

Mg²⁺ Sensing by an RNA Fragment: Role of Mg²⁺ Coordinated Water Molecules

Antarip Halder,[†] Sunil Kumar,[†] Omar Valsson,[‡] and Govardhan Reddy^{*,†}

[†]*Solid State and Structural Chemistry Unit, Indian Institute of Science, Bengaluru 560012,
Karnataka, India*

[‡]*Max Planck Institute for Polymer Research, Ackermannweg 10, D-55128 Mainz, Germany*

E-mail: greddy@iisc.ac.in.

Phone: +91-80-22933533. Fax: +91-80-23601310

Abstract

RNA molecules selectively bind to specific metal ions to populate their functional active states making it important to understand their source of ion selectivity. In large RNA systems, metal ions interact with the RNA at multiple locations making it difficult to decipher the precise role of ions in folding. To overcome this complexity, we studied the role of different metal ions (Mg^{2+} , Ca^{2+} and K^+) in the folding of a small RNA hairpin motif (5'-ucCAAaga-3') using unbiased all-atom molecular dynamics simulations. The advantage in studying this small system is that it requires specific binding of a single metal ion to fold to its native state. We find that even for this small RNA, the folding free energy surface (FES) is multidimensional as different metal ions present in the solution can simultaneously facilitate folding. The FES shows that specific binding of a metal ion is indispensable for its folding. We further show that in addition to the negatively charged phosphate groups, spatial organization of electronegative nucleobase atoms drive the site specific binding of the metal ion. Even though the binding site cannot discriminate between different metal ions, RNA folds efficiently only in Mg^{2+} solution. We show that the rigid network of Mg^{2+} coordinated water molecules facilitate the formation of important interactions in the transition state. The other metal ions such as K^+ and Ca^{2+} cannot facilitate the formation of such interactions. These results allow us to hypothesize possible metal sensing mechanisms in large metallo-riboswitches and they also provide useful insights for the design of appropriate collective variables for studying large RNA molecules using enhanced sampling methods.

Introduction

Ribonucleic acid (RNA) molecules fold into competent structures to participate in a plethora of cellular processes such as protein synthesis,¹ catalysis,² regulation of gene expression,^{3,4} gene editing,^{5,6} etc. Metal ions are inextricably linked to the assembly of functional RNA structures.⁷⁻⁹ Metal ions renormalize the negative charge on the phosphate groups and decrease the persistence length of the RNA chain enabling the chain to collapse in the initial

stages of folding.¹⁰⁻¹² Metal ions are also found to affect the rate of conformational rearrangements,¹³⁻¹⁵ modify the kinetic free energy barriers,¹⁶ participate in the formation of transition states,¹⁷ alter the dominant folding pathways¹⁸⁻²⁰ and are shown to be intricately involved in the catalytic mechanism of RNAs such as ribozymes.²¹⁻²³

Mg²⁺ is the most ubiquitous metal ion, which is bound to the RNA in crystal structures.²⁴ Mg²⁺ in RNA folding are classified into two types: diffuse and site-binding, depending on its contribution to the stabilization of the RNA structure.²⁵ The diffusely bound Mg²⁺ binds to the polyanionic RNA chain nonspecifically and result in the overall collapse of the chain dimensions. Whereas, the site-bound Mg²⁺ is attracted to specially arranged binding pockets and aid in the stabilization of secondary and tertiary structures. Although recent simulations²⁶ on *Azoarcus* ribozyme show that Mg²⁺ condensation on the RNA chain is site specific and depends on the structure of the folded state.

There is an extensive effort to understand how metal ions influence RNA folding using various biophysical methods.²⁷⁻³¹ Due to the fluctuating nature of the ion atmosphere around the RNA, it is extremely difficult to study its influence on RNA folding using traditional techniques, which probe structure such as X-ray crystallography, NMR, or (cryo-) electron microscopy. These techniques only identify ions, which specifically bind to a group of RNA atoms as a part of their coordination shell and stabilize the RNA structure.³² Theoretical and computational techniques are playing an important role in complementing experiments to identify the role of ions in various other aspects of RNA folding.³³

Molecular dynamics (MD) simulations using atomistic³⁴⁻³⁸ and coarse-grained³⁹⁻⁴⁶ force fields have provided important insights into RNA folding. Nevertheless, a comprehensive understanding of how large RNAs selectively recognize specific metal ions and fold to their respective active conformations is far from complete. For example, metal ion sensing riboswitches^{47,48} recognize a specific type of metal ion (Mg²⁺,⁴⁹ Mn²⁺,^{50,51} Co²⁺,⁵² Ni²⁺,⁵² etc.) and depending on the metal ion concentration, they switch from unfolded to folded conformations to regulate the downstream gene expression. However, the source of selectiv-

ity towards specific ions in this type of large RNA systems (> 90 nt) is not understood at the molecular level. In large RNA systems, metal ions interact with the RNA at multiple locations making it difficult to decipher the precise role of site-bound ions in folding. To overcome this complexity and probe the source of metal ion specificity in RNA, we have identified a simple yet robust RNA tetraloop system that requires a single site bound metal ion for its folding and studied its unfolding/refolding mechanism using atomistic MD simulations.

RNA tetraloops are small three-dimensional hairpin motifs that are found ubiquitously in almost all functional RNAs and are associated with important structural and functional roles.⁵³ A tetraloop hairpin motif is composed of an A-form helical stem capped by four nucleotides arranged in a specific three-dimensional structure (Figure 1). Due to their small sizes and availability of ample experimental data, the tetraloops are considered to be ideal testbed for atomistic MD simulation studies on RNA.³³ There are only a limited number of nucleotide sequences that form these tetraloops and a majority of the known tetraloops either belong to GNRA or UNCG tetraloop category, where N is any residue and R is any purine residue (A/G).⁵⁴ Folded structures of RNA hairpin motifs with GNRA (or UNCG) tetraloop sequences are conserved and are thermally highly stable as the bases in the loops display ordered stacking interactions in addition to specific non-Watson-Crick base-pairs and backbone conformations.⁵⁵

In this work, we studied the folding of a special RNA tetraloop (L4 loop in M-box riboswitch⁴⁹), which despite not having the required sequence, folds into the shape of a GNRA tetraloop (Figure 1A). The L4 tetraloop is part of the magnesium sensing M-box riboswitch that switches conformations with the change in Mg^{2+} concentration and regulates downstream gene expression via termination of the transcription process. Crystal structure⁴⁹ of magnesium bound folded aptamer domain (PDB ID: 2QBZ) shows that a Mg^{2+} (Mg_{204}) is present inside the loop. This Mg^{2+} holds together the loop-closing bases (C69 and A72) via its coordinated water molecules (Figure 1B), and facilitates the L4 loop to fold into the shape of a conventional GNRA tetraloop (Figure 1C). This Mg^{2+} mediated inter-nucleotide

interaction compensates for the signature G:A *trans*S:H base pair of a conventional GNRA tetraloop (Figure 1D). Electrostatic potential surface for the folded conformation of a L4 tetraloop hairpin motif shows the presence of an anionic cavity at its center (Figure 1E). This is absent in the conventional GNRA tetraloop hairpin motif (Figure 1F). The anionic cavity facilitates cation binding enhancing the stability of the folded conformation, but the source of specificity towards Mg^{2+} remains ambiguous. The simple L4 tetraloop system with a single site for metal cation binding is an ideal system to probe the source of specificity towards selective metal ions in RNA fragments and also study the mechanism of tetraloop folding.

We have carried out all-atom MD simulations of L4 tetraloop hairpin motif in different metal ion concentrations and show that the binding of a metal ion is indispensable for its folding. The site specific binding of the metal ion is guided by the spatial organization of electronegative nucleobase atoms. However, the binding site cannot discriminate between different metal ions. We find that in a mixture of monovalent and divalent ions, the corresponding free energy surface (FES) becomes multi-dimensional as both the ions contribute towards folding. We designed appropriate collective variables (CVs) to compute the FES, which describe the complex folding mechanism of the L4 tetraloop. The FESs show that the L4 tetraloop folds efficiently only in Mg^{2+} solution. Further, the structures in the transition state ensemble (TSE) reveal that a hydrated Mg^{2+} is involved in the transition state (TS) where it plays a critical role in stabilizing the loop part of the motif. The subsequent formation of the A-form helix also requires crucial contribution from the Mg^{2+} coordinated water molecules. We argue that in the absence of rigid octahedral geometry of the water molecules in the first coordination shell of the metal ion, RNA cannot efficiently fold to the native state.

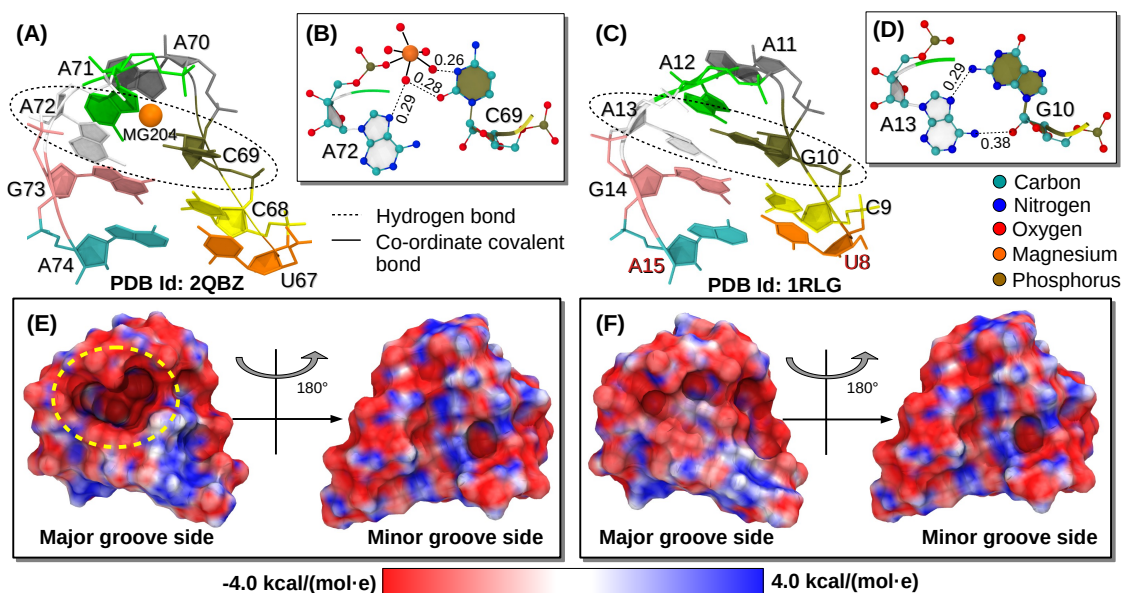


Figure 1: (A) Crystal structure⁴⁹ of the hairpin motif containing an A-form helix capped by the L4 tetraloop. The Mg²⁺ present inside the loop is shown in orange color. (B) The Mg²⁺ has water mediated interaction with C69 and A72 nucleobases. Hydrogen bonds are shown as broken lines in black. The interatomic distances between the hydrogen bond donor and acceptor atoms are given in nm. (C) Crystal structure of a GNRA tetraloop hairpin motif containing the same A-form helix. The terminal residues of the motif are manually edited (C8 → U8 & G15 → A15) to resemble the former hairpin motif. (D) Closing residues of the tetraloop form a G10:A13 *trans*S:H pair. Electrostatic potential surfaces of the hairpin motif containing the (E) L4 tetraloop and (F) GNRA tetraloop, respectively with views from the major and minor groove side. An anionic cavity present at the center of the L4 tetraloop is visible from the major groove side (encircled with a yellow broken line). No such cation binding cavity is present in the GNRA tetraloop. The electrostatic potential surfaces are calculated by solving the Poisson-Boltzmann equation⁵⁶ using the PBEQ Solver server⁵⁷ present in CHARMM-GUI.⁵⁸ The parameters used for the calculation are given in the SI.

Materials and Methods

Enhanced sampling methods have been used to study the folding of GNRA and UNGC loops along with other RNA hairpin motifs.^{34,35,59,60} These simulations showed that due to the drawbacks in the forcefields, the native conformations of tetraloops obtained from the crystal structures and NMR, do not correspond to the global minimum in the FES. MD simulations using replica-exchange technique reported folding of UNGC and GNRA tetraloops to their respective native structures.^{59,61,62} However, they are only accounted for a small fraction of the conformations populated ($\sim 10\%$ in Chen-García and $< 1\%$ in AMBER ff14 forcefields, respectively) in the ensemble with the lowest-temperature, $T = 277$ K.

Recent modifications to the AMBER ff14 force field made by Tan and co-workers alleviated some of the problems in the forcefields to study tetraloop folding.⁶³ In the simulations using this modified forcefield,⁶³ the population of the native tetraloop conformations increased to 40% in the ensemble with the lowest-temperature $T = 280$ K. Two other recent versions of AMBER force field promise further improvements in sampling the folded state of RNA tetraloops.^{64,65} These improvements in the forcefields will facilitate the study of the role of metal ions in the tetraloop hairpin motif formation using atomistic MD simulations.

MD simulations: In this work we have used GROMACS simulation program (version 5.1.2)^{66,67} and a modified AMBER ff14 force field⁶³ along with TIP4P-D water model⁶⁸ to study the role of different metal ions on RNA tetraloop folding. It is challenging for a force field to capture all the important physical properties of RNA.³³ The modified AMBER ff14 force field was shown to sample the RNA tetraloop conformations reasonably well⁶³ despite its drawbacks.⁶⁴ The folded structure of the eight nucleotide long L4 tetraloop hairpin motif (U67 to A74) is taken from the X-ray crystal structure of M-box riboswitch (PDB ID: 2QBZ⁴⁹). The sequence of the hairpin motif is U⁶⁷CCAAAGA⁷⁴ and the phosphate group at the 5'-end is removed. The hairpin is placed at the center of a cubic box of dimensions ≈ 6.3 nm, and is solvated with TIP4P-D⁶⁸ water molecules. Periodic boundary conditions

are applied in all the directions. The negative charges present on the RNA backbone are neutralized by adding seven K^+ ions. This system corresponds to $[K^+] \approx 45$ mM and $[Mg^{2+}] = 0$ mM. To study the role of Mg^{2+} and Ca^{2+} on the tetraloop folding, we prepared two additional systems, where the concentration of Mg^{2+} and Ca^{2+} is increased to 65 mM, respectively. To these systems appropriate number of Cl^- ions are added to maintain charge neutrality. Further details of these three systems are given in Table 1. The simulation box is initially subjected to energy minimization using steepest descent minimization algorithm with an energy cut-off of 1000 kJ/mol/nm.

The energy minimized systems are equilibrated in three steps: (i) position restraints are applied on the RNA atoms and ions using a harmonic potential with a force constant of 1000 kJ/(mol nm²) and only the solvent molecules are equilibrated for 2 ns in the NVT ensemble, (ii) position restraints are applied on the RNA atoms and ions and only the solvent molecules are equilibrated for 4 ns in the NPT ensemble and (iii) position restraints are applied only on the RNA atoms and both the solvent molecules and ions are equilibrated for 4 ns in the NPT ensemble. Finally, the equilibrated systems are subjected to production runs. The temperature T is maintained at 300 K using the stochastic velocity rescaling thermostat developed by Bussi *et al.*⁶⁹ (with time constant $\tau_t = 0.1$ ps) and the pressure P is maintained at 1 atm using Parrinello-Rahman barostat⁷⁰ (with time constant $\tau_p = 2.0$ ps). During the production run, all the bond lengths involving the hydrogen atoms are constrained using the LINCS algorithm.⁷¹ The simulation is advanced using the leap-frog integrator with a 2 fs time step. The long-range electrostatic interactions are computed using particle mesh Ewald summation (PME) method⁷² using 1 nm cutoff for both electrostatic and van der Waals interactions. For each system, we ran two independent trajectories each of length ≈ 10 μ s long, and the cumulative length of all the trajectories is ≈ 60 μ s (Figure S1). We saved the simulation trajectory every 5 ps for analysis. The first 500 ns of data from each trajectory are ignored and the rest of the data are used for analysis.

Table 1: Description of the three L4 tetraloop systems studied

Metal ion concentration	RNA sequence	Number of particles				
		Water	Mg	Ca	K	Cl
$[\text{Mg}^{2+}] = 0 \text{ mM}$	ucCAAAGA	8400	0	0	7	0
$[\text{Mg}^{2+}] = 65 \text{ mM}$	ucCAAAGA	8370	10	0	7	20
$[\text{Ca}^{2+}] = 65 \text{ mM}$	ucCAAAGA	8370	0	10	7	20

Computing the FES: The FES corresponding to the tetraloop folding is calculated using $G(\mathbf{s}) = -k_B T \ln(P(\mathbf{s}))$, where $P(\mathbf{s})$ is the probability distribution of the set of CVs \mathbf{s} , and k_B is the Boltzmann constant. We estimated $P(\mathbf{s})$ by histogramming the values of the CVs \mathbf{s} computed from the simulation data. The FESs are plotted using matplotlib tool.⁷³ To probe the contribution of metal ions on folding of the hairpin motif, the free energy is projected onto various CVs viz. eRMSD, end-to-end distance (R_{ee}), coordination number between metal ions and selected atoms of the RNA fragment (CN_K , CN_{Mg} & CN_{Ca}).

eRMSD is a metric for calculating the deviation between 2 RNA structures.⁷⁴ Unlike the conventional root mean square deviation (RMSD), which use the positions of all nucleic acid atoms to compute the deviation between 2 structures, eRMSD considers only the relative positions and orientations of nucleobases. eRMSD is shown to be more effective and less ambiguous descriptor compared to RMSD.^{34,75,76} The BARNABA software⁷⁷ is used to calculate the eRMSD values of the RNA conformations with respect to the structure obtained after the energy minimization of the crystal structure. R_{ee} is the distance between the C4' atoms of the terminal residues U67 and A74, and is computed using the post processing tools provided by GROMACS. The coordination number between different metal ions and a set of RNA atoms from the nucleobase and sugar for the four key nucleotides, C68, C69, A72 and G73 is calculated using the PLUMED library^{78,79} (version 2.5.2). Details of the parameters used are provided in the supplementary information (SI).

Spatial distribution of metal ions around RNA: To quantify the condensation of positively charged metal ions onto the negatively charged RNA, we calculated the local ion concentration (c^*) around the different polar atoms present in the RNA (O atoms in

phosphate, O atoms in sugar and, O and N atoms in nucleobases). Following earlier work,²⁶ c^* in molar units is defined as,

$$c^* = \frac{1}{N_A V_c} \int_0^{r_c} \rho(r) 4\pi r^2 dr, \quad (1)$$

where $\rho(r)$ is the number density of a specific type of ions at the distance r from a specific RNA atom, V_c is the spherical volume of radius r_c , and N_A is the Avogadro's number. The cutoff distance, r_c is set to 0.73 nm, which is the Bjerrum length of water at $T = 300$ K.

To study the distribution of different metal ions around the RNA residues in different basins of FES, we computed the spatial density maps using the volmap plugin of the VMD software.⁸⁰ Initially, all the frames belonging to a particular basin are aligned to a representative RNA structure of the basin, which is randomly selected. The alignment is performed using the Kabsch method⁸¹ implemented in VMD. To construct the spatial density maps, the metal ions are treated as spheres of radii equal to their respective van der Waals radii and a grid of size 0.01 nm^3 is constructed to divide the space around the RNA residues. A grid point is considered to be occupied and assigned a value equal to 1, if the sphere corresponding to a specific metal ion lies on the grid. Otherwise it takes a value equal to 0. The value at a grid point is averaged over all frames that belong to a specific basin. RNA residues, around which the occupancy of metal ions are found to be the highest in the spatial density maps, are further probed using the pair distribution functions, $g(r)$. The $g(r)$ between selected RNA residues and specific metal ions are calculated using VMD software with a distance cutoff of 2 nm and a grid spacing of 0.01 nm. Packages in VMD are also used for structural superposition⁸² and rendering of images.⁸³

Computing the TSE: To identify the TS structures that connect the folded and unfolded states, we have initially identified fragments of the simulation trajectory that either (a) leave the unfolded basin and enter the folded basin without returning to the unfolded basin or visiting any other basin, or (b) leave the folded basin and enter the unfolded basin without

returning to the folded basin or visiting any other basin. Each such transition path, TP , detected in our trajectories (shown in the SI) is labeled as ‘reactive’ (Figure S2,S3). We have used the theory proposed by Hummer^{84,85} to identify the TSs in tetraloop folding. According to this theory, TSs are those points in the configuration space that have the highest probability of equilibrium trajectories passing through them are reactive.⁸⁴ Given that the system is in configuration r , the probability of being on a TP is given by

$$p(TP|r) = \frac{p(r|TP)p(TP)}{p_{eq}(r)}, \quad (2)$$

where $p(TP)$ is the probability of a frame of the equilibrium trajectory to be on a TP, $p_{eq}(r)$ is the probability of a frame of the equilibrium trajectory to have the configuration r , and $p(r|TP)$ is the probability of a frame of the equilibrium trajectory to have the configuration r , given that it already belongs to a TP. The configuration r is defined using the two CVs eRMSD and R_{ee} . There are 46 and 97 tetraloop conformations in the simulation trajectories corresponding to $[Mg^{2+}] = 0$ mM and 65 mM, respectively that have the maximum $p(TP|r)$ value and are putative TS structures.

To verify that these conformations are TS structures, we further performed P_{fold} analysis.⁸⁶ In the P_{fold} analysis, with each putative structure as an initial starting conformation (RNA + ions + water), we spawned 100 simulation trajectories each with a different set of velocities for a time period of 1.6 ns to probe whether they land in the folded or unfolded basins with approximately equal probabilities. For each starting structure, P_{fold} is defined as the fraction of these trajectories that land in the folded state. We labeled a putative structure as a TS, if the P_{fold} computed from 100 independent simulation runs satisfies $0.4 \leq P_{fold} \leq 0.6$. Representative trajectories are shown in the SI (Figure S4C,D). Using this procedure, we have identified four conformations as the TS structures for each of the $[Mg^{2+}] = 0$ mM and 65 mM systems.

Results and Discussion

Folded Structure in the Simulations Deviates from the Crystal Structure: Two independent simulations of the L4 tetraloop hairpin motif performed at $T = 300$ K and $[\text{Mg}^{2+}] = 0$ mM for a cumulative time of $\approx 20 \mu\text{s}$ long show multiple transitions between the folded and unfolded conformations (Figure S1A,B). A total of 8 folding events are observed in both the trajectories. Analyses of the folding events using CN_K reveal active involvement of K^+ in L4 tetraloop folding. The folded conformations are short lived with an average lifetime of ≈ 2.5 ns. We also performed two independent simulations of the L4 tetraloop at $T = 300$ K and $[\text{Mg}^{2+}] = 65$ mM for a cumulative time of $\approx 20 \mu\text{s}$. We observed 10 folding events in these trajectories (Figure S1C,D). In the presence of Mg^{2+} , the average life time of the folded state increased to ≈ 8.5 ns indicating that Mg^{2+} stabilize the folded state compared to K^+ .

There are structural differences between the folded conformation of the L4 tetraloop obtained from the MD simulations and crystal structure (Figure S5). In the crystal structure,⁴⁹ the geometry of the L4 tetraloop at the termini is influenced by long range interactions, which are absent in the isolated system modeled in MD simulations (Figure S5A,B). For example, the helix of the L4 hairpin is coaxially stacked with the P4 helix (Figure S5C,D), and also A71 and A72 residues from the loop of the L4 hairpin are involved in hydrogen bonding interactions with the P5 stem and L5 loop, respectively (Figure S5E). Although the deviation between the crystal structure and the folded structure in the MD simulations is significant at the termini and at the orientation of the C69 residue (Figure S5A,B), the consecutive stacking interactions between the bases of A70, A71 and A72 remain relatively unperturbed (Figure S5F).

Site Specific Binding of a Metal Ion is Indispensable for the Folding of L4 Tetraloop: We constructed the FES of L4 tetraloop folding to probe the role of metal ions in the folding mechanism (Figure 2). We projected the FES onto the following CVs:

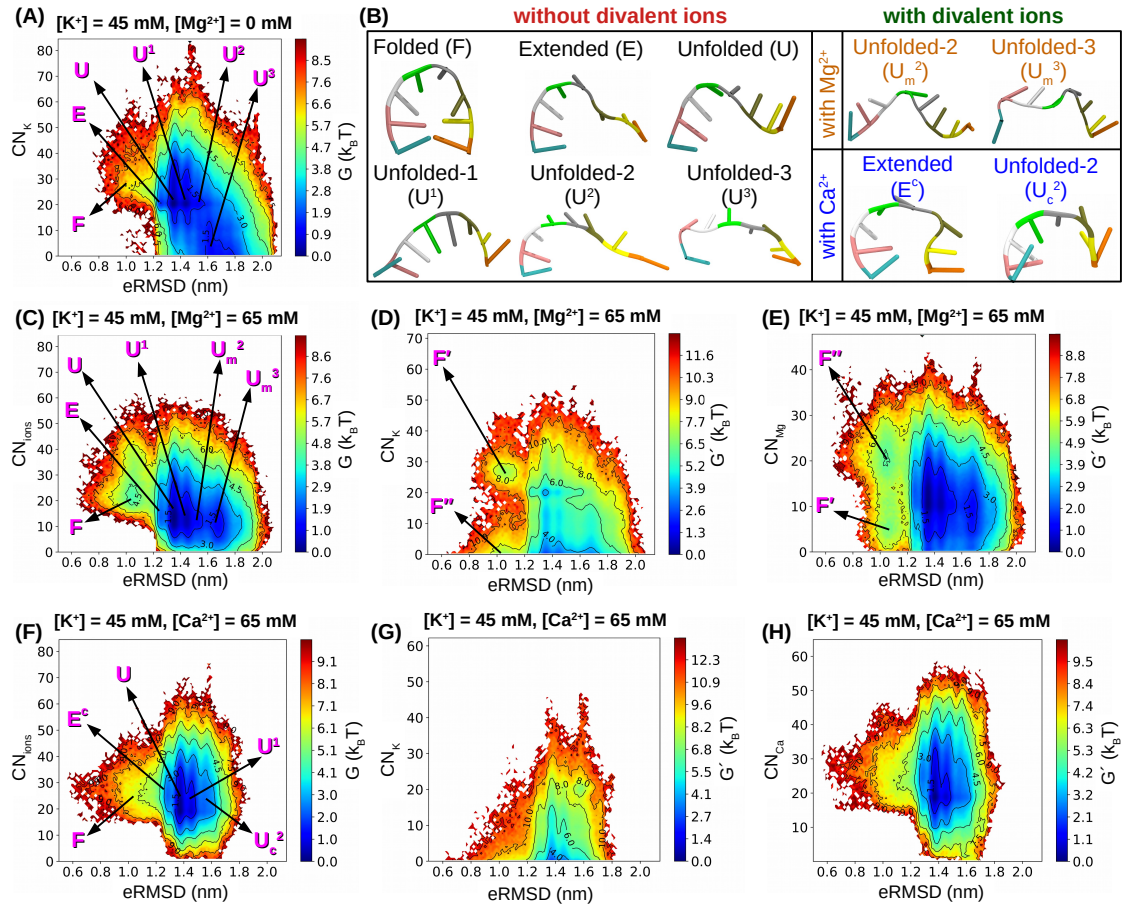


Figure 2: (A) FES of the L4 tetraloop projected onto CVs, eRMSD and CN_K at $[Mg^{2+}] = 0$ mM. Important basins in the FES are highlighted. (B) Representative structure of the RNA from different basins are shown in cartoon representation. (C) FES projected onto eRMSD and CN_{ions} ($= CN_K + CN_{Mg}$) at $[Mg^{2+}] = 65$ mM. In a solution containing K^+ and Mg^{2+} , FES is further projected onto eRMSD, CN_K and CN_{Mg} . (D) FES constructed using 3 CVs is shown in two dimensions, eRMSD and CN_K , after integrating out the third dimension, CN_{Mg} . (E) Similarly, FES is projected onto eRMSD and CN_{Mg} after integrating out CN_K . In these projections, instead of a single folded basin 'F', two folded basins F' and F'' are visible, which are due to the specific binding of K^+ and Mg^{2+} , respectively. (F) FES projected onto eRMSD and CN_{ions} ($= CN_K + CN_{Ca}$) at $[Ca^{2+}] = 65$ mM. In a solution containing K^+ and Ca^{2+} , FES is further projected onto eRMSD, CN_K and CN_{Ca} . (G) FES is shown in 2 dimensions, eRMSD and CN_K , after integrating out CN_{Ca} . (H) Similarly, FES is projected onto eRMSD and CN_{Ca} after integrating out CN_K . In a solution containing K^+ and Ca^{2+} , only Ca^{2+} contributed to L4 tetraloop folding.

eRMSD, CN_K and CN_{Mg} , where CN_K and CN_{Mg} are the coordination numbers of K^+ and Mg^{2+} with respect to the atoms of four nucleosides C68, C69, A72 and G73. Note that C69 and A72 form the loop closing pair, which is followed by the first base pair of the helix composed of C68 and G73 (Figure 1). Therefore, when a metal ion gets trapped inside the RNA loop, its corresponding value of coordination number increases. Low and high values of eRMSD correspond to folded and unfolded conformations, respectively.

The FES for both $[Mg^{2+}] = 0$ mM (projected onto eRMSD and CN_K) and 65 mM (projected onto eRMSD and $CN_{ions} (= CN_K + CN_{Mg})$) solutions shows 3 distinct basins, which correspond to (i) the folded state (basin ‘F’ at eRMSD ≈ 1.04 nm), (ii) a semi-folded state with extended end-to-end distance (basin ‘E’ at eRMSD ≈ 1.25 nm) and (iii) a broad unfolded basin (Figure 2A,C). The unfolded basin can be further classified into 4 basins. RNA conformations that populate the basin ‘U’ centered around eRMSD ≈ 1.35 nm, have all the eight bases stacked consecutively (Figure S5G). The overall backbone orientation resembles that of a single stranded helix, as reported for homopolymeric RNAs.⁸⁷ The other three basins correspond to different unfolded conformations of the the RNA and are centered around eRMSD ≈ 1.44 (basin U^1), ≈ 1.53 (basin U^2) and ≈ 1.67 nm (basin U^3), respectively (Figure 2A,C). In the presence of Mg^{2+} , the basins located at higher values of eRMSD (U^2 & U^3) get populated with RNA conformations that have noticeable variation in their backbone orientation (Figure 2B).

The FES shows that both Mg^{2+} and K^+ ions contribute to folding. When $[Mg^{2+}] = 0$ mM, the folded basin (F) at $CN_K \approx 27.5$ and eRMSD ≈ 1.04 nm is narrow and shallow (Figure 2A). The absence of a basin close to $CN_K \approx 0$ and low eRMSD values suggests that binding of K^+ within the region surrounded by C68-C69-A72-G73 residues is indispensable for the folding of L4 tetraloop. In the presence of Mg^{2+} , the FES projected onto eRMSD and $CN_{ions} (= CN_K + CN_{Mg})$ shows a broad and deep folded basin with $CN_{ions} \gtrsim 21$ and eRMSD ≈ 1.04 nm (Figure 2C). Even in this case, the absence of a basin close to $CN_{ions} \approx 0$ and low eRMSD values indicate that the folding of L4 tetraloop is driven by the specific

binding of a metal ion (either Mg^{2+} or K^+) inside the loop. Representative structures of the folded state with the metal ion ($\text{K}^+/\text{Mg}^{2+}$) bound at the center of the loop are illustrated in Figure S6.

Mg^{2+} Stabilizes the Folded Structure Compared to Other Metal Ions: To contrast the L4 tetraloop folding in the presence of Mg^{2+} and K^+ , we compared the FES at $[\text{Mg}^{2+}] = 0$ mM and 65 mM (Figure 2A,C). The folded state becomes more stable when Mg^{2+} is also present in the solution in addition to K^+ (Figure 2C). This is clear from the FES projected onto a single CV, eRMSD (Figure S7). The free energy difference between the folded and unfolded states is estimated³⁴ as $\Delta G_{FU} = -k_B T [\log(\sum_{i,F} P(\mathbf{s}_i)) - \log(\sum_{i,U} P(\mathbf{s}_i))]$, where F and U correspond to the folded and unfolded basins, respectively (Figure 2A,C). ΔG_{FU} for the L4 tetraloop decreases from 14.6 kJ/mol to 10.6 kJ/mol on increasing $[\text{Mg}^{2+}]$ from 0 mM to 65 mM. This value is significantly less than the ΔG_{FU} (≈ 16.1 kJ/mol) of a standard GNRA tetraloop at the same temperature in the presence of only monovalent counterions.^{34,35} However, contribution of K^+ in stabilizing the L4 tetraloop folded state when both K^+ and Mg^{2+} are present in the solution remains ambiguous from the FES shown in Figure 2C.

UV melting experiments⁸⁸ and computational studies⁸⁹ have established that when monovalent ions are also present in the solution along with Mg^{2+} , they compete in stabilizing RNA tertiary structures. At low Mg^{2+} concentration ($[\text{Mg}^{2+}] \approx 0$), RNA's stability primarily depends on the monovalent ion concentration, whereas at higher $[\text{Mg}^{2+}]$ it becomes independent of the concentration of the monovalent ion. To probe the contribution of K^+ and Mg^{2+} to L4 tetraloop folding when both K^+ and Mg^{2+} are present in the solution, we projected the FES onto three CVs, eRMSD, CN_K and CN_{Mg} . The FES shows that even in the mixture of K^+ and Mg^{2+} , both ions contribute to the folding of the tetraloop (Figure 2D,E). The two folded basins, F' and F'' , observed in the FES are due to the specific binding of K^+ and Mg^{2+} , respectively in the center of the hairpin motif formed in the folded state. However, the folded basin F'' is deeper compared to the folded basin F' showing that RNAs fold

efficiently in the presence of Mg^{2+} at the center of the loop compared to the K^+ .

We further studied whether other divalent ions such as Ca^{2+} are as efficient as Mg^{2+} in stabilizing the folded state of L4 tetraloop. We performed two independent simulations of the L4 tetraloop at $T = 300$ K and $[\text{Ca}^{2+}] = 65$ mM for a cumulative time of ≈ 20 μs long (Figure S1E,F). A total of 3 folding events are observed in both the trajectories. The FES projected onto eRMSD and CN_{ions} ($= \text{CN}_K + \text{CN}_{Ca}$) does not show any deep folded basin as observed for 65 mM Mg^{2+} solution (Figure 2F). This shows that the increase in the folded state stability in the presence of Mg^{2+} is not just due to the abundance of cations present in the medium. Also the RNA conformations that populate the semi-folded extended basin labeled ‘E’ (eRMSD ≈ 1.25 nm) is distinctly different from the conformations that populate the same basin in $[\text{Mg}^{2+}] = 0$ mM and $[\text{Mg}^{2+}] = 65$ mM systems (Figure 2B). Absence of a basin close to $\text{CN}_{ions} \approx 0$ and low eRMSD suggests that even in this case, the folding of L4 tetraloop is driven by the specific binding of metal ions (either Ca^{2+} or K^+) at the center of the loop. However, the FES projected onto eRMSD, CN_K and CN_{Ca} has no basin at eRMSD < 1.2 nm for either at a high value of CN_K (Figure 2G) or at $\text{CN}_{Ca} \approx 0$ (Figure 2H). This implies that in the presence of Ca^{2+} , the monovalent K^+ did not contribute towards folding and we explore the reasons below.

Ca^{2+} and Mg^{2+} Condense More Efficiently on the RNA Compared to K^+ : Monovalent metal ions like K^+ can actively participate in stabilizing RNA tertiary structures,⁹⁰ selectively chelate to specific sites on RNA (e.g. AA platform,²⁷ G-quadruplex,⁹¹ etc.) and even modulate the folding pathways of RNA.²⁰ Yet K^+ show limited participation in modulating the folding of L4 tetraloop, especially when the solution contains a mixture of divalent and monovalent ions. This can be rationalized from the difference in the ion condensation patterns around the RNA (Figure 3). It was shown earlier that in a mixture of divalent and monovalent ions, divalent ions are preferred over monovalent ions for condensation around the negatively charged RNA, due to their high charge to volume ratio.²⁶ We also observe that in the absence of divalent ions, the K^+ condense around both backbone (sugar-phosphate)

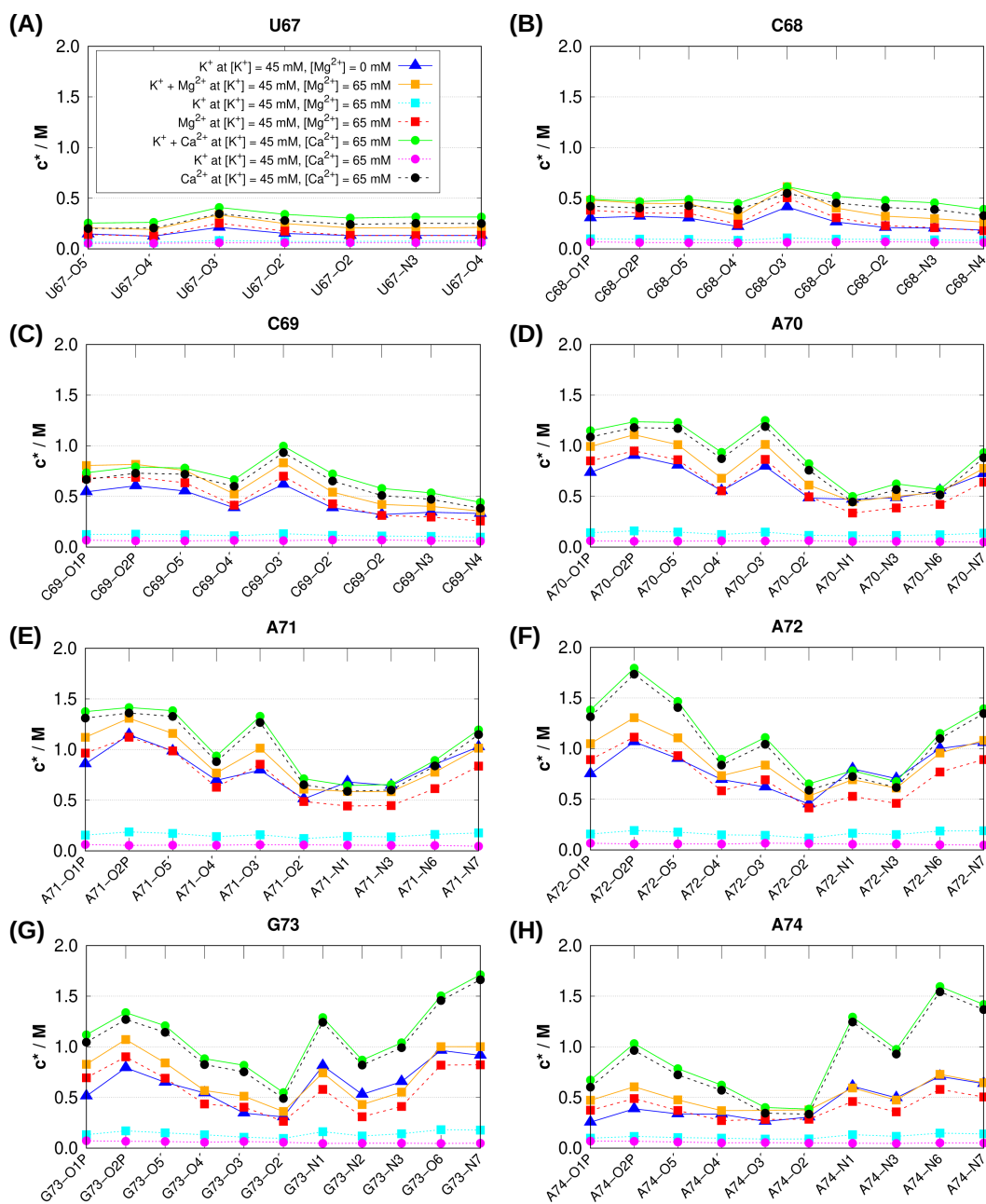


Figure 3: Fingerprint of ion coordination expressed as local ion concentrations (c^*) at different polar sites located on the eight RNA residues (A) U67, (B) C68, (C) C69, (D) A70, (E) A71, (F) A72, (G) G73 and (H) A74. Different binding sites on the RNA are annotated following the PDB nomenclature.

and nucleobase atoms (blue line in Figure 3). However, in the presence of divalent ions, K^+ concentration around the RNA atoms drops drastically (magenta and cyan lines in Figure 3). In the mixture of K^+ and Ca^{2+} , the K^+ have $c^* \approx 0$ at all the polar sites on RNA (magenta line in Figure 3), as only the Ca^{2+} condense around the polar sites of RNA ($c^* \gtrsim 0.5$ for most of the sites, black line in Figure 3). However, in the mixture of K^+ and Mg^{2+} , a noticeable concentration of K^+ is observed around the RNA atoms with relatively higher c^* values around the nucleobase atoms present at the Hoogsteen edge of the purine residues (Figure 3D-H). Therefore in a mixture of K^+ and Mg^{2+} both contribute towards folding with Mg^{2+} being the major contributor (Figure 3D,E). The important point to note here is that all the three metal ions (K^+ , Mg^{2+} and Ca^{2+}) prefer to condense around the Hoogsteen edge of A72/G73. Therefore, the ion binding site (as present in the crystal structure) cannot discriminate between different metal ions.

This observation is also in agreement with the experiments,^{49,92,93} which have shown that Mg^{2+} sensing M-box riboswitch from which the RNA fragment is extracted is not very selective for binding Mg^{2+} unlike other metal sensing riboswitches. This riboswitch is shown to compact and fold in the presence of other divalent metal ions such as Mn^{2+} . Crystals of M-box riboswitch grown in the presence of Mn^{2+} show that the binding pockets of Mg^{2+} are occupied by Mn^{2+} .^{92,93} Since the Mg^{2+} sensing M-box riboswitch has small affinity and selectivity of metal ions it is hypothesised that this riboswitch probably functions by binding to the most prevalent Mg^{2+} ion present in the cell.

Specific Site Bound Metal Ion is Intrinsic Component of the TS: Although, the L4 tetraloop folds into the shape of a conventional GNRA tetraloop in the presence of Mg^{2+} , its folding thermodynamics is significantly different from that of a conventional GNRA tetraloop. The rate limiting step in the unfolding of GNRA tetraloops is the breaking of the loop closing G:A *trans*S:H base pair⁹⁴ stabilized by two strong inter-base hydrogen bonds.⁹⁵ Whereas the crystal structure of L4 tetraloop suggests that stability of the loop closing C:A pair and hence, the sequential adenine-adenine stacks primarily depend on the presence of

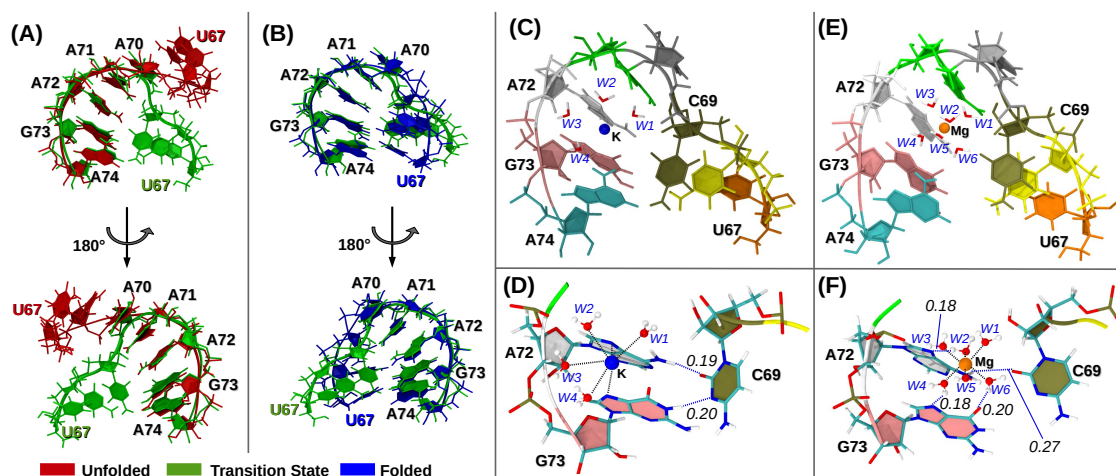


Figure 4: Representative structure from the TSE (green) is structurally aligned with a representative structure from the (A) folded basin (blue) and (B) unfolded basin (red). The alignment is performed with respect to only the A70-A74 residues. (C) Representative structure of the TSE with a K⁺ bound at the Hoogsteen edge of A72 and G73. The K⁺ is partially dehydrated and four water molecules present in its first coordination shell are labeled W1 - W4. (D) Details of the K⁺-RNA interactions are further illustrated. The K⁺ directly interacts with the N7 atoms of A72 and G73 and with four water molecules. These contacts are shown in black broken line. C69 forms two inter-base hydrogen bonds (shown in blue broken lines), one between O2 of C69 and N6 of A72, and the other between N3 of C69 and N1 of G73. (E) Representative structure from the TSE shows a hydrated Mg²⁺ bound at the Hoogsteen edge of A72 and G73. The six water molecules in the first coordination shell of the bound Mg²⁺ are labeled W1 - W6. (F) Details of the Mg²⁺-RNA interactions are highlighted. The W2, W4 and W6 water molecules form hydrogen bonds with the N7 atom of A72, N7 atom of G73 and O6 atom of G73, respectively. In addition there is an inter-base hydrogen bond between O2 of C69 and N6 of A72. These hydrogen bonds are shown in blue broken lines. Interaction between Mg²⁺ and water molecules are shown in black broken lines. For each hydrogen bond, the corresponding distance between the hydrogen and the acceptor is reported in nm. Metal ions are shown as spheres with radii equal to 30% of their respective van der Waals radius.

hydrated Mg^{2+} inside the loop (Figure 1). To probe at what stage of folding, the metal ion binds to the RNA and how it selects the specific site for binding, we further identified the TS connecting the folded and unfolded basins.

We have used the theory proposed by Hummer^{84,85} to identify the TSE using the CVs, eRMSD and R_{ee} . The FESs projected onto eRMSD and R_{ee} show that R_{ee} can also resolve the folded state and various unfolded states for both $[\text{Mg}^{2+}] = 0$ mM and $[\text{Mg}^{2+}] = 65$ mM (Figure S8). In the previous studies of GAGA and UUCG type tetraloop folding, eRMSD and R_{ee} were also used as CVs to construct the folding FES.^{34,35} To identify TSE, we analyzed the paths that connect the unfolded and folded basins of the L4 tetraloop in the simulation trajectories. In each path, the RNA fragment before transitioning to the folded basin hops between different unfolded basins and the extended basin. The final transition to the folded basin F takes place from either the unfolded basin U or from the extended basin E. In the absence of Mg^{2+} , 20% of the transitions to the folded state took place from the extended basin. Whereas, in the presence of Mg^{2+} it reduced to only 8%. Therefore we restricted our TS search to only the paths connecting the unfolded U and folded F basins (Figure S2,S3).

TSE analysis carried out for $[\text{Mg}^{2+}] = 0$ mM and 65 mM systems suggests that for both the systems, the value of $p(TP|r)$ is maximum at the same configuration r , which is located at eRMSD ≈ 1.2 nm and $R_{ee} \approx 1.8$ nm on the FES (Figure S4A,B). Further, P_{fold} analysis of the RNA conformations (along with the ion and water molecules) that have maximum value of $p(TP|r)$ helped in identifying the role of ions and water in the TS structures. The overall geometry of the TS structures are similar when folding is driven by the specific binding of K^+ or Mg^{2+} inside the tetraloop (Figure 4). The TS is primarily composed of an unfolded helix and a nearly folded loop with a site bound metal ion. However, the interactions between RNA and the metal ions change significantly when Mg^{2+} is added to the solution.

Hydrated Mg^{2+} Forms Key Interactions in the TS that Drive the Formation of the Loop: Structural superposition of (a) unfolded and TS structures (Figure 4A) and, (b) folded and TS structures (Figure 4B) collectively suggest that for the formation of the

folded hairpin motif from the single stranded helix conformation in the unfolded state will require breaking of the base-base stacking interactions between the consecutive C69 and A70 residues at the 5' end. This is followed by the formation of the A-form helix involving C68:G73 and U67:A74 base pairs. In DNA double helix, base-stacking is the main stabilizing factor.⁹⁶ Similarly in single stranded RNA helix, the strong base-stacking interactions⁹⁷ are important for structural stabilization. Therefore, during the L4 tetraloop folding, after the breakdown in the consecutive base-stacking interactions at the 5' end, the resulting alternate geometry has to be stabilized by other competing noncovalent interactions leading to folding. Structures in the TSE reveal that it involves a network of hydrogen bonding interactions that connect C69 to the stacked A72/G73 residues (Figure 4C,E). This network of hydrogen bonding interactions are further supported by metal ion coordination.

In the case of K^+ driven folding, a partially dehydrated K^+ is observed to interact directly with the Hoogsteen edges (N7 positions) of A72 and G73 (Figure 4D). Whereas, in the case of Mg^{2+} driven folding (Figure 4F), the water molecules present in the coordination shell of the Mg^{2+} interact with the Hoogsteen edges of both A72 (N7 position) and G73 (both N7 and O6 positions). The resulting electrostatic interactions between this cationic environment and the electronegative sites of the WC edge of C69 (O2 and N3) are crucial for the stabilization of the loop in the TS. Compared to other metal ions, Mg^{2+} has an extremely high charge to volume ratio due to its small ionic radius (Table S1), which makes it a preferable candidate for mediation of these electrostatic interactions.

Quantum chemical calculations suggest that metal ion coordination at the Hoogsteen edge of purines modulate the charge distribution and hence the hydrogen bonding potential of the respective WC edges.⁹⁸⁻¹⁰⁰ Generally, on metal ion binding at the Hoogsteen edge of a purine, the hydrogen bonding potential of the hydrogen bond donors (N6 in adenine, N1 and N2 in guanine) and acceptors (N1 in adenine, O6 in guanine) increases and decreases, respectively. However, in classical MD simulations with non-polarizable force fields we cannot capture such effects.

A-form Helix Formation Requires Rearrangement of Mg^{2+} Mediated Interactions: Comparison of the structures in the TSE and folded state shows that C69 and A72 interact with each other in *trans* orientation in the folded state (Figure 4B,S6B,D), whereas in the TSE they interact in *cis* orientation (Figure 4D,F). Therefore for the L4 tetraloop to fold, the C69 residue has to rearrange itself to interact with A72 in *cis* conformation. This requires breaking of the hydrogen bond between O2 of C69 and N6 of A72 present in the TS (Figure 4D,F) and formation of new hydrogen bonds between O2 of C69 and water molecules present at the first coordination shell of K^+/Mg^{2+} (Figure S6B,D).

For K^+ driven folding, this rearrangement results in the formation of new hydrogen bond between one K^+ coordinated water and O6 position of G73 in the folded state (Figure S6B). Also, N6 of A72 forms a hydrogen bond with the sugar oxygen O2' of C69 (Figure S6B). For Mg^{2+} driven folding (Figure S6D) the consequence of such rearrangement is reflected in elongation of the hydrogen bonds that connect the hydrated Mg^{2+} to the Hoogsteen edges of A72 and G73. This shows that the K^+/Mg^{2+} coordinated water molecules are crucial for rearrangement of C69 residue and subsequent formation of the A-form helix. Divalent cations like Ca^{2+} that lack the rigid octahedral orientation of the inner-shell water molecules¹⁰¹ cannot efficiently facilitate the formation of L4 tetraloop hairpin motif. Hence, the role of Mg^{2+} in the folding of L4 hairpin is very specific and other divalent cations are not as efficient in creating the same effect.

Spatial Organization of Electronegative Nucleobase Atoms Drive the Site-Specific Binding of Mg^{2+} : Mg^{2+} binds to specific sites in RNA and controls its folding.⁴² The specific binding was proposed to be driven by spatial arrangement of negatively charged phosphate groups in the folded RNA molecule and subsequent development of binding pockets with high anionic potential. However, using a coarse-grained model of *Azoarcus* ribozyme and MD simulations it was recently demonstrated that site specific binding of Mg^{2+} to RNA occurs even when the RNA is unfolded and the tertiary interactions that define the folded conformation are absent.²⁶

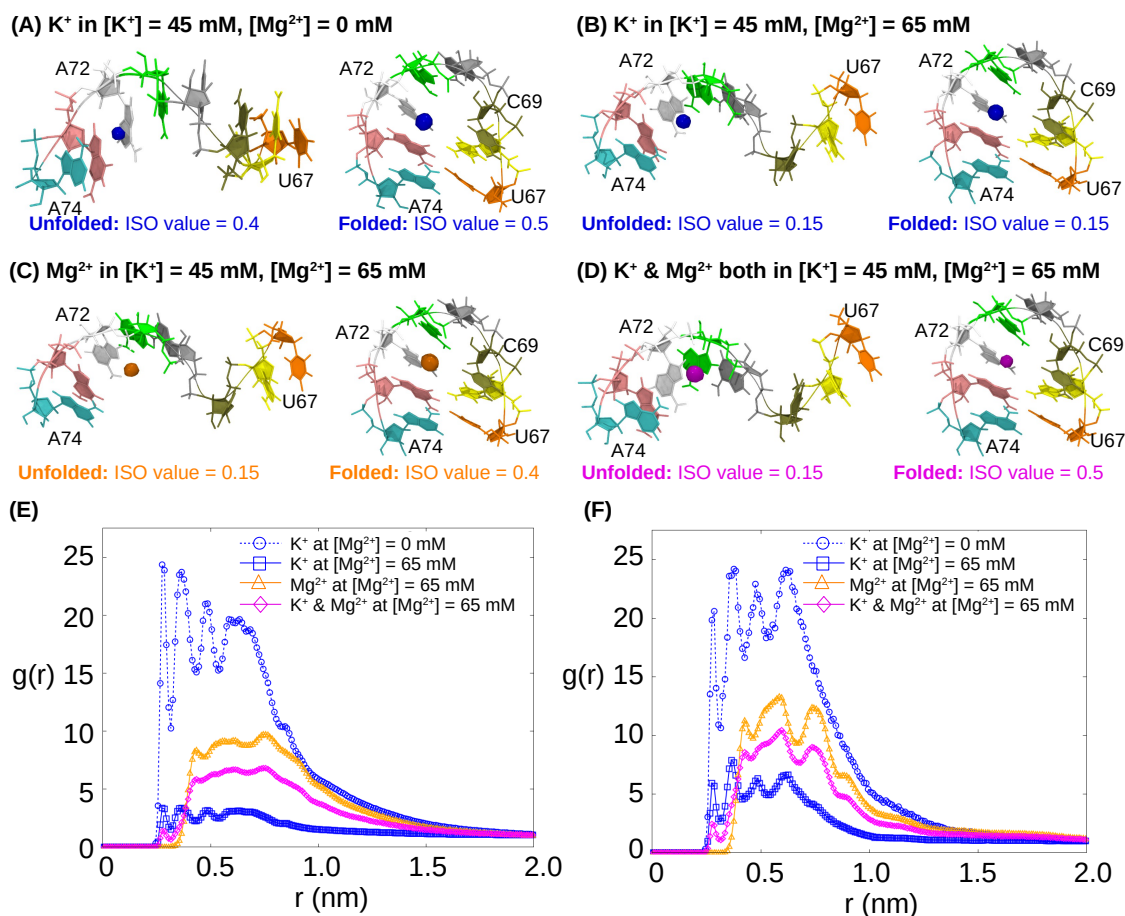


Figure 5: (A) Spatial density map of K⁺ around the unfolded and folded states of RNA are shown as blue isosurfaces at [Mg²⁺] = 0 mM. Respective ISO value of a isosurface indicates the fraction of time for which the area under the surface is occupied by K⁺. Similarly, at [Mg²⁺] = 65 mM, spatial density maps for (B) K⁺, (C) Mg²⁺ and (D) both K⁺ & Mg²⁺ around the unfolded and folded states of RNA are shown as isosurfaces in blue, orange and magenta, respectively. ISO values for each isosurfaces are mentioned below the respective illustrations. Maximum observed ISO values for each case are given in Table S2. (E) Pair distribution function ($g(r)$) between 2 groups of atoms for the RNA unfolded state. The first group contains atoms of A72 and G73 nucleotides of RNA, and the second group contains metal ions present in the solution. r is the distance between a pair of particles in the 2 groups. (F) $g(r)$ between 2 groups of K atoms for the RNA folded state. The first group contains atoms of C69 and A72 nucleotides of the RNA and the second group contains different metal ions present in the solution.

In our simulations, we also observe that Mg^{2+} binds to the Hoogsteen edge of A72 even when the L4 tetraloop is unfolded. The spatial density map of the metal ions around the RNA conformations in the unfolded state show that occupancy of the metal ions is maximum near the Hoogsteen edge of A72 (Figure 5A-5D). To probe the interaction between metal ions and the Hoogsteen edge of A72, we computed the pair distribution function ($g(r)$) between different metal ions and the RNA fragment (Figure 5E). Similar to the trends observed in Figure 3, the $g(r)$ peak height for Mg^{2+} is significantly higher than that for K^+ in $[\text{Mg}^{2+}] = 65$ mM solution. It is easier for the metal ions with smaller size such as Mg^{2+} to reach specific RNA binding pockets with high electronegative potentials, as proposed in earlier reports.^{26,42} Consistent with the earlier report¹⁰² we also observe that the position of the first peak in the $g(r)$ for K^+ is at a shorter distance ($r \approx 0.2$ nm) compared to that for Mg^{2+} ($r \approx 0.5$ nm). This indicates that K^+ directly interacts with the RNA, whereas Mg^{2+} interacts via the water molecules in its first coordination shell (diameter of water molecule is ≈ 0.28 nm). Interestingly, the occupancy of metal ions is found to be maximum near the Hoogsteen edge of A72 even for RNA conformations in the folded basin (Figure 5). We probed further to understand the reason for consistent electrostatic potential around the Hoogsteen edge of A72, which facilitates the binding of metal ions throughout the L4 tetraloop folding process.

Superposition of folded, unfolded and TS structures with respect to each other reveal that, throughout the folding process five consecutive purine residues at the 3' end (A70-A74) remain stacked with each other (Figure 4A-B). This brings the Hoogsteen edges of the five purines in a specific order where the Hoogsteen edge of A72 remains at the center. It is known that hydrogen bond acceptor sites of the Hoogsteen edges of purines (N7 in adenine and N7 and O6 in guanine) are strongly electronegative¹⁰³ and N7 of guanine has the highest proton affinity out of all the nucleobase atoms.¹⁰⁴ Therefore, we argue that in addition to the negatively charged backbone phosphate groups, spatial organization of electronegative nucleobase atoms critically contribute towards construction of the unique electrostatic potential that drives metal ion binding at specific sites. To find out whether it

is a general rule or specific to this L4 tetraloop hairpin motif, requires further investigation.

The results discussed above collectively suggest that binding of metal ions at specific site on RNA (hoogsteen edge of A72/G73) is indispensable for the folding of L4 tetraloop. Folding efficiency becomes maximum when the bound ion is Mg^{2+} . Such selective preference towards Mg^{2+} is determined by its involvement in the TS. In fact, hydrated Mg^{2+} are intrinsic components of the TS. Binding of hydrated Mg^{2+} at the Hoogsteen edge of A72/G73 provides the electrostatic stabilization necessary for the formation of the loop part of the motif. For the subsequent formation of the helix part of the motif, the Mg^{2+} coordinated water molecules actively mediate the rearrangement of the necessary inter-nucleotide interactions. Other divalent ions like Ca^{2+} that lack a rigid octahedral arrangement of their coordinated water molecules fail to facilitate these rearrangements efficiently. On the other hand, monovalent ions like K^+ has to undergo partial dehydration to facilitate folding. This is energetically costly^{105,106} and the subsequent folded state is also unstable. We further observe that the specific electronegative potential required for the binding of Mg^{2+} is provided by the spatial organization of the electronegative nucleobase atoms of selected nucleotides, in addition to the negatively charged backbone. Therefore, the CVs designed on the basis of coordination number between metal ions and atoms of selected nucleosides (nucleobase + sugar) successfully provide a comprehensive description of the free energy landscape.

Conclusions

Several aspects of RNA folding remains elusive despite extensive efforts.¹⁰⁷ Elucidating the exclusive contribution of metal ions in the formation of globular RNA structures is one of the major challenges.¹⁰⁸ Out of various metal ions, Mg^{2+} has the most versatile contribution in RNA's folding. Recent studies have unraveled active participation of Mg^{2+} in modulating RNA's catalytic activity,¹⁰⁹ conformational flexibility,¹¹⁰ and stability of folding intermediates.¹¹¹ In this paper we show that a site bound hydrated Mg^{2+} is an intrinsic component of

the TS of a unique RNA hairpin motif. Similar to the folding of the P4-P6 RNA domain,¹¹² the observed TS is an early TS which lacks majority of its native contacts. Formation of these missing native contacts takes place through water mediated interactions promoted by the water molecules present in the first coordination shell of the chelated Mg^{2+} . Mg^{2+} coordinated waters are known to form strong hydrogen bonds in RNA.¹¹³ The L4 tetraloop hairpin motif studied here utilizes such strong hydrogen bonding interactions to connect the distant nucleotides and stabilize the folded state. This is an important addition to the long standing discussion^{114,115} on the importance of water in RNA folding.

The larger goal of our study was to understand how large RNAs, like metal sensing riboswitches,⁴⁸ selectively recognize specific metal ions. Conventionally anionic phosphate groups are treated as building blocks of metal binding pockets in RNA.³⁶ However, the metal ion binding site present in the L4 tetraloop is composed of electronegative nucleobase atoms. This is not surprising as structural bioinformatics based analysis of available RNA crystal structures have revealed that out of the 13 known Mg^{2+} binding motifs in RNA, 8 have one or more nucleobase atoms present in the inner and/or outer shell of the Mg^{2+} .²⁴ Rather what is remarkable in this context is the fact that, this binding site cannot discriminate between different metal ions. Experimentally, it was shown that for Mg^{2+} sensing M-box riboswitch from which the RNA fragment is extracted is not very selective for binding Mg^{2+} .⁹² Based on this it was argued that M-box riboswitches can sense Mg^{2+} because they are the most prevalent divalent ions *in-vivo*.^{48,92} However, we propose an alternative hypothesis for the source of metal ion specificity in metallo-riboswitches. As we know, metal ion binding in riboswitches¹¹⁶ depends on the availability of a binding site with complementary electrostatic potential. In addition to that we propose the active involvement of the metal ion in the formation of the TS and stabilization of the folded state could also be an important factor.

We further observe that, even for a small RNA fragment like the L4 tetraloop hairpin motif, the folding FES becomes multidimensional especially when multiple metal ions are present in the solution. We have identified relevant CVs, which will aid in computing the FES

of large RNAs using enhanced sampling techniques. The L4 tetraloop (CAAA) folds in the shape of a conventional GNRA tetraloop, where two nucleobases (C69 and A72) connected by hydrated metal ions mimics the effect of an isosteric¹¹⁷ substitution of a noncanonical base pairs (i.e., the terminal G:A *trans*S:H pair present in a GNRA tetraloop). It is therefore important for the developers of RNA 2D¹¹⁸ and 3D¹¹⁹ structure prediction algorithms to take this check point into consideration. Interestingly, the nearest GNRA sequence of the L4 tetraloop is GAAA (Figure 1). Tertiary interaction between this GAAA tetraloop and its receptor motif has been used extensively as a model system in single molecule experiments to study various thermodynamic and kinetic aspects of RNA folding.¹²⁰ They reveal a complex dependence of the RNA folding dynamics on metal ion concentrations.^{121,122} Note that docking of the tetraloop with its receptor takes place via only the A-minor interactions involving the adenine residues of the tetralop, which are common in both GAAA and L4 tetraloops.¹²³ Also it is demonstrated that these adenine residues remain spatially constrained throughout the folding of L4 tetraloop. Therefore, we suggest that studying the tetraloop-receptor interaction where the GAAA tetraloop is substituted by the metal sensitive L4 tetraloop may reveal further information regarding involvement of metal ions in RNA folding.

Scope of this work is limited by relatively small system size and application of a fixed charge based force field. Study of the complete aptamer domain of metal sensing riboswitches using polarized force fields can further validate our hypothesis. Nevertheless, to the best of our knowledge, this is the first study that elucidates the folding free energy surface of a RNA hairpin using unbiased all atom molecular dynamics simulations.

Acknowledgement

This work was supported in part by a grant from the National Supercomputing Mission (NSM) through the grant MeitY/R&D/HPC/2(1)/2014. G.R. acknowledges the Alexander von Humboldt Foundation for support of his visit to the Max Planck Institute for Polymer

Research in Mainz. A.H. acknowledges the financial support from the DBT-RA Program in Biotechnology and Life Sciences. S.K. acknowledges research fellowship from Indian Institute of Science-Bangalore. The computations are performed using the TUE and Cray XC40 clusters at IISc. The authors thank Joseph F Rudzinski and Arghya Dutta for critical reading of the manuscript.

Supporting Information Available

Details of the coordination number calculation using PLUMED, list of parameters used for calculating the electrostatic potential surfaces in the PBEQ Solver server, Figure S1 – S8 and Table S1, S2. Figure S1: Time evaluation of eRMSD values in six equilibrium trajectories analyzed in this work. Figure S2: Illustration of transition paths in 0mM Mg²⁺ solution. Figure S3: Illustration of transition paths in 65mM Mg²⁺ solution. Figure S4: Results of the Hummer analysis performed to obtain the TSE. Figure S5: Differences between the crystal structure of the L4 tetraloop hairpin motif and its folded structure obtained in MD simulations. Figure S6: Details of metal-RNA interactions in the folded state. Figure S7: FES projected on single CV, eRMSD. Figure S8: FESs projected on two CVs, eRMSD and R_{ee} . Table S1: Comparison of chemical properties of different metal ions. Table S2: Maximum ISO values of the isosurfaces obtained from spatial density map analysis.

References

- (1) Lehninger, A. L.; Cox, M. M.; Nelson, D. L. *Lehninger principles of biochemistry*, 5th ed.; W.H. Freeman: New York, 2008.
- (2) Fedor, M. J.; Williamson, J. R. The catalytic diversity of RNAs. *Nat. Rev. Mol. Cell Biol.* **2005**, *6*, 399.

- (3) Morris, K. V., Ed. *RNA and the regulation of gene expression: a hidden layer of complexity*, 1st ed.; Caister Academic Press: Poole, United Kingdom, 2008.
- (4) Holoch, D.; Moazed, D. RNA-mediated epigenetic regulation of gene expression. *Nat. Rev. Genet.* **2015**, *16*, 71.
- (5) Mali, P.; Yang, L.; Esvelt, K. M.; Aach, J.; Guell, M.; DiCarlo, J. E.; Norville, J. E.; Church, G. M. RNA-guided human genome engineering via Cas9. *Science* **2013**, *339*, 823–826.
- (6) Jinek, M.; East, A.; Cheng, A.; Lin, S.; Ma, E.; Doudna, J. RNA-programmed genome editing in human cells. *Elife* **2013**, *2*, e00471.
- (7) Pyle, A. Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem.* **2002**, *7*, 679–690.
- (8) Draper, D. E. A guide to ions and RNA structure. *RNA* **2004**, *10*, 335–343.
- (9) Klein, D. J.; Moore, P. B.; Steitz, T. A. The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA* **2004**, *10*, 1366–1379.
- (10) Woodson, S. A. Compact but disordered states of RNA. *Nat. Struct. Mol. Biol.* **2000**, *7*, 349.
- (11) Thirumalai, D.; Lee, N.; Woodson, S. A.; Klimov, D. Early events in RNA folding. *Annu. Rev. Phys. Chem.* **2001**, *52*, 751–762.
- (12) Heilman-Miller, S. L.; Thirumalai, D.; Woodson, S. A. Role of counterion condensation in folding of the *Tetrahymena* ribozyme. I. Equilibrium stabilization by cations. *J. Mol. Biol.* **2001**, *306*, 1157–1166.
- (13) Fang, X.-w.; Pan, T.; Sosnick, T. R. Mg²⁺-dependent folding of a large ribozyme without kinetic traps. *Nat. Struct. Mol. Biol.* **1999**, *6*, 1091.

- (14) Fang, X.-w.; Pan, T.; Sosnick, T. R. A thermodynamic framework and cooperativity in the tertiary folding of a Mg^{2+} -dependent ribozyme. *Biochemistry* **1999**, *38*, 16840–16846.
- (15) Fang, X.-w.; Littrell, K.; Yang, X.-j.; Henderson, S. J.; Siefert, S.; Thiyagarajan, P.; Pan, T.; Sosnick, T. R. Mg^{2+} -dependent compaction and folding of yeast tRNA^{Phe} and the catalytic domain of the *B. subtilis* RNase P RNA determined by small-angle X-ray scattering. *Biochemistry* **2000**, *39*, 11107–11113.
- (16) Li, P. T.; Vieregge, J.; Tinoco, I. How RNA unfolds and refolds. *Annu. Rev. Bioch.* **2008**, *77*, 77–100.
- (17) Bokinsky, G.; Rueda, D.; Misra, V. K.; Rhodes, M. M.; Gordus, A.; Babcock, H. P.; Walter, N. G.; Zhuang, X. Single-molecule transition-state analysis of RNA folding. *Proc. Nat. Acad. Sci. U.S.A.* **2003**, *100*, 9302–9307.
- (18) Uchida, T.; Takamoto, K.; He, Q.; Chance, M. R.; Brenowitz, M. Multiple monovalent ion-dependent pathways for the folding of the L-21 *Tetrahymena thermophila* ribozyme. *J. Mol. Biol.* **2003**, *328*, 463–478.
- (19) Shcherbakova, I.; Gupta, S.; Chance, M. R.; Brenowitz, M. Monovalent ion-mediated folding of the *Tetrahymena thermophila* ribozyme. *J. Mol. Biol.* **2004**, *342*, 1431–1442.
- (20) Roca, J.; Hori, N.; Baral, S.; Velmurugu, Y.; Narayanan, R.; Narayanan, P.; Thirumalai, D.; Ansari, A. Monovalent ions modulate the flux through multiple folding pathways of an RNA pseudoknot. *Proc. Nat. Acad. Sci. U.S.A.* **2018**, *115*, E7313–E7322.
- (21) Hanna, R.; Doudna, J. A. Metal ions in ribozyme folding and catalysis. *Curr. Opin. Struct. Biol.* **2000**, *4*, 166–170.

- (22) DeRose, V. J. Metal ion binding to catalytic RNA molecules. *Curr. Opin. Struct. Biol.* **2003**, *13*, 317–324.
- (23) Bowman, J. C.; Lenz, T. K.; Hud, N. V.; Williams, L. D. Cations in charge: magnesium ions in RNA folding and catalysis. *Curr. Opin. Struct. Biol.* **2012**, *22*, 262.
- (24) 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.
- (25) Misra, V. K.; Draper, D. E. A thermodynamic framework for Mg^{2+} binding to RNA. *Proc. Nat. Acad. Sci. U.S.A.* **2001**, *98*, 12456–12461.
- (26) Hori, N.; Denesyuk, N. A.; Thirumalai, D. Ion condensation onto ribozyme is site-specific and fold-dependent. *Biophysical J.* **2019**, *116*, 2400–2410.
- (27) Basu, S.; Rambo, R. P.; Strauss-Soukup, J.; Cate, J. H.; Amaré, F.-D.; 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.
- (28) Goody, T. A.; Melcher, S. E.; Norman, D. G.; Lilley, D. M. The kink-turn motif in RNA is dimorphic, and metal ion-dependent. *RNA* **2004**, *10*, 254–264.
- (29) Das, R.; Travers, K. J.; Bai, Y.; Herschlag, D. Determining the Mg^{2+} stoichiometry for folding an RNA metal ion core. *J. Am. Chem. Soc.* **2005**, *127*, 8272–8273.
- (30) Moghaddam, S.; Caliskan, G.; Chauhan, S.; Hyeon, C.; Briber, R.; Thirumalai, D.; Woodson, S. A. Metal ion dependence of cooperative collapse transitions in RNA. *J. Mol. Biol.* **2009**, *393*, 753–764.
- (31) Bartova, S.; Alberti, E.; Sigel, R. K.; Donghi, D. Metal ion binding to an RNA internal loop. *Inorganica Chim. Acta* **2016**, *452*, 104–110.

- (32) Lipfert, J.; Doniach, S.; Das, R.; Herschlag, D. Understanding nucleic acid–ion interactions. *Annu. Rev. Biochem.* **2014**, *83*, 813–841.
- (33) Šponer, J.; Bussi, G.; Krepl, M.; Banáš, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jurečka, P.; Walter, N. G.; Otyepka, M. RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* **2018**, *118*, 4177–4338.
- (34) Bottaro, S.; Banáš, P.; Šponer, J.; Bussi, G. Free energy landscape of GAGA and UUCG RNA tetraloops. *J. Phys. Chem. Lett.* **2016**, *7*, 4032–4038.
- (35) Bottaro, S.; Banáš, P.; Šponer, J.; Bussi, G. Correction to “Free energy landscape of GAGA and UUCG RNA tetraloops”. *J. Phys. Chem. Lett.* **2018**, *9*, 1674–1675.
- (36) Cunha, R. A.; Bussi, G. Unraveling Mg²⁺–RNA binding with atomistic molecular dynamics. *RNA* **2017**, *23*, 628–638.
- (37) Smith, L. G.; Tan, Z.; Spasic, A.; Dutta, D.; Salas-Estrada, L. A.; Grossfield, A.; Mathews, D. H. Chemically accurate relative folding stability of RNA hairpins from molecular simulations. *J. Chem. Theo. Comput.* **2018**, *14*, 6598–6612.
- (38) Bao, L.; Wang, J.; Xiao, Y. Dynamics of metal ions around an RNA molecule. *Phys. Rev. E* **2019**, *99*, 012420.
- (39) Hyeon, C.; Thirumalai, D. Mechanical unfolding of RNA hairpins. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 6789–6794.
- (40) Denesyuk, N. A.; Thirumalai, D. Coarse-grained model for predicting RNA folding thermodynamics. *J. Phys. Chem. B* **2013**, *117*, 4901–4911.
- (41) Šulc, P.; Romano, F.; Ouldridge, T. E.; Doye, J. P. K.; Louis, A. A. A nucleotide-level coarse-grained model of RNA. *J. Chem. Phys.* **2014**, *140*, 235102.

- (42) Denesyuk, N. A.; Thirumalai, D. How do metal ions direct ribozyme folding? *Nat. Chem.* **2015**, *7*, 793.
- (43) Cragolini, T.; Laurin, Y.; Derreumaux, P.; Pasquali, S. Coarse-grained HiRE-RNA model for ab initio RNA folding beyond simple molecules, including noncanonical and multiple base pairings. *J. Chem. Theo. Comput.* **2015**, *11*, 3510–3522.
- (44) Hayes, R. L.; Noel, J. K.; Mandic, A.; Whitford, P. C.; Sanbonmatsu, K. Y.; Mohanty, U.; Onuchic, J. N. Generalized Manning condensation model captures the RNA ion atmosphere. *Phys. Rev. Lett.* **2015**, *114*, 258105.
- (45) Bell, D. R.; Cheng, S. Y.; Salazar, H.; Ren, P. Capturing RNA folding free energy with coarse-grained molecular dynamics simulations. *Sci. Rep.* **2017**, *7*, 1–14.
- (46) Denesyuk, N. A.; Hori, N.; Thirumalai, D. Molecular simulations of ion effects on the thermodynamics of RNA folding. *J. Phys. Chem. B* **2018**, *122*, 11860–11867.
- (47) Ferre-D'Ámare, A. R.; Winkler, W. C. The roles of metal ions in regulation by riboswitches. *Met. Ions Life Sci.* **2011**, *9*, 141–173.
- (48) Wedekind, J. E.; Dutta, D.; Belashov, I. A.; Jenkins, J. L. Metalloriboswitches: RNA-based inorganic ion sensors that regulate genes. *J. Biol. Chem.* **2017**, *292*, 9441–9450.
- (49) Dann III, C. E.; Wakeman, C. A.; Sieling, C. L.; Baker, S. C.; Irnov, I.; Winkler, W. C. Structure and mechanism of a metal-sensing regulatory RNA. *Cell* **2007**, *130*, 878–892.
- (50) Price, I. R.; Gaballa, A.; Ding, F.; Helmann, J. D.; Ke, A. Mn²⁺-sensing mechanisms of yybP-ykoY orphan riboswitches. *Mol. Cell* **2015**, *57*, 1110–1123.
- (51) Dambach, M.; Sandoval, M.; Updegrove, T. B.; Anantharaman, V.; Aravind, L.; Waters, L. S.; Storz, G. The ubiquitous yybP-ykoY riboswitch is a manganese-responsive regulatory element. *Mol. Cell* **2015**, *57*, 1099–1109.

- (52) Furukawa, K.; Ramesh, A.; Zhou, Z.; Weinberg, Z.; Vallery, T.; Winkler, W. C.; Breaker, R. R. Bacterial riboswitches cooperatively bind Ni²⁺ or Co²⁺ ions and control expression of heavy metal transporters. *Mol. Cell* **2015**, *57*, 1088–1098.
- (53) Woese, C. R.; Winker, S.; Gutell, R. R. Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proc. Nat. Acad. Sci. U.S.A.* **1990**, *87*, 8467–8471.
- (54) Bottaro, S.; Lindorff-Larsen, K. Mapping the universe of RNA tetraloop folds. *Biophys. J.* **2017**, *113*, 257–267.
- (55) Heus, H. A.; Pardi, A. Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **1991**, *253*, 191–194.
- (56) Im, W.; Beglov, D.; Roux, B. Continuum solvation model: Computation of electrostatic forces from numerical solutions to the Poisson-Boltzmann equation. *Comput. Phys. Commun.* **1998**, *111*, 59–75.
- (57) Jo, S.; Vargyas, M.; Vasko-Szedlar, J.; Roux, B.; Im, W. PBEQ-Solver for online visualization of electrostatic potential of biomolecules. *Nuc. Acids Res.* **2008**, *36*, W270–W275.
- (58) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.
- (59) Kührová, P.; Banáš, P.; Best, R. B.; Šponer, J.; Otyepka, M. Computer folding of RNA tetraloops? Are we there yet? *J. Chem. Theo. Comput.* **2013**, *9*, 2115–2125.
- (60) Haldar, S.; Kührová, P.; Banáš, P.; Spiwok, V.; Šponer, J.; Hobza, P.; Otyepka, M. Insights into stability and folding of GNRA and UNCG tetraloops revealed by microsecond molecular dynamics and well-tempered metadynamics. *J. Chem. Theo. Comput.* **2015**, *11*, 3866–3877.

- (61) Bergonzo, C.; Henriksen, N. M.; Roe, D. R.; Cheatham, T. E. Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA* **2015**, *21*, 1578–1590.
- (62) Chen, A. A.; García, A. E. High-resolution reversible folding of hyperstable RNA tetraloops using molecular dynamics simulations. *Proc. Nat. Acad. Sci. U.S.A.* **2013**, *110*, 16820–16825.
- (63) Tan, D.; Piana, S.; Dirks, R. M.; Shaw, D. E. RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc. Nat. Acad. Sci. U.S.A.* **2018**, *115*, E1346–E1355.
- (64) Kührová, P.; Mlýnský, V.; Zgarbová, M.; Krepl, M.; Bussi, G.; Best, R. B.; Otyepka, M.; Šponer, J.; Banáš, P. Improving the performance of the AMBER RNA force field by tuning the hydrogen-bonding interactions. *J. Chem. Theo. Comput.* **2019**, *15*, 3288–3305.
- (65) Cesari, A.; Bottaro, S.; Lindorff-Larsen, K.; Banáš, P.; Šponer, J.; Bussi, G. Fitting corrections to an RNA force field using experimental data. *J. Chem. Theo. Comput.* **2019**, *15*, 3425–3431.
- (66) Berendsen, H.; van der Spoel, D.; van Drunen, R. GROMACS: a message-passing parallel molecular dynamics implementation. *Comp. Phys. Commun.* **1995**, *91*, 43–56.
- (67) GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.
- (68) Piana, S.; Donchev, A. G.; Robustelli, P.; Shaw, D. E. Water dispersion interactions strongly influence simulated structural properties of disordered protein states. *J. Phys. Chem. B* **2015**, *119*, 5113–5123.

- (69) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
- (70) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: a new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
- (71) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.
- (72) 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.
- (73) Hunter, J. D. Matplotlib: a 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95.
- (74) Bottaro, S.; Di Palma, F.; Bussi, G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res.* **2014**, *42*, 13306–13314.
- (75) Bottaro, S.; Gil-Ley, A.; Bussi, G. RNA folding pathways in stop motion. *Nucleic Acids Res.* **2016**, *44*, 5883–5891.
- (76) Kührová, P.; Best, R. B.; Bottaro, S.; Bussi, G.; Šponer, J.; Otyepka, M.; Banáš, P. Computer folding of RNA tetraloops: identification of key force field deficiencies. *J. Chem. Theo. Comput.* **2016**, *12*, 4534–4548.
- (77) Bottaro, S.; Bussi, G.; Pinamonti, G.; Reißer, S.; Boomsma, W.; Lindorff-Larsen, K. Barnaba: software for analysis of nucleic acids structures and trajectories. *RNA* **2018**, 10.1261/rna.067678.118.
- (78) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comp. Phys. Commun.* **2014**, *185*, 604–613.
- (79) Bonomi, M. et al. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **2019**, *16*, 670–673.

- (80) Humphrey, W.; Dalke, A.; Schulten, K. VMD – Visual Molecular Dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.
- (81) Kabsch, W. A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Cryst. A* **1978**, *34*, 827–828.
- (82) Eargle, J.; Wright, D.; Luthey-Schulten, Z. Multiple alignment of protein structures and sequences for VMD. *Bioinformatics* **2006**, *22*, 504–506.
- (83) Stone, J. An efficient library for parallel ray tracing and animation. M.Sc. thesis, Computer Science Department, University of Missouri-Rolla, 1998.
- (84) Hummer, G. From transition paths to transition states and rate coefficients. *J. Chem. Phys.* **2004**, *120*, 516–523.
- (85) Best, R. B.; Hummer, G. Reaction coordinates and rates from transition paths. *Proc. Nat. Acad. Sci. U.S.A* **2005**, *102*, 6732–6737.
- (86) Du, R.; Pande, V. S.; Grosberg, A. Y.; Tanaka, T.; Shakhnovich, E. S. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998**, *108*, 334–350.
- (87) Seol, Y.; Skinner, G. M.; Visscher, K.; Buhot, A.; Halperin, A. Stretching of homopolymeric RNA reveals single-stranded helices and base-stacking. *Phys. Rev. Lett.* **2007**, *98*, 158103.
- (88) Shiman, R.; Draper, D. E. Stabilization of RNA tertiary structure by monovalent cations. *J. Mol. Biol.* **2000**, *302*, 79–91.
- (89) Shi, Y.-Z.; Jin, L.; Wang, F.-H.; Zhu, X.-L.; Tan, Z.-J. Predicting 3D structure, flexibility, and stability of RNA hairpins in monovalent and divalent ion solutions. *Biophys. J.* **2015**, *109*, 2654–2665.

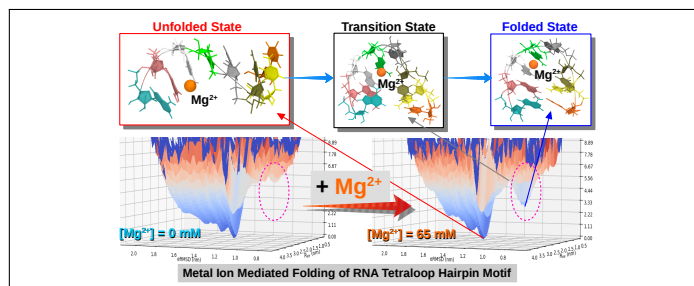
- (90) Lambert, D.; Leipply, D.; Shiman, R.; Draper, D. E. The influence of monovalent cation size on the stability of RNA tertiary structures. *J. Mol. Biol.* **2009**, *390*, 791–804.
- (91) Miserachs, H. G.; Donghi, D.; Börner, R.; Johannsen, S.; Sigel, R. K. Distinct differences in metal ion specificity of RNA and DNA G-quadruplexes. *J. Biol. Inorg. Chem.* **2016**, *21*, 975–986.
- (92) Ramesh, A.; Wakeman, C. A.; Winkler, W. C. Insights into metalloregulation by M-box riboswitch RNAs via structural analysis of manganese-bound complexes. *J. Mol. Biol.* **2011**, *407*, 556–570.
- (93) Wedekind, J. E.; Dutta, D.; Belashov, I. A.; Jenkins, J. L. Metalloriboswitches: RNA-based inorganic ion sensors that regulate genes. *J. Biol. Chem.* **2017**, *292*, 9441–9450.
- (94) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. J. *Nucleic acids: structure, properties and functions*, 1st ed.; Univ. Sci. Books: Sausalito, CA, 2000.
- (95) Mládek, A.; Sharma, P.; Mitra, A.; Bhattacharyya, D.; Šponer, J.; Šponer, J. E. Trans Hoogsteen/sugar edge base pairing in RNA. Structures, energies, and stabilities from quantum chemical calculations. *J. Phys. Chem. B* **2009**, *113*, 1743–1755.
- (96) Yakovchuk, P.; Protozanova, E.; Frank-Kamenetskii, M. D. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.* **2006**, *34*, 564–574.
- (97) Šponer, J.; Šponer, J. E.; Mládek, A.; Jurečka, P.; Banáš, P.; Otyepka, M. Nature and magnitude of aromatic base stacking in DNA and RNA: Quantum chemistry, molecular mechanics, and experiment. *Biopolymers* **2013**, *99*, 978–988.
- (98) Oliva, R.; Tramontano, A.; Cavallo, L. Mg²⁺ binding and archaeosine modification stabilize the G15–C48 Levitt base pair in tRNAs. *RNA* **2007**, *13*, 1427–1436.

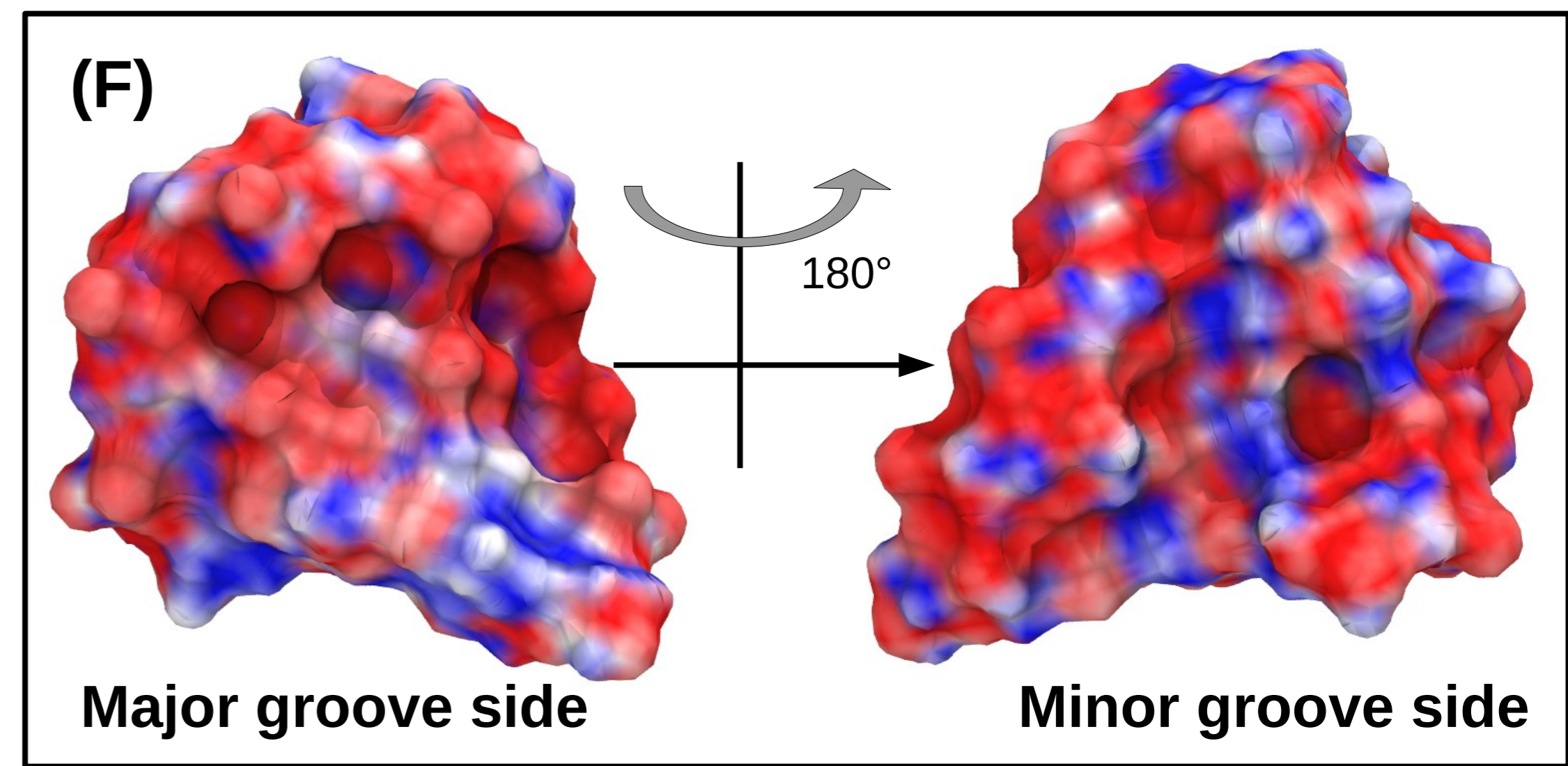
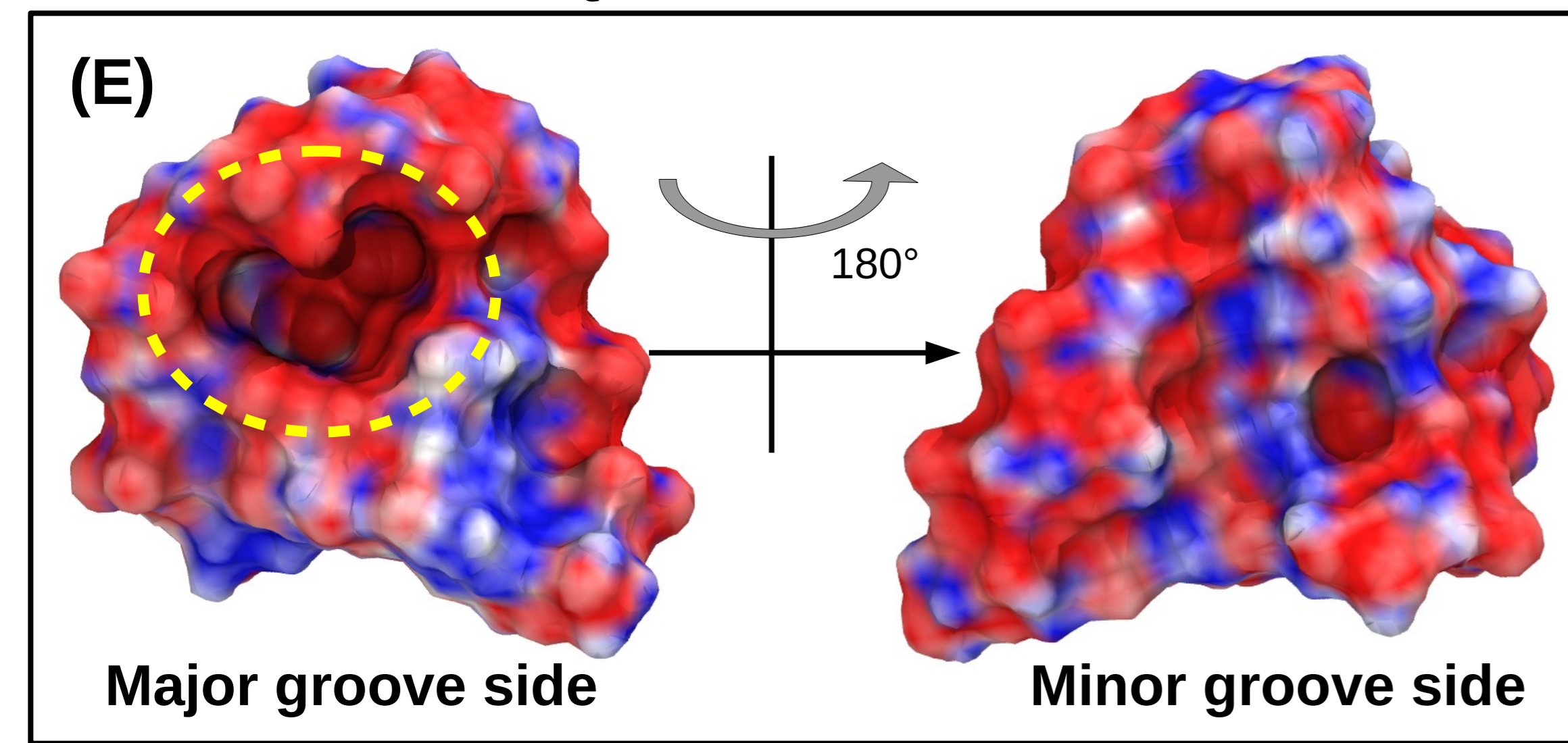
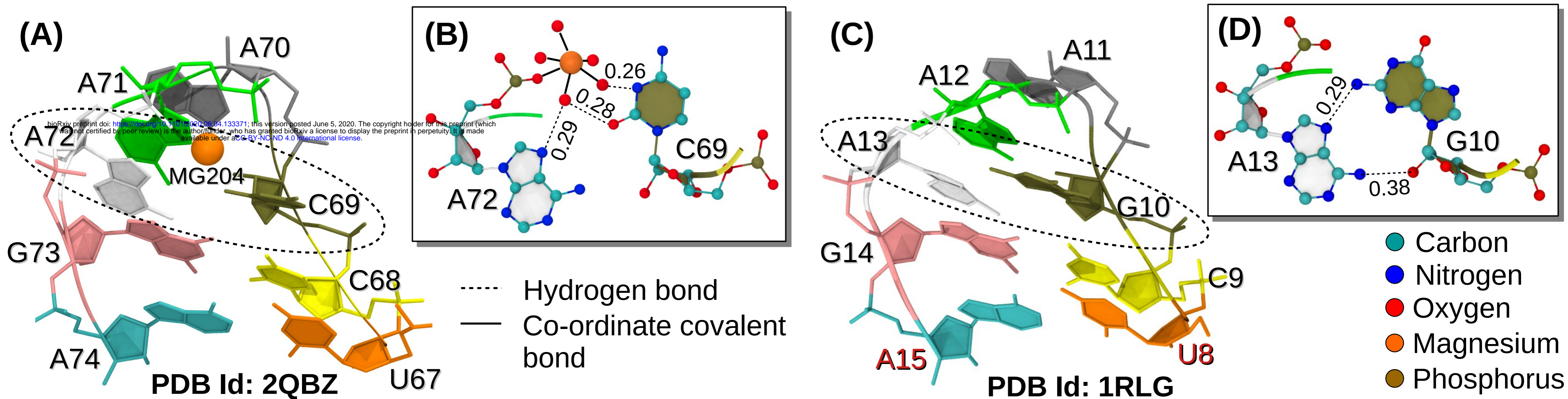
- (99) Oliva, R.; Cavallo, L. Frequency and effect of the binding of Mg^{2+} , Mn^{2+} , and Co^{2+} ions on the guanine base in Watson–Crick and reverse Watson–Crick base pairs. *J. Phys. Chem. B* **2009**, *113*, 15670–15678.
- (100) Halder, A.; Roy, R.; Bhattacharyya, D.; Mitra, A. Consequences of Mg^{2+} binding on the geometry and stability of RNA base pairs. *Phys. Chem. Chem. Phys.* **2018**, *20*, 21934–21948.
- (101) Ikeda, T.; Boero, M.; Terakura, K. Hydration properties of magnesium and calcium ions from constrained first principles molecular dynamics. *J. Chem. Phys.* **2007**, *127*, 074503.
- (102) Lammert, H.; Wang, A.; Mohanty, U.; Onuchic, J. N. RNA as a complex polymer with coupled dynamics of ions and water in the outer solvation sphere. *J. Phys. Chem. B* **2018**, *122*, 11218–11227.
- (103) Zhachkina, A.; Liu, M.; Sun, X.; Amegayibor, F. S.; Lee, J. K. Gas-phase thermochemical properties of the damaged base O6-methylguanine versus adenine and guanine. *J. Org. Chem.* **2009**, *74*, 7429–7440.
- (104) Halder, A.; Halder, S.; Bhattacharyya, D.; Mitra, A. Feasibility of occurrence of different types of protonated base pairs in RNA: A quantum chemical study. *Phys. Chem. Chem. Phys.* **2014**, *16*, 18383–18396.
- (105) Draper, D. E. A guide to ions and RNA structure. *RNA* **2004**, *10*, 335–343.
- (106) Draper, D. E.; Misra, V. K. RNA shows its metal. *Nat. Struct. Biol.* **1998**, *5*, 927–930.
- (107) Herschlag, D.; Bonilla, S.; Bisaria, N. The story of RNA folding, as told in epochs. *Cold Spring Harb. Perspect. Biol.* **2018**, *10*, a032433.
- (108) Auffinger, P.; Grover, N.; Westhof, E. In *Metal Ions in Life Sciences*; Sigel, A.,

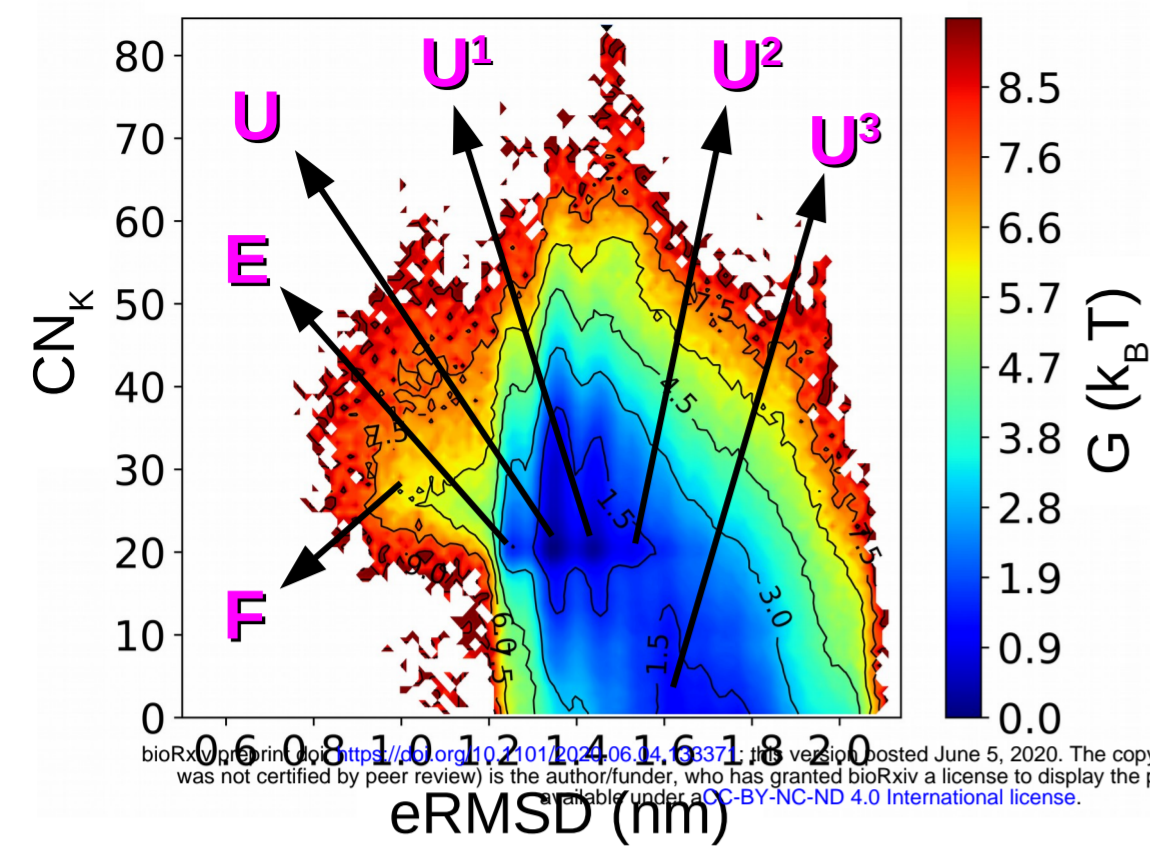
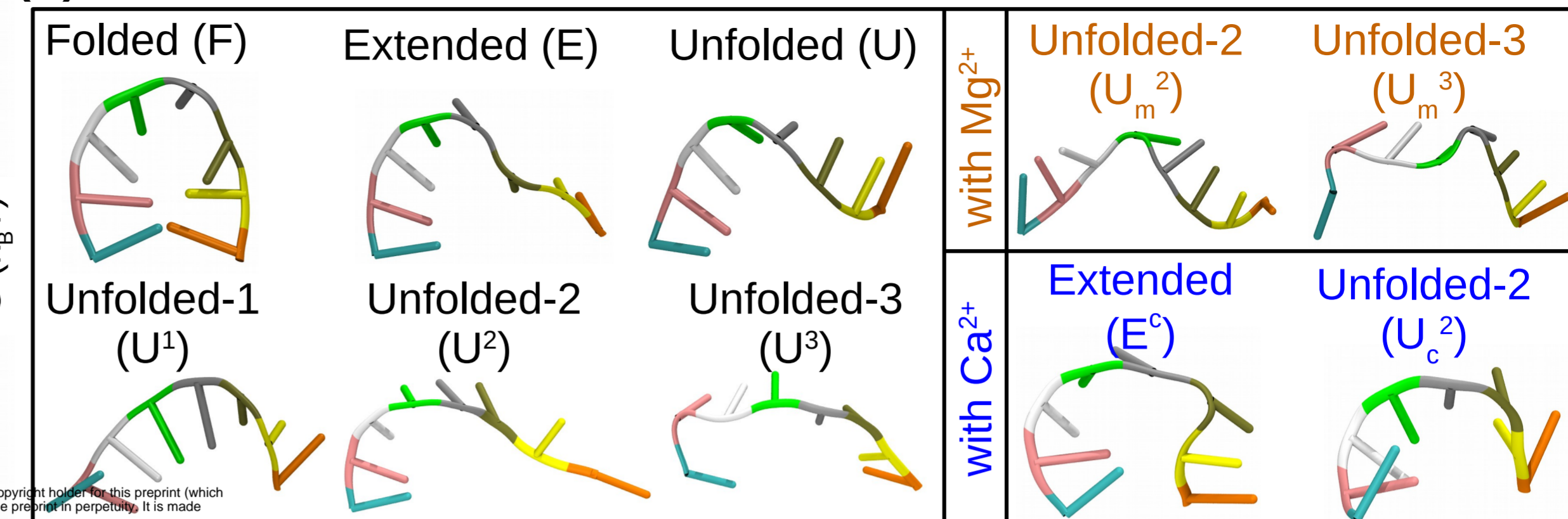
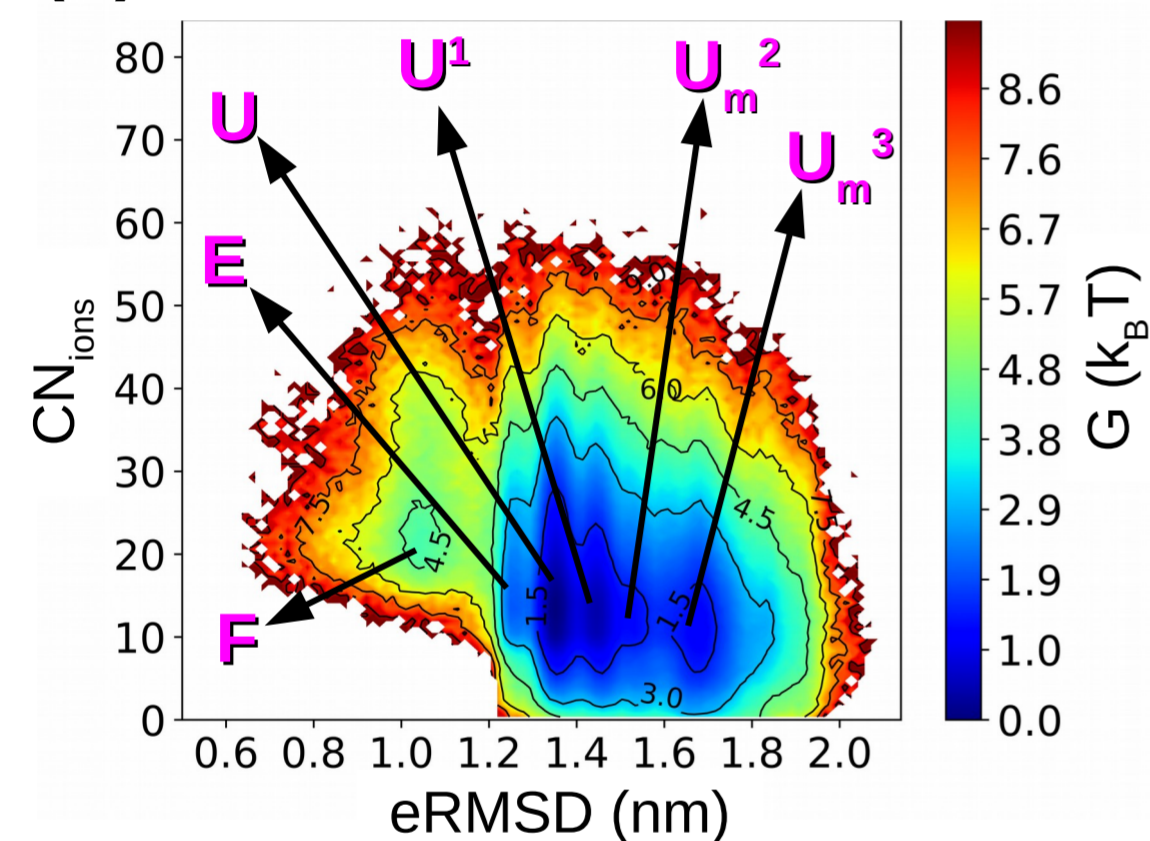
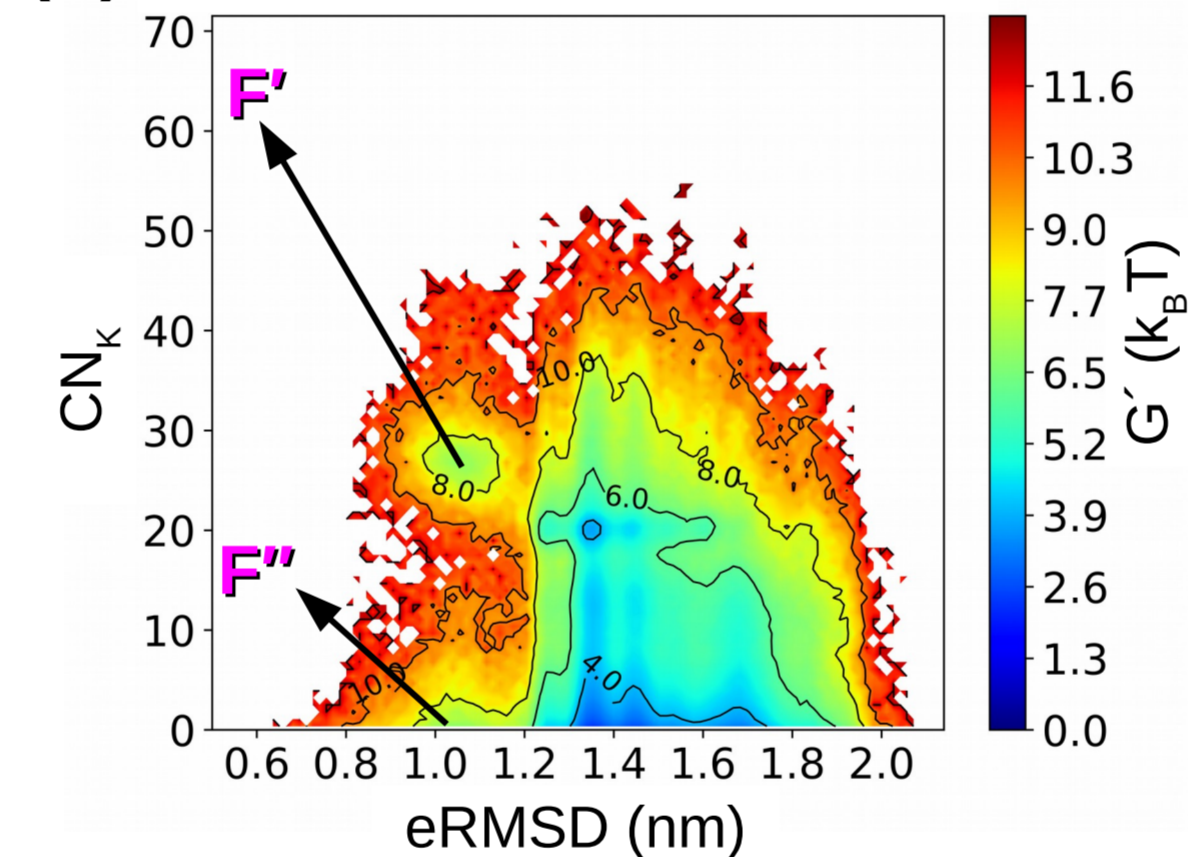
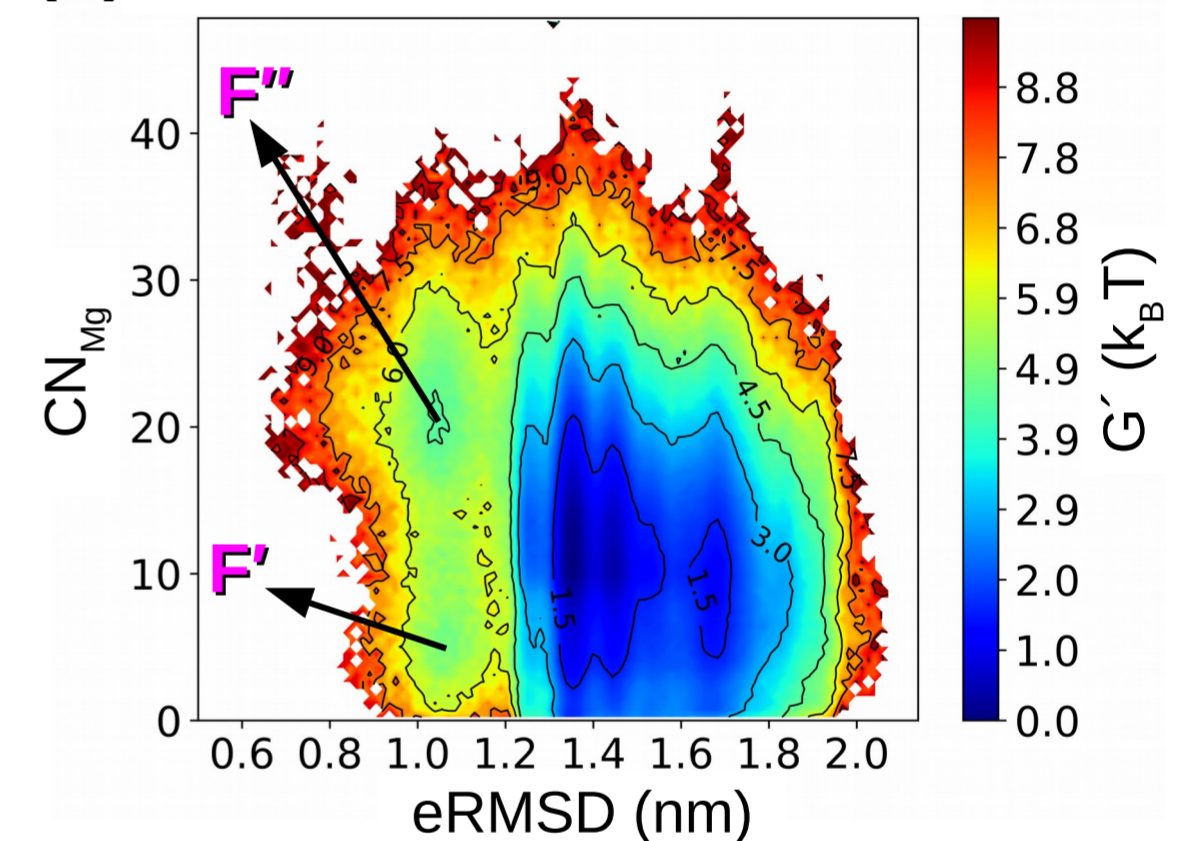
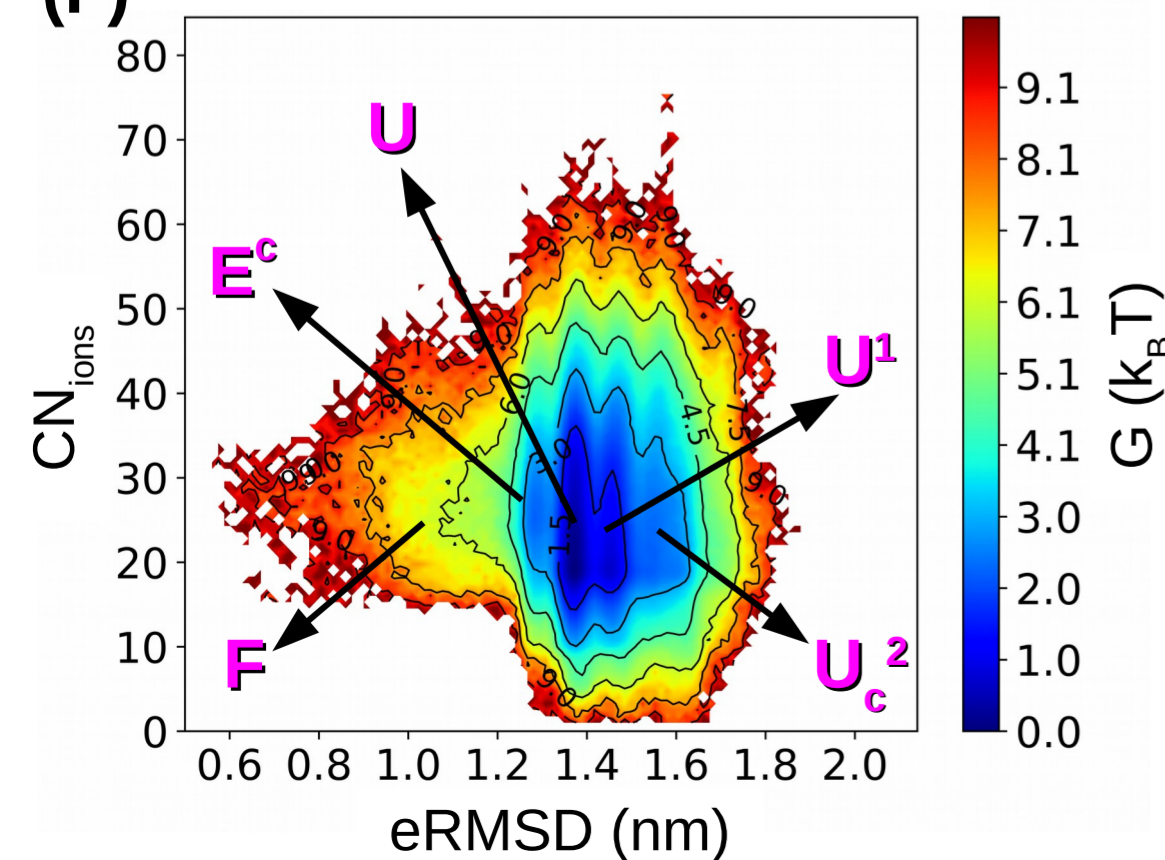
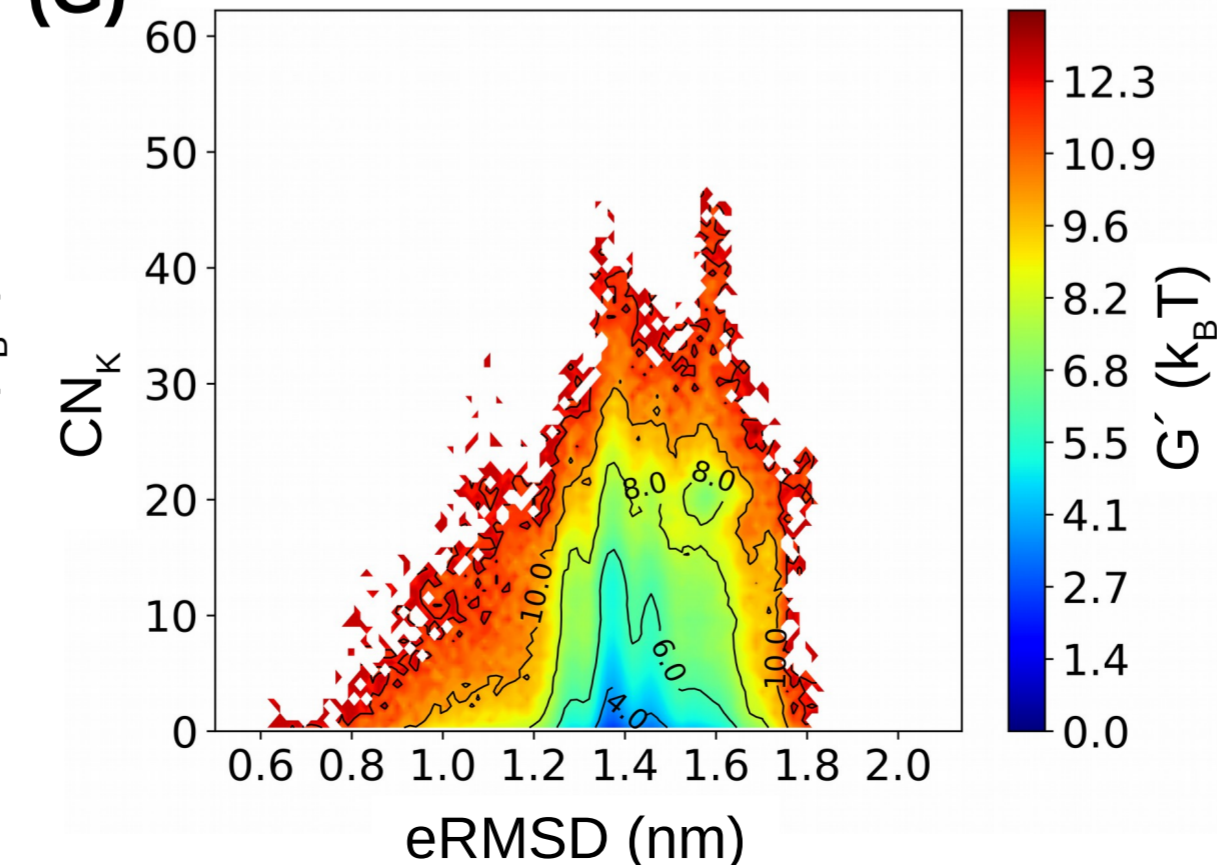
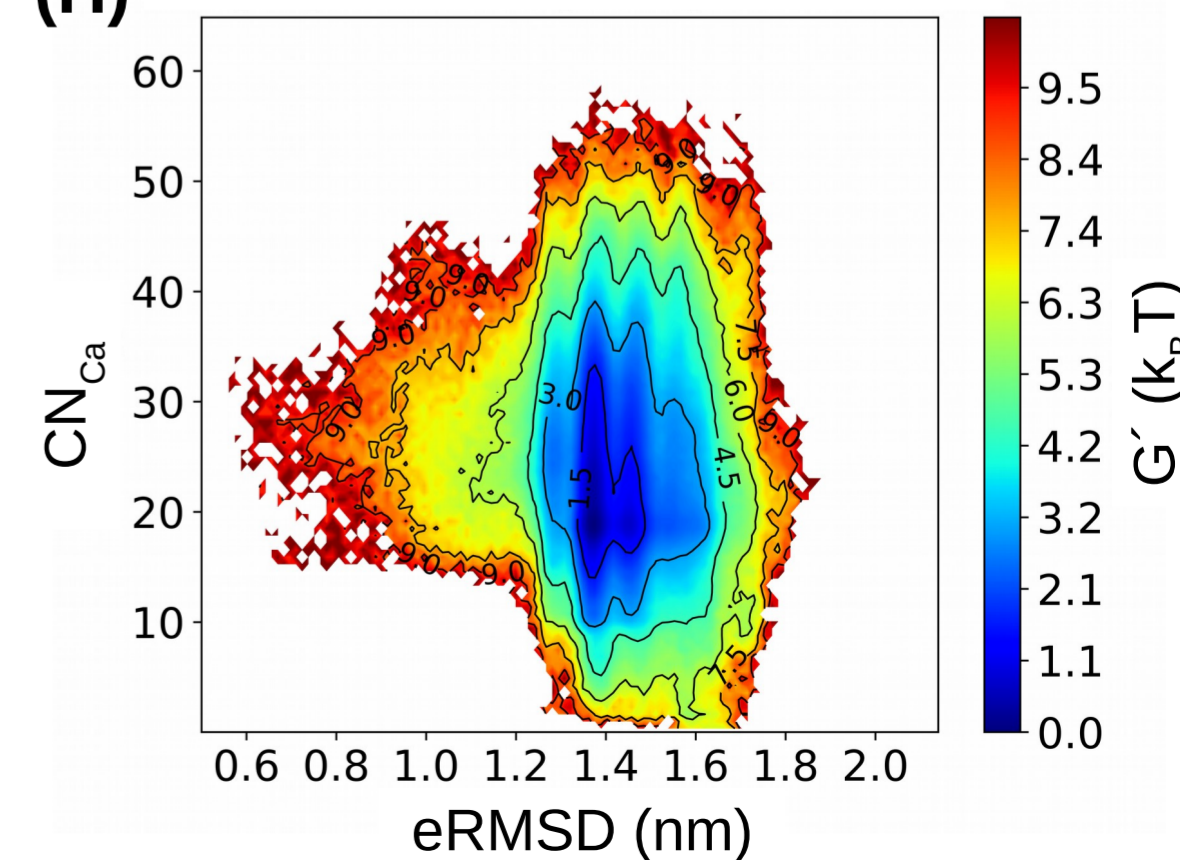
- Sigel, H., Sigel, R. K. O., Eds.; Royal Society of Chemistry,: London, United Kingdom, 2011; Chapter 1, pp 1–35.
- (109) Teplova, M.; Falschlunger, C.; Krasheninina, O.; Egger, M.; Ren, A.; Patel, D. J.; Micura, R. Crucial roles of two hydrated Mg^{2+} ions in reaction catalysis of the pistol ribozyme. *Angew. Chem. Int. Ed.* **2020**, *132*, 2859–2865.
- (110) Hetzke, T.; Vogel, M.; Gophane, D. B.; Weigand, J. E.; Suess, B.; Sigurdsson, S. T.; Prisner, T. F. Influence of Mg^{2+} on the conformational flexibility of a tetracycline aptamer. *RNA* **2019**, *25*, 158–167.
- (111) Plumridge, A.; Katz, A. M.; Calvey, G. D.; Elber, R.; Kirmizialtin, S.; Pollack, L. Revealing the distinct folding phases of an RNA three-helix junction. *Nucleic Acids Res.* **2018**, *46*, 7354–7365.
- (112) Silverman, S. K.; Cech, T. R. An early transition state for folding of the P4-P6 RNA domain. *RNA* **2001**, *7*, 161–166.
- (113) Yu, T.; Chen, S.-J. Hexahydrated Mg^{2+} binding and outer-shell dehydration on RNA surface. *Biophys. J.* **2018**, *114*, 1274–1284.
- (114) Sorin, E. J.; Rhee, Y. M.; Pande, V. S. Does water play a structural role in the folding of small nucleic acids? *Biophys. J.* **2005**, *88*, 2516–2524.
- (115) Templeton, C.; Elber, R. Why does RNA collapse? The importance of water in a simulation study of helix–junction–helix systems. *J. Am. Chem. Soc.* **2018**, *140*, 16948–16951.
- (116) Bédard, A.-S. V.; Hien, E. D.; Lafontaine, D. A. Riboswitch regulation mechanisms: RNA, metabolites and regulatory proteins. *BBA-Gene Regul. Mech.* **2020**, *1863*, 194501.

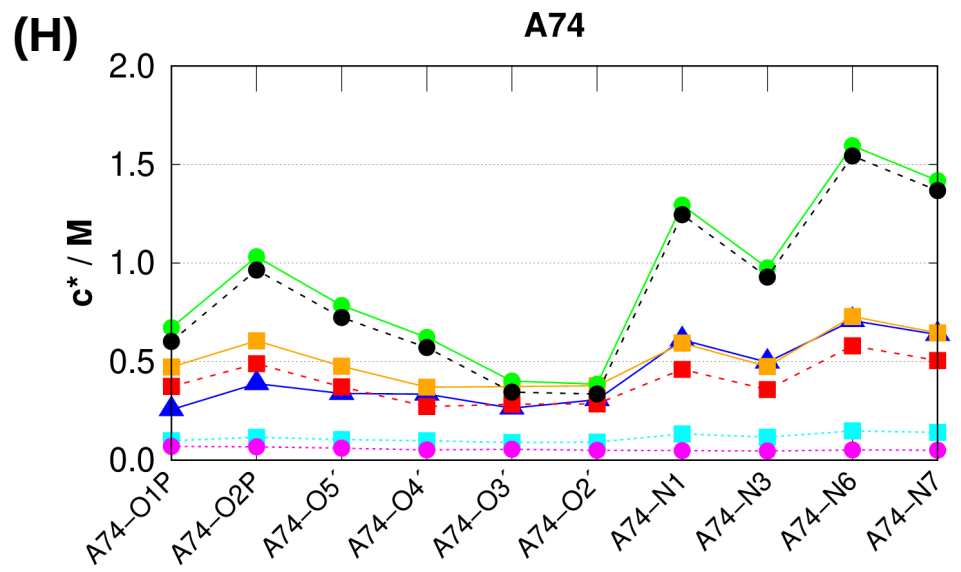
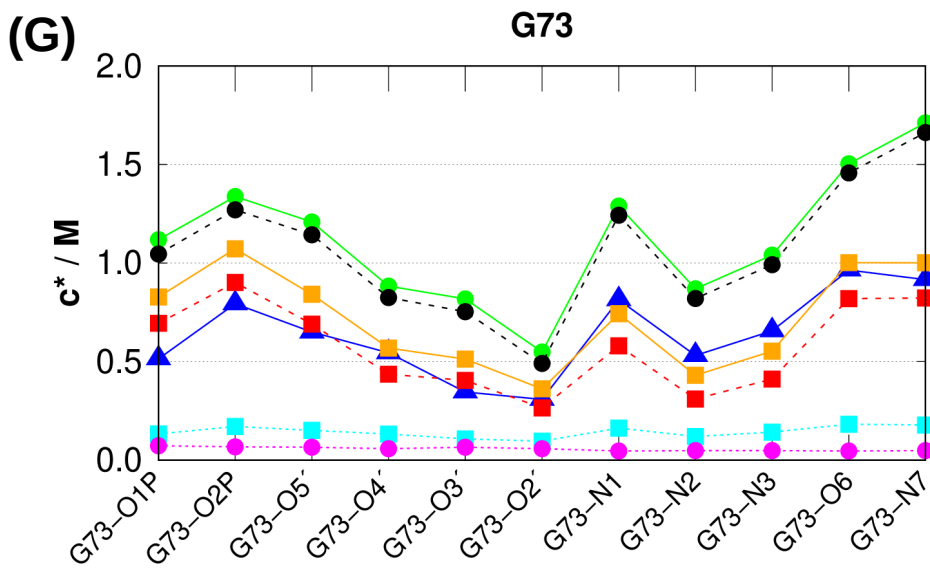
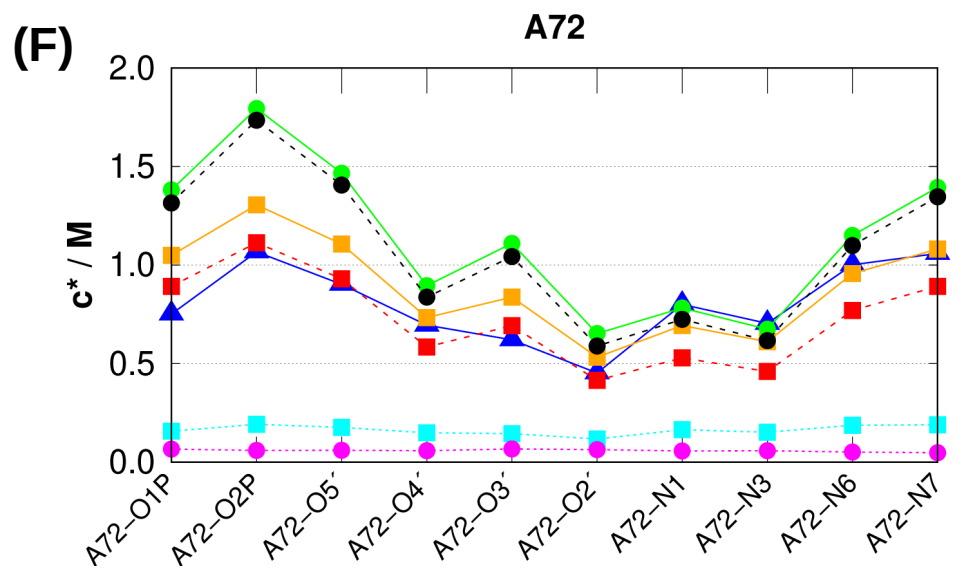
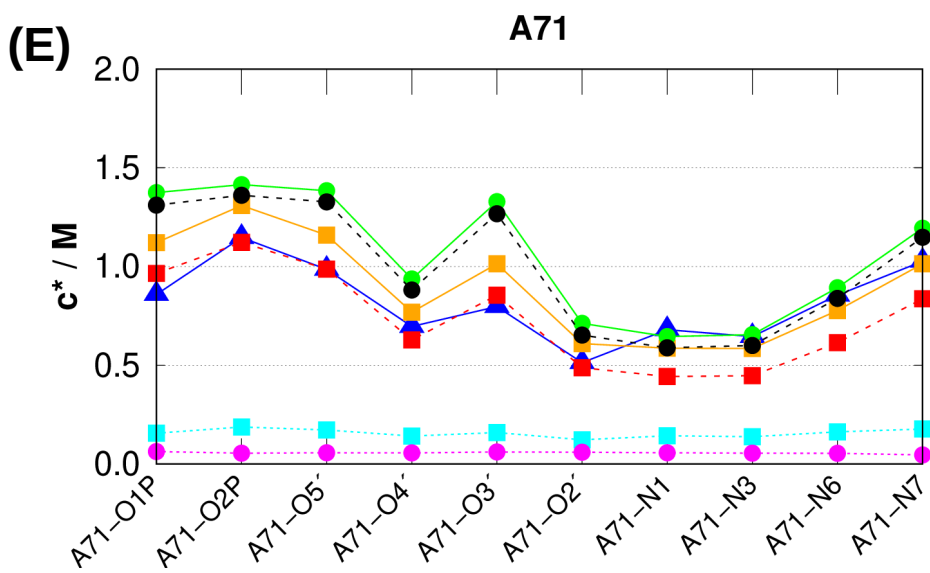
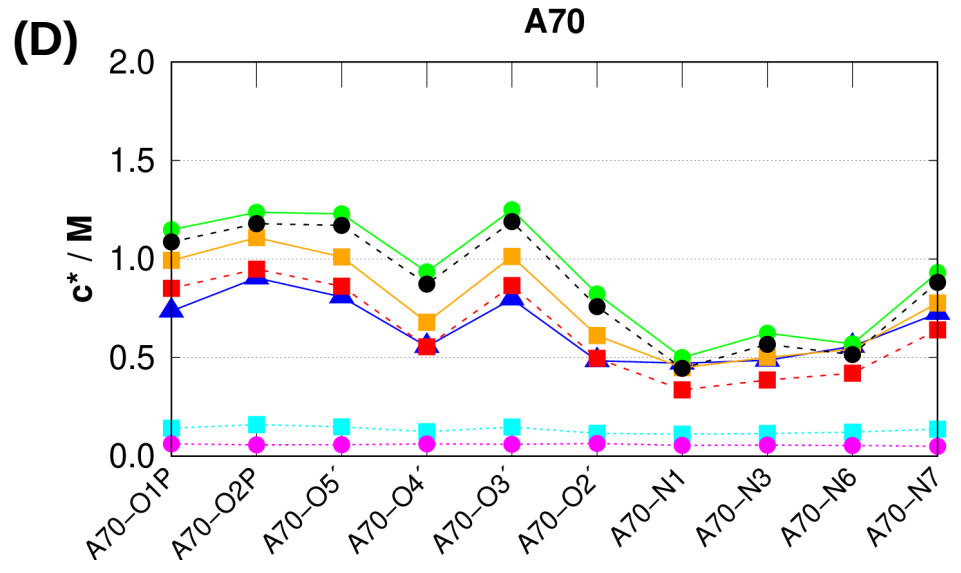
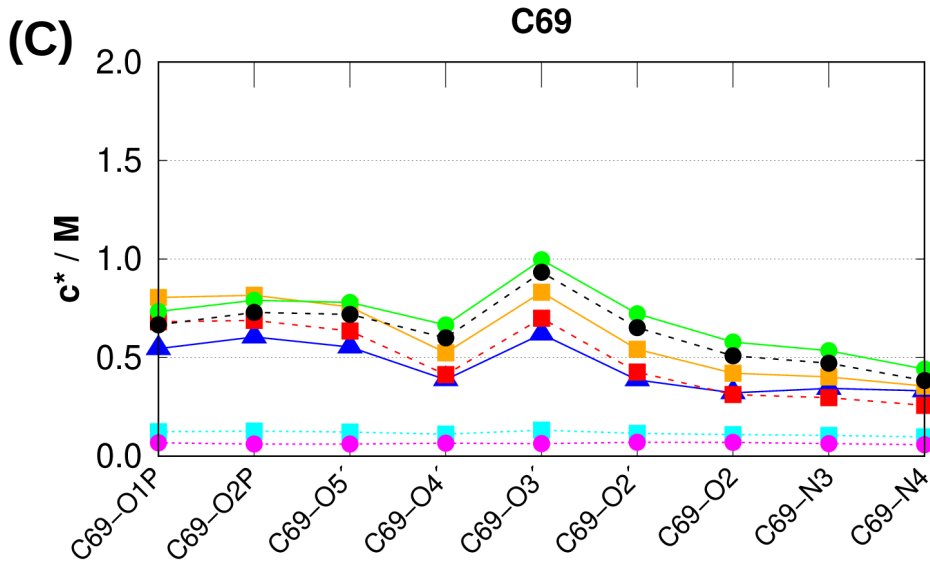
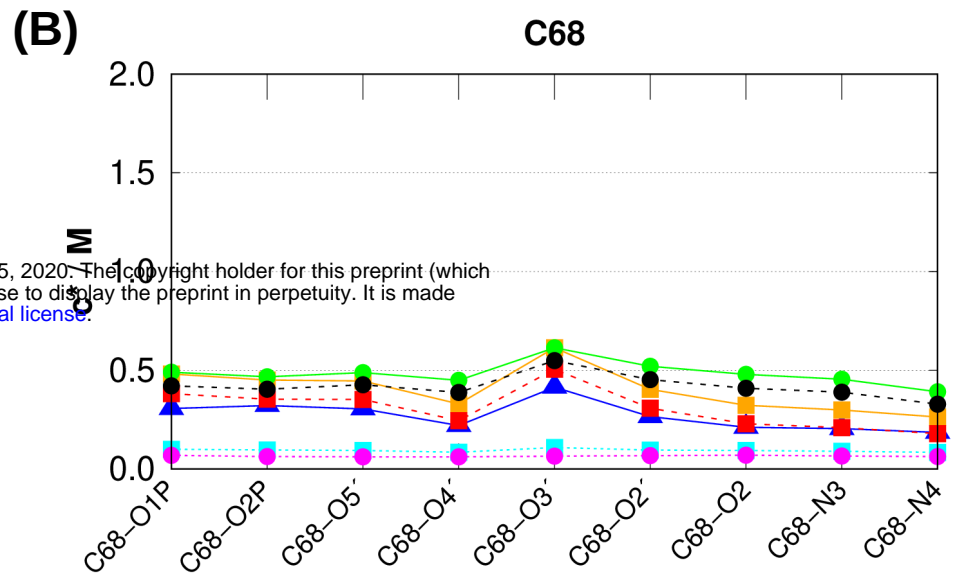
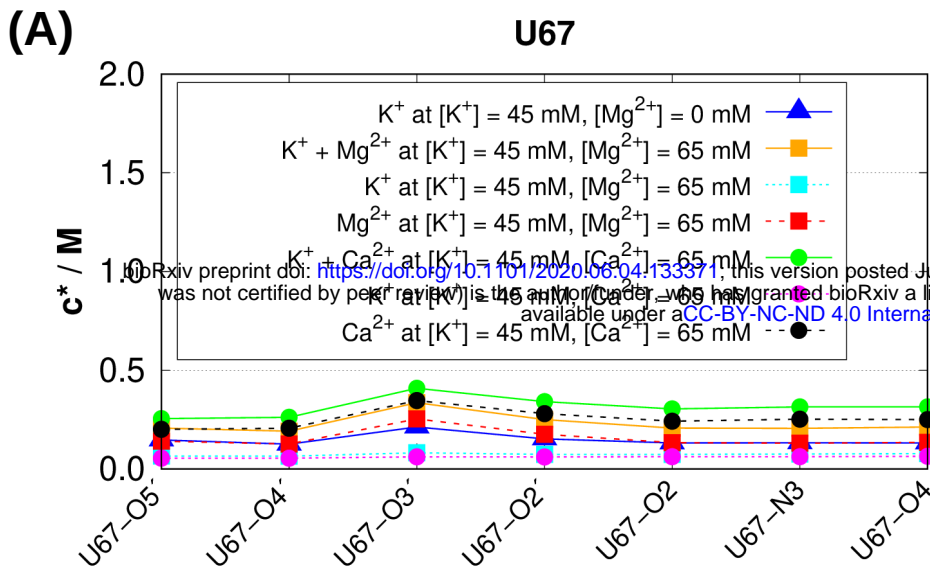
- (117) Stombaugh, J.; Zirbel, C. L.; Westhof, E.; Leontis, N. B. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* **2009**, *37*, 2294–2312.
- (118) Lorenz, R.; Wolfinger, M. T.; Tanzer, A.; Hofacker, I. L. Predicting RNA secondary structures from sequence and probing data. *Methods* **2016**, *103*, 86–98.
- (119) Miao, Z.; Westhof, E. RNA structure: Advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.* **2017**, *46*, 483–503.
- (120) Fiore, J. L.; Nesbitt, D. J. An RNA folding motif: GNRA tetraloop–receptor interactions. *Quart. Rev. Biophys.* **2013**, *46*, 223–264.
- (121) Hodak, J. H.; Fiore, J. L.; Nesbitt, D. J.; Downey, C. D.; Pardi, A. Docking kinetics and equilibrium of a GAAA tetraloop–receptor motif probed by single-molecule FRET. *Proc. Nat. Acad. Sci.* **2005**, *102*, 10505–10510.
- (122) Fiore, J. L.; Hodak, J. H.; Piestert, O.; Downey, C. D.; Nesbitt, D. J. Monovalent and divalent promoted GAAA tetraloop–receptor tertiary interactions from freely diffusing single-molecule studies. *Biophys. J.* **2008**, *95*, 3892–3905.
- (123) Butcher, S. E.; Dieckmann, T.; Feigon, J. Solution structure of a GAAA tetraloop receptor RNA. *EMBO J.* **1997**, *16*, 7490–7499.

Graphical TOC Entry

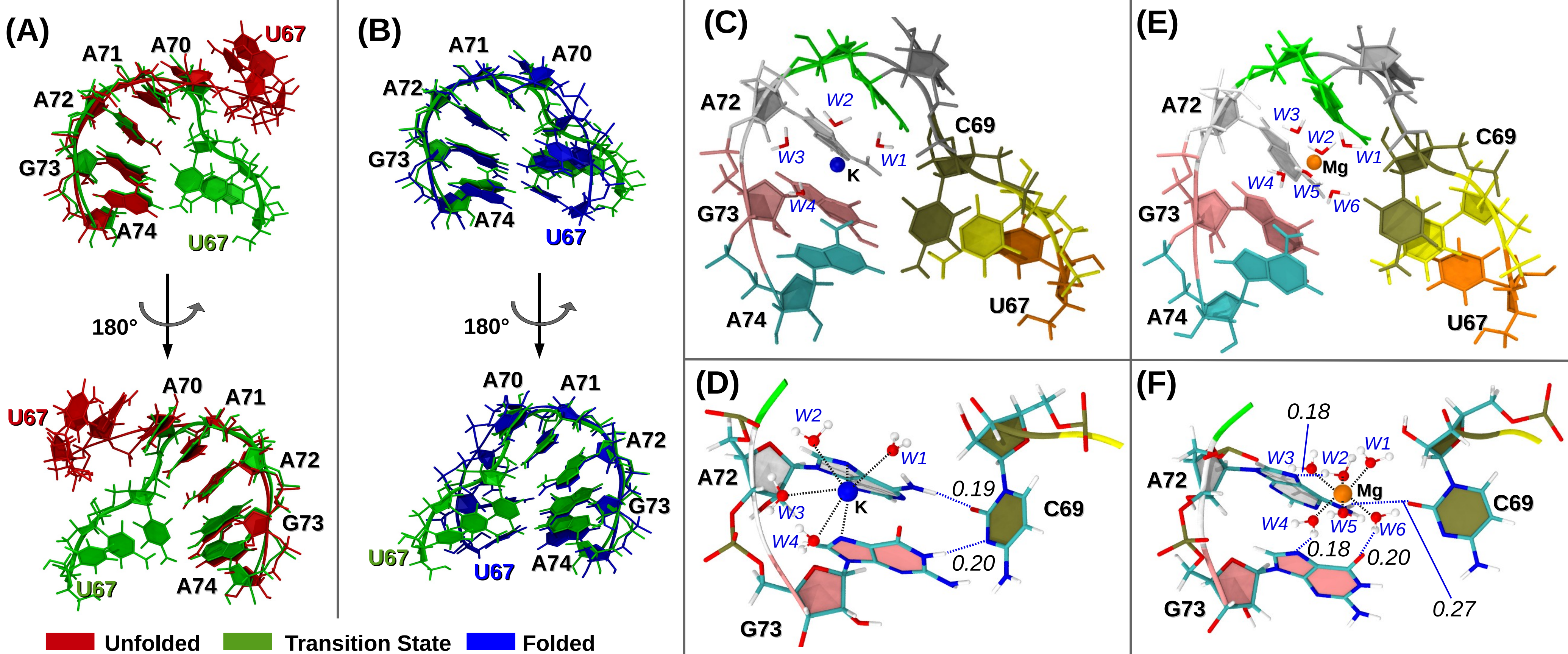


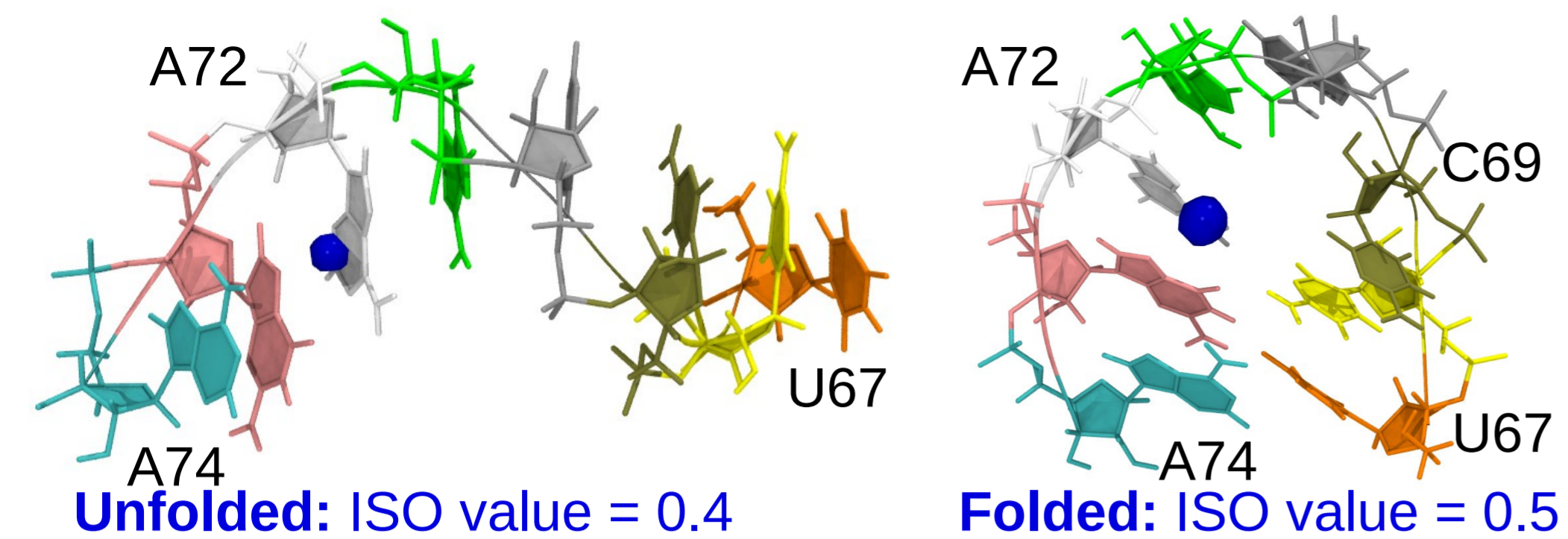
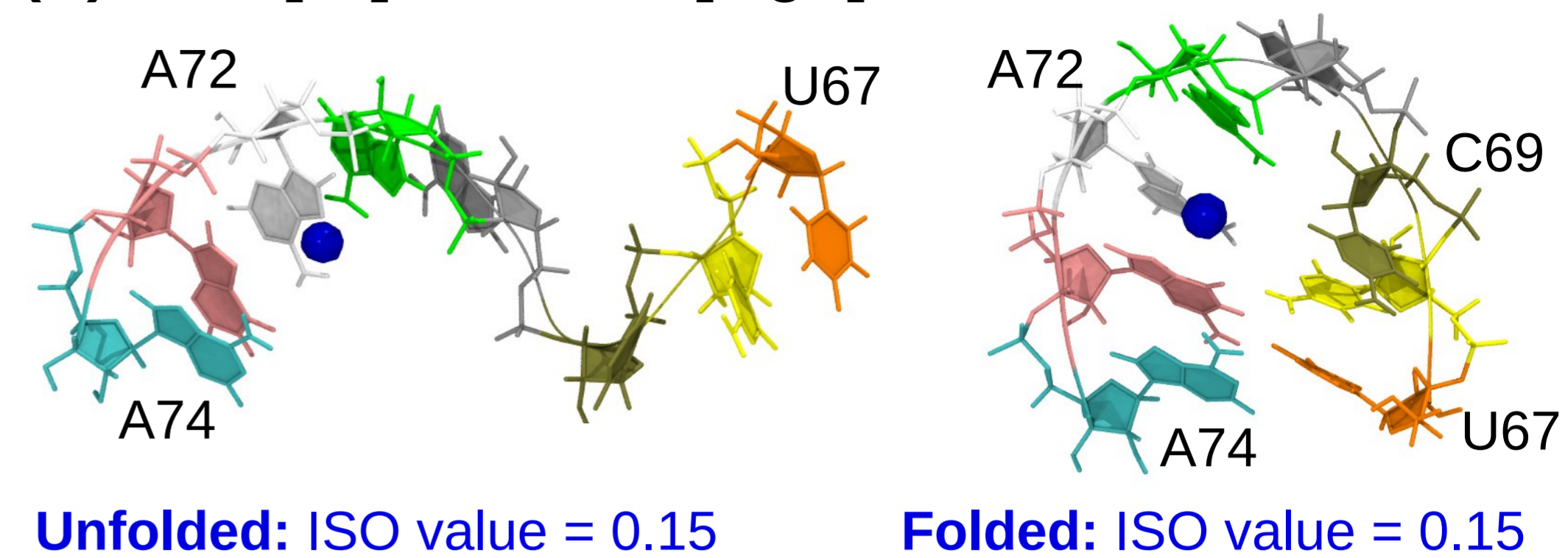
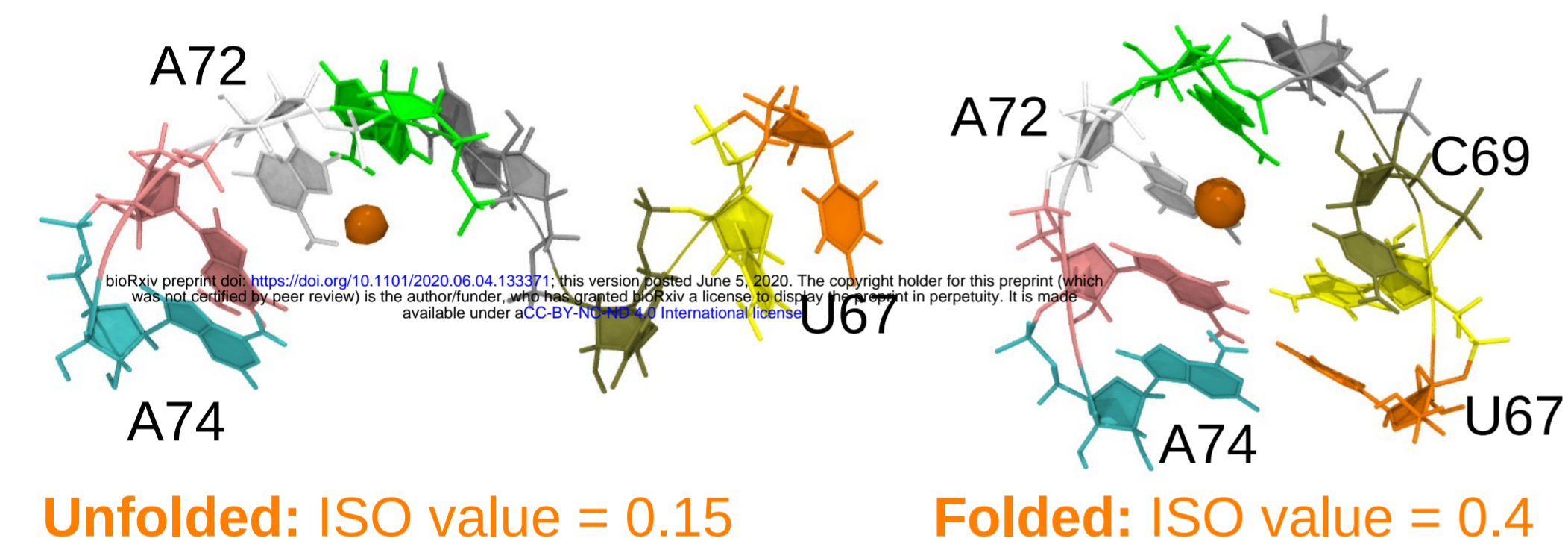
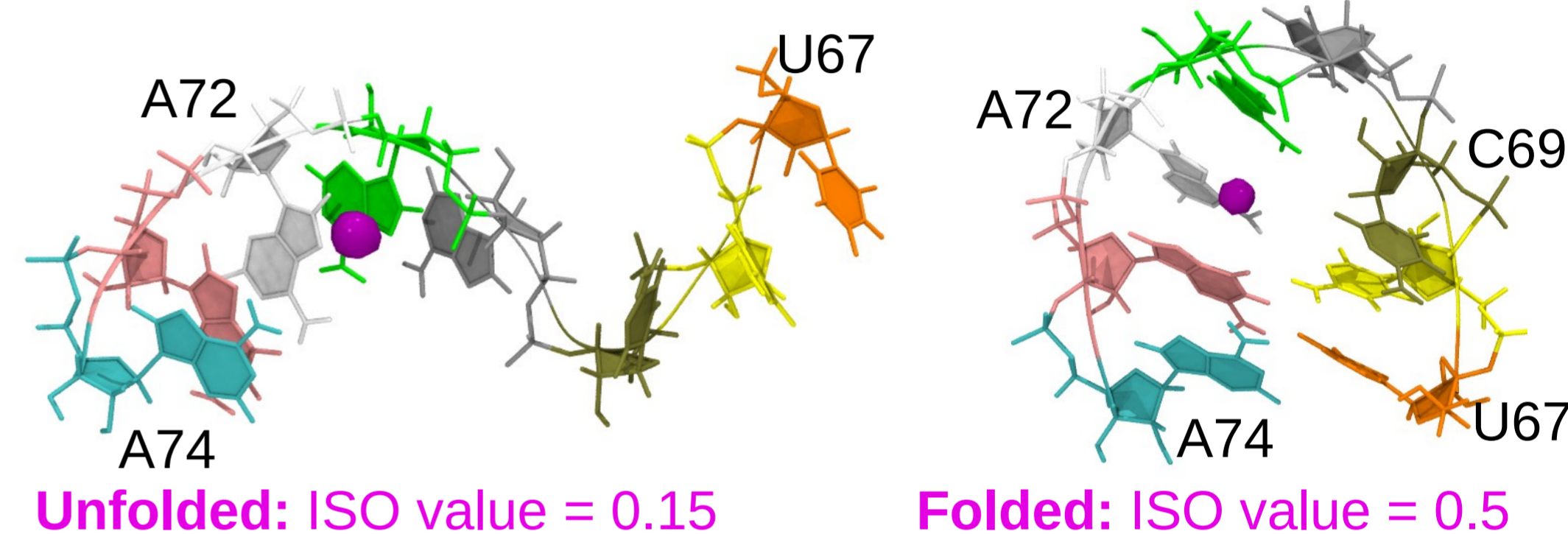
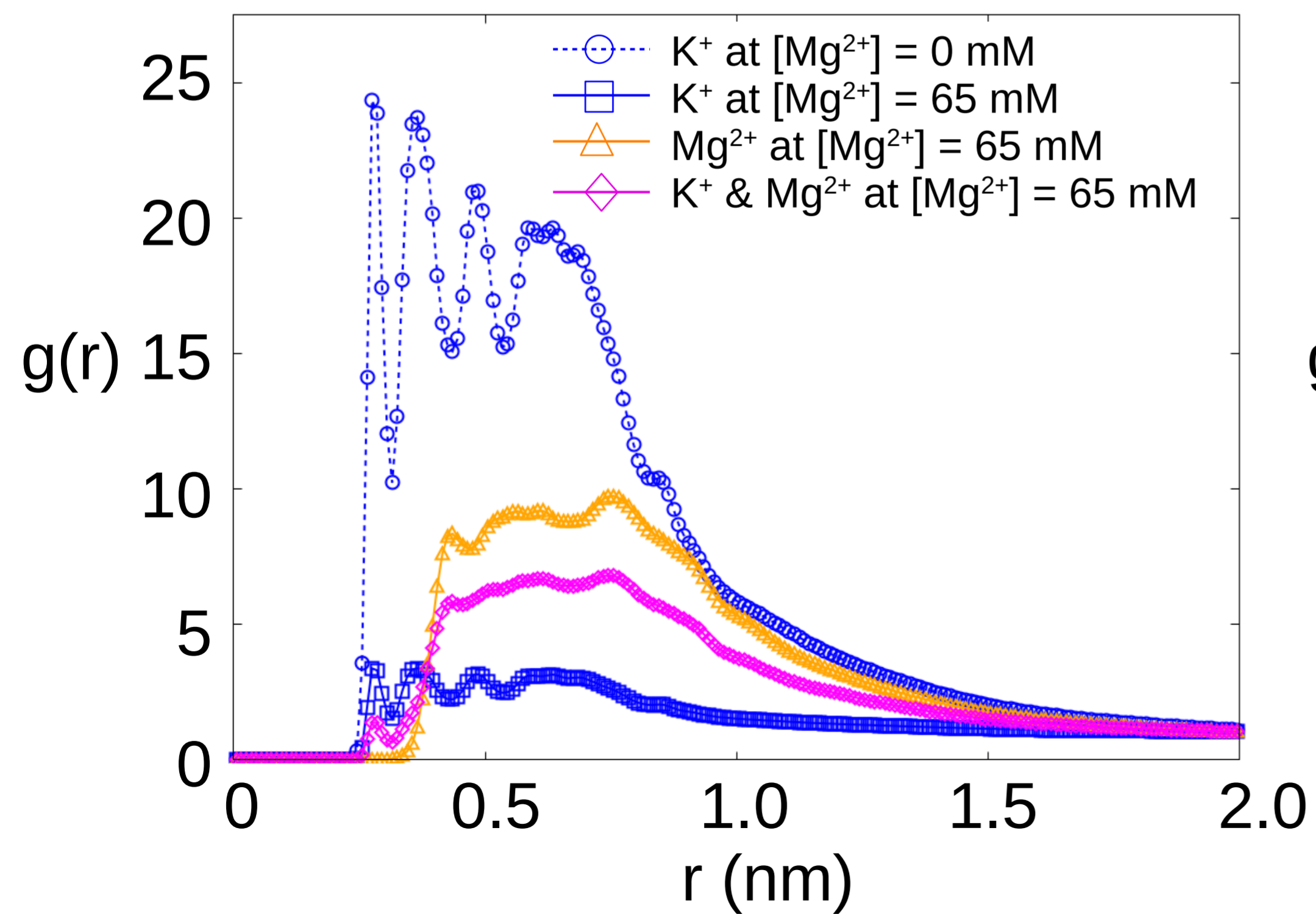


(A) $[K^+] = 45 \text{ mM}$, $[Mg^{2+}] = 0 \text{ mM}$ **(B)** **without divalent ions****(C)** $[K^+] = 45 \text{ mM}$, $[Mg^{2+}] = 65 \text{ mM}$ **(D)** $[K^+] = 45 \text{ mM}$, $[Mg^{2+}] = 65 \text{ mM}$ **(E)** $[K^+] = 45 \text{ mM}$, $[Mg^{2+}] = 65 \text{ mM}$ **(F)** $[K^+] = 45 \text{ mM}$, $[Ca^{2+}] = 65 \text{ mM}$ **(G)** $[K^+] = 45 \text{ mM}$, $[Ca^{2+}] = 65 \text{ mM}$ **(H)** $[K^+] = 45 \text{ mM}$, $[Ca^{2+}] = 65 \text{ mM}$ 



bioRxiv preprint doi: <https://doi.org/10.1101/2020.06.04.133371>; this version posted June 5, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



(A) K⁺ in [K⁺] = 45 mM, [Mg²⁺] = 0 mM**(B) K⁺ in [K⁺] = 45 mM, [Mg²⁺] = 65 mM****(C) Mg²⁺ in [K⁺] = 45 mM, [Mg²⁺] = 65 mM****(D) K⁺ & Mg²⁺ both in [K⁺] = 45 mM, [Mg²⁺] = 65 mM****(E)****(F)**