Molecular insights into the Stereospecificity of Arginine in RNA tetraloop folding

Amal Vijay and Arnab Mukherjee*

Department of Chemistry, Indian Institute of Science Education and Research, Pune-411008, India.

ABSTRACT: One of the possible hypotheses for the homochirality of amino acids in the context of the origin of life is that only a particular stereoisomer provides preferential stability to RNA folding by acting as a chemical chaperon. This study probes into the molecular understanding of such preferential stability for a small GAAA RNA tetraloop in the presence of chiral arginine amino acids using a combination of umbrella sampling and parallel bias metadynamics involving five collective variables to tackle the multi-dimensional free energy landscape for faster, better, and more efficient estimation with controlled sampling. Our results show that the free energetic stability of RNA differs significantly in the presence of D- and L-arginine, giving rise to different unfolding rates. Interestingly, the folding rates are not altered. We show that the origin of the chirality difference in RNA folding–unfolding dynamics is due to the differences in configurational diversity of RNA by adopting different unnatural conformations accompanied by different binding modes of D-arginine, and L-arginine towards the given RNA motif.

INTRODUCTION

The essential molecules of life are mostly found in one specific chirality. While proteins are made mostly of L-amino acids, nucleic acids often contain D-ribose sugars.[1-2] Many of the hypothesis that address this predominance stems from the theories of the origin of life.[3-5] One such hypothesis states that in the RNA world, preceding the modern "protein world," chiral amino acids worked as chaperons for RNA folding and functions.[6-8] It is also assumed that in the evolution of amino acid and protein-based biological machinery that succeeded RNA world, the interaction of amino acids and RNA motifs could also play a role in the above chiral predominance.[8-9]

RNA tetraloop hairpins are a special type of tree-dimensional motifs where the majority of bases in an RNA structure belong to the paired regions (Watson – Crick base pairing between purine and pyrimidine bases in a given sequence) forming the stem part of RNA while the unpaired region having a specific sequence forms the loop part of the RNA.[10] The stability, similarity, structural features and probable interactions between different hairpin structures are studied and compared using various experimental techniques such as NMR and various other absorption spectroscopic studies.[11-15] The significant feature of these types of motifs involves spontaneous transitions between native folded states to form various misfolded and elongated single-stranded structures.[16-19] The role and dynamics of these structural motifs in various recognition processes are extensively evaluated by various experimental methods for understanding various transition states and RNA folding pathways.[10, 20-25] The recent FRET experimental studies[26] on GAAA tetraloop structures show that the kinetics and thermodynamics of events associated with folding–unfolding equilibria is highly influenced by various complexities of amino acids.[27-28] The arguments based on theories related to chiral amino acid-based origin of life are validated to an extent by various studies that specify that the energetics and kinetics are altered in the presence of amino acids from the secondary to tertiary level of interactions in RNA motifs[27].

Along with experimental studies, the mechanistic aspects of folding–unfolding events in various tetraloop hairpin RNA motifs received significant attention by several computational methods in recent years.[16-19, 29-30] The molecular dynamics simulation with various enhanced sampling techniques show promising results in understanding the key intermediates and transition events involved in these processes.[16-18, 29, 31-38]

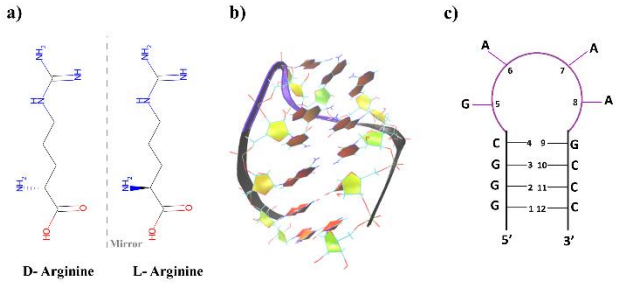

Figure 1. (a) Chemical structures of D- and L-arginines. (b) Structural representation of GAAA tetraloop. (c) Cartoon representation of the RNA tetraloop structure. The stem part is highlighted with black color and the loop part is highlighted with violet color.

It is evident from the literature that the many physical properties of biomolecules including various RNA motifs show marginal difference in the presence of chiral small molecules.[39-42] Even though the dynamics of various RNA tetraloop motifs are studied extensively using molecular dynamics simulations, the impact of chirality-dependent folding - unfolding studies using computational studies are not well explored. The molecular driving force and mechanistic aspects causing the specific effect of enantiomers on folding – unfolding events remain as a mystery. Molecular-level understanding of these processes is necessary to comprehend and solve many such scientific ambiguities related to enantiospecific dynamics of RNA motifs. These challenging aspects of chirality-dependent folding-unfolding problems and experimental evidence motivated us to provide molecular insights into the folding – unfolding of GAAA tetraloop in the presence D-arginine and L-arginine

using multidimensional free energy landscape aided by enhanced sampling techniques. We hope that our discussions on chiral dependencies on RNA dynamics will be beneficial for further small molecule dependent enantiospecific studies of various biomolecules and comprehend chiral amino acid-based theories on origin of life.

The present study aims to gain a molecular understanding of the preference of a particular chirality in amino acids on RNA tetraloop folding. For that, we constructed here multi-dimensional free energy landscapes of folding-unfolding equilibria of GAAA tetraloop (with sequence 5'-gggcGAAAgccc-3') in presence of D- and L-arginine. The choice of arginine amongst other amino acids is due to its diverse binding capabilities with RNA. Due to the presence of double-sided polar groups and long hydrophobic tail, arginine can bind RNA in the groove as well as stack with the base pairs with specific binding affinities.[43-45] It is to be noted that the binding affinity and stereospecificity of amino acids towards RNA motifs is highly dependent on the choice of interest of RNA motifs.

We have studied the chiral arginine dependent folding – unfolding landscape of GAAA tetraloop using multiple collective variables (CV). The CVs (discussed later) were defined in such a way that to understand the conformational flexibility of RNA in folded, misfolded, and unfolded states. We have used a combination of umbrella sampling[46] and parallel bias metadynamics[47] for our further simulations for controlled exploration of phase space related to RNA dynamics. Our results indicate that the misfolded states of RNA have equal free energy stabilities in presence of for both the stereoisomers, while the native folded state get much more stability in the presence of D – arginine. Further our single molecule binding studies targeting the native folded state of RNA revealed that the origin of extra folded stability difference in these stereoisomers are originated due to different binding modes and orientational effects caused by these molecules towards RNA motif.

METHODS

System setup. We took the initial structure of the tetraloop from the protein data bank (PDB ID – 1ZIF) and subsequently edited it to get the target 12-mer sequence of 5′-gggcGAAAgccc-3′. We have used ff99bsc0χOL$_5$ forcefield[48] for the RNA and previously reported amber based forcefield[49] for amino acids. The tetraloop structure was inserted into a cubical box of dimension 9 X 9 X 9 nm^3 and solvated using the TIP3P[50] water model and maintained the target concentration of 300 mM for the systems containing amino acids. For neutralizing the charge of the system, we used Na^+ and Cl^- ions. A schematic representation of the system is shown in Fig. 1

Equilibration and Simulation. Initially, each system was energy minimized using the steepest descent method[51] for 10000 steps, followed by heating it to 300 K in 200 ps using Berendsen thermostat and barostat[52] with a coupling constant of 0.6 ps for each. Restraints of 25 kcal/mol/Å^2 were applied on heavy atoms during the heating process. Thereafter, equilibration was carried out for 2 ns at constant temperature (300 K) and pressure (1 bar) without any restraints using the same thermostat and barostat with coupling constants of 0.2 ps each. Finally, 10 ns unrestrained NVT equilibrations were carried out using the Nosé-Hoover thermostat[53] with a coupling constant of 0.2 ps. During the simulation, the LINCS algorithm[54] was used to constrain all the bonds, and Particle Mesh Ewald (PME) method[55] was used for electrostatics. The distance cutoffs for the van der Waals (vdw) and long-range electrostatic interaction were kept at 10 Å. The time step for each simulation was 2 fs. All the molecular dynamics simulations and free energy calculations were carried out using GROMACS 2019.6[56-57] patched with plumed 2.6[58-59].

Design of the study. To understand the effect of chiral amino acids on RNA tetraloop stability, we have performed two different types of calculations – (a) free energy landscape of RNA folding and unfolding equilibrium in the presence of D- and L-arginine separately, (ii) free energy of binding of D- and L-amino acids separately. These two are discussed below:

(i) Construction of RNA folding-unfolding free energy landscape. The equilibrated structure in the production run was taken as a starting point for the construction of folding-unfolding free energy landscape.

We initially constructed a system containing RNA in the absence of any amino acids and considered well-tempered metadynamics[60] for achieving the proper sampling of the conformational space and understanding the probable minima states ranging from the native folded state to elongated unfolded state.

At first, we used the radius of gyration of RNA and native hydrogen bonds in the stem region of the RNA motif for well-tempered metadynamics simulation (section 1 and Figs. S1-S3 in the Supporting information (SI)). However, we identified that the given choice of collective variables failed to distinguish similar states specifically in the loop region, which consists of the sugar-backbone hydrogen bonds (Fig. S4 of SI). Reweighting the free energy surface with respect to ermsd showed encouraging results (Fig. S5 of SI). We then incorporated, modified, and used various combinations of CVs (Figs. S5) for well-tempered metadynamics limited to a maximum of two per combination to tackle the problems faced in defining the initially constructed reaction coordinate setup. Details of these simulations are provided in the SI.

However, we have observed the following discrepancies in the free energy estimates based on our micro-second time scale simulations. i) The free energy values and sampling are dependent on the choice of collective variables for biasing. ii) The system tries to exist in trapped false minima states, fails to provide converged free energy surfaces (Figs S6, S7 of SI). The main reason for this is poor choice of collective variables in the complex multidimensional landscape. This leads to improperly oversampling of the trapped state resulting in a high barrier for transitions to other states. The previous studies on similar RNA motifs also encountered similar issues especially related to a lack of convergence with respect to the choice of collective variables, technique, and force field parameters.[35] The understanding from our unreported well-tempered metadynamics simulations and the need for using multiple collective variables for chasing the multi-dimensional free energy landscape directed us to use a combination of two enhanced sampling techniques, umbrella sampling and parallel bias metadynamics.

The parallel bias metadynamics makes use of enhanced sampling along multiple CVs by applying low dimensional bias potential across multiple CVs simultaneously. This method was found to be very efficient for revealing the free energy estimates associated with various biological & chemical systems having high energy barriers.[47, 61-64]. While the umbrella sampling merits the control over a given CV to model the conformational transitions aided by a series of independent windows.[46, 65]

The advantage of using these combinations of two enhanced sampling methods is as follows. By applying the umbrella sampling potential over a given collective variable, we will have control & directionality over the reaction coordinate. Further, the minima states in each umbrella window can be well explored by using parallel bias metadynamics simulations, providing details about all probable local structural changes in each slice of a given umbrella window. So, the method takes advantage of metadynamics simulations and is expected to give an overview of relevant free energy minima states by controlled exploration of phase space.

We have used five CVs for the controlled sampling approach. Among the CVs used, the center of the mass distance between the stem regions was used as an umbrella sampling window coordinate. Further, in each window, the sampling is accelerated by parallel bias metadynamics by the other four collective variables (subsidiary CVs): radius of gyration, hydrogen bonds, ermsd and contacts. The protocol was carried out for each specific system of amino acids containing D- and L- stereoisomers. A potential harmonic restraint of 200 kJ/mol/nm^2 was applied to the umbrella sampling coordinate to make sure that the value of the reaction coordinate fluctuates around the successive positions where potential is applied. Windows were constructed from folded state to unfolded state from the umbrella sampling coordinate value range from 0.8 nm (folded state) to 4.6 nm (unfolded state) with a difference of 0.2 nm between the windows. For the parallel bias metadynamics, a bias factor of 10 and an initial hill height of 0.5 kJ/mol were used. Further, the Gaussian widths of 0.1, 0.4, 1, and 0.05 were used respectively for ermsd, hydrogen bonds, contacts, and radius of gyration, respectively. A total of 1 μs simulation was carried out per system by constructing 20 windows across the umbrella sampling coordinate and by simulating each window for 50 ns.

Finally, the free energy surface was constructed by combining the bias contribution from the umbrella sampling potential and the potential from the parallel bias metadynamics from each window using the weighted histogram analysis method (WHAM)[66]. The results are interpreted against the ermsd parameter and stem distance (umbrella sampling coordinate). The ermsd parameter is well suited to explain various interactive folded states, base pairing schemes, and base stacking schemes in the literature[30, 67].

(ii) Binding free energy of a single amino acid. The most favorable conformation of amino acid closest to the terminal residue of RNA from the cluster was taken for binding studies towards single molecular binding studies between RNA and D\L Arginine. We have used the center of the mass distance between the stem region of RNA and arginine as the biased reaction coordinate, referred to as binding distance.

(iii) Choice of collective variables. To facilitate the free energy calculations for both studies above, we have defined the collective variables based on an intuitive way of understanding the process and also based on reported studies[67] on similar RNA motifs, as these choices are heuristic. Since the goal of the present study is to understand the folding-unfolding thermodynamics in the presence of amino acids, our choice was governed by the need to observe the conformational change in the RNA (e.g., the stem distance, radius of gyration, contacts, ermsd[67]). Also, we studied binding free energy of a particular stereoisomer to the folded RNA using another set of CVs (binding distance, angvec). The CVs are discussed in more detail below.

a. Radius of gyration (R_g) of RNA. The radius of gyration was defined by,

$$R_g = \left(\frac{\sum_i^n m_i |r_i - r_{COM}|^2}{\sum_i^n m_i} \right)^{1/2}$$

and the position of the center of mass r_{COM} is defined by

$$r_{COM} = \frac{\sum_i^n r_i \, m_i}{\sum_i^n m_i}$$

The center of mass of heavy atoms was taken into consideration for measuring the radius of gyration. Since folded and unfolded RNA would have different overall sizes, this coordinate helps to distinguish these states.

b. Stem Distance. The RNA hairpin loop has specific hydrogen bonding pattern between two stem regions. Therefore, the center of the mass distance between the two stem regions helps monitoring the relative motions of the stem regions with respect to the loop part of the RNA. For the first stem, the part center of mass of the backbone and bases of residues 2, 3, and 4 were used. For the other stem region center of mass of the backbone and base of residues 9, 10, and 11 were considered.

c. Hydrogen bonds (HB). The number of hydrogen bonds between the purine and pyrimidine bases in two stem regions is defined by the switching function as follows.

$$HB = \sum_i \frac{1 - \left({}^{r_i}/_{r_0}\right)^n}{1 - \left({}^{r_i}/_{r_0}\right)^m}$$

The cutoff distance r_0 was set to 2.5 Å. The values for n and m parameters were set to 6 and 12, respectively. The index value for the value of *i* is chosen in such a way that the RNA forms a total of 10 hydrogen bonds (9 corresponds to hydrogen bonds between complementary bases formed by three base pairs in the stem region, and the rest is formed by the hydrogen bond in the loop structure.

To account for the relative loop motion, we also defined a hydrogen bond parameter between the phosphate and amide between the bases G (residue) and A (residue) at 5 and 8 positions in the loop structure (Fig. 1c). The switching function used here ensures that the calculated hydrogen bond values are continuous derivatives.

d. End-to-end distance. End-to-end distance is defined between the two terminal C1 atoms between the first and last residues in RNA. End-to-end distance will be able to comment on the extent of elongated structures formed by the RNA in various folding–unfolding events.

e. ermsd. ermsd[68] is a metric developed to measure the distances specifically for three-dimensional structures of nucleic acids. The usual RMSD considers only the relative positions, while the ermsd incorporates both the relative positions and orientational effects caused by the nucleobases in the RNA. A cutoff of 3.2 was used for ermsd calculation.

f. Contacts. The number of contacts is defined by the heavy atoms of the purine and pyrimidine bases between the two stem regions of the RNA. The benefit of defining contacts over hydrogen bonds is that by defining contacts, all possible orientations of bases will be incorporated in the possible ways giving possible contacts formed by the stem regions. The hydrogen bonds only give clear information about native structures. The contacts are defined by,

$$N_c = \sum_{i \in A} \sum_{i \in B} s_{ij}$$

$$s_{ij} = \frac{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^n}{1 - \left(\frac{r_{ij} - d_0}{r_0}\right)^m}$$

where $s_{ij} = 1$ if the contacts between the atoms i and j is formed between the independent stem units of RNA.

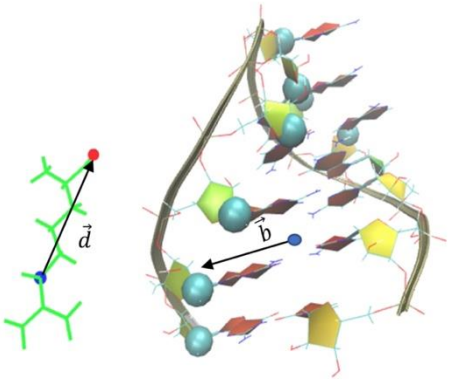

Figure 2. Representation of angvec collective variable for binding studies.

f. angvec. The angvec collective variable is defined as the angle between two vectors $\vec{b}$ and $\vec{d}$, where $\vec{b}$ is defined as the vector connecting center of mass of backbone and sugar region of residues 9, 10, 11, and 12 towards residues 1,2,3 and 4. $\vec{d}$ is defined as the vector connecting a nitrogen atom of the Guanidium end and the oxygen group of the c-alpha end of Arginine.

$angvec = \cos^{-1}[\vec{b}.\vec{d}/(|\vec{b}||\vec{d}|)]$. This CV was used in our previous studies of DNA intercalation.[69-71]

RESULTS & DISCUSSION

a) Free energy landscape of GAGA tetraloop in the presence of L-arginine and D-arginine. Although we used five collective variables (CV), we have chosen two of them, the stem distance and ermsd parameter, to discuss and understand the free energy landscape associated with the folding–unfolding equilibrium of GAAA tetraloop in the presence of L-arginine (Fig. 3a) and D-arginine (Fig. 3b). While stem distance measures the extent of unfolding, ermsd captures various native and non-native base pairing and stacking geometries. Based on the free energy values, as indicated by the color bar, we have identified three broad regions: folded state (stem distance < 1.8 nm and ermsd < 2.0 nm), elongated unfolded states (stem distance > 4.0 nm and ermsd > 2.0 nm), and the misfolded states (1.8≥ stem distance≤ 4.0 and ermsd > 2.0 nm) and denoted them by F, U, and M, respectively. The minimum free energy path connecting the folded state to the unfolded state through the misfolded state was calculated using MULE program[72] and is indicated by the black dotted lines in the Figs. 3a and 3b. Also, to get a clear picture of the three states, the free energy profile relative to the unfolded state was plotted against stem distance in Fig. 3c.

Figure 3c shows that GAAA is stable by -21.1 kcal/mol in the presence of D-arginine while its stability is only -12.2 kcal/mol in the presence of L-arginine. The free energy stability of the misfolded state is similar (between -17 – 18.0 kcal/mol) in presence of either of stereoisomers. However, the minimum in presence of L-arginine is flatter compared that that around D-isomer. Note that, the definition of misfolded state is with respect to the structure in presence of D-arginine. In case of L-arginine, the misfolded state forms the global minimum. Therefore, we can say that in presence of L-arginine, GAAA tetraloop does not fold to the hairpin structure at all. This shows the effect of stereospecificity in the given RNA sequence, and further the evidence for chirality-driven tetraloop folding-unfolding equilibria which correlates with the hypothesis on chiral amino acid dependent RNA interactions related to the origin of life.

In Fig. 3d, we have shown some representative structures of the folded (F), misfolded (M) and unfolded (U) states with their corresponding CV values. As expected, the stem distance (along with end-to-end distance and Rg) increases from F→M→U as also evident from Figs 3a-c. However, folded state is unique from the rest with a high number of hydrogen bonds and contacts between the stems of the tetraloop and a low value of ermsd. Both the misfolded and unfolded states have almost no contact and hydrogen bond within the stems. Misfolded states span a wide range of structures (due to broad minimum) differing mostly in terms of stem distance and end-to-end distance. The difference in the misfolded and unfolded state is mostly in the free energy values, which structurally correlates with stem distance, end-to-end distance, and Rg values. Therefore, beyond a certain stem distance, free energy rises fast and that we indicate here as the unfolded state. The free energy surfaces reweighted with respect to stem distance and other subsidiary collective variables are provided in the supplementary information (Fig. S8 of SI).

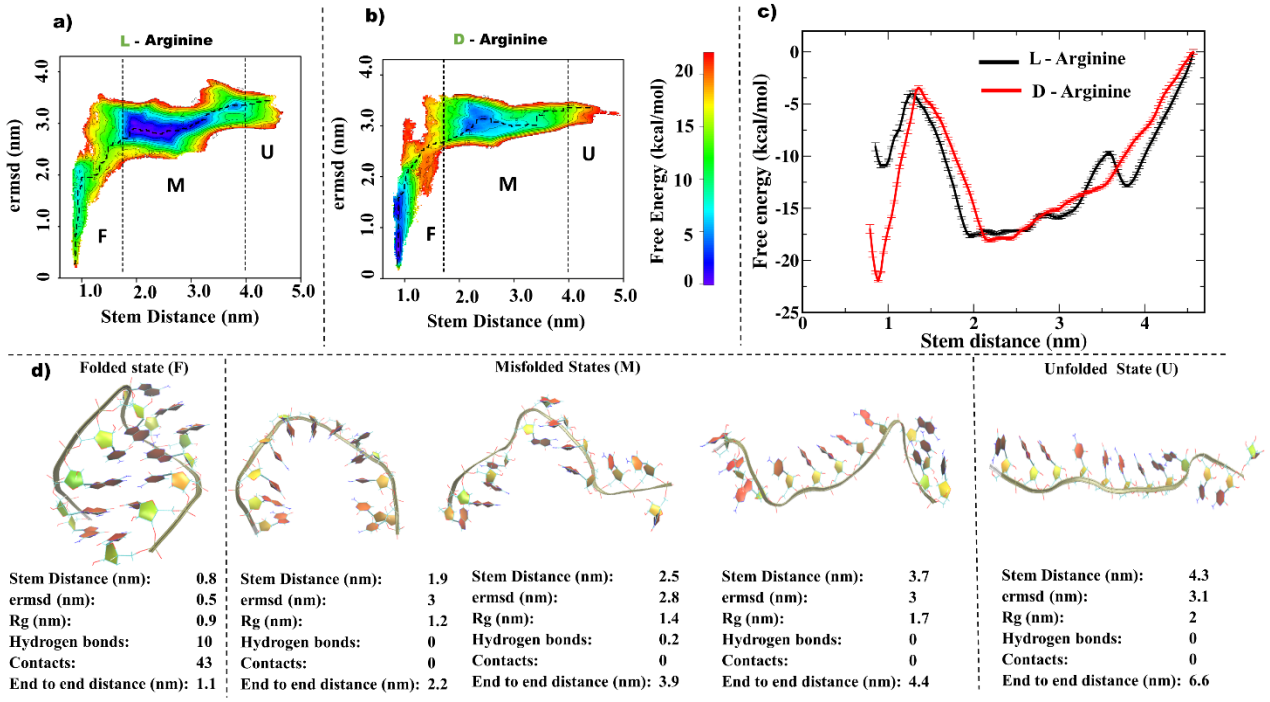

Figure 3. Free energy landscape of RNA folding-unfolding equilibria in the presence of (a) L-arginine and (b) D-arginine projected on ermsd and stem distance c) Comparison of free energy profiles projected only on stem distance for both L- and D-arginines. d) Structural representation for folded (F), misfolded (M) and unfolded states (U) states with corresponding collective variable values.

b) Influence of enantiomers in the folded state. The question that naturally comes is why L-arginine can't stabilize the folded state of GAAA tetraloop. First hypothesis was that interaction energy between RNA and stereoisomers are different. Therefore, we collected all the structures of the folded state (as defined earlier) and calculated the average number of hydrogen bond and average interaction energy between RNA and arginine. To our surprise, we find that both the quantities are same for L- and D-arginine, as shown in Fig. 4, ruling out any favorable interaction played by the D-arginine to stabilize the native folded state of GAAA tetraloop.

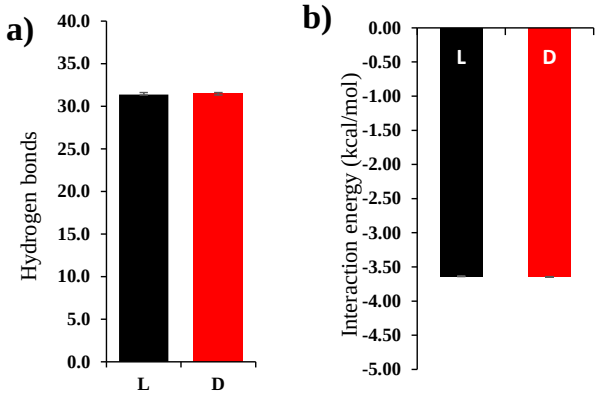

Figure 4. Histogram of the average number of (a) hydrogen bonds formed by RNA and stereoisomers of arginine, and (b) average interaction energy per molecule between RNA and stereoisomers of arginine in the cluster corresponding to the native state.

Therefore, our next hypothesis was that RNA structures may be different in the folded state in presence of D- and L-arginine (although they belong to the similar values of the CV), giving rise to similar interaction energy and hydrogen bond number. The reason for this hypothesis is that the nucleobases in the RNA form different types of interactions with respect to its complimentary/neighboring pair. These pairing schemes can be distinguished by the edge side of interaction forming the H-bonds. The possible edges of interactions are Watson-Crick edge, Hoogsteen/C-H edge and the Sugar edge. These base pairing schemes can be ideally achieved by cis/trans glycosidic rotation around the sugar backbone and the orientation of the H-bond formation between the nuclear bases. By identifying a given base, probable complementary base, and the type of glycosidic rotation, we can identify the ideal orientation of the state of a given residue of RNA. We have used the above-mentioned parameters based on the way portrayed by Leontis for geometric classification of RNA structures.[73-74]

As the stem region is expected to initiate the folding, we investigated the nature of base pairing in RNA and the glycosidic rotation around the sugar backbone for all the four base pairs in the stem region.

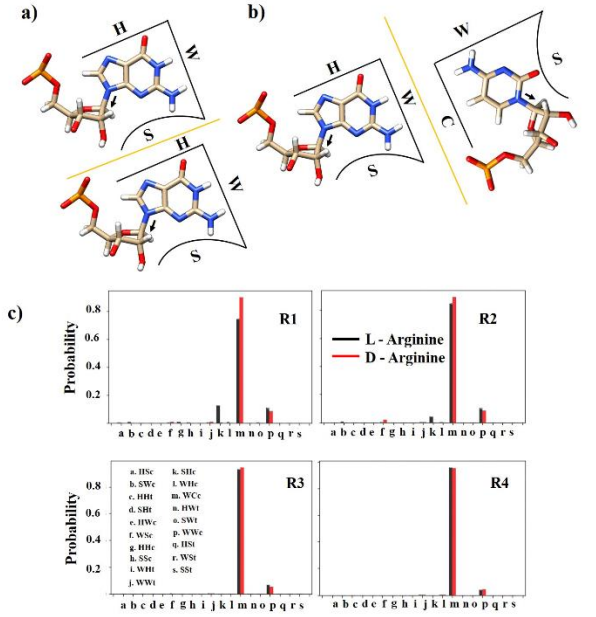

Figure 5. (a), (b) Representation of SHt and WCc states of RNA. The orientation of glycosidic bond is represented by black arrows. Yellow line represents the surface of interaction between the edges. W, H, C, and S represent Watson-Crick, Hoogsteen, C-H and sugar edges, respectively. (c) Probability distribution of different base pair interactions in the stem region of RNA in the native folded state. The three-letter code describing the states are described as per Leontis nomenclature and classification of RNA base pairs. R1, R2, R3 and R4 describes the residue junctions in the stem regions of RNA, starting from the 5' end. Note that for L-arginine, there is significant probability to form SHc structure for the R1 and somewhat R2 base pairs.

A representation of various possible combinations of base pairing schemes is shown in the Fig. S9 of SI. We have calculated the probability of the existence of these base paring orientations in RNA structures using Barnaba[75] program. Each possible complementary interactions between the two bases in the RNA structure is labelled by a three-letter code. The first letter indicates the probable donating edge for base pairing in the 5' end stem region of RNA. The second letter defines the probable acceptor edge of interaction in the complimentary residue/neighboring residue in the RNA. The third letter indicate the possible cis/trans mode of interactions by glycosidic rotation of the sugar backbone. For example, WHc means that 5' end of a given stem residue with Watson-Crick edge is interacting with neighboring Hoogsteen acceptor edge at neighboring end in cis orientation. A representation of SHc and WCc state is shown in the Fig. 5a and Fig. 5b respectively. The details regarding the three-letter nomenclature are discussed in Table S1, and Fig. S10 in SI section III. The probability of existence of the possible states at all four base pairs in the stem region is shown in the Fig. 5c. The R1, R2, R3 and R4 indicate the base pairs in the stem region starting from the terminal end of the RNA. In all these base pairs, most tend to exist in WCc state (Watson-Crick edge, C-H edge, cis orientation). Next is the WWc base pairs, with much less probability. Figure 5 shows that presence of D- or L- influence the geometry of only the terminal base pair (R1). While most of the structures preserve the WCc state in presence of D-arginine, there is a noticeable probability of forming SHc structures in presence of L-arginine. This indicates that the chirality effect is somehow present for the terminal base pair only.

c) Single molecule binding studies. Now that we understand that chirality of the amino acids affects RNA structure in the folded state (especially for the terminal base pair), we wanted to further explore the cause for it. Therefore, we explored specific binding mode of the amino acids to the RNA by calculating the free energy of a single D- and L-arginine with RNA separately. For each system, we took the amino acid closest to R1 as the starting point and performed well-tempered metadynamics simulations as a function of binding distance. Free energy surface for binding plotted with respect binding distance and angvec (see method) is shown in Fig. 6.

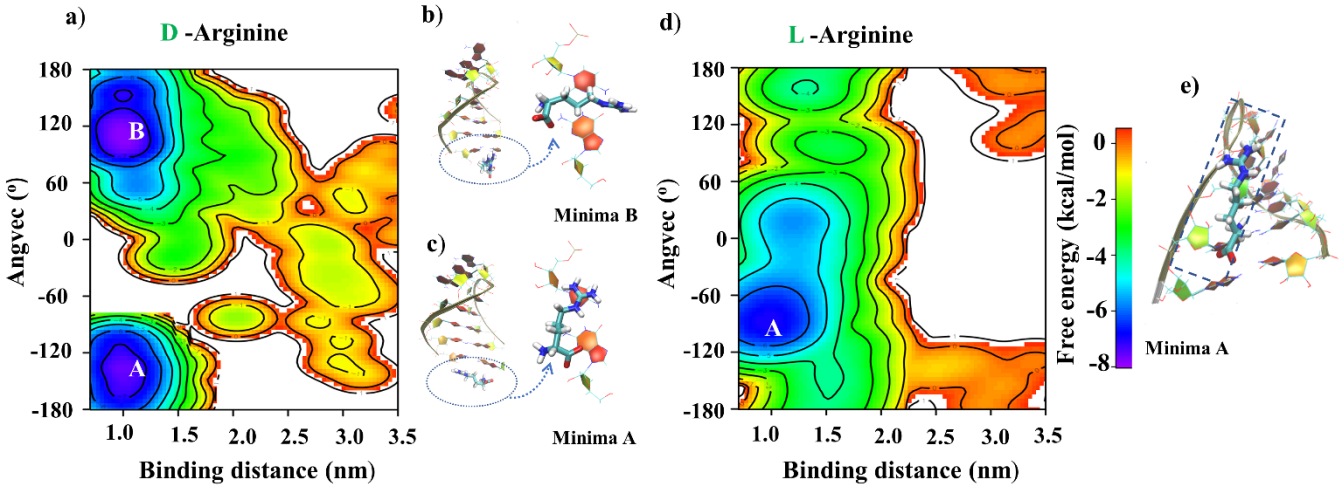

Figure 6. Free energy surface of binding of a single (a) D-arginine and (d) L-arginine with respect to binding distance and angvec. (b), (c) The representative structures at the minima for D-arginine bound to RNA. (e) The representative structure of the minimum for L-arginine bound to RNA.

We have observed two minima of around -7.8 kcal/mol at the bound state for D-arginine (denoted as A and B. in Fig. 6a) that correspond to different values of angvec. The representative structures of the minima are shown in Figs. 6b and 6c. Interestingly, we have observed a single minimum of lesser stability (-6.8 kcal/mol) in the bound state of for in L-

arginine (Fig. 6d). The corresponding minima structure is shown in Fig. 6e.

To analyze the origin of these different interaction behavior, we have probed further into the mode of interaction of arginine by analyzing the bound minima through three structural parts of arginine: (i) C-alpha (ii) alkyl (iii) guanidinium, as shown in the Fig. 7a. From Fig. 6e, we observe that C-alpha part of the L-arginine is close to the sugar-backbone junction of the RNA. This structural orientation provides the most favorable direction of nucleophilic interaction between the hydroxyl sugar part of RNA and the C-alpha region of the arginine. The similar type of observation was previously reported for trans-acylation reaction between the RNA motif and acylated L-amino acids indicating the favorable geometry for the nucleophilic attack between the hydroxyl group of sugar backbone and the carbonyl group of C-alpha terminal of L-amino acids.[76] A representation of this favorable interaction is shown in the Fig. 7b. This representation was in accordance with the observed minima state A for L-arginine as shown in the Fig. 6e. The D-arginine does not form this interaction with the sugar group of RNA, as the steric hindrance caused by the long alkyl chain in D-arginine would prevent this favorable nucleophilic interaction. Fig 7c shows the possible unfavorable mode of interaction in case of D-arginine, thus discouraging it to bind to the sugar-backbone part of the RNA. Indeed, we also have not observed the similar binding mode in our simulations. Further, to understand the preferred orientation of the C-alpha part (containing the stereocenter) for L- and D-arginine in presence of RNA, we have calculated and shown the distribution of dihedral angle θ (O_2, C_2, C_1, N) in Fig. 7d. The distribution of θ is very different for D- and L-arginine indicating that stereocenters indeed play an important role for interaction with RNA. θ peaks at -132° and 36° for D-arginine, while it peaks at -37° and 123° for L-arginine. It is evident from the Fig. 7d that the distribution is different for D-arginine and L-arginine supporting the favorable interaction shown in the Fig. 7b.

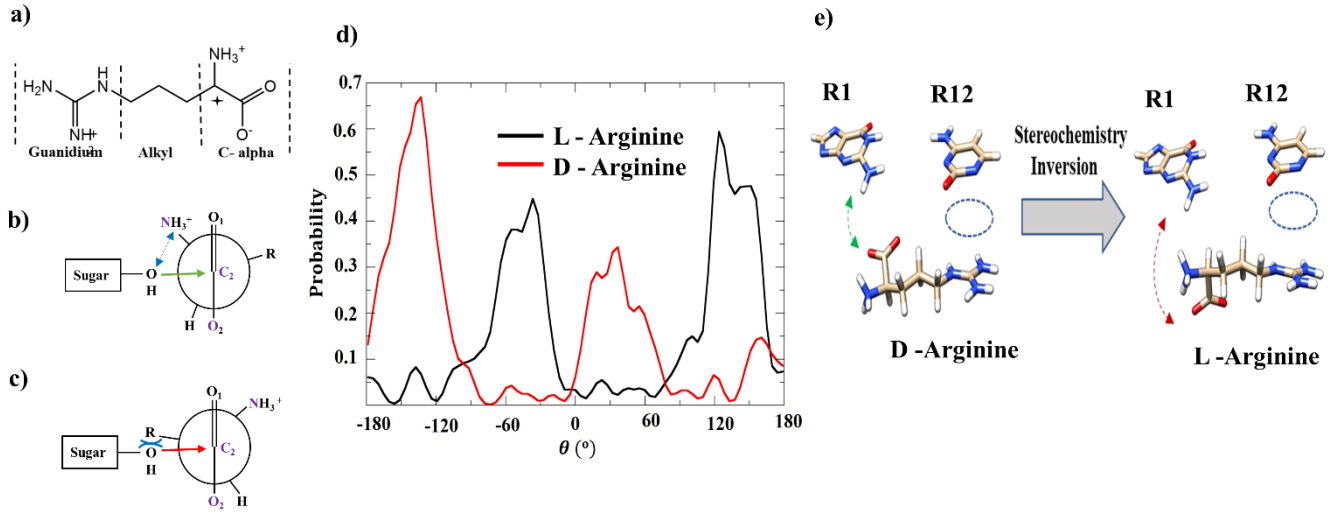

Figure 7. (a) Showing three parts of arginine to analyze the interactions. The blue star represents the stereocenter. Schematic representation of the minima states for (b) favorable nucleophilic interaction of RNA with L-arginine and (c) unfavorable nucleophilic interaction of D-arginine with RNA. (d) Probability distribution of dihedral angle θ (O_2, C_2, C_1, N) in the binding mode of L-arginine and D-arginine. (e) Representative structures of terminal base pair (R1-R12) of RNA and arginine. Stacking interactions are represented by blue dotted circles formed by the guanidinium group in enantiomers and residue 12 (R12) of RNA. The favorable, closest geometric preference between the carboxylate group in D-arginine and residue 1 (R1) of RNA is shown by green arrow. The distant unfavorable mode of interaction between the carboxylate group in L-arginine and R1 base is shown with red arrow.

Table 1. Interaction energy between the terminal residues and arginine enantiomers.

Sl No	System	Interaction energy (kJ/mol)	
		Stacking	Carboxyl - Amino interaction
1	D-arginine	-12.0 ± 0.046	-5.2 ± 0.016
2	L-arginine	-12.2 ± 0.037	-0.6 ± 0.016

In the context of the above discussion, we find that the two minima in the free energy surface of D-arginine binding to RNA correspond to two different types of stabilizing interactions. In the minima A, (Fig. 6c), the guanidium part of the arginine is found to form stacking interactions with terminal Cytosine base (residue 12) at the 3' end of RNA. At the same time the C-alpha part of arginine is found to stabilize the complementary base guanine (residue 1) by forming the stabilizing interactions between the carboxylate group in c-alpha part and amino terminal in the Guanine base part. It is important to note that the similar guanidium-nucleobase interaction was found to

be very relevant in monitoring the specificity of between D-arginine, and L-arginine as discussed in the literature.[44-45]

It is to be noted that the orientation in of D-arginine is aligned in such a way that the carboxylate group faces towards the nucleobase, increasing the binding affinity. We have inverted the stereochemistry at this stereo center to observe what might have prevented the L-arginine to form the similar binding mode. We have observed that the mirror image L-arginine, the carboxylate part of c-alpha group will be facing against the direction of RNA, failing to form the stabilizing interactions at the c-alpha end, as shown in the Fig. 7e. These orientational effect also reflected in the difference in the interaction energies between the terminal residues of RNA and polar ends of the enantiomers, specifically the amino – carboxylate interactions. A comparison of interaction energy profiles is shown in the Table 1, supporting the structural observations. Since the stabilizing stacking interactions of D-arginine (and not for L-arginine) was found with the terminal base pair of the RNA, it was assumed that this interaction would prevent the initiation of unfolding in RNA.

These specific individual binding modes contribute to the observation of existence of extra stabilization of D-arginine compared to L-arginine in the free energy landscape in the folding-unfolding studies previously discussed.

CONCLUSION

In this study, we have investigated the folding-unfolding equilibria of GAAA RNA in presence of D- and L-arginine. Using rigorous enhanced sampling method with multiple collective variables, we could construct the free energy landscape of GAAA RNA in presence of D- and L- stereoisomers of arginine.

Here we have combined rigorous umbrella sampling and parallel bias metadynamics to understand the specificity of amino acids (D- and L-arginine) to the folding-unfolding equilibria of a 12-mer GAAA RNA hairpin motif. We constructed the free energy landscape using multiple collective variables of RNA folding in presence of either D- or L-arginine. Our results show that the free energy surface of the folded state is stabilized only in presence of D-arginine and destabilized in presence of L-arginine. Interestingly, the misfolded states were of similar free energy with respect to the extended unfolded state.

To investigate the reason for stereochemical influence of amino acid on RNA folding, we also calculated the binding free energy of the D- and L-arginine to the folded structure. This result corroborated the previous one showing two different minima for the D-arginine while a single less stable minimum for L-arginine.

Further investigation revealed that the stabilization of D-arginine comes from two factors: the stacking interactions by the guanidinium group and electrostatic stabilization by the c-alpha part in the D-Arginine. In case of L-arginine, due to mirror symmetry, the electrostatic stabilization is not favorable. Therefore, in case of L-arginine, the stabilization majorly occurred due to the interaction of the sugar -backbone towards the arginine side groups. As the D-arginine stabilizes the terminal base pairs through multitude of interactions, it prevents unfolding of the RNA. In case of L-arginine, the side interaction does not help the stabilization.

Therefore, our results show the molecular origin of the chiral structures on the thermodynamics of RNA hairpin taking the example of a specific GAAA motif. This indicates that similar specific interaction between the RNA and chiral amino acids support the hypothesis of homochirality in the context of the origin of life.

This work can be extended to study more complex RNA motifs to see how chirality of amino acids affects RNA folding in general.

ASSOCIATED CONTENT

Supporting Information. Text and figures discussing different choices of collective variables for RNA folding in the absence of amino acid, reweighted free energy surfaces with respected to other collective variables, representation of edge interactions in the RNA folds.

AUTHOR INFORMATION

Corresponding Author

*E-mail: arnab.mukherjee@iiserpune.ac.in Phone: +91 20 2590 8051

Author Contributions

The manuscript was written through contributions of all authors.

Funding Sources

Department of Biotechnology (DBT), India (BT/PR34215/AI/133/22/2019) is acknowledged for partial funding support. AV (fellow id: 09/936(0209)/2019-EMR-I) thank CSIR for his fellowship.

ACKNOWLEDGMENT

Authors acknowledge the support and the resources provided by 'PARAM Brahma Facility' under the National Supercomputing Mission, Government of India at the Indian Institute of Science Education and Research, Pune.

ABBREVIATIONS

CV, collective Variable; US, umbrella sampling; Rg, radius of gyration; HB, hydrogen bonds

REFERENCES

1. Bonner, W. A., Chirality and life. *Orig Life Evol Biosph* **1995,** *25* (1-3), 175-90.
2. Mason, S. F., Origins of biomolecular handedness. *Nature* **1984,** *311* (5981), 19-23.
3. Morozov, L., Mirror symmetry breaking in biochemical evolution. *Orig Life* **1979,** *9* (3), 187-217.
4. Blackmond, D. G., The origin of biological homochirality. *Cold Spring Harb Perspect Biol* **2010,** *2* (5), a002147.
5. Breslow, R., A likely possible origin of homochirality in amino acids and sugars on prebiotic earth. *Tetrahedron Letters* **2011,** *52* (17), 2028-2032.
6. Morris, K. V.; Mattick, J. S., The rise of regulatory RNA. *Nat Rev Genet* **2014,** *15* (6), 423-37.
7. Joyce, G. F., The antiquity of RNA-based evolution. *Nature* **2002,** *418* (6894), 214-21.
8. Herschlag, D., RNA chaperones and the RNA folding problem. *J Biol Chem* **1995,** *270* (36), 20871-4.
9. Hoffman, M. M.; Khrapov, M. A.; Cox, J. C.; Yao, J.; Tong, L.; Ellington, A. D., AANT: the Amino Acid-Nucleotide Interaction Database. *Nucleic Acids Res* **2004,** *32* (Database issue), D174-81.
10. Woese, C. R.; Winker, S.; Gutell, R. R., Architecture of ribosomal RNA: constraints on the sequence of "tetra-loops". *Proceedings of the National Academy of Sciences* **1990,** *87* (21), 8467-8471.

11. Jucker, F. M.; Pardi, A., Solution structure of the CUUG hairpin loop: a novel RNA tetraloop motif. *Biochemistry* **1995,** *34* (44), 14416-27.

12. Jaeger, L.; Michel, F.; Westhof, E., Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *Journal of Molecular Biology* **1994,** *236* (5), 1271-1276.

13. Zhao, Q.; Huang, H. C.; Nagaswamy, U.; Xia, Y.; Gao, X.; Fox, G. E., UNAC tetraloops: to what extent do they mimic GNRA tetraloops? *Biopolymers* **2012,** *97* (8), 617-28.

14. Molinaro, M.; Tinoco, I., Jr., Use of ultra stable UNCG tetraloop hairpins to fold RNA structures: thermodynamic and spectroscopic applications. *Nucleic Acids Res* **1995,** *23* (15), 3056-63.

15. Baumruk, V.; Gouyette, C.; Huynh-Dinh, T.; Sun, J. S.; Ghomi, M., Comparison between CUUG and UUCG tetraloops: thermodynamic stability and structural features analyzed by UV absorption and vibrational spectroscopy. *Nucleic Acids Res* **2001,** *29* (19), 4089-96.

16. 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* (42), 16820-5.

17. Ma, H.; Proctor, D. J.; Kierzek, E.; Kierzek, R.; Bevilacqua, P. C.; Gruebele, M., Exploring the energy landscape of a small RNA hairpin. *J Am Chem Soc* **2006,** *128* (5), 1523-30.

18. Kuhrova, P.; Banas, P.; Best, R. B.; Sponer, J.; Otyepka, M., Computer Folding of RNA Tetraloops? Are We There Yet? *J Chem Theory Comput* **2013,** *9* (4), 2115-25.

19. 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 Theory Comput* **2018,** *14* (12), 6598-6612.

20. Fiore, J. L.; Nesbitt, D. J., An RNA folding motif: GNRA tetraloop-receptor interactions. *Q Rev Biophys* **2013,** *46* (3), 223-64.

21. Varani, G.; Cheong, C.; Tinoco, I., Structure of an unusually stable RNA hairpin. *Biochemistry* **1991,** *30* (13), 3280-3289.

22. Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D., Single nucleotide RNA choreography. *Nucleic Acids Res* **2006,** *34* (5), 1481-1491.

23. Correll, C. C.; Swinger, K., Common and distinctive features of GNRA tetraloops based on a GUAA tetraloop structure at 1.4 A resolution. *RNA* **2003,** *9* (3), 355-63.

24. Heus, H. A.; Pardi, A., Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science* **1991,** *253* (5016), 191-4.

25. Lescrinier, E.; Nauwelaerts, K.; Zanier, K.; Poesen, K.; Sattler, M.; Herdewijn, P., The naturally occurring N6-threonyl adenine in anticodon loop of Schizosaccharomyces pombe tRNAi causes formation of a unique U-turn motif. *Nucleic Acids Res* **2006,** *34* (10), 2878-86.

26. Preus, S.; Wilhelmsson, L. M., Advances in quantitative FRET-based methods for studying nucleic acids. *Chembiochem* **2012,** *13* (14), 1990-2001.

27. Sengupta, A.; Sung, H. L.; Nesbitt, D. J., Amino Acid Specific Effects on RNA Tertiary Interactions: Single-Molecule Kinetic and Thermodynamic Studies. *J Phys Chem B* **2016,** *120* (41), 10615-10627.

28. Nicholson, D. A.; Sengupta, A.; Sung, H. L.; Nesbitt, D. J., Amino Acid Stabilization of Nucleic Acid Secondary Structure: Kinetic Insights from Single-Molecule Studies. *J Phys Chem B* **2018,** *122* (43), 9869-9876.

29. Zerze, G. H.; Piaggi, P. M.; Debenedetti, P. G., A Computational Study of RNA Tetraloop Thermodynamics, Including Misfolded States. *J Phys Chem B* **2021,** *125* (50), 13685-13695.

30. Bottaro, S.; Lindorff-Larsen, K., Mapping the Universe of RNA Tetraloop Folds. *Biophys J* **2017,** *113* (2), 257-267.

31. Banas, P.; Hollas, D.; Zgarbova, M.; Jurecka, P.; Orozco, M.; Cheatham, T. E., 3rd; Sponer, J.; Otyepka, M., Performance of Molecular Mechanics Force Fields for RNA Simulations: Stability of UUCG and GNRA Hairpins. *J Chem Theory Comput* **2010,** *6* (12), 3836-3849.

32. Mohan, S.; Hsiao, C.; Bowman, J. C.; Wartell, R.; Williams, L. D., RNA tetraloop folding reveals tension between backbone restraints and molecular interactions. *J Am Chem Soc* **2010,** *132* (36), 12679-89.

33. Proctor, D. J.; Ma, H.; Kierzek, E.; Kierzek, R.; Gruebele, M.; Bevilacqua, P. C., Folding thermodynamics and kinetics of YNMG RNA hairpins: specific incorporation of 8-bromoguanosine leads to stabilization by enhancement of the folding rate. *Biochemistry* **2004,** *43* (44), 14004-14.

34. Chakraborty, D.; Collepardo-Guevara, R.; Wales, D. J., Energy landscapes, folding mechanisms, and kinetics of RNA tetraloop hairpins. *J Am Chem Soc* **2014,** *136* (52), 18052-61.

35. Haldar, S.; Kuhrova, P.; Banas, P.; Spiwok, V.; Sponer, 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 Theory Comput* **2015,** *11* (8), 3866-77.

36. Zhang, W.; Chen, S. J., Exploring the complex folding kinetics of RNA hairpins: II. Effect of sequence, length, and misfolded states. *Biophys J* **2006,** *90* (3), 778-87.

37. Pathak, A. K.; Bandyopadhyay, T., Water isotope effect on the thermostability of a polio viral RNA hairpin: A metadynamics study. *J Chem Phys* **2017,** *146* (16), 165104.

38. Halder, A.; Kumar, S.; Valsson, O.; Reddy, G., Mg(2+) Sensing by an RNA Fragment: Role of Mg(2+)-Coordinated Water Molecules. *J Chem Theory Comput* **2020,** *16* (10), 6702-6715.

39. Malyshko, E. V.; Semenova, E. V.; Bagrova, O. E.; Murtazina, A. R.; Tverdislov, V. A. Chiral Dualism as a Unifying Principle in Molecular Biophysics *Biophysica* [Online], 2021, p. 22-37.

40. Challier, L.; Miranda-Castro, R.; Barbe, B., *et al.*, Multianalytical Study of the Binding between a Small Chiral Molecule and a DNA Aptamer: Evidence for Asymmetric Steric Effect upon 3 '-versus 5 '-End Sequence Modification. *Anal Chem* **2016,** *88* (23), 11963-11971.

41. McConathy, J.; Owens, M. J., Stereochemistry in Drug Action. *Prim Care Companion J Clin Psychiatry* **2003,** *5* (2), 70-73.

42. Sharma, S.; Toupet, L.; Ahmad, M.; Arjmand, F., Synthesis, characterization, and crystal structure of RNA targeted L- and D-phenylalanine-(1,10-phen)-copper(II) conjugate complexes: comparative in vitro RNA binding profile of enantiomers and their biological evaluation by morphological studies and antibacterial activity. *Rsc Adv* **2016,** *6* (83), 79372-79382.

43. Yarus, M., A specific amino acid binding site composed of RNA. *Science* **1988,** *240* (4860), 1751-8.

44. Janas, T.; Widmann, J. J.; Knight, R.; Yarus, M., Simple, recurring RNA binding sites for L-arginine. *RNA* **2010,** *16* (4), 805-16.

45. Yarus, M.; Widmann, J. J.; Knight, R., RNA-amino acid binding: a stereochemical era for the genetic code. *J Mol Evol* **2009,** *69* (5), 406-29.

46. Torrie, G. M.; Valleau, J. P., Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977,** *23* (2), 187-199.

47. Pfaendtner, J.; Bonomi, M., Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics. *J Chem Theory Comput* **2015,** *11* (11), 5062-7.

48. Zgarbova, M.; Sponer, J.; Otyepka, M.; Cheatham, T. E., 3rd; Galindo-Murillo, R.; Jurecka, P., Refinement of the Sugar-Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J Chem Theory Comput* **2015,** *11* (12), 5723-36.

49. Horn, A. H., A consistent force field parameter set for zwitterionic amino acid residues. *J Mol Model* **2014,** *20* (11), 2478.

50. Mark, P.; Nilsson, L., Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *The Journal of Physical Chemistry A* **2001,** *105* (43), 9954-9960.

51. Deift, P.; Zhou, X., A Steepest Descent Method for Oscillatory Riemann--Hilbert Problems. Asymptotics for the MKdV Equation. *The Annals of Mathematics* **1993,** *137* (2).

52. 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. *The Journal of Chemical Physics* **1984,** *81* (8), 3684-3690.

53. Nosé, S., A molecular dynamics method for simulations in the canonical ensemble. *Molecular Physics* **2006,** *52* (2), 255-268.
54. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M., LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **1997,** *18* (12), 1463-1472.
55. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: AnN·log(N) method for Ewald sums in large systems. *The Journal of Chemical Physics* **1993,** *98* (12), 10089-10092.
56. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015,** *1-2*, 19-25.
57. Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J., GROMACS: fast, flexible, and free. *J Comput Chem* **2005,** *26* (16), 1701-18.
58. consortium, P., Promoting transparency and reproducibility in enhanced molecular simulations. *Nat Methods* **2019,** *16* (8), 670-673.
59. Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G., PLUMED 2: New feathers for an old bird. *Computer Physics Communications* **2014,** *185* (2), 604-613.
60. Barducci, A.; Bussi, G.; Parrinello, M., Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys Rev Lett* **2008,** *100* (2), 020603.
61. Prakash, A.; Fu, C. D.; Bonomi, M.; Pfaendtner, J., Biasing Smarter, Not Harder, by Partitioning Collective Variables into Families in Parallel Bias Metadynamics. *J Chem Theory Comput* **2018,** *14* (10), 4985-4990.
62. Fu, C. D.; Pfaendtner, J., Lifting the Curse of Dimensionality on Enhanced Sampling of Reaction Networks with Parallel Bias Metadynamics. *J Chem Theory Comput* **2018,** *14* (5), 2516-2525.
63. Prakash, A.; Sprenger, K. G.; Pfaendtner, J., Essential slow degrees of freedom in protein-surface simulations: A metadynamics investigation. *Biochem Biophys Res Commun* **2018,** *498* (2), 274-281.
64. Alamdari, S.; Sampath, J.; Prakash, A.; Gibson, L. D.; Pfaendtner, J. In *Efficient Sampling of High-Dimensional Free Energy Landscapes: A Review of Parallel Bias Metadynamics*, Foundations of Molecular Modeling and Simulation, Singapore, 2021//; Maginn, E. J.; Errington, J., Eds. Springer Singapore: Singapore, 2021; pp 123-141.
65. Kästner, J., Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2011,** *1* (6), 932-942.
66. 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. *Journal of Computational Chemistry* **1992,** *13* (8), 1011-1021.
67. Bottaro, S.; Di Palma, F.; Bussi, G., The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res* **2014,** *42* (21), 13306-14.
68. Poblete, S.; Bottaro, S.; Bussi, G., A nucleobase-centered coarse-grained representation for structure prediction of RNA motifs. *Nucleic Acids Res* **2018,** *46* (4), 1674-1683.
69. Sasikala, W. D.; Mukherjee, A., Molecular mechanism of direct proflavine-DNA intercalation: evidence for drug-induced minimum base-stacking penalty pathway. *J Phys Chem B* **2012,** *116* (40), 12208-12.
70. Mukherjee, A.; Lavery, R.; Bagchi, B.; Hynes, J. T., On the molecular mechanism of drug intercalation into DNA: a simulation study of the intercalation pathway, free energy, and DNA structural changes. *J Am Chem Soc* **2008,** *130* (30), 9747-55.
71. Wilhelm, M.; Mukherjee, A.; Bouvier, B.; Zakrzewska, K.; Hynes, J. T.; Lavery, R., Multistep drug intercalation: molecular dynamics and free energy studies of the binding of daunomycin to DNA. *J Am Chem Soc* **2012,** *134* (20), 8588-96.
72. Fu, H.; Chen, H.; Wang, X.; Chai, H.; Shao, X.; Cai, W.; Chipot, C., Finding an Optimal Pathway on a Multidimensional Free-Energy Landscape. *J Chem Inf Model* **2020,** *60* (11), 5366-5374.
73. Leontis, N. B.; Westhof, E., Geometric nomenclature and classification of RNA base pairs. *RNA* **2001,** *7* (4), 499-512.
74. Leontis, N. B.; Stombaugh, J.; Westhof, E., The non-Watson-Crick base pairs and their associated isostericity matrices. *Nucleic Acids Res* **2002,** *30* (16), 3497-531.
75. Bottaro, S.; Bussi, G.; Pinamonti, G.; Reisser, S.; Boomsma, W.; Lindorff-Larsen, K., Barnaba: software for analysis of nucleic acid structures and trajectories. *RNA* **2019,** *25* (2), 219-231.
76. Ando, T.; Takahashi, S.; Tamura, K., Principles of chemical geometry underlying chiral selectivity in RNA minihelix aminoacylation. *Nucleic Acids Res* **2018,** *46* (21), 11144-11152.