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 package for identifying, characterizing, and visualizing ligand-binding sites in RNA. *fpocketR* was optimized, through 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 use 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 high (≤3.5 Å) resolution, but also provides useful information with lower resolution structures and computational models. *fpocketR* is a powerful, freely available tool for discovery and analysis of ligand-binding 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.

RNA molecules of any significant length routinely fold to form (base paired) secondary structures and a subset of these base-paired regions fold further to form higher-order tertiary structures. Small molecules can preferentially engage pockets formed within these complex tertiary RNA structures (4, 11, 12). However, platforms for reliably identifying pockets in RNA that bind ligands with favorable physicochemical properties (colloquially called drug-like) are underdeveloped. Furthermore, the distinct physicochemical properties and potentially dynamic nature of a subset of RNAs (13–15) suggest that current tools and heuristics require refinement for evaluating RNA interactions with small-molecule ligands. Prior studies exploring ligand binding pockets in RNA have primarily used tools developed for proteins and have not prioritized RNA pockets capable of binding drug-like ligands (16, 17).

We previously attempted to identify pockets in RNA using a widely used and broadly successful open-source, geometry-based software for finding pockets in proteins, *fpocket* (18, 19). Using default parameters, *fpocket* extensively over-predicts pockets in RNA. In essence, the algorithm mischaracterizes the grooves in RNA, which result from simple duplex formation, as pockets. RNA grooves are likely too polar, solvent-exposed, and similar to each other to interact selectively with small molecules that have favorable physicochemical properties (12). We found, however, that *fpocket* can be optimized to detect RNA pockets capable of binding ligands with high druglikeness (QED) scores (20). Using these optimized parameters, 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 to improve the positive predictive value from 19% for *fpocket* to 78%. We built on this optimization of *fpocket* and developed a pocket finding package specifically for RNA called *fpocketR* (12).

Several major conclusions 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 than simpler motifs like RNA bulges and loops. In this study, we illustrate how *fpocketR* can be leveraged to identify pockets in both experimental structures and in predicted RNA models. We share examples of pocket finding in large RNAs, modeled three-dimensional structures, RNA ensembles, and dynamic RNAs. We also characterize the geometric features of ligands able to bind high-quality pockets. We anticipate that *fpocketR* will be broadly useful to identify and characterize ligand-binding pockets in diverse RNA molecules.

METHODS

Software availability. This paper 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. A detailed tutorial that reviews all arguments and options for *fpocketR* is available. *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 (18, 19). 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 (PDB: 6wlq) (21) and secondary structure predicted by Fold from the RNAstructure package (22). These sequence and secondary structure files were then used to generate a multiple sequence alignment with rMSA (23) using the NCBI nucleotide and RNAcentral databases (24). The multiple sequence alignment and predicted secondary structure were input to trRosettaRNA (25) 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 allow 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 (26). The library was not filtered and contained 159 molecules.

FDA-approved ligand library. FDA-approved (Phase 4) molecules (27) (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 (28) (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) (29). 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 maximized unique RNA-ligand complexes and contains approximately 160 RNA structures, many of which bind the same ligands (for example, TPP and 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 (30). 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 (31), quantifying the multidimensional dissimilarity relative to the distribution of FDA-approved ligands.

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 (22), VARNA (32), and R2DT (33).

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 a provided 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**). *fpocketR* automatically detects features of the RNA structure and bound ligands to simplify required inputs. Users can customize pocket finding by manually specifying which RNA chain(s) and ligand to analyze. Additional options allow users to adjust *fpocketR* parameters, manage output files, and customize the output tertiary and secondary structure figures.

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 (28). *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 (21, 34, 35). 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) (36). 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) (37). 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 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 size and structure class, 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. In particular, RNA-only structures solved using cryo-EM have, until recently (35), had resolutions in the 4 to 10 Å range (34). Modeling of RNA structure using physics-based or machine learning methods is rapidly improving but still often exhibits large deviations from accepted structures (38, 39). 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) (40), 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 (38). The models varied widely in accuracy, compared to the accepted structure, with root mean square deviations (RMSDs) between 2.0 and 28 Å and template modeling scores

(TMscores) (41) 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**). Next, we 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 best, 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, solved by cryo-EM (PDB 6wql) (12, 41). We generated a 20-model ensemble of the SAM-IV riboswitch using trRosettaRNA, among the most successful machine-learning-based RNA modeling programs (25, 42), 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*. Nonetheless, the performance of *fpocketR* with modeled structures for the preQ1 and SAM-IV riboswitches indicates that RNA modeling softwares 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 (13, 14). During translation, the ribosome undergoes large-scale conformational changes that include rotation of the small and large ribosomal subunits in a rachet-like mechanism and swiveling of the head domain in the small ribosomal subunit (43, 44). 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 (45).

We used *fpocketR* to search for pockets in rRNAs 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) (44) (**Fig. 5A**). 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 (46). 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 (46) and macrocyclic peptides viomycin (47) and capreomycin (48) (**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) (20). 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. 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, 16, 17). 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 (49). Diverse prior work has suggested that RNA ligands and pockets tend to be more rod-like and have fewer sphere-like shapes as compared to protein ligands and pockets (16, 17, 26).

We calculate NPR values for the 376 pockets detected by *fpocketR* among non-redundant complexes in the Hariboss RNA-ligand database (<160 kDa; n = 364) (28) and a single reference structure of the bacterial ribosome (PDB 7k00) (29). This analysis used all available RNA-ligand complexes including RNAs in complex with the same ligands and with ligands having 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) 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) with distribution (Wasserstein) distances of 0.038 and 0.024, respectively.

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 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 (50). 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 reliably detects, characterizes, and visualizes high-quality ligand-binding pockets in RNA, in both individual structures and multi-state ensembles. *fpocketR* is broadly useful for examining 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* successfully 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, often 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.

ACKNOWLEGEMENT

This work was supported by grants from the US National Institutes of Health (R35 GM122532 to K.M.W. and 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. (2009) The Centrality of RNA. *Cell*, **136**, 577–580.
2. Cech,T.R. and Steitz,J.A. (2014) The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell*, **157**, 77–94.
3. Thomas,J.R. and Hergenrother,P.J. (2008) Targeting RNA with small molecules. *Chem Rev*, **108**, 1171–1224.
4. Warner,K.D., Hajdin,C.E. and Weeks,K.M. (2018) Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov*, **17**, 547–558.
5. Umuhire Juru,A. and Hargrove,A.E. (2021) Frameworks for targeting RNA with small molecules. *J Biol Chem*, **296**, 100191.
6. Childs-Disney,J.L., Yang,X., Gibaut,Q.M.R., Tong,Y., Batey,R.T. and Disney,M.D. (2022) Targeting RNA structures with small molecules. *Nat Rev Drug Discov*, **21**, 736–762.
7. Kovachka,S., Panosetti,M., Grimaldi,B., Azoulay,S., Di Giorgio,A. and Duca,M. (2024) Small molecule approaches to targeting RNA. *Nat Rev Chem*, **8**, 120–135.
8. Fullenkamp,C.R., Liang,X., Pettersson,M. and Schneekloth,J. (2024) Outlook. In Schneekloth,J., Pettersson,M., Mannhold,R., Buschmann,H., Holenz,J. (eds), *RNA as a Drug Target*. John Wiley & Sons, Ltd, pp. 355–384.
9. Moellering,R.C. (2003) Linezolid: The first oxazolidinone antimicrobial. *Ann Intern Med*, **138**, 135–142.
10. Ratni,H., Scalco,R.S. and Stephan,A.H. (2021) Risdiplam, the First Approved Small Molecule Splicing Modifier Drug as a Blueprint for Future Transformative Medicines. *ACS Med Chem Lett*, **12**, 874–877.
11. Connelly,C.M., Moon,M.H. and Schneekloth,J.S. (2016) The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol*, **23**, 1077–1090.
12. Veenbaas,S.D., Koehn,J.T., Irving,P.S., Lama,N.N. and Weeks,K.M. (2025) Ligand-binding pockets in RNA, and where to find them. *bioRxiv*, 10.1101/2025.03.13.643147.
13. Mustoe,A.M., Brooks,C.L. and Al-Hashimi,H.M. (2014) Hierarchy of RNA functional dynamics. *Annu Rev Biochem*, **83**, 441–466.
14. Ganser,L.R., Kelly,M.L., Herschlag,D. and Al-Hashimi,H.M. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat Rev Mol Cell Biol*, **20**, 474–489.
15. Bosio,S., Bernetti,M., Rocchia,W. and Masetti,M. (2024) Similarities and Differences in Ligand Binding to Protein and RNA Targets: The Case of Riboflavin. *J Chem Inf Model*, **64**, 4570–4586.
16. Hewitt,W.M., Calabrese,D.R. and Schneekloth,J.S. (2019) Evidence for ligandable sites in structured RNA throughout the Protein Data Bank. *Bioorg Med Chem*, **27**, 2253–2260.
17. Zhou,T., Wang,H., Zeng,C. and Zhao,Y. (2021) RPocket: an intuitive database of RNA pocket topology information with RNA-ligand data resources. *BMC Bioinformatics*, **22**, 428.
18. Le Guilloux,V., Schmidtke,P. and Tuffery,P. (2009) Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics*, **10**, 168.
19. Schmidtke,P., Le Guilloux,V., Maupetit,J. and Tufféry,P. (2010) fpocket: Online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res*, **38**, W582–W589.

20. Bickerton,G.R., Paolini,G. V., Besnard,J., Muresan,S. and Hopkins,A.L. (2012) Quantifying the chemical beauty of drugs. *Nat Chem*, **4**, 90–98.

21. Kappel,K., Zhang,K., Su,Z., Watkins,A.M., Kladwang,W., Li,S., Pintilie,G., Topkar,V. V., Rangan,R., Zheludev,I.N., *et al.* (2020) Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nature Methods 2020 17:7*, **17**, 699–707.

22. Reuter,J.S. and Mathews,D.H. (2010) RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics*, **11**, 129.

23. Zhang,C., Zhang,Y. and Pyle,A.M. (2023) rMSA: A Sequence Search and Alignment Algorithm to Improve RNA Structure Modeling. *J Mol Biol*, **435**, 167904.

24. 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., *et al.* (2019) RNAcentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res*, **47**, D221–D229.

25. Wang,W., Feng,C., Han,R., Wang,Z., Ye,L., Du,Z., Wei,H., Zhang,F., Peng,Z. and Yang,J. (2023) trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nat Commun*, **14**, 7266.

26. 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., *et al.* (2022) R-BIND 2.0: An Updated Database of Bioactive RNA-Targeting Small Molecules and Associated RNA Secondary Structures. *ACS Chem Biol*, **17**, 1556–1566.

27. Zdrazil,B., Felix,E., Hunter,F., Manners,E.J., Blackshaw,J., Corbett,S., de Veij,M., Ioannidis,H., Lopez,D.M., Mosquera,J.F., *et al.* (2024) The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res*, **52**, D1180–D1192.

28. Panei,F.P., Torchet,R., Mé Nager,H., Gkeka,P. and Bonomi,M. (2022) HARIBOSS: a curated database of RNA-small molecules structures to aid rational drug design. *Bioinformatics*, **38**, 4185–4193.

29. Watson,Z.L., Ward,F.R., Méheust,R., Ad,O., Schepartz,A., Banfield,J.F. and Cate,J.H.D. (2020) Structure of the bacterial ribosome at 2 Å resolution. *Elife*, **9**, e60482.

30. Greg Landrum and Julie Penzotti (2010) RDKit https://www.rdkit.org/.

31. Flamary,R., Vincent-Cuaz,C., Courty,N., Gramfort,A., Kachaiev,O., Quang Tran,H., David,L., Bonet,C., Cassereau,N., Gnassounou,T., *et al.* (2025) POT Python Optimal Transport.

32. Darty,K., Denise,A. and Ponty,Y. (2009) VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics*, **25**, 1974–1975.

33. 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., *et al.* (2021) R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nat. Commun.*, **12**, 1–12.

34. Ma,H., Jia,X., Zhang,K. and Su,Z. (2022) Cryo-EM advances in RNA structure determination. *Sig Transduct Target Ther*, **7**, 1–6.

35. Haack,D.B., Rudolfs,B., Jin,S., Khitun,A., Weeks,K.M. and Toor,N. (2025) Scaffold-enabled high-resolution cryo-EM structure determination of RNA. *Nat Commun*, **16**, 1–17.

36. 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., *et al.* (2023) Structure, folding and flexibility of co-transcriptional RNA origami. *Nat Nanotechnol*, **18**, 808–817.

37. Sampedro Vallina,N., McRae,E.K.S., Hansen,B.K., Boussebayle,A. and Andersen,E.S.

(2023) RNA origami scaffolds facilitate cryo-EM characterization of a Broccoli–Pepper aptamer FRET pair. *Nucleic Acids Res*, **51**, 4613–4624.

38. Das,R., Kretsch,R.C., Simpkin,A.J., Mulvaney,T., Pham,P., Rangan,R., Bu,F., Keegan,R.M., Topf,M., Rigden,D.J., *et al.* (2023) Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins:Struct., Funct., Bioinf.*, **91**, 1747–1770.
39. Magnus,M. and Miao,Z. (2023) RNA 3D Structure Comparison Using RNA-Puzzles Toolkit. *Methods Mol. Biol.*, **2586**, 263–285.
40. Schroeder,G.M., Kiliushik,D., Jenkins,J.L. and Wedekind,J.E. (2023) 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*, **299**, 105208.
41. Zhang,K., Li,S., Kappel,K., Pintilie,G., Su,Z., Mou,T.C., Schmid,M.F., Das,R. and Chiu,W. (2019) Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat Commun*, **10**, 5511.
42. Bernard,C., Postic,G., Ghannay,S. and Tahi,F. (2024) State-of-the-RNArt: benchmarking current methods for RNA 3D structure prediction. *NAR Genom Bioinform*, **6**.
43. Mohan,S., Donohue,J.P. and Noller,H.F. (2014) Molecular mechanics of 30S subunit head rotation. *Proc Natl Acad Sci U S A*, **111**, 13325–13330.
44. Rundlet,E.J., Holm,M., Schacherl,M., Natchiar,S.K., Altman,R.B., Spahn,C.M.T., Myasnikov,A.G. and Blanchard,S.C. (2021) Structural basis of early translocation events on the ribosome. *Nature*, **595**, 741–745.
45. Paternoga,H., Crowe-McAuliffe,C., Bock,L. V., Koller,T.O., Morici,M., Beckert,B., Myasnikov,A.G., Grubmüller,H., Nováček,J. and Wilson,D.N. (2023) Structural conservation of antibiotic interaction with ribosomes. *Nat Struct Mol Biol*, **30**, 1380–1392.
46. Bulkley,D., Johnson,F. and Steitz,T.A. (2012) The Antibiotic Thermorubin Inhibits Protein Synthesis by Binding to Inter-Subunit Bridge B2a of the Ribosome. *J Mol Biol*, **416**, 571–578.
47. Zhang,L., Wang,Y.H., Zhang,X., Lancaster,L., Zhou,J. and Noller,H.F. (2020) The structural basis for inhibition of ribosomal translocation by viomycin. *Proc Natl Acad Sci U S A*, **117**, 10271–10277.
48. Stanley,R.E., Blaha,G., Grodzicki,R.L., Strickler,M.D. and Steitz,T.A. (2010) The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nature Structural & Molecular Biology 2010 17:3*, **17**, 289–293.
49. Sauer,W.H.B. and Schwarz,M.K. (2003) Molecular shape diversity of combinatorial libraries: A prerequisite for broad bioactivity. *J Chem Inf Comput Sci*, **43**, 987–1003.
50. Fan,C., Wang,X., Ling,T., Yang,Y. and Zhao,H. (2023) Characterizing RNA-binding ligands on structures, chemical information, binding affinity and drug-likeness. *RNA Biol*, **20**, 431–443.

FIGURE LEGENDS

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 transparent backbone and pockets. In the map of pockets in secondary structure space, the intensity of colored nucleotides correlates with 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) (36) and (B) an RNA engineered to include Pepper (orange) and Broccoli (yellow) ligand-binding aptamers (374 nts) (37). 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) (40) and in 153 models of the class I type III preQ1 riboswitch, from the CASP15 exercise (38). (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 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) (21) or generated by trRosettaRNA (25) 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 (44). (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) (46), viomycin (6lkq) (47), and capreomycin (8ceu) (48).

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, 0 indicates identical distribution.

Figure 1

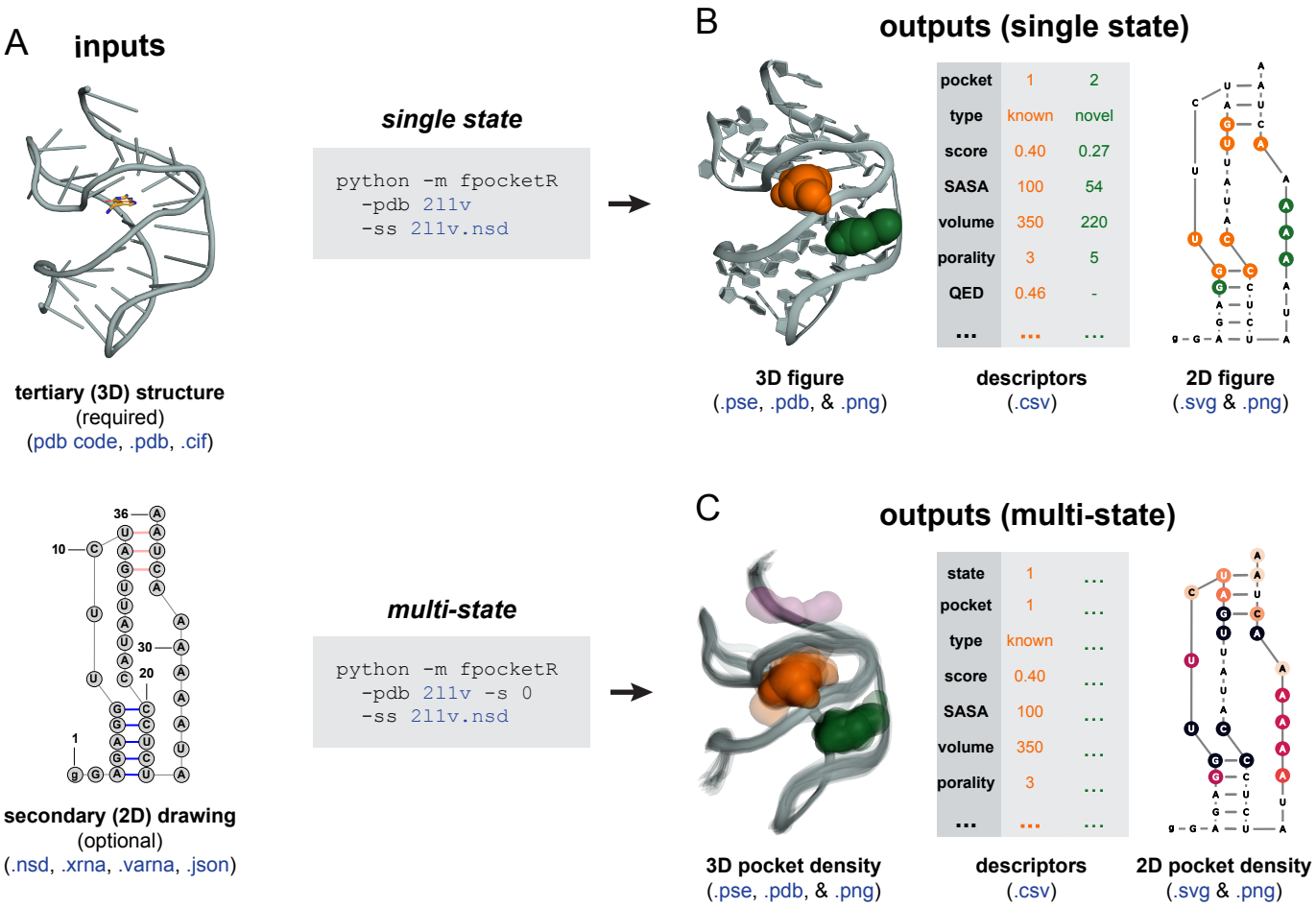

A
inputs
tertiary (3D) structure
(required)
(pdb code, .pdb, .cif)
secondary (2D) drawing
(optional)
(.nsd, .xrna, .varna, .json)
single state
python -m fpocketR
-pdb 2ll1v
-ss 2ll1v.nsd
multi-state
python -m fpocketR
-pdb 2ll1v -s 0
-ss 2ll1v.nsd
B
outputs (single state)
pocket 1 2
type known novel
score 0.40 0.27
SASA 100 54
volume 350 220
porality 3 5
QED 0.46 -
... ...
3D figure
(.pse, .pdb, & .png)
descriptors
(.csv)
2D figure
(.svg & .png)
C
outputs (multi-state)
state 1 ...
pocket 1 ...
type known ...
score 0.40 ...
SASA 100 ...
volume 350 ...
porality 3 ...
... ...
3D pocket density
(.pse, .pdb, & .png)
descriptors
(.csv)
2D pocket density
(.svg & .png)

Figure 2

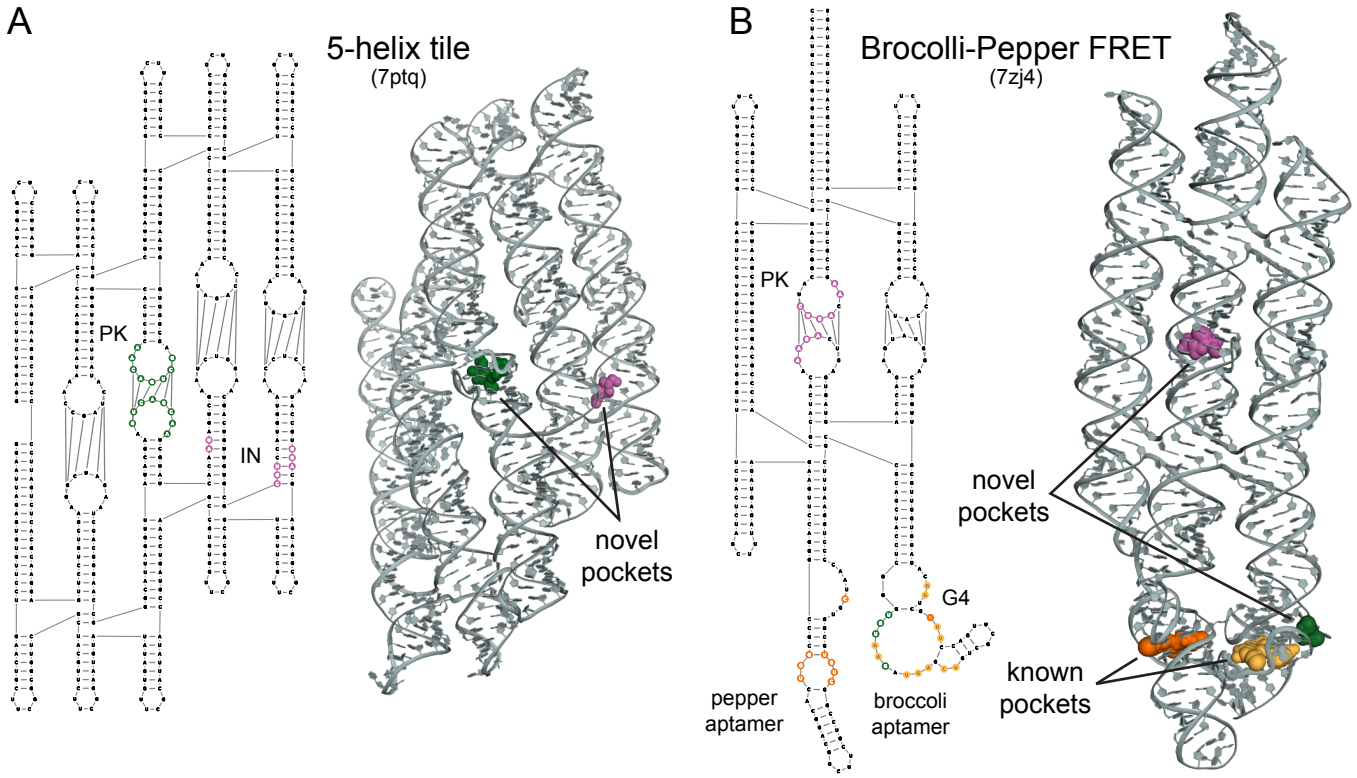

A
5-helix tile
(7ptq)
PK
IN
novel
pockets
B
Brocolli-Pepper FRET
(7zj4)
PK
G4
pepper
aptamer
broccoli
aptamer
novel
pockets
known
pockets

Figure 3

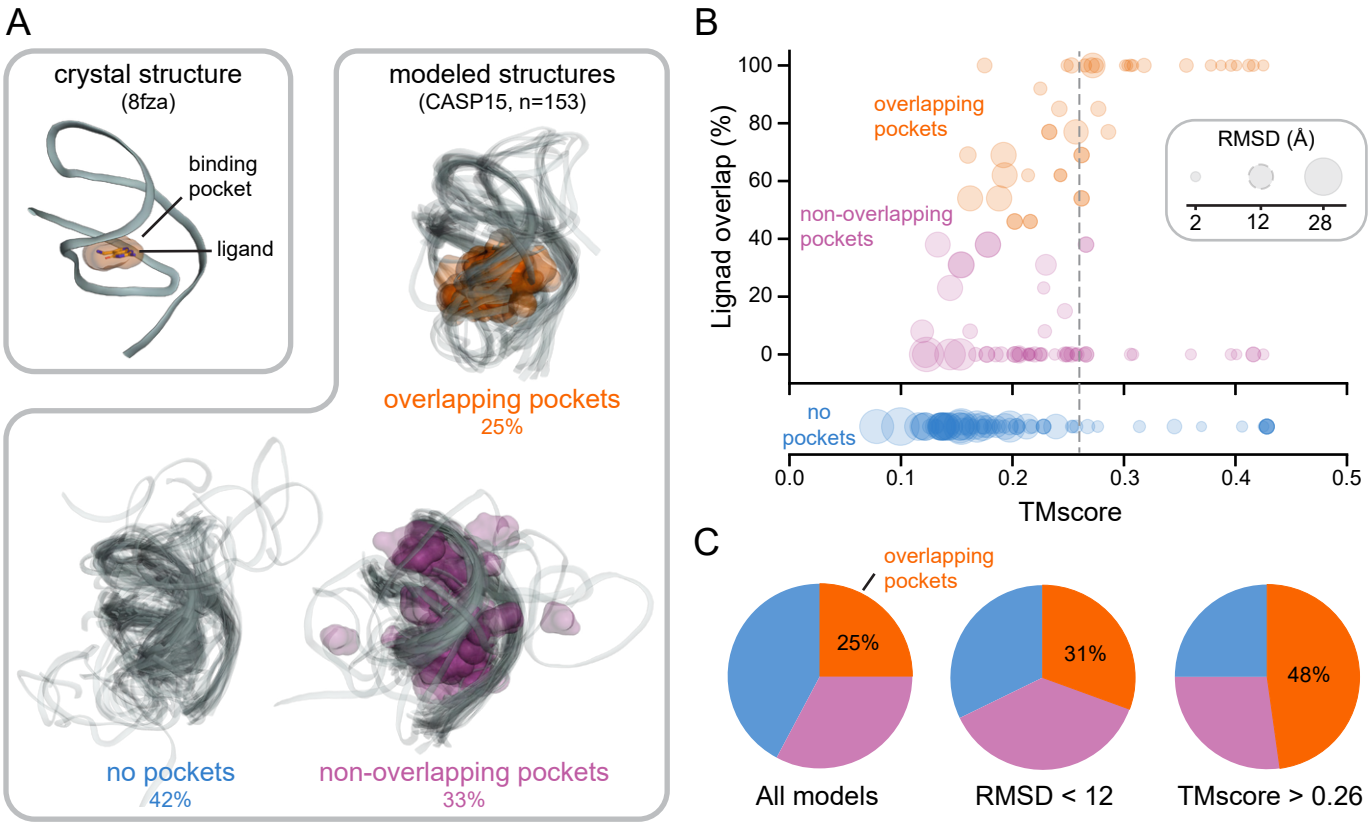

A
crystal structure
(8fza)
binding pocket
ligand
modeled structures
(CASP15, n=153)
overlapping pockets
25%
no pockets
42%
non-overlapping pockets
33%
B
Lignad overlap (%)
100
80
60
40
20
0
overlapping pockets
non-overlapping pockets
no pockets
RMSD (Å)
2
12
28
0.0
0.1
0.2
0.3
0.4
0.5
TMscore
C
overlapping pockets
25%
31%
48%
All models
RMSD < 12
TMscore > 0.26

Figure 4

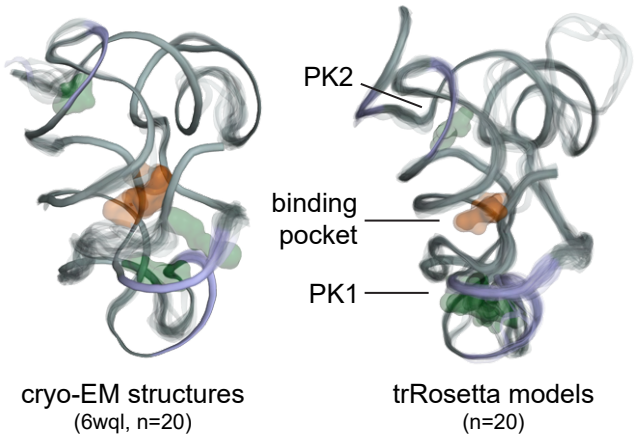

PK2
binding
pocket
PK1
cryo-EM structures
(6wql, n=20)
trRosetta models
(n=20)

Figure 5

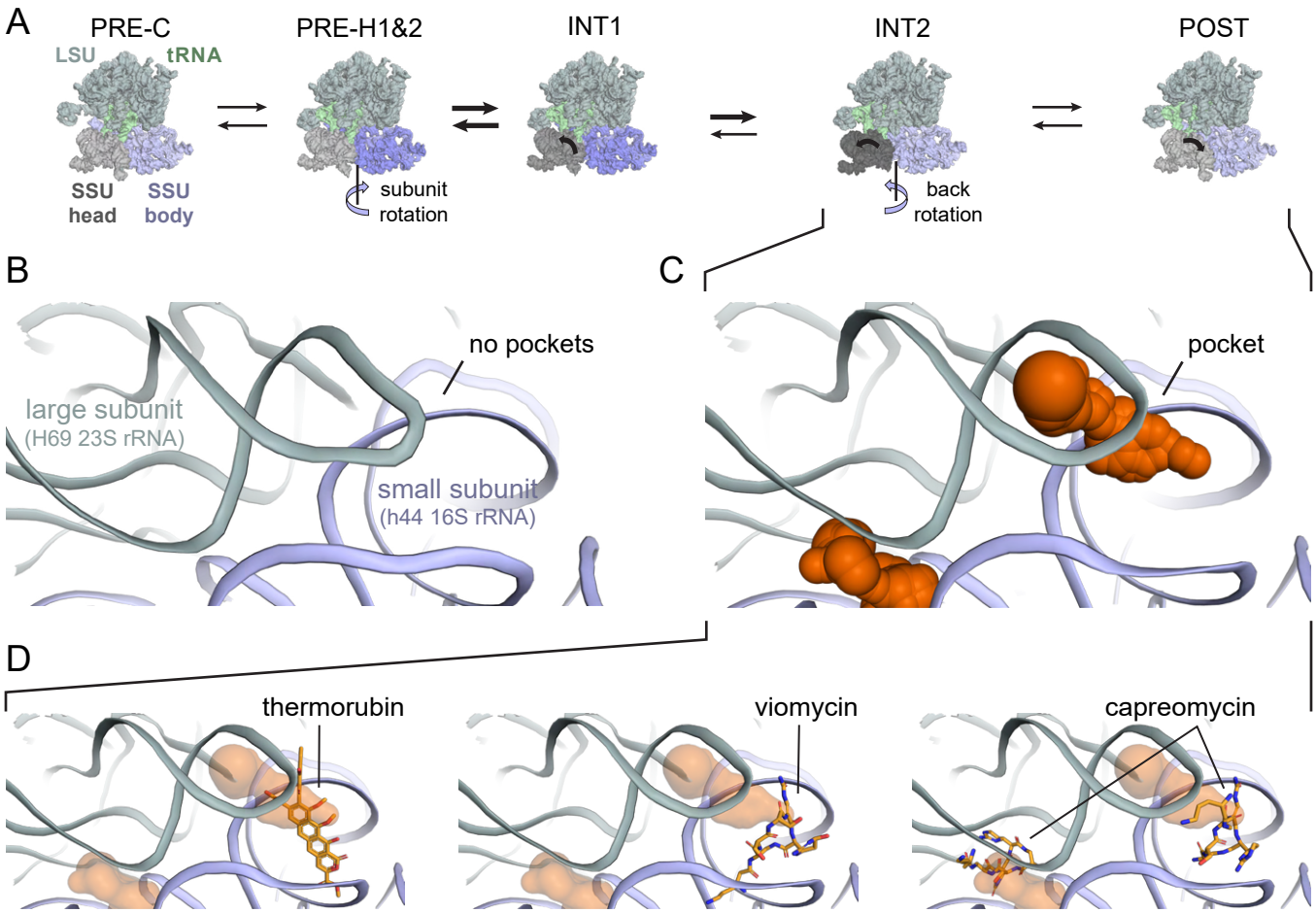
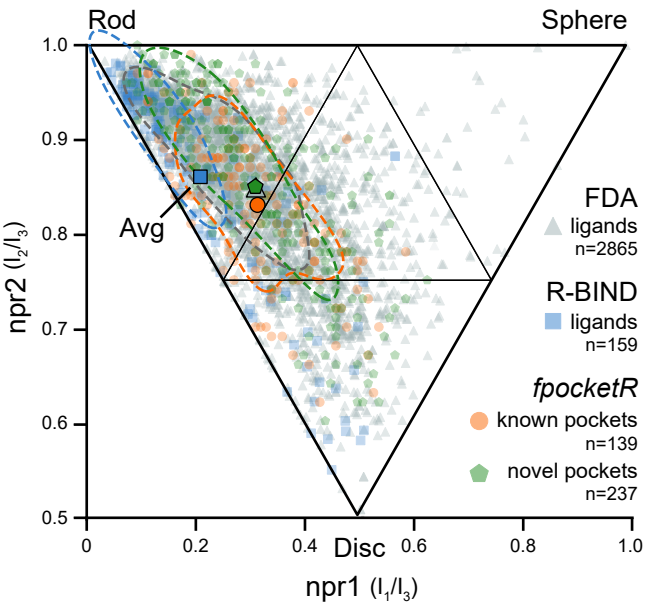

A
PRE-C
LSU
tRNA
SSU
head
SSU
body
PRE-H1&2
subunit
rotation
INT1
INT2
back
rotation
POST
B
no pockets
large subunit
(H69 23S rRNA)
small subunit
(h44 16S rRNA)
C
pocket
D
thermorubin
viomycin
capreomycin

Figure 6


Rod
Sphere
Disc
Avg
1.0
0.9
0.8
0.7
0.6
0.5
0
0.2
0.4
0.6
0.8
1.0
npr2 (I2/I3)
npr1 (I1/I3)
FDA
ligands
n=2865
R-BIND
ligands
n=159
fpocketR
known pockets
n=139
novel pockets
n=237