fpocketR: A platform for identification and analysis of ligand-binding pockets in RNA

Seth D. Veenbaas, Simon Felder, Kevin M. Weeks*

Department of Chemistry, University of North Carolina, Chapel Hill NC 27599-3290
* correspondence, weeks@unc.edu

Keywords: RNA-ligand interactions, RNA-targeted ligand discovery, RNA pockets, drug-likeness

Abstract

Small molecules that bind specific sites in RNAs hold promise for altering RNA function, manipulating gene expression, and expanding the scope of druggable targets beyond proteins. Identifying binding sites in RNA that can engage ligands with good physicochemical properties remains a significant challenge. *fpocketR* is a software and framework for identifying, characterizing, and visualizing ligand-binding sites in RNA. *fpocketR* was optimized, through a comprehensive analysis of currently available RNA-ligand complexes, to identify pockets in RNAs able to bind small molecules possessing favorable properties, generally termed drug-like. Here, we demonstrate multiple uses of *fpocketR* to analyze RNA-ligand interactions and novel pockets in small and large RNAs, to assess ensembles of RNA structure models, and to identify pockets in dynamic RNA systems. *fpocketR* performs best with RNA structures visualized at atomistic resolution but also provides useful information with lower resolution structures and computational models. *fpocketR* is a powerful, ligand-agnostic tool for discovery and analysis of targetable pockets in RNA molecules.

INTRODUCTION

RNA molecules are essential regulators of gene expression, template protein synthesis, and carry out diverse additional cellular functions.[1,2] As a result, RNAs lie upstream of nearly all biology. The vast scope of biological functions that can be modulated by altering RNA structure, translatability, or intermolecular interactions makes RNA an enticing target for small-molecule ligands.[3–8] The field has made progress in targeting RNA with a few successful human-devised small molecules, currently limited to linezolid, an antibiotic that binds the ribosome,[9] and risdiplam and branaplam, splicing modifiers that bind pre-messenger RNA.[10] Better tools for assessing RNA ligands and their binding sites would facilitate exploiting the full potential of targeting RNAs with small molecules.

Much like proteins, RNA molecules routinely fold to form (base-paired) secondary structures and occasionally form higher-order tertiary structures. Small molecules can preferentially engage pockets formed within these complex tertiary RNA structures.[4,11,12] However, physicochemical features of RNAs differ substantially from proteins: RNAs are generally more polar, more flexible, and more dynamic.[13–15] These differences complicate the direct transfer of protein-based tools and heuristics to RNA. In addition, relatively few, roughly only 50,[12] distinct RNA-ligand complexes have been visualized in which the ligand has good physicochemical properties (colloquially called drug-like). As a result of these two challenges, platforms for reliably identifying pockets in RNA that bind drug-like ligands are underdeveloped.

Foundational prior investigations have categorized small molecule interactions across diverse RNAs, based on comprehensive literature mining,[16,17] on large-scale experimental binding measurements,[18,19] on using available protein-focused software,[20,21] and on analysis of high-resolution structure databases.[21,22] These studies reinforce the idea that RNA motifs with complex structures are best able to form ligand-binding pockets, but have not focused on ligands with good physicochemical properties. Machine learning-based methods such as RLBind[23] and MultiModRLBP[24] were recently developed to predict the nucleotides in an RNA-ligand binding site. While powerful, these approaches typically do not return detailed geometric and physicochemical descriptors of complete binding pockets, limiting their ability to characterize ligand-binding environments in three dimensions. Finally, docking programs including AutoDock Vina,[25] Dock 6,[26] rDock,[27] and RLDOCK[28] have been widely applied to RNA.[29] Docking requires a candidate ligand(s) and usually relies on generic pocket-finding routines to define the search space. Docking program strengths lie in pose prediction and affinity estimation, but do not address the complementary challenge of identifying which RNAs and which pockets are inherently capable of binding drug-like ligands in a ligand-agnostic manner. Overall, assessing RNA-ligand

interactions is a rapidly growing area; nonetheless, characterization of RNA motifs specifically able to bind drug-like ligands remains incomplete.

We previously demonstrated that the open-source, geometry-based protein pocket finding software *fpocket*[30,31] could be adapted to identify RNA-binding pockets through parameter optimization. Our optimization was guided by a curated training library of RNAs in complex with drug-like ligands. In a benchmark, we were able to detect all known ligand-binding sites in a dataset of 32 RNA-ligand complexes, selected for containing drug-like ligands, and we improved the positive predictive value from 19% for *fpocket* to 78%.[12] We built on this optimization of *fpocket* and developed a pocket finding package specifically for RNA called *fpocketR*.[12] Several major conclusions, unique to this work, emerged from the original development of *fpocketR*.[12] First, *fpocketR* reliably detects pockets capable of binding ligands with good physicochemical properties. Second, *fpocketR* detected many novel, likely targetable, pockets in RNAs, an ability that was validated experimentally. Third, complex secondary structures, especially multi-helix junctions and pseudoknots, are an order of magnitude more likely to form pockets able to bind drug-like ligands than simpler motifs like RNA bulges and loops.

In this study, we explore multiple expanded uses for *fpocketR* by identifying pockets in large cryo-EM structures, computationally modeled RNA ensembles, and dynamic RNA systems. We also characterize the geometric features of RNA pockets for drug-like ligands. We anticipate that *fpocketR* will be a broadly useful ligand-agnostic method to identify and characterize ligand-binding pockets in diverse RNA molecules.

RESULTS AND DISCUSSION

fpocketR workflow. *fpocketR* is a straightforward and streamlined analysis pipeline for identifying, visualizing, and characterizing pockets and ligands in RNA structures via a well-documented command-line interface. The pipeline is executed with a single input containing a Protein Data Bank (PDB) accession code or a locally stored RNA structure file (**Fig. 1A**). *fpocketR* also accepts RNA secondary structure drawing templates in multiple formats, including NSD,[32] VARNA,[33] and R2DT.[34]

fpocketR runs the *fpocket* algorithm with optimized parameters,[12] characterizes pocket and ligand properties, ranks identified pockets by ligandability, and produces comprehensive output files. Outputs include a three-dimensional figure (and *pymol* session file) visualizing all pockets in the RNA tertiary structure, a detailed description of the properties of each pocket and ligand (**SI Table 2**), and a figure mapping the nucleotides that form each pocket onto an RNA secondary structure template (**Fig. 1B** and **SI Table 3**).

The behavior and features of *fpocketR* can be modified with optional arguments. A powerful feature of *fpocketR* enables analysis of multiple states in an RNA ensemble. Users can provide RNA ensembles from any source, including experimentally-determined structures, molecular dynamics simulations, or computational structure models. The multi-state analysis generates images that visualize the density of pocket formation in both tertiary and secondary structure formats (**Fig. 1C** and **SI Table 4**). Pocket finding can further be customized by manually adjusting pocket finding parameters and specifying which RNA chain(s) and ligand to analyze.

Identification of pockets in large RNAs. Most available non-ribosomal RNA-ligand complexes involve short RNAs (< 200 nt), and many are riboswitches or simple aptamers.[22] *fpocketR* was primarily trained on these small RNAs. However, application of cryo-electron microscopy (cryo-EM) has begun to expand the variety and sizes of solved RNA tertiary structures.[35–37] We therefore examined the ability of *fpocketR* to detect and characterize pockets in large RNA structures. For the following examples, we emphasize that *fpocketR* has been optimized to selectively find pockets capable of binding drug-like ligands.[12]

We identified pockets in two large, engineered RNA structures solved by cryo-EM. The first structure is a five-helix panel (544 nt), designed as an RNA origami scaffold (PDB 7ptq).[38] The second structure is a large RNA (374 nt), engineered to include the small-molecule aptamers for Broccoli and Pepper, that bind the ligands DFHBI-1T and HBC620, respectively, configured as a Förster resonance energy transfer pair (PDB 7zj4).[39] Between the two RNAs, *fpocketR* identified four novel pockets and both known pockets (**Fig. 2**). For the four novel pockets, two were formed by kissing-loop pseudoknots (PK), one was formed at the interface (IN) of two helices, and one was formed in the G-quadraplex (G4) of the Broccoli aptamer, adjacent to the DFHBI-1T binding site. The two known pockets overlapped exactly with the ligand binding sites for DFHBI-1T and HBC620. This analysis, first, demonstrates that *fpocketR* detects high-quality binding pockets across a diverse range of RNA sizes and structure classes, despite the limited scope of its training dataset. Second, this analysis reinforces the idea that pockets able to bind drug-like ligands tend to form selectively in regions of complex local tertiary structure (**Fig. 2**).[12]

Identification of pockets in low-resolution and modeled structures. Many RNA tertiary structures are solved at modest resolutions. Until recently,[37] RNA-only structures solved using cryo-EM had resolutions in the 4 to 10 Å range.[36] Modeling of RNA structure using physics-based or machine learning methods is rapidly improving, but still often exhibits large deviations from accepted

structures.[40,41] We investigated whether low-resolution and modeled structure ensembles can inform the ligandability of an RNA.

fpocketR detects one pocket in a high-resolution (2.3 Å) crystal structure of the class I type III preQ1 riboswitch (PDB 8fza),[42] correctly identifying the ligand binding site for the preQ1 ligand (**Fig. 3A**). We applied *fpocketR* to 153 models of the same riboswitch submitted to the CASP15 evaluation.[40] The models varied widely in accuracy, compared to the accepted structure, with root mean square deviations (RMSDs) between 2 and 28 Å and template modeling scores (TMscores) between 0.08 and 0.43. Of these models, 25% contained a pocket that overlapped the known preQ1 ligand binding site, 33% contained a pocket or pockets that did not overlap with the known ligand binding site, and 42% contained no pocket (**Fig. 3A**). We then compared the accuracy with which *fpocketR* detected the known ligand binding site in each model relative to the structural quality of the model. There is a modest correlation between pocket finding performance and the structural quality metrics RMSD and TMscore (**Fig. 3B**). The TMscore metric was the most informative, but still only modest, predictor of RNA ligandability, with one-half (48%) of the models with a TMscore over 0.26 containing a pocket overlapping the ligand binding site (**Fig. 3C**).

We previously showed that *fpocketR* identified the SAM ligand binding site in a subset of states for a ligand-free ensemble of the SAM-IV riboswitch, determined by cryo-EM (PDB 6wql).[12,43] We generated a 20-model ensemble of the SAM-IV riboswitch using trRosettaRNA, a successful machine-learning-based RNA modeling program,[40,44] and then identified pockets in the structures of the resulting ensemble using the multi-state analysis mode from *fpocketR*. Pockets in the cryo-EM and modeled ensembles both clustered in the same two regions of the SAM-IV riboswitch: in the SAM ligand binding site and in the PK-1 pseudoknot (PK) (**Fig. 4**).

Overall, higher-accuracy RNA structures improve the accuracy of pocket finding with *fpocketR*. The performance of *fpocketR* with modeled structures for the preQ1 and SAM-IV riboswitches indicates that RNA modeling software can produce RNA structures with sufficient quality to assess the ligandability of the RNA.

Identification of transient pockets in dynamic RNA structures. Changes in RNA structure and RNA dynamics are often critical for cellular function because structural alterations allow RNAs to respond to their environment and to interact with new partners.[45–47] During translation, the ribosome undergoes large-scale conformational changes that include (*i*) rotation of the small and large ribosomal subunits in a rachet-like mechanism and (*ii*) swiveling of the head domain in the small ribosomal subunit.[48,49] These

large-scale movements enable concerted movement of messenger and transfer RNAs. Several antibiotics, including spectinomycin, neomycin, and Hygromycin B, inhibit translocation by stabilizing transient conformational states of the ribosome.[50]

We used *fpocketR* to search for pockets in ribosomal RNAs across six conformational states of the *E. coli* 70S ribosome during translocation including: PRE-C (7n1p), PRE-H1 (7n2u), PRE-H2 (7n30), INT1 (7n2v), INT2 (7n2c), and POST (7n31) (**Fig. 5A**).[49] We then identified (transient) pockets present in only a subset of conformational states. We identified two pockets at the inter-subunit bridge, B2a, a conserved region located at the interface of helix 69 of the 23S rRNA, helix 44 of the 16S rRNA, and D-stems of the A- and P-site tRNAs.[51] The transient pockets at the B2a bridge are not present in the early stages of translocation (**Fig. 5B**) and only form after back rotation of the small ribosomal subunit in the INT2 and POST states (**Fig. 5C**). These transient pockets partially overlap with binding sites for the antibiotics thermorubin[51] and macrocyclic peptides viomycin[52] and capreomycin[53] (**Fig. 5D**). These antibiotics all have large masses (avg ~650 Da), a large number of hydrogen bond acceptors (mean 13) and donors (mean 11), and low drug-likeness scores (avg quantitative estimate of drug-likeness, QED = 0.08).[54] The (only) partial overlap of the pockets identified using *fpocketR* with binding sites visualized for these antibiotics likely reflects that large macrocyclic compounds do not require deep binding pockets, in contrast to the small molecules used to train *fpocketR*.

Intriguingly, this analysis suggests that compact, drug-like ligands could be devised to bind these pockets, creating a path towards new antibiotics. This exercise demonstrates that *fpocketR* can identify regions that transiently form pockets in concert with RNA-mediated conformational changes. In principle, transient pockets can be targeted with small molecules, and understanding state-specific pocket formation could inform RNA-targeted drug mechanisms.

Shapes of RNA pockets. Pockets in RNA, able to bind ligands with favorable physicochemical properties, are more polar and less hydrophobic compared to pockets in proteins.[12,20,21] An important question is whether these differences influence the molecular shape of ligands that bind RNA. Molecular shape can be evaluated, independent of molecule size, using normalized ratios of principal moments of inertia (termed NPR values) to categorize molecular shapes broadly as rod-, disc-, or sphere-like.[55] Diverse prior work has proposed that RNA ligands and pockets tend to be more rod-like and have fewer sphere-like shapes as compared to protein ligands and pockets.[17,20,21]

We calculated NPR values for the 376 pockets detected by *fpocketR* among non-redundant complexes in the Hariboss RNA-ligand database[22] (<160 kDa; n = 364) and a single reference structure of the

bacterial ribosome (PDB 7k00).[56] This analysis used most available RNA-ligand complexes and includes ligands with a wide range of drug-likeness (QED 0.05 to 0.93). We emphasize that, by using *fpocketR*, we are focusing on pockets able to bind ligands with favorable physicochemical properties.[12] RNA pockets appear to be mostly flat with shapes spanning the space from rod-like to disc-like. Both known (NPR1: 0.31, NPR1: 0.83) and novel (NPR1: 0.31, NPR2: 0.85) *fpocketR*-identified pockets have very similar distribution and average shape (**Fig. 6**). The distribution and average shape for both known and novel RNA pockets are strikingly similar to FDA-approved ligands (NPR1: 0.31, NPR2: 0.85).

RNA pockets detected by *fpocketR* are not as rod-like as the RNA-associated ligands assessed in prior studies (**Fig. 6**). The difference in shape between pockets detected using *fpocketR* versus ligands examined previously[16,17] likely reflects prior inclusion of long rod-like compounds that interact with RNA in the major and minor grooves and with sequence repeats. Ligands that bind in pockets are less rod-like than ligands that bind to repeat sequences.[57] Together, these analyses suggest that RNAs contain pockets with a wide range of shapes able to bind ligands with equally diverse shapes (**Fig. 6**). The shape of RNA-targeted ligands is certainly important for selective engagement with a specific RNA pocket but, in bulk, the shapes of RNA-targeted and protein-targeted pockets and their ligands are broadly similar.

CONCLUSIONS

fpocketR is a reliable ligand-agnostic platform that detects, characterizes, and visualizes high-quality ligand-binding pockets in RNA. We have demonstrated that *fpocketR* is broadly useful for examining flexible RNA ensembles, small and large RNAs, experimentally-determined structures and computational models, and dynamic conformational ensembles. *fpocketR* identifies pockets in complex structural regions of large RNAs despite being trained on short (< 200 nt) RNA riboswitches and aptamers. The rules for the local structural environments that create ligand-binding pockets, capable of binding a drug-like ligand, are thus independent of macromolecular size. This study also provides further support that complex RNA structures – multi-helix junctions, pseudoknots, and other idiosyncratic tertiary structures – are the best RNA targets for drug-like small molecule ligands.

Unsurprisingly, pocket finding is most successful when structures are solved at high resolution. However, *fpocketR* also identified known pockets in computational models for the preQ1 and SAM-IV riboswitches examined here. These results demonstrate that RNA structures generated by physics-based and machine-learning modeling can be used to inform pocket detection and ligandability.

Many biologically important RNA targets are dynamic and *fpocketR* provides dedicated and flexible tools for analyzing RNA ensembles. *fpocketR* identified transiently formed pockets at the interface between the small and large subunit rRNAs in the bacterial ribosome, which also overlapped with primary and secondary binding sites of translation-inhibiting antibiotics. The ability to identify RNA conformation states that selectively bind small-molecule ligands creates a powerful platform for guiding ligand (and drug) discovery and for defining binding and inhibition mechanisms.

fpocketR is optimized and validated to detect pockets capable of binding small molecules with favorable physicochemical properties, commonly termed drug-like.[12] The shape-space of these RNA pockets closely matches the shape-space of FDA-approved small-molecule drugs, indicating that RNA-specific shape is probably not an important property to consider when curating libraries for RNA-targeted small molecules. Instead, chemical property differentiators other than shape should be the primary focus of RNA-focused ligand design.

fpocketR enables robust pocket detection for local RNA regions able to bind drug-like ligands, using both experimental and modeled RNA structures. The performance of *fpocketR* has been examined with and broadly validated for holo (with ligand), apo (without ligand), synthetic, dynamic, low-resolution, and computationally modeled RNA structures. We anticipate that *fpocketR* will provide diverse and thought-provoking information about the location and properties of ligand-binding pockets in RNA.

METHODS

Software availability. This study uses two newly developed software packages *fpocketR* and *MolMetrics*. *fpocketR*, used for pocket discovery and analysis, is freely accessed at https://github.com/Weeks-UNC/fpocketR. *fpocketR* requires a Unix-based operating system (for example, Linux or MacOSX). Windows users can run *fpocketR* via Windows Subsystem for Linux (WSL2) or a virtual machine. This repository includes usage documentation, example high-throughput workflows, and an explanation of all available features. *MolMetrics* generates multiple molecular conformers and calculates molecular descriptors and principal moments of inertia for small-molecule libraries via an efficient command line interface. *MolMetrics* is a cross-platform Python (3.7+) tool and is available at https://github.com/Weeks-UNC/molmetrics.

Pocket finding. Pockets were identified, characterized, and visualized using *fpocketR* 1.2.0, which functions as a wrapper for *fpocket* v4.0.3.[30,31] RNA tertiary and secondary structures were input using the *fpocketR* arguments --pdb and --nsd, respectively. Multi-state analyses were performed by setting the --state argument to 0 (all states).

SAM-IV riboswitch 3D structure modeling. Structural models of the SAM-IV riboswitch were produced using the sequence from a reference structure[35] (PDB: 6wlq) and secondary structure predicted by Fold from the RNAstructure package.[32] These sequence and secondary structure files were then used to generate a multiple sequence alignment with rMSA[58] using the NCBI nucleotide and RNAcentral databases.[59] The multiple sequence alignment and predicted secondary structure were input to trRosettaRNA[44] to generate 20 structural models. The resulting structures were aligned using PyMOL 3.0 (pymol.org, Schrödinger LLC) and saved as multiple states within a single (PDB) structure file.

Multi-state analysis of the SAM-IV riboswitch. Ensembles for the SAM-IV riboswitch were analyzed using the *fpocketR* --state and --qualityfilter arguments. Pockets with a score less than 0.40 were omitted to facilitate direct comparison to previous studies.[12]

R-BIND library. The RNA-targeted BIoactive ligaNd Database (R-BIND) (v2.1) molecules (downloaded on January 14, 2025) includes organic small-molecule probes reported in the literature through December 2021.[17] The library was not filtered and contained 159 molecules.

FDA-approved ligand library. FDA-approved (Phase 4) molecules[60] (downloaded from CHEMBL on January 14, 2025) were filtered to exclude molecules with masses less than 140 amu or greater than 590 amu. The final library contained 2865 molecules.

Hariboss RNA-ligand complex library. Hariboss RNA-ligand complexes[22] (downloaded on January 14, 2025) were filtered to exclude redundant complexes and require RNAs be between 4 and 160 kDa (~15 – ~500 nts). A single high-resolution ribosome structure was added (PDB 7k00).[56] The final library contained 365 RNA-ligand complexes (**SI Table 1**) which bind to both low and high QED score ligands (avg. QED = 0.35). Our curated library maximizes unique RNA-ligand complexes and contains approximately 160 RNA structures, many of which bind the same ligands (for example, TPP or SAM). *fpocketR* detected 139 known pockets, which selectively overlapped ligands with higher QED scores (avg. QED = 0.44), and identified 237 novel pockets. Notably, pockets detected in the final curated Hariboss library have a nearly identical average shape (NPR1: 0.30, NPR2: 0.84) to the pockets detected in the RNAs used to test and train *fpocketR* (NPR1: 0.31, NPR2: 0.84).[12]

Principal moment of inertia analysis. Normalized principal ratios (NPRs) of principal moments of inertia were calculated for R-BIND and FDA-approved ligands using RDKit.[61] Low-energy

conformations (n=1000) were generated from SMILES strings using ETKDGv3. The lowest energy conformer was determined by optimizing geometries using the Universal Force Field. Boltzmann weighted average NPR values were calculated from all conformers within 3 kcal/mol of the lowest energy conformer. We note that the assessment of NPRs for a small molecule by RDKit does not account for the radius of the atoms in the molecule. In contrast, NPRs of RNA pockets calculated by *fpocketR* reflect a method that takes the radius of alpha cores into account.

fpocketR calculates the principal moments of inertia for pockets after converting each pocket from a cluster of alpha cores (alpha core = alpha sphere – 1.65 Å) into a single solid object with uniform density. The shape of RNA pockets detected with *fpocketR* is shifted to -0.07 on the NPR1 axis, relative to values provided by RDKit. We therefore normalized NPR1 values to the left diagonal (running between rod-like and disc-like vertices) to normalize NPR values calculated for pockets and ligands. Contour lines were generated using kernel density estimation at a 50% density level, and joint Wasserstein distances were computed between bivariate distributions using the Python Optimal Transport Theory library,[62] quantifying the multidimensional dissimilarity relative to the distribution of FDA-approved ligands.

ACKNOWLEGEMENT

We are indebted to Patrick Irving for providing thoughtful feedback on this work. This study was supported by grants from the US National Science Foundation (MCB-2027701 to K.M.W.) and National Institutes of Health (R21 AG084970 to S.F. and K.M.W.).

DISCLOSURE

K.M.W. is a founder at ForagR Medicines, Ribometrix, and A-Form Solutions.

REFERENCES

(1) Sharp, P. A. The Centrality of RNA. *Cell* **2009**, *136* (4), 577–580. https://doi.org/10.1016/j.cell.2009.02.007.

(2) Cech, T. R.; Steitz, J. A. The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* **2014**, *157* (1), 77–94. https://doi.org/10.1016/J.CELL.2014.03.008.

(3) Thomas, J. R.; Hergenrother, P. J. Targeting RNA with Small Molecules. *Chem Rev* **2008**, *108* (4), 1171–1224. https://doi.org/10.1021/CR0681546/ASSET/CR0681546.FP.PNG_V03.

(4) Warner, K. D.; Hajdin, C. E.; Weeks, K. M. Principles for Targeting RNA with Drug-like Small Molecules. *Nat Rev Drug Discov* **2018**, *17* (8), 547–558. https://doi.org/10.1038/nrd.2018.93.

(5) Umuhire Juru, A.; Hargrove, A. E. Frameworks for Targeting RNA with Small Molecules. *J Biol Chem* **2021**, *296*, 100191. https://doi.org/10.1074/JBC.REV120.015203.

(6) Childs-Disney, J. L.; Yang, X.; Gibaut, Q. M. R.; Tong, Y.; Batey, R. T.; Disney, M. D. Targeting RNA Structures with Small Molecules. *Nat Rev Drug Discov* **2022**, *21*, 736–762. https://doi.org/10.1038/s41573-022-00521-4.

(7) Kovachka, S.; Panosetti, M.; Grimaldi, B.; Azoulay, S.; Di Giorgio, A.; Duca, M. Small Molecule Approaches to Targeting RNA. *Nat Rev Chem* **2024**, *8* (2), 120–135. https://doi.org/10.1038/s41570-023-00569-9.

(8) Fullenkamp, C. R.; Liang, X.; Pettersson, M.; Schneekloth, J. Outlook. In *RNA as a Drug Target*; Schneekloth, J., Pettersson, M., Mannhold, R., Buschmann, H., Holenz, J., Eds.; John Wiley & Sons, Ltd, 2024; pp 355–384. https://doi.org/10.1002/9783527840458.CH13.

(9) Moellering, R. C. Linezolid: The First Oxazolidinone Antimicrobial. *Ann Intern Med* **2003**, *138* (2), 135–142. https://doi.org/10.7326/0003-4819-138-2-200301210-00015.

(10) Ratni, H.; Scalco, R. S.; Stephan, A. H. Risdiplam, the First Approved Small Molecule Splicing Modifier Drug as a Blueprint for Future Transformative Medicines. *ACS Med Chem Lett* **2021**, *12* (6), 874–877. https://doi.org/10.1021/ACSMEDCHEMLETT.0C00659/ASSET/IMAGES/LARGE/ML0C00659_0002.JPEG.

(11) Connelly, C. M.; Moon, M. H.; Schneekloth, J. S. The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol* **2016**, *23* (9), 1077–1090. https://doi.org/10.1016/j.chembiol.2016.05.021.

(12) Veenbaas, S. D.; Koehn, J. T.; Irving, P. S.; Lama, N. N.; Weeks, K. M. Ligand-Binding Pockets in RNA and Where to Find Them. *Proceedings of the National Academy of Sciences* **2025**, *122* (17), e2422346122. https://doi.org/10.1073/PNAS.2422346122.

(13) Chin, K.; Sharp, K. A.; Honig, B.; Pyle, A. M. Calculating the Electrostatic Properties of RNA Provides New Insights into Molecular Interactions and Function. *Nat Struct Biol* **1999**, *6* (11), 1055–1061. https://doi.org/10.1038/14940.

(14) Morgan, B. S.; Sanaba, B. G.; Donlic, A.; Karloff, D. B.; Forte, J. E.; Zhang, Y.; Hargrove, A. E. R-BIND: An Interactive Database for Exploring and Developing RNA-Targeted Chemical Probes. *ACS Chem Biol* **2019**, *14* (12), 2691–2700. https://doi.org/10.1021/ACSCHEMBIO.9B00631/SUPPL_FILE/CB9B00631_SI_004.XLSX.

(15) Bosio, S.; Bernetti, M.; Rocchia, W.; Masetti, M. Similarities and Differences in Ligand Binding to Protein and RNA Targets: The Case of Riboflavin. *J Chem Inf Model* **2024**, *64* (11), 4570–4586. https://doi.org/10.1021/ACS.JCIM.4C00420/ASSET/IMAGES/LARGE/CI4C00420_0008.JPEG.

(16) Morgan, B. S.; Forte, J. E.; Culver, R. N.; Zhang, Y.; Hargrove, A. E. Discovery of Key Physicochemical, Structural, and Spatial Properties of RNA-Targeted Bioactive Ligands. *Angew. Chem. Int. Ed.* **2017**, *56* (43), 13498–13502. https://doi.org/10.1002/ANIE.201707641.

(17) Donlic, A.; Swanson, E. G.; Chiu, L.-Y.; Wicks, S. L.; Juru, A. U.; Cai, Z.; Kassam, K.; Laudeman, C.; Sanaba, B. G.; Sugarman, A.; Han, E.; Tolbert, B. S.; Hargrove, A. E. R-BIND 2.0: An Updated Database of Bioactive RNA-Targeting Small Molecules and Associated RNA Secondary Structures. *ACS Chem Biol* **2022**, *17* (6), 1556–1566. https://doi.org/10.1021/ACSCHEMBIO.2C00224.

(18) Yazdani, K.; Jordan, D.; Yang, M.; Fullenkamp, C. R.; Calabrese, D. R.; Boer, R.; Hilimire, T.; Allen, T. E. H.; Khan, R. T.; Schneekloth, J. S. Machine Learning Informs RNA-Binding Chemical Space**. *Angewandte Chemie International Edition* **2023**, *62* (11), e202211358. https://doi.org/10.1002/ANIE.202211358.

(19) Nagasawa, R.; Onizuka, K.; Komatsu, K. R.; Miyashita, E.; Murase, H.; Ojima, K.; Ishikawa, S.; Ozawa, M.; Saito, H.; Nagatsugi, F. Large-Scale Analysis of Small Molecule-RNA Interactions Using Multiplexed RNA Structure Libraries. *Commun Chem* **2024**, *7* (1), 1–12. https://doi.org/10.1038/S42004-024-01181-8;SUBJMETA.

(20) Hewitt, W. M.; Calabrese, D. R.; Schneekloth, J. S. Evidence for Ligandable Sites in Structured RNA throughout the Protein Data Bank. *Bioorg Med Chem* **2019**, *27* (11), 2253–2260. https://doi.org/10.1016/j.bmc.2019.04.010.

(21) Zhou, T.; Wang, H.; Zeng, C.; Zhao, Y. RPocket: An Intuitive Database of RNA Pocket Topology Information with RNA-Ligand Data Resources. *BMC Bioinformatics* **2021**, *22* (1), 428. https://doi.org/10.1186/S12859-021-04349-4.

(22) Panei, F. P.; Torchet, R.; Mé Nager, H.; Gkeka, P.; Bonomi, M. HARIBOSS: A Curated Database of RNA-Small Molecules Structures to Aid Rational Drug Design. *Bioinformatics* **2022**, *38* (17), 4185–4193. https://doi.org/10.1093/BIOINFORMATICS/BTAC483.

(23) Wang, K.; Zhou, R.; Wu, Y.; Li, M. RLBind: A Deep Learning Method to Predict RNA–Ligand Binding Sites. *Brief Bioinform* **2023**, *24* (1), 1–12. https://doi.org/10.1093/BIB/BBAC486.

(24) Wang, J.; Quan, L.; Jin, Z.; Wu, H.; Ma, X.; Wang, X.; Xie, J.; Pan, D.; Chen, T.; Wu, T.; Lyu, Q. MultiModRLBP: A Deep Learning Approach for Multi-Modal RNA-Small Molecule Ligand Binding Sites Prediction. *IEEE J Biomed Health Inform* **2024**, *28* (8), 4995–5006. https://doi.org/10.1109/JBHI.2024.3400521.

(25) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J Comput Chem* **2010**, *31* (2), 455–461. https://doi.org/10.1002/JCC.21334;CTYPE:STRING:JOURNAL.

(26) Lang, P. T.; Brozell, S. R.; Mukherjee, S.; Pettersen, E. F.; Meng, E. C.; Thomas, V.; Rizzo, R. C.; Case, D. A.; James, T. L.; Kuntz, I. D. DOCK 6: Combining Techniques to Model RNA–Small Molecule Complexes. *RNA* **2009**, *15* (6), 1219–1230. https://doi.org/10.1261/RNA.1563609.

(27) Ruiz-Carmona, S.; Alvarez-Garcia, D.; Foloppe, N.; Garmendia-Doval, A. B.; Juhos, S.; Schmidtke, P.; Barril, X.; Hubbard, R. E.; Morley, S. D. RDock: A Fast, Versatile and Open Source Program for Docking Ligands to Proteins and Nucleic Acids. *PLoS Comput Biol* **2014**, *10* (4), e1003571. https://doi.org/10.1371/JOURNAL.PCBI.1003571.

(28) Sun, L. Z.; Jiang, Y.; Zhou, Y.; Chen, S. J. RLDOCK: A New Method for Predicting RNA-Ligand Interactions. *J Chem Theory Comput* **2020**, *16* (11), 7173–7183. https://doi.org/10.1021/ACS.JCTC.0C00798/ASSET/IMAGES/LARGE/CT0C00798_0007.JPEG.

(29) Zhou, Y.; Jiang, Y.; Chen, S. J. RNA–Ligand Molecular Docking: Advances and Challenges. *Wiley Interdiscip Rev Comput Mol Sci* **2022**, *12* (3), e1571. https://doi.org/10.1002/WCMS.1571;PAGEGROUP:STRING:PUBLICATION.

(30) Le Guilloux, V.; Schmidtke, P.; Tuffery, P. Fpocket: An Open Source Platform for Ligand Pocket Detection. *BMC Bioinformatics* **2009**, *10* (1), 168. https://doi.org/10.1186/1471-2105-10-168.

(31) Schmidtke, P.; Le Guilloux, V.; Maupetit, J.; Tufféry, P. Fpocket: Online Tools for Protein Ensemble Pocket Detection and Tracking. *Nucleic Acids Res* **2010**, *38* (SUPPL. 2), W582–W589. https://doi.org/10.1093/nar/gkq383.

(32) Reuter, J. S.; Mathews, D. H. RNAstructure: Software for RNA Secondary Structure Prediction and Analysis. *BMC Bioinformatics* **2010**, *11* (1), 129. https://doi.org/10.1186/1471-2105-11-129/FIGURES/3.

(33) Darty, K.; Denise, A.; Ponty, Y. VARNA: Interactive Drawing and Editing of the RNA Secondary Structure. *Bioinformatics* **2009**, *25* (15), 1974–1975. https://doi.org/10.1093/BIOINFORMATICS/BTP250.

(34) Sweeney, B. A.; Hoksza, D.; Nawrocki, E. P.; Ribas, C. E.; Madeira, F.; Cannone, J. J.; Gutell, R.; Maddala, A.; Meade, C. D.; Williams, L. D.; Petrov, A. S.; Chan, P. P.; Lowe, T. M.; Finn, R. D.; Petrov, A. I. R2DT Is a Framework for Predicting and Visualising RNA Secondary Structure Using Templates. *Nat. Commun.* **2021**, *12* (1), 1–12. https://doi.org/10.1038/s41467-021-23555-5.

(35) Kappel, K.; Zhang, K.; Su, Z.; Watkins, A. M.; Kladwang, W.; Li, S.; Pintilie, G.; Topkar, V. V.; Rangan, R.; Zheludev, I. N.; Yesselman, J. D.; Chiu, W.; Das, R. Accelerated Cryo-EM-Guided Determination of Three-Dimensional RNA-Only Structures. *Nat Methods* **2020**, *17* (7), 699–707. https://doi.org/10.1038/s41592-020-0878-9.

(36) Ma, H.; Jia, X.; Zhang, K.; Su, Z. Cryo-EM Advances in RNA Structure Determination. *Sig Transduct Target Ther* **2022**, *7* (1), 1–6. https://doi.org/10.1038/s41392-022-00916-0.

(37) Haack, D. B.; Rudolfs, B.; Jin, S.; Khitun, A.; Weeks, K. M.; Toor, N. Scaffold-Enabled High-Resolution Cryo-EM Structure Determination of RNA. *Nat Commun* **2025**, *16* (1), 1–17. https://doi.org/10.1038/s41467-024-55699-5.

(38) Mcrae, E. K. S.; Rasmussen, H. Ø.; Liu, J.; Bøggild, A.; Nguyen, M. T. A.; Sampedro Vallina, N.; Boesen, T.; Pedersen, J. S.; Ren, G.; Geary, C.; Andersen, E. S. Structure, Folding and Flexibility of Co-Transcriptional RNA Origami. *Nat Nanotechnol* **2023**, *18* (7), 808–817. https://doi.org/10.1038/s41565-023-01321-6.

(39) Sampedro Vallina, N.; McRae, E. K. S.; Hansen, B. K.; Boussebayle, A.; Andersen, E. S. RNA Origami Scaffolds Facilitate Cryo-EM Characterization of a Broccoli–Pepper Aptamer FRET Pair. *Nucleic Acids Res* **2023**, *51* (9), 4613–4624. https://doi.org/10.1093/NAR/GKAD224.

(40) Das, R.; Kretsch, R. C.; Simpkin, A. J.; Mulvaney, T.; Pham, P.; Rangan, R.; Bu, F.; Keegan, R. M.; Topf, M.; Rigden, D. J.; Miao, Z.; Westhof, E. Assessment of Three-Dimensional RNA Structure Prediction in CASP15. *Proteins:Struct., Funct., Bioinf.* **2023**, *91* (12), 1747–1770. https://doi.org/10.1002/PROT.26602.

(41) Magnus, M.; Miao, Z. RNA 3D Structure Comparison Using RNA-Puzzles Toolkit. *Methods Mol. Biol.* **2023**, *2586*, 263–285. https://doi.org/10.1007/978-1-0716-2768-6_16.

(42) Schroeder, G. M.; Kiliushik, D.; Jenkins, J. L.; Wedekind, J. E. Structure and Function Analysis of a Type III PreQ1-I Riboswitch from Escherichia Coli Reveals Direct Metabolite Sensing by the Shine-Dalgarno Sequence. *Journal of Biological Chemistry* **2023**, *299* (10), 105208. https://doi.org/10.1016/J.JBC.2023.105208.

(43) Zhang, K.; Li, S.; Kappel, K.; Pintilie, G.; Su, Z.; Mou, T. C.; Schmid, M. F.; Das, R.; Chiu, W. Cryo-EM Structure of a 40 KDa SAM-IV Riboswitch RNA at 3.7 Å Resolution. *Nat Commun* **2019**, *10* (1), 5511. https://doi.org/10.1038/s41467-019-13494-7.

(44) Wang, W.; Feng, C.; Han, R.; Wang, Z.; Ye, L.; Du, Z.; Wei, H.; Zhang, F.; Peng, Z.; Yang, J. TrRosettaRNA: Automated Prediction of RNA 3D Structure with Transformer Network. *Nat Commun* **2023**, *14* (1), 7266. https://doi.org/10.1038/s41467-023-42528-4.

(45) Ganser, L. R.; Kelly, M. L.; Herschlag, D.; Al-Hashimi, H. M. The Roles of Structural Dynamics in the Cellular Functions of RNAs. *Nat Rev Mol Cell Biol* **2019**, *20* (8), 474–489. https://doi.org/10.1038/s41580-019-0136-0.

(46) Bose, R.; Saleem, I.; Mustoe, A. M. Causes, Functions, and Therapeutic Possibilities of RNA Secondary Structure Ensembles and Alternative States. *Cell Chem Biol* **2024**, *31* (1), 17–35. https://doi.org/10.1016/J.CHEMBIOL.2023.12.010.

(47) Arney, J. W.; Laederach, A.; Weeks, K. M. Visualizing RNA Structure Ensembles by Single-Molecule Correlated Chemical Probing. *Curr Opin Struct Biol* **2024**, *88*, 102877. https://doi.org/10.1016/J.SBI.2024.102877.

(48) Mohan, S.; Donohue, J. P.; Noller, H. F. Molecular Mechanics of 30S Subunit Head Rotation. *Proc Natl Acad Sci U S A* **2014**, *111* (37), 13325–13330. https://doi.org/10.1073/pnas.1413731111.

(49) Rundlet, E. J.; Holm, M.; Schacherl, M.; Natchiar, S. K.; Altman, R. B.; Spahn, C. M. T.; Myasnikov, A. G.; Blanchard, S. C. Structural Basis of Early Translocation Events on the Ribosome. *Nature* **2021**, *595* (7869), 741–745. https://doi.org/10.1038/s41586-021-03713-x.

(50) Paternoga, H.; Crowe-McAuliffe, C.; Bock, L. V.; Koller, T. O.; Morici, M.; Beckert, B.; Myasnikov, A. G.; Grubmüller, H.; Nováček, J.; Wilson, D. N. Structural Conservation of Antibiotic Interaction with Ribosomes. *Nat Struct Mol Biol* **2023**, *30* (9), 1380–1392. https://doi.org/10.1038/s41594-023-01047-y.

(51) Bulkley, D.; Johnson, F.; Steitz, T. A. The Antibiotic Thermorubin Inhibits Protein Synthesis by Binding to Inter-Subunit Bridge B2a of the Ribosome. *J Mol Biol* **2012**, *416* (4), 571–578. https://doi.org/10.1016/J.JMB.2011.12.055.

(52) Zhang, L.; Wang, Y. H.; Zhang, X.; Lancaster, L.; Zhou, J.; Noller, H. F. The Structural Basis for Inhibition of Ribosomal Translocation by Viomycin. *Proc Natl Acad Sci U S A* **2020**, *117* (19), 10271–10277. https://doi.org/10.1073/PNAS.2002888117/SUPPL_FILE/PNAS.2002888117.SAPP.PDF.

(53) Stanley, R. E.; Blaha, G.; Grodzicki, R. L.; Strickler, M. D.; Steitz, T. A. The Structures of the Anti-Tuberculosis Antibiotics Viomycin and Capreomycin Bound to the 70S Ribosome. *Nat Struct Mol Biol* **2010**, *17* (3), 289–293. https://doi.org/10.1038/nsmb.1755.

(54) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the Chemical Beauty of Drugs. *Nat Chem* **2012**, *4* (2), 90–98. https://doi.org/10.1038/nchem.1243.

(55) Sauer, W. H. B.; Schwarz, M. K. Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *J Chem Inf Comput Sci* **2003**, *43* (3), 987–1003. https://doi.org/10.1021/CI025599W/SUPPL_FILE/CI025599WSI20030304_021848.PDF.

(56) Watson, Z. L.; Ward, F. R.; Méheust, R.; Ad, O.; Schepartz, A.; Banfield, J. F.; Cate, J. H. D. Structure of the Bacterial Ribosome at 2 Å Resolution. *Elife* **2020**, *9*, e60482. https://doi.org/10.7554/ELIFE.60482.

(57) Fan, C.; Wang, X.; Ling, T.; Yang, Y.; Zhao, H. Characterizing RNA-Binding Ligands on Structures, Chemical Information, Binding Affinity and Drug-Likeness. *RNA Biol* **2023**, *20* (1), 431–443. https://doi.org/10.1080/15476286.2023.2231708.

(58) Zhang, C.; Zhang, Y.; Pyle, A. M. RMSA: A Sequence Search and Alignment Algorithm to Improve RNA Structure Modeling. *J Mol Biol* **2023**, *435* (14), 167904. https://doi.org/10.1016/J.JMB.2022.167904.

(59) Sweeney, B. A.; Petrov, A. I.; Burkov, B.; Finn, R. D.; Bateman, A.; Szymanski, M.; Karlowski, W. M.; Gorodkin, J.; Seemann, S. E.; Cannone, J. J.; Gutell, R. R.; Fey, P.; Basu, S.; Kay, S.; Cochrane, G.; Billis, K.; Emmert, D.; Marygold, S. J.; Huntley, R. P.; Lovering, R. C.; Frankish, A.; Chan, P. P.; Lowe, T. M.; Bruford, E.; Seal, R.; Vandesompele, J.; Volders, P. J.; Paraskevopoulou, M.; Ma, L.; Zhang, Z.; Griffiths-Jones, S.; Bujnicki, J. M.; Boccaletto, P.; Blake, J. A.; Bult, C. J.; Chen, R.; Zhao, Y.; Wood, V.; Rutherford, K.; Rivas, E.; Cole, J.; Lauledekind, S. J. F.; Shimoyama, M.; Gillespie, M. E.; Orlic-Milacic, M.; Kalvari, I.; Nawrocki, E.; Engel, S. R.; Cherry, J. M.; Team, S.; Berardini, T. Z.; Hatzigeorgiou, A.; Karagkouni, D.; Howe, K.; Davis, P.; Dinger, M.; He, S.; Yoshihama, M.; Kenmochi, N.; Stadler, P. F.; Williams, K. P. RNAcentral: A

Hub of Information for Non-Coding RNA Sequences. *Nucleic Acids Res* **2019**, *47* (D1), D221–D229. https://doi.org/10.1093/NAR/GKY1034.

(60) Zdrazil, B.; Felix, E.; Hunter, F.; Manners, E. J.; Blackshaw, J.; Corbett, S.; de Veij, M.; Ioannidis, H.; Lopez, D. M.; Mosquera, J. F.; Magarinos, M. P.; Bosc, N.; Arcila, R.; Kizilören, T.; Gaulton, A.; Bento, A. P.; Adasme, M. F.; Monecke, P.; Landrum, G. A.; Leach, A. R. The ChEMBL Database in 2023: A Drug Discovery Platform Spanning Multiple Bioactivity Data Types and Time Periods. *Nucleic Acids Res* **2024**, *52* (D1), D1180–D1192. https://doi.org/10.1093/NAR/GKAD1004.

(61) Greg Landrum; Julie Penzotti. *RDKit*; https://www.rdkit.org/, 2010. https://www.rdkit.org/.

(62) Flamary, R.; Vincent-Cuaz, C.; Courty, N.; Gramfort, A.; Kachaiev, O.; Quang Tran, H.; David, L.; Bonet, C.; Cassereau, N.; Gnassounou, T.; Tanguy, E.; Delon, J.; Collas, A.; Mazelet, S.; Chapel, L.; Kerdoncuff, T.; Yu, X.; Feickert, M.; Krzakala, P.; Liu, T.; Fernandes Montesuma, E. POT Python Optimal Transport. March 20, 2025. https://github.com/PythonOT/POT.

FIGURES

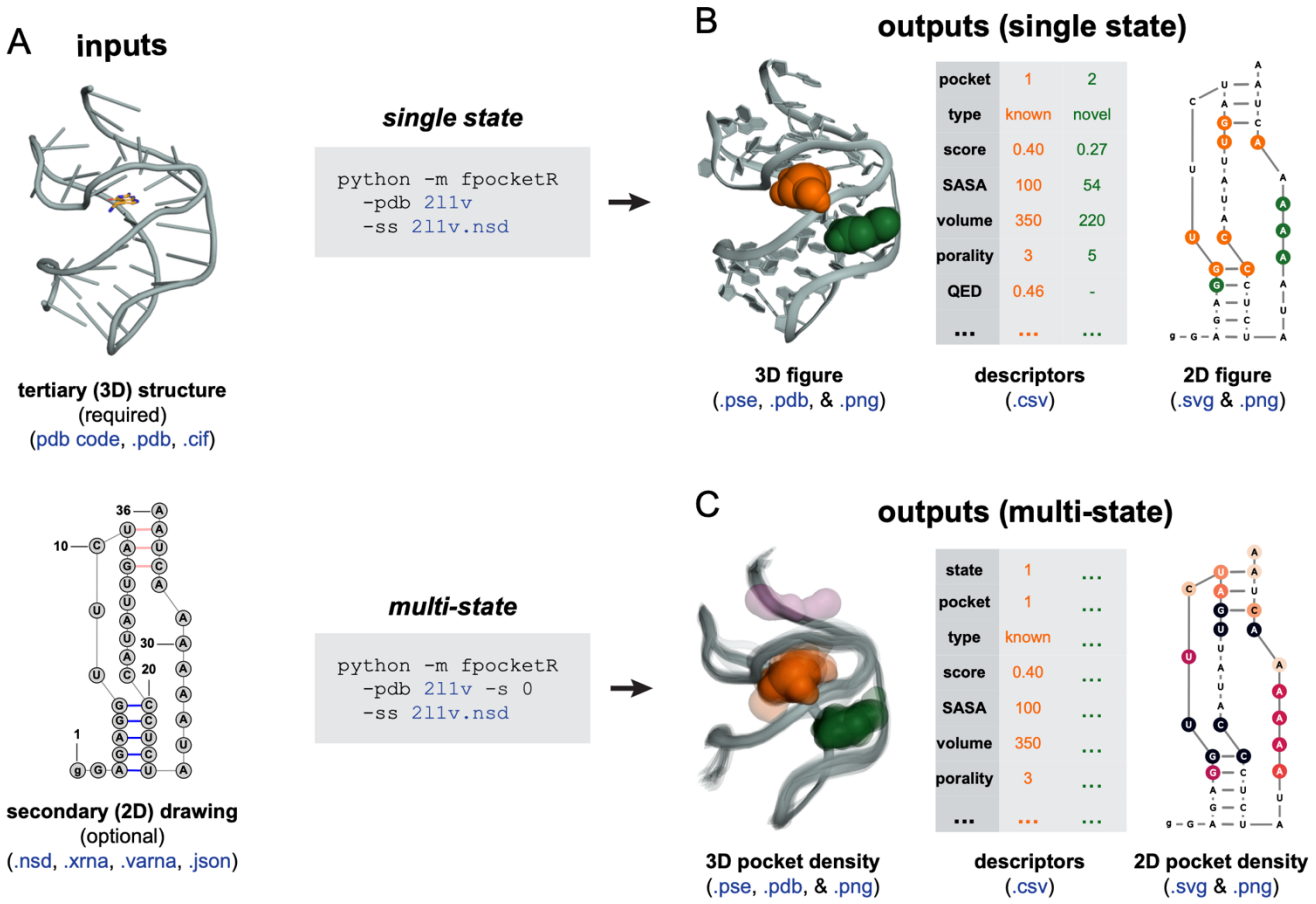


Figure 1. Pocket finding workflow for *fpocketR*, visualized with the class I preQ1 riboswitch (PDB: 2l1v). (A) Structural and command line inputs with accepted file formats. (B) Single-state analyses. Pockets are visualized as colored spheres and nucleotides that form pockets are highlighted in secondary structures. (C) Multi-state analyses. Map of tertiary structure pocket density displays all states of an RNA as a transparent backbone and pockets. In the map of pockets in secondary structure space, the intensity of colored nucleotides correlates with the frequency of participation in pockets.

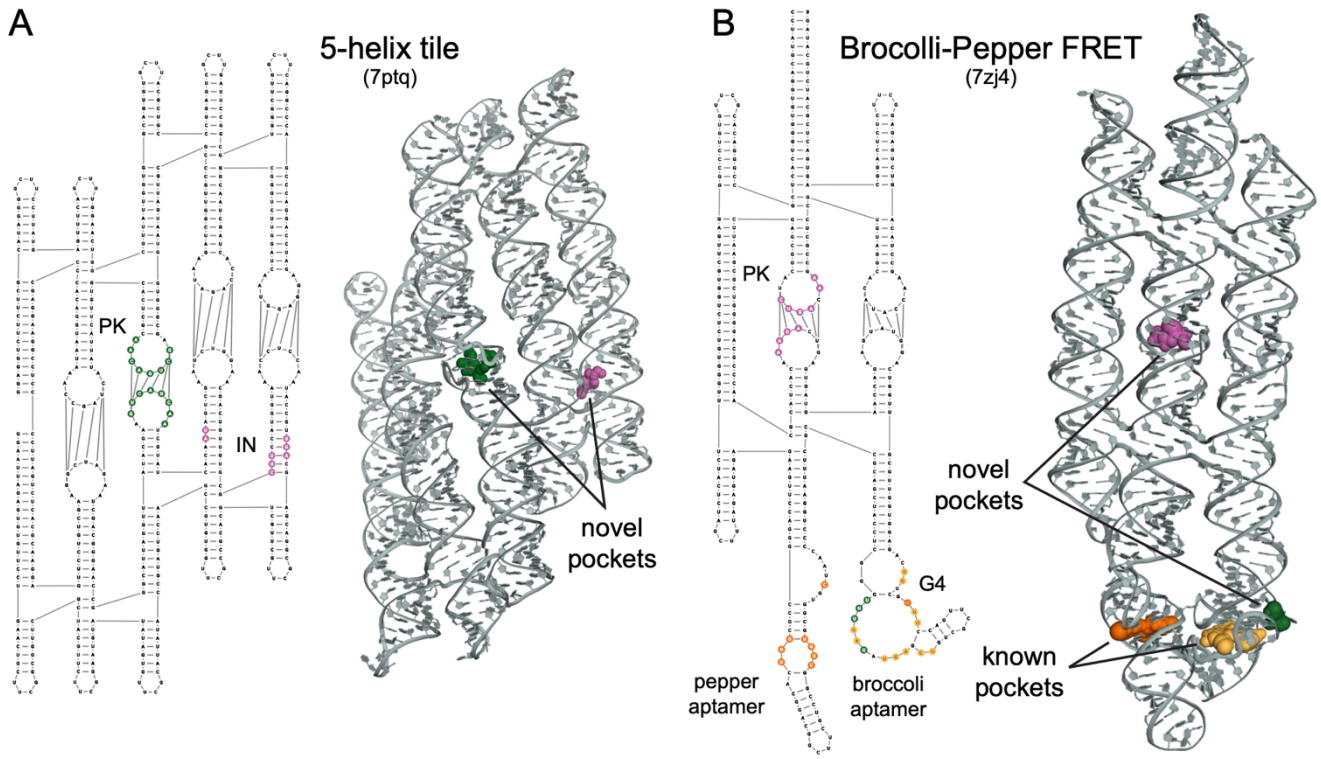


Figure 2. Pocket detection in large engineered RNAs. Pockets detected in (A) a five-helix RNA origami scaffold (544 nts)[38] and (B) an RNA engineered to include Pepper (orange) and Broccoli (yellow) ligand-binding aptamers (374 nts).[39] Novel RNA structures that form pockets are labeled as pseudoknots (PK), interfaces (IN) or G-quadraplex (G4).

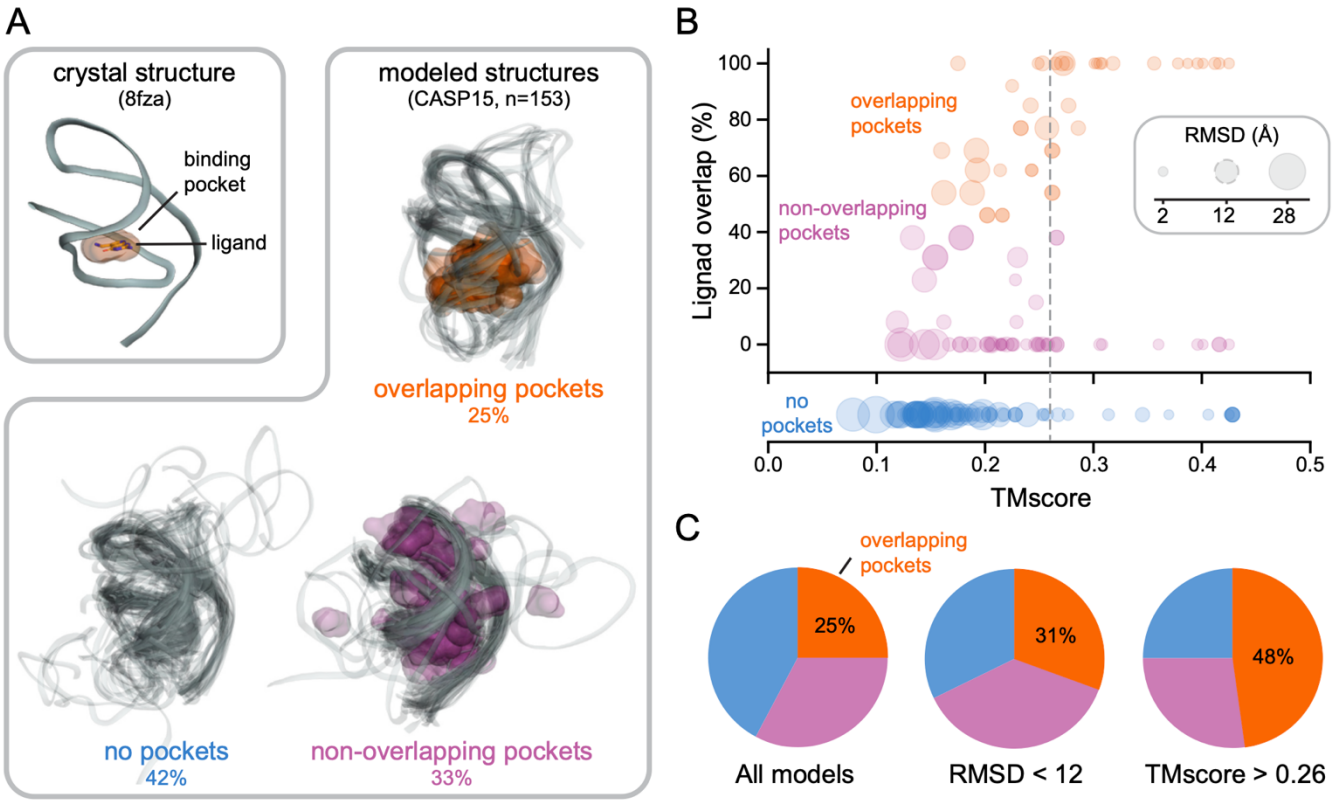


Figure 3. Pocket detection performance in modeled RNA structures. (A) Pockets detected in a reference structure (2.3 Å resolution; PDB 8fza)[42] and in 153 models of the class I type III preQ1 riboswitch, from the CASP15 exercise.[40] (B) Relationship between pocket performance (ligand overlap) and structural quality metrics, RMSD and TMscore. Dashed gray circular (RMSD) and vertical (TMscore) lines represent quality thresholds that produced the best selection for pockets overlapping the known ligand binding site. (C) Distribution of models with overlapping (orange), non-overlapping (pink), or no pocket (blue) for the indicated threshold of RMSD and TMscore.

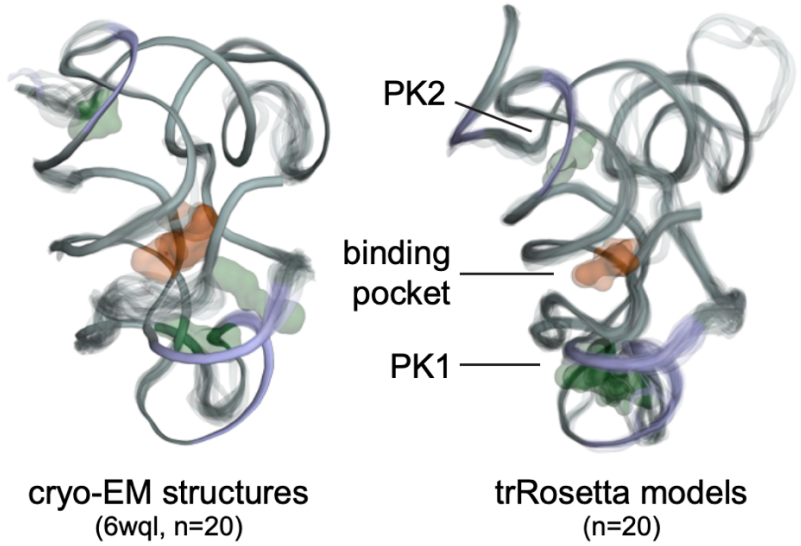


Figure 4. Pocket detection in multi-state ensembles. Map of pocket densities for ensembles of the SAM-IV riboswitch visualized by cryo-EM (PDB 6wlq)[35] or generated by trRosettaRNA [44] modelling.

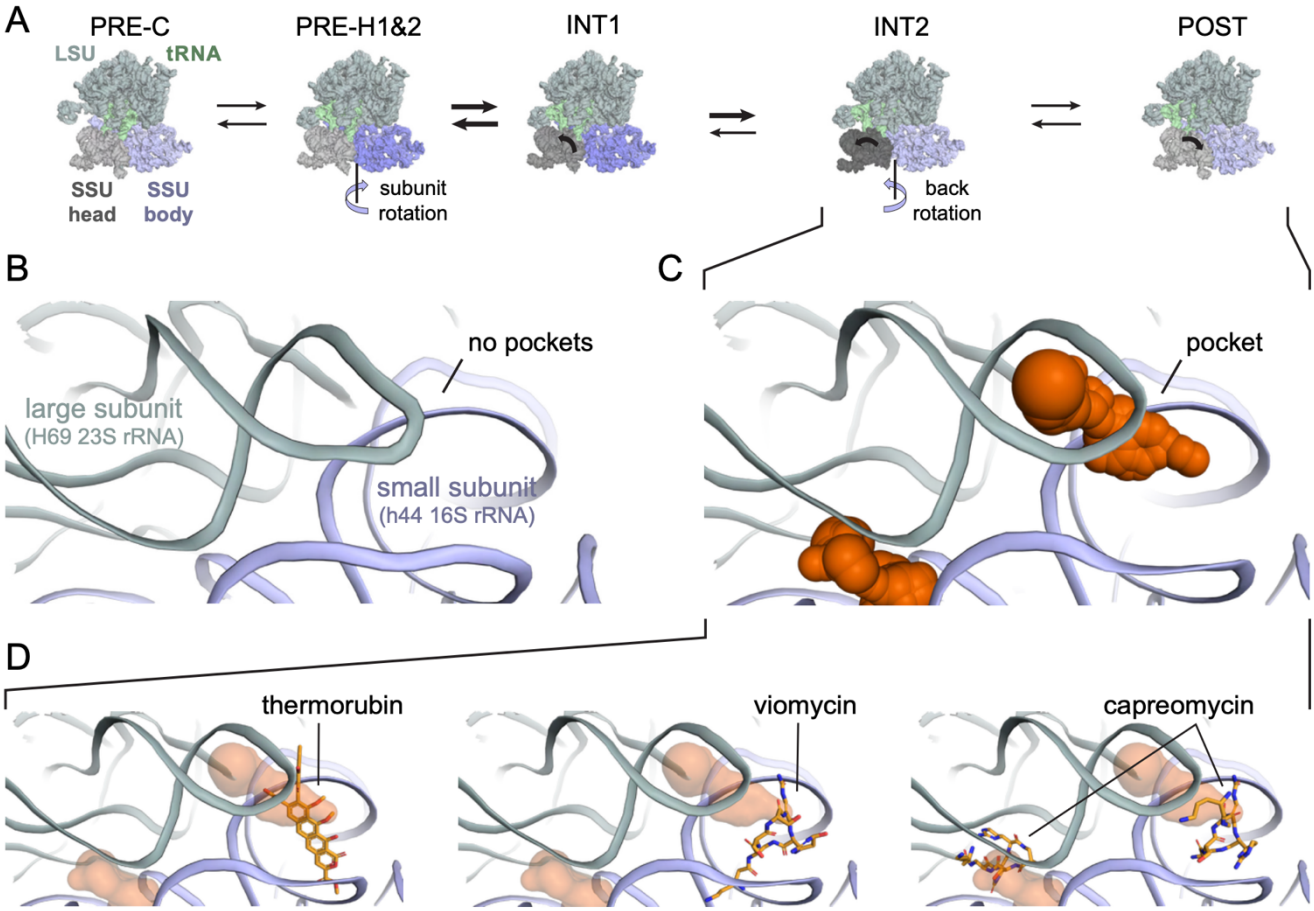


Figure 5. Identification of transient pockets at the interface of the large and small subunits of the bacterial ribosome, during translocation. (A) Overview of ribosome conformations at indicated stages of translocation.[49] (B) Detailed view of the B2a inter-subunit bridge, which contains no pockets at early states of translocation. (C) Pockets in the B2a inter-subunit bridge formed between H69 and h44 in the INT2 and POST states of translocation. (D) Pockets detected at the B2a bridge region partially overlap binding sites for the antibiotics thermorubin (4v8a),[51] viomycin (6lkq),[52] and capreomycin (8ceu).[53]

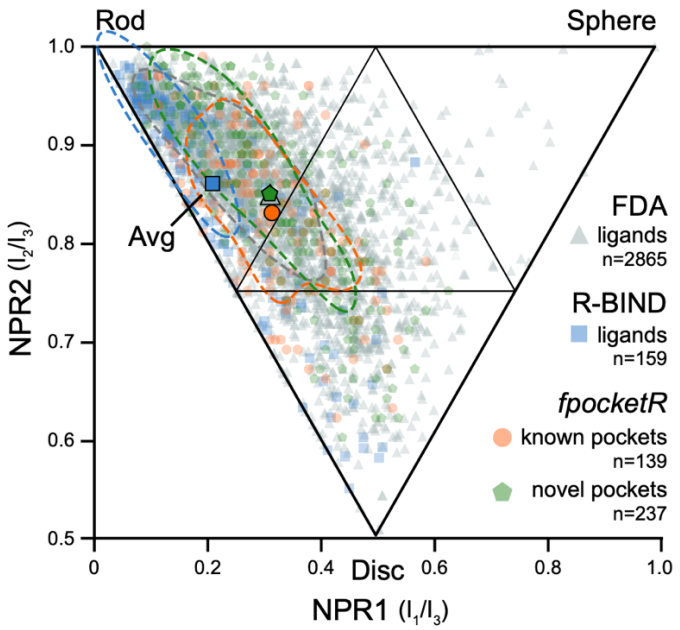
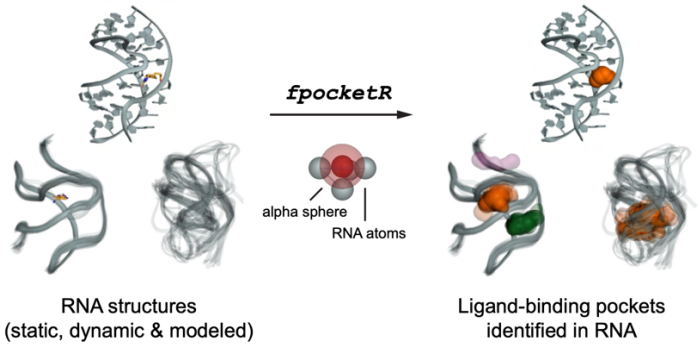


Figure 6. Comparison of shape-space for RNA-binding ligands versus RNA pockets. FDA-approved ligands (mostly protein binding), R-BIND ligands (exclusively RNA binding), and *fpocketR*-identified RNA pockets are shown. Normalized principal moment of inertia ratios (NPRs) for FDA-approved ligands (140-590 amu, n = 2865, grey triangles), R-BIND v2.1 small molecule ligands (n = 159, blue squares), and RNA pockets detected by *fpocketR* (known n = 139, orange circles; novel, n = 237, green pentagons). Each point represents the Boltzmann weighted average of NPRs for a single molecule using conformations within 3 kcal/mol of the lowest energy conformer. Large opaque symbols show the average shape for each category. Dashed lines represent the 50% contour line for each category. The distribution (Wasserstein) distances between FDA-approved ligands versus known RNA pockets, novel RNA pockets, and R-bind ligands are 0.038, 0.024, and 0.109, respectively; scale from 0-1, where 0 indicates identical distribution.

TOC Graphic


fpocketR
alpha sphere
RNA atoms
RNA structures
(static, dynamic & modeled)
Ligand-binding pockets
identified in RNA