) RMSD, (b) R_g, and (c) SASA.

In **Figure 5.3**, the structural parameters are calculated in terms of RMSD, R_g, and SASA. This gives information about the changes in the IDP structure. The IDP changes from the disordered to a particular ordered conformation under different temperatures. The two peaks of RMSD distribution at 290 K are mainly due to the existence of two different conformations. The presence of beta-sheet and α-helix is observed at the same time and later 3_{10}-helix or α-helix after 600 ns (**Figure 5.2(a)**).

At 300 K **Figure 5.3(a)** shows the broad distribution of the RMSD probability due to the formation of Beta-sheet towards the C-terminal side and random motion in

N-terminal side residues. In the case of 308 K and 320 K, the peptide was found to be a compact distribution of RMSD due to formation of the formation of α-helix at the N-terminal side. At 308 K, three peaks are seen due to changes in the α-helix position. The peak at 320 K is compact due to the formation of an extended α-helix structure. At 330 K, the different peaks are showing the formation of more than one conformation. The probability distribution of the radius of gyration shows the compactness of the peptide (**Figure 5.3(b)**). At 290 K, 308 K, and 320 K the peptide is more compact compared due to the formation of Alpha-helix. At 300 K and 330 K, the peptide is forming the beta-sheet and the beta bridge which is less compact compared to α-helix. Moreover, the structural rearrangements are characterized by the decrement in SASA. It shows the total surface accessibility by the water molecules (**Figure 5.3(c)**). At lower temperatures (290 K), the SASA value is more due to the interaction of solvent in the hydrophilic region (143-152) of the IDP. This region makes α-helix at higher temperatures, and therefore, the SASA reduces. It can be seen that the alpha helix structure has fewer SASA values than the B-sheet structure at 300 K and 330 K which is similar to the trend seen in RMSD and R_g value which further confirms that the a-helix structures are more compact.

5.4 REMD SIMULATION AND THERMODYNAMIC ANALYSIS OF CONVERGENCE

The REMD data convergence is checked by investigating the equal occupancy rule of replicas at each temperature shown in **Figure 5.4 (a)**. By REMD simulations we have obtained a probability distribution function of potential energy and structural parameters for a set of temperatures as shown in **Figure 5.4**. In **Figure 5.4(a)**, it can be seen that a significant amount of potential energy overlaps in the coverage of probability distribution between two adjacent temperatures T_i and T_{i+1}. This confirms that the 15% acceptance ratio is suitable for replica exchange. The root-mean-square deviation (RMSD) distribution as a function of temperature obtained from REMD simulation is also shown in **Figure 5.4(b)**. It observed that the average RMSD increases as temperature increases, indicating that the population of the changes in IDP conformations increases. The radius of gyration (R_g) distribution as a function of

temperature obtained from REMD simulation is shown in **Figure 5.4(c).** The average radius of gyration increases as temperature increases.

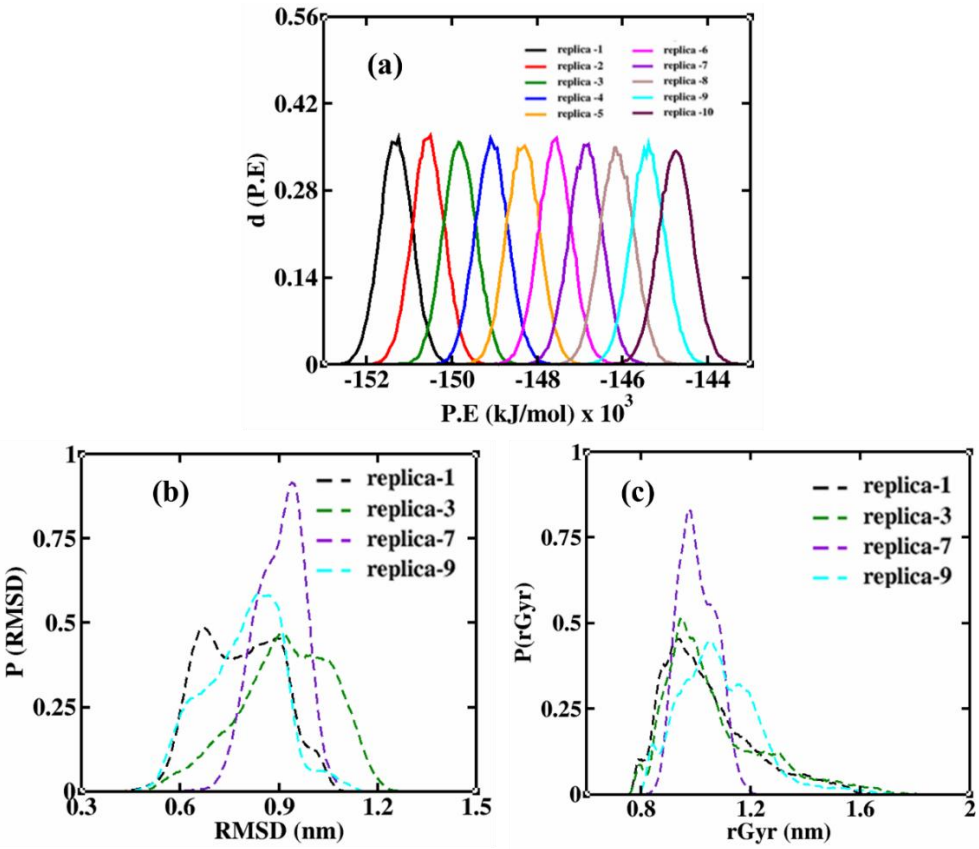


Figure 5.4: (a) The probability distribution functions of the potential energy obtained from the REMD simulation are shown for all 10 replicas. Significant overlaps are observed between neighboring temperatures. The distribution functions of the (b) RMSD and (c) radius of gyration obtained from REMD simulation.

The thermodynamics equilibrium achieved by each replica is plotted in **Figure 5.5(a).** The 1-D free-energy profile projected for all the replicas to obtain well-defined conformations is shown in **Figure 5.5(b)**. The time average secondary structure for the sheep prion peptide from the REMD simulation trajectory was calculated by DSSP analysis for each replica **Appendix XV**. A set of free-energy minima (replicas 1, 3, 7, and 9) can be defined along the free-energy profile. These minima represent the IDP stable conformations with a larger population of secondary structures during the REMD simulations. These structures are shown along the free-energy profile over the radius of gyration and RMSD reaction coordinates in **Figure 5.5 (c).** The

intrinsically disordered peptide encounters various disordered-to-ordered conformations transition states between extended and compact structures.

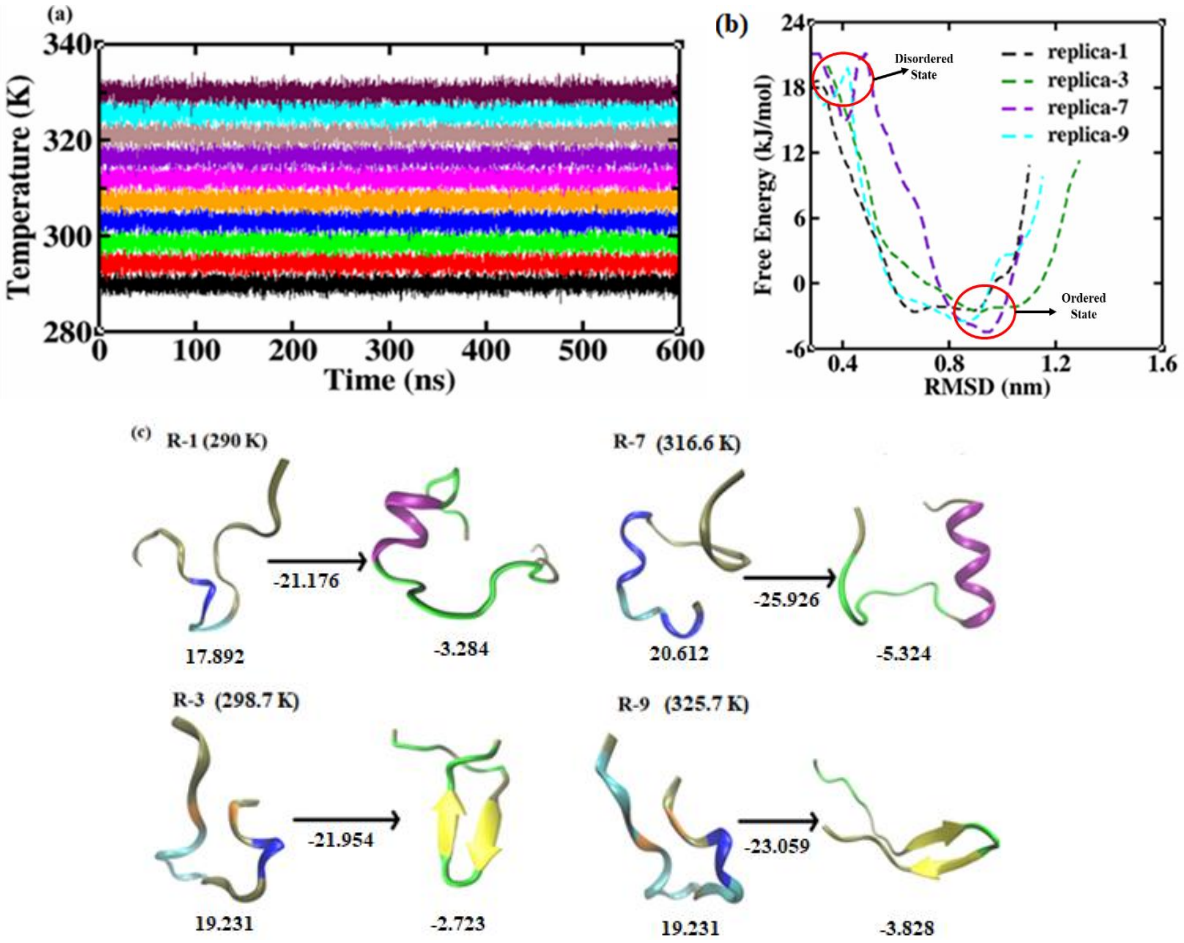


Figure 5.5: (a) Time average temperature evaluation of each replica and **(b)** 1-D free-energy profile projected for replica- 1, 3, 7 and 9 and **(c)** corresponding peptide structures showing disordered-to-ordered conformations with the free energy change (ΔΔG) in kJ/mol.

5.5 WATER CONTENT AND SECONDARY STRUCTURE PERCENTAGE IN DIFFERENT TEMPERATURES

The percentage of secondary structure content alpha-helix and beta-sheet with respect to temperature has been plotted in **Figure 5.6.** The water content per residue of the peptide with temperature was calculated and plotted it in the inset of **Figure 5.6 (a).** This will give an overview of how the hydration structure of the peptide varies with the change in the secondary structure and the temperature. It can be seen that in general, the water content is lesser in the case of the α-helix structure, and the water

content of the beta-sheet is found to be more (Nishimoto et al. 2019). It is found that at higher temperatures of 308 K and above, the secondary structures (α-helix and β-sheets) are mainly formed with the hydrophilic residues at the N-terminal; whereas at the lower temperature at 290 K and 300 K, they are formed involving the hydrophobic residues at the C-terminal. Therefore, at lower temperatures, the hydrophilic residues are free to interact with water resulting in higher water content at 290 K and 300 K. At 300 K temperature, the IDP sequence formed the beta-sheet with the higher water molecules compared to the beta-sheet formed at 330 K. At 308 K, 315 K, and 320 K temperatures the prion peptide formed in an alpha-helix state where the water molecules are very less compared to other temperatures.

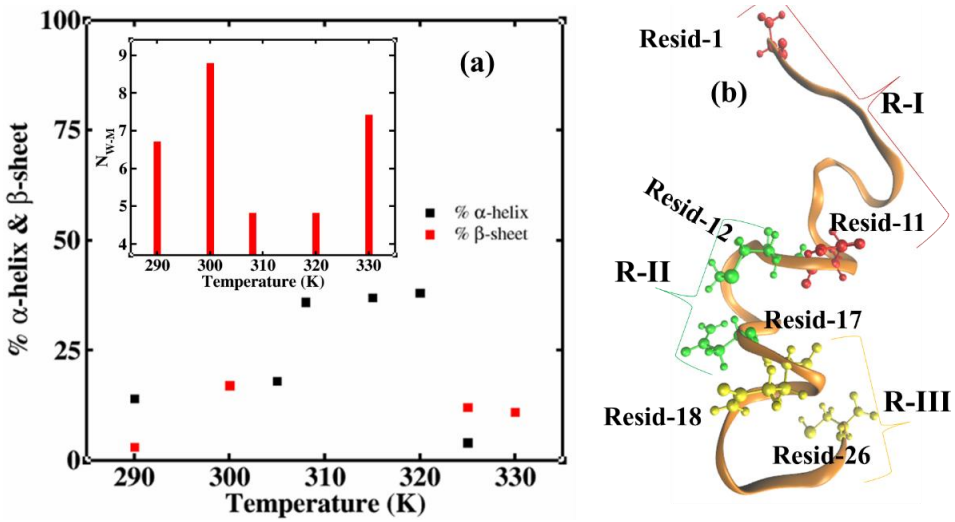


Figure 5.6: % of α-helix (Black dots) and β-sheet (Red dots) with temperature. Number of water molecules per residue with temperature visualized in the inset.

Therefore, it can be inferred from **Figure 5.6** that the percentage of α-helix increases with the loss of water near the protein surface, and the beta-sheet structures involve more water molecules near the protein surface visualized in the inset. Interestingly, the REMD simulation data shows a similar trend, the percentage of α-helix (replica-1, replica-7) was observed when the water content is lower around the peptide while the β-sheet/bridge (replica-3, replica-9) forms when the water content is more near the peptide surface (**Table 5.3).** Next, it will be interesting to examine the interaction of the residues with the temperature which guides the formation of the different secondary structures of the peptide at different temperatures. In the next section, the residue-residue contact map is discussed.

Table 5.3: Number of water content dependence of secondary structure distributions.

Replica No.	Water molecules	% of α-helix	% of β-sheet/bridge
Replica-1	185	9	1
Replica-3	240	0	7
Replica-7	140	18	0
Replica-9	243	0	11

5.6 CALCULATION OF CONTACT MAP

One of the most powerful features of cross-correlation is the ability to correlate pairs with other pairs of interest or external perturbations with the contact map. This allows precise identification of key interactions (between residues or groups of residues) responsible for some macroscopic quantity. We can easily find out the major and minor interaction networks between the residues by projecting the distances on residues using the given formula, $M_i(t) = \sum\left(1 - \frac{r_{ij}}{R_c}\right)$, where M_i is the correlation time series with a residue-residue (r_{ij}) distance and R_c is cut off. We have plotted the pair-wise residual contact maps with a cutoff distance of 0.45 nm, the cut-off was chosen to capture the non-covalent intra-peptide interactions including van der Waals, and hydrogen bonding. The anti-diagonal blue-colored region in **Figure 5.7** shows the residual contacts. It is clear from the contact map of the IDP that the temperatures have a significant effect on the residue-residue interaction profile of the prion peptide. The blue region parallel to the diagonal line is the signature of the helical conformation and the off-diagonal (cross line to diagonal) blue region indicates the β-sheet/bridge.

At 290 K (**Figure 5.7 (A)**), IDP shows more residual interaction in the hydrophobic and amphipathic regions (II and III - between residues 10-20 and 17-25) in the initial time frame (10 ns). After the alpha helix is made between residues 16-21, the interaction is reduced to mainly the C-terminal residues (residue no 16-25). The α-helix is stabilized throughout the simulation after 450 ns, and the alpha helix shifted towards C-terminal by making the residual interaction between res-22, and res-18 with the off-diagonal residues (residues 6-14). Also, it is clear from **Figure 5.13(a)** where it shows that regions I and III have more solvated compared to 300 K.

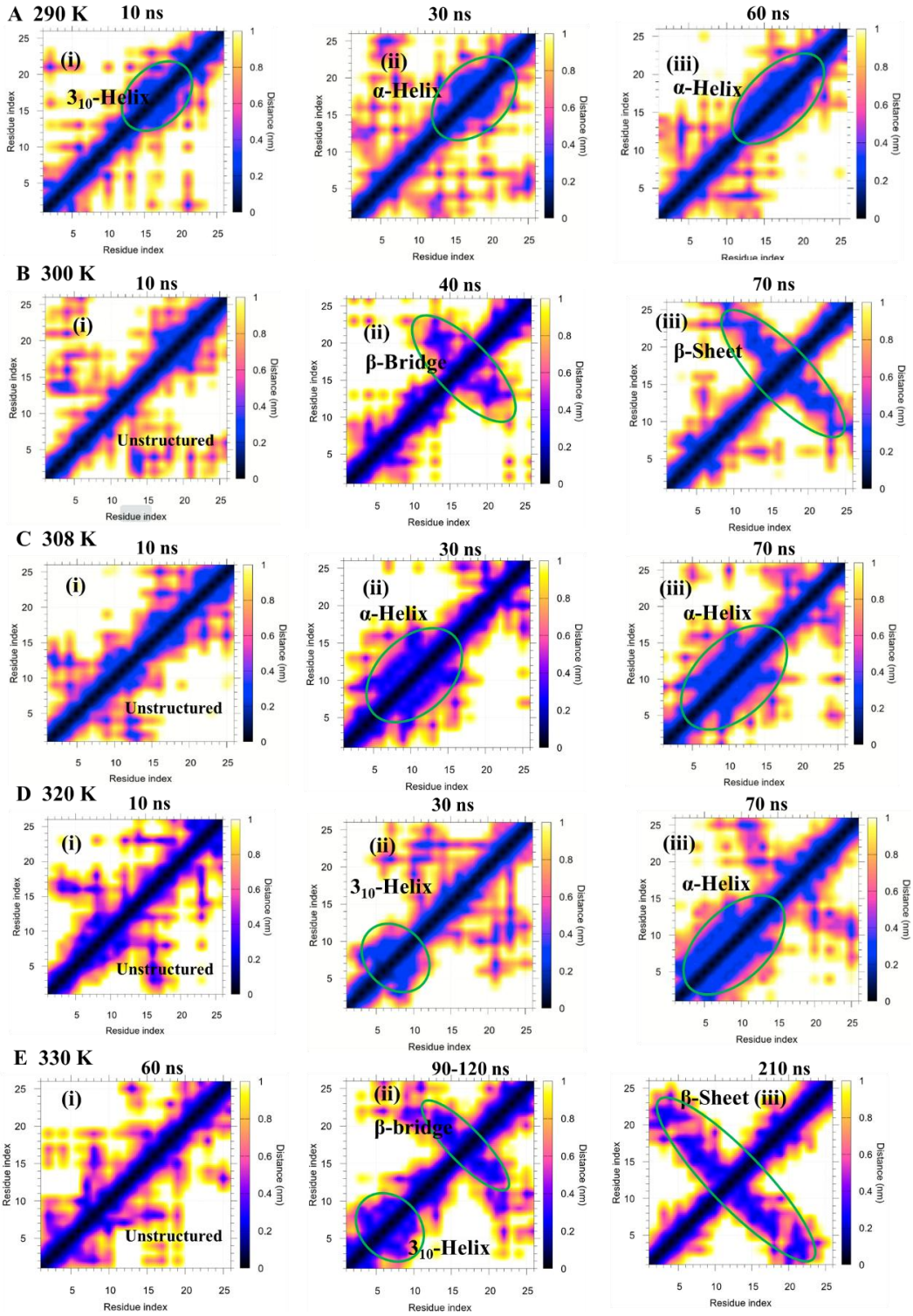


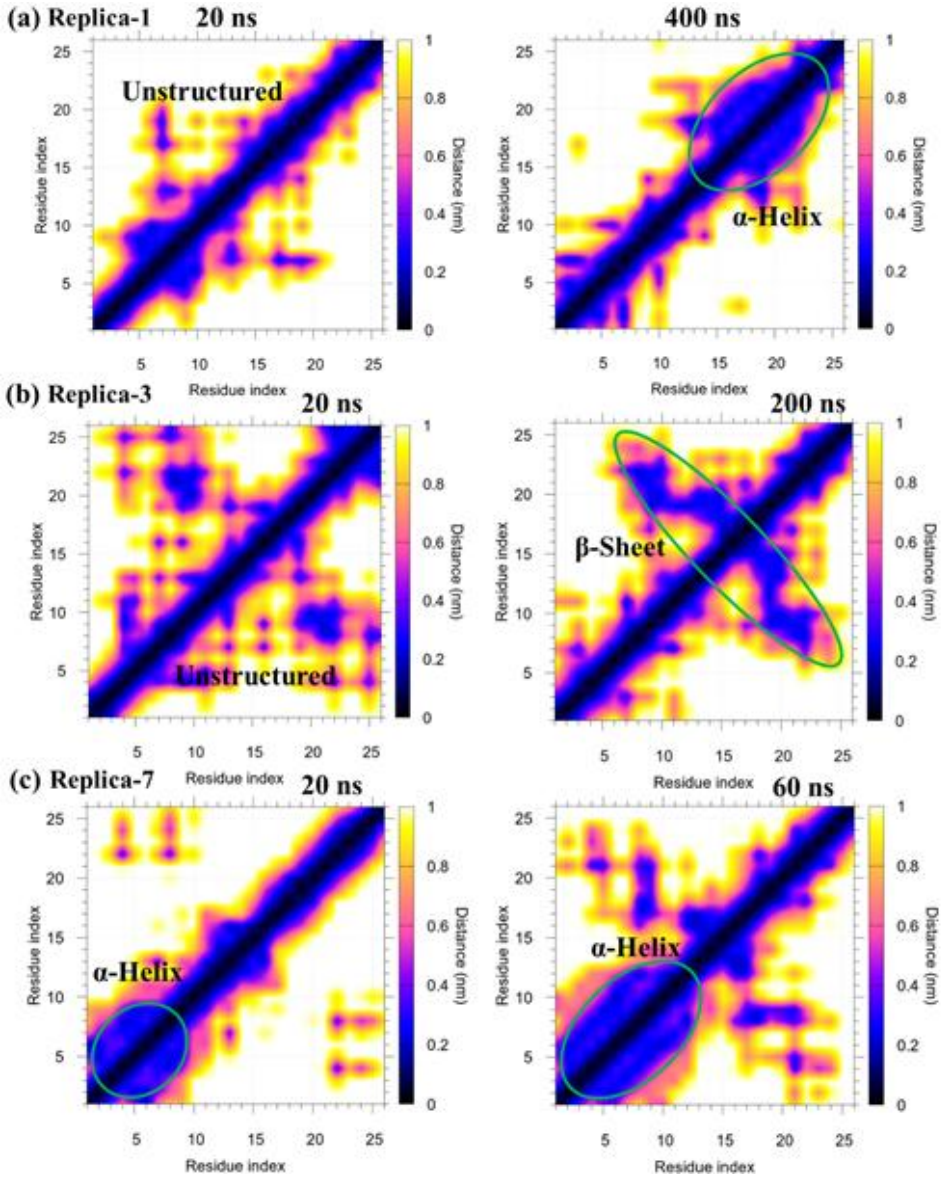
Figure 5.7: Visualization of a residue-residue contact map with different temperatures and time frames.

Experimentally, it has been reported that this peptide forms β-sheet at neutral pH. At 300 K, the hydrophobic core opens up exposing more residues to

neighbouring residues which results in the extension of the inter-residue interaction through non-covalent interactions. In **Figure 5.7 (B)**, IDP was found to have some interaction between the N-terminal residues (residues 1-10 R-I) and C-terminal residues (residues 20-25 R-III) initially. A stable bridge sheet is found to exist from 40 ns between the residues-14 to residue-18 of the IDP which extends to residues 12-21 (towards the C-terminal) after 70 ns and stabilized up to 1μs simulation. Also, it is found that, at 300 K temperature, the hydrophobic region (R-II) of the IDP molecule was solvated which indicated that the water density was found to more compared to amphipathic region (R-III). **Figure 5.13 (b)** shows that R-II is more solvated than 290 K, which may have extended the interaction between R-III. There are three proline residues towards the C-terminal ends in which one is located in the globular domain. Due to the higher mobility of these prolines, the residues from 12-21 (mostly in R-II) showed more flexible hinge which makes the residue more soluble in water thereby forming a stable beta sheet at 300 K. Thus, the secondary structure of peptide not only depends on the interactions of residues, but also on the hydration structure of involving residues. At 308 K, the inter-residue interactions are found to be mainly shifted from the C-terminal (10 ns) towards the N-terminal (30 ns) resulting to form a stable alpha helix between residue 2-11 (Region II). This helix is found to be stable throughout the simulation. After 70 ns, we observe the beginning of some extended interactions between the hydrophobic residues (Residues 18-22, R-II). Similar observations are seen for 320 K (**Figure 5.7 (D)**). Thus, the secondary structure of peptide is not only depending on the interactions of residues but also on the hydration structure of involved residues.

At 330 K, the IDP has a random intra-residual interaction compared to inter-residual interaction in the initial time frame (10 ns) and found to be unstructured (**Figure 5.7 (E)**). At 70 ns the IDP was found to be in the 3_{10}-helix state and after 120 ns the 3_{10}-helix state and a transient into β-sheet/bridge is found. After 210 ns, a stable β-sheet/bridge structure is observed almost throughout the peptide i.e., at the C-terminal residues as well as near the N-terminal end because of the random atomic fluctuations. The prion peptide was found to be more or less equally solvated in all three regions and the density of the water molecules is higher compared to other

temperatures. Nevertheless, the IDP was found to be in a small fraction of α-helix, 3_{10}-helix, and β-bridge conformation at different temperatures in various locations of the peptide. In conclusion, the structure of the IDP is found to be affected by the density of water at different temperatures showing different secondary structures (Nishimoto et al. 2019). We have calculated the residue contact map from the enhancement sampling simulation trajectory (REMD) and found a similar trend as observed in individual simulations in **Figure 5.8**. It will be more interesting to calculate the various prominent conformations from disorderliness conformations to orderliness conformations. Therefore, network analysis was done to find out the prominent conformation at various temperatures during the simulation.



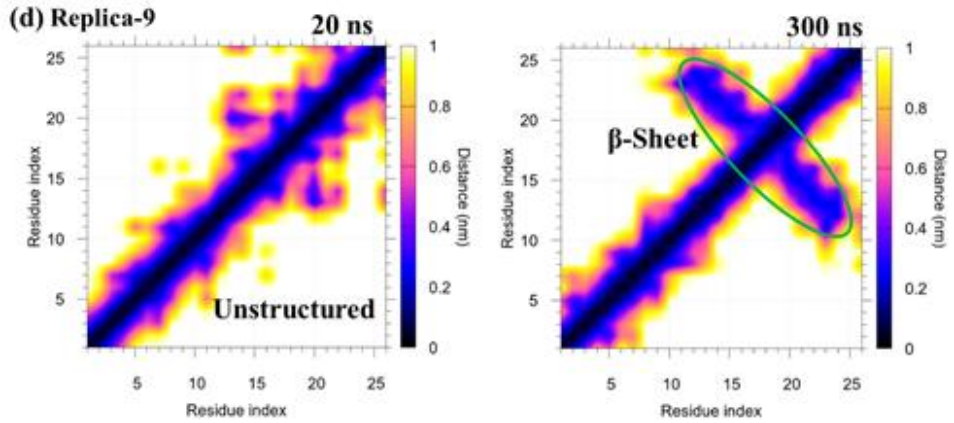


Figure 5.8: Visualization of residue-residue contact map for replica-1, 3, 7, and 9

5.7 PROMINENT CONFORMATIONS OF THE IDP

To get an overview of different conformations for the calculations, the most prominent conformations for each temperature were identified based on the frequency of its existence in the whole trajectory, and network analysis was done to find out how close or different these clusters are related to each other; *i.e* inter-connectivity between the conformations. A maximum of 250 conformations were considered for each analysis. Two conformations obtained from the simulation trajectory are considered to be connected if d (t, t′) < 2.5 Å. Since we compared the change in the conformation for all the temperatures, we preferred RMSD criteria 5.0 Å where we get two prominent conformations for 300 K. The conformational dynamics of the IDP change due to the thermal fluctuations in the side chain atoms of the peptide. The major conformations of prion peptide were extracted from 1μs trajectory for each temperature using the Gromos clustering method (Daura et al. 1999). The prominent structure has been chosen from the cluster in which the conformation exists more than 60 ps. The network distribution at different temperatures is given in **Figure 5.9** with communities represented in different colors. The prominent conformations are fewer as the temperature increases. This can be explained as, with the increase in temperature, the conformations are more random and transient. Therefore, the number of prominent stable structures reduces with temperature.

At lower temperatures, 290 K more stable prominent structures are found than the higher temperatures (**Figure 5.9 (a)**). The different color shows a different

community of structures. The nodes present in different communities have different structures. At higher temperatures, many communities with fewer connections are observed which explains that sharp differences in the structure happened during the structural transition/conformational evolution. It can be noticed here that for sampling the IDP conformations at different temperatures, the number of communities is constant from 308 k to 330 K.

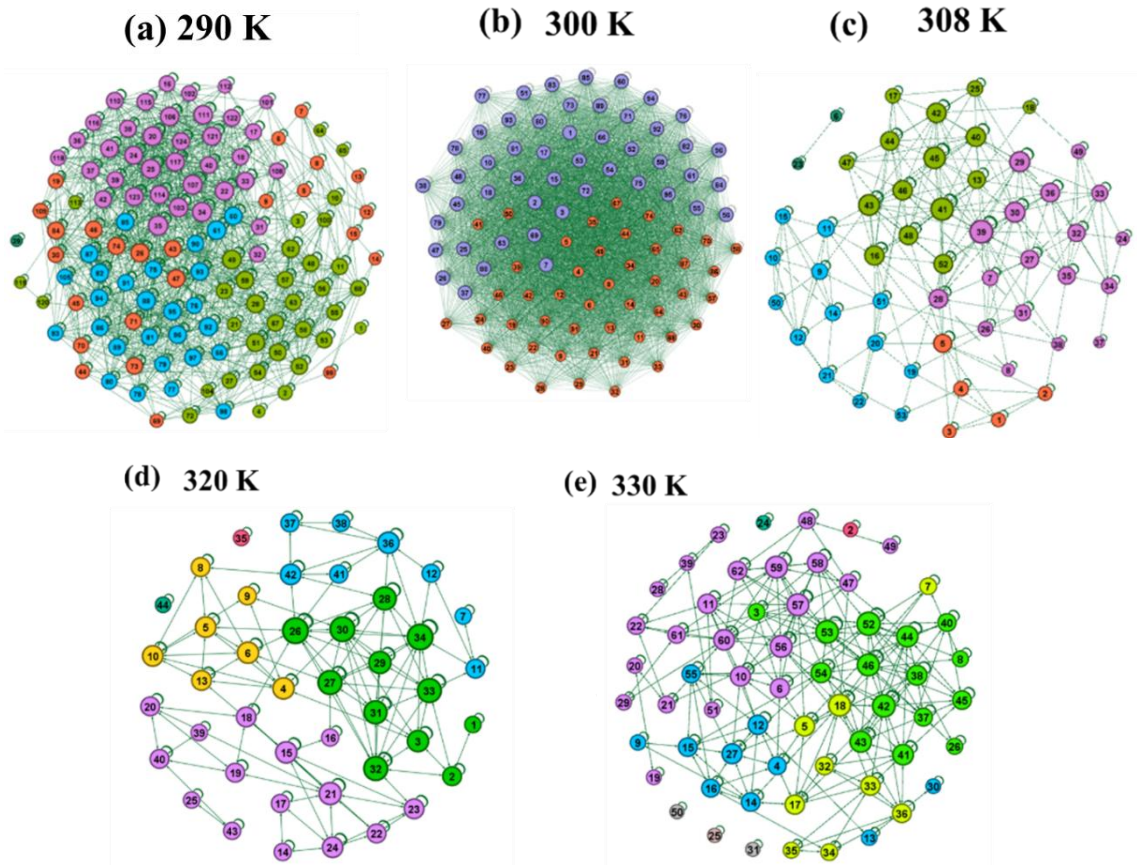


Figure 5.9: Network distributions of most prominent conformations of intrinsically disorder prion peptide for 1 μs simulation trajectory at different temperatures in TIP3P water model.

300 K showed only two communities which have more interconnections (**Figure 5.9 (b)**) which suggests that at this temperature, the structures are inter-convertible and related. The rise in temperature helped the easy transition between the structures. At this temperature, the formation of the beta-sheet is observed and there is no other changes in the structure except the side chain fluctuations. At 290 K, there is a change in the structure from a random coil to a 3_{10}-helix to the alpha helix. At

higher temperatures (308 K, 320 K, and 330 K) also such different structures are found, therefore there we have different communities. These are more distinct separated communities, and the connection between the communities was less (**Figure 5.9 (c-e)**). This indicates that structure was found to be less or moderately interchangeable between the communities. The free energy change in IDP conformations shows the stability of the prominent structure in each community. The free energy change was calculated by RMSD-based calculation (Anandakrishnan et al. 2019) which is given by, $\Delta G = -k_B T ln P(RMSD)$, where k_B is the Boltzmann constant (8.3145×10^{-3} kJ/mol) and P(RMSD) is the probability of conformations with RMSD. The change in the free energy of the prominent conformation is shown in **Table 5.4**.

Table 5.4: Free energy of the prominent conformation for intrinsically disordered prion peptides.

Temperature (K)	Color	Energy (kJ/mol)
290	Magenta	-1.042
	Green	1.016
	Cyan	1.082
	orange	2.172
300	Blue	-0.894
	red	1.483
308	Magenta	-0.816
	Green	-0.746
	Cyan	-0.065
	orange	0.346
320	Magenta	-0.868
	Green	-0.063
	Cyan	0.197
	Yellow	0.127
330	Magenta	-0.195
	Green	-0.825
	Cyan	1.042
	Yellow	2.195

It is found from **Table 5.4** that the negative ΔG values at every temperature of the IDP were found to be stable in conformation in the secondary structure such as α-helix, 3_{10}-helix, and β-sheet/bridge state. **Figure 5.10** shows that the magenta color

community shows the stable conformation (α-helix) in which the nodes are more connected to each other with -1.042 kJ/mol at 290 K. At 300 K, 308 K, and 320 K, the ΔG values are clustered between -0.89 to 0.062 kJ/mol, and this free energy shows that the disordered conformation of IDP is found to be in ordered alpha-helix and beta-sheet. At 300 K, the free energy of the blue community is found to be -0.894 kJ/mol which shows the stable beta-sheet conformation. At higher temperatures, 308 K, and 320 K, the IDP shows some stable conformation with the negative free energy changes -0.816 (Magenta), and -0.748 (green) kJ/mol in the obtained communities.

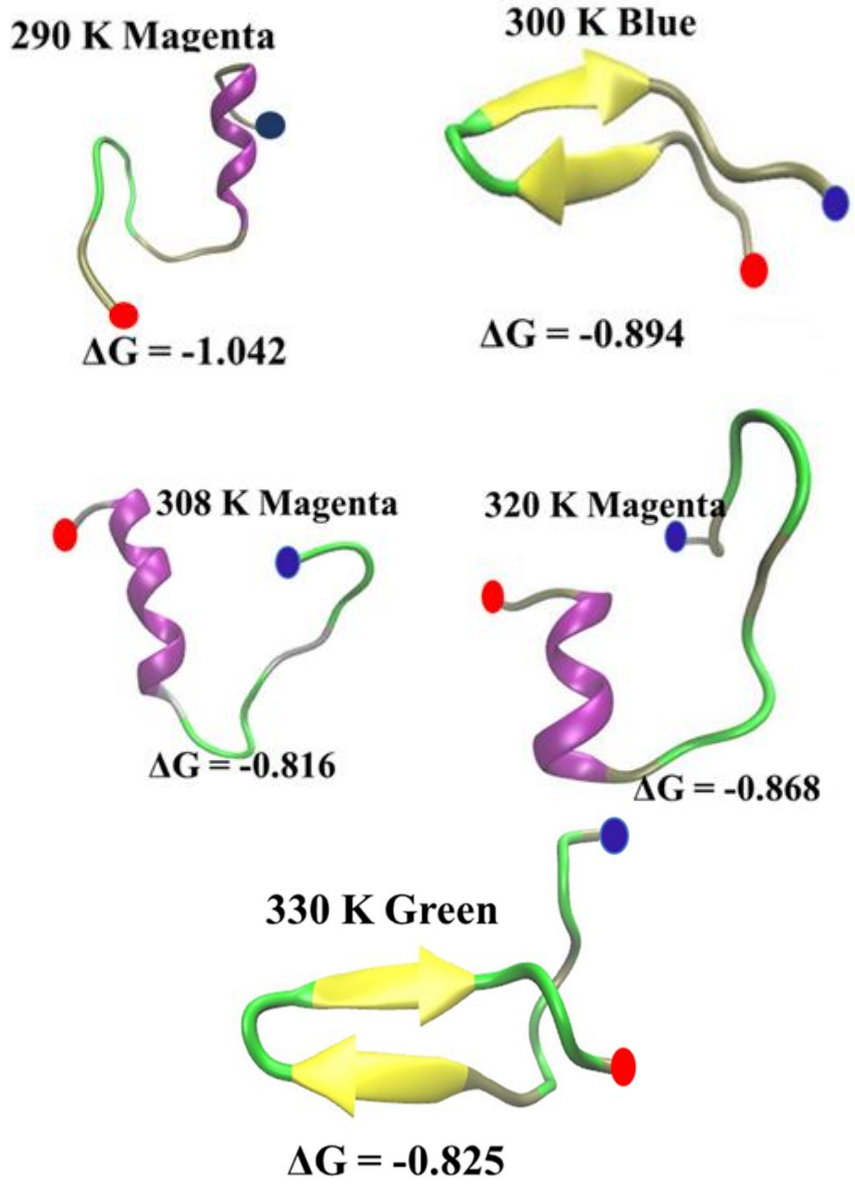
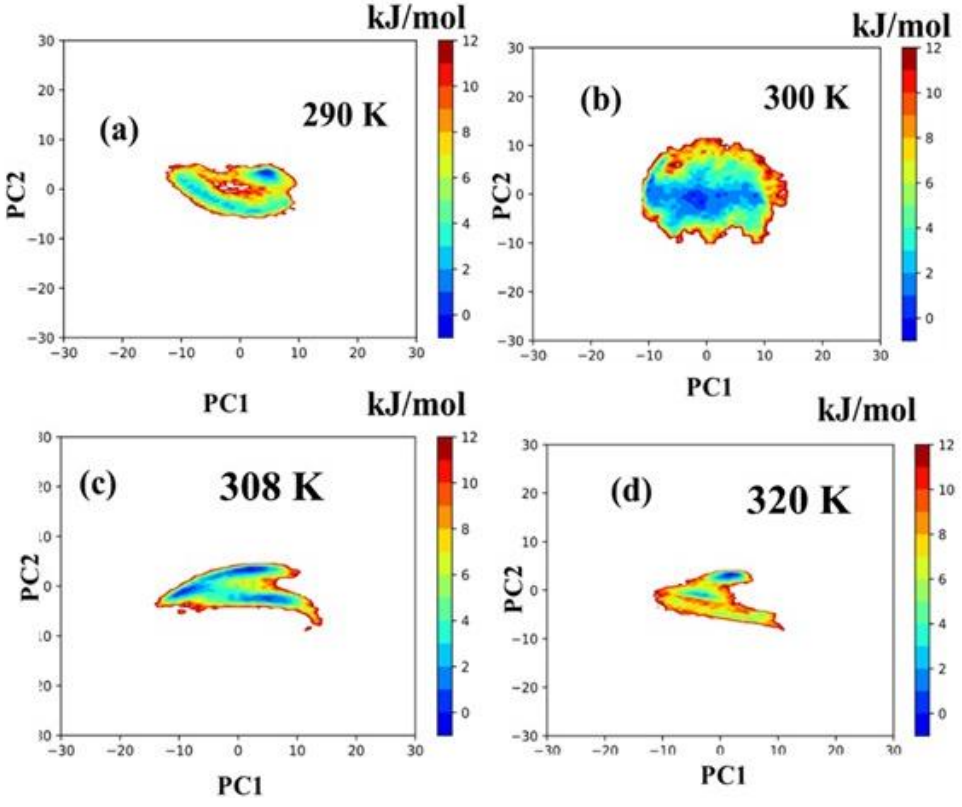


Figure 5.10: The stable conformation evolved in different communities at different temperatures with the lowest free energy changes. ΔG values in given in kJ/mol.

5.8 FREE ENERGY SURFACE (FES)

The principal component analysis discloses the high amplitude collective motion in MD simulation trajectories, through the eigenvectors of the covariance matrix of the positional fluctuations of the atoms, which was calculated on protein alpha carbons (C_α) using individual MD simulation trajectories. To describe the dimensionality of the crucial subspace, the fraction of total fluctuation/motions is described by reducing the dimension and computed as the sum of the eigenvalues included with the eigenvectors (Amadei et al. 1993). The cosine content (c_i) of the principal component (PC) can be extracted from covariance analysis and ranges between 0 and 1. The cosine value tells about the convergence of the sampling data from the MD simulation trajectory, if the cosine value is below 0.5 then the system converges and the higher value represents random motions (Hess 2000; Venugopal and Chakraborty 2022). The evaluation of cosine contribution for the first eigenvectors is sufficient to give a reliable idea of the protein behavior. The cosine content was calculated on the first principal components of every simulation trajectory (**Appendix XVI**). It is evident that the 1 μs simulation trajectory showed the convergence of the conformational sampling of every individual replicated simulation.



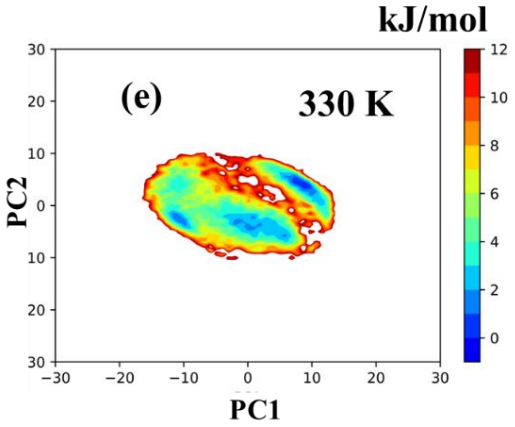


Figure 5.11: The free energy surface using as reaction coordinates of the principal components (PC1) along the (PC2) of the MD simulation trajectory at different temperatures.

The first two principal components i.e., PC1 and PC2 were considered as the reaction coordinate to generate the free energy surface of the intrinsically disordered prion peptide at different temperatures in **Figure 5.11**. It can be seen that the projections or sampling points of the prion peptide are widely spaced in the case of 300 K, and 330 K temperatures due to the formation of β-sheet/bridge, leading the more thermal fluctuation and the solvation of the IDP. In contrast, the FEL corresponding to alpha helix structures is more compact (290 K, 308 K, 320 K) **Figure 5.11 (a,c,d)**. The lower free energy corresponds to the structure is found to be more spread and leads to the extended beta-sheet conformation at 300 K. This can be observed from the **Figure 5.3(a-b)** the RMSD and R_g values were found to be higher at 300 K. The free energy surface value is comparatively higher which indicates a flat energy surface where the peptide conformations can float the entire phase space (Singh et al. 2022) obtained from our MD simulation **Figure 5.11 (b**). Moreover, the free energy surface of the PCA calculation suggested that at lower temperatures at 300 K, the more positional motion was found and reduced after forming a secondary structure α-helix at higher temperatures 308 K, and 320 K **Figure 5.11 (c-d).** The peptide is less compact at 300 K compared to 308 K, 320 K temperature therefore the water molecule are near to the hydrophobic residues of the peptide. At 330 K the free energy surface is more spread out due to the random motion of the residues **Figure 5.11 (e)**. The free energy surface calculated for REMD simulation trajectory from the

last 200 ns and found that the replica-1 (290 K), replica-3 (300 K), replica-7 (323 K), and replica-9 (329 K) found in the lesser wide-spaced leading with ordered structures in **Figure 5.12**. The replica-1 was found to be in α-helix toward the C-terminal side, replica-3, replica-9 β-sheet/bridge, and replica-7 were in α-helix toward the N-terminal side. Next, it will be interesting to see the solvation effects on the structural conformations of the prion peptide at different temperatures.

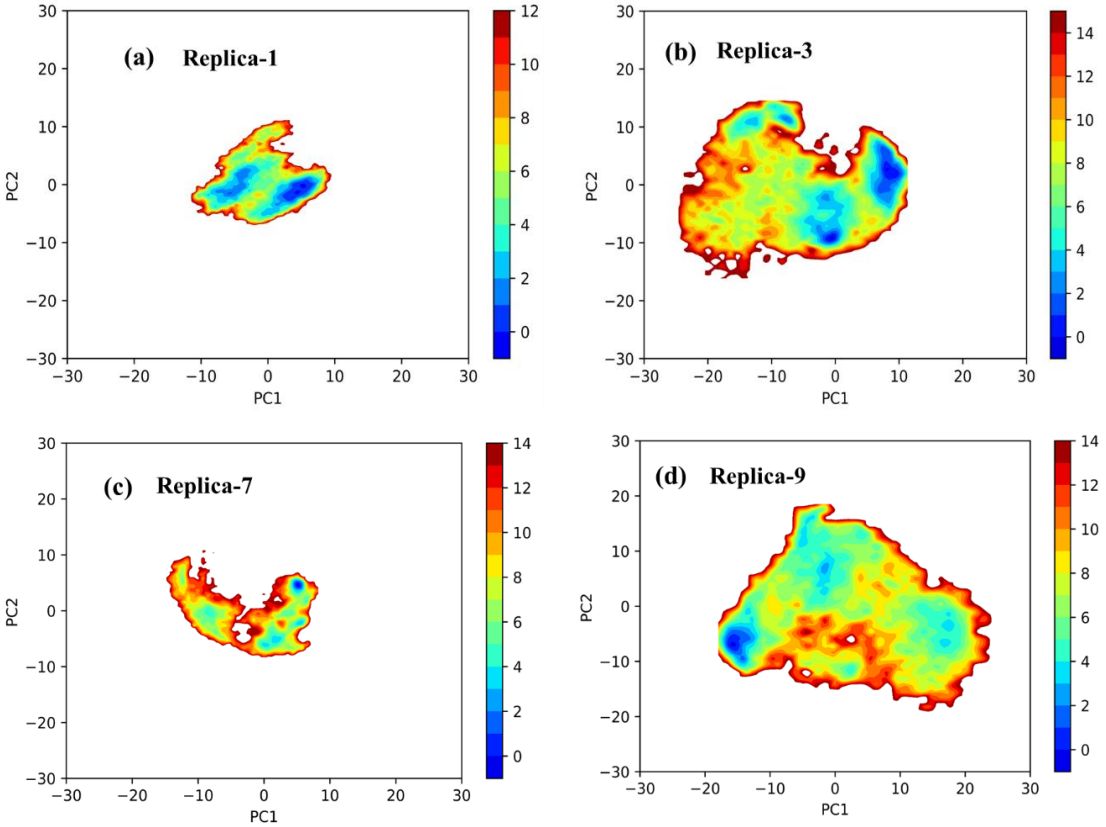


Figure 5.12: The free energy surface using as reaction coordinates of the principal components (PC1) along the (PC2) for replica-1, replica-3, replica-7, and replica-9 of the REMD simulation trajectory.

Additionally, the free energy surface can be explained based on the solvation structure of the peptide at different secondary structure regions, due to the fluctuation in the residues which are dependent on the solvation structure. Therefore, in the next section, the study of solvation structure has been done at various temperatures.

5.9 HYDRATION SHELL STRUCTURE

The hydration shell structure was quantified by calculating the radial distribution function (Nguyen and Pettitt 2015) (RDF) of water oxygen atoms around the oxygen atoms of protein **Figure 5.13**. We selected the peptide in three different regions hydrophilic (region-1), hydrophobic (region-2), and amphipathic (region-3) region. The position of the first two solvation shells was found to be at a distance $r \sim 2.8$ Å and 5.8 Å respectively from the protein surface.

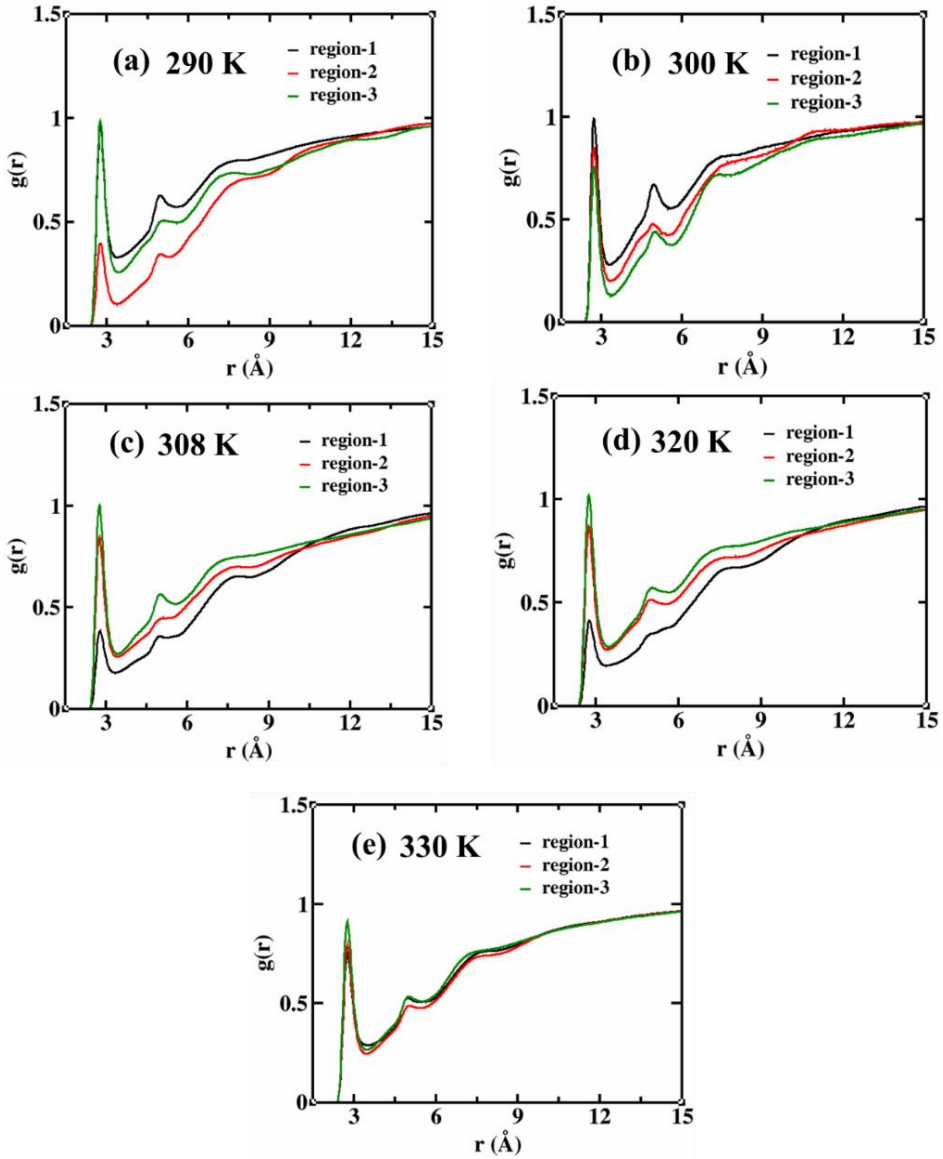


Figure 5.13: The RDF is a function of the distance **r** between an oxygen atom of water and the oxygen atoms of prion peptide at different temperatures in all three regions.

The first solvation shell shows a higher density compared to the second solvation shell. At lower temperatures (290 K) the hydrophilic region-1 and amphipathic region-3 were found to be more solvated compared to hydrophobic region-2 (**Figure 5.13 (a)**). At 300 K temperature, the hydrophobic region-2 of the IDP molecule was solvated which indicated that the water density was found to more compared to amphipathic region-3 (**Figure 5.13 (b)**). When we increased the temperature from 308 K to 320 K the peak height of the first solvation shell of hydrophilic region-1 decreased showing the dewetting of this region. The weak hydrogen bond between the peptide and the water molecules allows to development of the secondary structure here and which leads to a decrease in water density and solvation (**Figure 5.13 (c-d)**). In the case of 330 K temperature, the prion peptide was found to be more or less equally solvated in all three regions and the density of the water molecules become low compared to lower temperatures. The structure of the IDP is found to be affected by the density of water at different temperatures showing different secondary structures. Further, to find the correlation between the secondary structure formed and the number water content we calculated the number of hydrogen bonds and the interaction energy for some specific peptide regions in the next section.

5.10 NUMBER OF HYDROGEN BONDS AND INTERACTION ENERGY

In **Figure 5.14 (a)**, the number of peptide-peptide hydrogen bonds (with error bar) and the number of intra-peptide hydrogen bonds for interfacial water molecules with variations in the temperatures were plotted. The water molecules that are present within 5 Å from the carbonyl groups of the protein backbone were considered as interfacial water molecules. It is decided from the RDF of oxygen of protein and oxygen of water shown in **Figure 5.13**. It can be seen that the protein-water hydrogen bond is decreased and peptide-peptide increased when the IDP transient from disordered to ordered conformation. The protein is found to have lesser protein-water hydrogen bonds at 308 K, 315 K, and 320 K when it forms the α-helix and consequently, the number of intra-peptide hydrogen bonds increased at these temperatures in **Figure 5.14 (a).** This is because the number of residues (11 residues) involved in making the secondary structure at this temperature is more. The protein is found to be more hydrated when it folds into the β-sheet/bridge structure and

consequently, the peptide-peptide hydrogen bond is found to be less. It can be noted that the number of residues involved in the b-sheet is generally lesser than the alpha helix. To further clarify this issue, we further considered two different segments in the peptide that forms alpha helix and β-sheet at different temperature. The first segment we considered from residue 15 to 18 present at the C-terminal which formed alpha helix at 290 K, β-sheet at 300 K, and unstructured at 320 K, and the other segment we considered from residue 6 to 12 which was unstructured at 290 K, folded alpha helix at 308 K and 320 K, β-sheet at 330 K.

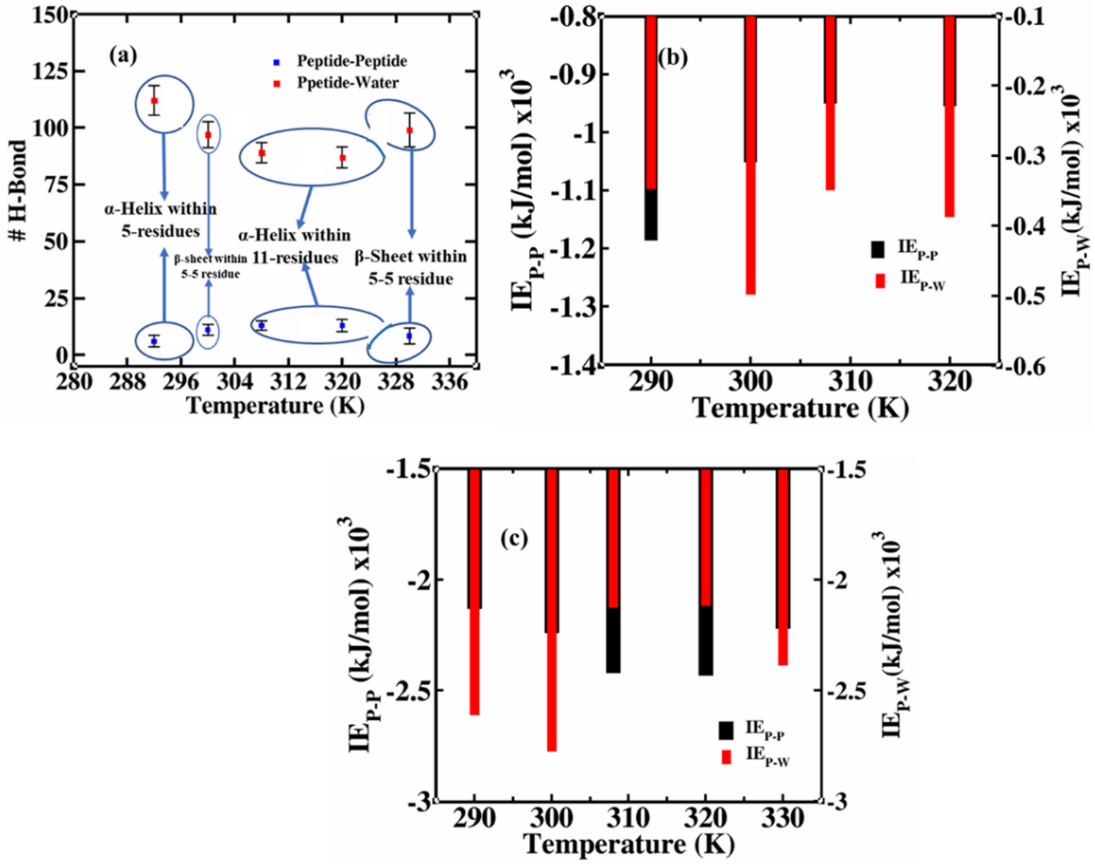


Figure 5.14: (a) Number of hydrogen bonds between the peptide-peptide and peptide-water (b) the interaction energy between the peptide-peptide and peptide-water of the IDP segment formed α-helix at 290 K, β-sheet at 300 K and unstructured at 320 K (c) the interaction energy between the peptide-peptide and peptide-water of the IDP segment formed α-helix at 308 K, 320 K whereas at 290 K it is unstructured and 330 K making beta-sheet.

It observed at 290 K when the peptide formed α-helix, it is less solvated. The beta-sheet structures are found to be more solvated. From **Figure 5.14 (b)** it is observed that the intra-peptide interaction energy where the IDP segment is forming the α-helix at 290 K was found more compared to β-sheet/bridge at 300 K and unstructured at 320 K forms in the same position of IDP segment. The interaction energy between the water-peptide was found to be less in ordered α-helix conformation compared to β-sheet/bridge and unstructured state.

The residues at the N-terminal are hydrophilic in nature, therefore the water-peptide interaction energy is found to be more. These residues were found to make alpha helix at 308 K and 320 K, where the water-peptide interaction energy is found to be less than the beta-sheet structure formed at 330 K. Further, the lifetime of the protein-water hydrogen bond from the time correlation function of continuous hydrogen bond were calculated as defined in Section 2.4.2 (ii).

Table 5.5: The lifetime (τ_{HB}) of continuous hydrogen bonds (ps) formed between protein-water and water-water in the presence of all salt solutions including the standard errors.

Temperature	hydrogen bond Life-time between peptide-peptide (ps)				hydrogen bond Life time bond between peptide-water (ps) (<5.5Å)			
	S1	S2	S3	**SD**	S1	S2	S3	**SD**
290 K	7.01	7.85	7.35	0.422	0.48	0.49	0.47	0.010
300 K	6.32	7.05	6.85	0.377	0.45	0.43	0.44	0.010
308 K	8.45	8.55	8.85	0.208	0.42	0.42	0.41	0.006
320 K	9.35	8.95	8.35	0.503	0.35	0.34	0.35	0.006
330 K	7.87	7.77	6.95	0.505	0.31	0.32	0.31	0.006

*S1, S2 and S3 are individual simulations.

In **Table 5.5**, the hydrogen bond between the intra-peptide and the peptide-water hydrogen bond in the interfacial water around the protein was calculated (Bandyopadhyay et al. 2005; Luzar and Chandler 1996; Singh and Chakraborty 2021). The interfacial molecules up to the second solvation shell (cut-off value of <5.5 Å) were considered for the calculation of protein-water hydrogen bond lifetime.

From **Table 5.5,** it is observed that the hydrogen bond lifetime of the peptide (donor)-peptide (acceptor) was found to be more when the IDP transient into the alpha-helix conformation of the IDP compared to the beta-sheet. The peptide is found to be in alpha helix structure at 290, 308, and 320 K temperatures. At 300 K and 330 K, it is found to be in a β-sheet structure. The hydrogen bond lifetime between the peptide and water molecules was however found to decrease with an increase in temperatures. At lower temperature (290 K) hydration layer around the sheep prion peptide have more hydrogen bond lifetime. At higher temperatures 325 K and 330 K, the lesser hydrogen bond lifetime of peptide water indicates a less stable hydration layer around the IDP surface. It is evident from the above discussion that the IDP is more sensitive under the thermodynamic conditions and residue-residue interaction network to change the conformation from a disordered to an ordered state.

5.11 CONCLUSION

Lastly, to conclude we have performed AMD and REMD simulations to obtain the conformational transitions of IDP from a disordered to an ordered state. It should be noted that the results of both atomistic molecular dynamics and REMD simulations showed similar results showing the robustness of the simulated trajectories. The multivariate statistical analysis and principal component analysis method efficiently showed the atomic motion responsible for the conformational transition of the IDP under different conditions. To the best of our knowledge, this is the first study to show the residue specificity involved in disordered-to-ordered structural transitions and vice versa. The hydrophobic region (C-terminal) of the IDP was found to be alpha-helix at 290 K which changed to beta-sheet at 300 K due to changes in fluctuations in the residues of the peptide. The RDF and the contact map of the residues showed that the peptide hydrophilic region (N-terminal) was found to be solvated more at lower temperatures whereas growth in the interaction was found to be in the case of hydrophobic residues. This led to the formation of the alpha helix at the C-terminal. With the temperature rise, the hydrophobic residue got more solvated, and the hydrophilic residues at the N-terminal got dewetted shifting the formation of the alpha helix at the N-terminal. At a very high temperature around 330 K, the random motion of the peptide was found to make a beta-sheet structure. The network

analysis correlated well with the structural transition of the peptides at a given temperature. The free energy corresponding to the stable conformation in a particular community was found to be the lowest. It is found the FEL for the alpha helix conformations is less widely spaced compared to the beta-sheet structures. The PCA free energy surface showed at 290 K, 308 K, 315 K, and 320 K are found to be compact to the less R_g and RMSD values. The intra-peptide number of hydrogen bond value was found more during the formation of alpha-helix and less in the beta-sheet with a higher peptide-peptide and lower peptide-water hydrogen bond lifetime. The interaction energy of peptide-peptide was found less due to the solvation of the peptide in the case of the beta-sheet. The peptide-peptide hydrogen bond lifetime of the alpha-helical structure was found to be more compared to the beta-sheet structure.

The REMD calculation showed a similar structure as reported by the AMD simulation. The higher temperatures (replica-9) indicate the formation of β-rich conformation of PrP^{Sc}-like structures and alpha-helix in replica-7 at 320 K. Replica 1 (290 K) and 7 (316 K) show the alpha helix structure and replica-3 (299 K), replica-9 (326 K) folded into beta-sheet. The antiparallel β-rich states a toxic oligomer is concerned in sheep prion peptide. Thus, in future directions to understand the mechanism of oligomerization of these disordered peptides, machine learning techniques can be used to understand the disordered-to-ordered conformations.

CHAPTER 6

TEMPERATURE-DEPENDENT CONFORMATIONAL EVOLUTION OF SARS COV-2 RNA GENOME BASED ON RMSD AND INTERHELICAL ANGLE CLUSTERING USING NETWORK ANALYSIS

Understanding the dynamics of the SARS CoV-2 RNA genome and its dependence on temperature is necessary to fight the current COVID-19 crisis. Computationally, the handling of large data is a major challenge in the elucidation of the structures of RNA. Network Analysis based on RMSD and interhelical orientation are analyzed to effectively sample different RNA conformations.

6.1 BACKGROUND

Large-scale conformational changes of biomolecules give necessary hints regarding their structure and activity at the cellular scale. Determining these structural changes has remained elusive both experimentally and theoretically. Over the past few decades, with the development of cryo-electron microscopy and other crystallography techniques, multiple snapshots of large and flexible systems were found. As this data got accumulated, theoretical methods became indispensable to understanding how different conformers interconvert to carry out biological functions (Huggins et al. 2019). The potential of pathogenic RNA viruses to adapt to various environmental factors poses a great threat to disease control and management. Current limitations in the experimental procedures to understand the RNA structure-dynamics-function relationship affects the process of vaccine and drug discovery which targets viral RNA genome (Rodrigues et al. 2020). This creates the need for theoretical investigation of the structure-function relation of this macromolecule.

The study of the folding-unfolding-misfolding mechanism and energy landscape of biomolecules gives meaningful information regarding the microstates of the system which governs its stabilization and biological functions. However, several challenges stand in using energy landscapes of macromolecular systems like RNA to relate with the structure and dynamics to function. One big problem in handling such a big system is the handling of the data. Because of this drawback, coarse-graining of

the biomolecules is becoming more popular at the expense of the atomistic detail of the simulations. On the other hand, atomistic molecular dynamic simulation can give us proper detail of a process but computation of the long trajectories of the macromolecular system is computationally prohibitive. In addition to that, long and multiple simulations are required to ensure proper sampling of the data. One very important characteristic of proper sampling of the data is we should not miss out on any important information which is happening during the process of the evolution of the trajectories; at the same time, we cannot focus on very minute details of such big systems because that will create unnecessarily large data which are of insignificant use. ENM (Elastic Network Model), on the other hand, is a coarse grain technique, where they map nucleotides onto a reduced set of particles that are bonded to all of their neighbors within a predefined cutoff radius by elastic potentials. These techniques are good at defining structures nearby the equilibrium states but cannot scan conformational evolutions of a system with frequent bond breaking and formation (Bahar et al. 2010; Fuglebakk et al. 2013; Haliloglu et al. 1997). Therefore, there is a need for an unbiased technique that can sample the conformations of large bio-molecular systems well and will also allow larger displacement transitions. To provide conformational flexibility to fold RNA helical stems, there exists at least one connecting segment comprising one or more unpaired nucleotides called linking nucleotides forming n-way junctions in the molecule. Three-way junctions are quite often found in functionally important RNA species (Butcher and Pyle 2011). Interhelical angles have been previously used for the clustering of RNA conformations using Euler angle calculation (Bailor et al. 2010; Dickson et al. 2014; Sim and Levitt 2011). But as it is known that the structure of the RNA is quite dynamic and is very sensitive to the change in temperature, the probability of retention of the native stems base pairs throughout the trajectory is quite low (Šponer et al. 2017). This poses the problem of consistency in mapping interhelical angles when the helix denatures or renatures. Therefore, we need an angle parameter that can sample configurations at higher temperatures and help us to compare those results with the lower temperature. In this chapter, network analysis method was introduced that can sample big data sets effectively to identify the most prominent conformations of the macromolecules like RNA, which existed during the process of

evolution/denaturation and can also show how these conformations are connected to each other.

6.2 SYSTEM SETUP

SARS-CoV-2 RNA FSE (PDB ID: 6XRZ (Zhang et al. 2021) was simulated in GROMACS (Abraham et al. 2015) 2018.4 using CHARMM27 (MacKerell et al. 1995) force fields at six different temperatures namely, 288 K, 303 K, 318 K, 333 K, 348 K, 363 K (**Table 6.1**). For detailed analysis, we compared the results of RNA simulations using two water models: the SPC/E model and the modified TIP3P model.

Table 6.1: Summary of the simulated system using CHARMM force field.

System	Temperature (K)	No of RNA	No of water	No of ions	Time (ns)
RNA in SPC/E water model	288	1	32669	87	250
	303				
	318				
	333				
	348				
	363				
RNA in mTIP3P water model	288	1	32669	87	250
	303				
	318				
	333				
	348				
	363				

The CHARMM27 force field has been validated for RNA, DNA, and oligonucleotide and higher conformational flexibility is observed with this force field. Overall the structural results obtained were similar to those from the CHARMM36 force field (Baker et al. 2011; Foloppe and MacKerell, Jr. 2000). These temperatures were selected based on the experimentally reported thermal denaturation temperature

of the SARS-CoV (Darnell et al. 2004). The simulations were solvated using SPC/E (Mark and Nilsson 2001) water model in a periodic cubic box with a distance of 9 Å from the box boundary. The RNA molecule is less stable in the TIP3P water model than in SPC/E at lower temperatures (Singh et al. 2021). Na^+ ions were added to neutralize the charge of the system. The energy was minimized by the steepest descent algorithm with a maximum of 50000 steps. LINCS constraints (Hess et al. 1997) were used to restrain the bond involving hydrogen atoms. NVT equilibrations were carried out for 10 ns using the V-rescale temperature coupling method (Bussi et al. 2007) with a time step of 2 fs. This was followed by NPT equilibrations of 10 ns with a time step of 2 fs at 1 atm using the Nose-Hoover thermostat (Martyna et al. 1992) and Parrinello-Rahman pressure coupling (Parrinello and Rahman 1981) scheme respectively. The particle mesh Ewald method (Essmann et al. 1995) was used to calculate the long-range electrostatic interactions with a Fourier spacing of 0.16 nm. The short-range van der Waals cut-off was fixed to 1.2 nm. Finally, NPT production runs were performed for 250 ns, until the RMSD converged for all the systems (**Appendix XVII**). The simulations were replicated and the results are provided wherever necessary.

6.3 RMSD CLUSTERING AND NETWORK ANALYSIS

The conformations for network analysis were obtained by an in-built clustering algorithm in GROMACS software using RMSD as the criterion. A cut-off value of 1.5 Å RMSD was selected to distinguish between 11500 structures extracted from the 250 ns trajectory saved at a time interval of 0.02 ns for each temperature. The selection of a proper cut-off is crucial as a lower cut-off value gives too many conformations, which may lead to irrelevant results; while a larger cut-off value neglects the finer details of the conformations. The total number of conformations increases due to the random motion of the atoms with the rise in temperature **(Figure 6.1, Appendix XVII)**.

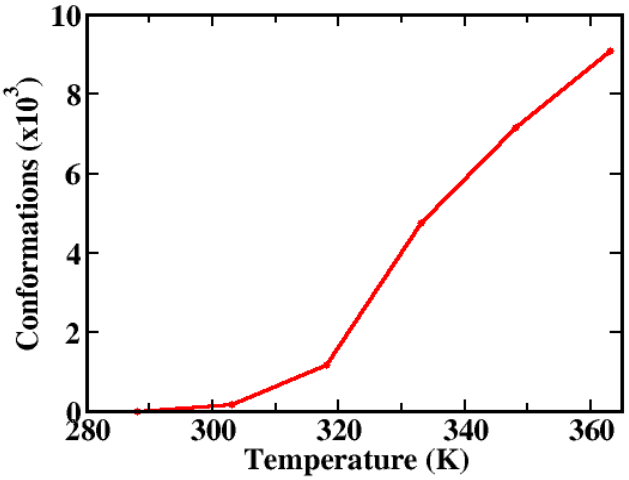


Figure 6.1: Distribution of the number of conformations with temperature.

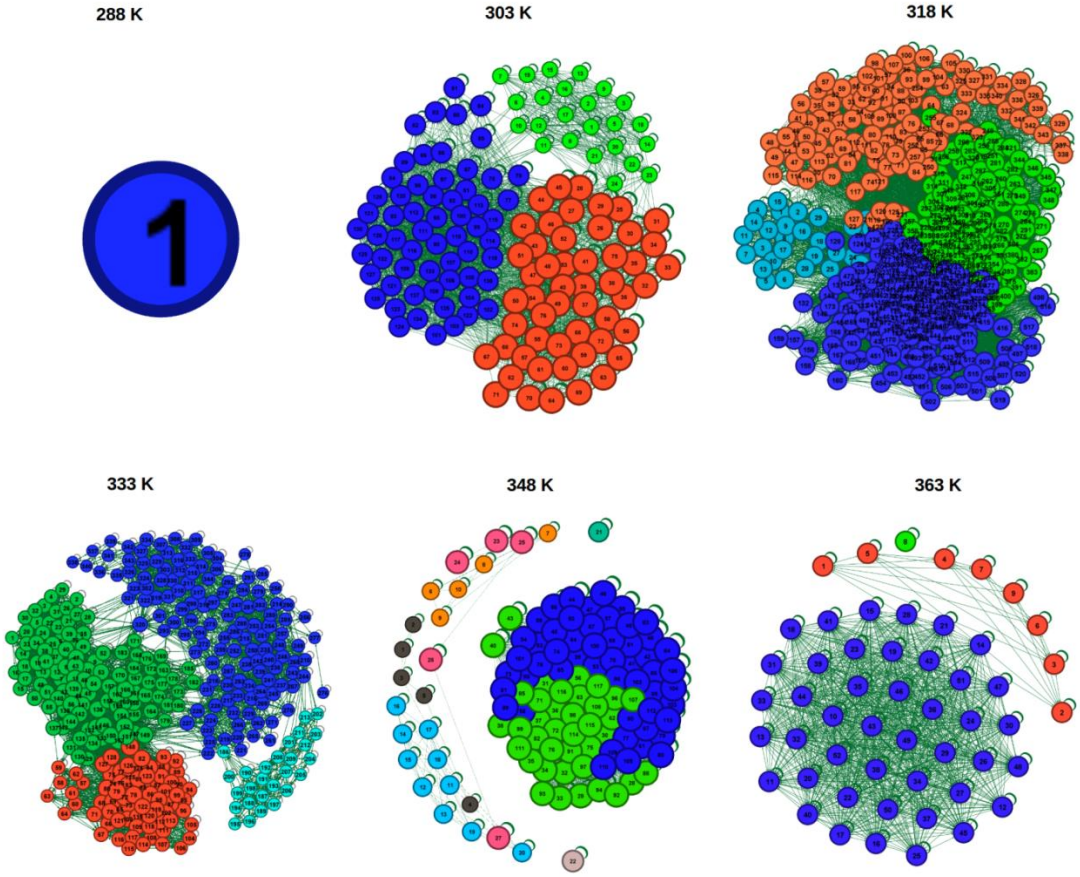


Figure 6.2: Network distributions of most prominent conformations at different temperatures (d(t, t') < 7.5 Å).

In this process, the presence of conformations that existed for smaller time scales and also for longer time scales were observed. To ensure the selection of

diverse conformations for the calculations, the most prominent conformations from each temperature were identified based on the frequency of their existence in the whole trajectory, and network analysis was performed. The conformations that existed for ≥120 ps were considered to determine the interconnectivity between the conformations. Two conformations are considered to be connected if d (t, t') < 7.5 Å. This criterion was chosen because the PDB ID taken has a resolution of around 6.9 Å. Therefore, it is better not to choose a cut-off value much lesser than 6.9 Å. The network analysis for both d(t, t') < 7.5 Å and < 6.5 Å were performed.

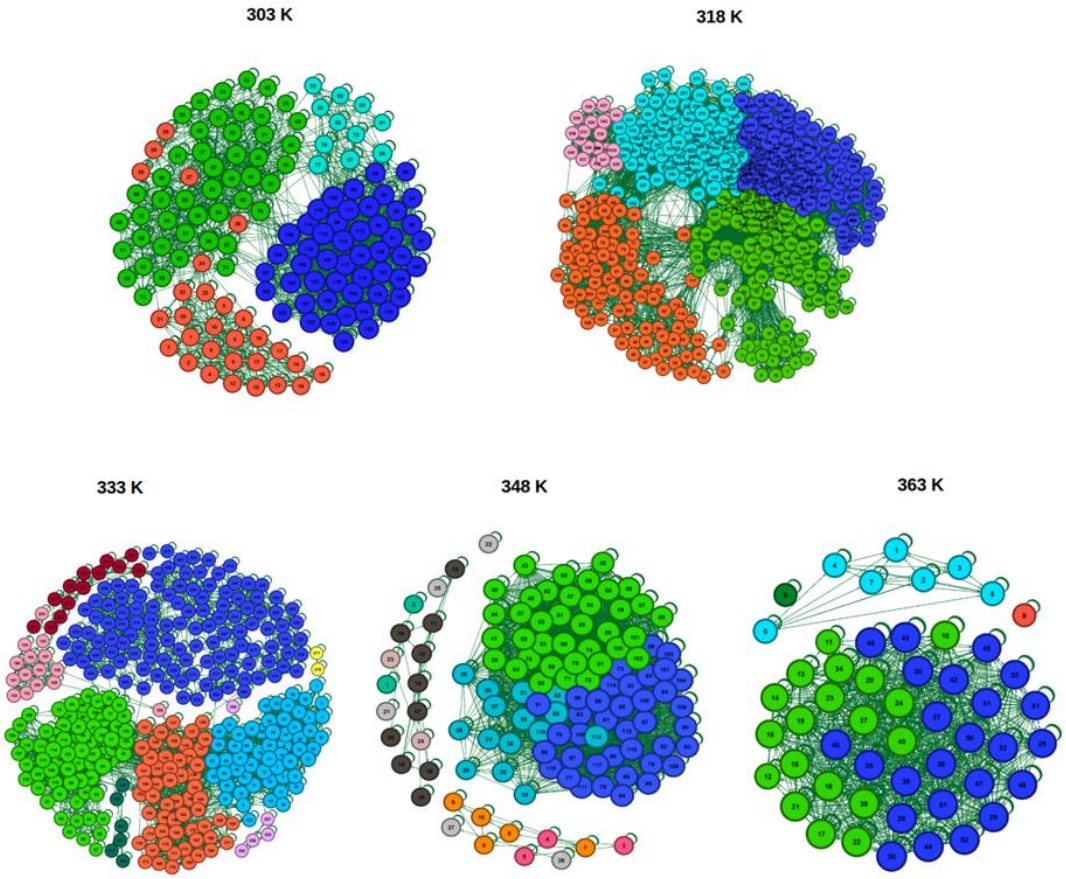


Figure 6.3: Network distribution of most prominent conformations at different temperatures (cut-off value: 6.5 Å).

The representations plotted using d(t, t') < 7.5 Å are shown in **Figure 6.2** which gives us the information about the evolution of the structures and their connectivity. Different colours represent different communities. At lower temperature, 288 K shows only one community which indicates that the conformation of the RNA molecule has not changed considerably in contrast to the higher

temperatures. With rise in temperature up to 348 K, there is an increase in the number of communities, suggesting different conformations. Change in the d(t, t') value to 6.5 Å did not change the trend of the analysis (**Figure 6.3**). This can be explained based on the following observations. We have seen in literature that, for 56nt (nucleotide) RNA pseudoknot, the RMSD between predicted & native structure is 7.4 Å (Shi et al. 2018) and for SARS Virus pseudoknot of 47nt is 7.4 Å (Hajdin et al. 2010). The standard deviations predicted for these calculations are found to be 1.8 ± 0.3 Å (Hajdin et al. 2010). Thus it can be concluded that the results obtained from changing the values of d (t_1, t_2) from 7.5 Å to 6.5 Å, does not affect the trend of analysis. We can see that the blue community obtained using d (t_1, t_2) = 7.5 Å is similar to the total of blue and aqua communities obtained using d (t_1, t_2) = 6.5 Å. As per the standard deviation, these two can be considered as similar. This is applicable to other cases too. Therefore, d(t, t') < 7.5 Å were fixed as the criteria. Selecting a much higher cut-off value reduces the number of different conformations and masking of finer detail information.

The separation between the communities gives information about the different conformations. These conformations are structurally different and can have the same or different energies. The free energy of the highest connected conformation corresponding to a particular community is given in **Table 6.2** and is calculated by $\Delta G = -k_B T ln(P_{rmsd})$ where P_{rmsd} is the probability distribution (Maisuradze et al. 2010). For instance, the cyan and blue communities at 318 K have similar ΔG values, meaning two very different conformations have similar energies. If a community is densely populated and well connected, it shows that conformation existed there is stable and will have lower energy. Scarcely populated and scattered communities at 348 K and 363 K (say, orange) have higher ΔG values, showing unstable conformations. Two overlapping communities can have similar ΔG values, as seen at 318 K and 348 K for cyan-blue-green and blue-green communities respectively. At 348 K and 363 K, there is a clear separation between the outer ring communities and inner ring communities. This suggests that there is a considerable change in the conformation of RNA at higher temperature, which is very much different from the initial reference structure. The RNA has formed some stable structures which cannot

revert to the initial structures. The blue community in this temperature is found to be stable during the simulation.

Table 6.2: Network analysis of conformations with respect to temperature.

Temperature (K)	SPC/E model		mTIP3P water model	
	Conformation ID	Energy (kcal/mol)	Conformation ID	Energy (kcal/mol)
288	1 (blue)	-1.71	1 (blue)	-2.570
303	24 (green) 47 (orange) 114 (blue)	1.22 -2.45 -2.43	75 (green) 194 (blue) 352 (cyan) 408 (red) 455 (magenta)	2.451 -1.728 -0.907 -2.178 1.810
318	179 (cyan) 184 (blue) 230 (orange) 243 (green)	-3.72 -3.70 -1.08 -3.30	3 (cyan) 53 (blue) 124 (red) 330 (green)	2.042 -2.590 -1.835 -1.827
333	128 (orange) 133 (green) 207 (cyan) 247 (blue)	-2.12 -2.84 -0.61 -2.63	8 (cyan) 25 (magenta) 35 (red) 80 (green) 164 (blue)	-1.209 0.084 -0.734 -2.235 -3.148
348	3 (brown) 10 (orange) 18 (cyan) 26 (pink) 56 (green) 86 (blue)	-0.78 3.26 -1.47 -0.78 -1.01 -1.13	28 (grey) 43 (blue) 61 (pink) 94 (green) 127 (magenta) 152 (cyan) 189 (yellow) 196 (red) 225 (orange)	-0.862 0.616 -0.023 -0.023 0.076 -0.008 -0.096 -1.435 2.052

363	2 (orange) 8 (green) 43 (blue)	9.94 3.62 -2.37	6 (magenta) 43 (cyan) 67 (red) 170 (green) 263 (blue)	-3.36 -4.171 -1.400 -2.443 -4.180

For the mTIP3P water model, we observed an increase in the number of conformations with the increase in the temperature up to 348 K. At 363 K, there is a decrease in the number of conformations. The conformation of the RNA was found to disrupt completely in this temperature and not much change in the structure was observed. The network distribution for the mTIP3P water model is shown in **Figure 6.4**. Overall, there is an increase in the number of conformations in each community compared to the SPC/E model. The free energy of prominent structures corresponding to each community is in **Table 6.2**. In the mTIP3P model, it is observed that the denaturation process of RNA structure is initiated from lower temperatures as low as 303 K.

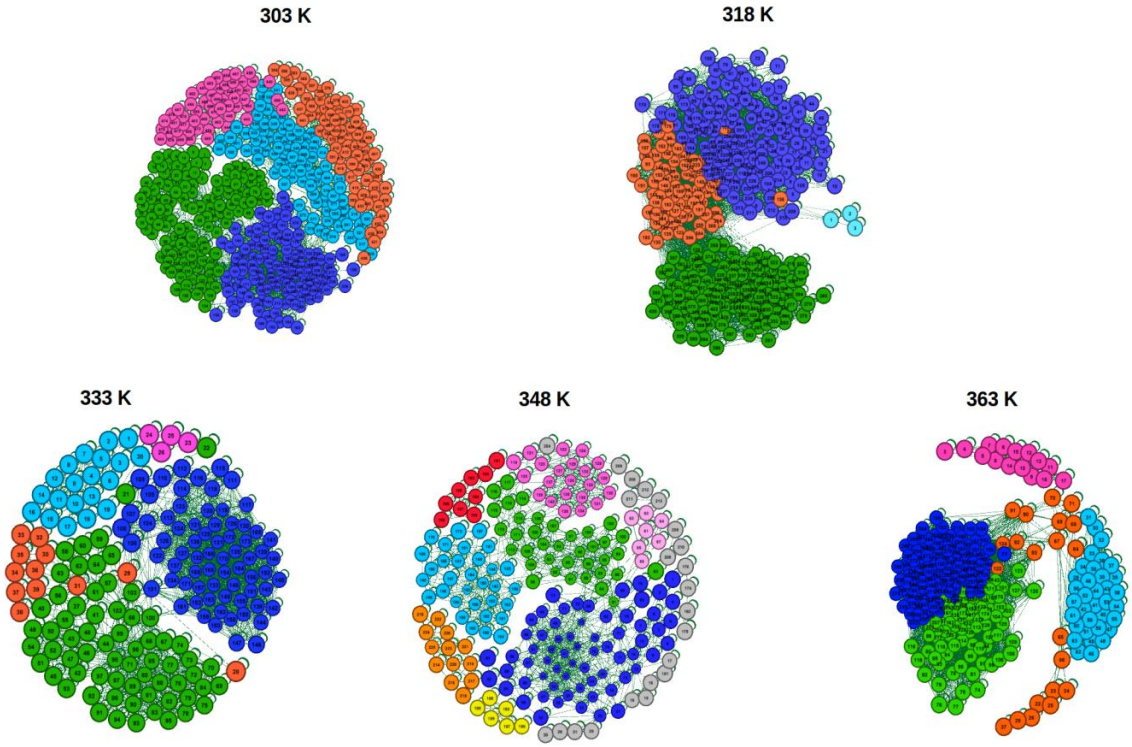


Figure 6.4: Network distribution of most prominent conformations at different temperatures in the mTIP3P water model. (cut-off value: 7.5 Å).

6.4 INTERHELICAL ANGLE METHODOLOGY

RNA helices were identified around the junction as H1, H2, and H3 as shown in **Figure 6.5**. We identified three angles around the junction. It is found that the 76th base is present at the center of the molecule/junction, and thus the phosphate atom of this base was chosen as the vertex point (O). Similarly, the phosphate atoms of the 58th (P), 8th (Q), and 86th (R) bases, which are the most distant bases of H1, H2, and H3 stems respectively, were taken as endpoints. The angle between these points was calculated as shown in **Figure 6.5**. For all conformations extracted, these three interhelical angles were thus calculated: ∠A = ∠POQ, ∠B = ∠QOR, and ∠C = ∠POR.

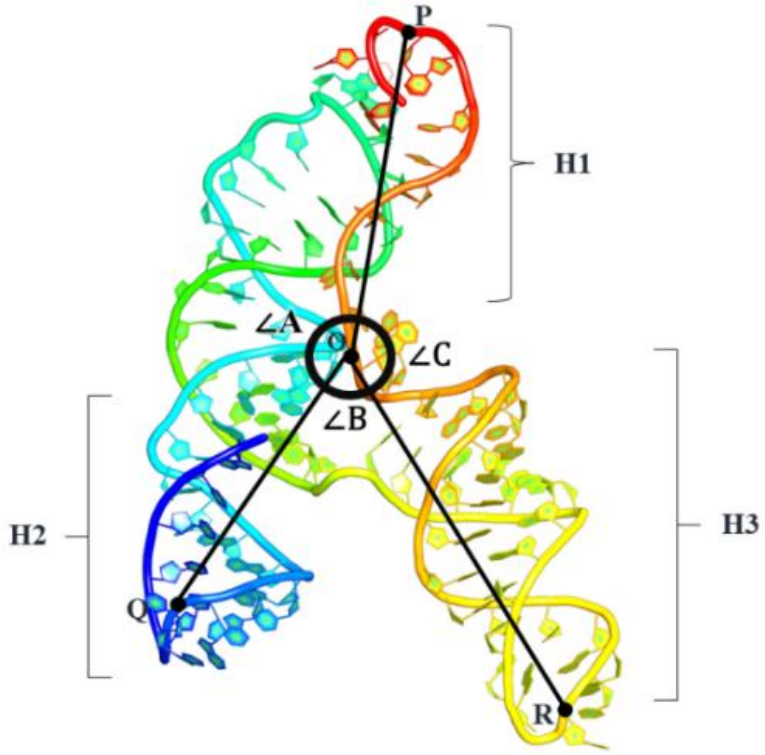


Figure 6.5: The angles ∠A, ∠B, and ∠C between three stems- H1, H2, and H3 are identified around the junction point of the SARS-CoV-2 RNA Genome.

Next, a network was constructed between the conformations with the help of nodes and edges. A particular conformation is denoted by a node. Two conformations are said to be related or connected if the difference between all the three respective angles of two conformations or nodes is less than 10°. We define ΔA, ΔB, and ΔC as

$\Delta A_{ij} = \angle A_i - \angle A_j$, $\Delta B_{ij} = \angle B_i - \angle B_j$ and $\Delta C_{ij} = \angle C_i - \angle C_j$ where i and j are the different conformations.

The structural evolution of the RNA based on the change in the ∠A, ∠B, and ∠C, thus it would be interesting to see their distribution with the temperature. In **Figure 6.6**, the probability distribution of the angles is plotted. The angle distribution of ∠A and ∠C mainly spread from 75° to 180°; however, ∠B is between 0-180°. This indicates that there are a greater number of relative orientations of the stem near the junction especially between H2- H3 stems. The angle distribution of ∠B and ∠C becomes broader with the increase in the temperature up to 348 K. For ∠A, however, we do not find any specific trend. Maximum variation is noticed in the case of 348 K. The distribution for 363 K is found to be narrower than 348 K, showing a more compact and less random structure in comparison to 348 K. In the case of 363 K, the RNA seems to have obtained a conformation where the H2 and H3 stems are closer together as compared to the native structure (**Figure 6.6(b)**), and is maintained as such for a longer period.

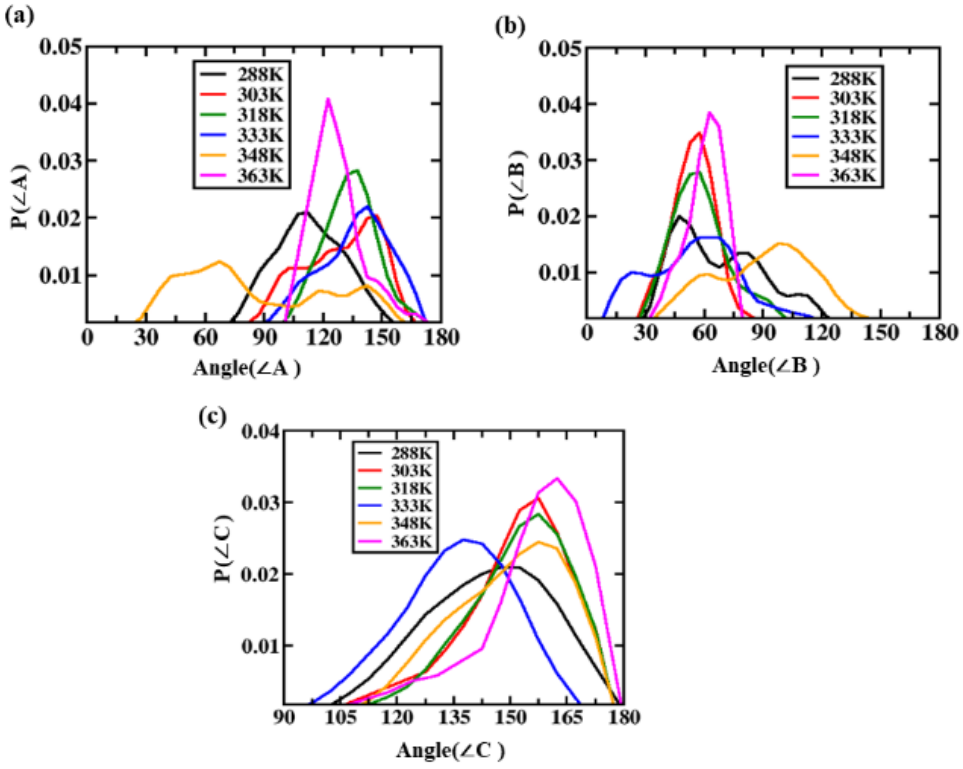


Figure 6.6: Distribution of angles **(a)** ∠A, **(b)** ∠B and **(c)** ∠C at six different temperatures.

If ΔA, ΔB, ΔC are less than 10°, two conformations are joined by edges, but if any one of the angle differences is more than 10°, the conformations remain unconnected. In this way, adjacency lists for each temperature were made using Python script. Four ranges of cutoff with angle differences of ±5°, ±10°, ±20°, and ±30° were tested to set the edge criteria. The Force Atlas layout (Jacomy et al. 2014) was implemented to visualize the networks. It is a force-directed layout that simulates a physical system in order to spatialize a network. In this algorithm, connected nodes tend to aggregate together and the ones not connected by edges are drawn apart. These forces create a movement that converges to a balanced state. The final layout of nodes helps in the interpretation of the data. In **Figure 6.7**, the network distribution for the conformations by using different angle difference cutoffs at 303 K is shown.

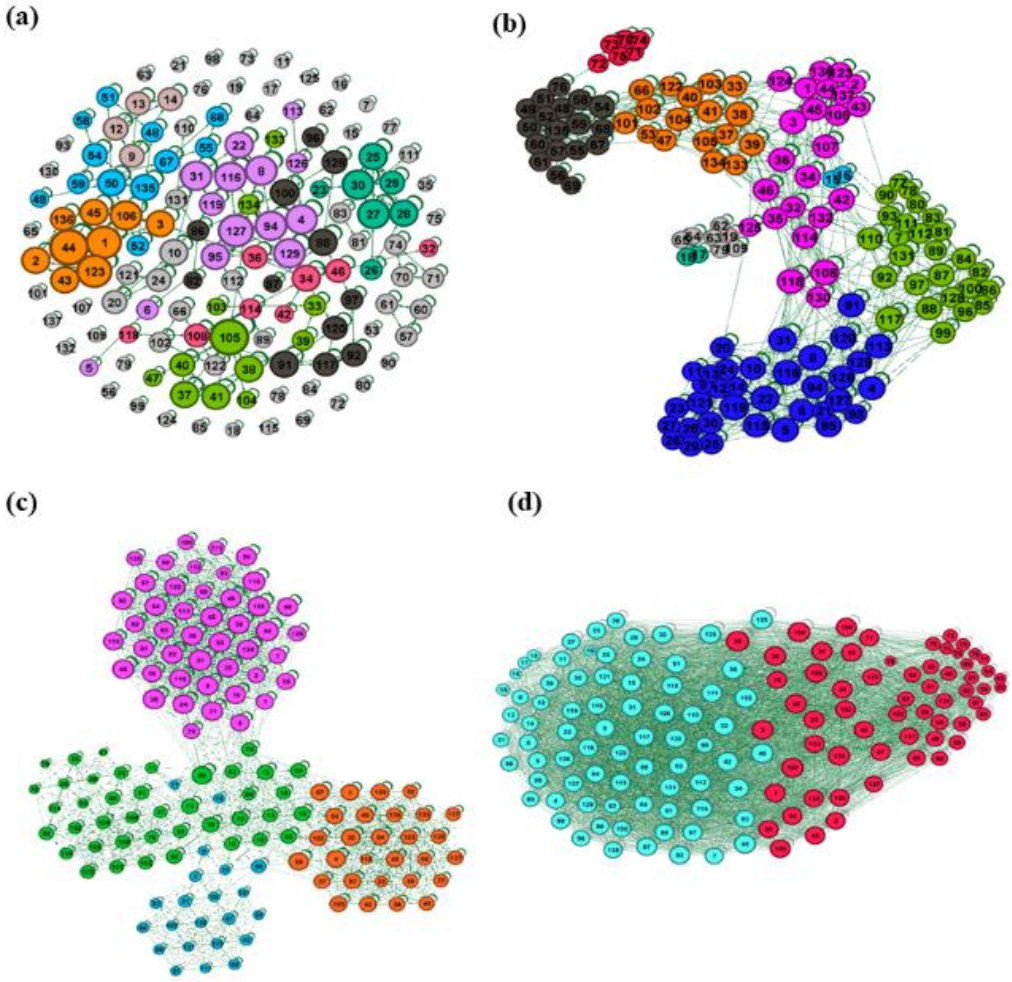


Figure 6.7: Network distribution of SARS-CoV-2 RNA FSE conformations with angle difference ΔA, ΔB, ΔC cut-off **(a)** ±5°, **(b)** ±10°, **(c)** ±20° and **(d)** ±30° at 303 K.

As it is evident from **Figure 6.7** at ±5°, many communities are found to be formed with very few edge connections and the communities are not well segregated. As we increase the angle cut off to ±30°, the total number of communities decreases, and the number of nodes in a community increases. The number of communities present at 303 K with cutoffs ±5°, ±10°, ±20°, and ±30° is respectively 50, 13, 4, and 2. Two broad and distinct communities were observed in the case of ±30° cutoff, which masked the smaller communities. The value of the module class percentage can be chosen as a parameter for selecting a good angle criterion. It describes the percentage of nodes (or configurations) present in a particular community. In **Appendix XIX** the module class percentages of each community at a different angle cutoff are shown.

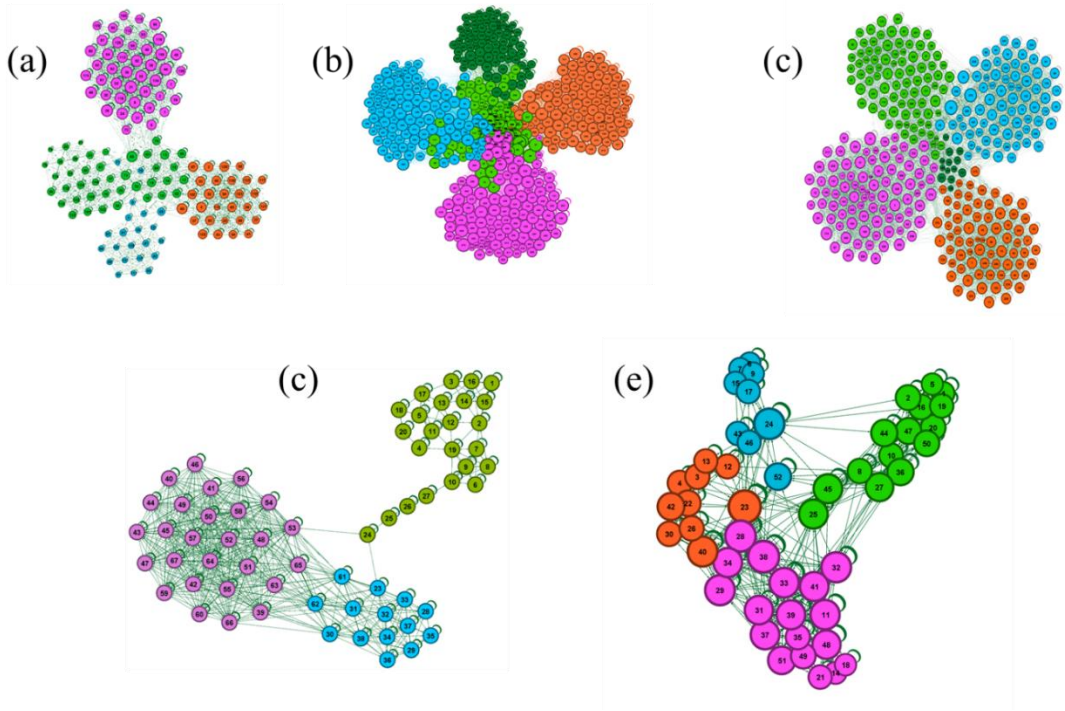


Figure 6.8: Network distribution of conformations with ±20° as angle cutoff at varioustemperatures: (a) 288 K, (b) 303 K, (c) 318 K, (d) 333 K, (e) 348 K, and (f) 363 K.

It can be seen that when the number of communities is high, the class percentages decrease, which is apparent in the case of a ±5° cut-off. A network with too many communities results in insignificant differences between the communities. It

is, therefore, recommended that at least two major communities present in the network distribution should have modularity above 10% which ensures sampling of the communities of considerable differences. The value of the class percentage of the largest community at 303 K at ±30° cut-off is 55% which is quite high. It is advisable that if two or more module classes show a class percentage higher than 30%, then a lower cut-off value should be considered. We did not consider ±5° and ±30 ° for conformation sampling since the module class percentages are not ideal, and will not yield proper information. Angle cut off of ±10°and ±20° shows better classification. In the case of ±20° cutoff, 4 major communities were found with a considerable module class percentage in all temperatures except 318 K, which shows that there is a less observable difference between the total number of communities (**Figure 6.8**). Therefore, we selected ±10° cutoff as our angle criteria and will be presenting the results corresponding to this cutoff (**Figure 6.9 and Appendix XX**).

From **Figure 6.9**, it can be seen that at 288 K, the number of communities was found to be 1 which increases to 6 visible major communities at 303 K, 7 at 318 K, 8 at 333 K, 9 at 348 K, and 5 at 363 K. The number of major communities increases with temperature and the maximum is seen at 348 K. It can be mentioned here that when we took RMSD as our criteria for sampling the RNA conformations, the number of communities also increased with the temperature, and the maximum is seen in case of 348 K (Singh et al. 2021). The distinct separated communities at 303 K show that the conformations are quite different from each other. The molecule is taking on new and diverse conformations. At 318 K, the communities are quite interlinked with each other which spread apart at 333 K, where the more connected communities are joined by the smaller intermediate conformations. At 348 K and 363 K, we however find discrete communities. The conformations are interlinked with each other at 318 K and 333 K; which however is less at 348 K and 363 K. At 363 K, the conformations of the RNA are mainly interlinked between 3 communities that together form a clique and separate from the smaller ones, which is again indicative of a particular angle range that is more prevalent. At higher temperature, the communities are found to be more disconnected.

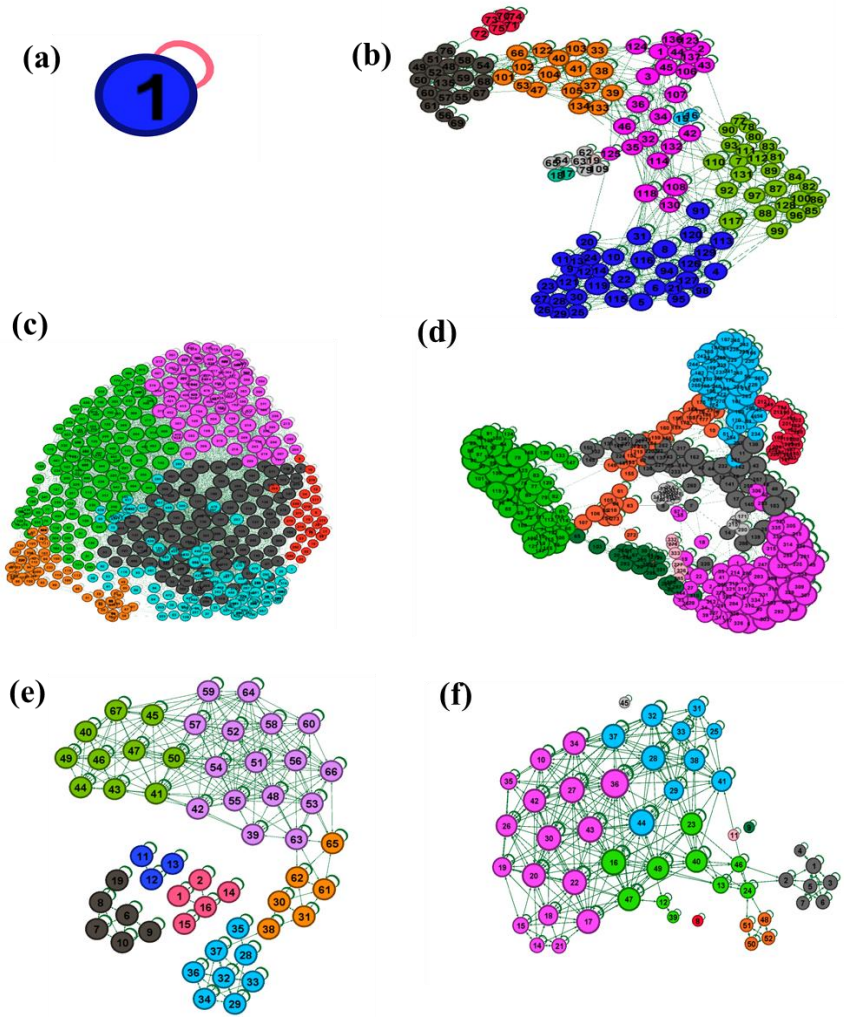


Figure 6.9: Network distribution of conformations with ±10° as angle cutoff at various temperatures: **(a)** 288 K, **(b)** 303 K, **(c)** 318 K, **(d)** 333 K, **(e)** 348 K, and **(f)** 363 K.

6.5 CHANGES IN MINOR AND MAJOR GROOVES

The RNA structural variation in each community with respect to temperature can be distinguished by base pair twists and the width of minor-major grooves. The common RNA A-form helix is characterized by a narrow major groove and a wider minor groove (Frugier and Schimmel 1997). The groove widths are measured from phosphate to phosphate, including the diameter of phosphate groups (El Hassan and Calladine 1998). It is evident from **Figure 6.10** that the representative conformations from each community are structurally different based on base pair groove width. It is

found that the number of identified base pairs varies within the communities for each temperature.

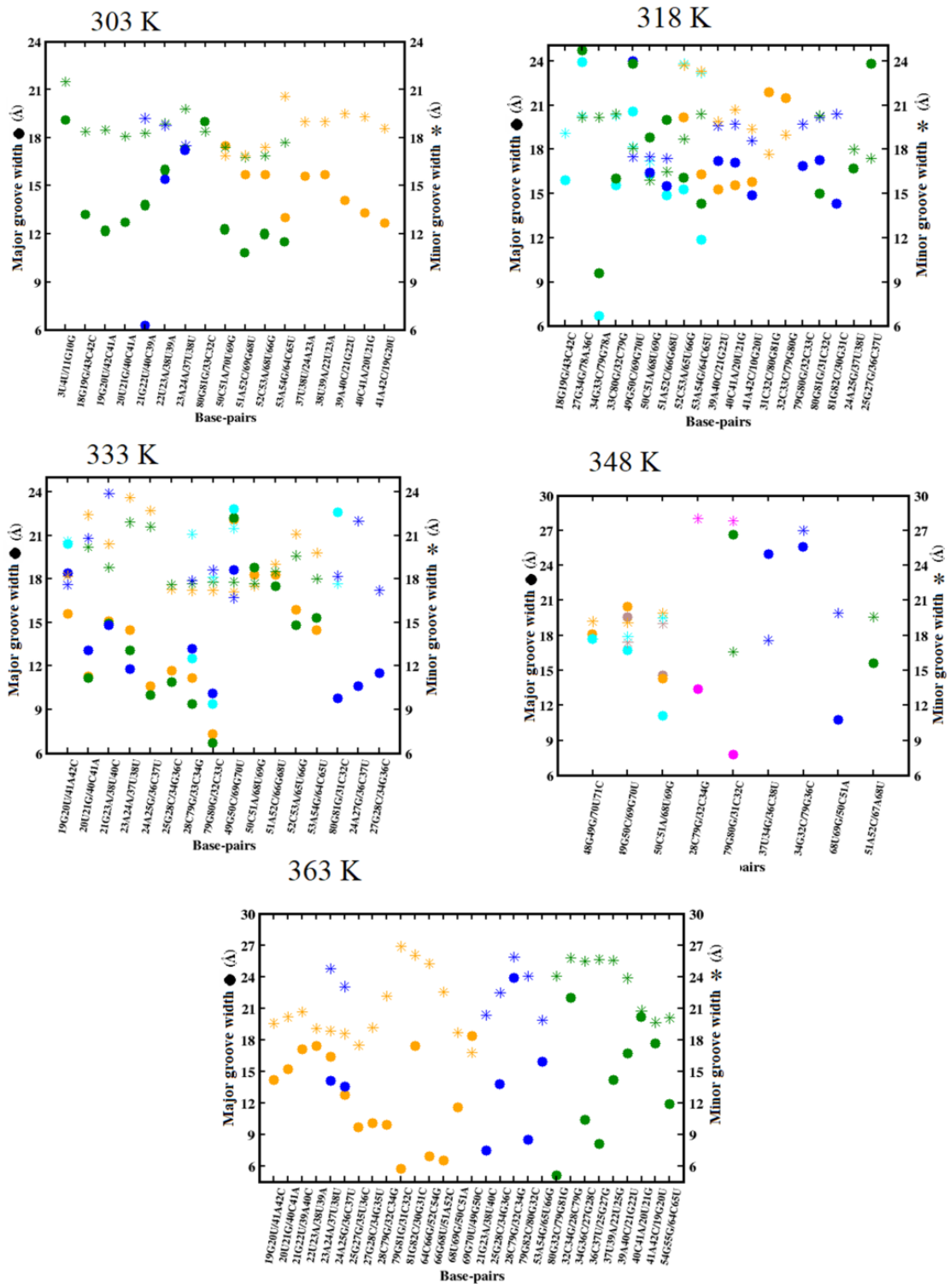
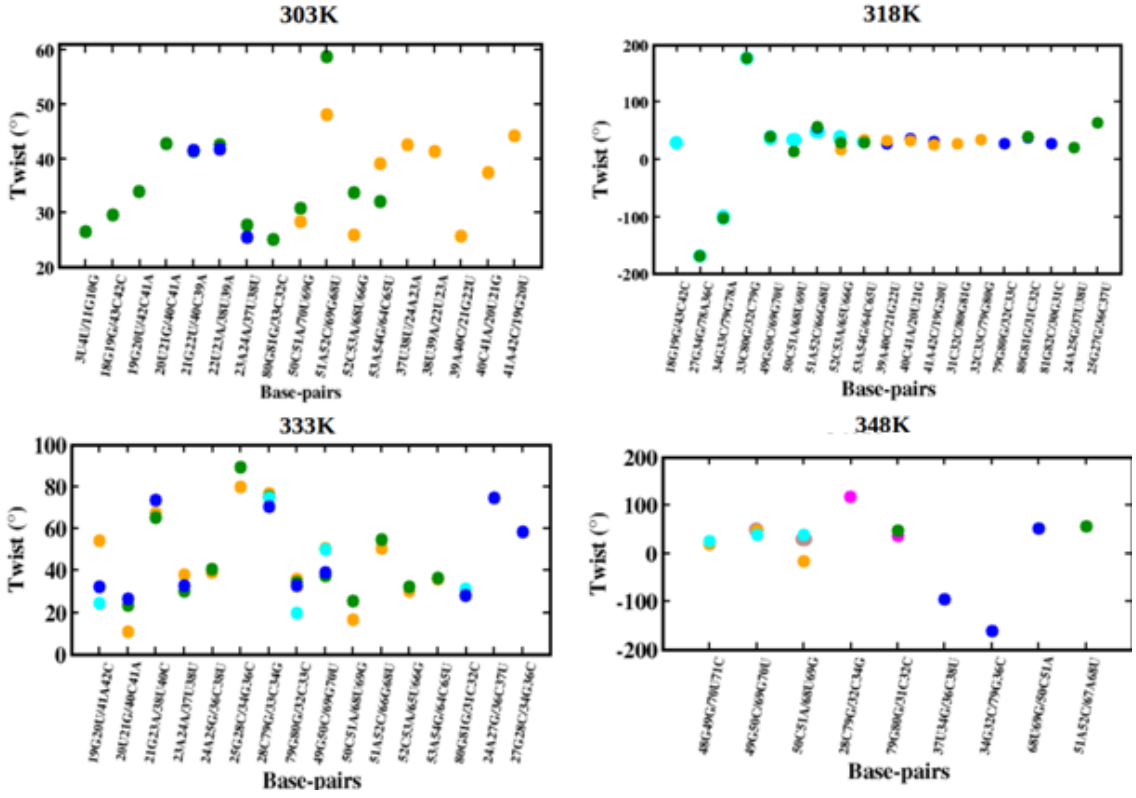


Figure 6.10: Scatter plot of groove width vs. base pairs for 5 different temperatures in SPC/E water model.

At 303 K, the three communities identified have different structures based on the groove width. At 318 K and 333 K, cyan community has a different structure while orange-green-blue communities have overlapping base pairs with variation in minor-major grooves. It can be seen that the major groove is wider for 318 K than 333 K indicating a structural change from common RNA A-form helix. For higher temperatures, 348 K and 363 K, the outer and inner ring communities have different minor and major groove positions, indicating a structural evolution. This is further supported by the twist angle, which measures the rotation of base pairs along the helical axis. The base pair twist angles are calculated and averaged with error based on the presence of minor-major grooves in the RNA structure. The standard twist angle for A-form of RNA is 32˚ ± 8 which is retained in lower temperatures, 288 K and 303 K. At higher temperatures, the twist angle is either more twisted or less twisted for the communities with large deviations in the values suggesting the evolution of conformations different from the A-form (**Figure 6.11**). The scatter plots show that the network distribution has effectively distinguished the communities having structural variation.



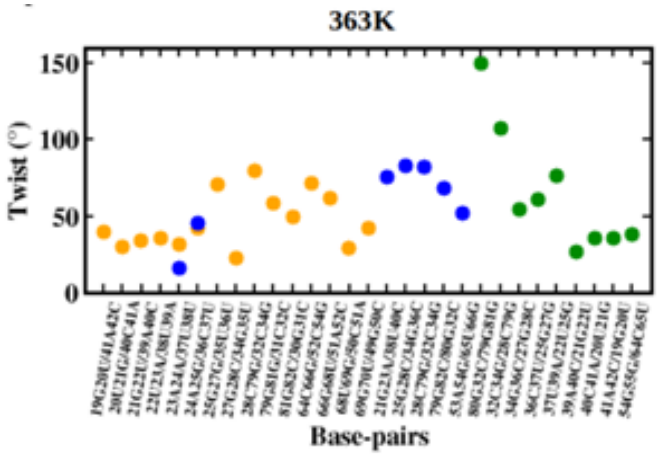


Figure 6.11: Distribution of base pair twist angle with respect to the presence of minor-major grooves in RNA structure.

For the mTIP3P water model, we observed an increase in the number of conformations with the increase in the temperature up to 348 K. At 363 K, there is a decrease in the number of conformations. The conformation of the RNA was found to disrupt completely in this temperature and not much change in the structure was observed. The network distribution for the mTIP3P water model is shown in Figure S6. Overall, there is an increase in the number of conformations in each community compared to the SPC/E model. The conformations from each community have different structures based on twist angle and minor-major groove distance as evident from **Appendix XXI**. In the mTIP3P model, it is observed that the denaturation process of RNA structure is initiated from lower temperatures as low as 303 K. At almost every temperature we found the structure is denatured (the helix structure is destroyed) at a longer simulation run which is evident from the major-minor groove distance. Experimentally, it is found that SARS coronavirus gets denatured and converted to be non-infectious above 330 K (Duan et al. 2003). Even though, SARS CoV-2 is genetically distant from SARS coronavirus, SARS CoV-2 shares 79% genomic identity with SARS CoV (Lu et al. 2020). Therefore, we focussed our calculations mainly on the SPC/E model in the following sections. The difference in the RNA structure is noticed due to the presence of different parameters in the SPC/E and mTIP3P water models.

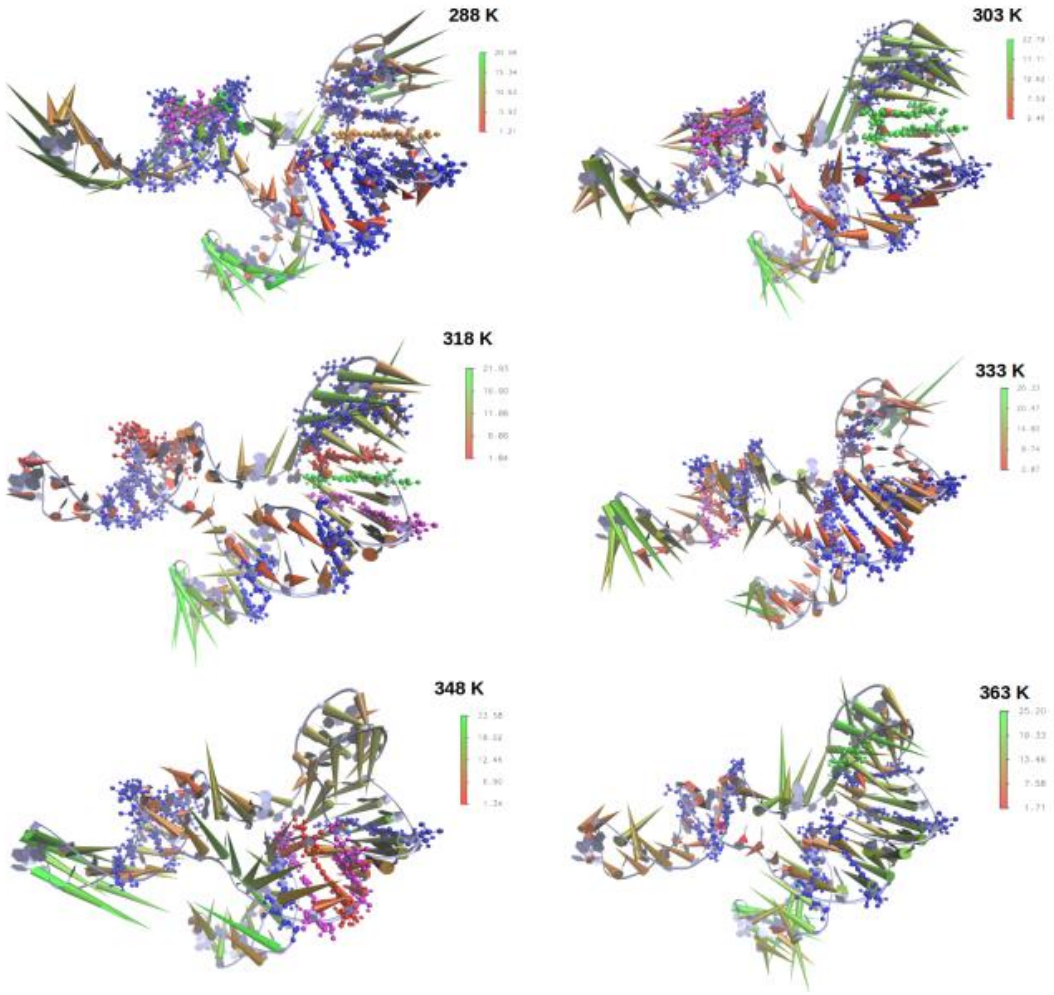


Figure 6.12: Porcupine plot showing the degree of fluctuation during the course of simulation.

Next, the thermal fluctuations in the RNA structures which lead to these conformational changes were identified (Haider et al. 2008). In **Figure 6.12**, the degree of magnitude and direction of the fluctuations in RNA by porcupine plots with respect to atom C1' is shown. The highlighted base pairs represent the variation of the native closed base pair distribution in the final conformation of RNA with temperature. Any change in the color of the base pairs with respect to the 288 K structure is due to a change in the distance between the base pairs. Lesser fluctuations and more native closed base pair at 288 K and 303 K indicate that native characteristics of RNA are preserved at this temperature. At 288 K, most of the closed base pairs (80%) lie in lower energy region-I (blue); whereas we can see the

emergence of different colors with the rise of the temperature. At higher temperatures (333 K and above) the number of the closed native base pairs is found to be reduced drastically due to higher fluctuations resulting to the breaking of bonds between the base pairs and the opening of stem-2 pseudoknot. The existence of the non-native base pair is seen at higher temperatures leading to partial stabilization (**Figure 6.13**).

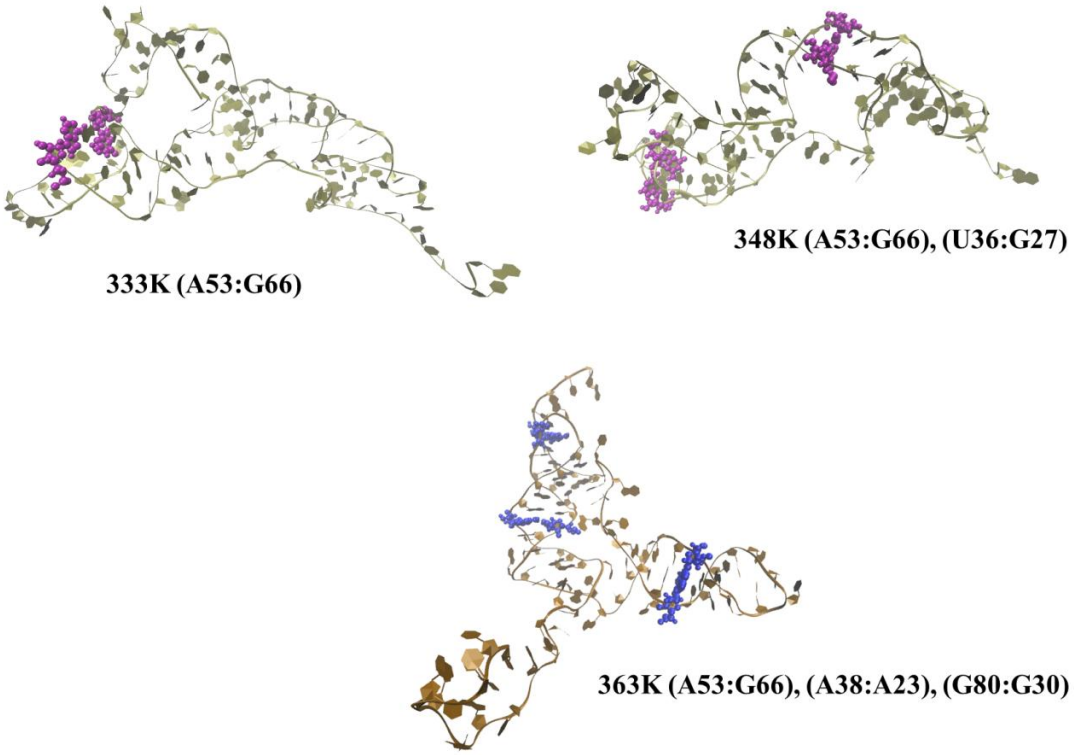


Figure 6.13: Mispairing/ non-native base pairs formed at **A)** 333 K **B)** 348 K and **C)** 363 K.

The distance criteria for closed base pairs can be further confirmed by the free energy surface plotted using N1-N3 and C1'-C1' distances of the base pairs in RNA as the reaction coordinates (**Figure 6.14**). The native structure of RNA and the final conformations of each temperature are considered in the calculation. The minimum energy region corresponds to the closed base pairs in RNA are divided into regions I to IV. It can be seen that at lower temperatures, 288 K (black dots) and 303 K (red dots), the base pairs lie mostly in the allowed region. At higher temperatures, such as 348 K (orange dots) and 363 K (magenta dots), the base pairs are scattered and mostly

lie outside the allowed region. This shows that the characteristic distance corresponding to a closed base pair corresponds to some specific N1-N3 and C1'-C1' distance ranges irrespective of temperature. However, the open base pair can have any distance range outside the allowed region.

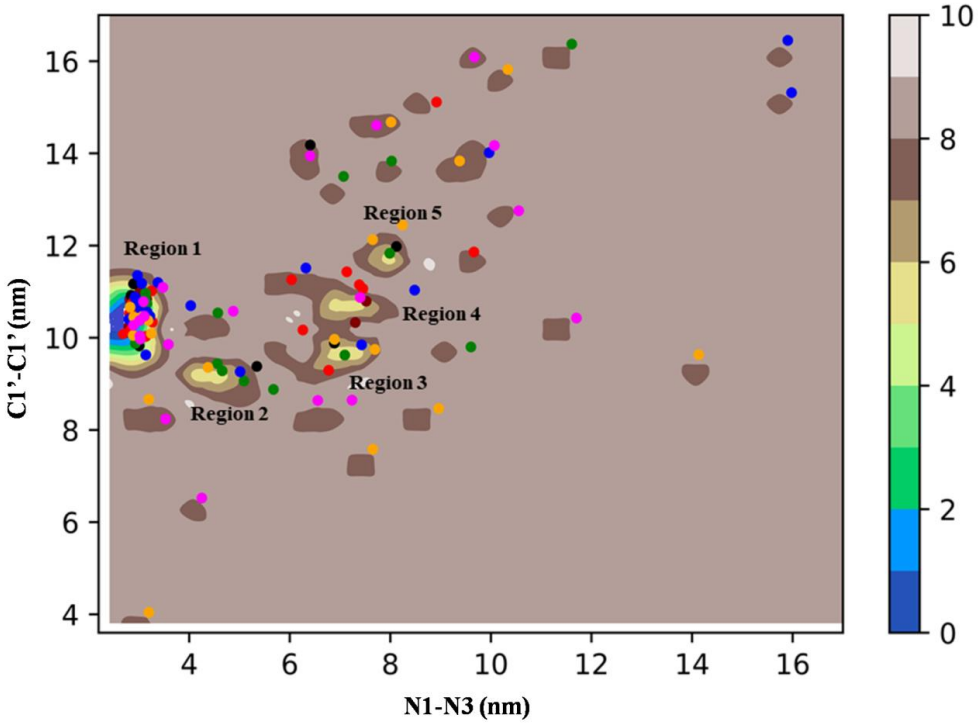


Figure 6.14: Free energy surface plotted using N1-N3 and C1'-C1' distance (nm) as reaction coordinates. The blue to yellow contour regions corresponds to the closed base pairs present in RNA which can be divided into region I to V. The base pair distance from the final structure of RNA for all the temperatures was plotted in different color dots to check the presence of native, non-native, and opened base pairs: native (maroon), 288 K (black), 303 K (red), 318 K (green), 333 K (blue), 348 K (orange) and 363 K (magenta).

6.6 SECONDARY STRUCTURE OF RNA

To study the variation in the secondary structure of the RNA during the simulation, the base RG76 that was taken as a vertex was used to calculate the chi angle and sugar angles at all temperatures (**Appendix XXII**). Chi angle (χ) is the backbone torsion angle between the nucleoside sugar and nitrogenous bases and sugar angles (V_0, V_1, V_2, V_3, V_4) are torsion angles about the 5 bonds of the ribose ring. The native RNA

structure was found to have χ = -110.56° and sugar angles as V_0=13.28°, V_1=-31.42°, V_2=-23.35°, V_3=-19.18°, V_4=-6.64°. We have plotted these distributions at different temperatures in **Figure 6.15.** It can be seen that the distribution of the χ angle for all temperatures is spread from -100° to -180°, corresponding to the *anti*-conformation, and in 333 K and 348 K χ angle is confined between -110° to -130° (**Figure 6.15(a)**). The sugar angles also show the same trend as seen in **Figure 6.15(b-f)**, where 333 K and 348 K conformations remain confined to a smaller range of angle deviation as opposed to other temperatures which are spread apart. It can be noted here that at these temperatures we observed the interlinking of the angles and rapid changes in the conformation were seen according to the conformational dynamics.

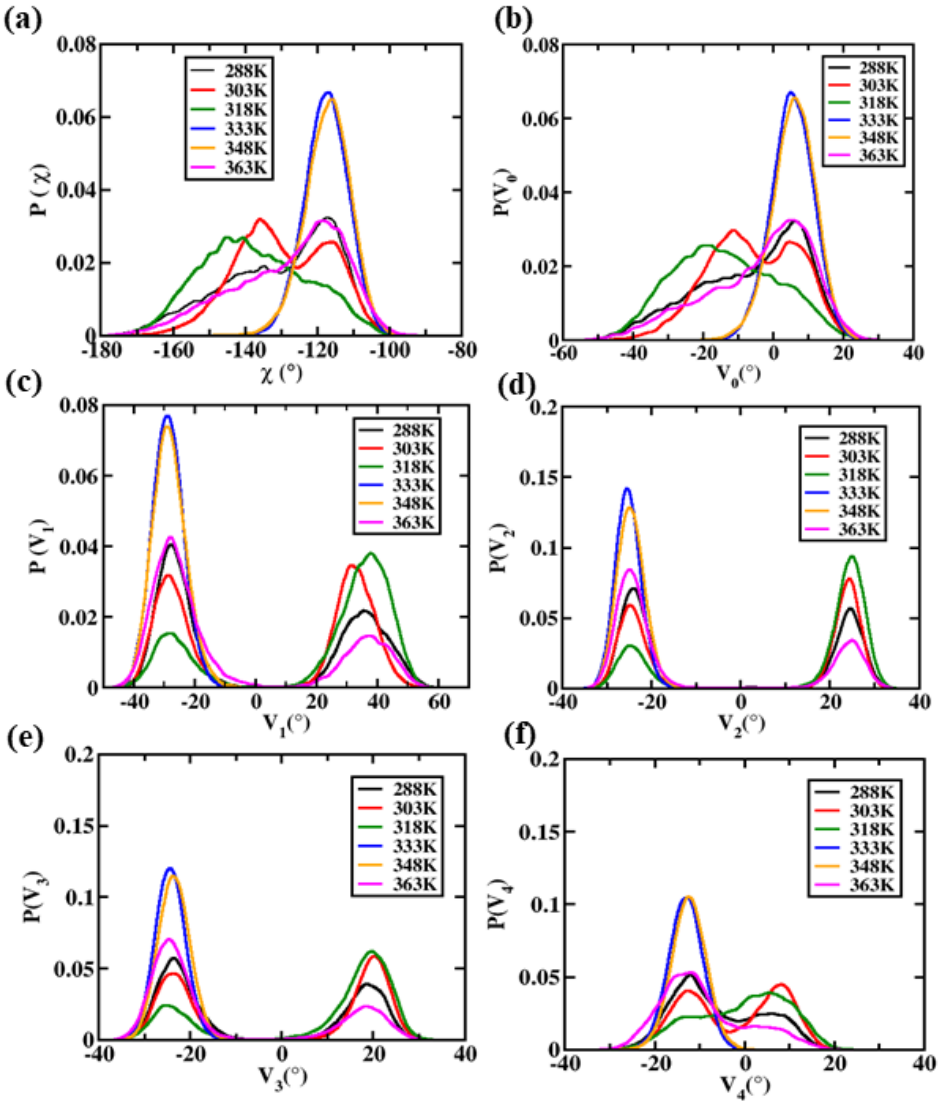


Figure 6.15. Distribution of **(a)** χ angle and sugar angles **(b)** V_0, **(c)** V_1, **(d)** V_2, **(e)** V_3, **(f)** V_4 of base RG76 of RNA at six different temperatures.

6.7 NATIVE BASE-PAIR OPENING/ DENATURATION OF RNA

Since the communities are mainly distributed based on the structural parameters, therefore there is every possibility that two conformations can have the same or different energies. The structure of the RNA structure is very dynamic therefore, it is better to study the multiple energy states accessible to RNA at a particular temperature. The energy range where these states exist will give an overview of how these states are different at different temperatures. In **Figure 6.16(a)**, we have shown the energy distribution of the most prominent structures with temperature. The free energy of the most prominent structures is calculated using $\Delta G = -k_B T ln\left(\frac{(1-f_o)}{f_o}\right)$ where k_B is Boltzmann constant, T is temperature, and f_o is the fraction of open bases.

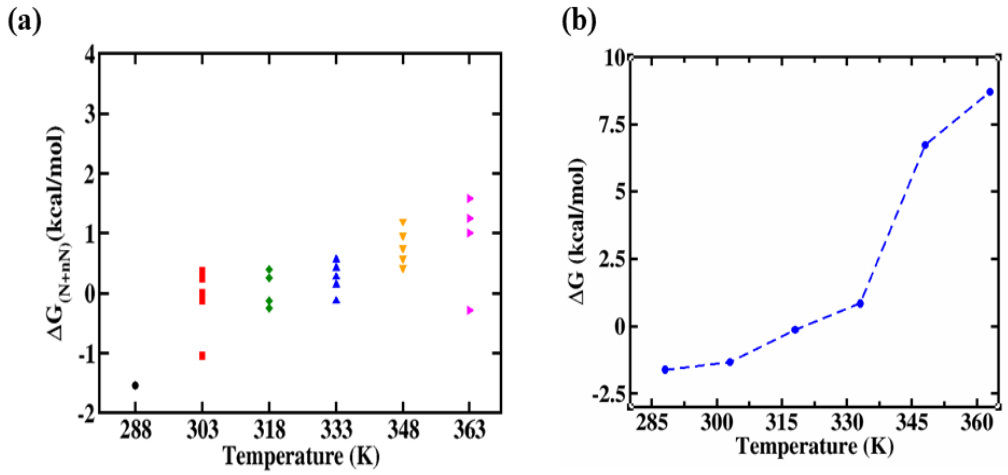


Figure 6.16: Scatter plot of free energy vs temperature **(a)** free energy of RNA from (native + non-native) base pairs for each community, **(b)** distribution of the total free energy with temperature.

The combined effects of native and non-native base pairing in RNA were considered to calculate the free energy for each community. Further, we calculated the total energy of the RNA for each temperature by taking the contribution from each community at a particular temperature using $\Delta G = \sum_i^n P_i E_i$ where, i = (1, 2, 3, …, n) community, P_i and E_i are the probability and energy of each community. **Figure 6.16 (b)** shows the distribution of the total free energy with temperature.

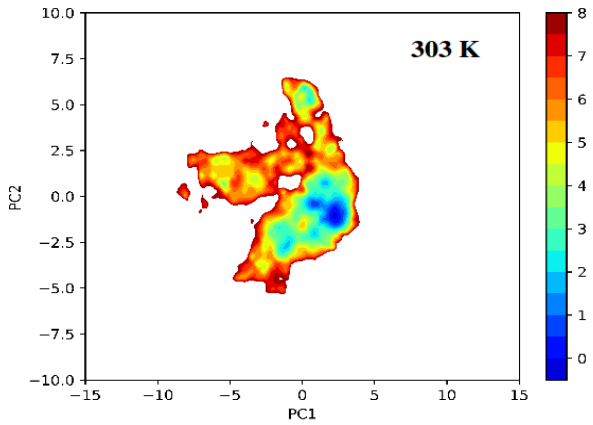
ΔG value calculated for 288 K is found to be -1.6 kcal/mol, while at 303 K, the most prominent structures have ΔG_{N+nN} value between the range -1.04 and 0.37 kcal/mol, which indicate that almost (50-60%) RNA structure is maintained its native structure at this temperature. At 318 K and 333 K, the ΔG_{N+nN} values are clustered between -0.27 to 0.37 and -0.11 to 0.57 kcal/mol respectively. As we go to higher temperatures, the free energy increases dramatically for 348 K and 363 K indicating that the RNA has been denatured considerably. A maximum of 95% native base pairs have opened up while 10-15% base pairs have formed newly at 348 K and 363 K respectively. The structure of the RNA has adapted new conformations which are prevalent at higher temperatures. At 363 K, the native conformation of RNA has refolded to other structures by forming non-native bonds which is evident from the energy of representative structures chosen from network distribution (**Figure 6.16 (a)**). Earlier, this kind of observation was found in the hyper stable RNA tetraloop under multiple temperature and pressure conditions (Miner et al. 2016). From **Figure 6.16 (b)**, the total free energy ΔG value at 303 K is found to be -1.32 kcal/mol which suggests that the structure is stable and is not denatured. It can be seen that the denaturation initiated from 318 K, where ΔG value is around 0. The total free energy value at 348 K and 363 K is found to be 6.75 and 8.71 kcal/mol respectively.

Therefore, ΔΔG value observed for RNA with 88 nt between 288 K and 363 K is found to be 10.31 kcal/mol. Experimentally, the denaturation energy of SARS CoV-NP (isolate BJ01) in urea and guanidine hydrochloride is found to be 3.6-7.2 kcal/mol and 4.1-7.8 kcal/mol respectively (Luo et al. 2004). Also, these values correlate well with the unfolding energy between folded and unfolded states of similar RNA structures such as *E.coli* α mRNA pseudoknot (~16 kcal/mol, 112 nt) determined by calorimetry/UV hyperchromicity (Gluick and Draper 1994) and human telomerase RNA telomerase (4.5-6.6 kcal/mol, 54 nt) determined by FRET analysis (Holmstrom and Nesbitt 2014; Theimer et al. 2003).

6.8 PRINCIPAL COMPONENT ANALYSIS

Principal component analysis (PCA) was carried out to determine the most important functional motions of the RNA leading to conformational evolution. PCA was

performed for C5' backbone atoms of RNA on the MD trajectory. The principal components, PC1 and PC2, which form the major motions, were taken as the reaction coordinates to construct the Free Energy Landscape (FEL). Both the components, PC1 and PC2, of evolved conformations showed high positive and negative correlations with the average structure for each temperature (**Figure 6.17**). The free energy minima represent the stable states of the system while the energy barriers indicate the presence of transient states. It can be seen that FEL constructed for each temperature shows a good correlation with the network distribution for prominent structures obtained from clustering analysis. The separation between the communities provides information about the conformations which are structurally different and can have the same or different energies. The interconnections between the communities are well defined in the free energy surface for each temperature. 303 K shows two well-defined lower and higher energy regions where the former corresponds to the inter-connected communities and the latter corresponds to the separated communities in network distribution. At 318 K, the presence of widely spread FEL supports the densely connected network seen in the network distribution. At the temperatures, 333 K and 348 K, the FEL is discrete and separated which shows the inter-conversion between the structures is unfavorable due to the high energy barrier between the conformations. At 363 K, the most prominent structures are connected to each other which corresponds to a low energy region in the FEL. The separated communities have higher energy and show a high negative correlation with the average structure.



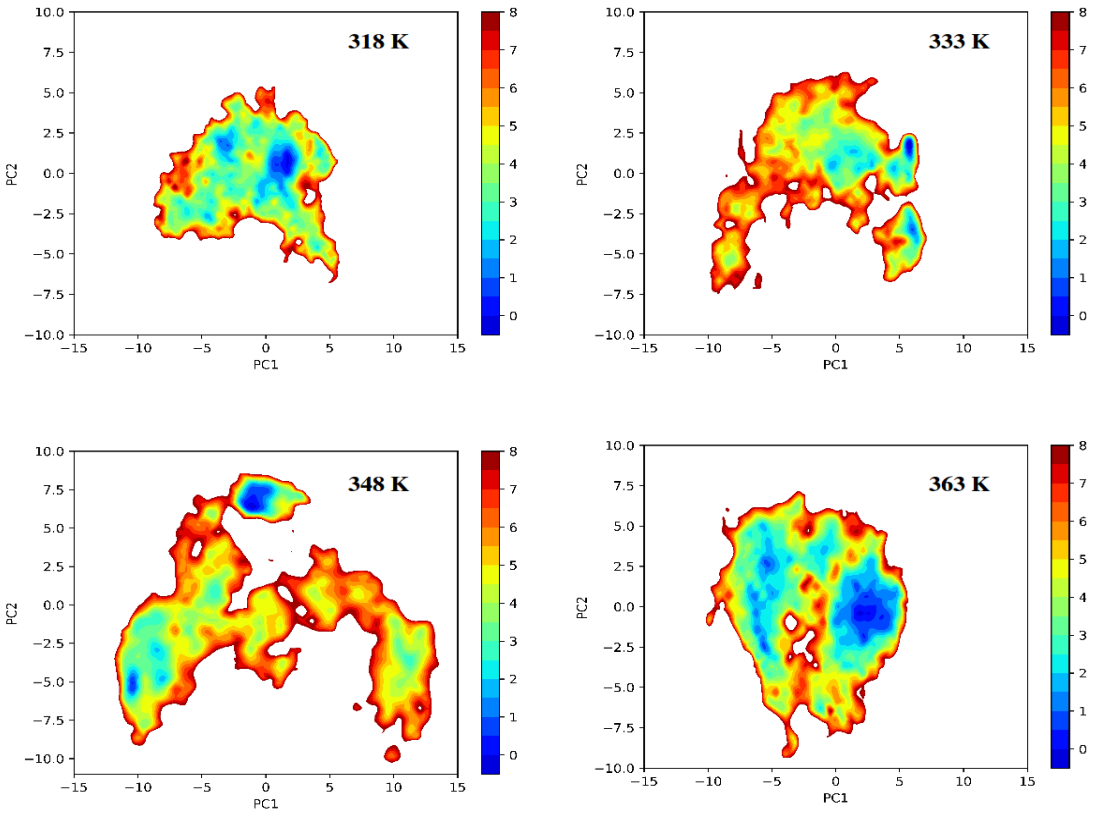


Figure 6.17: Gibbs free energy landscape for RNA at different temperatures. FEL obtained from MD trajectory for all the six systems using the reaction coordinates as the projection of C5' atoms onto the first two principal components.

6.9 HYDROGEN BOND LIFETIME OF BASE PAIRS

To see the stability of a conformation, we calculated the change in G: C base pair distance. Since these base pairs are present throughout the RNA, any change in their base pairing distance can give us information about the stability of a conformation. We focussed on the radial distribution function (RDF), g(r), of three intramolecular hydrogen bonds between guanine (G) and cytosine (C) namely, N_4H---O_6, N_1H_1---N_3 and N_2H---O_2. As it is evident from the figures the first peak mainly lies within 2.9 Å followed by the second peak around 3-4 Å. The presence of sharp peaks at low temperatures indicates strong hydrogen bonds between the G: C base pairs, whereas the reduced peak height at higher temperatures (348 K and 363 K) indicates lesser hydrogen bond strength **Figure 6.18**. The RDFs indicate the high stability of the G: C base pair at low temperatures as compared to high temperatures.

Therefore, we defined a G: C base pair as being unchanged if the distance between any of the three hydrogen bonds present in G: C is less than or equal to 2.9 Å and changed otherwise.

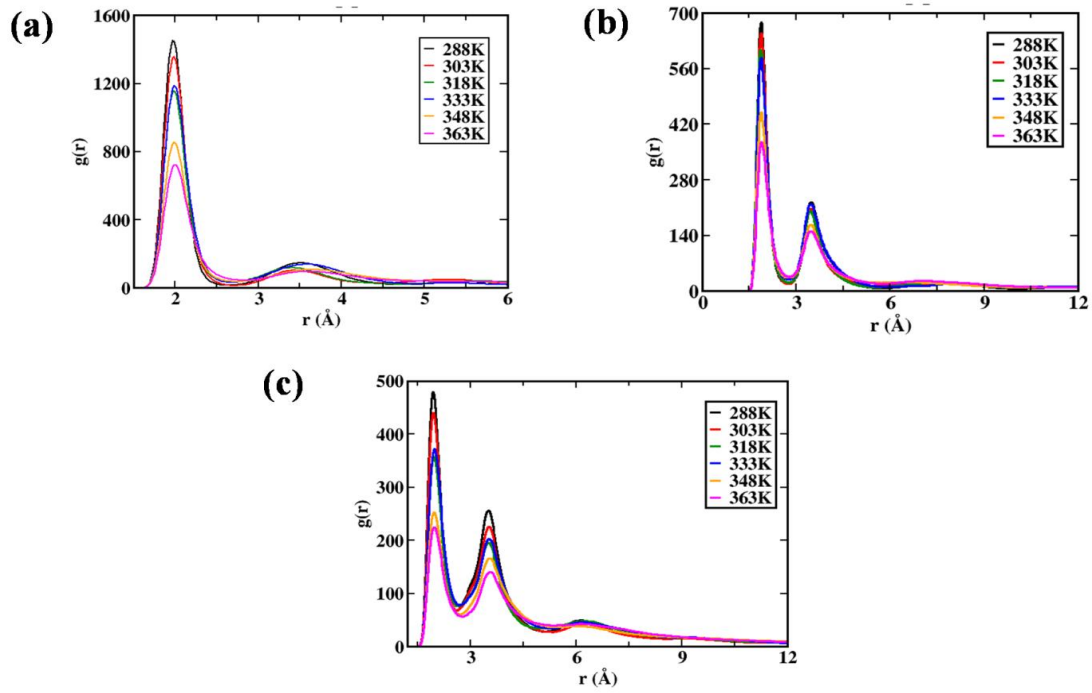


Figure 6.18: Radial distribution functions of different atoms of G:C base pairs with respect to temperature in SPC/E water model **(a)** N-H_1**(b)** O_2-H_2 and **(c)** O_6-H_4.

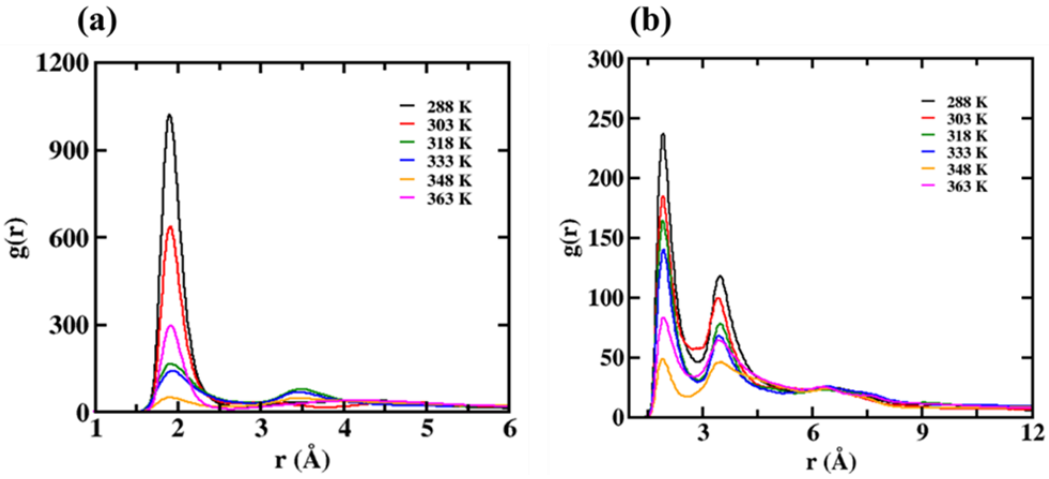


Figure 6.19: Radial distribution functions of different atoms of A: U base pairs with respect to temperature **(a)** N1-H3 **(b)** O4-H61/62.

Next, the base pair opening lifetime of A: U base pairs of RNA were calculated. The fluctuations of the RNA backbone induce changes in the hydrogen bond lifetime of A: U base pairs which provide an idea of RNA structural dynamics. The hydrogen bond lifetime is the main result that can prove the stability of the nucleic acid, protein, and peptide structures (Alves et al. 2021). The cut-off (2.5 Å) for calculating the hydrogen bond lifetime has been selected based on the radial distribution function of A: U base pair atoms N_1—H_3 (2.5 Å) and O_4—H_{61}/H_{62} (2.54 Å) as shown in **Figure 6.19**.

Table 6.3: The lifetime of the G:C and A: U base pair in RNA molecule.

Temperature	G: C Base-pair HB lifetime (ps)	A: U Base-pair HB lifetime (ps)
288 K	5850.30	4200.96
303 K	4429.83	3102.05
318 K	2723.55	2309.53
333 K	2096.19	2090.34
348 K	876.60	1095.43
363 K	1181.91	982.30

It is evident from **Table 6.3** that as the temperature increases, the $S_{G:C}$ value decreases, indicating frequent changes in the conformations. However, it can be noticed that at 363 K, the $S_{G:C}$ value is higher than 348 K indicating conformational stability due to the formation of non-native base pairs in the SPC/E model. The hydrogen bond lifetime of the A: U base pairs are given in **Table 6.3**. This value show a good agreement with the open pair lifetimes of DNA/RNA base pairs which are in the range of 10 ns to hundreds of nanosecond (Várnai et al. 2004). The base pair opening mechanism can be determined by the dynamics of the hydrogen bond. The hydrogen bond which has a shorter lifetime breaks initially while the higher lifetime shows comparative stability. The data shows that the lifetime of the base pair opening of RNA is decreased by rising the temperature. At low temperatures, 288 K and 303 K, the hydrogen bond dissociation time is higher compared to high temperatures 348

K and 363 K. The results suggest that at higher temperatures the base pairs opening is faster as compared to lower temperatures energetically favorable structures. It is found that the range of the timescale obtained from hydrogen bond dynamics is comparable to the conformational lifetime calculated at higher temperatures. Hence, it is clear that the sub-nanosecond value of the conformational lifetime is indicative of the structural dynamics of the base pairs. Further, the intercorrelations between the prominent structures were drawn using a configurational tree to check whether the prominent structures at each temperature are visited at multiple temperatures or not (**Figure 6.20**).

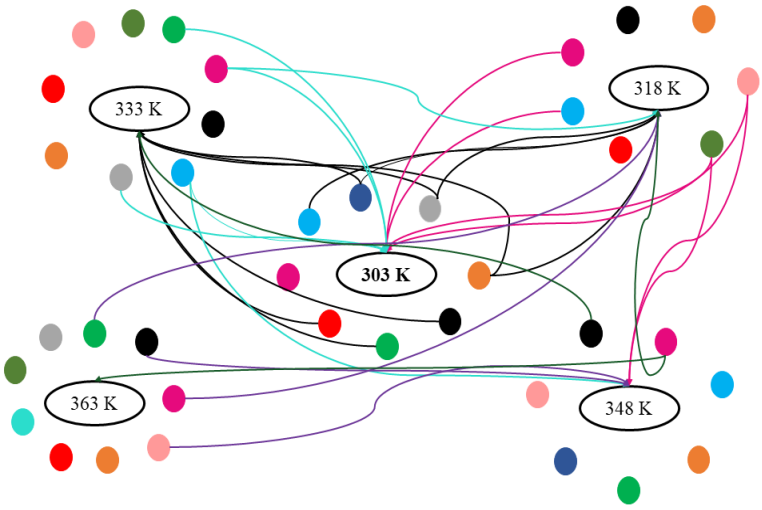


Figure 6.20: Configurational correlation tree showing the interconnection of prominent structures at different temperatures. Only the interconnections between the communities of different temperatures have been shown for clarity. The color of the communities is the same as shown in network distribution.

6.10 CONCLUSION

Finally, to conclude we presented the conformational evolution of the SARS CoV-2 by network analysis, which gives us information about the connectivity of the conformations and the existence of the most probable conformation of RNA at a particular temperature. The groove width and the twist based on minor-major grooves clearly show that the clustering method can efficiently distinguish the temperature-dependent conformational evolution. The ΔG value of the long-lived conformation shows stabilization of the RNA at 363 K with respect to 348 K due to the formation of

non-native contacts in the SPC/E water model. In this model, it is seen that in the temperature range from 303 K to 318 K, the ΔG value is comparable; the RNA shows similar activity in this region while at higher temperatures; the structure of RNA changes a lot due to the formation of the non-native contacts which determines its stability. The lifetime of the prominent conformations which have ≤ ±10° difference in their inter-helical angle was found to decrease with the increase in temperature initially up to 333 K and after which it increased. The maximum interlinking between the communities and minimum conformation lifetime are observed at 333 K and 348 K. This shows that maximum transitions between the conformations occurred at these temperatures. At higher temperatures, increased random motion of the atoms breaks the native bonds between the native base pairs, at the same time increasing the probability of formation of new non-native bonds due to an increase in collision frequency between the two residues which gives stability to the RNA structure at the higher temperature. The prominent structures were found to exist at multiple temperatures from the configurational correlation tree. The free energy values ΔΔG calculated between 288 K and 363 K match well with the experimentally reported unfolding energies of structurally similar RNA pseudoknots. The RNA conformation in the mTIP3P water model denatures at a lower temperature.

CHAPTER 7

SENSITIVITY OF THE RNA CONFORMATIONS ON THE WATER MODELS: EFFECT OF THE HYDRATION SHELLS

Various force fields and water model potentials influence significantly RNA conformations. The polyanionic nature of RNA attracts the water molecules and the counter ions which in turn affects their stability. The influence of the hydration shell on the SARS-CoV-2 RNA genome at different temperatures was analyzed using SPC/E water and modified TIP3P water in combination with different force fields CHARMM and AMBER.

7.1 BACKGROUND

The water is considered "an integral part of the nucleic acid structure" as it assists the folding of RNA and takes part in non-covalent interactions (Westhof 1988). The linear unbranched polymer of nucleotide units in RNA is composed of sugar units, phosphate groups, and nitrogen bases (A, C, G, and U) (Wang and Farhana 2022). The charged phosphate groups are the principal sites for hydration as well as for maintaining charge neutrality in an aqueous solution. The interactions of nucleic acids with water and solvent molecules are stronger than the proteins due to the polyanionic nature of RNA and DNA (Fingerhut 2021). The negative charges present on the backbone are shielded by positive ions (Na^+, Mg^{2+}) that help the RNA molecule to conserve its native conformation (Fischer et al. 2018). The electrostatic interaction with Mg^{2+} ions stabilizes RNA *via* inter-groove coordination of Mg^{2+} ions with the phosphate group leading to correlations in the structure of the sugar-phosphate backbone and ions (Fingerhut et al. 2021; Schauss et al. 2021). The electrical interactions in-between the charged molecular groups, water molecules and counter ions have a fundamental role in the stabilization of RNA. The negative-charged phosphate group of the RNA backbone induces the accumulation of the counterions in the first few solvation shells of the RNA. The distribution of K^+ and Cl^- ions around the surface of DNA oligomer also make influence on structural and dynamic

properties (Lavery et al. 2014). This affects the dynamical changes in the macromolecular structure of the biomolecule. The role of the solvent-separated ion pair on the stability of the folded structure of RNA has been discussed abundantly, but the role of intimate ion pairs and their effect on the hydration shell of RNA defining its 3D structure is largely unknown.

The interface of RNA in a biological system can be considered a complicated many-body problem. The phosphate groups of RNA-backbone form hydrogen bonds with hydration water as well as it gets polarized by the fluctuation in electric fields arising from the surrounding environment. The fluctuation in polarization is due to the first two hydration shells which are evident from the 2D-IR spectrum of DNA at different hydration shells and counter ion concentrations (Siebert et al. 2016). Only, a few studies have been done on the role of intimate ion pairs and the polarising effect of water molecules on RNA stabilization (Cuervo et al. 2014; Fingerhut et al. 2021). Therefore, there is a dire need of understanding interfacial electrostatics and their effect on the stability of the RNA system. The structural and conformational stability of RNA at the atomic level can be studied using atomistic molecular dynamics simulations. The effect of the electrostatic forces of ion-pair and its solvation shell can be studied by using different force fields and water model potential. The usage of non-polarizable water force- fields (FFs), by construction, cannot take into account the strong polarization. But the use of polarizable water models in studying the biomolecular system is affected by the fact that they are not fully compatible with the commonly used force-field parameters of biomolecules such as CHARMM, AMBER, etc. One way out is to perform the *ab-initio* molecular dynamics study but it also suffers from the drawback of huge computational expensiveness (Cassone et al. 2019, 2020). Considering this, the commonly used water models are non-polarisable water models such as SPC/E and TIP3P. However, the highly approximate nature of these kinds of water models and the indirect treatment of polarisation effects, the dynamics of the biomolecules such as RNA will depend on the interaction between the RNA-water, RNA-ion, and ion-water and these interactions are dependent on the force field parameters (Šponer et al. 2018). For binding of the cations, while *ab initio* molecular dynamics can help in studying the binding mechanism, local distribution of the

charge, and their impact on the water structure at the binding site; classical molecular dynamics is more useful in studying the long-term dynamics of the system such as residence time of the ions near the cavity of the RNA/DNA.

Water is described by several model potentials such as SPC/E, mTIP3P, TIP4P (Mark and Nilsson 2001), etc. having different parameterizations and different dielectric constants. Therefore, the use of different water model potentials can affect the structural and dynamical aspects of simulated biomolecules such as proteins (Venugopal et al. 2022) and RNA. Water molecules of different model potentials can be considered as a medium having different polarizability that can affect the structure and stability of the biomolecules. In this study, we performed MD simulations of the SARS CoV-2 RNA genome at six different temperature ranges between 288 K to 363 K with different water model potentials. We chose the commonly used force field like CHARMM27, AMBER-ff99bsc FFs, and the commonly used model potential of the water molecules such as SPC/E, and modified TIP3P to check their combination on the structure and dynamics of SARS CoV-2 RNA. This RNA genome has practical importance in drug discovery in the ongoing global pandemic. Efforts are continuing to develop efficient vaccines and drug candidates against the covid-19 pandemic using computer-aided methods (Das and Chakraborty 2020; El Hassab et al. 2020). Frameshift Stimulation Element (FSE) from the RNA genome helps the virus to compact the larger genetic information into shorter segments during translation (Neupane et al. 2020). Thus it can be considered an ideal target for the inhibition of RNA replication enzymes in viruses (Zhang et al. 2021). As the temperature and pressure rise, the folded potential energy of SARS CoV-2 in the aqueous atmosphere increases leading to instability. Experimentally, the coronavirus RNA genome was observed to be stable from 20°C to room temperature (Alfaro-Núñez et al. 2022; Liu et al. 2021) and it gets denatured at above 45°C. Given this experimental fact, a systematic study of this RNA can be done at a different temperature to check which combination of force fields and water model better describes the stability of the RNA with the experimental result and the effect of the interfacial water behind such stability.

7.2 SYSTEM SETUP

The same systems prepared in the previous chapter were used to analyze the effect the hydration shells on RNA conformations. Additionally, MD simulations were performed using AMBER99bsc force field in a modified TIP3P water model at six temperatures (288 K, 303 K, 318 K, 333 K, 348 K, 363 K) (**Table 7.1**). The simulations follow the same procedure as described in the previous chapter. The time evolution of potential energy, temperature, and pressure of the systems ensured the thermodynamic equilibrium (**Appendix XXIII**). The insights into the structural and dynamical properties of RNA and water have been analyzed by considering the last 200 ns simulation trajectory of the RNA-water system.

Table 7.1: Simulation details: N_{RNA} number of RNA, N_{WM} number of water molecules, ***a** is showing the temperature range (288, 303, 318, 333, 348, and 363 K).

System	N_{RNA}, N_{WM} and Water model	Force field	Simulation time
Set 1	1, 32669, SPC/E	CHARMM27	250 ns *a
Set 2	1, 32669, mTIP3P	CHARMM27	250 ns *a
Set 3	1, 32669, mTIP3P	AMBER99bsc	250 ns *a

7.3 PROBABILITY DISTRIBUTION OF STRUCTURAL PROPERTIES

The primary changes in structural properties, Root Mean Square Deviation (RMSD), and radius of gyration (R_g) of RNA genome with the rise in temperature are shown in **Figure 7.1**. The RMSD distribution for both water models SPC/E and mTIP3P in CHARMM force field are shown in **Figures 7.1 (a), (b),** and that of AMBER-mTIP3P is shown in **Figure 7.1 (c)** respectively. It is found that at lower temperatures such as 288 K and 303 K the RMSD of the RNA molecule has a broad distribution of probable structures and at higher temperatures (except 348 K) the sharp probability distribution of RMSD shows less collection of RNA structures. As the temperature rises to 348 K, the RMSD probability distribution shifts towards the higher value and becomes broader which shows more changes in RNA structures. At 363 K, the structure is less dynamic with respect to 348 K largely due to the confined change in conformation of the RNA genome.

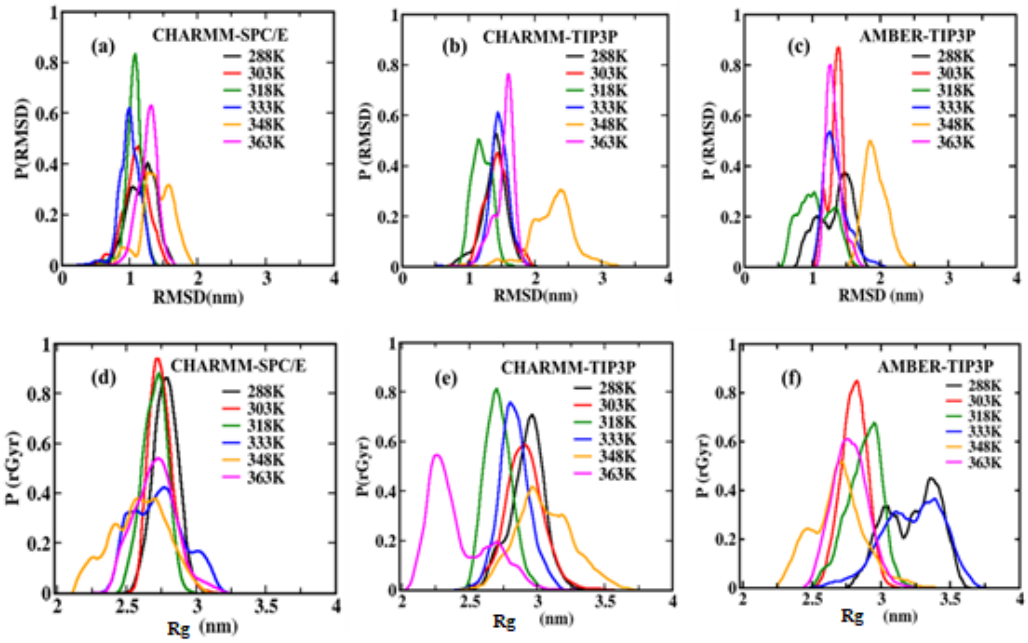


Figure 7.1: The probability distribution of RMSD (a), (b), (c) and R_g (d), (e), (f) for SPC/E and mTIP3P water models for the SARS-CoV-2 genome at various temperatures.

The mTIP3P water model is found to show higher RMSD distribution in the case of both the CHARMM and AMBER force fields compared to the SPC/E water model. This shows the RNA molecule undergoes more conformational changes in the mTIP3P water model than in the SPC/E water model. Both water models showed a broader and higher RMSD distribution at 348 K. To have further details about the stability of the RNA structure, we calculated R_g values in **Figures 7.1 (d), (e), and (f).** The measure of R_g value can give information regarding the compactness of the RNA structure. It can be seen that the SPC/E water model gives a better trend than the mTIP3P water model. It is found that for the CHARMM force field, the probability distribution is broader as we go towards higher temperatures (348 K and 363 K). The narrow peak of the probability distribution of the radius of gyration of the RNA at lower temperatures shows the compact conformation of the RNA. The compactness of the RNA molecule reduces due to the thermal fluctuations at higher temperatures. There is also the probability of the formation of misfolded structures at higher temperatures; which can give rise to a compact structure leading to a lesser R_g value

at the higher temperature. This phenomenon is mainly observed for 333 K and above in the case of CHARMM-SPC/E and at 363 K for CHARMM-mTIP3P. At 348 K, in CHARMM-mTIP3P, the high value and broader distribution of R_g value is found due to the base pair opening. In the case of AMBER-mTIP3P, R_g values are higher and broader at 288 K, and 333 K temperatures. At temperatures 348 K, and 363 K, R_g is low and has less broad distribution compared to 288 K and 333 K temperatures. To have more insights, we have calculated the native and non-native contacts of the base pairs at every temperature as shown in **Figure 7.2**. RNA molecules tend to form more non-native base pairs at high temperatures compared to the low temperatures as can be seen in **Figure 7.2**. The RNA structure showed 20 native base pairs for CHARMM-SPC/E and CHARMM-mTIP3P model whereas, only 18 native base pairs are observed for the AMBER-mTIP3P model at 288 K.

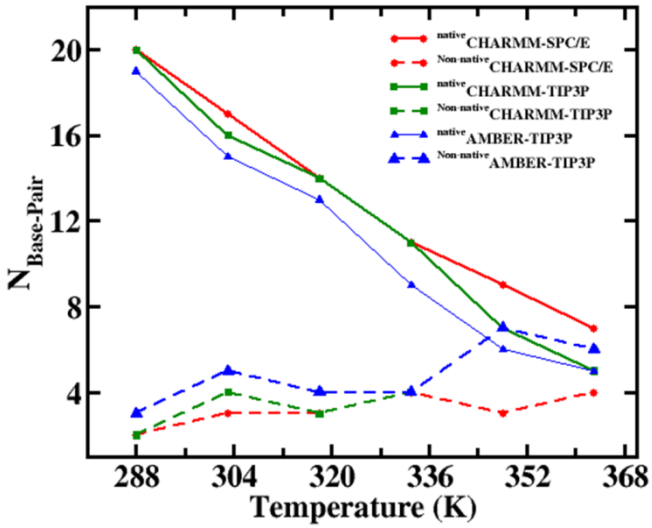


Figure 7.2: The scatter plot of the native and non-native base pairs with different temperatures.

The native base pairs are found to decrease with the rise in temperatures. At 363 K, 8 and 5 native base pairs are observed for SPC/E and mTIP3P water models respectively. The presence of non-native base pairs is less at lower (288 K, and 303 K) temperatures in SPC/E compared to mTIP3P. This difference is evident at the higher temperature in mTIP3P than the SPC/E water model. Therefore, there is a possibility of getting the RNA structure more compact due to the formation of non-

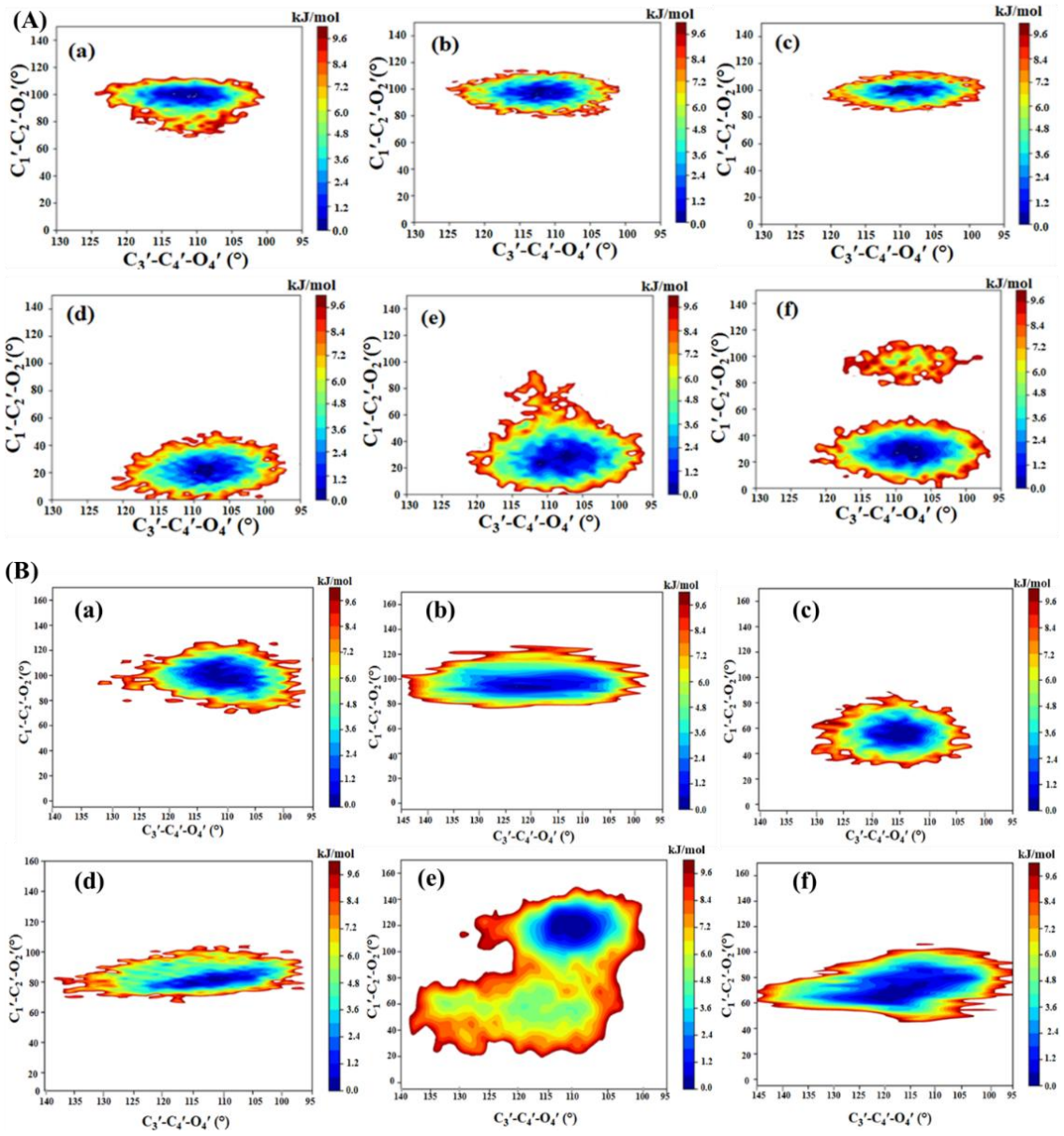
native base pairs at higher temperatures. This explains the observation of **Figure 7.1** at higher temperatures (363 K) where the RMSD and R_g values are smaller for 363 K than 348 K. In undergoing this course of transformation, the structural changes of RNA occurred from A conformation to A′ conformation' at higher temperatures. Further, we have calculated the change in the dihedral angles to have more insight into the conformational change of the RNA.

7.4 FREE ENERGY LANDSCAPE (FEL): THE CONFORMATIONAL CHANGES WITH BASE PAIRS

It will be interesting to study the changes in the RNA conformations in different water models based on the changes in dihedrals. The conformational analysis suggested that the base pair hydrogen bonding and base pair stacking in RNA can be distinguished using bond angles. In this case, two order parameters or reaction coordinated (RCs) are due to their suitability to explain the conformational changes.

Figure 7.3 shows 2D-FEL obtained between the backbone dihedral (C_3'-C_4'-O_4') and pentose sugar angle (C_1'-C_2'-O_2') near the base region i.e., C_2'-endo of RNA. Experimentally, it is observed that the conformational states of the RNA hairpins are isolated by free energy barriers ≥2 kcal/mol (Sarkar et al. 2009; Stancik and Brauns 2008). In general, the free energy surface significantly depends on the variation of bond angles in RNA conformation. At the lower temperatures (288 K, 303 K, and 318 K) in the CHARMM-SPC/E water model, the sugar pucker bond angle (C_1'-C_2'-O_2') is 107° in **Figure 7.3A (a-c),** which is almost same as for native conformation (Olson et al. 2001). At higher temperatures, it shifted towards a very lower value (20°) **Figure 7.3A (d-f)**. It can be seen that at 348 K, and 363 K the sugar pucker bond angle is more spread indicating conformational heterogeneity due to changes in base pairs of RNA molecules. The appearance of another minimum around 100° leads at 363 K is due to the non-native base pair that occurred due to the misfolding of the RNA leading to the stabilization of the molecule. In the case of the CHARMM-mTIP3P water model, the free energy landscape at 288 K, and 303 K showed a minima around 109° of sugar pucker bond angle in **Figure 7.3B (a-b).** It can be seen that in **Figure 7.3B (c-d)** at 318 K and 333 K, the conformation got changed and

shifted towards a lower value of 60°. It shows that the RNA molecule has changed from its native conformation. At 348 K and 363 K the FEL was found to be more spread out in both backbones (C_3'-C_4'-O_4') and pentose sugar angle (C_1'-C_2'-O_2') directions. The RNA molecule showed more dynamic behavior in CHARMM-mTIP3P compared to CHARMM-SPC/E. Similarly, a transient behavior is seen in AMBER-mTIP3P (**Figure 7.3C**). At Higher temperatures, the FEL is more spread out in both directions compared to CHARMM-SPC/E water. Hence, it is clear that in the CHARMM-mTIP3P the RNA molecule has more transient conformations. This can be related to the changes in hydrogen bonds in base pairs and stacking. Next, the changes in O2′-O2′ at every temperature have been calculated.



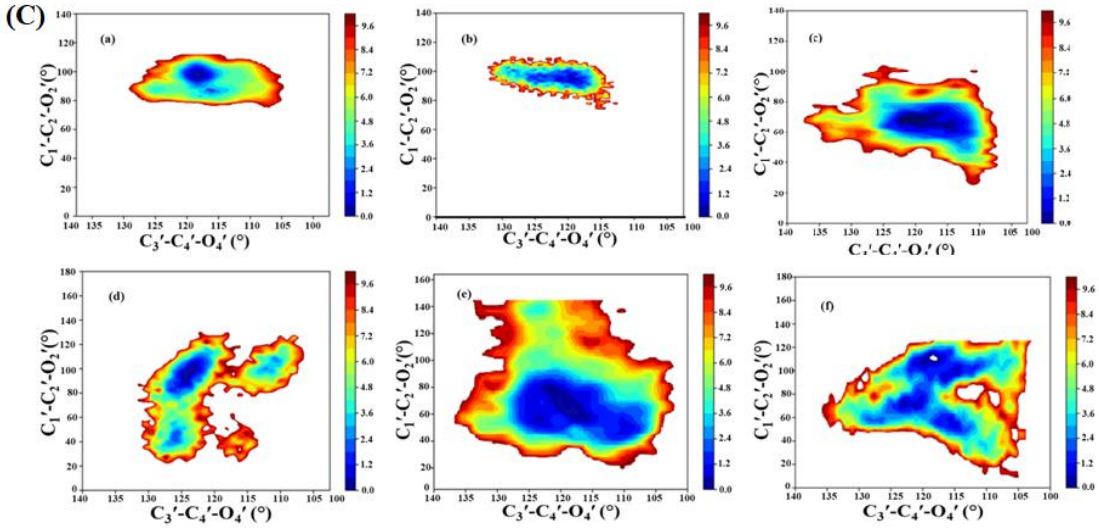


Figure 7.3: Free energy landscape of bond angles (A) CHARMM_SPC/E, (B) CHARMM-mTIP3P (C) AMBER-mTIP3P at different temperatures (**a** 288 K, **b** 303 K, **c** 318 K, **d** 333 K, **e** 348 K, **f** 363 K). The FEL is obtained by considering reaction coordinates as bond angles of sugar parker phosphate constituents.

7.5 CHANGES IN O2′- O2′ DISTANCE IN THE RNA

The 2′-hydroxyl group controls RNA conformational properties. The 2′-hydroxyl group orientation can alter the natural flexibility of the phosphodiester bond in the backbone of RNA (Denning and MacKerell 2012). **Figure 7.4,** shows the distribution of O2′···O2′ distances (sugar pucker) of the RNA genome using the CHARMM-SPC/E and CHARMM/AMBER-mTIP3P water models. It was found that at 288 K, the average distance between the O2′···O2′ was around r = 6.75 Å CHARMM-SPC/E water model (**Figure 7.4 (a)**). With the temperature rise, the probability distribution of the average distance between O2′···O2′ is shifted toward a higher value. At 333 K, 348 K, and 363 K we found broader distribution due to the formation of new native/non-native base pairs in the RNA molecule. A similar trend is observed in the case of the CHARMM-mTIP3P water model. The higher solvation of the RNA molecule in the TIP3P water model leads to the opening of the base pairs stacking resulting in the shifting of the probability distribution curves of the O2′-O2′ distance to the higher values for the CHARMM-mTIP3P. At higher temperatures, in RNA molecules, the nitrogenize bases do not have stable overlaps between neighboring base pairs in both the water models.

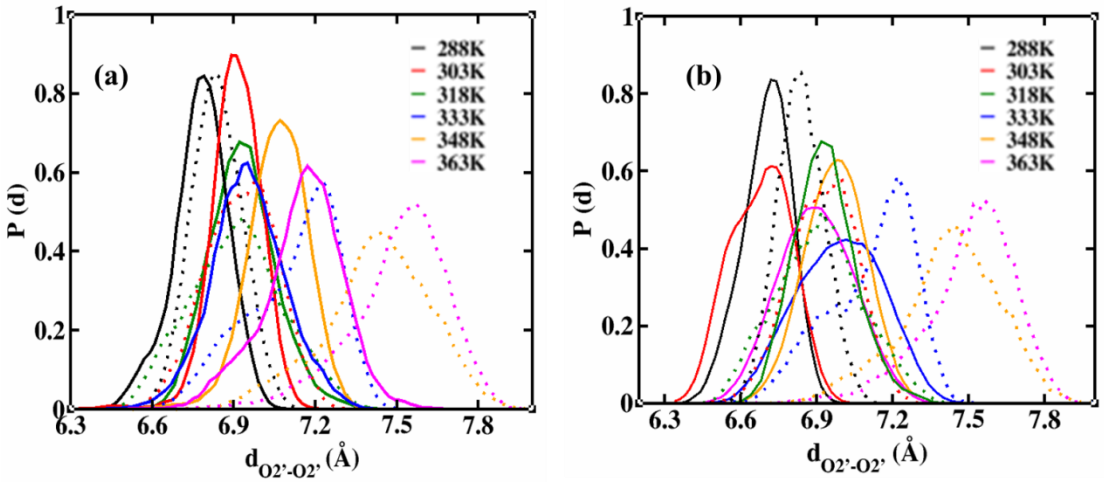


Figure 7.4: The probability distribution of the O2′···O2′ distance at various temperatures in (a) CHARMM-SPC/E (solid line), CHARMM-mTIP3P (dotted line), and (b) AMBER (solid line)/CHARMM-mTIP3P (dotted line) water models.

The probability distribution of hydrogen bonds between the O_2' of RNA and the hydrogen of the water molecules is depicted in **Figure 7.5**. It is evident that the distribution of hydrogen bonds decreases as the temperature raised from 288 K to 363 K i.e., the structural properties of water molecules around the surface is changed in both water models. In the case of the AMBER-mTIP3P system, **Figure 7.4 (b)** no definite trend is seen. At lower temperatures, the O2′···O2′ distance has a low value compared to higher temperatures. The probability distribution at 318 K, and 363 K shows a lower value of O2′···O2′ distance than 333 K and 348 K. The probability distribution O2′···O2′ distance has broader distribution at 333 K and 348 K temperatures in the AMBER-mTIP3P water model. It is evident that the selection of the water model and force field has a clear effect on the stability of the RNA structure. Therefore, in the next section, we have included the study of the solvation structure of RNA with the effect of different water models.

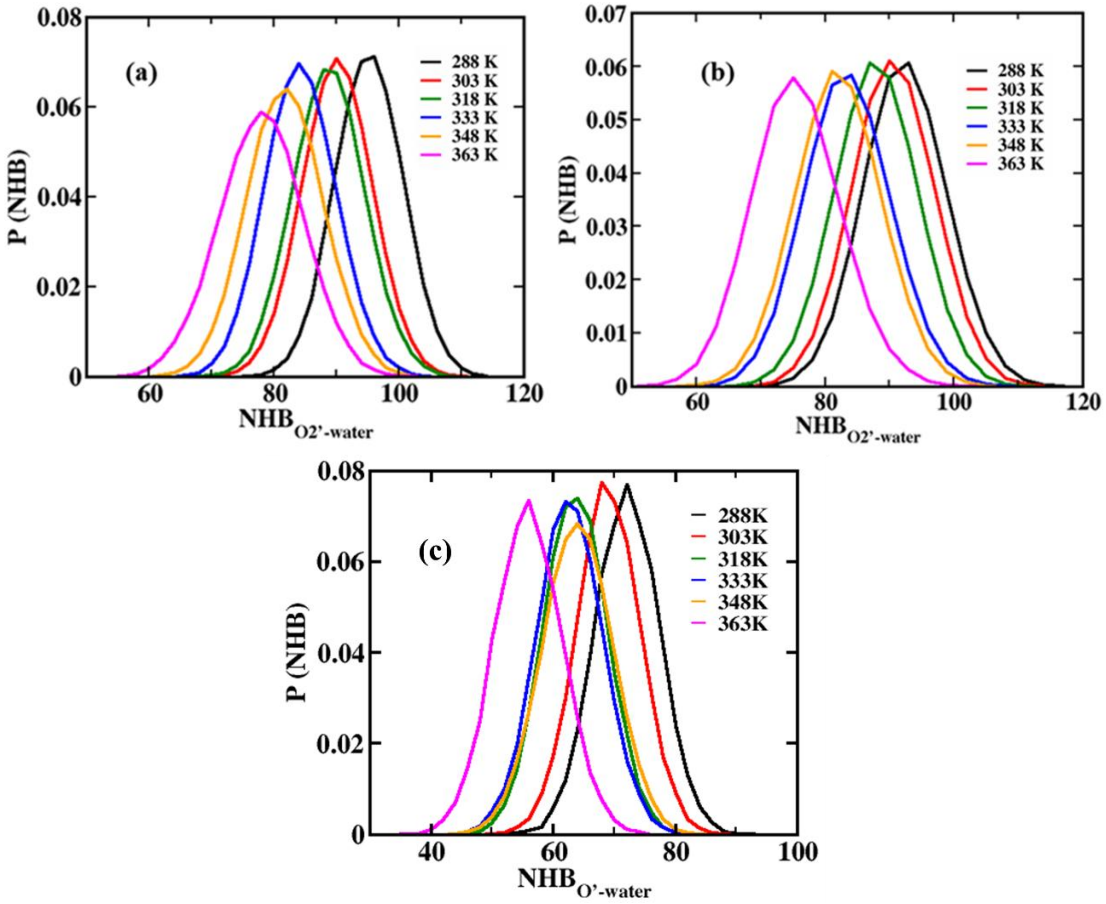


Figure 7.5: Probability distribution of hydrogen bond between O2′-water (a) CHARMM-SPC/E (b) CHARMM-mTIP3P and (c) AMBER-mTIP3P models.

7.6 EFFECT OF THE WATER MODEL AND FORMATION OF ION PAIR

CHARMM-SPC/E and CHARMM-mTIP3P results were chosen to further discuss the solvation structure of the RNA. The chosen water models have also a lot of differences in their dielectric constant. A solvent with a higher dielectric constant can solvate the polar groups more. The effect of temperature and force field on the polarization properties of water can be understood from the dipole moment value of the water and the dipole moment fluctuations of the RNA molecule. The calculated dipole moment of SPC/E water is 2.30 D and for mTIP3P water is 2.35 D. It is known that water molecule has higher dipole moment and dielectric constant which changes with temperature. The effect of a higher dielectric constant is, that it can screen out the charges effectively resulting in better solvation of the polar groups. Therefore, the changes in the dielectric constant of the water and its influence on the ion-pair

formation help to study the effect of the water model on the stability of the RNA molecule. With the rise in temperature, the dielectric constant of the water decreases for both the model potential (Cuervo et al. 2014) which is evident from **Figures 7.6 (a)**. The dielectric constant in mTIP3P is higher compared to the SPC/E water model. Therefore, the RNA molecule is supposed to be more polarised in the mTIP3P model, due to the interaction of the water molecule. The polar groups of the RNA genome are more solvated leading to the distortion of the structural parameters.

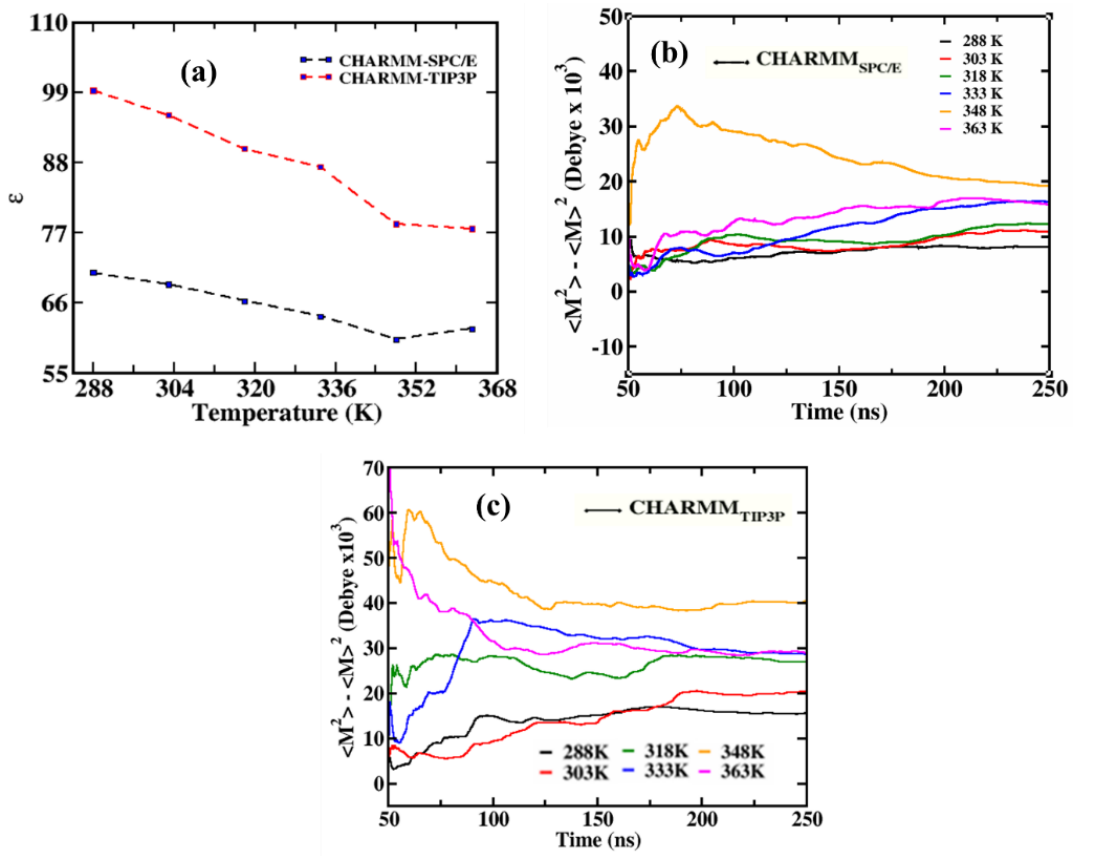


Figure 7.6: Figures (a) the dielectric constant of water and (b-c) are showing the dipole moment fluctuation for RNA molecules in CHARMM-SPC/E and CHARMM-mTIP3P respectively.

The dipole moment fluctuation of the RNA molecule is plotted in **Figures 7.6 (b) and (c).** It is found that with the increase in temperature, the polarisation value increases. This is because, with the increase in temperature, the compactness of the RNA decreases, as a result the structure of the RNA gets more solvated and hence more polarised. It can also be seen that, in both the models, at 348 K, the RNA

molecule has a higher polarization value, higher than 363 K. This can be related to the higher RMSD and R_g values of the RNA structure at this temperature, which suggests that the RNA is less compact and more accessible to the water molecules. At higher temperatures above 348 K, i.e., 363 K, the RMSD and R_g value decreases due to the formation of a stable misfolding structure of the RNA. Such a phenomenon has been observed for RNA tetraloop experimentally (Miner et al. 2016). This decreases the fluctuations in the dipole moment value of the RNA structure more. In mTIP3P, the RNA molecule has a higher dielectric constant at higher temperatures than the SPC/E water model and after 180 ns it got converged (Lavery et al. 2014) in both water models. The polar groups of the RNA genome are more solvated in the mTIP3P water model. As a result, in the SPC/E system, the positively charged Na^+ ion (counter-ion) will like to be found more around the negatively charged phosphate groups compared to the mTIP3P systems. The effect of a higher dielectric constant is, that it can screen out the charges effectively resulting in better solvation of the polar groups. Therefore, the changes in the dielectric constant of the water and its influence on the ion-pair formation help to study the effect of the water model on the stability of the RNA molecule. To find this, the coordination number of the oxygen atoms of the phosphate group around Na^+ ions for both water models we calculated and is given in **Table 7.2**.

Table 7.2: Coordination number of the oxygen atoms of the phosphate group of the RNA molecule around Na^+ ions within 6.0 Å cut-off (decided from RDF **Figure 7.8)** from the phosphate group.

System	CHARMM-SPC/E		CHARMM-mTIP3P	
	Na^+-$O_{(PO4^-)}$	$O_{(PO4^-)}$-$O_{(Water)}$	Na^+-$O_{(PO4^-)}$	$O_{(PO4^-)}$-$O_{(Water)}$
288 K	3	55	2	57
303 K	3	52	2	54
318 K	3	50	3	52
333 K	3	49	3	50
348 K	4	46	4	48
363 K	4	44	4	45

The coordination number of the oxygen atoms of the phosphate group of the RNA molecule around Na^+ ions at 288 K for SPC/E water is 3 and that of mTIP3P water is 2. The main difference in the coordination number was found at lower temperatures mainly at 288 K and 303 K. This coordination number increases with the temperature changes. The coordination number of the water molecules near the oxygen of the phosphate group of RNA is however more for the mTIP3P water model. This shows that the RNA molecule is more solvated in the mTIP3P model and has less contact ion pair. This coordination number decreases with a rise in the temperature. The dielectric constant of the water decreases in both force fields with the increase in temperature, therefore the probability of the formation of ion pair between the Na^+ ions and the phosphate group increase. At higher temperatures, the water molecules that are involved inside the cavity, however, move apart and the RNA structure becomes dewetted.

There is a tendency for the ions to be found inside the cavity of the big biomolecular systems like RNA (Pan et al. 2014). These ions are present along with water molecules and stabilize the RNA by water-mediated hydrogen bonds. The population of the Na^+ ions strongly interact with the oxygen of the phosphate group. It will be interesting to see the free-energy barrier heights of the monovalent cations Na^+ from the phosphate group of the RNA. Next, metadynamics were performed to calculate the free energy barrier.

7.7 METADYNAMICS

To calculate the free-energy barriers associated with the unbinding of Na^+ from the phosphate group of the RNA molecule, well-tempered metadynamics (Cassone et al. 2020) simulations have been conducted by the PLUMED-2.3.3 (Tribello et al. 2014) patch with the GROMACS.

The metadynamics has been performed at lower (288 K) and higher (363 K) temperatures for SPC/E and mTIP3P water models **Figure 7.7**. The inter-atomic distance between cationic specie (Na^+) and the nearest phosphate group of RNA was selected as a collective variable (CV) to perform the metadynamics. For CHARMM-SPC/E and CHARMM-mTIP3P water model systems, the initial Gaussian potential

heights of 0.5 kcal mol^{-1} which is deposited along the relevant CV(s) were considered. The deposition rate of the potential hills was set to 50 fs, corresponding to 100 time-steps, and the Gaussian width σ = 0.03 Å, 0.02 Å used for SPC/E and mTIP3P water model respectively. The bias factor has been chosen equal to 10.0, and 8.0 in the combination of simulated temperatures for SPC/E and mTIP3P respectively. At lower temperatures the bias-free energy (-23.0 kcal/mol) is more to unbound the cation from the stable site of RNA compared to a higher temperature (-17 kcal/mol) in the CHARMM-SPC/E water model. In the case of CHARMM-mTIP3P water model, at a lower temperature, the bias-free energy is found to be -20.0 kcal/mol and at a higher temperature, the bias-free energy is -11.0 kcal/mol. It is evident that the bias-free energy is more in SPC/E compared to the mTIP3P water model. Overall, at the lower temperature, the ions are near to the phosphate group of RNA molecule with higher bias-free energy compared to the higher temperature. This observation is more prominent in CHARMM-SPC/E compared to CHARMM-mTIP3P water model. The significant difference in the free energy at 363 K for SPC/E and mTIP3P may be attributed to difference in polarization and solvation of RNA at higher temperature. At 363 K, the higher dielectric constant of TIP3P water molecules guides the solvation of the RNA molecule which prevents strong Na^{+}-phosphate ion pair formation. This leads to weaker electrostatic interactions which results the lesser unbinding energy in mTIP3P water model compared to SPC/E water model.

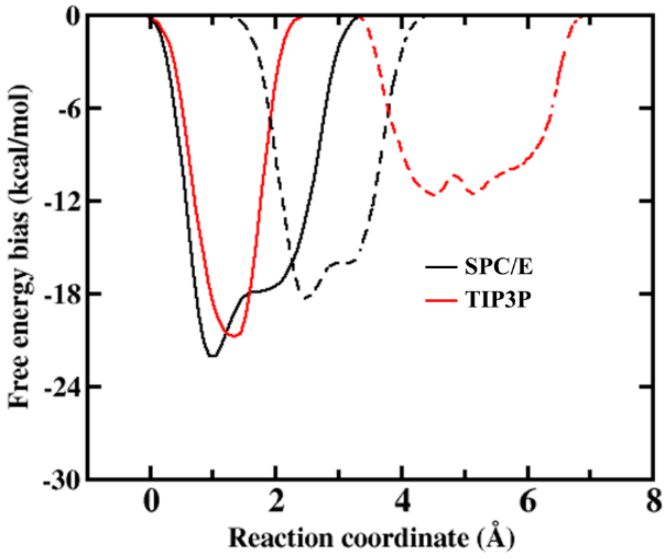


Figure 7.7. Free-energy bias deposited along the reaction coordinates (the inter-atomic distance between Na^+ - PO_4^- group) to unbind. The solid line shows 288 K and the dashed line is 363 K.

It is clear from the above discussions that the water models have important effects on the solvation structure of the biomolecule that can ultimately help in the stability of the biomolecule. Therefore, it is significant to study the hydration structure of the RNA in the presence of these ions. The distribution of the water molecules around the RNA is discussed in the next section.

7.8 RADIAL DISTRIBUTION FUNCTION

Solvation plays a crucial role in determining the secondary and tertiary structure of RNA molecules. It helps to stabilize the secondary structure of the RNA. The counter ions present in the cavity of the RNA impact directly the solvation structure of the RNA. Therefore, the intermolecular RDF of PO_4^--Na^+ and O_{1P}-O_w in both CHARMM-SPC/E and CHARMM-mTIP3P water model systems were calculated (**Figure 7.8**). The solvation peak heights of phosphate sodium RDFs are found to increase with the rise in temperature for both CHARMM-SPC/E and CHARMM-mTIP3P water models **Figures 7.8 (a), (b)**. However, the peak height of O_{1P}-O_w distribution was found to be decreased with the rise in the temperature for both CHARMM-SPC/E and CHARMM-mTIP3P water models **Figures 7.8 (c) and (d)**.

In this case, the first peak is found to be highest and sharp at around 2.75 Å, and the second and the third peaks are broader. The RDF of O_{1P}-O_w shows that the higher solvation shells vanish with the rise in the temperature indicating that the interaction of water with oxygens of the phosphate group vanishes as the temperature rises. From the two RDFs, the distribution of sodium ions near the phosphate group is more compared to the water molecules at higher temperatures. This suggests that the ion -ion pair interaction is more at higher temperatures. This situation is favored by the lower dielectric constant of the water molecules. The presence of water molecules and the counter ion near the phosphate group will affect the orientation of the water in the hydration shell of RNA molecules.

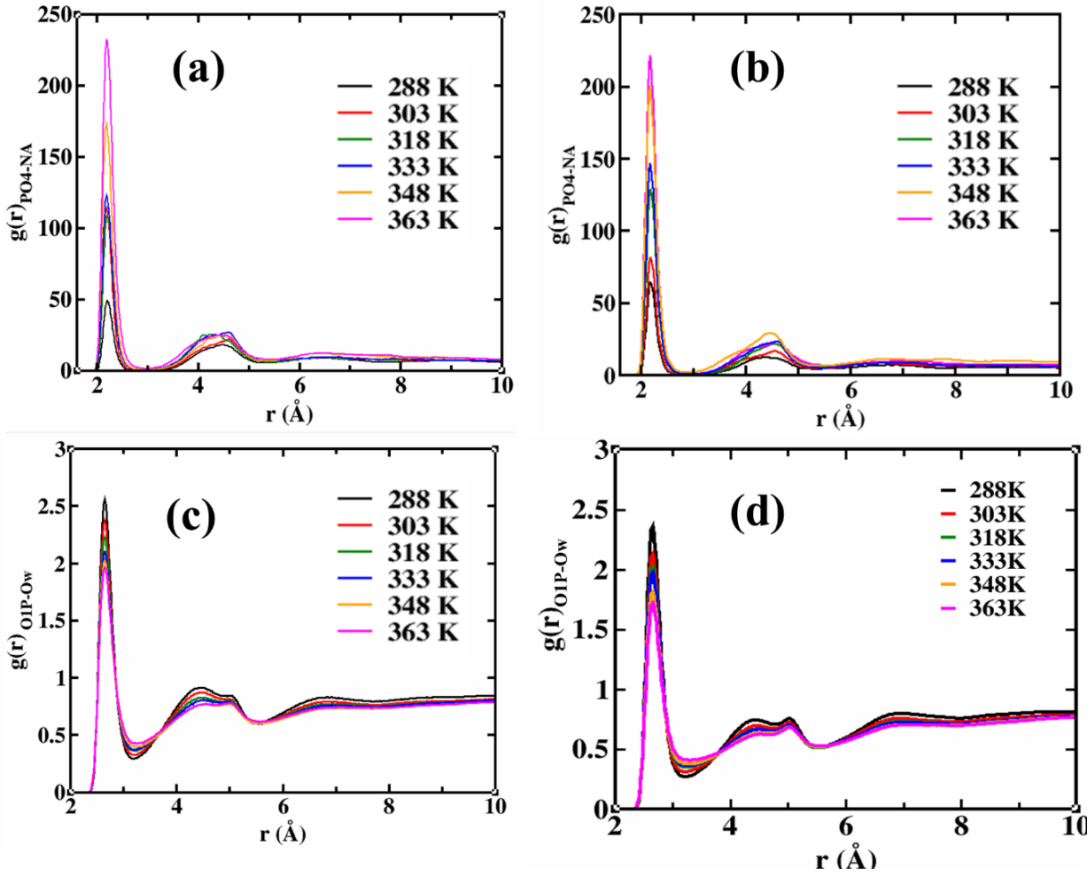


Figure 7.8: RDF of (a-b) phosphate group of RNA with Na^+, (c-d) O_{1P}-O_w at various temperatures in CHARMM-SPC/E (a, c) and CHARMM-mTIP3P (b, d) water models.

Further, the RDFs of O_w-Na in **Figure 7.9**, RG_N_2-O_w RG_O_6-O_w, RC_N_4-O_w, and RC_O_2-O_w in **Appendix XXIV** were plotted. It is found that the water structure is more disrupted around the nitrogen atom of the RNA molecules. RDFs corresponding to mTIP3P water models show a similar trend and are shown in **Appendix XXIV**. The arrangement of the water molecules, present at the interface will be affected by the presence of the contact ion pair near the RNA. Therefore, it will be worthwhile to check the arrangement of the interfacial water molecule and their effect on the stabilization of RNA.

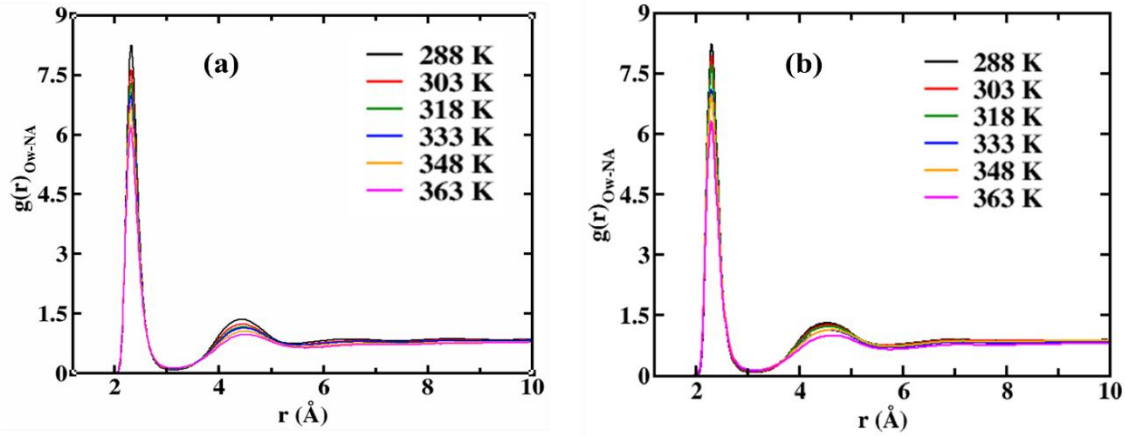


Figure 7.9: The radial distribution function of the (a) oxygen of water with Na^+ CHARMM-SPC/E (b) oxygen of water with Na^+ CHARMM-mTIP3P.

7.9 SOLVATION STRUCTURE OF THE RNA

The water structure near the RNA surface can be identified based on the local orientational tetrahedral parameter (S_g) which is based on the four nearest water oxygen (O_w) neighbours (Section 2.4.1 (viii)). The S_g parameter for the water molecules at the interface (cut-off value < 6.0 Å from the RNA base pairs; selected from the RDF **Appendix XXIV**) of the RNA genome is shown in **Figure 7.10**. It can be seen that the distribution of the S_g value near the RG and RC (**Figure 7.10 (a) and (b))** have broader distribution at 288 K, 303 K, and 318 K compared to 333 K, 348 K, and 363 K in CHARMM-SPC/E water model. These differences are less pronounced in the case of the mTIP3P water model. This can be explained as RG and RC contain more nitrogen and oxygen atoms as the electronegative group, they are more surrounded by the counter Na^+ ions. These ion pairs attract the water molecules to form stable trapped water molecules hydrogen bonds. The water molecules are trapped near the O2′ oxygen and hydrogen bond forms between the intracavity water molecules (**Figure 7.5**). It is already seen that at 288 K and 303 K, the coordination number of phosphate-Na^+ ion pairs are more (**Table 7.2).**

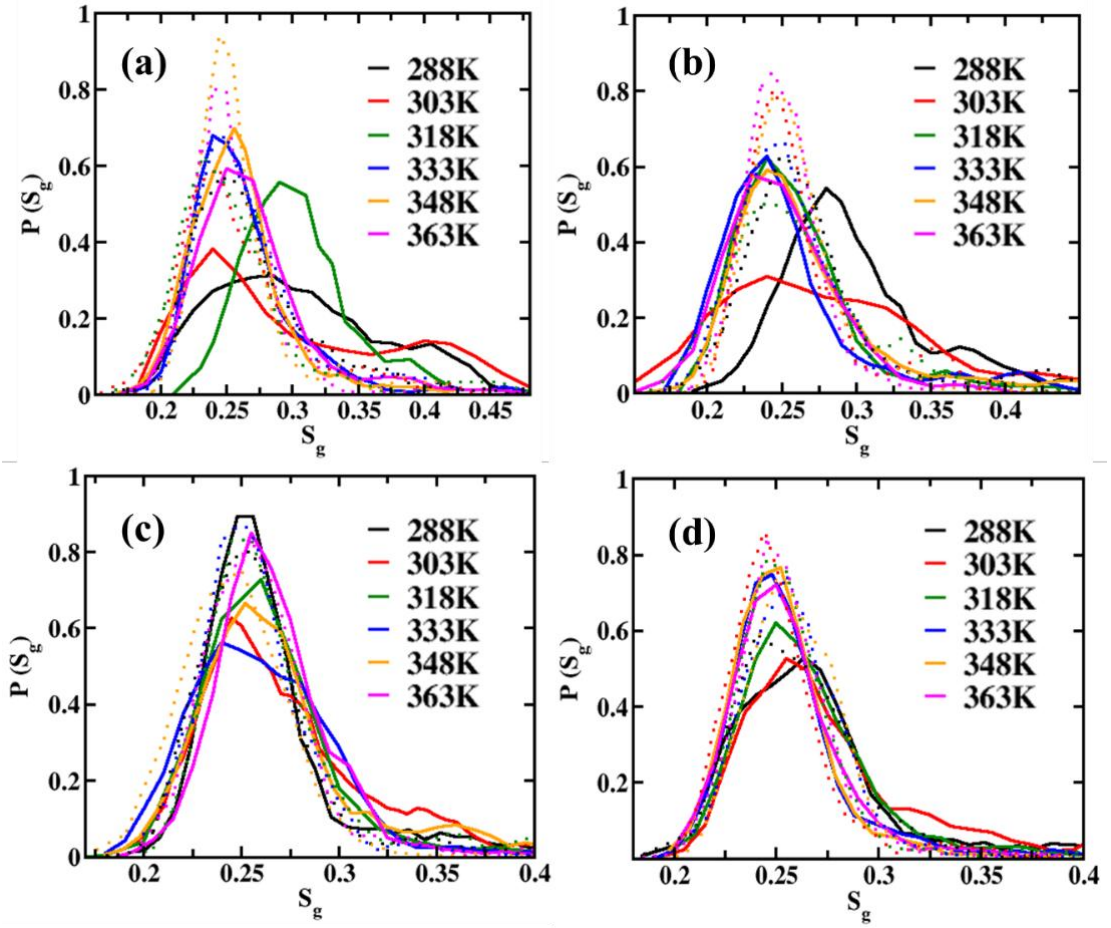


Figure 7.10: Local Orientational tetrahedral parameter of water near the interface of nucleic acids of RNA genome RG, RC, RA, and RU (a), (b), (c), and (d) respectively at different temperatures. The solid line for CHARMM-SPC/E and the dotted line for CHARMM-mTIP3P.

In the case of SPC/E, due to the lower dielectric constant, these ion-pair are more compact. These ion pairs attract a stable hydration layer near the RNA interface mainly at lower temperatures. This hydration layer is heterogeneous and consists of a non-tetrahedral (S_g > 0.35) structure of interface water. As a result, a broader distribution is seen in the SPC/E water model for lower temperatures. At higher temperatures, the hydration shell of the water molecules near the ion pairs is broken due to the random motion of water molecules. Therefore, S_g distribution is transformed to narrow and sharp, this indicates a uniform distribution solvation structure. In the case of the mTIP3P water model, the presence of such ion-pair is less due to a higher dielectric constant, and hence the hydration shell is more homogeneous (S_g <0.35).

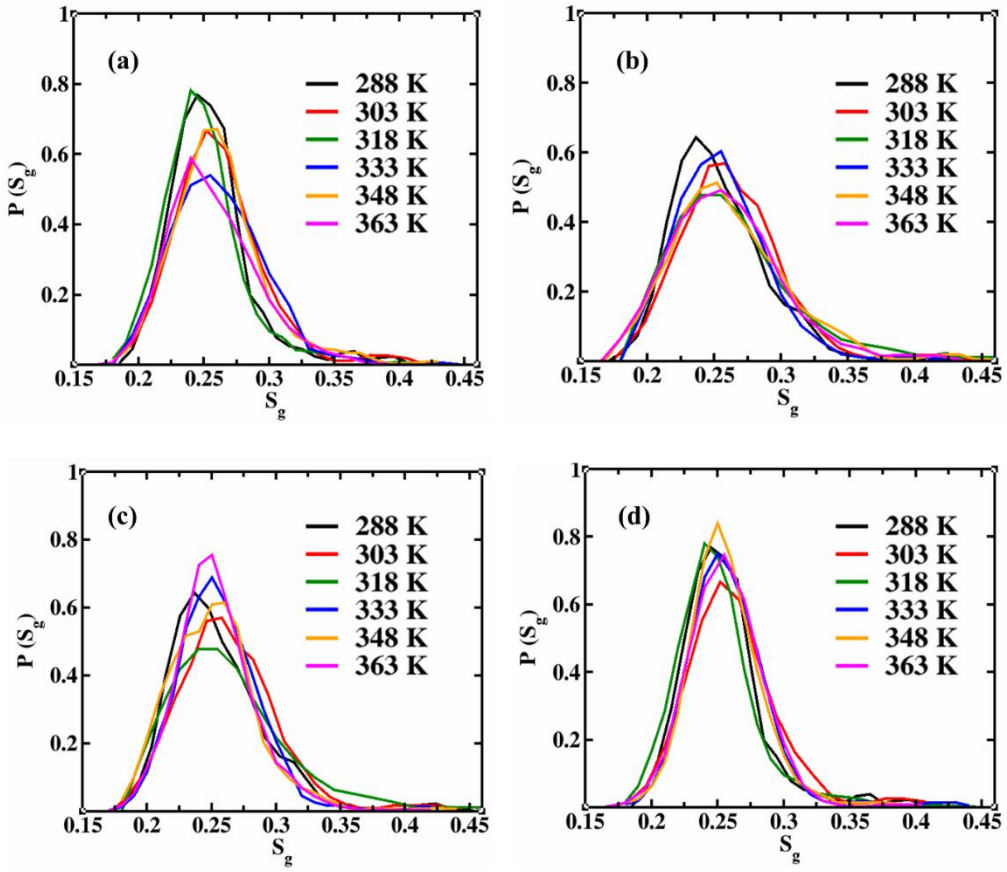


Figure 7.11: Orientational tetrahedral order Parameter of water molecules around the interface of nucleic acids of the SARS-CoV-2 RNA genome RG, RC, RA, and RU (a), (b), (c), and (d) respectively at different temperatures in AMBER-mTIP3P.

The S_g value for the AMBER-mTIP3P water model has been calculated and is given in **Figure 7.11**. The non-tetrahedral behavior of the interfacial water molecules are generally because of High-Density Water (HDW) having interstitial water molecules in their hydration shell (Biswas et al. 2018). The distribution of the interfacial water molecules for the mTIP3P water model is found to be more homogeneous in the mTIP3P water model and tetrahedral in comparison to SPC/E. Also, S_g parameter for the bulk water was calculated and shown in **Appendix XXV.**

To gain more insight, the voids distribution of the interface and bulk water at various temperatures were plotted. The void distribution (Biswas et al. 2018; Chakraborty and Chandra 2012) of the interface (**Figure 7.12a**) water molecules corresponding to higher temperatures have a tail region that clearly suggests lower homogeneity and less dense water molecules. It can be seen that the bulk water has a

lower void radius and the distribution becomes narrower; suggesting more dense water molecules and more homogeneity (**Figure 7.12b**). At lower temperature, it is observed that the SPC/E model have a lower void radius compared to the mTIP3P water model. This shows that at lower temperatures, the water molecules are higher coordinated in the SPC/E water model.

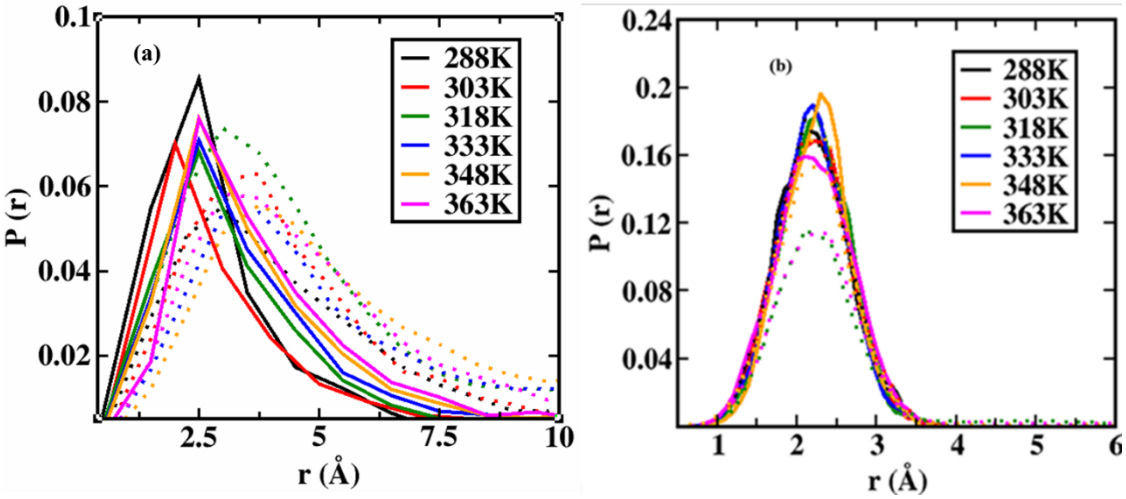
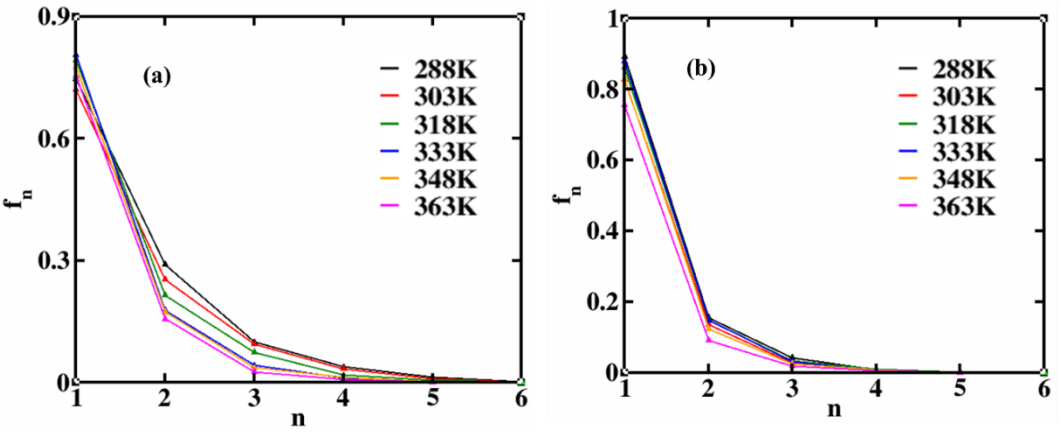


Figure 7.12: Void probability distribution of water molecules at different temperatures (a) interface (b) bulk. The solid line shows CHARMM-SPC/E and the dotted line for the CHARMM-mTIP3P water models system.

The fraction of the number of hydrogen-bonded water molecules are given in **Figure 7.13**. The fraction of higher coordinated water molecules is found to be more at lower temperatures in the SPC/E water model. Next, it will be interesting to see the dynamics of water molecules near the RNA interface and bulk water.



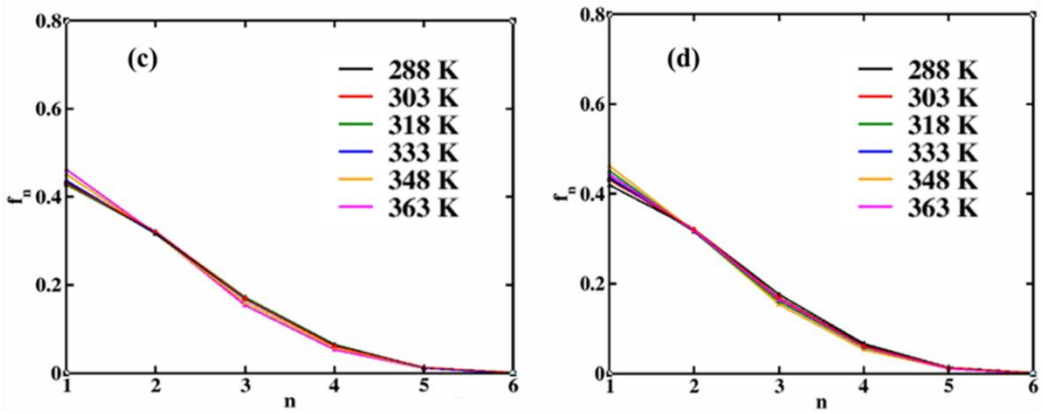
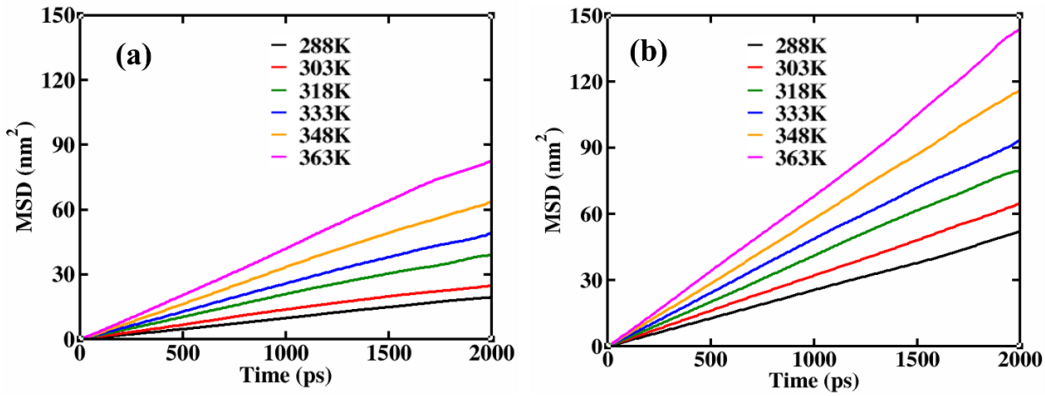


Figure 7.13: The fraction of water molecules having *n* number of hydrogen bonds in the interface (a-b) and bulk (c-d). (a-c) CHARMM-SPC/E and (b-d) CHARMM-mTIP3P water models.

7.10 DIFFUSION OF THE WATER MOLECULES

The diffusion coefficients are obtained by the fitting of linear equation of Mean Square Displacement and are calculated using equation 2.35. The MSD curve for bulk and interfacial water molecules is shown in **Figure 7.14**. The diffusion coefficient of the SPC/E water in bulk is found to be 2.82 $\times 10^{-5}$ cm^2/s, which correlates with the reported experimental value of 2.76 $\times 10^{-5}$ cm^2/s (Easteal et al. 1989). The diffusion of bulk water at different temperatures for this water model is found to be similar to the reported experimental values (Holz et al. 2000). The MSD values for mTIP3P water are found to be much higher (**Table 7.3**). The diffusivity of the water molecules increases with an increase in temperature for both water models. The strong hydrogen bonds of the water molecules near the RNA interface lower temperature leads to a lower value of diffusion coefficient. Therefore, it becomes important to calculate the strength of hydrogen bonded water molecules near the interface of RNA.



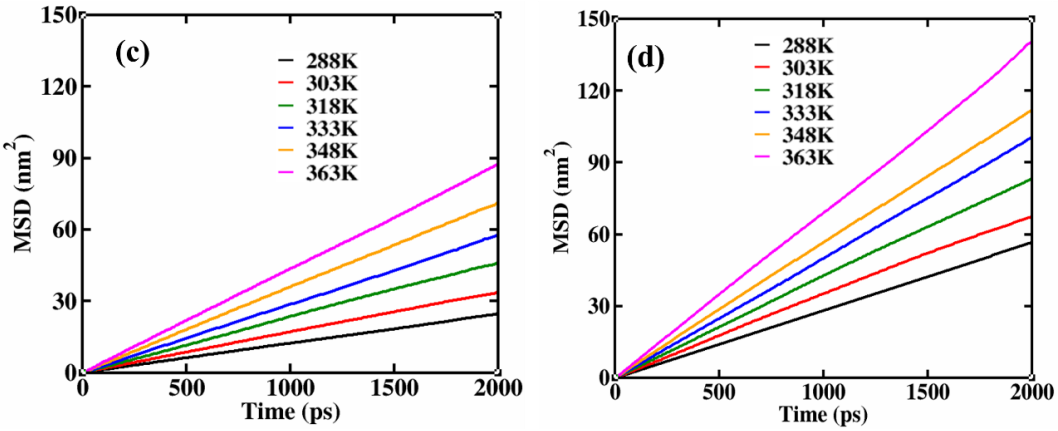


Figure 7.14: Mean Square Displacement of water molecules (a-b) interface around the SARS-CoV-2 RNA genome and (c-d) bulk at various temperatures for both CHARMM-SPCE (a-c) and CHARMM-mTIP3P (b-d) water models.

Table 7.3: Diffusion coefficient (D) of water molecules around the interface of RNA in SPC/E and mTIP3P water models, D = (value) x 10^{-5} cm^2/sec. Error reported as SD of individual simulations.

Temperature (K)	CHARMM- SPC/E		CHARMM- mTIP3P	
	Interface (< 6.0 Å)	Bulk (> 8.0 Å)	Interface (< 6.0 Å)	Bulk (> 8.0 Å)
288	0.99 ± 0.13	2.01 ± 0.01	3.01 ± 0.12	4.70 ±0.16
303	1.38 ± 0.12	2.82 ± 0.05	3.21 ± 0.15	5.21 ± 0.08
318	2.08 ± 0.20	3.62 ± 0.06	3.84 ± 0.18	6.54 ± 0.21
333	2.70 ± 0.16	4.74 ±0.12	4.91 ± 0.23	7.47 ± 0.17
348	2.88 ± 0.25	5.93 ± 0.14	5.15 ± 0.26	8.06 ± 0.32
363	3.82 ± 0.13	7.21 ± 0.02	6.47 ± 0.36	9.88 ± 0.42

7.11 HYDROGEN BOND LIFETIME

The hydrogen bond lifetime between the RNA and the water molecules is given in **Table 7.4**. Two molecules are hydrogen-bonded if the distance between the hydrogen bond acceptor and donor is < 2.5 Å (Section 2.4.2 (ii)). The value of hydrogen bond lifetime for the SPC/E water model is higher compared to the mTIP3P water model. The difference is more at lower temperatures. This shows the hydration shell near the RNA interface is more stable for the SPC/E model. At higher temperatures, the random motion of the water molecules around the RNA surface leads to the breakage

of hydrogen bonds leading to a lower hydrogen bond lifetime. The higher hydrogen bond lifetime in SPC/E can be attributed due to the presence of sodium ions at the RNA interface. Therefore, it will be interesting to check the residence time of the Na^+ ion on the interface of the RNA.

Table 7.4: The hydrogen bond lifetime of the RNA-water.

Temperature (K)	Interface Lifetime (ps) SPC/E	Interface Lifetime (ps) mTIP3P
288	2.42	1.46
303	1.63	0.93
318	1.58	0.78
333	0.97	0.73
348	0.91	0.61
363	0.53	0.48

7.12 SURVIVAL PROBABILITY

Survival probability (SP) gives an idea about the ion staying on the surface or inside complex environments such as peptides proteins, nucleic acids (Section 2.4.2 (iii)). The cut-off of 6.0 Å from the phosphate group of interfaces of RNA we selected to define the presence of sodium ions (decided from RDF **Figure 7.8**). The relaxation time was calculated for a long-time of 80 ps window duration. The calculated beta values are given in **Table 7.5**.

Table 7.5: The survival probability of Na^+ around the RNA.

Temperature (K)	CHARMM- SPC/E		CHARMM-mTIP3P	
	Time (ps)	β -value	Time (ps)	β -value
288	23.37	0.52	17.77	0.53
303	17.40	0.63	10.77	0.58
318	16.81	0.62	9.94	0.63
333	9.93	0.51	9.67	0.64
348	7.34	0.42	6.54	0.54
363	5.92	0.69	4.62	0.58

The beta values range from 0.4 – 0.7 indicates the relaxation of sodium ions near the surface of RNA is heterogeneous in nature. In the case of CHARMM-SPC/E, the residence time of the Na^+ ion near the RNA is more at lower temperatures compared to the CHARMM-mTIP3P water model. At higher temperatures, the SP value of Na^+ near the RNA molecule is decreased for both models. The higher residence time of the Na^+ ion near the RNA can also contribute to the higher hydrogen bond lifetime. To further clarify this, we calculated the interaction energies of the RNA-ion and RNA-water shown in **Figure 7.15**. In the case of CHARMM-SPC/E, the interaction energy of RNA-Na^+ is higher compared to CHARMM-mTIP3P. This indicates that the sodium ions are nearer to the RNA molecule in CHARMM-SPC/E compared to CHARMM-mTIP3P.

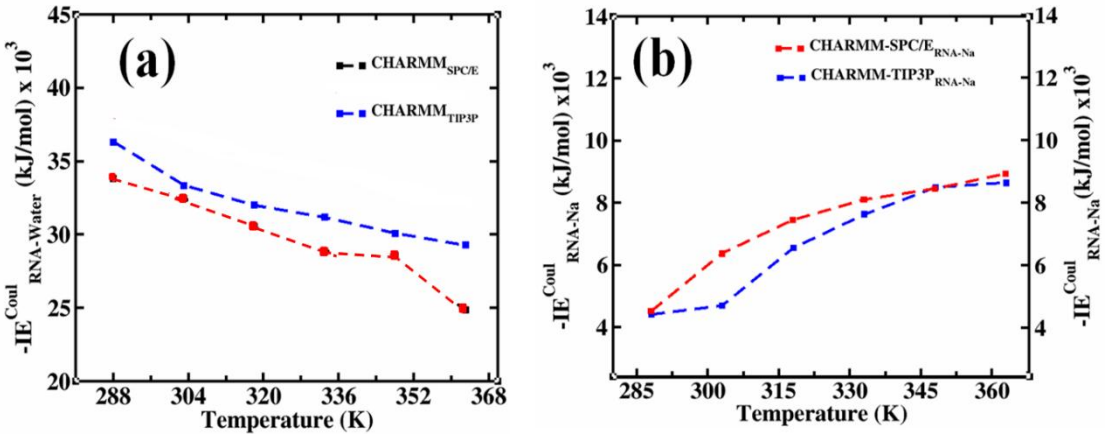


Figure 7.15: The interaction energies of the (a) RNA-water and (b) RNA-ion.

7.13 CONCLUSION

To conclude, we performed the MD simulation study of the SARS-Cov2 RNA genome at different temperatures with SPC/E and mTIP3P water model potentials to understand the effects on the structural properties of RNA, hydration shell, and stability. The RNA structure was found to be more dynamic in the mTIP3P water model. The RNA structure has undergone more structural changes leading to broader RMSD distribution in this model potential. An increase in temperature involves the breaking of native base pairs and the making of non-native base pairs affecting the stability of the RNA molecule. AMBER-mTIP3P combination showed the maximum

formation of non-native contacts. The dihedral angle distribution also showed maximum transient conformations for this water model. It is evident that the selection of the force field and water model has a significant role in maintaining the stable conformation of the RNA. Further, this study showed that the RNA molecule is more denatured and solvated in the mTIP3P water due to its higher dielectric constant. The main difference between the SPC/E and mTIP3P water models is the way by which the potential parameters for these water models were obtained. In the SPC/E model, the addition of a "self-polarisation term" (Leontyev and Stuchebrukhov 2010) leads to better performance of the potential model. We found a higher survival probability of Na^+ ions at the RNA interface for the SPC/E water model giving rise to higher interaction energy between RNA-Na^+ for the CHARMM-SPC/E water model. Due to the presence of the ions near the RNA interface, the solvation structure is also affected. We found the presence of inhomogeneous, non-tetrahedral water at the interface at 288 K and 303 K. These water molecules are higher coordinated and have lower void space than the bulk water molecules. They have a lower diffusion coefficient and have higher hydrogen bond lifetime. The effect is more prominent in the SPC/E water model. At higher temperatures, due to the random displacement of the molecules, the system becomes more homogeneous. In short, the RNA structure was found to be maximum stable in the SPC/E water model. This can be related to the existence of the ion-counter ion pair due to the lower dielectric constant of the SPC/E water model that in turn affects the hydration shell of the RNA interface.

CHAPTER 8

SUMMARY AND CONCLUSION

This chapter constitutes the summary of the current research work and the major conclusions obtained from the research. This also includes a brief scope regarding future work which can effectively give an idea of the biological process.

8.1 SUMMARY

This work addressed importantly the structure and dynamic properties of water near the hydration shell and bulk in the presence of metal ions. The interaction of metal ions with the intrinsic disorder peptides (IDPs) and their folding and unfolding in the presence of ions were studied. Insight to the residue specificity on the disordered to ordered transition at different temperatures. Another important work constitutes the structural evolution of the SARS-CoV-2 RNA genome at thermodynamic conditions. This overall study can help to obtain the 3D structure of the peptide and nucleic acids and help in drug delivery in the biological process. A detailed summary of the whole research work has been pointed out below.

- Molecular dynamics simulation of an antimicrobial peptide in the presence of metal ions with varying concentrations has been performed.
- The structure and dynamics of the prion peptide and the interfacial water have been done in the presence of monovalent and divalent metal ions and under different temperatures.
- Structural properties of water near interface and bulk are investigated by radial distribution functions, fraction of hydrogen bonds, and orientation profile whereas dynamical properties are studied by survival probability, hydrogen bond lifetime and diffusion coefficient. Thermodynamic properties are calculated by Kirkwood-Buff Integral and excess entropy.
- The secondary structure changes of the IDP are found to be more sensitive in the presence of ions.

- The influence of monovalent and divalent cations at different concentrations has been taken to calculate the free energy landscape of the unfolding/refolding process of the sheep prion peptide.
- To understand the factor affecting ions on prion peptide, the intra-peptide, and inter-peptide hydrogen bonds have been calculated. To understand the interaction of the ions with the prion peptide surface, preferential binding has been calculated.
- The dynamics of the folding/unfolding have been calculated in terms of hydrogen bond lifetime, diffusion coefficient, and survival probability.
- The conformational changes mainly in terms of the changes in secondary structure, contact map, surface exposure, hydrogen bonds, and interaction energy has been calculated.
- The clustering has been done to find out the conformations of the SARS CoV-2 RNA genome at different temperatures.
- Network analysis (a kind of Machine learning) has been used to see the changes in evolved conformations via clustering.
- The inactivation of the SARS CoV-2 RNA genome is evaluated in the presence of two different water models SPC/E and modified TIP3P at various temperatures.
- The hydrogen bond lifetime of G: C Base pairs have been calculated.
- A new methodology has been developed for the sampling of RNA: RMSD, and interhelical angle-based clustering.
- The Network analysis successfully helps to find out the prominent conformations of the RNA by making the edge connections between the nodes using a different algorithm.
- The CHARMM, AMBER force field, and TIP3P, SPC/E water models have been taken to see the influence on hydration water of RNA genome.
- The water structure around the RNA surface is evaluated by radial distribution function (RDF), the fraction of hydrogen-bonded water molecules, and the orientational tetrahedral order parameter.

- The dynamical properties such as diffusion coefficient survival probability, and hydrogen bond lifetime calculated and revealed the contribution of water molecules in stabilizing the binding sites of the RNA.

8.2 CONCLUSION

The major conclusions drawn from this work are as follows:

- Water molecules in the hydrophilic environment of an antimicrobial peptide are found to be more disruptive, and less dense containing broken hydrogen bonds compared to the hydrophobic environment.
- Tetrahedrality of water molecules near the protein surface is found to be more in the presence of high charge density Li^+ ions which preferentially binds with the protein surface and have less diffusion of water molecules compared to K^+ ion.
- IDP is found to be maximumly solvated at 0 M; where it forms a β-sheet structure. Li^+ ions were found to interact more with Asp144, Glu146, Asp147, and Asp152, residue whereas K^+ was found to interact with Cys167 on C-terminal ends.
- At higher ionic concentrations, the ion-water and ion-counter ion interactions increase, therefore the interaction of ions with protein decreases with a low survival probability value of ions near the protein surface.
- The residue-residue contact map showed segregated interactions between the residues which changed with the change in the temperature. This resulted to a different secondary structure.
- The hydrophobic region (C-terminal) of the IDP was folded into alpha-helix at 290 K and changed to beta-sheet at 300 K due to the fluctuations in the residues of the peptide.
- Network analysis was used to see the evolved conformations of RNA. At lower temperatures (303 K, 318 K), connections between the nodes are more, showing interchangeable conformations compared to higher temperatures (348 K, 363 K).

- Two types of network analysis are done based on RMSD and the interhelical angle of RNA. In angle-based clustering; the angle difference of ≤ ±10° gives a better sampling of RNA conformation compared to ±5°, ±20°, and ±30°.
- The non-native base pairs were found to be more in the mTIP3P water model compared to the SPC/E water model, the RNA molecule was found 80 % denatured at 348 K temperature in the mTIP3P water model.
- The CHARMM, AMBER force field, and mTIP3P, SPC/E water models have been used to see the influence of the hydration water on the RNA genome. The higher dielectric constant of the mTIP3P model screened out the attraction between ion pairs therefore the RNA molecule is more solvated in the mTIP3P water model compared to SPC/E.
- The formation of the ion-contact pair near the negatively charged phosphate group (Na^+-PO_4^-) leads to strong RNA-ion interaction and strong hydration shells in the SPC/E water model compared to mTIP3P. Further, it is confirmed by bias-free energy calculations using metadynamics.

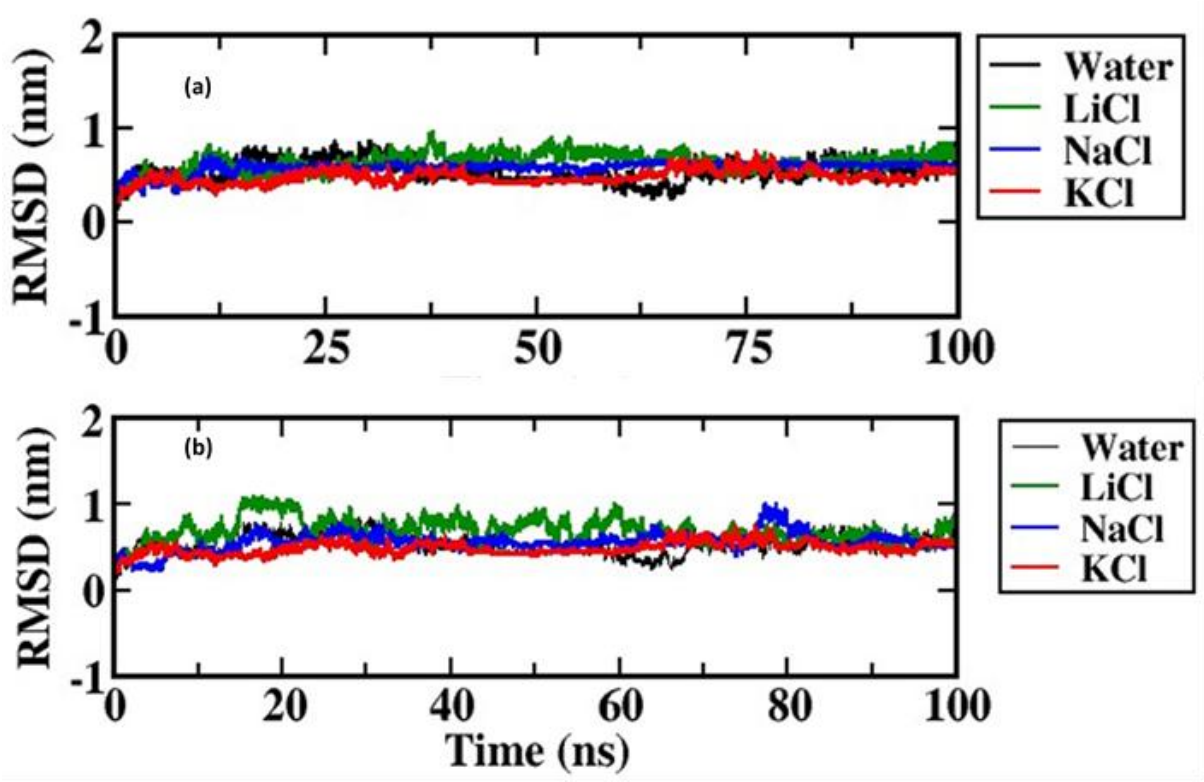
In atomistic molecular dynamic simulations methods metal ions and cosolvent or organic molecules can interact with the peptide and change its native conformations. It is found that the binding affinity of the metal ions towards the favorable amino acids or nucleic acids can change the functional property of the biomolecules. The proper interpretation of the non-covalent interactions under some thermodynamic conditions; such as hydrogen bonds, intramolecular interaction, and intermolecular interactions are responsible to alter the conformations of the molecules. In the future, the researchers can focus on investigating systems in which a nucleotide is not involved in base-pairing interactions under some conditions via higher-order sampling methods for example REMD and Umbrella sampling. A good study can be done based on the disordered protein by taking more ensembles and finding which residue is playing the major role in making the secondary structure. The base-flipping process in RNA can also be studied in association with peptides and some organic molecules binding to explore to what extent the molecules (Peptide, organic molecules) can affect the flipping of bases in RNA molecules.

APPENDIX

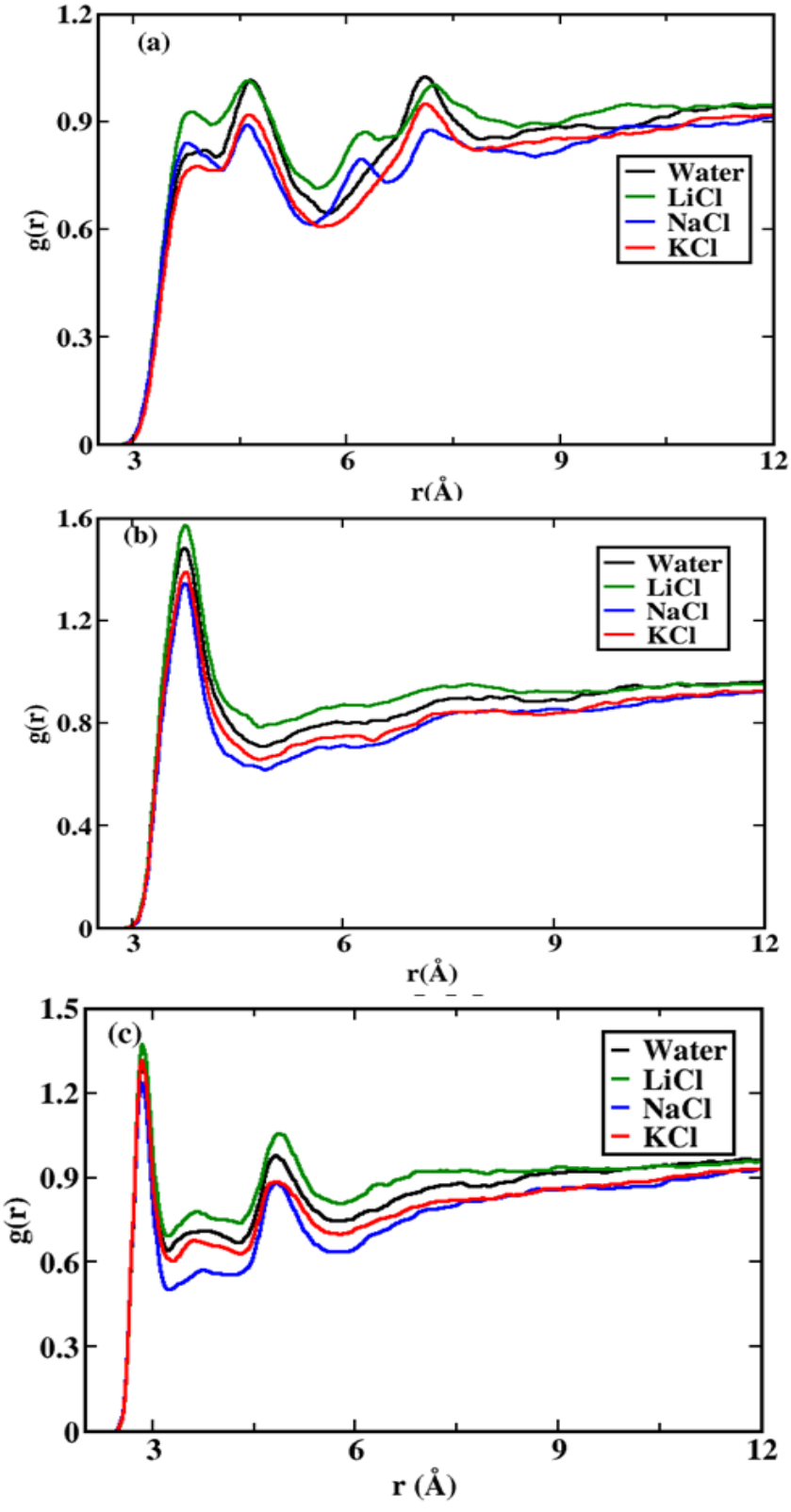
Appendix I: Values of Lennard-jones and electrostatic interaction potential parameters for LiCl, NaCl, KCl, and $MgCl_2$ CHARMM-FF and AMBER-FF.

CHARMM-FF Ions	**σ (Å)**	**ε(kJ/mol)**
LiCl	0.2312	0.00975
NaCl	0.2514	0.1962
KCl	0.3143	0.3640
$MgCl_2$	0.2114	0.0627
AMBER-FF Ions	**σ (Å)**	**ε(kJ/mol)**
LiCl	0.2026	0.0765
NaCl	0.3328	0.0116
KCl	0.3143	0.0014

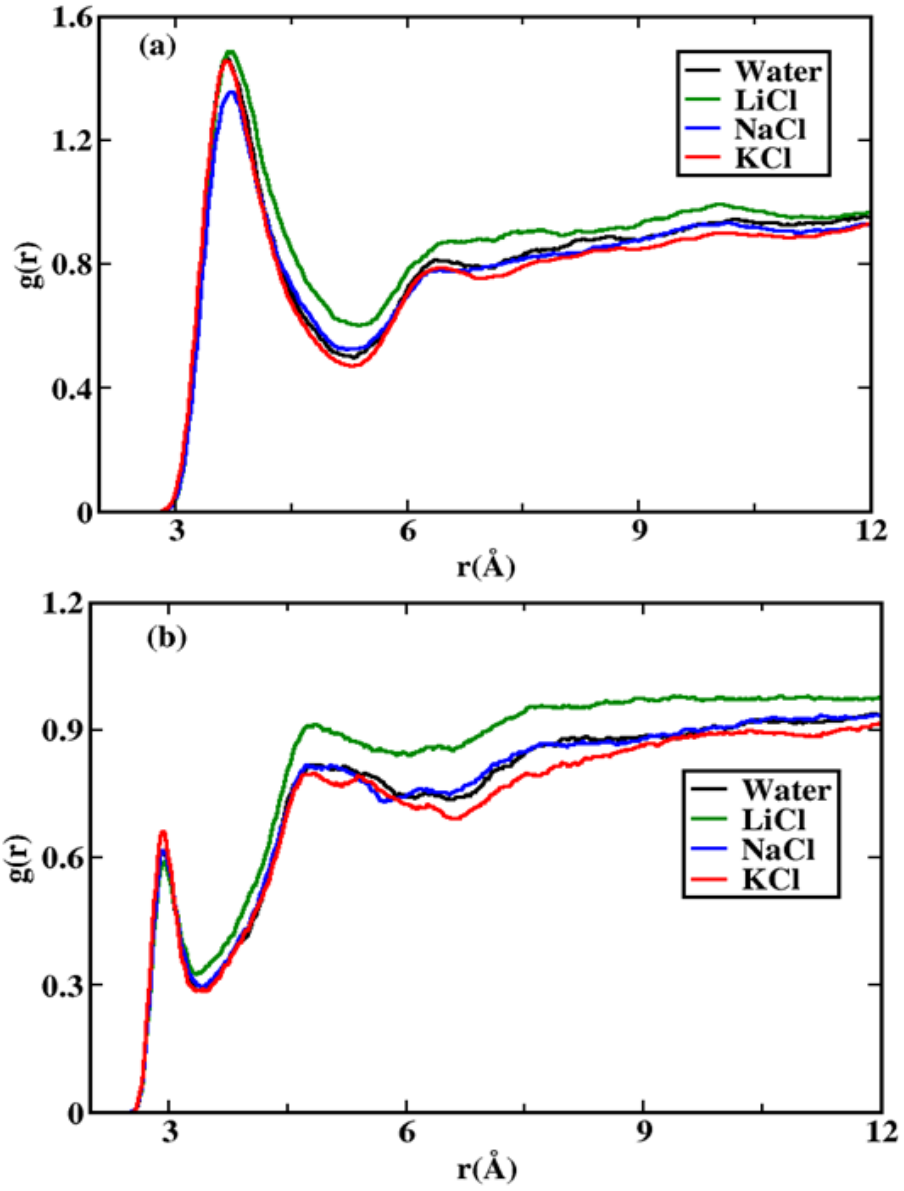
Appendix II: The RMSD graph for Cα of protein for (a) CHARMM-SPC/E (b) AMBER-TIP4P model systems.



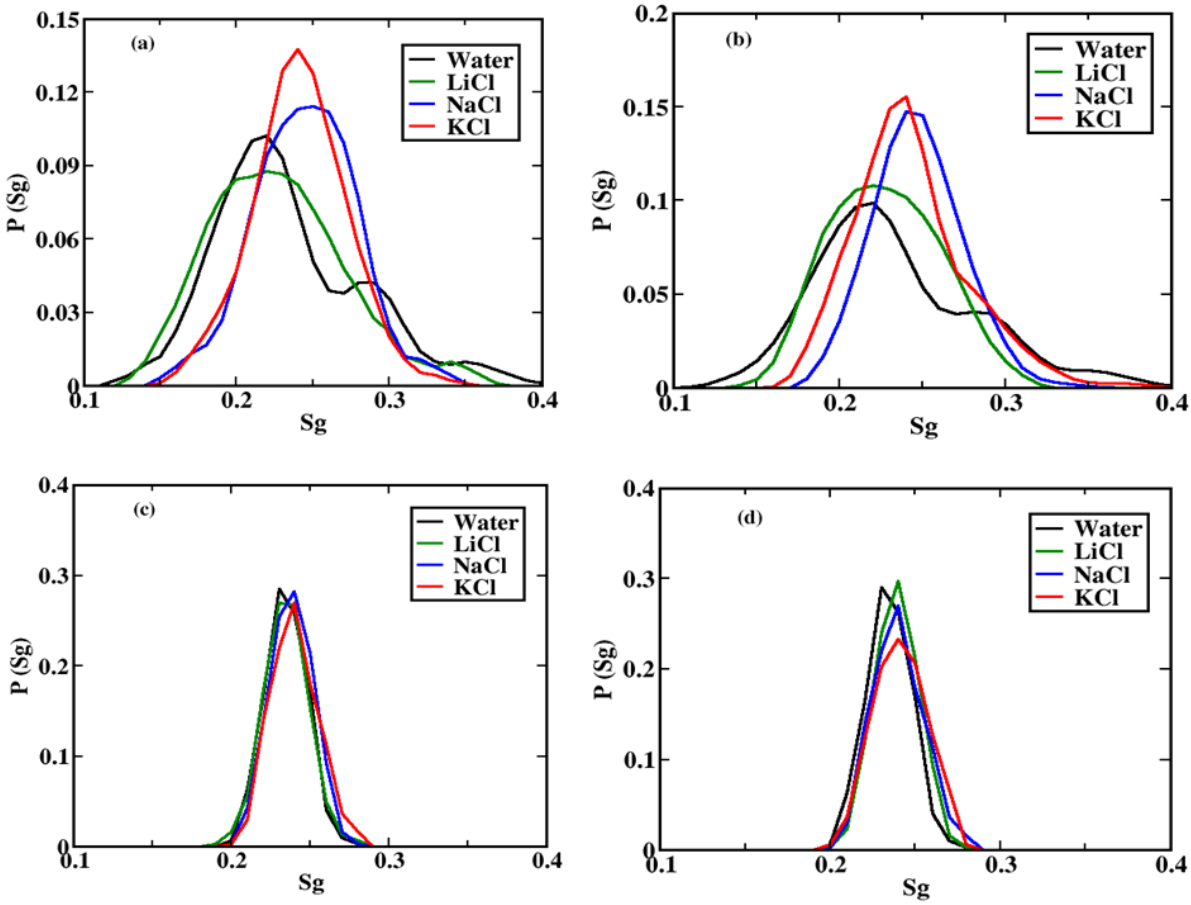
Appendix III: Radial Distribution functions g(r) of amino acid (ARG) with different salts for 1M concentration (a) C_γ-O_w (b) C_δ-O_w, (c) N_H-O_w for CHARMM-SPC/E model systems.



Appendix IV: Radial Distribution functions g(r) of amino acid (ALA) with different salts for 1M concentration (a) C_{β}-O_w (ALA), (b) N_H-O_w ALA for CHARMM-SPC/E model systems.



Appendix V: The orientation order parameter of water molecules within the distances [(a), (b) 4.3 Å] and [(c), (d) > 9 Å] from C_α of ALA and ARG amino acid residues of protein in presence of various salt solutions 1.0 M for AMBER-TIP4P model systems.



Appendix VI: Diffusion coefficient of water molecules around (< 4.3 Å) amino acids of Antimicrobial protein in water for CHARMM-SPC/E and AMBER-TIP4P model system, LiCl, NaCl and KCl, D = (value) x 10^{-5}cm2/sec.

SPCE	Time (ns)	Water	LiCl	NaCl	KCl
ALA	0-25	2.46 ±0.036	2.24 ±0.018	2.41 ±0.182	2.47 ±0.035
	25-50	2.40 ±0.032	2.18 ±0.011	2.31 ±0.012	2.38 ±0.042
	50-75	2.45 ±0.034	2.38 ±0.126	2.34 ±0.213	2.44 ±0.124
	75-100	2.49 ±0.140	2.32 ±0.017	2.38 ±0.029	2.35 ±0.033
ARG	0-25	2.49 ±0.024	2.32 ±0.113	2.38 ±0.031	2.46 ±0.024
	25-50	2.52 ±0.032	2.13 ±0.007	2.35 ±0.043	2.36 ±0.026
	50-75	2.45 ±0.023	2.21 ±0.021	2.25 ±0.021	2.32 ±0.032
	75-100	2.46 ±0.018	2.04 ±0.012	2.22 ±0.031	2.38 ±0.041
TIP4P	Time (ns)	Water	LiCl	NaCl	KCl
ALA	0-25	3.34 ±0.181	2.83 ±0.080	2.72 ±0.087	2.66 ±0.081
	25-50	3.26 ±0.197	2.87 ±0.196	2.68 ±0.067	2.64 ±0.190
	50-75	3.24 ±0.043	2.68 ±0.082	2.76 ±0.171	2.63 ±0.072
	75-100	3.32 ±0.243	2.90 ±0.099	2.61 ±0.231	2.67 ±0.157
ARG	0-25	3.12 ±0.087	2.51 ±0.240	2.62 ±0.028	2.57 ±0.055
	25-50	3.25 ±0.183	2.65 ±0.056	2.49 ±0.023	2.48 ±0.162
	50-75	3.22 ±0.192	2.61 ±0.103	2.55 ±0.136	2.53 ±0.076
	75-100	3.09 ±0.042	2.46 ±0.243	2.42 ±0.032	2.45 ±0.106

Appendix VII: Diffusion coefficient of water molecules around (> 9.0 Å) amino acids of antimicrobial protein in water for CHARMM-SPC/E and AMBER-TIP4P model system, LiCl, NaCl, and KCl, D = (value) x 10^{-5}cm2/sec.

SPCE	Time (ns)	Water	LiCl	NaCl	KCl
ALA	0-25	2.55 ±0.028	2.41 ±0.108	2.48 ±0.036	2.43 ±0.014
	25-50	2.51 ±0.022	2.43 ±0.021	2.39 ±0.027	2.33 ±0.143
	50-75	2.48 ±0.104	2.37 ±0.026	2.43 ±0.034	2.31 ±0.016
	75-100	2.50 ±0.031	2.35 ±0.183	2.35 ±0.018	2.37 ±0.033
ARG	0-25	2.57 ±0.029	2.38 ±0.035	2.39 ±0.026	2.51 ±0.026
	25-50	2.52 ±0.022	2.43 ±0.025	2.45 ±0.044	2.42 ±0.032
	50-75	2.49 ±0.016	2.46 ±0.017	2.50 ±0.042	2.48 ±0.013
	75-100	2.54 ±0.028	2.37 ±0.051	2.46 ±0.023	2.43 ±0.031
TIP4P	Time (ns)	Water	LiCl	NaCl	KCl
ALA	0-25	3.54 ±0.087	3.32 ±0.181	3.36 ±0.024	3.62 ±0.072
	25-50	3.58 ±0.128	3.27 ±0.019	3.33 ±0.047	3.54 ±0.182
	50-75	3.46 ±0.164	3.28 ±0.010	3.31 ±0.016	3.58 0.0824
	75-100	3.52 ±0.034	3.24 ±0.090	3.28 ±0.006	3.48 ±0.124
ARG	0-25	3.58 ±0.015	3.26 ±0.137	3.46 ±0.027	3.48 ±0.026
	25-50	3.56 ±0.152	3.19 ±0.071	3.32 ±0.023	3.52 ±0.057
	50-75	3.52 ±0.054	3.23 ±0.067	3.47 ±0.077	3.54 ±0.177
	75-100	3.51 ±0.036	3.17 ±0.031	3.29 ±0.035	3.46 ±0.156

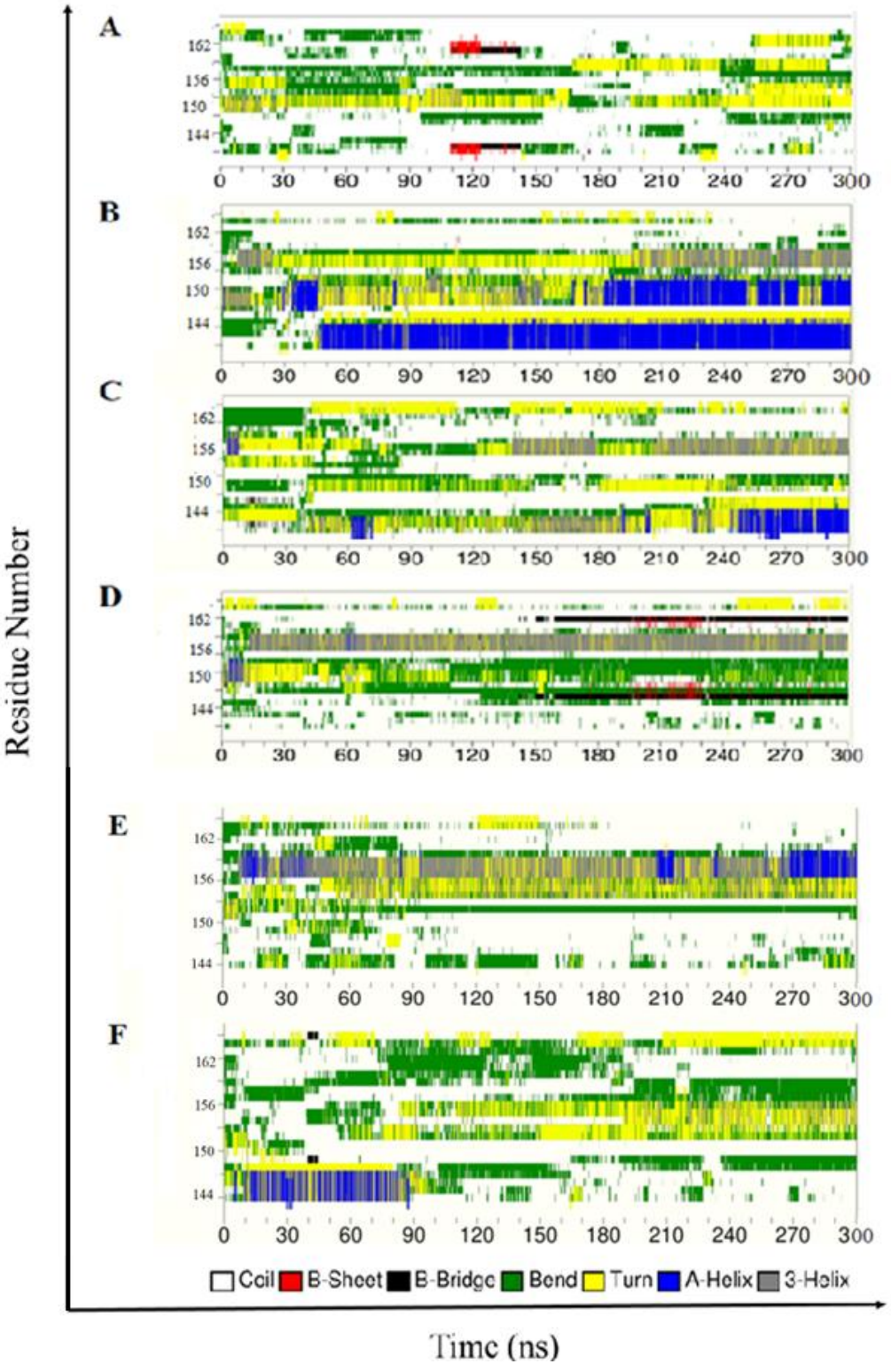
Appendix VIII: Hydrogen bond occupancy between amino acid residues in salts solutions at different concentrations.

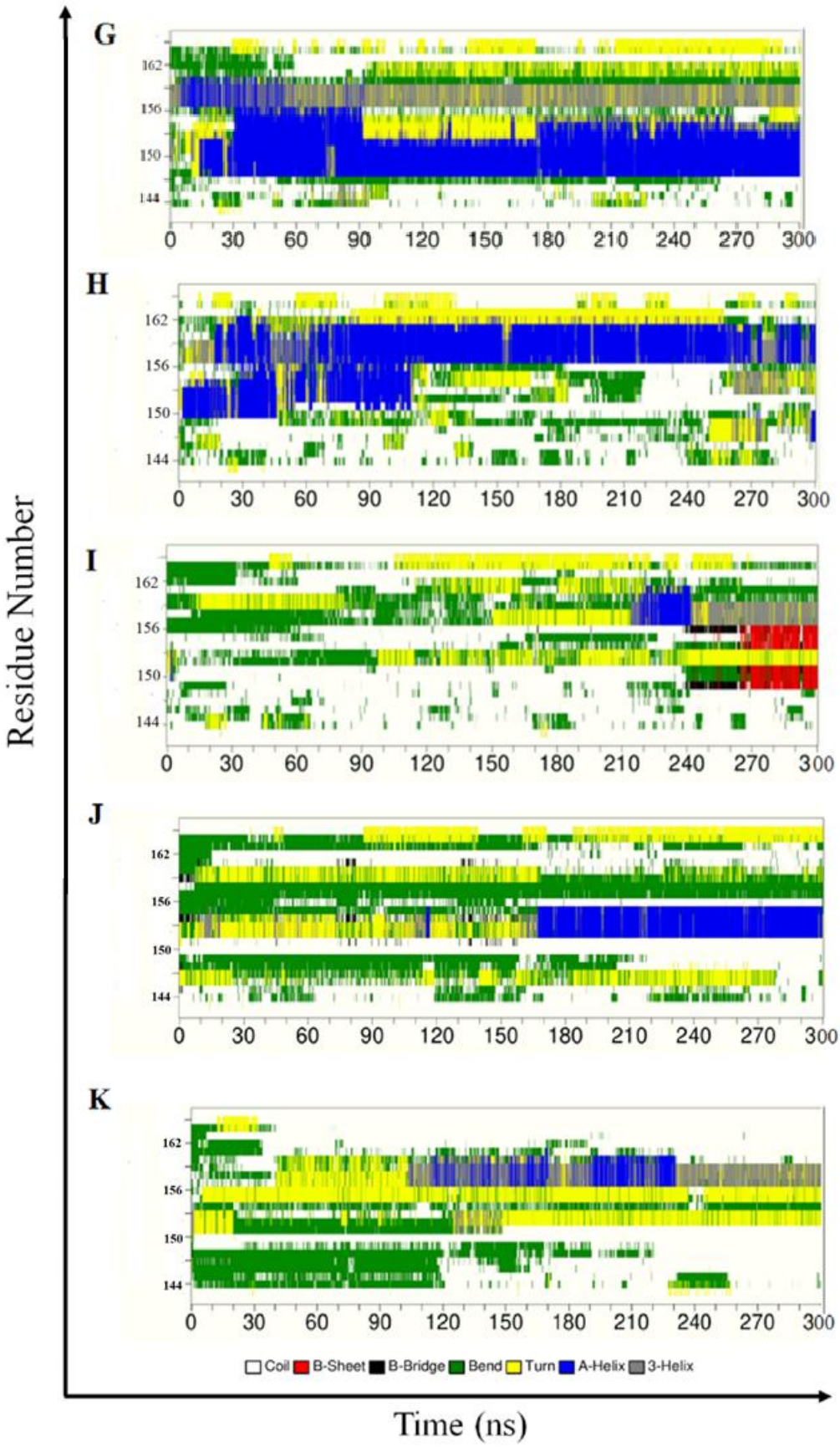
System	Hydrogen bonded residue	Percentage (%)
0.03 M LiCl	162TYR (HN) - 154MET(O)	82.6
	160GLN (HN) - 156ALA (O)	30.2
	159ASN (HN) - 156ALA(O)	45.1
	156ALA (HN) - 160GLN(O)	79.6
	154MET (HN) - 162TYR(O)	36.8
0.3 M LiCl	167CYS (HN) - 164ARG(O)	48.7
	163TYR (HN) - 153ASN(OD1)	66.0
	162TYR (HN) - 153ASN(OD1)	51.7
	159ASN (HN) - 156ALA(O)	60.4
	156ALA (HN) - 153ASN(O)	30.2
	155TYR (HN) - 151ARG(O)	51.2
	154MET (HN) - 150TYR(O)	78.2
	153ASN(D21) - 163TYR(O)	58.2
	153ASN (HN) - 149TYR(O)	78.1
	152GLU (HN) - 148ARG(O)	66.4
	151ARG (HN) - 147ASP(O)	64.3
	150TYR (HN) - 146GLU(O)	67.0
	149TYR (HN) - 145TYR(O)	72.0
	148ARG (HN) - 144ASP(O)	49.9
	147ASP (HN) - 143ASN(O)	30.8
0.6 M LiCl	167CYS (HN) - 146GLU(OE1)	31.2
	160GLN (HN) - 157TYR(O)	30.2
	159ASN (HN) - 156ALA(O)	56.3
	155TYR (HN) - 150TYR(O)	42.2
	153ASN (D21) - 155TYR(OH)	32.2
	153ASN (HN) - 149TYR(O)	61.1
	152GLU (HN) - 149TYR(O)	37.6
	148ARG(H11) - 145TYR(O)	35.7

	148ARG (HN) - 145TYR(O)	75.1
0.9 M LiCl	166VAL (HN) - 163TYR(O)	37.9
	163TYR (HN) - 160GLN(O)	36.5
	154MET (HN) - 149TYR(O)	38.6
	153ASN (HN) - 150TYR(O)	44.5
	152GLU (HN) - 148ARG(O)	54.4
	151ARG (HN) - 147ASP(O)	46.4
	150TYR (HN) - 147ASP(O)	31.3
0.3 M NaCl	155TYR(HN) - 151ARG(O)	39.3
	154MET(HN) - 151ARG(O)	49.6
	148ARG(HN) - 145TYR(O)	30.25
0.6 M NaCl	167CYS(HN) - 164ARG(O)	48.20
	166VAL(HN) - 163TYR(O)	31.20
	149TYR(HN) - 157TYR(OH)	32.51
0.9 M NaCl	166VAL(HN) - 163TYR(O)	41.00
	164ARG(HN) - 151ARG(O)	67.10
	153ASN(D21) - 149TYR(O)	43.10
	152GLU(HN) - 148ARG(O)	56.80
	151ARG(H11) - 162TYR(O)	66.60
	150TYR(HN) - 147ASP(O)	41.30
	146GLU(HN) - 143ASN(O)	70.00
0.3 M KCl	166VAL (HN) - 144ASP(O)	31.2
	160GLN (E21) - 158PRO(O)	62.2
	160GLN (E21) - 157TYR(O)	55.1
	153ASN (D21) - 149TYR(O)	31.4
	148ARG (HN) - 145TYR(O)	60.0
0.9 M KCl	166VAL (HN) - 144ASP(O)	81.1
	153ASN (HN) - 150TYR(O)	45.8
	148ARG (HN) - 145TYR(O)	30.2
	146GLU (HN) - 164ARG(O)	69.9
	150TYR (HH) - 155TYR(N)	76.4

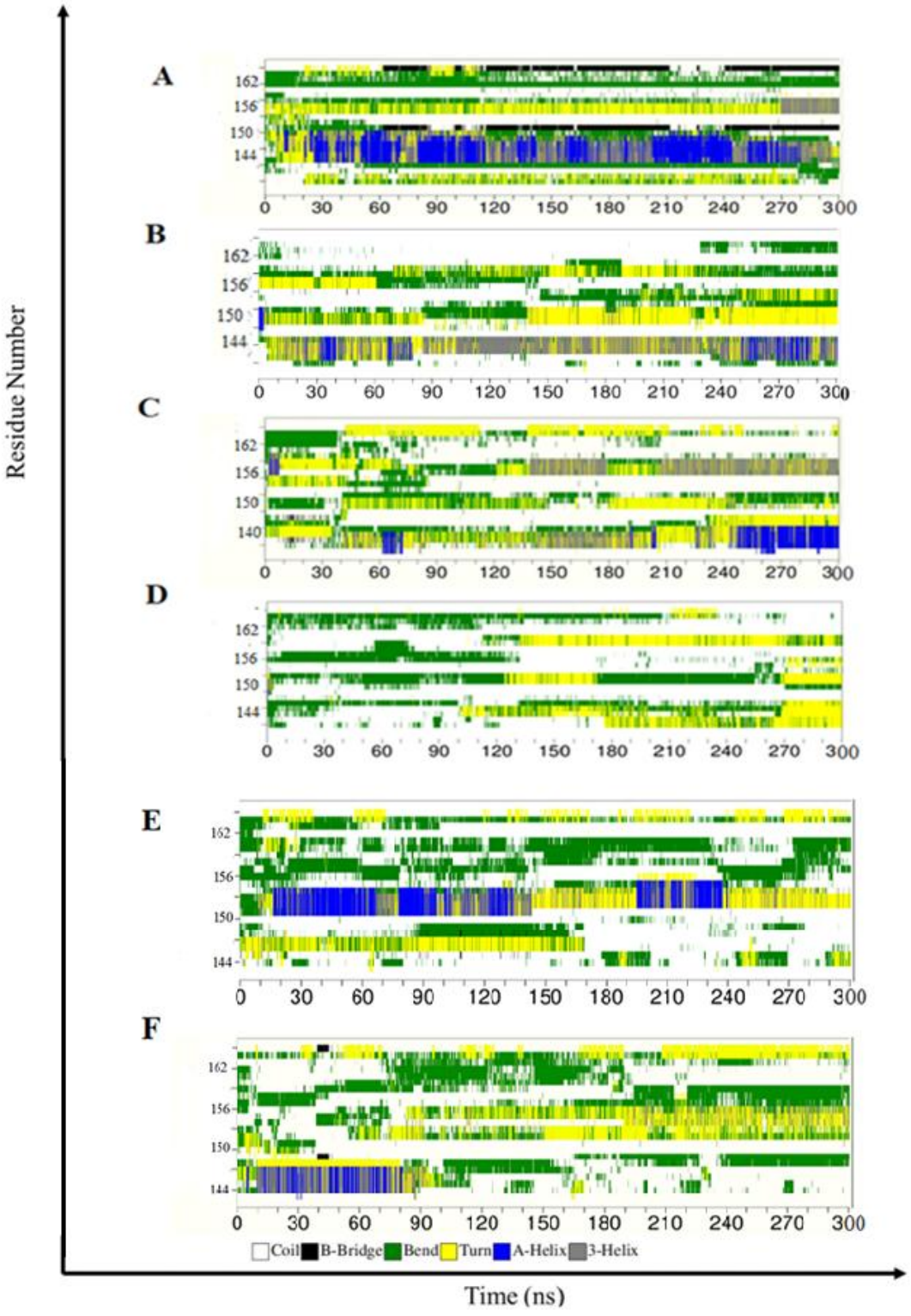
	143ASN (D21 - 153ASN(O)	31.2
0.1 M $MgCl_2$	167CYS(HN) - 164ARG(O)	30.0
	166VAL(HN) - 163TYR(O)	35.5
	162TYR (HN) - 160GLN(OE1)	33.7
	155TYR (HH) - 164ARG(N)	45.5
	155TYR (HN) - 161VAL(O)	88.3
	154MET (HN) - 151ARG(O)	35.6
	153ASN (HN) - 150TYR(O)	30.6
	153ASN (HN) - 149TYR(O)	30.4
	152GLU (HN) - 148ARG(O)	63.5
	151ARG (HN) - 147ASP(O)	64.9
	145TYR (HN) - 143ASN(OD1)	63.0
0.3 M $MgCl_2$	164ARG(H11) - 162TYR(O)	30.2
	157TYR (HH) - 153ASN(O)	36.1
	148ARG (H11) - 145TYR(O)	31.3
	148ARG (HN) - 145TYR(O)	32.3

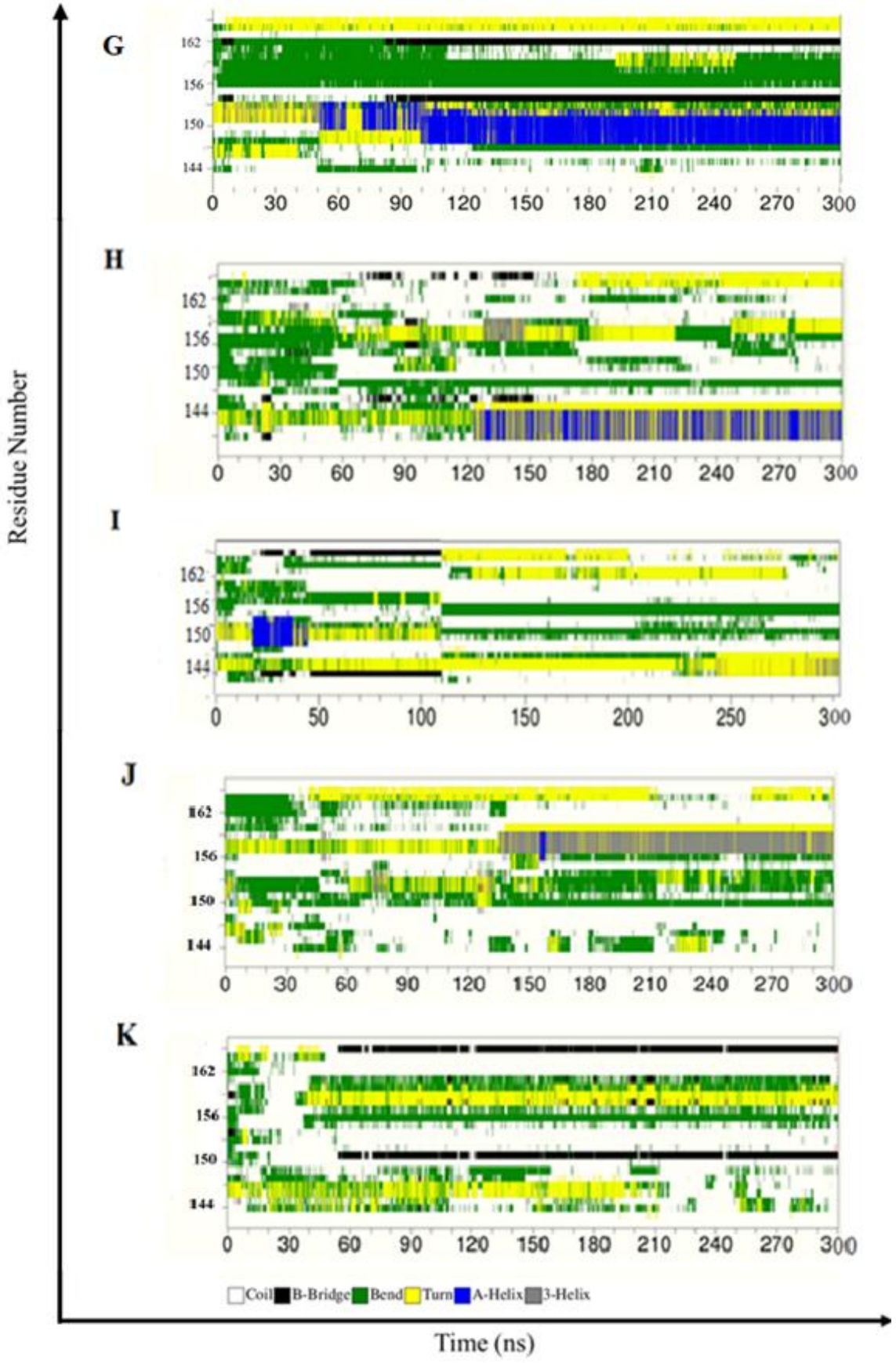
Appendix IX: 2 set; Time evaluation of Residue wise secondary structure composition throughout the 300 ns simulation trajectory. (A) 0 M, (B) 0.3 M LiCl, (C) 0.6 M LiCl, (D) 0.9 M LiCl, (E) 0.3 M NaCl, (F) 0.6 M NaCl, (G) 0.9 M NaCl, (H) 0.3 M KCl, (I) 0.9 M KCl, and (J) 0.1 M $MgCl_2$, (K) 0.3 M $MgCl_2$.



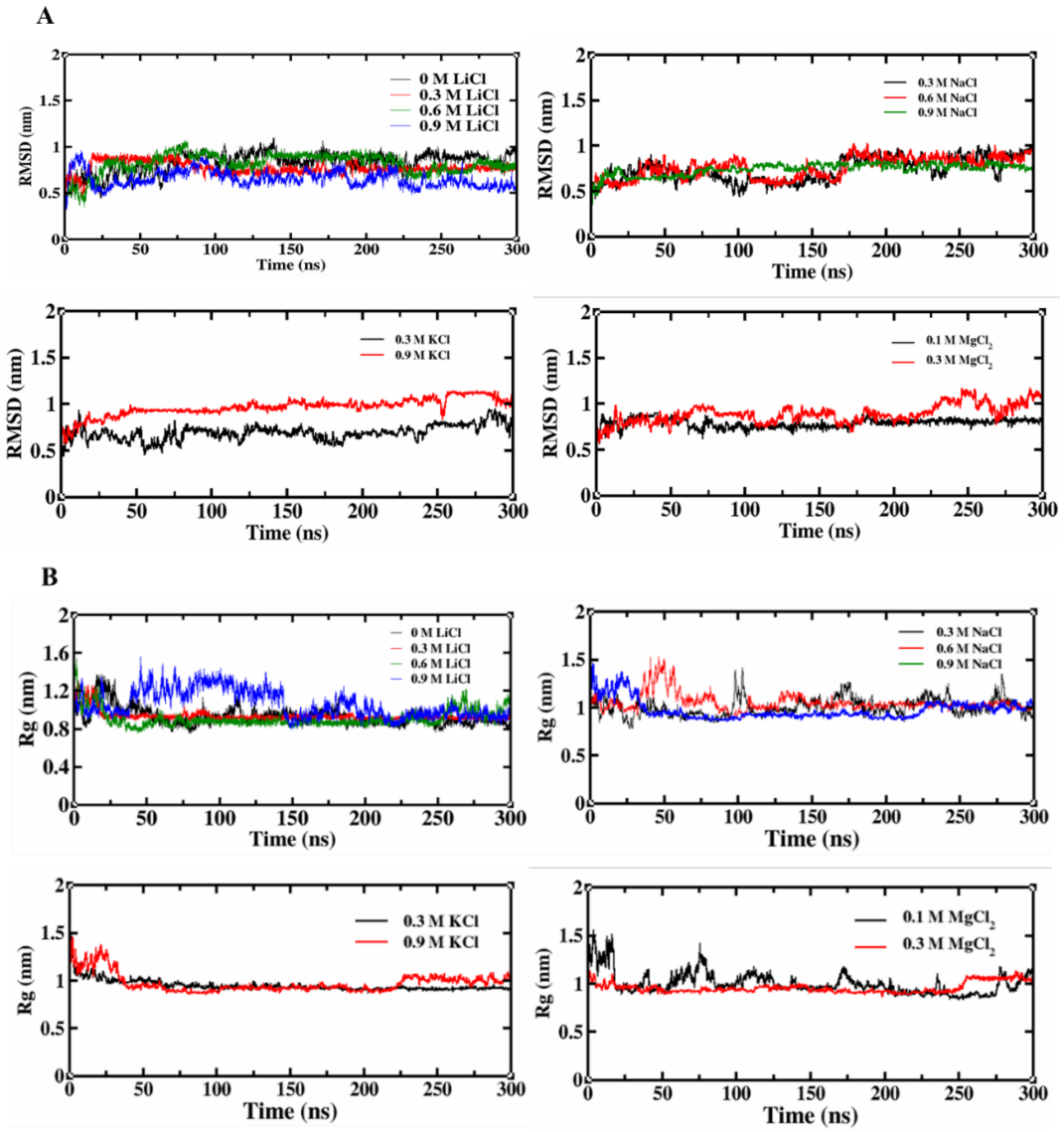

G
H
I
J
K
Residue Number
162
156
150
144
0
30
60
90
120
150
180
210
240
270
300
Coil
B-Sheet
B-Bridge
Bend
Turn
A-Helix
3-Helix
Time (ns)

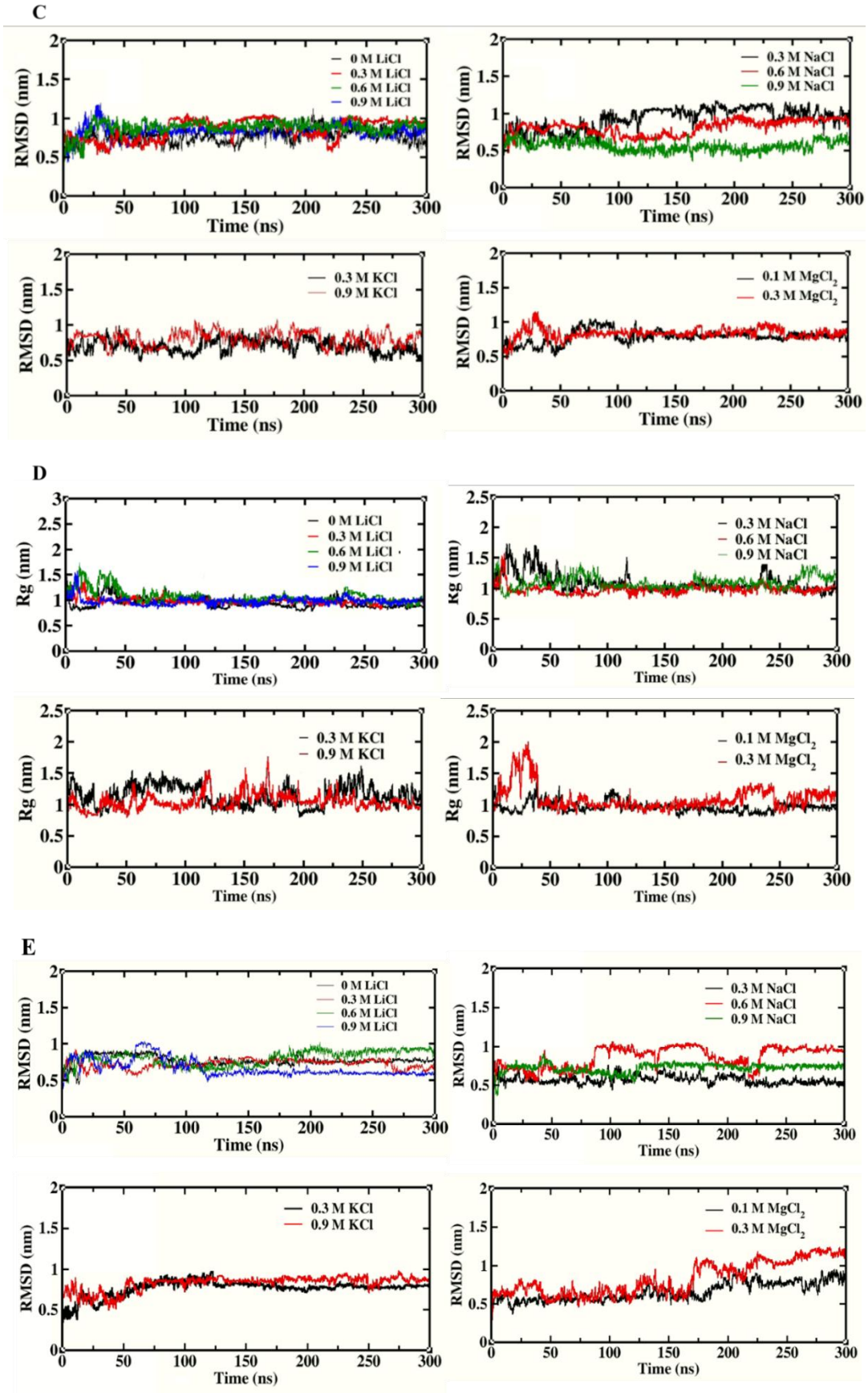
Appendix X: Third set; Time evaluation of Residue wise secondary structure composition throughout the 300 ns simulation trajectory. (A) 0 M, (B) 0.3 M LiCl, (C) 0.6 M LiCl, (D) 0.9 M LiCl, (E) 0.3 M NaCl, (F) 0.6 M NaCl, (G) 0.9 M NaCl, (H) 0.3 M KCl, (I) 0.9 M KCl, and (J) 0.1 M $MgCl_2$, (K) 0.3 M $MgCl_2$.



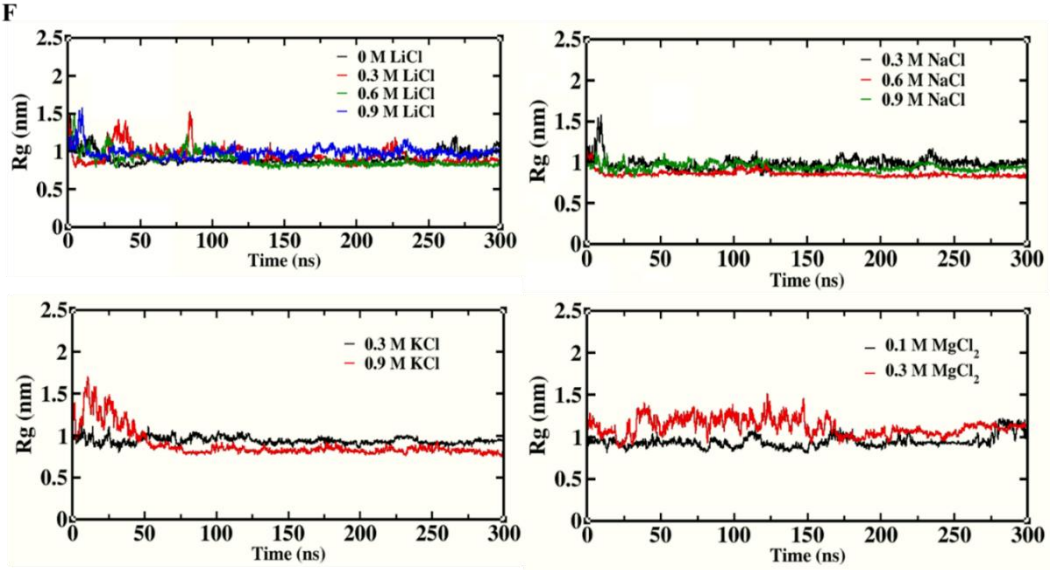

G
H
I
J
K
Residue Number
162
156
150
144
0 30 60 90 120 150 180 210 240 270 300
0 50 100 150 200 250 300
Coil B-Bridge Bend Turn A-Helix 3-Helix
Time (ns)

Appendix XI: Simulation Set-1- (A) Time evolution of RMSD profile of protein backbone and (B) radius of Gyration of protein over 300 ns trajectory. **Set-2-** (C) RMSD profile and (D) radius of Gyration. **Set-3-** (E) RMSD profile and (F) radius of Gyration.

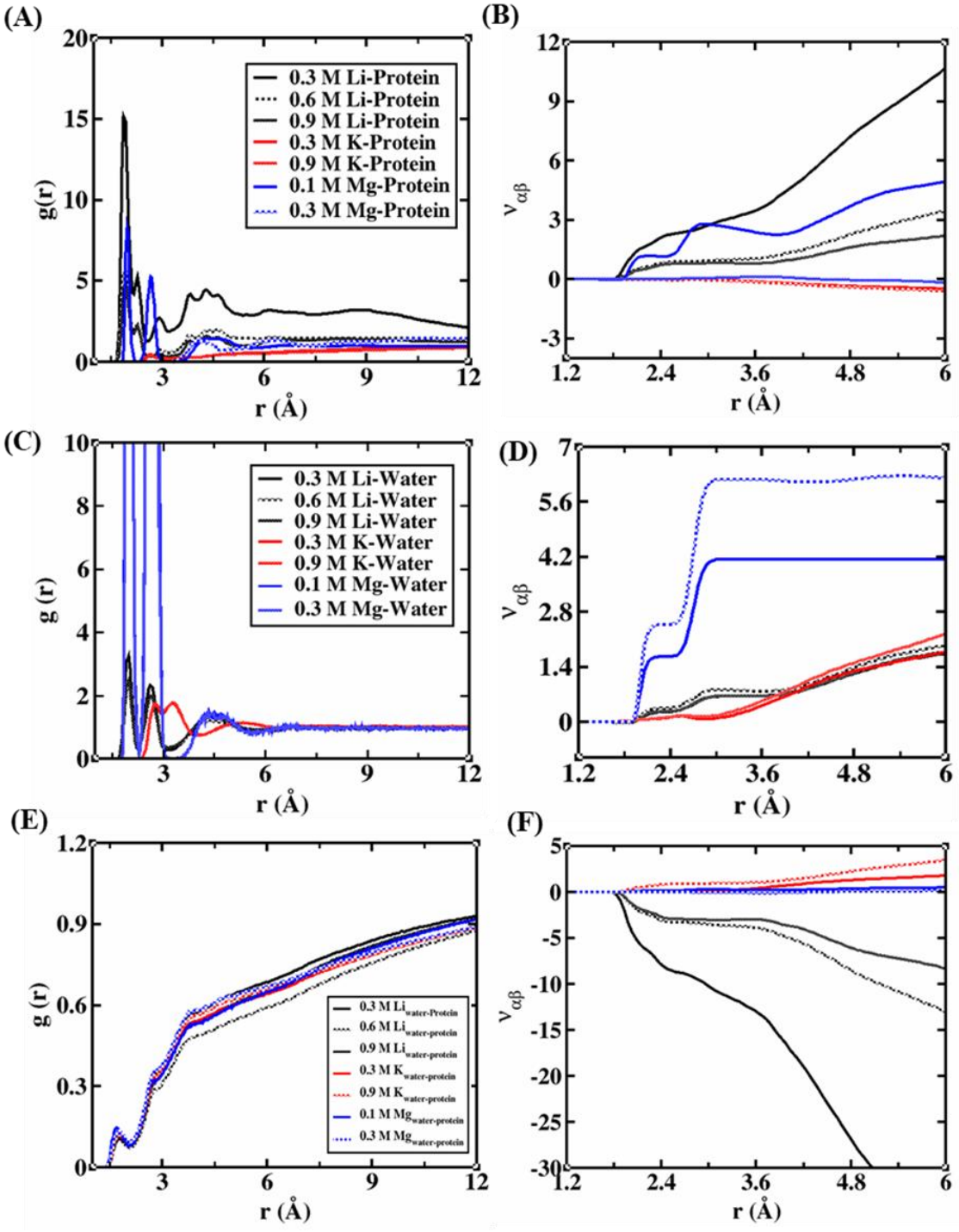



C
RMSD (nm)
Time (ns)
0 M LiCl
0.3 M LiCl
0.6 M LiCl
0.9 M LiCl
0.3 M NaCl
0.6 M NaCl
0.9 M NaCl
0.3 M KCl
0.9 M KCl
0.1 M MgCl2
0.3 M MgCl2
D
Rg (nm)
Time (ns)
E
RMSD (nm)
Time (ns)

F

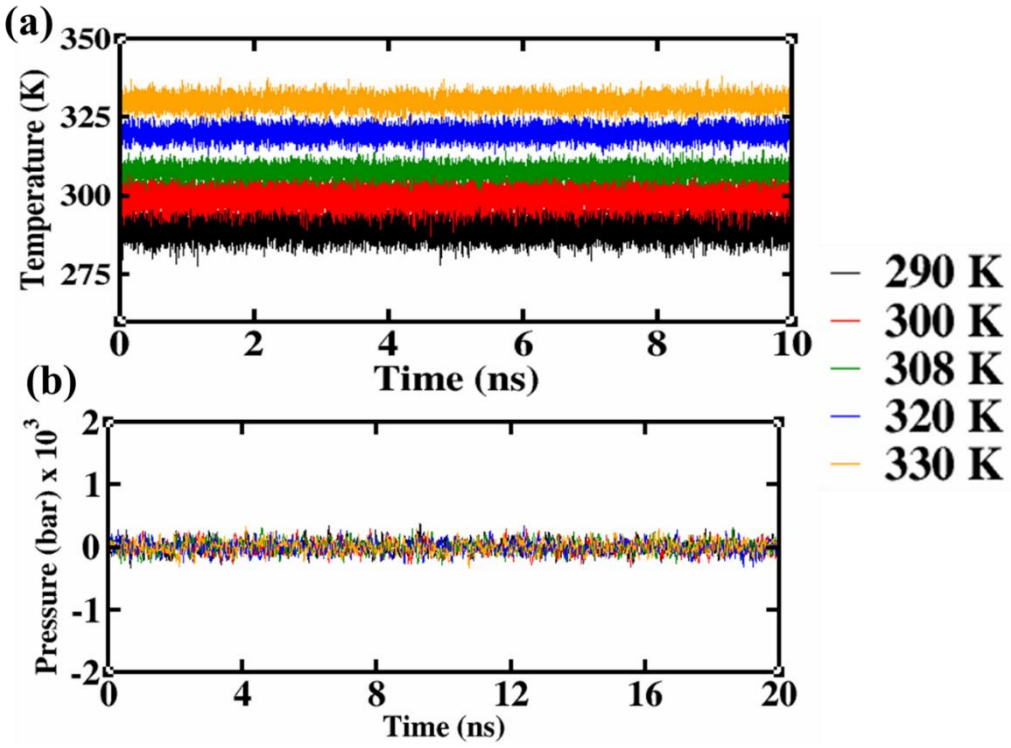

0 M LiCl
0.3 M LiCl
0.6 M LiCl
0.9 M LiCl
Rg (nm)
Time (ns)
0.3 M NaCl
0.6 M NaCl
0.9 M NaCl
Rg (nm)
Time (ns)
0.3 M KCl
0.9 M KCl
Rg (nm)
Time (ns)
0.1 M MgCl2
0.3 M MgCl2
Rg (nm)
Time (ns)

Appendix XII: Radial distribution function and Preferential binding coefficients (PBC) $v_{\alpha\beta}$ for (A) RDF of Protein-Ions (B) PBC of Protein-Ions (C) RDF of Ions-Water (D) PBC of Ions-Water (E) RDF of Protein-Water and (F) PBC of Protein-Water in different salts solutions.

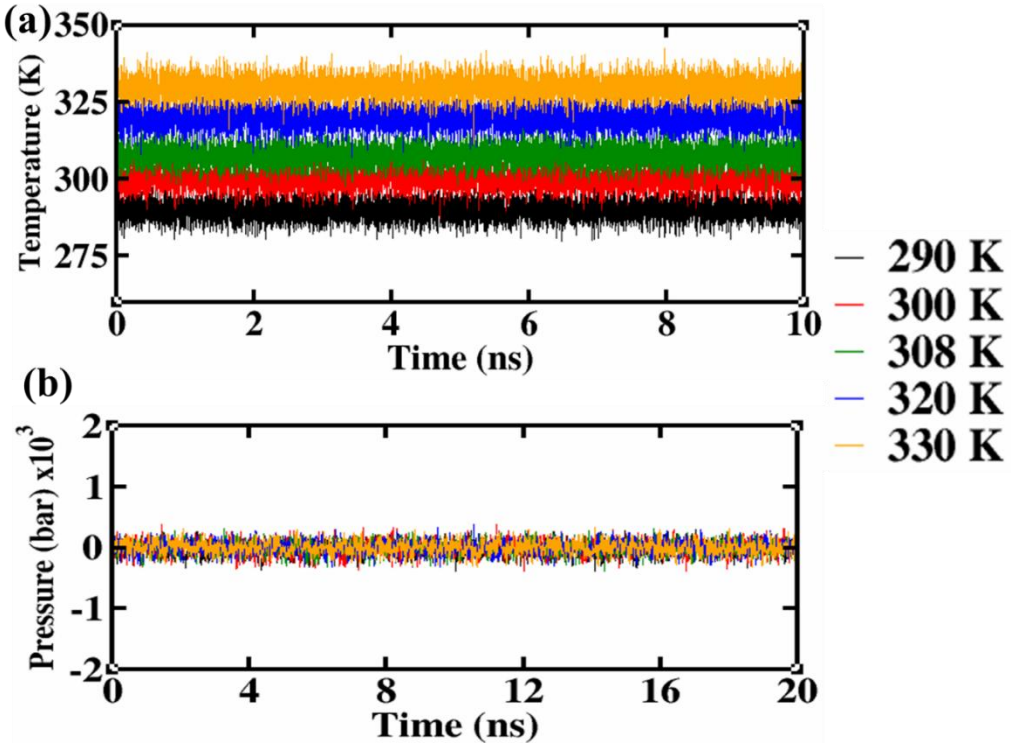


Appendix XIII: Thermodynamic equilibrium graph (a) Time evolution of Temperature (b) Pressure at different temperatures.

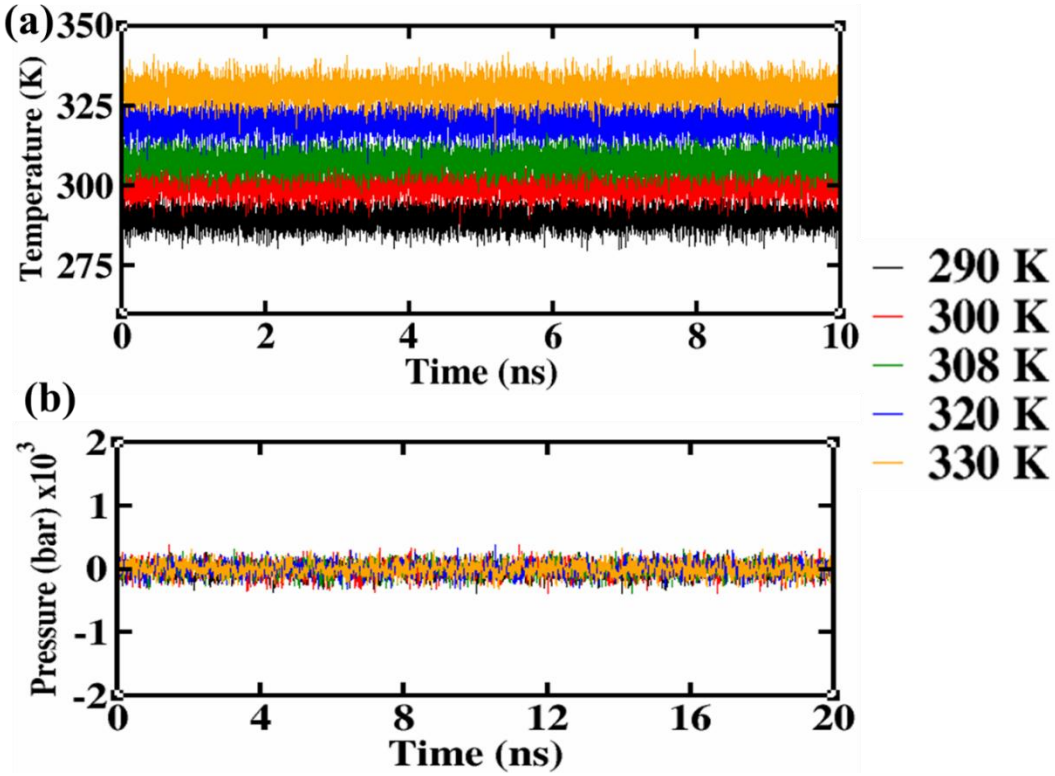
Set I



Set II

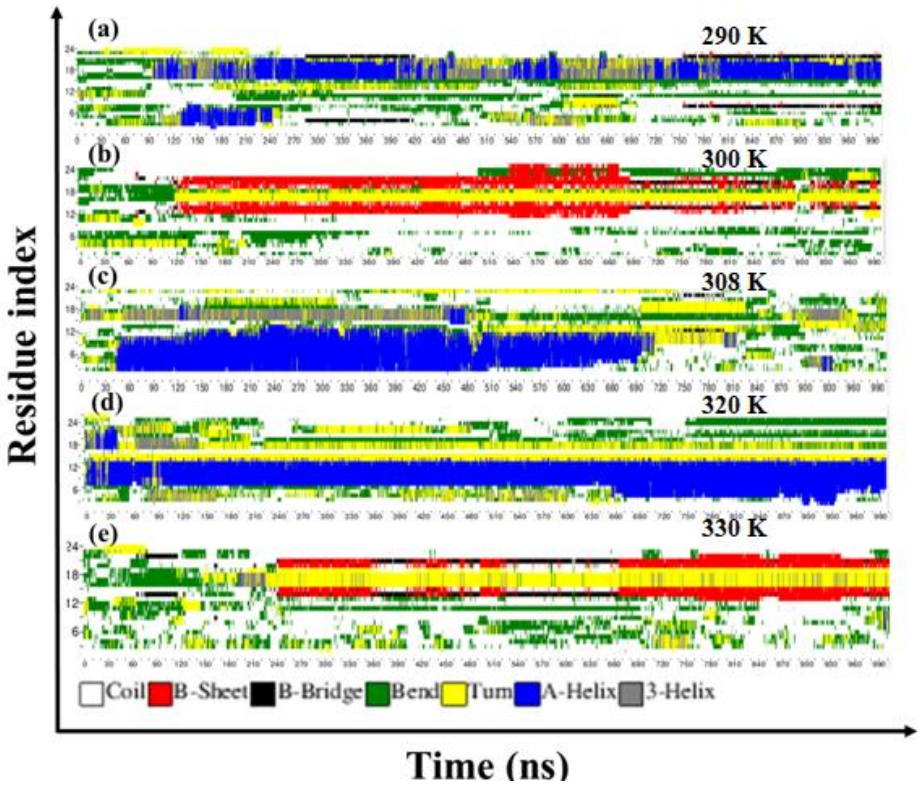


Set III

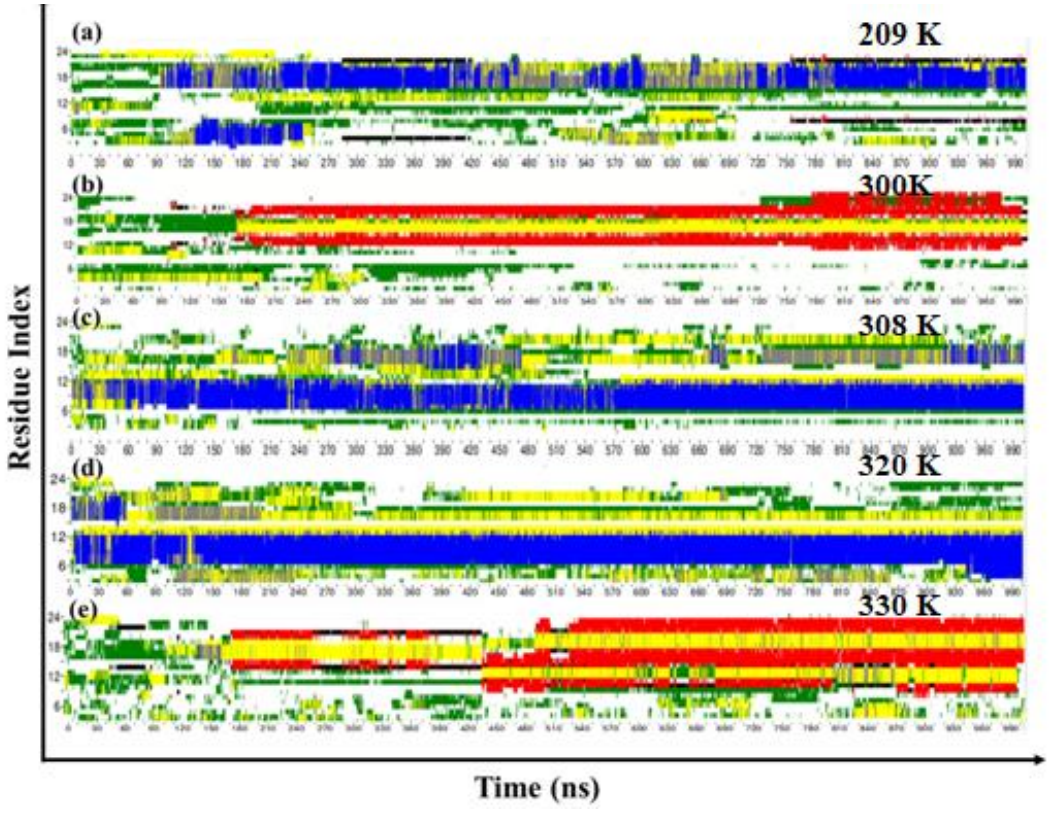

(a)
Temperature (K)
350
325
300
275
0
2
4
6
8
10
Time (ns)
290 K
300 K
308 K
320 K
330 K
(b)
Pressure (bar) x10^3
2
1
0
-1
-2
0
4
8
12
16
20
Time (ns)

Appendix XIV: Time evaluation of Residue wise secondary structure composition throughout the simulation trajectory at all 5 temperatures.

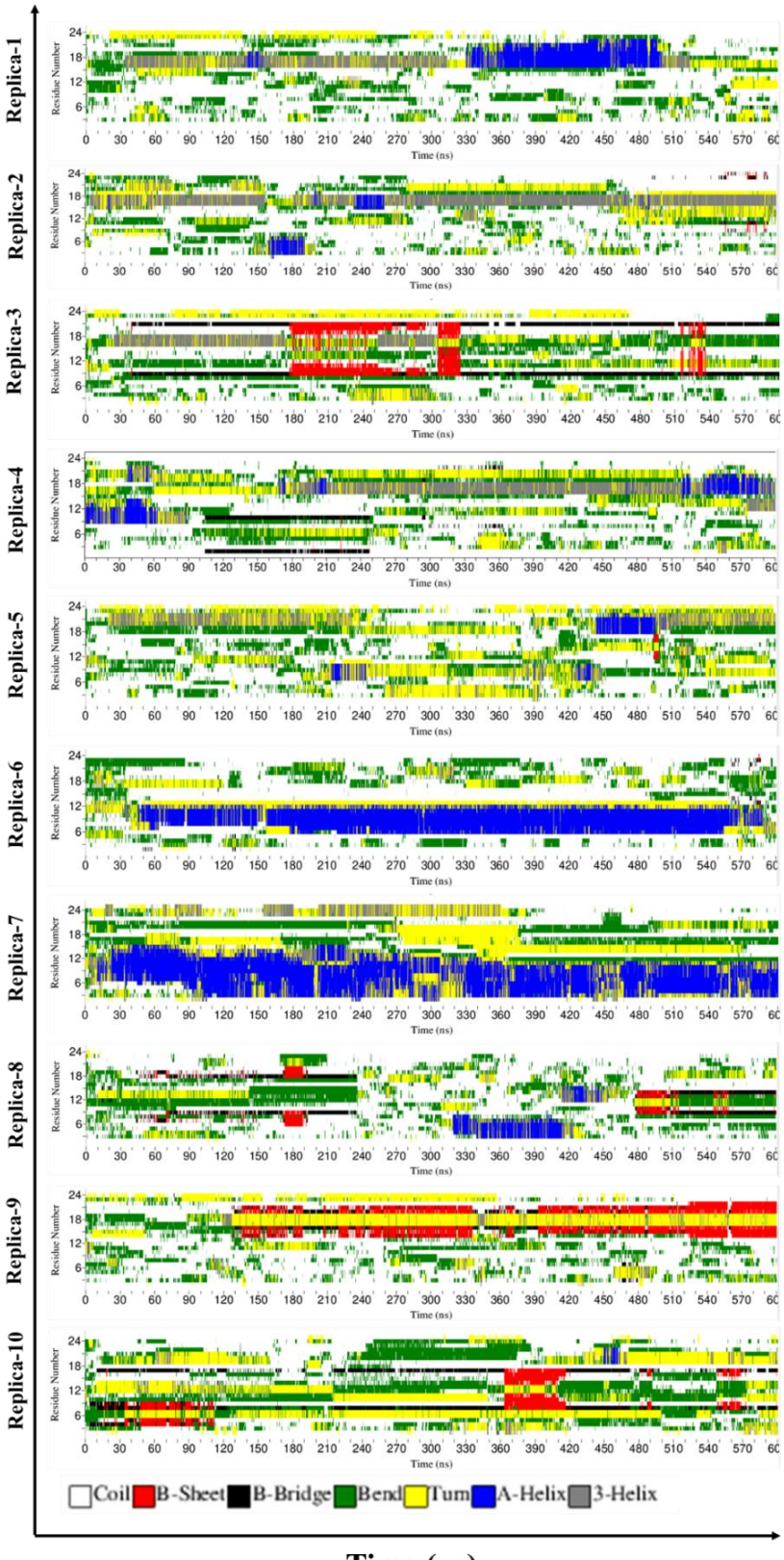
Set II



Set III



Appendix XV: Time evaluation of Residue wise secondary structure composition throughout the REMD simulation trajectory.

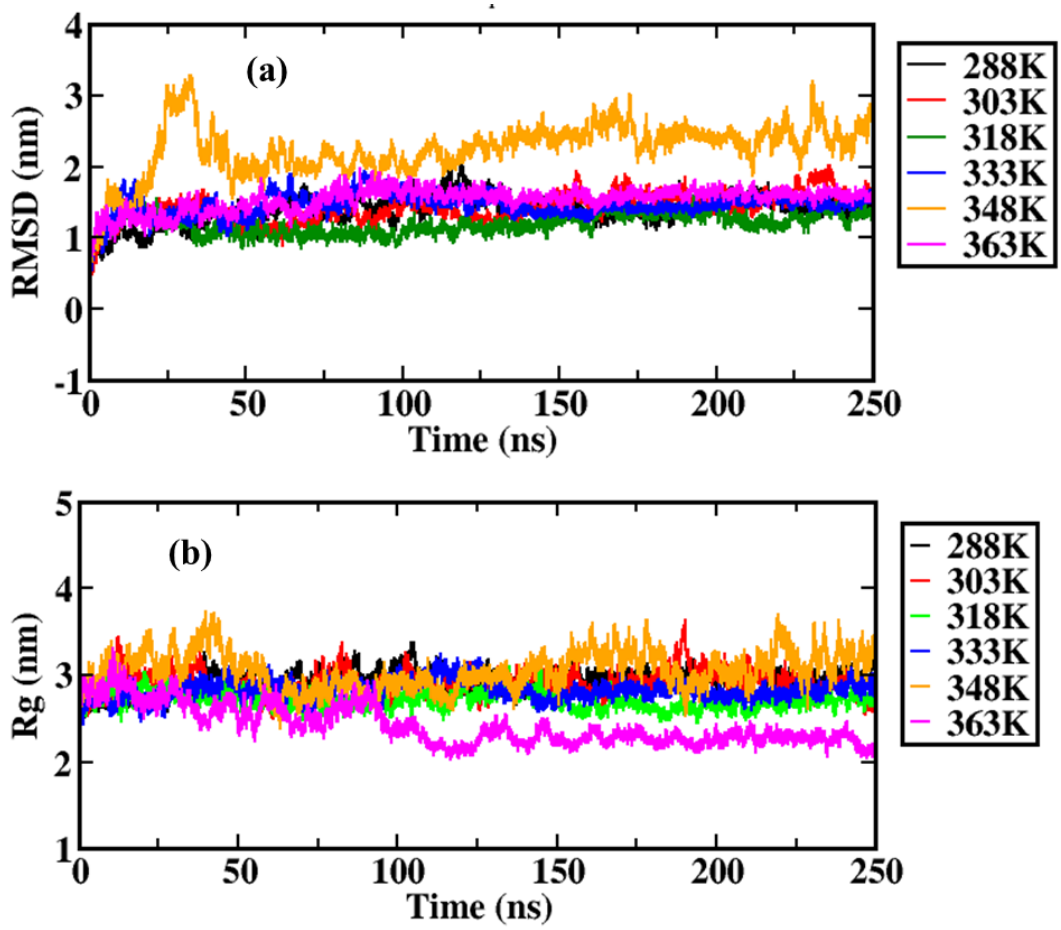


Appendix XVI: Cosine values calculated on the first principal components of every simulation trajectory for different temperature.

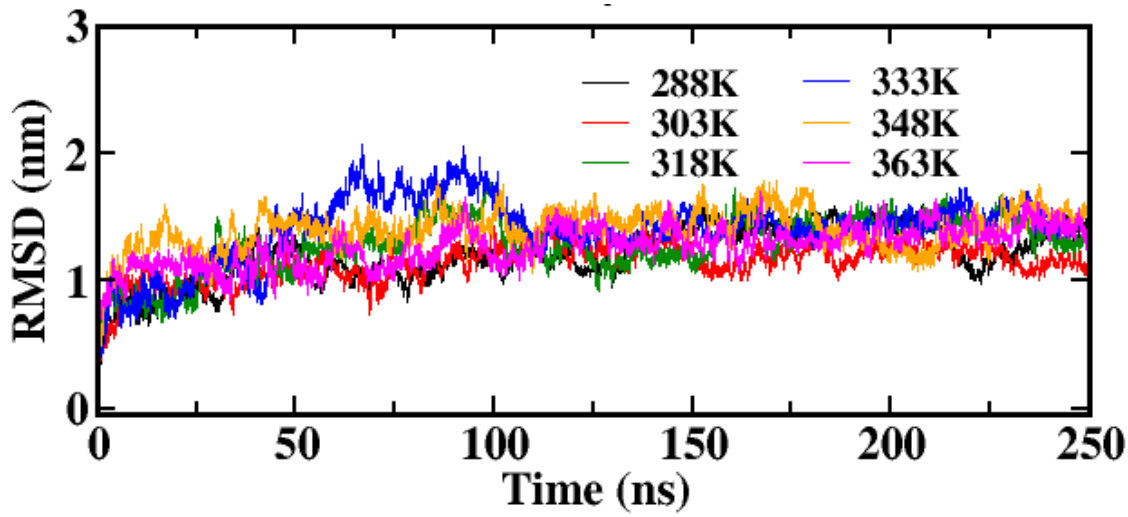
Temperature (K)	Cosine values
290	0.20
295	0.43
300	0.23
305	0.24
308	0.08
315	0.12
320	0.11
325	0.23
330	0.42

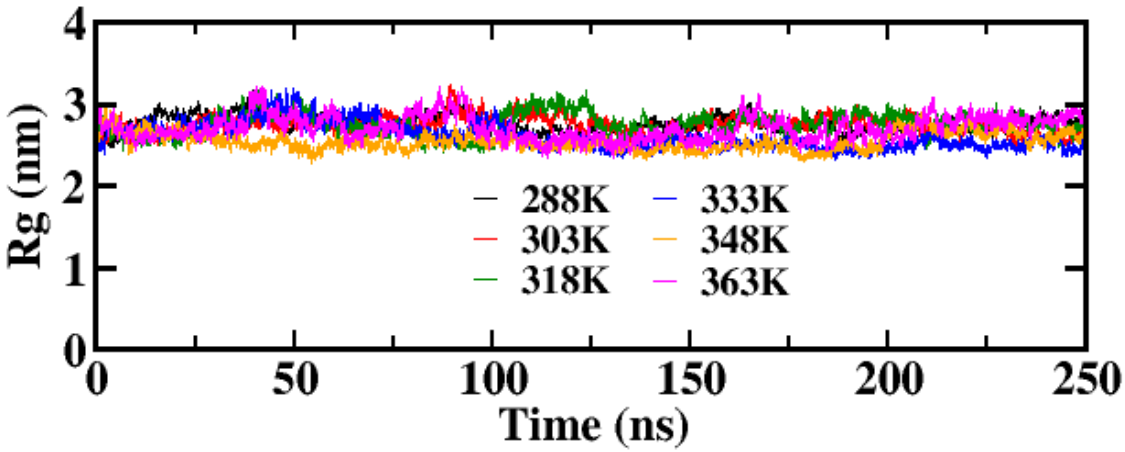
Appendix XVII: RMSD profile and Radius of gyration of heavy atoms for SARS-CoV-2 RNA Genome at six temperatures.

SPC/E water model, Set-I

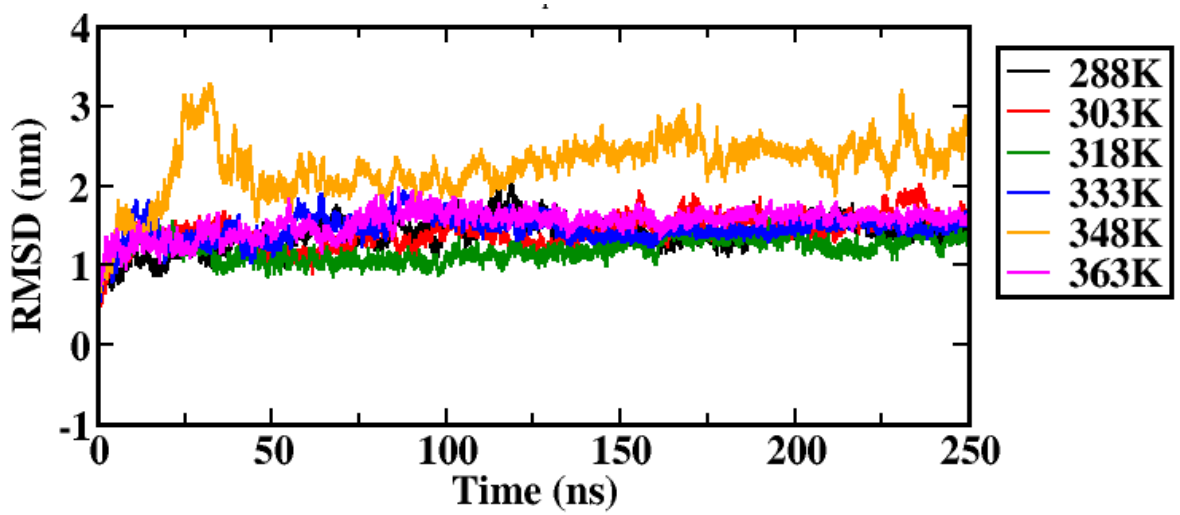


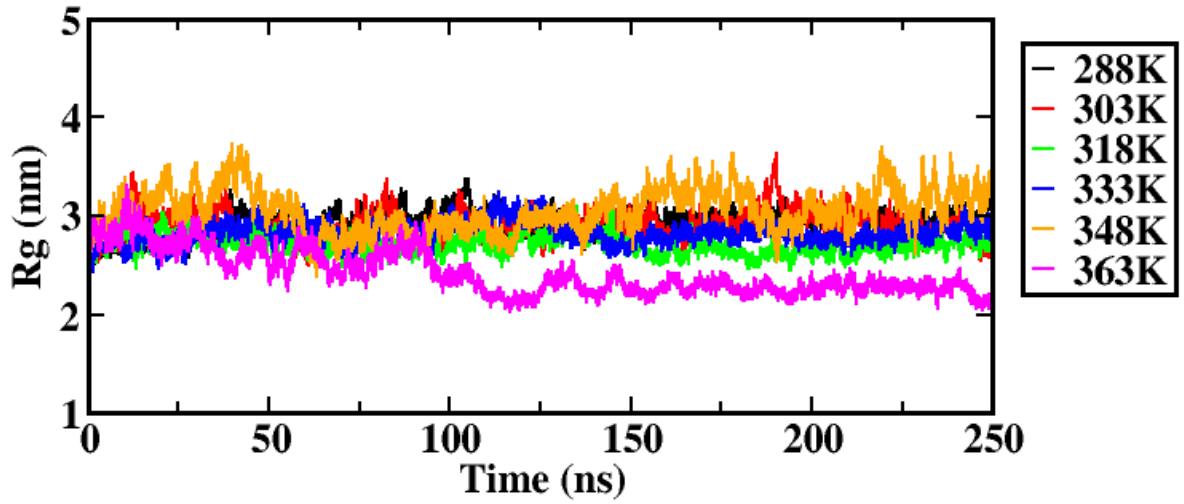
Set- II





mTIP3P water model





Appendix XVIII:. The number of conformations with temperatures in SPC/E water model.

Temperature	No. of conformations obtained with RMSD cut off 1.5 Å	No. of conformations retained for		
		>80 ps	>120 ps	>160 ps
288 K	1	1	1	1
303 K	182	149	137	130
318 K	1290	704	520	467
333 K	2768	619	344	180
348 K	7170	248	119	79
363 K	9093	90	52	38

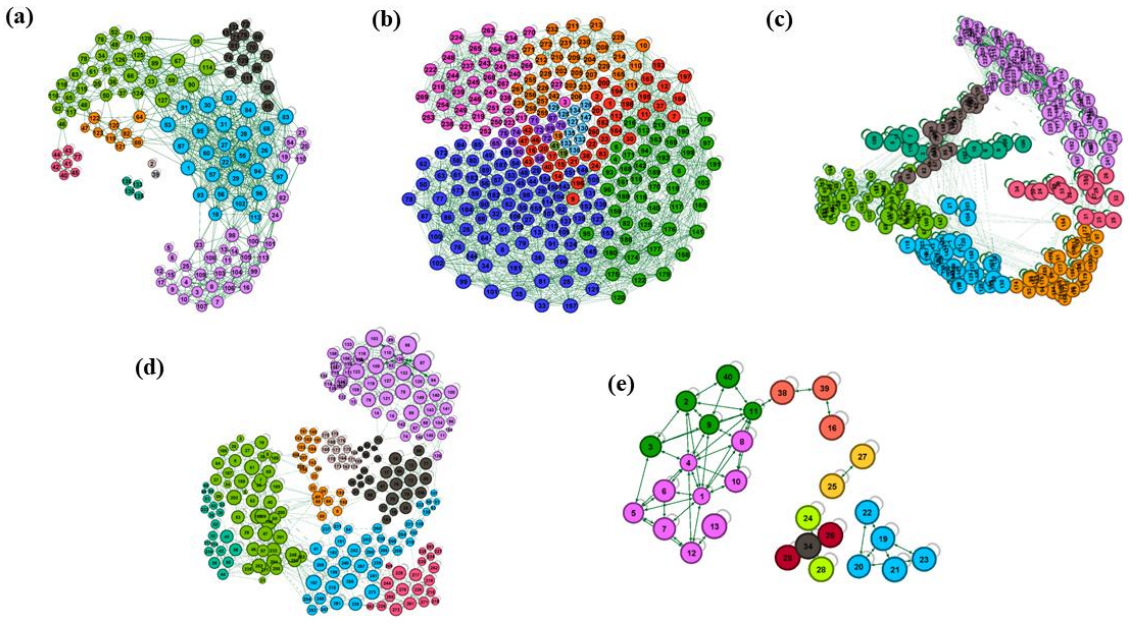
Appendix XIX: Community class percentage of the conformations at different angle cutoff at different temperatures.

Threshold	Temperature (K)	Total Number of communities	Number of communities with class percentage >10%	Top 5-module Class percentage
±5	**303**	50	1	10.22 8.76 8.76 8.76 6.57
	318	75	0	8.09 8.09 8.09 7.51 7.51
	333	154	0	8.14 6.1 5.23 4.65 3.78
	348	39	1	14.93 8.96 8.96 7.46 7.46
	363	30	2	17.31 13.46 9.62 3.85 3.85
±10	**303**	13	5	25.55 18.98 17.52 13.14 12.41
	318	9	5	23.31 19.85 19.85 19.27 11.95

	333	21	5	20.64 18.02 15.12 14.83 11.34
	348	7	4	26.67 25.19 20.00 11.85 6.67
	363	9	4	32.69 19.23 19.23 13.46 7.69
±20	**303**	4	4	32.85 29.93 21.17 16.06
	318	5	5	22.16 21 21 20.23 15.61
	333	5	4	26.45 25.29 22.97 21.51
	348	3	3	41.00 36.00 21.00
	363	4	4	34.62 28.85 19.23 17.31

±30	**303**	2	2	55.47 44.43
	318	3	3	40.27 30.44 29.29
	333	3	3	34.88 32.85 32.27
	348	4	3	38.81 34.33 23.88 1.49
	363	2	2	63.46 36.54

Appendix XX: Network distribution of conformations with ±10° as angle cutoff at various temperatures: (a) 303 K, (b) 318 K, (c) 333 K, (d) 348 K, and (e) 363 K for second set of simulation.



Appendix XXI: Network analysis of conformations with respect to temperature in mTIP3P water model. The base pair twist and width of minor-major grooves are given.

Temperature (K)	Conformation ID	Twist (˚)	Base pairs	P-P Distance	
				Minor	Major
288	1 (blue)	33.31 ± 24.83	20U21G/40C41A	18.1	15.0
			21G22U/39A40C	18.3	16.0
			22U23A/38U39A	17.7	18.3
			80G81G/31C32C	18.3	17.2
			48G49G/70U71C	17.4	25.6
			49G50C/69G70U	18.6	27.1
			50C51A/68U69G	21.0	22.3
			51A53A/65U68U	25.1	20.9
			53A54G/64C65U	29.8	18.4
303	75 (green)	-17.51 ± 83.13	44G43C/18G17C	24.3	14.4
			43C42C/19G18G	22.1	15.8
			42C20U/41A19G	22.3	16.6
			20U21G/40C41A	22.7	16.7
			79G80G/32C33C	18.7	9.5
			80G81G/31C32C	18.1	15.1
	194 (blue)	31.28 ± 7.28	79G80G/32C33C	19.5	9.3
			80G81G/31C32C	17.9	19.5
			81G82C/30G31C	17.5	15.7

	352 (cyan)	31.95 ± 2.36	79G80G/32C33C	18.3	11.4
			80G81G/31C32C	18.3	18.0
			81G82C/30G31C	18.4	13.6
	408 (red)	Not observed	Not observed	-	-
	455 (magenta)	30.45 ± 14.23	79G80G/32C33C	21.1	12.2
			80G81G/31C32C	17.4	17.8
			50C51A/67A69G	17.6	26.4
			51A52C/66G67A	18.7	23.1
318	3 (cyan)	56.77 ± 51.95	20U21G/40C41A	17.8	16.2
			21G22U/39A40C	18.7	16.2
			22U23A/38U39A	18.9	15.8
			23A24A/37U38U	18.2	15.4
			24A25G/36C37U	18.0	11.1
			25G34G/35U36C	17.5	6.8
			34G79G/33C35U	17.4	8.4
			80G79G/32C33C	18.6	16.9
			50C51A/68U69G	17.5	16.1
			51A52C/66G68U	19.2	18.2
			52C53A/65U66G	21.5	17.7
			53A54G/64C65U	20.3	15.7
	53 (blue)	20.42 ± 96.17	50C51A/68U69G	20.6	23.0
			51A66G/67A68U	17.8	20.9

			66G53A/65U67A	18.0	22.3
			54A53G/64C65U	18.9	20.3
			19G20U/41A42C	18.6	15
			20U21G/40C41A	18.6	17.6
			21G22U/39A40C	18.6	18.3
			22U23A/38U39A	18.7	18.2
			23A24A/37U38U	17.5	16.8
			24A25G/36C37U	16.3	11.1
			25G34G/35U36C	15.9	9.5
			35U79G/33C34G	17.6	7.4
			79G80G/32C33C	18.3	21.2
			80G81G/31C32C	19.0	19.5
	124 (red)	34.80 ± 4.20	19G20U/41A42C	17.1	17.9
			20U21G/40C41A	18.1	17.6
			21G22U/39A40C	19.4	17.5
			22U23A/38U39A	19.9	18.2
	330 (green)	43.31 ± 20.0	19G20U/41A42C	18.7	17.5
			79G80G/32C33C	17.8	12.4
			51A52C/66G68U	20.6	9.9
			52C53A/65U66G	21.8	10.4
			53A54G/64C65U	18.8	16.9
333	8 (cyan)	32.85	31C32C/80G81G	17.7	18.1

	25 (magenta)	Not observed	Not observed	-	-
	35 (red)	Not observed	Not observed	-	-
	80 (green)	33.10 ± 0.73	80G81G/31C32C	18.2	18.5
			81G82C/30G31C	18.0	17.1
	164 (blue)	59.71 ± 31.35	21G23A/38U40C	25.3	7.2
			23A25G/37U38U	23.1	5.2
			25G26U/36C37U	18.6	7.2
			26U27G/34G36C	16.8	9.4
			27G79G/33C34G	20.2	8.5
			79G80G/32C33C	22.0	11.5
348	28 (grey)	Not observed	Not observed	-	-
	43 (blue)	Not observed	Not observed	-	-
	61 (pink)	32.94	53A54G/64C65U	20.9	14.7
	94 (green)	Not observed	Not observed	-	-
	127 (magenta)	Not observed	Not observed	-	-
	152 (cyan)	45.98 ± 28.99	63A64C/54G55G	20.5	14
			64C65U/53A54G	19.9	15.8

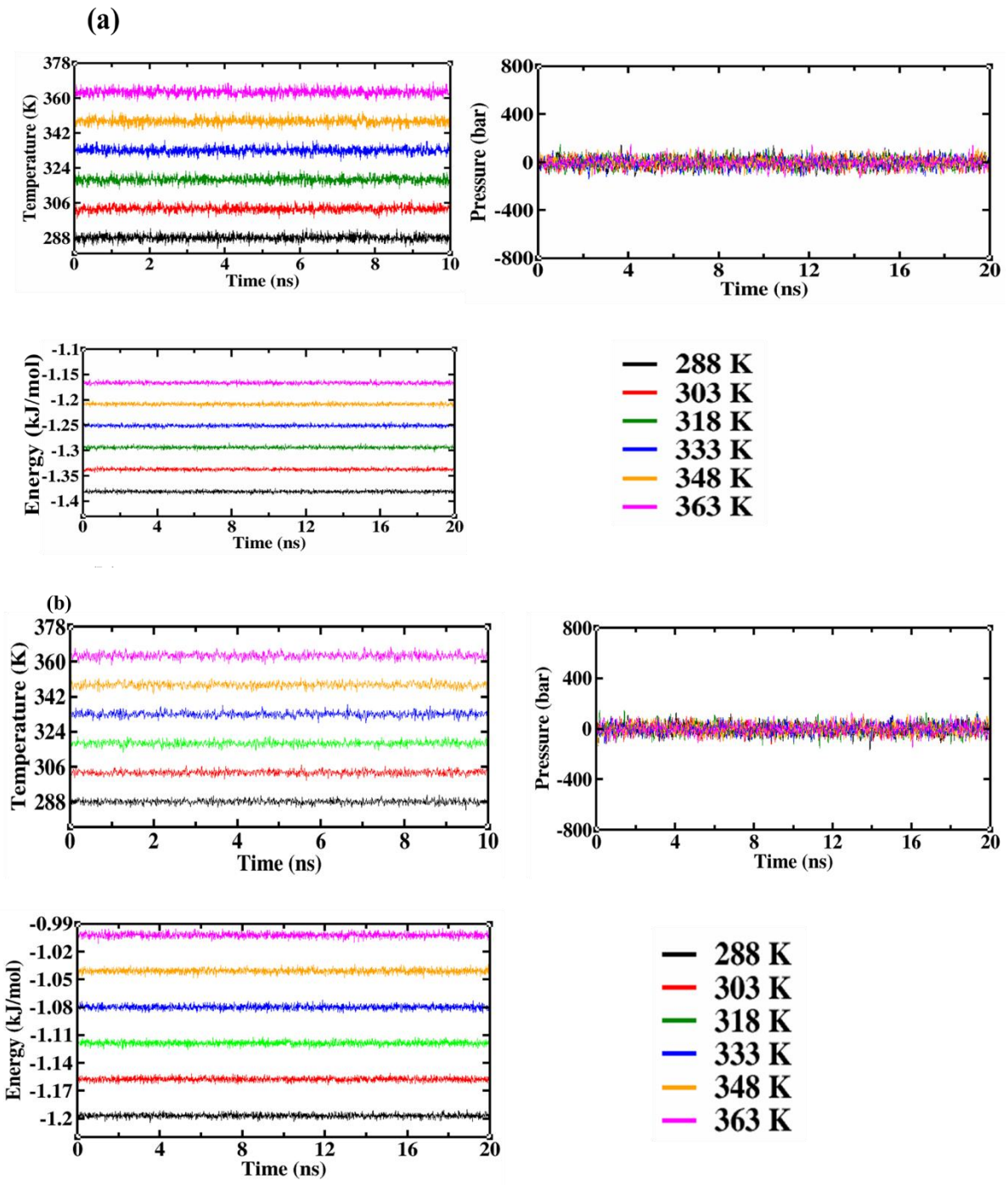
	189 (yellow)	Not observed	Not observed	-	-
	196 (red)	Not observed	Not observed	-	-
	225 (orange)	Not observed	Not observed	-	-
363	6 (magenta)	Not observed	Not observed	-	-
	43 (cyan)	Not observed	Not observed	-	-
	67 (red)	Not observed	Not observed	-	-
	170 (green)	Not observed	Not observed	-	-
	263 (blue)	Not observed	Not observed	-	-

Appendix XXII: Torsion angles of the most prominent structures for each community for base RG76 at six temperatures.

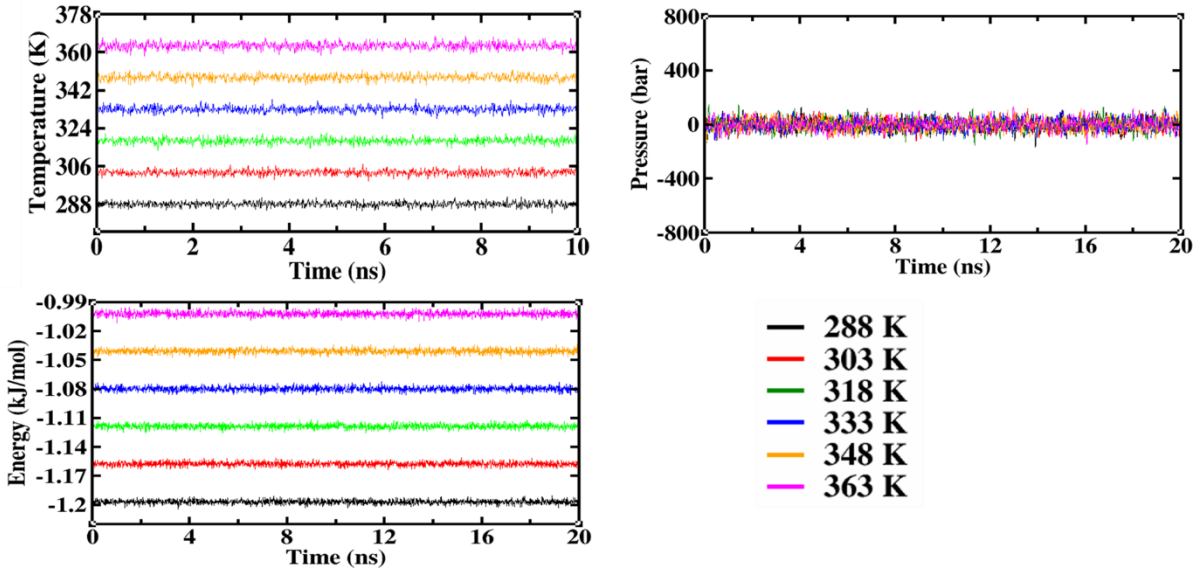
Temperature	Community	χ	V_0	V_1	V_2	V_3	V_4
288 K	-	-110.56	13.28	-31.42	-23.35	-19.18	-6.64
303 K	Blue	-128.51	-6.20	24.93	21.30	19.71	10.37
	Green	-103.44	21.01	-32.40	-19.71	-12.70	-0.19
	Magenta	-142.09	-10.42	-13.96	-19.90	-24.87	19.92
	Black	-121.89	-0.33	-28.63	-28.85	-29.41	-18.38
	Orange	-139.32	-16.07	33.91	24.45	20.69	6.54
	Red	-112.15	9.76	-29.49	-23.52	-20.50	-8.49
318 K	Green	-144.09	-20.04	38.16	26.65	21.75	5.56
	Magenta	-136.71	-16.45	39.13	28.74	22.76	7.61
	Cyan	-127.13	-5.63	-24.94	-28.51	-31.76	-22.02
	Black	-139.68	-12.36	30.60	22.83	18.41	6.51
	Orange	-145.22	-20.40	40.57	28.08	21.63	5.05
	Red	-145.68	-24.58	42.54	27.34	19.41	1.85
333 K	Magenta	-125.16	-1.23	-20.21	-20.96	-22.01	-14.36
	Green	-126.21	-3.14	-24.49	-26.39	-29.22	-19.52
	Cyan	-115.49	8.43	-27.45	-22.23	-20.24	-9.05
	Black	-122.56	-0.13	-22.29	-22.37	-23.71	-15.02
	Orange	-118.31	6.98	-27.24	-23.04	-21.17	-10.89
	Red	-122.88	-2.67	-26.52	-28.17	-30.35	-19.52
	Dark Green	-119.29	6.75	-32.93	-28.57	-28.30	-14.45

348 K	Magenta	171.2	0.75	-27.81	-27.33	-27.98	-16.76
	Green	-154.2	-6.09	-22.26	-28.00	-30.64	-21.15
	Orange	-161.11	0.78	-23.32	-22.83	-24.17	-15.03
	Cyan	-159.06	6.78	-31.11	-26.65	-25.32	-13.93
	Blue	-171.95	0.68	-22.72	-22.28	-22.82	-14.24
	Pink	-158.84	6.67	-32.08	-27.92	-26.06	-13.50
	Black	-169.5	6.28	-28.77	-22.27	-23.19	-12.15
363 K	Magenta	-116.33	2.41	-27.65	-26.05	25.53	-14.97
	Cyan	-116.04	11.55	-34.37	-27.08	-24.05	-10.20
	Green	-108.94	13.06	-35.23	-26.98	-24.24	-10.19
	Black	-120.19	5.83	-27.04	-23.65	-23.09	-12.62
	Orange	-152.83	-24.73	37.85	22.74	13.29	-2.26

Appendix XXIII:. Time evolution of the equilibrium properties of the second set of MD simulation in (a) CHARMM-SPC/E (b) CHARMM-mTIP3P and (c) AMBER-mTIP3P water models.

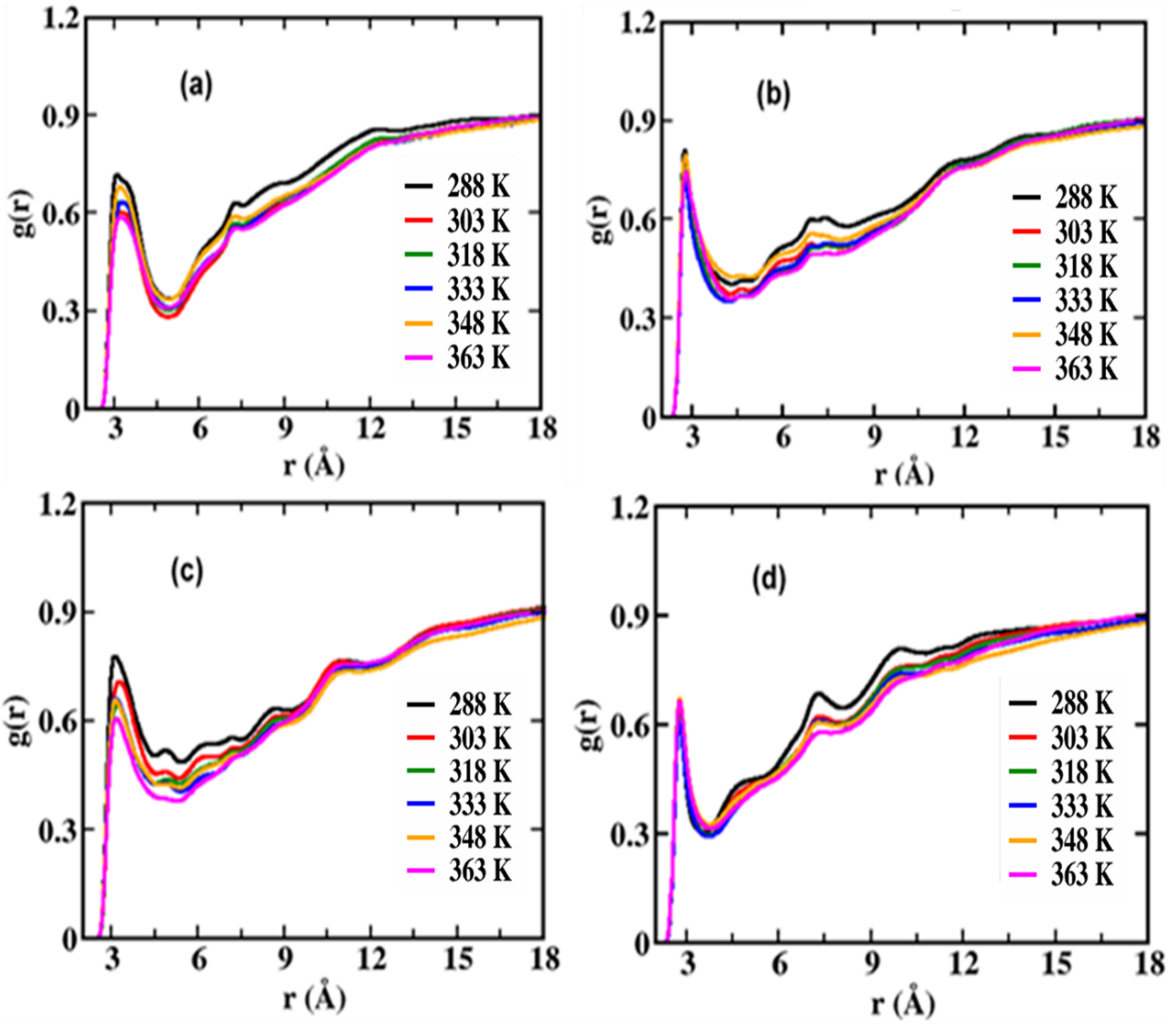


(c)

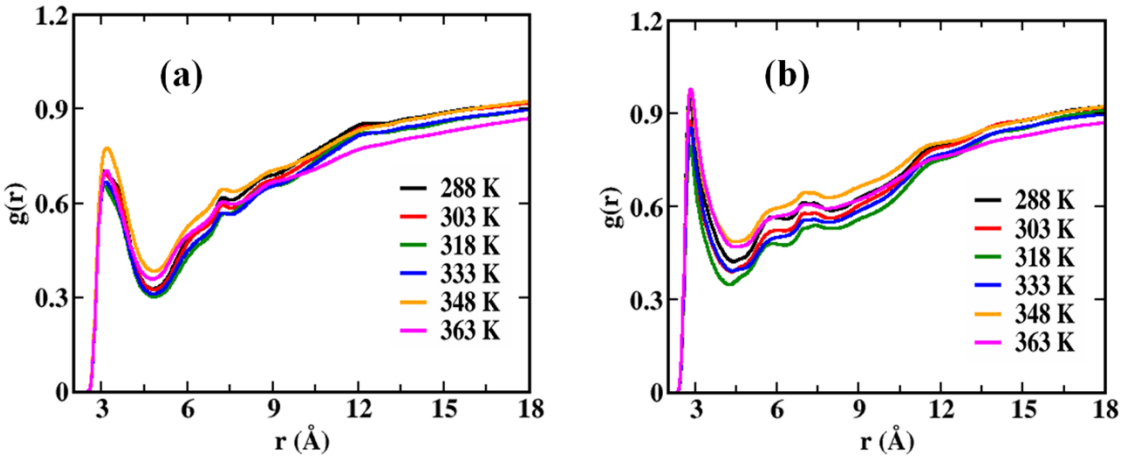

Temperature (K)
378
360
342
324
306
288
0
2
4
6
8
10
Time (ns)
Pressure (bar)
800
400
0
-400
-800
0
4
8
12
16
20
Time (ns)
Energy (kJ/mol)
-0.99
-1.02
-1.05
-1.08
-1.11
-1.14
-1.17
-1.2
0
4
8
12
16
20
Time (ns)
288 K
303 K
318 K
333 K
348 K
363 K

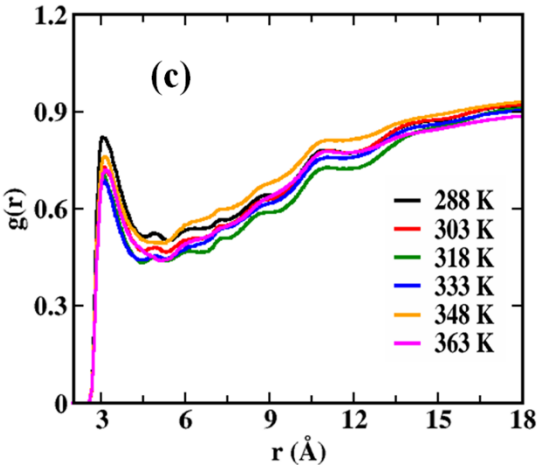
Appendix XXIV: Radial distribution functions of (a) RG_N_2-O_w (b) RG_O_6-O_w (c) RC_N_4-O_w, and (d) RC_O_2-O_w at various temperatures.

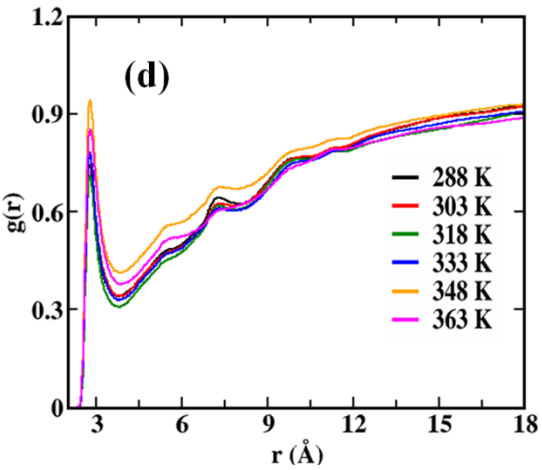
CHARMM-SPC/E water model



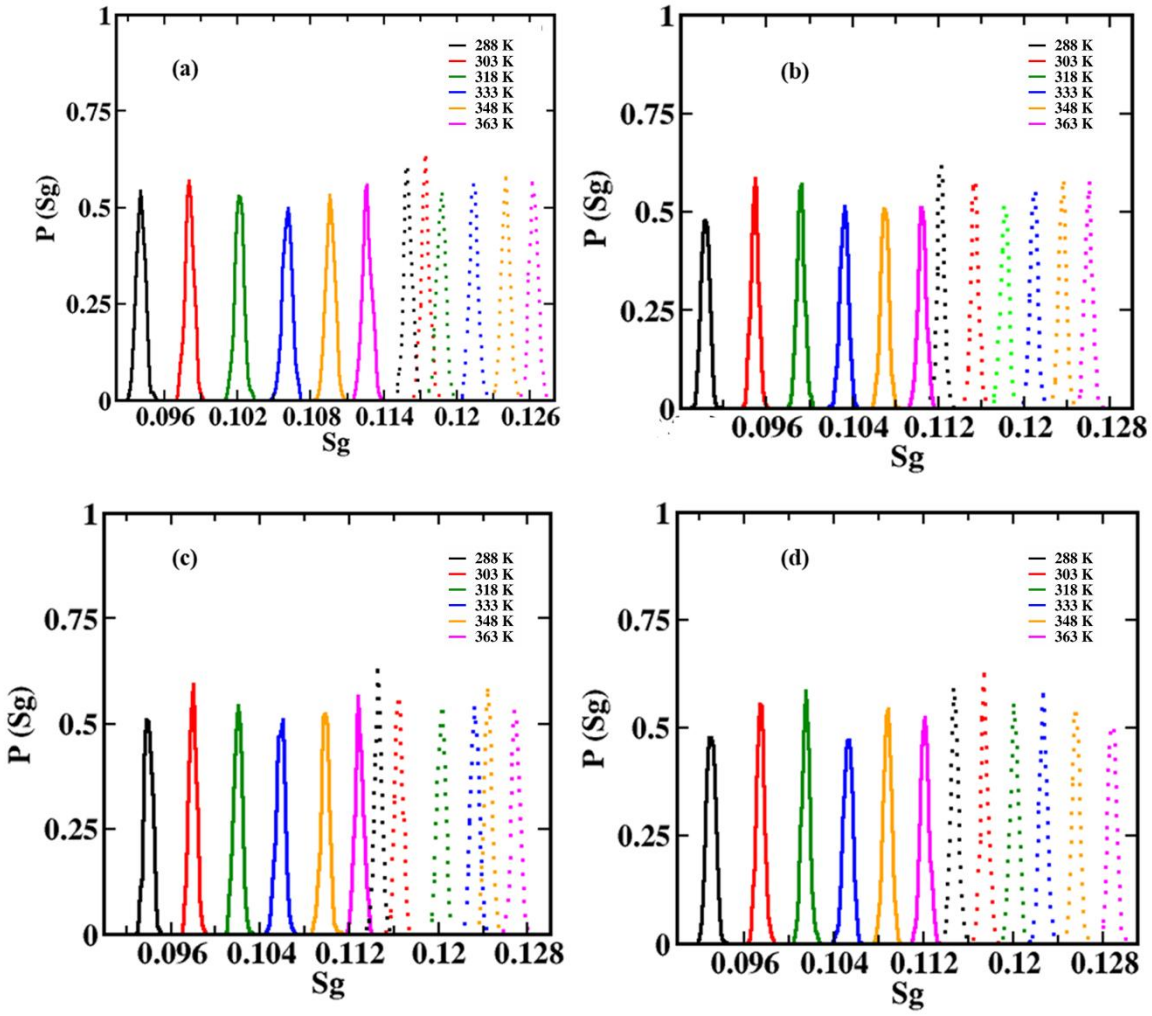
CHARMM-mTIP3P water model




(c)
g(r)
r (Å)
288 K
303 K
318 K
333 K
348 K
363 K


(d)
g(r)
r (Å)
288 K
303 K
318 K
333 K
348 K
363 K

Appendix XXV: Orientational tetrahedral order Parameter of water molecules present at the bulk (> 8 Å) nucleic acids of the SARS-CoV-2 RNA genome RG, RC, RA, and RU (a), (b), (c), and (d) respectively in the bulk region at different temperatures for both CHARMM-SPC/E (solid line) and CHARMM-TIP3P (dash line) water models.



REFERENCES

Abelein, A., Jarvet, J., Barth, A., Gräslund, A., and Danielsson, J. (2016). "Ionic Strength Modulation of the Free Energy Landscape of Aβ40 Peptide Fibril Formation." *J. Am. Chem. Soc.*, 138(21), 6893–6902.

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

Abramyan, T. M., Snyder, J. A., Thyparambil, A. A., Stuart, S. J., and Latour, R. A. (2016). "Cluster analysis of molecular simulation trajectories for systems where both conformation and orientation of the sampled states are important." *J. Comput. Chem.*, 37(21), 1973–1982.

Ageitos, J. M., Sánchez-Pérez, A., Calo-Mata, P., and Villa, T. G. (2017). "Antimicrobial peptides (AMPs): Ancient compounds that represent novel weapons in the fight against bacteria." *Biochem. Pharm.*, 133, 117–138.

Alfaro-Núñez, A., Crone, S., Mortensen, S., Rosenstierne, M. W., Fomsgaard, A., Marving, E., Nielsen, S. H., Jørgensen, M. G. P., Polacek, C., Cohen, A. S., and Nielsen, C. (2022). "SARS-CoV-2 RNA stability in dry swabs for longer storage and transport at different temperatures." *Transbound. Emerg. Dis*, 69(2), 189–194.

Allen, M. P., and Tildesley, D. J. (2017). *Computer Simulation of Liquids*. Oxford University Press.

Alves, E. D., Andrade, D. X. de, Almeida, A. R. de, and Colherinhas, G. (2021). "Atomistic molecular dynamics study on the influence of high temperatures on the structure of peptide nanomembranes candidates for organic supercapacitor electrode." *J. Mol. Liq.s*, 334, 116126.

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

Anandakrishnan, R., Izadi, S., and Onufriev, A. V. (2019). “Why Computed Protein Folding Landscapes Are Sensitive to the Water Model.” *J. Chem. Theory Comput.*, 15(1), 625–636.

Anderson, C. F., and Record, M. T. (1995). “Salt-Nucleic Acid Interactions.” *Annu. Rev. Phys. Chem.*, 46(1), 657–700.

Appadurai, R., Nagesh, J., and Srivastava, A. (2021). “High resolution ensemble description of metamorphic and intrinsically disordered proteins using an efficient hybrid parallel tempering scheme.” *Nat Commun*, 12(1), 958.

Araya-Secchi, R., Perez-Acle, T., Kang, S., Huynh, T., Bernardin, A., Escalona, Y., Garate, J.-A., Martínez, A. D., García, I. E., Sáez, J. C., and Zhou, R. (2014). “Characterization of a Novel Water Pocket Inside the Human Cx26 Hemichannel Structure.” *Biophys. J.*, 107(3), 599–612.

Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E., and Baranov, P. V. (2016). “Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use.” *Nucleic Acids Res*, gkw530.

Avbelj, F., and Moult, J. (1995). “Role of Electrostatic Screening in Determining Protein Main Chain Conformational Preferences.” *Biochemistry*, 34(3), 755–764.

Aziz, E. F., Ottosson, N., Eisebitt, S., Eberhardt, W., Jagoda-Cwiklik, B., Vácha, R., Jungwirth, P., and Winter, B. (2008). “Cation-Specific Interactions with Carboxylate in Amino Acid and Acetate Aqueous Solutions: X-ray Absorption and *ab initio* Calculations.” *J. Phys. Chem. B*, 112(40), 12567–12570.

Bahar, I., Lezon, T. R., Bakan, A., and Shrivastava, I. H. (2010). “Normal Mode Analysis of Biomolecular Structures: Functional Mechanisms of Membrane Proteins.” *Chem. Rev.*, 110(3), 1463–1497.

Bailor, M. H., Sun, X., and Al-Hashimi, H. M. (2010). “Topology Links RNA Secondary Structure with Global Conformation, Dynamics, and Adaptation.” *Science*, 327(5962), 202–206.

Baker, C. M., Anisimov, V. M., and MacKerell, A. D. (2011). “Development of CHARMM Polarizable Force Field for Nucleic Acid Bases Based on the Classical Drude Oscillator Model.” *J. Phys. Chem. B,* 115(3), 580–596.

Baldwin, R. L. (1996). “How Hofmeister ion interactions affect protein stability.” *Biophys J,* 71(4), 2056–2063.

Bandyopadhyay, S., Chakraborty, S., and Bagchi, B. (2005). “Secondary Structure Sensitivity of Hydrogen Bond Lifetime Dynamics in the Protein Hydration Layer.” *J. Am. Chem. Soc.,* 127(47), 16660–16667.

Bankura, A., Carnevale, V., and Klein, M. L. (2013). “Hydration structure of salt solutions from *ab initio* molecular dynamics.” *J. Chem. Phys.,* 138(1), 014501.

Bastian, M., Heymann, S., and Jacomy, M. (2009). “Gephi: An Open Source Software for Exploring and Manipulating Networks.”

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

Bell, I. H., Messerly, R., Thol, M., Costigliola, L., and Dyre, J. C. (2019). “Modified Entropy Scaling of the Transport Properties of the Lennard-Jones Fluid.” *J. Phys. Chem. B,* 123(29), 6345–6363.

Bendheim, P. E., Brown, H. R., Rudelli, R. D., Scala, L. J., Goller, N. L., Wen, G. Y., Kascsak, R. J., Cashman, N. R., and Bolton, D. C. (1992). “Nearly ubiquitous tissue distribution of the scrapie agent precursor protein.” *Neurology,* 42(1), 149–156.

Ben-Naim, A. (1987). “On the role of water in molecular recognition and self-assembly.” *Proc. Indian Acad. Sci. (Chem. Sci.),* 98(5–6), 357–377.

Berendsen, H. J. C., Grigera, J. R., and Straatsma, T. P. (1987). “The missing term in effective pair potentials.” *J. Phys. Chem.,* 91(24), 6269–6271.

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

Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. van, and Hermans, J. (1981). “Interaction Models for Water in Relation to Protein Hydration.” *Intermolecular Forces*, B. Pullman, ed., Dordrecht: Springer Netherlands, 331–342.

Berendsen, H. J. C., Spoel, D. van der, and Drunen, R. van. (1995). “GROMACS: A message-passing parallel molecular dynamics implementation.” *Comput. Phys. Commun.*, 91(1–3), 43–56.

Berntsson, E., Paul, S., Vosough, F., Sholts, S. B., Jarvet, J., Roos, P. M., Barth, A., Gräslund, A., and Wärmländer, S. (2021). “Lithium ions display weak interaction with amyloid-beta (Aβ) peptides and have minor effects on their aggregation.” *Acta Biochi.m Pol.*, 68(2), 169–179.

Bertho, G., Bouvier, G., Hoa, G. H. B., and Girault, J.-P. (2008). “The key-role of tyrosine 155 in the mechanism of prion transconformation as highlighted by a study of sheep mutant peptides.” *Peptides*, 29(7), 1073–1084.

Best, R. B., Zhu, X., Shim, J., Lopes, P. E. M., Mittal, J., Feig, M., and MacKerell, A. D. (2012). “Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ_1 and χ_2 Dihedral Angles.” *J. Chem. Theory Comput.*, 8(9), 3257–3273.

Bhowmick, A., Brookes, D. H., Yost, S. R., Dyson, H. J., Forman-Kay, J. D., Gunter, D., Head-Gordon, M., Hura, G. L., Pande, V. S., Wemmer, D. E., Wright, P. E., and Head-Gordon, T. (2016). “Finding Our Way in the Dark Proteome.” *J. Am. Chem. Soc.*, 138(31), 9730–9742.

Biswas, S., Chakraborty, D., and Mallik, B. S. (2018). “Interstitial Voids and Resultant Density of Liquid Water: A First-Principles Molecular Dynamics Study.” *ACS Omega*, 3(2), 2010–2017.

Biswas, S., and Mallik, B. S. (2014). “Effects of Temperature on the Structure and Dynamics of Aqueous Mixtures of *N* , *N* -Dimethylformamide.” *J. Chem. Eng. Data,* 59(10), 3250–3257.

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

Bruyn, E. de, Dorn, A. E., Zimmermann, O., and Rossetti, G. (2023). “SPEADI: Accelerated Analysis of IDP-Ion Interactions from MD-Trajectories.” *Biology,* 12(4).

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

Butcher, S. E., and Pyle, A. M. (2011). “The Molecular Interactions That Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks.” *Acc. Chem. Res.*, 44(12), 1302–1311.

Cacace, M. G., Landau, E. M., and Ramsden, J. J. (1997). “The Hofmeister series: salt and solvent effects on interfacial phenomena.” *Quart. Rev. Biophys.*, 30(3), 241–277.

Campo a*, M. G., and Raul Grigera, J. (2004). “Molecular Dynamics Simulation of OH^- in Water.” *Mol. Simul.*, 30(8), 537–542.

Campo, M. G. (2006). “Molecular dynamics simulation of glycine zwitterion in aqueous solution.” *J. Chem.Phys.,* 125(11), 114511.

Cassone, G., Chillè, D., Mollica Nardo, V., Giuffrè, O., Ponterio, R. C., Sponer, J., Trusso, S., Saija, F., and Foti, C. (2020). “Arsenic–nucleotides interactions: an experimental and computational investigation.” *Dalton Trans.*, 49(19), 6302–6311.

Cassone, G., Kruse, H., and Sponer, J. (2019). “Interactions between cyclic nucleotides and common cations: an *ab initio* molecular dynamics study.” *Phys. Chem. Chem. Phys.*, 21(15), 8121–8132.

Castro, T., Munteanu, F.-D., and Cavaco-Paulo, A. (2019). “Electrostatics of Tau Protein by Molecular Dynamics.” *Biomolecules*, 9(3), 116.

Cavalli, A., Salvatella, X., Dobson, C. M., and Vendruscolo, M. (2007). “Protein structure determination from NMR chemical shifts.” *Proc. Natl. Acad. Sci. U.S.A.*, 104(23), 9615–9620.

Chakraborty, D., and Chandra, A. (2011a). “An analysis of voids and necks in supercritical water.” *J. Mol. Liq.*, 163(1), 1–6.

Chakraborty, D., and Chandra, A. (2011b). “Hydrogen bonded structure and dynamics of liquid-vapor interface of water-ammonia mixture: An *ab initio* molecular dynamics study.” *J. Chem.Phys.*, 135(11), 114510.

Chakraborty, D., and Chandra, A. (2012). “Voids and necks in liquid ammonia and their roles in diffusion of ions of varying size.” *J. Comput. Chem.*, 33(8), 843–852.

Chamachi, N. G., and Chakrabarty, S. (2016). “Replica Exchange Molecular Dynamics Study of Dimerization in Prion Protein: Multiple Modes of Interaction and Stabilization.” *J. Phys. Chem. B*, 120(30), 7332–7345.

Chamachi, N. G., and Chakrabarty, S. (2017). “Temperature-Induced Misfolding in Prion Protein: Evidence of Multiple Partially Disordered States Stabilized by Non-Native Hydrogen Bonds.” *Biochemistry*, 56(6), 833–844.

Chaudhary, U., and Kaphle, G. C. (2023). “Molecular Dynamics Study of Peptide in Water at Different Temperature.” *J. Nep. Phys. Soc.*, 9(1), 83–91.

Chen, J., Liang, H., and Fernández, A. (2008). “Protein structure protection commits gene expression patterns.” *Genome Biol*, 9(7), R107.

Chen, Y., Okur, H. I., Liang, C., and Roke, S. (2017). “Orientational ordering of water in extended hydration shells of cations is ion-specific and is correlated directly with viscosity and hydration free energy.” *Phys. Chem. Chem. Phys.*, 19(36), 24678–24688.

Chopra, R., Truskett, T. M., and Errington, J. R. (2010). “On the Use of Excess Entropy Scaling to Describe the Dynamic Properties of Water.” *J. Phys. Chem. B*, 114(32), 10558–10566.

Collins, K. (2004). “Ions from the Hofmeister series and osmolytes: effects on proteins in solution and in the crystallization process.” *Methods*, 34(3), 300–311.

Collins, K. D. (1997). “Charge density-dependent strength of hydration and biological structure.” *Biophys. J.*, 72(1), 65–76.

Collins, K. D., and Washabaugh, M. W. (1985). “The Hofmeister effect and the behaviour of water at interfaces.” *Quart. Rev. Biophys.*, 18(4), 323–422.

Crehuet, R., Buigues, P. J., Salvatella, X., and Lindorff-Larsen, K. (2019). “Bayesian-Maximum-Entropy Reweighting of IDP Ensembles Based on NMR Chemical Shifts.” *Entropy*, 21(9), 898.

Cruz-León, S., and Schwierz, N. (2020). “Hofmeister Series for Metal-Cation–RNA Interactions: The Interplay of Binding Affinity and Exchange Kinetics.” *Langmuir*, 36(21), 5979–5989.

Cuervo, A., Dans, P. D., Carrascosa, J. L., Orozco, M., Gomila, G., and Fumagalli, L. (2014). “Direct measurement of the dielectric polarization properties of DNA.” *Proc. Natl. Acad. Sci.*, 111(35), E3624–E3630.

Curtis, R. A., Prausnitz, J. M., and Blanch, H. W. (1998). “Protein-protein and protein-salt interactions in aqueous protein solutions containing concentrated electrolytes.” *Biotechnol. Bioeng.*, 57(1), 11–21.

Dahanayake, J. N., and Mitchell-Koch, K. R. (2018). “Entropy connects water structure and dynamics in protein hydration layer.” *Phys. Chem. Chem. Phys.*, 20(21), 14765–14777.

Darden, T., York, D., and Pedersen, L. (1993). “Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems.” *J. Chem.Phys.*, 98(12), 10089–10092.

Darling, A. L., and Uversky, V. N. (2018). “Intrinsic Disorder and Posttranslational Modifications: The Darker Side of the Biological Dark Matter.” *Front. Genet.*, 9, 158.

Darnell, M. E. R., Subbarao, K., Feinstone, S. M., and Taylor, D. R. (2004). “Inactivation of the coronavirus that induces severe acute respiratory syndrome, SARS-CoV.” *J.Virol. Methods.*, 121(1), 85–91.

Das, B. K., and Chakraborty, D. (2020). “Epitope-Based Potential Vaccine Candidate for Humoral and Cell-Mediated Immunity to Combat Severe Acute Respiratory Syndrome Coronavirus 2 Pandemic.” *J. Phys. Chem. Lett.*, 11(22), 9920–9930.

Daura, X., Gademann, K., Jaun, B., Seebach, D., Gunsteren, W. F. van, and Mark, A. E. (1999). “Peptide Folding: When Simulation Meets Experiment.” *Angew. Chem. Int. Ed.*, 38(1–2), 236–240.

Davies, P., Moualla, D., and Brown, D. R. (2011). “Alpha-Synuclein Is a Cellular Ferrireductase.” *PLoS one*, 6(1), e15814.

Debnath, A., Mukherjee, B., Ayappa, K. G., Maiti, P. K., and Lin, S.-T. (2010). “Entropy and dynamics of water in hydration layers of a bilayer.” *J. Chem. Phys.*, 133(17), 174704.

Denisov, V. P., and Halle, B. (1996). “Protein hydration dynamics in aqueous solution.” *Faraday Disc.*, 103, 227.

Denning, E. J., and MacKerell, A. D. (2012). “Intrinsic Contribution of the 2′-Hydroxyl to RNA Conformational Heterogeneity.” *J. Am. Chem. Soc.*, 134(5), 2800–2806.

Dey, A., Yan, S., Schlick, T., and Laederach, A. (2024). “Abolished frameshifting for predicted structure-stabilizing SARS-CoV-2 mutants: Implications to alternative conformations and their statistical structural analyses.”

Dickson, A., Mustoe, A. M., Salmon, L., and Brooks, C. L. (2014). “Efficient in silico exploration of RNA interhelical conformations using Euler angles and WExplore.” *Nucleic Acids Res.*, 42(19), 12126–12137.

Dignon, G. L., Zheng, W., Kim, Y. C., Best, R. B., and Mittal, J. (2018). “Sequence determinants of protein phase behavior from a coarse-grained model.” *PLoS Comput. Biol.*, 14(1), e1005941.

Dilip, H. N., and Chakraborty, D. (2019). “Hydrophilicity of the hydrophobic group: Effect of cosolvents and ions.” *J. Mol. Liq.*, 280, 389–398.

Dilip.H.N., and Chakraborty, D. (2020). “Effect of cosolvents in the preferential binding affinity of water in aqueous solutions of amino acids and amides.” *J. Mol. Liq.*, 300, 112375.

Dill, K. A., Truskett, T. M., Vlachy, V., and Hribar-Lee, B. (2005). “Modeling Water, the Hydrophobic Effect, and Ion Solvation.” *Annu. Rev. Biophys. Biomol. Struct.*, 34(1), 173–199.

Dohnalová, H., Seifert, M., Matoušková, E., Klein, M., Papini, F. S., Lipfert, J., Dulin, D., and Lankaš, F. (2024). “Temperature-Dependent Twist of Double-Stranded RNA Probed by Magnetic Tweezer Experiments and Molecular Dynamics Simulations.” *J. Phys. Chem. B,* 128(3), 664–675.

Dong, Q., Stellwagen, E., and Stellwagen, N. C. (2009). “Monovalent Cation Binding in the Minor Groove of DNA A-Tracts.” *Biochemistry,* 48(5), 1047–1055.

Duan, Y., Wu, C., Chowdhury, S., Lee, M. C., Xiong, G., Zhang, W., Yang, R., Cieplak, P., Luo, R., Lee, T., Caldwell, J., Wang, J., and Kollman, P. (2003). “A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations.” *J. Comput. Chem.*, 24(16), 1999–2012.

Duka, A., and Ahearn, G. A. (2013). “L-leucine, L-methionine, and L-phenylalanine share a Na(+)/K (+)-dependent amino acid transporter in shrimp hepatopancreas.” *J. Comp. Physiol. B,* 183(6), 763–771.

Dyson, H. J., and Wright, P. E. (2005). “Intrinsically unstructured proteins and their functions.” *Nat. Rev. Mol. Cell. Biol.*, 6(3), 197–208.

Dzugutov, M. (1996). “A universal scaling law for atomic diffusion in condensed matter.” *Nature*, 381(6578), 137–139.

Easteal, A. J., Price, W. E., and Woolf, L. A. (1989). “Diaphragm cell for high-temperature diffusion measurements. Tracer Diffusion coefficients for water to 363 K.” *J. Chem. Soc., Faraday Trans. 1*, 85(5), 1091.

El Hassab, M. A., Shoun, A. A., Al-Rashood, S. T., Al-Warhi, T., and Eldehna, W. M. (2020). “Identification of a New Potential SARS-COV-2 RNA-Dependent RNA Polymerase Inhibitor via Combining Fragment-Based Drug Design, Docking, Molecular Dynamics, and MM-PBSA Calculations.” *Front. Chem.*, 8, 584894.

El Hassan, M. A., and Calladine, C. R. (1998). “Two distinct modes of protein-induced bending in DNA 1 1Edited by J. Karn.” *J. Mol. Biol.*, 282(2), 331–343.

Elfiky, A. A., Mahran, H. A., Ibrahim, I. M., Ibrahim, M. N., and Elshemey, W. M. (2022). “Molecular dynamics simulations and MM-GBSA reveal novel guanosine derivatives against SARS-CoV-2 RNA dependent RNA polymerase.” *RSC Adv.*, 12(5), 2741–2750.

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

Errington, J. R., Truskett, T. M., and Mittal, J. (2006). “Excess-entropy-based anomalies for a waterlike fluid.” *J. Chem.Phys.*, 125(24), 244502.

Espinoza-Fonseca, L. M. (2009). “Leucine-rich hydrophobic clusters promote folding of the N-terminus of the intrinsically disordered transactivation domain of p53.” *FEBS Letters*, 583(3), 556–560.

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

Ferner, J., Villa, A., Duchardt, E., Widjajakusuma, E., Wöhnert, J., Stock, G., and Schwalbe, H. (2008). “NMR and MD studies of the temperature-dependent dynamics of RNA YNMG-tetraloops.” *Nucleic Acids Res.*, 36(6), 1928–1940.

Fingerhut, B. P. (2021). “The mutual interactions of RNA, counterions and water–quantifying the electrostatics at the phosphate–water interface.” *Chem. Commun.*, 57(96), 12880–12897.

Fingerhut, B. P., Schauss, J., Kundu, A., and Elsaesser, T. (2021). “Contact pairs of RNA with magnesium ions-electrostatics beyond the Poisson-Boltzmann equation.” *Biophy. J.*, 120(23), 5322–5332.

Fink, A. L. (1998). “Protein aggregation: folding aggregates, inclusion bodies and amyloid.” *Fold Des*, 3(1), R9-23.

Fischer, N. M., Polêto, M. D., Steuer, J., and van der Spoel, D. (2018). “Influence of Na+ and Mg2+ ions on RNA structures studied with molecular dynamics simulations.” *Nucleic Acids Res.*, 46(10), 4872–4882.

Fisher, C. K., Huang, A., and Stultz, C. M. (2010). “Modeling Intrinsically Disordered Proteins with Bayesian Statistics.” *J. Am. Chem. Soc.*, 132(42), 14919–14927.

Fogarty, A. C., Duboué-Dijon, E., Sterpone, F., Hynes, J. T., and Laage, D. (2013). “Biomolecular hydration dynamics: a jump model perspective.” *Chem. Soc. Rev.*, 42(13), 5672.

Foloppe, N., and MacKerell, Jr., A. D. (2000). “All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data.” *J. Comput. Chem.*, 21(2), 86–104.

Frank, H. S., and Evans, M. W. (1945). “Free Volume and Entropy in Condensed Systems III. Entropy in Binary Liquid Mixtures; Partial Molal Entropy in Dilute Solutions; Structure and Thermodynamics in Aqueous Electrolytes.” *J. Chem.Phys.*, 13(11), 507–532.

Friedman, R., Nachliel, E., and Gutman, M. (2005). “Molecular Dynamics of a Protein Surface: Ion-Residues Interactions.” *Biophys. J.*, 89(2), 768–781.

Frugier, M., and Schimmel, P. (1997). “Subtle atomic group discrimination in the RNA minor groove.” *Proc. Natl. Acad. Sci.*, 94(21), 11291–11294.

Fuglebakk, E., Reuter, N., and Hinsen, K. (2013). “Evaluation of Protein Elastic Network Models Based on an Analysis of Collective Motions.” *J. Chem. Theory Comput.*, 9(12), 5618–5628.

Galamba, N. (2013). “Water’s Structure around Hydrophobic Solutes and the Iceberg Model.” *J. Phys. Chem. B*, 117(7), 2153–2159.

Garcia, A. E., and Hummer, G. (2000). “Water penetration and escape in proteins.” *Proteins*, 38(3), 261–272.

Gee, M. B., Cox, N. R., Jiao, Y., Bentenitis, N., Weerasinghe, S., and Smith, P. E. (2011). “A Kirkwood-Buff Derived Force Field for Aqueous Alkali Halides.” *J. Chem. Theory Comput.*, 7(5), 1369–1380.

Gil Pineda, L. I., Milko, L. N., and He, Y. (2020). “Performance of CHARMM36m with modified water model in simulating intrinsically disordered proteins: a case study.” *Biophys Rep*, 6(2–3), 80–87.

Glättli, A., Daura, X., and Van Gunsteren, W. F. (2003). “A novel approach for designing simple point charge models for liquid water with three interaction sites: Novel Simple Point Charge Models for Liquid Water.” *J. Comput. Chem.*, 24(9), 1087–1096.

Gluick, T. C., and Draper, D. E. (1994). “Thermodynamics of Folding a Pseudoknotted mRNA Fragment.” *J. Mol. Biol.*, 241(2), 246–262.

Goto, Y., Adachi, M., Muta, H., and So, M. (2017). “Salt-induced formations of partially folded intermediates and amyloid fibrils suggests a common underlying mechanism.” *Biophys. Rev.*, 10(2), 493–502.

Gunsteren, W. F. van, and Berendsen, H. J. C. (1987). “Groningen Molecular Simulation (GROMOS) Library Manual.” Biomos, Groningen, The Netherlands.

Gunsteren, W. F. van, Hünenberger, P. H., Mark, A. E., Smith, P. E., and Tironi, I. G. (1995). “Computer simulation of protein motion.” *Comput. Phys. Commun.*, 91(1–3), 305–319.

Haider, S., Parkinson, G. N., and Neidle, S. (2008). “Molecular Dynamics and Principal Components Analysis of Human Telomeric Quadruplex Multimers.” *Biophys. J.*, 95(1), 296–311.

Hajdin, C. E., Ding, F., Dokholyan, N. V., and Weeks, K. M. (2010). “On the significance of an RNA tertiary structure prediction.” *RNA*, 16(7), 1340–1349.

Haliloglu, T., Bahar, I., and Erman, B. (1997). “Gaussian Dynamics of Folded Proteins.” *Phys. Rev. Lett.*, 79(16), 3090–3093.

Hancock, R. E. W., and Sahl, H.-G. (2006). “Antimicrobial and host-defense peptides as new anti-infective therapeutic strategies.” *Nat. Biotechnol.*, 24(12), 1551–1557.

Hande, V., and Chakrabarty, S. (2022). “Size-Dependent Order–Disorder Crossover in Hydrophobic Hydration: Comparison between Spherical Solutes and Linear Alcohols.” *ACS Omega*, 7(3), 2671–2678.

Hess, B. (2000). “Similarities between principal components of protein dynamics and random diffusion.” *Phys. Rev. E*, 62(6), 8438–8448.

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

Hess, B., and Vegt, N. F. A. van der. (2009). “Cation specific binding with protein surface charges.” *Proc. Natl. Acad. Sci. U S A*, 106(32), 13296–13300.

Hockney, R. W., and Eastwood, J. W. (1988). *Computer simulation using particles*. Bristol [England] ; Philadelphia: A. Hilger.

Holmstrom, E. D., and Nesbitt, D. J. (2014). “Single-Molecule Fluorescence Resonance Energy Transfer Studies of the Human Telomerase RNA Pseudoknot: Temperature-/Urea-Dependent Folding Kinetics and Thermodynamics.” *J. Phys. Chem. B*, 118(14), 3853–3863.

Holz, M., Heil, S. R., and Sacco, A. (2000). “Temperature-dependent self-diffusion coefficients of water and six selected molecular liquids for calibration in accurate 1H NMR PFG measurements.” *Phys. Chem. Chem. Phys.*, 2(20), 4740–4742.

Hosseini, A. N., and Van Der Spoel, D. (2023). “Simulations of Amyloid-Forming Peptides in the Crystal State.” *Protein J.*, 42(3), 192–204.

Hribar, B., Southall, N. T., Vlachy, V., and Dill, K. A. (2002). “How Ions Affect the Structure of Water.” *J. Am. Chem. Soc.*, 124(41), 12302–12311.

Huan, Y., Kong, Q., Mou, H., and Yi, H. (2020). “Antimicrobial Peptides: Classification, Design, Application and Research Progress in Multiple Fields.” *Front. Microbiol.*, 11, 582779.

Huat, T. J., Camats-Perna, J., Newcombe, E. A., Valmas, N., Kitazawa, M., and Medeiros, R. (2019). “Metal Toxicity Links to Alzheimer’s Disease and Neuroinflammation.” *J. Mol. Biol.*, 431(9), 1843–1868.

Huggins, D. J., Biggin, P. C., Dämgen, M. A., Essex, J. W., Harris, S. A., Henchman, R. H., Khalid, S., Kuzmanic, A., Laughton, C. A., Michel, J., Mulholland, A. J., Rosta, E., Sansom, M. S. P., and Kamp, M. W. van der. (2019). “Biomolecular simulations: From dynamics and mechanisms to computational assays of biological activity.” *WIREs Comput. Mol. Sci.*, 9(3).

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

Huraskin, D., and Horn, A. H. C. (2019). “Alkali ion influence on structure and stability of fibrillar amyloid-β oligomers.” *J. Mol. Model.*, 25(2), 37.

Ibragimova, G. T., and Wade, R. C. (1998). “Importance of Explicit Salt Ions for Protein Stability in Molecular Dynamics Simulation.” *Biophys. J.*, 74(6), 2906–2911.

Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). “ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software.” *PLoS one*, (M. R. Muldoon, ed.), 9(6), e98679.

Jaeger, L. (1997). “The new world of ribozymes.” *Curr. Opin. Struct. Biol.*, 7(3), 324–335.

Jakobsson, E., Argüello-Miranda, O., Chiu, S.-W., Fazal, Z., Kruczek, J., Nunez-Corrales, S., Pandit, S., and Pritchet, L. (2017). “Towards a Unified Understanding of Lithium Action in Basic Biology and its Significance for Applied Biology.” *J. Membrane Biol.*, 250(6), 587–604.

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

Jephthah, S., Staby, L., Kragelund, B. B., and Skepö, M. (2019). “Temperature Dependence of Intrinsically Disordered Proteins in Simulations: What are We Missing?” *J. Chem. Theory Comput.*, 15(4), 2672–2683.

Jones, C. E., Abdelraheim, S. R., Brown, D. R., and Viles, J. H. (2004). “Preferential Cu2+ coordination by His96 and His111 induces beta-sheet formation in the unstructured amyloidogenic region of the prion protein.” *J Biol Chem*, 279(31), 32018–32027.

Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L. (1983). “Comparison of simple potential functions for simulating liquid water.” *J. Chem. Phys.*, 79(2), 926–935.

Jorgensen, W. L., and Tirado-Rives, J. (1988). “The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.” *J. Am. Chem. Soc.*, 110(6), 1657–1666.

Jucker, M., and Walker, L. C. (2013). “Self-propagation of pathogenic protein aggregates in neurodegenerative diseases.” *Nature*, 501(7465), 45–51.

Jucker, M., and Walker, L. C. (2015). “Amyloid-β pathology induced in humans.” *Nature*, 525(7568), 193–194.

Julió Plana, L., Nadra, A. D., Estrin, D. A., Luque, F. J., and Capece, L. (2019). “Thermal Stability of Globins: Implications of Flexibility and Heme Coordination Studied by Molecular Dynamics Simulations.” *J. Chem. Inf. Model.*, 59(1), 441–452.

Jungwirth, P. (2015). “Biological Water or Rather Water in Biology?” *J. Phys. Chem. Lett.*, 6(13), 2449–2451.

Kachala, M., Valentini, E., and Svergun, D. I. (2015). “Application of SAXS for the Structural Characterization of IDPs.” *Intrinsically Disordered Proteins Studied by NMR Spectroscopy*, Advances in Experimental Medicine and Biology, I. C. Felli and R. Pierattelli, eds., Cham: Springer International Publishing, 261–289.

Kang, H.-K., Kim, C., Seo, C. H., and Park, Y. (2017). “The therapeutic applications of antimicrobial peptides (AMPs): a patent review.” *J. Microbiol.*, 55(1), 1–12.

Khalak, Y., Baumeier, B., and Karttunen, M. (2018). “Improved general-purpose five-point model for water: TIP5P/2018.” *J. Chem. Phys.*, 149(22), 224507.

Khan, F. I., Lobb, K. A., and Lai, D. (2022). “The Molecular Basis of the Effect of Temperature on the Structure and Function of SARS-CoV-2 Spike Protein.” *Front. Mol. Biosci.*, 9, 794960.

Khan, S., Farooq, U., and Kurnikova, M. (2017). “Protein stability and dynamics influenced by ligands in extremophilic complexes – a molecular dynamics investigation.” *Mol. BioSyst.*, 13(9), 1874–1887.

Kührová, P., Mlýnský, V., Otyepka, M., Šponer, J., and Banáš, P. (2023). “Sensitivity of the RNA Structure to Ion Conditions as Probed by Molecular Dynamics Simulations of Common Canonical RNA Duplexes.” *J. Chem. Inf. Model.*, 63(7), 2133–2146.

Kumar, R., Schmidt, J. R., and Skinner, J. L. (2007). “Hydrogen bonding definitions and dynamics in liquid water.” *J. Chem. Phys.*, 126(20), 204107.

Laage, D., Stirnemann, G., and Hynes, J. T. (2009). “Why Water Reorientation Slows without Iceberg Formation around Hydrophobic Solutes.” *J. Phys. Chem. B*, 113(8), 2428–2435.

Lavery, R., Maddocks, J. H., Pasi, M., and Zakrzewska, K. (2014). “Analyzing ion distributions around DNA.” *Nucleic Acids Res.*, 42(12), 8138–8149.

Lee, S., Shek, Y. L., and Chalikian, T. V. (2010). “Urea interactions with protein groups: A volumetric study.” *Biopolymers*, 93(10), 866–879.

Lei, J., Sun, L., Huang, S., Zhu, C., Li, P., He, J., Mackey, V., Coy, D. H., and He, Q. (2019). “The antimicrobial peptides and their potential clinical applications.” *Am. J. Transl. Res.*, 11(7), 3919–3931.

Leontyev, I. V., and Stuchebrukhov, A. A. (2010). “Electronic Polarizability and the Effective Pair Potentials of Water.” *J. Chem. Theory Comput.*, 6(10), 3153–3161.

Levy, Y., and Onuchic, J. N. (2006). “WATER MEDIATION IN PROTEIN FOLDING AND MOLECULAR RECOGNITION.” *Annu. Rev. Biophys. Biomol. Struct.*, 35(1), 389–415.

Lewis, V., and Hooper, N. M. (2011). “The role of lipid rafts in prion protein biology.” *Front. Biosci. (Landmark Ed)*, 16(1), 151–168.

Li, H., Ngo, V., Da Silva, M. C., Salahub, D. R., Callahan, K., Roux, B., and Noskov, S. Yu. (2015). “Representation of Ion–Protein Interactions Using the Drude Polarizable Force-Field.” *J. Phys. Chem. B*, 119(29), 9401–9416.

Lin, B., Lopes, P. E. M., Roux, B., and MacKerell, A. D. (2013). “Kirkwood-Buff analysis of aqueous *N* -methylacetamide and acetamide solutions modeled by the CHARMM additive and Drude polarizable force fields.” *J. Chem.Phys.*, 139(8), 084509.

Liu, H., Fei, C., Chen, Y., Luo, S., Yang, T., Yang, L., Liu, J., Ji, X., Wu, W., and Song, J. (2021). “Investigating SARS-CoV-2 persistent contamination in different indoor environments.” *Environ. Res.*, 202, 111763.

Liu, S., Keten, S., and Lueptow, R. M. (2023). “Effect of molecular dynamics water models on flux, diffusivity, and ion dynamics for polyamide membrane simulations.” *J. Membrane Sci.*, 678, 121630.

Lu, H., Wang, Y., Wu, Y., Yang, P., Li, L., and Li, S. (2008). “Hydrogen-bond network and local structure of liquid water: An atoms-in-molecules perspective.” *J. Chem. Phys.*, 129(12), 124512.

Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., Bi, Y., Ma, X., Zhan, F., Wang, L., Hu, T., Zhou, H., Hu, Z., Zhou, W., Zhao, L., Chen, J., Meng, Y., Wang, J., Lin, Y., Yuan, J., Xie, Z., Ma, J., Liu, W. J., Wang, D., Xu, W., Holmes, E. C., Gao, G. F., Wu, G., Chen, W., Shi, W., and Tan, W. (2020). “Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding.” *The Lancet*, 395(10224), 565–574.

Luczkowski, M., De Ricco, R., Stachura, M., Potocki, S., Hemmingsen, L., and Valensin, D. (2015). “Metal ion mediated transition from random coil to β-sheet and aggregation of Bri2-23, a natural inhibitor of Aβ aggregation.” *Metallomics*, 7(3), 478–490.

Luo, H., Gauthier, M., Tan, X., Landry, C., Poupon, J., Dehouck, M.-P., Gosselet, F., Perrière, N., Bellivier, F., Cisternino, S., and Declèves, X. (2018). “Sodium Transporters Are Involved in Lithium Influx in Brain Endothelial Cells.” *Mol. Pharm.*.

Luo, H., Ye, F., Sun, T., Yue, L., Peng, S., Chen, J., Li, G., Du, Y., Xie, Y., Yang, Y., Shen, J., Wang, Y., Shen, X., and Jiang, H. (2004). “In vitro biochemical and thermodynamic characterization of nucleocapsid protein of SARS.” *Biophys. Chem.*, 112(1), 15–25.

Luzar, A. (2000). “Resolving the hydrogen bond dynamics conundrum.” *J. Chem. Phys.*, 113(23), 10663–10675.

Luzar, A., and Chandler, D. (1996). “Effect of Environment on Hydrogen Bond Dynamics in Liquid Water.” *Phys. Rev. Lett.*, 76(6), 928–931.

Machado-Vieira, R., Manji, H. K., and Zarate, C. A. (2009). “The role of lithium in the treatment of bipolar disorder: convergent evidence for neurotrophic effects as a unifying hypothesis.” *Bipolar Disord.*, 11(Suppl 2), 92–109.

MacKerell, A. D., Wiorkiewicz-Kuczera, J., and Karplus, M. (1995). “An all-atom empirical energy function for the simulation of nucleic acids.” *J. Am. Chem. Soc.*, 117(48), 11946–11975.

Maclean, D. S., Qian, Q., and Middaugh, C. R. (2002). “Stabilization of proteins by low molecular weight multi-ions.” *J. Pharm. Sci.*, 91(10), 2220–2229.

Mähler, J., and Persson, I. (2012). “A Study of the Hydration of the Alkali Metal Ions in Aqueous Solution.” *Inorg. Chem.*, 51(1), 425–438.

Mahoney, M. W., and Jorgensen, W. L. (2000). “A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions.” *J. Chem. Phys.*, 112(20), 8910–8922.

Maisuradze, G. G., Liwo, A., and Scheraga, H. A. (2010). “Relation between Free Energy Landscapes of Proteins and Dynamics.” *J. Chem. Theory Comput.*, 6(2), 583–595.

Makowski, J. A., Kensinger, A. H., Cunningham, C. L., Frye, C. J., Shine, M., Lackey, P. E., Mihailescu, M. R., and Evanseck, J. D. (2023). “Delta SARS-CoV-2 s2m Structure, Dynamics, and Entropy: Consequences of the G15U Mutation.” *ACS Phys. Chem Au*, 3(5), 434–443.

Mancinelli, R., Botti, A., Bruni, F., Ricci, M. A., and Soper, A. K. (2007). “Hydration of Sodium, Potassium, and Chloride Ions in Solution and the Concept of Structure Maker/Breaker.” *J. Phys. Chem. B*, 111(48), 13570–13577.

Mao, A. H., Crick, S. L., Vitalis, A., Chicoine, C. L., and Pappu, R. V. (2010). “Net charge per residue modulates conformational ensembles of intrinsically disordered proteins.” *Proc. Natl. Acad. Sci. U S A*, 107(18), 8183–8188.

Marcus, Y. (2009). “Effect of Ions on the Structure of Water: Structure Making and Breaking.” *Chem. Rev.*, 109(3), 1346–1370.

Mark, P., and Nilsson, L. (2001). “Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K.” *J. Phys. Chem. A*, 105(43), 9954–9960.

Martins de Oliveira, V., Godoi Contessoto, V. de, Bruno da Silva, F., Zago Caetano, D. L., Jurado de Carvalho, S., and Pereira Leite, V. B. (2018). “Effects of pH and Salt Concentration on Stability of a Protein G Variant Using Coarse-Grained Models.” *Biophys. J.*, 114(1), 65–75.

Martyna, G. J., Klein, M. L., and Tuckerman, M. (1992). “Nosé–Hoover chains: The canonical ensemble via continuous dynamics.” *J. Chem. Phys.*, 97(4), 2635–2643.

Megy, S., Bertho, G., Kozin, S. A., Debey, P., Hui Bon Hoa, G., and Girault, J.-P. (2009). “Possible role of region 152-156 in the structural duality of a peptide fragment from sheep prion protein.” *Protein Sci.*, 13(12), 3151–3160.

Megyes, T., Bálint, S., Grósz, T., Radnai, T., Bakó, I., and Sipos, P. (2008). “The structure of aqueous sodium hydroxide solutions: A combined solution x-ray diffraction and simulation study.” *J. Chem. Phys.*, 128(4), 044501.

Miner, J. C., Chen, A. A., and García, A. E. (2016). “Free-energy landscape of a hyperstable RNA tetraloop.” *Proc. Natl. Acad. Sci. USA*, 113(24), 6665–6670.

Miner, J. C., and García, A. E. (2017). “Equilibrium Denaturation and Preferential Interactions of an RNA Tetraloop with Urea.” *J. Phys. Chem. B*, 121(15), 3734–3746.

Miyamoto, S., and Kollman, P. A. (1992). “SETTLE: an analytical version of the SHAKE and RATTLE algorithm for rigid water models.” *J. Comput. Chem.*, 13(8), 952–962.

Mohammad, A., Al-Mulla, F., Wei, D.-Q., and Abubaker, J. (2021). “Remdesivir MD Simulations Suggest a More Favourable Binding to SARS-CoV-2 RNA Dependent RNA Polymerase Mutant P323L Than Wild-Type.” *Biomolecules*, 11(7), 919.

Mohid, Sk. A., Ghorai, A., Ilyas, H., Mroue, K. H., Narayanan, G., Sarkar, A., Ray, S. K., Biswas, K., Bera, A. K., Malmsten, M., Midya, A., and Bhunia, A. (2019). “Application of tungsten disulfide quantum dot-conjugated antimicrobial peptides in bio-imaging and antimicrobial therapy.” *Colloids Surf. B Biointerfaces*, 176, 360–370.

Mondal, J. (2019). “A brief appraisal of computational modeling of antimicrobial peptides’ activity.” *Drug. Dev. Res.*, 80(1), 28–32.

Mu, Y., Nguyen, P. H., and Stock, G. (2004). “Energy landscape of a small peptide revealed by dihedral angle principal component analysis.” *Proteins*, 58(1), 45–52.

Mukhopadhyay, S. (2020). “The Dynamism of Intrinsically Disordered Proteins: Binding-Induced Folding, Amyloid Formation, and Phase Separation.” *J. Phys. Chem. B*, 124(51), 11541–11560.

Nandi, N., and Bagchi, B. (1998). “Anomalous Dielectric Relaxation of Aqueous Protein Solutions.” *J. Phys. Chem. A*, 102(43), 8217–8221.

Nayar, D., and Chakravarty, C. (2013). “Water and water-like liquids: relationships between structure, entropy and mobility.” *Phys. Chem. Chem. Phys.*, 15(34), 14162.

Neupane, K., Munshi, S., Zhao, M., Ritchie, D. B., Ileperuma, S. M., and Woodside, M. T. (2020). “Anti-Frameshifting Ligand Active against SARS Coronavirus-2 Is Resistant to Natural Mutations of the Frameshift-Stimulatory Pseudoknot.” *J. Mol. Biol.*, 432(21), 5843–5847.

Newman, K. E. (1994). “Kirkwood–Buff solution theory: derivation and applications.” *Chem. Soc. Rev.*, 23(1), 31–40.

Nguyen, B. L., and Pettitt, B. M. (2015). “Effects of Acids, Bases, and Heteroatoms on Proximal Radial Distribution Functions for Proteins.” *J. Chem. Theory Comput.*, 11(4), 1399–1409.

Nick Pace, C., Scholtz, J. M., and Grimsley, G. R. (2014). “Forces stabilizing proteins.” *FEBS Letters*, 588(14), 2177–2184.

Niknamian, S. (2021). “On the Neglected Shifting Balance Theory, Bateson–Dobzhansky–Muller Model & Quantum Evolution Plus the Role of Mitochondrial Membrane Potential (Mmp) Impact on Covid-19.” *Open Access J. Oncol. Med.*, 4(3).

Nilsson, A., and Pettersson, L. G. M. (2015). “The structural origin of anomalous properties of liquid water.” *Nat. Commun.*, 6(1), 8998.

Nishimoto, T., Takahashi, Y., Miyama, S., Furuta, T., and Sakurai, M. (2019). “Replica exchange molecular dynamics simulation study on the mechanism of desiccation-induced structuralization of an intrinsically disordered peptide as a model of LEA proteins.” *Biophysics*, 16(0), 196–204.

Niskanen, J., Fondell, M., Sahle, C. J., Eckert, S., Jay, R. M., Gilmore, K., Pietzsch, A., Dantz, M., Lu, X., McNally, D. E., Schmitt, T., Vaz da Cruz, V., Kimberg, V., Gel’mukhanov, F., and Föhlisch, A. (2019). “Compatibility of quantitative X-ray spectroscopy with continuous distribution models of water at ambient conditions.” *Proc. Natl. Acad. Sci. USA*, 116(10), 4058–4063.

Okumura, H. (2023). “Perspective for Molecular Dynamics Simulation Studies of Amyloid-β Aggregates.” *J. Phys. Chem. B*, 127(51), 10931–10940.

Oliveira, L. B. A., and Colherinhas, G. (2020). “Can CHARMM36 atomic charges described correctly the interaction between amino acid and water molecules by molecular dynamics simulations?” *J. Mol. Liq.*, 317, 113919.

Olson, W. K., Bansal, M., Burley, S. K., Dickerson, R. E., Gerstein, M., Harvey, S. C., Heinemann, U., Lu, X.-J., Neidle, S., Shakked, Z., Sklenar, H., Suzuki, M., Tung, C.-S., Westhof, E., Wolberger, C., and Berman, H. M. (2001). “A standard reference frame for the description of nucleic acid base-pair geometry.” *J. Mol. Biol.*, 313(1), 229–237.

Pace, C. N., Fu, H., Lee Fryar, K., Landua, J., Trevino, S. R., Schell, D., Thurlkill, R. L., Imura, S., Scholtz, J. M., Gajiwala, K., Sevcik, J., Urbanikova, L., Myers, J. K., Takano, K., Hebert, E. J., Shirley, B. A., and Grimsley, G. R. (2014). "Contribution of hydrogen bonds to protein stability: Hydrogen Bonds and Protein Stability." *Protein Sci.*, 23(5), 652–661.

Pan, F., Roland, C., and Sagui, C. (2014). "Ion distributions around left- and right-handed DNA and RNA duplexes: a comparative study." *Nucleic Acids Res.*, 42(22), 13981–13996.

Pan, T. (1995). "Higher order folding and domain analysis of the ribozyme from Bacillus subtilis ribonuclease P." *Biochemistry*, 34(3), 902–909.

Pan, Y., Short, J. L., Newman, S. A., Choy, K. H. C., Tiwari, D., Yap, C., Senyschyn, D., Banks, W. A., and Nicolazzo, J. A. (2018). "Cognitive benefits of lithium chloride in APP/PS1 mice are associated with enhanced brain clearance of β-amyloid." *Brain Behav. Immun.*, 70, 36–47.

Park, S.-J., Kim, Y.-G., and Park, H.-J. (2011). "Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening." *J. Am. Chem. Soc.*, 133(26), 10094–10100.

Parrinello, M., and Rahman, A. (1980). "Crystal Structure and Pair Potentials: A Molecular-Dynamics Study." *Phys. Rev. Lett.*, 45(14), 1196–1199.

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

Pavlek, A.-M., Pem, B., and Bakarić, D. (2024). "The Signature of Fluctuations of the Hydrogen Bond Network Formed by Water Molecules in the Interfacial Layer of Anionic Lipids." *Biophysica*, 4(1), 92–106.

Persson, E., and Halle, B. (2008). "Cell water dynamics on multiple time scales." *Proc. Natl. Acad Sci.*, 105(17), 6266–6271.

Persson, F., and Halle, B. (2018). “Compressibility of the protein-water interface.” *J. Chem. Phys.*, 148(21), 215102.

Pierce, V., Kang, M., Aburi, M., Weerasinghe, S., and Smith, P. E. (2008). “Recent Applications of Kirkwood–Buff Theory to Biological Systems.” *Cell. Biochem. Biophys.*, 50(1), 1–22.

Poole, P. H., Sciortino, F., Essmann, U., and Stanley, H. E. (1992). “Phase behaviour of metastable water.” *Nature*, 360(6402), 324–328.

Porcelli, F., Casavola, A. R., Grottesi, A., Schiumarini, D., and Avaldi, L. (2024). “Probing the conformational dynamics of an Ago–RNA complex in water/methanol solution.” *Phys. Chem. Chem. Phys.*, 26(3), 2497–2508.

Preußke, N., Sönnichsen, F. D., and Leippe, M. (2023). “A guided tour through α-helical peptide antibiotics and their targets.” *Biosci. Rep.*, 43(5), BSR20230474.

Průša, J., and Cifra, M. (2020). “Dataset of molecular dynamics simulation trajectories of amino-acid solutions with various force fields, water models and modified force field parameters.” *Data in Brief*, 30, 105483.

Rahman, A., and Stillinger, F. H. (1971). “Molecular Dynamics Study of Liquid Water.” *J. Chem. Phys.*, 55(7), 3336–3359.

Remko, M., and Rode, B. M. (2006). “Effect of Metal Ions (Li^+, Na^+, K^+, Mg^{2+}, Ca^{2+}, Ni^{2+}, Cu^{2+}, and Zn^{2+}) and Water Coordination on the Structure of Glycine and Zwitterionic Glycine.” *J. Phys. Chem. A*, 110(5), 1960–1967.

Rezaei-Ghaleh, N. (2022). “Water Dynamics in Highly Concentrated Salt Solutions: A Multi-Nuclear NMR Approach.” *ChemistryOpen*, 11(6), e202200080.

Rodrigues, J. P. G. L. M., Barrera-Vilarmau, S., M. C. Teixeira, J., Sorokina, M., Seckel, E., Kastritis, P. L., and Levitt, M. (2020). “Insights on cross-species transmission of SARS-CoV-2 from structural modeling.” *PLoS Comput. Biol.*, (C. M. Deane, ed.), 16(12), e1008449.

Romero, C. P., Jeldres, R. I., Quezada, G. R., Concha, F., and Toledo, P. G. (2018). "Zeta potential and viscosity of colloidal silica suspensions: Effect of seawater salts, pH, flocculant, and shear rate." *Colloids Surf. A: Physicochem. Eng. Asp.*, 538, 210–218.

Rozhkov, S. P., and Goryunov, A. S. (2014). "Phase states of water-protein(polypeptide)-salt system and reaction to external environment factors." *Biophysics*, 59(1), 43–48.

Sanches, M. N., Knapp, K., Oliveira, A. B., Wolynes, P. G., Onuchic, J. N., and Leite, V. B. P. (2022). "Examining the Ensembles of Amyloid-β Monomer Variants and Their Propensities to Form Fibers Using an Energy Landscape Visualization Method." *J. Phys. Chem. B*, 126(1), 93–99.

Santuccione, A., Sytnyk, V., Leshchyns'ka, I., and Schachner, M. (2005). "Prion protein recruits its neuronal receptor NCAM to lipid rafts to activate p59fyn and to enhance neurite outgrowth." *J. Cell. Biol.*, 169(2), 341–354.

Saravi, S. H., and Panagiotopoulos, A. Z. (2021). "Individual Ion Activity Coefficients in Aqueous Electrolytes from Explicit-Water Molecular Dynamics Simulations." *J. Phys. Chem. B*, 125(30), 8511–8521.

Sarkar, K., Meister, K., Sethi, A., and Gruebele, M. (2009). "Fast Folding of an RNA Tetraloop on a Rugged Energy Landscape Detected by a Stacking-Sensitive Probe." *Biophys. J.*, 97(5), 1418–1427.

Sato, H., and Feix, J. B. (2006). "Peptide–membrane interactions and mechanisms of membrane destruction by amphipathic α-helical antimicrobial peptides." *Biochim. Biophys. Acta Biomembranes*, 1758(9), 1245–1256.

Schauss, J., Kundu, A., Fingerhut, B. P., and Elsaesser, T. (2021). "Magnesium Contact Ions Stabilize the Tertiary Structure of Transfer RNA: Electrostatics Mapped by Two-Dimensional Infrared Spectra and Theoretical Simulations." *J. Phys. Chem. B*, 125(3), 740–747.

Schlick, T., Zhu, Q., Dey, A., Jain, S., Yan, S., and Laederach, A. (2021). “To knot or not to knot: Multiple conformations of the SARS-CoV-2 frameshifting RNA element.” *J. Am. Chem. Soc*. 143(30), 11404-11422.

Sha, W., Wu, X., and Keong, K. G. (2011). “Molecular dynamics (MD) simulation of the diamond pyramid structure in electroless copper deposits.” *Electroless Copper and Nickel–Phosphorus Plating*, Elsevier, 82–103.

Shanker, S., and Bandyopadhyay, P. (2017). “How Mg2+ ion and water network affect the stability and structure of non-Watson-Crick base pairs in E. coli loop E of 5S rRNA: a molecular dynamics and reference interaction site model (RISM) study.” *J. Biomol. Struct. Dyn.*, 35(10), 2103–2122.

Shi, Y.-Z., Jin, L., Feng, C.-J., Tan, Y.-L., and Tan, Z.-J. (2018). “Predicting 3D structure and stability of RNA pseudoknots in monovalent and divalent ion solutions.” *PLoS Comput. Biol.*, (T. M. Przytycka, ed.), 14(6), e1006222.

Shinto, H., Morisada, S., and Higashitani, K. (2005). “Potentials of Mean Force for Hydrophilic–Hydrophobic Solute Pairs in Water.” *J. Chem. Eng. Jpn.*, 38(7), 465–477.

Siebert, T., Guchhait, B., Liu, Y., Fingerhut, B. P., and Elsaesser, T. (2016). “Range, Magnitude, and Ultrafast Dynamics of Electric Fields at the Hydrated DNA Surface.” *J. Phys. Chem. Lett.*, 7(16), 3131–3136.

Silva, G. A. (2004). “Introduction to nanotechnology and its applications to medicine.” *Surg. Neurol.*, 61(3), 216–220.

Silva, J. A. B. da, Moreira, F. G. B., Santos, V. M. L. dos, and Longo, R. L. (2011). “Hydrogen bond networks in water and methanol with varying interaction strengths.” *Phys. Chem. Chem. Phys.*, 13(2), 593–603.

Silvestrelli, P. L., and Parrinello, M. (1999). “Structural, electronic, and bonding properties of liquid water from first principles.” *J. Chem. Phys.*, 111(8), 3572–3580.

Sim, A. Y. L., and Levitt, M. (2011). “Clustering to identify RNA conformations constrained by secondary structure.” *Proc. Natl. Acad. Sci. USA*, 108(9), 3590–3595.

Singh, O., and Chakraborty, D. (2021). “Preferential binding affinity of ions and their effect on structure and dynamics of water near antimicrobial peptide.” *J. Mol. Liq.*, 344, 117789.

Singh, O., Kumar Das, B., and Chakraborty, D. (2022). “Influence of Ion Specificity and Concentration on the Conformational Transition of Intrinsically Disordered Sheep Prion Peptide.” *ChemPhysChem*, 23(16).

Singh, O., Venugopal, P. P., Mathur, A., and Chakraborty, D. (2021). “Temperature-Dependent Conformational Evolution of SARS CoV-2 RNA Genome Using Network Analysis.” *J. Phys. Chem. B*, 125(38), 10672–10681.

Singh, R. K., Chamachi, N. G., Chakrabarty, S., and Mukherjee, A. (2017). “Mechanism of Unfolding of Human Prion Protein.” *J. Phys. Chem. B*, 121(3), 550–564.

Sinha, S. K., and Bandyopadhyay, S. (2011). “Dynamic properties of water around a protein–DNA complex from molecular dynamics simulations.” *J. Chem. Phys.*, 135(13), 135101.

Smith, M. D., and Cruz, L. (2013). “Effect of Ionic Aqueous Environments on the Structure and Dynamics of the Aβ21–30 Fragment: A Molecular-Dynamics Study.” *J. Phys. Chem. B*, 117(22), 6614–6624.

Smyth, M. S. (2000). “x Ray crystallography.” *Mol. Pathol.*, 53(1), 8–14.

Song, B., Cho, J.-H., and Raleigh, D. P. (2007). “Ionic-Strength-Dependent Effects in Protein Folding: Analysis of Rate Equilibrium Free-Energy Relationships and Their Interpretation.” *Biochemistry*, 46(49), 14206–14214.

Soper, A. K., and Phillips, M. G. (1986). “A new determination of the structure of water at 25°C.” *Chem. Phys.*, 107(1), 47–60.

Soper, A. K., and Weckström, K. (2006). “Ion solvation and water structure in potassium halide aqueous solutions.” *Biophys. Chem.*, 124(3), 180–191.

Šponer, J., Bussi, G., Krepl, M., Banáš, P., Bottaro, S., Cunha, R. A., Gil-Ley, A., Pinamonti, G., Poblete, S., Jurečka, P., Walter, N. G., and Otyepka, M. (2018). “RNA Structural Dynamics As Captured by Molecular Simulations: A Comprehensive Overview.” *Chem. Rev.*, 118(8), 4177–4338.

Šponer, J., Krepl, M., Banáš, P., Kührová, P., Zgarbová, M., Jurečka, P., Havrila, M., and Otyepka, M. (2017). “How to understand atomistic molecular dynamics simulations of RNA and protein-RNA complexes?: Atomistic molecular dynamics simulations of RNA and protein-RNA complexes.” *WIREs RNA*, 8(3), e1405.

Stancik, A. L., and Brauns, E. B. (2008). “Rearrangement of Partially Ordered Stacked Conformations Contributes to the Rugged Energy Landscape of a Small RNA Hairpin.” *Biochemistry*, 47(41), 10834–10840.

Starr, F. W., and Sciortino, F. (2014). “‘Crystal-clear’ liquid–liquid transition in a tetrahedral fluid.” *Soft Matter*, 10(47), 9413–9422.

Strandberg, E., Wadhwani, P., Bürck, J., Anders, P., Mink, C., Van Den Berg, J., Ciriello, R. A. M., Melo, M. N., Castanho, M. A. R. B., Bardají, E., Ulmschneider, J. P., and Ulrich, A. S. (2023). “Temperature-Dependent Re-alignment of the Short Multifunctional Peptide BP100 in Membranes Revealed by Solid-State NMR Spectroscopy and Molecular Dynamics Simulations.” *ChemBioChem*, 24(4), e202200602.

Sugita, Y., and Okamoto, Y. (1999). “Replica-exchange molecular dynamics method for protein folding.” *Chem. Phys. Lett.*, 314(1–2), 141–151.

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

Szabo, S. T., Harry, G. J., Hayden, K. M., Szabo, D. T., and Birnbaum, L. (2016). “Comparison of Metal Levels between Postmortem Brain and Ventricular Fluid in Alzheimer’s Disease and Nondemented Elderly Controls.” *Toxicol. Sci.*, 150(2), 292–300.

Tachi, Y., Itoh, S. G., and Okumura, H. (2022). “Molecular dynamics simulations of amyloid-β peptides in heterogeneous environments.” *Biophysics*, 19(0).

Tamamizu-Kato, S., Kosaraju, M. G., Kato, H., Raussens, V., Ruysschaert, J.-M., and Narayanaswami, V. (2006). “Calcium-Triggered Membrane Interaction of the α-Synuclein Acidic Tail.” *Biochemistry*, 45(36), 10947–10956.

Tanemura, M., Ogawa, T., and Ogita, N. (1983). “A new algorithm for three-dimensional voronoi tessellation.” *J. Comput. Phys.*, 51(2), 191–207.

Tanner, N. K. (1999). “Ribozymes: the characteristics and properties of catalytic RNAs.” *FEMS Microbiol. Rev.*, 23(3), 257–275.

Tarek, M., and Tobias, D. J. (2000). “The Dynamics of Protein Hydration Water: A Quantitative Comparison of Molecular Dynamics Simulations and Neutron-scattering Experiments.” *Biophys. J.*, 79(6), 3244–3257.

Theimer, C. A., Finger, L. D., Trantirek, L., and Feigon, J. (2003). “Mutations linked to dyskeratosis congenita cause changes in the structural equilibrium in telomerase RNA.” *Proc. Natl. Acad. Sci.*, 100(2), 449–454.

Thomas, A. S., and Elcock, A. H. (2006). “Direct Observation of Salt Effects on Molecular Interactions through Explicit-Solvent Molecular Dynamics Simulations: Differential Effects on Electrostatic and Hydrophobic Interactions and Comparisons to Poisson−Boltzmann Theory.” *J. Am. Chem. Soc.*, 128(24), 7796–7806.

Tobias, D. J., and Hemminger, J. C. (2008). “Getting Specific About Specific Ion Effects.” *Science*.

Tobler, I., Gaus, S. E., Deboer, T., Achermann, P., Fischer, M., Rülicke, T., Moser, M., Oesch, B., McBride, P. A., and Manson, J. C. (1996). "Altered circadian activity rhythms and sleep in mice devoid of prion protein." *Nature*, 380(6575), 639–642.

Tompa, P. (2002). "Intrinsically unstructured proteins." *Trends Biochem. Sc.*, 27(10), 527–533.

Tõugu, V., Tiiman, A., and Palumaa, P. (2011). "Interactions of Zn(II) and Cu(II) ions with Alzheimer's amyloid-beta peptide. Metal ion binding, contribution to fibrillization and toxicity." *Metallomics*, 3(3), 250–261.

Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G. (2014). "PLUMED 2: New feathers for an old bird." *Comput. Phys. Commun.*, 185(2), 604–613.

Ussing, H. H. (1959). "The alkali metal ions in isolated systems and tissues." *The Alkali Metal Ions in Biology*, Berlin, Heidelberg: Springer Berlin Heidelberg, 1–195.

Uversky, V. N. (2009). "Intrinsically Disordered Proteins and Their Environment: Effects of Strong Denaturants, Temperature, pH, Counter Ions, Membranes, Binding Partners, Osmolytes, and Macromolecular Crowding." *Protein J.*, 28(7–8), 305–325.

Uversky, V. N. (2013). "A decade and a half of protein intrinsic disorder: Biology still waits for physics: Protein Intrinsic Disorder." *Protein Sci.*, 22(6), 693–724.

Uversky, V. N. (2014). "Introduction to Intrinsically Disordered Proteins (IDPs)." *Chem. Rev.*, 114(13), 6557–6560.

Uversky, V. N., and Fink, A. L. (2004). "Conformational constraints for amyloid fibrillation: the importance of being unfolded." *Biochim. Biophys. Acta - Proteins and Proteomics*, 1698(2), 131–153.

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C. (2005). "GROMACS: Fast, flexible, and free." *J. Comput. Chem.*, 26(16), 1701–1718.

Vancraenenbroeck, R., Harel, Y. S., Zheng, W., and Hofmann, H. (2019). “Polymer effects modulate binding affinities in disordered proteins.” *Proc. Natl. Acad. Sci.*, 116(39), 19506–19512.

Várnai, P., Canalia, M., and Leroy, J.-L. (2004). “Opening Mechanism of G·T/U Pairs in DNA and RNA Duplexes: A Combined Study of Imino Proton Exchange and Molecular Dynamics Simulation.” *J. Am. Chem. Soc.*, 126(44), 14659–14667.

Vedamuthu, M., Singh, S., and Robinson, G. W. (1994). “Properties of Liquid Water: Origin of the Density Anomalies.” *J. Phys. Chem.*, 98(9), 2222–2230.

Venugopal, P. P., and Chakraborty, D. (2022). “Molecular mechanism of inhibition of COVID-19 main protease by β-adrenoceptor agonists and adenosine deaminase inhibitors using *in silico* methods.” *J. Biomol. Struct. Dyn.*, 40(11), 5112–5127.

Venugopal, P. P., M, S., and Chakraborty, D. (2021). “Theoretical insights into molecular mechanism and energy criteria of PARP-2 enzyme inhibition by benzimidazole analogues.” *Proteins: Struct. Funct. Bioinform.*, 89 (8), 988-1004.

Venugopal, P. P., Singh, O., and Chakraborty, D. (2022). “Understanding the role of water on temperature-dependent structural modifications of SARS CoV-2 main protease binding sites.” *J. Mol. Liq.*, 363, 119867.

Verlet, L. (1967). “Computer ‘Experiments’ on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules.” *Phys. Rev.*, 159(1), 98–103.

Vitalis, A., and Pappu, R. V. (2009). “ABSINTH: A new continuum solvation model for simulations of polypeptides in aqueous solutions.” *J. Comput. Chem.*, 30(5), 673–699.

Voronoi, G. (1908). “Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Deuxième mémoire. Recherches sur les parallélloèdres primitifs.” *J. reine angew. Math.*, 1908(134), 198–287.

Wadsworth, J. D., Hill, A. F., Joiner, S., Jackson, G. S., Clarke, A. R., and Collinge, J. (1999). “Strain-specific prion-protein conformation determined by metal ions.” *Nat. Cell. Biol.*, 1(1), 55–59.

Wang, D., and Farhana, A. (2022). “Biochemistry, RNA Structure.” *StatPearls*, Treasure Island (FL): StatPearls Publishing.

Wang, J. C. (1969). “Variation of the average rotation angle of the DNA helix and the superhelical turns of covalently closed cyclic λ DNA.” *J. Mol. Biol.*, 43(1), 25–39.

Weiner, S. J., Kollman, P. A., Case, D. A., Singh, U. C., Ghio, C., Alagona, G., Profeta, S., and Weiner, P. (1984). “A new force field for molecular mechanical simulation of nucleic acids and proteins.” *J. Am. Chem. Soc.*, 106(3), 765–784.

Wen, J., Shen, X., Shen, H., and Zhang, F.-S. (2014). “Hofmeister series and ionic effects of alkali metal ions on DNA conformation transition in normal and less polarised water solvent.” *Mol. Phys.*, 112(20), 2707–2719.

Westhof, E. (1988). “Water: An Integral Part of Nucleic Acid Structure.” *Annu. Rev. Biophys. Biophys. Chem.*, 17(1), 125–144.

Wicky, B. I. M., Shammas, S. L., and Clarke, J. (2017). “Affinity of IDPs to their targets is modulated by ion-specific changes in kinetics and residual structure.” *Proc. Natl. Acad. Sci*, 114(37), 9882–9887.

Wiggins, P. (2008). “Life Depends upon Two Kinds of Water.” *PLoS one*, (S. Zhang, ed.), 3(1), e1406.

Wohl, S., Jakubowski, M., and Zheng, W. (2021). “Salt-Dependent Conformational Changes of Intrinsically Disordered Proteins.” *J. Phys. Chem. Lett.*, 12(28), 6684–6691.

Wolf, M. G., and Groenhof, G. (2012). “Evaluating nonpolarizable nucleic acid force fields: A systematic comparison of the nucleobases hydration free energies and chloroform-to-water partition coefficients.” *J. Comput. Chem.*, 33(28), 2225–2232.

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

Wright, P. E., and Dyson, H. J. (2015). “Intrinsically disordered proteins in cellular signalling and regulation.” *Nat. Rev. Mol. Cell. Biol.*, 16(1), 18–29.

Wu, F., Zhao, S., Yu, B., Chen, Y.-M., Wang, W., Song, Z.-G., Hu, Y., Tao, Z.-W., Tian, J.-H., Pei, Y.-Y., Yuan, M.-L., Zhang, Y.-L., Dai, F.-H., Liu, Y., Wang, Q.-M., Zheng, J.-J., Xu, L., Holmes, E. C., and Zhang, Y.-Z. (2020). “A new coronavirus associated with human respiratory disease in China.” *Nature*, 579(7798), 265–269.

Xia, T., SantaLucia, J., Burkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., and Turner, D. H. (1998). “Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson−Crick Base Pairs.” *Biochemistry*, 37(42), 14719–14735.

Yan, S., Zhu, Q., Hohl, J., Dong, A., and Schlick, T. (2023). “Evolution of coronavirus frameshifting elements: Competing stem networks explain conservation and variability.” *Proc. Natl. Acad. Sci. U.S.A.*, 120(20), e2221324120.

Yan, Z., Buldyrev, S. V., and Stanley, H. E. (2008). “Relation of water anomalies to the excess entropy.” *Phys. Rev. E*, 78(5), 051201.

Yeh, Y., and Mou, C.-Y. (1999). “Orientational Relaxation Dynamics of Liquid Water Studied by Molecular Dynamics Simulation.” *J. Phys. Chem. B*, 103(18), 3699–3705.

Zapałowski, M., and Bartczak, W. M. (2000). “Structural and dynamical properties of concentrated aqueous NaOH solutions: a computer simulation study.” *Comput. Chem.*, 24(3–4), 459–468.

Zerze, G. H., Best, R. B., and Mittal, J. (2015). “Sequence- and Temperature-Dependent Properties of Unfolded and Disordered Proteins from Atomistic Simulations.” *J. Phys. Chem. B*, 119(46), 14622–14630.

Zhang, K., Zheludev, I. N., Hagey, R. J., Haslecker, R., Hou, Y. J., Kretsch, R., Pintilie, G. D., Rangan, R., Kladwang, W., Li, S., Wu, M. T.-P., Pham, E. A., Bernardin-Souibgui, C., Baric, R. S., Sheahan, T. P., D'Souza, V., Glenn, J. S., Chiu, W., and Das, R. (2021). "Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome." *Nat. Struct. Mol. Biol.*, 28(9), 747–754.

Zhang, R., Jalali, S., Dias, C. L., and Haataja, M. P. (2024). "Growth kinetics of amyloid-like fibrils: An integrated atomistic simulation and continuum theory approach." *PNAS Nexus*, 3(2), 045.

Zhang, Z., Shi, Y., and Liu, H. (2003). "Molecular Dynamics Simulations of Peptides and Proteins with Amplified Collective Motions." *Biophys. J.*, 84(6), 3583–3593.

Zhou, H.-X. (2005). "Interactions of macromolecules with salt ions: An electrostatic theory for the Hofmeister effect." *Proteins: Struct. Funct. Bioinform.*, 61(1), 69–78.

Zou, Q., Bennion, B. J., Daggett, V., and Murphy, K. P. (2002). "The Molecular Mechanism of Stabilization of Proteins by TMAO and Its Ability to Counteract the Effects of Urea." *J. Am. Chem. Soc.*, 124(7), 1192–1202.

LIST OF PUBLICATIONS

1. **Singh, O.,** Das, B. K., and Chakraborty, D. (2022). "Influence of Ion Specificity and Concentration on the Conformational Transition of Intrinsically Disordered Sheep Prion Peptide." *ChemPhysChem*, 23(16).
2. **Singh, O.,** Venugopal, P. P., Mathur, A., and Chakraborty, D. (2022). "Exploring the multiple conformational states of RNA genome through interhelical dynamics and network analysis." *Journal of Molecular Graphics and Modelling*, 116, 108264.
3. **Singh, O.,** Venugopal, P. P., Mathur, A., and Chakraborty, D. (2021). "Temperature-Dependent Conformational Evolution of SARS CoV-2 RNA Genome Using Network Analysis." *Journal of Physical Chemistry B*, 125(38), 10672–10681.
4. **Singh, O.,** and Chakraborty, D. (2021). "Preferential binding affinity of ions and their effect on structure and dynamics of water near antimicrobial peptide." *Journal of Molecular Liquids*, 344(5).
5. **Singh, O.,** Venugopal, P. P., and Chakraborty, D. (2023). "Effect of Water Models on The Stability of RNA: Role of Counter-Ions." *Chemical Physics Impact*, 7, 100313.
6. Chakraborty, D., **Singh, O.,** and Divya P. (2024). "Study of Correlated Motions to Detect the Conformational Transitions of the Intrinsically Disordered Sheep Prion Peptide." *Journal of Chemical Information and Modeling*.

Other Publications

1. Das, B. K., **Singh, O.**, and Chakraborty, D. (2023). "Exploring the Barriers in the Aggregation of a Hexadecameric Human Prion Peptide through the Markov State Model." *ACS Chemical Neuroscience*, 14(19):3622-3645.
2. Sahu, J., **Singh, O.**, Chakraborty, D., Sadhu, K. (2023). "Growth Reaction of Gold Nanorod for Discriminating α-Helical Peptide from Mutated Peptide with Helix at Amino Terminus." *Chemistry an Asian Journal*, 18(8):e202300049.

3. Venugopal, P. P., **Singh, O.**, and Chakraborty, D. (2022). "Understanding the Role of Water on Temperature-dependent Structural Modifications of SARS CoV-2 Main Protease Binding Sites." *Journal of Molecular Liquids*, 363, 119867.

LIST OF CONFERENCES

1. **Singh, Omkar,** and D. Chakraborty "Characterization of biological water at interface of an antimicrobial peptide in presence of salt solutions" in MOLSIM at IIT Kanpur 2020.
2. **Singh, Omkar** and D. Chakraborty "Simulation and analysis of prion peptide structure in Lithium chloride solution" in virtual International Conference on Molecules to Materials at SVNIT –Surat 2020.
3. **Singh, Omkar** and D. Chakraborty "Structure evolution of SARS CoV-2 RNA Genome Using Molecular Dynamics Simulation: A Network Analysis" in 17th Theoretical Chemistry Symposium 2021**.**

BIODATA

Name: Omkar Singh

PhD Registration Number: 187081CY005

e-mail id: omkantnirala92@gmail.com

EDUCATIONAL QUALIFICATIONS

Examination	Institution	Board	Duration	Marks
PhD (CHEMISTRY)	National Institute of Technology, Surathkal	NITK	2018-2024	7.86/ 10 (Coursework)
MSc (CHEMISTRY)	Gurukula Kangri Viswavidhyalaya Haridwar	GKVH	2012-2014	73.95%
BSc (CHEMISTRY)	Hindu College Moradabad	MJPRU Bareilly	2009-2012	54.00%
HSC	SIC Narangpur	UP Board	2009	73.00%
SSC	MHMIC Asgaripur	UP Board	2007	65.95%

ACHIEVEMENTS

- **GATE 2016** qualified.
- **CSIR-NET** qualified **AIR-53**.
- Diploma in language "C"
- Awarded "**Quantumzyme Certified Professional**" in 2 days workshop on "Advanced Molecular Modelling and Simulation: SARS CoV-2 Case Study" organized by Quantumzyme LLP, Bangalore, India, 29th and 30th January 2021.

LIST OF PUBLICATIONS

1. **Singh, O.,** Das, B. K., and Chakraborty, D. (2022). "Influence of Ion Specificity and Concentration on the Conformational Transition of Intrinsically Disordered Sheep Prion Peptide." *ChemPhysChem*, 23(16).
2. **Singh, O.,** Venugopal, P. P., Mathur, A., and Chakraborty, D. (2022). "Exploring the multiple conformational states of RNA genome through interhelical dynamics and network analysis." *Journal of Molecular Graphics and Modelling*, 116, 108264.
3. **Singh, O.,** Venugopal, P. P., Mathur, A., and Chakraborty, D. (2021). "Temperature-Dependent Conformational Evolution of SARS CoV-2 RNA Genome Using Network Analysis." *Journal of Physical Chemistry B*, 125(38), 10672–10681.
4. **Singh, O.,** and Chakraborty, D. (2021). "Preferential binding affinity of ions and their effect on structure and dynamics of water near antimicrobial peptide." *Journal of Molecular Liquids*, 344(5).
5. **Singh, O.,** Venugopal, P. P., and Chakraborty, D. (2023). "Effect of Water Models on The Stability of RNA: Role of Counter-Ions." *Chemical Physics Impact*, 7, 100313.
6. Chakraborty, D., **Singh, O.,** and Divya P. (2024). "Study of Correlated Motions to Detect the Conformational Transitions of the Intrinsically Disordered Sheep Prion Peptide." *Journal of Chemical Information and Modeling*.

Other Publications

1. Das, B. K., **Singh, O.**, and Chakraborty, D. (2023). "Exploring the Barriers in the Aggregation of a Hexadecameric Human Prion Peptide through the Markov State Model." *ACS Chemical Neuroscience*, 14(19):3622-3645.
2. Sahu, J., **Singh, O.,** Chakraborty, D., Sadhu, K. (2023). "Growth Reaction of Gold Nanorod for Discriminating α-Helical Peptide from Mutated Peptide with Helix at Amino Terminus." *Chemistry an Asian Journal*, 18(8):e202300049.

3. Venugopal, P. P., **Singh, O.,** and Chakraborty, D. (2022). "Understanding the Role of Water on Temperature-dependent Structural Modifications of SARS CoV-2 Main Protease Binding Sites." *Journal of Molecular Liquids*, 363, 119867.

LIST OF CONFERENCES

1. **Singh, Omkar,** and D. Chakraborty "Characterization of biological water at interface of an antimicrobial peptide in presence of salt solutions" in MOLSIM at IIT Kanpur 2020.
2. **Singh, Omkar** and D. Chakraborty "Simulation and analysis of prion peptide structure in Lithium chloride solution" in virtual International Conference on Molecules to Materials at SVNIT –Surat 2020.
3. **Singh, Omkar** and D. Chakraborty "Structure evolution of SARS CoV-2 RNA Genome Using Molecular Dynamics Simulation: A Network Analysis" in 17th Theoretical Chemistry Symposium 2021**.**

WORKSHOPS ATTENDED

1. Two weeks of Summer school on "**Advanced Molecular Dynamic Simulation 2021**" Participated organized by KU Nepal and IISER Pune, India.
2. Two days workshop on **"Advanced Molecular Modelling and Simulation: SARS CoV-2 Case Study"** organized by Quantumzyme LLP, Bangalore, India, 29th and 30th January 2021.
3. Virtual Workshop on "**Molecular Dynamics Simulations for Beginners using LAMMPS**" on 13th-17th September 2021.

Place: Surathkal

Date: 01-05-2024