



Published in final edited form as:

J Chem Theory Comput. 2022 September 13; 18(9): 5672–5691. doi:10.1021/acs.jctc.2c00381.

SILCS-RNA: Toward a Structure-Based Drug Design Approach for Targeting RNAs with Small Molecules

Abhishek A. Kognole,

Anthony Hazel,

Alexander D. MacKerell Jr.

Computer Aided Drug Design Center, Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland Baltimore, Baltimore, Maryland 21201, United States

Abstract

RNA molecules can act as potential drug targets in different diseases as their dysregulated expression or misfolding can alter various cellular processes. Noncoding RNAs account for ~70% of the human genome and these molecules can have complex tertiary structures that present a great opportunity for targeting by small molecules. In the present study the site identification by ligand competitive saturation (SILCS) computational approach is extended to target RNA, termed SILCS-RNA. Extensions to the method include an enhanced oscillating excess chemical potential protocol for the grand canonical Monte Carlo calculations and individual simulations of the neutral and charged solutes from which the SILCS functional group affinity maps (FragMaps)

Corresponding Author: Alexander D. MacKerell Jr. *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland Baltimore, 20 Penn Street, Baltimore, Maryland 21201, United States.* alex@outerbanks.umaryland.edu.

Author Contributions

A.A.K. contributed to conceptualization, target identification and ligand database curation, code development, running simulations, analysis of results, and writing of the article. A.H. contributed to development of GCMC code. A.D.M. supervised the entire research including conceptualization, data analysis, and writing of the article.

Abhishek A. Kognole *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland Baltimore, 20 Penn Street, Baltimore, Maryland 21201, United States.*

Anthony Hazel *Department of Pharmaceutical Sciences, School of Pharmacy, University of Maryland Baltimore, 20 Penn Street, Baltimore, Maryland 21201, United States.*

Conflict of Interest

ADM Jr. is co-founder and CSO of SilcsBio LLC.

Supporting Information

The Supporting Information is available free of charge at <https://doi.org/10.1021/acs.jctc.2c00381>

si_001.pdf: Table S1. Description of programs available for RNA-ligand docking; Figure S1. Illustration of the active GCMC sub volume; Figure S2. Binding sites of FMN, TPP, THF and GUA; Figure S3. Binding sites of TAR, HCV and IVP; Table S2. Overlap coefficient (OC) each set of SILCS FragMaps; Table S3. Schemes for oscillating excess chemical potential; Table S4. LGFE scores from SILCS-MC Pose Refinement for the FMN RNA; Table S5. LGFE scores from SILCS-MC Pose Refinement for HCV RNA; Figure S4. Oscillating excess chemical potential and MAMY concentration in an example SILCS simulation; Figure S5. SILCS FragMaps for TPP, THF, GUA, TAR and IVP RNAs; Figure S6. Methylammonium nitrogen FragMaps around the FMN RNA; Table S6. Average LGFE scores of ligands from SILCS-MC; Figure S7. Phosphate containing ligands; Table S7. SILCS-MC results for phosphate containing ligands; Figure S8. Pocket identified by HS#14 in GUA; Figure S9. Different structures of TAR with small molecule interactions.

Additional Supporting information files:

- **si_002.xlsx:** The database for all seven RNA ligands with molecule details and references.
- **si_003.xlsx:** The results for SILCS-MC calculations for all the ligands for all RNAs using different metrics.
- **si_004.xlsx:** The results from SILCS-MC docking calculations in selected hotspots for FDA compounds.
- **si_005.pdf:** 2D structure images for all 104 ligands in the silcs-rna-database.

are calculated for subsequent binding site identification and docking calculations. The method is developed and evaluated against seven RNA targets and their reported small molecule ligands. SILCS-RNA provides a detailed characterization of the functional group affinity pattern in the small molecule binding sites, recapitulating the types of functional groups present in the ligands. The developed method is also shown to be useful for identification of new potential binding sites and identifying ligand moieties that contribute to binding, granular information that can facilitate ligand design. However, limitations in the method are evident including the ability to map the regions of binding sites occupied by ligand phosphate moieties and to fully account for the wide range of conformational heterogeneity in RNA associated with binding of different small molecules, emphasizing inherent challenges associated with applying computer-aided drug design methods to RNA. While limitations are present, the current study indicates how the SILCS-RNA approach may enhance drug discovery efforts targeting RNAs with small molecules.

Keywords

Computer-Aided Drug Design; Docking; Molecular Dynamics Simulations; Site-identification by ligand competitive saturation; ribonucleic acid; ions; HIV

Introduction

The last two decades have seen revolutionary advances in optimization of the drug discovery process and computer-aided drug design (CADD) approaches have contributed significantly in accelerating it.¹ While the majority of the focus has been on protein targets, there has been increasing interest in exploring RNA molecules as potential therapeutic targets.^{2–4} Due to their structural complexity and polyanionic nature, RNA molecules have been deemed highly challenging drug targets as compared to proteins, and hence they have received limited attention.^{5, 6} Recent advances in technologies to analyze small molecule binding to RNA targets have changed this perspective, generating drug-discovery efforts in this field.^{7, 8} Specifically, there is growing evidence that small molecules can be used to inhibit or alter the function of RNA molecules that exhibit well-defined binding sites for such small molecules to bind.^{9, 10}

Multiple ligand docking methods have been developed to predict RNA–ligand interactions (Table S1). Leclarc and Cedergren were one of the first to model RNA–ligand interactions using a 3D-SAR-based approach coupled with MD calculations to optimize small molecule poses.¹¹ RiboDock¹² implemented pseudoenergy functions to describe RNA–ligand interactions. It was further developed as rDock.¹³ DrugScore^{RNA}¹⁴ used distance dependent potentials trained with structural data; however, including NMR data for training did not improve the results. MORDOR¹⁵ was the first method to allow flexibility in both the RNA and ligands, where RNA and ligands were modeled using CHARMM27¹⁶ and AMBER,¹⁷ respectively. It used an induced-fit approach with RMSD restrained energy minimization. Further, small molecule–RNA docking was performed in the context of DOCK 6.¹⁸ In that study a comprehensive analysis of the anchor-and-grow approach was performed suggesting that the success rate drops suddenly beyond 12 rotatable bonds for a ligand. The Bujnicki group first developed a method in conjunction with DOCK 6 that

builds a model using a grid-based algorithm where they introduced a knowledge-based distance- and angle-dependent potential by defining two RNA atoms instead of one as ligand-interacting moieties, termed LigandRNA, with the model trained using 251 ligand–RNA structures including NMR data.¹⁹ Recently, that same group published AnnapuRNA which has further increased the accuracy and speed of predicting the binding poses.²⁰ RLDock is another recent program that implements a versatile scoring function that includes contributions such as changes in solvent accessible surface area and polarization.²¹ While these computational tools provide promising results, none of them have utilized the full potential of molecular dynamic simulations that can include the conformational flexibility of nucleic acids. Also, the majority of these approaches focus on docking in rigid experimental structures of the ligand–RNA complexes and lack the ability to provide relative binding affinities that can be compared with experiments.

The site identification by ligand competitive saturation (SILCS) technology is a unique platform for structure-based drug design for protein targets.^{22–24} The SILCS approach involves a series of grand canonical Monte Carlo-molecular dynamics (GCMC-MD) simulations of a target macromolecule in the presence of explicit water and solutes representing various functional groups and physiochemical properties.²³ Probability distribution analysis of selected atoms of the solutes in the full 3D region encompassing the target is then used to define Grid Free Energy (GFE) FragMaps specific to the target studied. Those GFE FragMaps may be used for binding site identification, database screening, ligand docking, and binding affinity evaluation of small molecules of potential therapeutic interest. Notably, the 3D GFE FragMaps represent qualitative characterization that can facilitate ligand design and the approach gives quantitative information on the contribution of individual atoms to ligand binding that is of further utility to the ligand design process. The SILCS method was initially presented in 2009 and has since been extended in the context of various aspects of CADD.^{25, 26} In the present study, the SILCS technology is extended to RNA molecules, termed SILCS-RNA, using data from seven RNA targets on which sufficient, though limited, experimental data is available to challenge the computational approach.

Methods

The SILCS approach was applied to seven RNA systems on which small molecule–RNA interaction data is available. Those systems are listed in Table 1, including all of the PDB structures used in this study. The SILCS protocol is comprised of three major stages: (1) *SILCS* grand canonical Monte Carlo-molecular dynamics (*GCMC-MD*) simulations, (2) *SILCS GFE FragMap* calculations and (3) *SILCS-Monte Carlo (MC)*, including *SILCS-Hotspots*, as shown in the workflow in Figure 1. The SILCS protocol as applied to proteins has been previously described in detail.²³ The basic overall approach is maintained in the present study with variations for use with RNA as described below.

RNA Preparation

A 3D structure of the macromolecule is required to start the SILCS calculations. With multiple PDB files available for each RNA as reported in Table 1, a best suited PDB

was selected based on availability, resolution and presence of bound ligand coordinates of the most relevant compound, such as the physiological or highest affinity compounds. These specific PDBs are shown in bold in Table 1. Except for 2LWK for IVP being a NMR structure, all other selected structures were determined using X-ray crystallography. For TAR, we chose 6CMN, which is an X-ray crystal structure of HIV-1 TAR with its RNA binding protein over other NMR structures that have small molecules bound given the high conformational variability of the RNA in NMR structures. The structures were stripped of bound ligands, structural waters, ions and other cosolvent molecules and a single asymmetric unit was prepared for each RNA using the *CHARMM-GUI*.⁶⁷ Terminal modified residues like GTP and CCC were removed from both PDBs 2GDI and 3SKI for TPP and GUA, respectively. All available PDB files were parsed for crystal Mg^{2+} ion positions. After alignment of all PDBs to the selected PDB, Mg^{2+} ions were retained on the RNA at important positions only for the SILCS simulations with charged solutes; SILCS calculations with neutral solutes did not have Mg^{2+} ions in the structure. In addition, SILCS simulations of the charged solutes were performed both without any additional ions and with crystal Mg^{2+} and neutralizing Na^+ .

SILCS GCMC-MD Simulations

SILCS simulations are performed using an oscillating excess chemical potential, μ_{ex} , GCMC approach²² to facilitate solute and water sampling in conjunction with MD simulations that account for target RNA flexibility as well as further enhance the water and solute sampling. The simulations involve 10 independent runs of SILCS GCMC-MD on each system. The initial simulations using the standard protocol are performed in the presence of explicit solvent with a collection of the following eight solutes at ~0.25 M: benzene (BENX), propane (PRPX), dimethylether (DMEE), methanol (MEOH), formamide (FORM), imidazole (IMIA), acetate (ACEY) and methylammonium (MAMY). In addition, to address the polyanionic nature of RNA molecules, two independent sets of 10 GCMC-MD simulations were run splitting the solutes into neutral and charged classes. The first 10 independent runs have the neutral solutes benzene (BENX), propane (PRPX), dimethylether (DMEE), methanol (MEOH), formamide (FORM) and imidazole (IMIA). The second 10 independent runs have the charged solutes acetate (ACEY), and methylammonium (MAMY). Previous studies have shown the hydration free energies of ACEY and MAMY to be in excellent agreement with experimental values.²² Water molecules are present in both cases at ~55 M with the solutes initially present at ~0.25 M.

Figure 1 shows the process flow diagram for SILCS simulations including the two independent sets of GCMC-MD simulations. GROMACS⁶⁸ is initially used to set up the systems where a PDB file for the RNA molecule is processed to generate the topology according to the CHARMM36 nucleic acid force field.^{69, 70} The RNA is overlaid with a simulation water box of dimensions determined based on a minimum of 15 Å from the RNA non-hydrogen atoms and includes the solutes at ~0.25 M. The systems are then subjected to a GCMC refilling where all of the solutes and water are deleted from GCMC active region. The GCMC active region is defined as the subvolume of the simulation box with a 15 Å margin from the edge of the box, such that it is defined at the maximum dimensions of the RNA (Figure S1). GCMC simulations are then used to insert the solutes and water around

the target RNA in a three-stage or two-stage process. In the revised protocol in Figure 1 for the neutral solutes, a three-stage refilling approach is used where, in the first stage, the solutes BENX, PRPX, and IMIA are inserted to a concentration of 0.35 M. This facilitates the sampling of buried binding pockets and intercalation sites by the larger solutes. This is followed by insertion of the other solutes in the second stage, FORM, MEOH, and DMEE to 0.25 M, and then, water molecules are inserted to 55 M in stage 3. For the charged solute SILCS simulations, the two-stage refilling process involves insertion of the solutes ACEY and MAMY in stage 1 and water in the second stage. The solute concentrations are calculated based on the number of solute molecules relative to the number of water molecules assuming a concentration of 55 M in the full simulation box. In every stage of the refilling process, the GCMC simulation is typically carried out for 200,000 steps. If desired concentrations of probes are not achieved by then, GCMC refilling simulations are continued with additional cycles of 200,000 steps. Following completion of the refilling process each system is subjected to a 5000-step steepest descent (SD) minimization and MD equilibration for 100 ps. Production involves 100 cycles of SILCS GCMC-MD on the 10 independent systems where one cycle consists of 200,000 steps of GCMC of all of the solutes and water followed by a 5000-step SD minimization, MD equilibration for 100 ps, and 1 ns of production MD simulation. Weak harmonic restraints with a force constant of 0.12 kcal/mol/Å² were applied during production MD on the C1' atoms of the sugar and N1/N3 atoms of the bases unless otherwise noted. In total, 1 μs of MD sampling is obtained over the 10 independent SILCS simulations for each neutral and charged system. The GCMC-MD simulations²² were performed using version 2020.2 (SilcsBio LLC) for the FMN and HCV systems during the testing of the standard SILCS protocol. In those initial tests, dimethylether (DMEE) was replaced with acetaldehyde (AALD). In the remaining SILCS-RNA calculations, the GCMC simulations were performed using in-house developed code (access described in the Supporting Information) and MD simulations were performed using GROMACS 2020.4.⁶⁸ The RNA molecules are simulated using the CHARMM36 force field⁶⁹, the solutes and ligands are treated with the CHARMM general force field (CGenFF)⁷¹ and the CHARMM-TIP3P model is used for water molecules^{72, 73}.

SILCS FragMaps

The 10 × 100 ns trajectories from the GCMC-MD simulations for the neutral and charged solutes sets were used to calculate the probability distribution of selected solute atoms that represent various functional groups (Figure 2). The probabilities were calculated in 1 Å × 1 Å × 1 Å cubic voxels encompassing the entire simulation system and then normalized with respect to number of frames from the MD simulations, total number of atoms of each respective atom type (n_{atoms}) and bulk occupancy (OCC_{bulk}) calculated based on average concentration of the fragment relative to the number of waters (equation 1). For example, n_{atoms} for benzene carbon atom (BENC) is 6. Normalized probability (P_{xyz}) is then converted to Grid Free Energies (GFE) based on a Boltzmann transformation using equation 2.

$$P_{xyz} = OCC_{xyz} / (n_{atoms} OCC_{bulk}) \quad \text{Eq. 1}$$

$$GFE_{xyz} = -RT \ln(P_{xyz}) \quad \text{Eq. 2}$$

The FragMaps are further evaluated to verify the convergence by calculating an overlap coefficient⁷⁴ between the first five (1–5) and last five (6–10) independent runs. Table S2 shows that high overlap coefficients (> 0.78) are obtained for all of the solute atoms, indicating that the simulations are well converged. For each type of solute atom representative of a specific functional group, the corresponding GFE_{xyz} value at each voxel is used to estimate ligand binding affinities, as described below and to define the contour levels for visualization of the FragMaps. For all subsequent calculations, the FragMaps for the charged (MAMY and ACEY) solutes are merged with the FragMaps of the neutral solutes. For water, the FragMaps from the neutral set are carried forward. Once the FragMaps are generated, the simulation data from GCMC-MD can be compressed and archived, as further calculations are solely based on the SILCS FragMaps along with the SILCS exclusion maps. The exclusion maps are defined as the regions of the RNAs in which there is no sampling by the water or solute non-hydrogen atoms based on the neutral solute simulation system.

SILCS-Monte Carlo

Docking of small molecules applies a Monte Carlo (SILCS-MC) sampling approach in the field of the SILCS FragMaps and the exclusion map. This involves each ligand molecule being processed through the CGenFF program that assigns a CGenFF atom type to each atom. The CGenFF atom types are then converted to types that correspond to a SILCS FragMap type based on an atom classification scheme (ACS), as previously described.²³ Here we used two ACSs recently described as Generic Apolar Scale (GAS21) and Specific Standard (SS21).²⁴ In this mapping, hydrogens are not considered, consistent with them not being included in the FragMap functional group definitions (Figure 2), and certain non-hydrogen atoms are nonclassified (NCLA type) and therefore are not considered in the GFE calculations. Once SILCS types are assigned, the ligand is overlaid on the FragMaps and the GFE score for the respective voxel corresponding to the position of each FragMap is assigned to that atom. The overall binding affinity is then approximated by summing over all of the atomic GFE contributions, yielding the ligand GFE (LGFE) score. The LGFE score along with the CGenFF intramolecular energy are then used as the Metropolis criteria for MC sampling of the translation, rotational, and dihedral degrees of freedom of the ligand. Two different types of SILCS-MC algorithms can be used, namely, *Pose Refinement* and *Docking* (previously²³ referred to as *local* and *exhaustive*, respectively), which differ based on the starting pose of the ligand, maximum step sizes for translation, rotations and dihedral angle sampling, and the number of MC steps to perform. In both cases, there is an initial MC relaxation at 300 K followed by simulated annealing (SA), with the details of the two protocols shown in Table 2. Pose refinement is initiated from known binding poses of ligands and allows a restricted sampling of the ligand in the binding site. This approach is useful to compare the predicted LGFE with the experimental binding affinity for ligands with structural data such as crystal or NMR structures. The docking algorithm performs a highly extensive sampling of ligand poses in the binding site to identify the best possible binding positions through initiation of the docking based on a random starting orientation of

the ligand in a sphere defining the binding site followed by additional MC sampling. In both cases following the initial MC sampling, simulated annealing is performed with smaller step sizes and the final estimates of the binding affinity are based solely on the LGFE score.

SILCS-Hotspots

The SILCS FragMaps can also be used to identify new potential ligand binding sites thorough SILCS-Hotspots analysis⁷⁵. In this approach, a set of fragments are exhaustively docked in the entire region occupied by the RNA using SILCS-MC docking. This involves partitioning each RNA into a set of overlapping regions and docking each fragment into each region. The docking includes random initial ligand placement within each region with $n_{CY}=1000$, $n_{MC}=10,000$, $dX=1$, $d\theta=180$, $d\phi=180$, $n_{SA}=40,000$, $dX_{SA}=0.2$, $d\theta_{SA}=9$, and $d\phi_{SA}=9$. Thousands of docked poses are generated for each fragment and then analyzed through two rounds of root-mean-square difference (RMSD) clustering. The first round involves clustering of each fragment type to identify local binding sites for each fragment, and the second round is clustering over all of the fragment types to identify binding sites occupied by one or more fragments. Each resulting hotspot may then be ranked based on a number of criteria, such as the average LGFE values over all fragments in each site or the number of fragments occupying a site. First, we used the ASTEX mini fragment database⁷⁶ to generate the hotspots for all seven RNA molecules followed by RMSD clustering using a radius of 3 Å for pose clustering and 6 Å for site clustering to get the hotspots. The hotspots were ranked based on average LGFE score and analyzed visually to identify new binding pockets and regions of potential interest on all seven RNA targets. These selected sites were further subjected to SILCS-MC *docking* (sphere radius 2 Å with GAS21) of ~380 FDA approved drug molecules to find the top 20 molecules with the most favorable LGFE. The docked poses of the top 20 molecules were then analyzed for relative change in solvent accessible surface area (rSASA). The rSASA is defined as the percent SASA of the ligand in the absence of the RNA molecules versus that in the presence of the RNA molecules such that rSASA of 100% would suggest that the ligand is fully hidden and is not accessible to solvent.

Results and Discussion

In the present study, the SILCS method is applied to seven diverse RNA targets. The RNA molecules were selected based on the different nature of their ligand binding sites as well as the availability of multiple ligands and experimental structures for each RNA, as summarized in Table 1. These include flavin mononucleotide riboswitch (FMN), thiamine pyrophosphate riboswitch (TPP), tetrahydrofolate riboswitch (THF), 2'-deoxyguanosine riboswitch (GUA), HIV-1 trans-activation response element (TAR), hepatitis C virus internal ribosomal entry site IIa element (HCV) and influenza A virus RNA promoter (IVP). The FMN riboswitch is an RNA element of prokaryotic mRNAs that plays an important role in flavin mononucleotide (FMN) biosynthesis and transport proteins.⁷⁷ The TPP riboswitch is a riboswitch domain that responds to the coenzyme thiamine pyro-phosphate, an active form of vitamin B1 which is an essential participant in many protein-catalyzed reactions.⁷⁸ A few of the ligands for the TPP riboswitch exhibit a positive charge. The THF riboswitch cooperatively binds to two molecules of THF to regulate the transport and metabolism of

folate in a number of bacteria. THF and its derivatives are essential in metabolic processes. The dG Riboswitch is a purine riboswitch. These riboswitches play important roles in genetic regulation of bacterial metabolism, with the dG riboswitch playing a critical role in the biosynthesis of deoxyguanosine. The HIV-1 TAR (trans-activation response) RNA element is attractive as a drug target because it exists in the 5'-noncoding regions of all viral mRNAs and resists mutations to maintain interactions with Tat (trans-activator of transcription) protein.^{79, 80} HCV-IRES-IIa is a small, enveloped, positive-stranded RNA virus from the *Flaviviridae* family. The 5'-UTR contains an internal ribosome entry site (IRES), which mediates the initiation of viral-RNA translation in a cap-independent manner. Stem II of the HCV-IRES is known to be important for IRES-dependent translation, as well as for HCV RNA replication.⁸¹ The influenza A virus is a member of the *Orthomyxoviridae* family, and its genome consists of eight single-stranded RNA-negative segments⁸². The influenza A virus RNA promoter is comprised of 13 nucleotides on the 5' terminus and 12 nucleotides on the 3' terminus of each viral RNA segment conserved throughout various human influenza A virus strains that combine to form the 3D structure of the promoter. Mutations in these conserved sequences negatively affect viral replication efficiency.

According to Warner et. al.⁶, the selected RNAs can be divided into two general groups based on the information content in the RNA structure and quality of the binding pocket. The four riboswitches, which may be considered high information content, exhibit complex 3D structures with multihelix junctions and large and buried binding pockets (Figure 3A–D and S2). For these RNAs, there are multiple crystal structures with bound ligands available. The FMN riboswitch, TPP riboswitch, and dG riboswitch binding sites are mostly solvent inaccessible, while two separate partially buried binding sites are present in the THF riboswitch. The other three RNAs are smaller, being dominated by secondary, duplex structures with limited tertiary contacts, leading them to be more flexible in nature, thereby sampling a wider range of conformation (Figure 3E–G), rendering low information content and poor pocket quality. The binding sites on these RNAs are usually small or shallow and can assume different local conformations based on the bound ligand. The binding site architectures of two of the RNAs, TPP riboswitch (Figure S3B) and HCV-IRES-IIa (Figure S3C), involve intercalation of the ligand. Based on available RNA–ligand databases^{83–86} and other available literature,^{37, 38, 40, 42, 43} we have curated a database of ligands targeting these RNAs. The Supporting Information file ct2c00381_si_002.xlsx provides the details of the small molecules, while the ct2c00381_si_005.pdf provides their 2D structures.

Standard SILCS Simulations Applied to RNA

Initial efforts involved the use of the standard SILCS simulation method, as developed for proteins and applied in numerous studies,^{87, 88} to the FMN and HCV systems. Given the polyanionic nature of the RNA, initial analysis focused on the number of positively charged MAMY ions in the system. As presented in Figure 4, where the initial number of ions in the systems corresponds to ~0.25 M, there is a sharp increase in the number of MAMY ions. Beyond the RNA–MAMY electrostatic contributions favoring insertions of the positive ion, the increase in concentration is magnified by the GCMC active region being smaller than the full simulation box and calculation of the solute concentration for adjustment of the oscillating chemical potential based on the GCMC active region. Accordingly, as

MAMY solutes are added to the active region, they subsequently diffuse throughout the entire simulation box during the MD simulations, keeping the concentration of the MAMY in the GCMC region relatively low. This leads to a delayed response of the chemical potential which is designed to shift to more favorable values in our oscillating chemical potential GCMC approach²³ as the number of each solute type in the system increases. As the chemical potential has not adequately compensated for the increase in MAMY into the active region, additional ions are inserted before saturation after approximately 40 GCMC-MD production-run cycles, with the final concentration of MAMY approaching ~0.8–1.2 M in the two systems. Most importantly, this influx of MAMY leads to the ion dominating the FragMaps surrounding the RNA and significantly decreasing sampling of the other solutes around the RNA that is reflected in the resulting FragMaps as discussed below.

As is well-known, the ion–atmosphere is critical in dictating the structure and dynamics of RNA molecules, the stability of their tertiary structures, and a range of structure–function relationships.^{89–90, 91} The ion atmosphere, as described by counterion condensation theory⁹², stabilizes the secondary structure of RNA and DNA while divalent ions, such as Mg^{2+} , interact specifically with RNA to mediate folding of the tertiary structure and are involved in catalytic mechanisms of ribozymes.^{93–96} Accordingly, 0.1 M NaCl was added to the SILCS simulations in addition to the standard solutes, though the NaCl was not subjected to GCMC sampling. The addition of 0.1 M NaCl buffer significantly reduces the total number of MAMY solutes added to each system during the production GCMC, although the final concentration is about twice that of the target concentration of 0.25 M (Figure 4). As a next step, the Mg^{2+} ions identified in the crystal structures of the FMN and HCV RNAs were included in the SILCS simulations in addition to the 0.1 M NaCl. The addition of the crystallographic Mg^{2+} ions (eight for FMN, six for HCV) led to a further decrease in the number of MAMY ions being added to the system, although the final concentrations still approached a range twice that of the 0.25 M target value. The NaCl and Mg^{2+} ions were not a part of GCMC sampling during these runs with the divalent cations restrained to their crystallographic binding sites. These results indicated the need for additional adjustment to the SILCS simulation protocol to account for the polyanionic nature of RNA.

Based on the excess insertions of the MAMY ions, it was evident that more robust control of the ion atmosphere was needed to maintain the concentration of MAMY in the SILCS simulations as well as give the neutral probe molecules a better opportunity to sample the critical pockets and surfaces of RNA molecules. Accordingly, the use of two separate SILCS simulations for the neutral and charged solutes was introduced (Figure 1), with the charged solute SILCS simulation only including 0.25 M acetate and methylammonium. In addition, to introduce improved control over the MAMY concentration as well as that of the other solutes, an enhanced μ_{ex} oscillation protocol was introduced in conjunction with the implementation of previously presented cavity biases^{22, 23} and configurational biases^{97, 98} in the GCMC methodology. Presented in Table S3 are details of the standard and enhanced μ_{ex} oscillation protocols with an example comparison in Figure S4. The new protocol significantly increases the change in μ_{ex} when the actual concentration of the individual solutes deviates from the target concentration. Notably, this leads to the concentration staying close to the target value of 0.25 M throughout the individual SILCS simulations. Applying this approach along with the inclusion of cavity and configurational biases into

the GCMC protocol leads to significant improvement in control of the concentration of the MAMY ions, as shown in Figure 4. Convergence of the SILCS FragMaps based on the overlap metric⁹⁹ is high with all overlap coefficients > 0.78, as shown in Table S2.

To evaluate the impact of the FragMaps obtained from the original SILCS and variations of the SILCS RNA protocols described above on LGFE scores, we performed SILCS-MC *pose refinement* calculations which dock the small molecules in the known binding site on the macromolecule. Pose refinement uses known binding poses of ligands and only performs a short and limited MC sampling. Table S4 shows the predicted LGFE values for various ligands for FMN with the results for HCV shown in Table S5. Applying the standard SILCS simulation protocol yields LGFE scores that are weakly favorable with FMN and unfavorable with HCV, results that are not consistent with the known binding affinities of these ligands being in the micromolar to nanomolar range (ct2c00381_si_002.xlsx). The addition of 0.1 M NaCl to the SILCS simulations leads to slightly more favorable LGFE scores in the FMN systems compared to no ions in the system, although the HCV scores become even more unfavorable. Inclusion of crystal Mg²⁺ ions along with 0.1 M NaCl leads to the average scores becoming slightly more favorable in the FMN system with the HCV scores still unfavorable. However, upon going to the new protocol with separate neutral and charged solute SILCS simulations, the LGFE scores for both the FMN and HCV systems become significantly more favorable with average values in the vicinity of -6 kcal/mol. While SILCS LGFE scores do not directly correspond to experimental binding affinities, as previously discussed^{23, 24}, for ligand-protein systems, they are typically within 2-3 kcal/mol of the experimental values. Accordingly, it was anticipated that the SILCS-RNA approach would yield similar results, as obtained using the new simulation protocol. Based on this improved convergence of the MAMY ion concentration and improved LGFE scores, the new protocol was applied throughout the remainder of the study, termed as SILCS-RNA.

SILCS-RNA FragMaps

The developed SILCS-RNA simulation protocol was applied to the remainder of the systems. Based on these simulations, analysis involved the spatial distribution of the GFE FragMaps in the various types of binding sites across the seven RNA systems. The nucleic acids, DNA and RNA, have been reported to interact with small molecules in multiple ways including covalent binders, major and minor groove binders, intercalators, multimode binders, quadruplex binders, and bulge/junction/phosphate backbone binders.¹⁰⁰ The seven RNAs studied here have different types of binding sites, and FragMaps acquired from the SILCS-RNA simulations need to capture the characteristics of the functional group distributions over the entire RNA as well as the local binding sites. Figure 5 shows the distribution of FragMaps around the FMN and HCV RNAs, with the FragMaps for the remaining RNAs shown in Figure S5. The overall FragMap distributions show a variety of different types of FragMaps distributed throughout the RNAs including apolar, hydrogen bond donor, and hydrogen bond acceptor FragMaps. This includes multiple pockets or buried regions where small molecules could potentially bind associated with the complex combination of the secondary and tertiary structural features of the RNAs. However, it should be noted that the SILCS FragMaps are still dominated by the positively charged methyl ammonium probes, as shown in Figure S6 where the MAMY FragMaps are shown

at -0.5 kcal/mol contour level, the same level as the FragMaps shown in Figure 5 with the exception of the MAMY maps at -2.0 kcal/mol. This further indicates the importance of the dual SILCS neutral and charged solute simulation protocol to obtain sampling of all classes of the SILCS solutes throughout the RNAs. Concerning anionic FragMaps, there were not any significant favorable GFE maps for acetate ions close to the RNA surface. Although this agrees with counterion condensation theory, one would anticipate sampling of negatively charged probes in the vicinity of the Mg^{2+} ions to represent the negatively charged moieties such as those found in the subset phosphate containing ligands present in the seven systems, as discussed below.

Detailed analysis of the FragMaps on the seven systems reveals favorable apolar FragMaps in the buried regions in the tertiary structures of the RNA as well as near the surface. The expected positive MAMY FragMaps are widely distributed throughout the RNAs even when viewed at a favorable GFE value of -2.0 kcal/mol versus -0.5 kcal/mol for the remaining FragMaps. Apolar maps are observed in the known ligand binding sites as well as around solvent exposed nucleobases, a feature common near sharp twist/turns in the backbone (Figure 5B and Figure S4), though such surface exposed regions are not generally of utility for ligand design. Around the helical regions, the polar, neutral FragMaps are more likely to occupy the minor grooves over the major grooves. Favorable generic donor and generic acceptor maps are also present in the grooves as well as in the ligand binding sites for all of the RNAs. As shown below, the FragMaps in the binding sites recapitulate the functional groups on the ligands binding to the different RNAs.

In addition to the FragMaps, an exclusion map is obtained from the SILCS simulations for each macromolecule. The exclusion map represents the grid points that were not accessible to solutes or water during SILCS simulations. These are represented as a sand-colored transparent surface in Figures 5 and Figure S4. In the case of FMN, analysis of both the FragMaps and the exclusion map indicates the binding site to be flexible, allowing the solutes to sample that site. During ligand design, the exclusion maps are used to represent the RNA surface rather than the solvent accessible surface based on the crystal structures, as they identify the extent to which the RNA can relax to allow for ligand binding “under” the traditional macromolecular surface.

Ligand Binding Site Characterization with SILCS FragMaps

Intercalators: The TPP riboswitch exhibits intercalation in the binding site (Figure 6I) which is occupied by the physiological ligand, thiamine pyrophosphate, in the crystal structure used to initiate the SILCS simulations. Though classical intercalators are relatively nonspecific against both DNA and RNA, when combined with specific H-bonding and electrostatic interactions, they can provide specificity as well as high affinity.¹⁰¹ To capture the contribution of stacking interactions driven by hydrophobic effects in SILCS, the aromatic probes benzene and imidazole need to sample the intercalation sites. In our initial application of SILCS to RNA, weak harmonic restraints were applied on only P atoms of the backbone while allowing unbiased dynamics for the sugars and bases; however, under these conditions, there were no favorable apolar or imidazole related FragMaps (GEHC, IMIN or IMIH) in the intercalation sites (not shown). This led to the use of weak harmonic

restraints with a force constant of $0.12 \text{ kcal/mol/\AA}^2$ on only the C1' atoms of the sugar and N1/N3 atoms of the bases on all nucleotides in the RNAs. These restraints maintain the conformation of the binding site in TPP, allowing for sampling of the apolar solutes in the site (Figure 6IA). In addition, H-bond donor, H-bond acceptor, and positive FragMaps are present in the intercalation site. Beyond the intercalation site in the region occupied by the remainder of the ligand, there are limited FragMaps, with no FragMaps present in the region of the diphosphate group. The 50 nM affinity of the ligand suggests that the phosphates may contribute to binding, motivating additional analysis of the regions of the binding sites that accommodate phosphates as presented below.

HCV-IRES-IIa RNA has been reported to bind benzimidazole derivatives with low-micromolar affinities.⁸¹ The ligands interact through intercalation of the heteroaromatic rings between two bases with the polar moieties coming off the heterocycles exposed to solvent (Figure S3C). SILCS FragMaps successfully capture these binding site characteristics as seen by the presence of the apolar, H-bond donor, H-bond acceptor, and methanol O FragMaps (Figure 6II). The presence of the polar maps is consistent with the HCV binding site featuring other bases offering hydrogen bonding opportunities on the inner side of the binding site. Additionally, FragMaps for positively charged methylammonium nitrogen are present in the binding site and at the exposed face of the site (Figure 6II). This pattern of FragMaps from the new SILCS-RNA protocol is consistent with significantly improved pose refinement results over the standard SILCS protocol yielding the LGFE scores comparable to low micromolar experimental binding affinities (Table S5).

Structural Pocket Binders: The FMN riboswitch binding site has been shown to bind the ribocil molecules at low nM affinities in U-shaped conformations (Figure S2A and ct2c00381_si_002.xlsx). One leg of the U-shape binds a hydrophobic region between the nucleobases from two strands (A48 and A85) which is also a primary location where the tricyclic moiety of flavin mononucleotide, the physiological ligand, binds. The other leg uses stacking interactions with G62 from a third strand. The FragMaps represent the two regions with highly favorable apolar maps along with multiple locations for donors and acceptors around them (Figure 7I - A, B, and C). The presence of both apolar and polar FragMaps in both ends of the U-shaped binding pocket is consistent with the high affinity of the studied ligands.

The dG riboswitch binds a guanine mononucleoside in an occluded binding site in the central region of the tertiary structure of the RNA (Figure 3D and Figure S2D). Consistent with the Gua base, the binding site is occupied by apolar, H-bond donor, and H-bond acceptor FragMaps (Figure 7II). Local alcohol FragMaps are in the vicinity of the sugar 2' hydroxyl group, though these are only evident at a GFE cutoff of -0.2 kcal/mol . Interestingly, the bottom of the binding site where the 5' hydroxyl of the mononucleoside is located is occupied by positive FragMaps. The presence of FragMaps throughout the binding site indicates that all regions of the ligand contribute to its binding affinity consistent with the reported binding affinity of 300 nM.

The THF riboswitch contains two binding sites, B1 and B2, as shown in Figure 3C. Both sites are only partially buried in the RNA with a significant portion of the ligands

highly exposed to the surrounding solvent. In both sites the heteroaromatic ring of the tetrahydrofolate ligand is in the binding pocket in the RNA with the diacid group in the B1 and B2 sites exposed to the solvent (Figure 7III and IV). While not resolved in the first site, the diacid group would also be solvent exposed based on the location of the amide linker that was resolved. With both sites, there are no apolar FragMaps in the pocket targeted by the heteroaromatic ring, though there is such a map in the vicinity of the phenyl ring in site B1. H-bond donor, H-bond acceptor, and alcohol FragMaps are present in the region of the pocket that binds the heteroaromatic, ring while no FragMaps are present in the region of the ligands that protrude beyond the RNA. The indicated lack of contribution of this region of the ligand is consistent with its relatively low affinity of 19 μM .

Minor and Major Groove Binding Sites: The wide and shallow nature of the minor groove in the A-form RNA has been recognized as an important player in mediating tertiary interactions.¹⁰² While both the major and minor groove in nucleic acids can offer distinct hydrogen bond donor/acceptor patterns¹⁰³, the minor groove in RNA offers additional interactions with 2'OH groups allowing non-canonical interactions to achieve complex RNA folds.¹⁰⁴ The SILCS FragMaps for HIV1-TAR (Figure 8III) capture the characteristics of multiple hydrogen bond donors and acceptors in the minor groove of the relatively helical RNA structure. Aminoglycosides have been shown to preferentially bind to the minor groove of the A-form double-stranded RNA¹⁰⁵, for example, neomycin-B.⁴⁷ Though the structure (PDB: 6CMN) used for SILCS simulations has significant differences compared to the NMR structure containing neomycin-B (PDB: 1QD3), the relevant FragMaps are in the minor groove binding site, B2 (Figure 8II). The major groove with α -helical geometry distorted by internal loops, bulged out nucleotides, or nonregular base pairs can offer interactions for small molecule binding.⁴⁷ For example, HIV1-TAR RNA exhibits a three-base bulge distorting the major groove that allows extra flexibility for induced-fit interaction with Tat.¹⁰⁶ FragMaps are present in this binding site, B1, (Figure 8 I), where analogues of bis-guanidine compound rbt203 (tar008) or the arginine rich peptide from Tat bind. A small molecule probe, mv2003 (tar012), was shown to induce dual interactions with the bases of the apical loop and with the arginine-binding pocket of TAR to cause major groove tightening that results in creating a new pocket for small molecules (PDB: 2L8H).⁴⁸ Such a drastic difference in conformation can limit the use of FragMaps from another structure. Interestingly, the starting structure used in this study exhibits the apical base A35 in a flipped state, and this creates a cavity for sampling of SILCS probes. Figure 8III, with the mv2003 ligand positioned based on structural alignment of PDB 6CMN with PDB 2L8H, shows favorable FragMaps in that cavity.

The binding site of IAV RNA promotor, where a novel scaffold binds⁶⁴, is in the major groove corresponding to a nearly helical region with high accessibility to solvent (Figure 8IV). While apolar, H-bond donor, and H-bond acceptor maps are present in the major groove they are not highly favorable with the GFE level for visualization being -0.4 kcal/mol, with the exception of the positive maps, in Figure 8IV. This is consistent with the relatively low affinity of the ligand targeting this RNA, with a binding affinity of 50 μM for the small molecule in PDB 2LWK.

Ligand Posing and Ranking with SILCS-RNA: Testing of the SILCS-RNA methodology used a database of all of the known ligands, 104 total, for the seven RNA targets curated as part of the present study (ct2c00381_si_002.xlsx). The available data for the binding affinities for these small molecules is very limited and inconsistent across the literature for the individual targets with variations of up to 20-fold in the affinities for some ligands reported in different studies. Nevertheless, evaluation of the overall ability of SILCS-RNA to predict the relative affinities and binding poses was undertaken by considering selected data (kd_best in ct2c00381_si_002.xlsx).

SILCS-MC pose refinement was performed for the ligands against the respective targets starting from known binding orientations. Initial ligand positioning was performed by aligning the experimental ligand–RNA PDB structures with the structure used for the SILCS simulations of each RNA. All other ligands were aligned to the experimentally determined orientation of the ligand in the structure used for the SILCS simulations using flexible alignment in MOE¹⁰⁷. For the posing, both the generic and specific ACSs were used to evaluate the LGFE scores. Details of the SILCS-MC sampling protocols are provided in Table 2. Using the SILCS-MC pose refinement method yields favorable LGFEs (ct2c00381_si_003.xlsx). However, the LGFE values are typically underpredicted with respect to the experimental binding free energies (Figure 9). With SILCS-MC docking, the LGFE scores typically become more favorable, consistent with the additional orientation and conformational relaxation over pose refinement, though the values are still typically underestimated except for some TAR ligands. While the SILCS LGFE scores are not direct estimates of experimental binding affinities, in the case of proteins, they are typically similar to the experimental values suggesting that, despite the adjustments made in the SILCS protocol to treat RNA, the FragMaps are still underrepresenting the favorable interactions of the respective functional groups with the RNA.

Pearson correlation with the experimental binding affinities over all of the RNAs and ligands revealed weak correlation with values of ~ 0.5 (Figure 9). This level of correlation occurs with both the SILCS-MC pose refinement and more rigorous docking methods. From analysis of the plots in Figure 9, it is evident that the level of correlation is associated with the relative affinities of the ligands over the different targets. Particularly, for GUA, the LGFE scores did not improve for docking, while, for FMN, TPP, and THF, the LGFE scores become favorable. As mentioned, some positively charged aminoglycoside ligands of TAR gain additional LGFE scores with higher contributions from positive MAMY maps. Correlation analysis of the individual targets generally led to very low or even anticorrelation associated with the small number of ligands against each target for which experiment data exists (ct2c00381_si_002.xlsx). For many targets, there are less than 10 ligands on which experimental data is available (Table 1), and in several cases, that data comes from different experimental studies. For example, with FMN, the experimental binding affinity data for the eight ligands comes from four separate studies. Another issue is the range of the experimental binding affinities. With IVP RNA, the range over the seven ligands is as low as 0.76 kcal/mol, and in the case of HCV, the range is 1.88 kcal/mol over seven ligands. However, the presence of the consistent, albeit low, correlation across the

studied targets does suggest that the SILCS FragMaps are representative of the energetic contributions to the binding of the different classes of ligand–RNA interactions.

Additional analysis of the SILCS-MC method included the average RMSD between the optimized ligand orientations with respect to starting poses for the different targets (Table 3). As expected, the pose refinement yields the smallest RMSD values, consistent with the local relaxation of the ligands in the field of the FragMaps. Full SILCS-MC docking leads to larger RMSD values as the ligands reorient to maximize their overlap with the SILCS FragMaps. Based on the analysis of the minimum RMSD values over the different targets and SILCS-MC methods, the results indicate that the SILCS-MC docking with GAS21 and a 5 Å radius would be the default approach to use in the absence of additional information to select the ideal docking approach. This selection is consistent with the average LGFE values in Table S6 showing the LGFE scores to become significantly more favorable upon going from pose refinement to docking with the more favorable values largely converged with the 5 Å docking radius.

Phosphate Containing Ligands:

There are eight phosphate containing ligands in the present database (Figure S7). The TPP riboswitch has been shown to bind five ligands with either mono- or diphosphate groups. Experimental binding affinity data shows that the phosphates contribute favorably to the affinities compared to their unphosphorylated counterparts (fnn001– fnn002 and tpp005 – tpp004, ct2c00381_si_002.xlsx). As stated above, there were no negatively charged favorable FragMaps within 6 Å from the surface of the RNA. Given the polyanionic nature of the RNA, we initially used neutral states of phosphate groups in the ligands where the anionic and hydroxyl oxygens were assigned to neutral GENA and MEOO FragMaps, respectively. However, the LGFE scores using the GAS21 with a 2 Å radius in SILCS-MC docking did not become significantly more favorable (ct2c00381_si_003.xlsx). LGFE scores were also calculated for the TPP riboswitch ligands in which the crystal Mg^{2+} ions and neutralizing Na^+ ions were included in the charged set of SILCS GCMC-MD simulations with the SILCS-MC performed with the standard negative assignment (ACEC FragMap type) to the ligand phosphates. This also did not lead to more favorable LGFE scores. Subsequent analysis of the GCMC-MD simulations showed that in some of the TPP simulation systems there were acetate solutes initially placed adjacent to the Mg^{2+} ions but that over the course of the 100 GCMC-MD cycles these were deleted from these regions and not reinserted, yielding unfavorable negative FragMaps (ie. $\text{GFE} > 0$ kcal/mol) in these regions. Motivated by this observation, additional SILCS simulations with charged solutes were performed by turning off the GCMC moves for both methylammonium and acetate while maintaining water GCMC moves during the production stage of the simulations. Using this approach, the acetate molecules are maintained in the region around positively charged Mg^{2+} ions, yielding favorable negative FragMaps. This was reflected in the pose refinement calculations with the phosphate groups of the ligands deprotonated to have –1 charge (Table S7). Significantly more favorable LGFE scores are maintained, bracketing the experimental binding free energies for the three ligands for which experimental data is available. While this approach was possible for the TPP riboswitch, which appears to be due to the binding pocket not being fully buried with access to the surrounding aqueous solution,

it did not work for FMN, as it has a much deeper pocket with both an inner and outer sphere of nucleobases,⁶ disallowing insertion of acetate solutes even during the initial refilling stage of the GCMC-MD protocol. These results suggest that, in systems in which phosphates or other negative charged moieties are present on the ligands, the omission of the GCMC on charged set may be of utility, though studies on additional ligand–RNA systems are needed to better validate the approach.

Atomic GFE Contributions to Ligand Binding: During ligand optimization, it is highly desirable to know the contributions of different portions of a molecule to changes in relative affinities associated with chemical modifications. The quantitative breakdown of LGFEs based on the atom-wise GFE contributions in SILCS facilitates this to supply information to optimize lead molecules efficiently. Figure 10 shows the atomic and group GFE contributions for three HCV ligands that differ subtly in the linkers to functional groups connected to the common benzimidazole structure. The differences in the experimental binding affinity of the three compounds is less than 0.5 kcal/mol with the LGFE scores differing by less than 0.9 kcal/mol with the docked orientations of the compounds being similar. While the binding affinities of the ligands are similar, a breakdown of the LGFEs clearly shows that the major contributions to differences in predicted binding affinities are coming from all regions of the molecules. The summed GFE score of the benzimidazole bicyclic groups has a maximum difference of 0.6 kcal/mol. The dimethylamine groups on the left side have the most apparent effect on the total LGFE with the groups making a contribution of -2.1 kcal/mol in the case of hcv007 to up to -4.3 kcal/mol in hcv002. The less favorable contribution of the left dimethylamine group in hcv007 is compensated by more favorable contributions of the central benzimidazole ring system and the right dimethylamine group, which makes an unfavorable contribution in hcv003. Thus, while the types of functional groups in the three compounds are similar, changes in the length between them and constraints associated with the presence or absence of additional rings leads to changes in the contributions of the individual groups that are larger than the differences in the LGFE scores. Such information can be leveraged to facilitate decision making on further ligand modifications during iterative cycles of ligand design.

Binding Site Identification with SILCS Hotspots: Once available, FragMaps may be used for various calculations including the identification of novel ligand binding sites, referred to as SILCS-Hotspots.⁷⁵ Previously only LigandRNA¹⁹ has mentioned an option to visualize the potential for different regions on a RNA surface to be considered as binding sites of small molecules. However, this capability was not further developed in its recent version AnnapuRNA²⁰. SILCS-Hotspots is an approach that utilizes SILCS-MC to comprehensively dock chemical fragments found in drug-like molecules in the entire 3D space of the macromolecule. This allows for identification of possible binding sites throughout the RNAs as well as ranking of the sites based on multiple criteria. Using the ASTEX mini frag database,⁷⁶ we performed SILCS-Hotspots analysis for the seven RNAs. Shown in Figure 11 are the identified hotspots for the seven targets. SILCS-Hotspots identifies a large number of weak binding sites with a cutoff of -2.0 kcal/mol for the average LGFE over all of the fragments in each hotspot. While many of the sites are located on the

surface of the RNA, the known ligand binding sites on the RNAs typically correspond to one of the most favorable hotspots. This is anticipated given that the SILCS simulations were initiated from the crystal structures that included the bound ligands, though SILCS-Hotspots is based on the SILCS FragMaps and exclusion map and does not directly use the crystal structures. In the case of TAR, site B2 is occupied by favorable hotspots, indicating the ability of the method to identify a binding site not occupied by a ligand in the structure used for the SILCS simulation. Additional analysis of the hotspots was therefore undertaken to see if additional putative binding sites on the RNAs could be identified.

As previously discussed with proteins, the most favorable hotspots do not necessarily correspond to binding sites for drug-like molecules.⁷⁵ While other quantitative criteria for ranking, such as the number of fragments per site, average fragment ligand efficiency per site, and lowest LGFE over all of the fragments in a site, may be applied it was shown that qualitative identification of hotspots that were adjacent to one or more additional hotspots was of utility for identifying putative druggable sites. Selection of these sites includes one or more adjacent hotspots without steric hindrance between them based on the exclusion maps that would preclude linking fragments between those sites. The success of this approach was verified by the identification of known allosteric and orthosteric binding sites in proteins⁷⁵ and the identification of novel allosteric sites in β -glucosidase A.¹⁰⁸

Visual analysis was thus undertaken to identify novel putative binding sites followed by the application of SILCS-MC docking on a database of 380 chemically diverse FDA approved compounds to quantitatively evaluate the suitability of the identified sites. The FDA compounds were selected, as they represent drug-like molecules such that their favorable interactions with the putative binding sites would indicate that the sites are suitable for the development of therapeutically relevant ligands. Shown in Table 4 are the results for the top 20 LGFE-ranked ligands in the putative binding sites indicated in Figure 11 including the known ligand binding sites for the target RNAs. Specific LGFE values for the top 20 compounds and the list of 380 compounds are provided in ct2c00381_si_004.xlsx. Analysis included the LGFE scores along with the relative solvent accessible surface area (rSASA) values to identify sites that are predicted to have favorable binding interactions in deep pockets in which bound ligands have low exposure to the surrounding aqueous solution. The known binding sites on the high-information-content RNAs (FMN, TPP, and GUA) have average LGFEs more favorable than -6 kcal/mol and average rSASA values greater than 80%. While the lower information content RNAs (HCV, TAR, and IVP) and THF with partially buried sites often have favorable average LGFE scores for the known sites, the rSASA scores are lower with only HCV having a value $> 75\%$ with the other sites all being $> 52\%$. The average LGFE and rSASA scores for the known sites were then used as guides to select putative binding sites. This yields potential sites on FMN, TPP, GUA, HCV, and IVP. HS#14 on GUA has both a favorable average LGFE of -7.4 kcal/mol and an average rSASA of 82%, with the pocket being nicely sequestered in the core of the RNA, as shown in Figure S8. High ranking sites are also seen on TPP and HCV with average LGFEs of approximately -6 kcal/mol and average rSASAs greater than 70%. In HCV, HS#2 is situated in the core of the deformed major groove, yielding a deep pocket which offers relatively favorable binding affinities. Additional interesting sites include HS#9 on FMN with a favorable average rSASA of 89%, though the average LGFE is -4.4 kcal/mol, and

HS#7 on IVP with a favorable average LGFE of -8.8 kcal/mol, though the site is relatively solvent exposed with an average rSASA of 47%. In IVP, HS#7 in the minor groove offers more favorable interactions as compared to the known binding site HS#6 situated in the major groove. Overall, these results indicate the potential for the SILCS methodology to identify novel binding sites on RNAs, sites that could be subjected to subsequent docking calculations and experimental validation.

Current Challenges in the Application of CADD to RNA: Identification of druggable RNA targets and specificity in targeting are the well-recognized challenges in the field of RNA small molecule drug design.¹⁰⁹ With developments in RNA sequencing methods, a better understanding of the biological role of long noncoding RNAs that were initially ignored or dismissed as “transcriptional noise” is now being achieved.^{110, 111} This increase in therapeutically interesting RNAs combined with structural methods is leading to a significant increase in the number of RNA-based systems on which CADD methods may be applied. However, as is clear from the present as well as previous studies (Table S1), significant challenges for CADD are present. Unlike for proteins²⁴, the available data for the small molecule binding affinity of RNA targets is insufficient to provide quantitative validation of the CADD approaches. Specifically, out of the 104 reported ligands for the seven RNAs in our study, which we selected based on the availability of 3D structures and multiple ligands, experimental affinity data is only available for 68 ligands. Moreover, that data was reported in 13 different studies. Thus, not only is there a sparsity of systems for model development, but the available data is also not consistent, as required for the validation of quantitative, predictive CADD methods. A particular challenge is the selection of a starting 3D structure for the discovery of new ligands using docking methods given the conformational heterogeneity of RNA. This is particularly true for the RNA with low-information-content binding sites. For example, with HIV1-TAR RNA, there are multiple ligand–RNA complexes that show significant differences in conformations based on the interacting partner ligand (Figure S9). While SILCS-RNA was able to identify the binding site in PDB 1UUI, 1UUD, and 1QD3 based on the SILCS simulations initiated with PDB 6CMN in the case of HIV1-TAR, it was not able to identify the sites in PDBs 1UTS and 2L8H. This points toward the sensitivity of RNA 3D structures to the bound ligands, representing a significant challenge in the identification of new ligands targeting RNA. To overcome this, the use of structures from RNA folding studies combined with MD simulations exploiting enhanced sampling⁹⁵ offers the potential to identify multiple conformations for docking studies. Such structures could then be used in the SILCS-Hotspots approach described above. We note that this goes beyond ensemble docking methods^{112, 113} used with proteins, as those methods typically focus on the generation of multiple structures in a conformational basin. Rather, it will be necessary for distinct conformational basins to be identified. In the context of SILCS-RNA, individual SILCS simulations, which each represent the local ensemble of conformations, will be required on each unique conformational basin for subsequent binding site identification, docking, and ligand design.

As shown in the present paper, current CADD methods are more suitable to ligand optimization (Table S1). The SILCS FragMaps can model the functional group distributions

in the known binding sites, especially in the regions occupied by polar, neutral groups such as heteroaromatic moieties. Importantly, the approach is appropriate for intercalation sites based on the use of the restraints on the C1', N1, and N3 atoms. Another problem directly associated with the polyanionic nature of RNA is the modeling of ligands with phosphates or other negatively charged groups. Despite the use of the split SILCS simulation protocol, it is a challenge to model the distribution of negative FragMaps in those regions, though the protocol that omitted GCMC for the charged groups during the GCMC-MD production portions of the simulations showed some success for the TPP riboswitch phosphate containing ligands but not for those in FMN and GUA.

Additional Methodological Considerations: As described above, molecular simulations of nucleic acids are highly challenging. In this section, we address additional details of the developed approach. Mg^{2+} ions are important for RNA to assume 3D structures⁹³ and are known to bind in well-defined locations on RNA molecules, as shown in structural studies^{6, 104}, including playing key roles in ligand–RNA interactions as seen in the TPP and FMN systems included in the present study. To understand the impact of Mg^{2+} on the SILCS-RNA approach, the method was applied in both the presence and absence of Mg^{2+} ions and with and without GCMC sampling of the ACEY and MAMY solutes. These different approach yielded mixed results due to challenges associated with the anionic solute, ACEY, sampling regions around the Mg^{2+} in the SILCS-RNA simulations. This clearly represents a simulation methodology challenge that will require further study.

The role of ionic strength significantly impacts RNA and DNA structure and dynamics.^{114, 115} In the present study, the ionic strength was varied from zero up to approximately 0.6 M when the solutes acetate and methylammonium were present along with NaCl. Notably, as weak harmonic restraints are applied to selected atoms in the RNA during the SILCS simulations, the impact of ionic strength, as well as the type of ions, will only minimally impact the conformational sampling of the RNA. However, as detailed in this study, even at the higher ionic strength there were challenges with MAMY ions dominating the sampling of the region in and around the RNA (Figure 4) as well as achieving ACEY sampling in the vicinity of Mg^{2+} ions. This again represents a significant challenge to the SILCS-RNA methodology that requires future consideration.

The applied empirical force field represents another consideration. The present study was based on the CHARMM36 RNA and CGenFF additive force field.^{69, 71} Alternative force fields that could be considered are the AMBER nucleic acid force fields¹¹⁶ and associated ions and, notably, the emerging polarizable force fields such as the Drude^{117, 118} and AMOEBA¹¹⁹ models, for which nucleic acid specific parameters have been presented. Indeed, the Drude FF has been shown to yield improved treatment of ions in the presence of DNA, and a successful Drude model of Mg^{2+} has been reported¹²⁰ but only applied in limited simulation studies of nucleic acids.¹¹⁷ Testing of alternate force fields represents an additional avenue for future studies. Questions to address include the following: Will an alternate additive or a polarizable force field better account for the issue with MAMY and ACEY sampling seen in the present study? Will alternate force fields lead to the identification of different binding sites? Will alternate force fields lead to significant differences in ligand scoring. While these questions remain open, we note that, in the SILCS

methodology, the force field combined with the GCMC-MD simulation approach directly impacts calculation of the FragMaps. However, as ligand docking and scoring are based on the overlap of the ligand atoms with the corresponding FragMaps, the docked poses and predicted binding affinities in the form of LGFEs are relatively insensitive to the underlying force field, which only impacts the conformational properties of the ligands during MC docking.

Conclusions

Presented is the extension of the SILCS approach to RNA, termed SILCS-RNA. Associated with the polyanionic nature of RNA, it was shown that a dual SILCS simulation protocol is required to allow for sufficient sampling of the apolar and polar neutral solutes in and around the RNA separately from the charged solutes acetate and methylammonium. The resulting SILCS FragMaps are shown to effectively map the functional group affinity pattern of known ligand binding sites, including intercalation sites. However, for binding regions occupied by phosphates or other negative groups, prudent selection of the inclusion of Mg^{2+} ions is required along with an altered SILCS simulation approach, though further work is needed in this area. With respect to quantitative ligand ranking, the SILCS LGFE scores do yield a low level of correlation when all of the ligands for all of the targets are considered. This indicates that the SILCS FragMaps are representative of the functional group affinity patterns associated with different classes of binding sites in the RNAs. The approach was not able to rank the ligands for the individual RNAs, though limitations in the number of ligands and potential inconsistency of the experimental binding affinities severely limit such validation. Notably, the SILCS-Hotspots approach shows success in identifying putative binding sites, though the conformational heterogeneity of RNA precludes identification of several of the known sites. Clearly, the present results and discussion as well as the overview of other CADD methods indicate the formidable challenges against the application of CADD methods on RNA. Thus, the guidelines for efforts toward RNA-targeted drug discovery reported by Warner et. al.⁶ are still very relevant such that high-quality pockets formed with unique RNA motifs are critical to apply state-of-the-art methodologies and results have to be interpreted in a specific manner to consider the effects of intercalation and electrostatic interactions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Financial support from the NIH (GM131710) to ADM Jr. Computational support from the University of Maryland Computer-Aided Drug Design Center and the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562.

References

1. Macalino SJ; Gosu V; Hong S; Choi S, Role of computer-aided drug design in modern drug discovery. *Arch Pharm Res* 2015, 38 (9), 1686–701. [PubMed: 26208641]

2. Mullard A, Small molecules against RNA targets attract big backers. *Nat Rev Drug Discov* 2017, 16 (12), 813–815. [PubMed: 29180732]
3. Disney MD, Targeting RNA with Small Molecules To Capture Opportunities at the Intersection of Chemistry, Biology, and Medicine. *J Am Chem Soc* 2019, 141 (17), 6776–6790. [PubMed: 30896935]
4. Petrone J; DeFrancesco L, Small molecules get the message. *Nat Biotechnol* 2018, 36 (9), 787–790. [PubMed: 30188529]
5. Crooke ST; Witztum JL; Bennett CF; Baker BF, RNA-Targeted Therapeutics. *Cell Metab* 2018, 27 (4), 714–739. [PubMed: 29617640]
6. Warner KD; Hajdin CE; Weeks KM, Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* 2018, 17 (8), 547–558. [PubMed: 29977051]
7. Mukherjee H; Blain JC; Vandivier LE; Chin DN; Friedman JE; Liu F; Maillet A; Fang C; Kaplan JB; Li J; Chenoweth DM; Christensen AB; Petersen LK; Hansen NJV; Barrera L; Kubica N; Kumaravel G; Petter JC, PEARL-seq: A Photoaffinity Platform for the Analysis of Small Molecule-RNA Interactions. *ACS Chem Biol* 2020, 15 (9), 2374–2381. [PubMed: 32804474]
8. Sheridan C, First small-molecule drug targeting RNA gains momentum. *Nat Biotechnol* 2021, 39 (1), 6–8. [PubMed: 33432225]
9. Stelzer AC; Frank AT; Kratz JD; Swanson MD; Gonzalez-Hernandez MJ; Lee J; Andricioaei I; Markovitz DM; Al-Hashimi HM, Discovery of selective bioactive small molecules by targeting an RNA dynamic ensemble. *Nat Chem Biol* 2011, 7 (8), 553–9. [PubMed: 21706033]
10. Rizvi NF; Howe JA; Nahvi A; Klein DJ; Fischmann TO; Kim HY; McCoy MA; Walker SS; Hruza A; Richards MP; Chamberlin C; Saradjian P; Butko MT; Mercado G; Burchard J; Strickland C; Dandliker PJ; Smith GF; Nickbarg EB, Discovery of Selective RNA-Binding Small Molecules by Affinity-Selection Mass Spectrometry. *ACS Chem Biol* 2018, 13 (3), 820–831. [PubMed: 29412640]
11. Leclerc F; Cedergren R, Modeling RNA-ligand interactions: the Rev-binding element RNA-aminoglycoside complex. *J Med Chem* 1998, 41 (2), 175–82. [PubMed: 9457241]
12. Morley SD; Afshar M, Validation of an empirical RNA-ligand scoring function for fast flexible docking using Ribodock. *J Comput Aided Mol Des* 2004, 18 (3), 189–208. [PubMed: 15368919]
13. Ruiz-Carmona S; Alvarez-Garcia D; Foloppe N; Garmendia-Doval AB; Juhos S; Schmidtke P; Barril X; Hubbard RE; Morley SD, rDock: a fast, versatile and open source program for docking ligands to proteins and nucleic acids. *PLoS Comput Biol* 2014, 10 (4), e1003571. [PubMed: 24722481]
14. Pfeffer P; Gohlke H, DrugScoreRNA--knowledge-based scoring function to predict RNA-ligand interactions. *J Chem Inf Model* 2007, 47 (5), 1868–76. [PubMed: 17705464]
15. Guilbert C; James TL, Docking to RNA via root-mean-square-deviation-driven energy minimization with flexible ligands and flexible targets. *J Chem Inf Model* 2008, 48 (6), 1257–68. [PubMed: 18510306]
16. MacKerell AD; Banavali N; Foloppe N, Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* 2000, 56 (4), 257–265. [PubMed: 11754339]
17. Wang J; Wolf RM; Caldwell JW; Kollman PA; Case DA, Development and testing of a general amber force field. *J Comput Chem* 2004, 25 (9), 1157–74. [PubMed: 15116359]
18. Lang PT; Brozell SR; Mukherjee S; Pettersen EF; Meng EC; Thomas V; Rizzo RC; Case DA; James TL; Kuntz ID, DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA* 2009, 15 (6), 1219–30. [PubMed: 19369428]
19. Philips A; Milanowska K; Lach G; Bujnicki JM, LigandRNA: computational predictor of RNA-ligand interactions. *RNA* 2013, 19 (12), 1605–16. [PubMed: 24145824]
20. Stefaniak F; Bujnicki JM, AnnapuRNA: A scoring function for predicting RNA-small molecule binding poses. *PLoS Comput Biol* 2021, 17 (2), e1008309. [PubMed: 33524009]
21. Sun LZ; Jiang Y; Zhou Y; Chen SJ, RLDOCK: A New Method for Predicting RNA-Ligand Interactions. *J Chem Theory Comput* 2020, 16 (11), 7173–7183. [PubMed: 33095555]
22. Lakkaraju SK; Raman EP; Yu W; MacKerell AD Jr., Sampling of Organic Solutes in Aqueous and Heterogeneous Environments Using Oscillating Excess Chemical Potentials in Grand Canonical-

- like Monte Carlo-Molecular Dynamics Simulations. *J Chem Theory Comput* 2014, 10 (6), 2281–2290. [PubMed: 24932136]
23. Ustach VD; Lakkaraju SK; Jo S; Yu W; Jiang W; MacKerell AD Jr., Optimization and Evaluation of Site-Identification by Ligand Competitive Saturation (SILCS) as a Tool for Target-Based Ligand Optimization. *J Chem Inf Model* 2019, 59 (6), 3018–3035. [PubMed: 31034213]
24. Goel H; Hazel A; Ustach VD; Jo S; Yu W; MacKerell AD Jr., Rapid and accurate estimation of protein-ligand relative binding affinities using site-identification by ligand competitive saturation. *Chem Sci* 2021, 12 (25), 8844–8858. [PubMed: 34257885]
25. Guvench O; MacKerell AD Jr., Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput Biol* 2009, 5 (7), e1000435. [PubMed: 19593374]
26. Raman EP; Yu W; Guvench O; Mackerell AD, Reproducing crystal binding modes of ligand functional groups using Site-Identification by Ligand Competitive Saturation (SILCS) simulations. *J Chem Inf Model* 2011, 51 (4), 877–96. [PubMed: 21456594]
27. Howe JA; Xiao L; Fischmann TO; Wang H; Tang H; Villafania A; Zhang R; Barbieri CM; Roemer T, Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the E. coli FMN riboswitch. *RNA Biol* 2016, 13 (10), 946–954. [PubMed: 27485612]
28. Serganov A; Huang L; Patel DJ, Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* 2009, 458 (7235), 233–7. [PubMed: 19169240]
29. Vicens Q; Mondragon E; Reyes FE; Coish P; Aristoff P; Berman J; Kaur H; Kells KW; Wickens P; Wilson J; Gadwood RC; Schostarez HJ; Suto RK; Blount KF; Batey RT, Structure-Activity Relationship of Flavin Analogues That Target the Flavin Mononucleotide Riboswitch. *ACS Chem Biol* 2018, 13 (10), 2908–2919. [PubMed: 30107111]
30. Howe JA; Wang H; Fischmann TO; Balibar CJ; Xiao L; Galgoci AM; Malinverni JC; Mayhood T; Villafania A; Nahvi A; Murgolo N; Barbieri CM; Mann PA; Carr D; Xia E; Zuck P; Riley D; Painter RE; Walker SS; Sherborne B; de Jesus R; Pan W; Plotkin MA; Wu J; Rindgen D; Cummings J; Garlisi CG; Zhang R; Sheth PR; Gill CJ; Tang H; Roemer T, Selective small-molecule inhibition of an RNA structural element. *Nature* 2015, 526 (7575), 672–7. [PubMed: 26416753]
31. Vicens Q; Mondragon E; Batey RT, Molecular sensing by the aptamer domain of the FMN riboswitch: a general model for ligand binding by conformational selection. *Nucleic Acids Res* 2011, 39 (19), 8586–98. [PubMed: 21745821]
32. Serganov A; Polonskaia A; Phan AT; Breaker RR; Patel DJ, Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* 2006, 441 (7097), 1167–71. [PubMed: 16728979]
33. Edwards TE; Ferre-D'Amare AR, Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. *Structure* 2006, 14 (9), 1459–68. [PubMed: 16962976]
34. Warner KD; Homan P; Weeks KM; Smith AG; Abell C; Ferre-D'Amare AR, Validating fragment-based drug discovery for biological RNAs: lead fragments bind and remodel the TPP riboswitch specifically. *Chem Biol* 2014, 21 (5), 591–5. [PubMed: 24768306]
35. Kulshina N; Edwards TE; Ferre-D'Amare AR, Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch. *RNA* 2010, 16 (1), 186–96. [PubMed: 19948769]
36. Thore S; Frick C; Ban N, Structural basis of thiamine pyrophosphate analogues binding to the eukaryotic riboswitch. *J Am Chem Soc* 2008, 130 (26), 8116–7. [PubMed: 18533652]
37. Trausch JJ; Batey RT, A disconnect between high-affinity binding and efficient regulation by antifolates and purines in the tetrahydrofolate riboswitch. *Chem Biol* 2014, 21 (2), 205–16. [PubMed: 24388757]
38. Huang L; Ishibe-Murakami S; Patel DJ; Serganov A, Long-range pseudoknot interactions dictate the regulatory response in the tetrahydrofolate riboswitch. *Proc Natl Acad Sci U S A* 2011, 108 (36), 14801–6. [PubMed: 21873197]
39. Wilt HM; Yu P; Tan K; Wang YX; Stagno JR, Tying the knot in the tetrahydrofolate (THF) riboswitch: A molecular basis for gene regulation. *J Struct Biol* 2021, 213 (1), 107703. [PubMed: 33571639]

40. Pikovskaya O; Polonskaia A; Patel DJ; Serganov A, Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nat Chem Biol* 2011, 7 (10), 748–55. [PubMed: 21841796]
41. Buck J; Wacker A; Warkentin E; Wohnert J; Wirmer-Bartoschek J; Schwalbe H, Influence of ground-state structure and Mg²⁺ binding on folding kinetics of the guanine-sensing riboswitch aptamer domain. *Nucleic Acids Res* 2011, 39 (22), 9768–78. [PubMed: 21890900]
42. Matyjasik MM; Hall SD; Batey RT, High Affinity Binding of N2-Modified Guanine Derivatives Significantly Disrupts the Ligand Binding Pocket of the Guanine Riboswitch. *Molecules* 2020, 25 (10).
43. Gilbert SD; Reyes FE; Edwards AL; Batey RT, Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure* 2009, 17 (6), 857–68. [PubMed: 19523903]
44. Belashov IA; Crawford DW; Cavender CE; Dai P; Beardslee PC; Mathews DH; Pentelute BL; McNaughton BR; Wedekind JE, Structure of HIV TAR in complex with a Lab-Evolved RRM provides insight into duplex RNA recognition and synthesis of a constrained peptide that impairs transcription. *Nucleic Acids Res* 2018, 46 (13), 6401–6415. [PubMed: 29961805]
45. Davis B; Afshar M; Varani G; Murchie AI; Karn J; Lentzen G; Drysdale M; Bower J; Potter AJ; Starkey ID; Swarbrick T; Aboul-ela F, Rational design of inhibitors of HIV-1 TAR RNA through the stabilisation of electrostatic “hot spots”. *J Mol Biol* 2004, 336 (2), 343–56. [PubMed: 14757049]
46. Murchie AI; Davis B; Isel C; Afshar M; Drysdale MJ; Bower J; Potter AJ; Starkey ID; Swarbrick TM; Mirza S; Prescott CD; Vaglio P; Aboul-ela F; Karn J, Structure-based drug design targeting an inactive RNA conformation: exploiting the flexibility of HIV-1 TAR RNA. *J Mol Biol* 2004, 336 (3), 625–38. [PubMed: 15095977]
47. Faber C; Sticht H; Schweimer K; Rosch P, Structural rearrangements of HIV-1 Tat-responsive RNA upon binding of neomycin B. *J Biol Chem* 2000, 275 (27), 20660–6. [PubMed: 10747964]
48. Davidson A; Begley DW; Lau C; Varani G, A small-molecule probe induces a conformation in HIV TAR RNA capable of binding drug-like fragments. *J Mol Biol* 2011, 410 (5), 984–96. [PubMed: 21763501]
49. Du Z; Lind KE; James TL, Structure of TAR RNA Complexed with a Tat-TAR Interaction Nanomolar Inhibitor that Was Identified by Computational Screening. *Chemistry & Biology* 2002, 9 (6), 707–712. [PubMed: 12079782]
50. Davidson A; Leeper TC; Athanassiou Z; Patora-Komisarska K; Karn J; Robinson JA; Varani G, Simultaneous recognition of HIV-1 TAR RNA bulge and loop sequences by cyclic peptide mimics of Tat protein. *Proc Natl Acad Sci U S A* 2009, 106 (29), 11931–6. [PubMed: 19584251]
51. Davidson A; Patora-Komisarska K; Robinson JA; Varani G, Essential structural requirements for specific recognition of HIV TAR RNA by peptide mimetics of Tat protein. *Nucleic Acids Res* 2011, 39 (1), 248–56. [PubMed: 20724442]
52. Borkar AN; Bardaro MF Jr.; Camilloni C; Aprile FA; Varani G; Vendruscolo M, Structure of a low-population binding intermediate in protein-RNA recognition. *Proc Natl Acad Sci U S A* 2016, 113 (26), 7171–6. [PubMed: 27286828]
53. Pham VV; Salguero C; Khan SN; Meagher JL; Brown WC; Humbert N; de Rocquigny H; Smith JL; D'Souza VM, HIV-1 Tat interactions with cellular 7SK and viral TAR RNAs identifies dual structural mimicry. *Nat Commun* 2018, 9 (1), 4266. [PubMed: 30323330]
54. Shortridge MD; Wille PT; Jones AN; Davidson A; Bogdanovic J; Arts E; Karn J; Robinson JA; Varani G, An ultra-high affinity ligand of HIV-1 TAR reveals the RNA structure recognized by P-TEFb. *Nucleic Acids Res* 2019, 47 (3), 1523–1531. [PubMed: 30481318]
55. Aboul-ela F; Karn J; Varani G, The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J Mol Biol* 1995, 253 (2), 313–32. [PubMed: 7563092]
56. Aboul-ela F; Karn J; Varani G, Structure of HIV-1 TAR RNA in the absence of ligands reveals a novel conformation of the trinucleotide bulge. *Nucleic Acids Res* 1996, 24 (20), 3974–81. [PubMed: 8918800]
57. Leeper TC; Athanassiou Z; Dias RL; Robinson JA; Varani G, TAR RNA recognition by a cyclic peptidomimetic of Tat protein. *Biochemistry* 2005, 44 (37), 12362–72. [PubMed: 16156649]

58. Dibrov SM; Ding K; Brunn ND; Parker MA; Bergdahl BM; Wyles DL; 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 (14), 5223–8. [PubMed: 22431596]
59. Paulsen RB; Seth PP; Swayze EE; Griffey RH; Skalicky JJ; Cheatham TE 3rd; Davis DR, Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proc Natl Acad Sci U S A* 2010, 107 (16), 7263–8. [PubMed: 20360559]
60. Zhao Q; Han Q; Kissinger CR; Hermann T; Thompson PA, Structure of hepatitis C virus IRES subdomain IIa. *Acta Crystallogr D Biol Crystallogr* 2008, 64 (Pt 4), 436–43. [PubMed: 18391410]
61. Lukavsky PJ; Kim I; Otto GA; Puglisi JD, Structure of HCV IRES domain II determined by NMR. *Nat Struct Biol* 2003, 10 (12), 1033–8. [PubMed: 14578934]
62. Locker N; Easton LE; Lukavsky PJ, HCV and CSFV IRES domain II mediate eIF2 release during 80S ribosome assembly. *EMBO J* 2007, 26 (3), 795–805. [PubMed: 17255934]
63. Dibrov SM; Johnston-Cox H; Weng YH; Hermann T, Functional architecture of HCV IRES domain II stabilized by divalent metal ions in the crystal and in solution. *Angew Chem Int Ed Engl* 2007, 46 (1–2), 226–9. [PubMed: 17131443]
64. Lee MK; Bottini A; Kim M; Bardaro MF Jr.; Zhang Z; Pellicchia M; Choi BS; Varani G, A novel small-molecule binds to the influenza A virus RNA promoter and inhibits viral replication. *Chem Commun (Camb)* 2014, 50 (3), 368–70. [PubMed: 24247110]
65. Bae SH; Cheong HK; Lee JH; Cheong C; Kainosho M; Choi BS, Structural features of an influenza virus promoter and their implications for viral RNA synthesis. *Proc Natl Acad Sci U S A* 2001, 98 (19), 10602–7. [PubMed: 11553808]
66. Lee MK; Bae SH; Park CJ; Cheong HK; Cheong C; Choi BS, A single-nucleotide natural variation (U4 to C4) in an influenza A virus promoter exhibits a large structural change: implications for differential viral RNA synthesis by RNA-dependent RNA polymerase. *Nucleic Acids Res* 2003, 31 (4), 1216–23. [PubMed: 12582241]
67. Jo S; Kim T; Iyer VG; Im W, CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* 2008, 29 (11), 1859–65. [PubMed: 18351591]
68. Abraham MJ; Murtola T; Schulz R; Páll S; Smith JC; Hess B; Lindahl E, GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* 2015, 1–2, 19–25.
69. Denning EJ; Priyakumar UD; Nilsson L; Mackerell AD Jr., Impact of 2'-hydroxyl sampling on the conformational properties of RNA: update of the CHARMM all-atom additive force field for RNA. *J Comput Chem* 2011, 32 (9), 1929–43. [PubMed: 21469161]
70. Hart K; Foloppe N; Baker CM; Denning EJ; Nilsson L; Mackerell AD Jr., Optimization of the CHARMM additive force field for DNA: Improved treatment of the BI/BII conformational equilibrium. *J Chem Theory Comput* 2012, 8 (1), 348–362. [PubMed: 22368531]
71. Vanommeslaeghe K; Hatcher E; Acharya C; Kundu S; Zhong S; Shim J; Darian E; Guvench O; Lopes P; Vorobyov I; Mackerell AD Jr., CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J Comput Chem* 2010, 31 (4), 671–90. [PubMed: 19575467]
72. Durell SR; Brooks BR; Ben-Naim A, Solvent-Induced Forces between Two Hydrophilic Groups. *The Journal of Physical Chemistry* 2002, 98 (8), 2198–2202.
73. Jorgensen WL; Chandrasekhar J; Madura JD; Impey RW; Klein ML, Comparison of simple potential functions for simulating liquid water. *The Journal of Chemical Physics* 1983, 79 (2), 926–935.
74. Lemkul JA; Lakkaraju SK; MacKerell AD Jr., Characterization of Mg(2+) Distributions around RNA in Solution. *ACS Omega* 2016, 1 (4), 680–688. [PubMed: 27819065]
75. MacKerell AD Jr.; Jo S; Lakkaraju SK; Lind C; Yu W, Identification and characterization of fragment binding sites for allosteric ligand design using the site identification by ligand competitive saturation hotspots approach (SILCS-Hotspots). *Biochim Biophys Acta Gen Subj* 2020, 1864 (4), 129519. [PubMed: 31911242]

76. O'Reilly M; Cleasby A; Davies TG; Hall RJ; Ludlow RF; Murray CW; Tisi D; Jhoti H, Crystallographic screening using ultra-low-molecular-weight ligands to guide drug design. *Drug Discov Today* 2019, 24 (5), 1081–1086. [PubMed: 30878562]
77. Vitreschak AG; Rodionov DA; Mironov AA; Gelfand MS, Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation. *Nucleic Acids Res* 2002, 30 (14), 3141–51. [PubMed: 12136096]
78. Winkler W; Nahvi A; Breaker RR, Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 2002, 419 (6910), 952–6. [PubMed: 12410317]
79. Berkhout B; Silverman RH; Jeang K-T, Tat trans-activates the human immunodeficiency virus through a nascent RNA target. *Cell* 1989, 59 (2), 273–282. [PubMed: 2478293]
80. Hamy F; Felder ER; Heizmann G; Lazdins J; Aboul-ela F; Varani G; Karn J; Klimkait T, An inhibitor of the Tat/TAR RNA interaction that effectively suppresses HIV-1 replication. *Proc Natl Acad Sci U S A* 1997, 94 (8), 3548–53. [PubMed: 9108013]
81. Seth PP; Miyaji A; Jefferson EA; Sannes-Lowery KA; Osgood SA; Propp SS; Ranken R; Massire C; Sampath R; Ecker DJ; Swayze EE; Griffey RH, 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 (23), 7099–102. [PubMed: 16279767]
82. Bottini A; De SK; Wu B; Tang C; Varani G; Pellicchia M, Targeting Influenza A Virus RNA Promoter. *Chemical Biology & Drug Design* 2015, 86 (4), 663–673. [PubMed: 25676805]
83. Morgan BS; Sanaba BG; Donlic A; Karloff DB; Forte JE; Zhang Y; Hargrove AE, R-BIND: An Interactive Database for Exploring and Developing RNA-Targeted Chemical Probes. *ACS Chem Biol* 2019, 14 (12), 2691–2700. [PubMed: 31589399]
84. Disney MD; Winkelsas AM; Velagapudi SP; Southern M; Fallahi M; Childs-Disney JL, Inforna 2.0: A Platform for the Sequence-Based Design of Small Molecules Targeting Structured RNAs. *ACS Chem Biol* 2016, 11 (6), 1720–8. [PubMed: 27097021]
85. Kumar Mishra S; Kumar A, NALDB: nucleic acid ligand database for small molecules targeting nucleic acid. *Database (Oxford)* 2016, 2016.
86. Mehta A; Sonam S; Gouri I; Loharch S; Sharma DK; Parkesh R, SMMRNA: a database of small molecule modulators of RNA. *Nucleic Acids Res* 2014, 42 (Database issue), D132–41. [PubMed: 24163098]
87. Goel H; Hazel A; Yu W; Jo S; MacKerell AD Jr., Application of Site-Identification by Ligand Competitive Saturation in Computer-Aided Drug Design. *New J Chem* 2022, 46 (3), 919–932.
88. Mousaei M; Kudaibergenova M; MacKerell AD Jr.; Noskov S, Assessing hERG1 Blockade from Bayesian Machine-Learning-Optimized Site Identification by Ligand Competitive Saturation Simulations. *J Chem Inf Model* 2020, 60 (12), 6489–6501. [PubMed: 33196188]
89. Woodson SA, Metal ions and RNA folding: a highly charged topic with a dynamic future. *Curr Opin Chem Biol* 2005, 9 (2), 104–9. [PubMed: 15811793]
90. Draper DE; Grilley D; Soto AM, Ions and RNA folding. *Annu Rev Biophys Biomol Struct* 2005, 34 (1), 221–43. [PubMed: 15869389]
91. Nguyen HT; Hori N; Thirumalai D, Theory and simulations for RNA folding in mixtures of monovalent and divalent cations. *Proc Natl Acad Sci U S A* 2019, 116 (42), 21022–21030. [PubMed: 31570624]
92. Manning GS, The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q Rev Biophys* 1978, 11 (2), 179–246. [PubMed: 353876]
93. Draper DE, Folding of RNA tertiary structure: Linkages between backbone phosphates, ions, and water. *Biopolymers* 2013, 99 (12), 1105–13. [PubMed: 23568785]
94. Pyle AM, Ribozymes: a distinct class of metalloenzymes. *Science* 1993, 261 (5122), 709–14. [PubMed: 7688142]
95. Kognole AA; MacKerell AD Jr., Contributions and competition of Mg(2+) and K(+) in folding and stabilization of the Twister ribozyme. *RNA* 2020, 26 (11), 1704–1715. [PubMed: 32769092]
96. Draper DE, A guide to ions and RNA structure. *RNA* 2004, 10 (3), 335–43. [PubMed: 14970378]
97. Shelley JC; Patey GN, A configuration bias Monte Carlo method for water. *The Journal of Chemical Physics* 1995, 102 (19), 7656–7663.

98. Woo HJ; Dinner AR; Roux B, Grand canonical Monte Carlo simulations of water in protein environments. *J Chem Phys* 2004, 121 (13), 6392–400. [PubMed: 15446937]
99. Raman EP; Yu W; Lakkaraju SK; MacKerell AD Jr., Inclusion of multiple fragment types in the site identification by ligand competitive saturation (SILCS) approach. *J Chem Inf Model* 2013, 53 (12), 3384–98. [PubMed: 24245913]
100. Sheng J; Gan J; Huang Z, Structure-based DNA-targeting strategies with small molecule ligands for drug discovery. *Med Res Rev* 2013, 33 (5), 1119–73. [PubMed: 23633219]
101. Hermann T, Strategies for the Design of Drugs Targeting RNA and RNA-Protein Complexes. *Angew Chem Int Ed Engl* 2000, 39 (11), 1890–1904. [PubMed: 10940979]
102. Batey RT; Rambo RP; Doudna JA, Tertiary Motifs in RNA Structure and Folding. *Angew Chem Int Ed Engl* 1999, 38 (16), 2326–2343. [PubMed: 10458781]
103. Hamilton PL; Arya DP, Natural product DNA major groove binders. *Nat Prod Rep* 2012, 29 (2), 134–43. [PubMed: 22183179]
104. Chen SJ, RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* 2008, 37 (1), 197–214. [PubMed: 18573079]
105. Arya DP, Aminoglycoside–Nucleic Acid Interactions: The Case for Neomycin. In *DNA Binders and Related Subjects*, Waring MJ; Chaires JB, Eds. Springer Berlin Heidelberg: Berlin, Heidelberg, 2005; pp 149–178.
106. Long KS; Crothers DM, Characterization of the solution conformations of unbound and Tat peptide-bound forms of HIV-1 TAR RNA. *Biochemistry* 1999, 38 (31), 10059–69. [PubMed: 10433713]
107. Molecular Operating Environment (MOE) 2018.0101 ed.; Chemical Computing Group ULC, 1010 Shrebrooke St. West, Suite #910, Montreal, QC, Canada H3A 2R7, 2022.
108. G. A.A; da Silva GF; Lakkaraju SK; Guimaraes BG; MacKerell AD Jr.; Magalhaes MLB, Insights into Glucose-6-phosphate Allosteric Activation of beta-Glucosidase A. *J Chem Inf Model* 2021, 61 (4), 1931–1941. [PubMed: 33819021]
109. Yu AM; Choi YH; Tu MJ, RNA Drugs and RNA Targets for Small Molecules: Principles, Progress, and Challenges. *Pharmacol Rev* 2020, 72 (4), 862–898. [PubMed: 32929000]
110. Mattick JS, The State of Long Non-Coding RNA Biology. *Noncoding RNA* 2018, 4 (3).
111. Deveson IW; Hardwick SA; Mercer TR; Mattick JS, The Dimensions, Dynamics, and Relevance of the Mammalian Noncoding Transcriptome. *Trends Genet* 2017, 33 (7), 464–478. [PubMed: 28535931]
112. Korb O; Olsson TS; Bowden SJ; Hall RJ; Verdonk ML; Liebeschuetz JW; Cole JC, Potential and limitations of ensemble docking. *J Chem Inf Model* 2012, 52 (5), 1262–74. [PubMed: 22482774]
113. Amaro RE; Baudry J; Chodera J; Demir O; McCammon JA; Miao Y; Smith JC, Ensemble Docking in Drug Discovery. *Biophys J* 2018, 114 (10), 2271–2278. [PubMed: 29606412]
114. Chen H; Meisburger SP; Pabit SA; Sutton JL; Webb WW; Pollack L, Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc Natl Acad Sci U S A* 2012, 109 (3), 799–804. [PubMed: 22203973]
115. De Costa NT; Heemstra JM, Evaluating the effect of ionic strength on duplex stability for PNA having negatively or positively charged side chains. *PLoS One* 2013, 8 (3), e58670. [PubMed: 23484047]
116. Aytenfisu AH; Spasic A; Grossfield A; Stern HA; Mathews DH, Revised RNA Dihedral Parameters for the Amber Force Field Improve RNA Molecular Dynamics. *J Chem Theory Comput* 2017, 13 (2), 900–915. [PubMed: 28048939]
117. Lemkul JA; MacKerell AD Jr., Polarizable force field for RNA based on the classical drude oscillator. *J Comput Chem* 2018, 39 (32), 2624–2646. [PubMed: 30515902]
118. Lemkul JA; Huang J; Roux B; MacKerell AD Jr., An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem Rev* 2016, 116 (9), 4983–5013. [PubMed: 26815602]
119. Zhang C; Lu C; Jing Z; Wu C; Piquemal JP; Ponder JW; Ren P, AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J Chem Theory Comput* 2018, 14 (4), 2084–2108. [PubMed: 29438622]

120. Lemkul JA; MacKerell AD Jr., Balancing the Interactions of Mg(2+) in Aqueous Solution and with Nucleic Acid Moieties For a Polarizable Force Field Based on the Classical Drude Oscillator Model. *J Phys Chem B* 2016, 120 (44), 11436–11448. [PubMed: 27759379]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

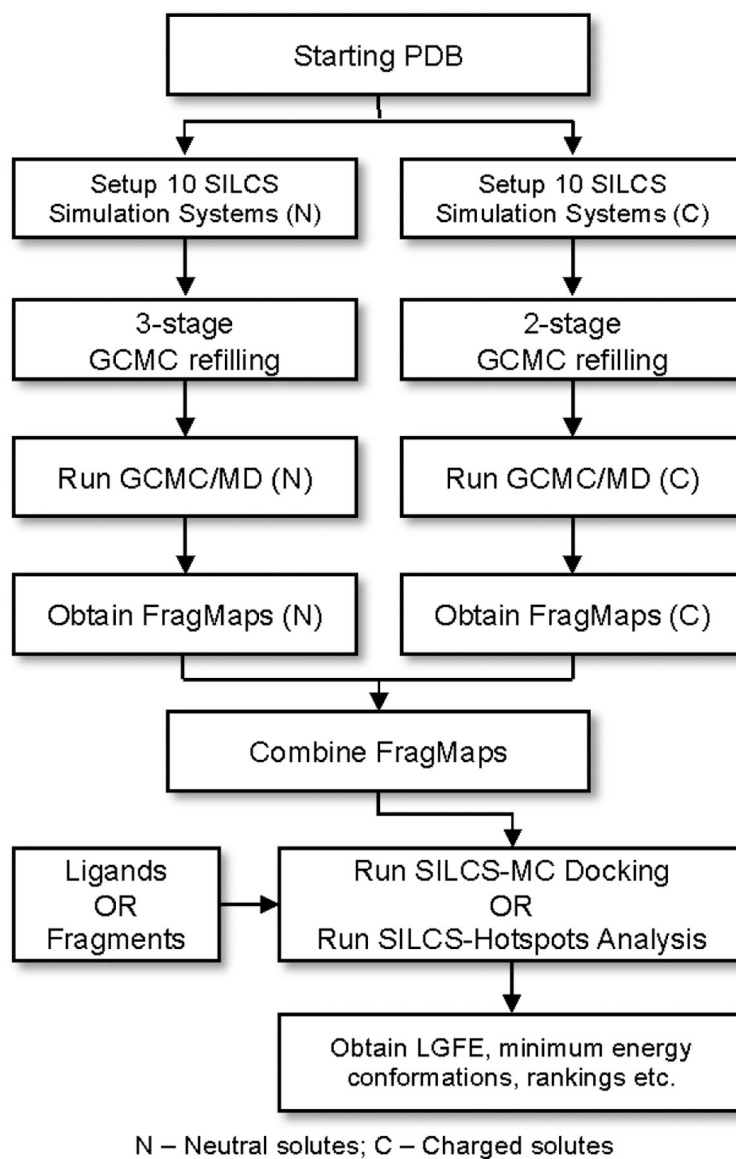


Figure 1. Process flow diagram for the SILCS-RNA protocol. Three-stage GCMC refilling is applied to the neutral (N) solute SILCS simulations, and two-stage filling is applied to the charged (C) solute SILCS simulations.



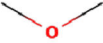
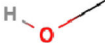

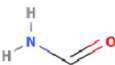


		SILCS type	Atom location	Major functional group represented
Neutral	 Benzene (BENX)	BENC	6 C on BENX	C atoms in aromatic ring
	 Propane (PRPX)	PRPC	3 C on PRPX	Aliphatic C atoms
	 Dimethyl ether (DMEE)	DMEO	O on DMEE	HB acceptor O atom from ethers
	 Methanol (MEOH)	MEOO	O on MEOH	HB donor and acceptor O atom from hydroxyls
	 Imidazole (IMIA)	IMIN	N on IMIA	HB acceptor N atom in planar rings
	 Formamide (FORM)	IMINH	N(H) on IMIA	HB donor N atom in rings
		GEHC	3 C on IMIA	Generic heterocycle carbon
		FORN	N on FORM	HB donor N atom in amides
Charged		FORO	O on FORM	HB acceptor O atom in carbonyl group
		FORC	C on FORM	Carbonyl C atom in aldehydes
	 Methyl ammonium (MAMY)	MAMN	N on MAMY	Positively charged N atom
	 Acetate (ACEY)	MAMC	C on MAMY	Conjugated C atom in cation
		ACEO	2 O on ACEY	Negatively charged O atom in carboxylates, anions etc
		ACEC	C on ACEY	Carbonyl C atom in carboxylates, P atom in phosphates etc

Figure 2. Neutral and charged solutes used in the SILCS-RNA approach. On the right, a description of different atoms from the solutes and the major functional groups represented by the respective atoms as used to define the FragMaps. The MAMC and ACEO maps are not used during the SILCS-MC docking to avoid overcounting. MAMC labeled atoms actually use the MAMN maps in specific cases with connected N⁺ atoms classified as NCLA (e.g., tpp001). ACEO maps are only used for visualization.

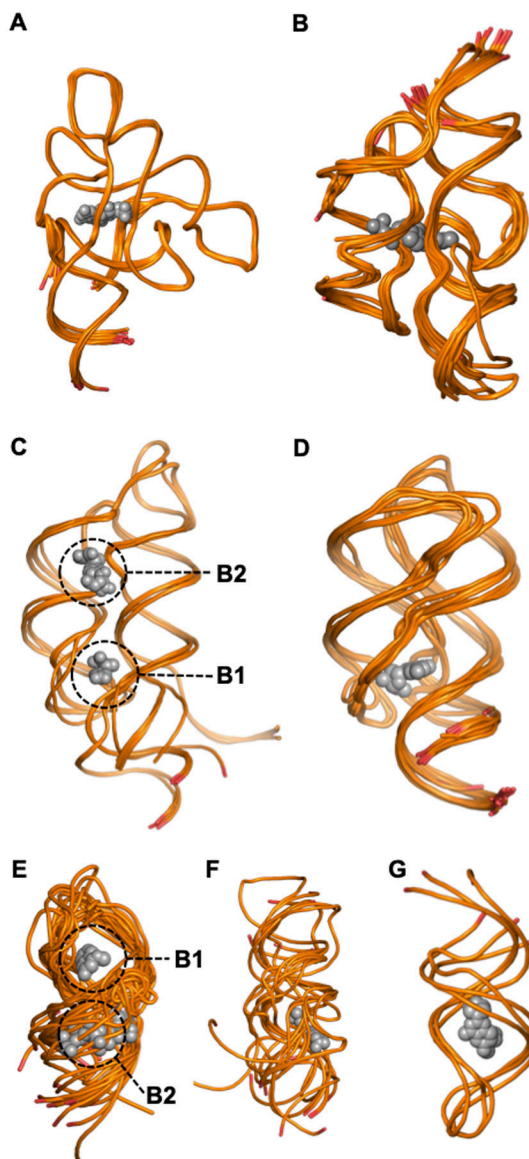


Figure 3. Conformational diversity on RNAs based on the available experimental crystal structures (Table 1). Gray spheres represent the experimentally identified ligands on each RNA. (A) FMN riboswitch structures aligned to PDB 5KX9. (B) TPP riboswitch structures aligned to PDB 2GDI. (C) THF riboswitch structures aligned to PDB 4LVV. (D) dG riboswitch structures aligned to PDB 3SKI. (E) HIV-1 TAR RNA structures aligned to PDB 6CMN. (F) HCV IRES IIa structures aligned to PDB 3TZR. (G) IAV promoter structures aligned to PDB 2LWK.

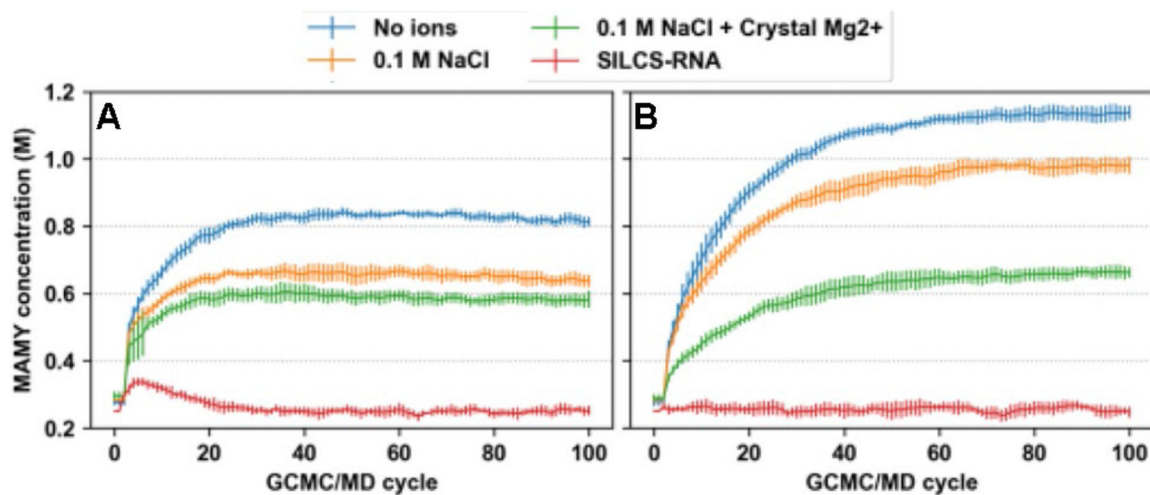


Figure 4. Average number of MAMY fragments present in the SILCS simulations as a function of the GCMC-MD cycles for (A) the FMN riboswitch and (B) the HCV-IRES-IIa RNA under different ion-atmosphere conditions. The error bars represent the standard deviation across 10 independent runs.

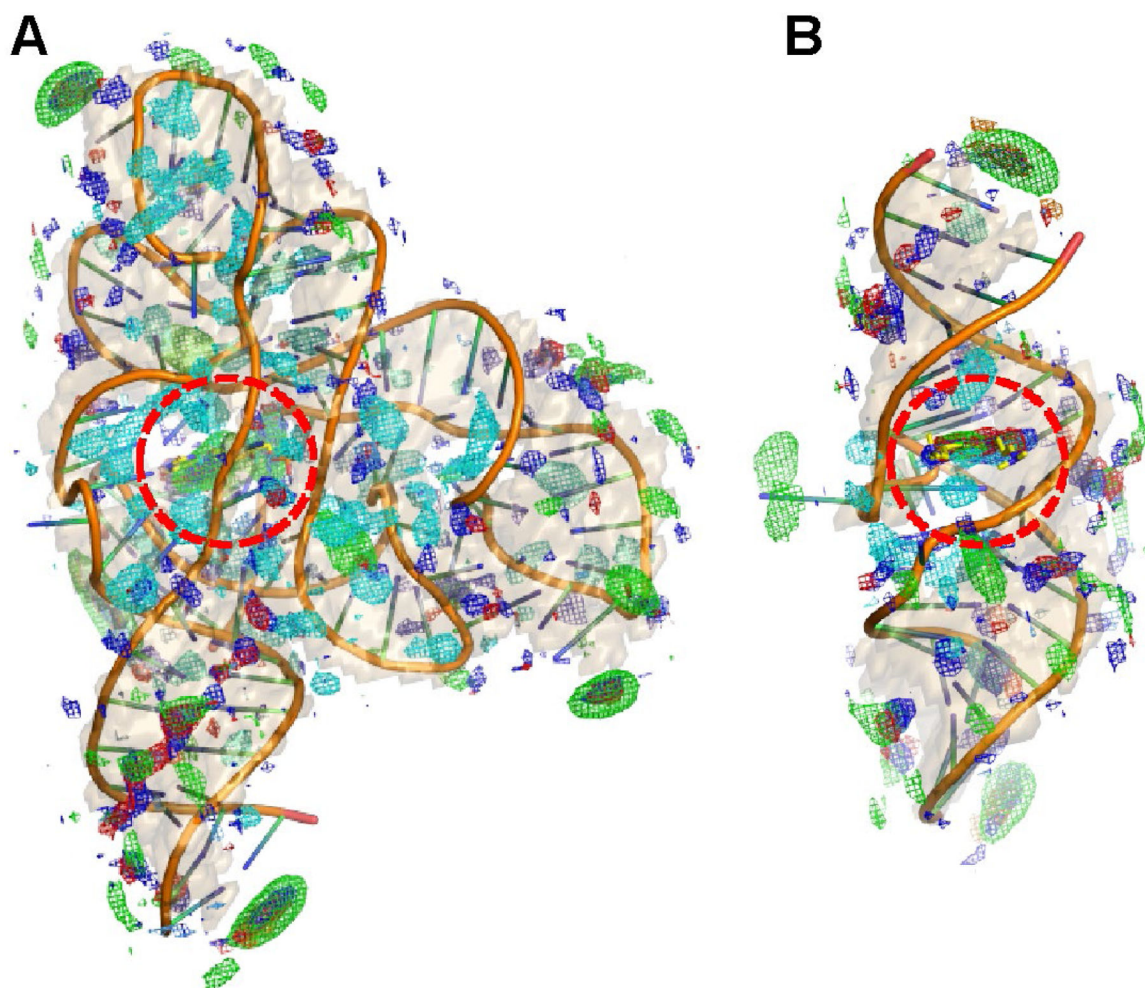


Figure 5. SILCS FragMaps for (A) FMN riboswitch (B) HCV-IRES-IIa RNA. The FragMaps are shown as colored mesh: generic apolar (light green, GFE level -0.5 kcal/mol), generic donor (blue, GFE level -0.5 kcal/mol), generic acceptor (red, GFE level -0.5 kcal/mol), methanol oxygen (light blue, GFE level -0.5 kcal/mol) and methylammonium ion (cyan, GFE level -2.0 kcal/mol). The sand-colored transparent surface represents the SILCS exclusion map. The flavin mononucleotide ligand (fmn001) and ISIS-11-b ligand (hcv001) are shown in the binding sites of the respective RNAs (yellow sticks) identified with red dotted circles.

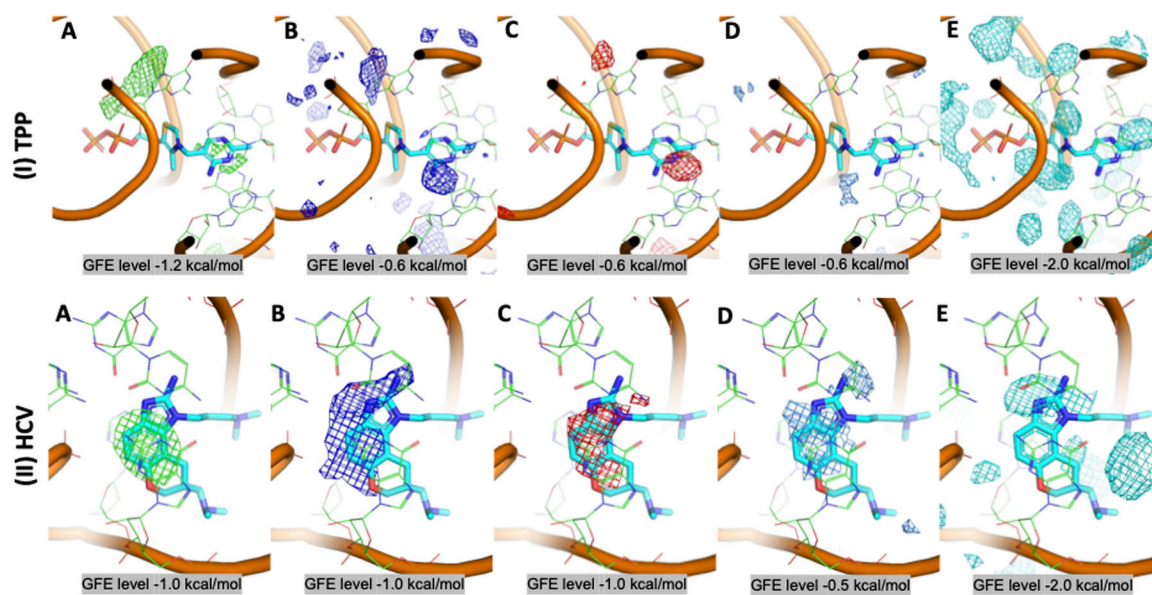


Figure 6. RNA intercalation site SILCS FragMaps for (I) TPP riboswitch and (II) HCV-IRES-IIa RNA. (A) Generic apolar maps in light green mesh, (B) generic donor maps in blue mesh, (C) generic acceptor maps in red mesh, (D) alcohol (MEOO) maps in light blue mesh, and (E) positive methylammonium nitrogen (MAMN) maps in cyan mesh with negative acetate oxygen (ACEO) maps in orange mesh.

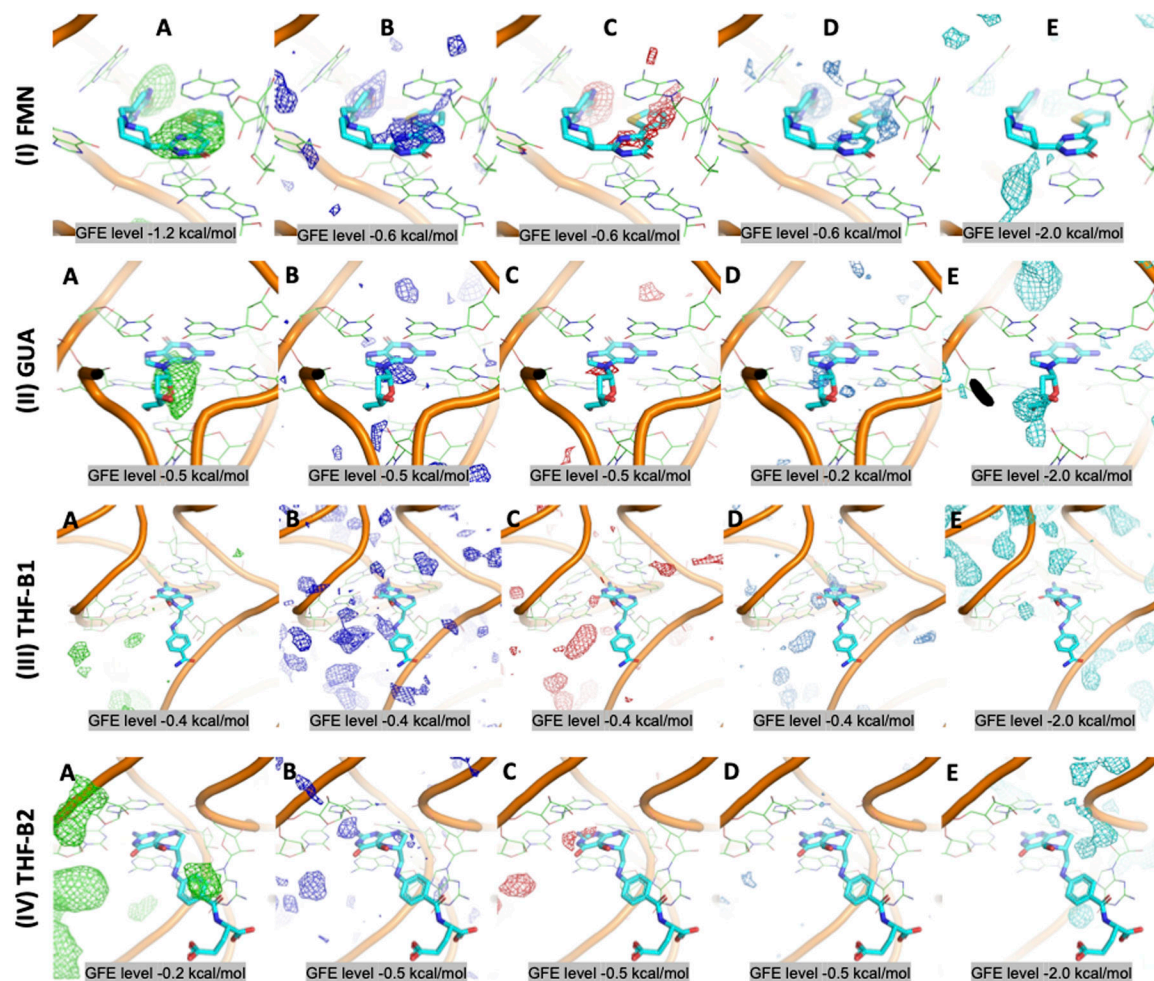


Figure 7. RNA structure pocket site SILCS FragMaps for (I) FMN riboswitch, (II) GUA riboswitch, and (III/IV) THF riboswitch binding sites B1 and B2, respectively. (A) Generic apolar maps in light green mesh, (B) generic donor maps in blue mesh, (C) generic acceptor maps in red mesh, (D) alcohol (MEOO) maps in light blue mesh, and (E) positive methylammonium nitrogen (MAMN) maps in cyan mesh with negative acetate oxygen (ACEO) maps in orange mesh.

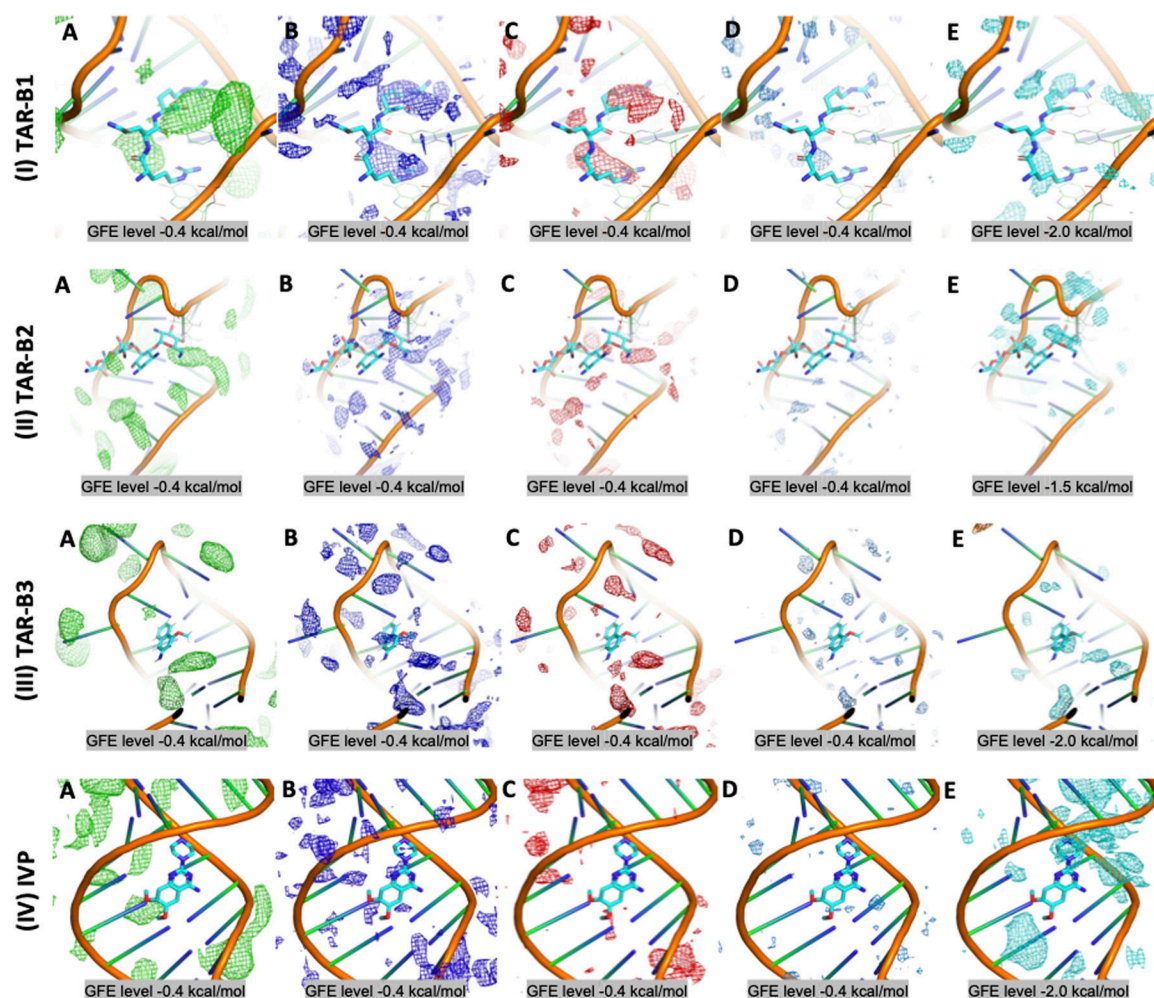


Figure 8.

RNA minor and major groove site SILCS FragMaps for the HIV1-TAR sites (I) B1 with ligand rbt203(tar008), (II) B2 with ligand neomycin-B(tar010), (III) B3 with ligand mv2003(tar012) and (IV) for the IAV RNA promotor. (A) Generic apolar maps in light green mesh, (B) generic donor maps in blue mesh, (C) generic acceptor maps in red mesh, (D) alcohol (MEOO) maps in light blue mesh, and (E) positive methylammonium nitrogen (MAMN) maps in cyan mesh with negative acetate oxygen (ACEO) maps in orange mesh. RNA molecules are shown in cartoon form and ligand molecules are shown as sticks.

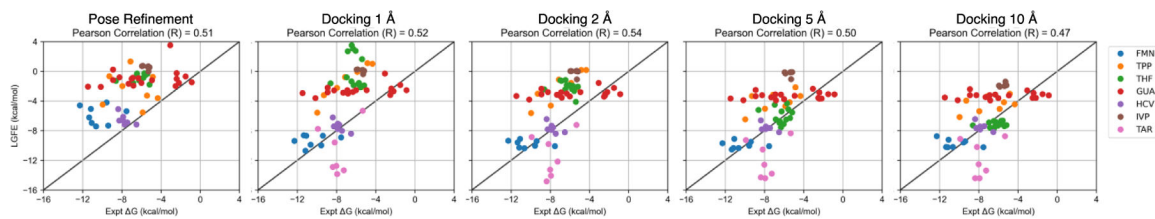
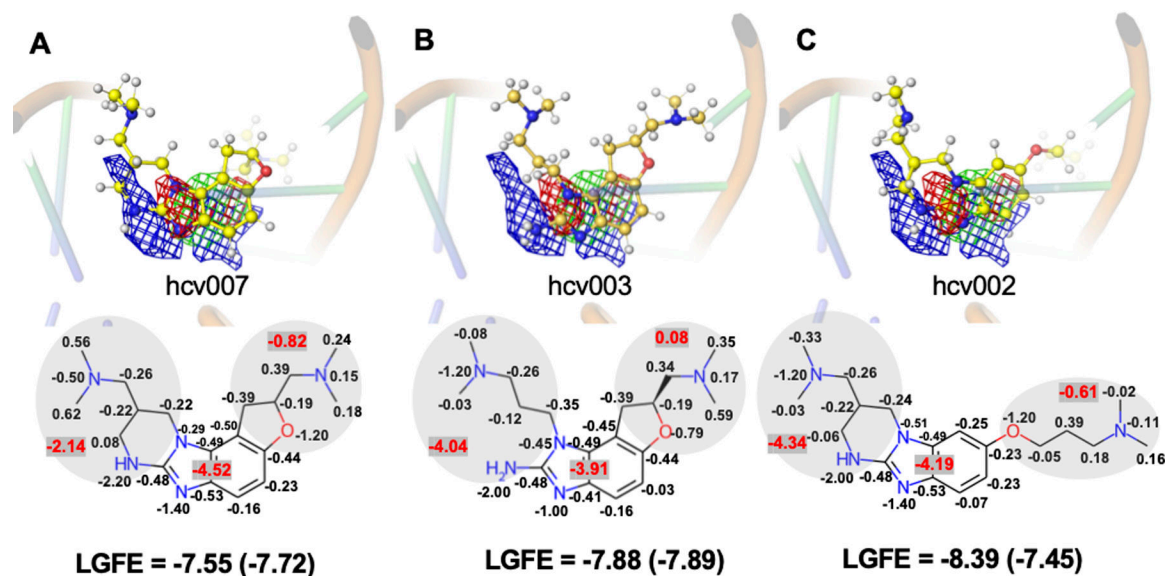


Figure 9. Correlation between experimental G_{bind} and LGFE scores predicted using SILCS-MC calculations with generic ACS (GAS21) for all of the RNA targets.

**Figure 10:**

Atomic and group GFE contributions to the total LGFE for benzimidazole derivatives shown to bind HCV-IRES-IIa RNA. (A) hcv007, (B) hcv003, (C) hcv002. The FragMaps in the top panels show generic apolar (light green mesh), generic donor (blue mesh), and generic acceptor (red mesh) maps at GFE level -1.2 kcal/mol. The ligand atoms carbon, nitrogen, oxygen, and hydrogen are shown as yellow, blue, red, and white colored spheres, respectively. The HCV RNA is shown in transparent cartoon form. SILCS-MC docking was performed with the GAS21 using a 2 \AA radius. Values in parentheses refer to experimental binding affinities.

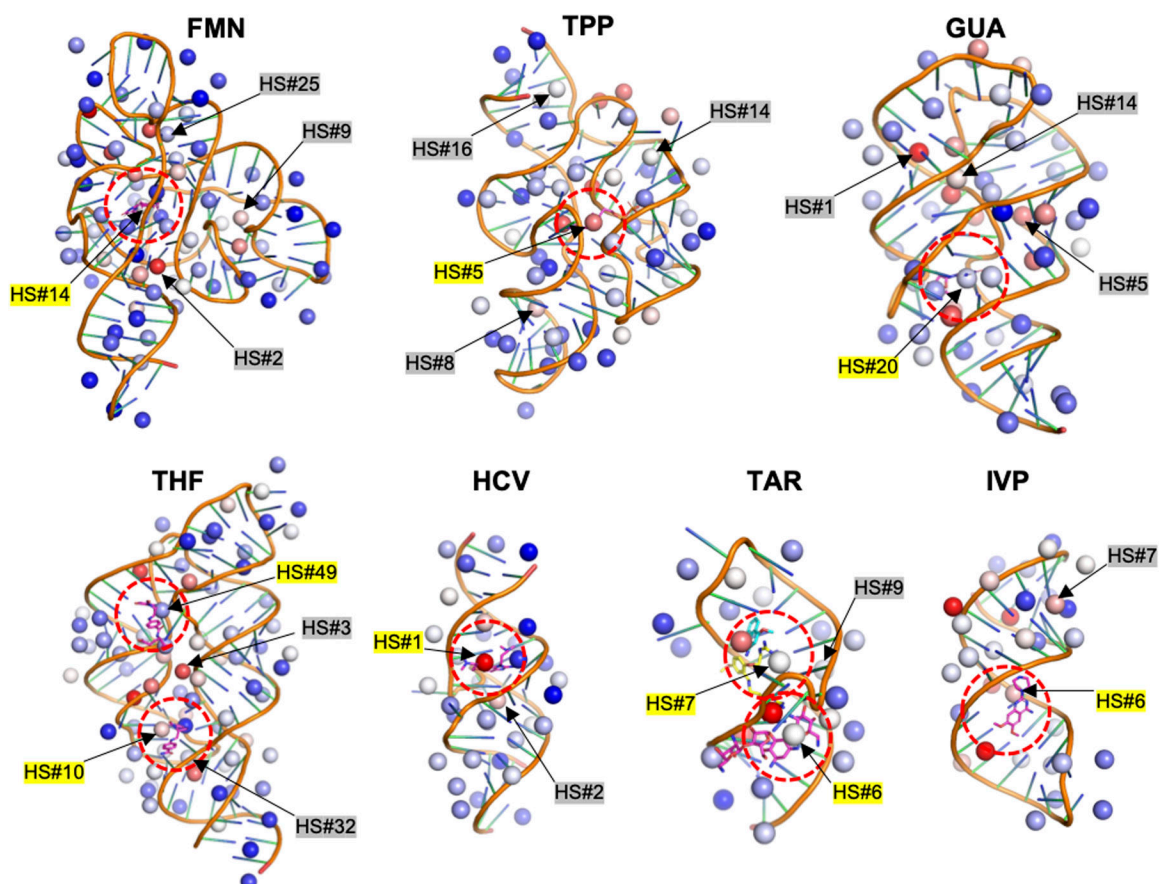


Figure 11. SILCS-Hotspots analysis for each RNA with the ASTEX mini fragment database. RNA is shown in cartoon form, the ligands from experimental structures are shown as sticks, and the hotspot locations are shown as spheres. The color scheme for the hotspots corresponds to a red–white–blue spectrum with red being highly favorable and blue being less favorable. The dashed red circles highlight the hotspots in the space around the known binding sites for the respective RNAs. The hotspots (HSs) selected for docking of FDA approved compounds are pointed out with hotspots in known binding sites highlighted with yellow.

Table 1.RNA Molecules Selected for This Study and Available Structural Data^a

Name (Short code)	# of nucleotides	PDB IDs available	# of ligands available ^a
FMN Riboswitch (FMN)	110	5KX9 ²⁷ , 3F4E ²⁸ , 3F4G, 3F4H, 3F2W, 3F2X, 3F2Y, 3F2Q, 3F2T, 3F30, 6DN1 ²⁹ , 6DN2, 6DN3, 5C45 ³⁰ , 6BFB ¹⁰ , 2YIE ³¹	13 (9,8)
TPP Riboswitch (TPP)	80	2GDI ³² , 2HOM ³³ , 2HOJ, 2HOK, 2HOL, 2HOO, 2HOP, 4NYA ³⁴ , 4NYB, 4NYC, 4NYD, 4NYG, 3K0J ³⁵ , 3D2G ³⁶ , 3D2V, 3D2X	11 (11,8)
THF Riboswitch (THF)	89	4LVV ³⁷ , 4LVW, 4LVX, 4LVY, 4LVZ, 4LW0, 3SUH, 3SUX ³⁸ , 3SUY, 3SD3, 7KD1 ³⁹	14 (7,14)
2'-Deoxyguanosine Riboswitch (GUA)	68	3SKI ⁴⁰ , 3SKL, 3SKR, 3SKT, 3SKW, 3SKZ, 3SLM, 3SLQ, 3RKF ⁴¹ , 6UC7 ⁴² , 6UBU, 6UC9, 6UC8, 3GAO ⁴³ , 3GOT, 3FO6, 3GER, 3GES, 3FO4, 3GOG, 3G4M	24 (14,19)
HIV-1 TAR RNA (TAR)	27	6CMN ⁴⁴ , 1UUI ⁴⁵ , 1UUD, 1UTS ⁴⁶ , 1QD3 ⁴⁷ , 2L8H ⁴⁸ , 1LVJ ⁴⁹ , 2KDQ ⁵⁰ , 2KX5 ⁵¹ , 5J2W ⁵² , 5J0M, 5J1O, 6MCE ⁵³ , 6D2U ⁵⁴ , 1ARJ ⁵⁵ , 1ANR ⁵⁶ , 2A9X ⁵⁷	30 (6,7)
HCV-IRES-IIa RNA (HCV)	36	3TZR ⁵⁸ , 2KU0 ⁵⁹ , 2KTZ, 2PN3 ⁶⁰ , 2PN4, 1P5M ⁶¹ , 2HUA ⁶² , 2NOK ⁶³	7 (2,7)
IAV RNA Promotor (IVP)	32	2LWK ⁶⁴ , 1JO7 ⁶⁵ , 1MFY ⁶⁶	6 (1,6)

^aThe PDB IDs in bold characters are the ones that were selected for SILCS-RNA calculations.^bValues in parenthesis represent number of ligands with structural data available and the number of ligands for which experimental binding affinities are available.

Table 2.Details of the SILCS-MC Protocols Applied in This Study^a

MC protocol	Initial ligand placement	n _{CY}	n _{MC}	dX	dθ	dφ	n _{SA}	dX _{SA}	dθ _{SA}	dφ _{SA}
pose refinement	aligned	50	1000	0.5	15	180	1000	0.2	9	9
docking	random	250	10,000	1	180	180	40,000	0.2	9	9

^an_{CY}: number of cycles, n_{MC}: number of MC steps, dX: step size of translation, dθ: step size of rotation, dφ: step size of dihedral angle rotation, n_{SA}: number of SA steps, dX_{SA}, dθ_{SA}, dφ_{SA}.: steps in SA

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3.

Average RMSD of Ligands Relative to the Initial Binding Orientations after SILCS-MC Pose Refinement and Docking Calculations.

system	pose refinement		docking							
	GAS21	SS21	1 Å GAS21	1 Å SS21	2 Å GAS21	2 Å SS21	5 Å GAS21	5 Å SS21	10 Å GAS21	10 Å SS21
FMN	2.64	2.50	6.15	6.26	5.94	5.91	6.01	6.24	6.05	6.20
TPP	2.36	3.71	6.78	7.60	6.72	7.76	4.87	8.70	5.79	8.55
THF	2.00	2.31	6.30	6.52	5.58	6.60	6.45	6.90	6.69	7.35
GUA	2.77	2.61	4.52	4.40	4.50	4.38	4.50	4.75	5.30	4.86
HCV	2.81	2.80	2.66	4.10	2.71	3.99	2.71	4.15	2.69	4.62
IVP	4.57	4.54	7.03	6.71	7.63	7.63	6.56	6.53	10.95	9.61
avg	2.86	3.08	5.57	5.93	5.51	6.04	5.18	6.21	6.25	6.87
st-dev	0.89	0.87	1.68	1.38	1.73	1.60	1.47	1.62	2.69	2.01

Table 4.

Average LGFE Scores and Relative Solvent Accessible Surface Areas (rSASAs) of the Top 20 FDA Approved Ligands Based on the LGFE Scores.^a

system	hotspot site no.	avg. LGFE	rSASA (%)
FMN	2	-0.17	63.4
FMN	9	-4.40	89.2
FMN	14	-13.30	88.9
FMN	25	-2.75	84.6
TPP	5	-6.71	81.6
TPP	8	-5.88	71.0
TPP	14	-6.06	79.7
TPP	16	-4.04	49.9
THF	3	-1.46	67.6
THF	10	-7.23	55.9
THF	32	-2.65	87.4
THF	49	-3.92	65.3
GUA	1	-3.48	73.4
GUA	5	-3.62	72.5
GUA	14	-7.46	82.2
GUA	20	-6.13	93.1
TAR	6	-7.48	53.9
TAR	7	-9.41	52.5
TAR	9	-7.94	42.9
HCV	1	-10.95	76.6
HCV	2	-5.62	85.0
IVP	6	-3.69	52.6
IVP	7	-8.79	47.2

^aExperimentally identified binding sites shown in bold text. rSASA is the solvent accessible surface area of each docked ligand in the presence and absence of the RNA presented as a percentage where 100% indicated the site to be fully occluded from the aqueous environment.