Mechanism of Ligand Binding to Theophylline RNA Aptamer

Sana Akhter,

Computational Biology Program and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66047, United States

Zhichao Tang,

Department of Medicinal Chemistry, University of Kansas, Lawrence, Kansas 66047, United States

Jinan Wang,

Computational Biology Program and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66047, United States

Mercy Haboro,

Department of Medicinal Chemistry, University of Kansas, Lawrence, Kansas 66047, United States

Erik D Holmstrom,

Department of Molecular Biosciences and Department of Chemistry, University of Kansas, Lawrence, Kansas 66045, United States

Jingxin Wang,

Department of Medicinal Chemistry, University of Kansas, Lawrence, Kansas 66047, United States

Yinglong Miao

Computational Biology Program and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66047, United States; Present Address: Department of Pharmacology and

Corresponding Authors: Jingxin Wang – *Department of Medicinal Chemistry, University of Kansas, Lawrence, Kansas 66047, United States;* wang.jingxin@ku.edu, **Yinglong Miao** – *Computational Biology Program and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66047, United States;* Present Address: Department of Pharmacology and Computational Medicine Program, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina 27599, United States (Y.M.); Yinglong_Miao@med.unc.edu.

The authors declare no competing financial interest.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.3c01454.

Example of input simulation parameter used in dua-lboost LiGaMD simulation, summary of simulated TEP and CFF systems, associated time courses free energy profiles (PDF)

TEP diffuses from outside into the pocket through a pathway via π–π stacking interaction with the G26 nucleotide in the aptamer's loop region (MP4)

TEP forms π-stacking interaction with the extended C27 in the "Out" conformation, as shown in the ligand trace (MP4)

As TEP passes through the stem of the aptamer, it forms stacking interactions with nucleotides A5 and C30 (MP4)

Ligand moves out of the binding pocket via stacking interaction with G26 (MP4)

CFF typically forms stacking interaction with the solvent-exposed C27 and diffuses inward through a gap between G26 and A28 into the RNA target site (MP4)

Ligand makes its way out into the solvent through the gap between G26 and A28 of the RNA, through which it first made its way inside (MP4)

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.3c01454

Computational Medicine Program, University of North Carolina – Chapel Hill, Chapel Hill, North Carolina 27599, United States (Y.M.)

Abstract

Studying RNA-ligand interactions and quantifying their binding thermodynamics and kinetics are of particular relevance in the field of drug discovery. Here, we combined biochemical binding assays and accelerated molecular simulations to investigate ligand binding and dissociation in RNA using the theophylline-binding RNA as a model system. All-atom simulations using a Ligand Gaussian accelerated Molecular Dynamics method (LiGaMD) have captured repetitive binding and dissociation of theophylline and caffeine to RNA. Theophylline's binding free energy and kinetic rate constants align with our experimental data, while caffeine's binding affinity is over 10,000 times weaker, and its kinetics could not be determined. LiGaMD simulations allowed us to identify distinct low-energy conformations and multiple ligand binding pathways to RNA. Simulations revealed a "conformational selection" mechanism for ligand binding to the flexible RNA aptamer, which provides important mechanistic insights into ligand binding to the theophylline-binding model. Our findings suggest that compound docking using a structural ensemble of representative RNA conformations would be necessary for structure-based drug design of flexible RNA.

Graphical Abstract
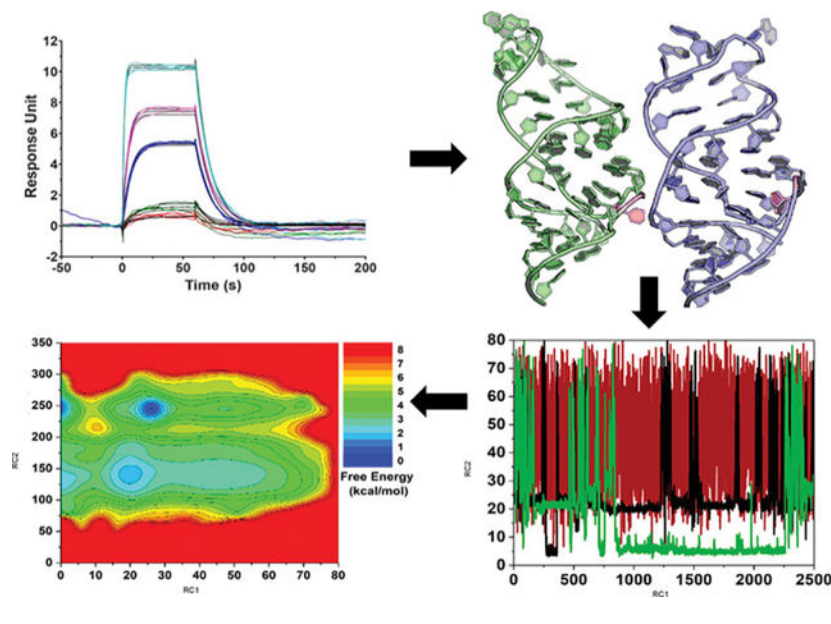
INTRODUCTION

RNAs undergo conformational transitions for various essential cellular functions such as RNA splicing, translation, and enzymatic activity.[1] In bacteria, archaea, fungi, yeast, and plants naturally occurring small molecules regulate gene expression through conformational changes in riboregulatory elements in mRNAs, namely, riboswitches.[2] Inspired by the interactions between RNAs and small molecules in nature, many research laboratories

have started to develop synthetic RNA-binding molecules to regulate RNA splicing and translation.[3–6] Fueled by these advances, RNA has also become an increasingly popular therapeutic target for drug development.[3,4,7,8] However, targeting RNA by small molecules is challenging because of its unique structural features (such as the polyanionic backbone) and high flexibility.[8,9] To better understand the nature of RNA-ligand interactions, researchers often study short single-stranded oligonucleotides called RNA aptamers, which specifically recognize small molecules and metabolites.[10]

The 33-nucleotide RNA structure of the theophylline (TEP)-binding aptamer RNA is one of the most studied RNA aptamers in the field.[11] The RNA aptamer has a binding affinity for TEP of $K_d \approx 0.4$ μM in the presence of divalent Mg^{2+} ions.[12] The TEP-binding pocket is formed by 14 highly conserved nucleotides.[11,13] The aptamer has the ability to discriminate between TEP and a closely related compound, caffeine (CFF), which only has one additional methyl group compared to TEP.[13] When compared to TEP, the binding affinity for CFF is more than 10,000 times less.[13] As a result of ligand interaction, the RNA experiences conformational changes.[12,13] Upon TEP binding, an unpaired base, C27, adopts a flipped conformation that is pointed away from the binding pocket.[12]

Conventional Molecular Dynamics (cMD) is a powerful technique for simulating biomolecular dynamics at an atomistic level.[4] Despite remarkable computational advances in the past two decades, cMD studies are typically limited to microsecond time scales and have not been able to sufficiently sample many biological processes of interest, such as small molecule binding and dissociation.[5,6,14] Many different enhanced sampling techniques have been developed to overcome the above challenge in bimolecular simulations.[5] Artificial intelligence-augmented Metadynamics simulations were applied to detect ligand dissociation from a riboswitch system for cognate and synthetic ligands.[15] Dissociation of acetylpromazine from the TAR-RNA binding pocket was simulated using nonequilibrium Steered MD.[16] We have applied Gaussian accelerated MD (GaMD) simulations to uncover the binding mechanism of small-molecule splicing modulators that interact with single-stranded RNA.[17] Recently, a new Ligand GaMD (LiGaMD) algorithm has been developed for more efficient sampling of ligand binding and dissociation processes, which allows for calculations of both the thermodynamics and kinetics of ligand binding.[6]

In this study, we performed all-atom enhanced sampling simulations using LiGaMD to study TEP and CFF binding to the TEP RNA aptamer. Microsecond-time scale LiGaMD simulations successfully captured the repetitive binding and unbinding of both methylxanthines into the experimentally determined ligand binding pocket, allowing us to characterize the free energy landscapes of the small-molecule RNA interactions. The kinetics and thermodynamics of interactions between the ligands and the RNA aptamer were also investigated by using an experimental SPR technique. LiGaMD simulations produced kinetic rate constants and binding free energies for the TEP that were in good agreement with our experimental results.

METHODS

Surface Plasmon Resonance (SPR).

SPR experiments were performed on a Biacore T200 (GE Healthcare) instrument at 25 °C using streptavidin-precoated SA sensor chips (GE Healthcare). The running buffer, which contained 20 mM HEPES, 100 mM NaCl, and 5 mM $MgCl_2$ at pH 7.3, was freshly made and used after being filtered through a 0.22 μm PVDF membrane. GenScript, Inc. synthesized the 3′-biotinylated TEP aptamer RNA (i.e., GGCGAUACCAGCCGAAAGGCCCUUGGCAGCGUCUU/3′Biotin/), which we reconstituted at 0.1 mM in molecular biology water. For immobilization of the biotinylated RNA, the sensor chip was first conditioned with 3 consecutive 1 min injections of a high salt solution (50 mM NaOH, 1 M NaCl) at a flow rate of 50 μL/min. Next, the biotinylated RNA was diluted in the running buffer (1 μM) and applied over the streptavidin sensor chip surface at a flow rate of 10 μL/min to achieve an immobilization level of about 1000 RU. Finally, alkyne-PEG-biotin (50 μM in running buffer) was injected (1 min, 10 μL/min) to block the remaining streptavidin surface binding sites. The BiaControl Software Wizard Kinetics methodology was used for the kinetic analysis. In order to titrate over the immobilized RNA (contact time: 1 min, flow rate: 50 μL/min), the TEP or CFF was dissolved in the running buffer to the necessary concentrations (0.032, 0.064, 0.32, 0.64, and 3.2 μM for the TEP, and 0.35, 0.7, 3.5, 7, and 35 mM for the CFF).

The data analysis and plotting were performed by using BiaEvaluation Software. All monitored resonance signals were subtracted with signals from a nonbinding reference channel. Equilibrium and rate constants (K_d, k_{on}, and k_{off}) were calculated using the BiaEvaluation Software Binding Affinity protocol with 1:1 fitting.

Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD).

Ligand Gaussian accelerated Molecular Dynamics (LiGaMD) is a computational method developed to enhance the sampling of ligand binding and dissociation in target receptors. Here, we briefly outline the algorithm of LiGaMD.[6]

We consider a scenario in which a ligand L binds to a receptor R in biological environment E. N atoms with their locations make up the system $r \equiv \left\{ \overrightarrow{r}_1, \cdots, \overrightarrow{r_N} \right\}$ and momenta $p \equiv \left\{ \overrightarrow{p_1}, \cdots, \overrightarrow{p_N} \right\}$. The Hamiltonian of the system can be expressed as

$$H(r, p) = K(p) + V(r)$$

(1)

where $K(p)$ and $V(r)$ are the system kinetic and total potential energies, respectively. Next, the potential energy is broken down into the following terms:

$$V(r) = V_{R,b}(r_P) + V_{L,b}(r_L) + V_{E,b}(r_E) + V_{RR,nb}(r_P) \\ + V_{LL,nb}(r_L) + V_{EE,nb}(r_E) + V_{RL,nb}(r_{PL}) \\ + V_{RE,nb}(r_{PE}) + V_{LE,nb}(r_{LE})$$

(2)

where $V_{R,b}$, $V_{L,b}$, and $V_{E,b}$ are the bonded potential energies in the receptor R, ligand L and environment E, respectively. The self-nonbonded potential energies in receptor R, Ligand L, and Environment E, respectively, are $V_{RR,nb}$, $V_{LL,nb}$ and $V_{EE,nb}$. R − L, R − E, and L − E's related nonbonded interaction energies are $V_{RL,nb}$, $V_{RE,nb,}$, and $V_{LE,nb,}$ respectively.[6] According to classical molecular mechanics force fields,[18] the nonbonded potential energies are calculated as

$$V_{nb} = V_{elec} + V_{vdW}$$

(3)

where V_{elec} and V_{vdW} are the system electrostatic and van der Waals potential energies.[6] Ligand binding mainly involves the nonbonded interaction energies of the ligand, $V_{L,nb}(r) = V_{LL,nb}(r_L) + V_{RL,nb}(r_{PL}) + V_{LE,nb}(r_{LE})$. We selectively provide a boost potential to the ligand nonbonded potential energy in LiGaMD as

$$\Delta V_{L,nb}(r) = \begin{cases} \frac{1}{2}k_{L,nb}(E_{L,nb} - V_{L,nb}(r))^2, V_{L,nb}(r) < E_{L,nb} \\ 0, V_{L,nb}(r) \geq E_{L,nb} \end{cases}$$

(4)

where $E_{L,nb}$ is the threshold energy for applying boost potential and $k_{L,nb}$ is the harmonic constant. The subscript of $\Delta V_{L,nb}(r)$, $E_{L,nb}$, and $k_{L,nb}$ is dropped in the section that follows for the sake of simplicity. The simulation parameters E and k are determined according to the GaMD enhanced sampling principles. [6]

When the system's maximum potential energy E is set to the lower bound ($E = V_{max}$), the effective harmonic force constant k_0 can be calculated as

$$k_0 = \min(1.0, k_0') = \min\left(1.0, \frac{\sigma_0}{\sigma_V} \frac{V_{max} - V_{min}}{V_{max} - V_{avg}}\right)$$

(5)

where V_{max}, V_{min}, V_{avg} and σ_V are the maximum, minimum, average, and standard deviation of the boosted system potential energy, and σ_0 is the user-specified upper limit of the standard deviation of ΔV (e.g., $10k_BT$) for proper reweighting.[6] The upper limit of the standard deviation of the first potential boost and second potential boost that allows for accurate reweighting in dual-boost simulations is referred to as σ_{0P} and σ_{0D}, respectively. [6] The harmonic constant is calculated as $k = k_0 \cdot \frac{1}{V_{max} - V_{min}}$ with $0 < k_0 \leq 1$. Alternatively, when the threshold energy E is set to its upper bound $E = V_{min} + \frac{1}{k}$, k_0 is set to

$$k_0 = k_0'' \equiv \left(1 - \frac{\sigma_0}{\sigma_V}\right)\frac{V_{max} - V_{min}}{V_{avg} - V_{min}}$$

(6)

if $k_0^{''}$ is found to be between 0 and 1. Otherwise, k_0 is calculated using eq 5.

In order to facilitate ligand binding to RNAs in MD simulations, additional ligand molecules might be added to the solvent.[19] This is based on the fact that the time needed for ligand binding is inversely proportional to the ligand concentration.[20] The higher the ligand concentration, the faster the ligand binds, provided that the ligand concentration is still within its solubility limits.[6,20] In addition to selectively boosting the bound ligand to accelerate its dissociation, another boost potential is applied to the unbound ligand molecules, RNA, and solvent to facilitate ligand rebinding.[6] The second boost potential is calculated using the total system potential energy other than the nonbonded potential energy of the bound ligand as

$$\Delta V_D(r) = \begin{cases} \frac{1}{2}k_D(E_D - V_D(r))^2, V_D(r) < E_D \\ \\ 0, V_D(r) \geq E_D \end{cases}$$

(7)

where E_D and k_D are the corresponding threshold energy for applying the second boost potential and the harmonic constant, respectively. This leads to dual-boost LiGaMD with the total boost potential $\Delta V(r) = \Delta V_{L,nb}(r) + \Delta V_D(r)$.[6]

Ligand Binding Free Energy Calculations from 3D Potential of Mean Force.

From the 3D potential of mean force (PMF) of ligand displacements from the target receptor, we calculate the ligand binding free energy as[21]

$$\Delta G^\circ = -\Delta W_{3D} - RT\ln\frac{V_b}{V_0}$$

(8)

where V_0 is the standard volume, $V_b = \int_b e^{-\beta W(r)}dr$ is the average sampled bound volume of the ligand with $\beta = 1/k_B T$, k_B is the Boltzmann constant, T is the temperature, and ΔW_{3D} is the depth of the 3D PMF. ΔW_{3D} can be calculated by integrating Boltzmann distribution of the 3D PMF $W(r)$ overall system coordinates except the x, y, z of the ligand:

$$\Delta W_{3D} = -RT\ln\frac{\int_u e^{-\beta W(r)}dr}{\int_u dr}$$

(9)

where $V_u = \int_u dr$ is the sampled unbound volume of the ligand. The exact definitions of the bound and unbound volumes V_b and V_u are not important as the exponential average cut-off contributions are far away from the PMF minima.[21] The PyReweighting tool kit (http://miao.compbio.ku.edu/PyReweighting/) includes a free Python script called PyReweighting-3D.py that can be used to compute the 3D PMF and related ligand binding

free energies. It functions for both enhanced sampling simulations using LiGaMD with energetic reweighting and cMD (without energetic reweighting).[6]

Ligand Binding Kinetics Obtained from Reweighting of LiGaMD Simulations.

One can record the time periods and compute their averages for the ligand found in the bound (τ_B) and unbound (τ_U) states from the simulation trajectories, assuming adequate sampling of repetitive ligand dissociation and binding.[20] The τ_B corresponds to residence time in drug design.[6,22] The binding and dissociation rate constants (k_{off} and k_{on}) for ligands were calculated as

$$k_{off} = \frac{1}{\tau_B}$$

(10)

$$k_{on} = \frac{1}{\tau_U \cdot [L]}$$

(11)

where [L] is the ligand concentration in the simulation system.

The ligand kinetics from the LiGaMD simulations are reweighted using Kramers' rate theory. According to Kramers' rate theory, the rate of a chemical reaction in the large viscosity limit is determined by[23]

$$k_R \cong \frac{2\pi w_m w_b}{\xi} e^{-\Delta F/k_B T}$$

(12)

where w_m and w_b are frequencies of the approximated harmonic oscillators (also referred to as curvatures of free energy surface[24,25] near the energy minimum and barrier, respectively, ξ is the frictional rate constant, and ΔF is the free energy barrier of transition. The friction constant ξ is related to the diffusion coefficient D with $\xi = k_B T / D$. The apparent diffusion coefficient D is calculated by dividing the kinetic rate calculated using the transition time series directly obtained from simulations by the probability density solution of the Smoluchowski equation.[26] The free energy barriers of ligand binding and dissociation are calculated from the original LiGaMD simulations in order to reweight ligand kinetics using Kramer's rate theory (reweighted, ΔF) and modified (no reweighting, ΔF^*) PMF profiles, similarly for curvatures of the reweighed (w) and modified (w^*, no reweighting) PMF profiles near the guest bound ("B") and unbound ("U") low-energy wells and the energy barrier ("Br"), and the ratio of apparent diffusion coefficients from LiGaMD simulations without reweighting (modified, D^*) and with reweighting (D).[20] The resulting numbers are then plugged into eq 9 to estimate accelerations of the ligand binding and dissociation rates during LiGaMD simulations, which allows us to recover the original kinetic rate constants.[23]

System Setup.

The NMR structure of the TEP-bound RNA complex (PDB ID: 1O15) was obtained for simulation.[27,28] To create the RNA-CFF complex, a methyl group was added to the N7 atom of TEP in the NMR structure, while keeping the coordinates of the rest of the receptor. The system was energy minimized to remove possible steric clashes. The RNA.OL3[29,30] force field was employed for RNA and GAFF2[31] for small molecules with the *tleap* module in the AMBER20 package.[32] To replicate the conditions of the solution structure in a TIP3P[33] water box using *tleap*, both systems were neutralized with 0.01 M $MgCl_2$[34] and 0.15 M NaCl.[32,35] A total of 3 ligands (one in the NMR bound conformation and 2 others placed randomly in the bulk solvent outside at a distance of >15 Å from the RNA surface) were included. The design is based on the fact that the time needed for ligand binding is inversely proportional to the ligand concentration.[20] The greater the ligand concentration, the less time is required for ligand binding, provided that the ligand concentration is within solubility limits.[20]

LiGaMD Simulation Protocol.

The GPU-accelerating program *pmemd.cuda* in AMBER20 was used to perform LiGaMD simulations.[32] The simulation system was minimized by using the steepest descent algorithm. The system was then heated from 0 to 300 K for 200 ps. It was further equilibrated using the NVT ensemble at 300 K for 800 ps and the NPT ensemble at 300 K and 1 bar for 1 ns with 1 kcal/mol/Å^2 constraints on the heavy atoms of the receptor and ligand, followed by a 2 ns short cMD without any constraint. The LiGaMD simulations proceeded with a 10 ns short cMD to collect the potential statistics, a 63 ns GaMD equilibration after adding the boost potential, and then three independent 2500 ns production runs for TEP. After a 63 ns equilibration run with similar parameters for cMD, five separate production runs for CFF were extended to 5000 ns.

The threshold energy for applying the ligand essential potential boost was also set to the upper bound in the LiGaMD simulation for both TEP and CFF. The selective boost potential was applied to the ligand. For the second boost potential applied to the system's total potential energy other than the ligand's essential potential energy, sufficient acceleration was obtained by setting the threshold energy to the lower bound. We aimed to observe ligand dissociation during equilibration while maintaining a low boost potential for accurate energetic reweighting; the $(\sigma_{0P}, \sigma_{0D})$ parameters were finally set to (4.0, 3.0 kcal/mol) for both systems (Figures S1 and S2). LiGaMD production simulation frames were saved every 0.4 ps for analysis.

Simulation Analysis.

The VMD and CPPTRAJ tools were used for simulation analysis.[36,37] The number of ligand dissociation and binding events (N_D and N_B) and the ligand binding and unbinding time periods (τ_B and τ_U) were recorded from individual simulations for TEP (Tables 1, S1 and S2) and CFF (Tables 2, S3 and S4) simulations. With high fluctuations, (τ_B and τ_U) were recorded for only time periods longer than 1 ns.[6] The 1D, 2D, and 3D PMF profiles as well as the ligand binding free energy were calculated through energetic reweighting of the LiGaMD

simulations. The distance between the N4 atom of the C8 residue of the nucleic acid and the N7 atom of TEP and CFF was used as a reaction coordinate for calculating 1D PMF. The ligand heavy atom RMSD relative to the NMR bound pose was explored as a reaction coordinate for calculating the 1D PMF profile. The bin size was set to 1.0 Å for the atom distances and RMSD calculations. 2D PMF profiles were calculated for conformational changes in RNA upon ligand binding. We used the RMSD of the ligand relative to the NMR structure and the Solvent Accessible Surface Area (SASA) of nucleotide C27 on the RNA. The bin size for the latter was set to 5.0 Å^2. The cutoff for the number of simulation frames in one bin was set to 500 for reweighing of 1D and 2D PMF profiles. The 3D PMF profiles of ligand displacements from the RNA host in the X, Y, and Z directions were further calculated from the LiGaMD simulations. The bin sizes were set to 1 Å in the X, Y, and Z directions. The cutoff of simulation frames in one bin for 3D PMF reweighting (ranging from 100 to 500 for three individual LiGaMD simulations) was set to the minimum number below which the calculated 3D PMF minimum will be shifted. Furthermore, the ligand binding free energies (ΔG) were calculated using the reweighted 3D PMF profiles and binding kinetic rates by $\Delta G = -RT \ln(k_{off}/k_{on})$, respectively.[20] The ligand dissociation and binding rate constants (k_{on} and k_{off}) were calculated from the LiGaMD simulations, with their accelerations analyzed using Kramers' rate theory.[20,24,38]

Furthermore, structural clustering of the trajectories was performed for each 2.5 μs (TEP) and 5 μs simulation (CFF) using the Hierarchical Agglomerative clustering algorithm[39] in CPPTRAJ.[37] The frames were sieved at a stride of 1000 s for clustering. The remaining frames were assigned to the closest cluster afterward. The RMSD cutoff for clustering was set to 3.5 Å. The resulting structural clusters were reweighted to obtain energetically significant binding pathways for the ligand. In addition, the ligand dissociation and binding rate constants (k_{on} and k_{off}) were calculated from the LiGaMD simulations, with their accelerations analyzed using the Kramers' rate theory as described above (Table S5).

RESULTS

SPR Analysis of Theophylline and Caffeine Binding to the RNA Aptamer.

The kinetics and thermodynamics of interactions between the ligands and the RNA aptamer were first examined by using an experimental SPR technique. Using the established protocol, we immobilized the biotinylated TEP RNA aptamer on a streptavidin SPR chip and titrated different concentrations of ligands in solution (Figure 1A). From kinetic fitting, we determined that $k_{on} = (1.92 \pm 0.37) \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$ and $k_{off} = (0.081 \pm 0.45) \text{ s}^{-1}$, resulting in a $K_d = 0.42$ μM. Notably, these values are consistent with the previously reported values.[2,40,41] For CFF, due to weak binding affinity to the TEP aptamer, k_{on} and k_{off} could not be determined accurately from kinetic fitting (Figure 1C). However, using a steady-state binding affinity analysis, we were able to obtain $K_d = 9.1$ mM (Figure 1D). Using the same analysis for TEP yields $K_d = 0.31$ μM (Figure 1B), which is quite consistent with the result from kinetic fitting. These results were also in agreement with previous observations that show the binding affinity for TEP is 10,000 times higher than it is for CFF.[40]

LiGaMD Captured Repetitive Theophylline Binding and Dissociation in RNA Aptamer.

LiGaMD equilibration simulations of TEP captured the complete dissociation of the bound ligand into the bulk solvent. This was followed by rebinding to the RNA target site within 63 ns. We observed that C27 in RNA showed two distinct conformations during ligand binding. These conformations alternated between one where C27 was solvent exposed, called the "Out" state, and the other where it was buried inside the ligand pocket, called the "In" state.

Following the LiGaMD equilibration, six independent 2500 ns production simulations ("Sim1″-"Sim6") (Table S1) were further performed on the RNA-TEP system with randomized initial atomic velocities. Three complete cycles of TEP unbinding and rebinding in three of the six LiGaMD simulations were observed ("Sim1″, "Sim2", and "Sim3" in Table S1). Once boosted out of the pocket, the ligand-bound back into the RNA target site with a minimum RMSD of as low as ~2.4 Å relative to the native NMR structure (PDB ID: 1O15) (Figure 2A).[27] During "Sim1″, TEP bound to the target site during ~1500–1525 ns and then dissociated into a bulk solvent (Figure 2B). Two other rebinding and dissociation events were observed at ~2200 and ~2400 ns, respectively. During "Sim2″, TEP bound to the RNA in ~200–500 ns (Figure 2C) and dissociated quickly, followed by another rapid rebinding event during ~500–1100 ns. The ligand again approached the RNA pocket at ~1800 ns with fast dissociation. In "Sim3″, TEP was observed to leave the pocket and reassociated with the binding site at ~550, ~1480, and ~1700 ns (Figure 2D). The other simulations in the set for TEP binding to the RNA aptamer were able to capture the ligand binding events clearly, such as rebinding in "Sim4" at ~500 ns, "Sim5" for ~1000–1029 ns, and "Sim6" at ~500 ns and again at ~2000 ns (Figure S3D–F). However, these simulations were not able to capture three complete cycles of dissociation and rebinding of the ligand to the pocket. In order to further analyze the TEP and estimate the ligand kinetics, the first three simulations were used. These simulations provided reasonable statistics for the analysis of ligand binding and dissociation.

Next, we explored the correlation between ligand binding and conformational changes in the RNA. RMSD of the ligand relative to the NMR structure and solvent accessible surface area (SASA) of C27 in the aptamer were used as reaction coordinates to calculate a 2D free energy profile (Figure 3). Six low-energy conformations were identified in the 2D free energy profile of the TEP-RNA system (Figure 3A). Namely, the "Bound/Out", "Bound/In", "Intermediate/Out", "Intermediate/In", "Unbound/Out", and "Unbound/In" states. When the C27 was solvent exposed, TEP diffused into the binding site, forming a "Bound/Out" complex with (TEP RMSD, C27 SASA) centered at (~2 Å, ~252 Å^2) (Figure 3B). C27 could also be packed behind TEP in the RNA pocket in the "Bound/In" state with the (TEP RMSD, C27 SASA) centered at (~3 Å, 120 Å^2) (Figure 3C). In the "Intermediate/ Out" state the ligand hovered near the RNA surface, making transient local contacts at an average RMSD of ~25 Å, while the C27 remained extended and sustained a high SASA of ~240 Å^2 (Figure 3D). In the "Intermediate/In" state, a low energy well was identified corresponding to ~20 Å TEP RMSD and ~135 Å^2 SASA of C27 (Figure 3E). When C27 was solvent exposed with no ligand bound to the aptamer, a low-energy "Unbound/Out" state was identified with the (TEP RMSD, C27 SASA) centered at (~50 Å, ~250 Å^2) (Figure 3F). In the "Unbound/In" state, the C27 was found to be fully buried inward with ~130 Å^2

SASA, and TEP was pushed out into the bulk solvent with ~50 Å RMSD (Figure 3G). As the RNA restructured itself to make room for the incoming ligand, we observed that flipping C27 from inside to outside the binding pocket played a key role in the recognition of TEP.

The Mg^{2+} ions stabilize the S-turn fold between nucleotides C22 and G26, resulting in a narrow fold with a 5 Å distance between G25-C22 (Figure S7A). This conformation allows C27 to flip outside and U24 to interact with A28 via π–π stacking, stabilizing the RNA pocket. In the ligand-unbound state, Mg^{2+} shifts toward the back of the fold, increasing the G25-C22 distance to up to ~12 Å and weakening the stacking interaction between U24 and A28 (Figure S7B), causing the RNA pocket to become larger. Mg^{2+} ions facilitate important RNA conformational changes.

LiGaMD Captured Binding and Dissociation of Caffeine to RNA.

In comparison, we performed longer LiGaMD simulations (i.e., 5000 ns) in order to capture both binding and unbinding of caffeine to the RNA.[28] The simulations were initiated similarly with CFF in the TEP binding pocket. The 63 ns equilibration simulation captured the dissociation and rebinding of CFF to the aptamer. CFF diffused into the TEP binding pocket during the ~5000 ns LiGaMD simulations with a binding conformation similar to TEP in the NMR structure (Figure 4A). In "Sim1″", CFF bound to the pocket only once at ~1000 ns (Figure 4B), stayed inside the pocket briefly for ~6.85 ns, and thereafter dissociated quickly. "Sim2" also captured CFF rebinding at ~2500 ns and subsequent dissociation at ~2542 ns (Figure 4C). In "Sim3″", we were able to capture 2 complete rebinding and 2 dissociation events of CFF at ~400 and ~800 ns (Figure 4D). During rebinding events, the CFF RMSD relative to the native bound pose reached a minimum of ~2.8 Å.

To further probe conformational changes in the CFF-RNA system, CFF RMSD and SASA of C27 were used to calculate 2D PMF (Figure 5A). Five low-energy states were identified, including the "Bound/Out", "Unbound/Out", "I1/In", "I2/In", and an "Unbound/In", for which the (CFF RMSD, C27 SASA) centered at (~2 Å, ~242 $Å^2$) (Figure 5B), (~50 Å, ~230 $Å^2$) (Figure 5C), (~22 Å, ~130 $Å^2$) (Figure 5D), (~10 Å, 140 $Å^2$) (Figure 5E), and (~50 Å, ~120 $Å^2$) (Figure 5F), respectively. A "Bound/Out" state was identified when CFF bound to the target site with high SASA of C27 in the range of ~230–260 $Å^2$ (Figure 5B). A low energy state referred to as "Unbound/Out," sampled CFF far away from RNA in the bulk solvent with RMSD centered at ~50 Å, and C27 SASA sampled at ~230 $Å^2$ (Figure 5C). In the "I1/In" intermediate state the CFF was observed to be located at RMSD ~22 Å near the stem of the aptamer, with a low C27 SASA sampled at ~130 $Å^2$ (Figure 5D). We recorded a second intermediate "I2/In" (Figure 5E), where CFF occupied a site ~10 Å away from the target TEP binding pocket nestled in a groove made by C9, A10, G11, C21, G25, and C20 nucleotides via π-stacking interactions in the loop region. In the "Unbound/In" state centered with CFF RMSD at ~50 Å, and C27 SASA ~120 $Å^2$ (Figure 5F), the TEP binding pocket was observed to be tightly packed, preventing CFF from making its way to lodge inside. The two unbound states differed from each other in their sampled C27 conformations.

Ligand Binding Kinetic Rates and Free Energies Calculated from LiGaMD Were Consistent with Experimental Data.

LiGaMD simulations captured repetitive binding and dissociation of TEP to the RNA aptamer, which allowed us to calculate the ligand binding kinetic rate constants. The time periods for the ligand found in the bound state (τ_B) and unbound (τ_U) states during the simulation replicates were recorded for each simulation event (Tables S2 and S4). Without reweighting, the binding rate constant ($k_{on}*$) of TEP to the RNA aptamer was directly calculated from LiGaMD trajectories as $(2.10 \pm 0.23) \times 10^{10}$ M^{-1} s^{-1}, whereas the dissociation rate constant ($k_{off}*$) was calculated as $(1.94 \pm 0.13) \times 10^6$ s^{-1}.

Next, we reweighted the LiGaMD simulations to calculate acceleration factors of TEP binding and dissociation processes (Table S5) and recovered the original kinetic rate constants using Kramers' rate theory. The dissociation free energy barrier (ΔF_{off}) increased to 5.65 ± 0.34 kcal/mol in the reweighted PMF profiles from 3.61 ± 0.26 kcal/mol in the modified PMF profiles (Table S5). The free energy barrier for ligand binding (ΔF_{on}) changed to 1.23 ± 0.65 kcal/mol in the reweighted profiles from 0.96 ± 0.53 kcal/mol in the modified PMF profile (Table S5). Curvatures of the reweighed (w) and modified ($w*$, no reweighting) free energy profiles were calculated near the ligand Bound ("B") and Unbound ("U") low-energy wells and the energy barrier ("Br"), along with their ratio of apparent diffusion coefficients calculated from LiGaMD simulations with reweighting (D) and without reweighting (modified, $D*$) (Table S5). The reweighted k_{on} was calculated to be $(2.72 \pm 0.23) \times 10^5$ M^{-1} s^{-1}, being consistent with the corresponding experimental value of $(1.92 \pm 0.37) \times 10^5$ M^{-1} s^{-1}. The reweighted k_{off} was calculated to be $(2.34 \pm 0.07) \times 10^{-3}$ s^{-1} compared to our experimental value of (0.081 ± 0.45) s^{-1}. From our reweighted kinetic parameters, we can also calculate a binding affinity ($K_d = k_{off}/k_{on} = 0.009$ μM) as well as the ligand binding free energy ($\Delta G = -RT \ln(k_{off}/k_{on}) = -6.35 \pm 1.38$ kcal/mol). Both of which are quite comparable with their corresponding experimental determined values ($K_d = 0.42$ μM and $\Delta G = -7.83$ kcal/mol) for (Table 1).

Multiple Ligand Binding and Dissociation Pathways Were Identified from LiGaMD Simulations.

We dwelled deeper into our LiGaMD simulation trajectories to understand the different pathways associated with ligand binding and dissociation of both TEP and CFF. TEP adopted three unique pathways of spontaneous binding to the RNA target site (Figure 6A–C). During all of our simulations, we observed that C27 could sample frequently between the "Out" and "In" conformations. In "Sim3″", when the aptamer was unbound with its C27 base fully extended toward the solvent (SASA ~ 250 Å^2), TEP diffused from outside into the pocket through a pathway via π–π stacking interaction with the G26 nucleotide in the aptamer's loop region (Figure 6A and Movie S1). As TEP approached the binding site, it oriented its five-membered ring toward the pocket such that the two-methyl group faced outside. This ensured that the N7 and N9 atoms of TEP were available to make interactions with C22 in the same plane (Figures 6A and 3C). In another binding pathway observed in Sim2 ("BP2") (Figure 6B), TEP formed π-stacking interaction with the extended C27 in the "Out" conformation, as shown in the ligand trace (Figure 6B and Movie S2). In the

third pathway ("BP3"), as TEP passed through the stem of the aptamer, it formed stacking interactions with nucleotides A5 and C30 (Figure 6C and Movie S3). A different nucleotide A28, located below the plane of C27, displayed a novel base-flipping mechanism as it guided the TEP to the target site through the gap between A28 and C27. In our LiGaMD simulations, the BP1 pathway had a dominant occurrence over BP2, followed by BP3, which was the least sampled by TEP.

The LiGaMD simulation trajectories were further used to study key insights into the dissociation of TEP from the aptamer. Following ligand binding, TEP started to rotate and change its orientation, causing the five-membered ring to reorient itself so that N7 and N9 faced the solvent. The ligand moved out of the binding pocket via stacking interaction with G26 (Figure 6D and Movie S4). When the ligand changed its orientation, G29 formed a non-native base pair with U6 and a hydrogen bond with the N3 atom of TEP. This caused the A5-G29 wobble base pair to undergo fluctuations, leading to the site becoming incompatible with accommodating the bound TEP. The movement of G29, accompanied by the inward movement of C27, pushed the TEP out of the binding site. This observation was also consistent with previous work in the literature, which describes the flipping mechanism of the C27 nucleotide in unbound aptamers.[12]

In our LiGaMD simulations of CFF, we observed a sequence of conformational transitions as the CFF associated with the aptamer. The extra methyl group on N7 made it difficult for CFF to bind to the RNA pocket. We observed only 1–2 binding and dissociation events in 4 out of the 8 3500–5000 ns LiGaMD simulations (Tables S3 and S4). CFF typically formed stacking interaction with the solvent-exposed C27 and diffused inward through a gap between G26 and A28 into the RNA target site (Figure 7A and Movie S5). CFF bound to the target site in an orientation similar to TEP, with its 5-member ring facing toward the pocket. This pathway was observed to be the BP2 taken up by TEP. For dissociation of CFF, we observed that the ligand made its way out into the solvent through the gap between G26 and A28 of the RNA, through which it first made its way inside (Figure 7B and Movie S6). The methyl group in the same plane of the ligand may not cause steric repulsion of the π–π stacking interaction between the planes.

We analyze the low-energy states of the ligand binding and dissociation pathways for TEP and CFF to gain a more intuitive understanding of the prerequisites for ligand binding (Figure S8). In the case of TEP binding (Figure S8A), when C27 is oriented outward in the solvent, we observed the Unbound/Out conformation. As the simulation progresses, TEP attempts to reach the binding site, leading to a decrease in the distance between RNA and TEP before binding occurs, sampling the Intermediate/Out state. Subsequently, the ligand enters the binding pocket through one of the three binding pathways, adopting the stable Bound/Out conformation; this ensures that the N7 and N9 atoms of TEP were available to make interactions with C22 in the same plane. This is followed by a conformational change in the TEP within the binding pocket as it rotates in the plane, ultimately leading to dissociation from the binding pocket. In the sampled Bound/In state, C27 transitions inward. This is followed by a conformational change in TEP within the binding pocket as it rotates in the plane, ultimately leading to dissociation from the binding pocket. In the sampled Bound/In state, C27 transitions inward. Upon binding, TEP undergoes rotation and

a change in orientation, causing the five-membered ring to face the solvent. The ligand that exists in the binding pocket through stacking interactions with G26 and G29 forms a non-native base pair with U6 while hydrogen bonding with TEP, rendering the site incompatible. On the other hand, the CFF binding pathway to RNA involves the initial Unbound/Out conformation of RNA, which allows CFF to diffuse toward the binding site via binding pathway 2 (BP2) to Bound/Out conformation. As C27 shifts inward represented by the SASA values in the range of ~120–150 $Å^2$, rendering the pocket unsuitable for accommodating the ligand, The Bound/Out state transitions to I1/In and eventually to an Unbound/In state as CFF dissociates. Additionally, we observe the binding of CFF to a different site approximately ~10 Å away from the binding pocket. With C27 positioned inward, the I2/IN state follows the dissociation pathway through the I1/In state, ultimately reaching the Unbound/In state (Figure S8B).

In our simulations, CFF spent the majority of its time prolonging stacking interactions with C27 outside of the pocket. This could be due to steric repulsions generated by the additional methyl group found on N7 of CFF, forcing it to adhere to the conserved site less frequently. When CFF acquired this pocket, it oriented itself so that the methyl group faced outward. This orientation allowed for the O2 atom of CFF to form an intermolecular hydrogen bond with the amino group on C22. TEP formed a hydrogen bond between N7 and the O4 atom of RNA U23. The nucleotides C8, A7, and G26 form a stable base triple to accommodate the ligand in the pocket (Figure S9B). Unlike TEP binding, the presence of CFF in the binding pocket disrupted the A7-C8-G26 base triple (Figure S9B). The disruption of this important base triple also aligned our simulation findings with literature explanations explaining the weaker interactions of CFF with the RNA aptamer.

CONCLUSIONS

To the best of our knowledge, we have successfully simulated repetitive ligand binding and dissociation to RNA using a novel LiGaMD-enhanced sampling method. MD is an effective method for simulating biomolecular structural dynamics; hundreds of microseconds of cMD simulations can be performed daily on an Anton supercomputer and GPUs. These simulations have been used to study biomolecular binding mechanisms despite the difficulty in capturing both ligand dissociation and binding processes. To simulate these processes and forecast binding kinetic rates, improved sampling techniques have been developed, including the Weighted Ensemble, mile-stoning method, GaMD, Metadynamics, Markov State Modeling, Random Acceleration MD, and scaled MD.[42,43] In LiGaMD simulations, the accelerations of ligand kinetic rates were precisely estimated using Kramers' rate theory. The simulations have allowed us to map the rugged energy landscape of RNA, investigate the RNA-methylxanthine interaction in detail, and evaluate the binding thermodynamics and kinetics of TEP.

Even with multiple copies of the ligand in the systems, no specific repulsive potential is applied between the ligand molecules in the LiGaMD simulations. As we examined the time courses of ligand RMSD in Figures 2 and 4, there appeared to be ligand aggregation for only brief periods of time (e.g., Lig2 and Lig3 of TEP in Figure 2B–D, and Lig2 and Lig3 of CFF in Figure 4B–D). However, the ligand molecules were still able to dissociate from each

other and bind to the RNA target site (Figures 2B–D and 4B–D). Therefore, the brief ligand aggregation did not significantly affect binding of the ligand to target RNA in the LiGaMD simulations.

The LiGaMD simulations captured significant conformational changes in the RNA during ligand binding. The structural alterations included the flipping of a C27 nucleotide, a crucial component of the RNA. Notably, binding of both TEP and CFF to these "Out" and "In" conformations sampled by C27 was investigated (Figures 3 and 5). The high flexibility of the RNA allowed for conformational rearrangement before complex formation (1). TEP, when bound to the aptamer, formed hydrogen bond interactions with nearby nucleotides, such as C22 in the binding pocket. During ligand binding, we also observed various conformations of the RNA secondary structure, with respect to the movement of the unpaired C27 nucleotide. The high affinity of TEP for the aptamer made it possible to capture repetitive binding and unbinding events LiGaMD simulations (Figures 2 and S3). We were able to quantify the association and dissociation kinetic rates of TEP. The LiGaMD simulation findings were consistent with our experimental data. From the kinetic data, the K_d value of RNA-aptamer complex was calculated as 0.42 μM. The binding free energy of TEP calculated from LiGaMD simulation was −6.35 ± 1.38 kcal/mol, which was consistent with the SPR experimental value of −7.83 kcal/mol calculated using K_d of 0.42 μM. The simulation predicted k_{on} (2.72 ± 0.23) × 10^5 M^{-1} s^{-1} and k_{off} (2.34 ± 0.07) × 10^{-3} s^{-1} were comparable to the experimental values (1.92 ± 0.37) × 10^5 M^{-1} s^{-1}, (0.081 ± 0.45) s^{-1}, respectively. These simulations allowed for precise estimates of the kinetic rate constants and the ligand binding free energy.

Three pathways were identified for TEP binding to the RNA aptamer. This was accompanied by a base flip of nucleotide A28 and outward movement of nucleotide C27 (BP3). Addition of a methyl group on the N7 atom in CFF caused steric repulsions. CFF adopted one of the TEP binding pathways (BP2). Because of the additional methyl group on the N7 atom of CFF, the dissociation of the two ligands also showed variations. LiGaMD simulations of CFF binding and dissociation have not yet converged. The calculated free energy profiles were not accurate for a direct comparison. More sufficient sampling would be needed to obtain converged simulation results and precisely compute the ligand binding free energies and kinetic rates of CFF. This can be potentially achieved through additional and longer simulations.

In summary, microsecond-time scale LiGaMD simulations have successfully captured repetitive TEP dissociation and binding to the RNA aptamer, allowing for accurate characterization of both binding free energy and kinetic rates of TEP that are consistent with the experimental data. LiGaMD simulations suggest a conformational selection mechanism of TEP binding to the RNA aptamer. Such a finding indicates that compound docking using a structural ensemble of representative RNA conformations ("ensemble docking")[44] would be necessary for structure-based drug design of flexible RNA. By analyzing each ligand binding pathway in detail, we have also determined that nucleotides C22, C27, G26, and A28 are the most critical for the recognition, binding, and dissociation of the ligands. Terminal base pairs in RNA duplex structures are highly dynamic,[45] and the observed differences in these regions of the two representative structures likely reflect those

dynamics. Furthermore, many RNA structural motifs are quite modular and fold over a wide range of structural contexts. This is particularly true for the theophylline aptamer.[46,47] It is important to note that RNA has unique chemical and structural properties compared with proteins. Because of its charge and high flexibility, RNA requires careful consideration when choosing the appropriate force field. Given the wide range of conformations that RNA molecules can adopt, it is possible that a force field meant for proteins is not accurate for RNA. Mg^{2+} ions stabilize the S-turn fold between nucleotides C22 and G26, allowing for a narrow fold and C27 flipping outside. As Mg^{2+} shifts toward the back of the fold, the G25-C22 distance increases and the stacking interaction weakens, causing the RNA pocket to become larger and C27 buried inside. These observations are particularly valuable for demonstrating RNA-small molecule interactions at an atomistic level. Therefore, our complementary simulations and experiments have provided important insights into the mechanisms of RNA-ligand interactions, facilitating rational drug design targeting the highly flexible RNA.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

Data Availability Statement

Molecular dynamics trajectories generated in this study are not publicly available as the data is over 100 GB in size. The data used to generate trajectories are placed in the publicly available repository Zenodo (https://zenodo.org/record/8330867). Outputs are available upon request and require several business days to share. All software packages used in this report are available in public repositories, including AMBER20 (http://ambermd.org/), CPPTRAJ (https://amber-md.github.io/cpptraj/CPPTRAJ.xhtml), and PyReweighting (https://github.com/MiaoLab20/PyReweighting).

REFERENCES

(1). Russell R; Zhuang X; Babcock HP; Millett IS; Doniach S; Chu S; Herschlag D Exploring the folding landscape of a structured RNA. Proc. Natl. Acad. Sci. U. S. A.' 2002, 99 (1), 155–160. [PubMed: 11756689]

(2). Menichelli E; Lam BJ; Wang Y; Wang VS; Shaffer J; Tjhung KF; Bursulaya B; Nguyen TN; Vo T; Alper PB; McAllister CS; Jones DH; Spraggon G; Michellys PY; Joslin J; Joyce GF; Rogers J Discovery of small molecules that target a tertiary-structured RNA. Proc. Natl. Acad. Sci. U. S. A.' 2022, 119 (48), No. e2213117119. [PubMed: 36413497]

(3). Rizvi NF; Santa Maria JP Jr; Nahvi A; Klappenbach J; Klein DJ; Curran PJ; Richards MP; Chamberlin C; Saradjian P; Burchard J Targeting RNA with small molecules: identification of selective, RNA-binding small molecules occupying drug-like chemical space. SLAS Discov.' 2020, 25 (4), 384–396. [PubMed: 31701793]

(4). Karplus M; McCammon JA Molecular dynamics simulations of biomolecules. Nature structural biology 2002, 9 (9), 646–652. [PubMed: 12198485]

(5). Spiwok V; Sucur Z; Hosek P Enhanced sampling techniques in biomolecular simulations. Biotechnology advances 2015, 33 (6), 1130–1140. [PubMed: 25482668]

(6). Miao Y; Bhattarai A; Wang J Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. J. Chem. Theory Comput.' 2020, 16 (9), 5526–5547. [PubMed: 32692556]

(7). Wang F; Zuroske T; Watts JK RNA therapeutics on the rise. Nat. Rev. Drug Discov 2020, 19 (7), 441–442. [PubMed: 32341501]

(8). Zhu Y; Zhu L; Wang X; Jin H RNA-based therapeutics: An overview and prospectus. Cell Death Dis.' 2022, 13 (7), 644. [PubMed: 35871216]

(9). Warner KD; Hajdin CE; Weeks KM Principles for targeting RNA with drug-like small molecules. Nat. Rev. Drug Discovery 2018, 17 (8), 547–558. [PubMed: 29977051]

(10). Keefe AD; Pai S; Ellington A Aptamers as therapeutics. Nat. Rev. Drug Discovery 2010, 9 (7), 537–550. [PubMed: 20592747]

(11). Wrist A; Sun W; Summers RM The theophylline aptamer: 25 years as an important tool in cellular engineering research. ACS Synth. Biol.' 2020, 9 (4), 682–697. [PubMed: 32142605]

(12). Warfield BM; Anderson PC Molecular simulations and Markov state modeling reveal the structural diversity and dynamics of a theophylline-binding RNA aptamer in its unbound state. PLoS One 2017, 12 (4), No. e0176229. [PubMed: 28437473]

(13). Zimmermann GR; Jenison RD; Wick CL; Simorre J-P; Pardi A Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. Nature structural biology 1997, 4 (8), 644–649. [PubMed: 9253414]

(14). Wang J; Arantes PR; Bhattarai A; Hsu RV; Pawnikar S; Huang YM; Palermo G; Miao Y Gaussian accelerated molecular dynamics: Principles and applications. Wiley Interdiscip. Rev.: Comput. Mol. Sci.' 2021, 11 (5), No. e1521. [PubMed: 34899998]

(15). Wang Y; Parmar S; Schneekloth JS; Tiwary P Interrogating RNA–small molecule interactions with structure probing and artificial intelligence-augmented molecular simulations. ACS Central Science 2022, 8 (6), 741–748. [PubMed: 35756372]

(16). Levintov L; Vashisth H Ligand recognition in viral RNA necessitates rare conformational transitions. J. Phys. Chem. Lett.' 2020, 11 (14), 5426–5432. [PubMed: 32551654]

(17). Tang Z; Akhter S; Ramprasad A; Wang X; Reibarkh M; Wang J; Aryal S; Thota SS; Zhao J; Douglas JT Recognition of single-stranded nucleic acids by small-molecule splicing modulators. Nucleic Acids Res.' 2021, 49 (14), 7870–7883. [PubMed: 34283224]

(18). (a)Vanommeslaeghe K; MacKerell AD Jr. CHARMM additive and polarizable force fields for biophysics and computer-aided drug design. Biochimica et biophysica acta 2015, 1850, 861. [PubMed: 25149274] (b)Duan Y; Wu C; Chowdhury S; Lee MC; Xiong GM; Zhang W; Yang R; Cieplak P; Luo R; Lee T; Caldwell J; Wang J; Kollman P A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. J. Comput. Chem.' 2003, 24 (16), 1999–2012. [PubMed: 14531054]

(19). (a)Dror RO; Pan AC; Arlow DH; Borhani DW; Maragakis P; Shan Y; Xu H; Shaw DE Pathway and mechanism of drug binding to G-protein-coupled receptors. Proc. Natl. Acad. Sci.' 2011, 108 (32), 13118–13123. [PubMed: 21778406] (b)Miao Y; Bhattarai A; Nguyen ATN; Christopoulos A; May LT Structural Basis for Binding of Allosteric Drug Leads in the Adenosine A1 Receptor. Sci. Rep.' 2018, 8 (1), 16836. [PubMed: 30442899]

(20). Wang J; Miao Y Ligand Gaussian accelerated molecular dynamics 2 (LiGaMD2): Improved calculations of ligand binding thermodynamics and kinetics with closed protein pocket. J. Chem. Theory Comput.' 2023, 19 (3), 733–745. [PubMed: 36706316]

(21). Buch I; Giorgino T; De Fabritiis G Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. Proc. Natl. Acad. Sci. U. S. A.' 2011, 108 (25), 10184–10189. [PubMed: 21646537]

(22). Ferruz N; De Fabritiis G Binding Kinetics in Drug Discovery. Mol. Inform 2016, 35 (6–7), 216–226. [PubMed: 27492236]

(23). Miao Y Acceleration of Biomolecular Kinetics in Gaussian Accelerated Molecular Dynamics. J. Chem. Phys.' 2018, 149 (7), No. 072308. [PubMed: 30134710]

(24). Doshi U; Hamelberg D Extracting Realistic Kinetics of Rare Activated Processes from Accelerated Molecular Dynamics Using Kramers' Theory. J. Chem. Theory Comput 2011, 7 (3), 575–581. [PubMed: 26596291]

(25). Frank AT; Andricioaei I Reaction Coordinate-Free Approach to Recovering Kinetics from Potential-Scaled Simulations: Application of Kramers' Rate Theory. J. Phys. Chem. B 2016, 120 (33), 8600–8605. [PubMed: 27220565]

(26). Hamelberg D; Shen T; McCammon JA Relating kinetic rates and local energetic roughness by accelerated molecular-dynamics simulations. J. Chem. Phys.' 2005, 122 (24), 241103. [PubMed: 16035738]

(27). Clore GM; Kuszewski J Improving the accuracy of NMR structures of RNA by means of conformational database potentials of mean force as assessed by complete dipolar coupling cross-validation. J. Am. Chem. Soc.' 2003, 125 (6), 1518–1525. [PubMed: 12568611]

(28). Zimmermann GR; Wick CL; Shields TP; Jenison RD; Pardi A Molecular interactions and metal binding in the theophylline-binding core of an RNA aptamer. Rna 2000, 6 (5), 659–667. [PubMed: 10836787]

(29). Fadrná E; Špa ková NA; Sarzynska J; Koca J; Orozco M; Cheatham TE III; Kulinski T; Sponer J Single stranded loops of quadruplex DNA as key benchmark for testing nucleic acids force fields. J. Chem. Theory Comput.' 2009, 5 (9), 2514–2530. [PubMed: 26616629]

(30). Zgarbová M; Otyepka M; Šponer JI; Mládek AT; Banáš P; Cheatham TE III; Jurecka P; et al. Refinement of the Cornell nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. J. Chem. Theory Comput.' 2011, 7 (9), 2886–2902. [PubMed: 21921995]

(31). He X; Man VH; Yang W; Lee TS; Wang J A fast and high-quality charge model for the next generation general AMBER force field. J. Chem. Phys.' 2020, 153 (11), 114502. [PubMed: 32962378]

(32). Salomon-Ferrer R; Case DA; Walker RC An overview of the Amber biomolecular simulation package. Wiley Interdisciplinary Reviews: Computational Molecular Science 2013, 3 (2), 198–210.

(33). Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML Comparison of simple potential functions for simulating liquid water. J. Chem. Phys.' 1983, 79 (2), 926–935.

(34). Li Z; Song LF; Li P; Merz KM Jr Systematic parametrization of divalent metal ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. J. Chem. Theory Comput.' 2020, 16 (7), 4429–4442. [PubMed: 32510956]

(35). (a)Joung IS; Cheatham TE III Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. J. Phys. Chem. B 2008, 112 (30), 9020–9041. [PubMed: 18593145] (b)Li P; Song LF; Merz KM Jr Systematic parameterization of monovalent ions employing the nonbonded model. J. Chem. Theory Comput.' 2015, 11 (4), 1645–1657. [PubMed: 26574374] (c)Sengupta A; Li Z; Song LF; Li P; Merz KM, Jr Parameterization of monovalent ions for the OPC3, OPC, TIP3P-FB, and TIP4P-FB water models. J. Chem. Inf. Model.' 2021, 61 (2), 869–880. [PubMed: 33538599]

(36). Humphrey W; Dalke A; Schulten K VMD: visual molecular dynamics. J. Mol. Graphics 1996, 14 (1), 33–38.

(37). Roe DR; Cheatham TE PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. J. Chem. Theory Comput 2013, 9 (7), 3084–3095. [PubMed: 26583988]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

(38). Xin Y; Doshi U; Hamelberg D Examining the limits of time reweighting and Kramers' rate theory to obtain correct kinetics from accelerated molecular dynamics. J. Chem. Phys.' 2010, 132 (22), 224101. [PubMed: 20550384]

(39). Ester M; Kriegel H-P; Sander J; Xu X A density-based algorithm for discovering clusters in large spatial databases with noise. Knowl. Discov. Data Min.' 1996, 96 (34), 226–231.

(40). Jenison RD; Gill SC; Pardi A; Polisky B High-resolution molecular discrimination by RNA. Science 1994, 263 (5152), 1425–1429. [PubMed: 7510417]

(41). Burley SK; Berman HM; Kleywegt GJ; Markley JL; Nakamura H; Velankar S Protein Data Bank (PDB): the single global macromolecular structure archive. Protein crystallography: methods and protocols 2017, 1607, 627–641.

(42). Wang J; Do HN; Koirala K; Miao Y Predicting biomolecular binding kinetics: A review. J. Chem. Theory Comput.' 2023, 19 (8), 2135–2148. [PubMed: 36989090]

(43). Conflitti P; Raniolo S; Limongelli V Perspectives on Ligand/Protein Binding Kinetics Simulations: Force Fields, Machine Learning, Sampling, and User-Friendliness. J. Chem. Theory Comput.' 2023, 19 (18), 6047–6061. [PubMed: 37656199]

(44). Amaro RE; Baudry J; Chodera J; Demir Ö; McCammon JA; Miao Y; Smith JC Ensemble Docking in Drug Discovery. Biophys. J.' 2018, 114 (10), 2271–2278. [PubMed: 29606412]

(45). Pinamonti G; Paul F; Noé F; Rodriguez A; Bussi G The mechanism of RNA base fraying: Molecular dynamics simulations analyzed with core-set Markov state models. J. Chem. Phys.' 2019, 150 (15), 154123. [PubMed: 31005065]

(46). Rai N; Ferreiro A; Neckelmann A; Soon A; Yao A; Siegel J; Facciotti MT; Tagkopoulos I RiboTALE: A modular, inducible system for accurate gene expression control. Sci. Rep.' 2015, 5 (1), 10658. [PubMed: 26023068]

(47). Roßmanith J; Narberhaus F Modular arrangement of regulatory RNA elements. RNA Biol.' 2017, 14 (3), 287–292, DOI: 10.1080/15476286.2016.1274853. [PubMed: 28010165]
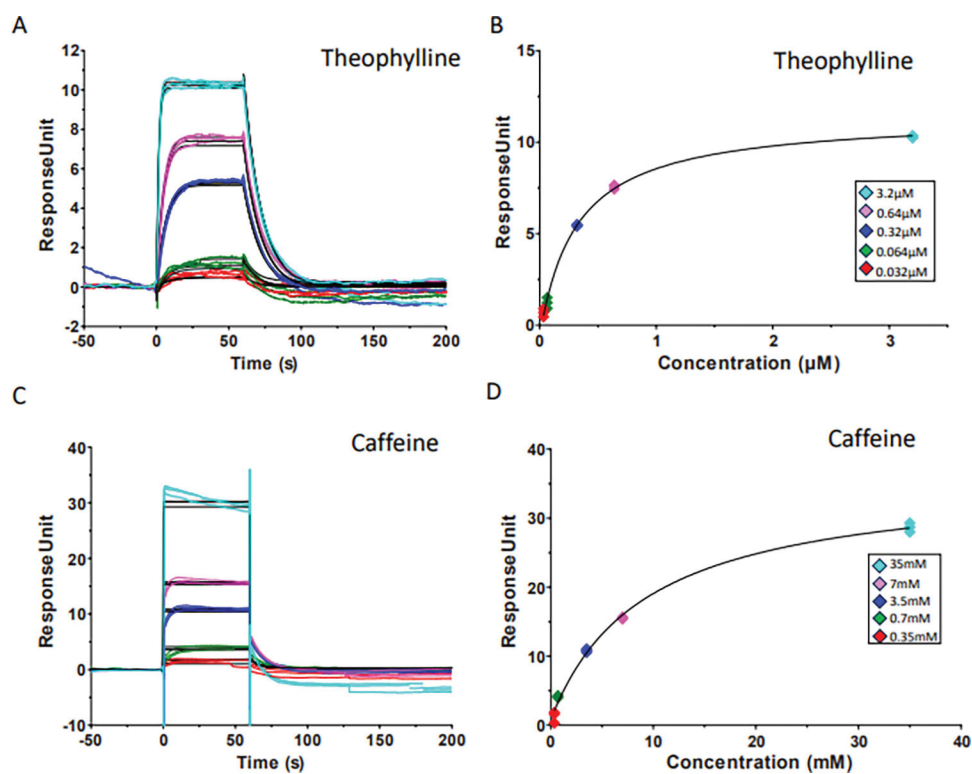
Figure 1.
Surface plasmon resonance (SPR) binding analysis of theophylline (A and B) and caffeine (C and D) with the RNA aptamer. Curve fitting was performed using Biocore T200 BioEvaluation software and shown as black line. Figure A and B were acquired from 1:1 kinetics fitting. Figure C and D were acquired from 1:1 steady state binding affinity fitting.
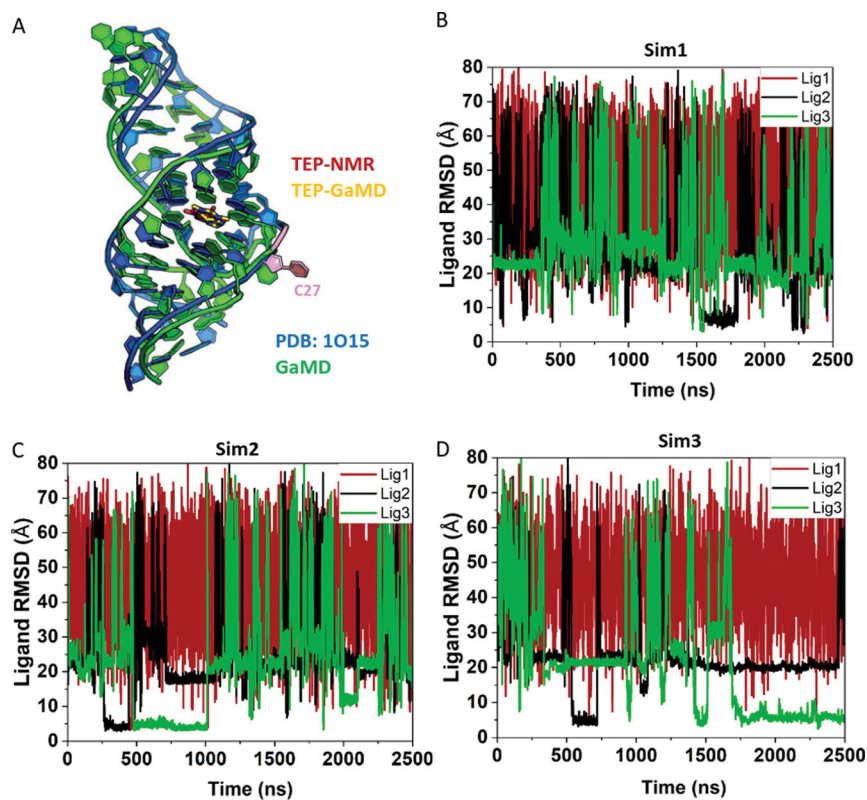
Figure 2.
LiGaMD simulations captured repetitive dissociation and binding of theophylline (TEP) to the RNA aptamer: (A) NMR structure (blue, PDB ID: 1O15) and GaMD predicted binding pose of TEP (sticks with C atoms colored in red and yellow, respectively) in the RNA aptamer. A key residue in the binding pocket, C27 is highlighted in light pink. (B–D) Time courses of the ligand heavy atom root-mean-square deviation (RMSD) relative to the NMR structure with the RNA aligned calculated from three independent 2.5 μs LiGaMD simulations of TEP binding to the RNA aptamer.
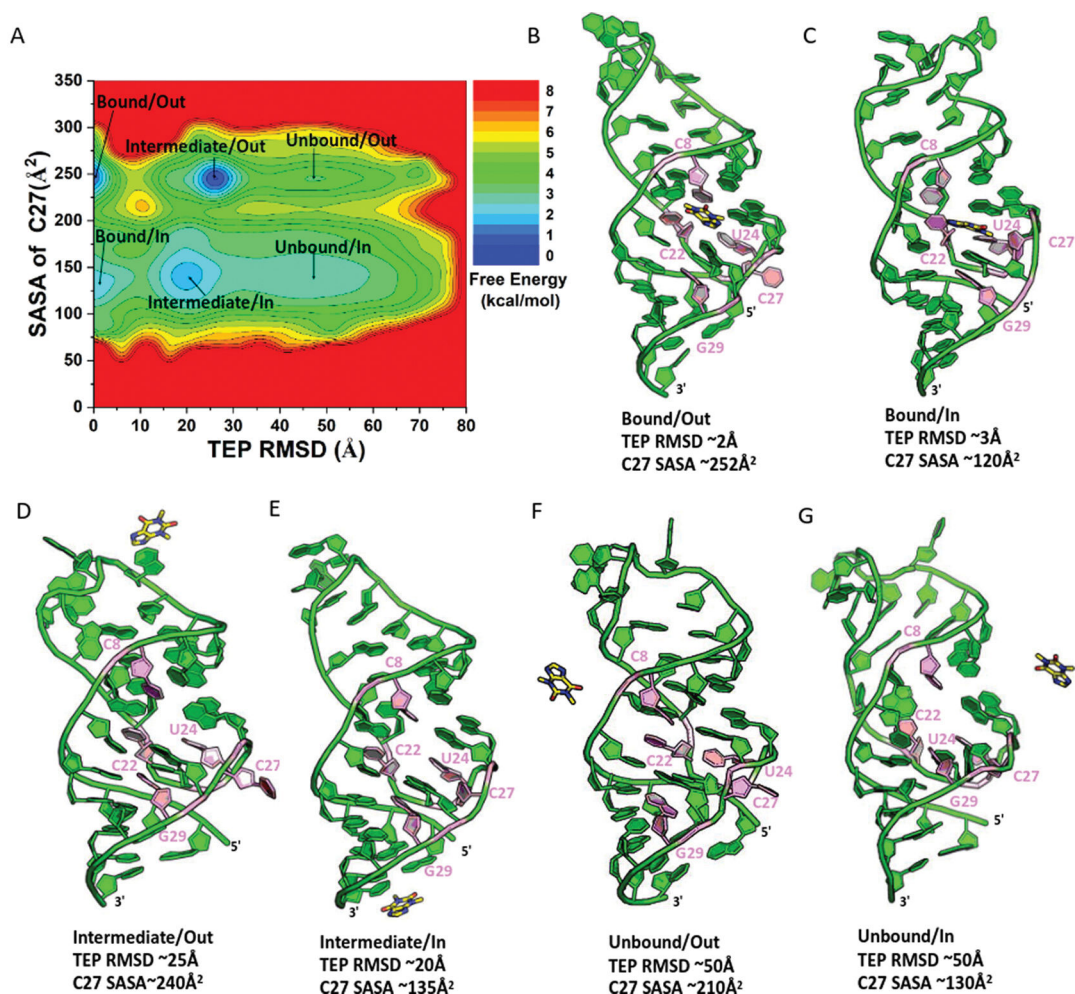
Figure 3.
Free energy profiles and low-energy conformational states of TEP binding to the RNA aptamer: (A) 2D potential of mean free energy (PMF) profile regarding the TEP RMSD (relative to its native bound pose in the NMR structure) and Solvent Accessible Surface Area (SASA) of nucleotide C27. (B–G) Low-energy states as identified from the 2D free energy profile from LiGaMD simulations, highlighted as (B) "Bound/Out", (C) "Bound/In", (D) "Intermediate/Out", (E) "Intermediate/In", (F) "Unbound/Out", and (G) "Unbound/In" states. The RNA is presented in cartoons in green, and the ligand in sticks with the C atoms colored in yellow. The C8, C22, U24, C27, and G29 nucleotides are highlighted in light pink. The 5′ and 3′ termini are indicated for the RNA aptamer.
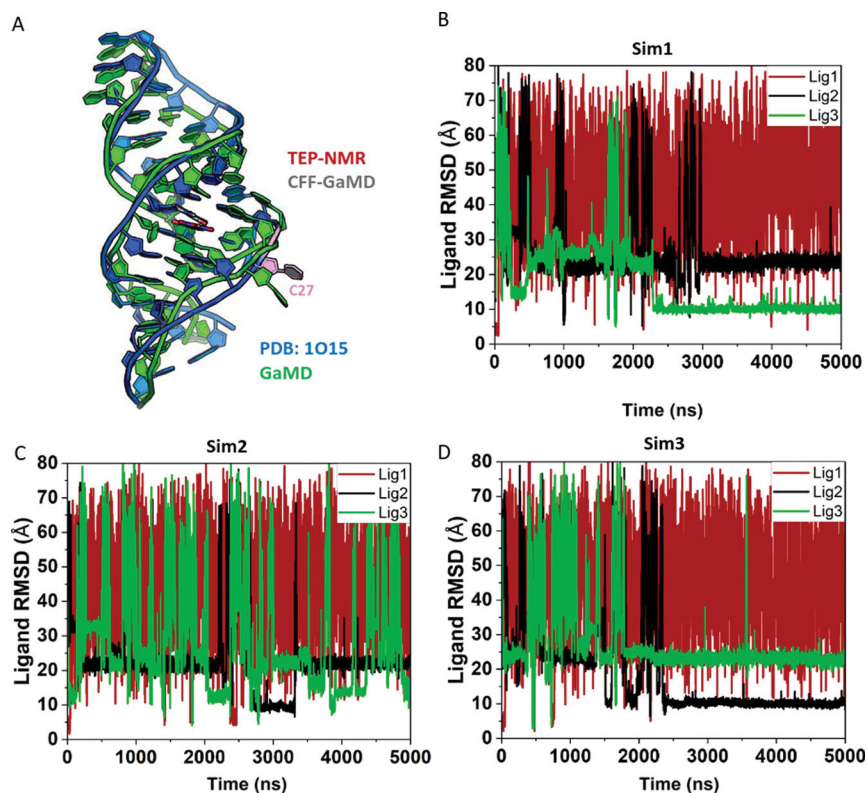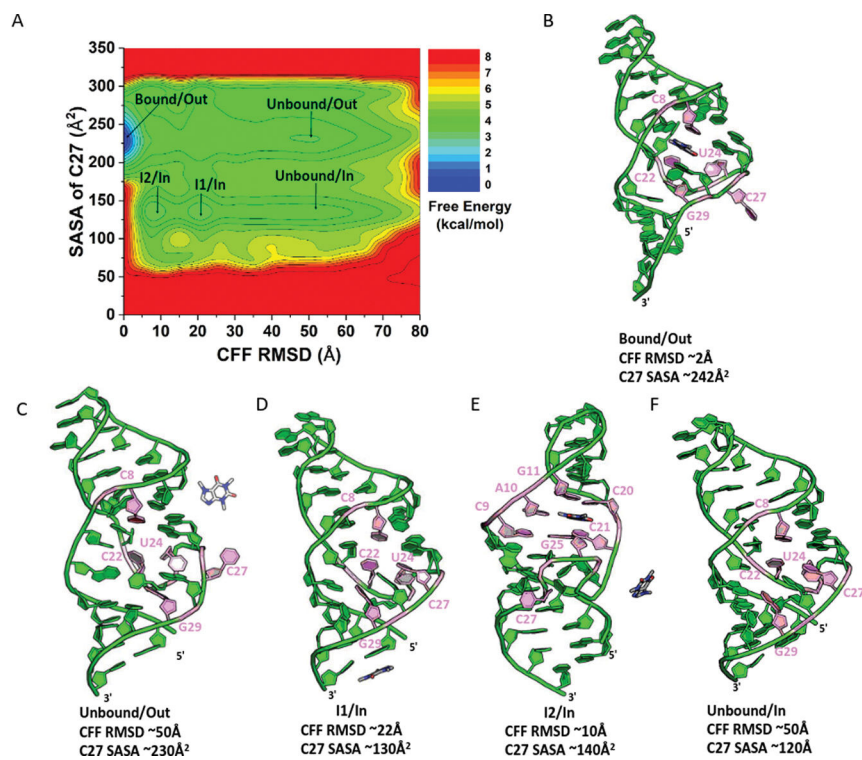
Figure 4.
LiGaMD simulations captured dissociation and binding of caffeine (CFF) to the TEP RNA: (A) the TEP-bound NMR structure (blue, PDB ID: 1O15) and GaMD predicted caffeine-bound pose (sticks with C atoms colored in red and gray, respectively) in the RNA aptamer. A key residue in the binding pocket, C27 is highlighted in light pink. (B–D) Time courses of the ligand heavy atom RMSD calculated from three independent 5 μs LiGaMD simulations of CFF binding to this site in the RNA.

Figure 5.
Free energy profiles and low-energy conformational states of CFF binding to RNA aptamer:
(A) 2D PMF energy profile regarding the CFF RMSD (relative to its bound pose in
the starting structure) and Solvent Accessible Surface Area (SASA) of nucleotide C27
(B–F) Low-energy states as identified from the 2D free energy profile (B) "Bound/Out",
(C) "Unbound/Out", (D) "I1/In", (E) "I2/In", and (F) "Unbound/In" states. The RNA is
presented in cartoons in green and the ligand in sticks with the C atoms colored in gray. The
C8, C22, U24, C27, and G29 nucleotides are highlighted in light pink. The 5′ and 3′ termini
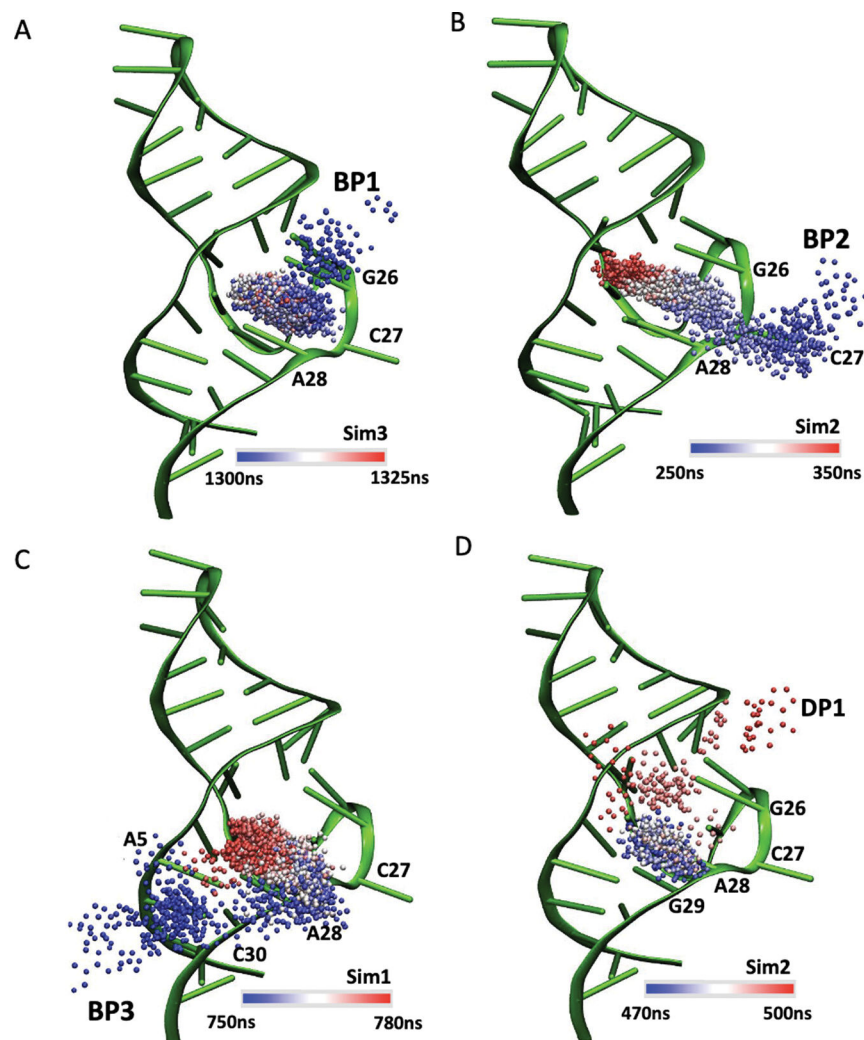are indicated for the RNA aptamer.

Figure 6.
Representative binding and dissociation pathways of TEP to the RNA aptamer revealed from LiGaMD: (A–C) Starting from outside in the solvent, TEP bound to the RNA target site through three distinct pathways to reach the binding pocket as shown with a bead trace in colored simulation time in a blue-white-red (BWR) color scale. (A) During "Sim3″, TEP approaches the aptamer through the loop region of the RNA, labeled as binding pathway 1 or "BP1". (B) In another pathway observed during a binding event in "Sim2″, TEP makes stacking interactions with the out C27 side chain before finally occupying the pocket, labeled as binding pathway 2 or "BP2". (C) A third pathway observed during a binding event in "Sim1″, captures TEP interacting with the RNA located at the stem of the RNA and before finally occupying the pocket. (D) "Sim2" captures the dissociation event labeled as dissociation pathway 1 or "DP1″ as TEP escapes from the RNA pocket captured during LiGaMD simulations shown in a blue-white-red (BWR) color scale.
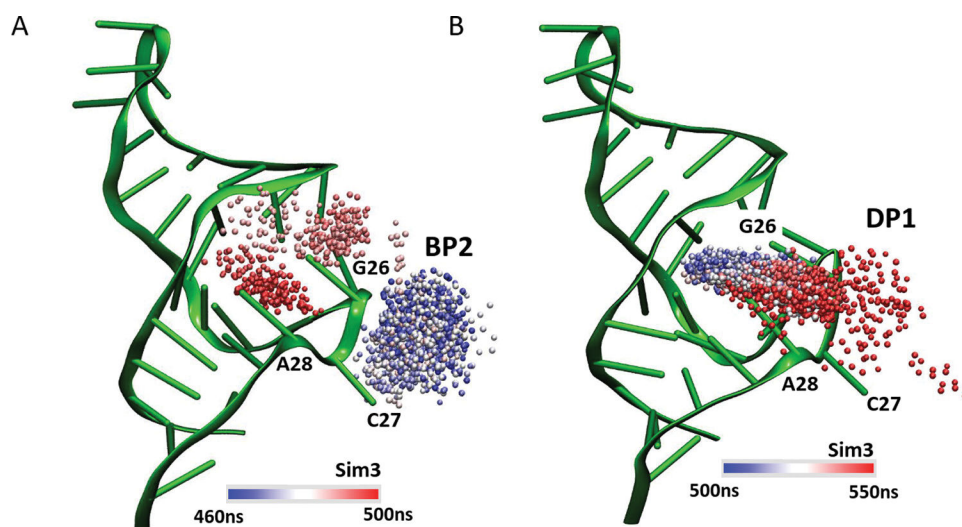
Figure 7.
Representative binding and dissociation pathway of caffeine to the RNA aptamer revealed from LiGaMD: (A) "Sim3" captured CFF binding in the pocket, starting from outside in the solvent as it approaches inward, shown with bead trace in colored simulation time in a blue-white-red (BWR) color scale, labeled as binding pathway 2, "BP2". (B) "Sim3" captures caffeine dissociation from the binding pocket, trace represented in a blue-white-red (BWR) color scale, labeled as Dissociation Pathway 2 or "DP2".

Table 1.

Summary of LiGaMD Simulations Performed on Theophylline (TEP) Binding to the RNA Aptamer[a]

ligand	ID	length (ns)	N_B	N_D	ΔV (kcal/mol)	G_{sim}(kcal/mol)	G_{exp}(kcal/mol)
TEP	Sim1	2500	3	3	23.11 ± 4.32	-6.35 ± 1.38	-7.83
	Sim2	2500	3	3	23.18 ± 4.26		
	Sim3	2500	3	3	23.18 ± 4.23		

[a]N_B and N_D are the numbers of ligand binding and dissociation events, respectively. ΔV is the total boost potential. G_{sim} and G_{exp} are the ligand binding free energies obtained from the LiGaMD simulations and experiments, respectively.

Table 2.

Summary of LiGaMD Simulations Performed on the Binding of Caffeine (CFF) to the RNA Aptamer[a]

ligand	ID	length (ns)	N_B	N_D	ΔV (kcal/mol)
CFF	Sim1	5000	1	1	21.11 ± 3.72
	Sim2	5000	1	1	21.34 ± 3.76
	Sim3	5000	2	2	21.45 ± 3.68
	Sim4	5000	1	1	20.56 ± 3.23
	Sim5	3500			21.10 ± 3.65
	Sim6	4000			21.07 ± 3.74
	Sim7	5000			21.15 ± 3.76
	Sim8	4000			21.15 ± 3.76

[a]N_B and N_D are the number of observed ligand binding and dissociation events, respectively. ΔV is the total boost potential.