

THE EFFECT OF ATOMISTIC INTERACTIONS ON RNA CONFORMATIONAL  
DISTRIBUTIONS, FOLDING AND LIGAND BINDING VIA MOLECULAR  
DYNAMICS SIMULATIONS AND DOCKING

by

Sayyed Hamed Sadat Hayatshahi

A dissertation submitted to the faculty of  
The University of Utah  
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Medicinal Chemistry

The University of Utah

December 2016

ProQuest Number:10247781

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



ProQuest 10247781

Published by ProQuest LLC (2017). Copyright of the Dissertation is held by the Author.

All rights reserved.

This work is protected against unauthorized copying under Title 17, United States Code  
Microform Edition © ProQuest LLC.

ProQuest LLC.  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106 – 1346

Copyright © Sayyed Hamed Sadat Hayatshahi 2016

All Rights Reserved

# The University of Utah Graduate School

## STATEMENT OF DISSERTATION APPROVAL

The dissertation of Sayyed Hamed Sadat Hayatshahi  
has been approved by the following supervisory committee members:

Thomas E. Cheatham, Chair 8/8/16  
Date Approved

Amy M. Barrios, Member 8/8/16  
Date Approved

Cynthia Burrows, Member 8/8/16  
Date Approved

Darrell R. Davis, Member 8/8/16  
Date Approved

Raphael Franzini, Member 8/8/16  
Date Approved

and by Eric W. Schmidt, Chair/Dean of

the Department/College/School of Medicinal Chemistry

and by David B. Kieda, Dean of The Graduate School.

## ABSTRACT

Advances in molecular dynamics (MD) simulation methodologies have enabled researchers to explore the conformational spaces of biological macromolecules more efficiently and quickly. Specifically, the development of enhanced sampling techniques has provided researchers with well-converged conformational ensembles of small macromolecules. It has been shown that converged simulations of small ribonucleic acids (RNA) such as tetranucleotides result in the population of experimentally unknown conformations, indicating RNA force field artifacts. However, although being imperfect, the current RNA force fields have also been useful in characterizing the varied interactions of ions and ligands with RNA. In this thesis, we analyze conformational ensembles of dinucleotide monophosphates generated with different force fields and water models with the aim of pinpointing force field problems. We also utilize the current force fields to demonstrate the preferential potassium binding to a buried ion-binding site in a ribosomal RNA molecule known as GTPase Associating Center (GAC) and also to elucidate an ion-dependent step in its unfolding pathway. We further show magnesium-independency of binding of a crystallographic 2-benzimidazole ligand to the Internal Ribosome Entry Site (IRES) of Hepatitis C virus (HCV), using MD simulations and docking. Our strategy is to assess simulation results with existing experimental data, and then also use simulation results to increase insight into RNA interactions and folding. These methods allow us to identify deficiencies of some current RNA force fields.

## TABLE OF CONTENTS

ABSTRACT.....	iii
Chapters	
1 INTRODUCTION.....	1
1.1 Scope of the presented research.....	1
1.2 Challenges in RNA simulations.....	2
Current perspectives for assessment and improvement of RNA	
1.3 simulations.....	5
2 CONSENSUS CONFORMATIONS OF DINUCLEOTIDE MONOPHOSPHATES DESCRIBED WITH WELL-CONVERGED MOLECULAR DYNAMICS SIMULATIONS.....	8
2.1 Introduction.....	8
2.2 Methods.....	14
2.3 Results.....	20
2.4 Conclusions.....	31
3 COMPUTATIONAL ASSESSMENT OF POTASSIUM AND MAGNESIUM ION BINDING TO A BURIED POCKET IN GTPASE-ASSOCIATING CENTER RNA.....	62
3.1 Introduction.....	62
3.2 Methods.....	67
3.3 Results.....	76
3.4 Conclusions.....	84
4 INVESTIGATING THE ION-DEPENDENCE OF THE FIRST UNFOLDING STEP OF GTPASE-ASSOCIATING CENTER RIBOSOMAL RNA .....	99
4.1 Introduction.....	99
4.2 Methods.....	102
4.3 Results.....	107
4.4 Conclusions.....	111
5 THE EFFECT OF CRYSTALLOGRAPHIC MAGNESIUM ON BINDING OF	

2-AMINOBENZIMIDAZOLE INHIBITORS TO THE HEPATITIS C VIRUS IRES RNA.....	125
5.1 Introduction.....	125
5.2 Methods.....	126
5.3 Results.....	130
5.4 Conclusions.....	133
6 CONCLUSION.....	140
6.1 Significant outcomes.....	140
6.2 Future directions.....	143
APPENDIX: SCRIPTS USED FOR BUILDING THE MODELS AND ANALYZING THE SIMULATIONS.....	146
REFERENCES.....	156

## CHAPTER 1

### INTRODUCTION

#### **1.1 Scope of the presented research**

There are limitations in experimental approaches in investigating the structure and dynamics of biological macromolecules and their interactions with drugs. Not all molecules can be easily crystallized to be studied using X-ray crystallography. Also when crystallization is done successfully, the crystal structures are not always high resolution, and it is difficult to identify interacting ions and water residues at certain resolutions (1-3). Nuclear magnetic resonance (NMR), as another powerful structural biology technique, has its own molecular size limitations (4) and might lead to ambiguous results in certain chemical exchange regimes (5). On the other hand, theoretical simulations aim to answer questions that cannot be investigated by these and other experimental methods. Among the simulation methods, molecular dynamics (MD) simulations can potentially provide detailed descriptions about the structure and dynamics of macromolecules and their interactions with ligands at an atomistic resolution. With continuing advances in MD simulation quality and rapidly increasing computational power, it is estimated that atomistic description of ligand binding to relatively large macromolecules will be possible in the coming decade although D.E. Shaw is more optimistic about the speed of the simulations in near future (6).

The focus of the research presented in this dissertation is to analyze the ability of MD



simulations to reveal structure and dynamics details of ribonucleic acids (RNA) and their interactions with ions and drugs. We start with investigating the conformational distributions and intra-molecular interactions of RNA dinucleotides (Chapter 2) as very small systems. We study such small RNA sequences to highlight the basic atomistic interactions that govern the conformational attributes of RNA molecules and to pinpoint the interactions that need to be tweaked to improve the reliability of the current models. Then, we apply the contemporary simulation technology to probe the role of ion interactions in unfolding pathway of a large ribosomal RNA (Chapters 3 and 4, respectively), and validate the results with existing experimental data. We also question and discuss a hypothesis about interdependence of ion binding and drug binding to an RNA drug target using molecular dynamics and docking approaches (Chapter 5).

Molecular dynamics and docking simulations of RNA have become more significant from a drug design viewpoint as the need for RNA targeting medicines is increasing (7). Viral RNA genomes and the translation apparatus in HIV, HCV, and Zika viruses as well as the bacterial ribosomal RNA are interesting drug targets that might help with efficient treatment of related diseases (7,8). Having reliable models that reproduce correct RNA conformational populations and predict RNA-drug interactions in biological conditions can significantly help with reducing the drug design cost and improving the efficacy of future medicines.

## **1.2 Challenges in RNA simulations**

Unfortunately, MD simulations have not evolved for RNA molecules as they have for proteins. The availability of enormous computational resources has enabled researches, to calculate the population of inaccurate conformations in fully-converged simulations for

some small RNA systems in recent years. For example, the GACC tetranucleotide has been shown to adopt experimentally unknown intercalated and inverted conformations in large populations in MD simulations using different force fields (9-11). The CCCC tetranucleotide which adopts an A-form conformation according to NMR experiments (12), populates two experimentally unknown conformations in the fully converged simulations (10). The UUCG tetraloop also adopts ladder-shaped structures in simulations at low temperatures which are also low population in experiment (10). The GAGA tetraloop is shown to populate a nonnative structure in which all purine bases stack on top of each other and also some single-stranded structures with high population of interaction between the 2' hydroxyl and phosphate nonbridging oxygen (13). The UUCG, GCAA and CUUG tetraloops are reported to fold to their hyperstable in molecular dynamics simulations using new force field parameters developed by Chen and Garcia, although these tetraloops failed to fold with the standard Amber force fields (14). A recent work reports that none of the state-of-the-art force fields can elucidate the folding of GAGA tetraloop (13).

There are different reasons that make RNA simulations particularly challenging. One reason is the complexity of RNA conformational space compared to protein. Part of this complexity comes from having many rotatable dihedral bonds per residue in RNA compared to few in proteins (Figure 1.1). Accordingly, few major secondary structure protein conformations can be described with locations on a simple two-dimensional Ramachandran plot with  $\Phi$  and  $\Psi$  dihedral values as each dimension. In contrast, in a study performed in 2008, each RNA dinucleotide unit within the experimentally known RNA structures were classified into 46 conformational suites based on its position on a seven-dimension dihedral space including  $\delta$ ,  $\epsilon$ ,  $\zeta$  of the preceding residue and  $\alpha$ ,  $\beta$ ,  $\gamma$  and

$\delta$  of the following residue (15). Furthermore, adding different configurations of the glycosidic bond in each nucleotide adds to the roughness of the conformational surface. The interdependence of these dihedral angles and the effect of atomistic nonbonded interaction on their configurations make it difficult to provide force field parameters that lead to accurate reproduction of real conformations in realistic relative populations.

The amphipathic nature of nucleic acids, which is resulted by negatively charged backbone phosphates on one hand and the hydrophobic bases on the other, contributes to the formation of their conformations. This fact highlights the importance of solvent and ion interactions in real nucleic acid samples, as certain conformations only exist in certain solvent/ion conditions. Therefore, correct and balance representation of these interactions in simulations is extremely important and determining. Computationally cheaper implicit solvent models, which have been used in protein simulations with acceptable reliability, are not as reliable in nucleic acid simulations (9,16). Also it has been shown that using different water models can significantly shift the RNA conformational distribution (17).

Folding of RNA causes accumulation of negative electrostatic charges on backbone phosphates in relatively small space, which should be neutralized by cations to provide stability for any nucleic acid conformation. This makes the solution RNA structure so sensitive to the ionic strength of its environment (18). Regarding the importance of ion-RNA interactions, one would expect that the correct parameterization of ions as well as the RNA atomic partial charges have determining roles in quality of simulations (19,20). There are different models presented divalent ions (21-24), which of course, have important roles in RNA conformational stability, folding and ligand interactions and will be discussed in the following chapters. Each of these ion models has been parameterized to reproduce some aspects of the ion interactions and different ions models are even

compatible with certain solvent representations (25); and therefore, there is no universal answer to the question of ion model choice in RNA simulations although such choice can potentially have a great effect on the resulting model accuracy.

### **1.3 Current perspectives for assessment and improvement of RNA simulations**

MD simulations have two major limitations. One reason is the lack of sufficient sampling. Short MD simulations do not sample the energy surface well enough to find conformations separated from the initial conformation with high energy barriers. The highest MD integration time step used in atomistic MD simulations of biological molecules with hydrogen mass repartitioning is 4 fs (26,27), and so with the state-of-the-art hardware technology (NVIDIA K80 GPUs), we can reach 448 ns per day of sampling for a dinucleotide monophosphate system in explicit water molecules with total of 3301 atoms. In contrast, the folding of RNA takes place on the millisecond to second time scale (28) and magnesium interaction with ATP has an exchange rate of 2.5 ms in-vitro (29). The gap between the simulation timescale and the folding timescale shows that complete sampling of the conformational space of a typical RNA molecule and its interactions with ions and ligands with straight MD simulation is not possible with the current hardware technology.

The other source of error in MD simulations is the inaccuracy of the force field parameters. According to a definition by M.A. Gonzalez, “a force field is a mathematical expression describing the dependence of the energy of a system on the coordinates of its particles” (30). The ideal parameters used in this expression would enable correct representation of potential energy of a system according to every set of atomic

coordinates. Therefore, any change in the force field parameters can potentially lead to major changes in the energetic surface of the system; and an inaccurate force field will result in generating wrong conformations and/or correct conformations in wrong relative populations with respect to the real system.

Development of enhanced sampling methods such as Replica-Exchange Molecular Dynamics (REMD) (31) and accelerated Molecular Dynamics (aMD) (32) together with accessing petascale computational resources have allowed us to study converged conformational ensembles of small RNA sequences in recent years (33). The ability to perform such extensive sampling has enabled us to assess the reliability of simulations in absence of the sampling problem at least for smaller systems. In fact, an unrealistic conformational distribution found in converged simulations of these systems can be indicative of force field problems. Therefore, a feasible current approach toward assessment of RNA simulations is to study conformational distributions of smaller RNA systems in converged simulations with regard to the available experimental data. The research presented here aims to follow this theme with generating and studying converged simulations of dinucleotide monophosphates and analyzing the effects of different force field parameters on the conformational distributions. It also tries to demonstrate the ability of the current RNA simulation protocols in answering questions about folding and ligand interactions of larger RNA systems and provide evidence from the available experimental data to analyze the reliability of these simulations.

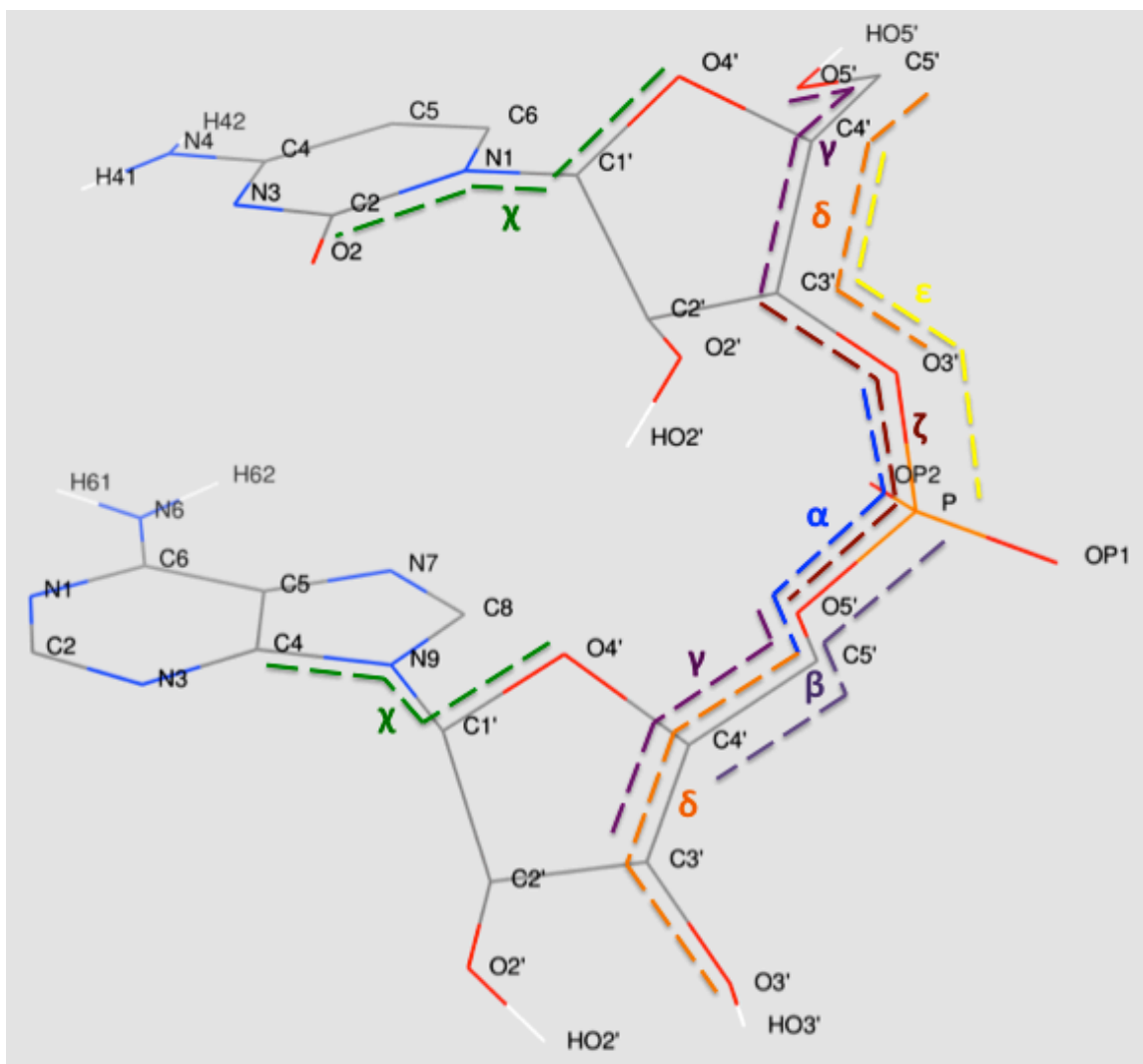


Figure 1.1. The structure of a dinucleotide monophosphate (CpA) in an A-form conformation is represented with atom names according to the standard nomenclature. Dashed lines represent rotatable dihedral angles with matching color Latin letter names. Each dihedral angle is made with connections between four atoms.

## CHAPTER 2

# CONSENSUS CONFORMATIONS OF DINUCLEOTIDE MONOPHOSPHATES DESCRIBED WITH WELL-CONVERGED MOLECULAR DYNAMICS SIMULATIONS

The research about dinucleotide monophosphates has been designed by Hamed S. Hayatshahi and Thomas E. Cheatham, III. Simulations and analysis are performed by H. S. Hayatshahi. The text is written and edited by Hamed S. Hayatshahi and Thomas E. Cheatham, III.

### **2.1 Introduction**

RNA molecules have crucial biological functions which are progressively being revealed. These functions include but are not limited to transferring the genetic code, contributing in the protein synthesis machinery and enzymatic activity. Many of these biological roles rely on adopting specific conformations and dynamics on certain timescales that facilitate intermolecular and intramolecular interactions required for the varied roles of RNA (34,35). These interactions range from local atomistic interactions between neighboring nucleotides to long-range interactions in larger tertiary structures. Understanding local atomistic interactions between neighboring nucleotides in RNA can provide insight into formation of conformations adopted by larger sequences as subsets of

different local motifs; the smallest RNA sequence unit is dinucleotide monophosphates (DNMPs), and here we analyze the conformational diversity of DNMPs with a variety of force fields.

The conformational distribution of DNMPs and their sequence specific structural dependencies are described in many experimental studies (36-42). For example, various nuclear magnetic resonance (NMR) studies have fit the data with an equilibrium of different conformational states (36). In one of these studies, proton NMR was utilized to analyze the structures of ApA, ApG, UpU, CpC, UpC and CpU (37). This study described the major conformations by identifying the  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  torsions made by phosphodiester connections (traditionally named as  $\omega_1$  and  $\omega_2$ .(36,37)). Xp- and -pX are used for the 5' and 3' end residues, respectively) as flexible dihedral torsions, which cause different orientations between bases. Accordingly, they proposed two stacked conformations and an extended conformation for DNMPs. One of the stacked conformations has the properties of the A-form RNA with g(-)-g(-) combination for  $\zeta_{Xp-}$  and  $\alpha_{-pX}$ , while the other is a left helical stacked conformation with g(+)-g(+)  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  combination. They also qualitatively described the sugar puckering at each nucleotide and its relationship to the base identity, stacking and temperature.

Another NMR study has proposed properties of A-form RNA for ApA and estimated the backbone dihedrals by measuring phosphorus-proton spin-spin couplings (38). In contrast, a crystal structure study suggested two distinct conformations for UpA, both of which are different from the A-form conformation.(39,40) One of these conformations is an unstacked extended form with g(-)-t combination of  $\zeta_{Xp-}$  and  $\alpha_{-pX}$ , and the other is a stacked form with g(+)-g(+) combination. These conformations are different from the g(-)-g(-) combination which is usually found in the double helical A-form (41). However,



both of these conformations share the C3'-endo sugar conformation and the anti configuration for the glycosidic bond as preferred in the A-form RNA. A more recent study has measured NMR scalar spin-spin couplings (J-coupling) for many dihedral angles in ApA, ApC, CpA and CpC.(42) However, calculation of the dihedral angles from this J-coupling data depends on the choice of the specific Karplus equations(43-46); this study used an average of these equations (42).

There are also experimental studies attempting to measure the stacking free energies of DNMPs (37,47,48). These results can potentially be used to roughly describe conformations of DNMPs and assess DNMP simulations as done in a recent theoretical study by R.F. Brown et al. (49). However, there are few uncertainties that should be considered in this approach. One uncertainty comes from having different stacking values that are generated via different experimental approaches. For example, the stacking free energies obtained by D. Frechet et al. (37) via UV absorbance during thermal melting differs significantly from those reported by C. Lee et al. (48) via an NMR approach based on the sugar-related coupling constants (37,48). Also interpreting the experimental data with a simple two-state model to describe the DNMP conformations as either “stacked” or “unstacked” is likely an inaccurate assumption as multiple states are likely populated (50). For example, the  $\zeta_{g(+)}-\alpha_{g(+)}$  conformation results in a stacked conformation in UpA which is a markedly different from the stacked A-form conformation (39,40). Also proposing any definition that distinguishes between the stacked and unstacked population in the simulation ensembles should be done with caution because the arbitrary distance or angle cutoffs used in such definition might not be consistent with the observations from a given experimental study.

Dinucleotides have also been studied with a bioinformatics approach by Richardson

et al. (15) where the available high resolution RNA structures were mined for different dihedral torsion combinations of  $\delta\epsilon\zeta\alpha\beta\gamma\delta$  in neighboring nucleotides (15). This analysis classified the dinucleotide structures into 46 sugar-to-sugar conformational “suites”. Although an outlier DNMP conformation that does not fit into these suite definitions is more likely to be unrealistic, it should be disregarded with caution because a free DNMP can potentially adopt conformations which are not frequently observed in the context of larger RNA sequences.

In spite of many approaches to describe the conformations of DNMPs, some of their important conformational elements have been described only qualitatively with high uncertainty. For example, determining the orientation of the glycosidic bond has been challenging due to technical issues and therefore mostly qualitative interpretations of the data have emerged (36,37).

Molecular dynamics (MD) simulations can potentially provide more insight into the conformations of DNMPs with the assumption of utilizing accurate force fields and adequate amount of sampling from the conformational space. Building on previous MD simulation studies of DNMPs, we now investigate additional and different force fields and also aim to achieve full convergence in the simulations (42,49). Our recent work in using various enhanced sampling methods applied to RNA tetranucleotide simulations has improved the sampling efficiency to a level of very reasonable convergence between independent runs, measured both in terms of structural and dynamic properties, so that the observed conformational distributions closely represent the correct Boltzmann-weighted populations for the specific force field studied (10,33). With sampling uncertainties eliminated, these simulations reveal anomalous or experimentally-unknown conformations with different force fields, a potential indication of force field problems

(9,10). To better understand the influence of different RNA force fields on the results, multiple different force field modifications were investigated. These include recent modifications to the traditional Amber ff99 (51) force field, including the parmbsc0  $\alpha/\gamma$  (52) and  $\chi_{OL3}$  (53) modifications (here referred to as ff14) and the Chen-Garcia modifications aimed to reduce base stacking (14). Although these decrease the base stacking interactions (14), they were found not to increase the experimentally known conformations in converged simulations of the GACC tetranucleotide (10). Another study has suggested new van der Waals parameters for phosphoric acid esters based on reproducing correct solvation energies (54). Those modifications were used in re-parametrization of RNA phosphates and found to increase the experimentally observed NMR conformations of GACC and also decreased the anomalous conformations of CCCC to some extent in converged simulations (10). Also, the recent OPC (55) water model has been found to further improve the GACC and CCCC results compared to the more traditional TIP3P(56) model (17). In spite of the force field caveat that none of the currently available RNA force fields provide a completely accurate modeling of tetranucleotides and tetraloops, well-converged simulations can still broaden our insight into various possible conformations that can be adopted by DNMPs and their stabilizing interactions. Rough judgement about accuracy of the generated conformations in the simulation ensembles can be done with existing experimental data while more elaborate assessment might need further quantitative experimental evidence.

Besides providing more insight into the DNMP conformations, simulations of DNMPs will also be helpful with optimizing the RNA force fields. DNMPs are the smallest sequences which encompass all RNA backbone dihedrals. The simplicity of these molecules, with fewer intramolecular interactions and degrees of freedom relative

to larger sequences, could facilitate better understanding of the role of each interaction in determining the DNMP conformations and their relative stability. This will potentially lead to identification of interactions and parameters that result in population of unwanted conformations. Even in the absence of essential experimental data for rigorous identification of unwanted DNMP conformations generated in simulations, conformational changes in these easy-to-converge simulations can be used for quick and computationally cheap assessment of any new force field modification and potentially prediction of its efficiency in simulations of larger RNA molecules.

In this study, we first highlight the importance of simulation convergence by comparing the results of the relatively converged unbiased regular MD simulations of ApA with less than fully converged simulations from previous work (49). Then we utilize a protocol for fast convergence of DNMP conformational distributions using temperature replica-exchange MD (TREMMD) as an enhanced sampling method. We apply this protocol to determine the converged conformational distributions of all 16 DNMPs with the Amber ff14 force field (ff99 (51) + parmbsc0  $\alpha/\gamma$  (52) +  $\chi_{OL3}$  (53) modifications) in TIP3P water. Distinct conformations generated for each DNMP in the simulation ensemble are identified and characterized in atomistic detail. We also test a selection of seven DNMPs with five other “water model / force field” combinations (i.e., TIP3P / ff14+vdW<sub>bb</sub> (54), TIP3P / ff99+Chen-Garcia (CG) (14), TIP3P / CHARMM36 (57,58), OPC (55) / ff14, OPC (55) / ff14+vdW<sub>bb</sub> (54)) and compare the effect of OPC water model and the tested force fields on populations of conformations of these DNMPs. We also provide qualitative assessment for each DNMP conformation with respect to the sugar-to-sugar conformational suites (15). Finally, we investigate the atomistic interactions and force field parameters that lead to stabilization of each conformation.

## 2.2 Methods

### 2.2.1 Simulations and scenarios

The three categories of simulations done in this work are as summarized in Table 2.1. All simulations were done with one neutralizing  $\text{Na}^+$  and 5 NaCl (~150 mM) ions, except for one ApA run of the second category which was done with only the neutralizing  $\text{Na}^+$  to see if the excess ions affect the conformational distributions.

First, ApA was simulated with regular MD independently in both TIP3P and TIP4P-EW (59) water with a 2 fs integration time step using the ff14 force field (ff99 (51)+ parmbsc0  $\alpha/\gamma$  (52) +  $\chi_{\text{OL3}}$  (53) modifications). The ApA simulation in TIP3P water was repeated with a 4 fs integration time step using hydrogen mass repartitioning (HMR) (26,27) to see if HMR affects the conformational distributions. The straight MD simulations were done with three independent copies starting from the same ApA structures with randomized ion positions.

Second, all 16 dinucleotide monophosphates (DNMPs) were simulated in explicit TIP3P water using ff14 force field (ff99 (51)+ parmbsc0  $\alpha/\gamma$  (52) +  $\chi_{\text{OL3}}$  (53) modifications) using temperature replica-exchange MD (TREMMD). Seven DNMPs (ApA, ApC, CpA, CpC, UpU, GpG, ApG) were also simulated with 5 other force field combinations (i.e., TIP3P / ff14+vdW<sub>bb</sub> (54), TIP3P / ff99+Chen-Garcia (CG) (14), TIPS3P / CHARMM36 (57,58) , OPC (55) / ff14, OPC (55) / ff14+vdW<sub>bb</sub> (54)) using TREMMD. Two independent runs of each TREMMD simulation were performed starting from different DNMP structures and randomized ion positions for convergence and error analysis. All TREMMD simulations in this work were performed with 4 fs integration time steps using hydrogen mass repartitioning (HMR) (26,27).

Third, ApA was simulated using the same TREMMD setup as in the second category in

TIP3P water with ff14 plus specific additional modifications or restraints. These simulations were performed to monitor the influence of specific force field parameter changes and altered atomistic interactions on the ApA conformational distribution.

### 2.2.2. Model setup and equilibration

To set up the ApA simulations with regular MD, the ApA structure was built and solvated in cubic boxes of either 1066 TIP4P-EW (59) (MD-TIP4PEW) or 1075 TIP3P (56) (MD-TIP3P and MD-TIP3P-HMR) waters with each side set to 33 Å using tLEaP. 6 Na<sup>+</sup> and 5 Cl<sup>-</sup> ions with Joung-Cheatham parameters (25) were added to reach an excess NaCl concentration of about 150 mM to neutralize the system (49). The ff99 (51) force field with parmbsc0  $\alpha/\gamma$  (52) and  $\chi_{OL3}$  (53) (ff14) modifications were used to parameterize the DNMP atoms. The ions were relocated to random positions by swapping with random waters using CPPTRAJ (60) in AmberTools. The second and third copies of the simulations were prepared by rerandomizing the ion positions in the simulation box. Hydrogen mass repartitioning (HMR) was used in simulations run with 4 fs integration time steps (MD-TIP3P-HMR and all TREMD simulations) (26,27). Changing the atomic mass for HMR simulations were done with *parmed.py* in AmberTools on the nucleic acids only.

TREMD-TIP3P-FF14 simulations were setup in a similar manner to the MD-TIP3P-HMR simulations. However, the representative structure of a highly populated cluster from the MD-TIP3P simulations of ApA (inverted conformation) was used as the initial ApA structure for the second independent MD-TIP4PEW run. The initial structures of the second independent MD-TIP4PEW runs for other DNMPs were built manually by setting the torsions to values similar to the inverted ApA structure. The same initial structures for

each system were used in all TREMD simulations. In all TREMD simulations of all DNMPs, the simulation box size was changed such that the same number of waters (1075) fit in the initial simulation box. TREMD-OPC-FF14 simulations were set up similar to the MD-TIP4PEW simulations except that the OPC (55) water model was used. Joung-Cheatham ions parameters for TIP4P-EW water were used for ions in simulations with OPC water model.

To setup the TREMD-TIP3P-VDW and TREMD-OPC-VDW simulations, after models were built using ff14 force field, *parmed.py* was used to set the radii of OP1 and OP2 atoms to 1.7493 Å, and also to set the radii of O3' and O5' atoms connected to phosphorus atom to 1.7718 Å.

To setup TREMD-TIP3P-CG simulations, systems were built in tLEaP with the ff99 (51) force field with modifications proposed by Chen and Garcia (14). The off-diagonal term between the base and water atoms were modified with *parmed.py*.

TREMD-TIP3P-C36 simulations were setup with building the DNMP structures in CHARMM36 and then converting the topology files to AMBER topology files using CHAMBER (61).

TREMD-MOD simulations were setup consistent with the TREMD-TIP3P-FF14 ApA simulations. The topology files were then changed with *parmed.py*. In TREMD-MOD-O2', the HO2' atom was neutralized with transferring its electrostatic charge (0.4186) to O2' atom in both residues changing the O2' charge to -0.1953. In TREMD-MOD-O5', the HO5' atom was neutralized with transferring its electrostatic charge (0.4295) to O5' atom in both residues changing the O5' charge to -0.1928. The setup for TREMD-REST simulations were done consistent with the TREMD-TIP3P-FF14 ApA simulations. The tLEaP, *parmed.py*, CPPTRAJ and CHARMM36 scripts that were used

to setup the systems are available in the Supporting Information.

All models were equilibrated with the same protocol starting with 1000 steps of steepest descent minimization with 5 kcal/mol-Å<sup>2</sup> positional restraints on DNMP heavy atoms. This was followed with 15 ps MD with the same positional restraints in NVT. Then 3 1000-steps steepest descent energy minimizations were performed with positional restraints of 2, 0.1 and 0 kcal/mol-Å<sup>2</sup> on DNMP heavy atoms, respectively. Then 5 ps of MD were done with 1 kcal/mol-Å<sup>2</sup> positional restraints on DNMP heavy atoms in NPT. This was followed by 1 ns MD with 0.5 kcal/mol-Å<sup>2</sup> positional restraints on DNMP heavy atoms and then 1 ns MD with 0.5 kcal/mol-Å<sup>2</sup> positional restraints on DNMP backbone heavy atoms and then 1 ns MD with no positional restraints, all in NPT. The integration time step was set to 1 fs in all MD steps except the last step which was done with 2 fs time step. Bonds to hydrogen atoms were constrained using SHAKE in all MD steps (62). Temperature was kept at 298.15 K with the weak-coupling algorithm (63) with 1 ps time constant in all MD steps. Periodic conditions and particle mesh Ewald (PME) with 8.0 Å cutoff and default parameters was applied to all steps (64). Simulations were done with *pmemd.cuda* of AMBER14 (65).

### 2.2.3 Running unbiased Molecular Dynamics (MD) simulations

The production MD-TIP4PEW, MD-TIP3P, MD-TIP3P-HMR and MD-OPC simulations were initiated from the equilibrated structures and elongated for 5 μs, 8 μs, 10 μs and 5 μs, respectively. At least 5 μs of simulation was found necessary to provide reasonable convergence (Figure 2.1). This was tested by monitoring the agreement between the conformational distributions and also via decrease in Kullback-Leibler divergence (33) (KLD) criteria between the two simulations. The MD-TIP3P and MD-



TIP3P-HMR were simulated for longer to see equivalent results were obtained. These simulations were done in NPT and temperature of 298.15 K, using Langevin thermostat (66,67) with collision frequency of  $5 \text{ ps}^{-1}$  and pressure relaxation time of 5 ps. SHAKE was used to constrain hydrogen-involved bonds. Periodic boundary conditions and particle-mesh Ewald (PME) with 8.0 Å cutoff and default parameters was used. Frames were written to the trajectories every 10 ps. These simulations were done with *pmemd.cuda* of AMBER 14 (65).

#### 2.2.4 Running TREMD simulations

The equilibrated structures at 298.15 K were used to initiate TREMD replicas at different temperatures. The replicas were spaced between 280.00 K and 396.40 K with an online tool, “<http://folding.bmc.uu.se/remd>” (68). Simulations were done with 18 replicas with temperatures set at 280.00, 286.04, 292.18, 298.41, 304.74, 311.17, 317.69, 324.31, 331.03, 337.85, 344.79, 351.82, 358.97, 366.22, 373.59, 381.07, 388.67 and 396.40. This temperature ladder with 1 ps exchange attempt interval resulted in replica exchange probability of 20-30% for all simulations. The 4 fs integration time step facilitated by HMR was used for all TREMD runs. The temperature in each replica was controlled with Langevin thermostat with collision frequency of  $2^{-1} \text{ ps}$ . All TREMD simulations were done in NVT. Periodic boundary conditions and particle mesh Ewald (PME) with 8.0 Å cutoff and default parameters was used and simulations were performed with *pmemd.MPI* of AMBER 14 (65). Simulations were performed for ~500 ns per replica at which time all convergence criteria were satisfied. Roundtrips between the replicas were monitored to ensure that the replicas are, in fact, shuffling between all of the different temperatures.

Additional TREMD-REST simulations were done with dihedral restraints to lock

related dihedral torsions into specified configuration ranges. In TREMD-REST- $\chi$ , 30 kcal/mol-Å<sup>2</sup> restraint was applied to the  $\chi$  torsion of the 3' adenine of ApA to keep it out of the high-anti region ( $-150 < \chi < -30$ ). In TREMD-REST-PUCKER, 30 kcal/mol-Å<sup>2</sup> restraint was used for  $\delta$  torsions of both adenine of ApA to lock the sugars at 3'-endo conformation. The MD input files and restraint definitions are provided in the Supporting Information.

### 2.2.5 Simulation analysis

All of the analyses were done using CPPTRAJ in AmberTools unless otherwise mentioned. The TREMD analyses were done at the 298.41 K replica as this is the closest to room temperature unless the objective was to compare with specific experimental data at other temperatures. To assess convergence of the simulations, the agreement between root mean squared deviations (RMSD) histograms of conformational distributions, the decrease in Kullback-Leibler divergence (KLD) between the principal component histograms from two independent runs (33), and also cluster populations versus time in each run were monitored. The A-form reference structures to which RMSD of coordinates were calculated, were prepared with DFT QM minimization of the 1a sugar-to-sugar suite (15) of each DNMP with restraining backbone torsions at 6-311G\* level (69,70) with Gaussian 09 (71).

Clustering was done on the MD trajectories of TREMD runs at 298.41 K using the *k-means* algorithm (72,73). Different clustering protocols were tried for each DNMP considering coordinates of different atoms and targeting different cluster numbers. Cluster trajectories were quantitatively and qualitatively analyzed in each attempt and protocols that resulted in the minimum number of clusters with the highest

conformational diversity were chosen as optimal clustering methods (Table 2.2).

To calculate the stacking free energies, DNMP structures in each trajectory were classified as either stacked or unstacked based on the definition presented by Brown et al. (49). Accordingly, DNMP structures with minimum distance of 4 Å between any two heavy atoms of the bases, and minimum distance of 5 Å between the center of mass of two bases, and angles between the normal to two bases between 0° and 45° or 135° and 180° were considered as stacked. Alternatively, structures that violate any of these criteria were counted as unstacked. Then the estimated free energies were calculated based on the relative populations with  $\Delta G = -RT \ln (P_s/P_u)$  equation in which  $P_s$  represents the stacked population and  $P_u$  represents the unstacked population. Sugar-to-sugar suites were assigned to the structures in simulation trajectories using *suitname* tool (15) and the results were analyzed using scripts generated in the lab.

## **2.3 Results**

### **2.3.1 Simulation convergence**

We performed unbiased MD simulation for ApA in TIP3P water (8 μs of MD-TIP3P simulations), TIP3P water with hydrogen mass-repartitioning (10+ μs of MD-TIP3P-HMR simulations), TIP4P-EW water (5 μs of MD-TIP4PEW simulations) and OPC water (5 μs of MD-OPC simulations), and each involved three independent sets of MD simulations with different initial conditions. The average mass-weighted heavy atom RMSD histograms of these three independent runs comparing the 2 fs and 4 fs integration time steps with TIP3P show very similar distributions, albeit not perfectly converged (Figure 2.1A). However, we note that the error bars (standard deviations between the three independent runs) are not very small either between the TIP3P runs with different

integration time steps or the runs with the different water models even after the ~5-10 micro-seconds of simulation. Figure 2.1 shows the KLD for each of the first 10 principal component projections between two of the three simulations in each set of simulations. The first 10 principal component modes represent more than 90% of the motions in each simulation. The KLD curves start to flatten after 4  $\mu$ s, however, they never completely flatten to zero which would suggest full convergence and agreement of the modes of motions between the two simulations (Figure 2.1). As an alternative indicator of convergence, analysis of the cluster populations from simulations clustered together with independent calculations of the clusters populations per simulations as a function of time suggests that the cluster populations fluctuate until even 11  $\mu$ s of the ApA simulation in the TIP3P water with HMR (Figure 2.2). These observations suggest that these straight MD simulations are not well-converged even on the ~5-10  $\mu$ s time scale. Table 2.3 also shows the stacking free energies of ApA for these runs according to the criteria defined by Brown et al. (49). The simulations in TIP4P-EW generated stacking energies much different from the previously reported values from MD simulations performed with the same conditions ( $-0.27 \pm 0.05$  vs.  $-1.64 \pm 0.10$  kcal/mol) (49). This suggests that the earlier simulations are not even close to convergence at the 1  $\mu$ s of sampling length performed in the previous work. However, even at 1  $\mu$ s sampling limit, the stacking energy in TIP4P-EW was  $-0.25 \pm 0.13$  kcal/mol which is still very different than the value reported in the work by Brown et al. (49) Beyond differences in sampling, this could be indicative of differences in the applied force fields or MD simulation protocols, noting that multiple AMBER force fields were investigated in preliminary work (unpublished) and none showed stacking values more favorable than ~1 kcal/mol.

The ApA simulations in TIP3P water with and without hydrogen mass-repartitioning

(HMR) show very similar RMSD distributions (Figure 2.1A) and similar stacking free energies (Table 2.3) although better matching within the bounds of the standard deviations would be expected in more converged conditions. Based on the reasonable similarity between the standard and HMR MD simulations, we used HMR in all of our TREMD simulations to speed up the simulations.

In contrast to the unbiased MD simulations, the TREMD simulations converged rather quickly. The KLD of the principal component histograms between two independent runs show little divergence by ~300-500 ns simulation time per replica (Figure 2.3). Accordingly, we used TREMD with HMR as a method for converging the conformational distributions of other DNMPs and with other water models and force fields.

All of the DNMPs were simulated with 18 replicas in TREMD simulations in TIP3P water using the ff14 force field. We also used five other water model and force field combinations as described in the Methods for ApA, ApC, CpA, CpC, UpU, GpG and ApG.

### 2.3.2 Structural properties of DNMPs in TREMD simulations

We simulated all DNMPs with 18 replicas TREMD in TIP3P water using ff12 force field. We also used 5 other water model/force field scenarios described in the Methods for ApA, ApC, CpA, CpC, UpU, GpG and ApG to enrich the conformational ensembles and to assess the performance of each scenario in terms of shifting the conformational distributions.

### 2.3.2.1 Stacking free energies

Table 2.4 summarizes the stacking free energies calculated from the 298.41 K replica based on the criteria defined by Brown et al. (49) In all of the tested DNMPs, Chen-Garcia (CG) modifications to ff99 and CHARMM36 simulations result in less stacking compared to the other simulations. Additionally, the vdW<sub>bb</sub> modifications slightly increase the stacked populations relative to the ff14 with both the TIP3P and OPC water models. The OPC water model usually results in slight increase of the stacked population with few exceptions. Occurrence of guanine in the Xp- position results in the highest stacking ratios, which is not expected from the experimental trend of stacking free energies reported by Frechet et al. (48) (Table 2.4) The stacking free energy values generated with the TIP3P/ff14 combination correlate with these experimental values with  $r^2=0.62$  correlation coefficient. This is higher correlation than what was estimated in the earlier work ( $r^2=0.51$ ) by Brown et al. (49).

### 2.3.2.2 Sugar conformation

The conformations of the sugar moieties were described with the  $\delta$  dihedral torsion. The  $\delta$  values between  $78^\circ$  and  $90^\circ$  were considered as C3'-endo sugar and  $\delta$  values between  $140^\circ$  and  $152^\circ$  were considered as C2'-endo sugar (Figure 2.4) (15). Different modifications of AMBER force fields result in similar sugar pucker distributions (Figure 2.5). The CHARMM36 results in significantly higher C3'-endo conformation for the Xp-residues. It also results in higher C2'-endo conformation for -pX residues. However, the Xp- residues in all DNMPs conform C3'-endo conformation in much higher populations than C2'-endo in all tested water / force field combinations; and in contrast, the -pX residues show relatively similar C3'endo and C2'-endo populations with slight preference

of one over the other depending on the DNMP and scenario. This agrees with Lee et al. (37) in which the C3'-endo is assigned as the dominant ribose conformation in all tested DNMPs (ApA, ApG, GpA, UpU, CpC, UpC and CpU) (37). They also observed that the shifting of the sugar pucker distributions of purine-purine DNMPs to C3'-endo upon their dimerization is more significant for the Xp- residues although ribose in both residues shift from C2'-endo to C3'-endo. All tested cases seem to qualitatively agree with this observation, although the augmented C2'-endo population in ApA and ApG which is generated by CHARMM36 contradicts with it. However, another NMR study by the same group on ApA has quantitatively reported the C2'-endo:C3'-endo ratio to shift from 70:30 to 50:50 in both residues upon dimerization (74). According to this observation, one would consider the Xp- residue sugar as too-rigidly-locked in the C3'-endo conformation, an artifact that is deteriorated by using CHARMM36.

### 2.3.2.3 $\chi$ dihedral torsions

The torsion around the glycosidic bond ( $\chi$ ) shows different distributions in Xp- vs. -pX positions and also show dependencies based on the base identity (Figure 2.6). At Xp- position, the syn orientation is highly populated for purines. This is reduced for pyrimidines with an increased population of high-anti orientation ( $-60^\circ$  to  $-150^\circ$ ). Increased anti orientations for Xp- pyrimidines qualitatively agrees with some experimental reports.(36) On the other hand, a proton relaxation time study has proposed dominantly syn orientation for  $\chi$  of the Xp- position (75). In our simulations, the syn orientation of  $\chi_{Xp-}$  in purines coincides with short distance between the 5' hydroxyl (HO5') and N3 atom on the base (Figure 2.7). So we hypothesized that this interaction stabilizes the syn orientation of  $\chi_{Xp-}$ . At -pX position, both purines and pyrimidines show

significant high-anti populations although syn orientation is also populated in purines. Different water / force field combinations show relatively similar  $\chi$  distributions except that the TIP3P / ff99 + CG results in less high-anti orientation in purines at the -pX position (Figure 2.8).

### 2.3.3 DNMP overall conformations

Figure 2.9 shows the conformational distributions of the DNMP ensembles generated at 298.41 K in TREMD simulations with TIP3P water and the ff14 force field. Figure 2.10 compares the distributions for a selection of DNMPs simulated with other water / force field combinations. We used k-means algorithm (72,73) to cluster these ensembles into different conformations. The number of target clusters (five to seven) and the considered atomic mask in clustering were tuned so that the lowest number of clusters with the highest diversity is obtained (Table 2.2). When different water model / force field combinations were tested, combined clustering was done on all trajectories generated with these water / force field combinations. Five distinct conformational clusters were obtained for most DNMPs which were named as *A-form*, *ladder*, *sheared*, *inverted* and *extended*. All cluster trajectories were post-processed to identify their unique characteristics and intermolecular interactions. Table 2.5 indicates the average mass-weighted heavy atom RMSD of each cluster trajectory, which approximately identifies the position of each conformation on Figure 2.9 and Figure 2.10. The same table also represents the average relative population of sugar-to-sugar suite outliers. Figure 2.11 shows the representative structures of these clusters for ApA. In fact, some dinucleotides resulted in more than one extended clusters as expected because of the flexibility of DNMPs in the extended form. However, all the extended forms are reported as one



conformation. Table 2.6 reports the population of different conformations generated for each DNMP. Sheared conformation was not detected as a distinct cluster for pyrimidine-pyrimidine DNMPs.

### 2.3.3.1 *A-form* conformation

The structures in the A-form cluster have similar attributes to the A-form RNA. This conformation is mainly characterized with g(-)-g(-) conformation of  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  (Figure 2.12). The average mass-weighted heavy atom RMSD to the QM-minimized A-form ( $1.17 \text{ \AA} \pm 0.21$  according to the Table 2.5) and the smallest average relative suite outlier population ( $12.67\% \pm 2.69$  according to Table 2.5) of this cluster over all DNMPs are smallest among all clusters. The ribose conformation of the Xp- residue is C3'-endo while that of the -pX residue is in equilibration between C3'-endo and C2'-endo with preference for the C3'-endo conformation (Figure 2.13).

### 2.3.3.2 *Ladder* conformation

The average mass-weighted heavy atom RMSD to the QM-minimized A-form structure for the ladder clusters over all DNMPs is  $1.59 \text{ \AA} \pm 0.33$  which is very close to that of the A-form clusters (Table 2.5). Ladder conformation also shares the same g(-)-g(-) conformation of  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  with the A-form (Figure 2.12). However, the ladder conformation backbone is different from the A-form with the  $\epsilon_{Xp-}$  torsion changed toward the anti orientation which is accompanied by close interaction between HO2'<sub>Xp-</sub> and O5'<sub>-pX</sub> atoms (Figure 2.14). Therefore, we hypothesized that this interaction stabilizes the ladder conformation. Also  $\chi_{Xp-}$  and  $\chi_{-pX}$  adopt syn and high-anti orientations in the ladder conformation, respectively. Not only in ladder clusters, but in sheared and inverted

clusters, the orientation of  $\chi$  is different based on the purine and pyrimidine identification of the residues (Figure 2.15). Distortion of  $\chi$  from the anti orientation is more severe when purines are present at either positions while pyrimidines prefer the anti and high-anti  $\chi$  configurations. The sugar conformations are similar to the A-form conformation except that there is small population of C2'-endo conformation of the Xp- residue (Figure 2.12).

### 2.3.3.3 Sheared conformation

This conformation which is missing in pyrimidine-pyrimidine DNMPs drives the base away from the stacking position. The average mass-weighted heavy atom RMSD to the QM-minimized A-form structure over all sheared clusters is  $2.43 \text{ \AA} \pm 0.16$ . 43.24%  $\pm 5.66$  of the sheared populations are sugar-to-sugar suite outliers (Table 2.5). A common distinguishing characteristic of structures in sheared conformation is their severe high-anti to g(-)  $\epsilon_{Xp-}$  relative to the moderate high-anti  $\epsilon_{Xp-}$  of the A-form conformation. This is correlated with population of C2'-endo sugar pucker of the Xp- residue (Figure 2.16). The  $\zeta_{Xp-}$  and  $\alpha_{pX}$  combination is diverse in sheared cluster populations with no clear base dependency (Figure 2.12). However, as for the ladder conformation, the  $\chi_{Xp-}$  and  $\chi_{-pX}$  adopt syn and high-anti orientations with some purine-vs-pyrimidine dependency (Figure 2.15). Regarding population of the C2'-endo sugar conformation of the Xp- position and significant shift of the  $\chi$  in the -pX position to high-anti orientation, we hypothesized that these factors regulate population of the sheared clusters. The sugar conformation of the -pX position is in equilibrium between C3'-endo and C2'-endo (Figure 2.12).

#### 2.3.3.4 *Inverted* conformation

The average heavy atom mass-weighted RMSD of the inverted clusters to the QM-minimized A-form has the largest value among the four conformations ( $2.78 \text{ \AA} \pm 0.15$  according to the Table 2.5). These clusters also contain high amount of sugar-to-sugar suite outlier structures ( $39.19\% \pm 8.10$  according to Table 2.5). The common characteristic of the structures with inverted conformation is anti orientation of the  $\epsilon_{Xp-}$ , which is similar to the ladder conformation (Figure 2.17). However, these structures are more clearly distinguished with their g(+)-g(+) combination of  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  torsions (Figure 2.12). The g(+) orientation of  $\zeta_{Xp-}$  correlates with anti orientation of  $\epsilon_{Xp-}$  and close interaction between  $\text{HO}2'_{Xp-}$  and  $\text{OP}1_{-pX}$  (Figure 2.17). This is in contrast with the ladder conformation in which the anti orientation of  $\epsilon_{Xp-}$  coincides with close interaction of  $\text{HO}2'_{Xp-}$  with  $\text{O}5'_{-pX}$  while  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  are in g(-) configuration (Figure 2.14 and Figure 2.12). Therefore,  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  seem to play the key role in changing the orientation of the residues relative to each other and forming the inverted conformation, an observation which is in agreement with experimental data (37). Based on these correlations, we hypothesized that the  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  provide degrees of freedom needed for formation of the inverted conformation and the interaction between  $\text{HO}2'_{Xp-}$  and  $\text{OP}1_{-pX}$  stabilizes these torsions at g(+)-g(+) orientation which leads to stabilization of the inverted conformation. The ribose conformation of the  $Xp-$  residue is C3'-endo in inverted clusters, while that of the  $-pX$  residue is in equilibration between C3'-endo and C2'-endo (Figure 2.12).

### 2.3.4 Stabilizing factors of different conformations

To test our hypotheses about the stabilizing factors of each DNMP conformation, we set up TREMD simulations of ApA at five different conditions in TIP3P water (TREMD-NEUT, TREMD-MOD and TREMD-REST simulations in Table 2.1). These simulations included ApA with no excess of NaCl (TREMD-NEUT), ApA with neutralized HO2' atom (TREMD-MOD-O2'), ApA with neutralized HO5' atom (TREMD-MOD-O5'), ApA with restrained  $\chi_{pX}$  (TREMD-RREST- $\chi$ ) and ApA with restrained  $\delta$  of both residues (TREMD-REST-PUCKER). To set up TREMD-NEUT, just one single Na<sup>+</sup> was added to the simulation box to neutralize the system and no extra ions were added. In TREMD-MOD-O2', the partial charge on HO2' atoms was transferred to O2' to retain the total charge of the molecule while neutralizing the HO2' atoms and knocking down its non-bonding interactions with other atoms. This was done to elucidate the role of HO2' interactions in stabilizing the ladder and inverted conformations. In TREMD-MOD-O5', the same strategy was used to neutralize HO5' atom to investigate the effect of its interaction with N3 on the  $\chi_{Xp}$  orientation. In TREMD-REST- $\chi$ ,  $\chi_{pX}$  was restrained from the high-anti orientation. In TREMD-REST-PUCKER,  $\delta$  of both residues were restrained to lock the sugar at C3'-endo conformation. The REMD-REST- $\chi$  simulations were specifically set up to understand the effect of  $\chi_{Xp}$  orientation and sugar conformation on the population of the sheared conformation. In all of these simulations, ff14 force field was used. Combined clustering was done on the ensembles resulted from these simulations and the simulation of ApA in regular conditions (TREMD-TIP3P-FF14) at 298.41 K. The effects of the restraints on the targeted properties are illustrated in the Figure 2.18.

Figure 2.19 shows the RMSD distribution of the ApA in the above-mentioned

TREMD simulations in addition to the unmodified and unrestrained TREMD simulation (TREMD-TIP3P-FF14). As it indicates, eliminating the excess NaCl does not change the distribution at all which implies that this much of excess NaCl (150 mM) does not contribute to the relative stability of any conformation. In contrast, the most severe change in conformational distribution seems to happen when the HO2' interactions are knocked down (TREMD-MOD-O2'). Neutralizing the HO2' atoms results in a dramatic increase in the A-form population and decrease in the ladder and inverted conformations (Table 2.7). This observation supports our hypotheses that the interaction between HO2'<sub>xp-</sub> and O5'<sub>-pX</sub> stabilizes the ladder conformation and that the interaction between HO2'<sub>xp-</sub> and OP1<sub>-pX</sub> stabilizes the inverted conformation.

Knocking down the HO5' interactions coincides with a significant shift of the  $\chi_{xp-}$  from syn to anti and high-anti orientations (Figure 2.20A). This supports our hypothesis that the interaction between this atom and the N3 atom on the base stabilizes the syn conformation of  $\chi_{xp-}$  as the distance between these atoms increase simultaneously (Figure 2.20B). Interestingly, neutralized HO5' significantly decreases populations of the inverted and sheared conformations and increases population of the A-form conformation. According to our analysis of trajectories, HO5' either interacts with N3<sub>xp-</sub> or the OP2<sub>-pX</sub> and the latter interaction is more frequent in the A-form clusters (data not shown). Both of these interactions are significantly decreased when HO5' is neutralized. Therefore, increasing the population of the A-form conformation should be an indirect result of HO5' neutralization and through the major shift of  $\chi_{xp-}$  from syn orientation, or silencing the HO5' interaction with OP2<sub>-pX</sub>. The latter is less probable since our analysis shows that such interaction is not less frequent in the A-form cluster of the ApA simulations relative to other clusters (data not shown). Therefore lack of this interaction

does not seem to contribute to the increase of the A-form conformation when HO5' is neutralized. Accordingly, the effect of HO5' neutralization on populations is through changing the  $\chi_{\text{pX}}$ . This implies that the conformation of the glycosidic bond affects the configuration of the backbone dihedrals and consequently contributes to changing the population of conformational clusters.

Restraining the  $\chi_{\text{pX}}$  from high-anti orientation (TREM-D-REST- $\chi$ ) significantly decreases the sheared and ladder conformations (Table 2.7). This happens with the cost of increasing the inverted conformation (Table 2.7). This reconfirms that the high-anti orientation of  $\chi_{\text{pX}}$  is a significant characteristic of the sheared and ladder conformations. Also locking the sugars at C3'-endo conformation decreases the sheared and extended conformations dramatically with no significant change in the ladder population (Table 2.7). This highlights the importance of pucker flexibility in formation of the sheared conformation as distinguished from the ladder conformation. Also, pucker flexibility plays a role in unfolding these four conformations to extended forms.

## **2.4 Conclusions**

In this work, we investigated the conformational diversity of RNA dinucleotide monophosphates (DNMPs) using molecular dynamics. Our convergence analysis showed that the full convergence of the conformational distribution of these small molecules might be attained with unbiased molecular dynamics only after several microseconds of sampling. Therefore, relying on sub-microsecond simulations of this kind for judging various properties of DNMPs such as stacking free energies, j-couplings and distribution of dihedral angles would be misleading. In contrast, we applied a TREMD protocol with 18 replicas that could fully converge the structural ensemble of DNMPs in ~500 ns per

replica of sampling. This simulation protocol provides a basis for objective assessment of characteristics of these molecules and could be used in future to assess new force field modifications in terms of shifting the conformational populations and reproducing experimental data. In fact, the effect of new modifications on DNMP conformations can be used for anticipating their effects on larger RNA molecules in a quick manner.

Using the TREMD protocol in TIP3P water and with ff14 AMBER force field, we distinguished four consensus conformations for all DNMPs in addition to their expected extended form. Using different water / force fields combinations did not add to these general conformations, but rather shifted the population distribution among them. We name these common conformations as A-form, ladder, sheared and inverted. Although previous experimental results have predicted certain characteristics of some of these conformations in some DNMPs (37-41), this work showed that they are consensus among all DNMPs regardless of their base identification. This comes with the exception that pyrimidine-pyrimidine DNMPs do not generate the sheared conformation.

In summary, different combinations of  $\epsilon_{Xp-}$ ,  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  torsions facilitate formation of different conformations. The  $\epsilon_{Xp-}$  would range from moderate high-anti values in A-form conformation to anti in ladder and inverted conformations and g(-) in the sheared conformation. The  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  determine the overall orientation of residues relative to each other and distinguish the inverted conformation from the ladder and A-form Figure 2.12. The  $\chi_{Xp-}$  torsion mostly adopts syn configuration in all clusters except the A-form. The high-anti  $\chi_{-pX}$  is highly populated in the sheared and ladder conformations.

Although identifying the same conformations with using different water / force field combinations minimizes the chance of missing any significant conformation of DNMPs, reality of the identified conformations and their relative populations should be judged

with caution. The results of this work show that the populations of these conformations are relatively sensitive to the choice of the force fields. Specifically, CHARMM36 and ff99 + CG generate very different distributions than ff14 (Figure 2.8). Therefore, this research would be well complemented with future experimental assessments that selectively and quantitatively target each of the identified conformations with the aim of identifying the conformations with little or no chance of existence *in-vitro*.

In spite of the lack of extensive quantitative experimental data for assessment of the identified conformations of DNMPs, there are still clues in experimental results in favor or disfavor of these conformations. The presence of the A-form like conformation has been reported by C. Lee et al and M. Tsuboi (37,38). The inverted conformation has been observed by the same group with an NMR approach as well as by J. Rubin et al. (39,40) as a crystal structure of UpA. This is while the inverted conformation clusters contain relatively high sugar-to-sugar outliers (Table 2.5). This might be reasonable for this conformation because occurrence of the inverted conformation with g(+)-g(+) conformation of the  $\zeta_{Xp-}$  and  $\alpha_{-pX}$  in the middle of a sequence would destroy helicity of the sequence. Hence, this local conformation might occur in larger sequences only in turns and loops while it can be widely adopted by DNMPs. In both A-form and inverted structures, the bases are in stacked orientation relative to each other and the sugar conformation of the Xp- residue adopts C3' endo conformation. This agrees with NMR observations that the base stacking correlates with increasing the C3'-endo sugar conformation (37).

To the best of our knowledge, the sheared and ladder conformations have not been described in previous experimental studies for DNMPs. These conformations have the common  $\chi_{-pX}$  high-anti characteristic. Preventing this dihedral angle from high-anti



orientation significantly decreases the populations of these conformations (Table 2.7). There are some reports claiming that the  $\chi$  conforms equilibrated amounts of syn and anti orientations with preference for the anti in nucleosides and nucleotides (76,77). One study reports it as dominantly syn in Xp- positions (75). However, a recent NMR study of dTpT shows that the  $\chi$  of both residues adopt anti orientation in this DNMP (78). This leaves us with uncertainty to judge the simulation outcomes for  $\chi$  based on the experimental data. On the other hand, high-anti configuration of the glycosidic bond has been observed in ladder-shaped conformations in other RNA simulations which are proven to be force field artifacts (10,53,79). Accordingly, we suspect that the formation of local ladder conformation with high-anti  $\chi$  might initiate the formation of ladder-shaped conformations in simulations of larger RNA molecules and therefore the DNMP ladder conformation might be the results of force field artifacts as well.

From a biochemical perspective, our results show how presence of the 2'-hydroxyl on ribose can potentially facilitate formation of different local conformations in RNA. Non-bonding interactions with this hydroxyl group play crucial roles in stabilizing the DNMP conformations. Specifically, the interaction between HO2'<sub>Xp-</sub> and O5'<sub>-pX</sub> is frequent in the ladder conformation (Figure 2.14) and the interaction between HO2'<sub>Xp-</sub> and OP1<sub>-pX</sub> stabilizes the inverted conformation (Figure 2.17). Neutralizing the HO2' atom significantly decreases these conformations through knocking down interactions with 2'-hydroxyl (Table 2.7). Also from a computational chemistry perspective, our results highlight the sensitivity of the RNA conformational ensembles to the force field parameters that govern the non-bonding interactions between this hydroxyl and other hydrogen bond acceptors. Our results show that not only neutralizing the HO2' has a dramatic effect on the conformation populations, but changing the vdW parameters of the

phosphate oxygens as done in vdW<sub>bb</sub> modifications can slightly shift the populations (Figure 2.8). However, changing the solvent did not have a significant effect on the populations of DNMPs (Figure 2.1 and Figure 2.8).

Knocking down the HO5' interactions affects the DNMP conformational populations as well. Because the HO5' only exists in the 5' end of a sequence, one would expect that its interactions with other atoms would have fewer significant effects on larger sequences. However, since we believe that this effect occurred through changing the  $\chi_{xp}$ -distribution, we expect that  $\chi$  parameters will play important roles in determining the conformation of DNMPs and larger RNA sequences.

With the TREMD approach that we used in this work, DNMPs are easy-to-converge molecules with which new RNA force field modifications can be quickly assessed. With this method and using some of the RNA current force fields, we identified the DNMP conformations. The conformational elements of the ensemble qualitatively agree with the available experimental data which predict the conformational ensemble of DNMPs to be relatively flexible and equilibrium of different conformations. However, new experimental data that selectively targets crucial features of these conformations will confirm or rule out certain identified conformations. In fact, tuning the MD simulations with more experimental evidence will reveal valuable conformational information about the DNMP molecules at atomistic details, and also will prepare the simulation tools to answer our questions about larger RNA sequences.

Table 2.1. Summary of simulations described in this chapter.

Name	Water model	Force field	Time step	Runs	System
MD-TIP4PEW	TIP4P-EW (59)	FF12 <sup>a</sup>	2 fs	3 <sup>b</sup>	ApA
MD-TIP3P	TIP3P (56)	FF12 <sup>a</sup>	2 fs	3 <sup>b</sup>	ApA
MD-TIP3P-HMR	TIP3P	FF12 <sup>a</sup>	4 fs	3 <sup>b</sup>	ApA
MD-OPC	OPC (55)	FF12 <sup>a</sup>	2 fs	3 <sup>b</sup>	ApA
TREMD-TIP3P-FF12	TIP3P	FF12 <sup>a</sup>	4 fs	2 <sup>c</sup>	All DNMPs
TREMD-TIP3P-VDW	TIP3P	FF12 <sup>a</sup> + vdW <sub>bb</sub> <sup>d</sup>	4 fs	2 <sup>c</sup>	Selection of DNMPs <sup>e</sup>
TREMD-TIP3P-CG	TIP3P	ff99+CG <sup>f</sup>	4 fs	2 <sup>c</sup>	Selection of DNMPs <sup>e</sup>
TREMD-TIPS3P-C36	TIPS3P <sup>g</sup>	CHARMM36 <sup>h</sup>	4 fs	2 <sup>c</sup>	Selection of DNMPs <sup>e</sup>
TREMD-OPC-FF12	OPC (55)	ff12 <sup>a</sup>	4 fs	2 <sup>c</sup>	Selection of DNMPs <sup>e</sup>
TREMD-OPC-VDW	OPC	ff12 <sup>a</sup> + vdW <sub>bb</sub> <sup>d</sup>	4 fs	2 <sup>c</sup>	Selection of DNMPs <sup>e</sup>
TREMD-NEUT	TIP3P	ff12 <sup>a</sup>	4 fs	2 <sup>c</sup>	ApA
TREMD-MOD-O2'	TIP3P	ff12 with The electrostatic charge on HO2' atom transferred to the O2'	4 fs	2 <sup>c</sup>	ApA
TREMD-MOD-O5'	TIP3P	ff12 with The electrostatic charge on HO5' atom transferred to the O5'	4 fs	2 <sup>c</sup>	ApA
TREMD-REST- $\chi^i$	TIP3P	ff12 <sup>a</sup>	4 fs	2 <sup>c</sup>	ApA
TREMD-REST-PUCKER <sup>j</sup>	TIP3P	ff12 <sup>a</sup>	4 fs	2 <sup>c</sup>	ApA

<sup>a</sup> ff99 (51) + parmbse0  $\alpha/\gamma$  (52) +  $\chi$ OL3 (53) modifications

<sup>b</sup> Independent runs prepared with randomizing ion positions.

<sup>c</sup> Independent runs prepared with different initial DNMP structures and randomizing ion positions.

<sup>d</sup> Bioorganic phosphate vdW radii modifications (54) were applied to backbone

Table 2.1 continued

---

oxygen atoms (O3', O5', OP1 and OP2).

<sup>e</sup> ApA, ApC, CpA, CpC, UpU, GpG and ApG

<sup>f</sup> vdW modifications were applied to nucleobase and nucleobase heavy atoms-water interactions plus modifications in  $\chi$  torsion (14).

<sup>g</sup> TIP3P water model implemented in CHARMM which has vdW parameters for water hydrogen

<sup>h</sup> CHARMM27 (57) + 2'-hydroxyl modification (58)

<sup>i</sup> Restrained simulations in which  $\chi_{\text{-px}}$  was prevented from high-anti orientation with restraint.

<sup>j</sup> Both sugar puckers were locked in C3'-endo conformation with restraint.

---

Table 2.2. Atomic mask and number of targeted cluster for clustering DNMP simulations

DNMP	Atomic mask	Number of targeted clusters in k-means algorithm
ApA	All heavy atoms	7
ApG	All heavy atoms	6
GpA	All heavy atoms	6
GpG	All heavy atoms	6
ApC	All heavy atoms	7
ApU	All heavy atoms	5
GpC	C5',C4',C3',O3',P,C2	6
GpU	All heavy atoms	6
CpA	All heavy atoms	5
CpG	C5',C4',C3',O3',P,C2	6
UpA	All heavy atoms	7
UpG	All heavy atoms	6
CpC	All heavy atoms	5
CpU	C5',C4',C3',O3',P,C2	5
UpC	All heavy atoms	6
UpU	All heavy atoms	5

---

Table 2.3. Average stacking free energies of ApA in different water models with regular MD. The error shows the standard deviation between three independent runs.

Simulation	Stacking free energy (kcal/mol)
TIP3P	$-0.22 \pm 0.02$
TIP3P HMR	$-0.18 \pm 0.02$
TIP4P-EW	$-0.27 \pm 0.05$
OPC	$-0.14 \pm 0.08$

Table 2.4. Average stacking free energies of DNMPs. The error shows the standard deviation between two independent runs. The experimental values are according to the work by Frechet et al. (48).

DMNP	Experiment	TIP3P / (TIPS3P)				OPC		
		ff12	ff12 + vdW <sub>bb</sub>	ff99+ CG	CHARMM36	ff12	ff12 + vdW <sub>bb</sub>	
ApA	-0.24 ± 0.02	-0.28 ± 0.01	-0.31 ± 0.02	0.00 ± 0.02	0.11 ± 0.01	-0.24 ± 0.02	-0.33 ± 0.02	
ApG	0.05 ± 0.11	-0.47 ± 0.02	-0.47 ± 0.02	-0.23 ± 0.00	-0.35 ± 0.01	-0.48 ± 0.02	-0.50 ± 0.00	
GpA	0.04 ± 0.71	-0.61 ± 0.00	-	-	-	-	-	
GpG	-0.01 ± 0.03	-0.80 ± 0.01	-0.89 ± 0.03	-0.34 ± 0.01	-0.30 ± 0.05	-0.87 ± 0.02	-0.95 ± 0.02	
ApC	0.15 ± 0.11	-0.26 ± 0.02	-0.39 ± 0.01	-0.08 ± 0.02	0.05 ± 0.01	-0.38 ± 0.01	-0.56 ± 0.03	
ApU	0.27 ± 0.11	-0.29 ± 0.04	-	-	-	-	-	
GpC	0.03 ± 1.88	-0.67 ± 0.01	-	-	-	-	-	
GpU	0.39 ± 0.10	-0.64 ± 0.04	-	-	-	-	-	
CpA	0.24 ± 0.12	-0.13 ± 0.01	-0.25 ± 0.03	0.46 ± 0.02	0.30 ± 0.00	-0.17 ± 0.00	-0.29 ± 0.00	
CpG	-0.08 ± 0.08	-0.36 ± 0.01	-	-	-	-	-	
UpA	0.06 ± 0.22	-0.23 ± 0.01	-	-	-	-	-	
UpG	0.39 ± 0.01	-0.29 ± 0.02	-	-	-	-	-	
CpC	0.39 ± 0.45	-0.31 ± 0.02	-0.50 ± 0.03	0.15 ± 0.01	0.18 ± 0.03	-0.45 ± 0.02	-0.59 ± 0.01	
CpU	0.19 ± 0.07	-0.37 ± 0.01	-	-	-	-	-	
UpC	0.53 ± 0.21	0.00 ± 0.03	-	-	-	-	-	
UpU	1.57 ± 0.45	0.15 ± 0.00	-0.09 ± 0.00	0.37 ± 0.01	1.01 ± 0.01	0.12 ± 0.01	-0.09 ± 0.02	

Table 2.5. Mass-weighted heavy atom RMSD and relative population of suite outlier frames averaged over cluster trajectories of all DMNPs for each cluster. Average RMSD values can be used to locate the approximate conformation in distribution plots of Figure 2.9 and Figure 2.10.

Conformational cluster	Average RMSD (Å)	Average suite outliers (%)
A-form	$1.17 \pm 0.21$	$12.67 \pm 2.59$
Ladder	$1.59 \pm 0.33$	$15.44 \pm 4.83$
Sheared	$2.43 \pm 0.16$	$43.28 \pm 5.66$
Inverted	$2.78 \pm 0.15$	$39.19 \pm 8.10$
Extended	$3.58 \pm 0.45$	$43.04 \pm 8.82$

Table 2.6. Relative populations of conformations generated for each DNMP in TREMD simulations at 298.41 K replica. For ApA, ApC, CpA, CpC, UpU, GpG and ApG, trajectories from different water model / force field scenarios were used, while for other DNMPs trajectories from TIP3P / ff12 were only used.

DMNP	A-form (%)	Ladder (%)	Sheared (%)	Inverted (%)	Extended (%)
ApA	18.7	29.1	8.5	27.4	16.3
ApG	16.6	17.8	19.9	38.2	7.5
GpA	31.4	23.8	7.2	34.2	3.4
GpG	34.1	17.6	8.0	31.8	8.5
ApC	38.4	14.3	7.2	18.2	21.9
ApU	41.1	17.5	10.0	19.9	11.5
GpC	47.3	17.3	4.1	24.4	7.0
GpU	56.5	17.4	4.7	12.7	8.6
CpA	24.2	28.7	14.8	12.9	19.4
CpG	40.2	21.7	17.7	12.0	8.3
UpA	50.9	13.5	18.5	6.5	10.6
UpG	49.5	18.8	19.1	5.7	7.1
CpC	45.5	22.4	-	11.3	20.8
CpU	55.9	18.6	-	7.5	18.1
UpC	27.1	36.1	-	11.0	25.7
UpU	21.1	37.8	-	15.1	25.9

Table 2.7. Relative population of different ApA conformations in different TREMD simulations at 298.41 K. TREMD-TIP3P-FF12 is unmodified and unrestrained simulation in TIP3P with ff12 with 150 mM excess NaCl. TREMD-NEUT is the same with no excess NaCl. In TREMD-MOD-O2' and TREMD-MOD-O5', HO2' and HO5' are neutralized respectively to knock-down their interactions with other atoms. In TREMD-REST- $\chi$ , the  $\chi_{\text{pX}}$  is prevented from adopting high-anti orientation (a range of  $-150^\circ$  to  $-30^\circ$ ). In TREMD-REST-PUCKER, both sugars are locked at the C3'endo conformation with restraining the  $\delta$  torsions.

Simulation	A-form (%)	Ladder (%)	Sheared (%)	Inverted (%)	Extended (%)
TREMD-TIP3P-FF12	$11.26 \pm 0.11$	$33.01 \pm 0.86$	$13.78 \pm 0.30$	$36.58 \pm 0.58$	$5.38 \pm 0.12$
TREMD-NEUT	$11.26 \pm 0.30$	$33.44 \pm 0.35$	$14.05 \pm 0.63$	$35.85 \pm 0.42$	$5.40 \pm 0.15$
TREMD-MOD-O2'	$40.57 \pm 0.45$	$18.75 \pm 0.07$	$16.30 \pm 0.52$	$18.19 \pm 0.52$	$6.21 \pm 0.52$
TREMD-MOD-O5'	$30.42 \pm 0.34$	$30.96 \pm 0.38$	$6.86 \pm 0.28$	$27.31 \pm 0.24$	$4.46 \pm 0.30$
TREMD-REST- $\chi$	$9.94 \pm 0.43$	$18.31 \pm 0.93$	$9.56 \pm 0.21$	$56.52 \pm 1.22$	$5.69 \pm 0.07$
TREMD-REST-PUCKER	$13.24 \pm 0.36$	$31.85 \pm 0.35$	$5.22 \pm 0.12$	$47.06 \pm 0.13$	$2.64 \pm 0.03$



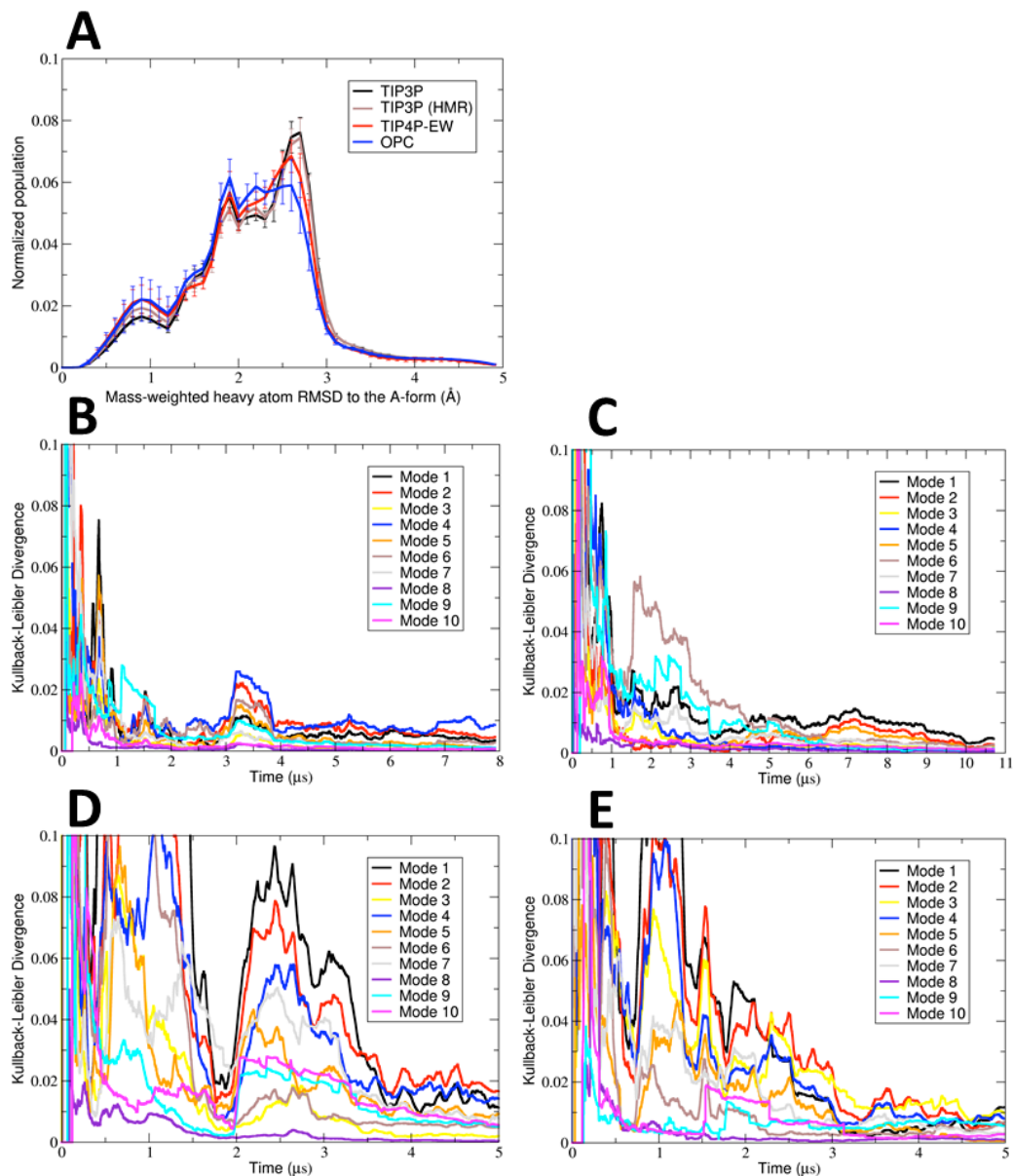


Figure 2.1. A) Distributions of the mass-weighted heavy atom RMSD to QM-minimized A-form for ApA in unbiased simulations with ff12 force field and different water models. The error bars represent the standard deviations between three independent runs. B) Kullback-leibler divergence (KLD) of the histograms of the principal component projections for the first ten principal components for two random unbiased simulations of ApA in TIP3P. C) The same for simulations in TIP3P with HMR. D) The same for simulations in TIP4P-EW. E) The same for simulations in OPC.

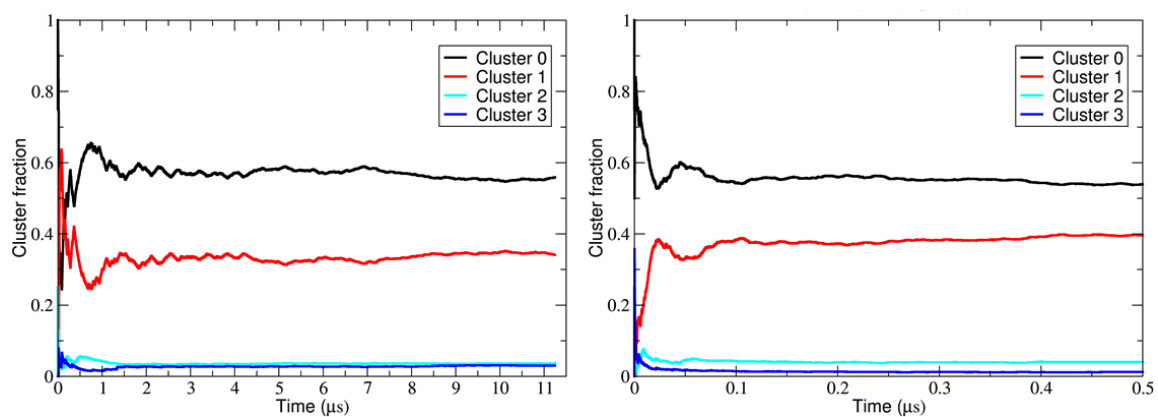


Figure 2.2. Left) Relative populations of the first four major clusters generated by an average linkage method for a random copy of the ApA simulation with regular MD in TIP3P and hydrogen mass repartitioning. Right) The same for the TREMD simulation.

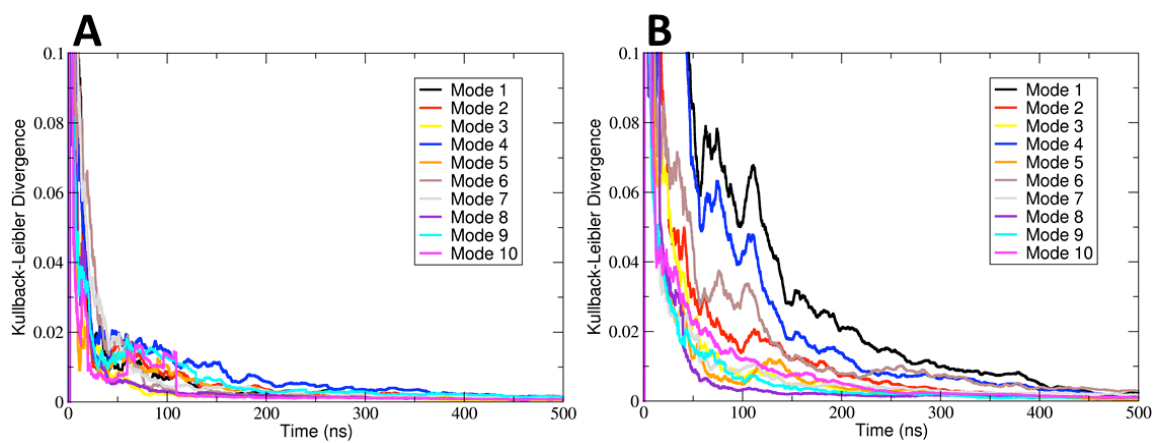


Figure 2.3. A) Kullback-leibler divergence (KLD) of the histograms of the principal component projections for the first ten principal components for two independent TREMD simulations of ApA in TIP3P. B) The same for simulation in OPT.

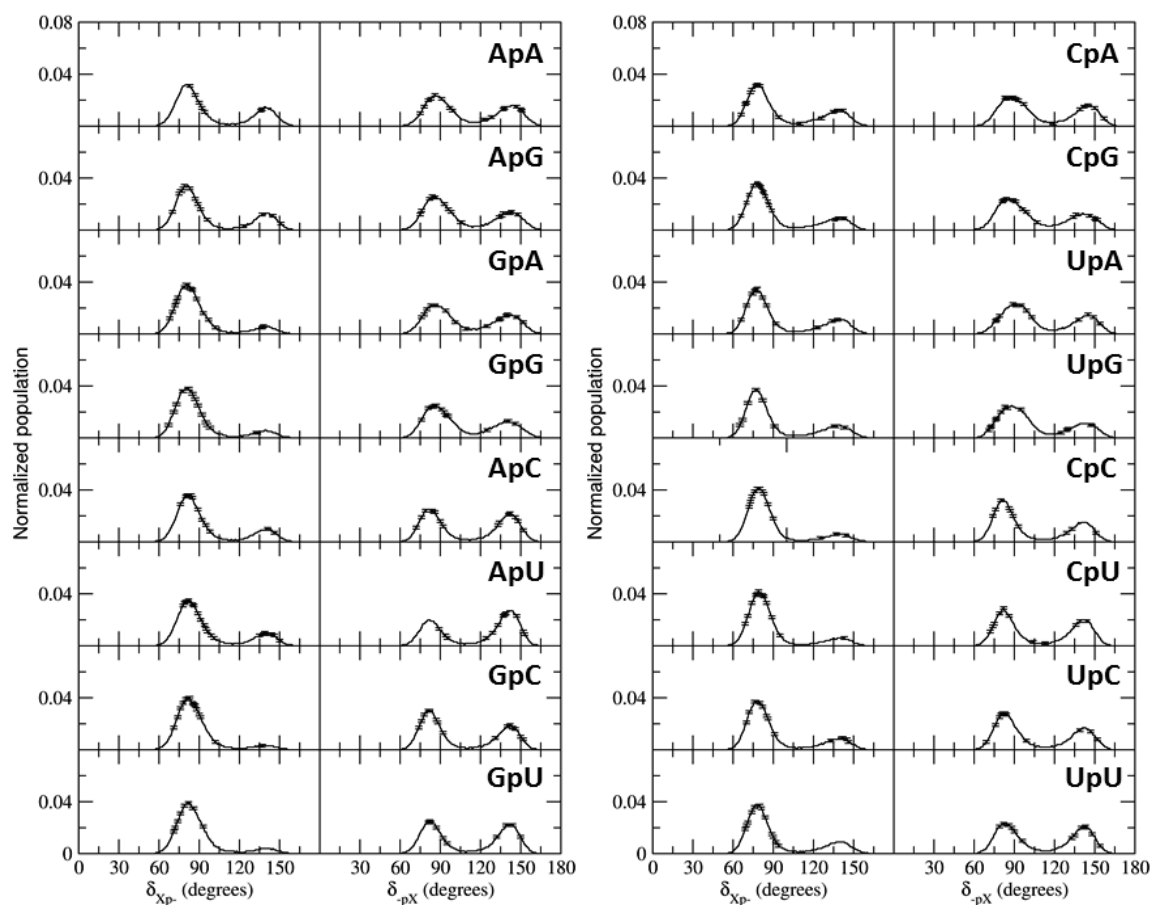


Figure 2.4. Histograms of  $\delta$  dihedral torsions for two residues of all DNMPs in TIP3P water and with ff14 force field at 292.18 K. The peaks at  $< 90^\circ$  represent C3'-endo and the peaks at  $> 140^\circ$  represent the C2'-endo sugar conformations. The error bars represent the standard deviation between two independent runs.

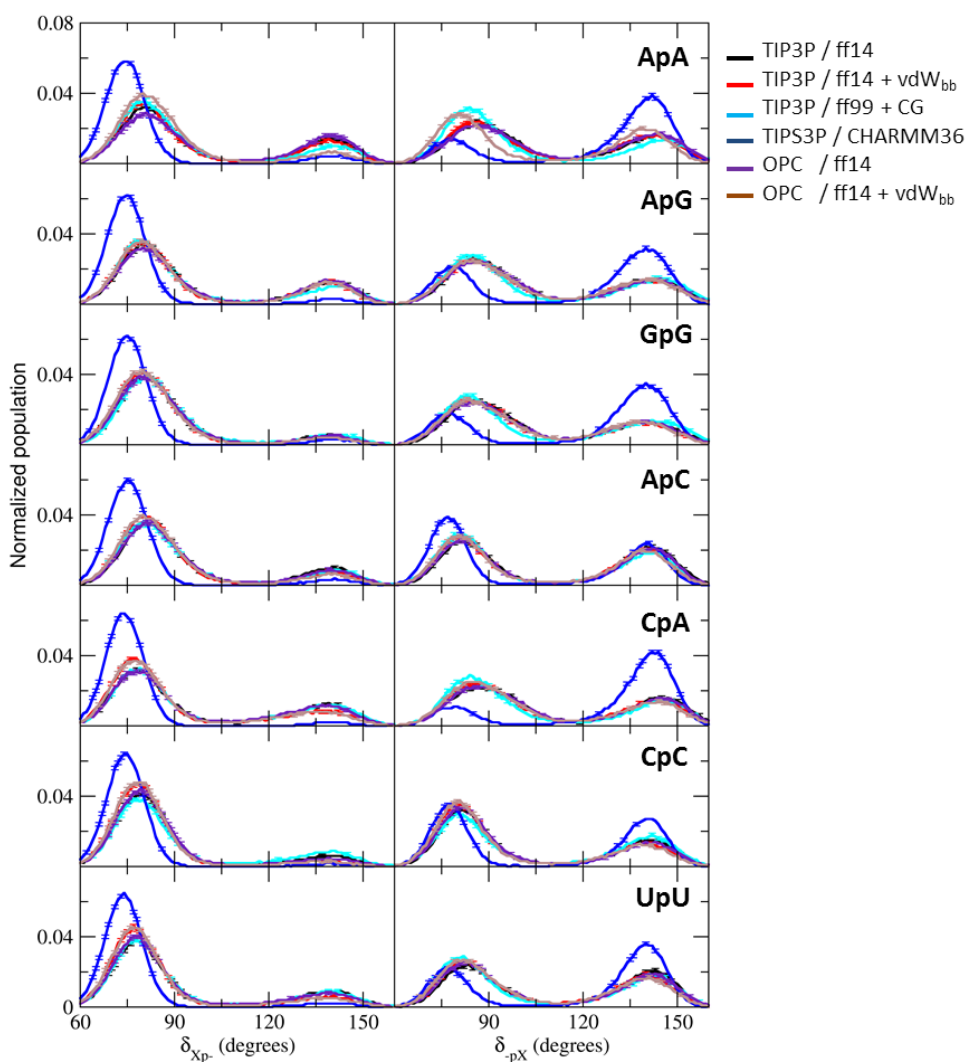


Figure 2.5. Histograms of  $\delta$  dihedral torsions for two residues of a selection of DNMPs with different water / force field combinations at 292.18 K. The peaks at  $< 90^\circ$  represent C3'-endo and the peaks at  $> 140^\circ$  represent the C2'-endo sugar conformations. A limited range containing data is shown to improve clarity. The error bars represent the standard deviation between two independent runs.

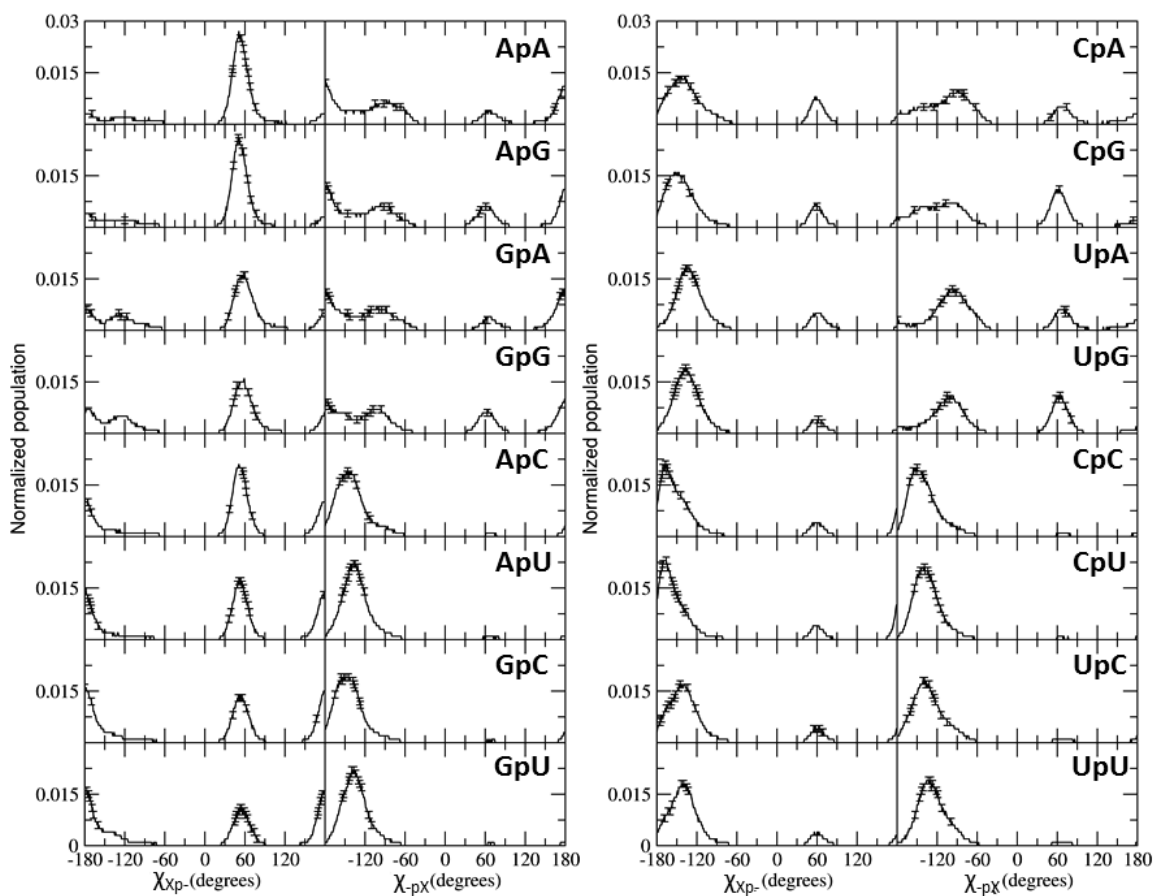


Figure 2.6. Histograms of  $\chi$  dihedral torsions for two residues of all DNMPs in TIP3P water and with ff14 force field at 292.18 K. The error bars represent the standard deviation between two independent runs.

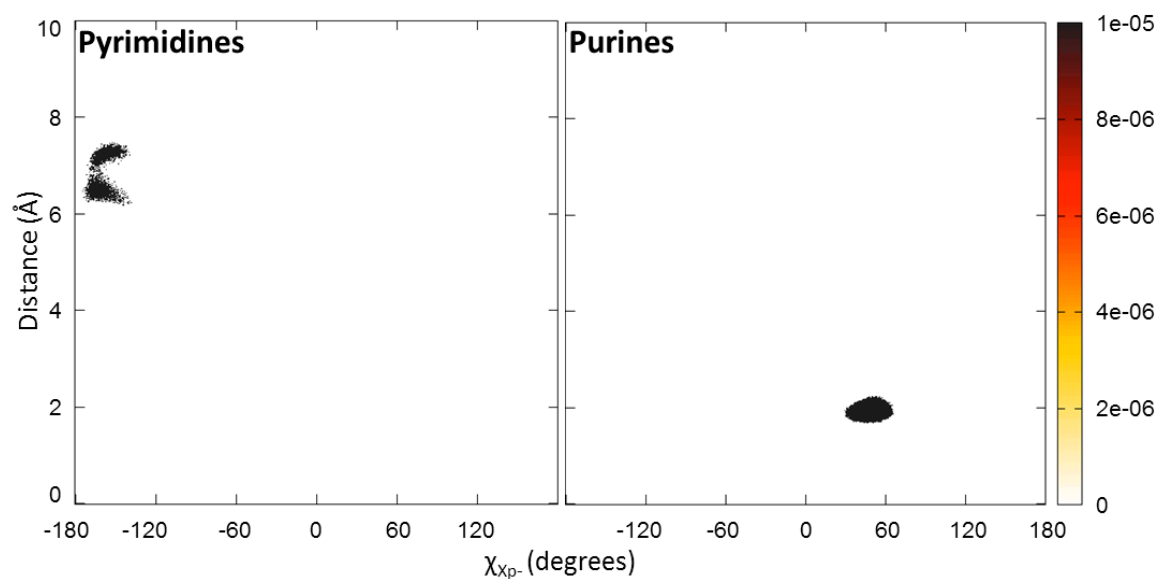


Figure 2.7. Correlation between  $\chi_{xp}$  and distance between HO5' and a base atom (N3 in purine and O2 in pyrimidines) in all TREMD runs. Colors in each plot represent normalized population over aggregated frames of the trajectories.

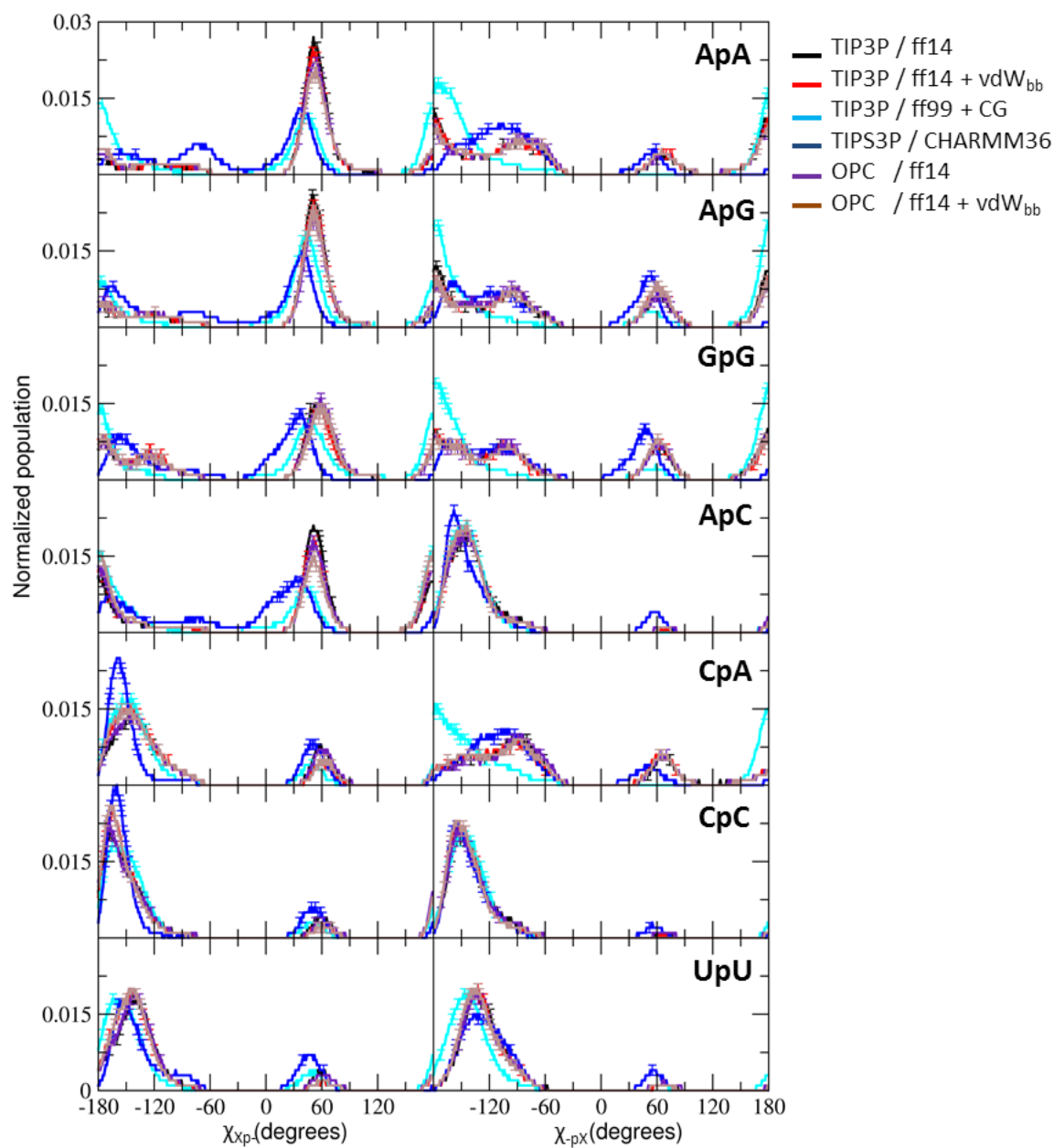


Figure 2.8. Mass-weighted heavy atom RMSD to the A-form distributions of all DNMPs with different tested scenarios in TREMD simulations at 298.41 K. The error bars represent the standard deviation between two independent runs.



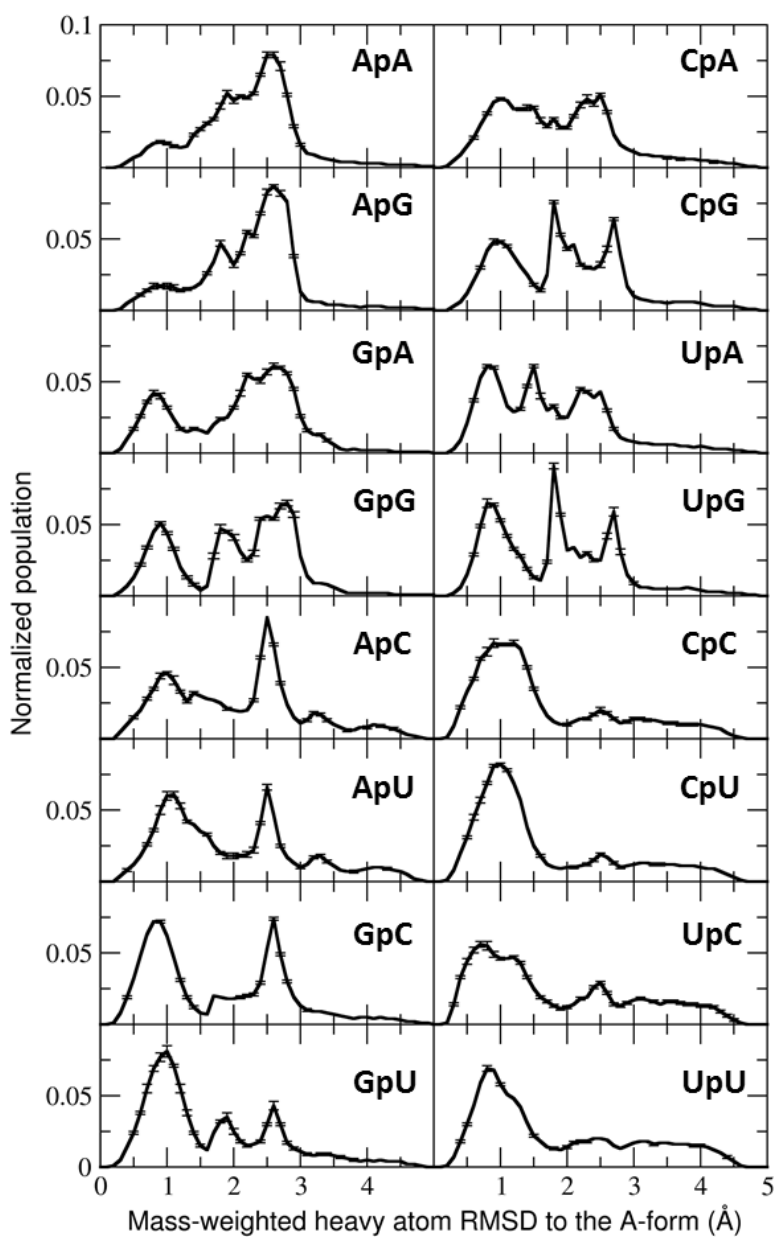


Figure 2.9. Mass-weighted heavy atom RMSD to the A-form distributions of all DNMPs in TIP3P water and ff14 force field in TREMD simulations at 298.41 K. The error bars represent the standard deviation between two independent runs.

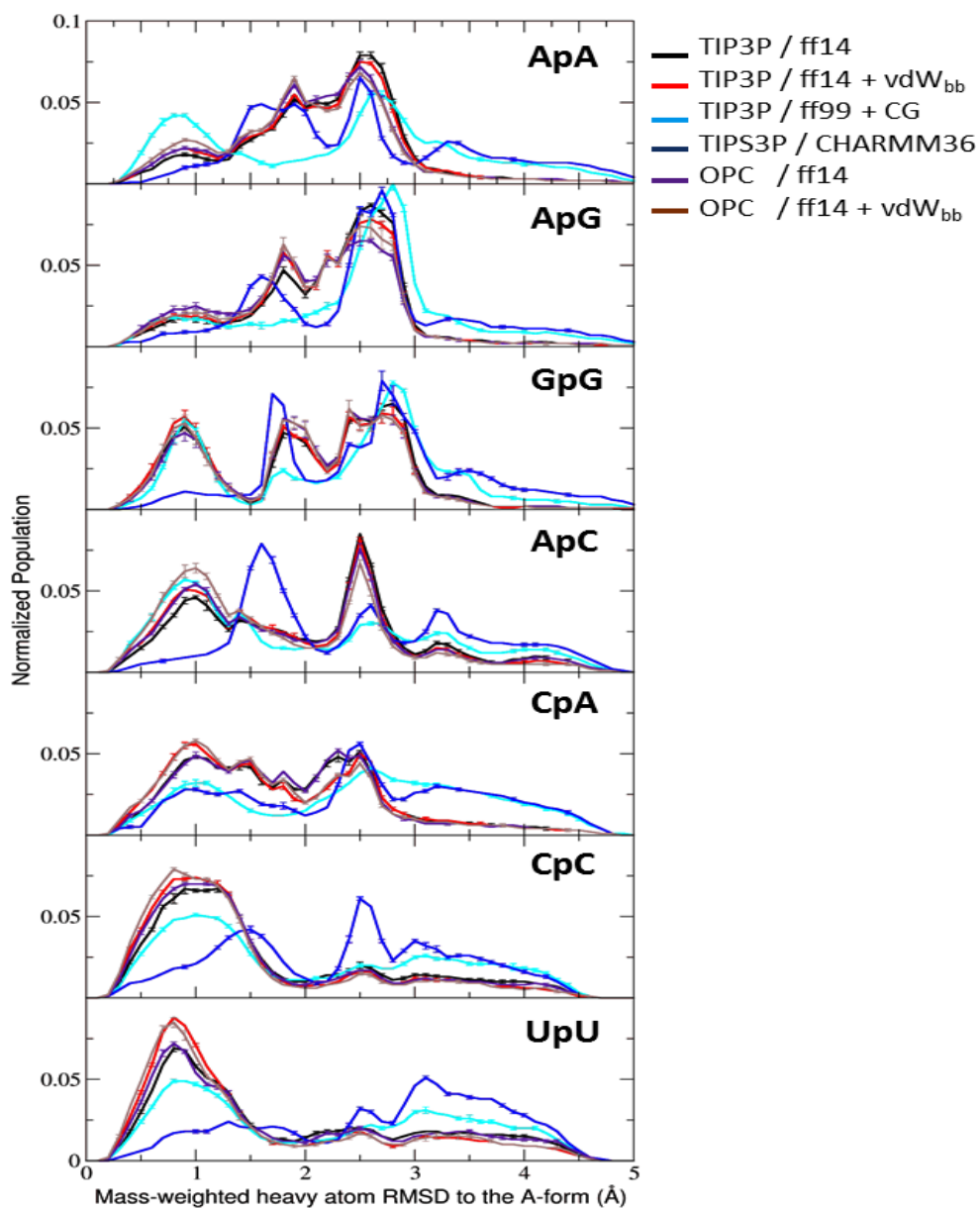


Figure 2.10. Mass-weighted heavy atom RMSD to the A-form distributions of a selection of DNMPs with different water and force field combinations in TREMD simulations at 298.41 K. The error bars represent the standard deviation between two independent runs.

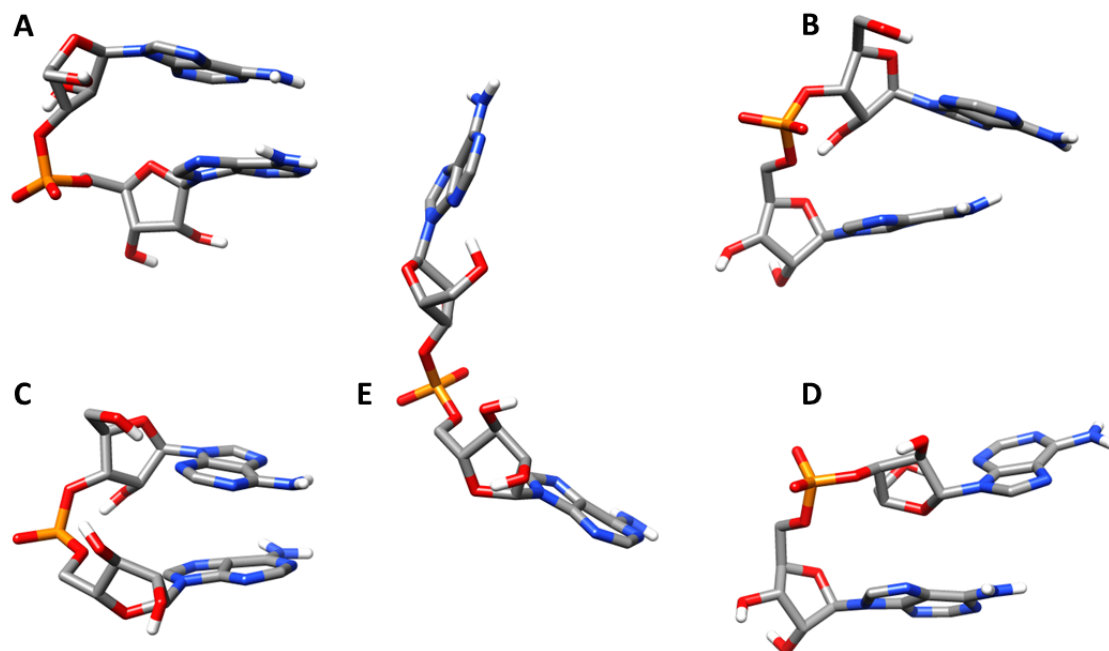


Figure 2.11. Representative structures of different conformational clusters of ApA, A) A-form cluster, B) Ladder cluster, C) Inverted cluster, D) Sheared cluster, E) Extended cluster.

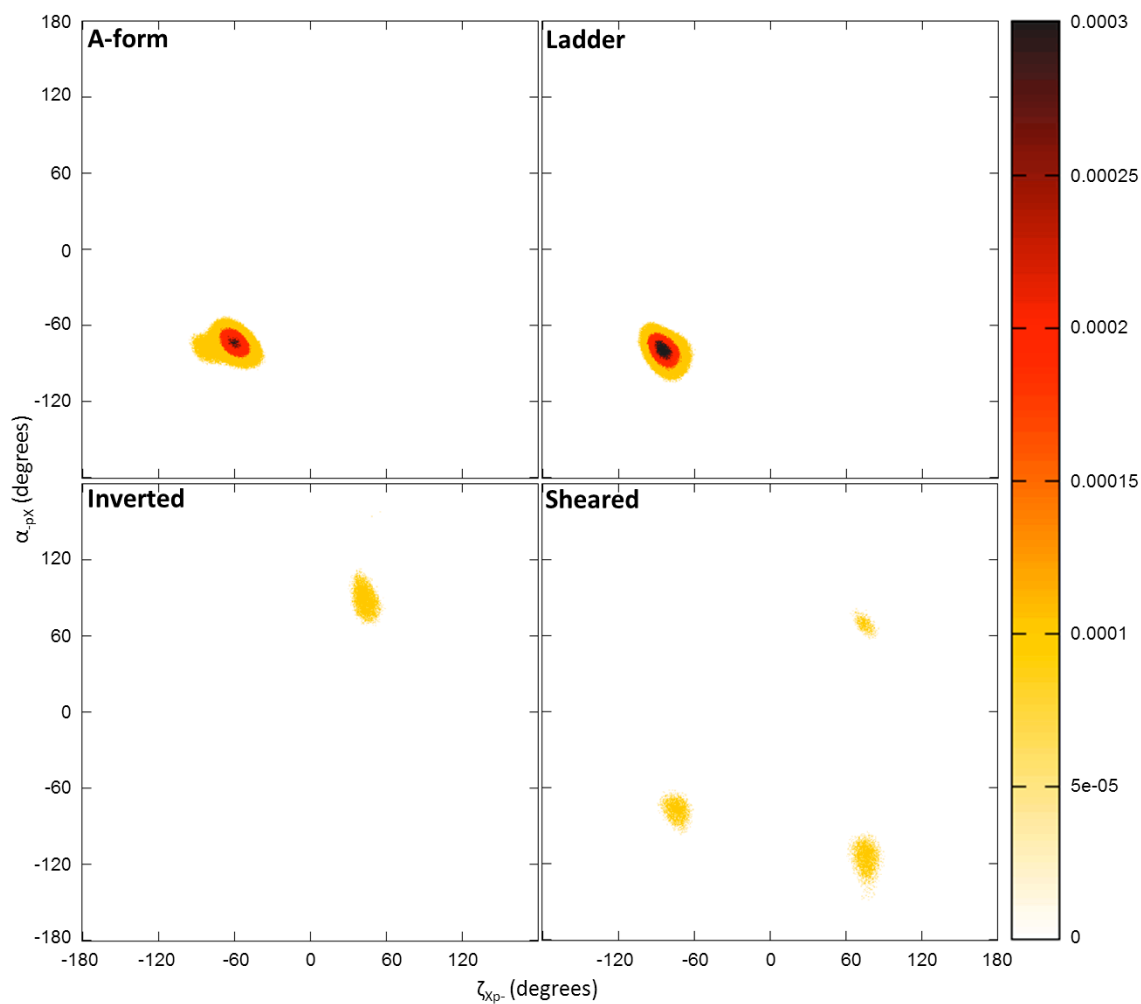


Figure 2.12. Correlations between  $\zeta_{xp}$ - and  $\alpha_{px}$  torsions in different conformational clusters in TREMD runs. Colors in each plot represent normalized population over aggregated frames of a cluster from all DNMPs.

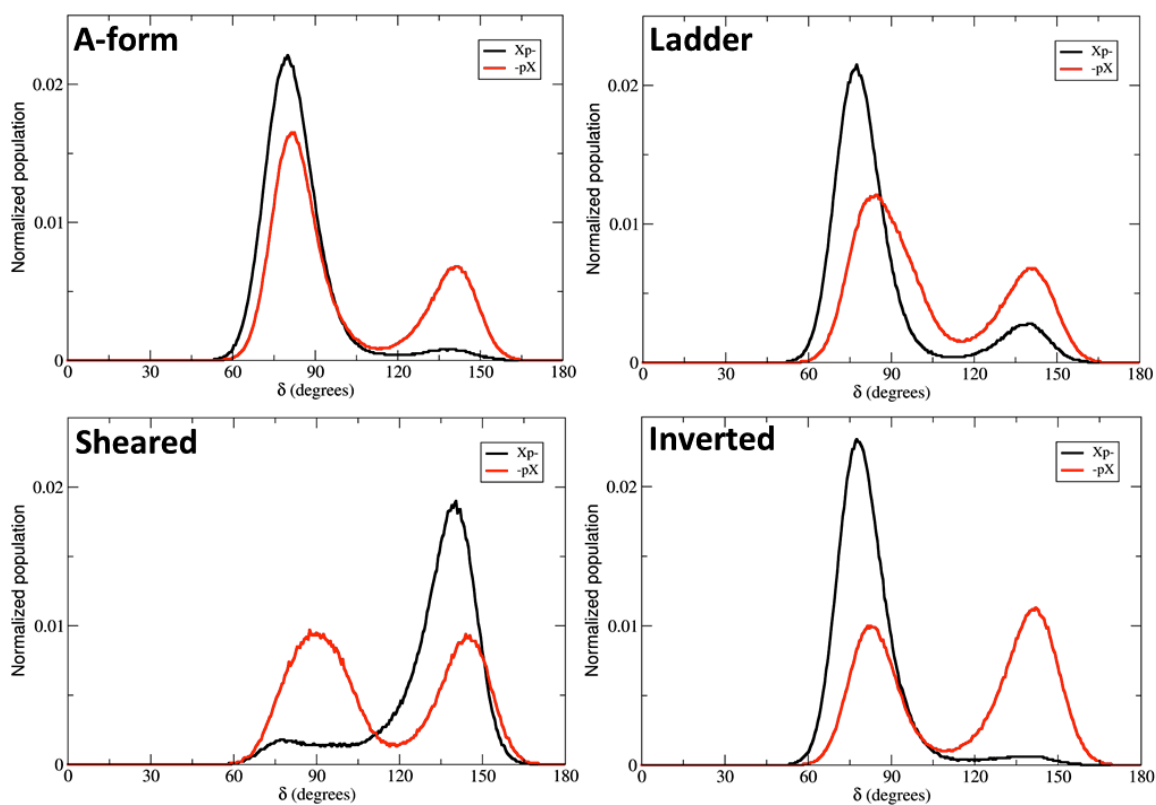


Figure 2.13. Histograms of  $\delta$  dihedral torsions for two residues of all DNMPs in the different conformation clusters. The peaks at  $< 90^\circ$  represent C3'-endo and the peaks at  $> 140^\circ$  represent the C2'-endo sugar conformations.

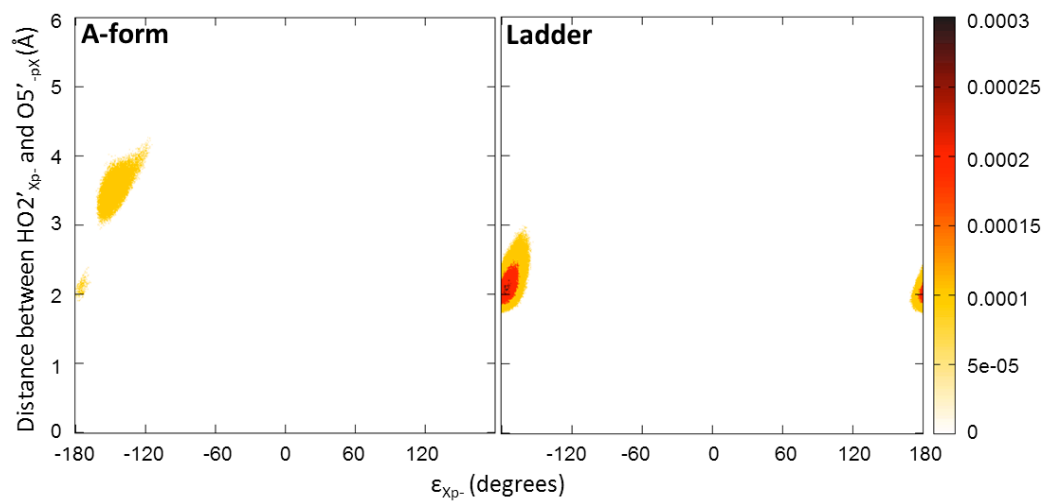


Figure 2.14. Correlation between the  $\epsilon_{Xp-}$  torsion and the distance between HO2'<sub>Xp-</sub> and O5'<sub>-pX</sub> atoms in the A-form and ladder conformational clusters in TREMD runs. Colors in each plot represent normalized population over aggregated frames of a cluster from all DNMPs.

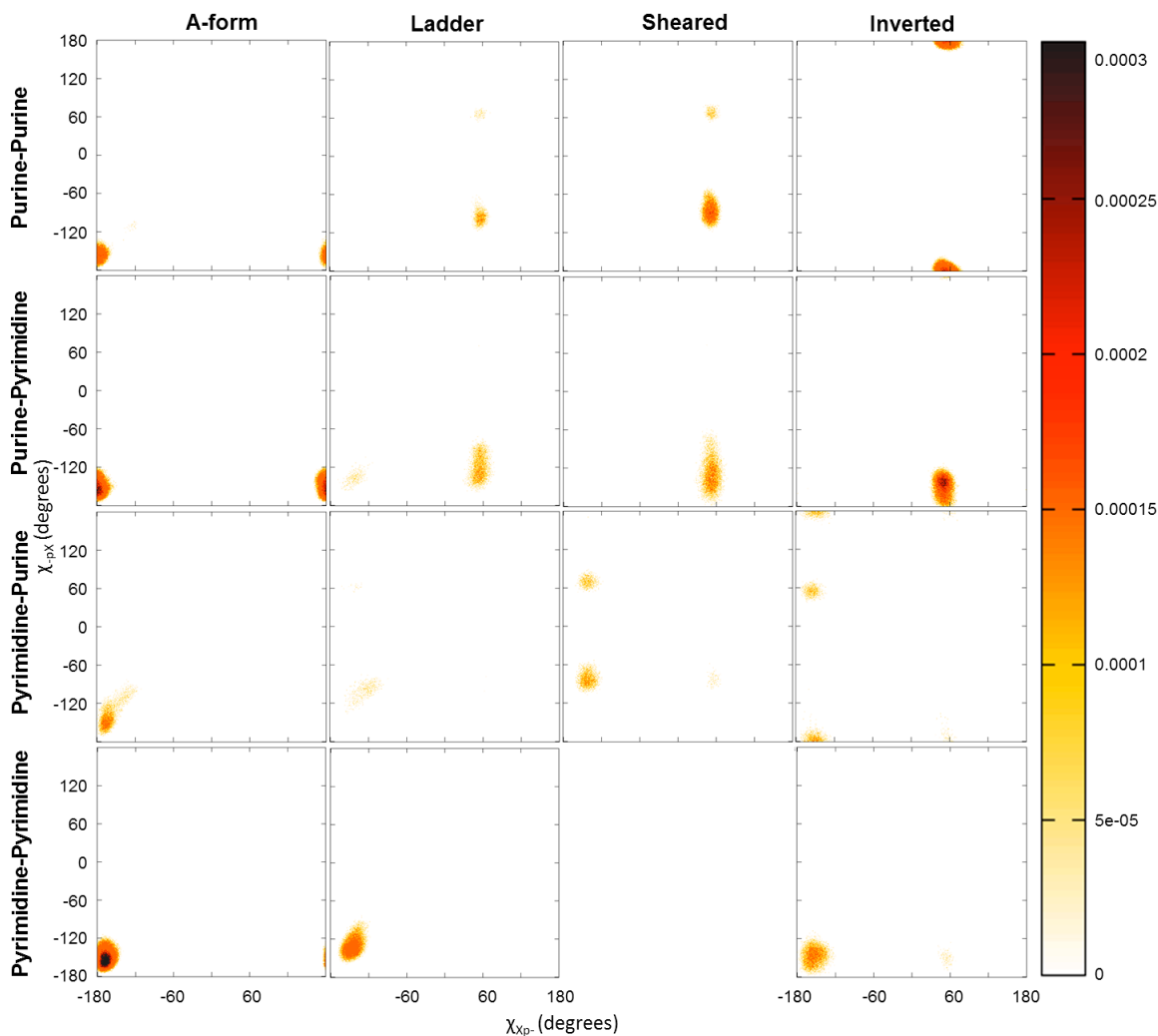


Figure 2.15. Correlations between  $\chi_{Xp}$ - and  $\chi_{pX}$  torsions in different conformational clusters in TREMD runs. Colors in each plot represent normalized population over aggregated frames of a cluster from one of the four classes of DNMPs (purine-purine, purine-pyrimidine, pyrimidine-purine and pyrimidine-pyrimidine).

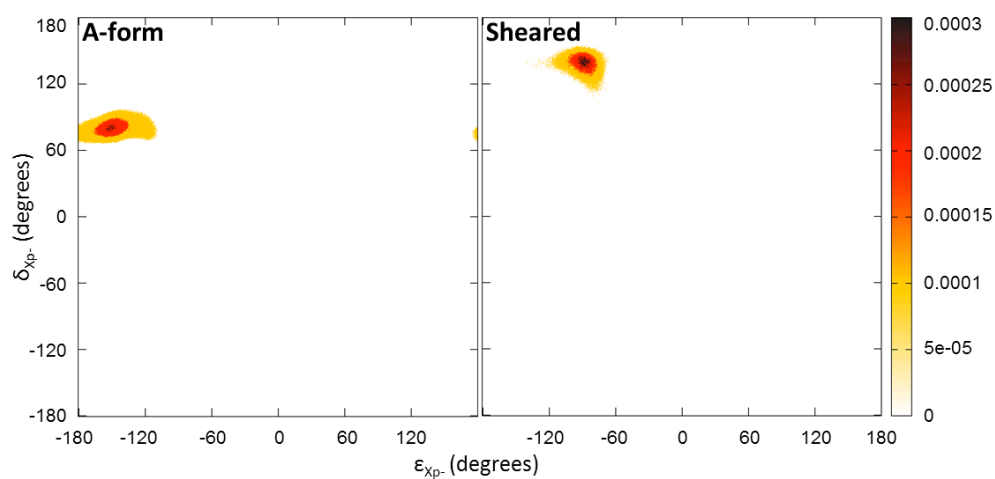


Figure 2.16. Correlations between  $\epsilon_{Xp-}$  and  $\delta_{Xp-}$  torsions in different conformational clusters in the A-form and sheared conformational clusters in TREMD runs.  $\delta_{Xp-}$  represents the conformation of the sugar in Xp- residue. Accordingly, the Xp- sugar in A-form conformation adopts C3'-endo while it adopts C2'-endo in the sheared conformation. Colors in each plot represent normalized population over aggregated frames of a cluster from all DNMPs.



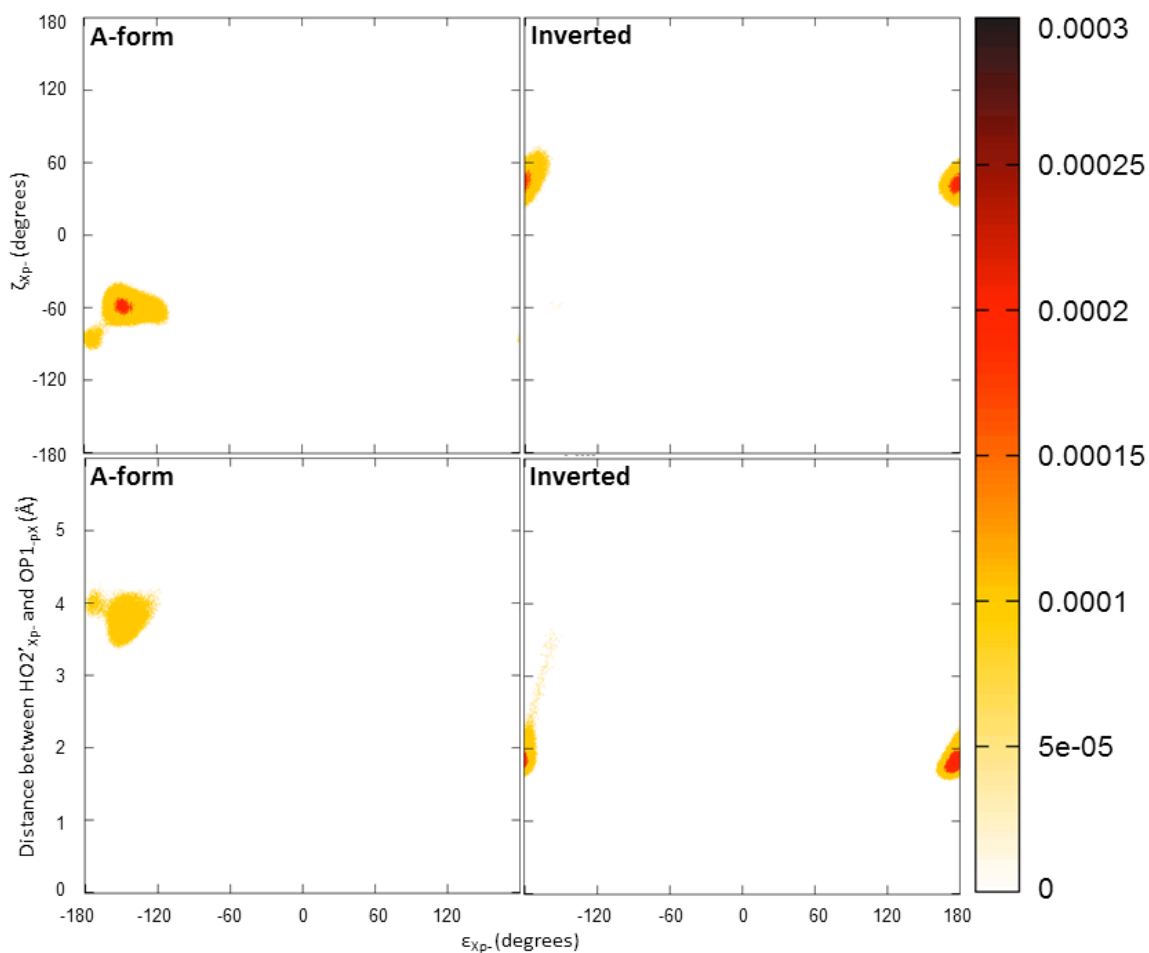


Figure 2.17. Correlation between  $\epsilon_{xp-}$  and  $\zeta_{xp-}$  torsions (TOP) and between  $\epsilon_{xp-}$  torsion and distance between HO2'<sub>xp-</sub> and OP1-pX atoms (BOTTOM) in the A-form and inverted conformational clusters in TREMD runs. Colors in each plot represent normalized population over aggregated frames of a cluster from all DNMPs.

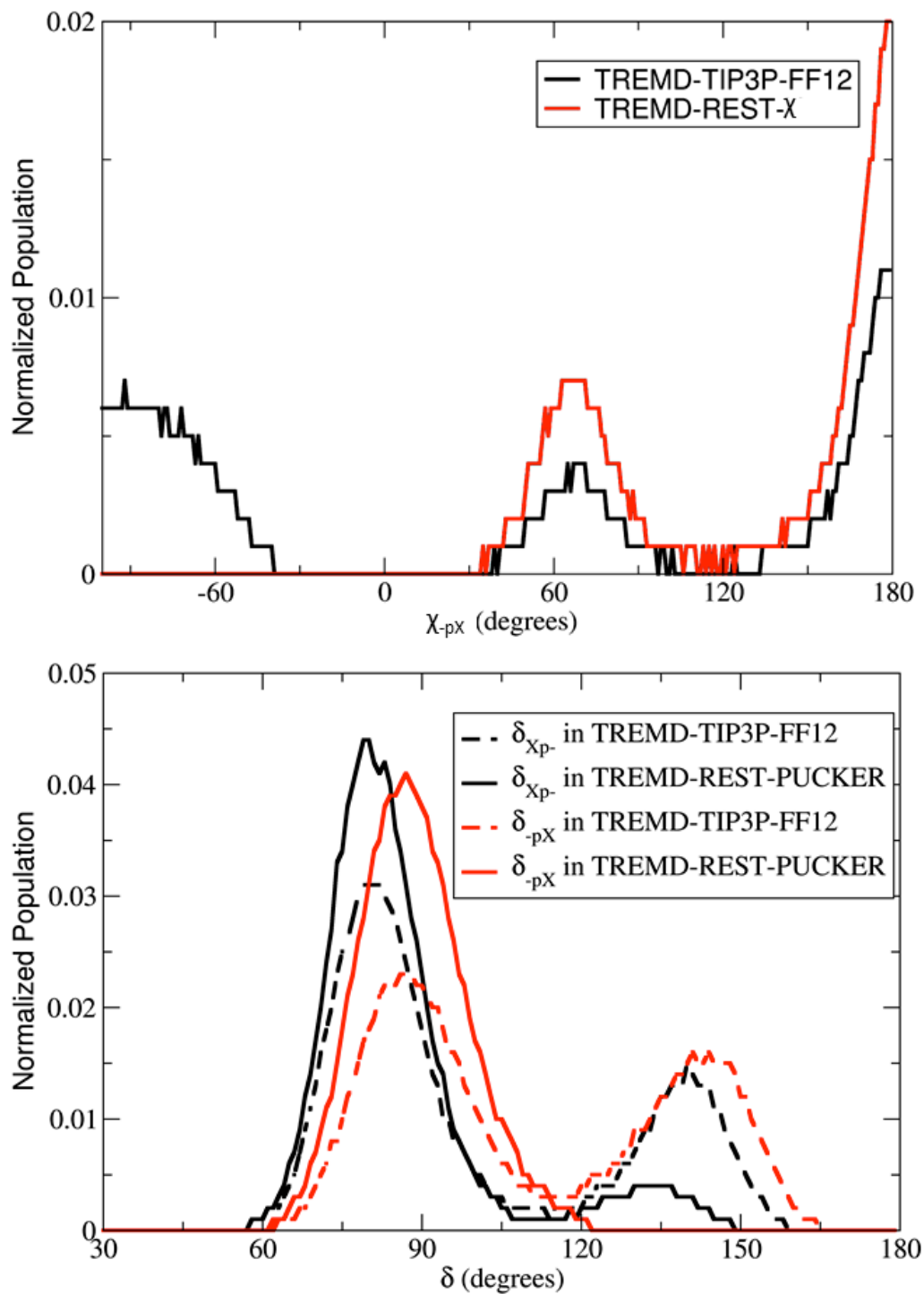


Figure 2.18. Effect of  $\chi_{pX}$  (top) and  $\delta$  (bottom) restraints on these dihedral torsions.

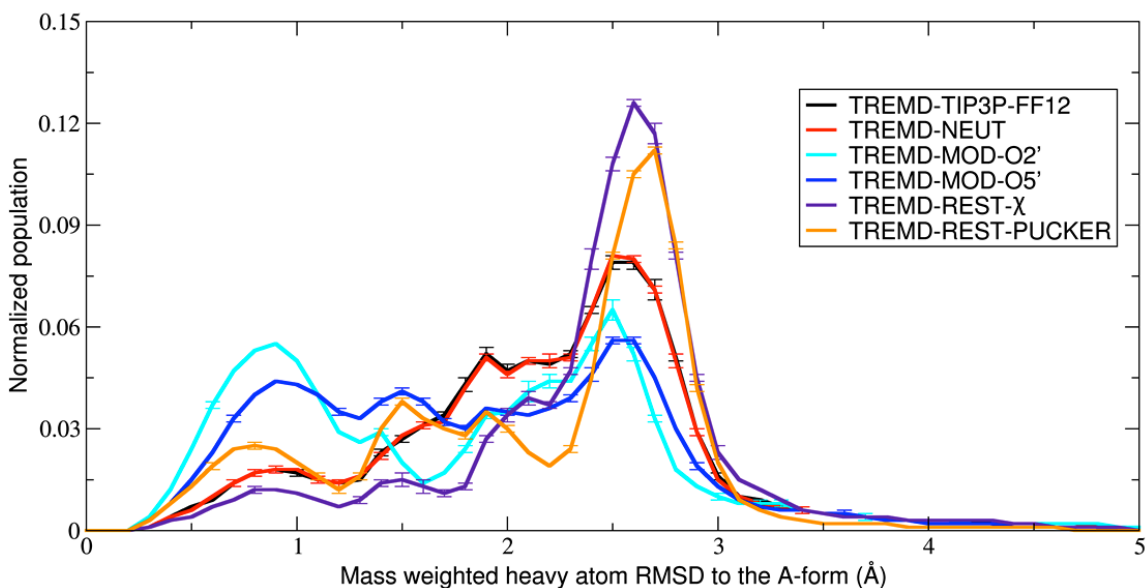


Figure 2.19. Mass-weighted heavy atom RMSD to the A-form distributions of ApA in different TREMD simulations at 298.41 K. TREMD-TIP3P-FF12 is unmodified and unrestrained simulation in TIP3P with ff12 with 150 mM excess NaCl. TREMD-NEUT is the same with no excess NaCl. In TREMD-MOD-O2' and TREMD-MOD-O5', HO2' and HO5' are neutralized respectively to knock-down their interactions with other atoms. In TREMD-REST- X , the  $\chi_{-pX}$  is prevented from adopting high-anti orientation (a range of  $-150^\circ$  to  $-30^\circ$ ). In TREMD-REST-PUCKER, both sugars are locked at the C3'endo conformation with restraining the  $\delta$  torsions.

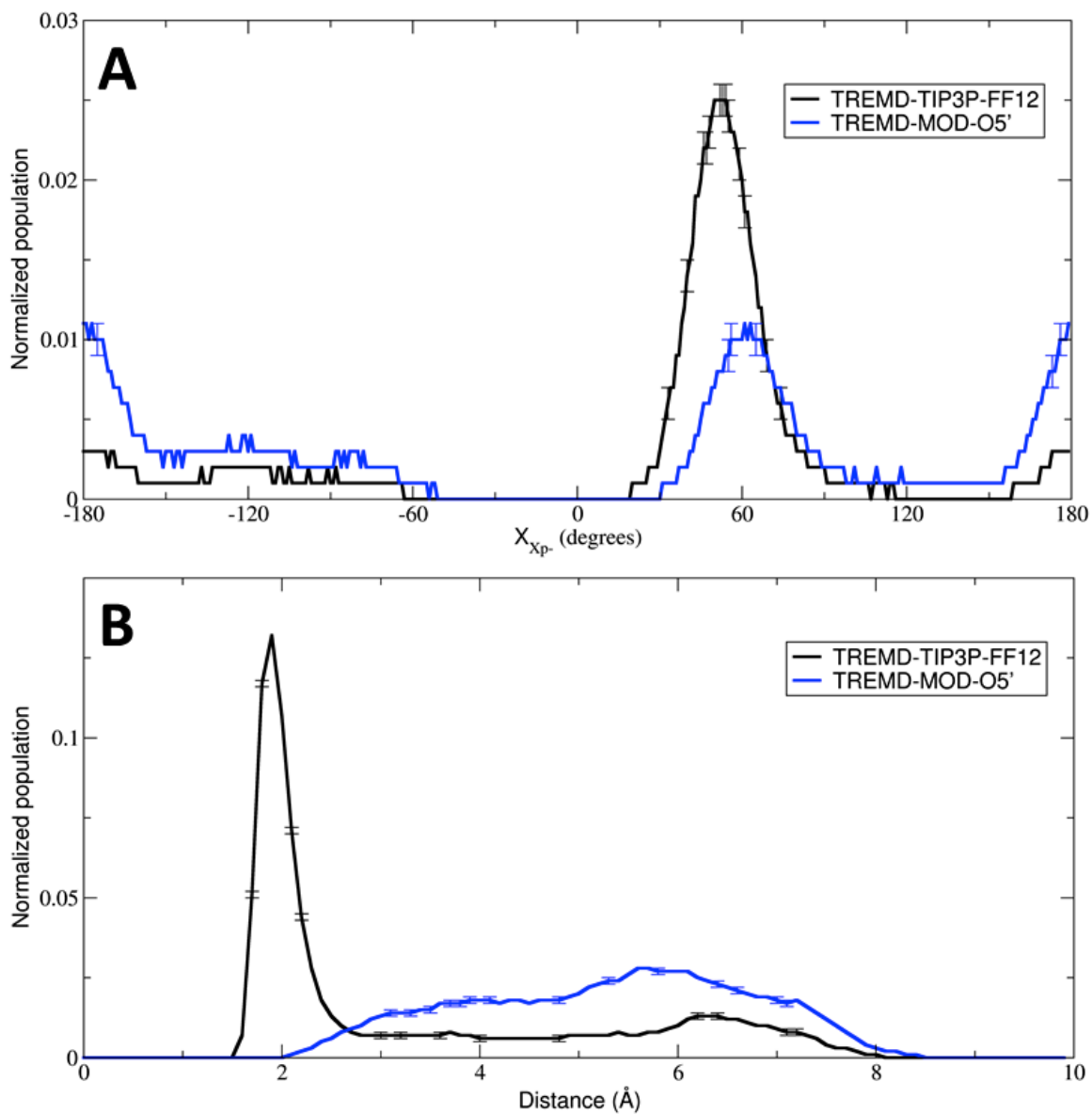


Figure 2.20. A) Distribution of the  $\chi_{Xp-}$  dihedral torsion of ApA, and B) distribution of distance between  $N3_{Xp-}$  and  $HO5'_{Xp-}$ , in normal TREMD simulation (TIP3P-FF12) and TREMD simulation in which the  $HO5'$  atom is neutralized (MOD-O5') at 298.41 K.

## CHAPTER 3

### COMPUTATIONAL ASSESSMENT OF POTASSIUM AND MAGNESIUM ION BINDING TO A BURIED POCKET IN GTPASE-ASSOCIATING CENTER RNA

The research project in this chapter has been designed by Hamed S. Hayatshahi, Thomas E. Cheatham, III and Daniel R. Roe. Part of the umbrella sampling simulations of  $K^+$  removal with no excess  $Mg^{2+}$  has been done by D. R. Roe and the rest of simulations and analyses were performed by H. S. Hayatshahi. The manuscript was primarily written by H. S. Hayatshahi and edited by D. R. Roe, T. E. Cheatham, III and Kathleen Hall.

#### **3.1 Introduction**

The folding and functioning of RNA molecules usually depends upon their interactions with monovalent and divalent cations (80-82). Diffusing cations are crucial, and, in general, they are the major driving force in stabilization of RNA structures via neutralizing the negative charges on RNA phosphate moieties (82). However in some cases, more complex tertiary structures are stabilized through the interaction of ions with specific RNA atoms, including bases and phosphates (82-85). Depending on the nature of the local RNA structure and the ion type, the interacting ions either “associate” with

RNA via hydrogen bonding to their first shell of waters (i.e., water-mediated interactions) or directly bind to one or more RNA atoms (82). The latter case, which is also referred to as “chelation,” is only possible after one or more water residues between the ion and RNA are removed from the first hydration shell surrounding both the ion and the RNA, a process which involves a high energetic barrier. Consequently, considering the large number of structurally characterized RNA molecules, there are fewer cases of identified direct ion-RNA interactions as compared to associated ions (86).

Some RNA molecules have been studied extensively by NMR and crystallography as model systems whose conformational and folding characteristics depend on specific ion binding as well as diffuse ion concentrations (87-91). A model system which has been studied extensively in terms of its ion interactions is the GTPase-Associating Center RNA (also known as 58-nucleotide RNA and here referred to as GAC RNA) which is a highly conserved fragment of the 23S ribosomal RNA which binds to the L11 ribosomal protein (92). Melting experiments have suggested that the tertiary structure formation of GAC RNA requires the presence of one bound monovalent ion ( $\text{NH}_4^+$  and  $\text{K}^+$  preferred) (93,94) and one or two bound divalent ions (with a preference for  $\text{Mg}^{2+}$ ) (95). There are two co-crystal structures of GAC RNA and L11 protein. These are a 2.6 Å resolution structure by Wimberly et. al. (92) (PDB code 1MMS) of GAC from eubacterium *Thermotoga maritima* and a 2.8 Å resolution structure by Conn et. al. (96) (PDB code: 1HC8) of GAC RNA from *E. coli* in complex with the C terminal domain of L11 protein from *Bacillus stearothermophilus*. The latter structure is a result of re-refinement of a previous structure from the same group (PDB code 1QA6) (97). In spite of residue differences at few positions, the RNA conformations are very similar in both structures (backbone RMSD of 0.6 Å, Figure 3.1.). An important feature of the RNA

tertiary structure conformation is that a bulged region containing nucleotides A1070, G1071 and C1072 is clamped between a U-turn on its 5' side, its 3' tail nucleotides and a hairpin C1092-G1099. Clamping this bulged region brings some phosphate and base oxygen atoms in relatively close proximity and results in the formation of a negatively charged area that is expected to interact with cations to maintain its stability. In both structures, three cations directly bind to oxygen atoms in this region with ion-oxygen distances matching direct chelation (Figure 3.2. and Table 3.1): an  $Mg^{2+}$  ion (Mg167 according to the nomenclature by Misra and Draper (98)) that directly bridges between the phosphate oxygen of A1073 and O4 of the U1094, an  $Mg^{2+}$  (Mg163) that binds to phosphate oxygen atoms of A1069 and A1070; and a cation in the buried region between those magnesium ions that directly interacts with the bulge phosphate oxygen atoms (Figure 3.2.). This third cation is identified as magnesium in the 1MMS structure (Mg58), but as potassium in the 1HC8 structure (K58). According to 1HC8 structure, the third cation is hypothesized to be the monovalent ion that is essential for final tertiary folding as characterized in the melting experiments (93,96). Since the RNA can be neutralized more effectively by cations with higher charges, one might expect that a bound  $Mg^{2+}$  ion may be a better candidate to stabilize the tight ion binding pocket in the GAC RNA, consistent with 1MMS structure (92). However, Conn et al. suggested that this stabilizing and bound ion is a  $K^+$  based on both its distances to the chelating oxygen atoms and an experiment in which they crystalized the GAC complex with thallium (which has a similar nominal ionic radius to potassium (91)) and found a bound thallium in the same position (96). However, simultaneous binding of three cations (Mg167, K/Mg58 and Mg163) to RNA in a relatively confined area (Figure 3.2.), as depicted in the crystal structures, at full occupancy could be considered unexpected due to

electrostatic repulsion except under very high salt concentrations (i.e., under crystallization conditions). Unfortunately, both crystal structures do not comment on the occupancy. Binding of Mg167 is supported by hydroxyl radical footprinting experiments(99) and nonlinear Poisson-Boltzmann calculations (98), while binding of Mg163 is not supported by noncrystallographic experimental data. Instead, according to nonlinear Poisson-Boltzmann calculations, Misra and Draper have concluded that although positions 163 and 167 have larger electrostatic potential in comparison to other magnesium sites in the 1HC8 crystal structure, binding to position 163 is energetically unfavorable due to the ion dehydration cost and repulsive force from the monovalent ion (98). Therefore, it can be hypothesized that occurrence of Mg163 near the K58 in 1HC8 or near Mg58 in 1MMS may be the result of crystallization conditions (i.e., very high salt concentrations), crystal packing, partial occupancy, or a combination of all three influences.

As reviewed previously, numerous experimental approaches applied to study RNA-ion interactions have led to quantitative measurement of the ion-dependent stability of many RNA conformations (87). These experimental approaches include monitoring the UV absorbance during thermal unfolding of tertiary RNA conformations (95,100,101), measurement of the force required to mechanically unfold certain RNA structures by using optical tweezers (102), calorimetric analysis of the ion binding events inferred to occur during the RNA unfolding/refolding procedure (103), and fluorescence titrations of proposed ion binding events at different temperatures (104). One potential limitation of these experimental methods is the difficulty of separating the binding to a single specific binding site from interactions between the background ion media and the RNA in titration experiments (99). A further confounding difficulty in both approaches that the



ion binding events are usually accompanied with major conformational changes in RNA (99). In other words, it is difficult to separate thermodynamics of ion binding from thermodynamic changes from folding (99,105). Molecular dynamics (MD) simulations can be a complementary approach to help overcome these limitations through the ability to separately investigate each and every ion-binding event at a very detailed atomistic level, albeit subject to limitations of the applied force field and effective sampling. In-silico molecular models enable us to artificially separate the influences by examining the ion binding event in a limited conformational space close to the RNA native structure and compare the intrinsic free energies of binding for different ions.

In this work, we used MD simulations in explicit solvent with different ion conditions, effectively an ion-competition experiment, to analyze the preferential ion binding at the K/Mg58 position depicted in two crystal structures of 1MMS and 1HC8. To further characterize preferential binding at this site, we also employed the umbrella sampling method to estimate the potential of mean force (essentially the free energy) of pulling of specific ions out of the pocket to bulk solvent at various salt concentrations. The umbrella sampling studies were performed in the presence and absence of a magnesium ion in the Mg163 position in order to analyze the influence this magnesium has on the ion binding in the adjacent K/Mg58 pocket, as well as to examine the hypothesis that it can coexist close to the other two cations at both the high magnesium concentrations used in crystallization and the lower magnesium concentrations used in most other solution biochemical assays.

## **3.2 Methods**

Table 3.2 summarizes the simulations performed in this project. Details about each simulation set and its preparation steps are presented in this section.

### **3.2.1 MD simulations to investigate preferential ion binding in the pocket**

#### **3.2.1.1 COMPETITION\_1 simulation set**

The crystal structure of GAC RNA with PDB code 1HC8 (96) was used as the initial structure. All of crystallographic ions and water residues, as well as the terminal GTP residue, were removed. The terminal GTP is paired with a uracil in the crystal structure and was removed since it is located about 30 Å from the monovalent ion binding site and does not closely interact with this site of interest in the absence of crystal packing. Hydrogen atoms were added to the remaining 57 nucleotides and the in-vacuo topology and coordinates were built using the tLEaP program in Amber 12 (106) with ff12SB force field which includes ff99 (51) with updated  $\chi$  (53) and  $\alpha/\gamma$  parmbsc0 (52) modifications. 2500 cycles of steepest descent minimization with the Hawkins, Cramer, Truhlar Generalized Born implicit solvent model (107,108) was performed using SANDER program in Amber 12. The resulting structure (with an RMSD of 0.37 Å to the initial structure) was solvated in a truncated octahedral TIP3P water (56) box such that the minimum distance of the residues to the edge of the box was 12.0 Å. Then 28 Mg<sup>2+</sup> ions with Allnér-Nilsson-Villa parameters (22) and 28 K<sup>+</sup> and Cl<sup>-</sup> ions with Joung-Cheatham parameters (25) were added to the box using tLEaP. The CPPTRAJ program (60) in Amber Tools was used to randomize ion positions in such a way that the distance of each ion to any RNA atom was not less than 6 Å and ions were separated with

minimum distance of 4 Å, causing the initial structures to lack ions in the positions previously occupied in the crystal structures. Ion position randomization was done 20 times in successive steps to generate 20 independent copies of the system. To avoid any initial bias for ion occupancy in the K/Mg58 pocket, the distance of the closest K<sup>+</sup> and Mg<sup>2+</sup> ions to the pocket were identified in each copy to make sure that almost half of the copies have each ion types as closest to the pocket.

Each copy of the system was equilibrated in four steps: 1) minimization, 2) heating at constant volume (NVT), 3) further heating at constant pressure (NPT), and 4) restrained MD at constant pressure. Details of the equilibration protocol are described in section 3.2.4.1. The production MD simulations were performed using *pmemd.cuda* from Amber 12 for 1 μs per copy with 2 fs time step in the NPT ensemble while RNA atoms were restrained with a 0.5 kcal/mol-Å<sup>2</sup> force constant to assess the ion binding with the RNA structure kept close to the crystal conformation. The temperature was held at 298.15 K with weak-coupling algorithm with time constant of 10 ps (63). SHAKE (62) was used to constrain bonds to hydrogen, and particle-mesh Ewald (109) with an 8.0 Å cutoff and default parameters was used to treat long-range electrostatics in the production simulations. Coordinates were saved to trajectory files every 5000 steps (10 ps intervals).

#### 3.2.1.2 COMPETITION\_2 simulation set

The last frames of the 20 independent COMPETITION\_1 simulations were used as initial structures for the COMPETITION\_2 simulations, which were run with the same protocol except with no restraints on the RNA for 500 ns to see if the ion binding preference is reproducible with potential subtle changes in RNA conformation.

### 3.2.2 HYDRATION simulation set

To analyze the effect of the ion position on its hydration shell, 20 copies of simulations with the same RNA/ions composition as above were built (named here as ion hydration simulations) except that the K58, Mg163 and Mg167 were positioned in their crystallographic sites. These three ions as well as the RNA atoms were fixed with 0.5 kcal/mol-Å<sup>2</sup> restraints during the course of 10 ns simulations with the same positional restraint used for the RNA atoms. These simulations were done to assess the ion hydration shell in the pocket exactly in the same pocket conformational situation as of the crystal structure.

### 3.2.3 Ion pull-out umbrella sampling simulations

#### 3.2.3.1 Preparing the initial structures for umbrella sampling simulations

Umbrella sampling was done to quantitatively compare the free energy of binding (potential of mean force) of different ions to the GAC monovalent ion binding site at three different ionic environments. Prior to the umbrella sampling simulation, we first performed unrestrained simulations of GAC RNA in three different ion environments to identify spontaneous ion leaving or unbinding events as follows:

- 1) MG0: RNA + 11 crystallographic Mg<sup>2+</sup> + neutralizing K<sup>+</sup> + 100 mM excess KCl
- 2) MG15: RNA + 11 crystallographic Mg<sup>2+</sup> + neutralizing K<sup>+</sup> + 100 mM excess KCl  
+ 15 mM excess MgCl<sub>2</sub>
- 3) MG80: RNA + 11 crystallographic Mg<sup>2+</sup> + neutralizing K<sup>+</sup> + 100 mM excess KCl  
+ 80 mM excess MgCl<sub>2</sub>

To prepare models, the crystal structure with PDB code 1HC8 was used as the initial

structure (96). Crystallographic  $K^+$  and  $Mg^{2+}$  ions as well as water residues were retained but the terminal GTP residue was removed (as discussed previously). Hydrogen atoms were added to the remaining 57 nucleotides. Using tLEaP program, the structure was solvated in a truncated octahedral box of TIP3P waters such that the minimum distance of the residues to the edge of the box was 12.8 Å. Then 33  $K^+$  ions were added to neutralize the whole system (RNA + crystallographic ions); and 26  $K^+$  and 26  $Cl^-$  were added to provide ~100 mM excess KCl. To prepare three sets of simulations with different ionic conditions, 0, 4 and 21  $Mg^{2+}$  as well as 0, 8 and 42  $Cl^-$  were added to the systems to make approximately 0, 15 and 80 mM excess  $MgCl_2$  concentrations, respectively (here referred to as MG0, MG15 and MG80 sets). The noncrystallographic ions were put in random positions using CPPTRAJ (60) as described in the COMPETITION\_1 section to prepare three (for the MG0 set) and six (for MG15 and MG80 sets) system copies. The MG0 set was equilibrated with the protocol discussed in section 3.2.4.1. The MG15 and MG80 sets were equilibrated in nine steps of minimizations and restrained MD simulations using a EQ2 equilibration protocol as described in section 3.2.4.2. Hydrogen mass repartitioning was used to facilitate a 4 fs time step in the final step of the equilibration as well as in the production simulations (26,27). This was done by transferring part of mass of the solute heavy atoms to the covalently-bonded hydrogen atoms using the parmed.py program from Amber 14 (65). For each copy of each set, the equilibration set was followed by about ~1  $\mu s$  of production molecular dynamics using the *pmemd.cuda* program (110) from Amber 14 as described for the production step in the former section but with no restraints.

The reaction coordinate for the umbrella sampling was chosen to be the distance between the ion and the center of mass of nine phosphate groups of A1070, C1072 and

A1073 (i.e., the phosphorus and two non-bridging oxygen atoms of each residue). We refer to this position as the reaction coordinate anchor point. Initial structures for the umbrella sampling windows were chosen differently in three sections along the ion-leaving path as follows:

1)  $\sim 2.5 \text{ \AA} - 6.9 \text{ \AA}$ : The bound potassium to the monovalent ion binding site was tracked during the above-mentioned unrestrained simulations (MG0, MG15 and MG80) and all ion exchange events were analyzed both visually and using CPPTRAJ (60). For each simulation set, the trajectory frames from one ion exchange event with the least conformational change in RNA near the ion binding site were chosen for generating the initial structures. The structures for initial umbrella windows with reaction coordinate values between the initial equilibrated distance ( $\sim 2.5 \text{ \AA}$ ) and  $6.9 \text{ \AA}$  were extracted from the chosen ion exchange path trajectory using CPPTRAJ (60).

2)  $0.5 \text{ \AA} - \sim 2.5 \text{ \AA}$ : The last frame of the umbrella with the reaction coordinate value of the initial distance in the ion exchange event was used as the initial structure for a set of consecutive 2 ns molecular dynamics simulations to direct the ion to  $0.5 \text{ \AA}$  on the reaction coordinate. The umbrella sampling simulations in this part of the path with low reaction coordinate values were conducted in serial fashion, with the last frame of each window simulation used as the initial frame of the next window simulation with  $0.1 \text{ \AA}$  lower reaction coordinate value. Each of these umbrellas was started with short equilibration simulations described in section 3.2.4.3.

3)  $7.0 \text{ \AA} - 20.0 \text{ \AA}$ : The initial frame of the umbrella window with reaction coordinate value of  $6.9 \text{ \AA}$  was used as the initial structure of a short molecular dynamics simulation in which the leaving  $\text{K}^+$  ion was gradually directed to the bulk solvent (reaction coordinate =  $20.0 \text{ \AA}$ ) in 130 ps as follows: the pseudo-angle involving the ion and the

phosphorus atoms of A1071 and C1072, as well as the pseudo-dihedral involving the ion and the phosphorus atoms of A1073, G1071, and C1072 were kept fixed at the initial value with 200 kcal/mol-Å<sup>2</sup> restraints, and the reaction coordinate value was increased from 6.9 Å to 20.0 Å with a 100 kcal/mol-Å<sup>2</sup> distance restraint. All other RNA heavy atoms were restrained in position with 20 kcal/mol-Å<sup>2</sup> force constant to avoid potential conformational disruption resulting from ion displacement near the RNA atoms. Then the initial structures for umbrellas within this distance frame were extracted using CPPTRAJ (60) as well to generate the path from 7.0 Å to 20.0 Å. The window interval for the entire reaction coordinate was 0.1 Å.

### 3.2.3.2 Running the umbrella sampling simulations

Initial structures generated in parts 2 and 3 of the path were equilibrated using the EQ3 protocol described in section 3.2.4.3. A total of 2 ns of production molecular dynamics in constant pressure and temperature was performed for each umbrella along the whole path, during which a Langevin thermostat (111) was used with collision frequency of 5 ps<sup>-1</sup> to hold temperature at 298.15 K; the SHAKE algorithm (62) was used to constrain the bonds involving hydrogen atoms; the direct space cutoff was set to 8.0 Å and particle-mesh Ewald (109) was used to calculate long-range interactions. A 2 fs time step was used for the production simulations and the reaction coordinate values were saved for every step. During equilibration and production simulations, the ion was held at the related umbrella value with 100 kcal/mol-Å<sup>2</sup> distance restraint. Section 2 of the path was sampled in serial fashion while other sections were sampled in parallel as independently running simulations. This way, the reaction coordinate between 0.5 and 20.0 Å was sampled at 0.1 Å intervals for a leaving K<sup>+</sup> at 0, 15 and 80 mM of excess

Mg<sup>2+</sup> with 2 ns simulation time per umbrella (US\_MG0\_K, US\_MG15\_K and US\_MG80\_K simulation sets).

To examine the behavior of Mg<sup>2+</sup> in the same position, the leaving ion was mutated from K<sup>+</sup> using parmed.py program in Amber 14, and the same initial coordinates for each umbrella window were used. When mutating the ion to Mg<sup>2+</sup>, a Mg<sup>2+</sup> located in bulk solvent was simultaneously mutated to K<sup>+</sup> to keep the system charge neutral. This way, the reaction coordinate between 0.5 and 20.0 Å was sampled at 0.1 Å intervals for a leaving Mg<sup>2+</sup> at 0, 15 and 80 mM of excess Mg<sup>2+</sup> with 2 ns simulation time per umbrella (US\_MG0\_MG, US\_MG15\_MG and US\_MG80\_MG simulation sets). For no excess Mg<sup>2+</sup> scenario, the reaction coordinate was also sampled for a leaving Na<sup>+</sup> ion with mutating the K<sup>+</sup> to Na<sup>+</sup> (US\_MG0\_NA set). The Na<sup>+</sup> simulation was performed to compare the affinity of the pocket for Na<sup>+</sup> and K<sup>+</sup> as a test for experimental agreement (93,94).

Two additional umbrella sampling simulations were conducted for excess Mg<sup>2+</sup> concentrations of 15 mM and 80 mM for either Mg58 or K58 leaving the binding pocket where Mg163 was moved to random positions far from the RNA using CPPTRAJ (60) in the initial structure of each umbrella (US\_MG15\_K\_163, US\_MG80\_K\_163, US\_MG15\_MG\_163, US\_MG80\_MG\_163 sets). To examine the effect of magnesium parameters on its free energy of binding, the Lennard-Jones parameters of all magnesium residues in the K<sup>+</sup>->Mg<sup>2+</sup> mutated system in the first scenario were changed to the parameters developed by Åqvist (21) as well as HFE and IOD parameters developed by Li et al. (23) (US\_MG0\_MG\_AQV, US\_MG0\_MG\_HFE and US\_MG0\_MG\_IOD sets).



### 3.2.3.3 Calculating the potential of mean force (PMF)

The histograms of the reaction coordinate values for the umbrellas were analyzed with CPPTRAJ (60) and extra umbrellas were initiated where the overlap between umbrellas was poor. The WHAM method (112,113) as implemented by Alan Grossfield<sup>1</sup> was used to generate the potential of mean force from the umbrella sampling simulations.

### 3.2.4 Equilibration protocols

#### 3.2.4.1 EQ1

This protocol was used for COMPETITION\_1, COMPETITION\_2, HYDRATION and MG0 simulations sets: 1) 1000 steps of steepest descent, followed by conjugate gradient energy minimization with 25 kcal/mol-Å<sup>2</sup> positional restraints on RNA atoms. 2) The systems were heated up to 150 K in 100 ps of constant volume MD simulation while RNA atoms were restrained with the same force constant. 3) The systems were heated up to 298.15 K in 100 ps of constant pressure MD simulations while RNA atoms were restrained with a 5 kcal/mol-Å<sup>2</sup> force constant. 4) A 5 ns MD simulation was performed for each copy in constant pressure with 0.5 kcal/mol-Å<sup>2</sup> positional restraint on RNA atoms. In the equilibration steps, time step was set to 1 fs, SHAKE (62) was used to constrain all bonds to hydrogen atoms and a Langevin thermostat was used with collision frequency of 2 ps<sup>-1</sup> to control temperature (66,67). Also, the direct space cutoff was set to 8.0 Å and particle-mesh Ewald (64) was used to calculate long-range interactions. The equilibration steps were performed using *pmemd.MPI* in Amber 12.

---

<sup>1</sup> <http://membrane.urmc.rochester.edu/sites/default/files/wham/doc.html>

### 3.2.4.2 EQ2

This protocol was used for MG15 and MG80 sets: 1) 1000 steps of steepest descent energy minimization with RNA heavy atoms restrained in position with 5 kcal/mol-Å<sup>2</sup> 2) 15 ps molecular dynamics with RNA heavy atoms restrained in position with 5 kcal/mol-Å<sup>2</sup> in constant volume and constant temperature of 298.15 K 3) 1000 steps of steepest descent energy minimization with RNA heavy atoms restrained in position with 2 kcal/mol-Å<sup>2</sup> 4) 1000 steps of steepest descent energy minimization with RNA heavy atoms restrained in position with 0.1 kcal/mol-Å<sup>2</sup> 5) 1000 steps of steepest descent energy minimization with no restraints. 6) 5 ps molecular dynamics in constant pressure and temperature with 1 kcal/mol-Å<sup>2</sup> force constant restraint on RNA heavy atoms 7) 1 ns molecular dynamics in constant pressure and temperature 0.5 kcal/mol-Å<sup>2</sup> force constant restraint on RNA heavy atoms 8) The same as in step 7 with restraint on backbone RNA atoms 9) The same as step 7 with no restraints. The molecular dynamics time steps were 1 fs for the steps before step 9 and 4 fs in step 9. SHAKE algorithm (62) was applied in steps 6 to 9.

### 3.2.4.3 EQ3

This protocol was used in umbrella simulations of the Umbrella Sampling in section 2 and 3 of the path: 1) 2000 cycle steepest descent energy minimization with 5 kcal/mol-Å<sup>2</sup> RNA heavy atom restraint 2) 50 ps equilibration with such 1 kcal/mol-Å<sup>2</sup> restraint. In section 3 of the path (7.0 Å – 20.0 Å), directing pseudo-angle and pseudo-dihedral restraints were also applied in equilibration phases.

### **3.3 Results**

#### **3.3.1 MD simulations exploring preferential ion binding in the pocket**

To assess occupancy of the monovalent binding site, molecular dynamic simulations were performed with different ion compositions to probe preferential ion binding. The GAC RNA was solvated in TIP3P water with total added 28  $\text{Mg}^{2+}$  (neutralizing amount) and 28  $\text{K}^+$  ions (plus 28 neutralizing  $\text{Cl}^-$ ). The ions were initially located at random positions within the simulation box with the minimum distance of 6 Å from any RNA atom and 4 Å from each other. Because of the long exchange rates for  $\text{Mg}^{2+}$  ions (29), it is not possible to converge the  $\text{Mg}^{2+}$  exchange events with a single simulation given the current simulation time scales we are able to achieve. Therefore, we increased the number of simulation copies and simulation length per copy to ideally improve the sampling within the capabilities of available resources. Twenty copies of the simulation were generated with initial placement of the ions at different random positions. 1  $\mu\text{s}$  of production simulation was performed for each copy as described in the Methods section. To better understand the influence of the ion binding on the tertiary structure, as opposed to the folding, the RNA crystal conformation was restrained with 0.5 kcal/mol-Å<sup>2</sup> positional restraints on all RNA atoms (COMPETITION\_1 set).

Figure 3.3 illustrates the proximity of the ion to the ion binding pocket by plotting the distance of the closest potassium and magnesium to the center of mass of the phosphate atoms of A1070, C1072 and A1073 in the tight ion binding pocket. The distance of the bound potassium to this center of mass point is 0.97 Å in the crystal structure (1HC8), which, as can be seen in Figure 3.3, is not rigidly maintained in the simulations due to fluctuations in the position of the ion and the residues in the binding pocket. Starting with an empty binding pocket, in most cases, the binding pocket became occupied by

potassium during the equilibrium stages, although due to the ion randomization in the initial structures, 8 copies had closer  $\text{Mg}^{2+}$  ion to the pocket and 12 copies had closer  $\text{K}^+$  ion to the pocket in their initial structures and the distances to the closest ions of different types were very similar. However, among 20 copies, the pocket was occupied by magnesium during only one simulation (shown as cyan). This ratio implies that the occupancy of the pocket by  $\text{K}^+$  is preferred over  $\text{Mg}^{2+}$ , but the difference between free energies of binding of two ions is not high enough to completely prevent  $\text{Mg}^{2+}$  from binding.

Simultaneous occupancy of K58 and Mg163 positions occurred in only two copies (shown as indigo and black). Interestingly, in one of these copies (black), the chelated magnesium stayed closer to the K58 pocket and the potassium pocket remained only partially occupied by a potassium that remained close to the pocket, but far from the bound magnesium. The buried pocket was also partially occupied in the simulation copy indicated as brown. The violet copy in the bottom of Figure 3.3 indicates occupancy of Mg167 site by magnesium while the buried pocket was occupied by potassium. In summary, the buried ion binding pocket was occupied by magnesium in only one of the simulation copy, and only two other copies showed simultaneous occupancy of K58 and Mg163 (Figure 3.4).

Similar ion-competition simulations were repeated for 500 ns but without any restraints on the RNA in order to capture any potential differences in the occupancy of the binding pocket caused by subtle RNA conformational changes (COMPETITION\_2 set). The results (Figure 3.5) show almost similar statistics for ion occupancy in the K58 and Mg163 sites although there is more flexibility in the position of the bound  $\text{K}^+$ , which is reasonable due to the greater flexibility of the RNA conformation.

The bound  $K^+$  is located within 2.7-3.5 Å distance from six oxygen atoms in the crystal structure (96) as illustrated in Figure 3.2. Figure 3.6 shows the distributions of the ion distances in the pocket to these oxygen atoms in the COMPETITON\_1 simulations to analyze if the ion retains its exact crystallographic position within the pocket or shifts position towards the sides. According to the short-distance major peaks in Figure 3.6, the location of the bound  $K^+$  during the RNA-restrained simulations was similar to the crystal structure with some flexibility in the binding pocket. Sharper short distance peaks for OP2 atoms of A1070 and C1072 compared to wider and bimodal distributions for other four atoms implies the tendency of the bound  $K^+$  to shift toward the Mg163 binding site within its own pocket while Mg163 is absent in its crystallographic position. This shift is also seen in Table 3.1, which compares the crystallographic distance of the  $K^+$  in the pocket to the pocket oxygen atoms with average of those distances generated by the simulations. The position of the bound  $K^+$  in the unrestrained competitive simulations is also similar to that of restrained simulations (Figure 3.7.) although the peaks for  $K^+$  distance from O3' of A1060 and O5' of C1072 are less sharp.

The absence of a bound  $Mg^{2+}$  in position of Mg163 in most simulation copies and the simultaneous shift of the bound potassium towards this position in these simulations imply that the crystallographic occurrence of the Mg163 might be a result of the crystal conditions. When Mg163 position lacks  $Mg^{2+}$ , the bound potassium might tend to shift towards this position rather than being in a middle of the buried binding pocket. However, the possibility of  $K^+$  being in different positions within the binding pocket cannot be justified based on the bimodal distribution of ion distances from C1072 and A1073. In fact, such observation can be a force field and/or convergence artifact related to the simulations. Currently no experimental evidence is available to support it.

In the crystal structure (96), the  $K^+$  binding pocket is isolated from the solvent by RNA. Therefore, it is feasible that the ion is partially or completely dehydrated. No water density is visible around the bound  $K^+$ . In solution and in simulations with TIP3P waters,  $K^+$  has 5/6 waters in its first hydration shell. During the ion competition simulations, the bound potassium was often only partially dehydrated with 3 or 4 water residues in its first solvation shell (Figure 3.8.), implying that the pocket can accommodate a potassium ion together with some of its first shell waters, allowing the bound potassium to move. The crystal structures do not contain any waters in the pocket, but fully dehydrating a potassium or magnesium upon binding is energetically costly, and hence, very unlikely.

For further investigation of the role of hydration, we set up 20 copies of short MD simulations (10 ns) using the 1HC8 structure to monitor the possible diffusion of the water molecules into the pocket (HYDRATION set). The RNA and K58, Mg163 and Mg167 were restrained to their crystallographic positions. The results show that (Table 3.1), the water molecules can leak into the pocket and directly interact with  $K^+$  (Figure 3.8). However, the comparison between two histograms in Figure 3.8. reveals that the greater chance of having more waters in the pocket and around the ion in ion competition simulations is probably a result of ion displacement from the pocket. Yet, the presence of 3-4 water residues in both simulations supports our hypothesis that the ion in the pocket is partially dehydrated even in its crystal conditions (noting that the lack of ion hydration in the crystal could be simply due to its omission during the refinement).

### 3.3.2 Ion pull-out umbrella sampling

Although the observed preference of potassium over magnesium in occupying the binding site in these simulations may be caused by thermodynamic favorability of

potassium in the pocket, one cannot rule out the possibility that the higher kinetic barrier for magnesium dehydration has prevented it from populating the pocket within the simulation timescale. To distinguish between these scenarios, we calculated the free energy needed to pull out different ions from the potassium binding site to the bulk solvent using umbrella sampling (see Methods for details). The crystallographic magnesium ions were retained in all experiments and 0  $\text{Mg}^{2+}$ , 15 mM  $\text{Mg}^{2+}$  and 80 mM  $\text{Mg}^{2+}$  were added to independent MD simulation experiments. 80 mM  $\text{Mg}^{2+}$  is similar to the crystallization conditions of the 1QA6 and 1HC8 structures (97).

A series of molecular dynamics simulations were performed during which the ion (Na, K, or Mg) was restrained to set positions along a reaction coordinate. The reaction coordinate was the distance of the ion to the center of mass of nine phosphate groups of A1070, C1072 and A1073 (phosphorus and two nonbridging oxygen atoms of each residue), referred to hereafter as the “anchor point.” The crystallographic distance on this reaction coordinate for the bound  $\text{K}^+$  is 0.97 Å in the 1HC8 crystal structure.

To prepare the initial structures for each umbrella, unrestrained MD was performed in three to six copies for each system (MG0, MG15 and MG80 sets) to capture ion exit events as the inserted ion in the pocket leaves the RNA. One ion exit was observed during 1  $\mu\text{s}$  simulation in 2 copies of MG0 and 4 copies of MG15 and MG80.

The trajectory frames in which the leaving  $\text{K}^+$  was less than 6.9 Å from the anchor point were extracted from a given simulation to be used as initial structures for related windows. After 6.9 Å, the ion was free and well exposed to the solvent, and was directed to 20 Å (bulk solvent) with a 130 ps simulation. Frames from this simulation were used to obtain the initial umbrella structures from 6.9 to 20 Å. Umbrellas were spaced 0.1 Å from each other. A total of 2 ns of molecular dynamics simulation was performed for

each umbrella. The ion was restrained using a harmonic distance restraint with 100 kcal/mol-Å<sup>2</sup> force constant for each umbrella. Such a large restraint was required especially at some distance ranges close to the RNA to keep the ion in the desired window. The ion was also directed from the lowest umbrella to 0.5 Å on the reaction coordinate with a series of 2 ns umbrella simulations in which the last frame of a simulation was used as the initial frame of the next.

This GAC RNA adopts its tertiary structure in 0.1 mM Mg<sup>2+</sup> with 100 mM K<sup>+</sup>; 5 mM Mg<sup>2+</sup> with 100 mM Na<sup>+</sup>; or 5 mM Mg<sup>2+</sup> with 100 mM NH<sub>4</sub><sup>+</sup> (114). Here, we replace K<sup>+</sup> in the crystallographic pocket with Mg<sup>2+</sup> and Na<sup>+</sup>, to look directly at the free energy of different ion occupancy in the pocket. If this pocket is the site of specific Mg<sup>2+</sup> or K<sup>+</sup> binding, then their binding free energies should correlate with experimental solution data (96,99). The sampling was repeated with mutating the leaving K<sup>+</sup> to Mg<sup>2+</sup> (in all scenarios) and Na<sup>+</sup> (in the scenario with no excess Mg<sup>2+</sup>). According to the experiments done by Wang Y.X. et. al., not only is a specific monovalent ion interaction needed for folding of a similar construct of the GAC RNA, but K<sup>+</sup> stabilizes its tertiary structure better than Na<sup>+</sup> (93,94).

To see if this preference is observed in the K<sup>+</sup> pocket of 1HC8 structure, we also set tested Na<sup>+</sup> in ion pull-out simulations. This would support the hypothesis that this pocket is the actual monovalent ion-binding pocket, which is expected from those experiments. Also, to examine the effect of Mg163 on the binding of each ion to the monovalent ion pocket, the Mg163 was moved to a far random position in the initial structure of each window and the simulations were repeated. The exception was in the system with no excess Mg<sup>2+</sup> in which Mg163 had already dissociated before the K<sup>+</sup> ion exchange event.

In the resulting PMF plots (Figure 3.9), the free energy of binding for each ion can be



calculated as the difference between the PMF value at 20 Å and in the pocket-related well at < 3 Å. All plots use a common and arbitrary zero of free energy of the free ion at 20.0 Å from the ion binding pocket. Slightly different minima on the reaction coordinate for different ions highlights the point that they have different binding modes within the pocket. The PMFs for binding in the absence of Mg163 suggest that the K<sup>+</sup> ion is binding tighter as the concentration of the magnesium increases, however, this is likely an artifact of the zero of the PMF; instead, likely it is destabilization of the unbound state (rather than tighter binding) due to the increased ionic density.

PMF calculations show that binding of Mg<sup>2+</sup> to the monovalent binding pocket is always energetically less favorable than binding of K<sup>+</sup>. Binding of Mg<sup>2+</sup> is energetically less favorable than binding of Na<sup>+</sup> in the case with no excess Mg<sup>2+</sup>. These data support the hypothesis that the pocket is specific for monovalent ions (96) and agree with the experimental observations that K<sup>+</sup> stabilizes the GAC RNA tertiary structure better than a bound Na<sup>+</sup> (93,94).

In presence of Mg<sup>2+</sup> in Mg163 position, it is unfavorable for Mg<sup>2+</sup> and K<sup>+</sup> ions to bind to the pocket. This agrees with the observation in ion-competition simulations that show crystallographic localization of Mg163 and K58 is mutually exclusive. However, in 80 mM Mg<sup>2+</sup> simulations, binding of potassium inside the pocket and in the bulk solvent is almost iso-energetic (when Mg163 occupies the crystallographic binding site).

In absence of Mg<sup>2+</sup> in Mg163 position and with no excess Mg<sup>2+</sup> ions, the binding of potassium to the pocket is about -2 kcal/mol favorable, the binding of sodium is almost iso-energetic inside the pocket and in the bulk water, and there is about +3 kcal/mol energetic penalty for the binding of magnesium. However, the free energy difference for having K<sup>+</sup> versus Mg<sup>2+</sup> in the absence of Mg<sup>2+</sup> in the Mg163 position gets smaller with

increasing bulk  $\text{Mg}^{2+}$  concentration. Therefore, mixed occupancy of  $\text{K}^+$  and  $\text{Mg}^{2+}$  in this pocket in the whole ensemble seems possible especially at higher  $\text{Mg}^{2+}$  concentrations.

Since the binding site is buried inside the RNA, the ions do not go through a smooth path when leaving it and so the PMF plots look rough. The small bumps on the PMF plots represent transient interaction of ions with other RNA atoms along their path. There are several other binding modes near the monovalent ion binding site at 3 and 4 Å on the reaction coordinate. These binding modes can be described as chelation of the ions to a nonbridging oxygen atom of U1097 on their path to the bulk solvent. However, when comparing the PMF plots of each ion type in different scenarios, it should be considered that the paths were generated from independent simulations, and therefore, the same points on the same ion type plots in different scenarios might not exactly represent the same ion-RNA interactions.

We expect the free energy values would vary to some extent with using different ion models (21-23). To test how the magnesium binding to this pocket might be affected by different magnesium models, a magnesium pulling experiment was repeated. We chose conditions with no excess magnesium scenario, using crystallographic initial structures.  $\text{Mg}^{2+}$  parameters included the traditional Åqvist parameters (21) and HFE and IOD parameters developed by Li et al. (23). The two latter sets are designed to reproduce correct hydration free energy and correct ion-oxygen distance, respectively. Ion dehydration energy and ion-oxygen interactions are two determining elements in ion binding to the GAC pocket (96).

PMF plots for these four magnesium models are shown in Figure 3.10. The HFE model causes an increase in the dehydration barrier for the ion to enter the binding site without a significant change in the binding free energy. The IOD and Allnér-Nilsson-

Villa models give very similar results. Åqvist parameters have a small effect on the dehydration barrier, but enhance binding energy by about -1 kcal/mol. The calculated PMFs for the different  $\text{Mg}^{2+}$  models vary in magnitude, but none exceed the binding free energy of  $\text{K}^+$  ions.

### **3.4 Conclusions**

Both ion competition molecular dynamics and ion pull out umbrella sampling simulations agree that binding of potassium is preferred over magnesium in the binding site (Figure 3.3, 3.5, 3.9). The simulations that facilitate competition between  $\text{K}^+$  and  $\text{Mg}^{2+}$  show very fewer  $\text{Mg}^{2+}$  binding. However,  $\text{Mg}^{2+}$  is not completely blocked from the monovalent pocket, which is in agreement with the PMF result as the difference of free energy of binding between  $\text{K}^+$  and  $\text{Mg}^{2+}$ , and is not huge in absence of Mg163. This supports the identification of potassium in this pocket as described by Conn et al. (96) rather than magnesium as suggested by Wimberly et al. (92). Also, preference of  $\text{K}^+$  over  $\text{Na}^+$  in this pocket in low  $\text{Mg}^{2+}$  concentration ion pull-out simulations agrees with the trend in melting experiments observed by Lu and Draper (114) and Wang et al. (93) where  $\text{K}^+/\text{Mg}^{2+}$  more efficiently stabilized the GAC tertiary structure than other monovalent and divalent ions (93,114). This agreement further supports the hypothesis that the K58 position is the monovalent ion-binding site.

Since Wimberly et al. have not discussed the details for  $\text{Mg}^{2+}$  identification in the K/Mg58 position as their crystal structure shows, it is possible that the ion has been misidentified in that position; while more elaborate experiments have led to accurate identification of  $\text{K}^+$  in that position later by Conn et al. (96) Further support for such conclusion comes from the fact that the closest coordinating oxygen atom in 1MMS (OP1

of A1073)(92) is located 2.6 Å away from the chelated  $\text{Mg}^{2+}$  (Table 3.1), while the mean experimental value for  $\text{Mg}^{2+}$ -O distances in RNA structures is 2.08 Å according to the Cambridge Structural Database (115,116) and distances longer than 2.4 Å are suspected to be misidentified (116). However, based on the results from the ion pull-out simulations and the free energy difference between  $\text{K}^+$  vs.  $\text{Mg}^{2+}$  for their binding to the pocket (Figure 3.9), it is not surprising if magnesium is observed in the pocket at higher magnesium concentrations which are used in crystallization conditions. But it is important to note that the free energy difference would dictate potassium binding at lower concentrations such as those which have been used in melting experiments; and therefore, it is very likely that the chelated monovalent ion described in those experiments have been actually bound to this pocket (93,94).

The MD simulations suggest that occupancy of the Mg163 and K58 positions by related ions are mutually exclusive at lower  $\text{Mg}^{2+}$  concentrations. Occupancy of the Mg163 site in the presence of potassium in the pocket is more likely to be caused by high magnesium concentration in crystallography conditions.

In the cell,  $\text{K}^+$  concentration is typically estimated at ~150 mM and free  $\text{Mg}^{2+}$  at 1-2 mM. Our pulling experiments suggest that in these solution conditions, the Mg163 site will be unoccupied, while K58 is stably bound. Regardless of magnesium concentration and magnesium Mg163 occupancy and utilized magnesium model, potassium binding is always more favorable than magnesium binding to the monovalent binding site. In the absence of an  $\text{Mg}^{2+}$  ion, the pocket will be more accessible to water molecules. Mobile water molecules could relax the pocket geometry, allowing more mobility of the nucleotides and also exchange of the bound  $\text{K}^+$ .

Table 3.1. Ion distances in the pocket. Distance of the closest  $K^+$  and the oxygen atoms in the monovalent pocket from 20  $\mu s$  aggregate trajectory of ion competition molecular dynamics (COMPETITION\_1 set) and 200 ns aggregate trajectory of ion hydration simulations (HYDRATION set) using the 1HC8 crystal structure. Errors are standard deviations.

Pocket oxygen atoms	Crystallographic distance according to 1HC8 ( $\text{\AA}$ )	Crystallographic distance according to 1MMS ( $\text{\AA}$ )	Average distance ( $\text{\AA}$ ) in ion competition simulations	Average distance ( $\text{\AA}$ ) in ion hydration simulations
A1069@O3'	3.2	3.0	$3.7 \pm 1.1$	$3.2 \pm 0.2$
A1070@OP2	2.7	2.8	$3.1 \pm 0.9$	$3.0 \pm 0.2$
C1072@OP1	2.9	2.9	$3.2 \pm 0.7$	$3.0 \pm 0.2$
C1072@OP2	3.3	3.4	$3.2 \pm 0.8$	$3.0 \pm 0.2$
C1072@O5'	3.5	3.8	$4.1 \pm 0.6$	$3.4 \pm 0.2$
A1073@OP1	2.8	2.6	$4.6 \pm 1.7$	$2.7 \pm 0.1$

Table 3.2. List of the simulations done in this work. \* US: Umbrella Sampling.

Set	Type	Condition	Length	Purpose
COMPETITION_1	Classic MD	28 K <sup>+</sup> , Mg <sup>2+</sup> , Cl <sup>-</sup> / restrained RNA	1 $\mu$ s per 20 copies	Ion competition in occupying K/Mg58 pocket
COMPETITION_2	Classic MD	28 K <sup>+</sup> , Mg <sup>2+</sup> , Cl <sup>-</sup> / unrestrained RNA	0.5 $\mu$ s per 20 copies	Ion competition in occupying K/Mg58 pocket
HYDRATION	Classic MD	28 K <sup>+</sup> , Mg <sup>2+</sup> , Cl <sup>-</sup> with ions in crystallographic sites of K58, Mg163 and Mg167/ restrained RNA and crystallographic ions	10 ns per 20 copies	Hydration of K58 in crystallographic position
MG0	US*	Crystallographic ions, neutralizing K <sup>+</sup> , 100 mM excess KCl, no excess MgCl <sub>2</sub>	1 $\mu$ s per 3 copies	Tracking leaving K <sup>+</sup>
MG15	US*	Crystallographic ions, neutralizing K <sup>+</sup> , 100 mM excess KCl, 15 mM excess MgCl <sub>2</sub>	1 $\mu$ s per 6 copies	Tracking leaving K <sup>+</sup>
MG80	US*	Crystallographic ions, neutralizing K <sup>+</sup> , 100 mM excess KCl, 80 mM excess MgCl <sub>2</sub>	1 $\mu$ s per 6 copies	Tracking leaving K <sup>+</sup>
US_MG0_K	US*	MG0 conditions / leaving K <sup>+</sup>	2 ns per 196 umbrellas	PMF calculation
US_MG15_K	US*	MG15 conditions / leaving K <sup>+</sup>	2 ns per 196 umbrellas	PMF calculation
US_MG80_K	US*	MG80 conditions / leaving K <sup>+</sup>	2 ns per 196 umbrellas	PMF calculation

Table 3.2 continued

Set	Type	Condition	Length	Purpose
US_MG0_MG	US*	MG0 conditions / leaving Mg <sup>2+</sup>	2 ns per 196 umbrellas	PMF calculation
US_MG15_MG	US*	MG15 conditions / leaving Mg <sup>2+</sup>	2 ns per 196 umbrellas	PMF calculation
US_MG80_MG	US*	MG80 conditions / leaving Mg <sup>2+</sup>	2 ns per 196 umbrellas	PMF calculation
US_MG0_NA	US*	MG0 conditions / leaving Na <sup>+</sup>	2 ns per 196 umbrellas	PMF calculation
US_MG15_K_163	US*	MG15 conditions / leaving K <sup>+</sup> and no Mg163	2 ns per 196 umbrellas	PMF calculation
US_MG80_K_163	US*	MG80 conditions / leaving K <sup>+</sup> and no Mg163	2 ns per 196 umbrellas	PMF calculation
US_MG15_MG_163	US*	MG15 conditions / leaving Mg <sup>2+</sup> and no Mg163	2 ns per 196 umbrellas	PMF calculation
US_MG80_MG_163	US*	MG80 conditions / leaving Mg <sup>2+</sup> and no Mg163	2 ns per 196 umbrellas	PMF calculation
US_MG0_MG_AQV	US*	MG0 conditions / leaving Mg <sup>2+</sup> with modified Mg <sup>2+</sup> parameters	2 ns per 196 umbrellas	PMF calculation
US_MG0_MG_HFE	US*	MG0 conditions / leaving Mg <sup>2+</sup> with modified Mg <sup>2+</sup> parameters	2 ns per 196 umbrellas	PMF calculation
US_MG0_MG_IOD	US*	MG0 conditions / leaving Mg <sup>2+</sup> with modified Mg <sup>2+</sup> parameters	2 ns per 196 umbrellas	PMF calculation

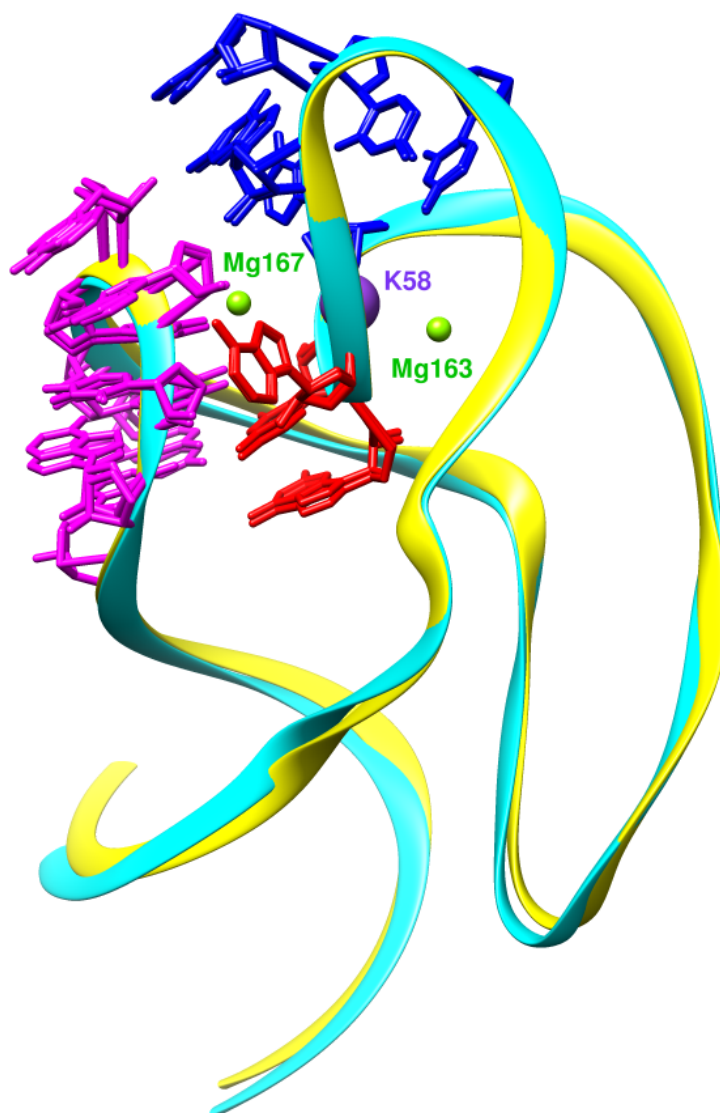


Figure 3.1. Overlay of the GAC RNA crystal structures 1HC8 (cyan ribbon) and 1MMS (yellow ribbon). The bulged region of A1070, G1071 and C1072 (red atoms and bonds) is clamped between the U-turns U1065-A1069 (blue atoms and bonds) and the hairpin C1092-G1099 (magenta atoms and bonds). Atoms and bonds of both structures are illustrated. The Mg163 and Mg167 (green spheres) have the same identity in two structures, but K58 (purple sphere) is identified as potassium in 1HC8 and as magnesium in 1MMS. Atoms of other residues and hydrogen atoms are not shown for clarity. The molecular graphics were generated with Chimera 1.9 (117).



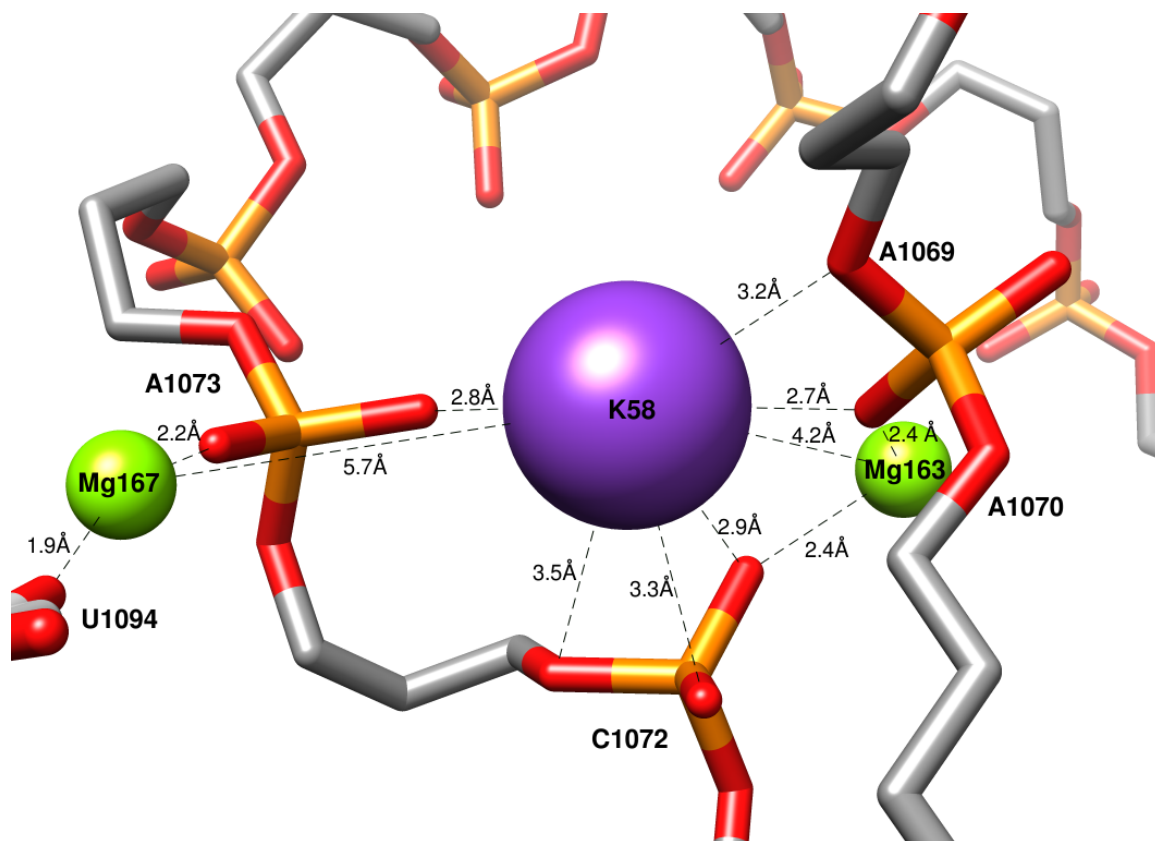


Figure 3.2. Potassium binding pocket according to Conn et al. (PDB entry 1HC8) (96). The potassium is represented as a purple sphere, and magnesium ions as green spheres. The nucleic acid backbone is in licorice representation, with phosphorus atoms in yellow, carbon atoms in grey and oxygen atoms in red. Base atoms are not shown for clarity, except U1094 whose O4 is bound to Mg167. The bound potassium is directly chelated to six phosphate oxygen atoms A1069@O3', A1070@OP2, C1072@O5', OP1, OP2 and A1073@OP1 (@ denotes atom). Mg163 directly bridges between A1070@OP2 and C1072@OP1 and is located at 4.2 Å distance from the bound potassium. Mg167 is directly bound to A1073@OP2 and U1094@O4 and is located at 5.7 Å distance from the bound potassium (OP1 and OP2 are nonbridging oxygen atoms of the nucleotides). The figure is generated with Chimera 1.9 (117).

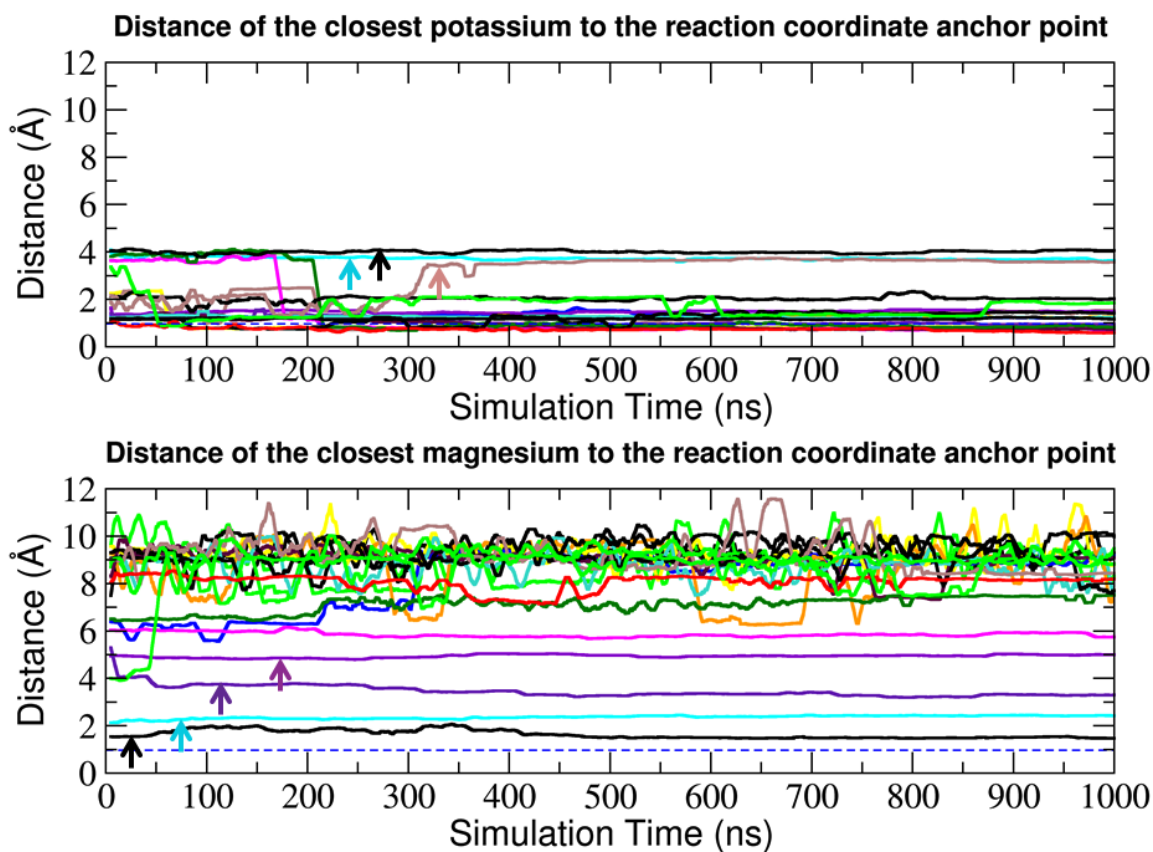


Figure 3.3. Distance of the closest  $K^+$  (top) and  $Mg^{2+}$  (bottom) to the center of mass of the phosphorus and two non-bridging oxygen atoms of A1070, C1072 and A1073 (named as the reaction coordinate anchor point) during the ion competition simulations with restrained RNA (COMPETITION\_1). Each line represents running averaged data over 10 ns windows for one of the 20 copies. The blue dashed line represents the crystallographic distance to the bound potassium. The arrows are markers for points of interest that are further discussed in the text.

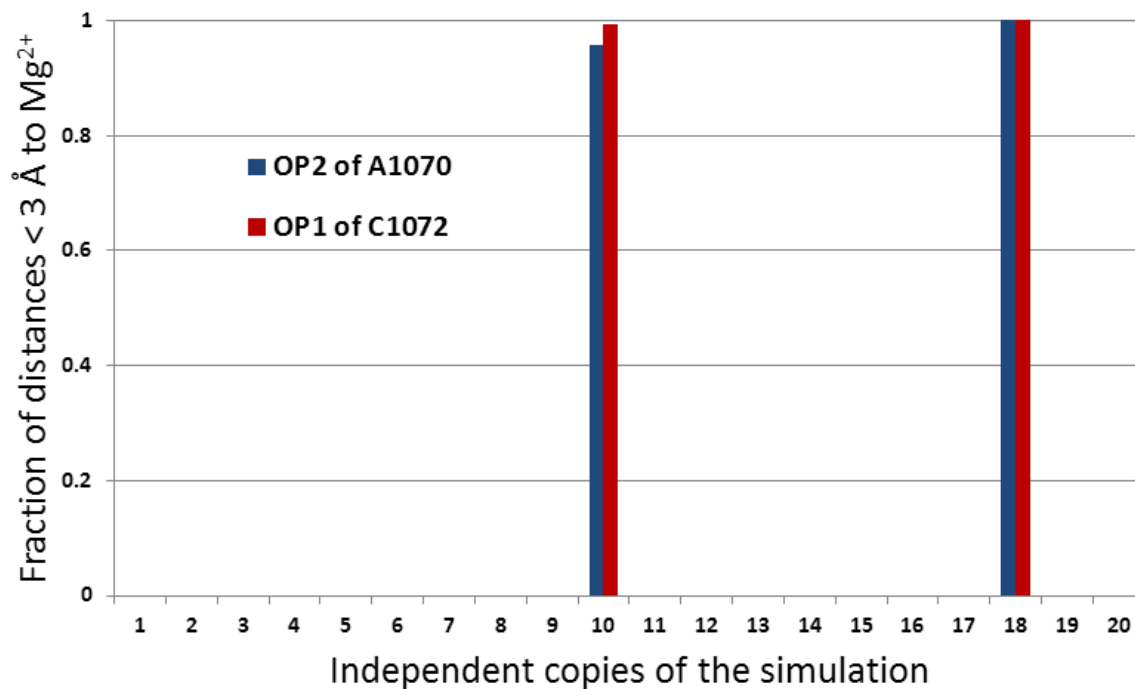


Figure 3.4. Occupancy of the Mg163 site as measured via fraction of direct binding (distances less than 3) between two oxygen atoms at Mg163 site and any  $\text{Mg}^{2+}$  ion in the 20 copies of restrained ion competition (COMPETITION\_1) simulations. Only 2 copies out of 20 show occupation of Mg163 with magnesium, while the site remains empty of magnesium in other copies. In copy 18, the  $\text{K}^+$  was shifted away from the Mg163 site. The fractions were calculated using hbond analysis in CPPTRAJ program.

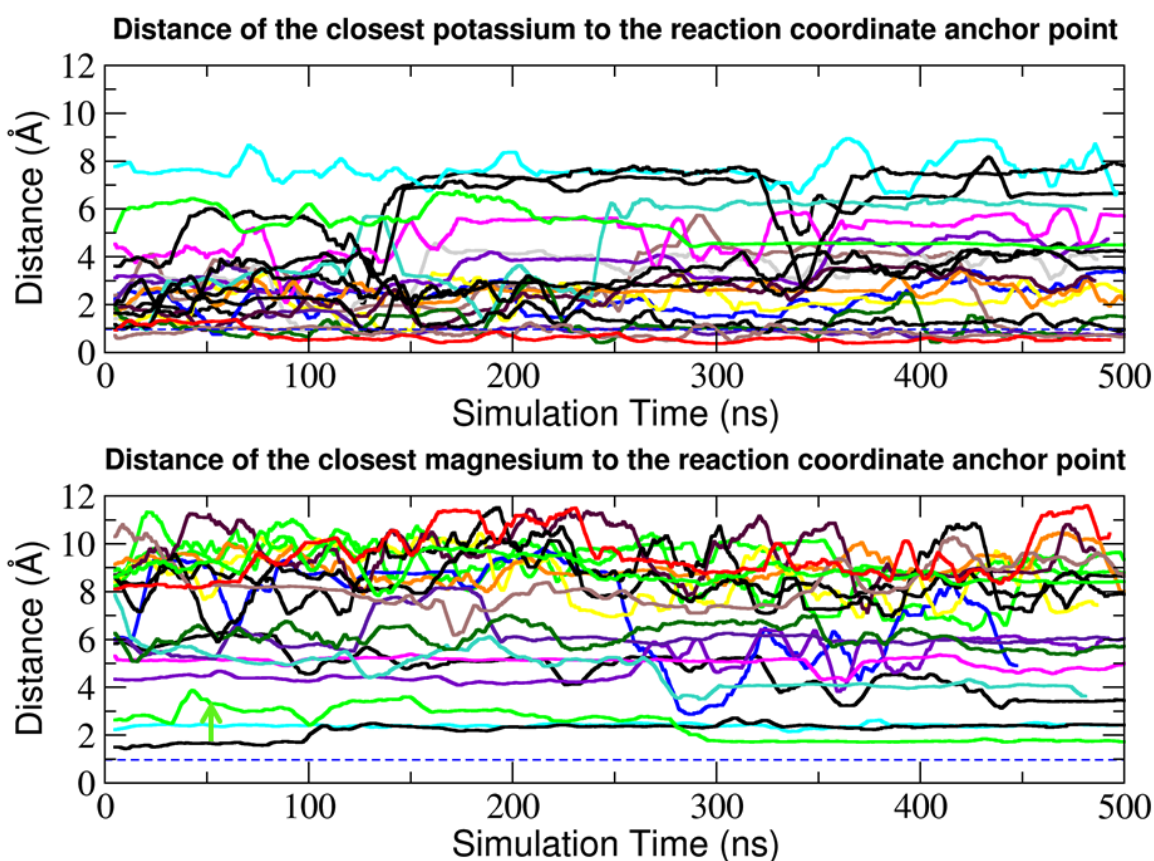


Figure 3.5. Distance of the closest  $K^+$  (top) and  $Mg^{2+}$  (bottom) to the center of mass of the phosphorus and two non-bridging oxygen atoms of A1070, C1072 and A1073 (named as reaction coordinate anchor point) during the ion competition simulations with unrestrained RNA. Each line represents running averaged data with 10 ns window for one of the 20 copies. The blue dashed line represents the crystallographic distance to the bound potassium. The simulation copy indicated by green arrow is the only one other than those discussed in RNA-restrained simulations which showed occupancy of Mg163 with a chelated magnesium ion. The bound potassium in K58 site is shifted away from the chelated magnesium in this copy as well.

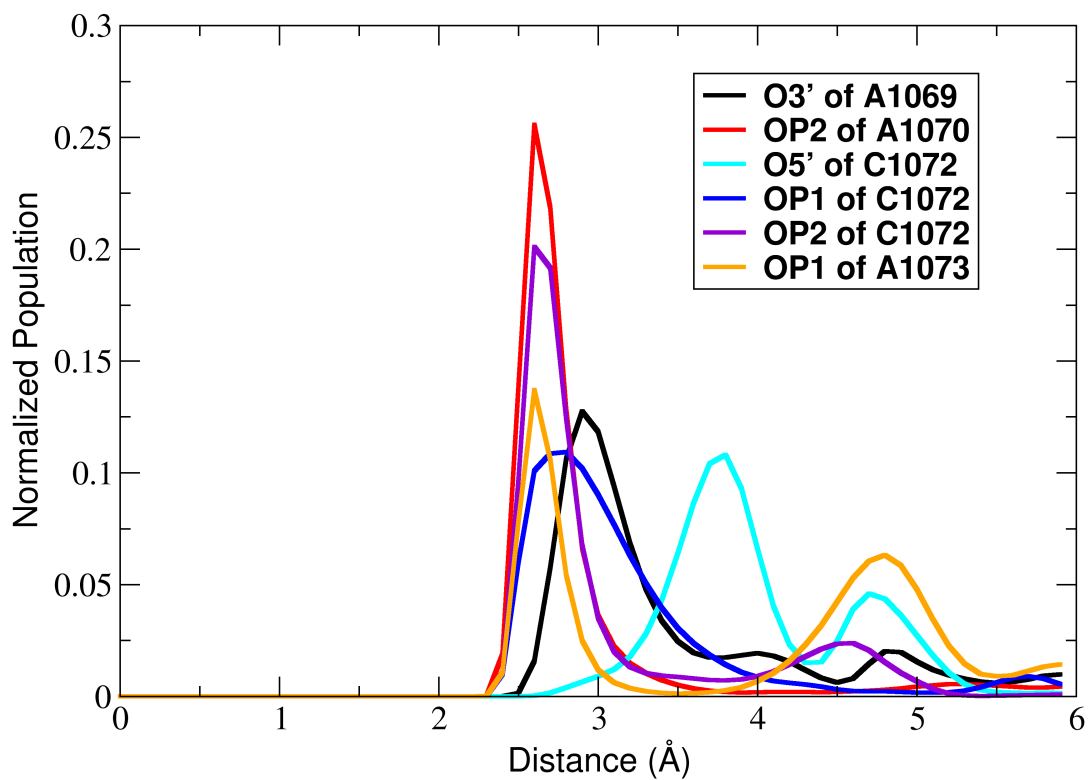


Figure 3.6. Histograms of distances of the bound  $K^+$  in the pocket to the six crystallographic interacting oxygen atoms in 20  $\mu$ s aggregate trajectory of RNA-restrained ion competition simulations (COMPETITION\_1).

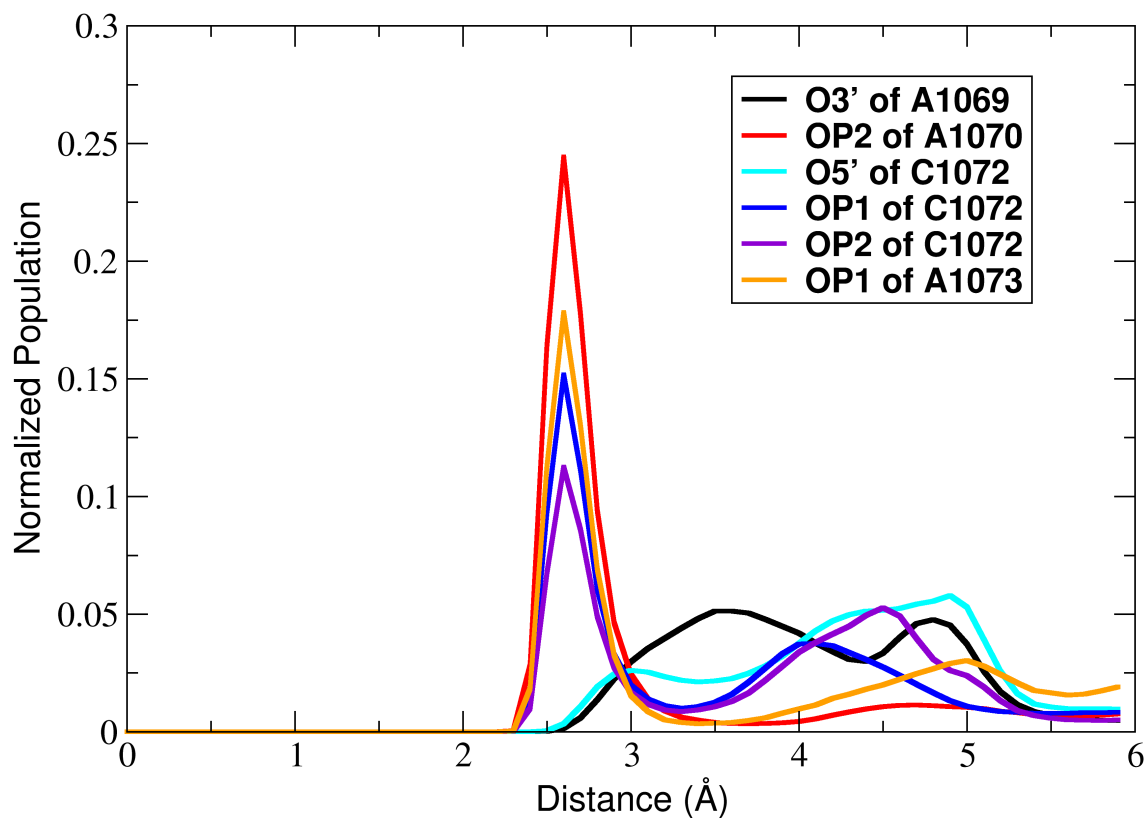


Figure 3.7. Histograms of distances of the bound  $K^+$  in the pocket to the six crystallographic interacting oxygen atoms in 20  $\mu s$  aggregate trajectory of unrestrained ion competition simulations (COMPETITION\_2).

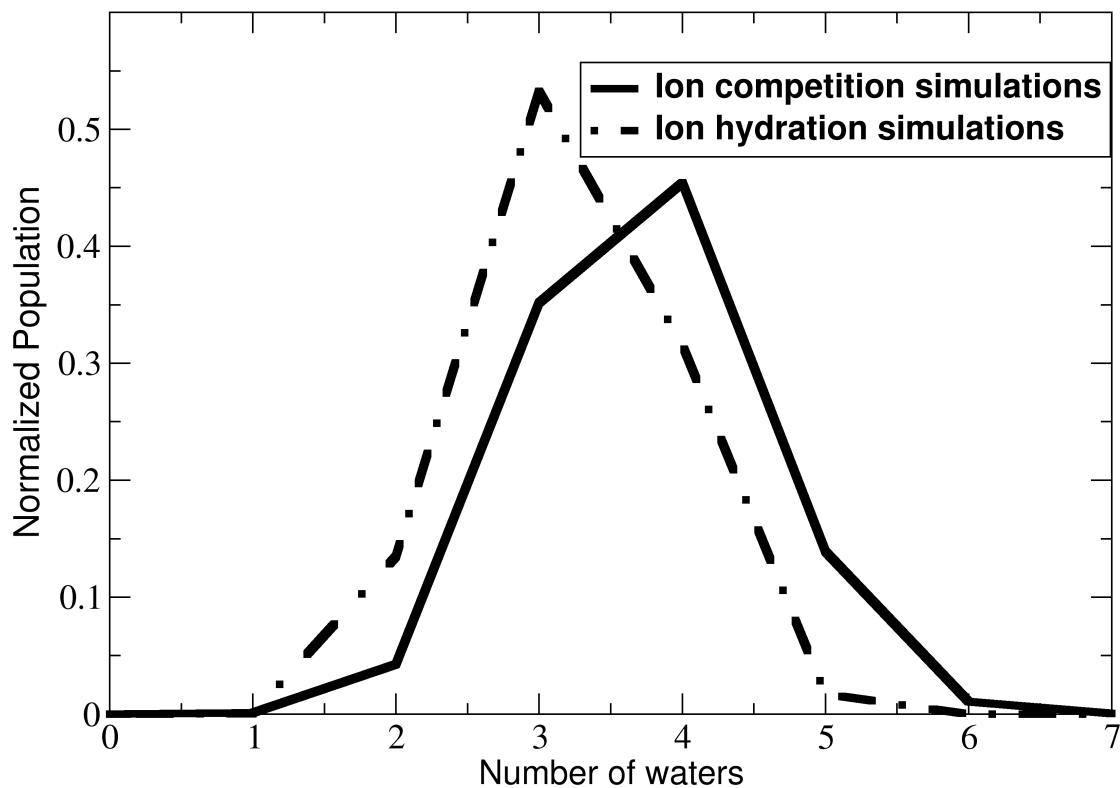


Figure 3.8. Distribution of waters in the first solvation shell of the bound  $K^+$  in 20  $\mu s$  aggregate trajectory of ion competition simulations and 200 ns aggregate trajectory of ion hydration simulations.

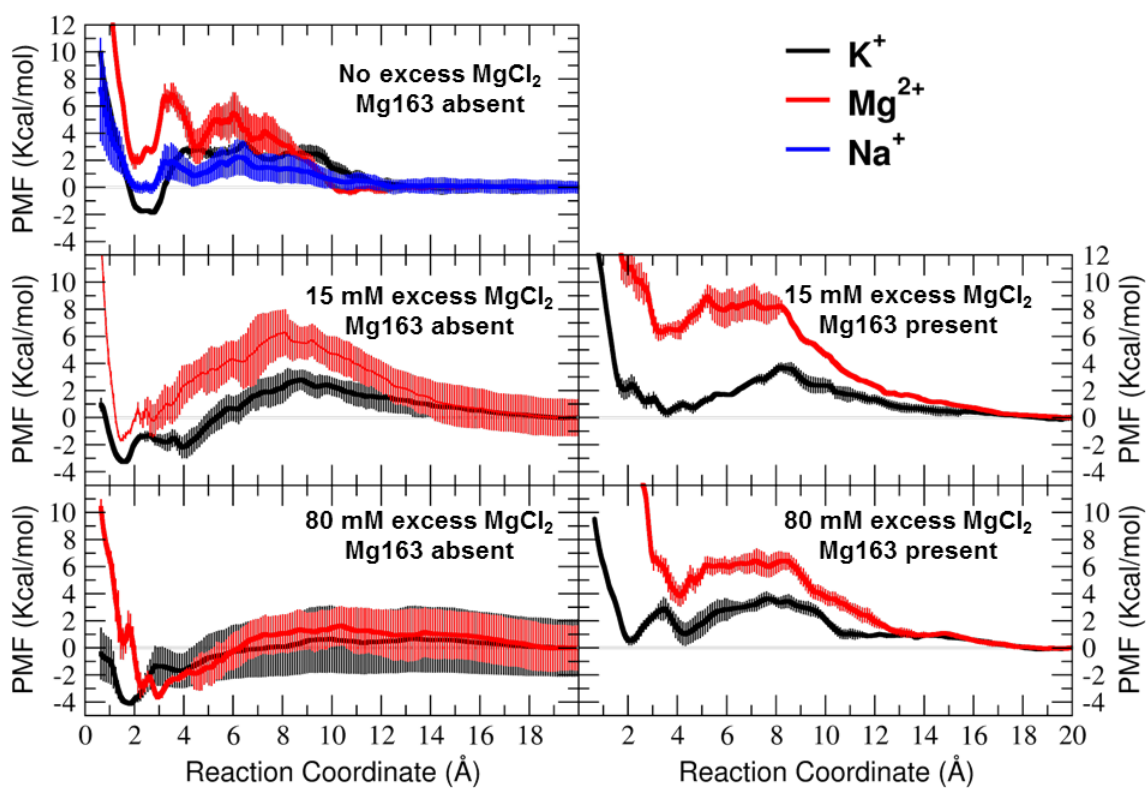


Figure 3.9. PMF profiles for pulling out  $K^+$ ,  $Mg^{2+}$  and  $Na^+$  from the GAC RNA monovalent ion binding site. Reaction coordinate represents the distance between the ion and anchor. The PMF plots were generated using the second 1 ns window in each umbrella. All the PMF plots were arbitrarily shifted to zero relative to their magnitude at 20.0 Å.



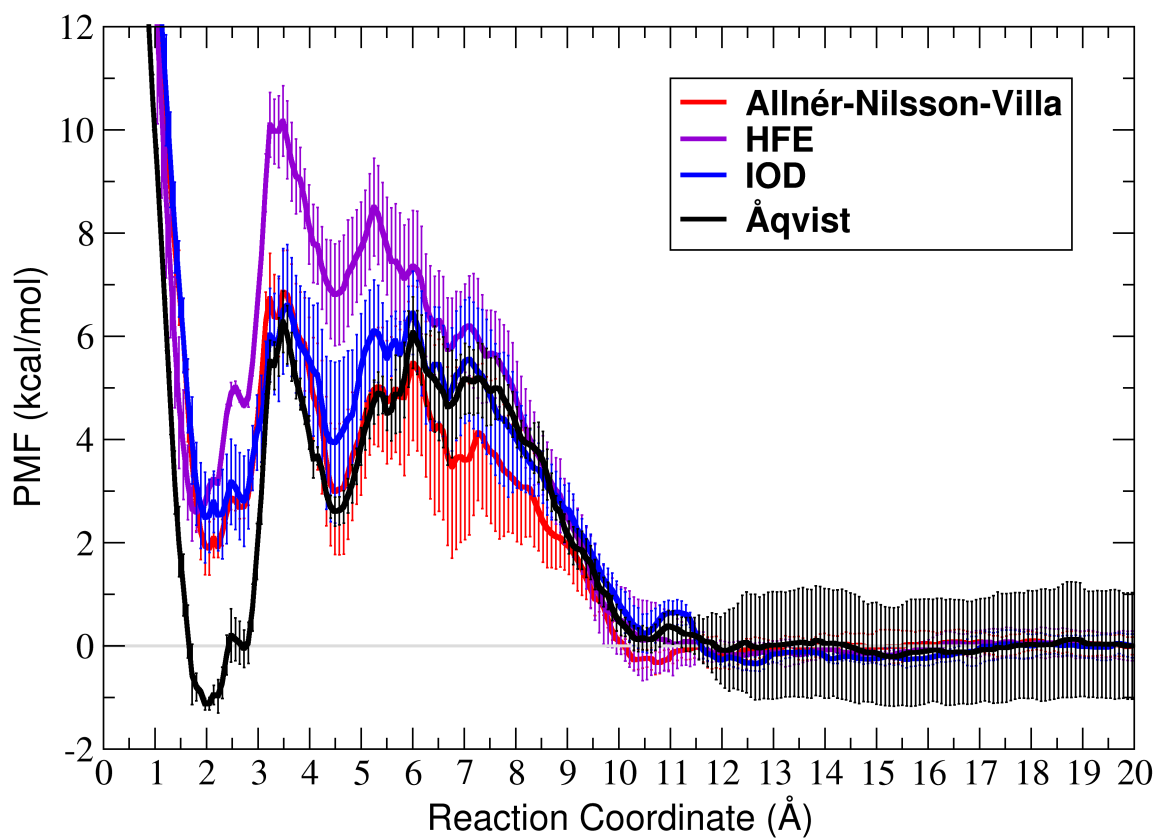


Figure 3.10. PMF profiles for pulling out  $Mg^{2+}$  from the GAC RNA monovalent ion binding site with different magnesium models. PMF plots are calculated with the same conditions described in Figure 3.9.

## CHAPTER 4

### INVESTIGATING THE ION-DEPENDENCE OF THE FIRST UNFOLDING STEP OF GTPASE-ASSOCIATING CENTER RIBOSOMAL RNA

This research project has been designed by Hamed S. Hayatshahi and Thomas E. Cheatham, III. The chapter was primarily written by H. S. Hayatshahi and edited by Christina Bergonzo and T. E. Cheatham, III.

#### **4.1 Introduction**

The folding of RNA molecules into their functional structures depends on their interaction with diffuse, associated, and bound cations that neutralize the RNA phosphate groups and trigger the formation of complex tertiary structures (80,82). In bulk solvent, diffuse “hydrated” cations can transiently interact with RNA atoms. In contrast, bound or chelated ions bind to the RNA atoms directly, a process that involves displacement of one or more water residues in the first hydration shell of the ion and RNA. This partial dehydration requires overcoming a considerable energetic barrier (~12 kcal/mol for removing one water molecule from the first shell of magnesium (22)) in the process of ion binding, especially for magnesium ((82).

The 58-nucleotide RNA in the GTPase-Associating Center (GAC RNA) has been extensively investigated to better understand RNA-ion interactions (92-99). This highly

conserved RNA structure interacts with the L11 protein in the 23S ribosomal subunit to facilitate GTP hydrolysis during transcription (118). Thermal unfolding experiments have shown that a bound monovalent ion (with preference for  $\text{NH}_4^+$  and  $\text{K}^+$  over other monovalent cations such as  $\text{Li}^+$ ,  $\text{Na}^+$ ,  $\text{Rb}^+$  and  $\text{Cs}^+$ ) is crucial in stabilizing the tertiary structure of the GAC RNA (93,94). According to UV absorbance experiments, thermal unfolding of GAC RNA involves two distinct transitions each at different temperatures (95). Higher concentrations of magnesium and other divalent ions increase the temperature at which the first transition occurs, while not affecting the second transition (95). These experiments, when combined with crystallography at high magnesium concentrations, indicate specific divalent ion binding sites, and imply the first unfolding transition of GAC RNA is affected by divalent ion concentration. The available consensus crystal structures of the GAC RNA (PDBs: 1MMS and 1HC8) (92,96), as well as a hydroxyl radical foot printing experiment, suggest that a chelated magnesium is directly bound to the O4 of U1094 and OP2 of A1073 that belong to two loops (1092-1098 and 1070-1073, respectively) which are in close proximity in the tertiary structure; this magnesium is also 5.7 Å away from a buried monovalent ion (Figure 4.1) (99).

Despite many experimental studies and theoretical calculations with nonlinear Poisson-Boltzmann methods (98), a clear model has not been devised to physically or structurally describe the transitions observed in thermal unfolding of GAC RNA. The simultaneous roles of diffuse and bound cations in stabilizing this RNA tertiary structure, as well as the presence of both monovalent and divalent ions in experiments, makes it difficult to generate models that specifically distinguish the role of one ion type from the other, and to probe the related structural changes during folding and unfolding transitions. Advances in molecular dynamics (MD) simulations have provided the means for physical

descriptions of conformational changes of RNA molecules as influenced by magnesium. For example, recent work in our lab demonstrated spontaneous conversion of the relatively dynamic magnesium free structure of the stem loop V (*SLV*) RNA to the magnesium-bound structure when  $\text{Mg}^{2+}$  was added, effectively stabilizing the structure (119). It was also shown that such folding effect is reproducible using different magnesium models although magnesium chelation properties were found to be slightly different (120). Related examples use MD to describe the influence of the absence and presence of bound ligand on the breaking up of specific kissing loop interactions in the *Add A*-riboswitch (121), and provide energetic insight into the migration of a specific catalytic magnesium from the C-site to the bridging site in both reactant and activated precursor states in the hammerhead ribozyme with all atom umbrella sampling simulations (122).

In this work, we use all atom MD simulations and umbrella sampling to probe the conformational changes related to the first unfolding transition of the GAC RNA, with regards to the role of a bound magnesium ion. Based on our initial observations in simulations in the absence of any ions, where loop interactions were rapidly destabilized, we hypothesized that the physical equivalent of the first transition event observed in the unfolding experiments (95) is separation of two loops, specifically the two loops from residues 1070-1073 and also 1092-1098. Given the presence of the chelated magnesium ion which bridges between these two loops in the crystal structures, we further hypothesized that this binding site plays the important role in the first GAC RNA unfolding transition as described by Bukhman and Draper (95). We show that the stability of the folded GAC RNA is increased by the presence of this chelated  $\text{Mg}^{2+}$  ion, and, despite increasing diffuse  $\text{Mg}^{2+}$  ion concentration, in the absence of the chelated ion,

the folded structure is less stable relative to the partially unfolded state.

## **4.2 Methods**

The simulations which were performed in this project are described in this section as follows.

### **4.2.1 Ion-free simulations (IF-R and IF-UnR)**

The GAC RNA in the crystal structure 1HC8 (96) was used as the initial structure for all simulations in this work. Hydrogen atoms were added to the RNA and all ions and crystallographic water residues were stripped using UCSF Chimera (117). The resulting structure was solvated using the tLEaP program in Amber 14 (65) in TIP3P water (56) in a truncated octahedral box with a minimum distance of 12 Å between its edges and the solute. The ff99 (51,123) force field with parmbsc0  $\alpha/\gamma$  (52) and  $\chi_{OL3}$  (53) modifications was used to build the topology. The structure was equilibrated in with a nine-step protocol as described in the following equilibration protocol section. Six production simulations were initiated from the same equilibrated structure, assigned different initial random velocities with random seeds to prevent synchronization (67) and extended to 10 ns in NPT conditions with pressure relaxation time of 5 ps (IF-UnR set). The time step for the production simulations was set to 2 fs and structures were saved to the trajectory file every 1 ps. The temperature was set to 298.15 K using Langevin algorithm (111) with 5 ps<sup>-1</sup> collision rate and SHAKE (62) was used to constrain the bonds to hydrogen atoms. The cut off for direct space interactions was set to 8.0 Å and particle-mesh Ewald was used for calculating the long-range interactions (64) with default parameters. *PMEMD.MPI* in Amber 14 was used to run the simulations. To prepare the IF-R set, six

production simulations as described above were repeated with 0.5 kcal/mol-Å<sup>2</sup> positional restraints on residues 1092-1098.

#### 4.2.2. Equilibration protocol

The solvated initial structure was energy minimized in 1000 steps with steepest descent while RNA heavy atoms were restrained in position with 5 kcal/mol-Å<sup>2</sup> force constant. 15 ps molecular dynamics was performed with 5 kcal/mol-Å<sup>2</sup> positional restraints on RNA heavy atoms in NVT conditions. 1000 steps of steepest descent energy minimization were performed with 2 kcal/mol-Å<sup>2</sup> positional restraints on RNA heavy atoms. The same energy minimization as in last step was repeated with decreased restraint weights of 0.1 kcal/mol-Å<sup>2</sup>, followed by the same minimization without restraints. 5 ps molecular dynamics was performed in NPT conditions with 1 kcal/mol-Å<sup>2</sup> restraints on RNA heavy atoms. SHAKE algorithm (62) was used to constrain bonds to hydrogen, and was applied from this step on. One nanosecond molecular dynamics in NPT conditions was done with 0.5 kcal/mol-Å<sup>2</sup> restraints on RNA heavy atoms. The same molecular dynamics was performed with the same restraints on backbone RNA atoms only. Then the same molecular dynamics was done with no restraints. This last step was not performed for equilibrating the umbrella trajectories. The molecular dynamics time steps were 1 fs for the steps prior to the last step and 2 fs in the last one. The weak-coupling algorithm was used to keep the temperature at 338.15 K in all molecular dynamics steps (63) with temperature coupling constant of 1 ps, except the IF-R and IF-UnR which were done at 298.15 K with the same thermostat. All steps were done with particle-mesh Ewald (64).

### 4.2.3 Umbrella sampling simulations

#### 4.2.3.1 Preparing the initial structures for the umbrella sampling simulations

In all umbrella sampling simulations in this work, the distance from the center of mass of phosphorus atoms of A1070 and A1073 to the center of mass of phosphorus atoms of U1094 and U1097 was used as the reaction coordinate (Figure 4.2). Umbrella sampling simulations were performed in parallel with 1 ns simulations for 321 umbrellas with 0.1 Å spacing between the reaction coordinate distances of 8.0 and 40.0 Å. The initial structures for the umbrellas were prepared in three different ways for three ranges along the reaction coordinate.

For low range distances (8.0-9.6 Å), which represent distances between the two loops closer than experimentally observed, simulations started from the initial crystal of the GAC RNA. The structure was prepared the same way as in ion-free simulations, except that the crystallographic bound  $K^+$  and the loop-loop bridging  $Mg^{2+}$  were retained, the system was neutralized by adding  $K^+$  and 1.6 M KCl (experimental KCl concentration in (95)) was also added. All ion positions (except the crystallographic  $K^+$  and  $Mg^{2+}$ ) were randomized using CPPTRAJ so that no ion was closer than 6 Å to any RNA atom and ions were not closer than 4 Å from each other. Joung-Cheatham parameters (25) were used for monovalent ions, and Allnér-Nilsson-Villa (22) parameters were used for  $Mg^{2+}$ . The system was equilibrated using the described equilibration protocol. Starting from the crystal reaction coordinate position (9.6 Å), the loops were pulled toward each other using 100 kcal/molÅ<sup>2</sup> distance restraint over 32 ps in NPT conditions. Other simulation parameters were set as in IF-R production simulations. The two closest frames with reaction coordinate values closest to each umbrella window value were extracted from the resulting trajectory using CPPTRAJ program (60) to be used as initial structures in

two independent runs.

For windows in the medium range distance (9.7-12.5 Å), the loops in the equilibrated structure as described above were also pulled away from each other using 100 kcal/mol-Å<sup>2</sup> distance restraint gradually in 60 ps in NPT conditions, while the charge of the chelated Mg<sup>2+</sup> was temporarily set to zero using parmed.py program in Amber suite, and its distance to O4 atom of the U1094 was fixed at the crystallographic distance using 100 kcal/mol-Å<sup>2</sup> distance restraint, so that the loops disengaged with the ion attached to one side while avoiding unwanted conformational disruptions. The initial structures for the umbrellas in this range were also extracted from the resulting trajectory using CPPTRAJ program as done for the low distance range.

The initial structures for umbrellas in the high range distance (12.6-40.0 Å) were extracted from one of the IF-R production simulations using CPPTRAJ program as done for the low and medium distance ranges. Where specified, the bound Mg<sup>2+</sup> was added manually to the O4 atom of U1094 for all US-boundMg runs.

The above methods were used to generate the initial structures for the umbrella sampling simulations sets indicated in Table 4.1.

#### 4.2.3.2 Running the umbrella sampling simulations

The initial structures in each umbrella for the US-IF, US-bound and the US-NeutK simulation sets without ion exclusion area were equilibrated using the described equilibration protocol. To initiate the umbrellas of US-unboundMg simulations, the bound Mg<sup>2+</sup> in the initial frame from the RNA backbone-restrained equilibration step of the corresponding US-boundMg umbrella was randomized, and then this step was repeated followed by the production simulation. Likewise, the US-NeutK simulations



with ion exclusion region were initiated in the same way from the corresponding simulations without the ion exclusion region. The ions were restrained from entering the exclusion area in the last equilibration step and production steps. Then a 1 ns production simulation was performed for each umbrella while the reaction coordinate distance was fixed with 100 kcal/mol-Å<sup>2</sup> restraint. This restraint weight provided desirable overlap between umbrellas and kept the sampled distance close to the umbrella distance (Figure 4.3). The temperature in all equilibration and production runs was set to 338.15 K to match the highest temperature for the first GAC RNA unfolding transition in melting experiments done by Bukhman and Draper (Bukhman and Draper 1997). In US-neutK simulations with ion-free zones, all K<sup>+</sup> ions were kept out of the zone with 10 kcal/molÅ<sup>2</sup> distance restraints during the equilibration and production simulations. The simulations were run with *PMEMD.CUDA* (110) and *PMEMD.MPI* from Amber 14 (65). Two independent copies of all umbrella sampling simulations were performed, each initiating from different initial structures which were extracted from the same trajectories.

#### 4.2.3.3 Calculating the potential of mean force (PMF)

The version of the WHAM method (112,124) implemented by Alan Grossfield<sup>2</sup> was utilized with a convergence tolerance of 10<sup>-9</sup> to generate the PMF plots. Each umbrella sampling window was run for 1 ns, and only 100-300 ps chunk of trajectories was used by WHAM to generate PMFs. The first 100 ps was not used because it caused much very high energetic barrier in the PMF plots which was significantly relaxed with removing this chunk of data from WHAM calculations (Figure 4.4). Also the data beyond 300 ps

---

<sup>2</sup> <http://membrane.urmc.rochester.edu/sites/default/files/wham/doc.html>

was not used in WHAM calculations to enable comparison of the free energy profiles close to the same energetic surface as the umbrella structures deviated from the initial structures significantly (Figure 4.5). The PMF plots reported here are averages of plots obtained from the two independent runs.

### **4.3 Results**

#### **4.3.1 The 1070-1073 and 1093-1098 loops dissociate in absence of ions**

According to the melting experiments (95), the first conformational changes in GAC RNA unfolding is dependent on the ion environment. To probe those changes, we first performed MD simulations under non-physical conditions where no ions were present in the simulation environment (specifically, the IF-UnR and IF-R simulations sets). The immediate and consistent major conformational change observed in the structure in six independent copies of the simulation was the separation of the two 1070-1073 and 1092-1098 loops, depicted in Figure 4.1A and Figure 4.1B, which happened spontaneously in the first few nanoseconds of the simulation (Figure 4.6A). Additionally, the U-turn conformation of the 1092-1098 loop was disrupted. Because of this, we further investigated whether the disengagement of loops was a consequence of the U-turn disruption by repeating the simulations with  $0.5 \text{ kcal/mol-Å}^2$  positional restraints on the heavy atoms of residue 1092-1098. In these restrained MD simulations the same loop-loop dissociation was observed as was seen in the unrestrained RNA simulations. Accordingly, we hypothesized that the distance between these loops can be used as a reaction coordinate describing the GAC RNA partial unfolding to its intermediate conformations, an event that seems to correlate with the first transition in previously described melting experiments because of its sensitivity to the presence of ions.

Umbrella sampling simulations were performed using the loop-loop distance as a reaction coordinate with no ions in the simulation environment to confirm the results of the short MD simulations (referred to as Umbrella Sampling-Ion Free simulations, US-IF). The reaction coordinate was chosen as the distance from the center of mass of phosphorus atoms of A1070 and A1073 to the center of mass of phosphorus atoms of U1094 and U1097. The resulting potential of mean force (PMF) along this reaction coordinate shows a significant drop in the calculated free energy as the two 1070-1073 and 1092-1098 loops are separated from each other, which is consistent with the very short time scale of the conformational change observed in unbiased MD simulations (Figure 4.6B).

#### 4.3.2 Presence of ions is essential for stability in the folded tertiary structure

To see the extent to which minimal amounts of ions can affect the stability of the tertiary structure, neutralizing potassium was added in random positions around the initial structure of each umbrella, and 1 ns of MD was performed after equilibration (US-neutK simulation set). Extending the umbrella simulations to 1 ns (in this run as well as other umbrella sampling runs) resulted in dramatic deviation of the structures from the initial structures (Figure 4.5). Therefore, the first 200 ps of data after discarding 100 ps for equilibration was used for PMF calculations. The black line in Figure 4.7 shows that adding neutralizing potassium ions reversed the free energy profile in favor of the relative stability of the folded state, at a reaction coordinate value of 9.6 Å. The structures obtained from the umbrellas close to the folded state, corresponding to the global minimum value of the PMF plot, show accumulation of the  $K^+$  ions near the magnesium

ion-binding site where the two loops closely interact (Figure 4.8).

#### 4.3.3 Chelation of a single $Mg^{2+}$ increases the relative stability of the folded GAC RNA at different $Mg^{2+}$ concentrations

Based on the fact that binding of a  $Mg^{2+}$  ion is involved in the first transition during thermal unfolding of the GAC RNA at different  $Mg^{2+}$  concentrations (95,99), we expect to see the folded structure (at 9.6 Å) is stabilized relative to the intermediate structure (at > 20 Å) in the presence of a chelated  $Mg^{2+}$  which bridges the two 1070-1073 and 1092-1098 loops. Therefore, we set up umbrella sampling simulations in neutralizing potassium with additional concentrations of diffuse  $MgCl_2$  (0, 10, 20, and 50 mM) while retaining the chelated magnesium at its crystallographic position (US-boundMg simulation sets, Figure 4.9, left). The initial equilibrium steps of each umbrella were performed with restraints on the bound magnesium so that it is not pulled into the closely located monovalent ion pocket (96). To analyze the effect of the bound  $Mg^{2+}$ , that ion was randomized to a position at minimum 6 Å away from any RNA atom prior to the last equilibration step in each umbrella, and the last equilibration step as well as the production step were repeated in the absence of the bound  $Mg^{2+}$  for a third set of independent simulations (US-unboundMg simulation sets, Figure 4.9, right). Figure 4.9 represents the PMF plots for the change along the reaction coordinate in presence (left) and absence (right) of the chelated  $Mg^{2+}$ .

#### 4.3.4 Description of the first unfolding transition

The simulations in this work provide insight into the possible conformational changes that are observed as the folded GAC RNA partially unfolds to an intermediate state. The

unfolding transition, which is described first by the unrestrained and restrained MD simulation and then further by the PMFs, initially involves disengagement of the 1070-1073 and 1092-1098 loops. According to the GAC crystal structures, these loops are tightly linked together by a chelated  $Mg^{2+}$  ion which bridges the O4 atom of the U1094 and OP2 atom of the A1073. This is in agreement with the PMF plots in Figure 4.9, where the energy barrier for unfolding is significantly higher in presence of chelated  $Mg^{2+}$ . However, there are other stabilizing interactions between RNA atoms that are unraveled during unfolding as well.

A major structural change that significantly contributes to RNA unfolding is the loss of contact between A1088 and U1060 (Figure 4.10B and Figure 4.11). In the crystal structure, A1088 from one strand intercalates into the adjacent helix and forms a Hoogsteen base pair with U1060 while stacking on top of C1079 (97). According to the thermal unfolding and L11 binding experiments performed on the GAC RNA mutants, a GAC RNA carrying A1088U mutation is described as lethal and is both unable to fold to the tertiary structure and to bind to the L11 protein (125). The described contacts of A1088 are broken immediately in the ion free simulations (IF-R) and therefore, they are missing in all umbrella initial structures which are extracted from those trajectories (Figure 4.11). This suggests that the base pairing of A1088 and U1060 and stacking between A1088 and C1079 are among the first key interactions that break during the unfolding pathway. This highlights the importance of these interactions and is consistent with the A1088U mutation experiments.

Another change in the RNA upon partial unfolding is the loss of a stacking interaction between G1071 and A1089 (Figure 4.10C and Figure 4.11). The backbone of G1071 contributes to the formation of the buried monovalent ion

pocket. Because of the proximity of A1089 and A1088 in sequence, loss of base stacking between A1089 and A1089 might potentially allow a concerted conformational change in the neighboring A1088 as described above. This stacking interaction is also lost in the IF-R simulations in absence of ions and early in the unfolding pathway (Figure 4.11). This is in agreement with a recent 2-aminopurine fluorescence experiment which reports that stacking of 2-aminopurine derivative of A1089 in GAC structure is dependent on the presence of  $Mg^{2+}$  (126).

Structural changes after 12.5 Å on the unfolding pathway include breaking of many interactions between these two parts of the RNA finishing with total disruption of the U-turn 1082-1086.

#### **4.4 Conclusions**

GAC RNA is an extensively studied RNA model system in terms of how the stability of its tertiary structure is affected by different ion interactions. However, there is little structural information about how these ions play their roles in the pathway during which this RNA unfolds. Here, we employed umbrella sampling molecular dynamics simulations to qualitatively describe an important step in the unfolding pathway of the GAC RNA, with an emphasis on the role of a bound magnesium ion. The results of unrestrained simulations in the absence of ions show that unfolding step is characterized by a loss of loop-loop interactions in the GAC RNA; and a chelated  $Mg^{2+}$  ion that bridges between the 1070-1073 and 1092-1098 loops can relatively stabilize the folded state. This is consistent with the first step of the multistep pathway for unfolding of GAC RNA which has been experimentally shown to be ion-sensitive and involve binding of a

divalent cation (95,99). We expect that dissociation of these loops is the first event in the GAC RNA unfolding pathway. Also we showed that separation of these loops in the unfolding pathway is accompanied by breaking of the A1088-U1060 Hoogsteen base pairing and A1088-C1079 stacking interactions and also the simultaneous loss of stacking between A1089 and G1071.

Increasing the stability of the folded tertiary structure relative to the intermediate structure could occur through either stabilizing the tertiary structure or destabilizing the intermediate structures in higher magnesium concentrations. We have shown stabilization of the GAC RNA is only possible in presence of cations that bridge between these two loops. A single  $Mg^{2+}$  ion with direct interactions with the OP2 atom of A1073 and O4 atom of U1094 can keep the loops in close proximity, stabilizing the folded state relative to the unfolded state. This magnesium ion is present in all available GAC crystal structures and its presence has been reconfirmed by hydroxyl radical foot printing experiments (92,96,99). Regarding the observation that the GAC RNA folds in presence of other divalent ions as well, using modified GAC RNA that selectively binds to some divalent ions can validate the proposed role for the chelated magnesium in the folding pathway in future experiments.

Table 4.1. Description of the umbrella sampling simulations performed in this project.

Simulation	Description	Ion content	Related figure
US-IF	The ions and water residues in the low and medium range distance initial structures were deleted and the RNA structures were resolvated as in IF-R simulation set with tLEaP.	No ion	Figure 4.6B
US-neutK	The ions and waters in all initial structures were deleted and the RNA structures were resolvated in water and neutralizing $K^+$ was added in random positions.	Neutralizing $K^+$ (56 ions)	Figure 4.7
US-boundMg	The ions except the bound $Mg^{2+}$ and water in all initial structures were deleted and the RNA structures were resolvated in water and neutralizing $K^+$ plus 0, 10, 20, and 50 mM excess $MgCl_2$ was added in random positions to prepare initial structures for systems with different $MgCl_2$ content and the bound $Mg^{2+}$ .	Bound $Mg^{2+}$ , Neutralizing $K^+$ (54 ions) plus 0, 20, 35 and 50 mM (0, 8, 14 and 18 $Mg^{2+}$ ions respectively), excess $MgCl_2$ .	Figure 4.9
US-unboundMg	The bound $Mg^{2+}$ was relocated to a random position with minimum of 6 Å distance from the RNA in the last frame of the partially equilibrated umbrella simulations in US-boundMg and the last equilibration step as well as the production simulations were repeated.	Neutralizing $K^+$ (54 ions) plus 0, 20, 35 and 50 mM (1, 9, 15 and 19 $Mg^{2+}$ ions respectively), excess $MgCl_2$ .	Figure 4.9



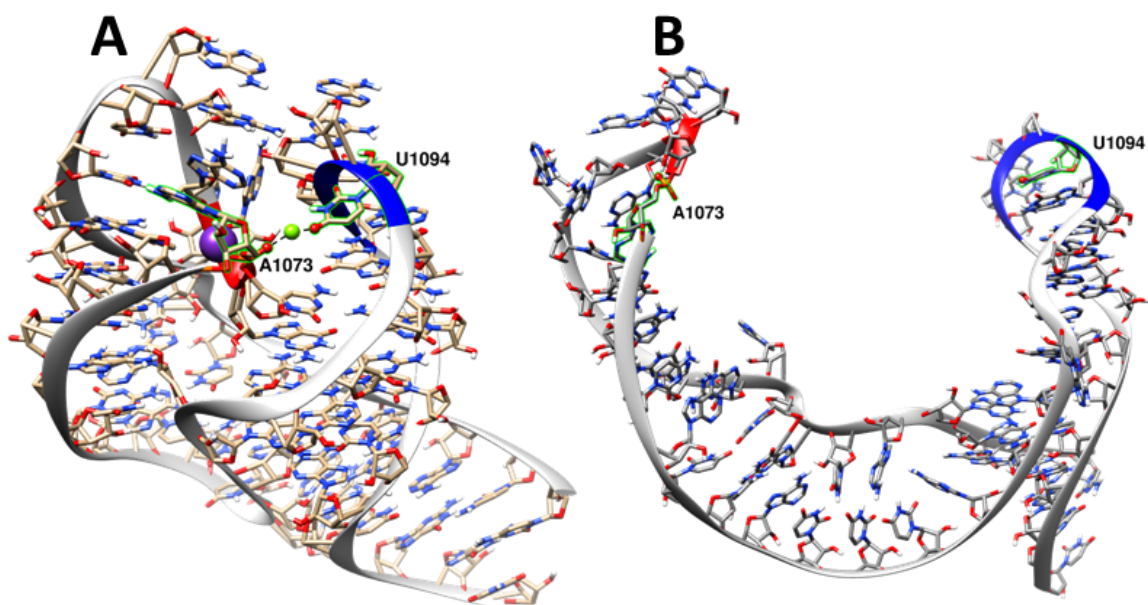


Figure 4.1. A) Tertiary structure of the GAC RNA according to the 1HC8 crystal structure (**96**). A magnesium ion (green sphere) bridges between the two 1070-1073 and 1093-1098 loops (red and blue ribbons, respectively) with direct interactions to O4 atom of U1094 and OP2 atom of A1073 (red spheres). The nearby bound potassium ion is shown as purple sphere. B) The GAC structure from the last frame of the 40.0 Å distance umbrella of the simulation in D in which the loops are completely disengaged.

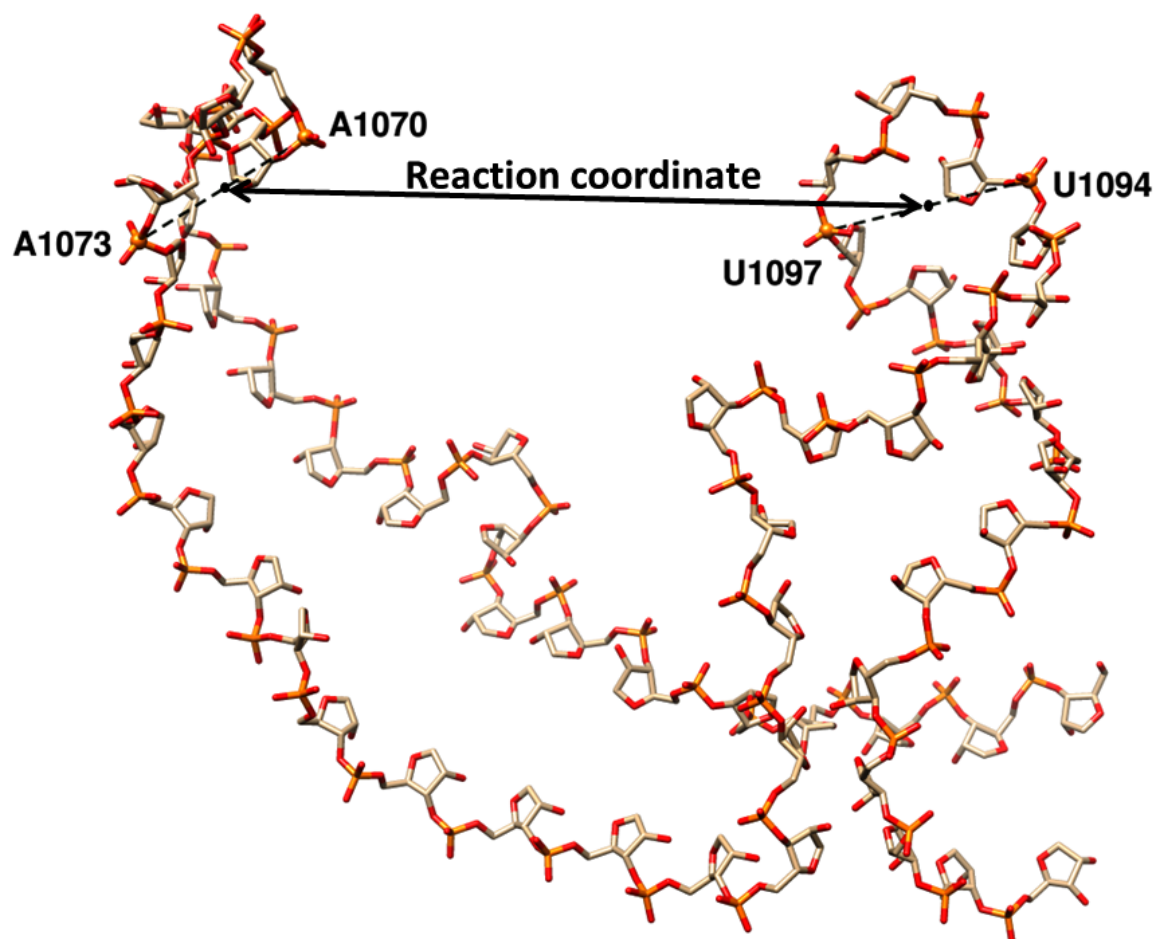


Figure 4.2. The reaction coordinate distance (double-headed arrow) used in umbrella sampling simulations which is the distance from the center of mass of phosphorus atoms of A1070 and A1073 to the center of mass of phosphorus atoms of U1094 and U1097, shown on the initial structure of the 40.0 Å umbrella window. The hydrogen and base atoms are not shown for clarity and the phosphorus atoms of the indicated four residues are shown with orange spheres.

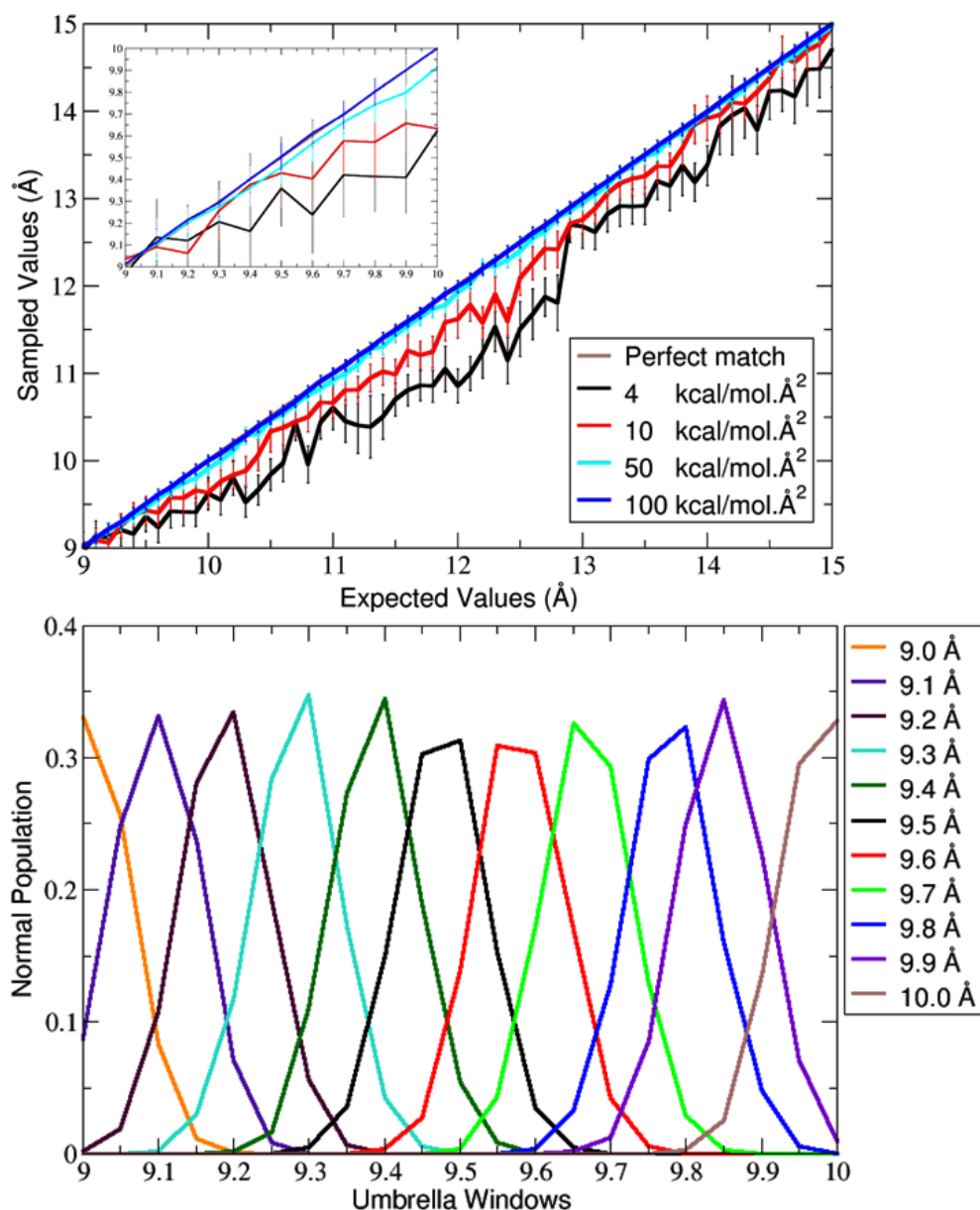


Figure 4.3. Top: Sampled vs. expected values for umbrellas close to the minimum PMF resulted from one copy of the US-boundMg simulations with no excess  $\text{MgCl}_2$  in presence of the bound  $\text{Mg}^{2+}$ , sampled with different force constants. Bottom: Histograms of the sampled values close to the minimum PMF for the system sampled with  $100 \text{ kcal/mol.}\text{\AA}^2$  force constant.

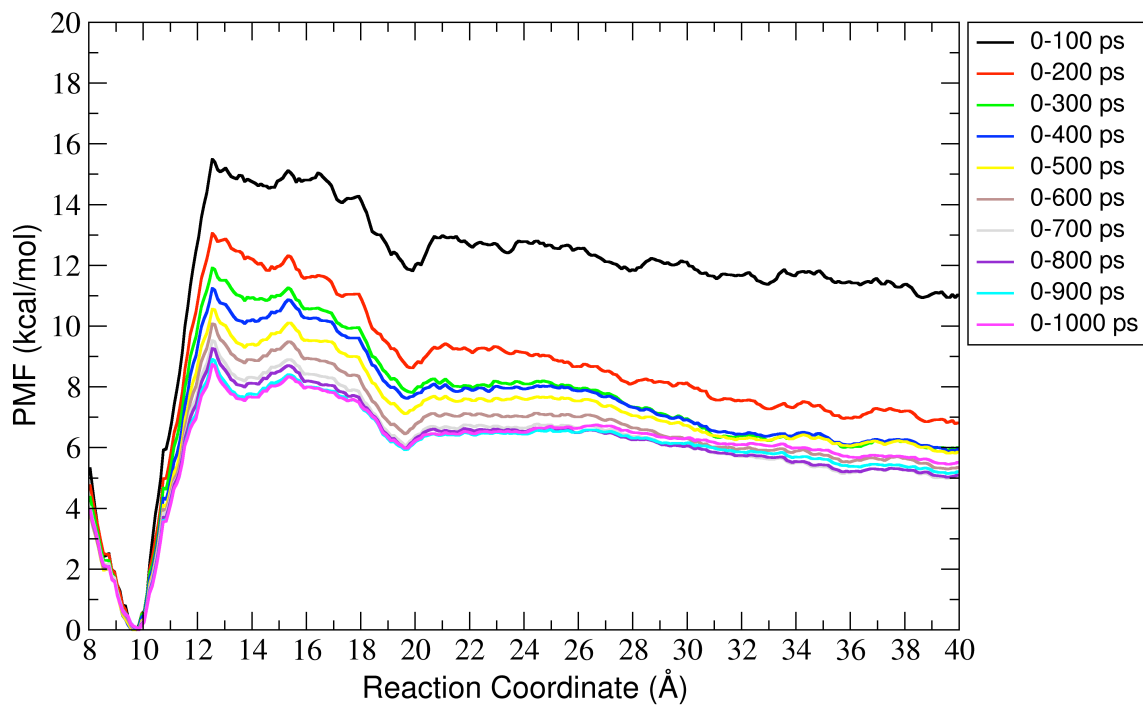


Figure 4.4. Evolution of the PMF plots in one copy of US-boundMg simulations with no excess  $\text{MgCl}_2$  in presence of the bound  $\text{Mg}^{2+}$ , with adding 100 ps chunks of data.

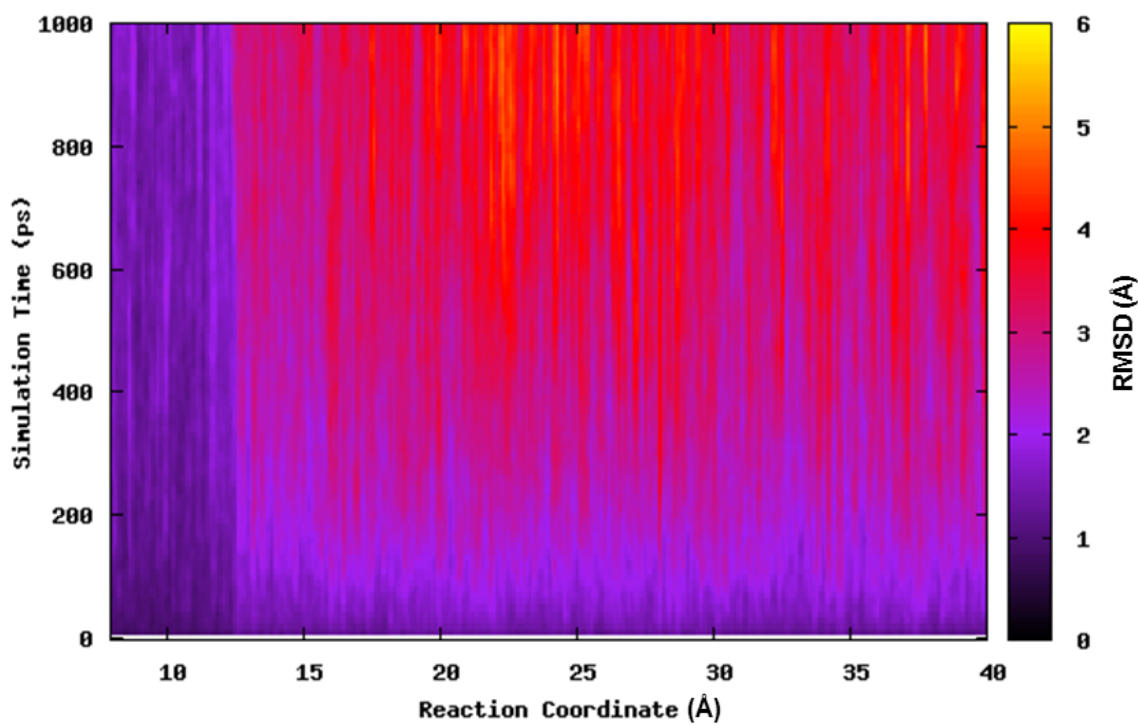


Figure 4.5. Root mean square deviation along umbrella trajectories with reference to the umbrella initial structure, in one copy of US-boundMg simulations with no excess  $\text{MgCl}_2$ .

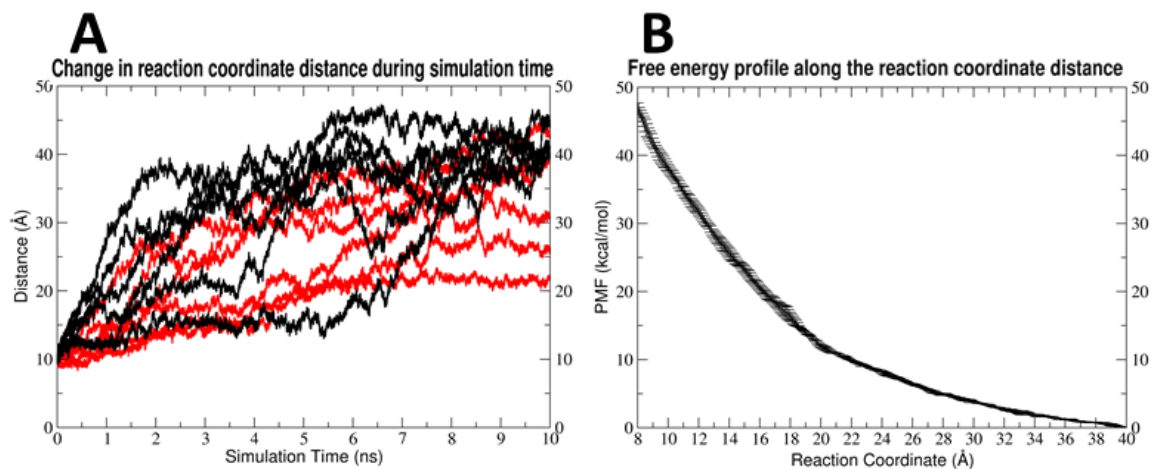


Figure 4.6. A) The distance from the center of mass of phosphorus atoms of A1070 and A1073 to the center of mass of phosphorus atoms of U1094 and U1097, in GAC RNA simulations in absence of any ions, which is used as the reaction coordinate in all umbrella sampling simulations here. Each black line represents an independent copy of unrestrained simulation, while each red line represents an independent copy of simulations with positional restraints on 1092-1098 residues. B) The PMF profile along the reaction coordinate described in (A) for partial unfolding of GAC RNA in absence of any ions.

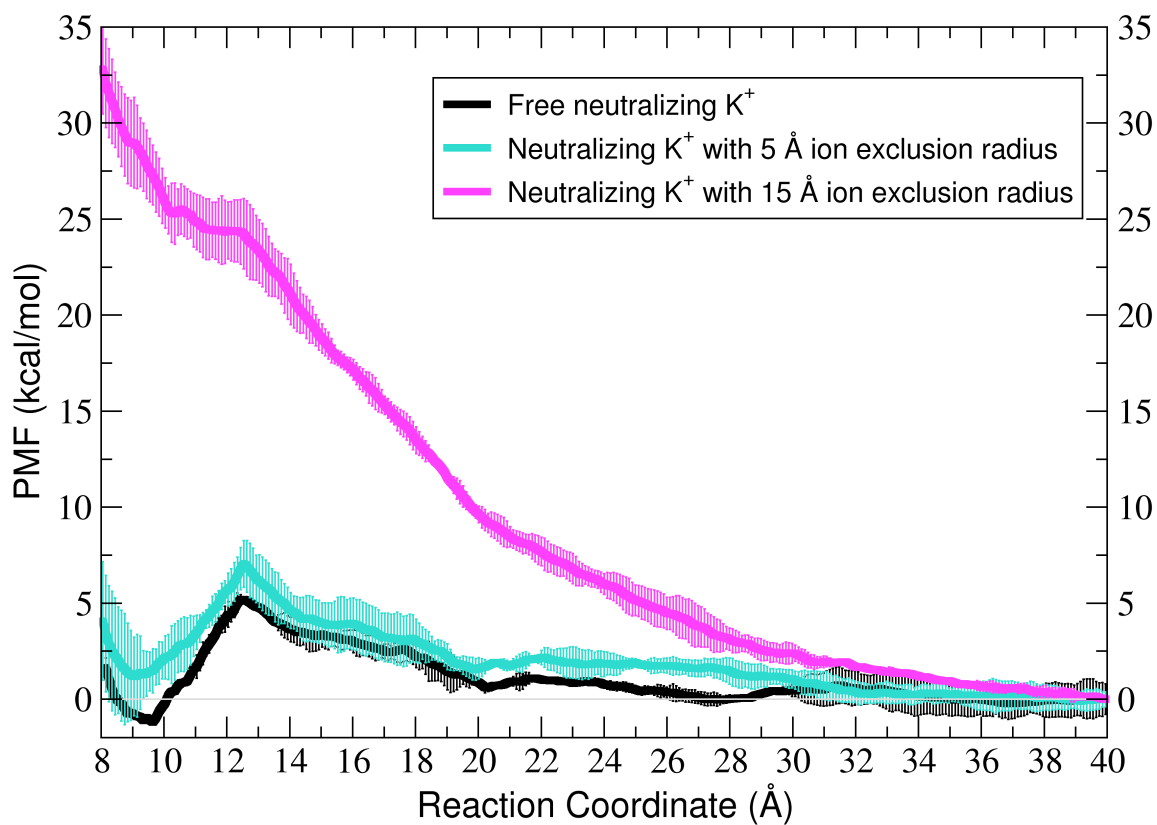


Figure 4.7. Free energy profiles along the loop-loop distance reaction coordinate as described in the text with free neutralizing potassium ions and ion exclusion zones around the magnesium binding site with ion exclusion radii of 5 and 15 Å. The zero point is arbitrarily adjusted to the PMF values at 40.0 Å. The plots represent the average PMF values between two independent runs and the error bars are calculated as the standard deviation between them.

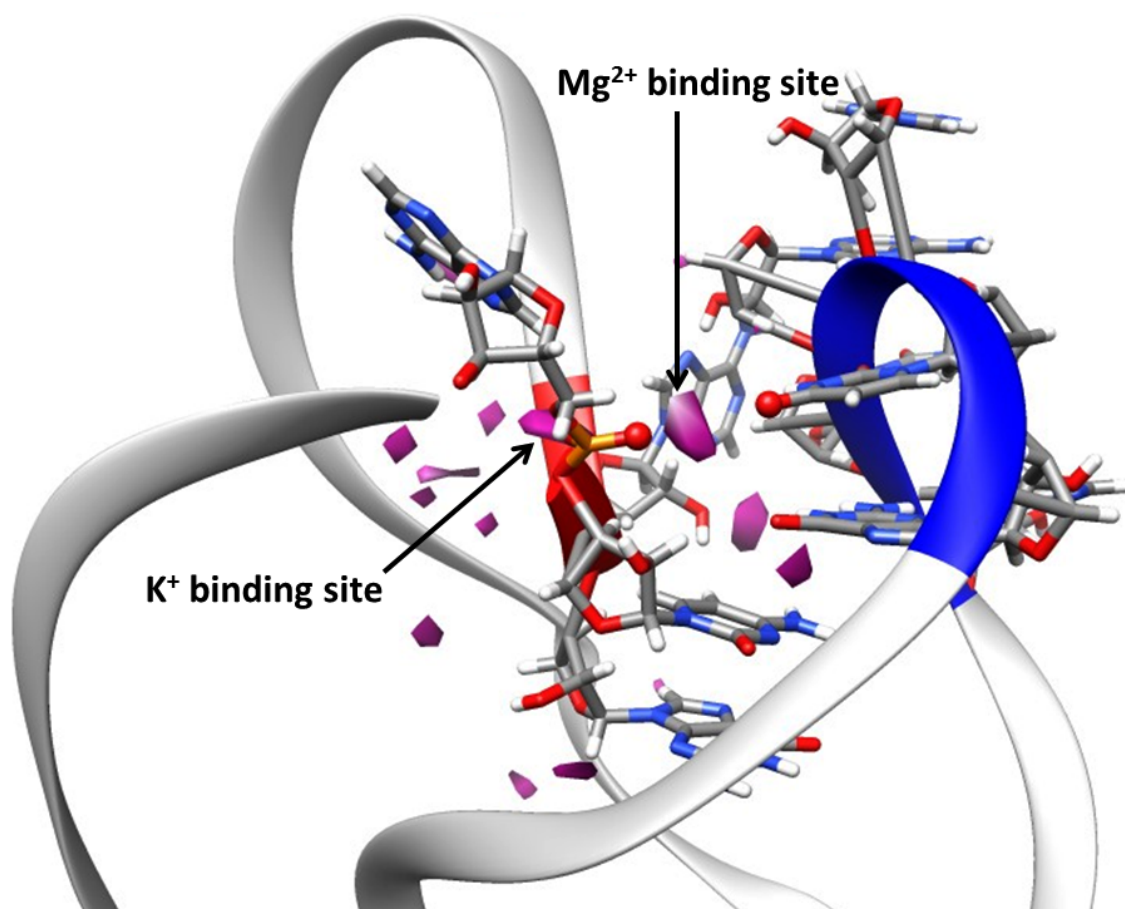


Figure 4.8. Part of the GAC RNA average structure over cumulative 2 ns production data of umbrellas 9.0-9.9 Å (close to the crystal structure with reaction coordinate value of 9.6 Å) in two independent sets of simulations with neutralizing potassium. The purple diamonds represent positions occupied by potassium 2000 times the evenly-distributed potassium would occupy. Potassium ions accumulate in the potassium and magnesium binding sites when the ions are free in simulations (black plot in Figure 4.7)



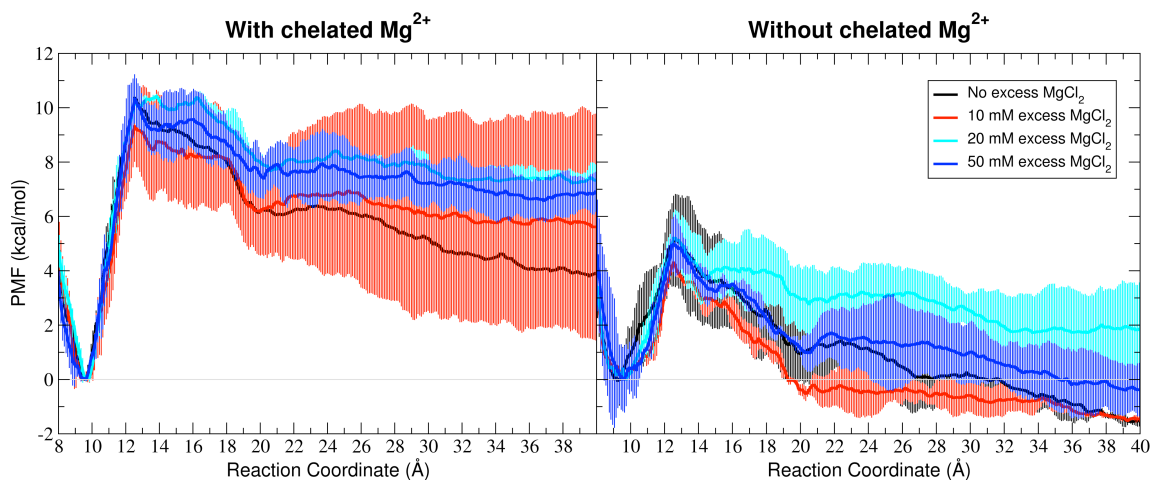


Figure 4.9. Free energy profiles along the loop-loop distance reaction coordinate with neutralizing potassium and increasing amount of excess magnesium chloride. Left) One  $\text{Mg}^{2+}$  ion is chelated in the crystallographic position as depicted in Figure 4.1A close to the global minimum and bound to only the O4 of U1094 in longer distances. Right) The bound  $\text{Mg}^{2+}$  is relocated to random positions with minimum of 6 Å distances from the RNA in partially equilibrated structures of corresponding simulations in (Left) and final equilibration step and production simulations are repeated. All the PMFs are zeroed at the closest local minimum to the crystallographic reaction coordinate distance (9.6 Å). The error bars are calculated as the standard deviation between two independent runs.

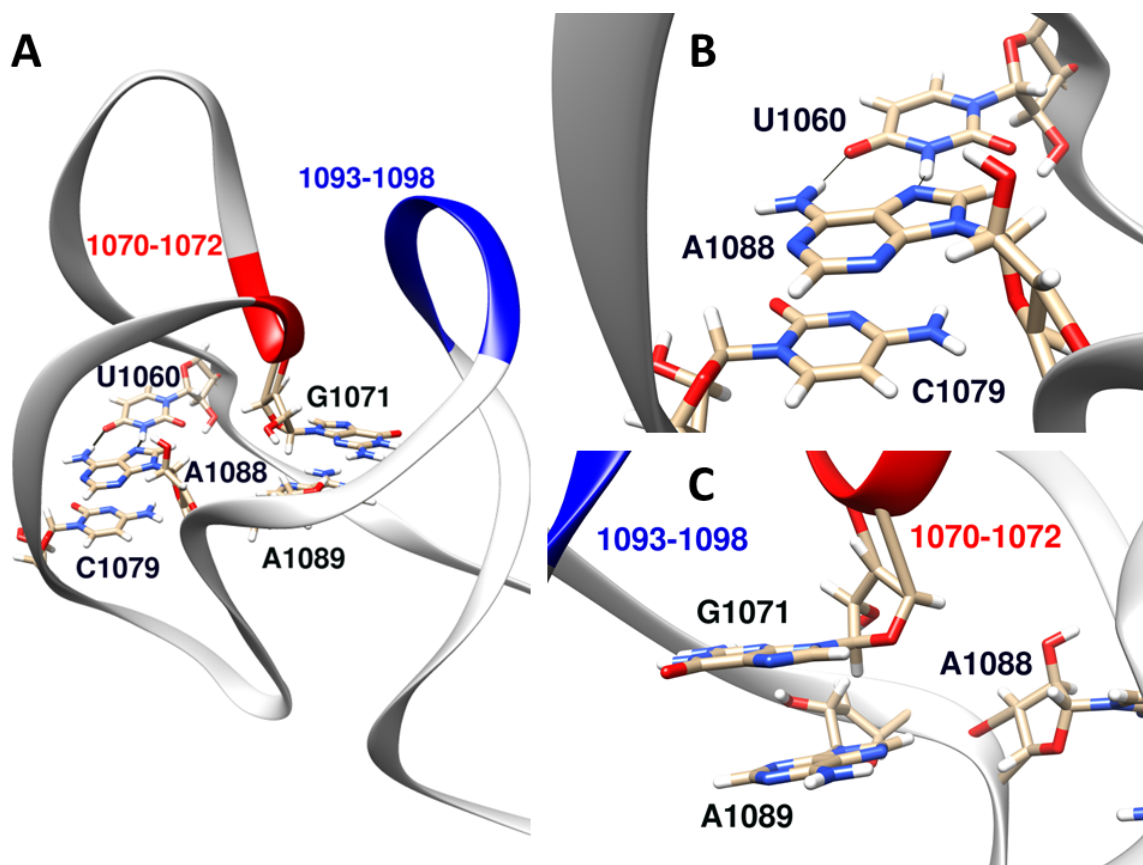


Figure 4.10. A) The global structure of GAC RNA according to the 1HC8 crystal structure (**96**), with important residues which undergo major structural changes during unfolding transition highlighted. B) Hoogsteen base pair between A1088 and U1060 (hydrogen bonds represented as solid black lines) and stacking interaction between A1088 and C1079 according to the 1HC8 crystal structure (**96**). C) Stacking interaction between G1071 and A1089 according to the 1HC8 crystal structure (**96**).

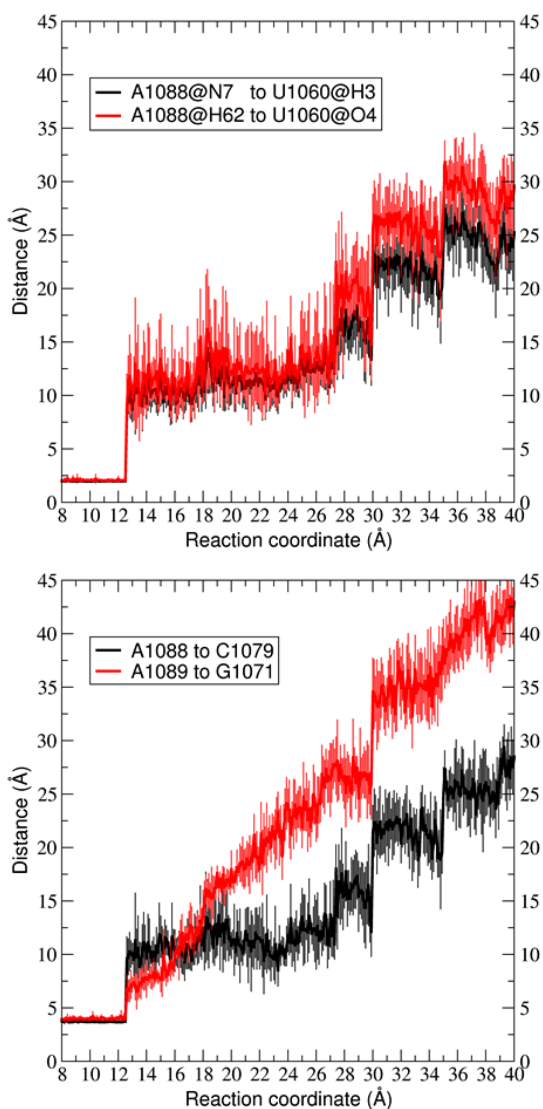


Figure 4.11. (Top) Hydrogen bond distances between A1088 and U1060 along the reaction coordinate. (Bottom) Distances between the center of masses of base atoms of A1088 and C1079; and between the center of masses of base atoms of A1089 and G1071 along the reaction coordinate which are representative of the stacking interactions between these two pairs of bases. The data is gathered from combined umbrella samplings with bound and unbound bridging  $\text{Mg}^{2+}$ , with error bars representing the standard deviations between different runs.

## CHAPTER 5

### THE EFFECT OF CRYSTALLOGRAPHIC MAGNESIUM ON BINDING OF 2-AMINOBENZIMIDAZOLE INHIBITORS TO THE HEPATITIS C VIRUS IRES RNA

This research project was designed by Hamed S. Hayatshahi and Thomas E. Cheatham, III in collaboration with Darrell R. Davis. Ligands were prepared for docking and one copy of simulations in absence of magnesium were set up by Niel M. Henriksen. All other model building, running and analysis were done by H.S. Hayatshahi. The chapter was written by H.S. Hayatshahi and revised by T.E. Cheatham.

#### **5.1 Introduction**

Hepatitis C is a widely spread global health problem with an estimated 3% prevalence (127,128) that is caused by an RNA virus called HCV (129). The special translation mechanism in HCV has provided the opportunity for development of RNA-targeting translation inhibitors which can specifically attack the viral protein synthesis machinery (130). This translation mechanism includes recognition of an Internal Ribosome Entry Site (IRES) on the 5'-UTR (Untranslated Region) of the viral RNA by the ribosome instead of recognition of a 5' RNA CAP (131). It has been shown that some 2-aminobenzimidazole inhibitors can interact with the subdomain IIa of the IRES region

and interfere with its recognition by the ribosome (130). According to the crystal structures of this domain in absence (132), and in presence (127) of one of these inhibitors, the inhibitor changes the RNA conformation by preventing formation of a kink in the RNA structure (Figure 5.1).

The inhibitor-bound crystal structure (PDB code 3TZR) contains crystallographic  $Mg^{2+}$  ions around the ligand binding site. It is also claimed that the ligand binding to the pocket is dependent on the presence of these crystallographic ions (127). The binding enhancement in presence of divalent cations is unexpected as the 2-aminobenzimidazole inhibitors are also +2 and +3 charged ligands (133). Therefore, we aimed to assess this hypothesis with docking and molecular dynamics simulations.

## **5.2 Methods**

### **5.2.1 Ligand preparation for docking**

A set of analogues that had been previously shown to bind to the HCV-IRES subdomain IIa (Figure 5.2) (130) were used for docking. Force field parameters for the ligands were developed in previous work and used in the docking studies presented here (133). Basically, the ligand preparation involved quantum mechanical optimization at HF/6-31G\* level (69,70), development of initial electrostatic charges by RESP fitting (19), application of the GAFF force field (134), 50 ns of in-vacuo molecular dynamics (MD) with a generalized Born implicit solvent potential (igb=1 with SHAKE on hydrogens (62), and a 2 fs time step at 400 K (135) in AMBER (106) to conformationally sample a more relevant ensemble (107) followed by clustering the trajectories into 20 clusters with PTRAJ (60) using the average linkage method and selecting the representative structures with > 2% occupancies. These structures were then

re-optimized at the same quantum-mechanical level and structures within 0.5 kcal/mol energy difference from the structure with minimum energy (total of 6 to 14 conformations per compound) were then reoptimized in a multiconformation and multi-orientation RESP charge fitting enabled by the RED program (136). QM calculations and MD simulations were performed with the Gaussian 9 (71) and Amber 12 (106), respectively.

### 5.2.2 Receptor preparation for docking

The HCV-IRES subdomain IIa RNA crystal structure (137) with PDB code 3TZR was loaded into UCSF Chimera program after deletion of all residues except the nucleotides and the  $Mg^{2+}$  ions. The receptor with no hydrogen was used in *DOCK 6.5* (138) to generate the box in which the inhibitors would be docked. The receptor with hydrogen atoms was later used for docking. The chosen box was sufficiently large to encompass all of the RNA structure. Three RNA receptors were prepared with different  $Mg^{2+}$  ion positions or properties, specifically one where the crystallographic  $Mg^{2+}$  ions were retained, the second where the ions were moved to random positions far from the RNA structure, but still in the docking box, and the third where the ions were retained in their crystallographic positions, but their charges were set to zero.

### 5.2.3 Docking

*DOCK 6.5* (138) was used with its flexible ligand capability for docking the representative structures of the prepared inhibitors with force field parameters as described above on the prepared RNA models. The grid score was calculated for each ligand in the ensemble and the RMSD of the docked poses against the experimental pose

was calculated with UCSF Chimera (117).

#### 5.2.4 Molecular dynamics

The crystal structure of HCV-IRES subdomain IIa (PDB: 3TZR (137)) was used as the starting point to prepare the initial structure for MD simulations. The initial structure preparation had been done previously to facilitate comparison with an NMR structure (133). This included modification of the RNA duplex into a hairpin with removing the 3' dangling base and changing the C:G base pair at the lower stem to G:C and then adding a UUCG tetraloop to the opposite stem. The RNA parameters were generated based on AMBER's ff12SB force field which includes ff99 (51) with more recent backbone  $\alpha/\gamma$  and  $\chi$  modifications (52,53) using tLEaP program in Amber 12 (106). The "j5r" inhibitor atomic partial charges were generated as previously discussed with the remaining force field parameters from GAFF with some missing torsion parameters inferred with the Parmcheck program (134,139). The initial structures were energy minimized in implicit Hawkins, Cramer, Truhlar Generalized Born solvent model (107) for 2500 cycles. Then the models were solvated in TIP3P water (56). The system was neutralized with  $K^+$  and ~200 mM extra KCl was added. Then three sets of simulations were generated with different salt properties: 1) Only 200 mM excess KCl, 2) 200 mM excess KCl plus six crystallographic  $Mg^{2+}$  and neutralizing  $Cl^-$ , 3) 200 mM excess KCl plus six  $Mg^{2+}$  ions in random positions and neutralizing  $Cl^-$ . Parameters developed by Allnér, Nilsson and Villa(22) were used for  $Mg^{2+}$  ions and those developed by Joung and Cheatham were used for  $K^+$  and  $Cl^-$  in TIP3P water (25). Four copies of each simulation set were generated by randomizing the noncrystallographic ions with the CPPTRAJ program (60) such that the ions were not closer to RNA atoms than 6 Å and separated from each other

with minimum distance of 4 Å. The systems were equilibrated applying the following protocol: 1) They were energy-minimized for 1000 cycles using steepest-descent method followed by 1000 cycles using conjugate gradient method with 25 kcal/mol-Å<sup>2</sup> positional restraints on RNA and ligand atoms. 2) Then they were heated from 10 K to 150 K using Langevin thermostat (66,67) with 2 ps<sup>-1</sup> collision frequency in 100 ps of MD with the same positional restraints at constant volume. 3) Then the systems were heated to 298 K in 100 ps with 5 kcal/mol-Å<sup>2</sup> positional restraints on the same atoms in constant pressure using Langevin thermostat (66,67) with 2 ps<sup>-1</sup> collision frequency. 4) 2 ns of constant pressure MD with 0.5 kcal/mol-Å<sup>2</sup> positional restraints on the same atoms was performed as the final equilibration step. The production MD simulations were done at 298 K with the weak coupling algorithms for thermostat and barostat (63). More details about the equilibration simulation parameters were discussed in Henriksen et al. (133). However in this work, part of the mass was transferred from RNA hydrogen atoms to the attached heavy atoms to facilitate 4 fs integration time steps during the production simulations (27). The production simulations were run on GPUs of Blue Waters Petascale Resource and also the NSF XSEDE Stampede machine at the Texas Advanced Supercomputer Center with the *CUDA* version of Amber 14 (65) *PMEMD CUDA* version of Amber 14 for 800 ns during which the coordinates were saved every 10 ps in to the trajectories.

### 5.2.5 Simulation analysis

Amber's CPPTRAJ program (60) was used to generate the RMSD data and ion density maps from the MD trajectories. To obtain the ion densities, the trajectory frames were imaged, centered and RMS-fit to the residues near the binding site (G52, A53, A109, G110, C111 and the ligand according to numbers in PDB code 3TZR). Then the



number of ions was counted in grid spaces (150 bins in each direction) with dimensions of 0.5 Å x 0.5 Å x 0.5 Å all over the simulation box in each frame and were added along the trajectory frames. Grid spaces with contour level of ion densities higher than 100 times the evenly-distributed ions were visualized using UCSF Chimera (117). The contour level was set as follows according to previous work by Cheatham and Kolman (140).

$$contour = \frac{100 \times V_g \times N_i \times N_f}{V_b} \quad (5.1)$$

In this equation,  $V_g$  is the grid volume (here 0.125 Å<sup>3</sup>),  $N_i$  is the number of ions (here 6 Mg<sup>2+</sup>),  $N_f$  is the number of trajectory frames and  $V_b$  is the volume of the simulation box in Å<sup>3</sup>.

## **5.3 Results**

### **5.3.1 Docking**

To assess the effect of the magnesium positions and magnesium charge on the ligand affinities as estimated by docking, the ligand ensembles were docked on to three HCV-IRES subdomain IIa RNA models: 1) with crystallographic Mg<sup>2+</sup>, 2) with randomized Mg<sup>2+</sup> and 3) with crystallographic Mg<sup>2+</sup> but with the ions neutralized. In the presence of the crystallographic Mg<sup>2+</sup> ions, all of the docked ligands produced less favorable docking scores compared to the scenario with Mg<sup>2+</sup> ions at random positions (Figure 5.3). Also with neutralizing Mg<sup>2+</sup>, docking poses closer to the experimental pose were observed (Figure 5.4), with the exception of j1a and j2a that tend to bind to different positions

along the RNA helix. More favorable docking scores and the observation of docked binding poses closer to the experimental pose in absence of crystallographic  $Mg^{2+}$  suggest either that steric hindrance with crystallographic  $Mg^{2+}$  blocks the ligands from the binding site or that the electrostatic repulsion between crystallographic  $Mg^{2+}$  and ligands prevents them from approaching the receptor.

To assess whether the binding was inhibited by  $Mg^{2+}$  steric interactions as opposed to electrostatics, the ions charges on the crystallographic  $Mg^{2+}$  were neutralized and the docking experiments were repeated. With the neutral magnesium ions in crystallographic positions, the docking scores were recovered and even improved to some extent from the situation in which the ions were in random positions (Figure 5.3). Also the docking poses including those for j1a and j2a got closer to the experimental pose (Figure 5.4). This suggests that the electrostatic repulsion from the positive charge on  $Mg^{2+}$  ions prevent the ligands docking to the experimental binding site.

Figure 5.5 specifically highlights the docking results for the crystallographic ligand (j5r). In presence of crystallographic  $Mg^{2+}$ , none of the docking poses are close to the experimental binding site. However, in presence of randomized  $Mg^{2+}$  and neutral crystallographic magnesium, most of the top scored poses are in the experimental pocket.

### 5.3.2 Molecular dynamics

In order to study the localization of the  $Mg^{2+}$  ions and their effect on presence of the bound ligand, four independent sets of MD simulations under three different conditions were performed, specifically 1) removing the crystallographic  $Mg^{2+}$  ions from the system, 2) retaining the  $Mg^{2+}$  ions in their crystallographic positions and 3) placing them initially in random positions away from the crystallographic binding sites. All simulations

were initiated from the RNA crystal structure coordinates with neutralizing  $K^+$  plus 200 mM extra KCl. System parametrization and equilibration were done as discussed in the methods section.

The binding pocket (residues G52, A53, A109, G110, C111 and the ligand) root mean squared deviation (RMSD) time series with reference to the starting structure all look very similar (Figure 5.6) regardless of  $Mg^{2+}$  presence and location. The characteristic crystallographic contacts between ligand and RNA remained unchanged in all sets except for copy 2 of the set with initial crystallographic  $Mg^{2+}$  in which the hydrogen bond between the ligand and O6 of G110 was lost and G110 was slightly shifted out. The RMSD spikes to  $\sim 2$  Å in this simulation copy indicating loss of stacking between the ligand and A53.

To assess the location of magnesium ions during simulations, the RNA structures were RMS fit to a common reference frame and the volume was divided into cubic  $0.125 \text{ \AA}^3$  volume elements in which magnesium ions were counted in each frame and summed over all frames of the aggregate trajectories of each set (Figure 5.7). The results imply that all  $Mg^{2+}$  ions moved to other positions in the MD simulations, positions that are different from where they were initially located in the crystal structure (Figure 5.7A). The  $Mg^{2+}$  ion located at the entry of the binding pocket shifted to positions closer to the G107 and C108 nonbridging phosphate oxygen and was also only minimally populated in simulations initiated from randomized magnesium (Figure 5.7B). The ions highly populated in a region confined with the residues C57, U61 and G107 in both sets of simulations in the presence of  $Mg^{2+}$  ions. The changes in magnesium positions can potentially be a result of repulsion from the ligand positive charges, attraction from the RNA backbone leading to ion binding events that are not converged on the MD time

scale applied, and absence of crystal packing interactions in these solution phase simulations. The highly populated ion region is also consistent with a second and lower affinity binding site for the examined ligands that was occupied in the docking experiments (Figure 5.5). Docking in presence of crystallographic  $Mg^{2+}$  ions directed all ligands to this second binding pocket (RMSD =  $\sim 10$  Å in Figure 5.4).

#### **5.4 Conclusions**

The hypothesis that the binding of a set of 2-benzimidazole inhibitors to the subdomain IIa of the IRES RNA of hepatitis C virus is dependent on the presence of crystallographic  $Mg^{2+}$  ions was assessed and invalidated in this research project using docking and molecular dynamics simulations. The docking experiments using receptors with crystallographic  $Mg^{2+}$  led to docking of the ligand in a non-crystallographic pocket and much less favorable docking score than docking experiments using receptors with far random  $Mg^{2+}$ . Both the crystallographic ligand positions as well as the docking scores were retained in docking experiments in which the crystallographic  $Mg^{2+}$  charge was set to zero. This implies that the positive charge of the ions in crystallographic positions prevents the ligands from entering the crystallographic pocket. The MD simulations in presence and absence of the crystallographic  $Mg^{2+}$  ions resulted in relatively similar RNA-ligand conformations. However, the crystallographic RNA-ligand interactions were disturbed in one of the simulation copies which were initiated with crystallographic  $Mg^{2+}$  ions. Docking of the ligand in a non-crystallographic pocket in presence of  $Mg^{2+}$  ions introduced a second inhibitor binding site which was previously unknown. Unlike the crystallographic binding site which is shown here to be inhibited by  $Mg^{2+}$ , binding to the second pocket is predicted to be independent from the presence of  $Mg^{2+}$ .

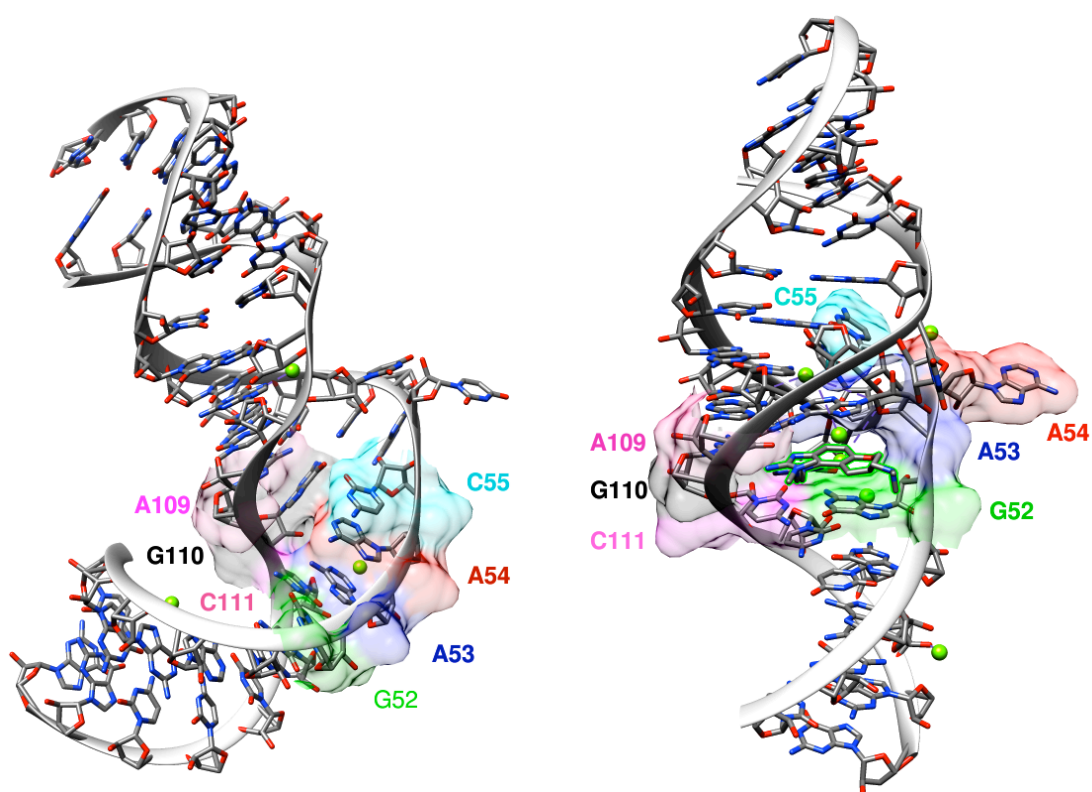


Figure 5.1. Structures of HCV IRES subdomain IIa RNA in the absence (left with PDB code 2NOK (*132*)) and presence (right with PDB code 3TZR (*137*)) of the crystallographic inhibitor. RNA residues in the binding pocket are labeled with matching color surface. The inhibitor is highlighted green in the right figure. The crystallographic magnesium ions are shown as green spheres.

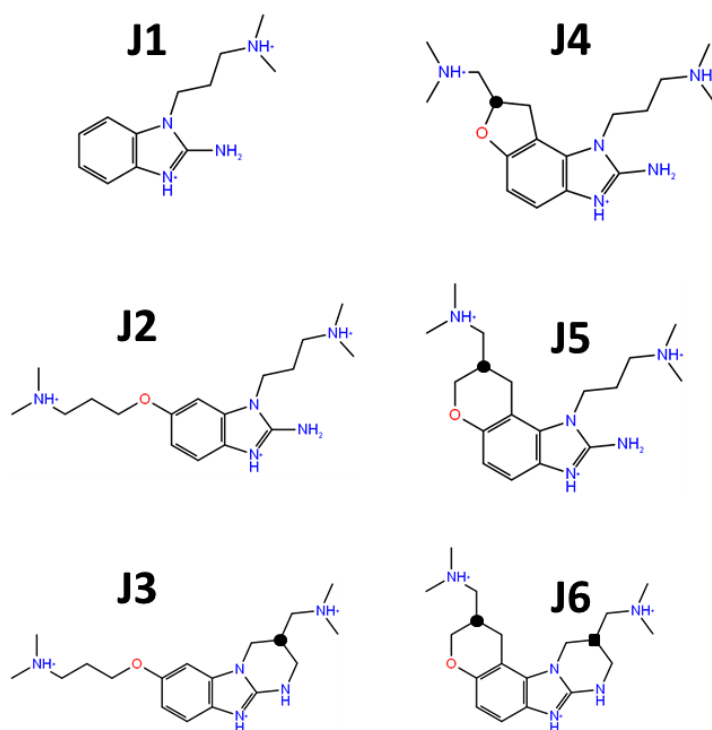


Figure 5.2. Structures of the ligands prepared and used for docking on to the HCV-IRES RNA.<sup>(130)</sup> J1 and J2 have no chiral centers so were called as j1a and j2a in the context. Each of J3, J4 and J5 has one chiral center which is represented as black sphere and makes j3r/s, j4r/s and j5r/s. J6 has two chiral centers represented as a black sphere and a black square which make j6r, j6s, j6q, j6t. J5 is the crystallographic ligand in crystal structure 3TZR.

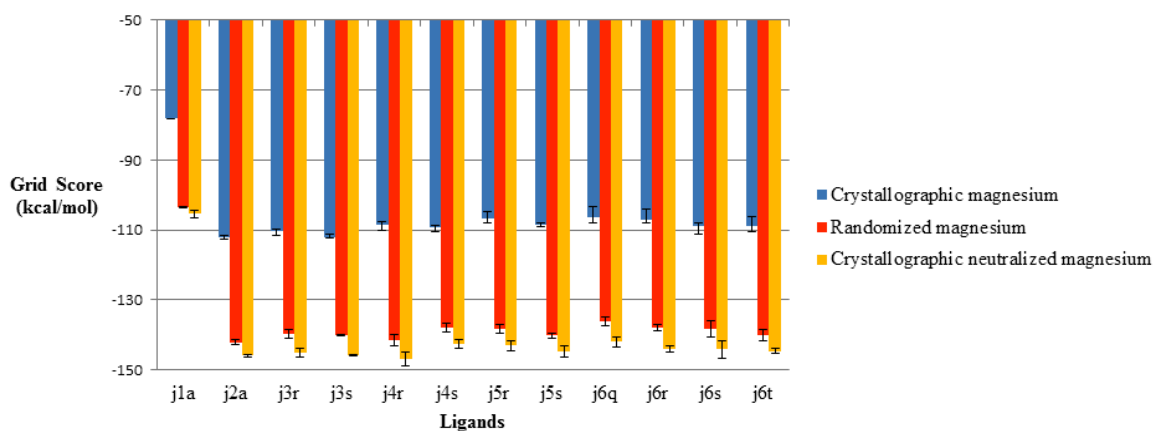


Figure 5.3. Average grid scores produced by *DOCK 6.5* for different ligands when docked on receptors with crystallographic Mg<sup>2+</sup> (blue), randomized Mg<sup>2+</sup> (red) and crystallographic neutral magnesium (yellow). j5r/s is the crystal structure ligand. The error bars shown are standard deviations for docking the different ligand conformations.

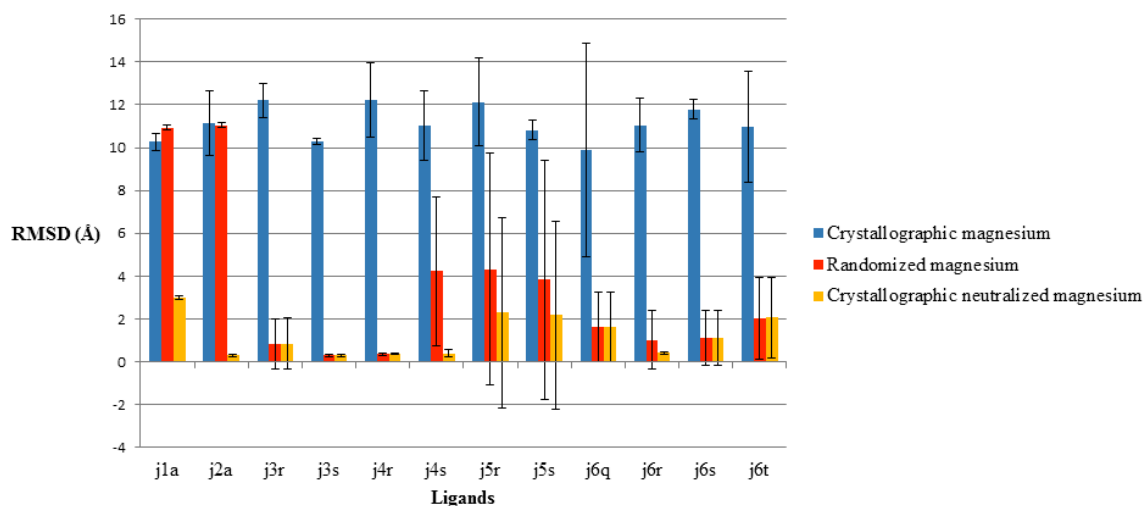


Figure 5.4. Root mean squared deviation of the benzimidazole moiety of the docked ligands to the experimental pose when docked on receptors with crystallographic Mg<sup>2+</sup> (blue), randomized Mg<sup>2+</sup> (red) and crystallographic neutral magnesium (yellow). j5r/s is the crystal structure ligand. Error bars display the observed standard deviations observed when docking the different ligand conformations.

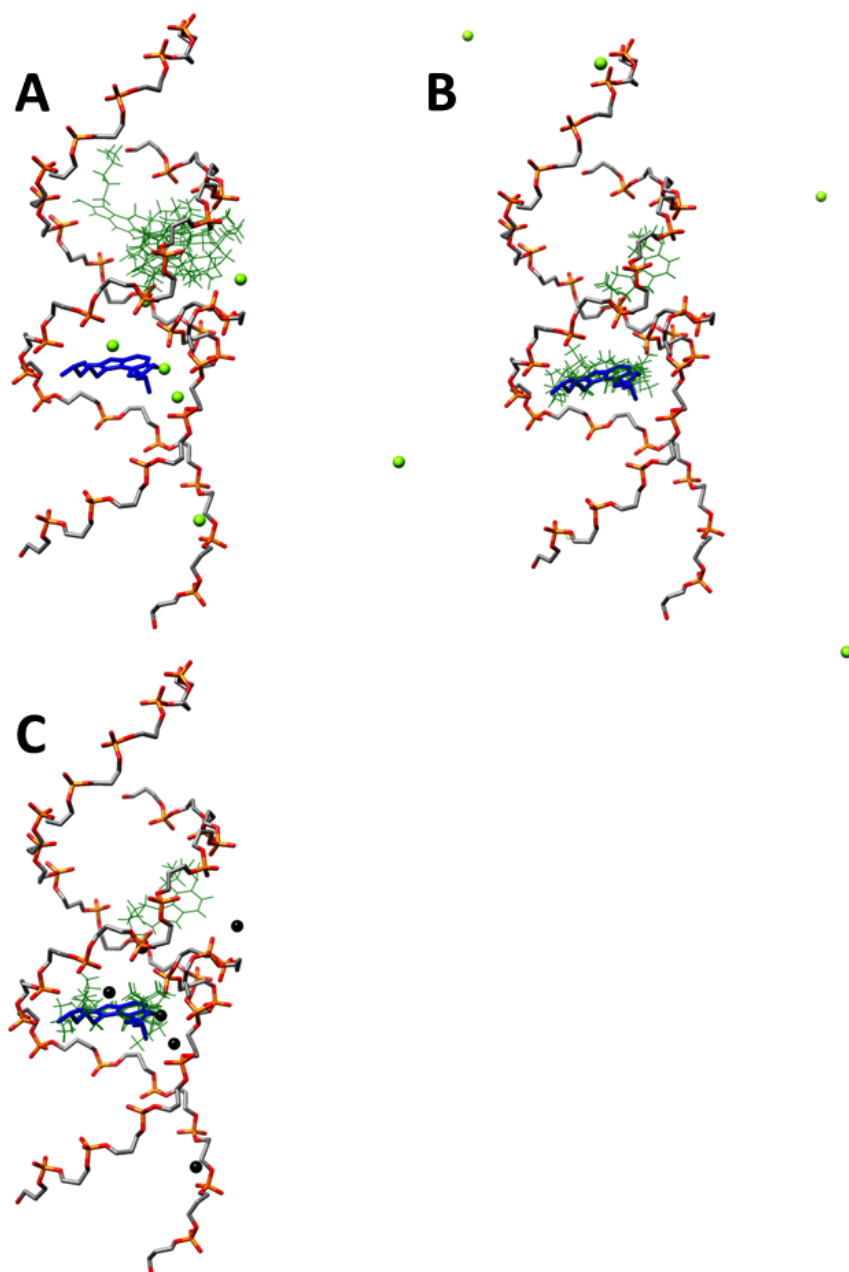


Figure 5.5. The six observed j5r docking poses with highest DOCK scores (green ligands) versus the j5r experimental pose (blue ligand) on receptors with A) crystallographic  $Mg^{2+}$ , B) randomized  $Mg^{2+}$  and C) crystallographic neutral magnesium ions. Normal  $Mg^{2+}$  ions are shown as green spheres and neutral magnesium is shown as black spheres.



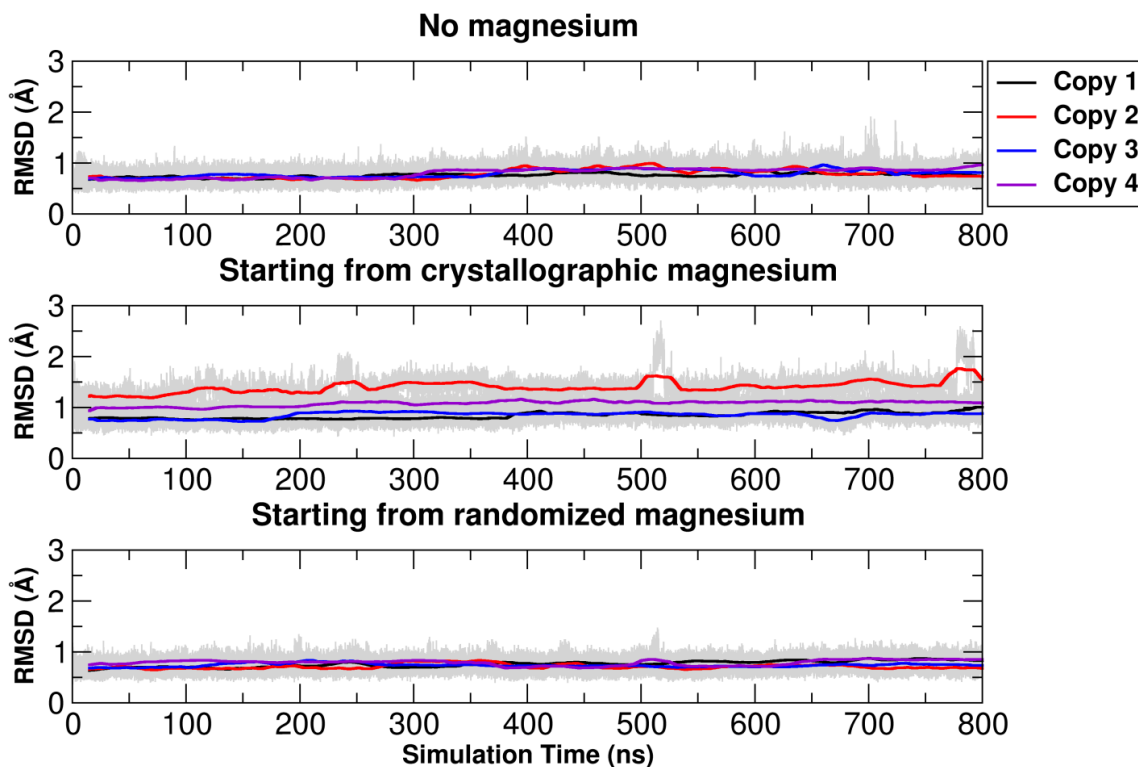


Figure 5.6. Mass weighted Root Mean Square Deviation (RMSD) of the binding pocket (residues G52, A53, A109, G110, C111 and ligand in the PDB code 3TZR) in simulations without  $\text{Mg}^{2+}$ , with crystallographic  $\text{Mg}^{2+}$  and with initially randomized  $\text{Mg}^{2+}$ , with reference to the crystal structure. Solid lines with different colors represent running averages with 20 ns windows for different copies of the simulation.

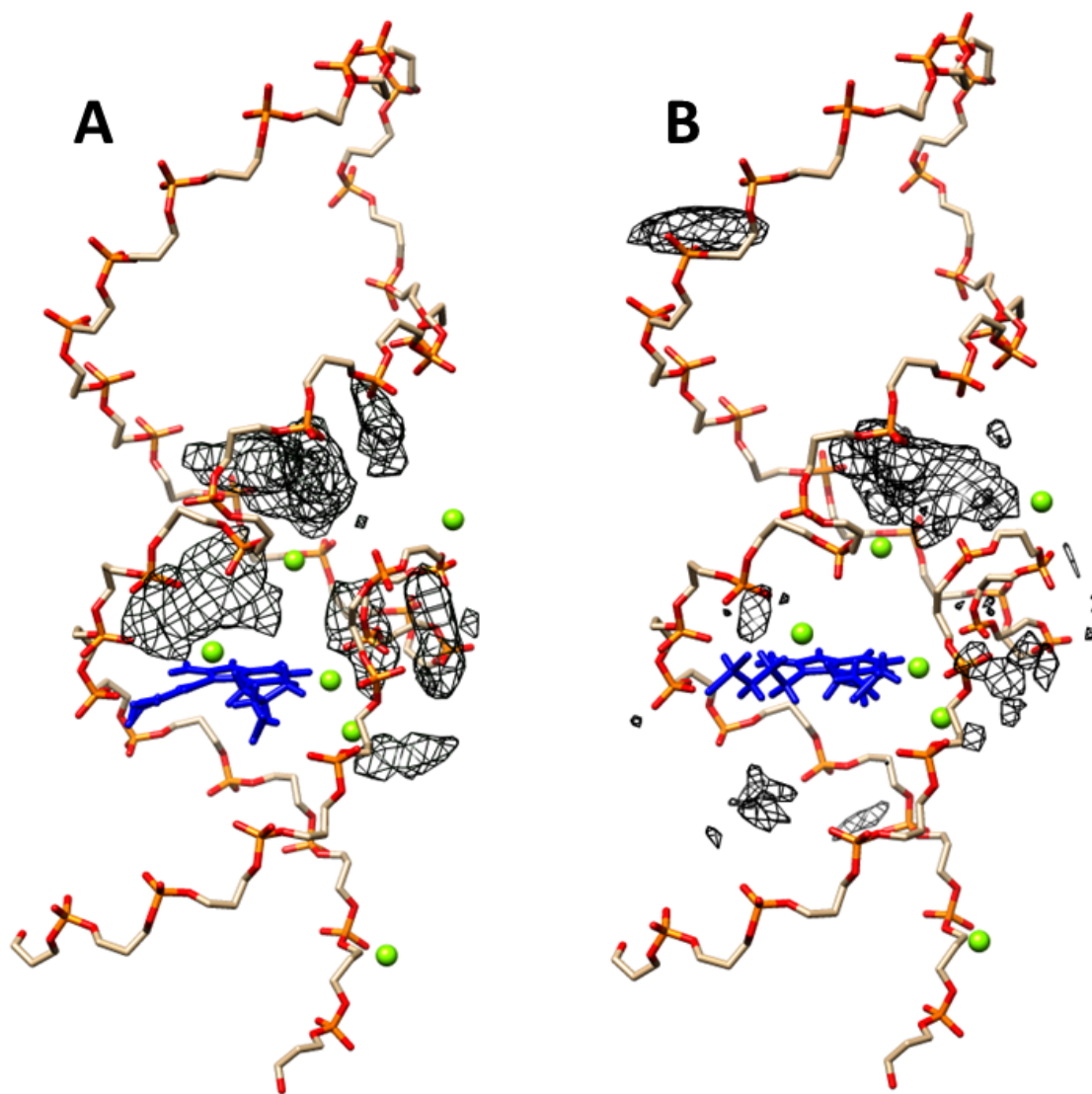


Figure 5.7. Average structure of RNA backbone and ligand atoms over  $\sim 3.5 \mu s$  of aggregate MD trajectory data initiated from A) crystallographic  $Mg^{2+}$  and B) randomized  $Mg^{2+}$ . The positions of crystallographic  $Mg^{2+}$  are shown with green spheres, the ligand is shown in blue and the area that was occupied by  $Mg^{2+}$  more than 100 times the evenly-distributed  $Mg^{2+}$  would occupy is shown with black mesh.

## CHAPTER 6

### CONCLUSION

In this last chapter, we first summarize results from the research presented in Chapters 2-5 and then present possible directions for future research in those projects.

#### **6.1 Significant outcomes**

Conformational distributions of dinucleotide monophosphates were studied in the research presented in Chapter 2. It was demonstrated that classic unbiased MD simulations of these small RNA molecules do not converge even in  $\sim 11 \mu\text{s}$ . Therefore, force field assessments based on such sub- $\mu\text{s}$  simulations can be completely flawed in these systems. In contrast, temperature replica-exchange MD (TREMMD) with 18 replicas as implemented in this research can converge the conformational distributions in less than 500 ns simulation per replica. It was then shown that different dinucleotides adopt consensus A-form, ladder, inverted and sheared conformations in addition to small fractions of extended conformations. We characterized each noncanonical conformation with specific dihedral configurations and/or non-bonding intra-molecular interactions. The 3' high-anti glycosidic dihedral angle and increased fraction of C2'-endo puckers are attributes of ladder and shear conformations. Gauche plus  $\zeta$  dihedral angle and  $(\text{HO}2')_{\text{Xp}}-(\text{OP}1)_{\text{pX}}$  interaction are attributes of the inverted conformation and  $(\text{HO}2')_{\text{Xp}}-(\text{O}5')_{\text{pX}}$

interaction is another characteristic of the ladder conformation. Since the inverted and shear conformations result in the highest suite outlier rates, and since the ladder conformations with high-anti glycosidic dihedral angles are seen in incorrect conformations of bigger RNA molecules, it was concluded that these conformations are less likely to be real or correctly populated. However, more rigorous judgement about the conformations will be a subject for future experimental research. Finally, the effect of different non-bonding interactions and certain dihedral angle configurations on dinucleotide conformations were shown. It was shown that the HO2' and HO5' nonbonding interactions with hydrogen bond acceptors in the backbone and base, respectively, have significant influence on the conformational space of the ApA dinucleotide.

The research project presented in Chapter 3 provides an example of how current protocols of RNA simulations can be used to analyze uncertainties in experimental structural data and how simulation data can be assessed based on the available experimental data. A cation within a highly negative and buried pocket in GTPase Associating Center (GAC) RNA has been identified as  $K^+$  and  $Mg^{2+}$  in two different X-ray crystallography experiments. Here, the cation was shown to be  $K^+$  with using unbiased as well as umbrella sampling MD simulations. The trend of the binding energies for potassium, sodium and magnesium to the pocket in low magnesium concentration supported a hypothesis that described this pocket as the monovalent binding site of the GAC RNA which had been identified in early GAC RNA biochemical experiments (93). Our simulations also predicted water molecules to be located inside the pocket which have not been identified in the crystal structures probably due to refinement issues. The occurrence of a bound magnesium ion close to the buried pocket in the crystal structures

was analyzed with the simulations and was concluded that binding of such magnesium ion is probably the result of high magnesium concentration in crystallization conditions.

The unfolding pathway of the GAC RNA was studied and described in Chapter 4 with regards to the role of different types of ion interactions with RNA. The first step in unfolding pathway of the GAC RNA has been described in experimental literature as very sensitive to the ion concentrations (93,95,141). Here, we examined the sensitivity of the folded GAC RNA to the absence of ions around the RNA and observed that in such extreme conditions the RNA is partially unfolded to intermediate structures in a way that a characterizing loop-loop interaction is lost. To analyze the hypothesis that the loss of this loop-loop interaction is the event equivalent to the first transition in the unfolding pathway of RNA, we measured the free energy needed to separate these loops at different ion conditions via umbrella sampling simulations. We observed that this loop-loop interaction is very sensitive to the presence of cations in the interaction site. Also we showed that a bound  $Mg^{2+}$  that bridges between the loops according to the crystal structures and can stabilize the folded state relative to the intermediate states in all  $Mg^{2+}$  concentrations, an observation in agreement with evidence from melting experiments (95). Therefore, it is very likely that the GAC RNA initiates its thermal unfolding with disengagement of these loops.

Chapter 5 is also an example of assessment of the experimental structures with computational approaches. In this chapter, the hypothesis stating that the binding of a 2-benzimidazole inhibitor to the IRES RNA of Hepatitis C virus is dependent to the existence of the bound magnesium ions was questioned and invalidated using docking and MD simulations. It was demonstrated that not only binding of the experimentally assessed inhibitor, but also binding of the other inhibitors in the same family to this RNA

is not promoted by the existence of the crystallographic magnesium ions. Instead, the positive electrostatic charge of these ions prevents the inhibitors from binding to the crystallographic pocket. However, it was shown that there is another binding pocket on this RNA to which binding of these inhibitors is independent of existence of the magnesium ions.

## **6.2 Future directions**

Regarding the fast convergence of dinucleotide monophosphates with the implemented TREMD method, our simulation protocol for these molecules can be used as an approach for future assessment of any new force field modification. Although lack of experimental data for dinucleotides might still prevent us from making conclusive claims about the conformations generated by each force field modification, the relative changes in intramolecular and intermolecular interactions and dihedral torsion configurations can be tracked quantitatively upon modifying any force field term. The trend observed in dinucleotide conformations upon tweaking any force field term can be used to predict the changes in larger RNA systems. In fact, this dinucleotide analysis platform will be used as a pretesting protocol for future force field modifications. Also, the emergence of new NMR data for dTpT (78) which rules out existence of some dihedral torsion configurations is an encouragement for performing similar experiments on RNA dinucleotides to provide necessary data for assessment of dinucleotide conformations illustrated in our simulations.

In the dinucleotide research presented in Chapter 2, the non-bonding interactions between hydroxyl groups and hydrogen bond acceptors were found to be very determining in changing the conformation distributions. Also, it was shown that the

simulation results are very sensitive to the small changes in van-der-Waals term of the phosphate oxygens. Regarding these observations and the fact that the van-der-Waals parameters and partial charges of nucleic acids have been determined in the mid 1990s without considering the Particle Mesh Ewald (PME) conditions in the simulations (142), it would be wise to reconsider parameterization of these force field terms. One approach for tweaking the partial charges would be using dinucleotide suite structures in a multi-conformation and multi-orientation RESP charge fitting with RED program (136). This approach can potentially result in development of more accurate partial charges that shift the conformational populations towards experimentally correct conformations. However, it should be considered that this approach gives the same weight to different dinucleotide conformational suites when calculating the charges, while these conformations are not equally populated in RNA sequences. Therefore, it is predictable that using all suites in the multi-conformation and multiorientation RESP charge fitting can potentially result in partial charges that shift the RNA populations from highly populated conformations (e.g., A-form). This potential pitfall should be resolved with using selections of suites in charge calculations instead of all.

In Chapters 2 and 3, the interactions of cations with the GAC RNA and their role in the unfolding pathway of this molecule were studied. It is known that the C-terminal domain (CTD) of the L11 protein can also stabilize the folded state of the GAC RNA in absence of magnesium ions (99). Although the GAC crystal structures illustrate the protein conformation as well as its interaction with RNA, the dynamics of this interaction and its effects on the RNA folding and function is still unknown. Investigating the dynamics of the RNA in presence and absence of the protein will provide some insight into the role of the L11 in stability of the RNA and hopefully the role of each partner in

the GAC complex. Also, previous NMR experiments have solved the L11 structure in solution and have estimated the interactions of the more flexible N-terminal domain (NTD) of L11 with the RNA (*143*). The NTD of the L11 is connected to its CTD with a flexible linker and possesses different orientations with respect to the CTD in presence and absence of the RNA. Interchange between the linker conformations exhibit motions in nanosecond-picosecond time scale which can be studied with MD simulations. Also the energetic scheme of the slower global motion of the NTD of L11 can be studied with umbrella sampling simulations with the hope of providing insight into the functional role of the GAC complex and its GTPase activity.

On the other hand, the GAC RNA adopts a relatively complex tertiary structure stabilized by many short-range and long-range interactions. The dynamics of the loop structures is still unknown. It will be interesting to investigate the dynamics of the free loops vs. the loops in the global tertiary RNA structure. Difference or similarity of the dynamics and conformations of local structures in these two conditions can answer more questions about the RNA folding pathway.

The research presented about the HCV IRES RNA in Chapter 5 led to identification of a new ligand binding site. Although the position of this new binding site is roughly identified, the details about the interactions between the ligand and the RNA within this binding site are still unknown. Future docking and molecular dynamics studies of the ligands which specifically bind to this pocket will reveal more details about their binding pose and help with designing more potent drug molecules that can target one or both of the pockets.



## APPENDIX

### SCRIPTS USED FOR BUILDING THE MODELS AND ANALYZING THE SIMULATIONS

## A.1 Scripts related to Chapter 2

### Sample scripts for setting up the systems

Sample tLEaP script (Amber14) for setting up TREMD-TIP3P-FF12 simulations:

```
source leaprc.ff12SB
loadoff atomic_ions.lib
loadamberparams frcmod.tip3p
loadamberparams frcmod.ionsjc_tip3p
m = sequence (144)
check m
solvatebox m TIP3PBOX 7.35 iso 2
addions m Na+ 0
addions m Na+ 5 Cl- 5
saveamberparm m full.topo non-randomized.crds
check m
quit
```

Sample parmed.py script for preparing a parmtop file for TREMD-TIP3P-VDW simulations from the TREMD-TIP3P-FF12 parmtop file:

```
parm ff12.topo.hmr
changeLJSingleType :2@OP1 1.7493 0.2100
changeLJSingleType :2@OP2 1.7493 0.2100
changeLJSingleType :2@O5' 1.7718 0.1700
changeLJSingleType :1@O3' 1.7718 0.1700
addLJType @O4' radius 1.6837 epsilon 0.1700
outparm vdw.topo.hmr
quit
```

frcmod file used in tLEaP for preparing the TREMD-TIP3P-CG simulations:

```
MASS
CA 12.01      0.360
CB 12.01      0.360
CK 12.01      0.360
CM 12.01      0.360
CQ 12.01      0.360
CP 12.01      0.360
CS 12.01      0.360
N* 14.01      0.530
NA 14.01      0.530
N2 14.01      0.530
NC 14.01      0.530
NB 14.01      0.530
O  16.00      0.434
```

## DIHE

```
OS-CT-N*-CK 1 0.760994 50.0 1.0
OS-CT-N*-CM 1 0.760994 50.0 -1.0
OS-CT-N*-CM 1 0.505314 74.0 2.0
```

## NONB

```
CA 1.812596614 0.068800
CB 1.812596614 0.068800
CK 1.812596614 0.068800
CM 1.812596614 0.068800
CQ 1.812596614 0.068800
CP 1.812596614 0.068800
CS 1.812596614 0.068800
N* 1.732800787 0.136002
NA 1.732800787 0.136002
N2 1.732800787 0.136002
NC 1.732800787 0.136002
NB 1.732800787 0.136002
O 1.634259844 0.168000
```

Sample parmed.py script for off diagonal vdW modifications of TREMD-TIP3P-CG simulations:

```
parm ff99.topo
printLJTypes @%CA
printLJTypes @%NA
printLJTypes @%O
changeLJPair @%CA @%OW 3.67622 0.102933
changeLJPair @%NA @%OW 3.592215 0.144722
changeLJPair @%O @%OW 3.429413 0.160848
outparm CG.topo
quit
```

Sample post-processing analysis scripts with CPPTRAJ

Sample script for combined clustering:

```
parm tip3p/ff12sb/run1/build/full.topo.hmr [tip3p-ff12sb]
parm tip3p/dacP-ff12sb/run1/build/full.topo.hmr [tip3p-dacP-ff12sb]
parm tip3p/chen-garcia/run1/build/full.topo.hmr [tip3p-chen-garcia]
parm tip3p/charmm36/run1/build/full.topo.hmr [tip3p-charmm36]
parm opc/ff12sb/run1/build/full.topo.hmr [opc-ff12sb]
parm opc/dacP-ff12sb/run1/build/full.topo.hmr [opc-dacP-ff12sb]

trajin tip3p/ff12sb/run1/traj.1.01 remdtraj remdtrajtemp 298.41 trajnames \
tip3p/ff12sb/run1/traj.1.02,tip3p/ff12sb/run1/traj.1.03,tip3p/ff12sb/run1/traj.1.04,t
ip3p/ff12sb/run1/traj.1.05,tip3p/ff12sb/run1/traj.1.06,tip3p/ff12sb/run1/traj.1.07,ti
```

```
p3p/ff12sb/run1/traj.1.08,tip3p/ff12sb/run1/traj.1.09,tip3p/ff12sb/run1/traj.1.10,ti
p3p/ff12sb/run1/traj.1.11,tip3p/ff12sb/run1/traj.1.12,tip3p/ff12sb/run1/traj.1.13,ti
p3p/ff12sb/run1/traj.1.14,tip3p/ff12sb/run1/traj.1.15,tip3p/ff12sb/run1/traj.1.16,ti
p3p/ff12sb/run1/traj.1.17,tip3p/ff12sb/run1/traj.1.18 parm [tip3p-ff12sb]
```

```
trajin tip3p/ff12sb/run2/traj.1.01 remdtraj remdtrajtemp 298.41 trajnames \
tip3p/ff12sb/run2/traj.1.02,tip3p/ff12sb/run2/traj.1.03,tip3p/ff12sb/run2/traj.1.04,t
ip3p/ff12sb/run2/traj.1.05,tip3p/ff12sb/run2/traj.1.06,tip3p/ff12sb/run2/traj.1.07,ti
p3p/ff12sb/run2/traj.1.08,tip3p/ff12sb/run2/traj.1.09,tip3p/ff12sb/run2/traj.1.10,ti
p3p/ff12sb/run2/traj.1.11,tip3p/ff12sb/run2/traj.1.12,tip3p/ff12sb/run2/traj.1.13,ti
p3p/ff12sb/run2/traj.1.14,tip3p/ff12sb/run2/traj.1.15,tip3p/ff12sb/run2/traj.1.16,ti
p3p/ff12sb/run2/traj.1.17,tip3p/ff12sb/run2/traj.1.18 parm [tip3p-ff12sb]
```

*# Reading trajectories from other scenarios*

```
autoimage origin
strip :WAT,Na+,Cl-
rmsd :1-2&!@H= first mass out test.dat \
    cluster :1,2&!@H= summary clust-summary.dat repout rep repfmt pdb \
    clusterout ctraj clusterfmt netcdf \
    kmeans clusters 7 sieve 200 random \
    savepairedist pairedist CpptrajPairDist summaryhalf split.dat \
    splitframe \
    50000,100000,150000,200000,250000,300000,350000,400000,450000, \
    500000,550000,600000
```

Sample script for KDE analysis:

```
parm run1/build/full.topo.hmr [traj]
trajin run1/traj.1.01 remdtraj remdtrajtemp 298.41 trajnames \
run1/traj.1.02,run1/traj.1.03,run1/traj.1.04,run1/traj.1.05,run1/traj.1.06,run1/traj.1.
07,run1/traj.1.08,run1/traj.1.09,run1/traj.1.10,run1/traj.1.11,run1/traj.1.12,run1/tra
j.1.13,run1/traj.1.14,run1/traj.1.15,run1/traj.1.16,run1/traj.1.17,run1/traj.1.18,
```

```
trajin run2/traj.1.01 remdtraj remdtrajtemp 298.41 trajnames \
run2/traj.1.02,run2/traj.1.03,run2/traj.1.04,run2/traj.1.05,run2/traj.1.06,run2/traj.1.
07,run2/traj.1.08,run2/traj.1.09,run2/traj.1.10,run2/traj.1.11,run2/traj.1.12,run2/tra
j.1.13,run2/traj.1.14,run2/traj.1.15,run2/traj.1.16,run2/traj.1.17,run2/traj.1.18,
```

```
autoimage origin
rmsd :1,2&!@H= first
average avg.rst restart
createcrd crd1
run
reference avg.rst [avg]
crdaction crd1 rms ref [avg] :1-2&!@H=
crdaction crd1 matrix covar :1-2&!@H= name DinucCovar
runanalysis diagmatrix DinucCovar out evecs.dat vecs 20
```

```
crdaction crd1 projection T1 modes evecs.dat beg 1 end 10 :1-2&!@H= \
  crdframes 1,50000 out T1.dat
```

```
crdaction crd1 projection T2 modes evecs.dat beg 1 end 10 :1-2&!@H= \
  crdframes 50001,100000 out T2.dat
```

```
kde T1:1 kldiv T2:1 klout KL-PC.agr bins 400 name TREMD-1
kde T1:2 kldiv T2:2 klout KL-PC.agr bins 400 name TREMD-2
kde T1:3 kldiv T2:3 klout KL-PC.agr bins 400 name TREMD-3
kde T1:4 kldiv T2:4 klout KL-PC.agr bins 400 name TREMD-4
kde T1:5 kldiv T2:5 klout KL-PC.agr bins 400 name TREMD-5
kde T1:6 kldiv T2:6 klout KL-PC.agr bins 400 name TREMD-6
kde T1:7 kldiv T2:7 klout KL-PC.agr bins 400 name TREMD-7
kde T1:8 kldiv T2:8 klout KL-PC.agr bins 400 name TREMD-8
kde T1:9 kldiv T2:9 klout KL-PC.agr bins 400 name TREMD-9
kde T1:10 kldiv T2:10 klout KL-PC.agr bins 400 name TREMD-10
kde T1:1 out kde-PC.agr bins 400 name KDE1-1
kde T2:1 out kde-PC.agr bins 400 name KDE1-2
hist T1:1,*,*,*,200 out pca.hist.agr normint name PC1-RUN1
hist T1:2,*,*,*,200 out pca.hist.agr normint name PC2-RUN1
hist T1:3,*,*,*,200 out pca.hist.agr normint name PC3-RUN1
hist T1:4,*,*,*,200 out pca.hist.agr normint name PC4-RUN1
hist T1:5,*,*,*,200 out pca.hist.agr normint name PC5-RUN1
hist T2:1,*,*,*,200 out pca.hist.agr normint name PC1-RUN2
hist T2:2,*,*,*,200 out pca.hist.agr normint name PC2-RUN2
hist T2:3,*,*,*,200 out pca.hist.agr normint name PC3-RUN2
hist T2:4,*,*,*,200 out pca.hist.agr normint name PC4-RUN2
hist T2:5,*,*,*,200 out pca.hist.agr normint name PC5-RUN2
run
modes eigenval name evecs.dat out modes-fraction.dat
```

### **A.2 Scripts related to Chapter 3**

#### Sample scripts for setting up the systems

Sample tLEap script (Amber 12) for setting up COMPETITION simulations:

```
source leaprc.ff12SB
loadoff ions08.lib
loadamberparams frcmod.ionsjc_tip3p
loadamberparams MG_villa.frcmod
Mg = loadmol2 MG.mol2
m = loadpdb min-gb.pdb
check m
solvateoct m TIP3PBOX 12 iso
addions m Mg+ 28
```

```

addions m K+ 28 Cl- 28
saveamberparm m full.topo non-randomized.crds
quit

```

Sample molecular dynamics input file for COMPETITION simulations:

```

Production run: 40 ns
&cntrl
  irst = 1, ntx = 5, temp0 = 298.15,
  ntp=5000, ntwr=5000, ntwx=5000, ntxo=2, ioutfm=1,
  ntf=2, ntc=2, nstlim=20000000, dt=0.002, iwrap=1,
  ntp=1, ntt=1, tautp=10.0, taup=10.0, cut=8.0,
/

```

Sample molecular dynamics input file for ion pull-out simulations:

```

&cntrl
  imin = 0, nstlim = 100000, dt=0.002,
  irst = 1, ntx = 5, ig = -1,
  ntwx = 5000, ioutfm = 1, ntp = 5000, ntwr = 50000, ntxo = 2,
  iwrap = 1, nscm = 1000,
  ntc = 2, ntf = 2, ntb = 2, cut = 8.0,
  ntt = 3, gamma_ln = 5, temp0 = 298.15, teMPI = 298.15,
  ntp = 1, taup = 5.0,
  nmropt = 1,
&end
&wt
  TYPE="DUMPFREQ", istep1=10,
&end
&wt
  TYPE="END",
&end
DISANG=pull.disang
DUMPAVE=dist.dat

```

Sample restraint file for pull-out production simulations (pull.disang mentioned in the input file):

```

&rst iat=1834,-1,0,
  r1=0.0, r2=12.400, r3=12.400, r4=999.0, rk2=100.0, rk3=100.0,

  IGR2(1)=581,IGR2(2)=582,IGR2(3)=583,IGR2(4)=648,IGR2(5)=649,IGR2(6)=6
  50,IGR2(7)=679,IGR2(8)=680,IGR2(9)=681, nstep1=0, nstep2=0,
&end

```

Sample restraint file for directing ion to make the umbrella initial structures from 6.9 to 10.0 Å in umbrella sampling:

```

&rst iat=1834,-1,0,
  r1=0.0, r2=6.9, r3=6.9, r4=999.0, rk2=100.0, rk3=100.0,
  r1a=0.0, r2a=20.0, r3a=20.0, r4a=999, rk2a=100.0, rk3a=100.0,

IGR2(1)=581,IGR2(2)=582,IGR2(3)=583,IGR2(4)=648,IGR2(5)=649,IGR2(6)=6
50,IGR2(7)=679,IGR2(8)=680,IGR2(9)=681, nstep1=0, nstep2=130000, ifvari=1,
&end
&rst
  iresid=0, iat=581,648,1834, r1=0.000, r2=65.0, r3=65.0, r4=360.0, rk2=200,
rk3=200,
&end
&rst
  iresid=0, iat=679,581,648,1834, r1=0.000, r2=340.0, r3=340.0, r4=360.0,
rk2=200, rk3=200,
&end

```

### Sample postprocessing scripts with CPPTRAJ

Sample ion randomizing script for COMPETITION simulations (different random seeds for ‘randomizeions’ were used for different simulations):

```

parm full.topo
trajin non-randomized.crds
trajout rand.crds restart
randomizeions :K+,Mg+,Cl- around :1-57 by 6.0 \
  overlap 4.0 noimage seed 5

```

Sample script for checking the closest  $K^+$  to the K/Mg58 pocket in the initial structures:

```

parm full.topo
trajin rand.crds
solvent :K+
closest 1 :21@P,OP1,OP2 first \
  closestout closest-K-in-rands.dat
run

```

Sample closest potassium identification script:

```

parm ../full.topo
trajin ../m1/md/traj.1
# Reading trajectories from all 20 copies (m1..m20)
autoimage origin
rmsd rmsdRNA :1-57 first mass
solvent :K+
closest 1 :19@P,OP1,OP2,:21@P,OP1,OP2,:22@P,OP1,OP2 \
  center first closestout closest-K-RC.dat \

```

```

outprefix closest-K-RC
run
writedata dist_closest-K-RC.dat CLOSEST_00003[Dist]

```

Sample ion hydration shell counting script:

```

parm ../full.topo
trajin ../m1/md/traj.1
# Reading trajectories from all 20 copies (m1..m20)
watershell :58@K+ watershell-K.dat lower 3.2 upper 4.5
run
hist WS_00000[lower],0,8,1 out shell_1st-K.dat norm

```

hbond analysis for occupation of Mg163 site:

```

parm ../full.topo
trajin ../md/traj.1
autoimage origin
hbond out mg-163.out nointramol solvout mg-163.dat \
  solventdonor :Mg+ acceptormask :19@OP2,:21@OP1 \
  series mg-163 printatomnum dist 3.0

```

### **A.3 Scripts related to Chapter 4**

#### Sample script for running the simulations

Sample restraint file for umbrella sampling simulations:

```

&rst iat=-1,-1, nstep1=0, nstep2=0,
  r1=0.0, r2=08.000, r3=08.000, r4=999.0, rk2=100.0, rk3=100.0,
  IGR1=581,679,
  IGR2=1356,1452,
&end

```

#### Sample CPPTRAJ scripts

Sample script for displacing the bound magnesium

```

parm build/full.topo
trajin c0-b/md/w-08.000/step7.rst
randomizeions :58 around :1-57 by 6.0 overlap 4.0 noimage
trajout step7.rst restart

```



The script used for grid analysis related to Figure 4.7:

```

parm neut-KCl-run1/exclusion-c0-0A/build/full.topo

trajin neut-KCl-run1/exclusion-c0-0A/md/w-09.000/prod_2.traj
trajin neut-KCl-run1/exclusion-c0-0A/md/w-09.000/prod_3.traj
trajin neut-KCl-run2/exclusion-c0-0A/md/w-09.000/prod_2.traj
trajin neut-KCl-run2/exclusion-c0-0A/md/w-09.000/prod_3.traj
  # Reading trajectories from umbrellas 09.100 to 09.900
autoimage origin
rmsd :1-57&!@H= first mass
average avg.pdb pdb :1-57
run
parm avg.pdb [nowat]
reference avg.pdb parm [nowat] [avg]
rmsd ref [avg] mass :1-57&!@H=
grid k.xplor 150 0.5 150 0.5 150 0.5 :K+

```

#### **A.4 Scripts related to Chapter 5**

##### Sample dock input script

ligand_atom_file	all-confs.mol2
limit_max_ligands	no
skip_molecule	no
read_mol_solvation	no
calculate_rmsd	no
use_database_filter	no
orient_ligand	yes
automated_matching	yes
receptor_site_file	../rec/selected_spheres.sph
max_orientations	10000
critical_points	no
chemical_matching	no
use_ligand_spheres	no
use_internal_energy	yes
internal_energy_rep_exp	12
flexible_ligand	yes
user_specified_anchor	no
limit_max_anchors	no
min_anchor_size	5
pruning_use_clustering	yes
pruning_max_orients	500
pruning_clustering_cutoff	100
pruning_conformer_score_cutoff	25
use_clash_overlap	no

write_growth_tree	no
bump_filter	yes
bump_grid_prefix	../../rec/grid
max_bumps_anchor	5
max_bumps_growth	5
score_molecules	yes
contact_score_primary	no
contact_score_secondary	no
grid_score_primary	yes
grid_score_secondary	no
grid_score_rep_rad_scale	1
grid_score_vdw_scale	1
grid_score_es_scale	1
grid_score_grid_prefix	../../rec/grid
dock3.5_score_secondary	no
continuous_score_secondary	no
descriptor_score_secondary	no
gbsa_zou_score_secondary	no
gbsa_hawkins_score_secondary	no
amber_score_secondary	no
minimize_ligand	yes
minimize_anchor	yes
minimize_flexible_growth	yes
use_advanced_simplex_parameters	no
simplex_max_cycles	500
simplex_score_converge	0.1
simplex_cycle_converge	1.0
simplex_trans_step	1.0
simplex_rot_step	0.1
simplex_tors_step	5.0
simplex_anchor_max_iterations	1000
simplex_grow_max_iterations	1000
simplex_grow_tors_premin_iterations	100
simplex_random_seed	378
simplex_restraint_min	no
atom_model	all
vdw_defn_file	vdw_AMBER_parm99.defn
flex_defn_file	flex.defn
flex_drive_file	flex_drive.tbl
ligand_outfile_prefix	pose
write_orientations	no
num_scored_conformers	1000
write_conformations	yes
cluster_conformations	no
rank_ligands	yes
max_ranked_ligands	500

## REFERENCES

1. Davis, A. M.; Teague, S. J.; Kleywegt, G. J., Application and Limitations of X-Ray Crystallographic Data in Structure-Based Ligand and Drug Design. *Angew. Chem. Int. Ed. Engl.* **2003**, *42*, 2718-2736.
2. Acharya, K. R.; Lloyd, M. D., The Advantages and Limitations of Protein Crystal Structures. *Trends Pharmacol. Sci.* **2005**, *26*, 10-14.
3. Davis, A. M.; St-Galley, S. A.; Kleywegt, G. J., Limitations and Lessons in the Use of X-Ray Structural Information in Drug Design. *Drug Discov. Today* **2008**, *13*, 831-841.
4. Wider, G.; Wuthrich, K., Nmr Spectroscopy of Large Molecules and Multimolecular Assemblies in Solution. *Curr. Opin. Struct. Biol.* **1999**, *9*, 594-601.
5. Kleckner, I. R.; Foster, M. P., An Introduction to Nmr-Based Approaches for Measuring Protein Dynamics. *Biochim. Biophys. Acta* **2011**, *1814*, 942-968.
6. Borhani, D. W.; Shaw, D. E., The Future of Molecular Dynamics Simulations in Drug Discovery. *J. Comput. Aided Mol. Des.* **2012**, *26*, 15-26.
7. Drysdale, M. J.; Lentzen, G.; Matassova, N.; Murchie, A. I.; Aboul-Ela, F.; Afshar, M., Rna as a Drug Target. *Prog. Med. Chem.* **2002**, *39*, 73-119.
8. McKnight, K. L.; Heinz, B. A., Rna as a Target for Developing Antivirals. *Antivir. Chem. Chemother.* **2003**, *14*, 61-73.
9. Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., III, Reliable Oligonucleotide Conformational Ensemble Generation in Explicit Solvent for Force Field Assessment Using Reservoir Replica Exchange Molecular Dynamics Simulations. *J. Phys. Chem. B* **2013**, *117*, 4014-4027.
10. Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E., III, Highly Sampled Tetranucleotide and Tetraloop Motifs Enable Evaluation of Common Rna Force Fields. *RNA* **2015**, *21*, 1578-1590.
11. Yildirim, I.; Stern, H. A.; Tubbs, J. D.; Kennedy, S. D.; Turner, D. H., Benchmarking

Amber Force Fields for Rna: Comparisons to Nmr Spectra for Single-Stranded R(GACC) Are Improved by Revised X Torsions. *J. Phys. Chem. B* **2011**, *115*, 9261-9270.

12. Tubbs, J. D.; Condon, D. E.; Kennedy, S. D.; Hauser, M.; Bevilacqua, P. C.; Turner, D. H., The Nuclear Magnetic Resonance of Cccc RNA Reveals a Right-Handed Helix, and Revised Parameters for Amber Force Field Torsions Improve Structural Predictions from Molecular Dynamics. *Biochemistry (Mosc)*. **2013**, *52*, 996-1010.

13. Kührová, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Sponer, J.; Otyepka, M.; Banáš, P., Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput.* **2016**, *12*, 4534–4548.

14. Chen, A. A.; Garcia, A. E., High-Resolution Reversible Folding of Hyperstable Rna Tetraloops Using Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16820-16825.

15. Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; HersHKovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M., Rna Backbone: Consensus All-Angle Conformers and Modular String Nomenclature (an Rna Ontology Consortium Contribution). *RNA* **2008**, *14*, 465-481.

16. Cheatham, T. E., III; Case, D. A., Twenty-Five Years of Nucleic Acid Simulations. *Biopolymers* **2013**, *99*, 969-977.

17. Bergonzo, C.; Cheatham, T. E., III, Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem. Theory Comput.* **2015**, *11*, 3969-3972.

18. Thirumalai, D.; Woodson, S. A., Kinetics of Folding of Proteins and RNA. *Acc. Chem. Res.* **1996**, *29*, 433-439.

19. Bayly, C. I.; Cieplak, P.; Cornell, W.; Kollman, P. A., A Well-Behaved Electrostatic Potential Based Method Using Charge Restraints for Deriving Atomic Charges: The Resp Model. *J. Phys. Chem.* **1993**, *97*, 10269-10280.

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

21. Åqvist, J., Ion-Water Interaction Potentials Derived from Free Energy Perturbation Simulations. *J. Phys. Chem.* **1990**, *94*, 8021-8024.

22. Allnér, O.; Nilsson, L.; Villa, A., Magnesium Ion–Water Coordination and Exchange in Biomolecular Simulations. *J. Chem. Theory Comput.* **2012**, *8*, 1493-1502.
23. Li, P.; Roberts, B. P.; Chakravorty, D. K.; Merz, K. M., Rational Design of Particle Mesh Ewald Compatible Lennard-Jones Parameters for +2 Metal Cations in Explicit Solvent. *J. Chem. Theory Comput.* **2013**, *9*, 2733-2748.
24. Li, P.; Merz, K. M., Taking into Account the Ion-Induced Dipole Interaction in the Nonbonded Model of Ions. *J. Chem. Theory Comput.* **2014**, *10*, 289-297.
25. Joung, I. S.; Cheatham, T. E., III, Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112*, 9020-9041.
26. Feenstra, K. A.; Hess, B.; Berendsen, H. J. C., Improving Efficiency of Large Time-Scale Molecular Dynamics Simulations of Hydrogen-Rich Systems. *J. Comput. Chem.* **1999**, *20*, 786-798.
27. Hopkins, C. W.; Le Grand, S.; Walker, R. C.; Roitberg, A. E., Long-Time-Step Molecular Dynamics through Hydrogen Mass Repartitioning. *J. Chem. Theory Comput.* **2015**, *11*, 1864-1874.
28. Pan, T.; Sosnick, T. R., Intermediates and Kinetic Traps in the Folding of a Large Ribozyme Revealed by Circular Dichroism and Uv Absorbance Spectroscopies and Catalytic Activity. *Nat. Struct. Biol.* **1997**, *4*, 931-938.
29. Misawa, K.; Lee, T. M.; Ogawa, S., A Study on the Exchange Rate of Magnesium with Atp. *Biochim. Biophys. Acta* **1982**, *718*, 227-229.
30. González, M. A., Force Fields and Molecular Dynamics Simulations. *JDN* **2011**, *12*, 169-200.
31. Sugita, Y.; Okamoto, Y., Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141-151.
32. Hamelberg, D.; Mongan, J.; McCammon, J. A., Accelerated Molecular Dynamics: A Promising and Efficient Simulation Method for Biomolecules. *J. Chem. Phys.* **2004**, *120*, 11919-11929.
33. Roe, D. R.; Bergonzo, C.; Cheatham, T. E., III, Evaluation of Enhanced Sampling Provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. *J. Phys. Chem. B* **2014**, *118*, 3543-3552.

34. Narberhaus, F., Rna Biology: An Introduction. By Gunter Meister. *ChemBioChem* **2011**, *12*, 2700-2700.
35. Hall, K. B., Rna in Motion. *Curr. Opin. Chem. Biol.* **2008**, *12*, 612-618.
36. Davies, D. B., Conformations of Nucleosides and Nucleotides. *Prog. Nucl. Magn. Reson. Spectrosc.* **1978**, *12*, 135-225.
37. Lee, C. H.; Ezra, F. S.; Kondo, N. S.; Sarma, R. H.; Danyluk, S. S., Conformational Properties of Dinucleoside Monophosphates in Solution: Dipurines and Dipyrimidines. *Biochemistry (Mosc)*. **1976**, *15*, 3627-3639.
38. Tsuboi, M.; Takahashi, S.; Kyogoku, Y.; Hayatsu, H.; Ukita, T.; Kainosho, M., Phosphorus-Proton Spin-Spin Coupling and Conformation of a Dinucleoside Phosphate. *Science* **1969**, *166*, 1504-1505.
39. Rubin, J.; Brennan, T.; Sundaralingam, M., Crystal Structure of a Naturally Occurring Dinucleoside Monophosphate: Uridyl (3',5') Adenosine Hemihydrate. *Science* **1971**, *174*, 1020-1022.
40. Rubin, J.; Brennan, T.; Sundaralingam, M., Crystal and Molecular Structure of a Naturally Occurring Dinucleoside Monophosphate. Uridyl-(3'-5')-Adenosine Hemihydrate. Conformational "Rigidity" of the Nucleotide Unit and Models for Polynucleotide Chain Folding. *Biochemistry (Mosc)*. **1972**, *11*, 3112-3128.
41. Sundaralingam, M., Stereochemistry of Nucleic Acids and Their Constituents. Iv. Allowed and Preferred Conformations of Nucleosides, Nucleoside Mono-, Di-, Tri-, Tetraphosphates, Nucleic Acids and Polynucleotides. *Biopolymers* **1969**, *7*, 821-860.
42. Vokáčová, Z.; Buděšínský, M.; Rosenberg, I.; Schneider, B.; Šponer, J.; Sychrovský, V., Structure and Dynamics of the Apa, Apc, Cpa, and Cpc Rna Dinucleoside Monophosphates Resolved with Nmr Scalar Spin-Spin Couplings. *J. Phys. Chem. B* **2009**, *113*, 1182-1191.
43. Yokoyama, S.; Inagaki, F.; Miyazawa, T., Advanced Nuclear Magnetic Resonance Lanthanide Probe Analyses of Short-Range Conformational Interrelations Controlling Ribonucleic Acid Structures. *Biochemistry (Mosc)*. **1981**, *20*, 2981-2988.
44. Munzarová, M. L.; Sklenář, V., Dft Analysis of Nmr Scalar Interactions across the Glycosidic Bond in DNA. *J. Am. Chem. Soc.* **2003**, *125*, 3649-3658.
45. Haasnoot, C. A. G.; de Leeuw, F. A. A. M.; Altona, C., The Relationship between Proton-Proton Nmr Coupling Constants and Substituent Electronegativities—I.

*Tetrahedron* **1980**, *36*, 2783-2792.

46. Wijmenga, S. S.; van Buuren, B. N. M., The Use of Nmr Methods for Conformational Studies of Nucleic Acids. *Prog. Nucl. Magn. Reson. Spectrosc.* **1998**, *32*, 287-387.

47. Ezra, F. S.; Lee, C.-H.; Kondo, N. S.; Danyluk, S. S.; Sarma, R. H., Conformational Properties of Purine-Pyrimidine and Pyrimidine-Purine Dinucleoside Monophosphates. *Biochemistry (Mosc).* **1977**, *16*, 1977-1987.

48. Frechet, D.; Ehrlich, R.; Remy, P.; Gabarro-Arpa, J., Thermal Perturbation Differential Spectra of Ribonucleic Acids. Ii. Nearest Neighbour Interactions. *Nucleic Acids Res.* **1979**, *7*, 1981-2001.

49. Brown, R. F.; Andrews, C. T.; Elcock, A. H., Stacking Free Energies of All DNA and Rna Nucleoside Pairs and Dinucleoside-Monophosphates Computed Using Recently Revised Amber Parameters and Compared with Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 2315-2328.

50. Davis, R. C.; Tinoco, I., Temperature-Dependent Properties of Dinucleoside Phosphates. *Biopolymers* **1968**, *6*, 223-242.

51. Wang, J.; Cieplak, P.; Kollman, P. A., How Well Does a Restrained Electrostatic Potential (Resp) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049-1074.

52. Perez, A.; Marchan, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M., Refinement of the Amber Force Field for Nucleic Acids: Improving the Description of Alpha/Gamma Conformers. *Biophys. J.* **2007**, *92*, 3817-3829.

53. Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P., Refinement of the Cornell Et Al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem. Theory Comput.* **2011**, *7*, 2886-2902.

54. Steinbrecher, T.; Latzer, J.; Case, D. A., Revised Amber Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **2012**, *8*, 4405-4412.

55. Izadi, S.; Anandakrishnan, R.; Onufriev, A. V., Building Water Models: A Different Approach. *J. Phys. Chem. Lett.* **2014**, *5*, 3863-3871.

56. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935.

57. Foloppe, N.; MacKerell, J. A. D., All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86-104.
58. Denning, E. J.; Priyakumar, U. D.; Nilsson, L.; Mackerell, A. D., Jr., Impact of 2'-Hydroxyl Sampling on the Conformational Properties of Rna: Update of the Charmm All-Atom Additive Force Field for Rna. *J. Comput. Chem.* **2011**, *32*, 1929-1943.
59. Horn, H. W.; Swope, W. C.; Pitner, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T., Development of an Improved Four-Site Water Model for Biomolecular Simulations: Tip4p-Ew. *J. Chem. Phys.* **2004**, *120*, 9665-9678.
60. Roe, D. R.; Cheatham, T. E., III, Ptraj and Cpptraj: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9*, 3084-3095.
61. Crowley, M. F.; Williamson, M. J.; Walker, R. C., Chamber: Comprehensive Support for Charmm Force Fields within the Amber Software. *Int. J. Quantum Chem.* **2009**, *109*, 3767-3772.
62. Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C., Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327-341.
63. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R., Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684-3690.
64. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G., A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577-8593.
65. Case, D.; Babin, V.; Berryman, J.; Betz, R.; Cai, Q.; Cerutti, D.; Cheatham, T. E., III; Darden, T.; Duke, R.; Gohlke, H.; Gusarov, S.; Homeyer, N.; Janowski, P.; Kaus, J.; Kolossvár, I.; Kovalenko, A.; Lee, T.; LeGrand, S.; Luchko, T.; Luo, R.; Madej, B.; Merz, K.; Paesani, F.; Roe, D. R.; Roitberg, A.; Asagui, C.; Salomon-Ferrer, R.; Saeabra, G.; Simmerling, C.; Smith, W.; Swails, J.; Walker, R.; Wang, J.; Wolf, R.; Wu, X.; Kollman, P. A., Amber 14. **2014**.
66. Uberuaga, B. P.; Anghel, M.; Voter, A. F., Synchronization of Trajectories in Canonical Molecular-Dynamics Simulations: Observation, Explanation, and Exploitation. *J. Chem. Phys.* **2004**, *120*, 6363-6374.
67. Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E., Bad Seeds Sprout Perilous



Dynamics: Stochastic Thermostat Induced Trajectory Synchronization in Biomolecules. *J. Chem. Theory Comput.* **2009**, *5*, 1624-1631.

68. Patriksson, A.; van der Spoel, D., A Temperature Predictor for Parallel Tempering Simulations. *Phys. Chem. Chem. Phys.* **2008**, *10*, 2073-2077.

69. Hehre, W. J.; Ditchfield, R.; Pople, J. A., Self—Consistent Molecular Orbital Methods. Xii. Further Extensions of Gaussian—Type Basis Sets for Use in Molecular Orbital Studies of Organic Molecules. *J. Chem. Phys.* **1972**, *56*, 2257-2261.

70. Hehre, W. J., *Ab Initio Molecular Orbital Theory*; Wiley: New York, 1986; pp 576.

71. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*, Gaussian, Inc.: Wallingford, CT, USA, 2009.

72. Steinhaus, H., Sur La Division Des Corps Matériels En Parties. *Bull. Acad. Polon. Sci. Cl. III.* **1957**, *4*, 801-804.

73. MacQueen, J. In *Some Methods for Classification and Analysis of Multivariate Observations*, Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics, Berkeley, Calif., 1967; University of California Press: Berkeley, Calif., 1967; pp 281-297.

74. Lee, C.-H.; Evans, F. E.; Sarma, R. H., The Gain in Conformational Purity and Loss in Flexibility as a Result of 3',5' Polymerization between the Component Mononucleotides — a 300 Mhz 1h and 40.5 Mhz 31p Nmr Comparative Study of the Dynamic Solution Conformation of Dinucleoside Monophosphates and Their Component Monomers. *FEBS Lett.* **1975**, *51*, 73-79.

75. Chachaty, C.; Zemb, T.; Langlet, G.; Son, T. D.; Buc, H.; Morange, M., A Proton-Relaxation-Time Study of the Conformation of Some Purine and Pyrimidine 5'-Nucleotides in Aqueous Solution. *Eur. J. Biochem.* **1976**, *62*, 45-53.

76. Hruska, F. E., Long-Range Spin–Spin Coupling in Pyrimidine Nucleosides. *Canad. J. Chem.* **1971**, *49*, 2111-2118.
77. Davies, D. B.; Danyluk, S. S., Nuclear Magnetic Resonance Studies of 5'-Ribo- and Deoxyribonucleotide Structures in Solution. *Biochemistry (Mosc)*. **1974**, *13*, 4417-4434.
78. Nganou, C.; Kennedy, S. D.; McCamant, D. W., Disagreement between the Structure of the DtpT Thymine Pair Determined by Nmr and Molecular Dynamics Simulations Using Amber 14 Force Fields. *J. Phys. Chem. B* **2016**, *120*, 1250-1258.
79. Mlýnský, V.; Banáš, P.; Hollas, D.; Réblová, K.; Walter, N. G.; Šponer, J.; Otyepka, M., Extensive Molecular Dynamics Simulations Showing That Canonical G8 and Protonated A38h<sup>+</sup> Forms Are Most Consistent with Crystal Structures of Hairpin Ribozyme. *J. Phys. Chem. B* **2010**, *114*, 6642-6652.
80. Woodson, S. A., Compact Intermediates in RNA Folding. *Annu. Rev. Biophys.* **2010**, *39*, 61-77.
81. Heilman-Miller, S. L.; Pan, J.; Thirumalai, D.; Woodson, S. A., Role of Counterion Condensation in Folding of the Tetrahymena Ribozyme. II. Counterion-Dependence of Folding Kinetics. *J. Mol. Biol.* **2001**, *309*, 57-68.
82. Draper, D. E., A Guide to Ions and RNA Structure. *RNA* **2004**, *10*, 335-343.
83. Woodson, S. A., Metal Ions and Rna Folding: A Highly Charged Topic with a Dynamic Future. *Curr. Opin. Chem. Biol.* **2005**, *9*, 104-109.
84. Auffinger, P.; Grover, N.; Westhof, E., Metal Ion Binding to RNA. *Met. Ions Life Sci.* **2011**, *9*, 1-35.
85. Auffinger, P.; D'Ascenzo, L.; Ennifar, E., Sodium and Potassium Interactions with Nucleic Acids. In *The Alkali Metal Ions: Their Role for Life*, Sigel, A.; Sigel, H.; Sigel, O. R. K., Eds. Springer International Publishing: Cham, 2016; pp 167-201.
86. Draper, D. E., Rna Folding: Thermodynamic and Molecular Descriptions of the Roles of Ions. *Biophys. J.* **2008**, *95*, 5489-5495.
87. Tan, Z. J.; Chen, S. J., Importance of Diffuse Metal Ion Binding to Rna. *Met. Ions Life Sci.* **2011**, *9*, 101-124.
88. Kieft, J. S.; Tinoco, I., Jr., Solution Structure of a Metal-Binding Site in the Major Groove of Rna Complexed with Cobalt (Iii) Hexammine. *Structure* **1997**, *5*, 713-721.

89. Colmenarejo, G.; Tinoco, I., Jr., Structure and Thermodynamics of Metal Binding in the P5 Helix of a Group I Intron Ribozyme. *J. Mol. Biol.* **1999**, *290*, 119-135.
90. Basu, S.; Rambo, R. P.; Strauss-Soukup, J.; Cate, J. H.; Ferre-D'Amare, A. R.; Strobel, S. A.; Doudna, J. A., A Specific Monovalent Metal Ion Integral to the Aa Platform of the RNA Tetraloop Receptor. *Nat. Struct. Biol.* **1998**, *5*, 986-992.
91. Basu, S.; Strobel, S. A., Biochemical Detection of Monovalent Metal Ion Binding Sites within RNA. *Methods* **2001**, *23*, 264-275.
92. Wimberly, B. T.; Guymon, R.; McCutcheon, J. P.; White, S. W.; Ramakrishnan, V., A Detailed View of a Ribosomal Active Site: The Structure of the L11-RNA Complex. *Cell* **1999**, *97*, 491-502.
93. Wang, Y. X.; Lu, M.; Draper, D. E., Specific Ammonium Ion Requirement for Functional Ribosomal Rna Tertiary Structure. *Biochemistry (Mosc)*. **1993**, *32*, 12279-12282.
94. Shiman, R.; Draper, D. E., Stabilization of RNA Tertiary Structure by Monovalent Cations. *J. Mol. Biol.* **2000**, *302*, 79-91.
95. Bukhman, Y. V.; Draper, D. E., Affinities and Selectivities of Divalent Cation Binding Sites within an Rna Tertiary Structure. *J. Mol. Biol.* **1997**, *273*, 1020-1031.
96. Conn, G. L.; Gittis, A. G.; Lattman, E. E.; Misra, V. K.; Draper, D. E., A Compact Rna Tertiary Structure Contains a Buried Backbone-K<sup>+</sup> Complex. *J. Mol. Biol.* **2002**, *318*, 963-973.
97. Conn, G. L.; Draper, D. E.; Lattman, E. E.; Gittis, A. G., Crystal Structure of a Conserved Ribosomal Protein-Rna Complex. *Science* **1999**, *284*, 1171-1174.
98. Misra, V. K.; Draper, D. E., A Thermodynamic Framework for Mg<sup>2+</sup> Binding to Rna. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 12456-12461.
99. Leipply, D.; Draper, D. E., Evidence for a Thermodynamically Distinct Mg<sup>2+</sup> Ion Associated with Formation of an RNA Tertiary Structure. *J. Am. Chem. Soc.* **2011**, *133*, 13397-13405.
100. Qiu, H.; Kaluarachchi, K.; Du, Z.; Hoffman, D. W.; Giedroc, D. P., Thermodynamics of Folding of the RNA Pseudoknot of the T4 Gene 32 Autoregulatory Messenger Rna. *Biochemistry (Mosc)*. **1996**, *35*, 4176-4186.
101. Nixon, P. L.; Giedroc, D. P., Equilibrium Unfolding (Folding) Pathway of a Model

H-Type Pseudoknotted RNA: The Role of Magnesium Ions in Stability. *Biochemistry (Mosc)*. **1998**, *37*, 16116-16129.

102. Viereg, J.; Cheng, W.; Bustamante, C.; Tinoco, I., Measurement of the Effect of Monovalent Cations on Rna Hairpin Stability. *J. Am. Chem. Soc.* **2007**, *129*, 14966-14973.

103. Rialdi, G.; Levy, J.; Biltonen, R., Thermodynamic Studies of Transfer Ribonucleic Acids. I. Magnesium Binding to Yeast Phenylalanine Transfer Ribonucleic Acid. *Biochemistry (Mosc)*. **1972**, *11*, 2472-2479.

104. Romer, R.; Hach, R., Trna Conformation and Magnesium Binding. A Study of a Yeast Phenylalanine-Specific Trna by a Fluorescent Indicator and Differential Melting Curves. *Eur. J. Biochem.* **1975**, *55*, 271-284.

105. Stein, A.; Crothers, D. M., Equilibrium Binding of Magnesium(Ii) by Escherichia Coli Trnafmet. *Biochemistry (Mosc)*. **1976**, *15*, 157-160.

106. Case, D.; Darden, T.; Cheatham, T. E., III; Simmerling, C.; Wang, J.; Duke, R.; Luo, R.; Walker, R.; Zhang, W.; Merz, K., Amber 12. *University of California, San Francisco* **2012**.

107. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Pairwise Solute Descreening of Solute Charges from a Dielectric Medium. *Chem. Phys. Lett.* **1995**, *246*, 122-129.

108. Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G., Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *J. Phys. Chem.* **1996**, *100*, 19824-19839.

109. Darden, T.; York, D.; Pedersen, L., Particle Mesh Ewald: An N· Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

110. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C., Routine Microsecond Molecular Dynamics Simulations with Amber on Gpus. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878-3888.

111. Loncharich, R. J.; Brooks, B. R.; Pastor, R. W., Langevin Dynamics of Peptides: The Frictional Dependence of Isomerization Rates of N-Acetylalanyl-N'-Methylamide. *Biopolymers* **1992**, *32*, 523-535.

112. Roux, B., The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Commun.* **1995**, *91*, 275-282.

113. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *J. Comput. Chem.* **1995**, *16*, 1339-1350.
114. Lu, M.; Draper, D. E., Bases Defining an Ammonium and Magnesium Ion-Dependent Tertiary Structure within the Large Subunit Ribosomal Rna. *J. Mol. Biol.* **1994**, *244*, 572-585.
115. Allen, F. H., The Cambridge Structural Database: A Quarter of a Million Crystal Structures and Rising. *Acta Crystallogr. B* **2002**, *58*, 380-388.
116. Zheng, H.; Shabalin, I. G.; Handing, K. B.; Bujnicki, J. M.; Minor, W., Magnesium-Binding Architectures in Rna Crystal Structures: Validation, Binding Preferences, Classification and Motif Detection. *Nucleic Acids Res* **2015**, *43*, 3789-3801.
117. Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E., Ucsf Chimera--a Visualization System for Exploratory Research and Analysis. *J. Comput. Chem.* **2004**, *25*, 1605-1612.
118. Sergiev, P. V.; Bogdanov, A. A.; Dontsova, O. A., How Can Elongation Factors Ef-G and Ef-Tu Discriminate the Functional State of the Ribosome Using the Same Binding Site? *FEBS Lett.* **2005**, *579*, 5439-5442.
119. Bergonzo, C.; Hall, K. B.; Cheatham, T. E., III, Stem-Loop V of Varkud Satellite Rna Exhibits Characteristics of the Mg(2+) Bound Structure in the Presence of Monovalent Ions. *J. Phys. Chem. B* **2015**, *119*, 12355-12364.
120. Bergonzo, C.; Hall, K. B.; Cheatham, T. E., III, Divalent Ion Dependent Conformational Changes in an Rna Stem-Loop Observed by Molecular Dynamics. *J. Chem. Theory Comput.* **2016**, *12*, 3382-3389.
121. Allnér, O.; Nilsson, L.; Villa, A., Loop-Loop Interaction in an Adenine-Sensing Riboswitch: A Molecular Dynamics Study. *RNA* **2013**, *19*, 916-926.
122. Panteva, M. T.; Giambaşu, G. M.; York, D. M., Force Field for Mg<sup>2+</sup>, Mn<sup>2+</sup>, Zn<sup>2+</sup>, and Cd<sup>2+</sup> Ions That Have Balanced Interactions with Nucleic Acids. *J. Phys. Chem. B* **2015**, *119*, 15460-15470.
123. Cheatham, T. E., III; Cieplak, P.; Kollman, P. A., A Modified Version of the Cornell Et Al. Force Field with Improved Sugar Pucker Phases and Helical Repeat. *J. Biomol. Struct. Dyn.* **1999**, *16*, 845-862.
124. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A., The

Weighted Histogram Analysis Method for Free-Energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **1992**, *13*, 1011-1021.

125. Maeder, C.; Conn, G. L.; Draper, D. E., Optimization of a Ribosomal Structural Domain by Natural Selection. *Biochemistry (Mosc)*. **2006**, *45*, 6635-6643.

126. Rau, M. J.; Welty, R.; Tom Stump, W.; Hall, K. B., Formation of Tertiary Interactions During Rna Gtpase Center Folding. *J. Mol. Biol.* **2015**, *427*, 2799-2815.

127. Pol, S.; Vallet-Pichard, A.; Corouge, M.; Mallet, V. O., Hepatitis C: Epidemiology, Diagnosis, Natural History and Therapy. *Contrib. Nephrol.* **2012**, *176*, 1-9.

128. Mohd Hanafiah, K.; Groeger, J.; Flaxman, A. D.; Wiersma, S. T., Global Epidemiology of Hepatitis C Virus Infection: New Estimates of Age-Specific Antibody to Hcv Seroprevalence. *Hepatology* **2013**, *57*, 1333-1342.

129. Rosen, H. R., Clinical Practice. Chronic Hepatitis C Infection. *N. Engl. J. Med.* **2011**, *364*, 2429-2438.

130. Seth, P. P.; Miyaji, A.; Jefferson, E. A.; Sannes-Lowery, K. A.; Osgood, S. A.; Propp, S. S.; Ranken, R.; Massire, C.; Sampath, R.; Ecker, D. J.; Swayze, E. E.; Griffey, R. H., Sar by Ms: Discovery of a New Class of Rna-Binding Small Molecules for the Hepatitis C Virus: Internal Ribosome Entry Site Iia Subdomain. *J. Med. Chem.* **2005**, *48*, 7099-7102.

131. Kieft, J. S.; Zhou, K.; Jubin, R.; Doudna, J. A., Mechanism of Ribosome Recruitment by Hepatitis C Ires Rna. *RNA* **2001**, *7*, 194-206.

132. Dibrov, S. M.; Johnston-Cox, H.; Weng, Y.-H.; Hermann, T., Functional Architecture of Hcv Ires Domain Ii Stabilized by Divalent Metal Ions in the Crystal and in Solution. *Angew. Chem. Int. Ed.* **2007**, *46*, 226-229.

133. Henriksen, N. M.; Hayatshahi, H. S.; Davis, D. R.; Cheatham, T. E., Structural and Energetic Analysis of 2-Aminobenzimidazole Inhibitors in Complex with the Hepatitis C Virus Ires Rna Using Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2014**, *54*, 1758-1772.

134. Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A., Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157-1174.

135. Wu, X.; Brooks, B. R., Self-Guided Langevin Dynamics Simulation Method. *Chem. Phys. Lett.* **2003**, *381*, 512-518.

136. Dupradeau, F.-Y.; Pigache, A.; Zaffran, T.; Savineau, C.; Lelong, R.; Grivel, N.; Lelong, D.; Rosanski, W.; Cieplak, P., The R.E.D. Tools: Advances in Resp and Esp Charge Derivation and Force Field Library Building. *Phys. Chem. Chem. Phys.* **2010**, *12*, 7821-7839.
137. Dibrov, S. M.; Ding, K.; Brunn, N. D.; Parker, M. A.; Bergdahl, B. M.; Wyles, D. L.; Hermann, T., Structure of a Hepatitis C Virus Rna Domain in Complex with a Translation Inhibitor Reveals a Binding Mode Reminiscent of Riboswitches. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 5223-5228.
138. Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D., Dock 6: Combining Techniques to Model RNA–Small Molecule Complexes. *RNA* **2009**, *15*, 1219-1230.
139. Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A., Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Model.* **2006**, *25*, 247-260.
140. Cheatham, T. E., III; Kollman, P. A., Insight into the Stabilization of a-DNA by Specific Ion Association: Spontaneous B-DNA to a-DNA Transitions Observed in Molecular Dynamics Simulations of D[Acccgcggt]2 in the Presence of Hexaamminecobalt(III). *Structure* **1997**, *5*, 1297-1311.
141. Laing, L. G.; Gluick, T. C.; Draper, D. E., Stabilization of Rna Structure by Mg Ions. Specific and Non-Specific Effects. *J. Mol. Biol.* **1994**, *237*, 577-587.
142. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A., A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
143. Lee, D.; Walsh, J. D.; Yu, P.; Markus, M. A.; Choli-Papadopoulou, T.; Schwieters, C. D.; Krueger, S.; Draper, D. E.; Wang, Y. X., The Structure of Free L11 and Functional Dynamics of L11 in Free, L11-Rrna(58 Nt) Binary and L11-Rrna(58 Nt)-Thiostrepton Ternary Complexes. *J. Mol. Biol.* **2007**, *367*, 1007-1022.
144. Michel, J.; Harker, E. A.; Tirado-Rives, J.; Jorgensen, W. L.; Schepartz, A., In Silico Improvement of Beta3-Peptide Inhibitors of P53 X Hdm2 and P53 X Hdmx. *J. Am. Chem. Soc.* **2009**, *131*, 6356-6357.