Exploring the Energy Landscape of Riboswitches Using Collective Variables Based on Tertiary Contacts

Jigneshkumar Dahyabhai Prajapati [1], José N. Onuchic [2,3] and Karissa Y. Sanbonmatsu [1,4]*

*1 - **Theoretical Biology and Biophysics,** Los Alamos National Laboratory, Los Alamos, NM 87545, United States*
*2 - **Center for Theoretical Biological Physics,** Rice University, Houston, TX 77005, United States*
*3 - **Departments of Physics and Astronomy,** Chemistry, and Biosciences, Rice University, Houston, TX 77005, United States*
*4 - **New Mexico Consortium,** Los Alamos, NM 87544, United States*

Correspondence to Karissa Y. Sanbonmatsu: Theoretical Biology and Biophysics, Los Alamos National Laboratory, Los Alamos, NM 87545, United States. *kys@lanl.gov (K.Y. Sanbonmatsu)@karissascience (K.Y. Sanbonmatsu)*
https://doi.org/10.1016/j.jmb.2022.167788
Edited by Philip C. Bevilacqua

Abstract

Messenger RNA regulatory elements, such as riboswitches, can display a high degree of flexibility. By characterizing their energy landscapes, and corresponding distributions of 3D configurations, structure–function relationships can be elucidated. Molecular dynamics simulation with enhanced sampling is an important strategy used to computationally access free energy landscapes characterizing the accessible 3D conformations of RNAs. While tertiary contacts are thought to play important roles in RNA dynamics, it is difficult, in explicit solvent, to sample the formation and breakage of tertiary contacts, such as helix-helix interactions, pseudoknot interactions, and junction interactions, while maintaining intact secondary structure elements. To this end, we extend previously developed collective variables and metadynamics efforts, to establish a simple metadynamics protocol, which utilizes only one collective variable, based on multiple tertiary contacts, to characterize the underlying free energy landscape of any RNA molecule. We develop a modified collective variable, the tertiary contacts distance (Q_{TC}), which can probe the formation and breakage of all or selectively chosen tertiary contacts of the RNA. The SAM-I riboswitch in the presence of three ionic and substrate conditions was investigated and validated against the structure ensemble previously generated using SAXS experiments. This efficient and easy to implement all-atom MD simulation based approach incorporating metadynamics to study RNA conformational dynamics can also be transferred to any other type of biomolecule.

1. Introduction

Understanding the structure and function of RNA has been a key challenge in the life sciences for decades.[1,2] Improving our understanding towards how RNAs can fold into different shapes and structures to carry out their numerous functions is especially timely now, in light of the numerous vaccines and therapeutic strategies based on messenger RNAs.[3,4]

Energy landscapes have been highly useful and effective in describing the folding and function of biomolecules.[5] Collective variables have played an important role in revealing order parameters that best describe barriers in the free energy landscape.[6] In particular, collective variables based on

the total number of native contacts of a biomolecule have led to a great number of insights into protein folding.[6] Recent efforts using structure-based potentials (native-contact based) have obtained near exhaustive sampling with atomistic resolution.[7–15] A significant difference between energy landscapes of proteins and of RNA[7,11–15] is that RNA landscapes are thought to be more rugged, containing more kinetic traps. Specifically, the landscapes may have several intermediate states separated by energy barriers higher than thermal energy.[16–18] For systems with such high degrees of freedom (DOFs), i.e., $3N$ with N = number of atoms, standard molecular dynamics (MD) simulations are not necessarily ergodic in the timescale achievable with current computation resources. Enhanced sampling MD simulations are techniques which can help one gain insight into the folding process and also improve understanding of the underlying thermodynamics.

In the past, replica exchange based MD (REMD) techniques, such as temperature-REMD[19–23] and replica exchange with solute tempering (REST2),[24] have extensively been utilized to explore conformation dynamics of biomolecules.[23,25] Implementing these methods is more straight forward considering a priori knowledge on folding process not required; however, in the case of RNA, helix and base pair melting present important challenges. Enhanced sampling methods based on collective variables, e.g., metadynamics,[26–34] umbrella sampling[35] and adaptive biasing force method,[36] facilitate acceleration of sampling along two to three predefined degrees of freedom, and in turn, allow estimates of free energy surfaces as a function of these. Great progress has been made in the prediction of secondary structure and tertiary structure predictions for RNA molecules.[37–41] Since these systems consist of an extremely large number of degrees of freedom, defining appropriate collective variables for a specific RNA system and applying these enhanced sampling techniques remains a major challenge. In particular, without proper collective variables the convergence of the free energy profiles is more difficult. Moreover, the interplay of key governing factors of RNA dynamics, such as base-pairing, tertiary bridges formation and mono/di-valent ions influence, is still not completely understood at atomistic scale and would be clarified by better understanding of the energy landscape surfaces.[42,43] The factors tend to result in rugged energy landscapes. In particular, our previous experimental studies have shown that tertiary contacts are important for the operation of the SAM-I riboswitch.[44,45] Thus, sampling the formation and breakage of tertiary contacts in explicit solvent molecular dynamics simulations is critical for obtaining the corresponding energy landscapes.[44] We show below that, while using the RMSD as a collective variable does not fully sample tertiary contacts, a collective variable based on tertiary contacts is able to sample formation and breakage of tertiary contacts.

Thus, we present a simple simulation strategy which can facilitate investigation of riboswitch free energy landscapes, as well as any other RNA, in general. Riboswitches are regulatory elements embedded in the 5' untranslated region of mRNA and typically have two domains, the aptamer domain, which binds to a specific metabolite, and an expression platform which regulates the gene expression in response to a conformation change in aptamer.[48,49] In this study, we have investigated the conformation transitions of the aptamer domain of the SAM-I riboswitch of *Thermoanaerobacter tengcongensis* (Figure 1(A)), which regulates expression of genes affiliated with sulphur and methionine metabolism through transcription termination upon binding to S-adenosylmthionine (SAM) metabolite.[50–53] The SAM bound structure of the SAM-I aptamer domain from *Thermoanaerobacter tengcongensis* is shown in Figure 1(B), and the unbound form adopts a similar secondary structure architecture, but has a 3D ensemble of configurations significantly more extended than the bound state.[47,46] The SAM-I aptamer domain consists of four helices, P1-P4, connected through a four way junction, with a compact tertiary fold featuring a pseudoknot and kink-turn. This ligand bound structure is also similar to that determined from bacterium *Bacillus subtilis yitJ*.[54] We refer to this conformation as crystallographic or native closed state hereafter. Using small angle X-ray scattering (SAXS), the SAM-I aptamer domain was demonstrated to undergo Mg^{2+} ion dependent compaction, adopting conformations closely similar to the native closed state at higher Mg^{2+} concentrations (Figure 1(B)).[47] The ensemble of structures from SAXS experiments in the ligand-free ensemble are shown in Figure 2(A), where some conformations in closed configurations have P1 and P3 helices in contact, while, in open configurations in this ensemble, P1 and P3 do not interact. Several other experimental studies also confirmed that Mg^{2+} ions, especially at physiological concentration, help to facilitate spontaneous folding of the SAM-I aptamer domain into well-structured, pre-organized form, ready to bind ligand.[54–56,45,13,57,15,58] These studies also revealed that the addition of SAM swiftly pushes SAM-I into the fully compact form, a thermodynamically more stable closed state.

Following these studies, we have employed all-atom explicit solvent metadynamics molecular dynamics simulations to investigate the SAM-I riboswitch in presence of three conditions in this study, i) 150 mM KCl, ii) 150 mM KCl with 7.6 mM $MgCl_2$ and iii) 150 mM KCl, 7.6 mM $MgCl_2$ with a SAM molecule bound to the binding site. These conditions are chosen to be consistent with the previous SAXS experiments.[47] SAM-I is progressively expected to achieve the more compact form in the respective conditions. We present a metady-

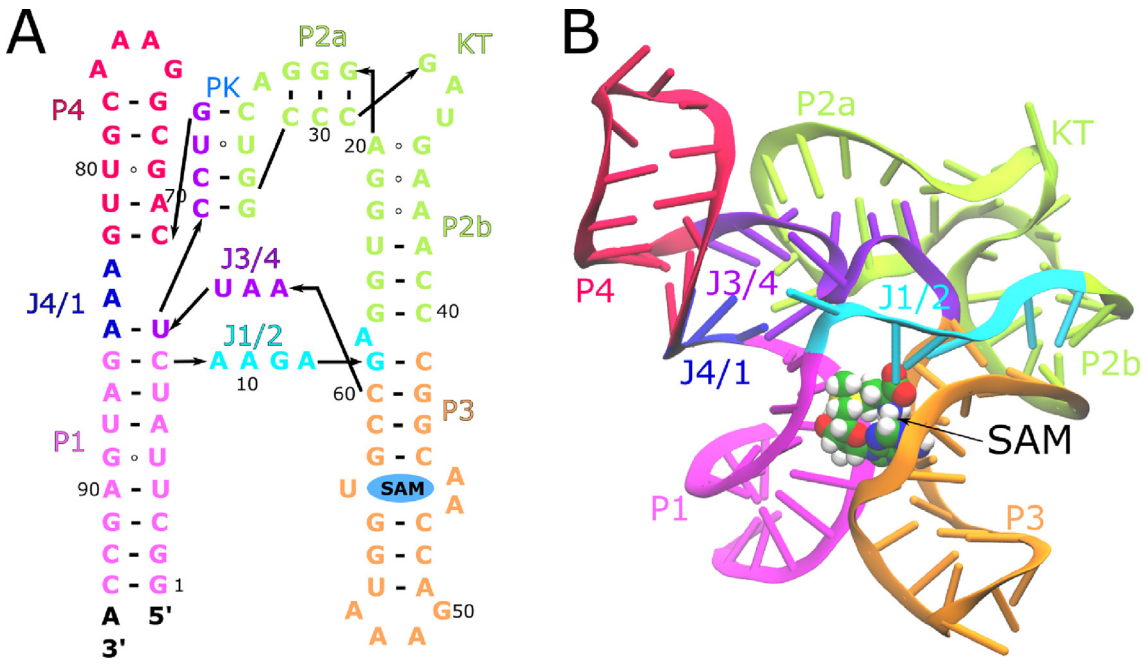

Figure 1. (A) Secondary structure of the aptamer domain of SAM-I riboswitch from *Thermoanaerobacter tengcongensis* bacterium, highlighting the tertiary architecture.[46] Different structural domains are colored as well as labelled as follows: P, helices; J, joining loops; PK, pseudoknot; KT, kink-turn. (B) Three dimensional structure of SAM-I aptamer domain in cartoon representation is depicted with the SAM ligand in van der Waals representation (PDB ID: 3IQR).[47]

namics based approach to investigate the energetics of conformation dynamics of the SAM-I riboswitch. We have conducted simulations with total time scale of $\approx 55\ \mu s$. As our previous experimental studies have shown that tertiary contacts play a key role in the operation of the SAM-I riboswitch, sampling the formation and breakage of tertiary contacts is critical.[44] The primary aim of this study is to design a collective variable to accelerate the dynamics of RNA molecules in order to derive meaningful transition pathways, sampling formation and breakage of tertiary contacts without significantly disrupting the secondary structure (i.e., helical structure of the helices). The commonly used collective variable RMSD captures the fluctuations of radius of gyration. The collective variable Q (fraction of total native contacts) captures folding of the RNA from a completely denatured state and obtains near exhaustive sampling of tertiary contacts for structure-based potentials.[7] For the case of explicit solvent, to specifically sample tertiary contact breakage and formation while maintaining secondary structure, important for functional operation of the riboswitch, we have designed a single collective variable, "tertiary contacts distance (Q_{TC})", that combines the four major tertiary contact regions and can enhance the formation or breakage of tertiary contacts in RNA. This variable has the advantage over a simple mean average of tertiary contacts in that it produces full sampling of all tertiary contacts while restricting tertiary contact breakage distances (distance between broken tertiary contact regions) to reasonable values, relative to experimentally determined ensembles. This enables us to describe meaningful conformation transitions in three conditions, producing configurations consistent with the SAXS based structure ensemble. Moreover, we have derived the underlying thermodynamics which is in accordance with observations made during previous experiments.

2. Theory

Here, we have employed the well-tempered metadynamics technique and the multiple-walker metadynamics technique. In the well-tempered metadynamics approach,[59–61] a history-dependent bias potential ($V(\mathbf{z}, t)$) which is a sum of Gaussian hills, is added during simulation along the predefined M number of collective variables $\mathbf{z}$. This potential is given as

$$V(\mathbf{z}, t_N) = \sum_{i=0}^{N} h \exp\left(-\frac{V(\mathbf{z}(t_i), t_{i-1})}{k_B \Delta T}\right) \exp\left(-\sum_{j=1}^{M} \frac{(\mathbf{z}_j - \mathbf{z}_j(t_i))^2}{2\delta \mathbf{z}_j^2}\right), \quad (1)$$

where $\delta\mathbf{z} = (\delta\mathbf{z}_1, \ldots, \delta\mathbf{z}_M)$ are the widths of Gaussian hills, and $\mathbf{z}(t_i) = (\mathbf{z}_1(t_i), \ldots, \mathbf{z}_M(t_i))$ are previously visited points along the collective variable space at time interval of t_i and h is the initial height of the Gaussian. To assure a smooth convergence of the bias potential in well-tempered scheme, height is progressively

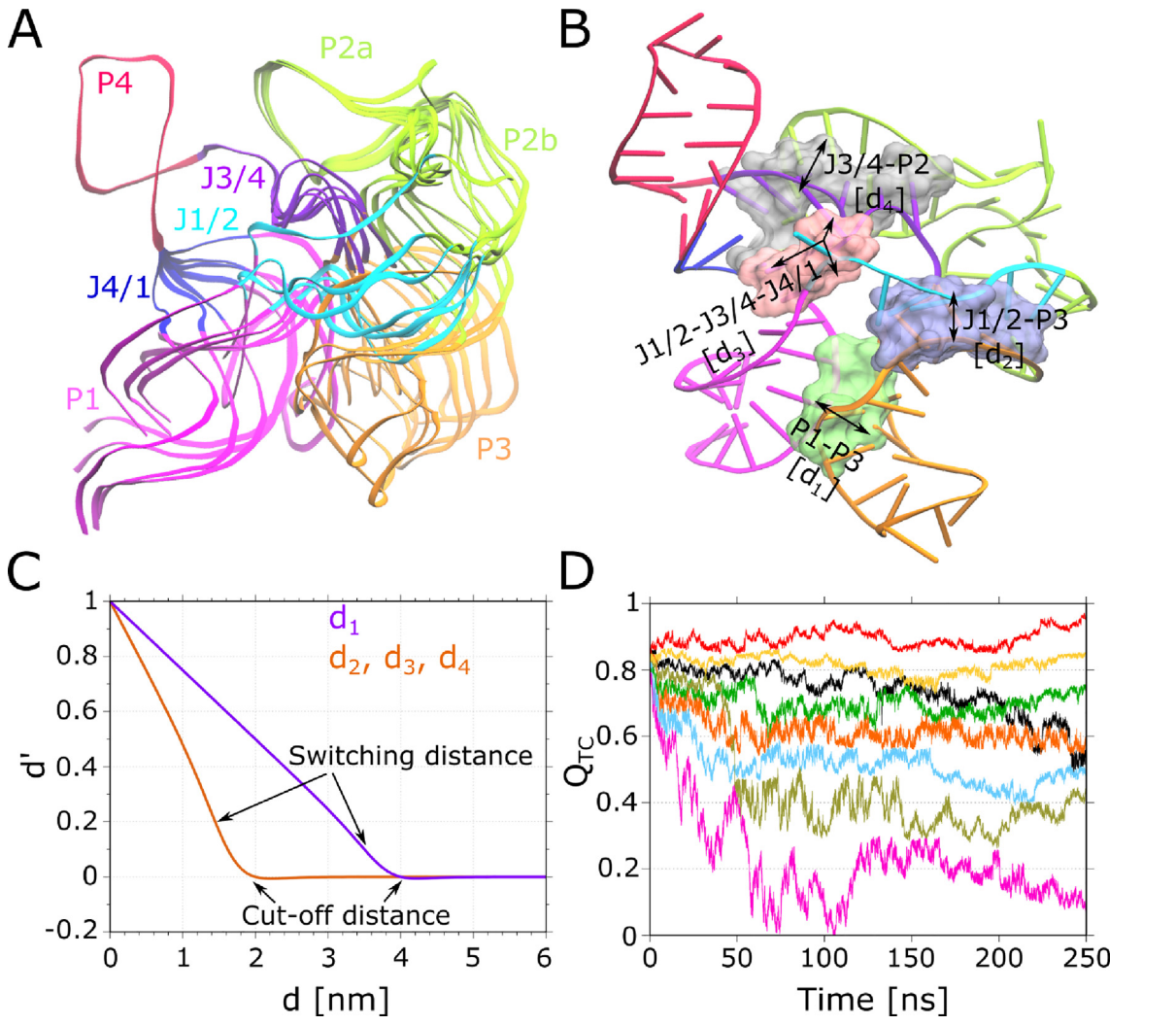

Figure 2. (A) Superimposition of seven conformations of SAM-I based on the SAXS experiments.[47] (B) Four tertiary contacts forming regions along the SAM-I structure are illustrated as surface and labelled according to the secondary structure elements involved. Additional labels d_1 to d_4 correspond to collective variables used during WTmetaD simulations. (C) Depiction of the normalization of collective variables d_1 to d_4 to new ones d'_1 to d'_4 using Eq. 6. (D) Time evolution of collective variable Q_{TC} for eight walkers during a WTmetaD simulation performed for the system having a SAM bound SAM-I in the presence of 150 mM KCl and 7.6 mM $MgCl_2$.

decayed using a scaling factor $\exp\left(-\frac{V(\mathbf{z}(t_i),t_{i-1})}{k_B\Delta T}\right)$. The k_B denotes the Boltzmann constant and the tuning temperature ΔT limits the sampling of collective variable space up to energy barriers of height $k_B(T+\Delta T)$, where T is the system temperature. Together with well-tempered approach, we also used multiple-walker metadynamics technique which runs multiple interacting replicas in parallel to speedup the bias potential construction,[62] collectively, denoted as WTmetaD simulations hereafter. All simulations in this study were conducted using PLUMED 2.7.2[63] patched with GROMACS 2021.4,[64] following the protocol established previously.[65–67]

Upon completion of simulations, one dimensional (1D) free energy surfaces are estimated according to typical metadynamics protocols, using the Tiwary-Parrinello reweighting scheme[68]

$$F(\mathbf{z}) = -\frac{1}{\beta}\log\frac{\sum_t \delta(\mathbf{z}(t)-\mathbf{z})\; e^{\beta(V(\mathbf{z},t)-c(t))}}{\sum_t e^{\beta(V(\mathbf{z},t)-c(t))}}, \tag{2}$$

where $\beta = 1/k_BT$, δ is the Dirac delta function, and $c(t)$ the time-dependent offset, which can be estimated by[68,65]

$$c(t) = \frac{1}{\beta}\log\frac{\int d\mathbf{z}\, e^{-\beta F(\mathbf{z})}}{\int d\mathbf{z}\, e^{-\beta(F(\mathbf{z})+V(\mathbf{z},t))}}. \tag{3}$$

Subsequently, free energy profiles for n number of independent runs $(F_i(z);\ i=1,..,n)$ are determined and the average profile $F_{avg}(z)$ is calculated using following expression[69]

$$F_{avg}(z) = -\frac{1}{\beta}\log\left(\frac{1}{n}\sum_{i=1}^{n} e^{-\beta F_i(z)}\right). \tag{4}$$

The error is calculated using standard error propagation scheme.

3. Results and Discussion

Following the system setup of the SAM-I riboswitch in three different ionic and substrate conditions (see section Materials and Methods), we choose collective variables, an important step in WTmetaD simulations. Collective variables are functions of atomic coordinates which describe the process under investigation in low dimensional space and should have the ability to distinguish initial, final and intermediate metastable states as well as, in the case of metadynamics, promote transitions between states. In the case of RNA, it is not clear which collective variables perform the best to elevate the sampling of the molecule with the size of SAM-I.[25] Moreover, how realistic conformation dynamics can be achieved for such molecules using these collective variables has not been benchmarked. Therefore, we compare routinely used collective variables with collective variables based on tertiary contacts defined in this study. Importantly, we aim to find a single collective variable which is easy to implement and capable of producing realistic dynamics for any RNA in the context of metadynamics.

The SAM-I aptamer has four regions where tertiary contacts are formed (Figure 2(B)). The structure ensemble of SAM-I from SAXS experiments shown in Figure 2(A) hints that these tertiary contacts need to be broken to achieve the conformation transitions. Moreover, it was also illustrated in our previous wetlab experiments that stability in the tertiary contacts mainly controls the collapse and the subsequent switching from aptamer to expression platform conformation in SAM-I.[44] In principle, this criteria may be sufficient to instigate the conformation changes in any RNA. Thus, we seek a collective variable that can bias tertiary contact formation and breakage in RNAs. In this direction, we have employed four collective variables in this study. First, we start with two of the most commonly used collective variables utilized for the conformation sampling of RNA molecules previously: the root mean squared deviation (RMSD) and total number of native contacts (Q).[6] RMSD was calculated using all atoms of the SAM-I molecule and a WTmetaD simulation was performed using eight walkers for a duration of 250 ns each, resulting in total timescale of 2 μs. Note that for all four collective variables tested in this investigation, the simulation time remains constant and the system setup of SAM-I containing SAM in binding pocket placed in ionic solution of 150 mM KCl and 7.6 mM $MgCl_2$ was utilized. The second collective variable Q which is also known as fraction of native contacts[6] is calculated using all native tertiary contact pairs of RNA in this study. When two atoms i and j from residues R_i and R_j present on different secondary domains form a contact, provided the distance between these atoms is below 0.45 nm, Q for N number of pairs of native contacts is defined as[6,70,71]

$$Q(X) = \frac{1}{N}\sum_{[i,j]}\frac{1}{1+exp\ [\alpha(r_{ij}(X)-\lambda r_{ij}^{0})]}, \quad (5)$$

where α is a smoothing factor set to 50 nm^{-1} and λ accounts for the fluctuations during the contact formation, which is taken to be 1.8. r_{ij}^{0} is the distance between atoms i and j in the native state, while $r_{ij}(X)$ is the distance between atoms i and j for instantaneous state X. The third and fourth collective variables are based upon the principle of biasing the distance between tertiary contacts. At first, we define four preliminary collective variables, d_1 to d_4, respectively for four contact forming regions of SAM-I (see Figure 2(B)). These collective variables correspond to inter-domain distances as follow: d_1, P1 and P3; d_2, J1/2, J3/4 and J4/1; d_3, J1/2 and P3; d_4, J3/4 and P2. For each collective variable, first the center of mass for each domain is calculated (e.g., P1 and P3 in the case of d_1) choosing only atoms involved in tertiary contact formation, i.e., those present within 0.45 nm, followed by the distance between them. Tackling all four collective variables at once during WTmetaD simulations is difficult, because the bias potential construction becomes computationally very expensive as the cost grows exponentially with addition of each collective variable. Thus, we define the third collective variable d_{mean}, which calculates the mean of collective variables d_1 to d_4. This collective variable was seen to perform well (discussed below), but controlling the sampling of underlying four collective variables was not possible. In other words, if the simulation oversamples one of the collective variable and the other three are not even being sampled, there is no way other than adding half-harmonic upper walls on individual collective variables to assure controlled sampling. However, this makes the free energy estimation process extremely tedious at later stages as the metadynamics and harmonic biases are added on different collective variables. To tackle this, by taking the inspiration from the collective variables Q and d_{mean} as well as the coordination number,[72] we design the fourth collective variable, the tertiary contacts distance, Q_{TC}. First, collective variables d_1 to d_4 are transformed to d_1' to d_4'. Here, original collective variables are normalized such that the renormalized variables range from 0 to 1 (Figure 2(C)). The mean of these new coordinates is used to obtain Q_{TC}:

$$Q_{TC}(X) = \frac{1}{N}\sum_{i} d_i' = \frac{1}{N}\sum_{i}\frac{\kappa(c_i - d_i(X))}{\kappa c_i + exp\ [\alpha(d_i(X)-\sigma_i)]}, \quad (6)$$

where the sum is over N, the number of normalized distances (or switching functions) d_i', which are calculated as follows. For collective variable number i, c_i is the cut-off distance. That is, c_i is the value of d_i at which its sampling needs to be restricted, where $d_i(X)$ is the instantaneous value of collective variable d_i for configuration X. The switching

distance, σ_i, is the value of d_i from where the normalized collective variable d_i' will be smoothly switched to zero while reaching the cut-off distance. How smoothly the switching occurs depends on the smoothing parameter α, taken to be 5 nm^{-1} for all four collective variables. In the case of SAM-I, to prevent sampling of the collective variable d_1 at 4 nm, c_1 is set to 4 nm and σ_1 to 3.5 nm, while for remaining three collective variables c_i is set to 2 nm and σ_i to 1.5 nm to prevent sampling at 2 nm, where $i = 2, 3, 4$. These cut-off distances for four collective variables are set based on the conformation changes observed in SAXS experiments (Figure 2(A)). The κ is a constant with a value of 1 nm^{-1} always, which makes the numerator and denominator dimensionless. Similar to Q, Q_{TC} is normalized such that $0 < Q_{TC} < 1$. The graphical depiction of the normalization of all four collective variables can be seen in Figure 2(C). Notably, the values of collective variables d_1' to d_4' go slightly below zero in the switching region, which is possible to tackle by taking the absolute of numerator in Eq. 6, i.e., $|\kappa(c_i - d_i(X))|$. Nevertheless, the normalization will not be perfect in both cases.

The collective variable Q_{TC} is superior compared to d_{mean} in a way that it will allow to restrict the sampling of underlying preliminary collective variables in a more controlled fashion, and mathematically, describes the conformation dynamics in a more meaningful manner, helping one to better differentiate the states of interest. Here, the SAXS-based structure ensemble provided the initial guesses for the cut-off distances for the respective collective variables. However, for the RNAs having no such conformation dynamics known, we recommend to set cut-off at 2 to 3 nm, which will be suitable for investigating the tertiary contacts mediated minor conformation switching. More complex conformation changes involving long-ranged tertiary bridges and base-pair formation or breaking would require inclusion of the preliminary collective variables for each interaction, as well as setting up large enough cut-off distances, intuitively based on the 3D structure inspection and anticipated structural changes. In general, defining large cut-off distances will be more beneficial than smaller, as this will allow sampling of more conformational space, but will be computationally more expensive. For future studies, based on the empirical fitting of Eq. 6, we highly recommend using $\sigma = c - 0.01$ nm and $\alpha = 100$ nm^{-1}, which will be suitable for all values of $c \in [0.1, \infty]$ nm. As mentioned above, users only need to define cut-off distance c appropriately for the preliminary collective variables. In future, for RNAs with no priori knowledge on conformation dynamics, it would be ideal to adaptively control the cut-off distances for each preliminary collective variables during the simulation, where ongoing conformational changes will dictate the cut-off distances on the fly.

We performed four metadynamics simulation studies, each with eight walker simulations. In study 1, we biased the metadynamics using RMSD. In study 2, we biased the metadynamics using Q. In study 3, we biased the metadynamics using d_{mean}. In study 4, we biased the metadynamics using Q_{TC} (see section Materials and Methods for more details on simulation parameters). In every simulation, we tracked the evolution of all four collective variables. In study 1, when the RMSD was biased, sampling with maximum RMSD values of $\approx 1.7nm$ was achieved (see Figure S1A in SI), much lower than the upper limit set to 3 nm. In study 1 (RMSD), analyzing the evolution of the other three collective variables shows that tertiary contacts are not broken. Visualization of trajectories indicates the deformation of secondary structure of SAM-I, e.g., melting of helices, suggesting RMSD is not an optimal collective variable. This behavior did not change when only SAM-I backbone atoms were used to calculate RMSD (data not shown). In study 2, with Q being biased, we witnessed complete sampling in the range of 0 to 1, indicating all tertiary contacts in four regions are diminished (Figure S1B in SI). However, evolution of other collective variables suggests that in most of the trajectories, tertiary contacts forming domains do not move further after detaching from each other. In other words, the distance between them did not increase soon after the detachment, producing structures less extended in comparison to the SAXS results. This inspired us to find a collective variable which can elevate the distance between the tertiary contacts forming elements to recapitulate the SAXS configurational ensemble. In study 3, we show that biasing the collective variable d_{mean} clearly satisfies this requirement. In study 3 (d_{mean}), as can be seen in Figure S1C in SI, the sampling of all four collective variable was improved compared to two previous studies, where RMSD (study 1) and Q (study 2) were biased. The only remaining issue with d_{mean} was that the underlying collective variables d_1 to d_4 were not sampled evenly. In study 3, we restricted the sampling of d_{mean} to 2.5 nm, with anticipation of limiting sampling of d_1 to 4 nm, limiting the sampling of d_2 to 2 nm, limiting the sampling of d_3 to 2 nm, and limiting the sampling of d_4 to 2 nm. However, we observed exploration up to values of $d_1 \approx 6, d_2 \approx 1, d_3 \approx 3$ and $d_4 \approx 1nm$ for respective collective variables during a simulation (Figure S2A in SI). In study 4, biasing with Q_{TC}, a complete sampling in the range of 0 to 1 was achieved, as shown in Figure 2(D). Moreover, biasing Q_{TC} produced excellent sampling of the remaining three collective variables (RMSD, Q, d_{mean}) (Figure S1D in SI). Most importantly, we

could restrict the sampling of underlying collective variables (d_1, d_2, d_3, d_4) more precisely using the collective variable Q_{TC}, as can be seen in Figure S2B in SI.

Next, we have performed the WTmetaD simulations biasing the collective variable Q_{TC} for three experimental condition systems: (i) 150 mM KCl, 0 mM $MgCl_2$, without SAM; (ii) 150 mM KCl, 7.6 mM $MgCl_2$, without SAM (supplementary video 1); and (iii) 150 mM KCl, 7.6 mM $MgCl_2$, with SAM bound at binding site (supplementary video 2). For each system, 2 μs long simulations were performed eight times. Upon completion of simulations, eight one dimensional (1D) free energy surfaces as a function of collective variable Q_{TC} were estimated using Eq. 2, and the average free energy profiles with Eq. 4. Final free energy surfaces are shown in Figure 3(A). In all three scenarios, the lowest free energy minimum is observed at Q_{TC} = 0.82, which corresponds to the crystallographic state shown on Figure 1(B) having all tertiary contacts intact. This state is undoubtedly the most probable state in all three conditions. In the presence of only 150 mM KCl (system (i)), the RNA is more extended, and several metastable states are observed. All four tertiary contacts are broken at the end of the simulations, i.e., $Q_{TC} \approx 0$. From the native state, the free energy barrier to achieve the extended state is approximately 18 kcal/mol, assuming there are no substates along the transition pathway. In the presence of 150 mM KCl and 7.6 mM $MgCl_2$ (system (ii)), the free energy profile looks similar until Q_{TC} = 0.5, where a subsequently a steep rise in the energy occurs. The energy barrier to attain the extended state is ≈ 30kcal/mol, which is 12 kcal/mol higher compared to system (i). This is consistent with previous studies showing that Mg^{2+} ions facilitate compaction of the SAM-I aptamer, producing conformations closer to the crystallographic form. In the presence of both salts and SAM in the

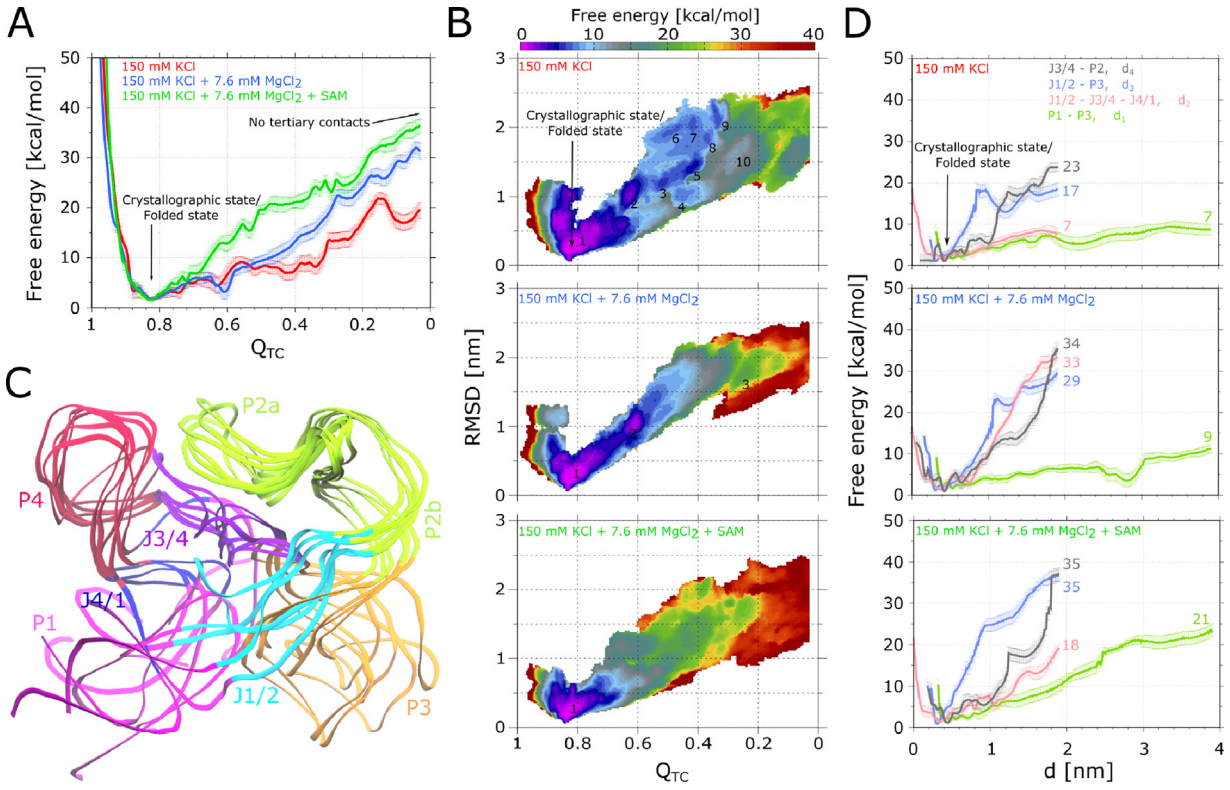

Figure 3. (A) 1D free energy surfaces as a function of collective variable Q_{TC} from WTmetaD simulations performed for SAM-I in three different ionic/ligand conditions. Error bars indicate the statistical error determined after averaging eight independent runs in each case. (B) Three panels depict 2D free energy landscapes with respect to collective variables Q_{TC} and RMSD determined from same three sets of simulations. Metastable states found along the landscapes are numbered. (C) Superimposition of the representative conformations of SAM-I from six metastable states (1–5 and 10) identified along the 2D free energy surface of the system having only 150 mM KCl. (D) 1D free energy profiles with respect to collective variables d_1 to d_4, which represent the four tertiary bonds forming areas, derived for all three SAM-I setups and depicted in three panels. The numbers indicate the energy difference between the native state ($d_{1-4} \approx 0.5$ nm) and the upper bound of each collective variable ($d_1 \approx 3.9$ nm; $d_{2-4} \approx 1.9$ nm). Note that in all panels, free energy estimates for few upper/lower bound values of collective variables are not depicted, because they do not seem to be converged due to the boundary effect.

binding cavity, the unfolding barrier was approximately 35 kcal/mol higher, which is 5 kcal/mol more with respect to the same ionic conditions but without SAM at binding site. More interestingly, energies were much higher throughout the landscape when compared with two previous scenarios, suggesting escape from the native state is much less likely. Again in agreement with previous outcomes, binding of SAM further drives the SAM-I aptamer closer to the native state.

To gain better understanding of metastable states, we have estimated 2D free energy surfaces with respect to Q_{TC} and RMSD using Eq. 2, where δ is the 2D Dirac delta function. First, eight 2D free energy surfaces were estimated from eight independent runs, and subsequently average free energy landscapes were determined using Eq. 4. Resulting free energy profiles for all three systems are shown in Figure 3(B). In each case, the lowest energy native state can be seen at Q_{TC} = 0.82 and RMSD = 0.25 nm. For system (i) (150 mM KCl, 0 mM $MgCl_2$, without SAM), 10 prominent metastable states are visible as SAM-I approaches the extended form, highlighted along the 2D free energy surface in Figure 3. Metastable states with RSMD less than 1.7 nm, i.e., substates 1, 2, 3, 4, 5 and 10, have similar conformations as seen with the SAXS derived structure ensemble (see Figs. 3C and 2A for the comparison). One can see the large scale changes in P1, P3, J1/2, and J4/1, while remaining regions remain stable. For substates 7 to 9, we see further deviations in the mobile domains of the SAM-I aptamer (Figure S3A in SI), which is much larger than observed in SAXS data. For system (ii) (150 mM KCl, 7.6 mM $MgCl_2$, without SAM), less collective variable space with respect to RMSD is being sampled. The free energy surface is more compact compared to the previous scenario and only two prominent minima are visible. Interestingly, only the area along the free energy profile where minima 1 to 5 were seen in the previous case, i.e., $Q_{TC} \in [0.4,1]$ and RMSD $\in$ [0,1.7] nm, is energetically favorable. Six representative conformations found from this region (Figure S3B in SI) are similar to those obtained in the previous system (i) and SAXS based structure ensemble. This is interesting, considering the SAXS experiments were conducted in the same ionic conditions. Collectively, all these outcomes also support the previous claim that the SAM-I aptamer domain in the $MgCl_2$ solution remains as an ensemble of structures having a fold closer to the native state. This study reveals that same is not necessarily true in the absence of $MgCl_2$, as the SAM-I aptamer can quite easily access the conformations much different than the native state, as shown in Figure S3A in SI. In system (iii) (150 mM KCl, 7.6 mM $MgCl_2$, with SAM bound at binding site), we observed only one minimum affiliated to the native state on the free energy surface, while other all regions are energetically much less favourable. This again shows that SAM binding is the key step for attaining the native fold of the SAM-I aptamer. Interestingly, the sampling of collective variable space with respect to RMSD is relatively higher compared to the previous case (ii) having no SAM, but the same ionic environment. We find that in this case, fewer chelated (inner shell) Mg^{2+} ions along the SAM-I surfaces are observed (Figure S4 in SI), especially on P1 and J1/2. These same regions display more fluctuations as shown in Figure S3B and C in SI. On the other hand, there are no dramatic changes in outer and inner shell K^+ ions as well as the outer shell Mg^{2+} ions, indicating inner sphere Mg^{2+} ions have a prominent role in the RNA dynamics. From the theoretical perspective, these outcomes indicate that Mg^{2+} binding with the RNA also needs to be biased to further improve the free energy estimates.

At the next stage, we have estimated the free energy landscapes using the above mentioned reweighting procedure with respect to collective variables d_1 to d_4 to understand how the interactions between the tertiary bridges forming domains vary, shown in Figure 3(D). The collective variables correspond to inter-domain distances as follow: d_1, P1 and P3; d_2, J1/2, J3/4 and J4/1; d_3, J1/2 and P3; d_4, J3/4 and P2. In the presence of 150 mM KCl (system i), the free energy barrier observed for breaking the tertiary bridges in these four regions are, 7, 7, 17 and 23 kcal/mol, respectively. That confirms that P1 and P3, and the three-way junction (J1/2-J3/4-J4/1) dissociation is much easier, as observed in the SAXS ensemble. Whereas, tertiary bridges between J3/4 and P2 (PK) as well as J1/2 and P3 are more difficult to interrupt. With the addition of 7.6 mM Mg^{2+} ions (system ii), the energy barrier for P1-P3 deformation increased slightly, i.e., 9 kcal/mol, while for the remaining regions increased dramatically to 33, 29 and 34 kcal/mol, suggesting that the three-way junction, which was much weaker during the previous condition (i), is strengthened with addition of Mg^{2+} ions. In the third setup (system iii) having SAM at the binding site in addition of potassium and magnesium, we observed a dramatic rise in the energy barrier to 21 kcal/mol for P1-P3 disengagement, confirming that SAM strongly enhances the binding affinity between P1 and P3 and helps to attain the closed state. The free energy barriers have also slightly gone higher for J1/2-P3 and J3/4-P2 dissociation to 35 kcal/mol in both cases, indicating that the stability of the P1-P3 bridge increases the rigidity of the whole SAM-I aptamer structure. One notable exception is the three-way junction J1/2-J3/4-J4/1, where the energy barrier for the disruption is reduced to 18 kcal/mol. As mentioned

above, there was a loss in the number of chelated Mg^{2+} ions near the loop J1/2 and the stem P1 (Figure S4B in SI), causing instability in this region. In the presence of chelated Mg^{2+} ions, these regions are more likely to remain rigid.

Taken together, our molecular dynamics trajectories suggest the following mechanism for SAM ligand recognition by the SAM-I riboswitch. When P1, P3 and J1/2 dissociate from each other, binding of SAM with the P3 helix, and to large extent with J1/2, was observed to remain intact (Figure S4C in SI and Supplementary video 2), suggesting that the ligand is first recognised by the P3 helix having proper secondary structure, and subsequently by J1/2. In this manner, the P1 helix is stabilized through interactions with ligand bound P3 and J1/2 to form a stable junction, facilitating formation of a terminator element in the expression platform region.

Previous studies collectively illustrated the Mg^{2+} ion dependent partial compaction of the SAM-I aptamer domain, i.e., Mg^{2+} ions help fold the SAM-I riboswitch into a well-structured form which looks similar to the ligand bound native state.[54–56,45,13,57,15,58] SAXS experiments revealed SAM-I in $MgCl_2$ solution adapts ensemble of states including the ligand bound form and few closely similar conformations.[47] These studies also demonstrated that SAM pushes SAM-I into the fully compact form. Taken together, the outcomes from systems (i)-(iii) in this study lead to similar conclusions. SAM-I aptamer can adopt the SAM-bound native state like conformations in the presence of only KCl solution (system i), but the probability is very low to remain in that form. The RNA will have diverse conformations ranging from native state like forms to extended conformations with no tertiary contacts. The pseudoknot will most likely to form, while remaining domains are expected to remain loose. The insertion of Mg^{2+} ions (system ii) plays an important role in driving the SAM-I aptamer towards the native form. Again, SAM-I adopts a variety of states, but not as diverse as seen with no Mg^{2+} ions. We observed that the chelated Mg^{2+} ions impose a significant rigidity throughout the SAM-I structure. We also confirm that SAM has a significant influence on enforcing the SAM-I aptamer toward the native state by bridging the P1 and P3 with more affinity. This also has an indirect effect on improving the rigidity of other regions of SAM-I.

4. Conclusions

In this study, we put forward a metadynamics simulations based strategy to study the energy landscapes of RNAs using the SAM-I riboswitch as an example. The collective variable, tertiary contacts distance Q_{TC}, developed in this study, is demonstrated to work efficiently to bias the tertiary bridging in RNA, and is shown to perform better then traditionally used collective variables like RMSD and total fraction of native contacts for the case of explicit solvent metadynamics. The collective variable d_{mean} is also found to work nicely, and in future can be used for cases where one wants to induce large scale conformation transitions, especially for systems with no priori knowledge. In the case of SAM-I, we only biased the sampling of tertiary bridges using Q_{TC} to understand the local conformation changes in a better fashion, but this collective variable is also capable of elevating the sampling of all other contacts in RNAs, e.g., canonical base pairing, to study global conformation dynamics. However, this will be more difficult to achieve and the quality of the outcomes will be needed to validate in a more systematic manner. This collective variable also allows one to choose selective tertiary bridges for biasing, if all are not required, as well as making it feasible to restrict sampling space for a particular bridge. The primary goal with developing the Q_{TC} was to achieve the more meaningful transition pathways as well as the underlying energetic of the RNA folding or conformation switching processes and not solely to achieve more sampling power. Moreover, identifying and differentiating the meaningful states during the RNA conformation switching, especially for riboswitches, has remained challenging. The Q_{TC} is demonstrated to have the potential to identify such states of interest. For the SAM-I aptamer, we demonstrate that inducing the sampling of tertiary bridge breaking was sufficient to obtain a picture on its conformation transitions, which we validated using SAXS experiments based structure ensemble. Most importantly, we could derive the underlying free energy landscapes and the energy barriers for overall structural transitions of SAM-I as well as for local regions. The overall mechanisms for conformation

Table 1 Parameters used during the WTmetaD simulations performed using four different collective variables. δ_z denote the Gaussian width, h the Gaussian height, ΔT the tuning temperature, t the bias deposition time interval, Uwall is the value of the collective variable at which the half-harmonic restraint will be activated, and k is the force constant for this restraint.

Collective variable	δ_z	h [kJ/mol]	ΔT [K]	t [ps]	Uwall	k [kJ mol^{-1} nm^{-2}]
RMSD	0.01 [nm]	1	8700	2	3 [nm]	800
Q	0.005	1	8700	2	-	-
d_{mean}	0.01 [nm]	1	8700	2	2.5 [nm]	800
Q_{TC}	0.0025	1	8700	2	-	-

changes observed here are also in excellent agreement with the previous studies. Overall, the strategy is easy to implement, computationally cost effective and might potentially be transferred to other RNA molecules or biomolecules to explore their conformation space.

5. Materials and Methods

5.1. System setup and MD simulations

We started with the crystal structure of the SAM-I aptamer domain bound with SAM from *Thermonaerobacter tengcongensis* organism (PDB ID: 3IQR).[47] The mutated residue 94 was changed back from adenine to guanine. Subsequently, we have built systems in three different conditions, i) in presence of 150 mM KCl and absence of SAM, ii) in presence of 150 mM KCl, 7.6 mM $MgCl_2$ and absence of SAM, iii) in presence of 150 mM KCl, 7.6 mM $MgCl_2$ and SAM present in the binding pocket. For the first and second systems, SAM was removed from the binding site, while for second and third systems five Ba^{2+} ions bound to SAM-I were replaced with Mg^{2+}. Then systems were solvated with the TIP3P water molecules and ions were added to match the above mentioned ionic conditions. During the equilibration stage, ions are equilibrated. Here, the bulk is defined as exclusion of RNA as well the surrounding water within the distance of 1.2 nm. To achieve the bulk ionic concentration of 150 mM KCl and 7.6 mM $MgCl_2$, we have iteratively changed the number of ions after following the below described equilibration protocol. We repeated the process until the bulk concentration of ions reaches closed to the desired ones. All the resulting systems were contained approximately 104,000 atoms.

Following the energy minimization step, all systems were equilibrated using the following strategy.[65–67] First, equilibration was conducted in the NVT ensemble for 1 ns with a time step of 1 fs using velocity rescaling thermostat[73] to keep the temperature at 300 K. At this step, all atoms of RNA and ligand were fixed applying the position restraints with a force constant of $k = 1000$ kJ mol^{-1} nm^{-2}. Then, 2 ns equilibration in NPT ensemble was conducted with 1 fs time step applying the position restraints same as previous step, where Parrinello-Rahman barostat[74] to maintain the pressure at 1 bar, and Nosé-Hoover thermostat[75] for keeping the temperature at 300 K were used. In following all steps, equilibration was conducted in NPT ensemble. At next step, 10 ns equilibration was carried out using 2 fs time step by keeping the remaining settings same as the previous step. Subsequently, the system was equilibrated during 20 ns where the position restraints were only applied to backbone atoms of RNA and all atoms of the SAM molecule. Then, another 100 ns equilibration was conducted with the position restraints kept only on the phosphorus atoms of RNA and all atoms of the SAM molecule. At last, 100 ns simulation was performed without any position restraints.

All MD simulations were performed using the GROMACS package 2021.4.[64] We have used amber ff99bsc0χ_{OL3} force field parameters for the RNA molecule.[76–79] For the SAM molecule, the force field parameters were obtained from the previous study.[80] Force field parameters for the ions K^+, Cl^- and Mg^{2+} (*MicroMg*) optimized with TIP3P water were taken from the recent studies.[81,82] The cut-off for the short-range electrostatics and the van der Waals interactions was set to 1.2 nm, and the particle-mesh Ewald method[83] was utilized for the long-range electrostatics with a grid size of 0.12 nm. Hydrogen containing bonds were constrained using the LINCS algorithm.[84]

5.2. WTmetaD simulations

The metadynamics simulations were performed using the PLUMED 2.7.2 plugin[63] patched to GROMACS package 2021.4.[64] WTmetaD simulations based strategy developed in the previous studies to investigate the antibiotics/substrates transport across the bacterial channels is implemented here.[65–67] A total of 8 walkers were used in WTmetaD simulations, and the simulation time was set to 250 ns for each walker, resulting in total time scale of 2 μs. Each calculation was carried out on the Chicoma supercomputer located at Los Alamos National Laboratory by distributing 8 walkers on 8 nodes, each containing 4 NVIDIA A100 GPUs and 1 AMD EPYC 7713 processor having 64 core @ 2 GHz. The computation time for each 2 μs simulation was approximately 3 days. For the SAM-I system with SAM bound in binding pocket and having 150 mM KCl and 7.6 mM $MgCl_2$, four independent WTmetaD simulations for timescale of 2 μs were performed having four collective variables (RMSD, Q, d_{mean} and Q_{TC}) being biased respectively. For all three SAM-I systems, simulations were conducted biasing the collective variable Q_{TC}, and repeated eight times for the same time duration of 2 μs. The free energy profiles estimated over the time revealed that the landscapes had very minimal variations from 1.5 to 2 μs in all three SAM-I setups (Figure S5 in SI). That suggests 2 μs timescale is enough to achieve converged free energy surfaces for these setups. All the metadynamics simulations related parameters are indicated in Table 1.

CRediT authorship contribution statement

Jigneshkumar Dahyabhai Prajapati: Conceptualization, Methodology, Visualization, Writing – original draft. **José N. Onuchic:** Writing – review & editing. **Karissa Y. Sanbonmatsu:** Writing – review & editing.

DATA AVAILABILITY

Data will be made available on request.

DECLARATION OF COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this pap

Acknowledgements

We acknowledge generous support from the National Institutes of Health (NIH) [RO1-GM110310 to J.N.O., RO1-GM110310 to K.Y.S.]. Generous allocation of computational resources on the Chicoma supercomputer by Los Alamos National Laboratory Institutional Computing is gratefully acknowledged.

Appendix A. Supplementary Data

Supplementary data includes computational results as well as the following videos depicting the conformation dynamics of SAM-I in the absence and presence of SAM molecule, respectively. Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.jmb.2022.167788.

Received 24 May 2022;
Accepted 7 August 2022;
Available online 11 August 2022

Keywords:
SAM-I Riboswitch;
metadynamics;
free energy;
collective variable;
contact distance

References

1. Li, P.T., Vieregg, J., Tinoco, I., et al., (2008). How rna unfolds and refolds. *Ann. Rev. Biochem.* **77** (1), 77–100. https://doi.org/10.1146/annurev.biochem.77.061206.174353.
2. Mortimer, S.A., Kidwell, M.A., Doudna, J.A., (2014). Insights into rna structure and function from genome-wide studies. *Nat. Rev. Genet.* **15** (7), 469–479. https://doi.org/10.1038/nrg3681.
3. Damase, T.R., Sukhovershin, R., Boada, C., Taraballi, F., Pettigrew, R.I., Cooke, J.P., (2021). The limitless future of rna therapeutics. *Front. Bioeng. Biotechnol.* **9** https://doi.org/10.3389/fbioe.2021.628137.
4. Kulkarni, J.A., Witzigmann, D., Thomson, S.B., Chen, S., Leavitt, B.R., Cullis, P.R., van der Meel, R., (2021). The current landscape of nucleic acid therapeutics. *Nat. Nanotechnol.* **16** (6), 630–643. https://doi.org/10.1038/s41565-021-00898-0.
5. Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G., (1995). Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **21** (3), 167–195. https://doi.org/10.1002/prot.340210302. URL http://www.ncbi.nlm.nih.gov/pubmed/7784423.
6. Clementi, C., Nymeyer, H., Onuchic, J.N., (2000). Topological and energetic factors: What determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. *J. Mol. Biol.* **298** (5), 937–953. https://doi.org/10.1006/jmbi.2000.3693.
7. Whitford, P.C., Schug, A., Saunders, J., Hennelly, S.P., Onuchic, J.N., Sanbonmatsu, K.Y., (2009). Nonlocal helix formation is key to understanding s-adenosylmethionine-1 riboswitch function. *Biophys. J.* **96** (2), L7–L9. https://doi.org/10.1016/j.bpj.2008.10.033. URL https://www.ncbi.nlm.nih.gov/pubmed/19167285.
8. Whitford, P.C., Noel, J.K., Gosavi, S., Schug, A., Sanbonmatsu, K.Y., Onuchic, J.N., (2009). An all-atom structure-based potential for proteins: bridging minimal models with all-atom empirical forcefields. *Proteins* **75** (2), 430–441. https://doi.org/10.1002/prot.22253. URL https://www.ncbi.nlm.nih.gov/pubmed/18837035.
9. Ratje, A.H., Loerke, J., Mikolajka, A., Brunner, M., Hildebrand, P.W., Starosta, A.L., Donhofer, A., Connell, S.R., Fucini, P., Mielke, T., Whitford, P.C., Onuchic, J.N., Yu, Y., Sanbonmatsu, K.Y., Hartmann, R.K., Penczek, P.A., Wilson, D.N., Spahn, C.M., (2010). Head swivel on the ribosome facilitates translocation by means of intra-subunit trna hybrid sites. *Nature* **468** (7324), 713–716. https://doi.org/10.1038/nature09547. URL https://www.ncbi.nlm.nih.gov/pubmed/21124459.
10. Whitford, P.C., Geggier, P., Altman, R.B., Blanchard, S.C., Onuchic, J.N., Sanbonmatsu, K.Y., (2010). Accommodation of aminoacyl-trna into the ribosome involves reversible excursions along multiple pathways. *RNA* **16** (6), 1196–1204. https://doi.org/10.1261/rna.2035410. URL https://www.ncbi.nlm.nih.gov/pubmed/20427512.
11. Hayes, R.L., Noel, J.K., Mandic, A., Whitford, P.C., Sanbonmatsu, K.Y., Mohanty, U., Onuchic, J.N., (2015). Generalized manning condensation model captures the rna ion atmosphere. *Phys. Rev. Lett.* **114** (25), 258105. https://doi.org/10.1103/PhysRevLett.114.258105. URL https://www.ncbi.nlm.nih.gov/pubmed/26197147.
12. Roy, S., Lammert, H., Hayes, R.L., Chen, B., LeBlanc, R., Dayie, T.K., Onuchic, J.N., Sanbonmatsu, K.Y., (2017). A magnesium-induced triplex pre-organizes the sam-ii riboswitch. *PLoS Comput. Biol.* **13** (3), e1005406. https://doi.org/10.1371/journal.pcbi.1005406. URL https://www.ncbi.nlm.nih.gov/pubmed/28248966.
13. Hayes, R.L., Noel, J.K., Whitford, P.C., Mohanty, U., Sanbonmatsu, K.Y., Onuchic, J.N., (2014). Reduced model captures mg2+-rna interaction free energy of riboswitches. *Biophys. J.* **106** (7), 1508–1519. https://doi.org/10.1016/j.bpj.2014.01.042.
14. Roy, S., Onuchic, J.N., Sanbonmatsu, K.Y., (2017). Cooperation between magnesium and metabolite controls collapse of the sam-i riboswitch. *Biophys. J.* **113** (2), 348–359. https://doi.org/10.1016/j.bpj.2017.06.044. URL https://www.ncbi.nlm.nih.gov/pubmed/28746845.

15. Roy, S., Hennelly, S.P., Lammert, H., Onuchic, J.N., Sanbonmatsu, K.Y., (2019). Magnesium controls aptamer-expression platform switching in the sam-i riboswitch. *Nucleic Acids Res.* **47** (6), 3158–3170. https://doi.org/10.1093/nar/gky1311.
16. Ma, H., Proctor, D.J., Kierzek, E., Kierzek, R., Bevilacqua, P.C., Gruebele, M., (2006). Exploring the energy landscape of a small rna hairpin. *J. Am. Chem. Soc.* **128** (5), 1523–1530. https://doi.org/10.1021/ja0553856.
17. Chen, S.-J., (2008). Rna folding: Conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* **37**, 197–214. https://doi.org/10.1146/annurev.biophys.37.032807.125957.
18. Sun, T.-T., Zhao, C., Chen, S.-J., (2018). Predicting cotranscriptional folding kinetics for riboswitch. *J. Phys. Chem. B* **122** (30), 7484–7496. https://doi.org/10.1021/acs.jpcb.8b04249.
19. Sugita, Y., Okamoto, Y., (1999). Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **314** (1), 141–151. https://doi.org/10.1016/S0009-2614(99)01123-9.
20. Garcia, A.E., Sanbonmatsu, K.Y., (2001). Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins* **42** (3), 345–354. https://doi.org/10.1002/1097-0134(20010215)42:3<345::aid-prot50>3.0.co;2-h. URL https://www.ncbi.nlm.nih.gov/pubmed/11151006.
21. Garcia, A.E., Sanbonmatsu, K.Y., (2002). Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc. Natl. Acad. Sci. USA* **99** (5), 2782–2787. https://doi.org/10.1073/pnas.042496899. URL https://www.ncbi.nlm.nih.gov/pubmed/11867710.
22. Sanbonmatsu, K.Y., Garcia, A.E., (2002). Structure of met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins* **46** (2), 225–234. https://doi.org/10.1002/prot.1167. URL https://www.ncbi.nlm.nih.gov/pubmed/11807951.
23. Yang, S., Onuchic, J.N., García, A.E., Levine, H., (2007). Folding time predictions from all-atom replica exchange simulations. *J. Mol. Biol.* **372** (3), 756–763. https://doi.org/10.1016/j.jmb.2007.07.010.
24. Wang, L., Friesner, R.A., Berne, B.J., (2011). Replica exchange with solute scaling: A more efficient version of replica exchange with solute tempering (rest2). *J. Phys. Chem. B* **115** (30), 9431–9438. https://doi.org/10.1021/jp204407d.
25. Mlýnský, V., Bussi, G., (2018). Exploring rna structure and dynamics through enhanced sampling simulations. *Curr. Opin. Struct. Biol.* **49**, 63–71. https://doi.org/10.1016/j.sbi.2018.01.004.
26. Laio, A., Parrinello, M., (2002). Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA* **99** (20), 12562.
27. Bussi, G., Laio, A., Parrinello, M., (2006). Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.* **96** (9), 090601. https://doi.org/10.1103/PhysRevLett.96.090601. URL https://www.ncbi.nlm.nih.gov/pubmed/16606249.
28. Bussi, G., Gervasio, F.L., Laio, A., Parrinello, M., (2006). Free-energy landscape for beta hairpin folding from combined parallel tempering and metadynamics. *J. Am. Chem. Soc.* **128** (41), 13435–13441. https://doi.org/10.1021/ja062463w. URL https://www.ncbi.nlm.nih.gov/pubmed/17031956.
29. Barducci, A., Bussi, G., Parrinello, M., (2008). Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100** (2), 020603. https://doi.org/10.1103/PhysRevLett.100.020603. URL https://www.ncbi.nlm.nih.gov/pubmed/18232845.
30. Branduardi, D., Bussi, G., Parrinello, M., (2012). Metadynamics with adaptive gaussians. *J. Chem. Theory Comput.* **8** (7), 2247–2254. https://doi.org/10.1021/ct3002464. URL https://www.ncbi.nlm.nih.gov/pubmed/26588957.
31. Gil-Ley, A., Bottaro, S., Bussi, G., (2016). Empirical corrections to the amber rna force field with target metadynamics. *J. Chem. Theory Comput.* **12** (6), 2790–2798. https://doi.org/10.1021/acs.jctc.6b00299. URL https://www.ncbi.nlm.nih.gov/pubmed/27153317.
32. Kuhrova, P., Best, R.B., Bottaro, S., Bussi, G., Sponer, J., Otyepka, M., Banas, P., (2016). Computer folding of rna tetraloops: Identification of key force field deficiencies. *J. Chem. Theory Comput.* **12** (9), 4534–4548. https://doi.org/10.1021/acs.jctc.6b00300. URL https://www.ncbi.nlm.nih.gov/pubmed/27438572.
33. Cunha, R.A., Bussi, G., (2017). Unraveling mg(2+)-rna binding with atomistic molecular dynamics. *RNA* **23** (5), 628–638. https://doi.org/10.1261/rna.060079.116. URL https://www.ncbi.nlm.nih.gov/pubmed/28148825.
34. Mlynsky, V., Bussi, G., (2017). Understanding in-line probing experiments by modeling cleavage of nonreactive rna nucleotides. *RNA* **23** (5), 712–720. https://doi.org/10.1261/rna.060442.116. URL https://www.ncbi.nlm.nih.gov/pubmed/28202709.
35. Kumar, S., Rosenberg, J., Bouzida, D., Swendsen, R., Kollman, P., (1992). The Weighted Histogram Analysis Method for Free-energy Calculations on Biomolecules. I. The Method. *J. Comput. Chem.* **13** (8), 1011–1021.
36. Darve, E., Rodriguez Gomez, D., Pohorille, A., (2008). Adaptive biasing force method for scalar and vector free energy calulations. *J. Chem. Phys.* **128** (14), 144120. https://doi.org/10.1063/1.2829861.
37. Yao, J., Reinharz, V., Major, F., Waldispuhl, J., (2017). Rna-moip: prediction of rna secondary structure and local 3d motifs from sequence data. *Nucleic Acids Res.* **45** (W1), W440–W444. https://doi.org/10.1093/nar/gkx429. URL https://www.ncbi.nlm.nih.gov/pubmed/28525607.
38. Dallaire, P., Major, F., (2016). Exploring alternative rna structure sets using mc-flashfold and db2cm. *Methods Mol. Biol.* **1490**, 237–251. https://doi.org/10.1007/978-1-4939-6433-8_15. URL https://www.ncbi.nlm.nih.gov/pubmed/27665603.
39. Reinharz, V., Major, F., Waldispuhl, J., (2012). Towards 3d structure prediction of large rna molecules: an integer programming framework to insert local 3d motifs in rna secondary structure. *Bioinformatics* **28** (12), i207–i214. https://doi.org/10.1093/bioinformatics/bts226. URL https://www.ncbi.nlm.nih.gov/pubmed/22689763.
40. Parisien, M., Major, F., (2012). Determining rna three-dimensional structures using low-resolution data. *J. Struct. Biol.* **179** (3), 252–260. https://doi.org/10.1016/j.jsb.2011.12.024. URL https://www.ncbi.nlm.nih.gov/pubmed/22387042.
41. Yang, S., Parisien, M., Major, F., Roux, B., (2010). Rna structure determination using saxs data. *J. Phys. Chem. B* **114** (31), 10039–10048. https://doi.org/10.1021/jp1057308. URL https://www.ncbi.nlm.nih.gov/pubmed/20684627.

42. Sponer, J., Bussi, G., Krepl, M., Banáš, P., Bottaro, S., Cunha, R.A., Gil-Ley, A., Pinamonti, G., Poblete, S., Jurecka, P., et al., (2018). Rna structural dynamics as captured by molecular simulations: A comprehensive overview. *Chem. Rev.* **118** (8), 4177–4338. https://doi.org/10.1021/acs.chemrev.7b00427.
43. Yu, T., Chen, S.-J., (2018). Hexahydrated mg2+ binding and outer-shell dehydration on rna surface. *Biophys. J.* **114** (6), 1274–1284. https://doi.org/10.1016/j.bpj.2018.01.040.
44. Hennelly, S.P., Sanbonmatsu, K.Y., (2011). Tertiary contacts control switching of the sam-i riboswitch. *Nucleic Acids Res.* **39** (6), 2416–2431. https://doi.org/10.1093/nar/gkq1096.
45. Hennelly, S.P., Novikova, I.V., Sanbonmatsu, K.Y., (2013). The expression platform and the aptamer: Cooperativity between mg2+ and ligand in the sam-i riboswitch. *Nucleic Acids Res.* **41** (3), 1922–1935. https://doi.org/10.1093/nar/gks978.
46. Montange, R.K., Batey, R.T., (2006). Structure of the s-adenosylmethionine riboswitch regulatory mrna element. *Nature* **441** (7097), 1172–1175. https://doi.org/10.1038/nature04819.
47. Stoddard, C.D., Montange, R.K., Hennelly, S.P., Rambo, R.P., Sanbonmatsu, K.Y., Batey, R.T., (2010). Free state conformational sampling of the sam-i riboswitch aptamer domain. *Structure* **18** (7), 787–797. https://doi.org/10.1016/j.str.2010.04.006.
48. Winkler, W.C., Breaker, R.R., (2003). Genetic control by metabolite-binding riboswitches. *Chembiochem* **4** (10), 1024–1032.
49. Nudler, E., Mironov, A.S., (2004). The riboswitch control of bacterial metabolism. *Trends Biochem. Sci.* **29** (1), 11–17. https://doi.org/10.1016/j.tibs.2003.11.004.
50. Grundy, F.J., Henkin, T.M., (1998). The s box regulon: A new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol. Microbiol.* **30** (4), 737–749. https://doi.org/10.1046/j.1365-2958.1998.01105.x.
51. Epshtein, V., Mironov, A.S., Nudler, E., (2003). The riboswitch-mediated control of sulfur metabolism in bacteria. *Proc. Nat. Acad. Sci.* **100** (9), 5052–5056. https://doi.org/10.1073/pnas.0531307100.
52. McDaniel, B.A.M., Grundy, F.J., Artsimovitch, I., Henkin, T. M., (2003). Transcription termination control of the s box system: Direct measurement of s-adenosylmethionine by the leader rna. *Proc. Nat. Acad. Sci.* **100** (6), 3083–3088. https://doi.org/10.1073/pnas.0630422100.
53. Winkler, W.C., Nahvi, A., Sudarsan, N., Barrick, J.E., Breaker, R.R., (2003). An mrna structure that controls gene expression by binding s-adenosylmethionine. *Nat. Struct. Mol. Biol.* **10** (9), 701–707. https://doi.org/10.1038/nsb967.
54. Lu, C., Ding, F., Chowdhury, A., Pradhan, V., Tomsic, J., Holmes, W.M., Henkin, T.M., Ke, A., (2010). Sam recognition and conformational switching mechanism in the bacillus subtilis yitj s box/sam-i riboswitch. *J. Mol. Biol.* **404** (5), 803–818. https://doi.org/10.1016/j.jmb.2010.09.059.
55. Heppell, B., Blouin, S., Dussault, A.-M., Mulhbacher, J., Ennifar, E., Penedo, J.C., Lafontaine, D.A., (2011). Molecular insights into the ligand-controlled organization of the sam-i riboswitch. *Nat. Chem. Biol.* **7** (6), 384–392. https://doi.org/10.1038/nchembio.563.
56. Hayes, R.L., Noel, J.K., Mohanty, U., Whitford, P.C., Hennelly, S.P., Onuchic, J.N., Sanbonmatsu, K.Y., (2012). Magnesium fluctuations modulate rna dynamics in the sam-i riboswitch. *J. Am. Chem. Soc.* **134** (29), 12043–12053. https://doi.org/10.1021/ja301454u.
57. Manz, C., Kobitski, A.Y., Samanta, A., Keller, B.G., Jäschke, A., Nienhaus, G.U., (2017). Single-molecule fret reveals the energy landscape of the full-length sam-i riboswitch. *Nat. Chem. Biol.* **13** (11), 1172–1178.
58. Manz, C., Kobitski, A.Y., Samanta, A., Nienhaus, K., Jäschke, A., Nienhaus, G.U., (2021). Exploring the energy landscape of a sam-i riboswitch. *J. Biol. Phys.* **47** (4), 371–386. https://doi.org/10.1007/s10867-021-09584-7.
59. Barducci, A., Bussi, G., Parrinello, M., (2008). Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **100** (2), 020603. https://doi.org/10.1103/PhysRevLett.100.020603.
60. Dama, J.F., Parrinello, M., Voth, G.A., (2014). Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.* **112** (24), 240602. https://doi.org/10.1103/physrevlett.112.240602.
61. Tiwary, P., Dama, J.F., Parrinello, M., (2015). A perturbative solution to metadynamics ordinary differential equation. *J. Chem. Phys.* **143** (23), 234112. https://doi.org/10.1063/1.4937945.
62. Raiteri, P., Laio, A., Gervasio, G.L., Micheletti, C., Parrinello, M., (2006). Efficient reconstruction of complex free energy landscapes by multiple walkers metadynamics. *J. Chem. Phys. B* **110** (8), 3533–3539. https://doi.org/10.1021/jp054359r.
63. Tribello, G.A., Bonomi, M., Branduardi, D., Camilloni, C., Bussi, G., (2014). PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **185** (2), 604–613. https://doi.org/10.1016/j.cpc.2013.09.018.
64. Hess, B., Kutzner, C., Spoel, D.V.D., Lindahl, E., (2008). GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.* **4** (3), 435–447.
65. Prajapati, J.D., Solano, C.J.F., Winterhalter, M., Kleinekathöfer, U., (2017). Characterization of ciprofloxacin permeation pathways across the porin ompc using metadynamics and a string method. *J. Chem. Theory Comput.* **13**, 4553–4566. https://doi.org/10.1021/acs.jctc.7b00467.
66. Prajapati, J.D., Solano, C.J.F., Winterhalter, M., Kleinekathöfer, U., (2018). Enrofloxacin permeation pathways across the porin ompc. *J. Phys. Chem. B* **122** (4), 1417–1426. https://doi.org/10.1021/acs.jpcb.7b12568.
67. Golla, V.K., Prajapati, J.D., Joshi, M., Kleinekathöfer, U., (2020). Exploration of free energy surfaces across a membrane channel using metadynamics and umbrella sampling. *J. Chem. Theory Comput.* **16** (4), 2751–2765. https://doi.org/10.1021/acs.jctc.9b00992.
68. Tiwary, P., Parrinello, M., (2014). A time-independent free energy estimator for metadynamics. *J. Phys. Chem. B* **119** (3), 736–742. https://doi.org/10.1021/jp504920s.
69. Atkovska, K., Hub, J.S., (2017). Energetics and mechanism of anion permeation across formate-nitrite transporters. *Sci. Rep.* **7** (1), 12027. https://doi.org/10.1038/s41598-017-11437-0.
70. Hardin, C., Eastwood, M.P., Prentiss, M.C., Luthey-Schulten, Z., Wolynes, P.G., (2003). Associative memory hamiltonians for structure prediction without homology: alpha/beta proteins. *Proc. Natl. Acad. Sci. USA* **100** (4), 1679–1684. https://doi.org/10.1073/pnas.252753899. URL https://www.ncbi.nlm.nih.gov/pubmed/12554830.

71. Best, R.B., Hummer, G., Eaton, W.A., (2013). Native contacts determine protein folding mechanisms in atomistic simulations. *Proc. Nat. Acad. Sci.* **110** (44), 17874–17879. https://doi.org/10.1073/pnas.1311599110.
72. Tsai, S.-T., Smith, Z., Tiwary, P., (2019). Reaction coordinates and rate constants for liquid droplet nucleation: Quantifying the interplay between driving force and memory. *J. Chem. Phys.* **151** (15), 154106. https://doi.org/10.1063/1.5124385.
73. Bussi, G., Donadio, D., Parrinello, M., (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126** (1), 014101. https://doi.org/10.1063/1.2408420.
74. Parrinello, M., Rahman, A., (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52** (12), 7182–7190.
75. Cheng, A., Merz, K.M., (1996). Application of the Nosé-hoover chain algorithm to the study of protein dynamics. *J. Phys. Chem.* **100** (5), 1927–1937.
76. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K. M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., et al., (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **117** (19), 5179–5197. https://doi.org/10.1021/ja00124a002. arXiv:https://doi.org/10.1021/ja00124a002.
77. Wang, J., Cieplak, P., Kollman, P.A., (2000). How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **21** (12), 1049–1074.
78. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham III, T.E., Laughton, C.A., Orozco, M., (2007). Refinement of the amber force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.* **92** (11), 3817–3829. https://doi.org/10.1529/biophysj.106.097782.
79. Zgarbová, M., Otyepka, M., Šponer, J., Mládek, A., Banáš, P., Cheatham III, T.E., Jurecka, P., (2011). Refinement of the cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7** (9), 2886–2902. https://doi.org/10.1021/ct200162x.
80. Saez, D.A., Vöhringer-Martinez, E., (2015). A consistent s-adenosylmethionine force field improved by dynamic hirshfeld-i atomic charges for biomolecular simulation. *J. Comput.-Aided Mol. Des.* **29** (10), 951–961. https://doi.org/10.1007/s10822-015-9864-1.
81. Mamatkulov, S., Schwierz, N., (2018). Force fields for monovalent and divalent metal cations in tip3p water based on thermodynamic and kinetic properties. *J. Chem. Phys.* **148** (7), 074504. https://doi.org/10.1063/1.5017694.
82. Grotz, K.K., Cruz-León, S., Schwierz, N., (2021). Optimized magnesium force field parameters for biomolecular simulations with accurate solvation, ion-binding, and water-exchange properties. *J. Chem. Theory Comput.* **17** (4), 2530–2540. https://doi.org/10.1021/acs.jctc.0c01281.
83. Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., Pedersen, L.G., (1995). A smooth particle mesh ewald method. *J. Chem. Phys.* **103** (19), 8577–8593. https://doi.org/10.1063/1.470117.
84. Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., (1997). LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **18** (12), 1463–1472. https://doi.org/10.1002/(sici)1096-987x(199709)18:12<1463::aid-jcc4>3.3.co;2-l.