



HAL
open science

Targeting RNA with Small Molecules using State-of-the-Art Methods Provides Highly Predictive Affinities of Riboswitch Inhibitors

Narjes Ansari, Chengwen Liu, Florent Hedin, Jérôme Hénin, Jay Ponder, Pengyu Ren, Jean-Philip Piquemal, Louis Lagardère, Krystel El Hage

► **To cite this version:**

Narjes Ansari, Chengwen Liu, Florent Hedin, Jérôme Hénin, Jay Ponder, et al.. Targeting RNA with Small Molecules using State-of-the-Art Methods Provides Highly Predictive Affinities of Riboswitch Inhibitors. 2024. hal-04742743v2

HAL Id: hal-04742743

<https://hal.science/hal-04742743v2>

Preprint submitted on 29 Oct 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Targeting RNA with Small Molecules using State-of-the-Art Methods Provides Highly Predictive Affinities of Riboswitch Inhibitors

Narjes Ansari,^{*,†} Chengwen Liu,^{†,‡} Florent Hédin,[†] Jérôme Hénin,[¶] Jay W. Ponder,^{†,§} Pengyu Ren,^{‡,†} Jean-Philip Piquemal,^{||,†} Louis Lagardère,^{*,||,†} and Krystal El Hage^{*,†}

[†]*Qubit Pharmaceuticals, 75014 Paris, France*

[‡]*The University of Texas at Austin, Department of Biomedical Engineering, TX, USA*

[¶]*Université Paris Cité, CNRS, Laboratoire de Biochimie Théorique, 75005 Paris, France*

[§]*Washington University in Saint Louis, Department of Chemistry, Saint Louis, MO, USA*

^{||}*Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005 Paris, France*

E-mail: narjesa@qubit-pharmaceuticals.com; louis.lagardere@sorbonne-universite.fr;
krystal.elhage@qubit-pharmaceuticals.com

Abstract

Targeting RNA with small molecules represents a promising yet relatively unexplored avenue for the design of new drugs. Nevertheless, challenges arise from the lack of computational models and techniques able to accurately model RNA systems, and predict their binding affinities to small molecules. Here, we tackle these difficulties by developing a tailored state-of-the-art approach for absolute binding free energy calculations of RNA-binding small molecules. To do so, we combine the advanced

AMOEBA polarizable force field, which accounts for both accurate multipolar electrostatics and many-body effects, to the newly developed Lambda-Adaptive Biasing Force (Lambda-ABF) scheme associated to refined restraints allowing for efficient sampling. Furthermore, to capture the free energy barrier associated to challenging RNA conformational changes, we apply machine learning to identify effective collective variables in order to use them into further enhanced sampling simulations based on an evolution of metadynamics. Applying this computational protocol to a complex Riboswitch-like RNA target, we demonstrate quantitative predictions. These results pave the way for the routine application of free-energy simulations in RNA-targeted drug discovery, thus providing a significant reduction in their failure rate.

Introduction

RNA-targeting therapeutics offer an attractive alternative to reach traditionally undruggable proteins and expand the druggable target space in the human genome. Indeed, RNA can adopt three-dimensional (3D) structures, conferring varied functional roles in human biology as well as disease-related dysfunction. They are unique molecules capable of interacting with all three major forms of biological macromolecules: proteins, DNAs, and RNAs. Due to their diverse mechanisms of action, including gene silencing, splice modulation, and protein interaction, RNA therapeutics offer a versatile platform for drug development.¹⁻⁴

Therapeutic strategies to target RNA include nucleotide-based agents such as (i) antisense oligonucleotides (ASOs), small interfering RNA (siRNA), and microRNAs (miRNAs) that can directly target messenger RNAs (mRNAs) and noncoding RNAs (ncRNAs) through Watson-Crick base pairing, (ii) clustered regularly inter-spaced short palindromic repeats (CRISPR) gene editing that can directly modify target RNA sequences to treat specific disorders, (iii) and small molecules that recognize RNA structures.² While ASOs and CRISPR editing have been invaluable to the field of chemical biology, their applications in therapeutics

lacks efficient delivery strategies and presents significant adverse reactions.^{5,6} These faced difficulties are due to their physico-chemical properties (size, charge,...), off-target effects, interactions with the immune system, rapid clearance and nuclease degradation. Alternatively, small molecules offer a promising alternative with potential for oral bio-availability and blood–brain barrier penetration. In particular, the extensive medicinal chemistry knowledge allows to systematically optimize pharmacokinetics and potency. Moreover, functional RNAs or RNA motifs are highly structured and form binding pockets or clefts that are accessible by small molecules.

Overall, targeting RNA with small molecules is emerging as a promising drug discovery approach. Such molecules interact with various target RNA substrates, including RNA structural motifs (such as hairpins, bulges, and internal loops), RNA enzymes (ribozymes), and specific RNA sequences.^{7,8} Several binders have been identified and exhibit various modes of action (MOAs), from simple binding to direct cleavage or recruitment of endogenous nucleases. These molecules modulate diverse biological processes, such as inhibiting bacterial and viral translation (e.g. ribocil and a riboswitch in *Escherichia coli*,⁹ 2-aminobenzimidazole derivatives and the hepatitis C internal ribosome entry site (IRES),¹⁰ respectively), and directing splicing by acting as molecular glues with cellular proteins (e.g., branaplam¹¹ and the FDA-approved risdiplam¹²).

Challenges arise when tackling these systems throughout the drug discovery process and more specifically at the early stages, where the computer-aided drug design (CADD) tools and simulation approaches are essential to characterize the target. Understanding its structural dynamics as well as its mechanism of action and interplay with other partners will guide the implementation of a pertinent inhibition or modulation mechanism. However, current methods face significant challenges in addressing the key obstacles associated with structure determination, accurate computational models, and reliable affinity prediction methods,

which are the key elements required to guide the initial molecular design and rational drug optimization phases.

RNA's chemical composition is distinct from proteins due to a highly electronegative surface potential associated to a limited buried surface area. It is highly charged, dynamic, and can adopt different conformations. In addition, divalent metal ions play an important role in its structural stability. Since RNA is also surrounded by ions and polarizable water molecules, it makes its targeting very challenging without having a reliable computational model that encompasses an accurate treatment of electrostatics.^{13,14} This can be addressed by using second-generation force fields that account for many-body polarization effects.^{15,16} Indeed, their inclusion in simulation play a key role in predicting biophysical properties governing the interactions between a target, its environment, and a potential drug candidate.¹⁷ This is the case of AMOEBA, an advanced multipolar polarizable force field that employs atomic induced dipoles to include polarization effects and atomic multipoles, up to quadrupoles, to represent the anisotropy of permanent electrostatics.^{18,19} These non-additive effects are essential for accurately ranking affinities of small molecules.^{13,20,21} Although polarizable force fields were traditionally considered as computationally slow, limiting their early adoption, recent advances in computing hardware, especially GPUs (Graphics Processing Unit) have greatly improved the situation.²² Further couplings with enhanced sampling techniques strongly increased the molecular dynamics (MD) timescales accessible to these simulations.^{14,23-28}

Computing free energies of binding *in silico* has become a routinely used tool in drug discovery projects.^{29,30} Among the existing techniques, alchemical ones, where the transformation of interest is made through a non-physical path, has proven to be both efficient and reliable, especially for relative free energies of binding (RBFE) between compounds.³¹ Computing Absolute Binding Free Energies (ABFE) is more challenging but also more directly useful for drug discovery as it does not rely on the knowledge of a reference compound and is

not limited to ligands sharing a similar binding mode. Computing binding affinities of small molecules bound to RNA or DNA with such methods remain a daunting challenge when compared to proteins.³²⁻³⁶ Given the particular nature of RNA, it is considerably more sensitive to the protocol adopted than in the case of small molecule-protein complexes, especially when the ligand is large, charged, with highly flexible, elongated structure. In addition, both the RNA target and the ligand remain exposed to the solvent and the counterions, in contrast to ligand-protein complexes in which the ligand is in a partly buried protein cavity and interacts with only a limited number of "discrete" water molecules without intrusions from a counterion, which is common in RNA systems.

Furthermore, the shape of the ligand and its binding mode, i.e. intercalation or groove binding, will impact the overall protocol used to get ABFEs, especially in terms of the choice and the number of restraints used to keep the ligand in place during simulations and facilitate convergence.^{37,38} Other aspects such as the discretization of the alchemical path (i.e. the so-called "lambda windows" used in standard methods) and the sampling time are critical factors that should be taken into account when developing a tailored protocol within a well-defined workflow for ABFE calculations for RNA (and DNA) small molecule binders.^{27,32,39} Moreover, the free energy barrier associated with the significant conformational change that may exist between the Apo and Holo structures of the target is notoriously challenging to estimate.⁴⁰

A recently developed technique, the Lambda-Adaptive Biasing Force (Lambda-ABF) approach, grounded on λ -dynamics combined with a newly developed multiple-walker adaptive biasing force bias, bypasses the discretization of the alchemical path and enables efficient sampling of orthogonal degrees of freedom through the variable nature of the alchemical variable. In combination with a recently introduced restraint scheme,³⁸ it has shown improved performance compared to more standard methods.²⁷

In this work, the inhibition of the hepatitis C internal ribosome entry site (HCV-IRES)

IIa subdomain by 2-aminobenzimidazole derivatives is taken as a case study to address the aforementioned challenges and to illustrate how advanced and state-of-the-art computational techniques can be employed to compute the affinity of a class of small molecules for their RNA target with high accuracy.

The HCV-IRES is essential for the initiation of viral protein synthesis and adopts well-defined folds that are potential targets for antiviral translation inhibitors. The 3D structure of the IRES domain IIa in complex with a benzimidazole translation inhibitor at 2.2 Å resolution (PDB ID 3TZR) was determined and provides insights into inhibiting its function.^{10,41} This system is chosen for its complexity that makes it one of the most challenging systems for atomistic simulations and absolute binding affinity calculations. First, the ligand-binding site has an intricate architecture, organized by a metal spine at the back of the cavity. Additionally, it contains three magnesium ions as intrinsic structural components (Fig. 1). Moreover, the inhibitors are 2-aminobenzimidazole derivatives containing 2 to 3 positive charges, as well as several aromatic and non-aromatic cycles with lengthy arms (see Fig. 2).⁴² Finally, the comparison of the structures of bound and unbound RNA shows that the RNA undergoes a dramatic ligand-induced conformational adaptation from an L-shaped to an extended form, creating a deep pocket that resembles the substrate binding site in riboswitches. Also, the aforementioned three magnesium ions undergo adaptive reorganization upon binding of the benzimidazole ligand.⁴¹

In order to comprehend the binding mode of these inhibitors and reproduce their binding affinities, we have resorted to a computational model that accounts for the underlying effects of polarization and electron anisotropy, AMOEBA,^{18,43,44} recently parameterized for DNA and RNA using quantum chemistry (QC) methods¹⁹ and tested in several oligonucleotide simulations.^{32,33,45} We have also used the massively parallel GPU-accelerated molecular dynamics Tinker-HP software package^{22,23,46,47} coupled to the Colvars library.^{27,48} We have put

in place a novel state-of-the-art approach for ABFE calculations tailored for Riboswitch-like RNA using the Lambda-ABF scheme²⁷ combined with positional, orientational, and conformational restraints.³⁸ We also propose a strategy to estimate the free energy difference associated to the conformational change between the Apo and the Holo structure of the Riboswitch through advanced simulation techniques leveraging machine learned collective variables.⁴⁹

Methods

System preparation

The initial input for the simulation comprises the 3D X-ray structure of Riboswitch-like RNA in the presence of a benzimidazole derivative (ligand 4 shown in Fig. 1 and Fig. 2d), obtained from the Protein Data Bank (PDB ID: 3TZR, resolution 2.21 Å). Using this structure as a starting point, we generated the input for all other ligands illustrated in Fig. 2. Concerning, ligands 5 and 6 which are diastereomers they were designed to be conformationally restricted benzimidazoles to increase the binding, and the desired conformation was separated by HPLC.⁴² Hence, for our simulations the desired 3D conformation was generated based on ligand 4 (SS0 in 3TZR). For consistency, we used a 30-mer RNA, as depicted in Fig. 1b, with all compounds. In the work of Seth *et al.*, ligands 4, 5, and 6 were measured with a 40-mer construct, while ligands 1, 2, and 3 were measured with the 29-mer construct. However, this does not significantly affect the K_D value or the ranking, given that the results for compound 3 reported in Tables 2 and 3 of the paper⁴² are consistent. Compound 3 was tested with both the 29-mer and 40-mer RNA and it was found to have a slightly stronger binding to the 29-mer ($K_D=10$ μM in the 40-mer compared to 17 μM in the 29-mer). Given the experimental measurement error, we expect the results to be within the same range.

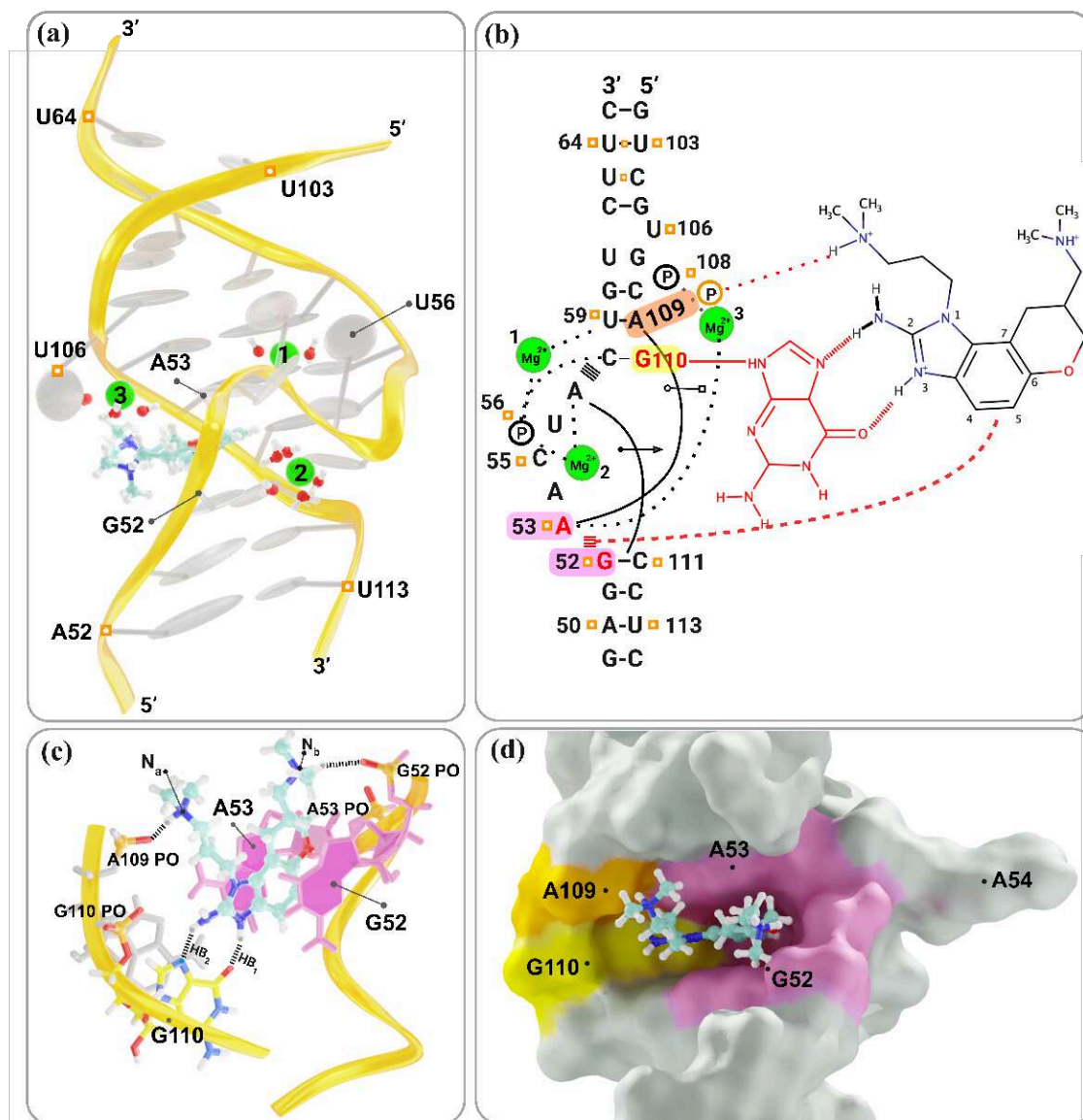


Figure 1: **Structural features of the Riboswitch-like RNA-Benzimidazole complex.** (a) Overall view of the Riboswitch-like RNA complex with Benzimidazole (ligand 4) after equilibration. The ligand is depicted in fluorescent blue. Three Mg²⁺ ions, shown in green, are coordinated by surrounding water molecules. (b) Schematic representation of interactions within the ligand-binding site. Hydrogen bonds are depicted by dashed lines. Non-Watson–Crick base pairs are shown with solid lines and symbols according to Leontis et al.⁵⁰ Stacked lines (\equiv) denote base stacking and intercalation of the ligand, which is colored pink. Residues that interact with the benzimidazole are in red. (c) Detailed view of the ligand-binding site. The bases of G52 and A53, which form the intercalation site for the benzimidazole scaffold, are shown in pink. The purine base of G110 is colored yellow. Hydrogen bond interactions between G110 and the ligand are depicted with dashed lines and labeled as HB₁ and HB₂. Possible hydrogen bonds between the NH⁺ group of the ligand and A109 and G52 are also shown by dashed lines. (d) Surface representation highlighting the ligand-binding pocket. The interaction regions of A53 and G52 are colored pink, G110 is shown in yellow, and A109 is depicted in orange.

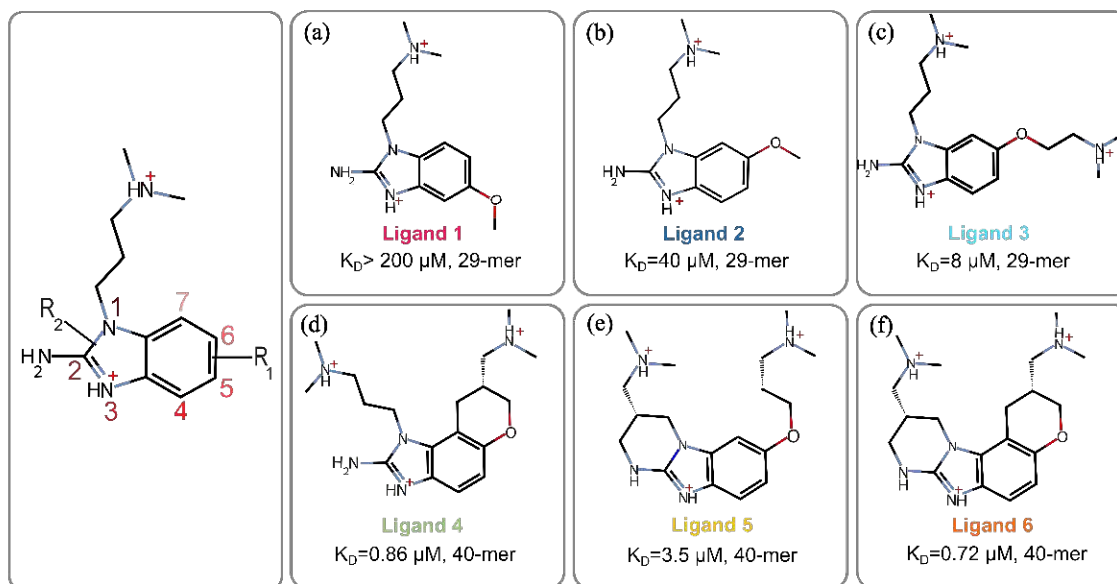


Figure 2: **2D chemical structures of translation inhibitors.** Left: 2D structure of the scaffold with common atoms labeled 1 to 7. The potential growing regions are highlighted as R_1 and R_2 . Panels (a) to (f) display the 2D structures of ligands 1 through 6. For each ligand, the dissociation constant (K_D) and the corresponding size of RNA are reported from Ref. 42.

AMOEBA Force field parameters

The systems studied here are highly charged and dynamic, while being surrounded by ions and polarizable water molecules. To accurately model these systems, we employed the AMOEBA force field, which accounts for polarization and anisotropy through a multipolar representation of electrostatics. AMOEBA has been recently parameterized for DNA and RNA based on QC methods and validated in several oligonucleotide simulations.¹⁹ Additionally, we used the Tinker-HP software on GPUs⁴⁶ for efficient computational performance.

Generating AMOEBA Force Field Parameters for Small Molecules: Ligand parameterization is performed using QC calculations and the Poltype package⁵¹ which facilitate the automated generation of AMOEBA parameters. The process begins with an SDF file defining the atom coordinate, bond order, ionic state, etc. Prior to geometry optimization, one conformation was selected from 500 conformers generated by the RDKit tool. Structure characteristics

including intra-molecular hydrogen bonds, radius of gyration, solvent-accessible surface area were considered during the selection. Key torsion angles were kept frozen to the values of the selected conformer during subsequent MP2/6-31G(d) geometry optimization to maintain the preferred conformation. Two single-point (SP) calculations were performed on the optimized structure using the Psi4 software:⁵² MP2/6-311G(d,p) for low-level and MP2/aug-cc-pvtz for high-level. Electron density from the low-level SP was used with the GDMA program⁵³ (with switch value 0) to generate initial atomic multipoles (charges, dipoles, and quadrupoles). Electron density from high-level SP was employed to calculate the electrostatic potential (ESP) on defined grids surrounding the molecule. The POTENTIAL program within Tinker 8 software⁵⁴ further refined the dipoles and quadrupoles by fitting them to reproduce the ESP. Valence (bond, angle, stretch-bend, and opbend), van der Waals (VDW), and torsion terms were matched using SMARTS patterns and assigned from published parameter databases.⁵⁵ Notably, VDW parameters for aromatic rings were assigned using those from a previous study⁵⁶ due to their improved interactions with DNA and RNA. For any missing torsion parameters, a conventional spin-and-fit procedure was used. This involved restrained optimization using the xtb program⁵⁷ with GFN-2 model⁵⁸ and SP energy calculations at ω b97x-d⁵⁹/6-311+G(d) level in Psi4 software.⁵² To minimize computational cost and potential steric clashes during rotating the bond of the whole ligand, a fragment molecule containing the fitted torsions was extracted from the ligand and used for torsion parameter fitting.

MD Simulations

All systems were neutralized by adding the appropriate number of counterions (K^+) in a water box of $80 \text{ \AA} \times 80 \text{ \AA} \times 80 \text{ \AA}$ such that the distance between the FLAP and its periodic images is at least 16 \AA . The systems contain $\sim 52,000$ atoms and were subject to NVT and NPT simulations for equilibration before production.

The systems underwent initial minimization using Tinker-HP "minimize" program, followed by a gradual heating process in 250 ps intervals. The temperature was increased from 50 K to 100 K, 200 K, and 300 K under the NVT ensemble, with RNA-ligand atoms restrained using a force constant of 10 kcal/(mol·Å²). During this step, MD simulations were conducted using the Verlet integrator with a time step of 1.0 fs. Subsequently, the systems were equilibrated in the NPT ensemble at 1 atm using the same restraints with force constants of 10 and 5 kcal/(mol·Å²), each for 0.5 ns, using the RESPA integrator⁶⁰ and a time step of 2 fs. To relax the water molecules around the ions at 300 K, a 10 ns restrained MD simulation was run for each system with a force constant of 1 kcal/(mol·Å²), followed by simulations where the restraint was applied only to phosphate (P) atoms and Mg²⁺ ions with a force constant of 1 kcal/(mol·Å²), run for 5 ns. The ligands were stabilized by employing distance-to-bound-configuration (DBC) restraints³⁸ through the Colvars library,⁴⁸ constraining their root mean square deviation (RMSD) from the bound configuration. For DBC definition, all the ligands' heavy atoms and the RNA heavy atoms within a 4 Å radius of the ligand were included. Temperature and pressure were controlled by the Bussi⁶¹ thermostat and Berendsen⁶² barostat, respectively. Van der Waals interactions utilized a 9 Å cut-off, while electrostatic interactions were treated by particle mesh ewald (PME)⁶³ with a real-space cutoff of 7 Å. Induced dipoles were calculated with a Preconditioned Conjugate Gradient (PCG) solver with a convergence tolerance of 1×10^{-5} Debye.⁶⁴

The polyanionic nature of RNA attracts water molecules and counterions, influencing its stability. Additionally, Mg²⁺ ions can shed their hydration shells and establish strong interactions with specific sites in the RNA, promoting local structural rearrangements. To mitigate artifacts during the production phase, positional restraints were applied to the phosphate (P) atoms and Mg²⁺ ions with a force constant of 1 kcal/(mol·Å²). To prevent fraying, distance restraints were added to enforce the hydrogen bonds between the terminal base pairs. The partially restrained simulations were run for 40 ns for production.

Distance-to-Bound-Configuration (DBC) coordinate for Binding Pose Determination and Restraint Definition

Accurate alchemical absolute free energy calculations require the use of restraints to ensure proper convergence. Typically, these restraints involve harmonic restraints between centers of mass of the ligand and the binding site (positional restraints)⁶⁵ or more sophisticated methods like 'Boresch' restraints,⁶⁶ which act on distances, angles, and dihedrals (positional and orientational restraints). In this study, we employed a more refined approach using DBC collective variables. DBC measures the RMSD of ligand coordinates relative to the moving frame of the receptor's binding site.³⁸ This metric captures positional, orientational and conformational deviation of the ligand in a single collective variable. The restraint is flat-bottomed with a starting value that matches the distribution characteristic of the binding mode. This distribution can be efficiently monitored using the Colvars Dashboard⁶⁷ within the VMD software.⁶⁸ Depending on the binding scenario, this distribution will be narrow for poses that are tightly bound, while it will be broader for systems with a diverse set of loosely bound configurations. The primary goal of these restraints is to restrict sampling during the decoupling process to those configurations that are relevant in the fully coupled state and thus facilitate convergence.

Lambda-ABF for Absolute Binding Free Energy (ABFE) Simulations

As stated above, we have resorted here to alchemical free energy simulations in which the ligand is progressively “alchemically” decoupled from its environment, first in complex with the Riboswitch host, and separately from the bulk solvent. The standard free energy of binding is obtained through a thermodynamic cycle. Such techniques require the sampling of

the alchemical Hamiltonians. State-of-the-art methods typically use a predefined λ schedule with simulations at fixed λ values and a post-processing stage in which free energy differences are computed using estimators such as Thermodynamic Integration (TI)⁶⁹ or Free Energy Perturbation (FEP).⁷⁰ However, these fixed- λ simulations can hinder the relaxation of orthogonal degrees of freedom that may benefit from a variable λ exploration which can be made possible through additional enhanced sampling techniques such as Hamiltonian Replica Exchange⁷¹ or Expanded Ensemble methods.⁷² Here, we used the newly developed Lambda-ABF approach²⁷ which enables efficient relaxation of orthogonal degrees of freedom due to the variable nature of λ , as well as ergodic sampling thanks to the ABF bias. Its sampling efficiency has been shown to be superior compared to more traditional techniques. We utilized its implementation within Tinker-HP, in combination with the Colvars library,⁴⁸ which facilitates user-friendly convergence estimation without requiring post-processing, and allows integration with other CV-based methods.

For calculating the free energy surface of our systems at T=300 K, we used four walkers with the Lambda-ABF method and decoupled VDW and Electrostatics (ELE) interactions sequentially in separate legs. Each walker was run for 50 ns per leg. We ran three independent replicas for each setup. As for the free energy cost associated with the release of the DBC restraint, it is computed in the gas phase through TI by progressively releasing it to a compatible harmonic distance restraint which can then be computed analytically.^{27,38,73}

During the alchemical decoupling, the phosphate atoms of the RNA backbone and the Mg²⁺ ions were restrained with a slight force constant of 1 kcal/(mol·Å²). It allows to, (i) prevent any destructureation, (ii) measure the ligand binding contribution independently from any Apo-Holo conformation changes by fixing the endpoint of the thermodynamic cycle.

Concerning the ion concentration used, it is well known that it can significantly impact the results of binding affinities,⁷⁴ especially for charged systems such as RNA-ligand com-

plexes.⁷⁵ Reproducing all experimental conditions is crucial to get meaningful results in free energy simulations. However, methodological issues can arise when computing electrostatics using standard methods like PME in periodic boundary conditions. Although finite size effects diminish with larger simulation boxes, they may still be non-negligible in practice.⁷⁶ A notable issue occurs with PME in non-neutralized boxes which happen by default when scaling down electrostatics of ligands during alchemical simulations. In the reference experimental setup for mass spectrometry (MS) based binding assay that was used to measure the K_D , the incorporation of a physiological ion concentration was not evoked.⁴² Given the positively charged nature of all the ligands reported here (2 to 3 positive charges) we cannot scale down the charge of anions to maintain a neutral charge of the simulation box in this setup. We have thus employed two distinct protocols for ABFE calculations:

- *Neutralized setup (Neut.):* K^+ counter-ions were added to neutralize the system, which led to non neutral boxes in the full decoupling state (total charge of 2 to 3, depending on the ligand). This setup allows us to assess the impact of non neutral boxes and the absence of physiological ions on binding affinities.
- *Physiological Ion Concentration (Phys.):* 0.15 Molar of KCl were added to the system 2 to 3 Cl^- ions had their charge scaled down during ligand scaling (neutral boxes maintained). This setup aims at replicating physiological conditions more closely and evaluating the effect of physiological ionic strength on the computed binding affinities.

OPES Explore

OPES Explore⁷⁷ is a CV-based enhanced sampling technique which is an evolution of metadynamics.⁷⁸ It aims at broadening the system's sampling to a target probability distribution known as the Well-tempered distribution, defined as

$$p^{\text{tg}}(\mathbf{s}) \propto [P(\mathbf{s})]^{\frac{1}{\gamma}}$$

where $P(\mathbf{s})$ represents the unbiased marginal distribution of chosen collective variables (CVs), and γ is the bias factor controlling the broadening.

To achieve this, Gaussian kernels are employed to reconstruct $p^{\text{tg}}(\mathbf{s})$, which in turn determines the bias potential $V(\mathbf{s})$ through the following recursive strategy at step n :

$$V(\mathbf{s}) = (\gamma - 1)k_{\text{B}}T \log \left(\frac{p^{\text{tg}}(\mathbf{s})}{Z} + \epsilon \right)$$
$$p^{\text{tg}}(\mathbf{s}) = \frac{1}{n} \sum_{k=1}^n G_k(\mathbf{s}, \mathbf{s}_k)$$

where k_{B} is the Boltzmann constant, T is the temperature set by the thermostat, Z is a normalization factor, and $G_k(\mathbf{s}, \mathbf{s}_k)$ represents the Gaussian kernel deposited at step k .

The initial Gaussian kernel width, often denoted as **SIGMA**, typically corresponds to the standard deviation of \mathbf{s} in the initial basin. The bias factor γ is typically set using $\gamma = \Delta E / (k_{\text{B}}T)$, where ΔE is the barrier parameter. The regularization term ϵ ensures stability and is related to ΔE through $\epsilon = e^{-\Delta E / (k_{\text{B}}T(1-1/\gamma))}$.

The barrier parameter ΔE sets a limit on the maximum bias energy that OPES⁷⁹ can introduce into the system. It should be large enough to facilitate transitions away from the initial basin while preventing the system from accessing irrelevant high-energy states that are difficult to reverse.

For further details, refer to the original literature introducing the OPES method.^{77,79}

Enhanced CV Design using Deep-LDA

Deep-LDA⁴⁹ is a way to define CV that correspond to a transition between two conformational states. It is rooted in Linear Discriminant Analysis (LDA),⁸⁰ a classical technique in classification tasks. Here, it is tailored to differentiate between Apo and Holo states observed in unbiased system simulations.

Standard LDA optimizes Fisher's ratio $\frac{\mathbf{w}^T \mathbf{S}_b \mathbf{w}}{\mathbf{w}^T \mathbf{S}_w \mathbf{w}}$, where \mathbf{w} defines the discriminant direction $s(R) = \mathbf{w}^T \mathbf{d}(R)$. Here, \mathbf{S}_b and \mathbf{S}_w represent between- and within-class scatter matrices

respectively, calculated from descriptor distributions in Apo and Holo states.

Deep-LDA extends LDA by employing a neural network (NN) to process N_d descriptors \mathbf{d} , optimizing LDA within the last hidden layer \mathbf{h} . The resulting Deep-LDA CV $s = \mathbf{w}^T \mathbf{h}$ offers a non-linear, expressive projection that has proven effective in various applications.⁸¹⁻⁸³ To enhance usability in simulations, we apply a cubic transformation, $s_w = s + s^3$,^{81,84} ensuring smoother distributions.

Results

In this section, we first assess the binding mode of the studied small molecules to their RNA target, followed by computing their binding affinities. Finally, we explore the conformational changes between the Apo and Holo states of the RNA.

Binding mode analysis

The RNA architecture of the ligand-binding pocket consists of the phosphate of U56, which is rotated into the RNA helix major groove, and two magnesium ions anchored between the phosphate oxygen atoms and the bases of A57 and U59 (Fig. 1b). The extreme contortion of the RNA backbone that directs the U56 phosphate into the helix interior is further pinned in place by hydrogen bonds between the C55 phosphate and 2'-hydroxyl groups of the flanking residues A54 and U56. The elaborate network of backbone and metal interactions forces the base of A54 to project away from the RNA whereas residues C55 and U56 tightly pack in the RNA helix major groove as well as against the backbone of C58 and U59 (Fig. 1d). The third magnesium ion closes part of the solvent exposed part of the ligand site by bridging the RNA strands with interactions at the base of A53 and the C108 phosphate. The top and bottom part of the binding pocket are stabilized by base triples which form through cross-bracing interactions along the RNA helix with A53 docking at the Hoogsteen edge of A109 and A57 interacting with the sugar edge of C111. Additional hydrogen bonds sta-

bilize the distorted internal loop in the benzimidazole complex and include interactions of the U56 phosphate with the C58 base and of the C55 2'-hydroxyl with the A53 base (Fig. 1b).

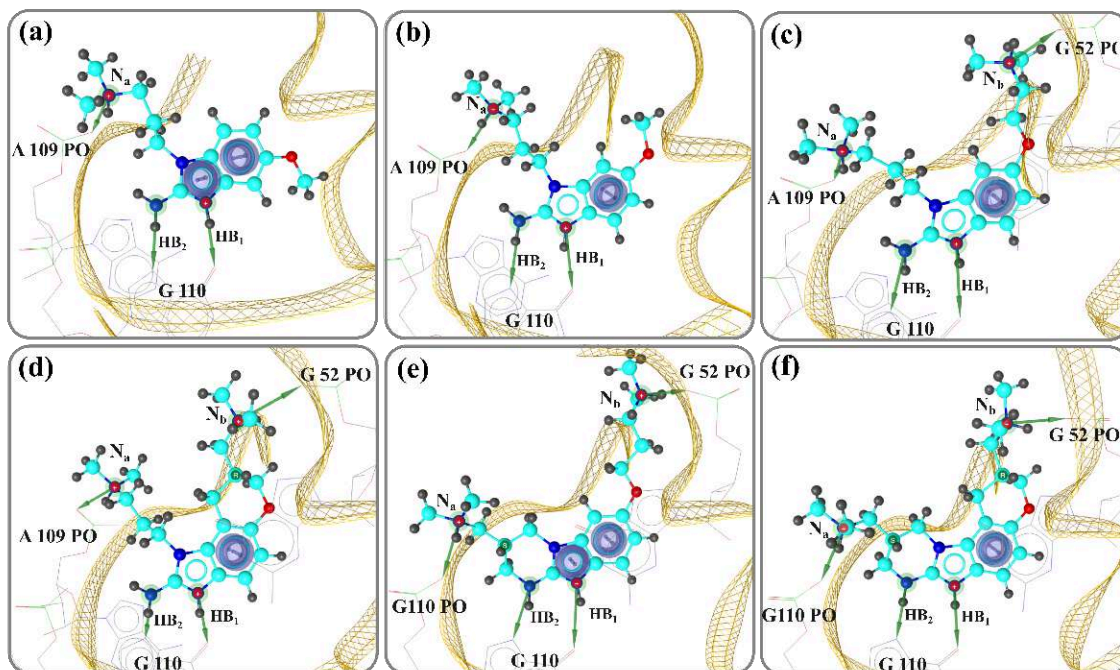


Figure 3: 3D representation of the binding modes of the studied ligands and their interactions with the environment. Ligands are depicted in licorice style and colored in cyan, while the phosphate backbone is represented by a yellow meshed ribbon. Green arrows indicate hydrogen bond donors, and purple circles indicate π - π stacking interactions between the G52, A53 bases and the ligand. The positively charged nitrogen (N) atoms are shown in red with a superscript (+). The chirality labels are displayed as **S** or **R** for chiral centers. Panels (a) to (c) represent the aforementioned information for ligands 1 through 6, respectively.

Understanding this intricate architecture is key to analyzing how the ligands interact with the RNA structure. It is well-known that the contribution of hydrogen bonding and stacking interactions are predominate in RNA recognition.⁸⁵ Following the detailed relaxation process described in the simulation setup, we analyzed the binding modes obtained from the 40 ns MD simulations. Figure 3 illustrates the binding modes of all ligands within the target RNA. Three primary interactions are identified as crucial for stabilizing the ligand within the binding pocket: (i) a π - π stacking interaction occurs between the ligand and the nucleic bases G52 and A53, (ii) two hydrogen bonds (labeled HB₁ and HB₂) formed between the

central core of the ligand (benzimidazole) and G110 (see Fig. 1c), and (iii) an additional ionic interaction is established between the NH^+ moiety of the dimethylazanium group at the end of each arm of the ligand and the non-bridging phosphate oxygen atoms (PO) of A109 (or G110) and the PO of G52 (or A53) on the opposite strand (depicted as N_a and N_b , respectively in Fig. 3).

With these key interactions established, their evolution as a function of time was computed throughout the trajectory and for each ligand; the normalized distributions are shown in Fig. 4. For the π - π stacking interactions, we computed the distance between the center of mass (labeled as COM) of the purine rings of G52 and A53, and the COM of the benzimidazole scaffold of the ligand (referred to as LIG_{COM} in Fig. 4). For the distances between G52_{COM} and LIG_{COM} , the results reveal that for most ligands, the distribution is narrow and centers around 4.1 to 4.4 Å, consistent with the typical π - π interaction range of 3.3 to 4.4 Å. In contrast, for ligands 1 and 2, the distribution is broader, with ligand 1 centering around 4.8 Å while showing a shoulder at 4.0 Å. As for ligand 2, the distribution centers around 4.3 Å with a shoulder at 5 Å. This is concomitant with the fact that ligand 1 is a non-binder, while ligand 2 is the weakest binder in the list. Similarly, the distribution of distances between A53_{COM} and LIG_{COM} shows a comparable trend for all ligands. Here, the distribution also centers around a narrow range, specifically between 3.6 to 3.8 Å for ligands 3 to 6. This indicates that the overall trend in interaction distances is consistent with the π - π stacking interactions observed with G52_{COM} , albeit with a slightly shorter average distance for the interactions involving A53_{COM} .

Next, the hydrogen bond interactions, HB_1 and HB_2 , were analyzed. For these interactions, ligand 1 exhibits a broader distribution, suggesting a weaker and more transient interaction. In comparison, ligand 6 exhibits a narrower distribution, centered around 2 Å with the highest intensities, reflecting strong and consistent interactions. Ligands 3, 4, and 5 show a peak similar to ligand 6 for HB_2 , although with slightly lower intensity for HB_1 . Ligand 2

shows a similar peak as the other ligands; however, it has a tail at low probability, suggesting occasional weaker interactions.

Finally, regarding the ionic interaction, the N_a atom of ligands 2, 3, and 4 forms a hydrogen bond with the phosphate (PO) atoms of A109, as indicated by a sharp peak around 2 Å. Ligand 6 is unique in forming an HB with the PO atom of G110. For the second arm of the ligand (N_b atom), ligands 4, 5, and 6 alternate between the PO atom of G52 and that of A53, as evidenced by multiple peaks in the distribution, one of which is centered around 2 Å. Ligands 1, 2, and 3 are also the only ligands that do not form a HB with the PO atom of G52, nor with A53, except for a rare HB with the PO atom of G52 at 2 Å for ligand 3.

Binding Affinity Calculations

To accurately predict the binding affinities of extended charged ligands (bearing 2 to 3 charges) bound to a long 30-mer riboswitch-like RNA with three structural Mg^{2+} ions as intrinsic components in the vicinity of the binding pocket, we have developed a specific protocol utilizing the Lambda-ABF method and DBC restraints.

As free energy is a state function, the computed ΔG s are fundamentally independent of the intermediate states sampled during an alchemical simulation. This allows for tailored modifications in these states to ensure better sampling and convergence as it is routinely done in restraining the ligand to keep it in the binding site while scaling down its non-bonded interactions with the environment. As mentioned earlier, different types of restraints exist that can either be positional, or positional and orientational. However, these two methods suffer from distinct limitations,^{27,38} especially when considering elongated ligands containing two flexible arms. What is required in such systems is positional, orientational as well as conformational restraints, that can be defined through the newly developed DBC collective variable, which is the RMSD of the ligand in the moving reference frame of the binding site.³⁸ Therefore, to facilitate sampling and maintain the ligand in the binding site, we employed DBC restraints between the benzimidazole ring of all ligands and heavy atoms from

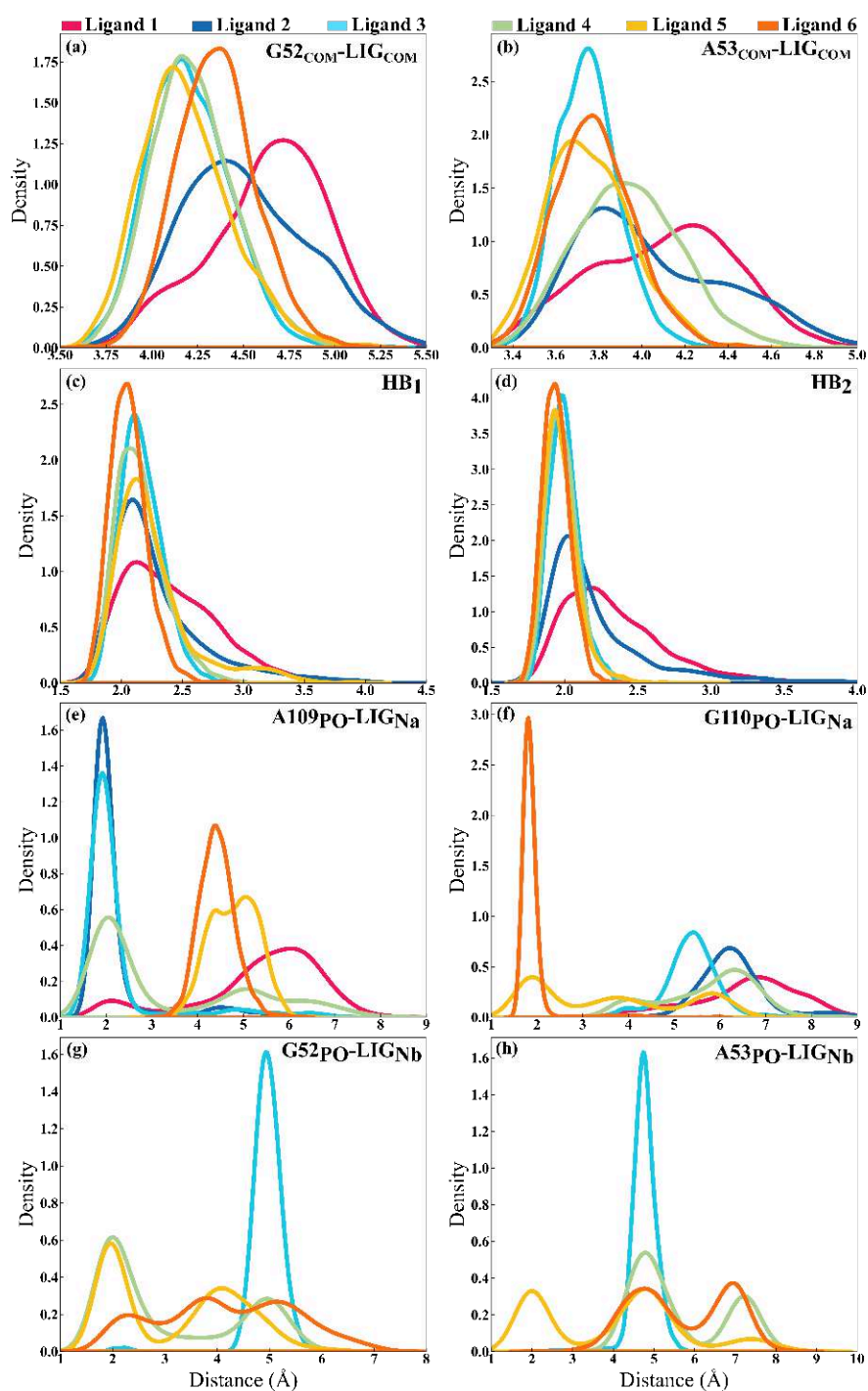


Figure 4: **Normalized distributions of key interaction quantities for six ligands.** Panels (a) to (h) show the distributions for distances between (a) $G52_{COM}$ and LIG_{COM} , (b) $A53_{COM}$ and LIG_{COM} , (c) HB_1 distances, (d) HB_2 distances, distances between (e) $A109_{PO}$ and LIG_{Na} , (f) $G110_{PO}$ and LIG_{Na} , (g) LIG_{Nb} and $G52_{PO}$, (h) LIG_{Nb} and $A53_{PO}$. The colors represent different ligands as indicated in the legend.

the RNA binding site. The selected atoms are highlighted in SI Fig. S15.

Besides, it is well known that starting alchemical simulations from Holo structures may lead to incomplete relaxation towards the Apo structure.⁸⁶ This incomplete relaxation can result in a significant free energy difference, potentially causing errors of several kcal/mol in the final prediction. Typically, to address this, all predicted affinities are usually shifted by a common value to account for this discrepancy.⁸⁶ This adjustment assumes that all systems experience the same degree of relaxation from the Holo to the Apo state. To ensure the validity of this assumption, facilitate the sampling and maintain structural stability of the RNA-ligand complex, we implemented slight restraints with a force constant of 1 kcal/(mol·Å²) on the phosphate (P) atoms of RNA and Mg²⁺ ions. These adjustments were designed to avoid impacting the fully bound state, ensuring the integrity of the complex throughout the simulations. This approach allows us to independently estimate and account for the relaxation effect, thus improving the accuracy of our absolute binding affinity calculations.

In the case of our system, it is known that benzimidazole inhibitors can induce a conformational change characterized by a widened interhelical angle in the IRES subdomain IIa.^{41,87,88} This conformation facilitates the undocking of subdomain IIb from the ribosome, leading to the inhibition of IRES-driven translation in HCV-infected cells. As illustrated in Fig. 5, the L-shaped RNA structure in the Apo state is known to adopt a straighter configuration.^{41,87,88} Apart from the restructuring of several nucleic bases, this conformational change also includes a rearrangement of Mg²⁺ ion positions (Fig. 5a and b). To mitigate any potential bias from such known conformational changes, we applied the restraints described earlier. This approach helps ensure that such known rearrangements are properly accounted for, thereby improving the reliability of our ABFE calculations.

To obtain the free energy surface for each system at $T = 300$ K, we ran three replicas, each utilizing four walkers. Calculations were conducted in both the complex (bound state) and

solvent (ligand in the bulk) phases. For the complex phase, simulations ran for approximately 50 ns for both VDW and ELE legs. In contrast, a 20 ns run was enough for convergence in the solvent phase for both VDW and ELE components. Raw data from each replica are provided in the Supplementary Information (SI), Tables S1 and S2. All convergence plots are provided in SI, see Figs. S1 to S24. The free energy cost associated with the release of the DBC restraint is readily computed in the gas phase using TI.^{38,73} This is achieved by gradually releasing the restraint to a compatible harmonic distance restraint, the cost of which can be computed analytically (see Table S3 in SI).

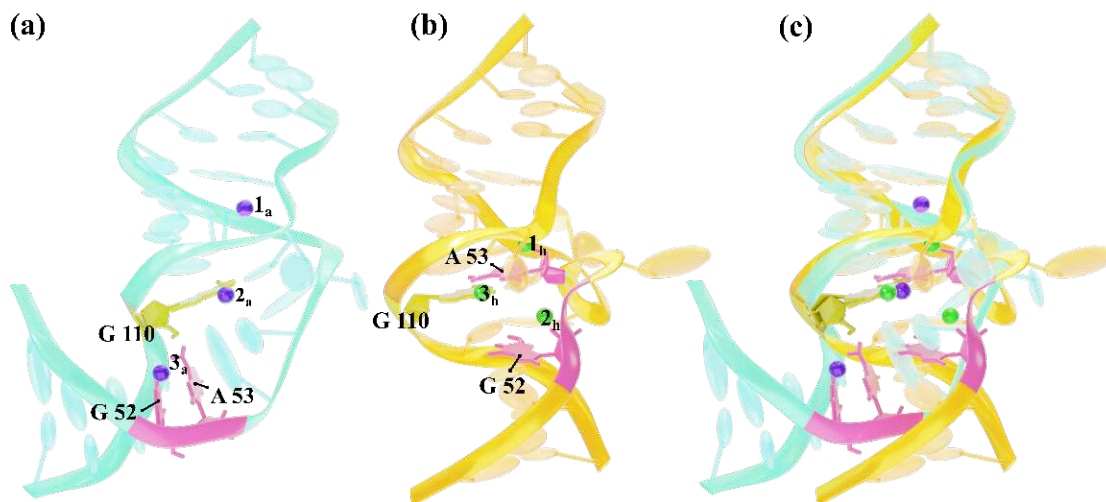


Figure 5: **Conformational Differences between Apo and Holo Forms of RNA.** (a) L-shaped structure of the Apo state (PDB ID: 2nok) and (b) straight structure of the Holo state (PDB ID: 3tzt). (c) Alignment of the Apo and Holo structures. The positions of Mg^{2+} ions in each structure are illustrated, denoted as i_a and i_h for the Apo and Holo states respectively, where $i = 1 - 3$. The key bases within the binding pocket are highlighted, with G110 in dark yellow and G52 and A53 in pink.

As described in Computational Details, we have employed two distinct setups for ABFE calculations, one consisting of a neutralized system (Neut.) and a second including a physiological ion concentration (Phys.). The computed ΔG values in neutralized condition ($\Delta G_{\text{Comp-Neut}}$) and in the presence of physiological ion concentration ($\Delta G_{\text{Comp-Phys}}$) are presented in Table 1 and compared with the experimental data. The raw absolute binding

Table 1: The values of K_D , experimental, and computed ΔG for each ligand. Computed ΔG_{Comp} are reported for two different conditions (in kcal/mol): neutralized complex ($\Delta G_{\text{Comp-Neut}}$) and including physiological ion concentration ($\Delta G_{\text{Comp-Phys}}$). For Ligands 4, 5, and 6, the experimental ΔG_{Exp} is measured on a 40-mer RNA, while for the rest, the 29-mer is used. The computed ΔG_{Comp} and errors represent the mean and standard error of the mean from three replicas for each ligand. The raw data for computed ΔG is reported in parentheses.

Ligand	K_D	ΔG_{Exp}	$\Delta G_{\text{Comp-Neut}}(\Delta G_{\text{Comp-raw}})$	$\Delta G_{\text{Comp-Phys}}(\Delta G_{\text{Comp-raw}})$
1	>200	>-5.05	-0.07 (-10.43) \pm 0.92	-0.76 (-12.08) \pm 0.88
2	40	-6.01	-6.61 (-16.96) \pm 0.68	-6.67 (-18.00) \pm 0.83
3	8	-6.69	-6.53 (-16.88) \pm 1.54	-6.58 (-17.90) \pm 0.75
4	0.86	-8.27	-7.13 (-17.48) \pm 0.65	-8.42 (-19.75) \pm 1.35
5	3.5	-7.44	-7.21 (-17.56) \pm 0.81	-5.66 (-16.98) \pm 0.19
6	0.72	-8.39	-9.60 (-19.95) \pm 1.54	-9.75 (-21.08) \pm 0.05

free energies ($\Delta G_{\text{Comp-raw}}$) consistently exhibit stronger values (favorable) compared to the experimental results. In order to account for the Apo-Holo conformational change, the raw binding free energy ($\Delta G_{\text{Comp-raw}}$) is converted to the final binding free energy ΔG_{Comp} by adding a constant shift, which is calculated as follows $\Delta G_{\text{shift}} = \frac{\sum_i G_i^{\text{Exp}} - \sum_i G_i^{\text{Comp}}}{N}$,⁸⁶ so that the mean of computed free energies is equal to the mean of experimental free energies. This shift yield a value of 10.35 ± 0.91 kcal/mol for the neutral setup and 11.33 ± 1.19 kcal/mol in the presence of physiological ion concentration. Despite this slight difference, the results are consistent within the error margins.

The correlation plots are illustrated in Fig. 5. Upon including the systematic shift between the calculated and experimental binding free energies to accommodate the Apo-Holo protein reorganization free energy, we observe a correlation between computed and experimental results across the entire dataset of 0.7 (Pearson r), with a root mean square error (RMSE) of 0.81 kcal/mol and a mean absolute error (MAE) of 0.70 kcal/mol for the neutralized setup (Fig. 5a). For the setup with a physiological ion concentration, a correlation of 0.7 with an RMSE of 1.05 kcal/mol and an MAE of 0.84 kcal/mol was observed (Fig. 5b). It's worth noting that ligand 1 was excluded from the analysis as it is predicted to be a non-binder in

agreement with the experimental measurement which showed no binding at 200 μM . These results show a solid reproduction of the binding mode and a good ranking, with ligand 6 and ligand 4 identified as the most potent and ligand 1 as a non-binder.

Eventhough the experimental K_D measurements were done with 2 different constructs, this shouldn't have a large effect on the affinities, as seen for the example of Compound 3, where $K_D=10\ \mu\text{M}$ in 40-mer compared to 17 μM in the 29-mer,⁴² which translate into 0.3 kcal/mol of difference, or on ranking the selected compounds in our study, it nonetheless might slightly effect the observed correlations. This is also without considering the error on the experimentally measured affinities.

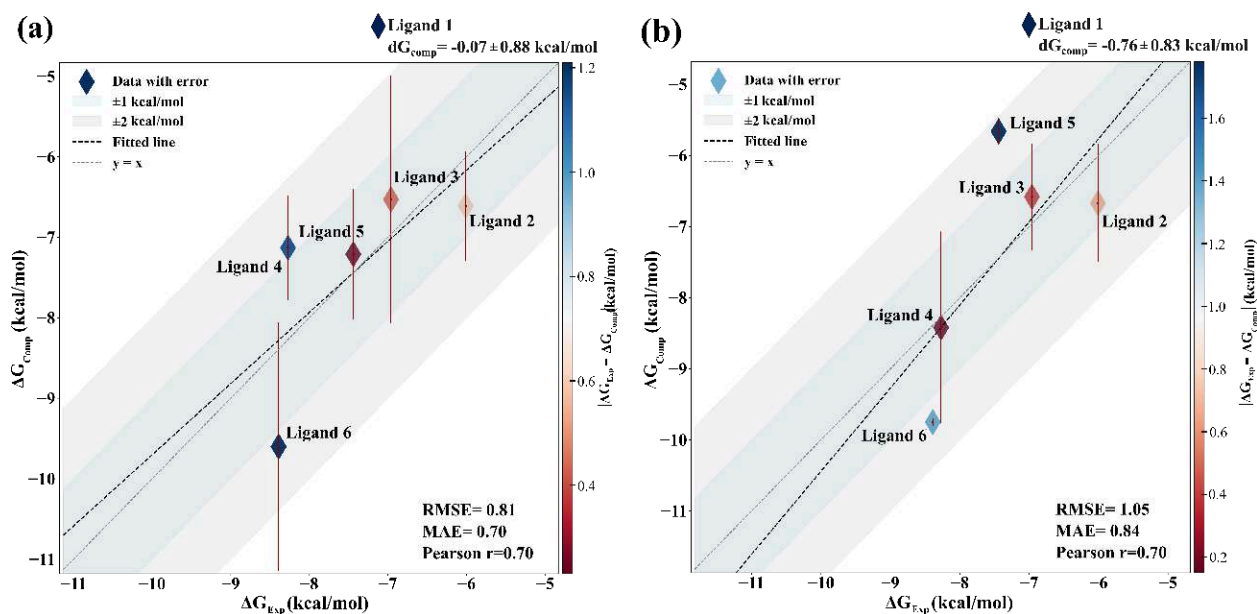


Figure 6: **Experimental vs Computed ΔG values.** (a) Neutralized complex, and (b) complex with physiological ion concentration. Computed ΔG and errors are the mean and standard error of the mean from three replicas for each ligand. The dark shaded region spans ± 1 kcal/mol; the lighter region spans ± 2 kcal/mol. The color bar represents the absolute value difference between experimental and computed values. Pearson r, RMSE, and MAE are reported in the right bottom corner of each plot. The ΔG value of non-binder ligand 1 is reported on the top corner of each panel.

Apo-Holo conformational change

As mentioned in the above subsection, the ΔG_{shift} resulting from the energy difference between the computed ΔG and experimental values is estimated to be between 10.35 and 11.35 ± 1.0 kcal/mol, primarily reflects the Apo-Holo conformational change. To substantiate these findings, employing more direct computational methods is crucial. Given the substantial free energy barrier associated with the Apo-Holo transition, capturing this difference accurately requires advanced sampling techniques. As illustrated in Fig. 5, the Apo-Holo conformational change involves complex collective motions that are challenging to capture with standard methods. To address this, we employed the OPES-Explore^{77,79} enhanced sampling method in conjunction with Deep-LDA,⁴⁹ a machine learning-based technique specifically designed to identify effective CVs for complex transitions.

To identify the Deep-LDA CV, we conducted an unbiased simulation of 40 ns from both Apo and Holo structures, incorporating 47 descriptors that included distances and angles between various P atoms and three Mg²⁺ ions. Subsequently, we performed exploratory OPES simulations, each lasting 10 to 15 ns, with applied biases ranging from 2 to 15 kcal/mol on the Deep-LDA CV. Notably, a transition from the Apo to the Holo state was observed around 10 ns when the applied bias exceeded 13 kcal/mol (see Fig. S26 and S27 in SI), which aligns well with the energy change estimated from our Lambda-ABF calculations. However, the conformational changes of specific bases is not captured by the P and Mg atoms in our Deep-LDA CV as well as the still approximate nature of the OPES strategy.

Despite the inherent challenges in achieving precise free energy calculations given the system's complexity, our approach provides a valuable qualitative approximation of the free energy barrier associated with the Apo-Holo conformational change.

Discussion

The complexity and dynamic nature of RNA, combined with its critical role in various biological processes, makes it a very difficult target in drug discovery. However, with these challenges, it becomes essential to harness the latest computational techniques in order to predict the binding affinities between drug candidates and the target of interest with high precision. In this study, we present a state-of-the-art approach for calculating ABFEs of RNA-binding small molecules for application in drug discovery, showcasing its effectiveness in a challenging RNA system: the HCV-IRES Riboswitch-like RNA. This system presents a highly charged and flexible RNA structure, an intricate binding site with three magnesium ions as intrinsic components, ligands with multiple positive charges and several aromatic and non-aromatic cycles with lengthy arms, as well as a substantial ligand-induced conformational change between the Apo and Holo states.

All in all, by combining the computational performance of the GPU-accelerated Tinker-HP package to enhanced sampling and machine learning, while leveraging the accuracy of the AMOEBA force field, we are able to model challenging drug discovery targets such as RNA and to compute absolute binding affinities to guide drug design with the Lambda-ABF methodology. This synergy not only overcomes the limitations of available free energy methods but also establishes a new standard in computational modeling of RNA-ligand interactions. The present work represents a new achievement in computational methodologies for drug design, where the complexities of RNA and other challenging targets can be addressed with unparalleled accuracy and speed, thus guiding the development of RNA-targeted therapeutics with precision and confidence.

Supporting Information

Additional figures S1-S24 showing the convergence plots of all alchemical simulations; tables S1 and S2, show the electrostatic and van der Waals decomposition of the Potential of

Mean Force (PMF). Illustration of the atoms involved in the definition of the DBC restraint (Figure S25) as well as the contribution of these restraints to the free energy reported in Table S3. Figures S26 and S27 report the RMSD of the biased trajectory corresponding to the Apo/Holo conformational change and the associated PMF.

Data Availability

Raw data generated during the study is available upon request from the authors (E-mail: krystel.elhage@qubit-pharmaceuticals.fr).

Code Availability

The code used during the study is available on GitHub:

<https://github.com/TinkerTools/tinker-hp>,

<https://github.com/TinkerTools/poltype2>,

<https://github.com/Colvars/colvars>

Author Contributions

N.A., C.L., L.L., K.E.H. performed research; N.A., C.L., F.H., J.W.P., P.R., J.H., J.-P.P., L.L., K.E.H. investigation; N.A., L.L., K.E.H. analyzed the results and wrote the paper with input from all authors; L.L. developed new code and theoretical tools; K.E.H conceived and supervised the project, and acquired funding.

Competing interests

J.W.P., P.R., J-P.P., and L.L. are co-founders and shareholders of Qubit Pharmaceuticals. All other authors declare no competing interests.

Acknowledgment

We thank the Grand Equipement de Calcul Intensif (GENCI), Institut du Développement et des Ressources en Informatique (IDRIS), Centre Informatique de l'Enseignement Supérieur (CINES), France, for supporting this work under the grant no A0140714153 and A0160714153 (K.E.H.).

References

- (1) Harries, L. W. RNA Biology provides new therapeutic targets for human disease. Front. Genet. **2019**, 10.
- (2) Zhu, Y.; Zhu, L.; Wang, X.; Jin, H. RNA-based therapeutics: an overview and prospectus. Cell Death & Dis. **2022**, 13, 644.
- (3) Yu, A.-M.; Choi, Y. H.; Tu, M.-J. RNA drugs and RNA targets for small molecules: principles, progress, and challenges. Pharmacol. Rev. **2020**, 72, 862–898.
- (4) Lightfoot, H. L.; Smith, G. F. Targeting RNA with small molecules—A safety perspective. British Journal of Pharmacology **2023**,
- (5) Bennett, C. F. Therapeutic antisense oligonucleotides are coming of age. Annu. Rev. Med. **2019**, 70, 307–321.
- (6) Luther, D. C.; Lee, Y. W.; Nagaraj, H.; Scaletti, F.; Rotello, V. M. Delivery approaches for CRISPR/Cas9 therapeutics in vivo: advances and challenges. Expert. Opin. Drug Deliv. **2018**, 15, 905–913.
- (7) Wu, P. Inhibition of RNA-binding proteins with small molecules. Nat. Rev. Chem. **2020**, 4, 441–458.

- (8) Childs-Disney, J. L.; Yang, X.; Gibaut, Q. M.; Tong, Y.; Batey, R. T.; Disney, M. D. Targeting RNA structures with small molecules. Nat. Rev. Drug Discov. **2022**, 21, 736–762.
- (9) Howe, J. A. et al. Selective small-molecule inhibition of an RNA structural element. Nat. **2015**, 526, 672–677.
- (10) Dibrov, S. M.; Parsons, J.; Carnevali, M.; Zhou, S.; Rynearson, K. D.; Ding, K.; Garcia Segal, E.; Brunn, N. D.; Boerneke, M. A.; Castaldi, M. P.; Hermann, T. Hepatitis C virus translation inhibitors targeting the internal Ribosomal entry site. J. Med. Chem. **2014**, 57, 1694–1707, PMID: 24138284.
- (11) Palacino, J. et al. SMN2 splice modulators enhance U1-pre-mRNA association and rescue SMA mice. Nat. Chem. Biol. **2015**, 11, 511–517.
- (12) Ratni, H. et al. Discovery of Risdiplam, a Selective Survival of Motor Neuron-2 (SMN2) Gene Splicing Modifier for the Treatment of Spinal Muscular Atrophy (SMA). J. Med. Chem. **2018**, 61, 6501–6517, PMID: 30044619.
- (13) Gresh, N.; Hage, K. E.; Goldwaser, E.; de Courcy, B.; Chaudret, R.; Perahia, D.; Narth, C.; Lagardère, L.; Lipparini, F.; Piquemal, J.-P. In Quantum Modeling of Complex Molecular Systems; Rivail, J.-L., Ruiz-Lopez, M., Assfeld, X., Eds.; Springer International Publishing: Cham, 2015; pp 1–49.
- (14) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable force fields for biomolecular simulations: recent advances and applications. Annu. Rev. Biophys. **2019**, 48, 371–394.
- (15) Shi, Y.; Ren, P.; Schnieders, M.; Piquemal, J.-P. Reviews in Computational Chemistry Volume 28; John Wiley and Sons, Ltd, 2015; Chapter 2, pp 51–86.

- (16) Melcr, J.; Piquemal, J.-P. Accurate biomolecular simulations account for electronic polarization. Front. Mol. Biosci. **2019**, 6, 143.
- (17) El Khoury, L. et al. Computationally driven discovery of SARS-CoV-2 Mpro inhibitors: from design to experimental validation. Chem. Sci. **2022**, 13, 3674–3687.
- (18) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio Jr, R. A., et al. Current status of the AMOEBA polarizable force field. J. Phys. Chem. B **2010**, 114, 2549–2564.
- (19) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA polarizable atomic multipole force field for nucleic acids. J. Chem. Theory Comput. **2018**, 14, 2084–2108.
- (20) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. Journal of Chemical Theory and Computation **2007**, 3, 1960–1986, PMID: 18978934.
- (21) El Hage, K.; Piquemal, J.-P.; Hobaika, Z.; Maroun, R. G.; Gresh, N. Substituent-modulated affinities of halobenzene derivatives to the HIV-1 integrase recognition site. Analyses of the interaction energies by parallel quantum chemical and polarizable molecular mechanics. J. Phys. Chem. A **2014**, 118, 9772–9782, PMID: 25230384.
- (22) Adjoua, O.; Lagardère, L.; Jolly, L.-H.; Durocher, A.; Very, T.; Dupays, I.; Wang, Z.; Inizan, T. J.; Célerse, F.; Ren, P., et al. Tinker-HP: Accelerating molecular dynamics simulations of large complex systems with advanced point dipole polarizable force fields using GPUs and multi-GPU systems. J. Chem. Theory Comput. **2021**, 17, 2034–2053.
- (23) Lagardère, L.; Aviat, F.; Piquemal, J.-P. Pushing the limits of multiple-time-step strategies for polarizable point dipole molecular dynamics. J. Phys. Chem. Lett. **2019**, 10, 2593–2599.

- (24) Jaffrelot-Inizan, T.; Célerse, F.; Adjoua, O.; El Ahdab, D.; Jolly, L.-H.; Liu, C.; Ren, P.; Montes, M.; Lagarde, N.; Lagardère, L.; Monmarché, P.; Piquemal, J.-P. High-Resolution Mining of SARS-CoV-2 Main Protease Conformational Space: Supercomputer-Driven Unsupervised Adaptive Sampling. Chem. Sci. **2021**, 12, 4889–4907.
- (25) Célerse, F.; Inizan, T. J.; Lagardère, L.; Adjoua, O.; Monmarché, P.; Miao, Y.; Derat, E.; Piquemal, J.-P. An Efficient Gaussian-Accelerated Molecular Dynamics (GaMD) Multilevel Enhanced Sampling Strategy: Application to Polarizable Force Fields Simulations of Large Biological Systems. Journal of Chemical Theory and Computation **2022**, 18, 968–977, PMID: 35080892.
- (26) Célerse, F.; Lagardère, L.; Derat, E.; Piquemal, J.-P. Massively Parallel Implementation of Steered Molecular Dynamics in Tinker-HP: Comparisons of Polarizable and Non-Polarizable Simulations of Realistic Systems. Journal of Chemical Theory and Computation **2019**, 15, 3694–3709, PMID: 31059250.
- (27) Lagardère, L.; Maurin, L.; Adjoua, O.; El Hage, K.; Monmarché, P.; Piquemal, J.-P.; Hénin, J. Lambda-ABF: simplified, portable, accurate, and cost-effective alchemical free-energy computation. J. Chem. Theory Comput. **2024**, 20, 4481–4498, PMID: 38805379.
- (28) Blazhynska, M.; Lagardère, L.; Liu, C.; Adjoua, O.; Ren, P.; Piquemal, J.-P. Water–glycan interactions drive the SARS-CoV-2 spike dynamics: insights into glycan-gate control and camouflage mechanisms. Chem. Sci. **2024**, 15, 14177–14187.
- (29) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J., et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. J. Am. Chem. Soc. **2015**, 137, 2695–2703.
- (30) Zhang, C.-H.; Stone, E. A.; Deshmukh, M.; Ippolito, J. A.; Ghahremanpour, M. M.;

- Tirado-Rives, J.; Spasov, K. A.; Zhang, S.; Takeo, Y.; Kudalkar, S. N., et al. Potent noncovalent inhibitors of the main protease of SARS-CoV-2 from molecular sculpting of the drug perampanel guided by free energy perturbation calculations. ACS Cent. Sci. **2021**, 7, 467–475.
- (31) Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical free energy methods for drug discovery: progress and challenges. Curr. Opin. Struct. Biol. **2011**, 21, 150–160.
- (32) El Hage, K.; Ribaudou, G.; Lagardère, L.; Ongaro, A.; Kahn, P. H.; Demange, L.; Piquemal, J.-P.; Zagotto, G.; Gresh, N. Targeting the major groove of the palindromic d(GGCGCC)₂ sequence by oligopeptide derivatives of anthraquinone intercalators. J. Chem. Inf. Model. **2022**, 62, 6649–6666.
- (33) Gresh, N.; El Hage, K.; Lagardère, L.; Brégier, F.; Godard, J.; Piquemal, J.-P.; Perrée-Fauvet, M.; Sol, V. Enforcing local DNA kinks by sequence-selective trisintercalating oligopeptides of a tricationic porphyrin: a polarizable molecular dynamics study. ChemPhysChem **2024**, 25, e202300776.
- (34) El Hage, K.; Mondal, P.; Meuwly, M. Free energy simulations for protein ligand binding and stability. Mol. Sim. **2018**, 44, 1044–1061.
- (35) Rasouli, A.; Pickard IV, F. C.; Sur, S.; Grossfield, A.; Işık Bennett, M. Essential considerations for free energy calculations of RNA-small molecule complexes: lessons from theophylline-binding RNA aptamer. bioRxiv **2024**, 2024–08.
- (36) Abramyan, A. M.; Bochicchio, A.; Wu, C.; Damm, W.; Langley, D. R.; Shivakumar, D.; Lupyan, D.; Wang, L.; Harder, E.; Oloo, E. O. Accurate physics-based prediction of binding affinities of RNA and DNA targeting ligands. **2024**,
- (37) Clark, F.; Robb, G.; Cole, D. J.; Michel, J. Comparison of receptor–ligand restraint

- schemes for alchemical absolute binding free energy calculations. J. Chem. Theory Comput. **2023**, 19, 3686–3704.
- (38) Salari, R.; Joseph, T.; Lohia, R.; Hénin, J.; Brannigan, G. A streamlined, general approach for computing ligand binding free energies and its application to GPCR-bound cholesterol. J. Chem. Theory Comput. **2018**, 14, 6560–6573, PMID: 30358394.
- (39) Zhang, H.; Gattuso, H.; Dumont, E.; Cai, W.; Monari, A.; Chipot, C.; Dehez, F. Accurate estimation of the standard binding free energy of Netropsin with DNA. Molecules **2018**, 23.
- (40) Gapsys, V.; Yildirim, A.; Aldeghi, M.; Khalak, Y.; Van der Spoel, D.; de Groot, B. L. Accurate absolute free energies for ligand–protein binding based on non-equilibrium approaches. Commun. Chem. **2021**, 4, 61.
- (41) Dibrov, S. M.; Ding, K.; Brunn, N. D.; Parker, M. A.; Bergdahl, B. M.; Wyles, D. L.; Hermann, T. Structure of a hepatitis C virus RNA domain in complex with a translation inhibitor reveals a binding mode reminiscent of riboswitches. Proc. Natl. Acad. Sci. U.S.A. **2012**, 109, 5223–5228.
- (42) Seth, P. P.; Miyaji, A.; Jefferson, E. A.; Sannes-Lowery, K. A.; Osgood, S. A.; Propp, S. S.; Ranken, R.; Massire, C.; Sampath, R.; Ecker, D. J.; Swayze, E. E.; Griffey, R. H. SAR by MS: discovery of a new class of RNA-binding small molecules for the Hepatitis C virus: internal ribosome entry site IIa subdomain. J. Med. Chem. **2005**, 48, 7099–7102, PMID: 16279767.
- (43) Wu, J. C.; Chattree, G.; Ren, P. Automation of AMOEBA polarizable force field parameterization for small molecules. Theor. Chem. Acc. **2012**, 131, 1–11.
- (44) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable atomic multipole-based AMOEBA force field for proteins. J. Chem. Theory Comput. **2013**, 9, 4046–4063.

- (45) Yang, X.; Liu, C.; Kuo, Y.-A.; Yeh, H.-C.; Ren, P. Computational study on the binding of Mango-II RNA aptamer and fluorogen using the polarizable force field AMOEBA. Front. Mol. Biosci. **2022**, 9.
- (46) Lagardère, L.; Jolly, L.-H.; Lipparini, F.; Aviat, F.; Stamm, B.; Jing, Z. F.; Harger, M.; Torabifard, H.; Cisneros, G. A.; Schnieders, M. J., et al. Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. Chem. sci. **2018**, 9, 956–972.
- (47) Jolly, L.-H.; Duran, A.; Lagardère, L.; Ponder, J. W.; Ren, P.; Piquemal, J.-P. Raising the performance of the Tinker-HP molecular modeling package [Article v1.0]. Living J. Comput. Mol. Sci. **2019**, 1, 10409.
- (48) Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. Mol. Phys. **2013**, 111, 3345–3362.
- (49) Bonati, L.; Rizzi, V.; Parrinello, M. Data-driven collective variables for enhanced sampling. J. Phys. Chem. Lett. **2020**, 11, 2998–3004.
- (50) Leontis, N. B.; Stombaugh, J.; Westhof, E. The non-Watson–Crick base pairs and their associated isostericity matrices. Nucleic Acids Res. **2002**, 30, 3497–3531.
- (51) Walker, B.; Liu, C.; Wait, E.; Ren, P. Automation of AMOEBA polarizable force field for small molecules: Poltype 2. J. Comput. Chem. **2022**, 43, 1530–1542.
- (52) Turney, J. M. et al. Psi4: an open-source ab initio electronic structure program. WIREs Comput. Mol. Sci. **2012**, 2, 556–565.
- (53) Stone, A. J.; Alderton, M. Distributed multipole analysis: methods and applications. Mol. Phys. **1985**, 56, 1047–1064.
- (54) Rackers, J. A.; Wang, Z.; Lu, C.; Laury, M. L.; Lagardère, L.; Schnieders, M. J.;

- Piquemal, J.-P.; Ren, P.; Ponder, J. W. Tinker 8: Software tools for molecular design. J. Chem. Theory Comput. **2018**, 14, 5273–5289.
- (55) Ren, P.; Wu, C.; Ponder, J. W. Polarizable Atomic Multipole-Based Molecular Mechanics for Organic Molecules. J. Chem. Theory Comput. **2011**, 7, 3143–3161.
- (56) Zhang, C.; Bell, D.; Harger, M.; Ren, P. Polarizable Multipole-Based Force Field for Aromatic Molecules and Nucleobases. J. Chem. Theory Comput. **2017**, 13, 666–678, PMID: 28030769.
- (57) Bannwarth, C.; Caldeweyher, E.; Ehlert, S.; Hansen, A.; Pracht, P.; Seibert, J.; Spicher, S.; Grimme, S. Extended tight-binding quantum chemistry methods. WIREs Comput. Mol. Sci. **2021**, 11, e1493.
- (58) Bannwarth, C.; Ehlert, S.; Grimme, S. GFN2-xTB—An accurate and broadly parametrized self-consistent tight-binding quantum chemical method with multipole electrostatics and density-dependent dispersion contributions. J. Chem. Theory Comput. **2019**, 15, 1652–1671.
- (59) Chai, J.-D.; Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. Phys. Chem. Chem. Phys. **2008**, 10, 6615–6620.
- (60) Tuckerman, M.; Berne, B. J.; Martyna, G. J. Reversible multiple time scale molecular dynamics. J. Chem. Phys. **1992**, 97, 1990–2001.
- (61) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. J. Chem. Phys. **2007**, 126.
- (62) Berendsen, H. J.; Postma, J. v.; Van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. J. Chem. Phys. **1984**, 81, 3684–3690.

- (63) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. J. Chem. Phys. **1995**, 103, 8577–8593.
- (64) Lagardère, L.; Lipparini, F.; Polack, E.; Stamm, B.; Cancès, E.; Schnieders, M.; Ren, P.; Maday, Y.; Piquemal, J.-P. Scalable evaluation of polarization energy and associated forces in polarizable molecular dynamics: II. Toward massively parallel computations using smooth particle mesh Ewald. Journal of Chemical Theory and Computation **2015**, 11, 2589–2599, PMID: 26575557.
- (65) Laury, M. L.; Wang, Z.; Gordon, A. S.; Ponder, J. W. Absolute binding free energies for the SAMPL6 cucurbit[8]uril host–guest challenge via the AMOEBA polarizable force field. J. Comput.-Aided Mol. Des. **2018**, 32, 1087–1095.
- (66) Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: a quantitative approach for their calculation. J. Phys. Chem. B **2003**, 107, 9535–9551.
- (67) Hénin, J.; Lopes, L. J.; Fiorin, G. Human learning for molecular simulations: the Collective Variables Dashboard in VMD. J. Chem. Theory Comput. **2022**, 18, 1945–1956.
- (68) Humphrey, W.; Dalke, A.; Schulten, K. VMD: visual molecular dynamics. J. Mol. Graph. **1996**, 14, 33–38.
- (69) Straatsma, T. P.; McCammon, J. A. Multiconfiguration Thermodynamic Integration. J. Chem. Phys. **1991**, 95, 1175–1188.
- (70) Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Non-polar gases. J. Chem. Phys. **1954**, 22, 1420–1426.
- (71) Jiang, W.; Roux, B. Free energy perturbation Hamiltonian replica-exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. J. Chem. Theory Comput. **2010**, 6, 2559–2565.

- (72) Lyubartsev, A.; Martsinovski, A.; Shevkunov, S.; Vorontsov-Velyaminov, P. New approach to Monte Carlo calculation of the free energy: Method of expanded ensembles. J. Chem. Phys. **1992**, 96, 1776–1783.
- (73) Santiago-McRae, E.; Ebrahimi, M.; Sandberg, J. W.; Brannigan, G.; Hénin, J. Computing absolute binding affinities by streamlined alchemical free energy perturbation (safep)[article v1. 0]. Living J. Comput. Mol. Sci. **2023**, 5, 2067–2067.
- (74) Thomas, J. R.; Hergenrother, P. J. Targeting RNA with small molecules. Chem. Rev. **2008**, 108, 1171–1224.
- (75) Far, S.; Kossanyi, A.; Verchère-Béaur, C.; Gresh, N.; Taillandier, E.; Perrée-Fauvet, M. Bis- and Tris-DNA intercalating porphyrins designed to target the major groove: synthesis of acridylbis-arginyl-porphyrins, molecular modelling of their DNA complexes, and experimental tests. Eur. J. Org. Chem. **2004**, 2004, 1781–1797.
- (76) Petrov, D.; Perthold, J. W.; Oostenbrink, C.; de Groot, B. L.; Gapsys, V. Guidelines for free-energy calculations involving charge changes. J. Chem. Theory Comput. **2024**, 20, 914–925.
- (77) Invernizzi, M.; Parrinello, M. Exploration vs convergence speed in adaptive-bias enhanced sampling. J. Chem. Theory Comput. **2022**, 18, 3988–3996.
- (78) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. WIREs Comput. Mol. Sci. **2011**, 1, 826–843.
- (79) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potentials to Probability Distributions. J. Phys. Chem. Lett. **2020**, 11, 2731–2736.
- (80) Welling, M. Fisher Linear Discriminant Analysis; 2005.
- (81) Rizzi, V.; Bonati, L.; Ansari, N.; Parrinello, M. The role of water in host-guest interaction. Nat. Commun. **2021**, 12, 93.

- (82) Ansari, N.; Rizzi, V.; Carloni, P.; Parrinello, M. Water-Triggered, Irreversible Conformational Change of SARS-CoV-2 Main Protease on Passing from the Solid State to Aqueous Solution. J. Am. Chem. Soc. **2021**, 143, 12930–12934.
- (83) Ansari, N.; Rizzi, V.; Parrinello, M. Water regulates the residence time of Benzamidine in Trypsin. Nat. Commun. **2022**, 13, 5438.
- (84) Bjelobrk, Z.; Piaggi, P. M.; Weber, T.; Karmakar, T.; Mazzotti, M.; Parrinello, M. Naphthalene crystal shape prediction from molecular dynamics simulations. Cryst. Eng. Comm. **2019**, 21, 3280–3288.
- (85) Padroni, G.; Patwardhan, N.; Schapira, M.; Hargrove, A. Systematic analysis of the interactions driving small molecule–RNA recognition. RSC Med. Chem. **2020**, 11, 802–813.
- (86) Chen, W.; Cui, D.; Jerome, S. V.; Michino, M.; Lenselink, E. B.; Huggins, D. J.; Beautrait, A.; Vendome, J.; Abel, R.; Friesner, R. A., et al. Enhancing hit discovery in virtual screening through absolute protein–ligand binding free-energy calculations. J. Chem. Inf. Model. **2023**, 63, 3171–3185.
- (87) Parsons, J.; Castaldi, M. P.; Dutta, S.; Dibrov, S. M.; Wyles, D. L.; Hermann, T. Conformational inhibition of the hepatitis C virus internal ribosome entry site RNA. Nat. Chem. Biol. **2009**, 5, 823–825.
- (88) Dibrov, S. M.; Parsons, J.; Carnevali, M.; Zhou, S.; Ryneerson, K. D.; Ding, K.; Garcia Sega, E.; Brunn, N. D.; Boerneke, M. A.; Castaldi, M. P., et al. Hepatitis C virus translation inhibitors targeting the internal ribosomal entry site: Miniperspective. J. Med. Chem. **2014**, 57, 1694–1707.