Molecular understanding of Three-Way Junction Prohead RNA (3WJ-pRNA) stability

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Pratishruti Panda

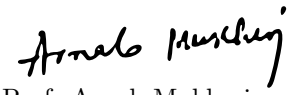
Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

May, 2024

Supervisor: Prof. Arnab Mukherjee

Certificate

This is to certify that this dissertation entitled *Molecular understanding of Three-Way Junction Prohead RNA (3WJ-pRNA) stability* towards the partial fulfillment of the BS-MS dual degree program at the Indian Institute of Science Education and Research, Pune, represents study/work carried out by Pratishruti Panda at Indian Institute of Science Education and Research under the supervision of Prof. Arnab Mukherjee, Department of Chemistry, Indian Institute of Science Education and Research, Pune , during the academic year 2023-2024.



Prof. Arnab Mukherjee

Committee:

Prof. Arnab Mukherjee

Dr. Susmita Roy

Dedicated to all my teachers...

Declaration

I hereby declare that the matter embodied in the report entitled *Molecular understanding of Three-Way Junction Prohead RNA (3WJ-pRNA) stability* are the results of the work carried out by me at the Department of Chemistry, Indian Institute of Science Education and Research, Pune, under the supervision of Prof. Arnab Mukherjee and the same has not been submitted elsewhere for any other degree. Wherever others contribute, every effort is made to indicate this clearly, with due reference to the literature and acknowledgment of collaborative research and discussions.

Pratishruti Panda

Acknowledgements

At the culmination of this academic checkpoint, I would like to express my gratitude to those who have been instrumental in shaping my research experience and personal growth.

Firstly, I would like to express my heartfelt gratitude to my supervisor, **Prof. Arnab Mukherjee** for his support and motivation . They are indispensable for the start of my research journey through my MS project. He is an inspiration to give my best in every aspect of academic pursuit and beyond. I would like to thank **Dr. Ashwani Sharma** of IISER Tirupati for the initial conceptualization of the research problem. I am grateful to **Dr. Susmita Roy**, expert of my Thesis Advisory Committee (TAC), whose insightful feedback and rigorous work have given direction to my work. Special thanks to **Dr. Amrita Hazra** for giving me the opportunity to collaborate and for enriching lessons of chemical biology.

I would like to thank IISER, Pune for the state-of-the-art computing facilities of PARAM BRAHMA, the enriching academic environment, and sports facilities. All of it has collectively enhanced my academic and personal development during my time here.

I owe a lot to my mentor, *Uddipan Bhattacharya*, for his continuous support and guidance in all phases of this project. I extend gratitude to my collaborators, *Prathith Bhargav* and *Ateek Shah*. I thank *Ashish* and *Bikirna* for discussions and *Parth* for lending resources. The collaborative atmosphere within the CCB lab has been a driving force to learn new things. I extend my thanks to all my labmates, past and present: *Ashish*, *Amal*, *Bikirna*, *Gopi*, *Sarath*, *Sonali*, *Parth*, *Prathith*, *Mansi*, *Uddipan*, and *Rahul*, for their camaraderie and engaging discussions.

A heartfelt thanks to my friends; *Ateek*, *Vidushi*, *Sujal*, *Lokendra*, *Prathith*, *Shreelekha*, *Raghav*, and others, who have been my support and encouraged me in different phases of this journey. I want to acknowledge the friendships made beyond classrooms and labs. The volleyball court has been a second home to me at IISER. I thank *Ateek*, *Shreelekha*, *Shuvam Kant*, *Abhishek*, *Salam Maheshwori*, and *Gopi*, who have added richness and balance to my life outside of academics. I thank all my teammates from different teams who have taught me to celebrate wins and learn from losses.

I thank my teachers from school; *Nilima ma'am*, *Debashish sir*, *Pramod sir* and *Panchanan sir* for their teachings and encouragement through tough times. Lastly, I offer heartfelt thanks to my family; my father, mother, and brother, for their unwavering patience, support, and for consistently bestowing the best upon me. Their encouragement has been a constant source of strength.

Abstract

Ribonucleic Acids (RNA) are one of the most pluri-potent chemical species. They play important roles of regulatory networks and catalyzing chemical reactions. With the advent of advanced molecular tools, RNA nanotechnology has gained attention. Its applications in biomedicine are noteworthy, offering fast, flexible and targeted drug delivery. They also have advantage of low immunogenicity compared to proteins.

One promising RNA motif is the ϕ29 bacteriophage three-way junction motif from prohead RNA (3WJ-pRNA). It is well-known for self-assembly and stability, making it suitable for biomedical applications. But little is understood about the specific mechanism responsible for its extraordinary thermodynamic stability. This study attempts to look into the dynamics of 3WJ-pRNA and its binding to fluorogenic ligands through molecular dynamics (MD) . Major factors affecting RNA stability are i) base pairing, ii) conformational entropy, iii) the effect of cations, and iv) hydration. Here, it was also observed that GFP-mimic fluorogen, HBC showing strong stacking interactions to the core of 3WJ RNA to form a stable complex. This can make way for experimental validation to use 3WJ pRNA to form RNA-fluorophore complexes, which have emerged as versatile tools for cellular imaging, surpassing traditional fluorescent proteins in brightness and application flexibility. Further study of 3WJ pRNA and its binding capabilities can pave way for the development of novel therapeutic, diagnostic and imaging tools based on ribo nucleic acids .

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Ribo nucleic Acids (RNA) are a vital component in molecular biology and its relevance has been growing due to its role in gene-regulation networks and catalysing cellular reactions [14]. Messenger RNA (mRNA) acts as the intermediary carrier of genetic information from DNA, the genetic material of cells. Transfer RNA (tRNA) facilitates the translation of this information into protein sequences, which form major catalyst and structural components of the cell. The translation process is facilitated by the ribosome, (complex of RNA and proteins), which ensures the precise pairing of mRNA with cognate aminoacylated tRNAs. Further, the discovery of ribozymes in the 1980s highlighted RNA's ability to catalyze specific chemical reactions, akin to protein enzymes. This discovery revolutionized biology, demonstrating RNA's dual capacity to store genetic information and catalyze chemical reactions. This provided support for the RNA World hypothesis, suggesting that ribose nucleic acids like molecules may have been the earliest forms of life on Earth [29]. According to this hypothesis, primitive cells were made of self replicating RNA molecules. Only later in evolutionary time, DNA having greater stability and durability than RNA became the predominant the genetic material and protein enzymes replaced RNA-based ribozymes as biocatalysts because their higher abundance and diversity of monomers.

With the advent of RNA nanotechnology, RNA have gained significant attention. They have tremendous applications in fields such as biomedicine. These nanoparticles, being multivalent offer targeted drug delivery capabilities. They also offer control over their size and shape and are biologically compatible, thereby mitigating toxicity concerns, commonly associated with polymer-based nanoparticles. Additionally, they exhibit low immunogenicity compared to protein therapies, making them suitable for treatment of chronic diseases. Despite these advantages, challenges remain in ensuring the stability of RNA molecules and nanoparticles, particularly at low concentrations, which is crucial for their biomedical applications. The three-way junction prohead RNA (3WJ-pRNA) from ϕ29 bacteriophage has emerged as a promising candidate for such applications due to its self-assembly properties and thermodynamic stability. Novel nanomaterials have been developed based on pRNA three way junction in which cell receptor-binding ligands, aptamers, siRNAs, and ribozymes are fused to individual arms of this junction RNA motif ([20]). These

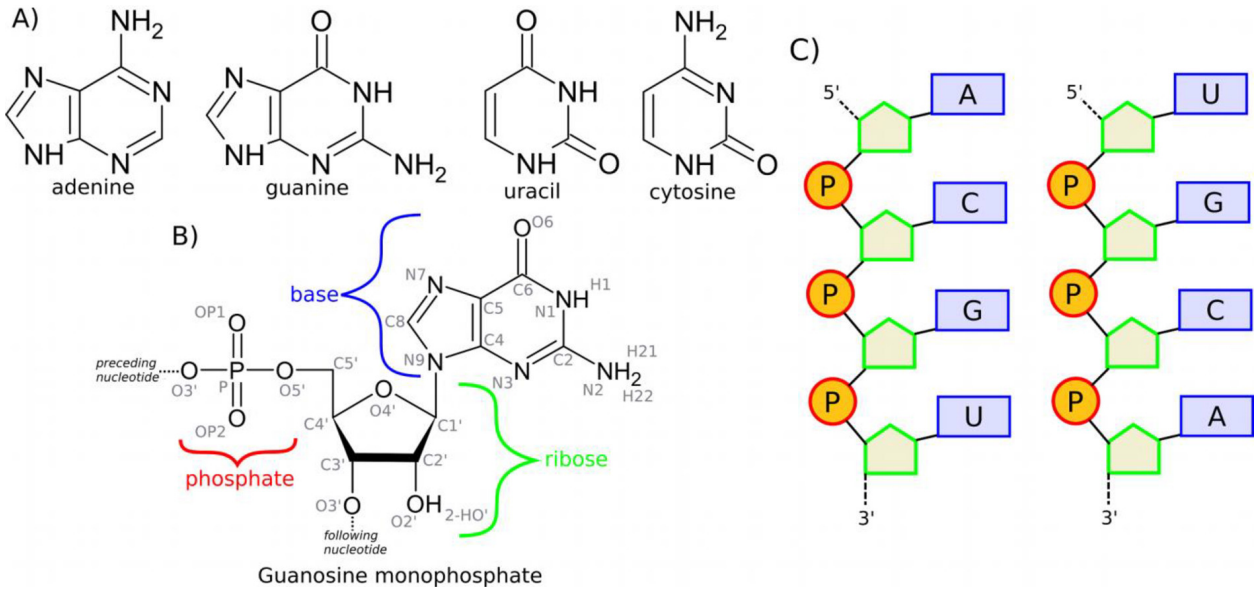


Figure 1.1: Composition of RNA - Nucleobases(A,G,U,C), Sugar and Phosphate groups [29]

motifs cab fold into their authentic structures. Their are multifunctionalised to target specific cancer cell binding, cell entry, gene silencing, catalytic function.

The prohead RNA (pRNA) plays a pivotal role in the ϕ29-like bacteriophage DNA packaging motor, one of the most powerful molecular motors. It is situated between the connector and ATPase motor proteins and forma a critical interface within the motor complex which drives the encapsidation of viral DNA. Figure 1.1 illustrates the spatial arrangement of these components. The ATPase, powered by ATP hydrolysis, converts chemical energy into mechanical energy necessary for DNA packaging. Although the specific function of pRNA in the motor remains unclear, current models propose that it provides structural support, communicates between motor components, and coordination of ATPase activity and may also be involved in stress absorption. The pRNA's unique role in the 29 motor reflects the adaptability of nucleic acids. The presence pRNA, for packaging activity which is unique to this family of bacteriophages while being absent in all other descendants of ϕ29. This can be interpreted as an evidence of transition from an RNA-only world to an RNA-protein world in descendants.

ϕ29 pRNA consists of two distinct domains connected by a flexible linker. Domain I is found to be critical for prohead binding,where helices A-F contribute to the structural integrity and function of pRNA. Domain II,on the other hand is hypothesied to play a role in phage morphogenesis and is thus, conserved among ϕ29-like phages. shown in fig 2.. The secondary structure of pRNA is remarkably conserved, despite low sequence similarity among 29-like phages. . Several regions within Domain I are functionally significant as they mediate interactions essential for packaging activity in the cell. The three-way junction (3WJ) motif, pivotal for pRNA function, exhibits varying conformations during the packaging process, indicating its dynamic role in motor function.(shown where?) The assembly of pRNA on the prohead is facilitated by magnesium ions.

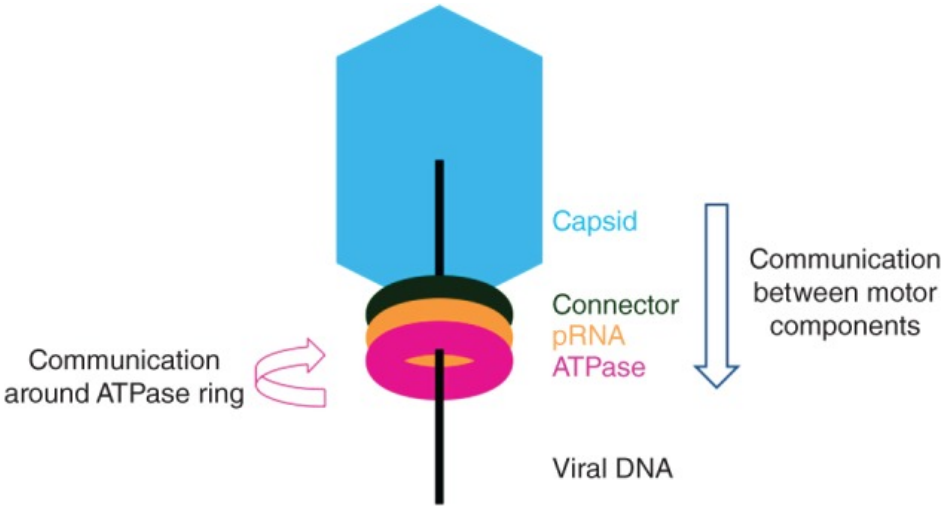


Figure 1.2: Assembly of ATPase, pRNA, Connector protein in viral capsid.

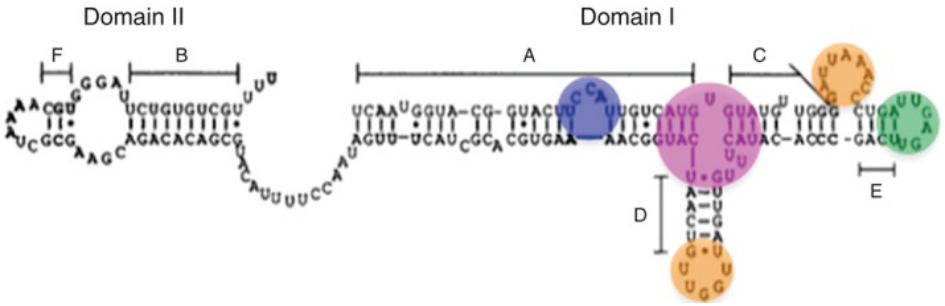


Figure 1.3: Secondary Structure of pRNA. Helices CE and D mediate self-association of pRNA molecules in vivo and in vitro. C and D correspond respectively to H2 and H3 of 4kz2 crystal structure.

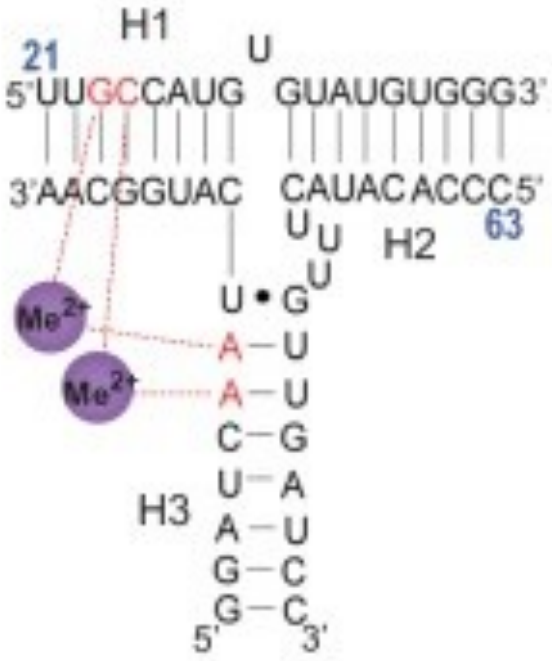


Figure 1.4: Secondary structure of 3WJ-pRNA obtained from ϕ29 bacteriophage

1.1 Structure of Three way junction pRNA

The core region of RNA folds with low Gibbs Free energy (G) and guides the folding of the entire RNA molecule. The crystal structure of bacteriophage ϕ29 pRNA-3WJ motif was obatined from RCSB databank with PDB ID: 4kz2 ([38]). The structure of the core region is supported by "coaxial stacking" of three individual RNA helices [38]. Three short ssRNA chains; chain A, B, and C ($\leq$ 20 mer) are assembled with high affinity of association in an entropy driven one step formation process [4]. Two divalent metal ions that coordinate 4 nucleotides of the RNA fragments at nucleotides G3 of Chain A and A8 of chain C, and C4 of Chain A and A7 of chain B. The 5' end of Chain A and 3' end of Chain C join to form helix H1. 5' end of Chain B and 3' end of Chain A form helix H2. The core is comprised of triple U bulge of Chain B and H3 is formed by the remaining nucleotides at 5' end of Chain B and 3' end of Chain A. These helices form distorted A-helix geometry due to chelating metal ions. These nanoparticles based on 3WJ-pRNA are shown to be resistant to high-concentration urea denaturation In addition to extraordinary thermal stability , it also exhibits mechanical anisotropy (shows differential mechanical stability in different pulling directions) which can be controlled by varying bound Mg^{+2} ions [30]. Mg^{+2} ions can reduce the influence of the negative charges in the RNA backbone by reducing repulsive electrostatic forces. Crystal structures and spectroscopic studies provide insights into the structure of pRNA. Being such a pluripotent chemical species, the functions of three way junction pRNA are linked to its structure and dynamics. As the three-way junction (3WJ) motif shows varying conformations during its self assembly, it is imperative to study its dynamics. Despite extensive research and knowledge about three-way junctions pRNA, there is still a lack of understanding

regarding the specific mechanism responsible for their high thermodynamic stability. Through computer simulations, structural dynamics of the junction pRNA motif can be studied with great temporal and spatial resolution.

1.2 Fluorescence light up aptamer complexes

Cellular imaging techniques are essential for visualization of components of molecular and cellular organisation to understand biological phenomenon. The use of Green fluorescent protein (GFP) for cellular imaging in live cells was transformative across various research domains. Over the past decade, new techniques involving the use of "genetically encoded biosensors" are being developed. These include RNA aptamers capable of binding with specific "fluorogenic ligands" and show fluorescence upon binding. Examples include malachite green (MG) aptamer, DFHBI (GFP-derivative) binding to Spinach aptamer and biotin binding to Mango aptamer [22]. These find application in visualizing localization and transport live in cells. These versatile complexes, exceeding the brightness of green fluorescent protein and its variants, have diverse structures. Structural analysis reveals that these RNA have uncorrelated architectures [32]. Most common mechanism suggested for these complexes is by restraining excited states of the ligand into a planar conformation. Thus, the fluorophore binding sites of these RNA usually have a G-quartet architecture, which gives an extensive molecular framework to arrest small molecule conformational change. But, G-quartet based RNA aptamers are easily degradable by nucleases in vivo and are not naturally excreted from the body. In this study we look at the three way junction based RNA-fluorophore complexes, which can potentially function as these "genetically encoded imaging platforms". The junction RNA being resistance to nucleases and capable of renal excretion, offer advantages over existing G-quadruplex-based aptamers and fluorescent proteins. The nascent GFP comprises of sequence $\text{Ser}^{65} - Tyr^{66}\text{-Gly}^{67}$, which constitutes the fluorophore region. It undergoes autocatalytic intramolecular cyclization to give 4-hydroxybenzlidene imidazolinone (HBI) [17]. HBI, when chemically synthesised does not show fluorescence and resembles denatured GFP. However, upon re−folding, it can recover fluorescence. The fluorophore interacts with residues of GFP, which restricts the intramolecular motions of ligand upon excitation. As this interaction closes one of the major deactivation pathways of the fluorophore in excited state, it shows fluorescence to dissipate energy in its excited state. Here, fluorogenic ligands based on HBI were screened. The following ligands were used for further studying complexation with the three-way junction pRNA:

DFHBI: HBI derivate, 3,5-DiFluoro-4-HydroxyBenzylidene Imadazolinone (abbreviated as DFHBI) remains exclusively in its phenolate form because of decrease in pk_a. It has been crystallised in complex with Spinach aptamer in cis isomer, which possess a G quaruplex for DFHBI to bind [13].

DFHBI-1T:
(3,5-difluoro-4-hydroxybenzylidene)-2-methyl-3-(2,2,2-trifluoroethyl)-3,5-dihydro-4H-imidazol-4-one (abbreviated as DFHBI-1T) is another membrane-permeable fluorogenic probe. It has been crytsallised with Brocolli aptamer in cis isomer form [16].

HBC: (4-((2-hydroxyethyl)(methyl)amino)-benzylidene)-cyanophenylacetonitrile (abbreviated as HBC) is a synthetic dye extracted from GFP-like fluorophore. It has electron acceptor and donor sites. Because of these

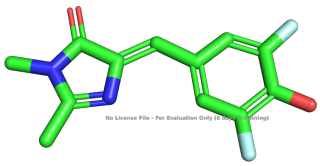


Figure 1.5: Structure of DFHBI [(5Z)-5-(3,5-difluoro-4-hydroxybenzylidene)-2,3-dimethyl-3,5-dihydro-4H-imidazol-4-one] obtained from crystal structure (PDB id: 4ts2)

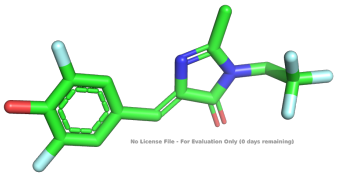


Figure 1.6: Structure of DFHBI-1T [(5Z)-5-(3,5-difluoro-4-hydroxybenzylidene)-2-methyl-3-(2,2,2-trifluoroethyl)-3,5-dihydro-4H-imidazol-4-one] obtained from crystal structure (PDB id: 7l0z)

sites being structurally rigid, it is easier to lock the ligand (i.e. constraint intramolecular motion). It has been crystallised with its cognate RNA aptamer called "Pepper" [10]. It shows a emission maximum at 530 nm.

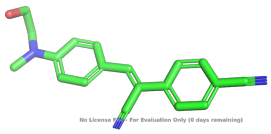


Figure 1.7: Structure of HBC [(4-((2-hydroxyethyl)(methyl)amino)-benzylidene)-cyanophenylacetonitrile] obtained from crystal structure (PDB id: 7eoh)

Chapter 2

Methods

Before delving into the work, I would like to present the basics of techniques and analysis parameters used in this study. There are several computational approaches to study dynamics and ligand binding of system of interest.

2.1 Molecular Docking

Molecular docking is a method to predict non-covalent binding between macromolecules or, of a macromolecule receptor and small molecule ligands. It attempts to predict the best conformations in which ligand is bound to receptor and gives binding scores for comparative analysis. Here, docking has been used to try to predict the bound conformation of FAD and its analogues to FAD-binding proteins, in the case of unavailability of the experimentally determined holo crystal structures. The broader objective of predicting small molecule (FAD and FAD-analogues; FGD, FUD, and FCD) binding to proteins is to use this method to screen proteins across the whole proteome of an organism (in this case *E.coli*) which bind better to FAD analogues as compared to their cognate ligands (proteins which are not experimentally shown to bind to FAD,$\approx$ 4400 in number) or FAD (for FAD binding proteins , 70 in number). For studies here, search-based docking using AutoDock Vina (version 1.1.2) was conducted [34]. Vina uses a iterated local search global optimizer for local minimisation of structures. The conformation-dependent component of AutoDock Vina is designed to operate with a general functional form given by:

$$V(q) = \sum_{i,j} V_{ij}(q) \tag{2.1.1}$$

where the summation accounts for all pairs of atoms capable of relative motion, excluding 1–4 interactions[34]. Each atom i is assigned a type t_i, and set of interaction functions V_{ij} (based on the interatomic distance r_{ij}) are calculated. The binding affinity (also called docking score) is calculated from the inter-molecular interactions of the best ranked conformation.

Table 2.1: Simulations performed

	Temperature	Ion concentration	Solvent
SIM1	300K	5 mM Mg^{+2}	TIP3P water
SIM2	400K	5 mM Mg^{+2}	TIP3P water
SIM3	300K	0 mM Mg^{+2}	TIP3P water
SIM4	400K	0 mM Mg^{+2}	TIP3P water
SIM5	300K	5mM Mg^{+2}	UREA (4M)

2.2 Molecular Dynamics

A Molecular Dynamics simulation (MD) refers to a computer simulation of atoms or molecules treated as point particles, with their interactions among themselves modeled according to various types of potentials, parametrised on experimental data and /or ab-initio calculations. Thus, the trajectory of atoms being simulated is obtained by integrating Newton's equations of motion for the system of particles in the box. For an n particle system, the force equation would be as follows:

$$F_i = -\nabla_i V(\mathbf{X}_1, \ldots, \mathbf{X}_n) = m_i \frac{d^2\mathbf{X}_i}{dt^2} = m_i \mathbf{a}_i(t) \tag{2.1}$$

where $\mathbf{X}_i$ represents the position of the ith atom and V the potential due to the 'bonded' and 'non-bonded' interactions between atoms parameterized as in the force field used. MD simulations were performed using GROMACS 2022.4 ([2]). hyperref For this study, AMBER-ff99 force field with parmbsc0 and χOL [11], [37] was used. For monovalent ions (Na^+ and Cl^-) , parameters proposed by Joung and Cheatham III [19] were used and parametrisation by Allnér et al. was used for magnesium ions[3]. Simulations were performed 3WJ RNA, the structure obtained from the RCSB Databank (PDB ID: 4kz2) with atomic resolution of 3.05 Å. The system was solvated in TIP3P water within a cubic simulation box of dimensions of $105Å^*105Å^*105Å$ and sodium ions (Na) and chloride ions (Cl) were added to maintain a physiological ion concentration of 156 mM. 5 mM of MgCl(2) was added. The system was initially equilibrated using the steepest descent algorithm and subsequently, an NVT equilibration was performed at 300 K for 5 ns, using a V-rescale thermostat [8]. The system was further equilibrated at a constant pressure of 1 atm and a constant temperature of 300 K for 10 ns using a Berendsen barostat. Nonbonded interactions were calculated using the Verlet cutoff scheme, and electrostatics using particle-mesh Ewald [1]. The cutoff was initially set to 10 Å. An unrestrained molecular dynamics simulation was run for 200 ns, with a timestep of 2 fs at 300 K and 1 atm pressure. The V-rescale thermostat and Parrinello-Rahman barostat [23]were used to maintain the system at constant temperature and pressure during the simulation.

MD simulations have inherent limitations because results depend upon the accuracy of the forcefield.

2.3 Analysis parameters

The thermodynamic stability of RNA is dependent on the following factors[31] :

2.3.1 Base pairs interactions

Nucleobases contribute by hydrogen bonds between base pairs and stacking interactions between neighbouring base pairs. The complementarity relationships between base pairs (donor-acceptor pairs) lead to high selectivity and stability. The aromatic rings of nucleobases, being electron-rich, can interact with other nucelobases via stacking of aromatic rings in them. To analyse these interactions in trajectories of 3wj-pRNA, root mean square deviation, root mean square fluctuations and $\epsilon RMSD$ values were calculated using Barnaba [6]. $\epsilon RMSD$ is a metric of change in stacking interactions and relative positions/orientations of base pairs in three-dimensional structures of nucleic acids. It is helpful to distinguish between structures whose RMSD is low but they show a different internal network of basepair interactions. The G vectors ($\vec{G}$) are computed for each nucleobase using the rescaled distance vectors ($\tilde{\vec{r}}$), incorporating trigonometric functions and a heaviside step function. ϵRMSD is determined by averaging the squared differences between $\vec{G}$ vectors of corresponding base pairs from two structures (α and β), and taking the square root of the average. The formula is given by: [7]

$$\epsilon RMSD = \frac{1}{N} \sum |\vec{G}(\tilde{\vec{r}}_{\alpha jk}) - \vec{G}(\tilde{\vec{r}}_{\beta jk})|^2$$

An $\epsilon RMSD$ value of greater than 0.7 can be considered structurally different. Multiple peaks in $\epsilon RMSD$ distribution would mean more than one stable structures even if the parameter may not give accurate results when applied on intermediate states in an MD trajectory.

2.3.2 Effect of Cations

Mechanical and thermal stability of 3wj-pRNA has shown to depend on the bound magnesium ions on the RNA.[30] The visiting frequency of Mg ions to phosphate group of each residue is calculated. It tells us about the sites to which cations(in this case Mg) binds to as well as non-bonded ionic interactions. This distance cutoff for interaction is determined by the Radial Distribution Function (RDF) of Mg2+ ions around phosphate groups and inner, and outer sphere interactions of Mg with RNA. The visiting frequency (denoted as ν)is mathematically defined as:

$$\nu = \int_0^t \gamma d\tau$$

Here, $\gamma = 1$ when an Mg^{2+} ion is within the cut-off distance 0.65 Åof a specific phosphate group, and $\gamma = 0$ otherwise. t represents the number of frames in the trajectory.

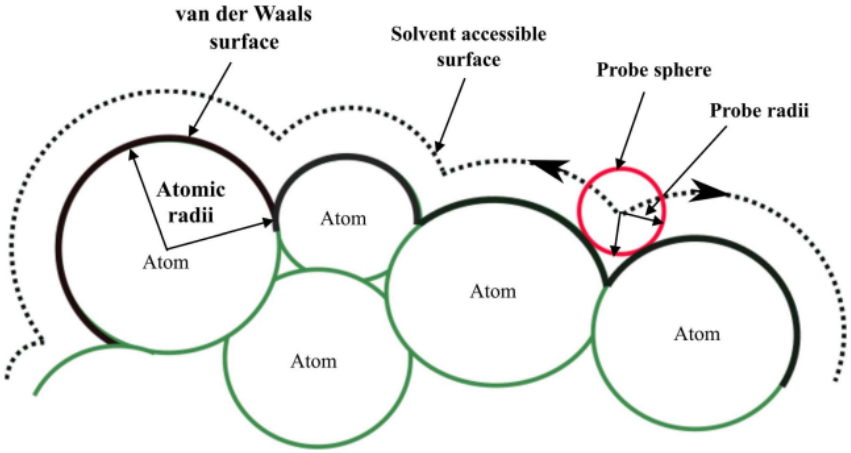


Figure 2.1: Solvent Accessible Surface Area or SASA is defined as the van der Waals surface area created by probe spheres (dummy water molecules) by rolling over the macromolecule surface. The formula was first given by Lee and Richards for protein surfaces. [27]

2.3.3 Hydration

Solvent plays an important effect on adoption of particular tertiary structure of RNA and can thus, influence its functional variety. Solvent Accessible Surface Area (abbreviated as SASA) is a 1D representation of the same [28], showing correlations with hydrophobicity, molecular weight and transfer free energies.

Chapter 3

Results and Discussion: Three-way junction pRNA

The analysis of SIM1, SIM2, SIM3, and SIM4 included an investigation into ion dependency and solvent interaction. The following observations were drawn by analysing simulation performed in this study [table 2.1].

To quantify the residue wise contribution to flexibility of RNA, conformational entropy for each residue was calculated. The values presented are quasi-harmonic entropy for 54 residues calculated using Schlitter's method [26]. For 3wj RNA in 5mM magnesium ions in solution at 300K, the entropy for terminal residues (0-2,17-20,37-39 and 53) are high as is expected (figure 3.1). It is interesting to note that residues near the core region (7-8,2830 form part of triple U bulge) have high conformational entropy indicating flexibility of the core region.

In 3.2, we observe higher $\epsilon RMSD$ values in simulations at higher temperature. An $\epsilon RMSD$ value of greater than 0.7 represents change in conformation of the bases in RNA. From figure 3.2, 3wj RNA is stable at higher temperature (400K) in the presence of Mg^{+2} ions. In the absence of Mg^{+2} ions, high $\epsilon RMSD$ value indicates that the continues to change conformations and is not stable after 200 ns simulation. The RNA shows stable structure in SIM3 and in presence of 4M Urea.

High values of root-mean-squared-deviations and fluctuations are evident in simulations without magnesium. In SIM4, RMSF increased up to 3 nm as shown in figure 3.3. The increased fluctuations are likely to be an effect of absence of Magnesium. Similar effects have also been reported by [9] where they show breaking of H-bonds in the core region in higher temperature conditions without Magnesium. The $\epsilon RMSD$ value showed significant changes in SIM4, indicating conformational alterations due to higher temperature, particularly pronounced in the absence of Magnesium.

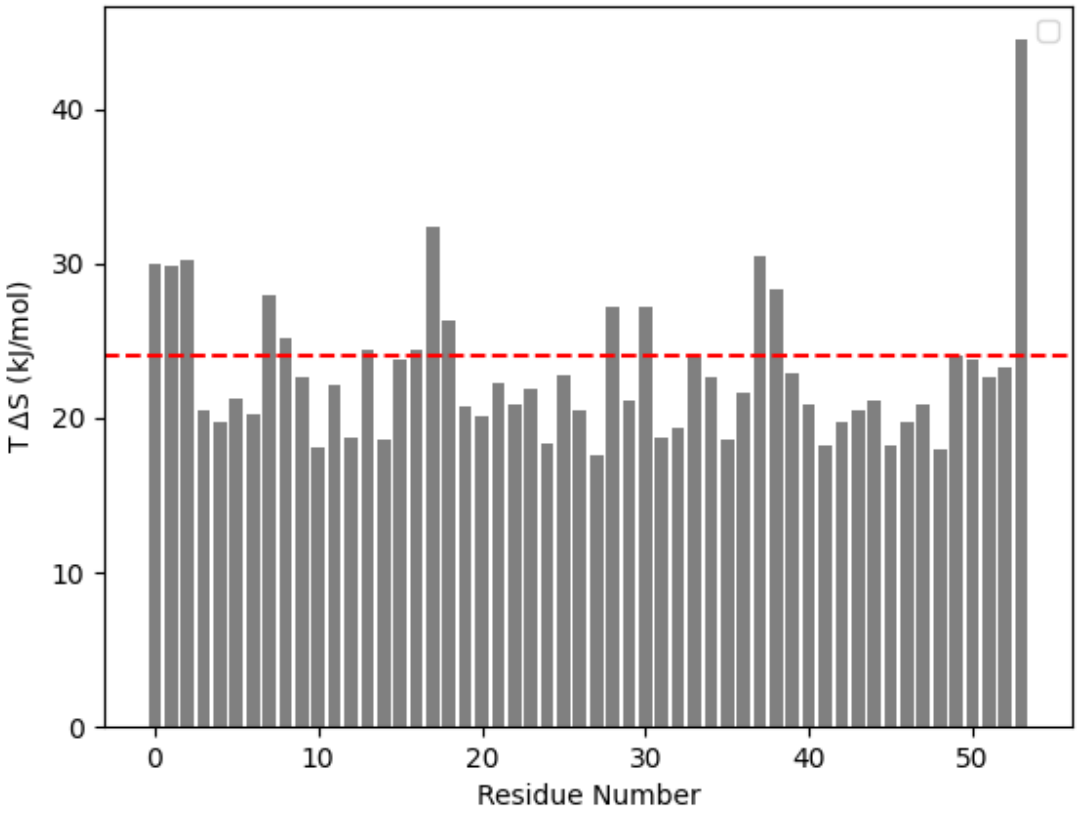


Figure 3.1: Quasiharmonic entropy for each residue of 3wj-pRNA in SIM1(in presence of 5mM $MgCl_2$ at 300K)

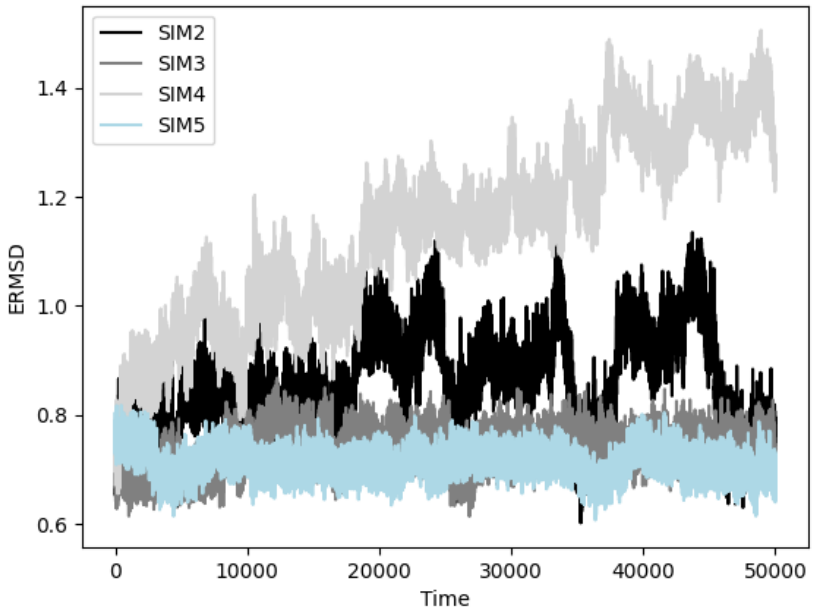


Figure 3.2: eRMSD values for SIM2, SIM3, SIM4, SIM5 with respect to structure obtained in SIM1 for the last 50 ns simulation time.

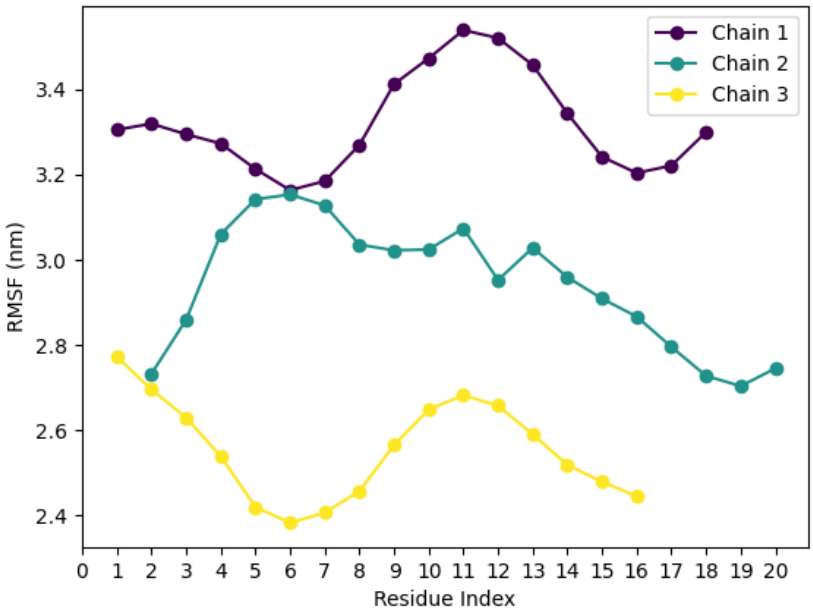


Figure 3.3: Root Mean Square Fluctuations of 3WJ RNA in simulation in absence of Magnesium ions at 400K

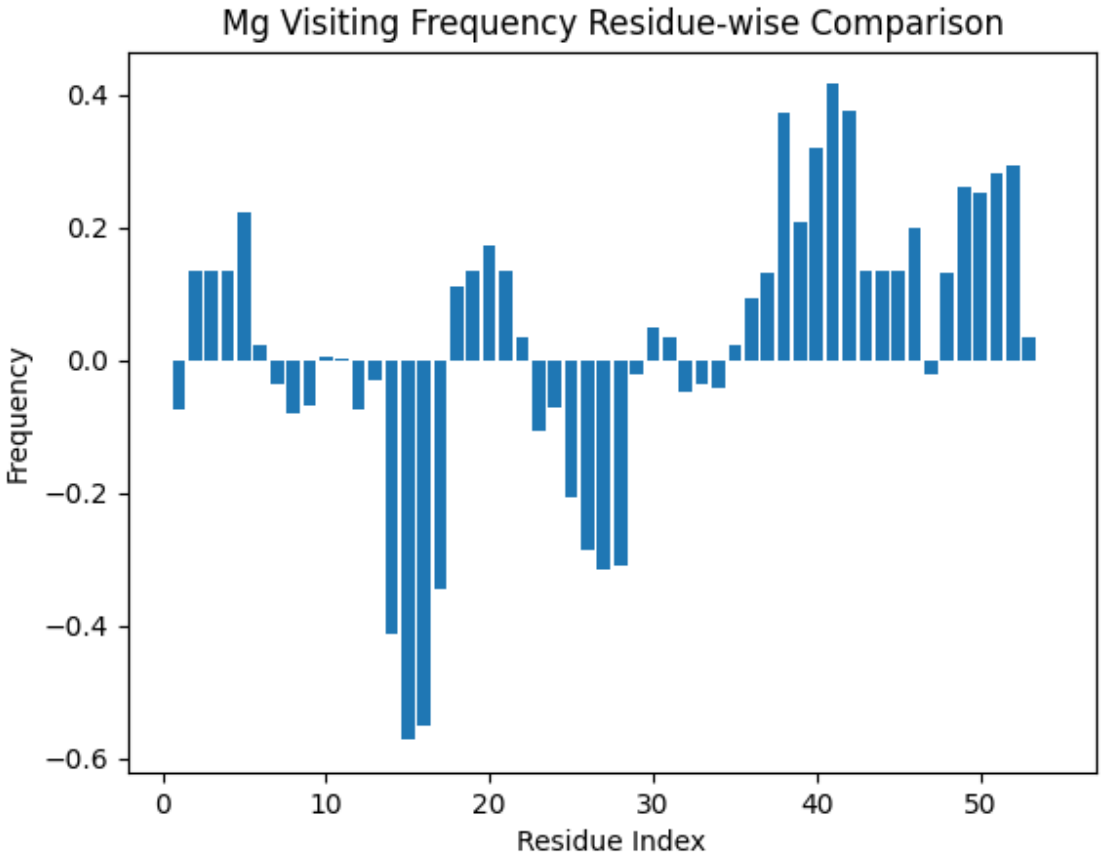


Figure 3.4: Change in Magnesium Visiting Frequency for each residue of 3WJ RNA at 400 K (SIM2) as compared to that at room temperature, 300K (SIM1).

Changes in Mg^{+2} visiting frequency : We observed an increase in visiting frequency ν for SIM1 when cutoff is set to 0.35 $\mathring{A}$ vs that at 0.65 $\mathring{A}$. This leads us to conclude that most Mg^{+2} ions interact with phosphate backbone of the RNA via their second hydration shell. At temperature 400K, visiting frequency ν (at a cutoff distance at 6.5 $\mathring{A}$ from RNA phosphate group) significantly increases for chain C. This correlates to lower fluctuations of the chain as shown in figure 3.3. As Helix 1 of the 3wj RNA is made from 5' end of chain A (residues 0-8) and 3' end of chain C (residues 45-53), we can conclude that the visiting frequency for H1 increases more than other helices. The Mg^{+2} visiting frequency, ν reduces for residues at the core of three way junction (triple U bulge: U10,U11,U12 in Chain B) which correlates with increased fluctuation in those residues.

Change in visiting frequency ν in presence of Urea in solution: As shown in figure 3.5, we do not observe a stark change in ν in residues of chain B while at the core of 3WJ (U10,U11) the mean squared fluctuation [shown in figure 3.6] values show a rise as compared to that in SIM1. We also observe an increase in the Magnesium visiting frequency for H1 while that remains constant for other helices. From figure 3.7, we can infer that the accessible surface area of RNA increases for chains B and C while it remains same for chain A.

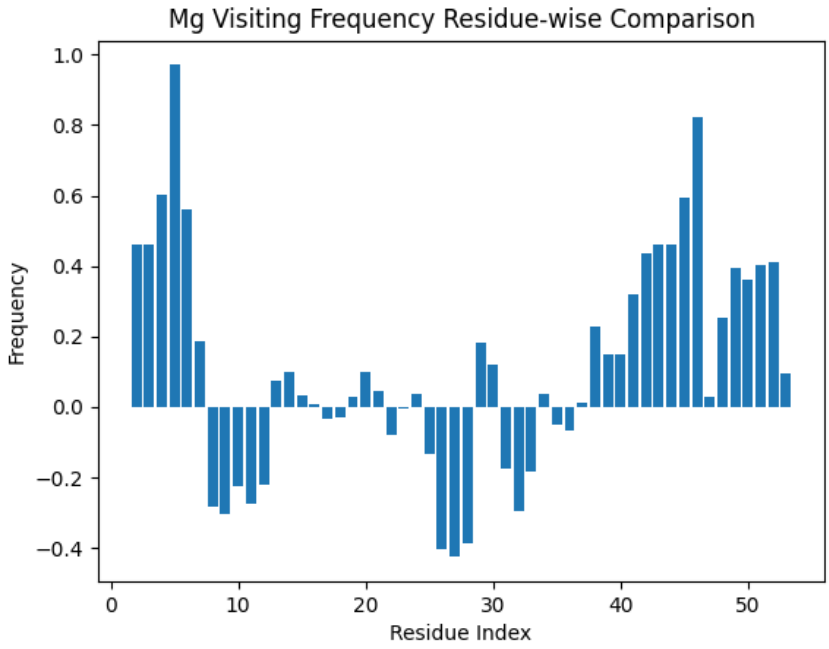


Figure 3.5: Change in Mg^{+2} visiting frequency in presence of Urea (SIM5) in comparison with SIM1

Figure 3.8 also shows that the buried surface area of 3WJ RNA increases in presence of urea in solution. This highlights that urea reduces inter-residue interactions in the RNA, possible by forming stacking interactions with the nucleobases and thereby, exposing them to solvent.

Changes in Solvent Accessible Surface Area (SASA) in other simulations: SASA changes between SIM1 and SIM2 are smaller and the buried surface area also remains similar. We do not observe much changes in SASA owing to temperature differences in the presence of Magnesium. At 300K, SASA remains comparable in simulation without Magnesium and with Magnesium. The Buried Surface Area also converges to 400 square $\mathring{A}$ for both these cases. (SIM1 and SIM3). In absence of Magnesium ions, SASA differs significantly between simulations at 300K and that at 400K. While it increases for chain C , it increases for chain A. The buried surface area increases at higher temperature and it also indicates that RNA residues are more exposed in the absence of Mg^{+2} at higher temperature.

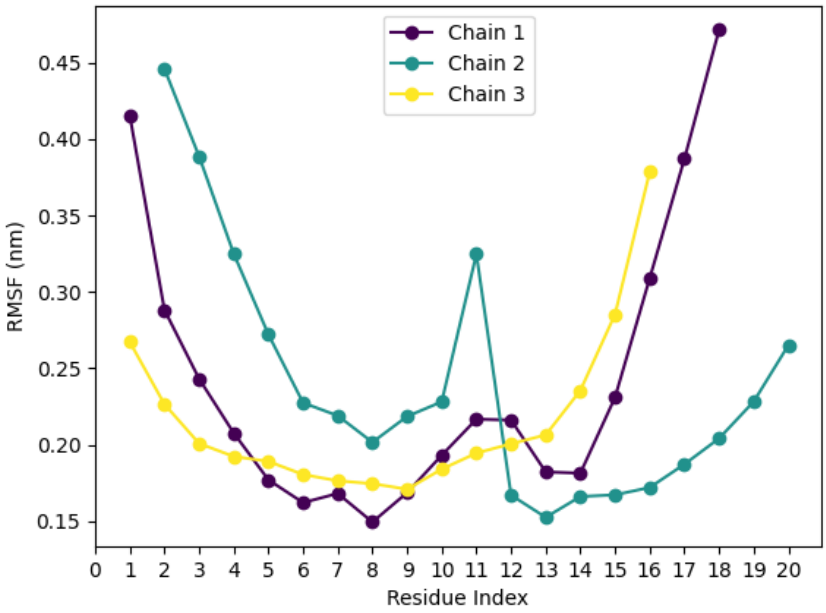


Figure 3.6: Root mean square fluctuations of RNA in SIM5 (in presence of 4M urea)

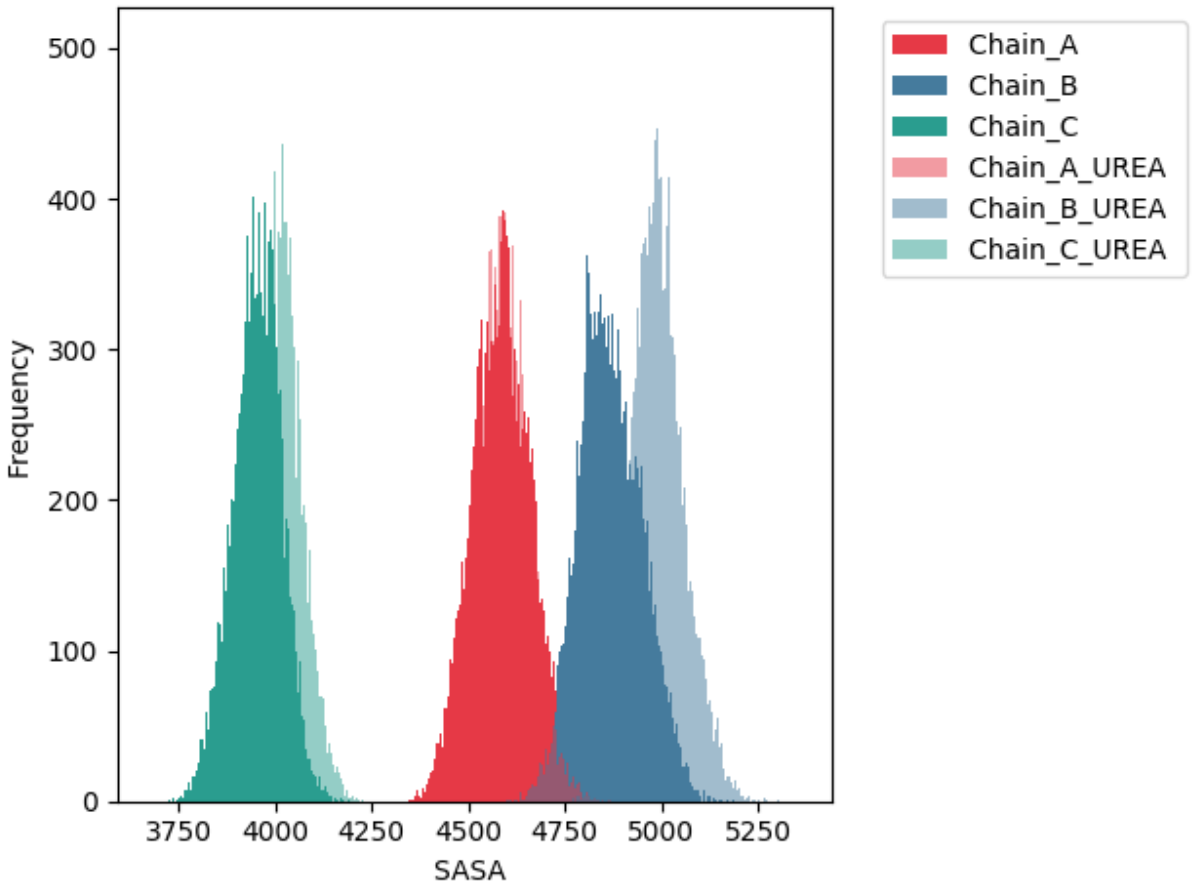


Figure 3.7: Solvent Accessible Surface Area changes in SIM1 and SIM5. Light color corresponds to SASA in SIM5 (in presence of 4M urea) and the darker to SIM1 (pure water solvent)

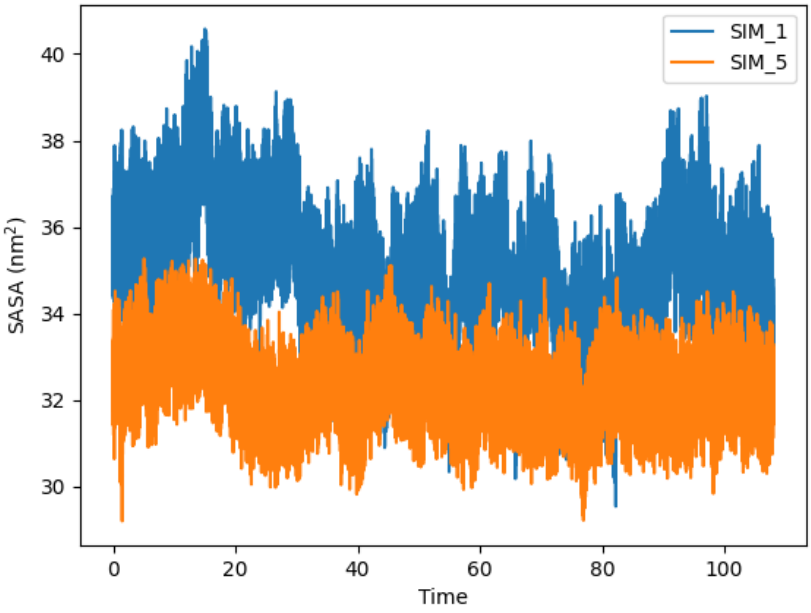


Figure 3.8: Buried Surface Area of 3wj RNA in presence of UREA as compared to SIM1. Buried Surface Area is calculated by removing the actual solvent accessible area of RNA from the sum of that of all three chains.

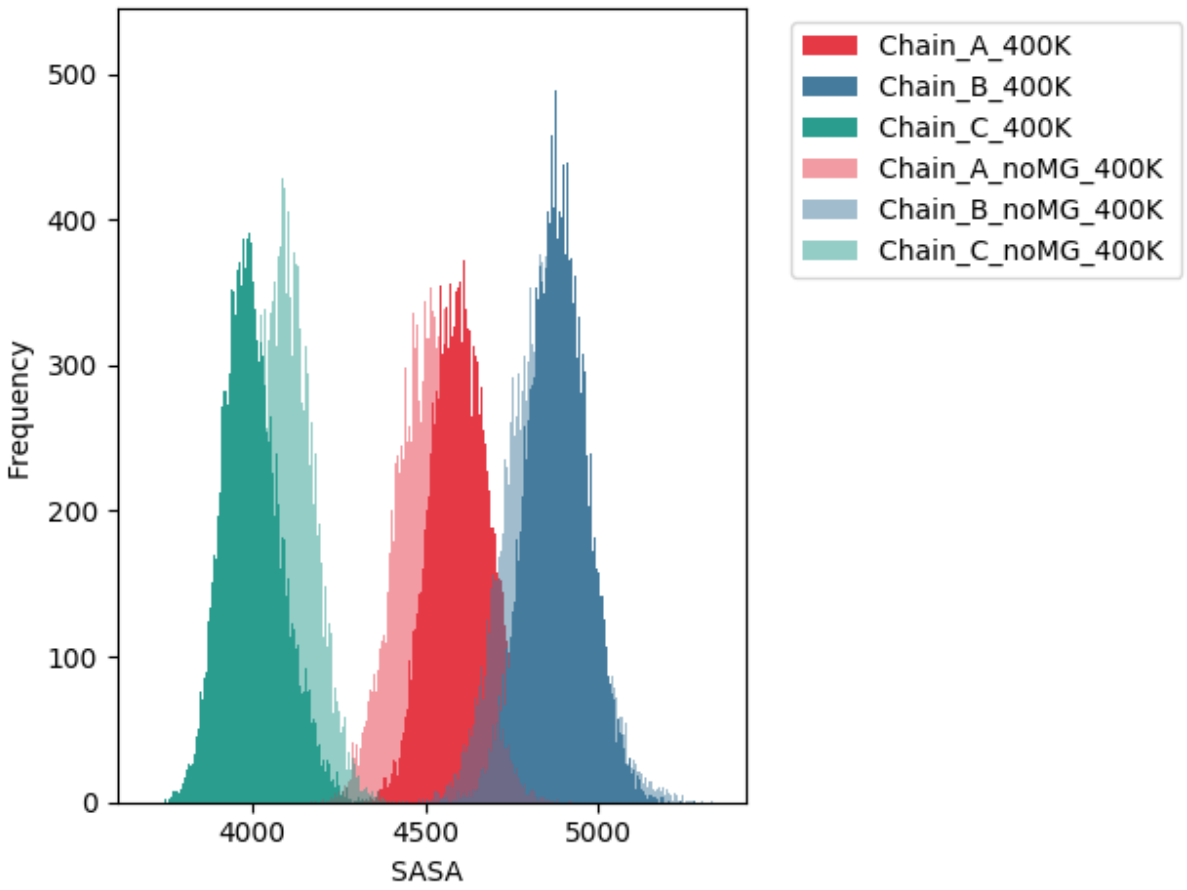


Figure 3.9: Solvent Accessible Surface Area histogram for Simulations 2 and 4. The mean SASA for chain C increases while that for chain A reduces. As both the chains have Mg^{+2} binding sites, the source of this differential behaviour of these chains is intriguing.

Chapter 4

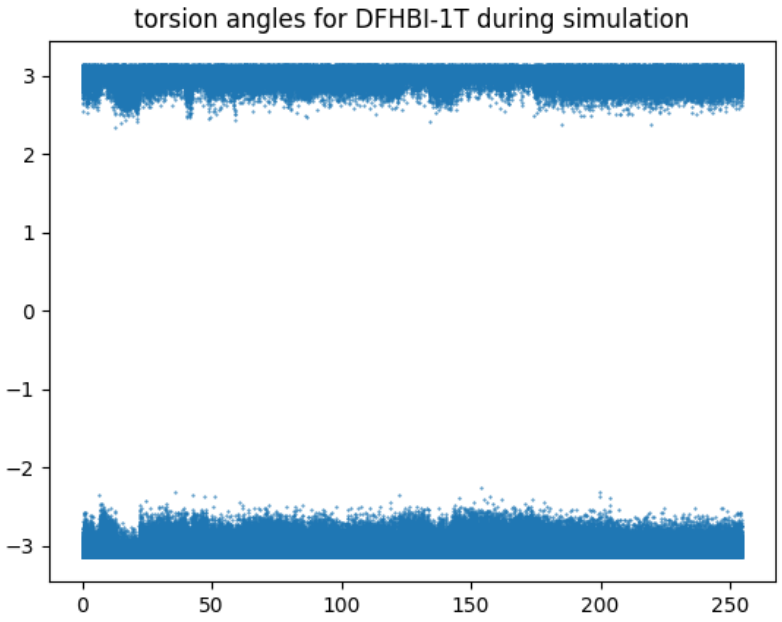
Results and Discussion : 3WJ RNA– Ligand systems

In this section, we are going to investigate the interaction of 3wj- pRNA with GFP-mimic fluorogenic ligands.

4.1 System Preparation

For this study, three Green Fluorescent Protein (GFP) mimics were used, namely DFHBI, DFHBI-1T, HBC . MD simulations were performed on a system comprising 3WJ RNA, obtained from the RCSB Databank (PDB ID: 4kz2), complexed with ligands. The ligands were obtained from a crystal structure in complex with its cognate RNA aptamers; DFHBI(cis-isomer) and DFHBI-1T in complex with Spinach aptamer (PDB ID: 4st2 [[36]]and 7l0z [[18]] respectively) and HBC in complex with pepper RNA (PDB ID: 7eoh [[25]]). The simulations were initialised from the best binding pose of ligands when restricted to bind to the core of 3WJ RNA. The binding poses were obtained by docking using vina 1.1.2.[34]. This system was solvated in TIP3P water within a cubic simulation box of dimensions of $105\mathring{A} * 105\mathring{A} * 105\mathring{A}$, and sodium (Na) and chloride ions (Cl) were added to maintain a physiological ion concentration of 156 mM. The ligand molecules were parameterized using the AnteChamber module of AmberTools [35]. The system was initially equilibrated using the steepest descent algorithm, followed by NVT equilibration at 300 K for 5 ns, using a V-rescale thermostat . The system was further equilibrated at a constant pressure of 1 atm and a constant temperature of 300 K for 10 ns using a Berendsen barostat.An unrestrained molecular dynamics simulation was run for 100 ns, with a timestep of 2 fs at 300 K and 1 atm pressure. The V-rescale thermostat [8] and Parrinello-Rahman barostat [23] were used to maintain the system at a constant temperature and pressure during the simulation.

Figure 4.1: Torsion angles for DFHBI-1T. Values near to π represent cis isomer and $-\pi$ represents trans isomer.



4.2 3WJ RNA in complex with Ligands

DFHBI unbinds from the center of the RNA and subsequently binds to one of the arms of the RNA . Similar behavior is also observed for DFHBI-1T. Only HBC remains bound to the core of 3WJ and exhibits maximum interaction energy with RNA (as shown in the figure).

Further investigation was conducted into the conformational changes of DFHBI and DFHBI-1T. Torsion angles was calculated for cis and trans conformations of DFHBI and DFHBI-1t. Values near to π radians correspond to cis and those near $-\pi$ to trans conformations of DFHBI and DFHBI-1T This reveals that despite the ligands binding to 3WJ RNA, the interaction does not restrict the isomerization of the ligands.(shown in figure 4.1). It means that fluorescence will not be exhibited by this binding.

4.3 Interactions of HBC with 3WJ RNA

4.3.1 Stacking of HBC to the core of 3WJ RNA

Simulation trajectory of 250 ns was used to analyse the binding of HBC to 3WJ RNA. The analysis were performed using Plumed software (version 2.8.2) [33][5]. The following parameters were used to check stacking of rings of HBC to residue U10 and U12 at the core of 3WJ RNA:

Distances d_1 and d_2 :The stacked configuration was defined as the distance between the center of masses of

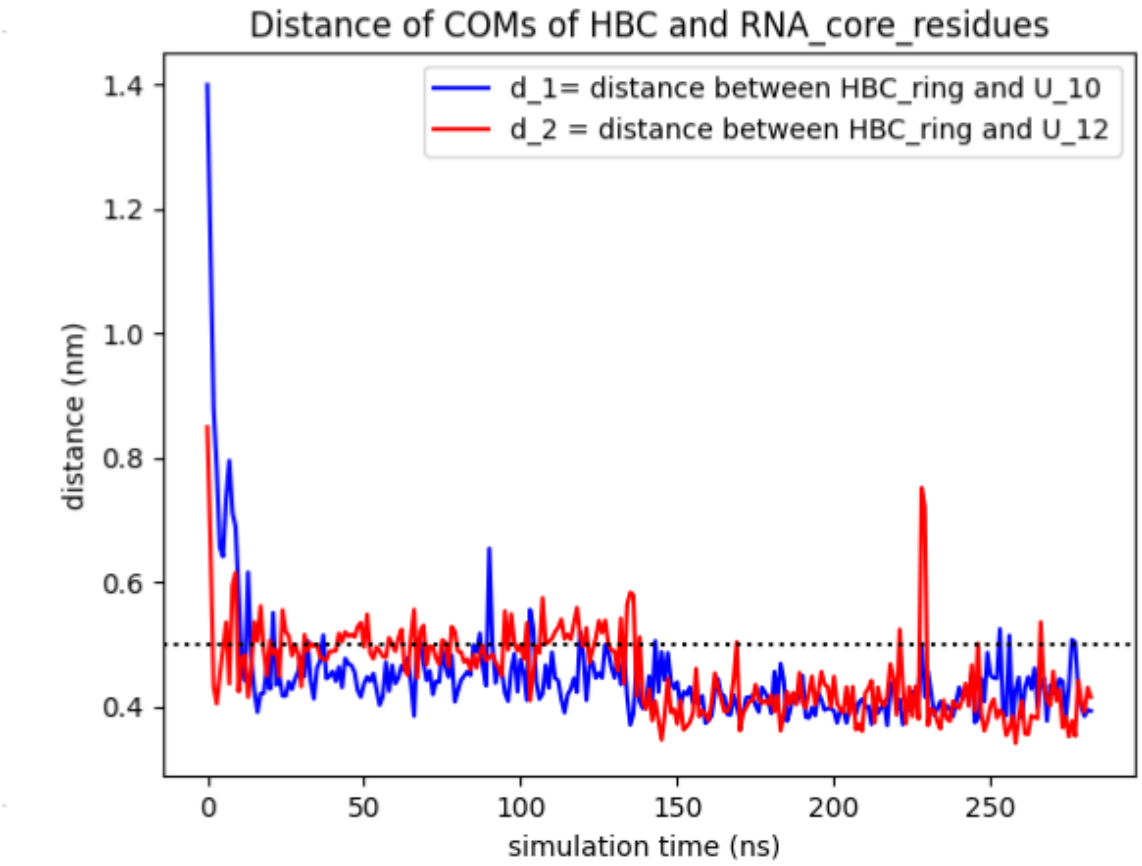


Figure 4.2: Distance between center of masses of residues U10 and U12 to center of mass of ligand.

the rings of the residue and HBC being less than 5 Å. Specifically, d_1 represented the distance between the center of masses of residue U10 and the phenyl ring of HBC, while d_2 denoted the distance between the center of masses of residue U12 and the phenyl ring of HBC. The simulation data indicated that within the first 30 ns, the distance reduced to less than 5 Å and further decreased to around 4 Å after 150 ns.

Stacking angles θ_1 and θ_2: Stacked configuration is defined when angles between the vector perpendicular to plane of nucleobase and that of ligand are either less than 40 or greater than 140 . Figure shows that initially in the best docked conformation, HBC rings are parallel to U10 and U12 but not within the stacking distance. The ligand then moves inwards at the core of junction, in the space created by residue U11 and after 150 ns of trajectory time stacks with the abovementioned residues. The stacking of HBC to core of 3WJ RNA is shown over the last 100 ns in figure. It shows a peak at 160 where the distance between ligand at residues is less than 4 Å .

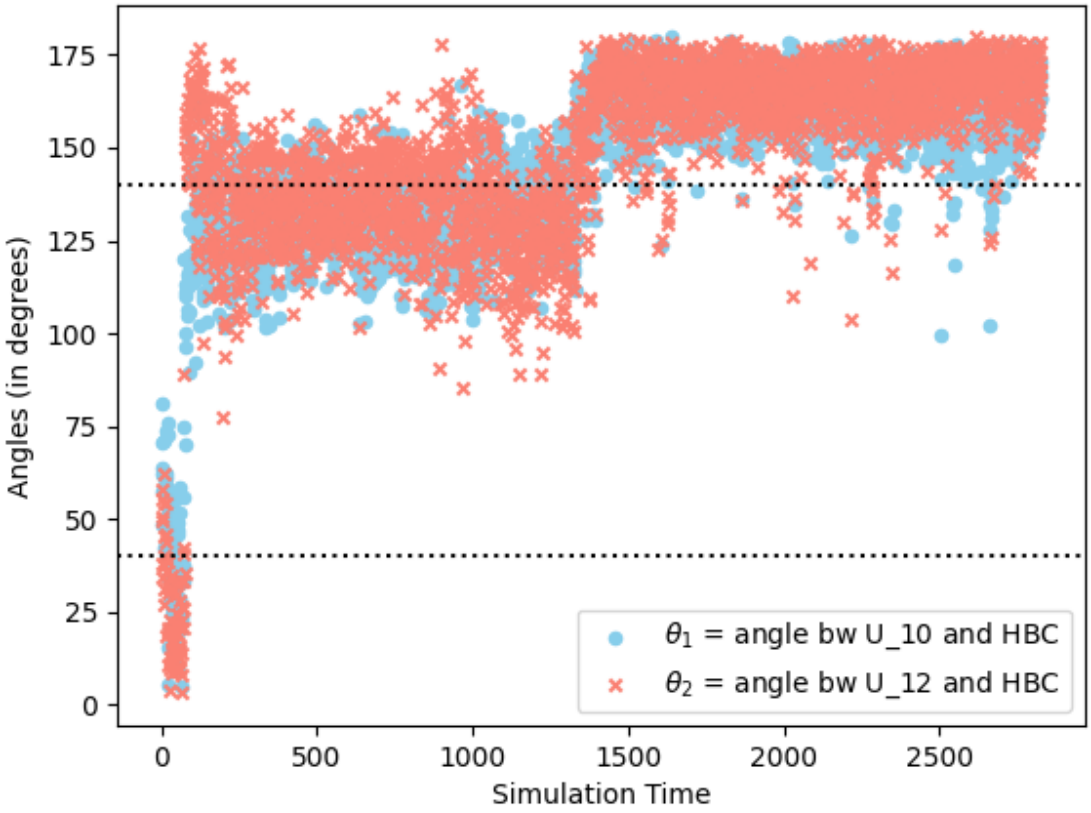


Figure 4.3: Stacking Angles 'θ_1' and 'θ_2' over the duration of simulation time.

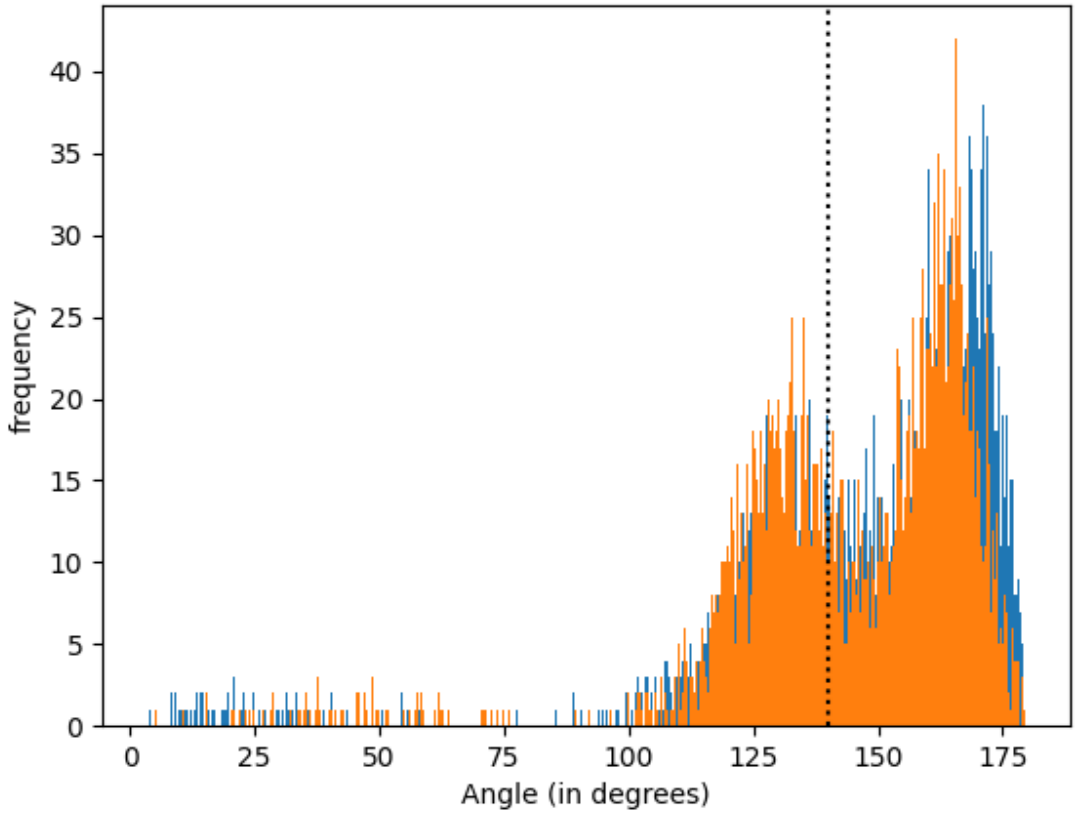


Figure 4.4: Distribution of parameters θ_1 and θ_2.

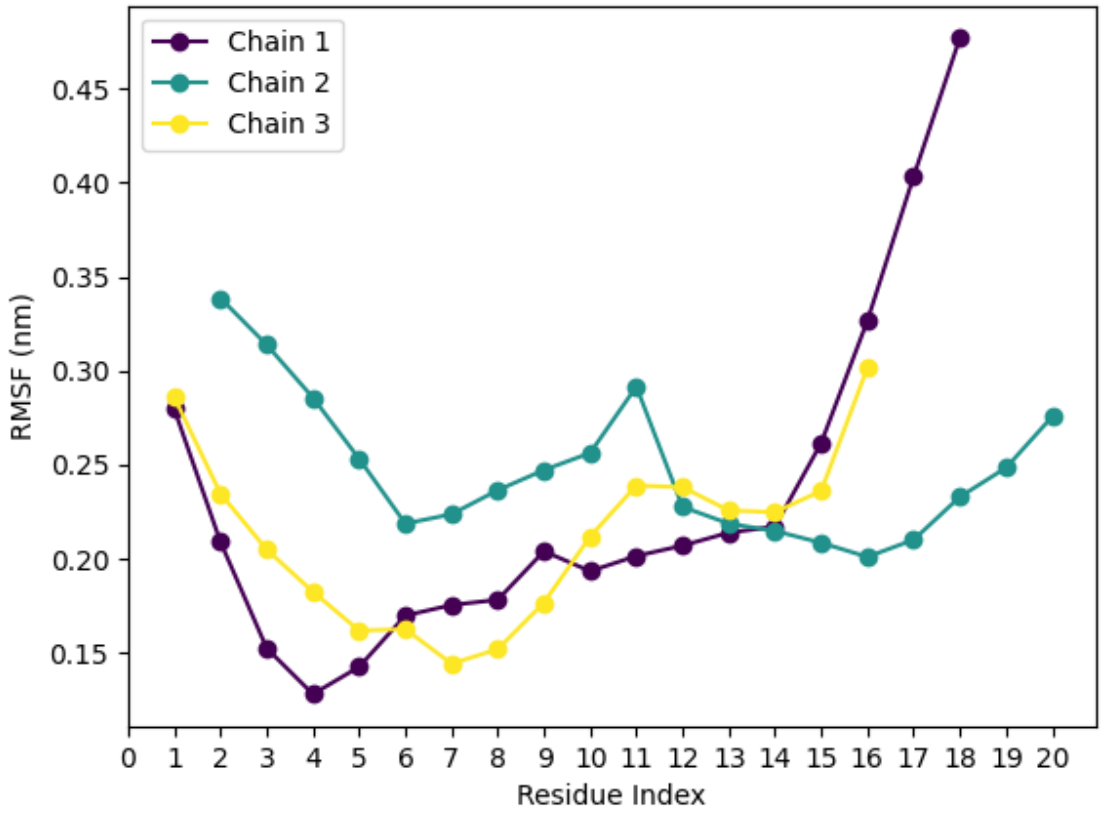


Figure 4.5: Root Mean Square Fluctuations of RNA residues upon binding to ligand HBC. We observe a marked reduction in fluctuations for chains 1 and 2, while that of chain 3 remains equal to that of 3WJ RNA in identical simulation conditions.

4.3.2 Dynamic properties of 3WJ RNA upon ligand binding

Residue wise fluctuations of RNA reduce owing to ligand binding(as shown in figure 4.5). $\epsilon RMSD$ values shown in figure 4.6 indicates that there is a slight change in conformation of 3WJ RNA bound to ligand when compared to unbound RNA in similar simulation conditions but it remains constant throughout the simulation. This indicates stability of RNA upon ligand binding. This study can be further extended by experimental collaboration. It also opens avenues to study existing RNA-aptamer ligand complexes through molecular dynamics simulations. Such studies can also guide the design of fluorogenic ligands and make them versatile to different RNA architectures.

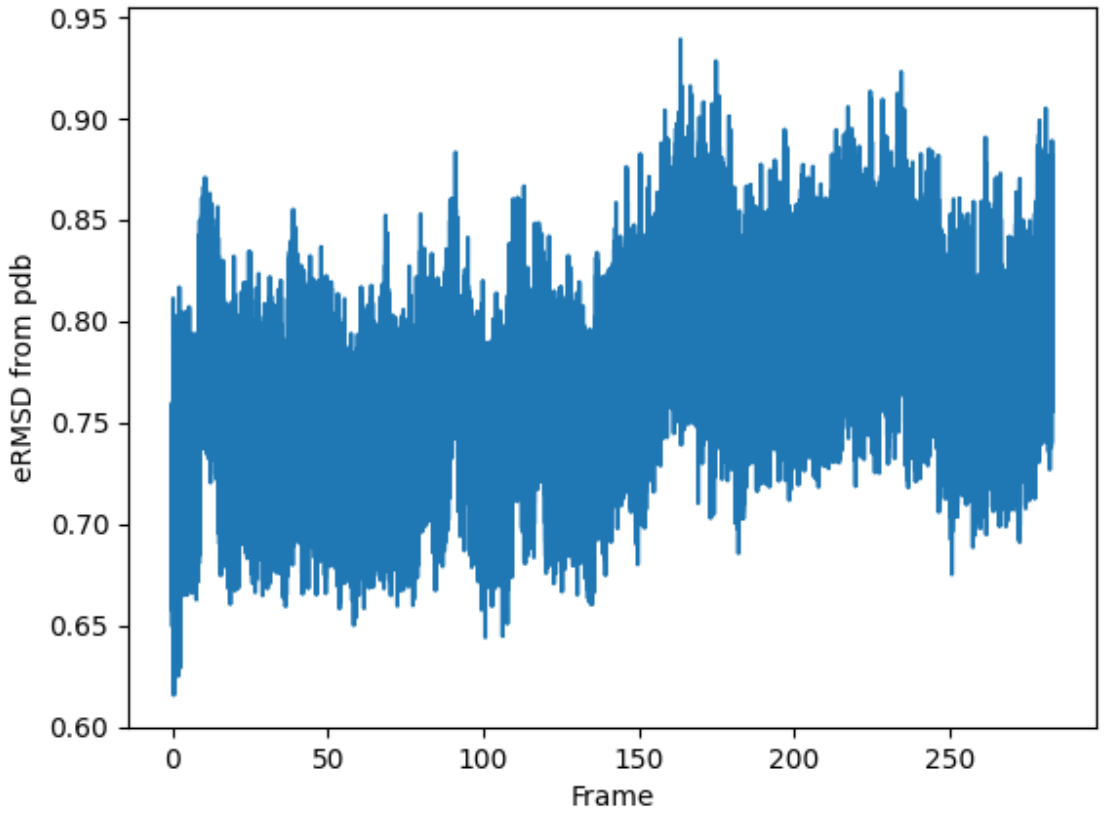


Figure 4.6: Evolution of $\epsilon RMSD$ values of RNA when bound to Ligand HBC

Chapter 5

Addendum I: A Study of binding of FAD analogues to *E.coli* proteome

5.1 Introduction

The biosynthesis of Flavin adenine dinucleotide (FAD) analogues and their role in cell physiology are crucial aspects in understanding cellular responses to oxidative stress. Reactive Oxidative Species (ROS) can cause damage to nucleic acids, proteins, lipids, and membranes, disrupting essential cellular functions and leading to adverse physiological effects [21]. To counteract oxidative stress, cells utilize glutathione (GSH). The effectiveness of the glutathione-based defense system depends on the presence and functioning of flavins, particularly FAD. In addition to maintaining redox balance in cells, FAD analogues also act as active cofactors for Escherichia coli glutathione reductase. Approximately 1% of eukaryotic and prokaryotic proteins are predicted to contain a FAD-binding domain. These domains are classified into three main families based on structural similarities and conserved sequence motifs: ferredoxin reductase (FR)-type, Rossmann fold, and p-cresol methylhydroxylase (PCMH) type FAD-binding domains [15].
To study the dependency of GSH on FAD our collaborator, Ateek Shah from Dr. Amrita Hazra's Lab at IISER, Pune, further investigated the impact of these analogues on cellular physiology under oxidative stress. Through genomic swapping of the essential gene ribF with a LoopSwap mutant gene, ribF25, they could produce of FAD analogues (FGD, FCD, and FUD) in E.coli. The mutant strain expressing ribF25, called LoopSwap was found to show similar growth and survival capabilities as the wild-type strain, even under oxidative stress induced by hydrogen peroxide.

Experimental data suggests that the LoopSwap mutant displays resistance against aminoglycoside antibiotics, such as kanamycin and gentamicin. Most resistance to aminoglycosides is caused by bacterial inactivation by intracellular enzymes .(*Manuscript under preparation*).

To investigate further, we analyzed the docking of FAD and its analogues with FAD binding proteins in *E. coli*. The aim of study was to identify potential targets in the cell through which FAD analogues contribute to antibiotic tolerance.

5.2 Methods

Ligand structures were generated using ChemDraw 3D. Protein structures were obtained from github repository prepared by Prathith, a collaborator on this project. The dataset contains structures of 70 FAD-binding proteins in *E coli* K-12 strain. Water molecules, metal atoms, and other ligands crystallized with the protein were removed. Protein and ligand structures were then converted into PDBQT format using AutoDockTools.

All docking calculations were performed using the AutoDock Vina 1.1.2 scoring function [34]. The exhaustiveness parameter was set to 32, and nine best binding poses were collected for each protein-FAD analogue pair. Blind docking calculations were conducted thrice for each protein-FAD analogue pair to account for the stochasticity associated with the docking method. The root mean square deviation values were caluclated using GROMACS module 'rms' [2].

5.2.1 Blind Docking

The grid box was centered around the center of mass of the protein. Structures generated from AlphaFold were found to be centered around the origin (0,0,0). For experimentally obtained structures, the geometric center was determined using BioPandas [24] (by collaborator, Prathith) and used as the center of the grid box. Grid sizes were obtained using the GROMACS 'editconf' module to create a cubic box around the protein with zero padding. Further analysis on blind-docked structures was done by Prathith, collaborator on this project.

5.2.2 Specific Docking

Docking was restricted to the binding region in holo-FAD binding proteins. A grid around the binding site in such proteins was created with a grid length of 25 Å(length of FAD in binding conformation + 3 angstroms on each side of ligand). Specific docking was performed to calculate the binding of FAD analogues and their conformations at the binding site.

5.3 Results and Discussion

The best docking scores, obtained by blind docking of FAD and its analogues give comparable mean docking scores and standard deviation values , shown in table 5.1. This preliminary evidence supporting the hypothesis is provided, that all FADs have comparable docking scores, indicating similar binding to the FAD Binding Proteins. But blind docking does not ensure correct binding site.
To validate results of blind docking, 11 FAD-binding proteins present in holo (ligand-bound form) were used. Upon further analysis of best binding conformation of FAD to these FAD-binding proteins present in *E. coli*, we observed that FAD bound to them at variable binding sites and variable conformations . Similar binding were observed for other analogues too. We concluded that blind docking using AutoDock Vina is not able to

Table 5.1: Docking Scores obtained from blind docking of FAD and its analogues with FAD-binding proteins.

	FAD	FUD	FGD	FCD
MEAN	-9.525	-9.267	-9.179	-9.190
S.D.	1.292	1.123	0.917	1.073
MAX VALUE	-13.8	-12.5	-11.4	-12.2
MIN VALUE	-6.8	-7.2	-7.6	-6.9

replicate binding sites and binding conformations for any of the FAD or its analogues.

Further, we applied specific docking to validate FAD binding to 11 holo-FAD binding proteins. The results of top conformations obtained from this protocol are shown in table:

From table 5.2 we observe comparable better docking scores FAD to FAD binding proteins and binding conformations within rmsd 2.5 $\mathring{A}$for 10 out 11 such proteins. Further, using the same protocol FAD analogues were docked to these FAD-binding proteins. Results of top conformations are shown in the table below:

From table 5.3 , we can conclude that FAD analogues show similar binding affinity to FAD-binding proteins on the validation set of 10 out of 11 such holo proteins. We excluded protein with uniProt ID : P244232 owing to mismatch in the nucleoside region of FAD(RMSD $\approx$ 3.5 Åvalue of docked ligand with respect to crystal structure. We further plan to validate these binding affinities from other docking techniques for example, diffusion models like Diffdock [12]

Table 5.2: Docking scores and root mean square deviations obtained via specific docking of FAD to 11 FAD-binding proteins. Proteins are represented by uniProt IDs.

Protein	RCSB PDB-ID	Docking Score (kcal/mol)	RMSD (wrt crystal structure) in Å
P77324	5G5G	-12.97	1.081
P28861	2XNJ	-10.37	2.492
P07003	3EY9	-13.9	0.985
P00363	1KF6	-14	1.4038
P77212	6KYY	-12.1	1.523
P42593	1PS9	-12.35	2.046
P35340	4O5Q	-13.37	0.84
P24232	1GVH	-11.85	3.4906
P10902	1KNR	-15.23	0.9499

Table 5.3: Docking scores obtained via specific docking of FAD and its analogues to 11 FAD-binding proteins. All docking scores are expressed in units kcal/mol. Proteins are represented by uniProt ID.

Protein	Docking score with FUD	Docking score with FGD	Docking score with FCD	Docking score with FAD
P77324	-13.2	-13.4	-13.2	-12.9
P28861	-10.4	-9.9	-8.6	-10.4
P07003	-13.4	-13.1	-12.2	-13.9
P00363	-13.1	-13.8	-13.1	-14
P77212	-10.6	-13.3	-12.7	-12.1
P42593	-13.1	-13.3	-12.5	-12.4
P35340	-13	-14.0	-12.7	-13.4
P10902	-15.2	-15.1	-15.2	-15.2
P08142	-14.3	-13.9	-13.5	-13.2
P0AC41	-12.1	-12.1	-11.3	-11.4

Bibliography

[1] (????). Equipartition and the Calculation of Temperature in Biomolecular Simulations .

[2] (2022). GROMACS Documentation Release 2022.4 GROMACS development team. Technical report.

[3] Allnér O, Nilsson L, Villa A (2012). Magnesium ion-water coordination and exchange in biomolecular simulations. Journal of Chemical Theory and Computation 8(4), 1493–1502.

[4] Binzel DW, Khisamutdinov EF, Guo P (2014). Entropy-driven one-step formation of phi29 pRNA 3WJ from three RNA fragments. Biochemistry 53(14), 2221–2231.

[5] Bonomi M, Bussi G, Camilloni C, Tribello GA, Banáš P, Barducci A, Bernetti M, Bolhuis PG, Bottaro S, Branduardi D (2019). Promoting transparency and reproducibility in enhanced molecular simulations. Nature Methods 16(8), 670–673.

[6] Bottaro S, Bussi G, Pinamonti G, Boomsma W, Lindorff-Larsen K (????). Barnaba: Software for Analysis of Nucleic Acids Structures and Trajectories .

[7] Bottaro S, Di Palma F, Bussi G (2014). The role of nucleobase interactions in RNA structure and dynamics. Nucleic Acids Research 42(21), 13306–13314.

[8] Bussi G, Donadio D, Parrinello M (2007). Canonical sampling through velocity rescaling. Journal of Chemical Physics 126(1).

[9] Chen L, Li J (2022). Mg2+ Ions Regulating 3WJ-PRNA to Construct Controllable RNA Nanoparticle Drug Delivery Platforms. Pharmaceutics 14(7).

[10] Chen X, Zhang D, Su N, Bao B, Xie X, Zuo F, Yang L, Wang H, Jiang L, Lin Q, Fang M, Li N, Hua X, Chen Z, Bao C, Xu J, Du W, Zhang L, Zhao Y, Zhu L, Loscalzo J, Yang Y (2019). Visualizing RNA dynamics in live cells with bright and stable fluorescent RNAs. Nature Biotechnology 37(11), 1287–1293.

[11] Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. Technical report.

[12] Corso G, Deng A, Fry B, Polizzi N, Barzilay R, Jaakkola T (2024). Deep Confident Steps to New Pockets: Strategies for Docking Generalization .

[13] Dao NT, Haselsberger R, Khuc MT, Phan AT, Voityuk AA, Michel-Beyerle ME (2021). Photophysics of DFHBI bound to RNA aptamer Baby Spinach. Scientific Reports 11(1).

[14] Doudna JA, Cech TR (2002). Site-specific RNA self-cleavage The chemical repertoire of natural ribozymes. Technical report.

[15] Dym O, Eisenberg D (2001). Sequence-structure analysis of FAD-containing proteins. Protein Science 10, 1712.

[16] Filonov GS, Moon JD, Svensen N, Jaffrey SR (2014). Broccoli: Rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution. Journal of the American Chemical Society 136(46), 16299–16308.

[17] Ho HO (????). A HO HO N N O O O DMHBI HBI (in GFP). Technical report.

[18] Jeng SCY, Trachman Iii RJ, Weissenboeck F, Truong L, Link KA, Jepsen MDE, Knutson JR, Andersen ES, Ferré-D'amaré AR, Unrau PJ (2021). Fluorogenic aptamers resolve the flexibility of RNA junctions using orientation-dependent FRET .

[19] Joung IS, Cheatham TE (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. Journal of Physical Chemistry B 112(30), 9020–9041.

[20] Liu X, Duan D, Wang Y, Liu J, Duan D (2022). Advancements in 3WJ-based RNA nanotechnology and its application for cancer diagnosis and therapy.

[21] Markkanen E (2017). Not breathing is not an option: How to deal with oxidative DNA damage.

[22] Ouellet J (2016). RNA fluorescence with light-Up aptamers.

[23] Parrinello M, Rahman A (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics 52(12), 7182–7190.

[24] Raschka S (2017). BioPandas: Working with molecular structures in pandas DataFrames. The Journal of Open Source Software 2(14), 279.

[25] Rees HC, Gogacz W, Li NS, Koirala D, Piccirilli JA (2022). Structural Basis for Fluorescence Activation by Pepper RNA. ACS Chemical Biology 17(7), 1866–1875.

[26] Schlitter J (1993). Estimation of absolute and relative entropies of macromolecules using the covariance matrix. Chemical Physics Letters 215(6).

[27] Solayman M, Litfin T, Singh J, Paliwal K, Zhou Y, Zhan J (2022). Probing RNA structures and functions by solvent accessibility: an overview from experimental and computational perspectives.

[28] Solayman M, Litfin T, Singh J, Paliwal K, Zhou Y, Zhan J (2022). Probing RNA structures and functions by solvent accessibility: an overview from experimental and computational perspectives.

[29] Sponer J, Bussi G, Krepl M, Banas P, Bottaro S, Cunha RA, Gil-Ley A, Pinamonti G, Poblete S, Jurečka P, Walter NG, Otyepka M (2018). RNA structural dynamics as captured by molecular simulations: A comprehensive overview.

[30] Sun Y, Di W, Li Y, Huang W, Wang X, Qin M, Wang W, Cao Y (2017). Mg 2+ -Dependent High Mechanical Anisotropy of Three-Way-Junction pRNA as Revealed by Single-Molecule Force Spectroscopy . Angewandte Chemie 129(32), 9504–9508.

[31] Takahashi S, Sugimoto N (2020). Stability prediction of canonical and non-canonical structures of nucleic acids in various molecular environments and cells. Chemical Society Reviews 49(23), 8439–8468.

[32] Trachman RJ, Ferré-D'Amaré AR (2019). Tracking RNA with light: selection, structure, and design of fluorescence turn-on RNA aptamers.

[33] Tribello GA, Bonomi M, Branduardi D, Camilloni C, Bussi G (2013). PLUMED 2: New feathers for an old bird .

[34] Trott O, Olson AJ (2010). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. Journal of Computational Chemistry 31(2), 455–461.

[35] Wang J, Wang W, Kollman PA, Case DA (????). Antechamber, An Accessory Software Package For Molecular Mechanical Calculations Correspondence to. Technical report.

[36] Warner KD, Chen MC, Song W, Strack RL, Thorn A, Jaffrey SR, Ferré-D'Amaré AR (2014). Structural basis for activity of highly efficient RNA mimics of green fluorescent protein. Nature Structural and Molecular Biology 21(8), 658–663.

[37] Zgarbová M, Otyepka M, Šponer J, Mládek A, Banáš P, Cheatham TE, Jurečka P (2011). Refinement of the Cornell et al. Nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. Journal of Chemical Theory and Computation 7(9), 2886–2902.

[38] Zhang H, Endrizzi JA, Shu Y, Haque F, Sauter C, Shlyakhtenko LS, Lyubchenko Y, Guo P, Chi YI (2013). Crystal structure of 3WJ core revealing divalent ion-promoted thermostability and assembly of the Phi29 hexameric motor pRNA. RNA 19(9), 1226–1237.