

**COMPUTATIONAL (FPOCKETR) AND EXPERIMENTAL (FRAG-MAP) DETERMINATION OF RNA
STRUCTURES THAT ENGAGE DRUG-LIKE SMALL-MOLECULE LIGANDS**

Seth D. Veenbaas

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment
of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry in the College
of Arts and Sciences.

Chapel Hill
2025

Approved by:

Kevin M. Weeks

Jeffrey Aubé

Alain Laederach

Dorothy A. Erie

Huong T. Kratochvil

Marc A. ter Horst

© 2025
Seth D. Veenbaas
ALL RIGHTS RESERVED

ABSTRACT

Seth D. Veenbaas: Computational (fpocketR) and experimental (Frag-MaP) determination of RNA structures that engage with small-molecule ligands
(Under the direction of Kevin M. Weeks)

RNAs are critical regulators of gene expression, and complex secondary and tertiary structures often mediate their functions. Structured regions in RNA can selectively interact with small molecules – via well-defined ligand binding pockets – to modulate the regulatory repertoire of an RNA. However, the broad potential to modulate biological function intentionally via RNA-ligand interactions remains unrealized due to challenges in identifying compact RNA motifs with the ability to bind ligands with good physicochemical properties (often termed drug-like). Here, we devise fpocketR, a software package 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. We experimentally confirmed the ligandability of novel pockets detected with fpocketR using a fragment-based approach introduced here, Frag-MaP, that detects ligand-binding sites in cells. Analysis of pockets detected by fpocketR and validated by Frag-MaP reveals dozens of newly identified sites able to bind drug-like ligands, supports a model for RNA secondary structural motifs able to bind quality ligands, and creates a broad framework for understanding the RNA ligand-ome. Lastly, we demonstrate fpocketR as a powerful discovery tool to analyze RNA-ligand interactions and novel pockets in small and large RNAs, to assess ensembles of experimentally determined and computationally predicted RNA structure models, and to identify pockets in dynamic RNA structural ensembles. Together, fpocketR and Frag-MaP create a powerful framework for understanding the potential of ligands to bind and manipulate cellular transcriptomes.

To my sweet Caroline, I couldn't have done this without you.

Thank you for getting me through graduate school with a smile on my face.

ACKNOWLEDGEMENTS

I have always been curious and scientifically minded, but I never imagined attending graduate school or being able to complete the daunting task of writing a dissertation. Completing my doctoral dissertation is my proudest, so far, and I owe it to a vast village of family, friends, and mentors who have loved, invested in, and believed in me.

The first members of my village are my parents and grandparents, who have always supported my endeavors, encouraged me to do hard things, and taught me to be a good citizen of the world. I am a curious and independent learner because of the relentless efforts of my Mom. My father taught me how to be positive, patient, creatively solve problems, and prepare for every challenge. My grandma and Cia taught me that failure was nothing to fear and did not define who I am. All these virtues served me well in graduate school.

I have been fortunate to share many of my successes and challenges in graduate school with my fiancée, Caroline Aufgebauer. I knew it was true love when you turned down a day of skiing to help me edit manuscripts. Thank you for bringing balance to my life, putting a smile on my face every day, and tackling life alongside me.

My years in graduate school have given me a deeper appreciation for the support and companionship of friends who join me on fun adventures and pitch in to help with the challenges of life. To my childhood friends (Avery Brinkley, Nicole Khattar, Christina Payton, James Payton, Andrew Henry, Stephen Hauck, and Nick Cowan), it means a lot to me that our friendship has endured from across the country. To my ski friends (Avery Brinkley (one day you'll be cool enough to ski), Adam Dew, Daniel Harper, Andrew Stallings, Ryan Sherrier, and Ben Kruse (I'm glad you have retired your snowboard)), the respite of our trips and the support of your friendship have gotten me through some of the most stressful parts of graduate school. To my non-science friends (Kristen Hefner, Kyle Lampart, and Christian Hunt), thank you for dragging me out of the lab and celebrating life with me. To the 540 crew (Ryan Sherrier, Ben

Kruse, and Caroline Aufgebauer), you made grad school infinitely more fun with banter, laughter, and many late nights.

I need to thank all the members of the Weeks and Laederach Laboratories for sharing their knowledge and skills, mentoring me through graduate school, and guiding my doctoral research with thoughtful feedback. I would especially like to thank Jordan Koehn for hosting me during my rotation, Patrick Irving for mentoring me throughout my tenure at UNC, and Catherine and Bree for never letting me get too lost in the lab. It has been a pleasure to work hard, play hard, and travel across the world with all the graduate students of the Weeks and Laederach Laboratories, and I hope our friendships last a lifetime.

Throughout my life, I have had countless teachers, coaches, and mentors whose life lessons continue to impact me. Thank you to Mrs. Villanueva, Jeff Abbot, Mrs. Bailey, and Chris DeMarchi for life lessons that stuck. Thank you to the faculty at Milligan College and Point Loma Nazarene University for teaching me to be a thoughtful skeptic and curious scholar. I am especially grateful to my undergraduate research advisor, Dr. Matthieu Rouffet, and academic advisor Dr. Katherine Maloney for preparing me to attend graduate school.

Finally, to the UNC Chemistry Department, my committee (Jeff Aubé, Alain Laederach, Dorothy Erie, Abigail Knight, Huong Kratochvil, and Marc ter Horst), and especially Kevin Weeks: Thank you for your mentorship, pushing me to be a better scientist, and fostering a collaborative environment.

Disclosure

Kevin M. Weeks is a founder at ForagR Medicines, Ribometrix, and A-Form Solutions.

TABLE OF CONTENTS

LIST OF TABLES	xi
LIST OF FIGURES.....	xii
LIST OF ABBREVIATIONS AND SYMBOLS	xiv
Chapter 1 : PRINCIPLES OF RNA THERAPEUTICS.....	1
RNA centrality	1
RNA structure and pockets	1
RNA as a therapeutic target.....	3
Challenges in RNA drug discovery	8
Research overview.....	10
Perspective	11
Chapter 2 : DEVELOPMENT OF FPOCKETR: IDENTIFYING LIGAND BINDING POCKETS IN RNA.....	13
Introduction	13
Results.....	14
Geometry-based analysis reveals differences in protein versus RNA pockets	14
A curated structure database of drug-like ligands bound to RNA.....	16
fpocketR: fpocket optimized for RNA.....	17
Discussion	17
Materials and methods	19
Curation of RNA-ligand complexes containing ligands with favorable (drug-like) physicochemical properties.....	19
Pocket evaluation metrics	20
Multivariate optimization of fpocket.....	20
fpocketR usage	21
Pocket visualization.....	21
Data availability.....	22

Chapter 3 : DEVELOPMENT OF FRAG-MaP AND VALIDATION OF FPOCKETR	23
Introduction	23
Results.....	24
Frag-MaP: Experimental validation of fpocketR predictions	24
Pocket ligandability: validation of sites for drug-like small molecules.....	26
Detection of pockets in unliganded (apo) structures.....	27
Tripling the known universe of RNA pockets able to bind drug-like ligands	29
RNA secondary structures that harbor pockets	32
Discussion	35
Materials and methods	36
Classifying pocket structural motifs.....	36
In-cell probing for Frag-MaP.....	36
Click labeling and RNA enrichment.....	37
MaP reverse transcription.	37
Library preparation and sequencing.	38
Analysis of Frag-MaP sequencing data.	38
Data availability.....	39
Chapter 4 : APPLICATIONS OF FPOCKETR: FINDING LIGAND-BINDING POCKETS IN RNA.....	40
Introduction	40
Materials and methods	41
Pocket finding.....	41
SAM-IV riboswitch 3D structure modeling	41
Multi-state analysis of the SAM-IV riboswitch.....	42
R-BIND library	42
FDA-approved ligand library	42
Hariboss RNA-ligand complex library	42
Principal moment of inertia analysis	43
Results.....	43
fpocketR workflow	43
Identification of pockets in large RNAs	45

Identification of pockets in low-resolution and modeled structures	47
Identification of transient pockets in dynamic RNA structures.....	49
Shapes of RNA pockets	52
Discussion	54
Data availability.....	55
APPENDIX A : SUPPLEMENTAL INFORMATION FOR CHAPTER 2	56
Supplemental figures	56
Supplementary Tables	58
APPENDIX B : SUPPLEMENTAL INFORMATION FOR CHAPTER 3.....	63
Materials and Methods	63
General chemistry information.....	63
General synthesis procedure for fully functionalized fragment probes.....	63
Synthesis of linezolid probe	64
Synthesis of methyl probe.....	64
Synthesis of 6-Aminoquinoxaline probe.....	65
Comparison of protein and RNA pocket detection and properties.....	65
Supplementary Figures	66
Supplementary Tables	88
APPENDIX C : FRAG-MAP PROTOCOL	95
Introduction	95
Probe cells and extract RNA	95
Partial Fragmentation of total RNA.....	98
Monarch RNA Cleanup kit	98
Click chemistry with alkyne-RNA & azide-oligo primer using 50% DMSO denaturant.....	99
RNA-Alkyne enrichment	100
Mutational profiling RT (DMS-optimized).....	101
Second Strand Synthesis	103
cDNA purification - AMpure beads.....	104
Qubit	104
DNA Library Prep (Adapter Ligation) - NEBNEXT Ultra II	105

Qubit	108
Agilent Tape Station.....	108
Pool and dilute library	109
Sequence Frag-MaP library (NextSeq)	110
APPENDIX D : SUPPLEMENTAL INFORMATION FOR CHAPTER 4.....	112
Supplemental Tables.....	112
REFERENCES.....	116

LIST OF TABLES

Table A.1. Training set library used to optimize fpocketR.....	58
Table A.2. Test set library used for evaluating fpocketR optimizations.....	59
Table A.3. Parameter combinations tested during 3 rounds of fpocketR optimization.....	60
Table A.4. Pocket detection performance of fpocket and fpocketR in a small RNAs.....	61
Table A.5. Pocket characteristics for RNA-ligand versus protein-ligand complexes.....	62
Table B.1. Detection of ligand binding pockets in paired apo and holo RNA structures.....	88
Table B.2. Pocket characteristics for small RNA library.....	89
Table B.3. Pocket characteristics for the <i>E. coli</i> 16S rRNA.....	91
Table B.4. Pocket characteristics for the <i>E. coli</i> 23S rRNA.....	92
Table B.5. Pocket characteristics for the group II intron.....	94
Table D.1. PDB codes for non-redundant Hariboss RNA-ligand complex library.....	112
Table D.2. Description of all pocket characteristics generated by fpocketR.....	113
Table D.3. Output files used and generated by fpocketR.....	114
Table D.4. Output files used and generated by fpocketR multistate analysis.....	115

LIST OF FIGURES

Figure 1.1. RNA structure.....	2
Figure 1.2. RNA-ligand interactions can have profound impacts on biological function.....	4
Figure 1.3. Expanding the druggable landscape of the human genome with RNA-target therapeutics.....	5
Figure 1.4. Protein and RNA bound to flavin mononucleotide (FMN).....	7
Figure 1.5. New methods for identifying RNA-ligand interactions.	12
Figure 2.1. Geometry-based pocket finding approach and optimization for detecting sites in RNAs that bind ligands with favorable physicochemical properties.....	15
Figure 2.2. Pockets detected in RNA structures with fpocketR.	18
Figure 3.1. Experimental detection of RNA-ligand binding sites by Frag-MaP.	25
Figure 3.2. Pockets detected by fpocketR in the bacterial rRNA, compared to small-molecule antibiotic binding sites.....	28
Figure 3.3. Detection of ligand-binding sites in apo (ligand-free) sites and in RNA ensembles.	30
Figure 3.4. Scheme for mapping RNA pockets, visualized in three-dimensional structures, onto secondary structure diagrams.....	31
Figure 3.5. Representative illustration of pockets formed by local versus long-range interactions, detected in a group II intron RNA.	33
Figure 3.6. Secondary structural motifs that form pockets able to bind ligands with (drug-like) high-quality physicochemical properties.....	34
Figure 4.1. Pocket finding workflow for fpocketR.....	44
Figure 4.2. Pocket detection in large engineered RNAs.....	46
Figure 4.3. Pocket detection performance in modeled RNA structures.....	48
Figure 4.4. Pocket detection in multi-state ensembles.	50
Figure 4.5. Identification of transient pockets at the interface of the large and small subunits of the bacterial ribosome, during translocation.....	51
Figure 4.6. Comparison of shape-space for RNA-binding ligands versus RNA pockets.....	53
Figure A.1. Parameter space evaluated during fpocketR optimization.	56
Figure A.2. Pockets detected in the SAM-III riboswitch using fpocketR	57
Figure B.1. The ribosomal pocket-ome. fpocketR detects 46 and 52 pockets in the classical (non-rotated) state of the 23S rRNA for E. coli and B. subtilis	66
Figure B.2. Frag-MaP sites identified in the B. subtilis 23S rRNA using the (A) linezolid, (B) quinoxaline, and (C) 206 functionalized probes.	70

Figure B.3. Distinct binding regions of ribosomal antibiotics near the peptidyl transferase center in the bacterial 23S rRNA.....	71
Figure B.4. Secondary structural motifs that form pockets in our curated non-redundant RNA-ligand library, featuring ligands with high QED values.....	76
Figure B.5. Secondary structural motifs that form pockets in each domain of the <i>E. coli</i> 16S and 23S rRNAs.....	87

LIST OF ABBREVIATIONS AND SYMBOLS

%	percent
°C	degrees Celsius
13C	NMR carbon-13 nuclear magnetic resonance
1H	NMR proton nuclear magnetic resonance
5-TAMRA	5-Carboxytetramethylrhodamine
Å	angstrom
apo	ligand-free
A-site	aminoacyl-site
ATCC	American Type Culture Collection
avg	arithmetic average
cDNA	complementary DNA
cm	centimeter
con loop	consecutive loop
cryo-EM	cryogenic electron microscopy
d	doublet
D	cluster distance (Å)
DCM	dichloromethane
dd	doublet of doublets
DFAME	3-[4-(difluoromethyl)phenyl]-2-[(2-methylpropan-2-yl)oxygenylamino]propanoic acid
DFHBI	3,5-difluoro-4-hydroxybenzylidene imidazolinone
DFHBI-1T	(Z)-4-(3,5-difluoro-4-hydroxybenzylidene)-2-methyl-1-(2,2,2-trifluoroethyl)-1H-imidazol-5(4 H)-one
DIPEA	N,N-Diisopropylethylamine
dis	distant long-range interaction
DMHBI	3,5-dimethoxy-4-hydroxybenzylidene imidazolone
DMS	dimethyl sulfate
DMSO	dimethyl sulfoxide

DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDC-HCl	N-Ethyl-N'-(3-dimethylaminopropyl)carbodiimide hydrochloride
eRMSD	estimated root mean square deviation
ESI-QTOF	electrospray ionization quadrupole time of flight
FDA	food and drug administration
EtOAc	ethyl acetate
FMN	flavin mononucleotide
Frag	small molecule fragment
g	gravitational constant
G4	G-quadruplex
GEO	gene expression omnibus
HBC	benzylidene-cyanophenyl
HIV-1	human Immunodeficiency virus-1
HOBt	1-Hydroxybenzotriazole hydrate
holo	ligand-bound
HRMS	high resolution mass spectrometry
Hz	Hertz
i	number of alpha spheres
J	coupling constant
J	joules
kcal	kilocalorie
LSU	large ribosomal subunit
M	molar
m	minimum alpha sphere radius
M	maximum alpha sphere radius
m/z	mass to charge ratio
MaP	mutational profiling
Methanol-d4	deuterated methanol
mg	milligram

MHJ	multihelix junction
MHz	megahertz
min	minute
mL	milliliter
mM	millimolar
mmol	millimole
mRNA	messenger RNA
MW	molecular weight
n	sample size
ncRNA	noncoding RNA
NEB	new england biolabs
ng	nanogram
NMR	nuclear magnetic resonance
NPR	normalize principal ratio
nt	nucleotides
OD600	optical density at 600 nm
PDB	protein data bank
pH	potential of hydrogen
PK	pseudoknot
pM	picomolar
PPDA	pyrimido[4,5-d]pyrimidine-2,4-diamine
ppm	parts per million
ppv	positive-predictive value
PreQ1	pre-queuosine1
P-site	peptidyl-site
PTC	peptidyl transfer center
PTLC	preparative thin layer chromatography
QED	quantitative estimate of drug-likeness
RMSD	root mean square deviation
RNA	ribonucleic acid

RNP	ribonuleoprotein
rpm	revolutions per minute
rRNA	ribosomal RNA
RT	reverse transcriptase
S	Svedberg unit
SAH	S-Adenosyl homocysteine
SAM	S-Adenosyl methionine
SASA	solvent accessible surface area
sens	sensitivity
SHAPE	selective 2'-hydroxyl acylation and primer extension
SSU	small ribosomal subunit
TAL2	theophylline aptamer ligand 2
TAR	trans-activation response element
TBTA	tris(benzyltriazolylmethyl)amine
TLC	thin layer chromatography
TM-score	template modeling score
TPP	thiamine pyrophosphate
Tris	tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
TWEEN-20	polyoxyethylene (20) sorbitan monolaurate
UV	ultraviolet
YO3-biotin	oxazole yellow biotin
ZMP	5-amino-4-imidazole carboxamide ribose-5'-monophosphate
δ	chemical shift
μg	microgram
μL	microliter
μM	micromolar
$\pi-\pi$	pi-pi interaction

CHAPTER 1: PRINCIPLES OF RNA THERAPEUTICS

RNA centrality

RNA is the primary conduit of genetic information transfer in all cells. Messenger RNAs (mRNAs) are essential intermediaries between DNA and proteins. RNAs also serve as multifaceted information carriers, regulators, and catalysts for nearly all cellular functions (1, 2). Strikingly, approximately 70% of the human genome is transcribed into non-coding RNAs (ncRNAs) (3). ncRNAs are not translated into proteins but can themselves catalyze biochemical reactions (4), silence chromosomes (5, 6), and fine-tune gene expression (7, 8). Critically, RNA operates upstream of nearly all cellular processes. No other macromolecules cover such a vast scope of biological functions, making RNA a unique and useful target for modulating biology (9).

RNA structure and pockets

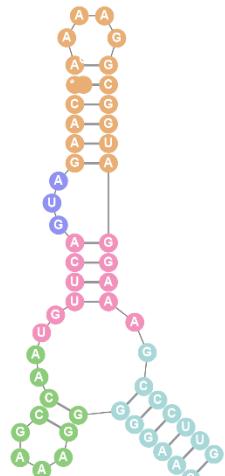
RNA is a biopolymer made of ribonucleic acid building blocks, almost identical to the deoxyribonucleic acid building blocks of DNA. The sequence of nucleic acids (A, U, G, C) is what allows RNA (and DNA) to carry genetic information (**Figure 1.1A**). In a cellular context, DNA is persistently accompanied by a complementary DNA molecule, largely limiting DNA to form simple and predictable double-helix structures via intermolecular Watson-Crick base pairs (10). RNA molecules are not typically accompanied by complementary pairs (or are “single-stranded”), allowing RNA to form intramolecular interactions that create diverse secondary structures. RNA secondary structures are formed by intramolecular interactions, including base-stacking, Watson-Crick base pairing, and prevalent “non-canonical” base pairing that pair nucleotides far away in the RNA sequence (FIG. 1A) (11). The secondary structure motifs discussed in this work include bulges, loops, consecutive bulges/loops, multi-helix junctions, and pseudoknots (**Figure 1.1B**). The base stacking patterns at junctions of unpaired nucleotides, non-canonical interactions, and long-range base pairing can further transform RNA secondary into an even more complex three-dimensional tertiary structure (12). Finally, RNAs can interact

A

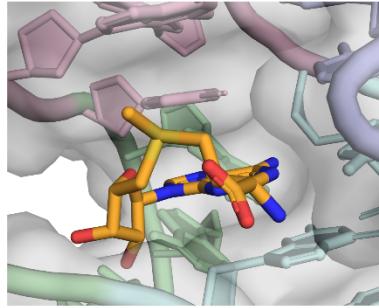
GUUCCGAAAGGAUGGC GGAAACGCCAGAUGCCUUGUAACCGAAAGGGGGAAC

primary structure

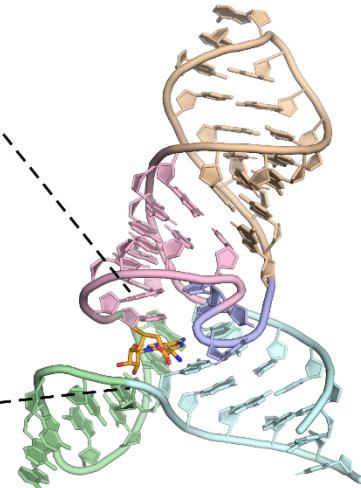
sequence of SAM-III riboswitch (PDB: 3e5c)



secondary structure



RNA-ligand binding interaction



tertiary structure

B

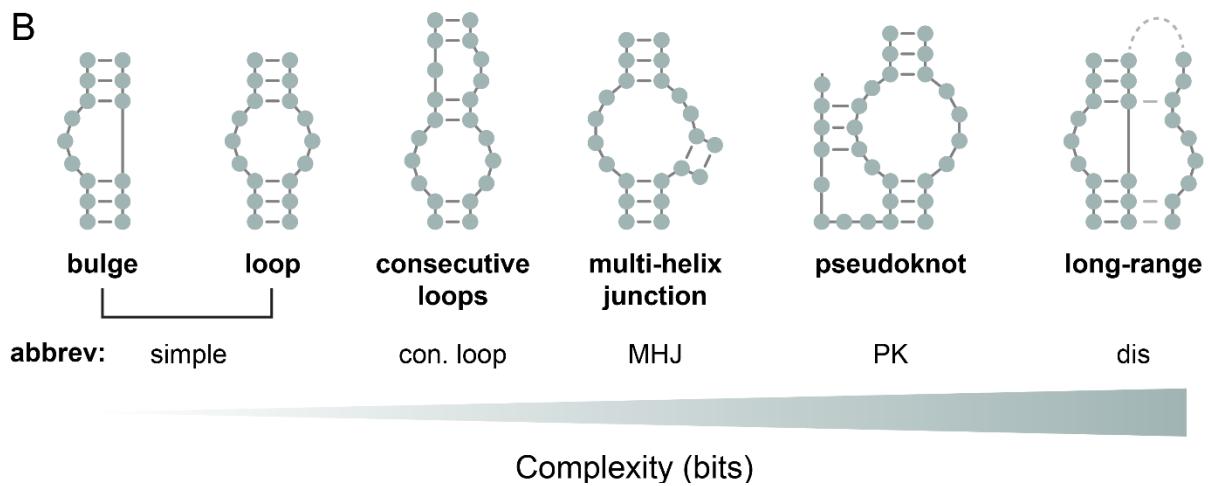


Figure 1.1. RNA structure (A) is comprised of a primary structure (sequence). An RNA can form intermolecular base pairs to bring nucleotides that are far away in sequence near to one another in secondary structure. The combination of all the intermolecular interactions in an RNA molecule can lead to a complex three-dimensional tertiary structure capable of forming cavities or pockets that can interact with small-molecule ligands such as S-adenosyl-L-methionine (SAM). (B) Common RNA secondary structural motifs and their abbreviations used in this work sorted by their complexity (information content).

with other RNAs and proteins to form quaternary complexes, often referred to as ribonucleoproteins (RNPs). The structural and catalytic properties of RNA are often central to the function of RNP complexes, as demonstrated in the ribosome, spliceosome, and telomerase (13, 14). All levels of RNA structure contribute to its versatile functionality and enable essential roles in storing genetic information, catalysis, ligand sensing, and macromolecular assembly.

A subset of RNAs fold into tertiary structures that contain pockets or clefts capable of interacting with small-molecule ligands. Riboswitches are a diverse class of naturally occurring small-molecule aptamers that have evolved as gene regulatory platforms in the 5' untranslated region of bacterial coding RNAs (**Figure 1.1A**) (15). Upon binding, ligand-induced conformational changes in the riboswitch structure shift the structural equilibrium of the riboswitch, resulting in *cis* regulation of gene expression (**Figure 1.2A**) (16, 17). The ribosome is a massive (2.5 MDa) RNP complex and ribozyme essential for protein production. Because of its biological importance, the bacterial ribosome is a primary target for small-molecule antibiotics. Several known binding sites engage small-molecule antibiotics in both the 23S and 16S rRNAs that act by inhibiting myriad ribosomal mechanisms (e.g., initiation, peptidyl transfer, termination, chain elongation) (**Figure 1.2B**) (18). Riboswitches and ribosomal antibiotics are both profound examples of how RNA structure informs function and how an entire biological system can be regulated by a simple small molecule.

RNA as a therapeutic target

The vast majority of FDA-approved small-molecule therapeutics target disease-related proteins. The current pharmacopoeia can target ~700 of the ~3,000 known disease-related proteins (19). However, proteins only represent ~1.5% of the human genome, meaning that scientists have only managed to target ~0.04% of the human genome with small molecules (**Figure 1.3**) (20, 21). Furthermore, many disease-related proteins have proven difficult to selectively target because they lack well-defined ligand-binding pockets (22).

RNAs are central to cellular functions and, similarly to proteins, their function can respond to interactions with small molecules. Coding RNAs (~1.5 % of the human genome) are directly upstream of all disease-related proteins. If coding mRNAs could be therapeutically targeted, their downstream disease-related proteins could be modulated, providing an alternative route for targeting challenging

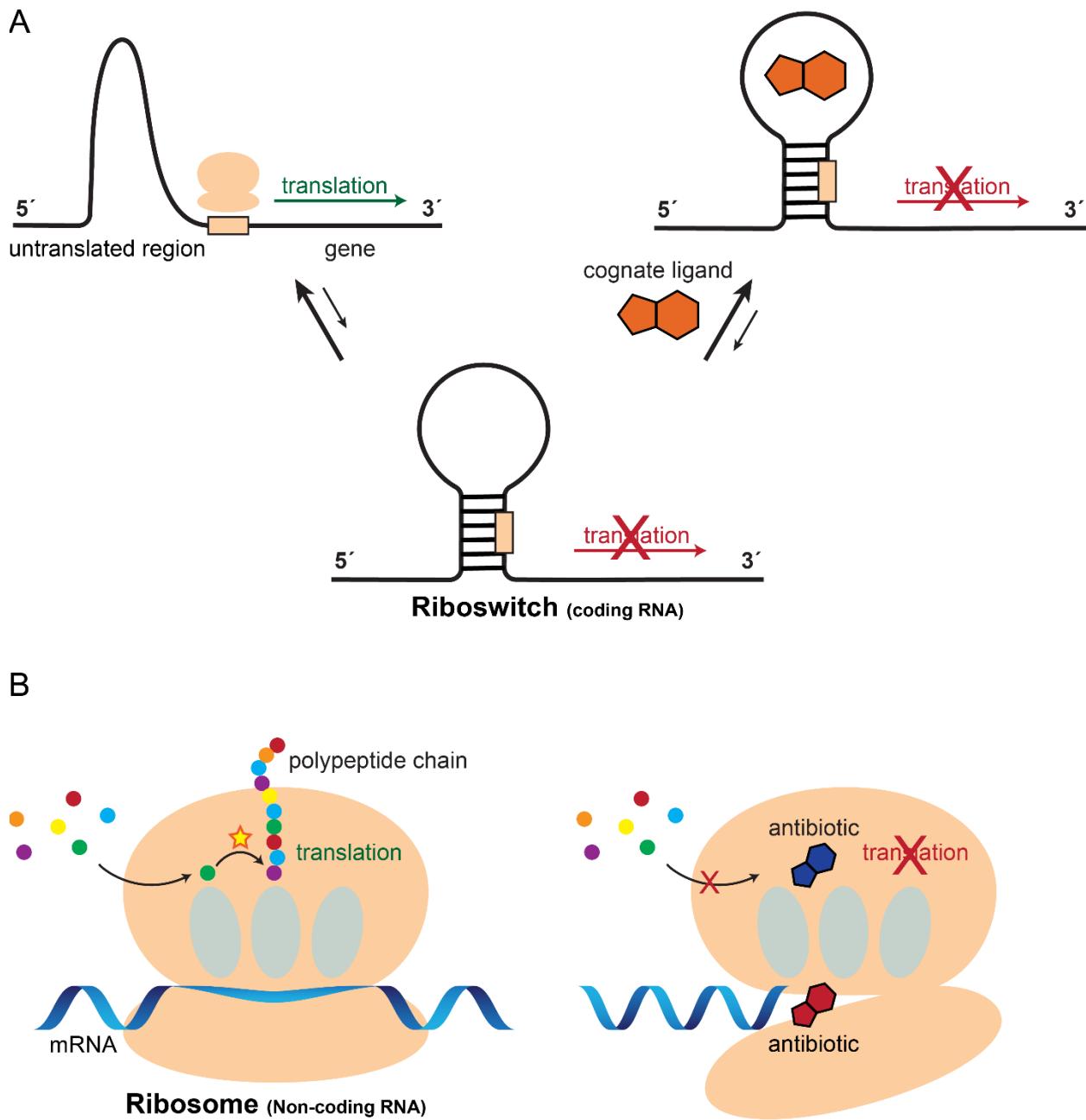


Figure 1.2. RNA-ligand interactions can have profound impacts on biological function, especially protein expression. (A) Bacterial riboswitches prototypically exist in an equilibrium of structure(s) that allow translation and structure(s) that prevent translation. The interactions from a cognate ligand can bind to one of these structures, selectively shifting the equilibrium of the RNA structures and affecting protein expression levels of the downstream gene(s). (B) The ribosome is a ribozyme essential for catalyzing protein synthesis and a primary target for small-molecule antibiotics. Ribosomal antibiotics (e.g., linezolid, capreomycin) have multiple binding sites (e.g., peptidyl transferase center, b2a inter-subunit bridge) and modes of action (e.g., inhibiting peptidyl transfer, translocation) (18).

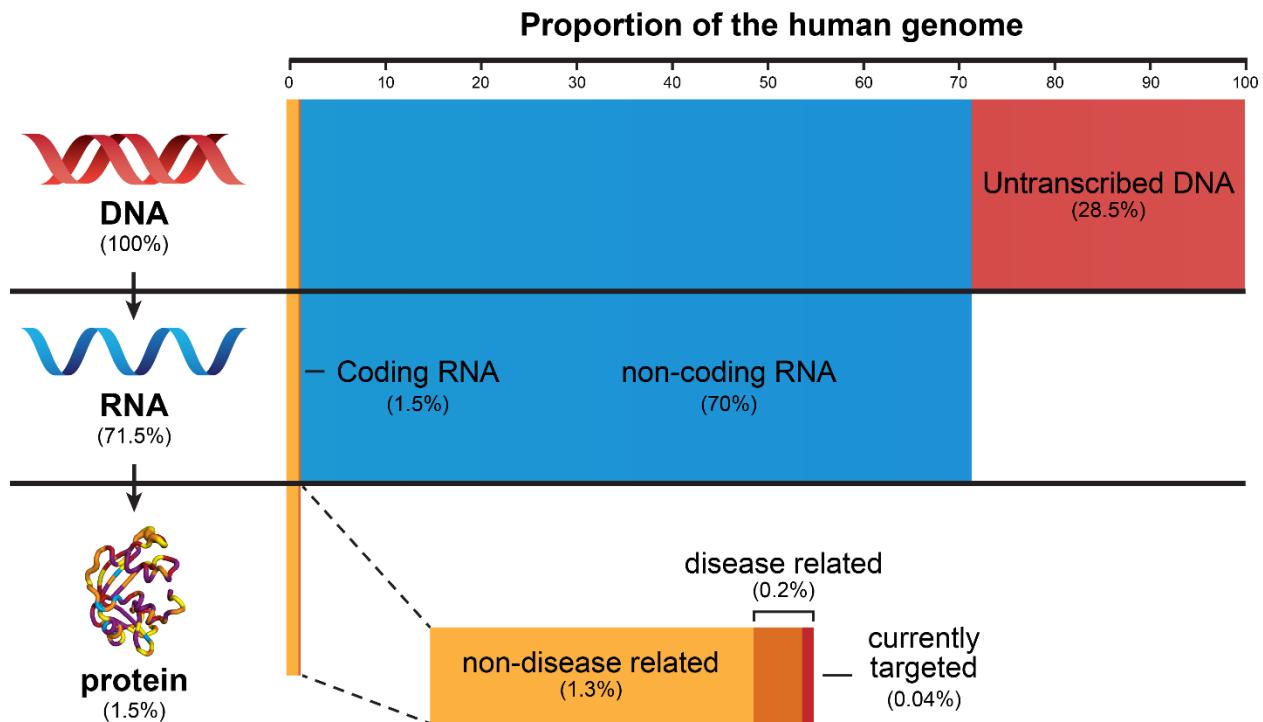


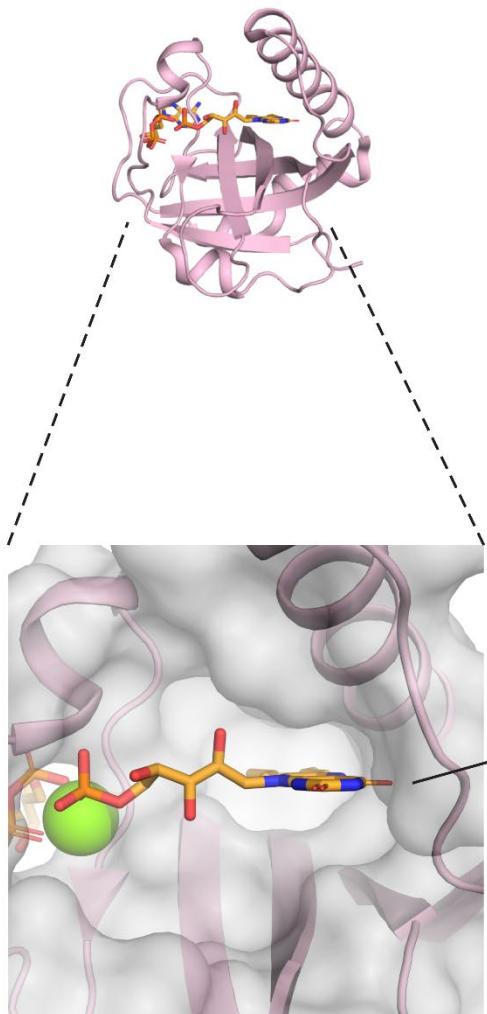
Figure 1.3. Expanding the druggable landscape of the human genome with RNA-target therapeutics. The information in the human genome is stored in DNA, transcribed into RNA, and some of that RNA (1.5%) is eventually translated into about 20,000 human proteins. There are about 3,000 disease-related proteins (orange) representing about 0.2% of the human genome. FDA-approved drugs target a small fraction of these disease-related proteins, representing an exceedingly small sliver (red) of the human genome (~0.05%). In contrast, >70% of the human genome is transcribed into coding and non-coding RNAs (blue and yellow), presenting a vast scope of potential therapeutic targets that are upstream of essentially all biological processes.

therapeutic targets (9). Additionally, a majority of the transcriptome (~70% of the human genome) encodes non-coding RNAs (**Figure 1.3**), many of which have important disease-related functions, such as microRNAs, intronic RNAs, and long non-coding RNAs (23). Targeting microRNAs and intronic RNAs represents the most promising near-term opportunity, as their biological mechanisms are the most well-understood. In the long term, we will likely uncover myriad applications for RNA-targeted small-molecule therapeutics as we continue to investigate the function and mechanism of the estimated 15,000 long non-coding RNAs in the human transcriptome (20).

Fundamentally, the physical and chemical interactions that allow proteins to bind to drug-like small molecules also exist in RNA. RNA pockets interact with small molecule ligands via base stacking, Van der Waals interactions, hydrogen bonding, and in some cases, metal ion-mediated binding (24). Additionally, pockets in RNA have similar bulk properties to proteins, with the notable exception that RNA molecules are significantly more polar (25, 26). These fundamental similarities allow both proteins and RNA to recognize the same small-molecule ligands. The protein riboflavin kinase (**Figure 1.4A**) (27) and RNA FMN riboswitch (**Figure 1.4B**) (28) both recognize the flavin mononucleotide ligand via similarly shaped pockets and strikingly similar intermolecular interactions, including stacking/Van der Waals, hydrogen-bonding, and ion coordination (9). Good protein and RNA ligand binding pockets have properties such as large volume, buriedness, low solvent accessibility, and sufficient chemical complexity to interact with drug-like ligands (29).

RNA-ligand interaction can affect biological functions through either competitively binding to a protein-binding motif or stabilizing/changing RNA structure in a similar mechanism to a riboswitch aptamer (30). These interactions can lead to either upregulation or downregulation of key cellular processes, making RNA an attractive therapeutic target (9, 31). Despite the promise of RNA therapeutics, the examples of successful human-designed RNA-targeted drugs are largely limited to linezolid, an antibiotic that binds the ribosome (32), and risdiplam and branaplam, splicing modifiers that bind pre-messenger RNA (33). These examples demonstrate how rationally designed small molecules can selectively target RNA to achieve therapeutic outcomes. However, the limited examples of RNA-targeted small molecules also highlight the current lack of knowledge and tools needed to reproducibly target RNAs.

A



B

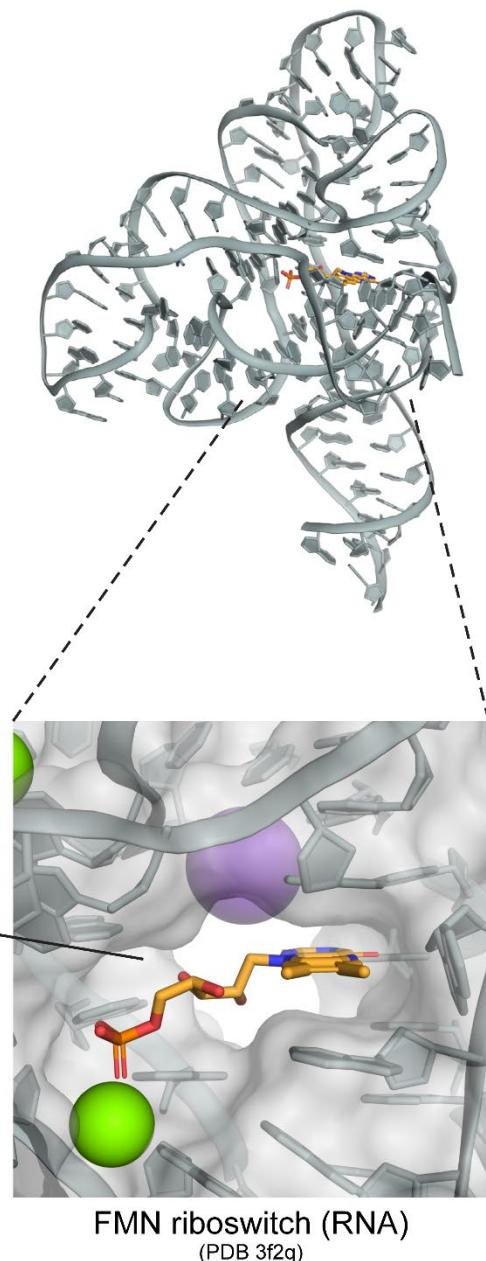


Figure 1.4. Protein and RNA bound to flavin mononucleotide (FMN).¹ (A) Riboflavin kinase protein (PDB 1nb9) and (B) FMN riboswitch (PDB 3f2q) can both form selective interactions with the FMN ligand. In both macromolecules, FMN has a similar pose and modes of interaction (H-bonding, stacking, and Van der Waals).

¹ Figure adapted from K. D. Warner, C. E. Hajdin, K. M. Weeks, Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* **17**, 547–558 (2018).

Challenges in RNA drug discovery

Targeting RNA with small molecules presents unique challenges because many available drug discovery tools were optimized for proteins and often fail to account for the distinct physicochemical properties of RNA molecules (9). High-resolution structural characterization methods are largely limited to “well-behaved” RNA. As a result, less than 1% of PDB deposits are RNA-only structures (34). Computational approaches to structure determination also face challenges in accurately generating force fields and scoring RNA-ligand interactions because of the polarity of the polyanionic backbones and the lack of hydrophobic pockets in RNA molecules (35). Furthermore, many biophysical techniques are not practical for RNA because they rely on large ligand-induced conformational changes, have significant challenges when applied to larger RNAs, or require prohibitively large quantities of pure RNA (36). In summary, three key challenges in RNA-targeted drug discovery are: 1) developing high-throughput methods to resolve RNA 3D structures at high resolution, 2) improving detection of transient or cryptic ligand-binding pockets in RNA, and 3) validating RNA-ligand interactions in cellular environments. Overcoming these hurdles will require innovative tools tailored to RNA’s unique chemistry and structural diversity.

A central challenge in RNA-targeted drug discovery is resolving tertiary structures and RNA-ligand complexes. Only a few dozen RNA-ligand complexes, involving small molecules with favorable (drug-like) physicochemical properties, have been visualized at high resolution. While RNA secondary structures are routinely mapped using chemical probing (e.g., SHAPE-MaP, DMS-MaP) and computational tools (37, 38), tertiary structure determination has historically lagged due to RNA’s high flexibility, electrostatic repulsion, and conformational heterogeneity (39). Recent advances in cryo-electron microscopy (cryo-EM) and machine learning-based computational modeling have begun to address this gap. Innovations like RNA-based scaffolds are enabling higher resolution cryo-EM imaging of flexible RNAs by stabilizing transient conformations and overcoming traditional size limitations (34, 40). Tools like trRosettaRNA and RhoFold+ predict 3D RNA structures by leveraging deep learning-trained evolutionary couplings and geometric constraints (41, 42). Together, these structural determination methods have significantly expanded the number of published RNA tertiary structures, capturing dynamic

ensembles of riboswitches, viral RNAs, and spliceosomes in multiple functional states. However, both approaches typically produce lower-resolution models (~2–10 Å for cryo-EM; ~2–12+ Å for computational predictions) (41–43) compared to X-ray crystallography (~1.5–4 Å) (39), limiting their ability to resolve small-molecule ligands or subtle backbone rearrangements. Computational models remain less reliable for RNAs with low sequence conservation or conformational diversity. Downstream analytical tools (e.g., molecular docking, simulation, and pocket finding software) that wish to take advantage of structures generated using cryo-EM and computational modeling need to accommodate the resolution limits and inherent structural flexibility.

Identifying targetable structures in RNA remains a major challenge for drug discovery. Tools like PocketFinder, SiteMap, and fpocket, designed to find deep, hydrophobic pockets in proteins, struggle with RNA's flatter, more charged surfaces (9). Newer RNA-focused methods, such as the ROBIN database and SHAMAN molecular dynamics software, are advancing our ability to map RNA-ligand interactions, especially in dynamic ensembles (44, 45). However, these tools still rely heavily on simple RNA structures (e.g., stem-loops) bound to non-drug-like molecules. Non-drug-like molecules such as aminoglycosides often bind RNA's major/minor grooves through hydrogen bonds rather than occupying defined pockets (46). Compounds that bind to RNA grooves or intercalate between bases as a simple bulge/loop often lack the pharmacokinetic properties (e.g., oral bioavailability, cell permeability) needed for therapeutics (9). To move beyond academic probes, the field must prioritize two goals: 1) developing computational tools that detect transient RNA pockets in dynamic structures and 2) designing compact, drug-like molecules that balance potency with favorable pharmacokinetic properties.

Validating RNA-ligand interactions is technically challenging. Biophysical techniques like ITC and SPR can characterize RNA-ligand interactions by measuring binding affinity, kinetics, and structural changes with high precision (47). However, these methods require purified RNA and are low-throughput, making them impractical for screening large libraries or studying interactions across multiple RNAs simultaneously (31, 36). In contrast, sequencing-based approaches like Chem-CLIP-Frag-Map and Pearl-Seq theoretically enable high-throughput discovery by identifying RNA-ligand interactions transcriptome-wide in a single experiment (48, 49). However, in practice, these tools can be noisy due to their reliance on biotin pull-downs and low sensitivity due to their reverse transcription stop (RT-stop) experimental

readout. Furthermore, they require gene-specific follow-up experiments to resolve the RNA-ligand binding sites at nucleotide resolution. Mutational profiling (MaP) is an alternative approach to traditional RT-stop that relies on error-prone reverse transcription to encode the location of chemical adducts on RNA as mutations that can be identified via next-generation sequencing. MaP has better signal-to-noise ratios and sensitivity than RT-stop because it directly measures chemical adducts on RNA molecules (50, 51). Integrating MaP into ligand discovery experiments could enable high-confidence, transcriptome-wide discovery of RNA-targetable sites with nucleotide resolution.

Despite significant progress in RNA structural determination, key gaps persist in identifying RNA structures that can be targeted by drug-like small-molecule ligands. Addressing these challenges requires optimizing computational tools for RNA-specific pocket detection and advancing high-throughput experimental methods. This work focuses on bridging these gaps through the development of fpocketR, an RNA-optimized pocket detection algorithm, and Frag-MaP, a MaP-based transcriptome-wide ligand mapping approach.

Research overview

The primary goal of this research is to develop tools to determine what types of RNA structures are likely to engage with drug-like small-molecule ligands. Unlike existing approaches, this work is differentiated by three critical innovations. First, we optimize pocket detection algorithms explicitly for RNA's unique physicochemical properties, including its high polarity, charged backbone, and dynamic conformations. Second, we train and validate these tools on a carefully curated dataset of RNA-ligand complexes with drug-like small molecules, rather than non-drug-like ligands (e.g., aminoglycosides, groove binders, intercalators). Third, we use both computational and MaP-based technologies to validate the ligandability of RNA pockets. Together, these advances aim to bridge a critical gap in RNA-targeted drug discovery by identifying and characterizing RNA structures capable of engaging therapeutics with drug-like properties.

In chapter 2, I discuss how I optimized fpocketR to detect pockets in RNA structures using a curated library of RNA-ligand complexes with drug-like ligands. fpocketR is a software package for detecting, characterizing, and visualizing RNA pockets.

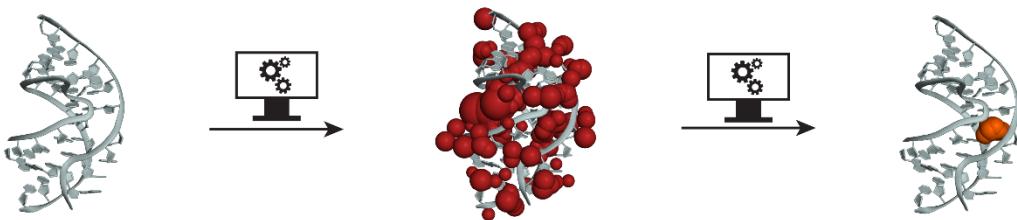
In chapter 3, I outline how I experimentally and computationally validated fpocketR as a discovery tool for detecting pockets targetable with drug-like ligands. I first developed Frag-MaP to experimentally determine ligandable sites in RNA in cells. Frag-MaP uses fully functionalized fragment probes to chemically crosslink RNA-ligand interactions and MaP to identify the RNA-ligand interactions with nucleotide precision. I further computationally validated the ability of fpocketR to detect pockets for specifically drug-like ligands, and in apo structures. Lastly, I analyzed which RNA secondary structural motifs are most likely to form ligand-binding pockets.

In chapter 4, I explain the workflow of fpocketR and apply the software to challenging applications in RNA drug discovery. I analyze ligand binding sites in large out-of-class RNAs, computationally modeled RNA, and dynamic RNA ensembles. Lastly, I analyze the shape space of drug-like RNA ligands.

Perspective

Manipulation of RNA biology using small-molecule ligands is a promising, yet currently elusive, challenge. Progress in targeting RNAs would be accelerated by a deeper understanding of the types and prevalence of RNA structures that selectively bind drug-like ligands. Here we devise independent computational (fpocketR) and experimental (in-cell Frag-MaP) strategies to detect ligand-binding sites for drug-like small-molecules in RNA tertiary structures (**Figure 1.5**). We use these tools to define RNA structure classes best targeted by small molecules, demonstrate the ligandability of computationally detected pockets, and explore the viability of using computationally modeled RNA structures for drug discovery. Going forward, fpocketR and Frag-MaP create a powerful framework for understanding the potential of ligands to bind and manipulate cellular transcriptomes.

fppocket-R: Computationally identify RNA binding pockets



Frag-MaP: Experimentally determine ligand binding sites

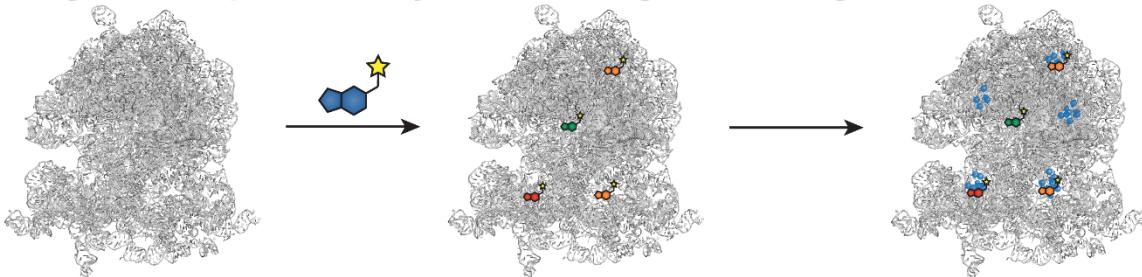


Figure 1.5. New methods for identifying RNA-ligand interactions. fppocketR is a Python package optimized to detect pockets for drug-like ligands in RNA structures. Frag-MaP is a chemical-probing technique that uses MaP to identify the binding site of photocrosslinkable functionalized fragment probes with nucleotide precision.

CHAPTER 2: DEVELOPMENT OF FPOCKETR: IDENTIFYING LIGAND BINDING POCKETS IN RNA²

Introduction

A subset of RNAs fold into complex secondary and tertiary structures capable of forming ligand-binding pockets that can engage small molecules with high affinity and specificity. RNA molecules lie upstream of most biological functions, both encoding proteins and broadly regulating gene expression. Small molecules that bind to and modulate the function of RNA thus have the potential to regulate diverse processes, including the levels of otherwise “undruggable” protein targets lacking well-defined ligand binding pockets (9, 22, 52). Further, non-coding RNAs present a vast scope of unexploited targets to manipulate biological processes (53). However, the promise of selectively targeting RNA with small molecules remains incompletely realized. Our current molecular understanding of RNA-ligand interactions is limited to a few classes of highly structured RNAs that bind to ligands, including riboswitches (54), ribozymes (55), and ribosomal RNAs (rRNAs) (18). The lack of an understanding of the types of RNA structures that form high-affinity and selective interactions with small-molecule ligands remains a major challenge for ligand discovery efforts targeting RNA.

Pockets in RNA have previously been computationally detected using approaches originally developed for finding pockets in proteins, including V3 (26, 56) and mkgrid (57, 58), which are rolling probe-based approaches, and PocketFinder (25, 59), which is energy-based. Additionally, the ROBIN database uses machine learning to characterize RNA-binding ligands (44) and SHAMAN uses molecular dynamics and fragment docking to identify small-molecule binding sites (45). These studies suggest that ligand-binding pockets in RNA and protein molecules have similar bulk properties of volume, buriedness, and solvent accessibility (25, 26, 44, 58). In contrast, specific properties that differentiate RNA and protein pockets include RNA pockets tend to be less hydrophobic and bind more rod-like ligands (25, 26, 58, 60), features which are generally viewed unfavorably for medicinal chemistry (61). Identification of these

² This chapter previously appeared as an article in *Proceedings of the National Academy of Sciences of the United States of America*. The original citation is as follows: S.D. Veenbaas, J.T. Koehn, P.S. Irving, N.N. Lama, & K.M. Weeks, Ligand-binding pockets in RNA and where to find them, Proc. Natl. Acad. Sci. U.S.A. 122, e2422346122 (2025).

unfavorable features may reflect that prior work tended to include (nearly) all ligand-bound RNA structures – containing a preponderance of simple organic molecules and less-complex stem-loop RNA structures – in contrast to focusing on RNA structures that bind to ligands with favorable (often called drug-like) physicochemical properties. Prior studies also used algorithms and scoring functions designed and validated for proteins, with the notable exception of SHAMAN which was innovatively built with the purpose of addressing RNA dynamics.

We posited, first, that differences in physicochemical properties between proteins and RNAs require that pocket-finding algorithms be optimized specifically for RNA structures. Second, pockets that bind ligands with favorable physicochemical properties (and are plausibly drug-like) should be prioritized to best understand RNA-ligand interactions in the context of structure-based ligand discovery (9, 62).

Here we introduce fpocketR, a framework for identifying ligandable pockets in RNA. fpocketR is a software package optimized to identify, visualize, and characterize pockets in RNA, built around the open-source pocket detection software fpocket (63). We optimized fpocketR, focusing on the limited examples of RNAs with known complex structures that bind small-molecule ligands with favorable physicochemical properties. This work represents the first attempt in the field to optimize pocket finding for the unique properties and characteristics of RNA structures and produced fpocketR as a powerful framework for discovering and characterizing pockets in RNA that are likely to bind drug-like small molecule ligands.

Results

Geometry-based analysis reveals differences in protein versus RNA pockets

We used the open-source, geometry-based pocket finding framework, fpocket, to detect ligandable pockets in RNA. fpocket places alpha spheres throughout the structure of a biomolecule, where every alpha sphere is in contact with the center point of exactly four atoms in the biomolecule (**Figure 2.1A**). The radius of each alpha sphere reflects the local curvature of the biomolecule (63), allowing solvent accessible cavities to be identified as groups of appropriately sized alpha spheres. Ligand-binding pockets are detected by clustering groups of alpha spheres that are close in three-dimensional space. Finally, pockets are characterized and scored based on their physical and electrostatic properties. The default fpocket algorithm is widely used and successful for detecting ligandable pockets in proteins (63, 64).

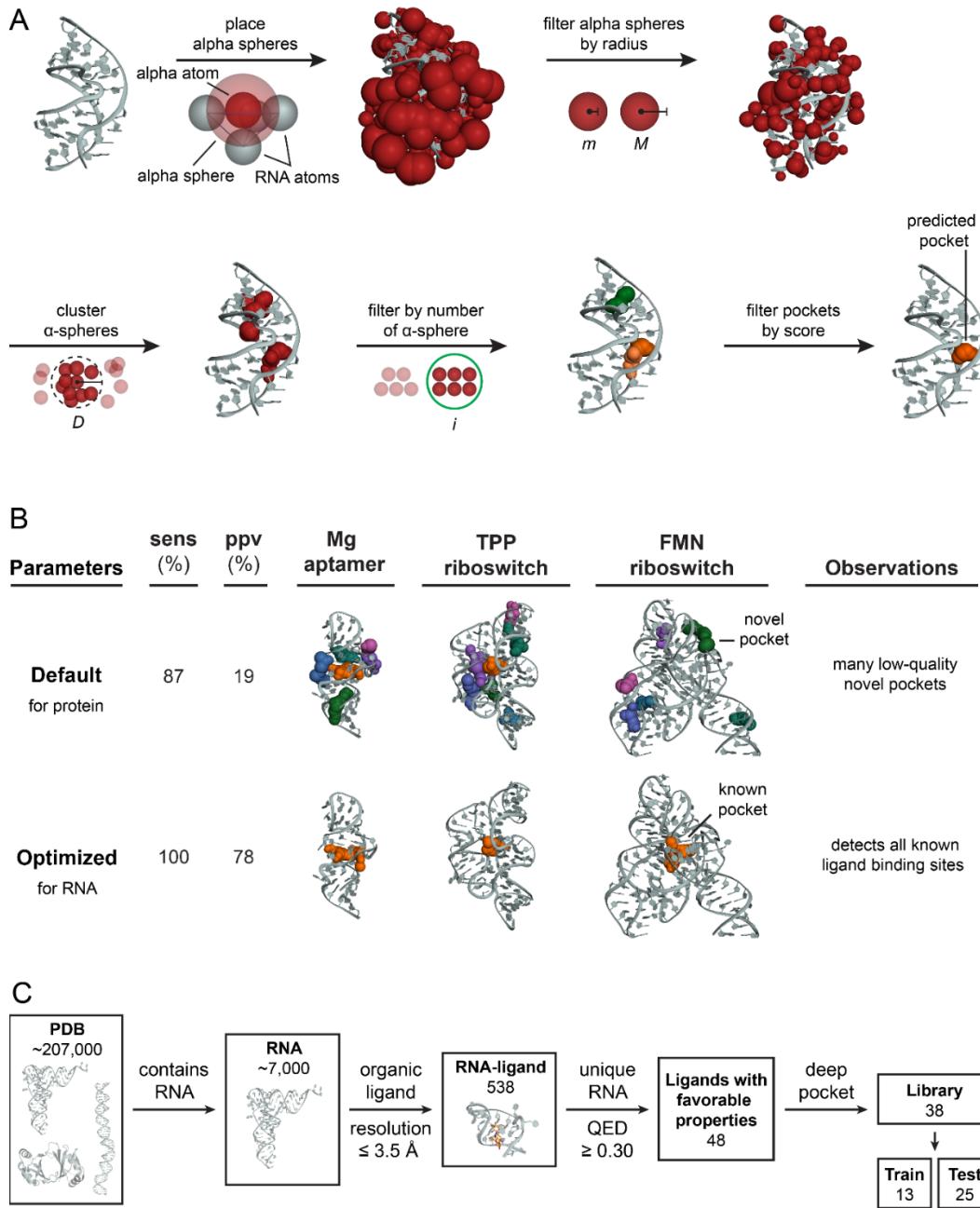


Figure 2.1. Geometry-based pocket finding approach and optimization for detecting sites in RNAs that bind ligands with favorable physicochemical properties. (A) Illustration of the fpocket algorithm and use of alpha spheres to detect pockets (PDB: 1F27) (65). (B) fpocket performance before and after optimization using a curated library of RNAs in complex with high QED ligands. Distinct pockets are indicated by differently colored alpha spheres. (Examples shown, PDB: 1Q8N, 2GDI, 5KX9) (66–68). (C) Curation and abundances of drug-like RNA-ligand complexes in the Protein Data Bank. RNA-ligand complexes were filtered from the HARIBOSS database (58). QED quantifies eight fundamental properties of a small molecule ligand into a single metric of drug-likeness (62).

The fpocket algorithm was optimized to detect pockets in proteins and tuned for the properties of amino acids (**Figure 2.1A**). When fpocket is applied to protein structures, the highest scoring pocket corresponds to the known ligand-binding site 83% of the time (63). We applied fpocket to a curated library of RNA-ligand complexes, selected for inclusion of drug-like ligands. Using the default, protein-optimized parameters, fpocket detects many (incorrect) pockets for each RNA structure, and the highest scoring pocket overlaps with the known ligand-binding site only 63% of the time. More worryingly, many pockets detected using the default parameters are located on the RNA surface or in non-selective and solvent-exposed grooves, making these "pockets" poor targets for selective, small-molecule ligands (**Figure 2.1B**). The discrepancy in fpocket performance between protein and RNA structures, first, reveals that a refinement of the algorithm is required to reliably detect RNA pockets capable of binding ligands with favorable physicochemical properties and, second, emphasizes there are fundamental differences between the ligand-facing surfaces of RNA and protein.

A curated structure database of drug-like ligands bound to RNA

To optimize fpocket for RNA, we curated a library of short (≤ 200 nt), high-resolution (≤ 3.5 Å) RNA-ligand complexes from the 538 structures available in the non-redundant version of the HARIBOSS database (**Figure 2.1C**) (58). We selected non-redundant RNAs containing ligands with favorable physicochemical properties, defined as a quantitative estimate of drug-likeness (QED) score (62) of ≥ 0.3 . The QED score is an evolution of Lipinski's rules that ranks small molecules based on eight fundamental physicochemical properties, including molecular mass, hydrophobicity, number of rotatable bonds, and number of hydrogen bond donors and acceptors. With the QED metric, overall favorable molecular properties can compensate for one or more less desirable features to provide a holistic view of ligand quality in a single value that ranges from 0 (worse) to 1 (best). Requiring a QED score of ≥ 0.3 naturally excludes ligands that bind to or intercalate with RNA non- or semi-specifically.

In the following text, for simplicity, we refer to ligands with QED (drug-likeness) scores ≥ 0.3 as drug-like RNA-binding ligands. However, we note that few of the small molecules discussed here, or known to bind RNA generally, are true drugs. By these definitions, we identified 48 high-resolution structures of RNA-ligand interactions in which the small molecule ligand is plausibly drug-like (**Figure 2.1C**). The few RNAs in this list emphasize that we have much to learn about the ligandability of RNA.

fsocketR: fsocket optimized for RNA

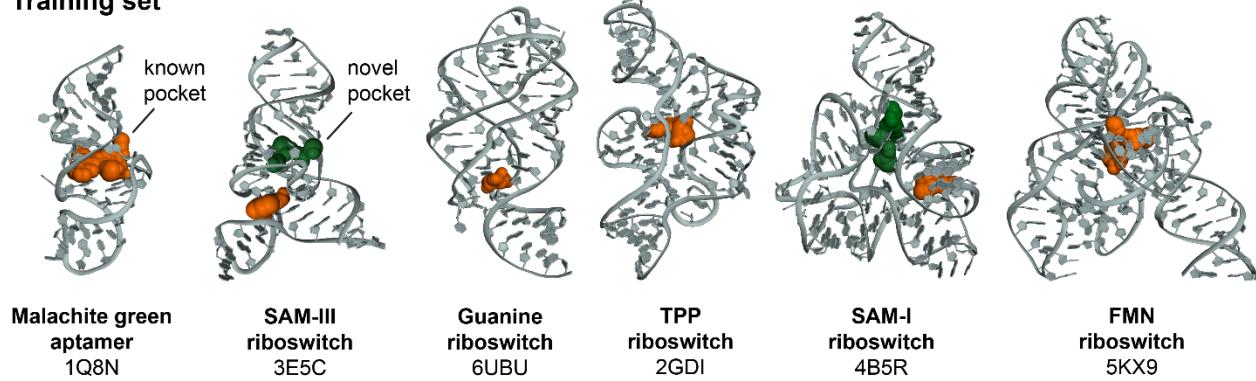
Of the 48 RNA-ligand complexes, 38 had deep, well-defined pockets. We divided these 38 complexes into a training set ($n = 13$), used for optimization of fsocketR (**Table A.1**), and a test set ($n = 25$), for validation (**Figure 2.1C, Table A.2**). Four core parameters (m, M, D, i) in the fsocket algorithm strongly affect RNA pocket detection by controlling alpha sphere size and clustering (**Figure 2.1A**). We systematically evaluated 1820 combinations of these four parameter values in a three-round multi-objective optimization to maximize sensitivity (sens) and positive predictive value (ppv) for detecting the known ligand-binding sites in our RNA-ligand training set (**Figure A.1, Table A.3**).

Optimization of fsocket yielded parameters that accurately detect binding sites for drug-like ligands in RNA (**Figure 2.1B**). We incorporated these optimized parameters into fsocketR, a custom python package that detects, characterizes, and visualizes pockets in RNA. The optimized parameters ($m = 3.0, M = 5.7, D = 1.65$, and $i = 42$) used in fsocketR differ markedly from the default fsocket parameters ($m = 3.4, M = 6.2, D = 2.4$, and $i = 15$). Critical changes were to require both smaller alpha spheres and larger numbers of alpha spheres to comprise a pocket. Using these changes, the probability that the top-scoring pocket overlapped with a known ligand binding site increased from 63% (fsocket) to 92% (fsocketR) representing a nearly five-fold reduction in false positives (**Table A.4**). Furthermore, fsocketR identified 100% of known ligand-binding sites in the training and test sets with a minimum positive predictive value of 78% (**Figure 2.1B** and **Figure 2.2**). fsocketR also detected eight novel pockets that do not overlap known ligand-binding sites. These novel pockets appear to be high quality (buried, sufficient volume, supporting π - π stacking) (**Figure 2.2**) and may bind cryptic ligands. For example, we detected a novel pocket in the SAM-III riboswitch, buried in the RNA structure between a multi-helix junction and a bulge, that is less solvent-exposed than the known ligand-binding site (**Figure A.2**). fsocketR significantly outperformed fsocket in detecting binding sites for drug-like ligands in RNA structures (**Figure 2.2B**; **Table A.4**).

Discussion

A very small number of compounds, with plausibly drug-like physicochemical properties, have ever been visualized bound to compact RNA motifs at high resolution. A few additional drug-like (and actual drug) compounds have been visualized bound to the ribosome. This study substantially expands

Training set



Test set

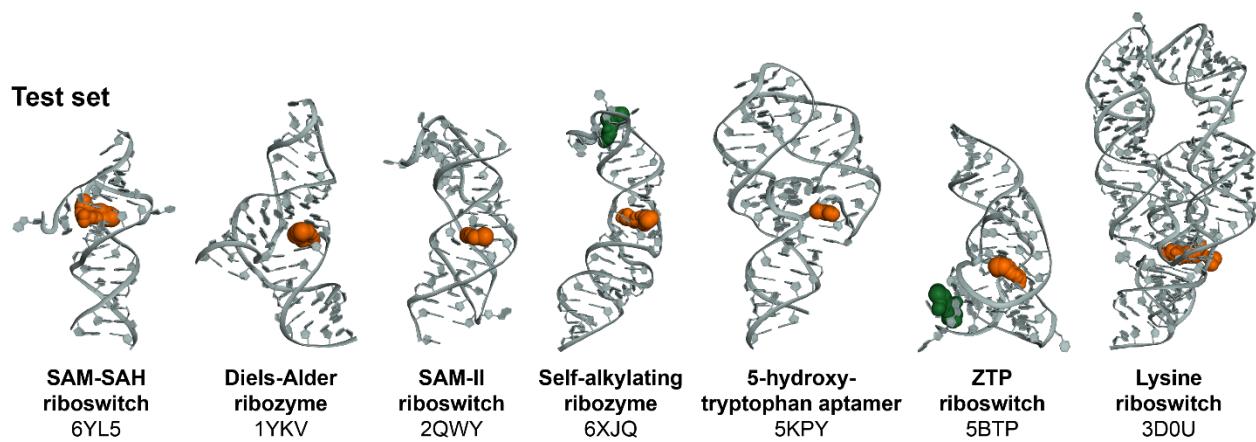


Figure 2.2. Pockets detected in RNA structures with fpocketR. Selected examples from the training and test sets are shown. Known and novel pockets are indicated in orange and green, respectively.

the number of identified pockets able to bind drug-like ligands. In principle, RNA-targeted ligand discovery could open broad opportunities for manipulating gene expression, especially for targeting difficult-to-ligand proteins and non-coding RNAs. A critical starting point is to understand the rules that underlie RNA targets containing pockets able to bind small-molecule ligands with good physicochemical features.

The approach taken here is unique and distinctly informative because fpocketR was optimized stringently using RNAs in complex with drug-like ligands (**Figure 2.1**). We used the QED framework (62), which integrates eight physical features strongly associated with druglikeness into a single metric. We required ligands to have QED values greater than 0.3, which tends to eliminate non-selective RNA binders. Our analysis of antibiotic binding pockets in the ribosome found that fpocketR selectively detected pockets that bind drug-like ligands and, critically, distinguishes between pockets that bind drug-like versus nondrug-like ligands.

When applied to RNA, fpocket (mis)identified numerous near- or non-pockets and did not accurately rank accepted pockets. Ultimately, fpocket, with default parameters, did not perform well with RNA, emphasizing fundamental differences between the ligand-facing surfaces of proteins and RNA. We identify three broad differences between pockets in RNA and protein. First, pockets detected in RNA (using fpocketR) and protein (using either algorithm) have broadly similar physical properties with the critical difference that pockets in RNA are more polar and less hydrophobic (**Table A.5**). Second, electrostatic properties significantly influence the pocket scoring functions used to rank pockets by their quality, and the unique electrostatics of RNA cause true pockets to be mis-ranked using current algorithms. Third, the unique and ubiquitous presence of solvent-exposed major and minor grooves on the surface of RNA create pseudo-pocket-like channels. As a result, false-positive pockets must be excluded more rigorously by imposing stricter geometric requirements that limit the solvent-exposure of pockets and reduce the reliance on pocket scoring functions to rank pockets.

Materials and methods

Curation of RNA-ligand complexes containing ligands with favorable (drug-like) physicochemical properties

RNA structures used to optimize fpocketR were obtained from the HARIBOSS non-redundant RNA-ligand database (58). RNA-ligand complex structures were required to (*i*) bind to a drug-like small

molecule ligand (QED score ≥ 0.3), (ii) be high-resolution ($\leq 3.5 \text{ \AA}$), (iii) be a unique RNA in our data set, and (iv) ≤ 200 nucleotides (this requirement removes ligands that bind the ribosome). The quantitative estimate of drug-likeness (QED) score (62) was calculated using the QED module in the rdkit library (69). We emphasize that the QED metric is an open source and highly useful measure of fundamental physicochemical properties: molecular mass, hydrophobicity, hydrogen bond donors and acceptors, polarizable surface area, number of rotatable bonds, number of aromatic rings, and structural alerts (62). As defined by these criteria, there were 48 known RNA-ligand complexes with plausible drug-like ligands, that bind compact RNA sites. These RNA-ligand complexes were manually evaluated in PYMOL (70) to assess the solvent accessibility and cavity depth of the ligand binding pocket. Ultimately, 38 unique RNA structures were chosen and randomly split into a training set ($n=13$) and test set ($n=25$).

Pocket evaluation metrics

Only pockets with an fpocket score ≥ 0 were analyzed. Pockets identified by fpocket were required to meet three criteria to be classified as a known pocket that binds a small-molecule ligand. The geometric center criterion is met if the geometric center of a pocket is within 4.5 \AA of any atom of the ligand. The ligand overlap criterion is met if at least 25% of the alpha spheres in a pocket overlap the ligand (within 3 \AA). The pocket overlap criterion is met if at least 25% of the atoms in the ligand overlap with a pocket (within 3 \AA). Pockets that do not meet all three criteria are classified as novel pockets. However, we note that several novel pockets are occupied by biomolecules and ions (**Table B.2 - Table B.5**). Software used to evaluate these criteria are provided (see data and software availability sections).

Multivariate optimization of fpocket

We optimized fpocket 4.0.3 (63) for detecting pockets in RNA by systematically adjusting values of six core parameters (M, m, D, i, A, p). Parameters A and p filter pockets based on polarity and, in preliminary analyses, were not effective at improving pocket detection in highly polar RNA molecules. The remaining four parameters (m, M, D, i) control the minimum (m) and maximum (M) alpha sphere radii, clustering distance (D), and minimum cluster size (i) (**Figure 2.1A**) and were systematically optimized in 3 rounds of multivariate optimization using our training set of RNA-ligand complexes ($n = 13$). We evaluated fpocket performance to detect ligand binding sites in a large parameter space (1820 parameter

combinations) by performing a coarse, medium, and fine round of optimization (**Table A.3**). In each round, we evaluated 5 values for each parameter with progressively smaller increments between the values. We started with the default fpocket parameters and their flanking values and used the best performing parameters from each round of optimization to generate values for the subsequent round. The best performing parameters were selected by calculating a pareto set that maximized sensitivity and positive predictive value (ppv) and by applying some intuition intended to avoid overtraining (71). We selected the parameters --m 3.0, --M 5.7, --D 1.65, and --i 42 to be the default parameters for fpocketR but high-quality results can also be achieved with several alternative parameters within the following ranges: --m 2.9-3.1, --M 5.5-5.8, --i 36-44, --D 1.5-1.7 (**Figure A.1**).

fpocketR usage

RNA pockets were detected using fpocketR. fpocketR is a wrapper for fpocket 4.0.3 (63) that provides options for advanced analysis and visualization of pockets detected in RNA structures. RNA pockets were detected using the input --pdb option for a .pdb or .cif file, the --chain option (maximum of 2 RNA chains), and the newly defined fpocketR parameters (--m 3.0, --M 5.7, --D 1.65, --i 42). Multistate NMR ensembles were analyzed using the --state option. Known binding ligands were specified using the --ligand and --ligandchain options, allowing fpocketR to differentiate known pockets (visualized in output files as orange to yellow) from novel pockets (green to pink). Visualizations for pockets in secondary structure space were generated by inputting a secondary structure drawing using the --nsd and --connectpocket options.

Pocket visualization

Figures visualizing predicted pockets were generated using the fpocketR software package. fpocketR generates tertiary structure figures using the PyMOL API (70) and secondary structure figures using RNavigate (72). fpocketR visualizes pockets as a group of alpha-atoms (computed by subtracting 1.65 Å from the fpocketR-generated alpha sphere radius) (**Figure 2.1A**). Pockets are visualized in secondary structures by coloring the nucleotides that contact alpha spheres from a pocket. RNA secondary structure drawing templates (.nsd) were generated from PDB structures using RNApdbee (73) manually edited in StructureEditor (74), and input into fpocketR using the --nsd option.

Data availability

All data and software generated or used in this work are freely available. The fpocketR datasets reported in this study and fpocketR software suite and user manual are available at https://github.com/Weeks-UNC/2025_Veenbaas and <https://github.com/Weeks-UNC/fpocketR>, respectively.

CHAPTER 3: DEVELOPMENT OF FRAG-MAP AND VALIDATION OF FPOCKETR³

Introduction

Only a few dozen RNA-ligand complexes, involving small molecules with favorable (drug-like) physicochemical properties, have been visualized at high resolution. Manipulation of RNA biology using small molecule ligands is thus a promising, yet currently elusive, challenge. Progress in targeting RNAs would be accelerated by a deeper understanding of the types and prevalence of RNA structures that selectively bind drug-like ligands. A set of computational and experimental RNA-specific tools are needed to identify and characterize these potential sites of RNA-ligand interactions.

There is intense debate in the RNA-ligand field regarding what kinds of specific RNA structures are capable of harboring pockets able to bind ligands with favorable physicochemical properties. The current dialogue centers on the relative importance of targeting simple step-loop or bulge-containing motifs, which may have stronger current biological validation, versus targeting complex RNA motifs, which can potentially form more selective interactions with small molecules (9, 31, 75, 76). Most recent work has focused on the former class of simple RNA targets.

Here we introduce Frag-MaP, a robust experimental tool for identifying ligandable pockets in RNA. Frag-MaP is a reliable tool that uses mutational profiling (MaP) technology (50, 51) to find crosslinking sites between RNA binding pockets and photo-crosslinking ligands. Frag-MaP is a compliment to our predictive software fpocketR. We used Frag-MaP to experimentally validate the ligandability of pockets detected by fpocketR. We computationally validated that fpocketR can be used as a discovery tool to identify novel pockets that specifically bind drug-like ligands in both apo (ligand-free) and (low resolution) dynamic RNA tertiary structures. Finally, we analyzed pocket-forming RNA secondary structures identified using fpocketR and propose a broad model for understanding RNA secondary structures that specifically interact with favorable (drug-like) small-molecule ligands. Going forward,

³ This chapter previously appeared as an article in *Proceedings of the National Academy of Sciences of the United States of America*. The original citation is as follows: S.D. Veenbaas, J.T. Koehn, P.S. Irving, N.N. Lama, & K.M. Weeks, Ligand-binding pockets in RNA and where to find them, Proc. Natl. Acad. Sci. U.S.A. 122, e2422346122 (2025).

fpocketR and Frag-MaP create a powerful framework for understanding the potential of ligands to bind and manipulate cellular transcriptomes.

Results

Frag-MaP: Experimental validation of fpocketR predictions

We applied fpocketR to the *E. coli* (PDB: 7K00) and *B. subtilis* (PDB: 7AS8) ribosomes (77, 78) and identified 46 and 52 pockets in the 23S rRNAs, respectively (**Figure B.1**). The ribosomal pocket-ome identified in this analysis includes most known ligand-binding sites for drug-like ligands, including binding sites for the antibiotics linezolid, sparsomycin, and spectinomycin. However, most predicted pockets are novel, suggesting the universe of RNA pockets capable of binding drug-like ligands might be quite large.

We therefore experimentally evaluated if the novel pockets identified by fpocketR bind drug-like small molecules in cells. We developed Frag-MaP, which leverages tri-functionalized probes containing a small-molecule fragment (MW ≤ 300) linked to a photocrosslinkable diazirine and “clickable” alkyne handle (**Figure 3.1A**) (79, 80). Probes penetrate cells and crosslink to nucleotides proximal to their RNA-ligand binding sites, yielding RNA-ligand adducts. Unique to our approach, Frag-MaP exploits mutational profiling (50, 51) to identify ligand-binding sites with low sequencing and library-preparation biases, high sensitivity, and single-nucleotide resolution (see methods for discussion). RNA-ligand adducts induce mutations in the cDNA during reverse transcription, which are quantified through massively parallel sequencing. Frag-MaP sites are identified as nucleotides that interact with a fragment probe with significantly higher mutation rates compared to a control probe bearing a simple methyl group (**Figure 3.1B**).

We first validated the Frag-MaP strategy using a known RNA-ligand interaction based on a functionalized linezolid probe (**Figure 3.1B**). Linezolid is a fragment- and drug-like (QED = 0.79) antibiotic with activity against Gram-positive bacteria and binds with low micromolar affinity to a pocket in the peptidyl transferase center (PTC) in the 23S rRNA (81, 82). The pocket in which linezolid binds is robustly predicted by fpocketR (**Figure 3.1F**). The acetamidomethyl group is oriented toward the exit tunnel and provides a vector to functionalize linezolid without interfering with binding. We treated *B. subtilis* cells with the linezolid probe and detected four Frag-MaP sites as positions highly reactive toward the linezolid probe (mutation rates: 3-18%) relative to the fragment-less control probe (median mutation rate: 0.5%)

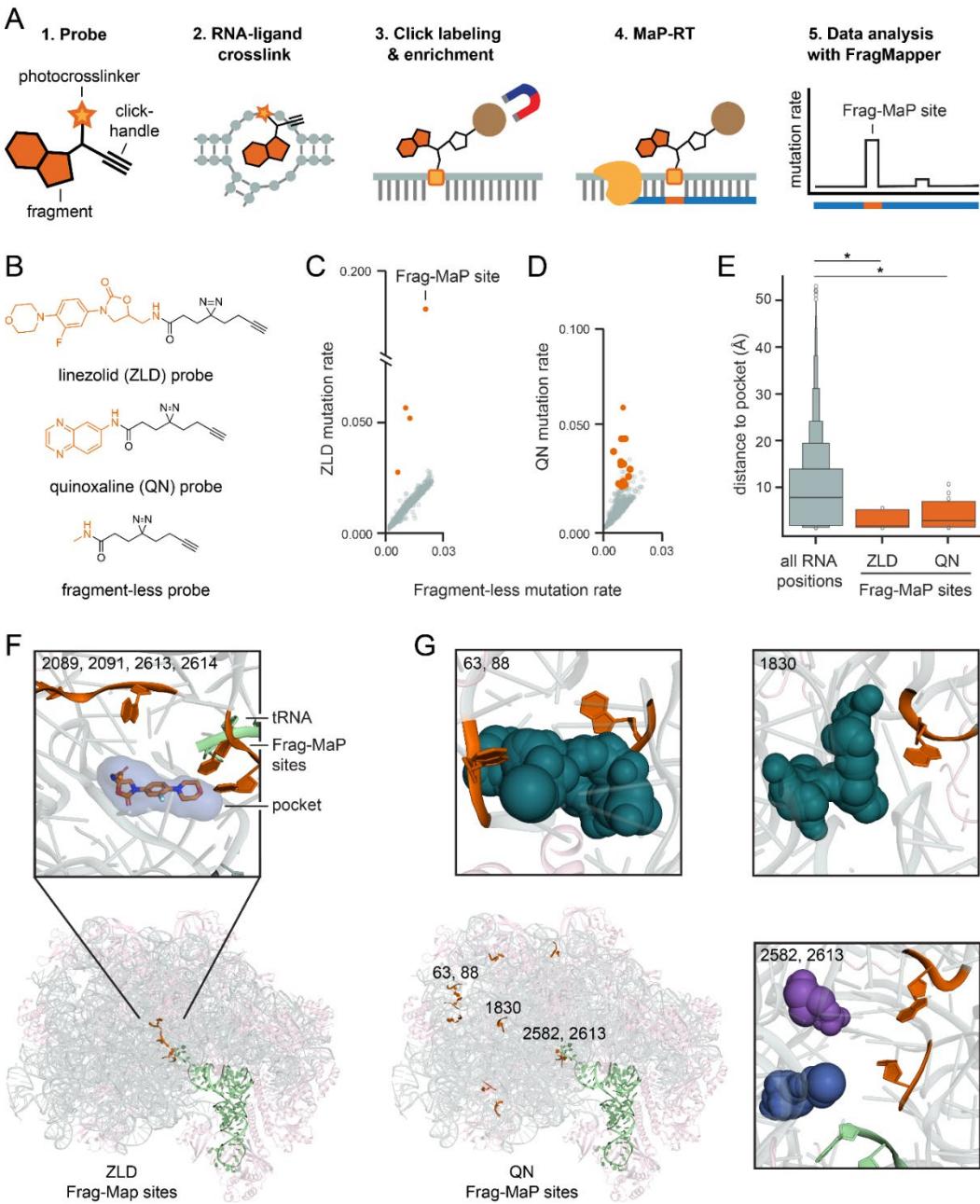


Figure 3.1. Experimental detection of RNA-ligand binding sites by Frag-MaP. (A) Frag-MaP experimental scheme. (B) Structure of High-QED probes used in Frag-MaP experiments. Mutation rate of Frag-MaP sites (orange, labeled) detected in the 23S rRNA for (C) linezolid and (D) quinoxaline probes. (E) Distance to a pocket predicted by fpocketR from all nucleotides in the 23S rRNA compared to Frag-MaP sites for linezolid and quinoxaline. *, p < 0.03; Mann Whitney test. (F) Visualization of all Frag-MaP sites detected using the linezolid-based probe. Frag-MaP sites are orange; linezolid is shown bound in the peptidyl transferase center (PDB: 7AS8) (78). (G) Most reactive Frag-MaP sites detected in the 23S rRNA using the quinoxaline-based probe. Pockets detected using fpocketR are shown as spheres.

(**Figure 3.1C**, **Figure B.2A**). The four Frag-MaP sites detected for the linezolid probe (nucleotides 2089, 2091, 2613, 2614) are hundreds of nucleotides apart in primary sequence but lie within 7 Å of the linezolid binding site in three-dimensional space (**Figure 3.1F**, **Figure B.2A**). Frag-MaP thus detects RNA-ligand binding sites with nucleotide resolution in cells, with high sensitivity and specificity.

We next screened for new ligand-binding sites in the *B. subtilis* 23S rRNA using 6-aminoquinoxaline (QED = 0.57) (**Figure 3.1B**) and fragment 206 (QED = 0.81) probes (**Figure 3.1C**). The quinoxaline probe yielded ten Frag-MaP sites with mutation rates up to 6% (**Figure 3.1D**). Some of these sites were tens of nucleotides apart from each other in sequence, but clustered at compact locations in the 23S rRNA (**Figure 3.1G**). Strikingly, 9 out of 10 Frag-MaP sites occurred at pockets predicted by fpocketR with a median distance of 4.5 Å (**Figure 3.1E**, **Figure B.2B**). Probe 206 yielded an additional 4 Frag-MaP sites, 3 of which are near pockets (median distance of 5 Å) (**Figure B.2C**). Both probes bound to pockets in functionally interesting areas within the large ribosomal subunit. The quinoxaline probe bound to a pocket formed by a pseudoknot within the exit tunnel (nucleotides: 63 and 88) and to three nucleotides in the L1 stalk (nucleotides: 2145, 2154, 2198). Both the quinoxaline (nucleotides: 2582, 2613) and 206 (nucleotides: 2613, 2614) probes also bound at the P-site within the PTC, adjacent to the linezolid binding site. These results demonstrate the ligandability and relevance of pockets detected by fpocketR.

Almost all (89%) ligand binding sites detected by Frag-MaP are close in three-dimensional space (median distance 4.8 Å) to pockets detected by fpocketR. Frag-MaP sites are significantly closer to pockets identified by fpocketR than expected by chance and occur at residues with both high and low solvent accessibility (**Figure 3.1E**). The juxtaposition of pockets – predicted by fpocketR and validated experimentally by Frag-MaP – emphasizes that pockets detected by fpocketR are reflective of authentic, in-cell ligand-binding sites for drug-like small-molecule ligands.

Pocket ligandability: validation of sites for drug-like small molecules

In a second validation strategy, we evaluated 17 RNA-ligand complexes comprised of antibiotics bound to bacterial rRNAs to assess the capacity of pockets detected by fpocketR to bind drug-like ligands. The 17 small-molecule ligands bind the 30S and 50S rRNAs at ten distinct binding sites, span multiple chemical classes, and have QED scores ranging from 0.01 (not drug-like) to 0.89 (highly drug-

like) (**Figure 3.2A**, **Figure B.3**). This is a rigorous test of fpocketR as rRNAs are significantly larger than and not evolved to bind small-molecule ligands compared to the RNA aptamers and riboswitches used for training fpocketR.

The known ligand-binding site was detected as a pocket by fpocketR for 7 of the 17 RNA-ligand complexes, across 4 distinct binding sites. fpocketR differentiated ligand-binding sites based on the drug-likeness of the bound ligands. fpocketR detected six of the seven pockets for ligands with a QED score above the training threshold of 0.3 (**Figure 3.2A**). In addition, fpocketR detected the pocket for the highly conjugated ligand thermorubin that forms extensive π - π interactions with rRNA at the junction of the 50S and 30S ribosomal subunits. This pocket appears able to bind drug-like ligands, even though the ligand in this case is not drug-like.

There are clear structural differences between ligand-binding sites that interact with druglike and non-druglike ligands (**Figure 3.2B**). Pockets that bind drug-like ligands are buried, generally encapsulate their ligands, and form both hydrogen bonding and π - π interactions. By contrast, RNA motifs that interact with less drug-like ligands, such as aminoglycosides and the extended edeine chain, are characterized by a high number of hydrogen bonding interactions and are located at solvent-exposed RNA surfaces (**Figure 3.2B**). All low-QED (and thus non-drug-like) ligands in our analysis are natural products.

Apparently, bacteria can evolve anti-ribosome antibiotics that do not follow human-compiled rules for drug-likeness and, further, can target ribosome regions that do not contain conventional pockets.

In sum, based on both direct experimental interrogation with Frag-MaP (**Figure 3.1**) and by evaluating drug-like and non-drug-like molecules that bind the ribosome (**Figure 3.2**), we validate that fpocketR selectively identifies pockets in RNA capable of binding ligands with drug-like physicochemical properties. We now use fpocketR to understand features of ligand-binding pockets in RNA.

Detection of pockets in unliganded (apo) structures

We used fpocketR to examine binding sites in both apo (ligand-free) and dynamic structures. We analyzed paired ligand-free (apo) and ligand-bound (holo) RNA structures for ten RNA-ligand complexes spanning a wide-range of QED scores (0.23 to 0.89; **Table B.1**). Most of these structures were determined by X-ray crystallography, so crystal packing partially stabilized the apo state. fpocketR detected the known ligand binding site in 9 of the 10 holo complexes, only failing to detect the ligand

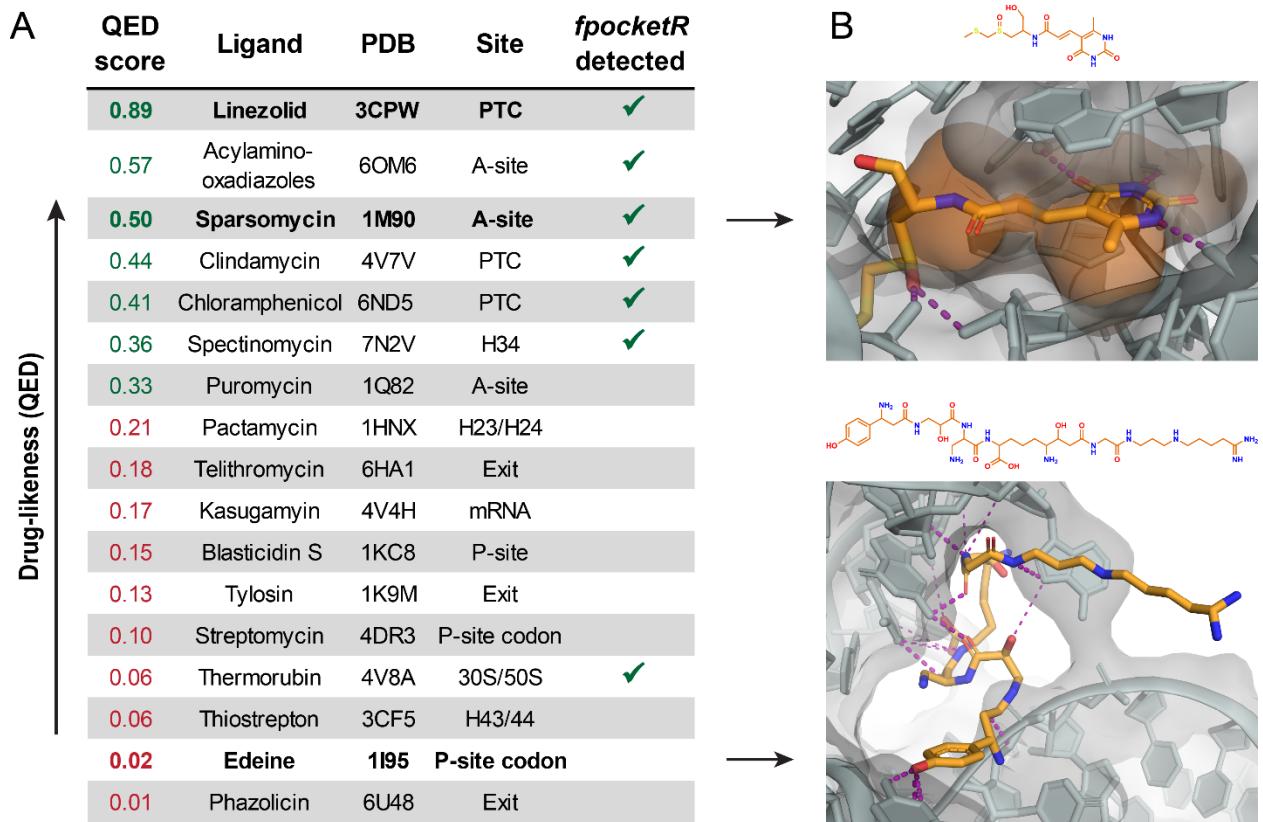


Figure 3.2. Pockets detected by *fsocketR* in the bacterial rRNA, compared to small-molecule antibiotic binding sites. (A) Ligand binding sites, listed by (QED) drug-likeness score. Sites detected with *fsocketR* indicated with check marks. (B) Contrasting examples of a drug-like ligand (sparsomycin) that binds a well-defined pocket (orange) and a non-drug-like ligand (edeine) that binds a non-pocket-like structure. An illustration of ligands bound near the PTC is provided in Fig. S5.

binding site for the least drug-like ligand, phosphoribosyl pyrophosphate (QED 0.23). fpocketR also detected the known ligand binding site in 9 of 10 apo (ligand-free) structures, including RNAs with global ligand-induced conformation changes (for example, the TPP riboswitch). The size, shape, and position of predicted pockets are often affected by ligand-induced conformational changes in the RNA structure, meaning that pockets in apo structures do not always perfectly align with known ligand binding sites (**Figure 3.3**). However, fpocketR-detected pockets in the apo structures reliably determined the local capacity of an RNA to bind a small-molecule ligand.

RNA molecules often populate structural ensembles, and the underlying structural dynamics influence biological function and response to ligand binding (83). This intrinsic dynamism in RNA structure also means that RNA structures in ensembles, especially for smaller RNAs, are visualized at modest resolution. We visualized pockets in a carefully refined NMR-based model of the HIV-1 TAR RNA hairpin (84). Of the 20 structures modeled for the TAR RNA, seven (35%) contained pockets overlapping the binding site for the known ligand (**Figure 3.3B**). Similarly, an ensemble of structures was reported for the apo state of the SAM-IV riboswitch (determined by cryo-EM; 3.7 Å resolution) (85). In this case, pockets formed in four states (20%) (**Figure 3.3B**). Novel pockets were observed for a single state for both RNAs. Thus, in these two examples of unliganded RNA structures with multiple states, pockets formed in only a minor subset of conformations but were concentrated at known ligand-binding sites. Small molecule binding then selects for the subset of structures with a ligandable pocket. Importantly, fpocketR finds ligandable sites even in ensembles based on modest-resolution structure models.

Tripling the known universe of RNA pockets able to bind drug-like ligands

Non-redundant, high-resolution structures for RNAs are limited, and are especially sparse for RNAs bound to ligands with good (drug-like) physicochemical properties. Prior to this study, roughly 50 pockets capable of binding drug-like ligands had ever been visualized; these are those found in small RNAs and in the ribosomal RNAs (**Figure 3.2**). Based on fpocketR, this work substantially expands this universe to 138 and could be made larger by extending the analysis developed here to additional RNAs.

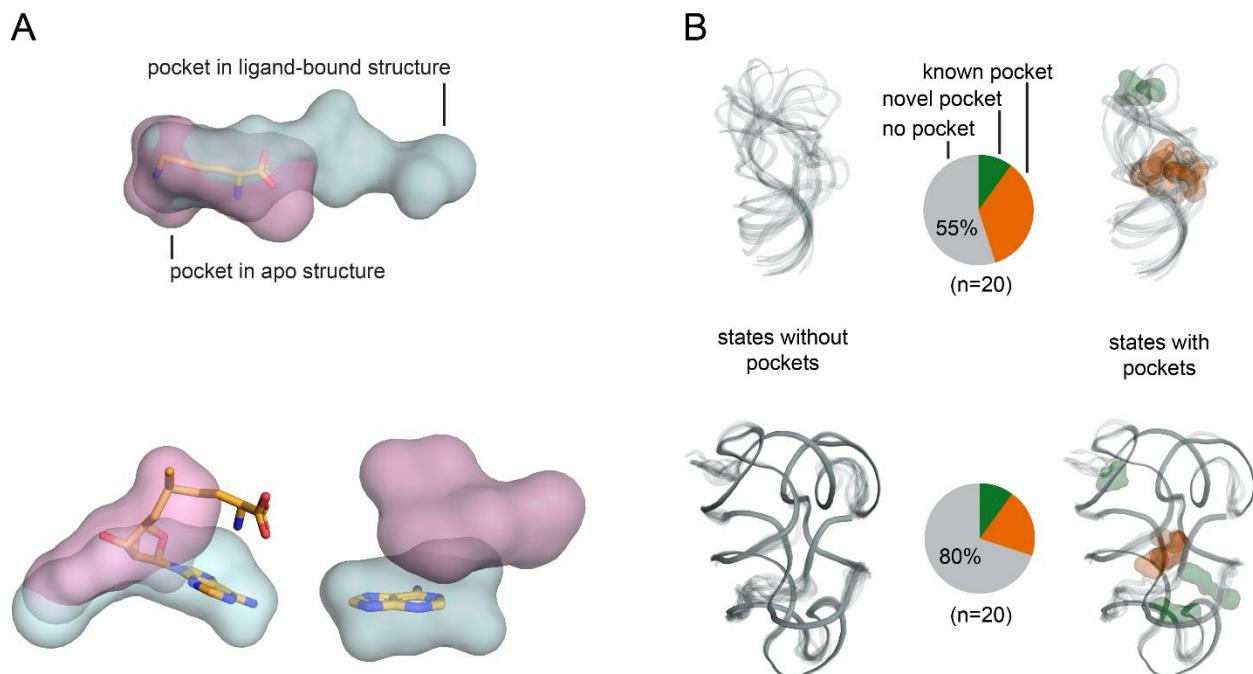


Figure 3.3. Detection of ligand-binding sites in apo (ligand-free) sites and in RNA ensembles. (A) Pockets detected by fpocketR in apo (pink) and ligand-bound (blue) RNA structures for the lysine riboswitch, SAM-I riboswitch, and adenine riboswitch (PDB: 3D0U & 3D0X, 3IQN & 3IQP, 5SWE & 5E54) (86–88). (B) Pockets detected in individual RNA states for ensembles of HIV-TAR, measured by NMR (PDB: 7JU1) (84), and the SAM-IV riboswitch, measured by cryo-EM (PDB: 6WLQ) (85). Known and novel pockets, detected by fpocketR, are orange and green, respectively. Structures are depicted as transparent backbones; pockets are shown as transparent surfaces.

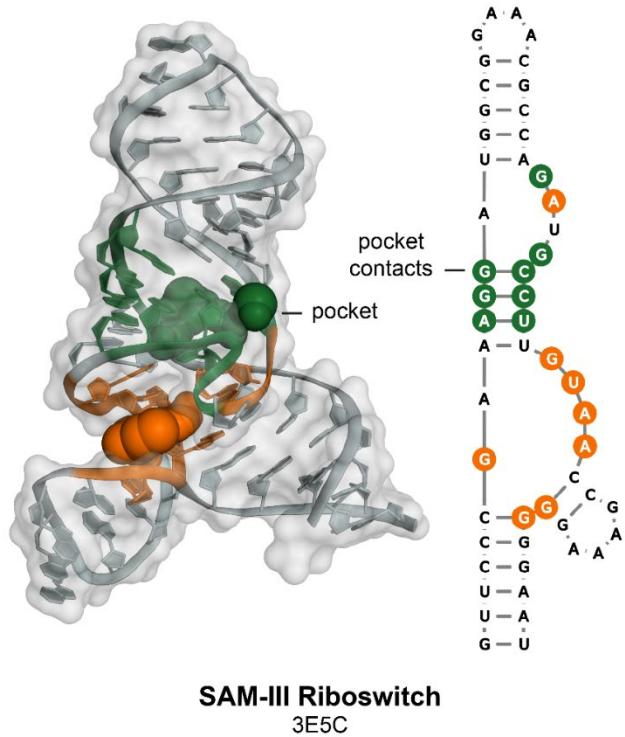


Figure 3.4. Scheme for mapping RNA pockets, visualized in three-dimensional structures, onto secondary structure diagrams. Nucleotides are colored to match the pocket they contact, in both tertiary and secondary structure images (PDB: 3E5C) (89). In this structure, both pockets are local.

RNA secondary structures that harbor pockets

We examined RNA structures from the training set, test set, bacterial rRNAs, and a group II intron and identified all nucleotides that form pockets detected by fpocketR, and thus likely to bind drug-like ligands. We then annotated the nucleotides that form pockets in secondary structure diagrams (**Figure 3.4**). In larger multi-domain RNAs, such as the ribosome or group II intron, pockets form both from nucleotides localized in secondary structure space (local interactions) and from nucleotides that become juxtaposed via through-space interactions involving multiple domains (long-range interactions) (**Figure 3.5, Figure B.4, Figure B.5, Table B.2 - Table B.5**). Pockets formed by long-range interactions in large RNA structures would be difficult or impossible to identify from a secondary structure model alone. Thus, we focused our analysis on local pockets. Small RNAs (< 200 nucleotides) form local pockets exclusively. A majority (66%) of pockets in multi-domain RNAs also occur in local secondary structures (**Figure 3.5C**), which emphasizes that readily characterized secondary structure motifs form pockets in both small and large RNAs. We assigned each pocket detected by fpocketR to a class defined by the most complex secondary structural motif that contributes to pocket formation. Secondary structure motifs that formed pockets varied in complexity and included (i) simple loops and bulges, (ii) consecutive loops or bulges separated by a contact distance of 5 nucleotides or less, (iii) multi-helix junctions, and (iv) pseudoknots.

Simple bulges and loops are, as expected, extremely common, and we observed 402 in our dataset; only 3% of these motifs formed a pocket. In addition, the observed pockets typically did not form independently, and instead tended to form in large, multi-domain RNAs, where the global architecture could stabilize a local pocket in a simple structure (**Figure 3.6A**). Pockets were slightly more common among consecutive loops (4% of which formed pockets; among 273 examples).

Pockets were much more common among complex structures, multi-helix junctions, and pseudoknots, based on our all-pockets dataset (48% and 63% of which formed pockets, respectively) (**Figure 3.6A**). Across our full dataset, the number of multi-helix junction and pseudoknot motifs are artificially increased, due to the inclusion of many riboswitch and aptamer motifs. Therefore, we additionally determined the local structure among only our novel pockets. Simple motifs and consecutive loops formed novel pockets very rarely (3% and 1%, respectively). Multi-helix junctions and pseudoknots formed novel pockets at high rates (39% and 53%, respectively) (**Figure 3.6A**), comparable to those in

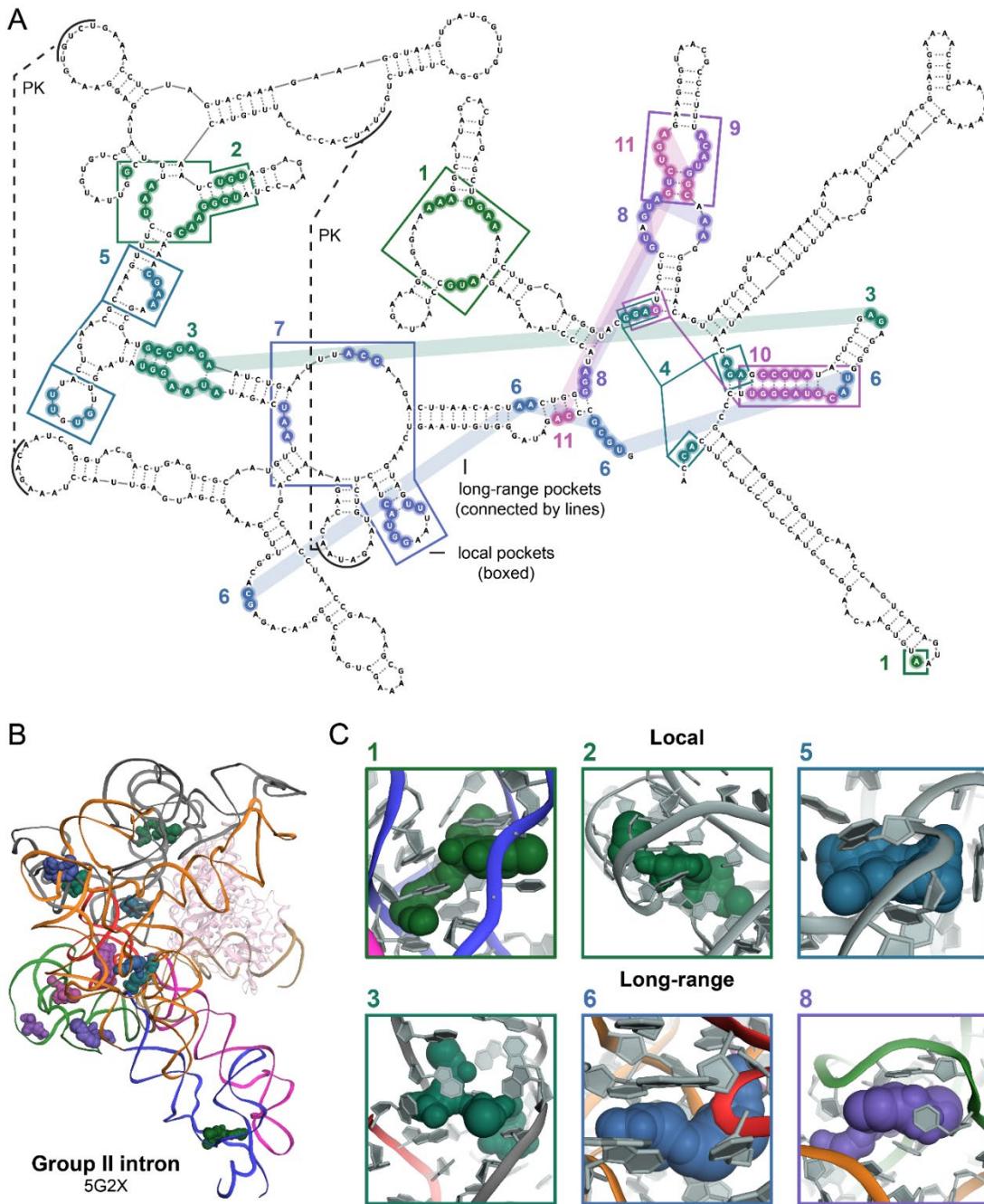


Figure 3.5. Representative illustration of pockets formed by local versus long-range interactions, detected in a group II intron RNA. Pockets formed by local domains are boxed. (A) Pockets are identified by number and color, and superimposed on the RNA secondary structure. (B) Pockets are visualized in the three-dimensional RNA structure (colored by domain). Complex contains a reverse transcriptase cofactor (pink) (PDB: 5G2X) (90). (C) Examples of pockets formed by local versus long-range RNA interactions. RNA backbone is colored by domain.

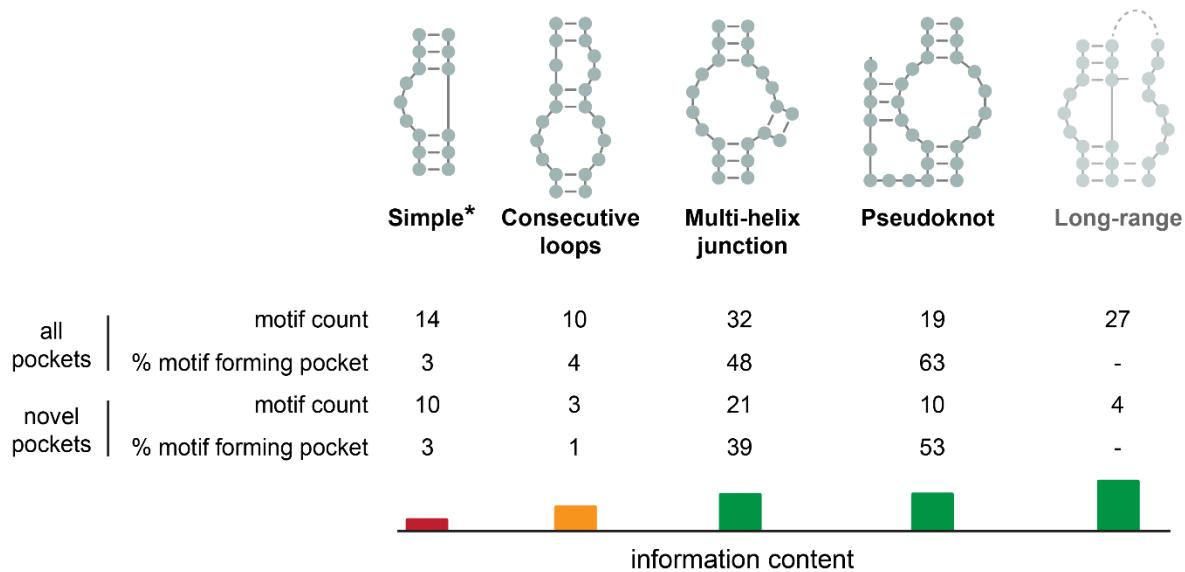
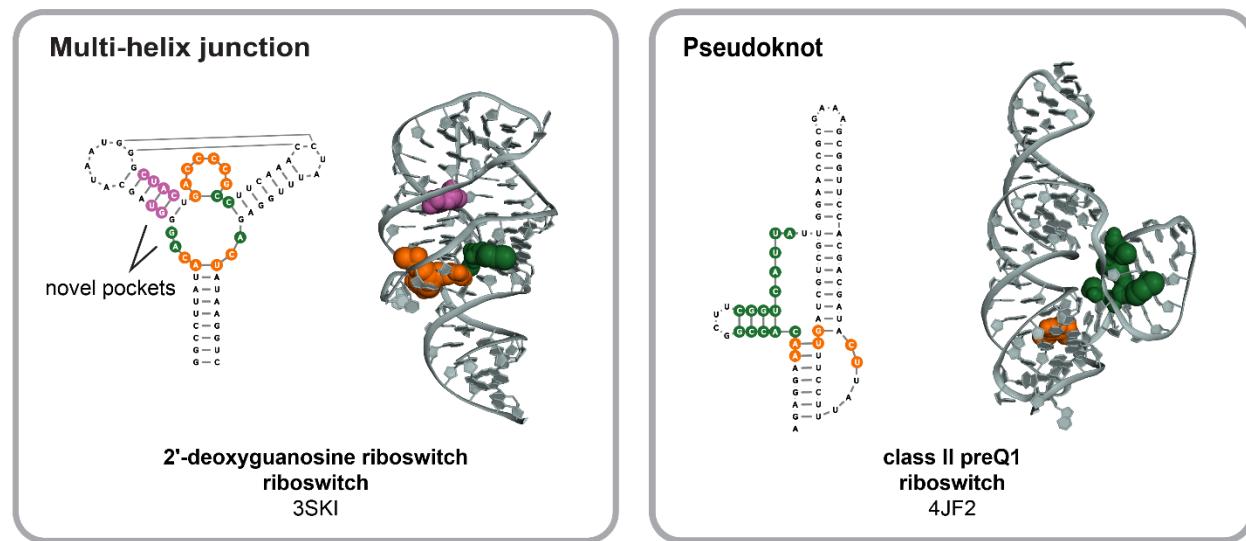
A**B**

Figure 3.6. Secondary structural motifs that form pockets able to bind ligands with (drug-like) high-quality physicochemical properties. (A) Likelihood of secondary motifs to form pockets, in ascending order of complexity and information content. *, simple pockets were observed primarily in large multi-domain RNAs. (B) Examples of novel pockets in RNA secondary structural motifs most likely to form pockets. Known pockets are orange; novel pockets are green and magenta.

the all-pockets dataset. This analysis emphasizes that any given simple motif is unlikely to contain a pocket able to bind small molecules with good physicochemical properties; but that many complex multi-helix junction and pseudoknot motifs will.

Discussion

We validated predictions made by fpocketR with Frag-MaP, which is unique relative to current approaches for detecting RNA-ligand interactions in complex mixtures (48, 80, 91, 92). Frag-MaP leverages mutational profiling to identify RNA-ligand binding sites with high sensitivity and independently from biases introduced by biotin pulldown and DNA library preparation steps (51, 93). Here, we used Frag-MaP in a simple way, to detect fragment binding sites in abundant rRNAs. Ligand-binding sites identified by Frag-MaP consistently surrounded pockets detected by fpocketR, demonstrating the general ligandability of novel pockets detected by fpocketR. A powerful future application will be to use Frag-MaP as an in-cell RNA tertiary structure discovery tool because ligand-binding pockets form preferentially and specifically in regions with complex RNA structures.

Analyses enabled by fpocketR emphasize that many pockets are created by nucleotides that are localized in RNA secondary structure. Local RNA secondary structures harbor nearly all the pockets in small RNAs (< 200 nucleotides) and form most pockets in large multi-domain RNAs. Pockets in secondary structures are primarily formed by complex motifs, including multi-helix junctions and pseudoknots, compared to simpler RNA structures such as bulges and loops. Complex structural motifs both have a much higher rate of forming drug-like pockets and also have higher information content, which facilitates selective interactions with drug-like ligands across RNA transcriptomes (9) (**Figure 3.6A**). Despite these clear advantages of targeting complex structures, a significant component of current work directed at discovery of RNA-targeting small molecules focuses on RNA motifs with simple structures (reviewed in (9, 31, 75, 76)). Our study indicates that simple bulge and consecutive loop motifs contain pockets only infrequently (**Figure 3.6A**). These observations emphasize that RNA secondary structure is a useful tool to estimate RNA targetability and that RNAs with complex structural motifs are likely the best targets for drug-like small molecules (**Figure 3.6B**).

De novo design of small-molecule ligands that target RNA and modulate function remains challenging. This work provides critical computational and experimental tools for identifying and

understanding features of RNA structures that form targetable pockets, specifically able to bind drug-like small molecule ligands with high potential for therapeutic development.

Materials and methods

Classifying pocket structural motifs.

Pocket-forming secondary structures were classified by the types of RNA structure in contact with alpha spheres. The classes identified and used in this work were, in order of structural complexity: simple structures, consecutive loops, multi-helix junctions, pseudoknots, and long-range structures. Specific structural definitions were: Simple structure, pocket only contacts nucleotides in an apical loop, internal loop, or bulge. Consecutive loops, pocket contacts a pair of loops/bulges within a contact distance of 5 nucleotides from each other (contact distance is defined as the shortest path length through the secondary structure graph between two nucleotides). Multi-helix junction, pocket contacts single-stranded nucleotides in the junction or base-paired nucleotides within 3 nucleotides of the multi-helix junction. Pseudoknot, pocket contacts base-paired nucleotides involved in a non-nested interaction, or single-stranded nucleotide located between non-nested helices. Long-range structure, pocket contacts motifs in a structure that have a contact distance of >15 nucleotides. If a structure fit the criteria for multiple motif classifications, the structure was assigned to the most complex structural motif. Structures (4 observed) containing g-quadruplexes were not analyzed.

In-cell probing for Frag-MaP.

Bacillus subtilis subsp. *subtilis* strain 168 (ATCC) was grown on LB agar plates at 30 °C. A 5 mL culture was inoculated with a single colony and grown overnight in LB media at 30 °C, 225 rpm. A 2 mL aliquot of the culture was diluted (1:25) with LB media to create a subculture that was grown at 30 °C, 225 rpm to OD₆₀₀ ~0.5. The cells were pelleted at 3,000 xg for 10 minutes, washed once with PBS, pelleted again, and then resuspended in a final volume of 25 mL PBS. Resuspended cells (5 mL) were then incubated with 200 µM fragment probe or control probe for 30 minutes in the dark at 30 °C, 225 rpm. Treated cells (5 mL) were then transferred to 6-well plates, placed on ice, and exposed to 3 J/cm² of 365-nm-wavelength UV light over 9 minutes (Analytik Jena UVP CL-1000 equipped with five 8-W F8T5 black lights) at 10 cm from the light source. Cells were collected and pelleted at 3,000 xg for 10 minutes, resuspended in 250 µL lysis buffer (30 mM Tris pH 7.0, 10 mM EDTA, 10 mg/mL lysozyme), and

incubated at 25 °C for 30 minutes. Total RNA was then extracted (1 mL of TRIzol; Invitrogen) and purified with on-column DNase treatment (Monarch Total RNA Miniprep; NEB).

Click labeling and RNA enrichment.

Total RNA was chemically fragmented at 94 °C for 2 minutes to yield RNAs with an average size of ~200 nucleotides (Magnesium RNA Fragmentation Module; NEB). RNA fragments were purified (silica-based SPE, using Monarch RNA Cleanup Kit; NEB). The cross-linked RNA fragments were conjugated to (magnetic) azide-containing beads via Cu(I)-catalyzed azide–alkyne cycloaddition. The cycloaddition was performed with 0.5 µM of heat-denatured RNA, 0.5 µM medium magnify azide click beads (Bang Laboratories, Inc.), 240 µM copper(II)-TBTA complex (Lumiprobe), 500 µM fresh ascorbic acid (Sigma), and 50% v/v DMSO, in a final reaction volume of 50 µL. Reactions were sparged with nitrogen, incubated at 40 °C for 30 minutes, and quenched with 1 µL 0.5 M EDTA. Cross-linked RNA was enriched by resuspending the RNA four times in wash buffer (1M NaCl, 10 mM Tris pH 7.5, 1 mM EDTA, 0.01% TWEEN-20) and a final time in nuclease-free water, the supernatant from each wash was removed via pipette after immobilizing the bead-conjugated RNA using a magnetic separation rack (NEB).

MaP reverse transcription.

Mutational profiling provides high sensitivity, and a significant decrease in false positives compared to other methods because it directly detects chemical adducts on RNA (50, 51). Frag-MaP directly detects ligand crosslinking at RNA binding sites using mutational profiling, which differentiates it from pull-down seq-based techniques such as Chem-CLIP-Frag-Map (49). Unlike competing methods, Frag-MaP is a single experiment that can simultaneously identify the RNA target and directly detect the binding site with nucleotide precision. RNA was subjected to mutational profiling reverse transcription with random nonamer primers for rRNA analysis, a 5 min 90 °C denaturation step was added (37). Bead-conjugated RNA was mixed with 100 ng of random nonamer primer (NEB), and 20 nmol of dNTPs (NEB), and denatured at 90 °C for 5 minutes followed by incubation at 4 °C for 2 minutes. MaP-RT buffer (6 mM MnCl₂, 1 M betaine, 50 mM Tris (pH 8.0 at 25 °C), 75 mM KCl, and 10 mM fresh DTT) was added to the RNA solution and incubated at 25 °C for 2 minutes. SuperScript II Reverse Transcriptase (1 µL, 200 units; Invitrogen) was added and the reverse transcription reaction was performed according to the following

temperature program: 25 °C for 10 min, 42 °C for 90 min, 10 × [50 °C for 2 min, 42 °C for 2 min], 72 °C for 10 min (37). The resulting cDNA was purified via magnetic pull-down using a magnetic separation rack (NEB).

Library preparation and sequencing.

Double-stranded DNA was prepared from the reverse transcription reaction product (NEBNext Ultra II Non-directional RNA Second Strand Synthesis Module; NEB) and DNA sequencing libraries were prepared (NEBNext Ultra II DNA Library Prep with Sample Purification; NEB). Libraries were then quantified (Qubit high-sensitivity dsDNA assay; Invitrogen and High sensitivity D1000 screentape; Agilent) and pooled as an equimolar sequencing library. The library was sequenced using 2x110 paired-end sequencing on an Illumina NextSeq 1000 instrument (P2 200 cycle, v3 chemistry; Illumina).

Analysis of Frag-MaP sequencing data.

ShapeMapper (v2.1.5) was used to align reads to the 23S rRNA and calculate per-nucleotide reactivities using the --random-primer-len option. The location of Frag-MaP sites was determined using the FragMapper analysis module written for RNavigate (v1.0) (72). FragMapper can be used to quantify experiments executed in both random-primed and gene-specific modes (38). These modes enable Frag-MaP analysis (*i*) of RNA-ligand interactions transcriptome-wide without prior knowledge of the RNA target or (*ii*) of targeted gene regions to yield very high-quality confirmation of hits from transcriptome-wide screens. FragMapper requires ShapeMapper profiles for an RNA treated with a fragment probe and a fragment-less control probe. First, FragMapper filters out nucleotides that do not meet a minimum read depth threshold or are hyperreactive to the fragment-less (non-selective control) probe (mutation rate > 2.5%). Second, a modified z-score is calculated for each nucleotide based on the difference (delta) in mutation rates between the fragment-treated and methyl-treated (fragment-less) samples. Nucleotides with a significantly higher mutation rate for the fragment-treated sample (modified z-score > 30, z-score - standard error > 5, and delta mutation rate > 1%) are identified as Frag-MaP sites. Our analysis focused on RNA-only pockets; Frag-MaP sites detected near proteins in our reference structure (PDB: 7AS8) were not analyzed (**Figure B.2i**). Three Frag-MaP sites (2145, 2154, and 2198) were identified in the L1 stalk which is a highly conserved component of the ribosome but not resolved in crystal structures for the

B. subtilis LSU. However, we found that the Frag-MaP sites in the L1 stalk were indeed close to pockets by analyzing a trRosettaRNA model for the *B. subtilis* L1 stalk (eRMSD: 1.7 Å) (41) and a crystal structure of the L1 Stalk from Haloarcula marismortui (PDB: 5ml7) (94). Frag-MaP sites were visualized in RNA tertiary structures using PyMOL (70).

Data availability

All data and software generated or used in this work are freely available. The fpocketR datasets reported in this study and fpocketR software suite and user manual are available at https://github.com/Weeks-UNC/2025_Veenbaas and <https://github.com/Weeks-UNC/fpocketR>, respectively. FragMapper software is available as an analysis module in RNavigate, <https://github.com/Weeks-UNC/RNavigate>. Frag-MaP datasets have been deposited in the Gene Expression Omnibus (GEO) database, <https://www.ncbi.nlm.nih.gov/geo> (accession no. GSE276279).

CHAPTER 4: APPLICATIONS OF FPOCKETR: FINDING LIGAND-BINDING POCKETS IN RNA⁴

Introduction

RNA molecules are essential regulators of gene expression, template protein synthesis, and carry out diverse additional cellular functions (95, 96). 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 (9, 29, 31, 75, 76, 97). 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 (32), and risdiplam and branaplam, splicing modifiers that bind pre-messenger RNA (33). 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 (9, 98, 99). 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 (83, 100, 101) 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 (25, 26).

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 (63, 102). 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

⁴ This chapter previously appeared as an article in *bioRxiv*. The original citation is as follows: S. D. Veenbaas, S. Felder, K. M. Weeks, fpoCKETR: A platform for identification and analysis of ligand-binding pockets in RNA. *bioRxiv* 2025.03.25.645323 (2025). <https://doi.org/10.1101/2025.03.25.645323>.

polar, solvent-exposed, and similar to each other to interact selectively with small molecules that have favorable physicochemical properties (99). We found, however, that fpocket can be optimized to detect RNA pockets capable of binding ligands with high druglikeness (QED) scores (62). 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 (99).

Several major conclusions emerged from the original development of fpocketR (99). 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.

Materials and methods

Pocket finding

Pockets were identified, characterized, and visualized using fpocketR 1.2.0, which functions as a wrapper for fpocket v4.0.3 (63, 102). 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) (43) and secondary structure predicted by Fold from the RNAstructure package (103). These sequence and secondary structure files were then used to generate a multiple sequence alignment with rMSA (104) using the NCBI nucleotide and RNACentral databases (105). The multiple

sequence alignment and predicted secondary structure were input to trRosettaRNA (41) 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 (99).

R-BIND library

The RNA-targeted Bloactive ligaNd Database (R-BIND) (v2.1) molecules (downloaded on January 14, 2025) includes organic small-molecule probes reported in the literature through December 2021 (60). The library was not filtered and contained 159 molecules.

FDA-approved ligand library

FDA-approved (Phase 4) molecules (106) (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 (58) (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) (77). The final library contained 365 RNA-ligand complexes (**Table D.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) (99).

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 (69). 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 (107), quantifying the multidimensional dissimilarity relative to the distribution of FDA-approved ligands.

Results

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 (**Figure 4.1A**). fpocketR also accepts RNA secondary structure drawing templates in multiple formats, including NSD (103), VARNA (108), and R2DT (109).

fpocketR runs the fpocket algorithm with optimized parameters (99), 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 (**Table D.2**), and a

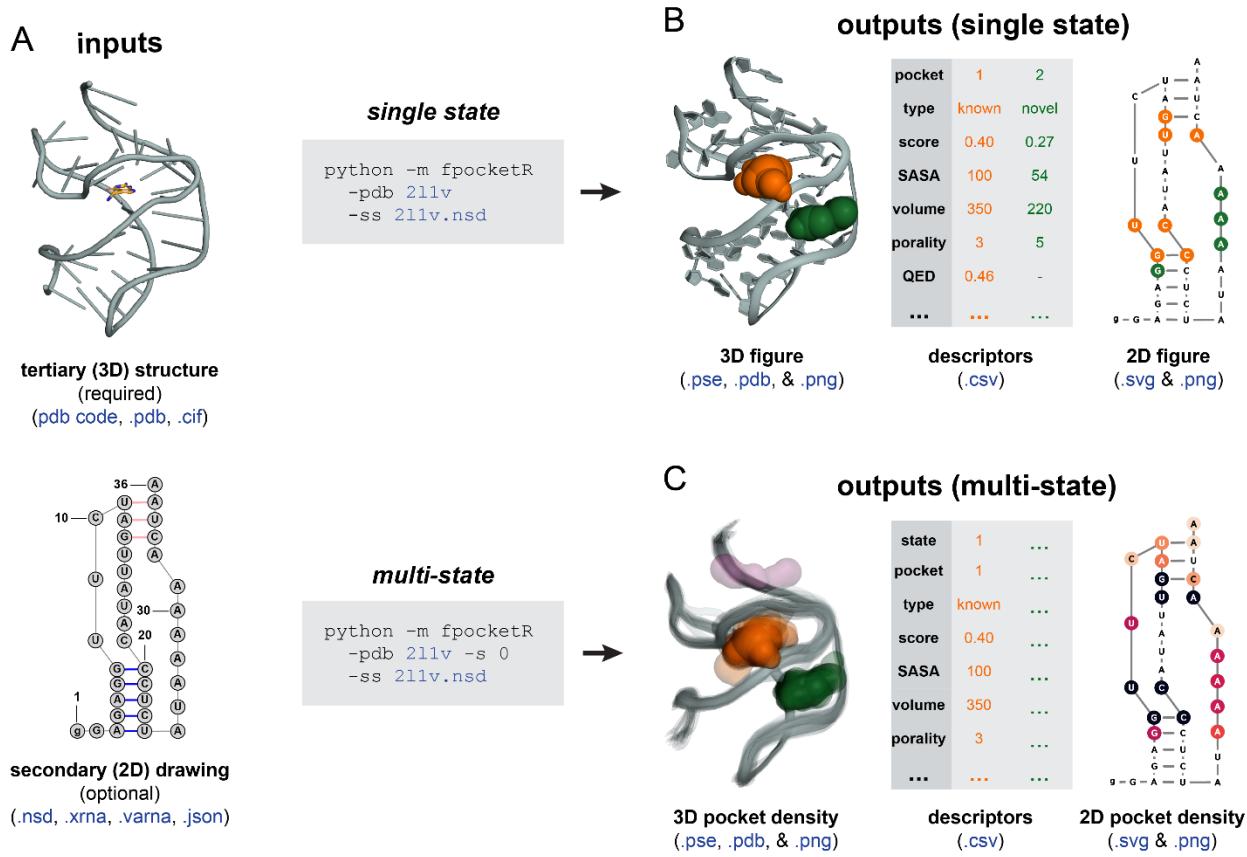


Figure 4.1. Pocket finding workflow for fpocketR, visualized with the class I preQ1 riboswitch (PDB: 2l1v). (A) Structural and command line inputs with accepted file formats. (B) Single-state analyses. Pockets are visualized as colored spheres and nucleotides that form pockets are highlighted in secondary 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 mapping the nucleotides that form each pocket onto a provided RNA secondary structure template (**Figure 4.1B** and **Table D.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 (**Figure 4.1C** and **Table D.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 (58). 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 (34, 40, 43). 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 (99).

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) (110). 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) (111). Between the two RNAs, fpocketR identified four novel pockets and both known pockets (**Figure 4.2**). 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-quadruplex (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

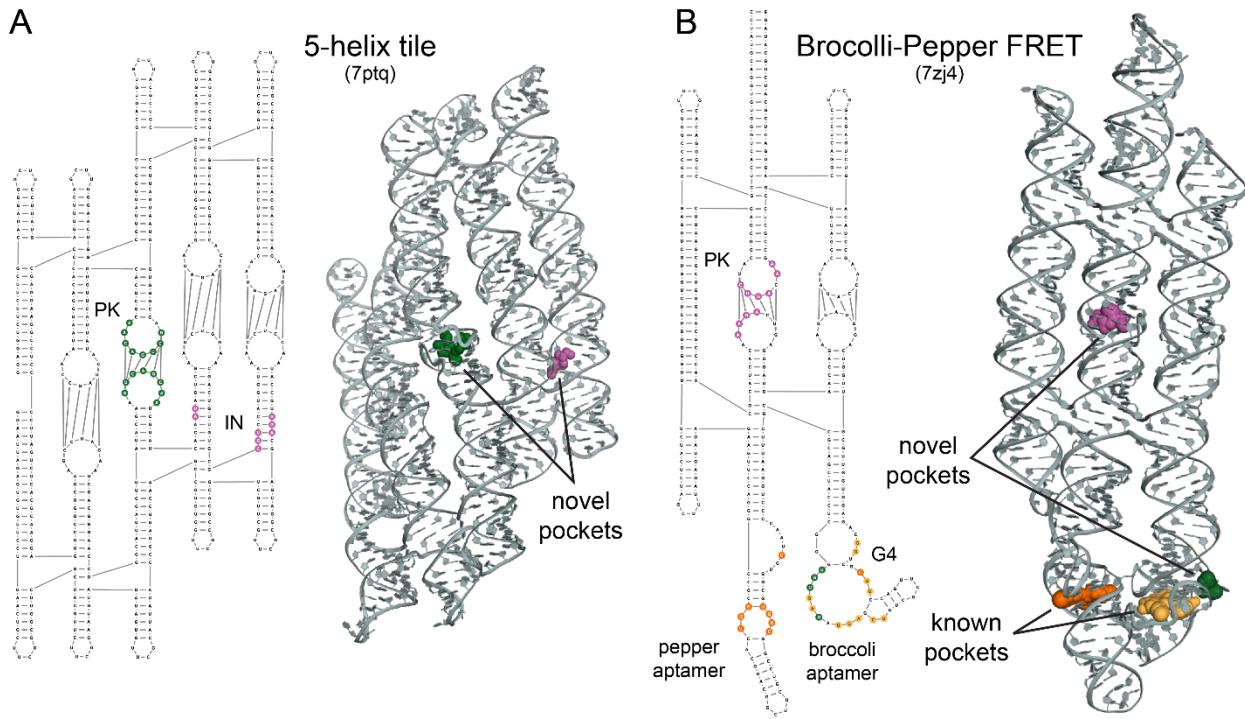


Figure 4.2. Pocket detection in large engineered RNAs. Pockets detected in (A) a five-helix RNA origami scaffold (544 nts) (110) and (B) an RNA engineered to include Pepper (orange) and Broccoli (yellow) ligand-binding aptamers (374 nts) (111). Novel RNA structures that form pockets are labeled as pseudoknots (PK), interfaces (IN) or G-quadruplex (G4).

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 (**Figure 4.2**) (99).

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 (40), 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 (112, 113). 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) (114), correctly identifying the ligand binding site for the preQ1 ligand (**Figure 4.3**). We applied fpoCKETR to 153 models of the same riboswitch submitted to the CASP15 evaluation (112). 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 (TM-scores)(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 (**Figure 4.3A**). 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 TM-score (**Figure 4.3B**). The TM-score metric was the best, but still only modest, predictor of RNA ligandability, with one-half (48%) of the models with a TM-score over 0.26 containing a pocket overlapping the ligand binding site (**Figure 4.3C**).

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) (85, 99). We generated a 20-model ensemble of the SAM-IV riboswitch using trRosettaRNA, among the most successful machine-learning-based RNA modeling programs (41, 115), and then identified pockets in the structures of the resulting ensemble using the multi-state analysis mode from fpoCKETR. Pockets in the

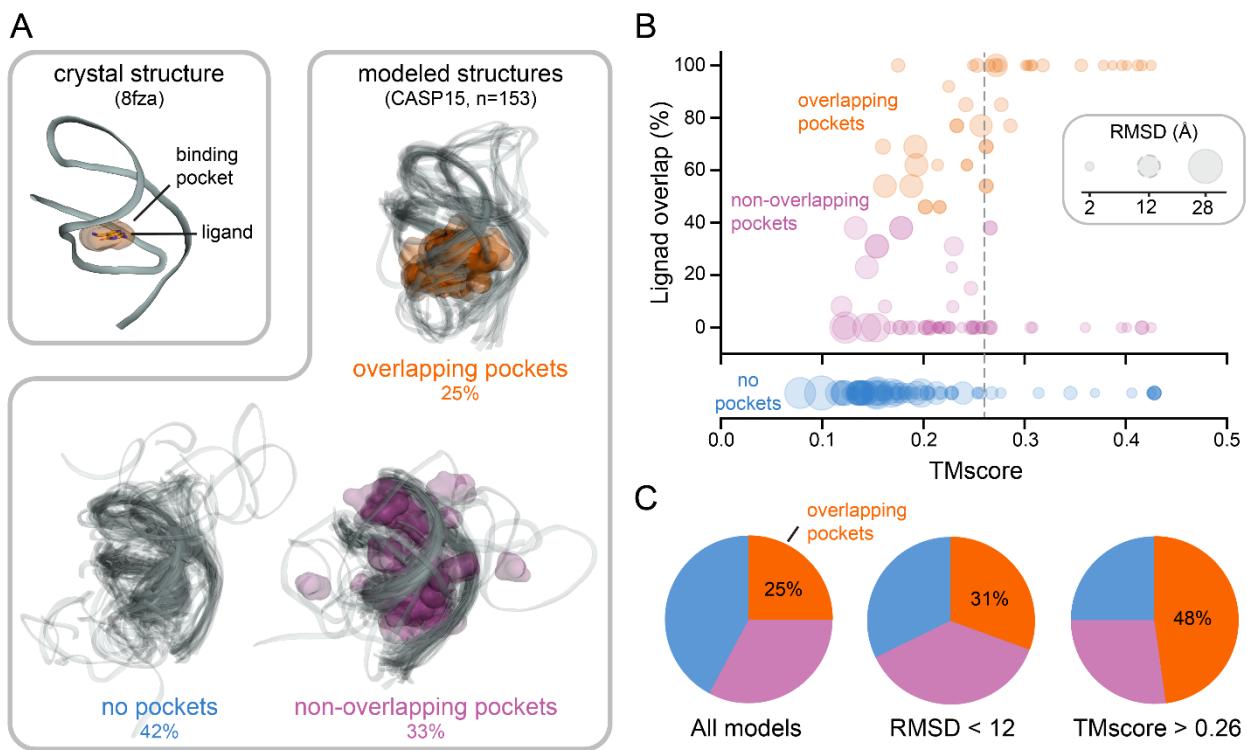


Figure 4.3. Pocket detection performance in modeled RNA structures. (A) Pockets detected in a reference structure (2.3 Å resolution; PDB 8fza) (114) and in 153 models of the class I type III preQ1 riboswitch, from the CASP15 exercise (112). (B) Relationship between pocket performance (ligand overlap) and structural quality metrics, RMSD and TM-score. Dashed gray circular (RMSD) and vertical (TM-score) 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 TM-score.

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) (**Figure 4.4**).

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 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 (83, 100). During translation, the ribosome undergoes large-scale conformational changes that include rotation of the small and large ribosomal subunits in a ratchet-like mechanism and swiveling of the head domain in the small ribosomal subunit (116, 117). 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 (118).

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) (117) (**Figure 4.5A**). 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 (119). The transient pockets at the B2a bridge are not present in the early stages of translocation (**Figure 4.5B**) and only form after back rotation of the small ribosomal subunit in the INT2 and POST states (**Figure 4.5C**). These transient pockets partially overlap with binding sites for the antibiotics thermorubin (119) and macrocyclic peptides, viomycin (120) and capreomycin (121) (**Figure 4.5D**). 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) (62). 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.

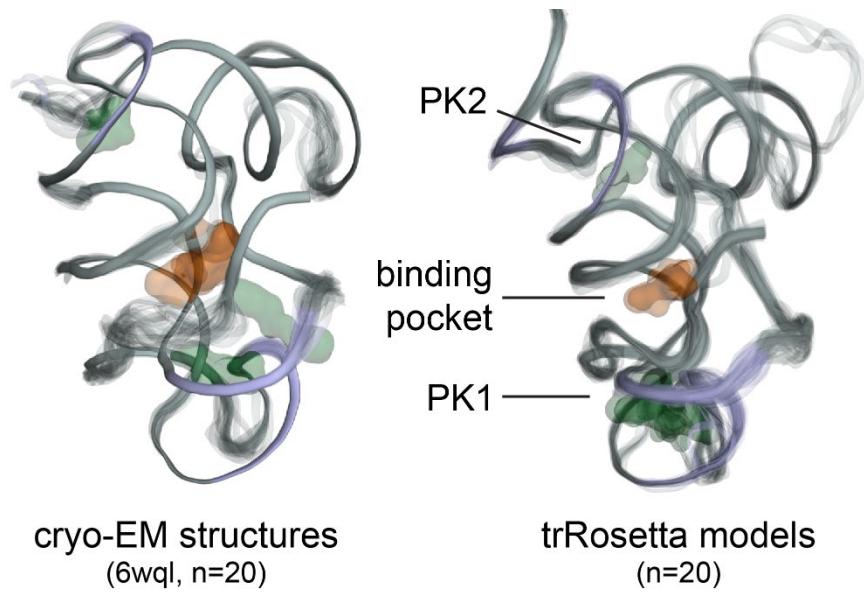


Figure 4.4. Pocket detection in multi-state ensembles. Map of pocket densities for ensembles of the SAM-IV riboswitch visualized by cryo-EM (PDB 6wql) (43) or generated by trRosettaRNA (41) modelling.

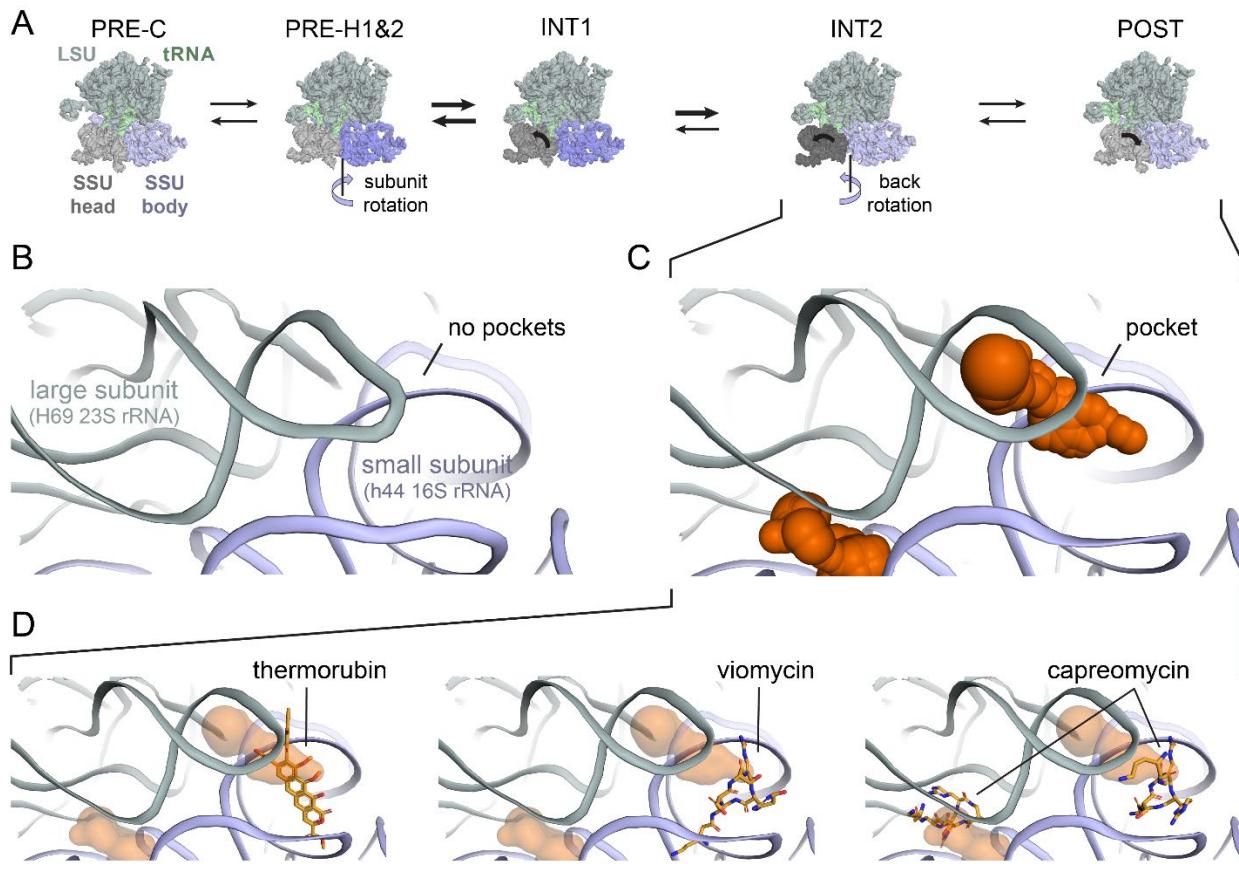


Figure 4.5. Identification of transient pockets at the interface of the large and small subunits of the bacterial ribosome, during translocation. (A) Overview of ribosome conformations at indicated stages of translocation (117). (B) Detailed view of the B2a inter-subunit bridge, which contains no pockets at early states of translocation. (C) Pockets in the B2a inter-subunit bridge are 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) (119), viomycin (6lkq) (120), and capreomycin (8ceu) (121).

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 (25, 26, 99). 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 (122). 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 (25, 26, 60).

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) (58) and a single reference structure of the bacterial ribosome (PDB 7k00) (77). 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 (99). 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 (**Figure 4.6**). 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 (**Figure 4.6**). 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

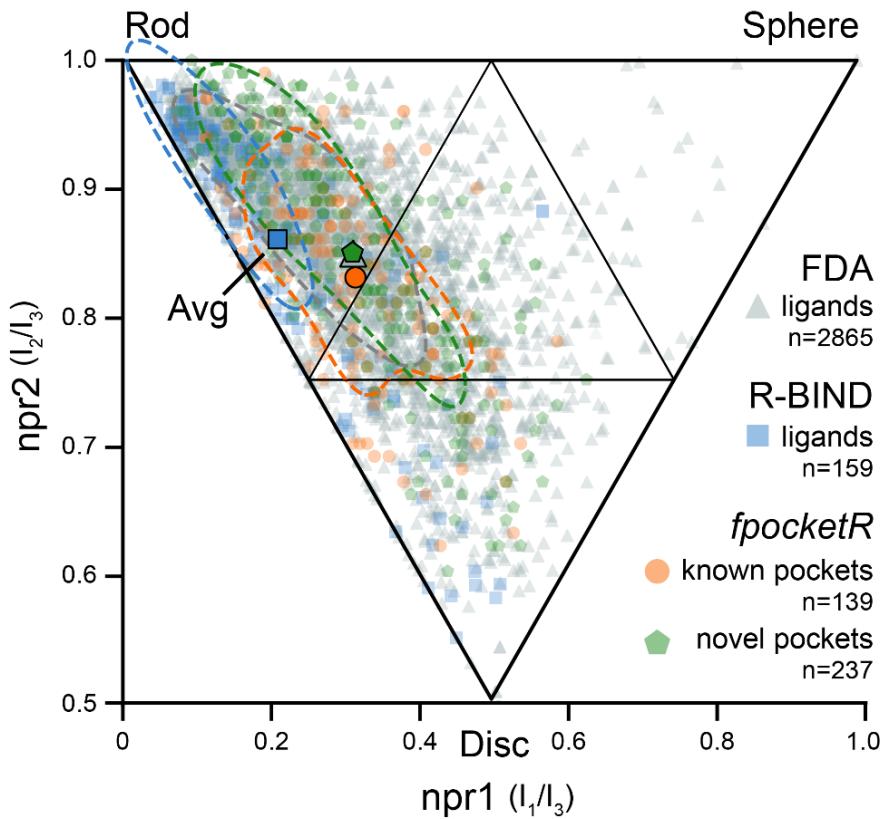


Figure 4.6. Comparison of shape-space for RNA-binding ligands versus RNA pockets. FDA-approved ligands (mostly protein binding), R-BIND ligands (exclusively RNA binding), and fpocketR-identified RNA pockets are shown. Normalized principal moment of inertia ratios (NPRs) for FDA-approved ligands (140-590 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.

rod-like than ligands that bind to repeat sequences (123). Together, these analyses suggest that RNAs contain pockets with a wide range of shapes able to bind ligands with equally diverse shapes (**Figure 4.6**). 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.

Discussion

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 (99). 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.

Data 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) and an x86 CPU architecture. 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>.

APPENDIX A: SUPPLEMENTAL INFORMATION FOR CHAPTER 2

Supplemental figures

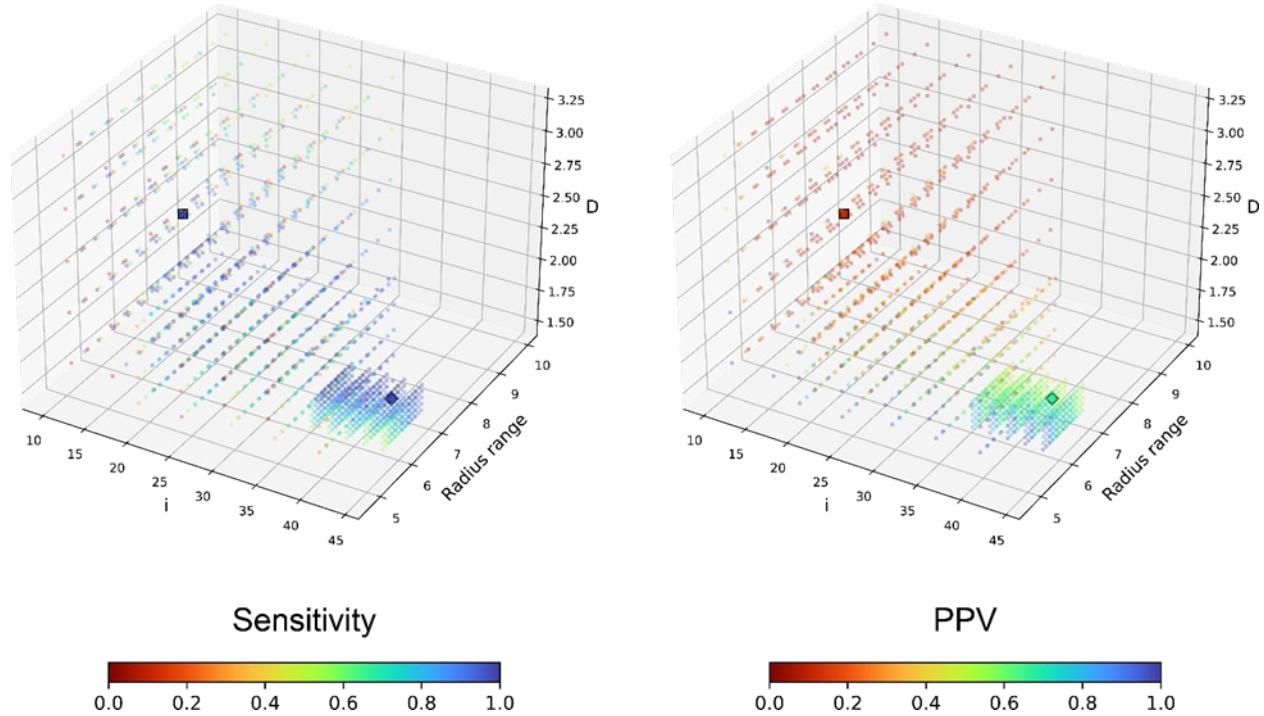


Figure A.1. Parameter space evaluated during focketR optimization. Each point represents a unique combination of parameters (i , m , M , and D). Parameters m and M which control alpha sphere radius were collapsed to a single value representing radius range ($M^{1.3} - m$) to allow visualization in three-dimensions. Points are colored to reflect the sensitivity and positive predictive value (ppv) for detecting known ligand binding sites in the training set (n=13). The default ($i=15$, $m=3.4$, $M=6.2$, $D=2.4$) and optimized ($i=42$, $m=3.0$, $M=5.7$, $D=1.65$) parameters are depicted as squares and diamonds respectively

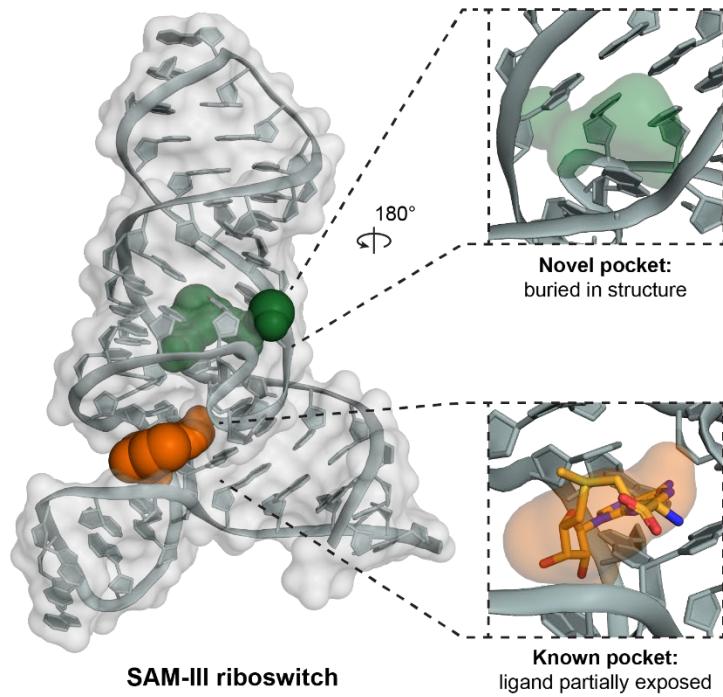


Figure A.2. Pockets detected in the SAM-III riboswitch using fpocketR (PDB: 3E5C) (89). The known pocket is formed in a three-way multi-helix junction and overlaps with most of the native SAM ligand but leaves the methionine tail exposed. The novel pocket forms between a three-way multi-helix junction and a bulge. Although the novel pocket is smaller in volume than the known pocket, it is more deeply buried in the RNA structure and has less solvent exposed surface area.

Supplementary Tables

Table A.1. Training set library used to optimize fpocketR.

PDB code	reference	length	ligand name	QED
1F27	(65)	30	Biotin	0.49
8D2B	(124)	33	TAL2	0.83
1Q8N	(66)	38	Malachite green	0.75
7ELR	(125)	45	Xanthine	0.45
3E5C	(89)	53	SAM (III)	0.34
6FZ0	(126)	53	SAM (V)	0.34
3NPQ	(127)	54	SAH	0.35
6UBU	(128)	67	Guanine	0.46
2GDI	(67)	80	TPP	0.79
6LAS	(129)	93	SAM (VI)	0.34
4B5R	(130)	94	SAM (I)	0.34
4RZD	(131)	101	PreQ1 (III)	0.46
5KX9	(68)	112	FMN	0.33
Average		66	-	0.48

Table A.2. Test set library used for evaluating fpocketR optimizations.

PDB code	reference	length	ligand name	QED
7FJ0	(132)	20	KG022	0.63
6VA4	(133)	21	MIP	0.70
1LVJ	(134)	31	Acetylpromazine	0.76
3Q50	(87)	33	Pre-Q1 (I)	0.46
6YL5	(135)	35	SAH (SAM-SAH)	0.35
2KTZ	(136)	38	Isis-11	0.87
6UP0	(137)	38	YO3-biotin	0.66
6XB7	(138)	41	DMA-135	0.40
6GZR	(139)	48	5-TAMRA	0.38
1YKV	(140)	49	Ethanoanthracene	0.64
7EOH	(141)	49	HBC	0.68
8EYU	(142)	49	DFAME	0.68
2QWY	(143)	52	SAM (II)	0.34
7OAW	(144)	52	DMHBI+	0.62
8HB3	(145)	55	Nicotinamide riboside	0.45
6XJQ	(146)	58	2,3-disubstituted epoxide	0.38
3SKI	(147)	68	2'-Deoxy-guanosine	0.51
5OB3	(148)	69	DFHBI	0.86
5KPY	(149)	71	5-Hydroxytryptophan	0.62
4LX5	(150)	71	PPDA	0.55
5BTP	(151)	75	ZMP	0.31
4JF2	(152)	77	Pre-Q1 (II)	0.46
7KVT	(153)	83	DFHBI-1T	0.67
7DWH	(154)	102	SAM	0.34
3D0U	(86)	161	Lysine	0.46
Average		58	-	0.55

Table A.3. Parameter combinations tested during 3 rounds of fpocketR optimization.

Round	<i>m</i>	<i>M</i>	<i>i</i>	<i>D</i>
Default	3.4	6.2	15	2.4
1	2.6 3.0 3.4 3.8 4.2	5.4 5.8 6.2 6.6 7.0	10 15 20 25 30	1.6 2.0 2.4 2.8 3.2
2	2.6 2.8 3.0 3.2 3.4	5.0 5.2 5.4 5.6 5.8	20 25 30 35 40	1.6 1.8 2.0 2.2 2.4
3	2.8 2.9 3.0 3.1 3.2	5.4 5.5 5.6 5.7 5.8	36 38 40 42 44	1.5 1.55 1.6 1.65 1.7
Optimized	3.0	5.7	42	1.65

Note: The optimal parameters from each round are highlighted in orange.

Table A.4. Pocket detection performance of fpocket and fpocketR in a small RNAs.

parameters	avg # of pockets	sens	rank 1 sens	ppv
fpocket (<i>protein optimized</i>)	4.7	87%	63%	19%
fpocket-R (<i>RNA optimized</i>)	1.3	100%	92%	78%

Table A.5. Pocket characteristics for RNA-ligand versus protein-ligand complexes.

library	parameters	pocket type	score	# alpha spheres	SASA	volume	hydro-phobic density	hydro-phobicity score	polarity score
RNA (n=38)	fpocket-R (RNA optimized)	all	0.34	84	117	351	8.17	-1.1	4.6
		rank 1	0.38	93	119	369	9.85	-1.0	4.6
		known	0.36	90	116	361	9.45	-1.8	4.5
	fpocket (protein optimized)	all	0.02	38	134	472	1.72	1.2	4.5
		rank 1	0.23	61	146	560	4.92	-0.8	5.2
		known	0.17	96	206	750	7.79	-1.8	6.3
	fpocket-R (RNA optimized)	all	0.35	64	95	318	19.89	24.2	7.7
		rank 1	0.49	77	103	370	21.12	25.6	8.8
		known	0.38	77	119	401	19.10	22.7	9.2
protein (n=200)	fpocket (protein optimized)	all	0.03	32	103	342	9.59	18.1	5.7
		rank 1	0.28	83	169	635	19.81	26.6	10.6
		known	0.16	84	191	676	21.10	24.6	10.8

APPENDIX B: SUPPLEMENTAL INFORMATION FOR CHAPTER 3

Materials and Methods

General chemistry information.

Reactions were carried out in amber glass sample vials. All reagents, starting materials, and solvents (including dry solvents) were obtained from commercial suppliers and used without further purification. Thin-layer chromatography (TLC) was performed using commercial silica gel 60 F₂₅₄ coated aluminum-backed sheets; products were visualized with UV light. Purification was carried out by automated flash chromatography (Selekt; Biotage) using normal phase columns (Sfär; Biotage). All NMR (¹H and ¹³C) spectra were recorded at 400 MHz with a dual carbon/proton cryoprobe, or at 600 MHz with a dual carbon/proton cryoprobe. NMR samples were recorded in CDCl₃ or CD₃OD. Chemical shifts are reported in parts per million (ppm) and referenced to the center line of residual solvent (for CDCl₃, δ 7.26 ppm for ¹H NMR and 77.16 for ¹³C NMR; for CD₃OD, δ 3.31 for ¹H NMR and 49.0 ppm for ¹³C NMR). Coupling constants are reported in Hertz (Hz). High-resolution mass spectrometry was performed on an Agilent 1200 series analytical high performance liquid chromatography system coupled to an Agilent 6520 Accurate Mass quadrupole time of flight (Q-TOF) mass spectrometer and an electrospray ion source (ESI). Fragment 206 was obtained from Enamine.

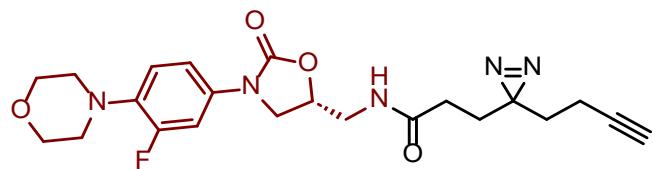
General synthesis procedure for fully functionalized fragment probes.

All compounds were synthesized following a general amide coupling procedure (79). 3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)propanoic acid (1 eq, 60 mM), DIPEA (3.0 eq), EDC-HCl (1.5 eq) and HOBt (1.5 eq) were dissolved in DCM and added to a commercially available amine (1.1 eq). Reaction mixtures were stirred at 20 °C overnight and monitored by TLC. The crude product was diluted in DCM (10 mL), washed with saturated aqueous NH₄Cl (10 mL), and then washed with saturated aqueous NaHCO₃ (10 mL). The organic layer was dried over anhydrous Na₂SO₄ and solvent removed by rotary evaporation under reduced pressure. The crude products were purified by PTLC or flash column chromatography (Biotage).



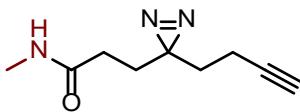
Synthesis of linezolid probe: (R)-3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)-N-((3-(3-fluoro-4-morpholinophenyl)-2-oxooxazolidin-5-yl)methyl)propanamide.

Prepared using the general synthesis procedure and (R)-5-(aminomethyl)-3-(3-fluoro-4-morpholinophenyl)oxazolidin-2-one (39 mg, 0.13 mmol). Purified by SiO₂ flash chromatography on a Biotage (DCM/MeOH, 100:0 to 99:1), yielding 13.2 mg of a white solid (25%). ¹H NMR (600 MHz, Methanol-d₄) δ: 7.51 (dd, *J* = 14.7, 2.6 Hz, 1H), 7.18 (ddd, *J* = 8.8, 2.6, 1.1 Hz, 1H), 7.07 (t, *J* = 9.2 Hz, 1H), 4.80 (dddd, *J* = 9.0, 6.2, 5.3, 4.2 Hz, 1H), 4.13 (t, *J* = 9.0 Hz, 1H), 3.85 (t, *J* = 6.3 Hz, 1H), 3.84 (t, *J* = 4.4 Hz, 4H), 3.62 (dd, *J* = 14.5, 5.2 Hz, 1H), 3.51 (dd, *J* = 14.5, 4.0 Hz, 1H), 3.06 (t, *J* = 4.4 Hz, 4H), 2.26 (t, *J* = 2.7 Hz, 1H), 2.07 – 2.00 (m, 2H), 1.95 (td, *J* = 7.5, 2.7 Hz, 2H), 1.71 (td, *J* = 7.5, 3.0 Hz, 2H), 1.58 – 1.52 (m, 2H). ¹³C NMR (151 MHz, MeOD) δ: 137.3, 135.2, 135.1, 130.2, 126.7, 115.4, 110.7, 108.5, 108.3, 86.9, 83.6, 73.5, 70.4, 67.9, 52.4, 52.4, 49.4, 49.3, 49.1, 49.0, 48.9, 48.7, 48.6, 42.9, 33.3, 30.9, 29.7, 28.8, 13.8. HRMS (ESI-QTOF) calculated [M+H]⁺ = 444.2042; observed m/z = 444.2057.



Synthesis of methyl probe: 3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)-N-methylpropanamide.

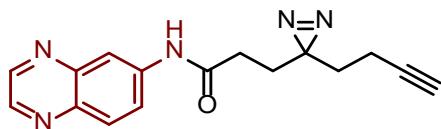
Prepared using the general synthesis procedure and methylamine HCl (8.9 mg, 0.13 mmol). Purified by SiO₂ flash chromatography on a Biotage (Hexanes/EtOAc, 7:3 to 1:1) yielding 15.0 mg of colorless solid (70%). ¹H NMR (400 MHz, Chloroform-d) δ: 5.51 (s, 1H), 2.81 (d, *J* = 3.9 Hz, 3H), 2.05 – 1.97 (m, 3H), 1.92 (ddd, *J* = 8.0, 6.2, 1.8 Hz, 2H), 1.85 (ddd, *J* = 8.7, 6.3, 1.8 Hz, 2H), 1.65 (t, *J* = 7.4 Hz, 2H). HRMS (ESI-QTOF) calculated [M+H]⁺ = 180.1131; observed m/z = 180.1130.



Synthesis of 6-Aminoquinoxaline probe: 3-(3-(but-3-yn-1-yl)-3H-diazirin-3-yl)-N-(quinoxalin-6-yl)propenamide.

Prepared using the general synthesis procedure and quinoxaline-6-amine (19 mg, 0.13 mmol).

Purified by PTLC (Pentane/EtOAc, 1:4), yielding 4.5 mg of a pale-yellow oil (13%). ¹H NMR (400 MHz, Methanol-*d*₄) δ 8.79 (dd, *J* = 19.2, 1.9 Hz, 2H), 8.52 (d, *J* = 2.3 Hz, 1H), 8.02 (d, *J* = 9.1 Hz, 1H), 7.92 (dd, *J* = 9.1, 2.4 Hz, 1H), 2.34 – 2.24 (m, 3H), 2.06 (td, *J* = 7.5, 2.7 Hz, 2H), 1.89 (dd, *J* = 8.4, 6.9 Hz, 2H), 1.66 (t, *J* = 7.4 Hz, 2H). ¹³C NMR (101 MHz, Methanol-*d*₄) δ 173.16, 146.77, 144.98, 144.63, 141.85, 141.00, 130.43, 125.49, 117.34, 83.60, 70.35, 33.50, 31.93, 29.38, 28.89, 13.86. HRMS (ESI-QTOF) calculated [M+H]⁺ = 294.1349; observed m/z = 294.1348.



Comparison of protein and RNA pocket detection and properties.

Protein structures (*n* = 200) were randomly selected from a curated list of protein-ligand complexes (155) and acquired from the PDB (156). All RNA structures in the training and test sets of this work were used. Pocket prediction and characterization for both protein-ligand and RNA-ligand complexes were performed using fpocketR with default (or fpocket) parameters defined as (*m* = 3.4, *M* = 6.2, *D* = 2.4, and *i* = 15) and optimized (or fpocketR) parameters defined as (*m* = 3.0, *M* = 5.7, *D* = 1.65, and *i* = 42).

Supplementary Figures

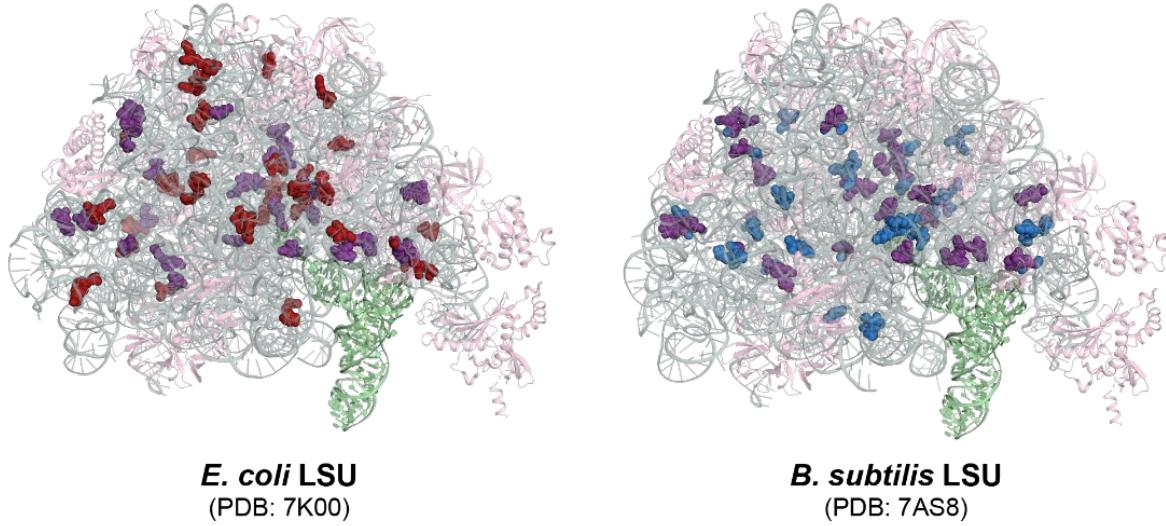
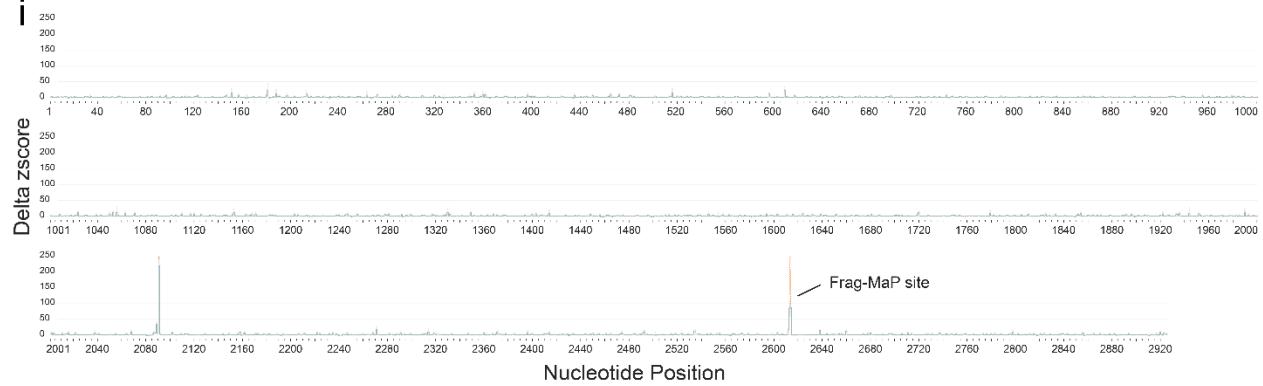
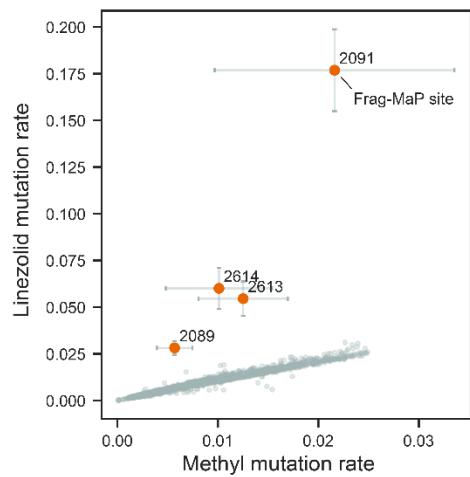
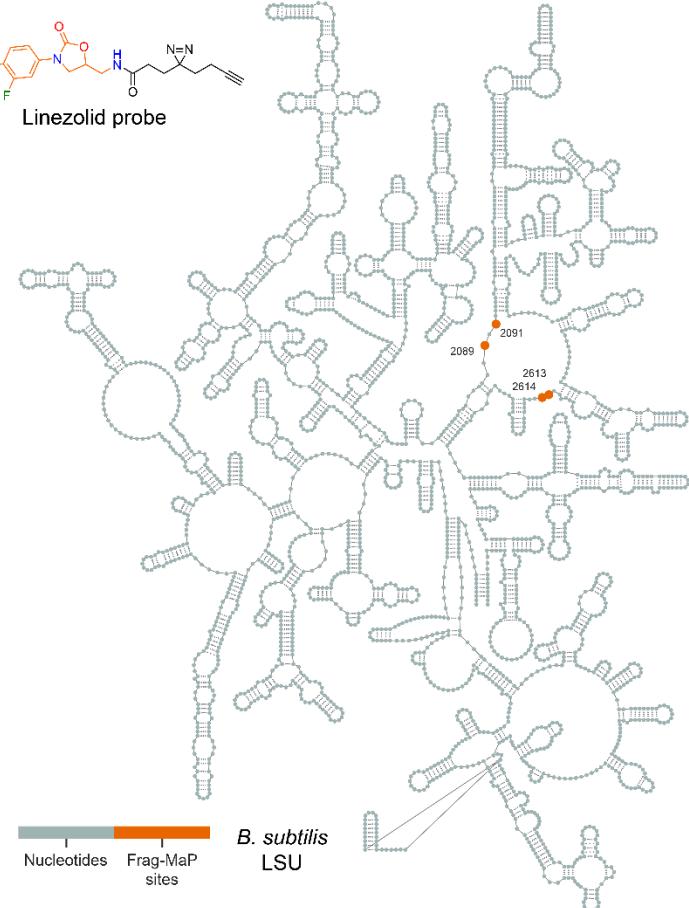
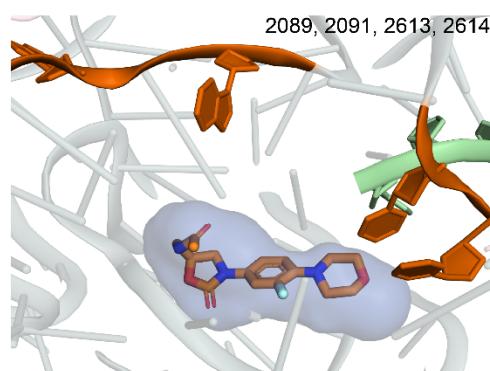
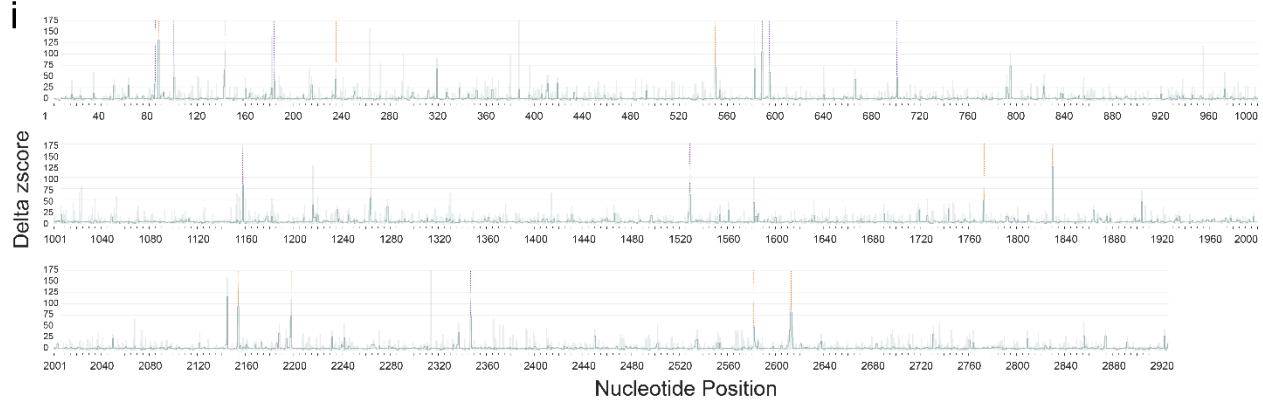
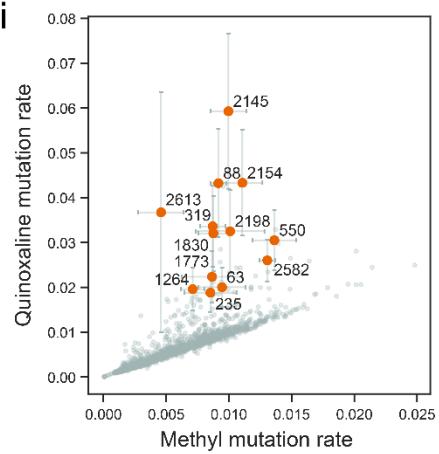
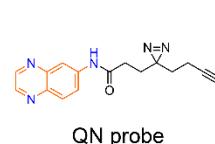
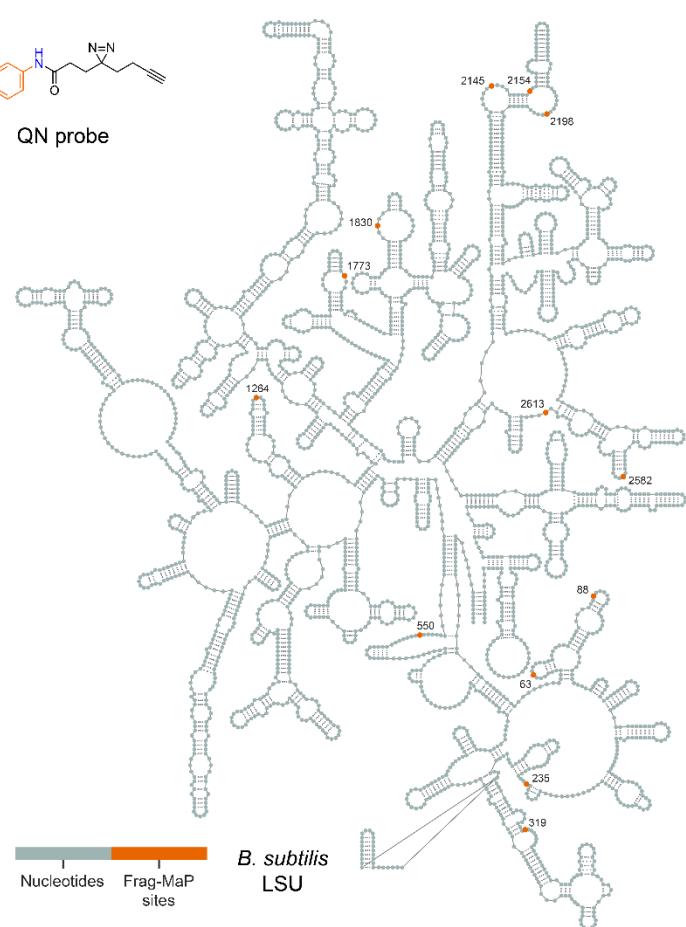
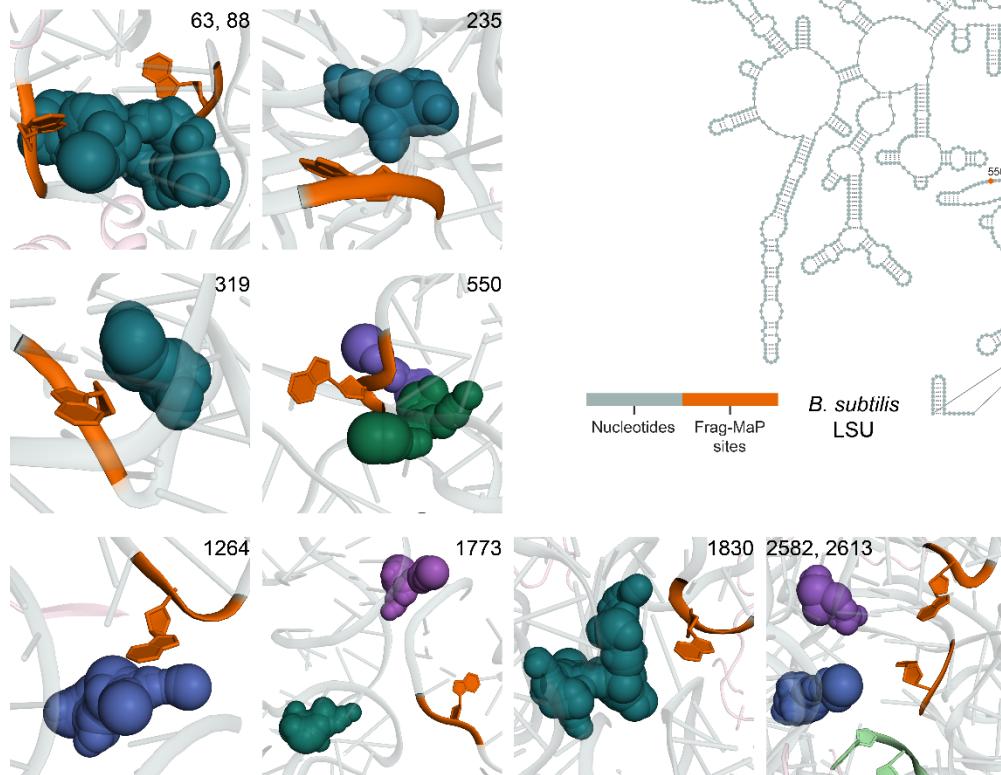


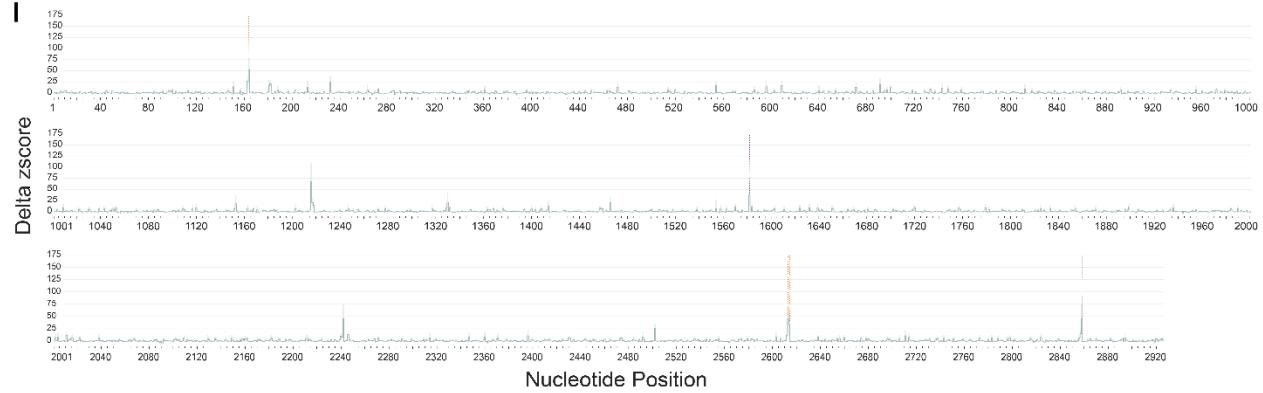
Figure B.1. The ribosomal pocket-ome. fpocketR detects 46 and 52 pockets in the classical (non-rotated) state of the 23S rRNA for *E. coli* and *B. subtilis* (77, 78), respectively. Of these, 21 pockets overlap between the two species (in purple). Ribosomal proteins are pink; A-site and P-site tRNAs are green.

A**i****ii****iii****iv****v**

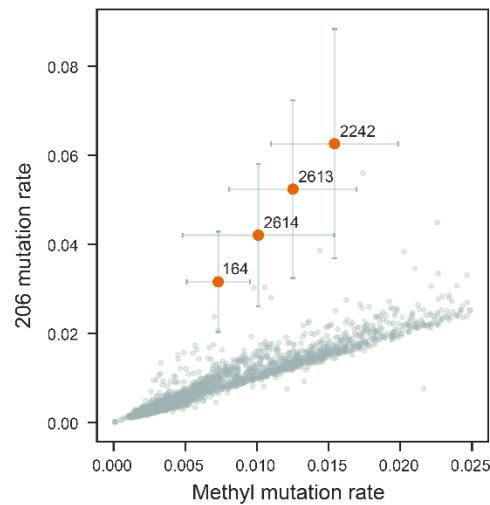
B**i****ii****iii****iv****v**

C

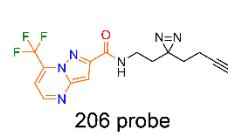
i



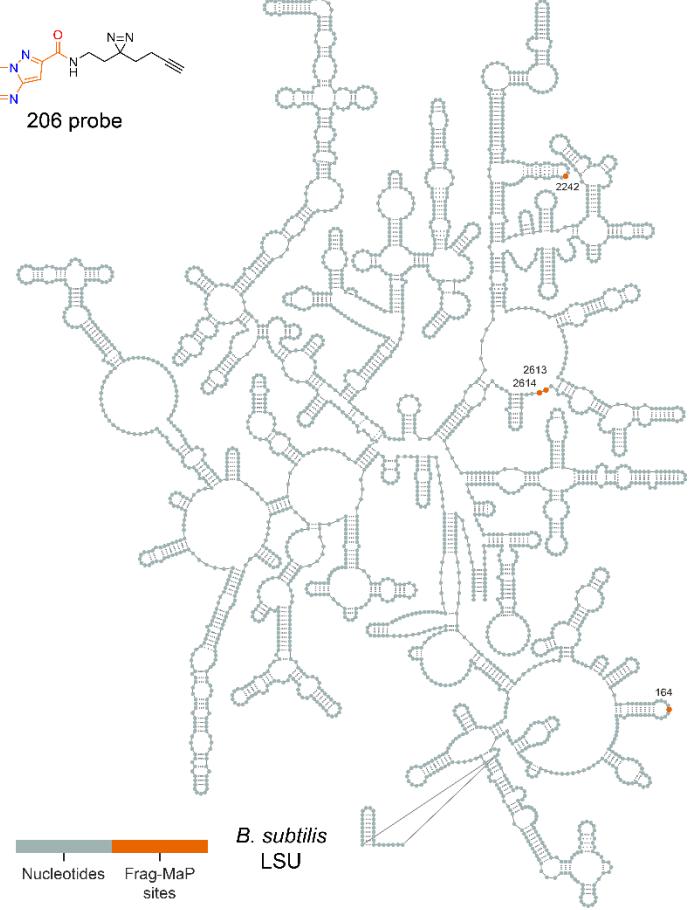
ii



iii



iv



v

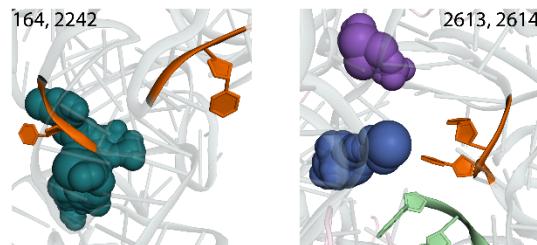


Figure B.2. Frag-MaP sites identified in the *B. subtilis* 23S rRNA using the (A) linezolid, (B) quinoxaline, and (C) 206 functionalized probes. (3 panels, *above*). For each panel: (i) Frag-MaP profiles displaying the z-score of the difference in mutation rate between the fragment and methyl probes for each nucleotide. Frag-MaP sites analyzed in this work are depicted in orange. Frag-MaP sites at RNA-protein interactions or in structurally unvisualized regions, not analyzed in this work, are purple. Error is shown as the standard error of the mean. (ii) Comparison of the per-nucleotide mutation rate between the fragment and methyl probes. (iii) Structure of the functionalized fragment probe. (iv) Location of Frag-MaP sites (orange) within the secondary structure of the *B. subtilis* large subunit (LSU). (v) Proximity of Frag-MaP sites (orange) to pockets predicted by fpocketR (colored spheres) (PDB: 7AS8) (78).

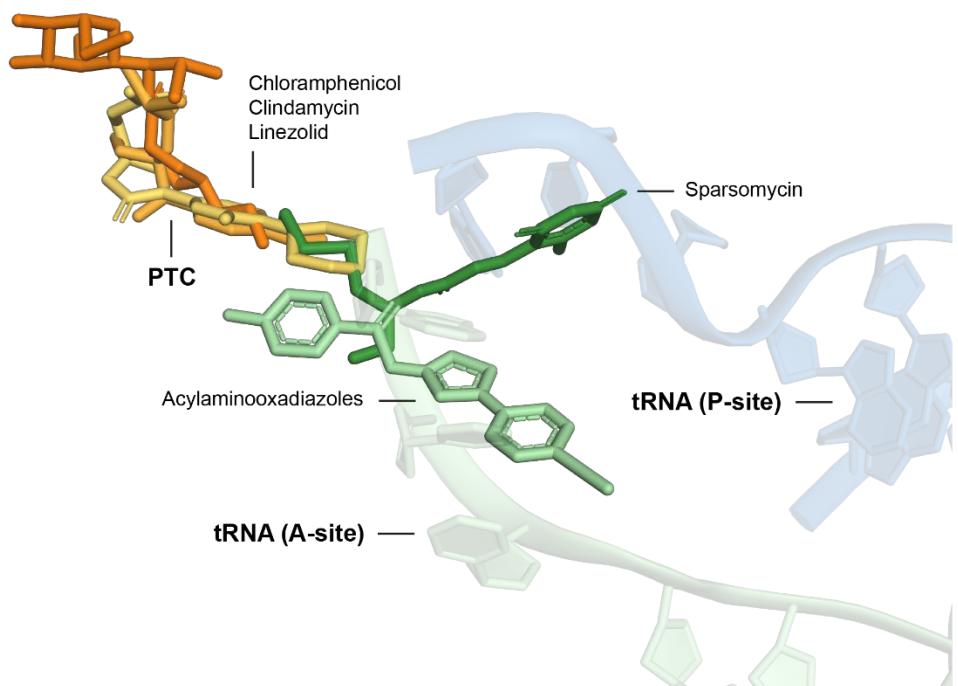
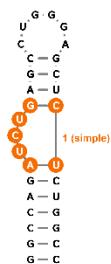


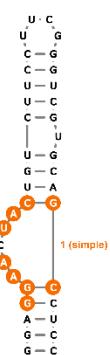
Figure B.3. Distinct binding regions of ribosomal antibiotics near the peptidyl transferase center in the bacterial 23S rRNA. The binding positions of chloramphenicol, clindamycin, and linezolid superimpose in the peptidyl transferase center. Sparsomycin and the acylaminooxadiazoles bind in an adjacent region overlapping the A-site tRNA. PDB codes are 6ND5, 4V7V, 3CPW, 1M90, and 6OM6 (82, 157–160)

Simple structures

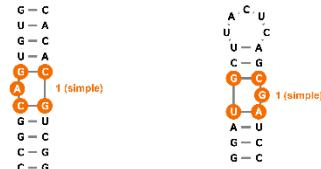
1lvj_state3



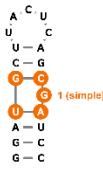
2ktz



6va4

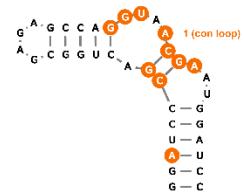


7fj0

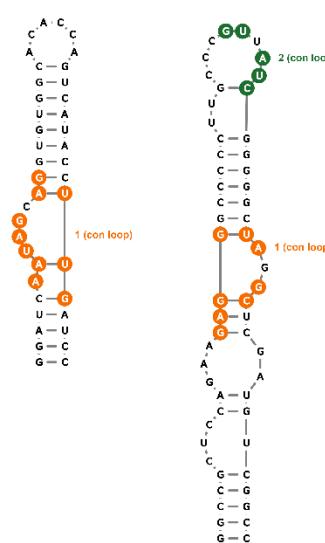


Consecutive Loops

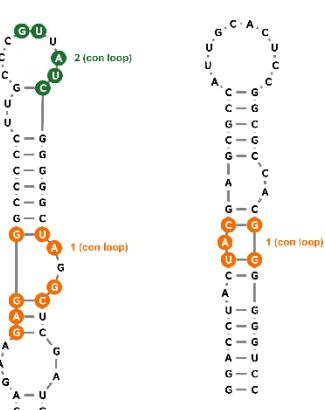
1q8n



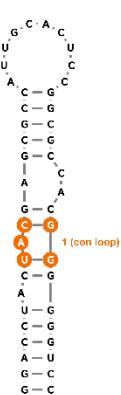
6xb7_state3



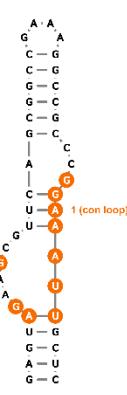
6xjq



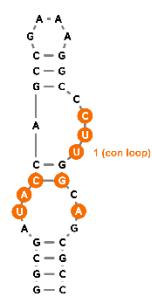
7dwh



7elr

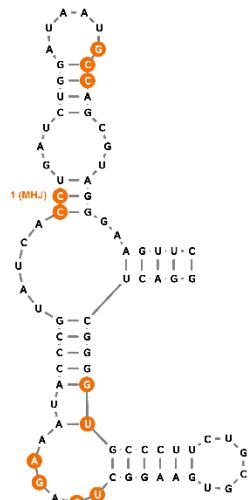


8d2b

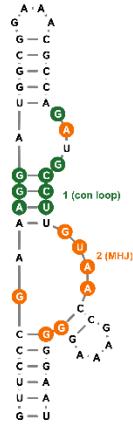


Multi-helix Junctions

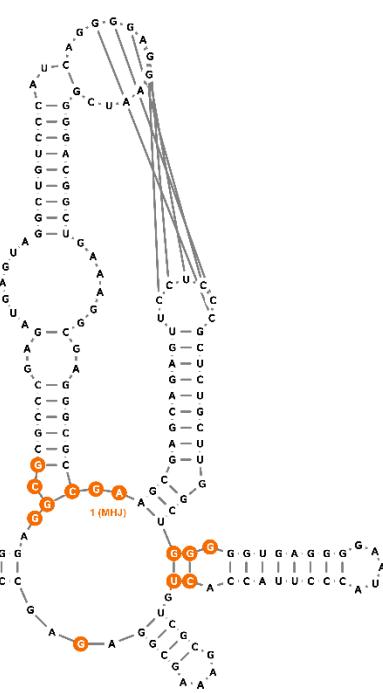
2gdi



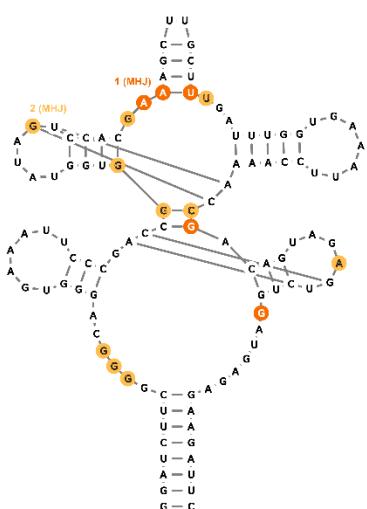
3e5c



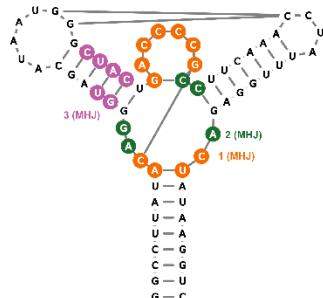
3d0u



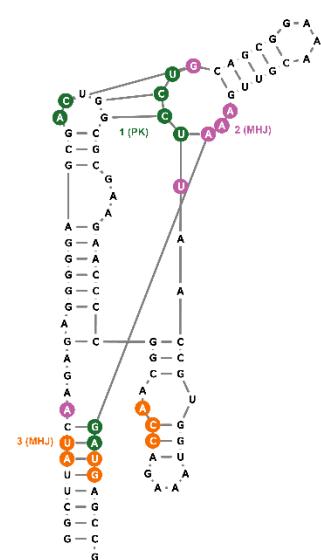
5kx9

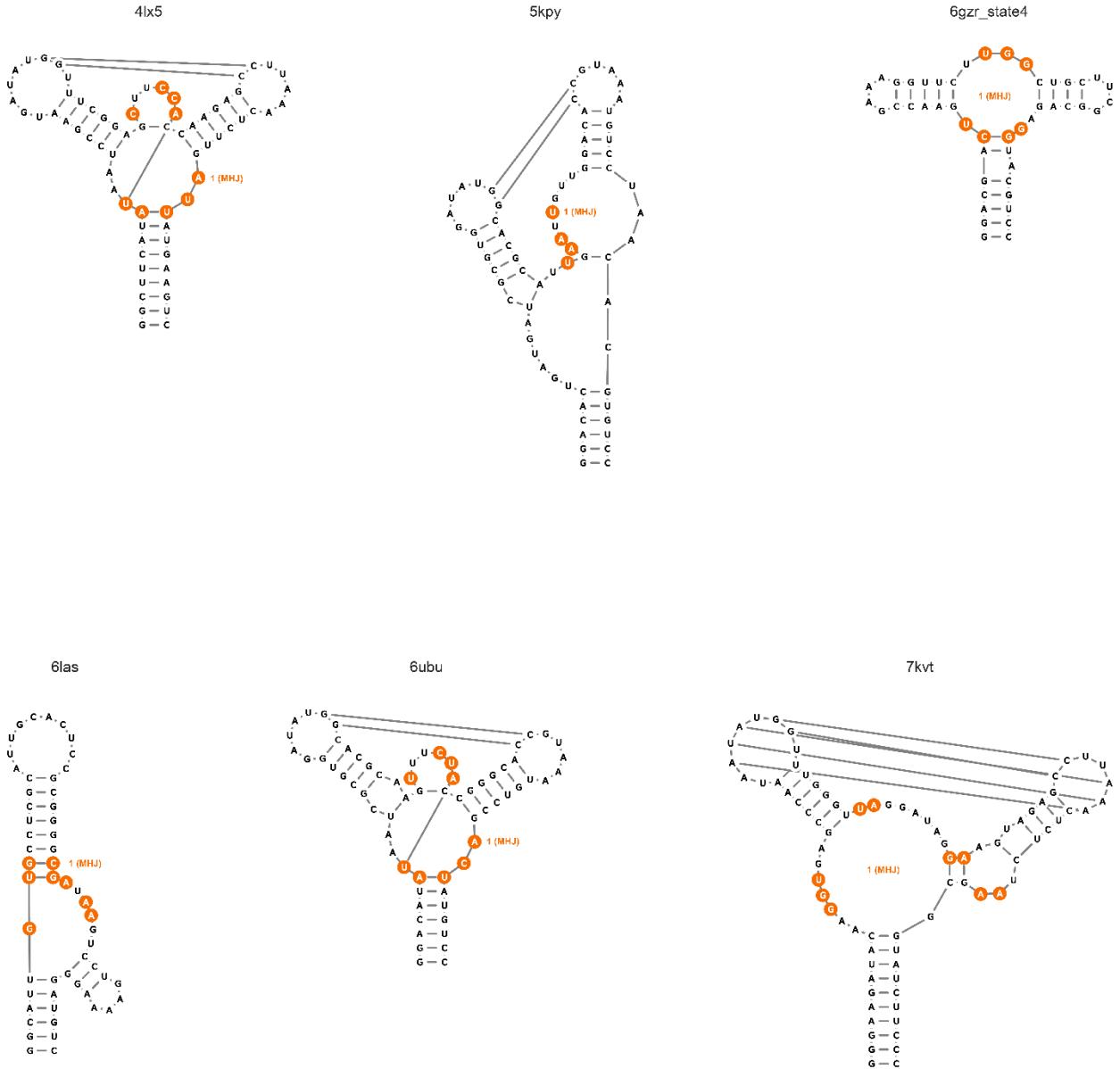


3ski



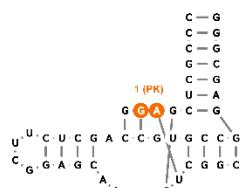
4b5r



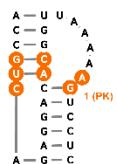


Pseudoknots

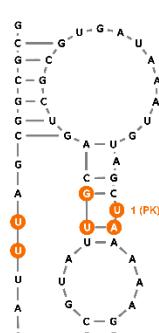
1ykv



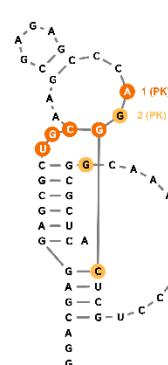
1f27



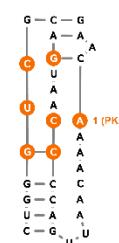
2qwy



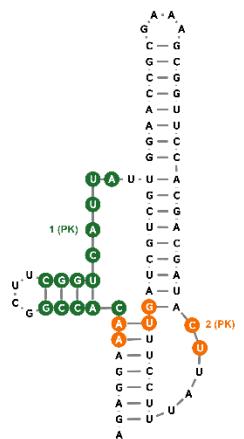
3npq



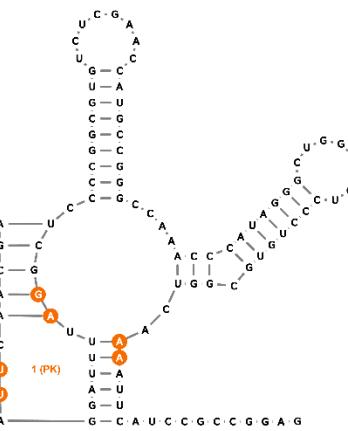
3q50



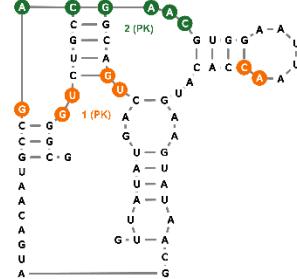
4jf2



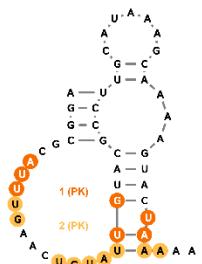
4rzd



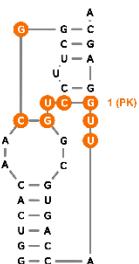
5btp



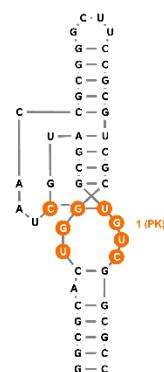
6fz0



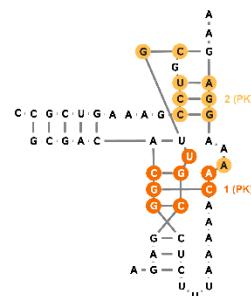
6yl5



7eoh



8hb3



G-quadruplexes

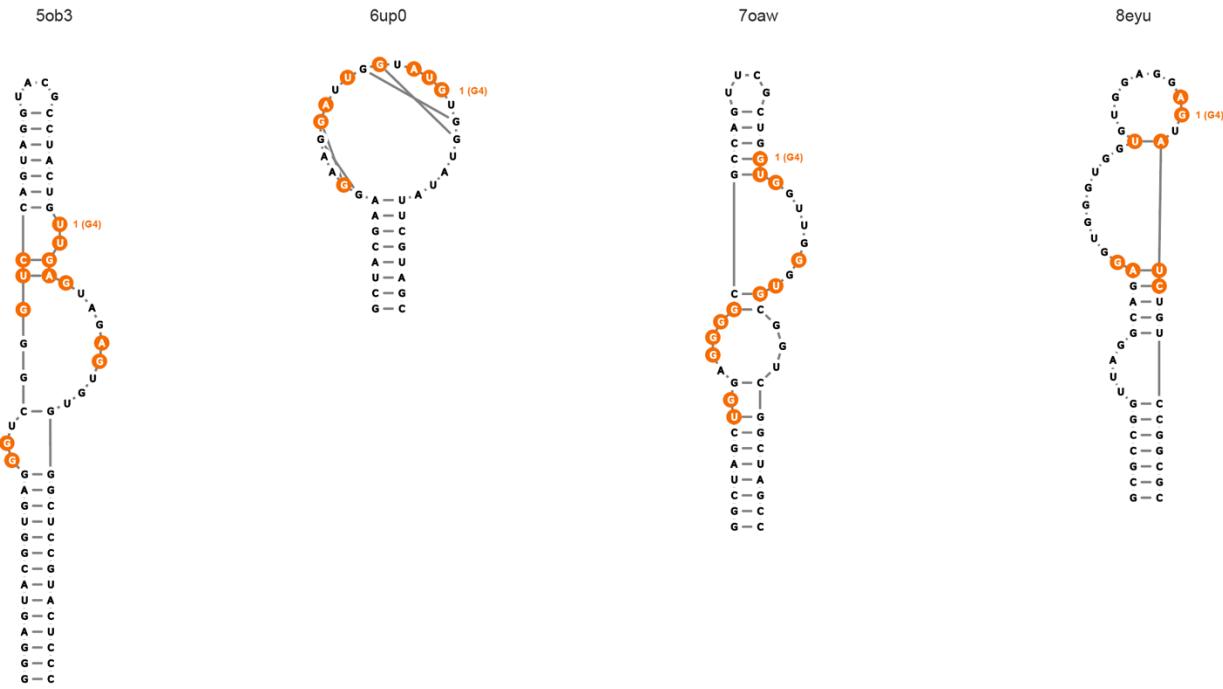
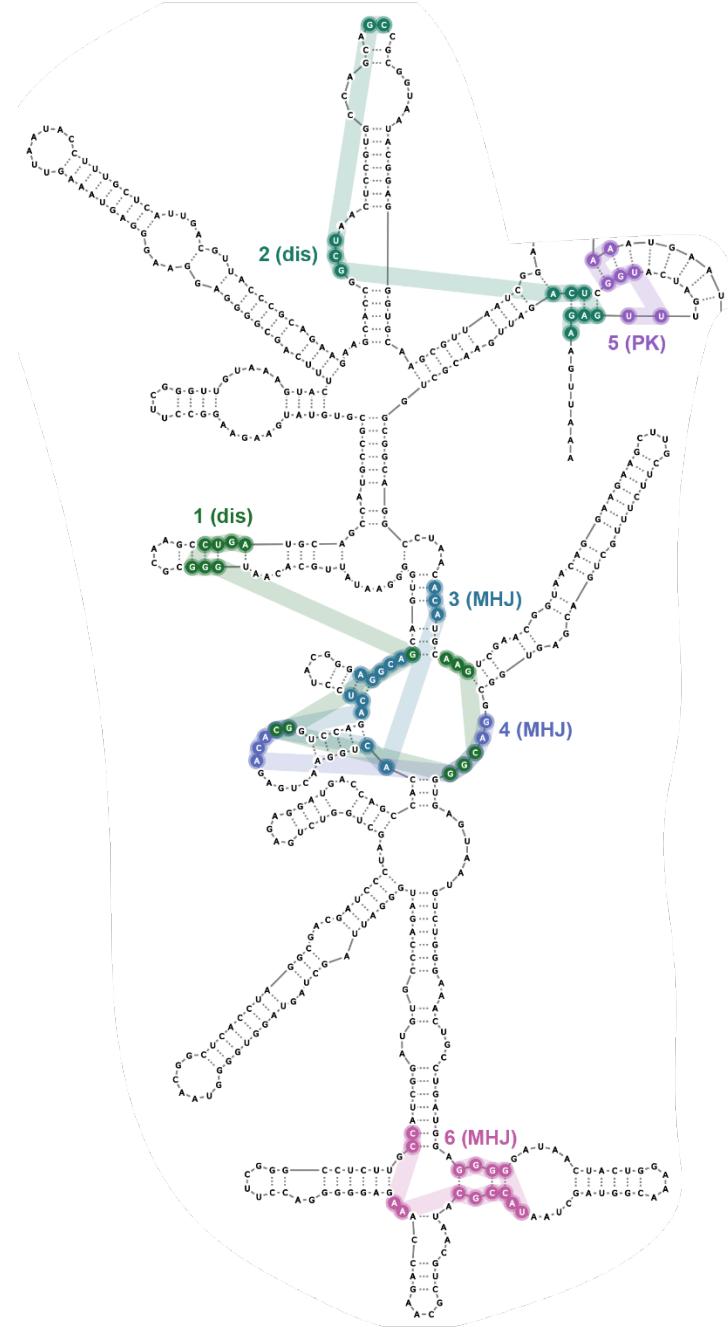
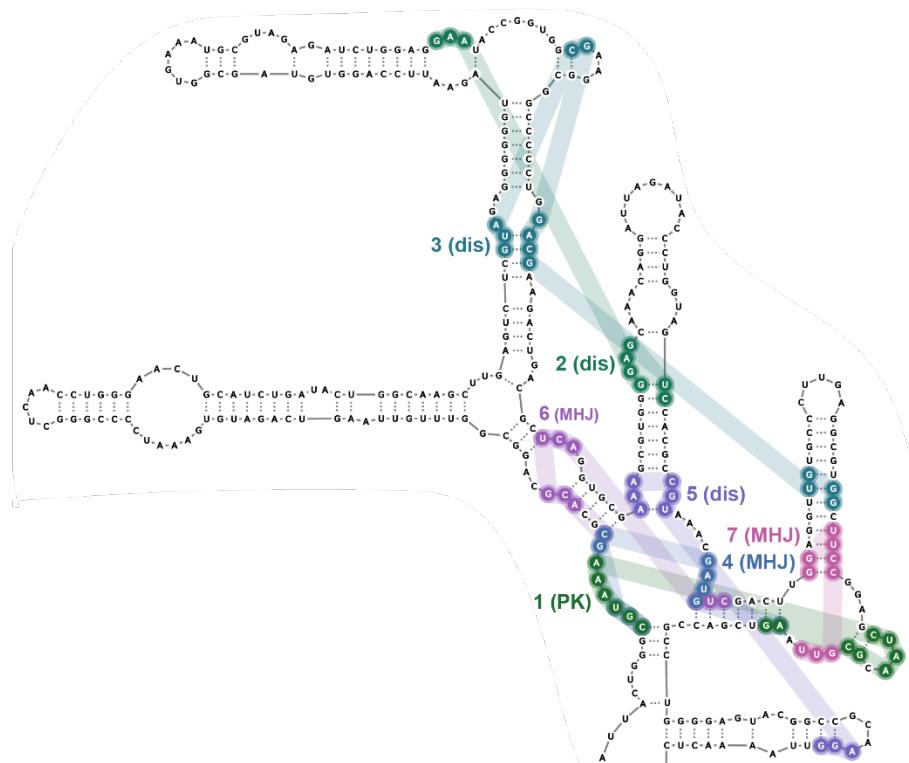


Figure B.4. Secondary structural motifs that form pockets in our curated non-redundant RNA-ligand library, featuring ligands with high QED values. (5 panels, above). Pockets that overlap a known binding site of a drug-like ligand or an RNA-protein-residue contact in the structure are colored orange to yellow; the remaining pockets are considered novel and are colored green to pink. Structures are classified as: (simple) simple structure, (con loop) consecutive loops, (MHJ) multi-helix junction, (PK) pseudoknot, or (dis) distant long-range interactions. PDB codes for all structures are provided in the Supporting Tables (below).

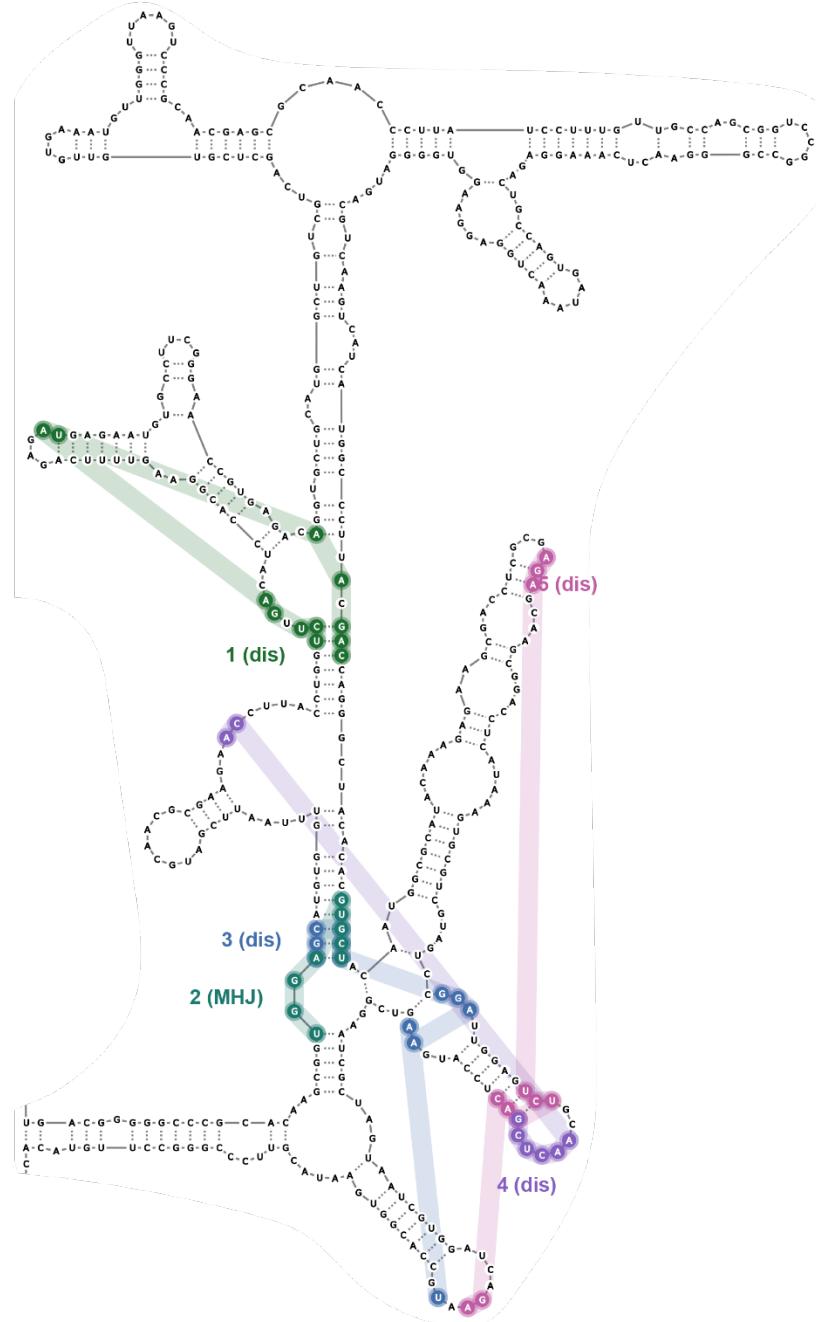
16S rRNA Domain 5' (7K00)



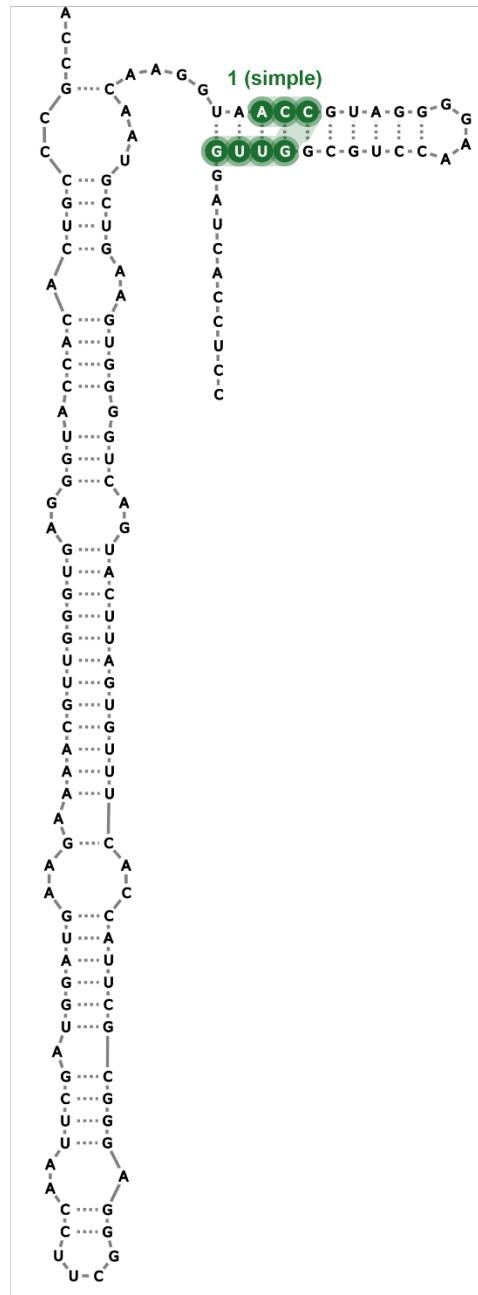
16S rRNA Domain C (7K00)



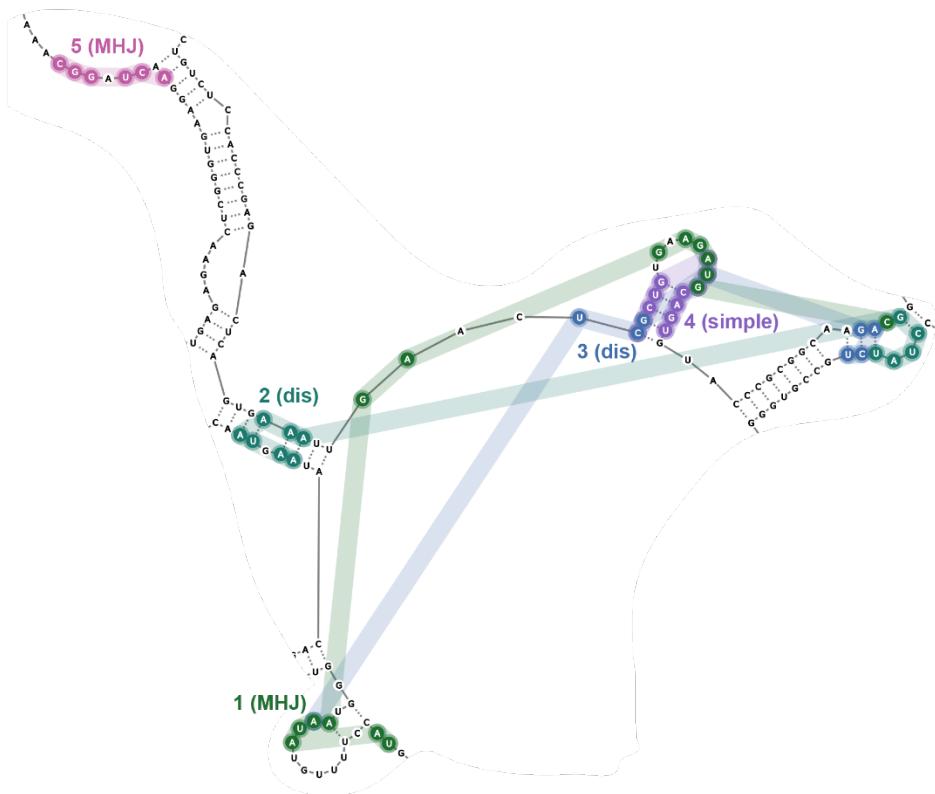
16S rRNA Domain 3'M (7K00)



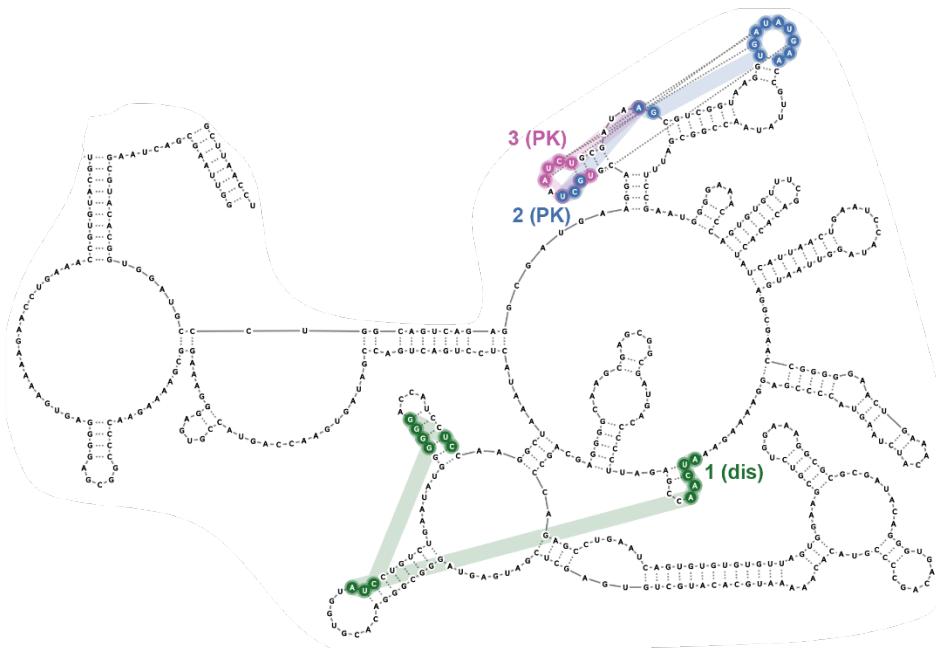
16S rRNA Domain 3'm (7K00)



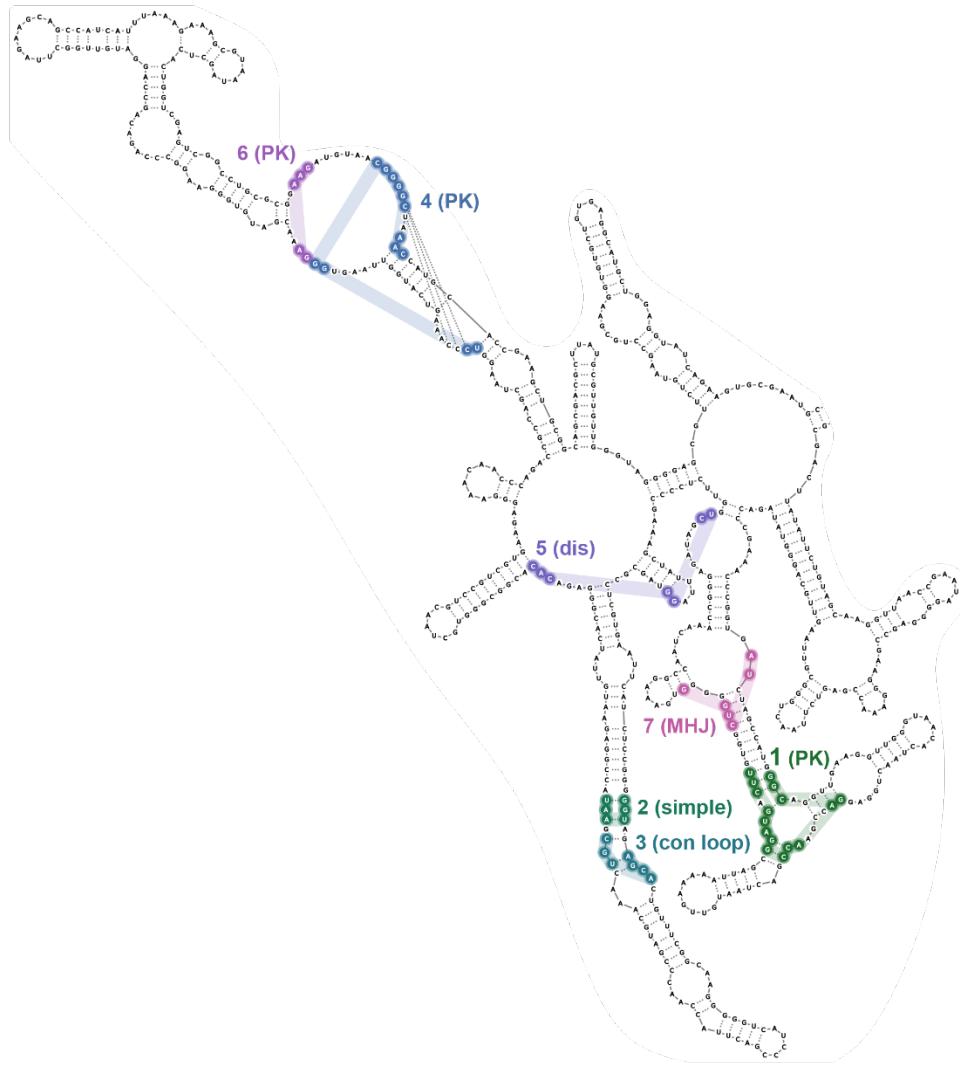
23S rRNA Domain 0 (7K00)



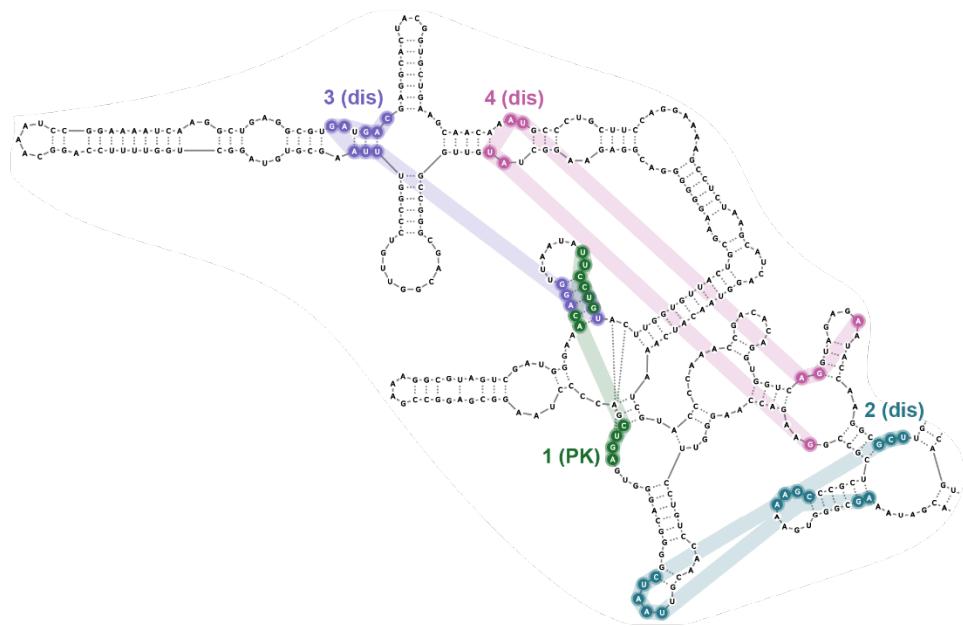
23S rRNA Domain 1 (7K00)



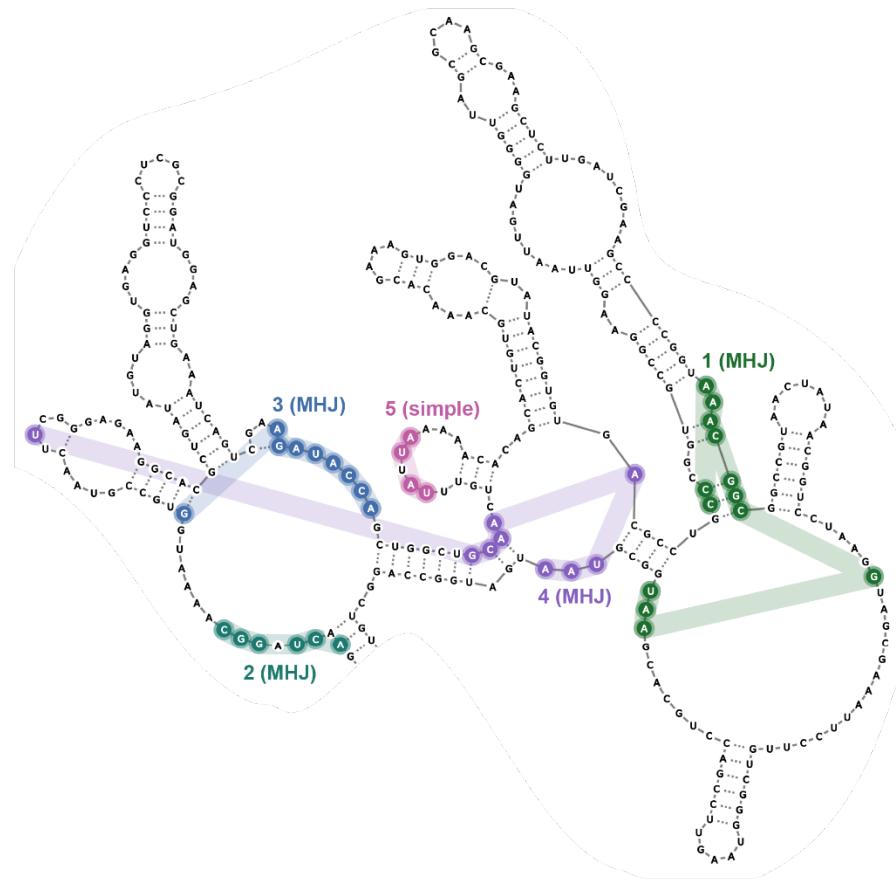
23S rRNA Domain 2 (7K00)



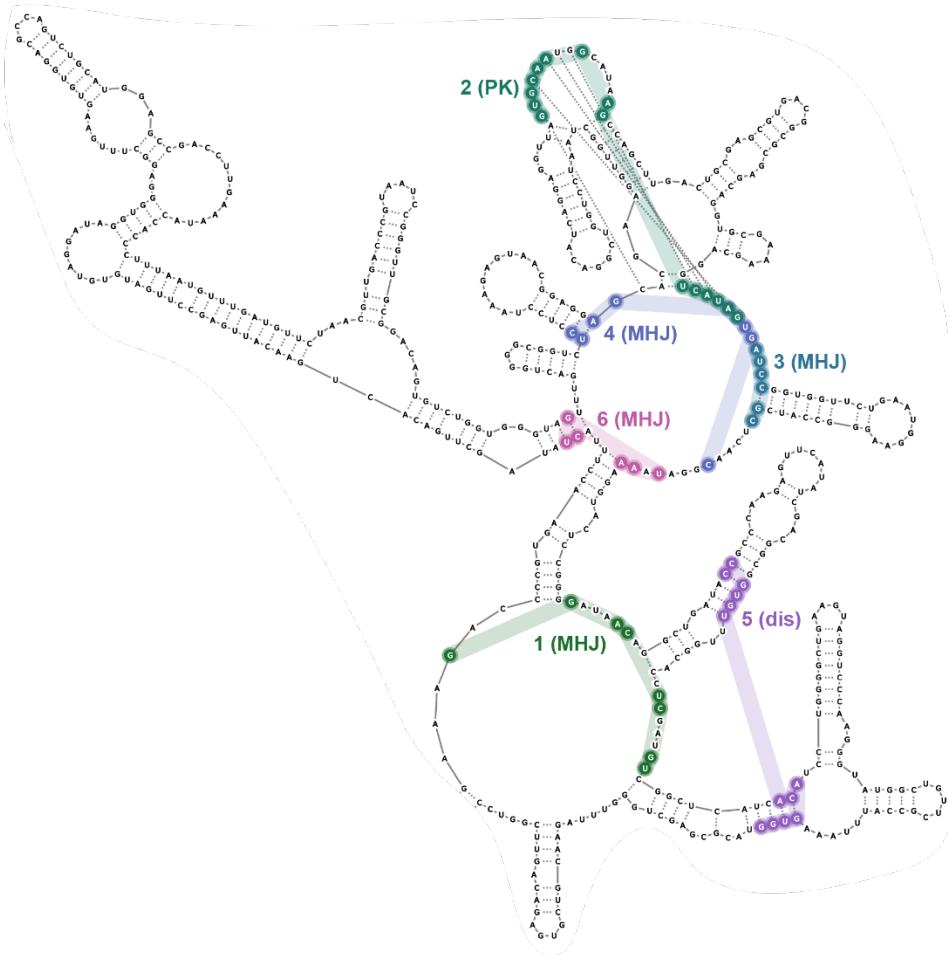
23S rRNA Domain 3 (7K00)



23S rRNA Domain 4 (7K00)



23S rRNA Domain 5 (7K00)



23S rRNA Domain 6 (7K00)



Figure B.5. Secondary structural motifs that form pockets in each domain of the *E. coli* 16S and 23S rRNAs. (11 panels, above). The rRNA reference structure was solved at 1.98 Å using cryo-EM (PDB: 7K00) (77). Nucleotides that contact the same pocket are annotated with a number and matching color. Structures are classified as: (simple) simple structure, (con loop) consecutive loops, (MHJ) multi-helix junction, (PK) pseudoknot, or (dis) distant long-range interactions.

Supplementary Tables

Table B.1. Detection of ligand binding pockets in paired apo and holo RNA structures.

name	QED	state	PDB code	reference	resolution (Å)	known binding site
linezolid	0.89	apo	7k00	(77)	1.98	detected
		holo	3cpw	(82)	2.70	detected
FMN	0.72	apo	6wjr	(161)	2.70	-
		holo	5kx9	(68)	2.90	detected
adenine	0.53	apo	5e54	(88)	2.30	detected
		holo	5swe	(88)	3.00	detected
pre-Q1	0.46	apo	6vuh	(162)	2.00	detected
		holo	6vui	(162)	2.98	detected
lysine	0.43	apo	3d0x	(86)	2.95	detected
		holo	3d0u	(86)	2.80	detected
TPP	0.35	apo	8f4o	(163)	3.10	detected
		holo	2gdi	(67)	2.05	detected
SAM-I	0.34	apo	3iqp	(164)	2.90	detected
		holo	3iqn	(164)	2.70	detected
glmS	0.29	apo	2gcs	(165)	2.10	detected
		holo	2z74	(166)	2.20	detected
self-alkylating epoxide	0.25	apo	6xjz	(146)	2.49	detected
		holo	6xjq	(146)	1.71	detected
PRPP	0.23	apo	6dnr	(167)	2.90	detected
		holo	6ck5	(168)	2.49	-

Table B.2. Pocket characteristics for small RNA library.

PDB code	pocket	structure type	structure class	occupied
1F27	1	local	PK	biotin
1LVJ	1	local	simple	acetyl promazine
1Q8N	1	local	con. loops	malachite green
1YKV	1	local	PK	ethanoanthracene
2GDI	1	local	MHJ	TPP
2KTZ	1	local	simple	Isis-11
2QWY	1	local	PK	SAM
3D0U	1	local	MHJ	lysine
3E5C	1	local	con. loops	Sr
3E5C	2	local	MHJ	SAM
3NPQ	1	local	PK	SAH
3NPQ	2	local	PK	SAH
3Q50	1	local	PK	PreQ1
3SKI	1	local	MHJ	2'-deoxy-guanosine
3SKI	2	local	MHJ	-
3SKI	3	local	MHJ	-
4B5R	1	local	PK	Ba
4B5R	2	local	MHJ	-
4B5R	3	local	MHJ	SAM
4JF2	1	local	PK	Cs
4JF2	2	local	PK	PreQ1
4LX5	1	local	MHJ	PPDA
4RZD	1	local	PK	PreQ1
5BTP	1	local	PK	ZMP
5BTP	2	local	PK	RNA (dimer)
5KPY	1	local	MHJ	5-hydroxy-L-tryptophan
5KX9	1	local	MHJ	ribocil-D
5KX9	2	local	PK	ribocil-D
5OB3	1	local	g-quadruplex	DFHBI
6FZ0	1	local	PK	SAM
6FZ0	2	local	PK	SAM
6GZR	1	local	MHJ	5-TAMRA
6LAS	1	local	MHJ	SAM
6UBU	1	local	MHJ	guanine
6UP0	1	local	g-quadruplex	YO3-biotin
6VA4	1	local	simple	MIP
6XB7	1	local	con. loops	DMA-135
6XJQ	1	local	con. loops	2,3-disubstituted epoxide
6XJQ	2	local	con. loops	protein
6YL5	1	local	PK	SAH
7DWH	1	local	con. loops	SAM
7ELR	1	local	con. loops	xanthine

7EOH	1	local	PK	HBC
7FJ0	1	local	simple	KG022
7KVT	1	local	MHJ	DFHBI-1T
7OAW	1	local	g-quadruplex	DMHBI+
8D2B	1	local	con. loops	TAL2
8EYU	1	local	g-quadruplex	DFAME
8HB3	1	local	PK	NNR
8HB3	2	local	PK	NNR

Table B.3. Pocket characteristics for the *E. coli* 16S rRNA. The rRNA reference structure was solved at 1.98 Å via cyro-EM (PDB: 7K00) (77).

domains	pocket	type	structure class	occupied
1	1	long-range	long-range	-
	2	long-range	long-range	-
	3	local	MHJ	-
	4	local	MHJ	-
	5	local	PK	-
	6	local	MHJ	Mg
	7	local	con. loops	Mg
	8	local	PK	-
	9	local	simple	-
	10	long-range	long-range	-
	11	local	MHJ	Mg
	12	long-range	long-range	-
2	1	local	PK	-
	2	long-range	long-range	-
	3	long-range	long-range	-
	4	local	MHJ	-
	5	long-range	long-range	-
	6	local	MHJ	-
	7	local	MHJ	-
	8	local	MHJ	-
	9	local	MHJ	protein
	10	local	simple	-
3	1	long-range	long-range	-
	2	local	MHJ	-
	3	long-range	long-range	-
	4	long-range	long-range	-
	5	long-range	long-range	-
	6	long-range	long-range	protein
	7	local	MHJ	Mg
	8	long-range	long-range	-
	9	long-range	long-range	-
	10	long-range	long-range	-
	11	long-range	long-range	-
	12	local	MHJ	Mg
4	1	local	simple	

Table B.4. Pocket characteristics for the *E. coli* 23S rRNA. The rRNA reference structure was solved at 1.98 Å via cyro-EM (PDB: 7K00) (77).

domain	pocket	type	structure class	occupied
0	1	local	MHJ	RNA
	2	long-range	long-range	protein
	3	long-range	long-range	-
	4	local	simple	protein
	5	local	MHJ	RNA
	6	long-range	long-range	protein
	7	long-range	long-range	RNA
1	1	long-range	long-range	-
	2	local	PK	-
	3	local	PK	-
	4	local	MHJ	protein
	5	local	simple	RNA
2	1	local	MHJ	-
	2	local	simple	RNA
	3	local	Con. Loops	-
	4	local	PK	-
	5	long-range	long-range	RNA
	6	local	PK	-
	7	local	MHJ	-
	8	local	simple	RNA
3	1	local	PK	-
	2	long-range	long-range	-
	3	long-range	long-range	-
	4	long-range	long-range	-
	5	local	MHJ	-
	6	local	MHJ	-
	7	local	MHJ	-
4	1	local	MHJ	-
	2	local	MHJ	-
	3	local	MHJ	-
	4	local	MHJ	-
	5	local	simple	RNA
5	1	local	MHJ	linezolid
	2	local	PK	-
	3	local	MHJ	protein
	4	local	MHJ	-
	5	long-range	long-range	RNA
	6	local	MHJ	RNA
	7	long-range	MHJ	RNA
	8	local	MHJ	-
6	1	local	MHJ	-

2	long-range	long-range	-
---	------------	------------	---

Table B.5. Pocket characteristics for the group II intron. The group II intron reference structure was solved at 3.8 Å via cryo-EM (PDB: 5G2X) (90).

pocket	type	structure class	occupied
1	local	MHJ	-
2	local	MHJ	-
3	long-range	long-range	-
4	local	MHJ	-
5	local	MHJ	-
6	long-range	long-range	-
7	local	MHJ	-
8	long-range	long-range	-
9	local	simple	-
10	local	simple	-
11	long-range	long-range	-

APPENDIX C: FRAG-MAP PROTOCOL

Introduction

Small-molecule ligands can engage RNA structure in pockets or cavities formed by RNA secondary and tertiary structure. Frag-MaP is an experimental technique to identify ligand binding sites in RNA with nucleotide precision. Frag-Map experiments use fully functionalized fragment probes that contain a small-molecule fragment, a diazirene crosslinking group, and a terminal alkyne. The fragment engages RNA binding sites, the diazirene group crosslinks the probe to the RNA binding site, and the alkyne is used to enrich cross-linked RNA. The location ligand binding site is encoded in cDNA by mutational profiling (MaP). A sequencing library is generated from the cDNA and sequenced using next-generation sequencing. The sequencing data is analyzed using the Frag-Mapper analysis module in RNavigate (72) to identify Frag-Map sites (nucleotides with significantly elevated mutation rates induced by fragment crosslinking).

The following protocol is written to analyze *B. subtilis* total RNA. However, Frag-MaP has been successfully performed in several cell lines, including mammalian cell lines (e.g., T47D). The following protocol used random nonamer primers during the reverse transcription step, technically making the experiment transcriptome-wide. In practice, we have had difficulty getting sufficient read coverage for RNAs that are not highly abundant using a random priming approach. It is possible that mRNA can be successfully analyzed using a higher input of mRNA (~100 µg), but we have not validated this approach. However, our lab has successfully identified mRNA-ligand interactions using gene-specific reverse transcription primers designed to target mRNAs of interest.

Probe cells and extract RNA

Probe *B. subtilis* cells with fully functionalized fragment probes and extract ligand-crosslinked total RNA for the Frag-MaP experiment.

NOTE: All experiments should start fresh by streaking *B. subtilis* on plates.

- *B. subtilis* does not grow well from glycerol stocks or passaged cell pellets.
- *B. subtilis* does not store well at 4 °C because it enters a spore form.

Day 1: Plate cells

Materials:

- *B. subtilis* str. 168 glycerol stock

- LB Agar plate

Protocol:

1. Streak a plate with *B. subtilis* str. 168 to form single colonies.
2. Incubate the plate at 30 °C overnight.

Day 2: Overnight culture

Materials:

- LB media (5 mL)
- Culture tube

Protocol:

1. Inoculate 5 mL LB media with a single colony from plate.
2. Grow cells overnight at 30 °C, 225 rpm.

Day 3: Probe cells and extract RNA (~10 hours)

Materials:

- LB media (50 mL)
- PBS buffer (100 mL)
- 50 mM ligand stock in DMSO (20 uL per sample)
- Sterile 6-well plastic plate (1 well per sample)
- Lysozyme (60 mg)
- 1M Tris pH 7 (180 uL)
- 0.5 M EDTA (120 uL)
- Chloroform
- 100% ethanol
- Qiogen RNAeasy Plus Universal Mini kit

Protocol:

1. Inoculate 60 mL LB media with 3 mL of cultured cells and grow cells to 0.50 OD₆₀₀.
NOTE: Expect cell growth to take ~2.5 hours.
 - Record: Time point, OD₆₀₀
2. Buffer exchange cells.
 - Gently pellet cells at 3200g for 10 min at 4 °C.
 - Carefully remove supernatant.
 - Wash cell pellet with 50 mL PBS buffer, repeat centrifugation, and carefully remove supernatant.
 - Resuspend cells in 50 mL PBS buffer.
3. Probe cells with functionalized ligands.
 - Combine 4,980 uL of cultured media with 20 uL of photo-reactive ligand (50 mM stock in DMSO) and pipette mix. Avoid prolonged light exposure.

NOTE: final [ligand] is 200 uM. 250X dilution.

- Incubate ligand-treated cells for 30 minutes at 30 °C and 225 rpm.

NOTE: Keep the solution in the dark.

4. Cross-link functionalized ligands to RNA.
 - Transfer 5 mL cell solution to a sterile 6-well plastic plate.
 - Uncover the plate and place it on ice.
 - Irradiate with 365 nm light source and cross-link with 3 J/cm².
5. Create a cell pellet.
 - Centrifuge to pellet the cells. 8,000xg for 10 min. at 4 °C.
 - Discard supernatant.
6. Prepare fresh lysis buffer.
 - (30 mM Tris pH 7, 10 mM EDTA, 10 mg/mL lysozyme):

Reagent	Amount
lysozyme	50 mg
1M Tris pH 7	150 uL
0.5M EDTA	100 uL
RNase inhibitor murine	50 uL
NF H ₂ O	4700 uL
TOTAL	5,000 uL

7. Lysis cell pellet with lysozyme buffer.
 - Add 250 uL of buffer lysis to each cell pellet.
 - Incubate for 30 min. at RT.
 8. Extract and purify total RNA from cells.
 - Use 1000 uL of Tri-reagent
 - Mix aggressively to aid in lysis
 - Lyse for 15 minutes
 - Add 250 uL of chloroform
 - Vortex and incubate at RT for 3-5 minutes
 - Centrifuge at 15000xg for 15 minutes at 4 C
- NOTE: Keep sample on ice until aqueous layer is recovered.
- Add 100% ethanol to each recovered aqueous layer in equal parts (1:1)
 - Follow Direct-zol RNA Miniprep Plus Kit instructions.
 - Elute from column with 50 uL of NFW

NOTE: Yields of ~50-60 ug total RNA have been achieved per 5 mL culture probed.

9. Check RNA purity.

- Measure absorbance ratio (A260/280 and A260/230) on nanodrop.
- Record: Culture, Sample ID, Yield (μg), [RNA] (ng/μL), A260/280, A260/230

Storage:

- RNA can be stored overnight at 4 °C for ~ 1 week.
- RNA can be flash-frozen in aliquots and stored at -80 °C for ~1 year.

Sample Details:

RNA provided from Frag-MaP Cell Probing experiment using *B. subtilis* cells and fully functionalized fragment probes at a concentration of 200 μM.

- Record: Sample ID, Replicate, Ligand, [RNA] (ng/μL), i7 index, i5 index

Partial Fragmentation of total RNA

Fragment RNA to around 200-250 nt average length.

Materials:

- Total RNA (13.5 μg)
- [NEBNext® Magnesium RNA Fragmentation Module](#)

Protocol:

1. Mix the following components in a sterile PCR tube:

Component	Volume
Purified total RNA (~13.5 μg)	27 μL
RNA Fragmentation Buffer (10X)	3 μL
Total	30 μL

2. Incubate in a preheated thermal cycler for 2 minutes at 94°C.
3. Transfer tube to ice immediately.
4. Add 3 μl 10X RNA Fragmentation Stop Solution.
5. Cleanup with Monarch RNA cleanup kit.

More Information:

[NEBNext® Magnesium RNA Fragmentation Module Protocol](#)

Monarch RNA Cleanup kit

Remove buffer solution and very small fragments from total RNA.

Materials:

- [NEB Monarch® Spin RNA Cleanup Kit \(10 μg\)](#)

Protocol:

1. Follow manufacturer [protocol](#).
2. Elute in 12.5 µL

More Information:

[NEB Monarch® Spin RNA Cleanup Kit Protocol](#)

Click chemistry with alkyne-RNA & azide-oligo primer using 50% DMSO denaturant

This protocol is intended to add steric bulk to increase MaP-RT mutation efficiency at fully functionalized fragment crosslink sites.

Materials:

- Total RNA-alkyne (~ 1 µg/µL)
- [Bangs Labs Click Beads, Magnefy™ Azide, Medium - 1.0µm \(50 µmol/mL\)](#)
- DMSO (anhydrous)
- Ascorbic Acid (10 mM)
- Cu(II): TBTA (10 mM)

Protocol:

1. Prepare a fresh 10 mM Ascorbic acid stock 20X:
 - Combine reagents.
 - Sparge solution with N2 and use immediately.

Reagents	Amount
Ascorbic acid	9.0 mg
NF Water	5.0 mL
TOTAL	5.0 mL

2. Dilute RNA:

- For each sample, add RNA-alkyne and water to a strip of PCR tubes.

Reagents	[Reagent]	Final [RNA]	Amount
RNA-alkyne – 10X	1000 ng/µL	~ 0.4 µM	12 µL
Nuclease free H2O	-	-	- µL
Subtotal	-	-	12 µL

3. Heat denature RNA:

- Incubate RNA at 95 °C for 2 min.
- Snap cool on ice for 2 min.

4. Add Magnefy™ Azide beads to tubes:

Reagents	[Reagent]	[Final]	Amount
RNA containing soln.	-	-	12 µL
Magnify™ Azide beads–5X	50 µmol/mL	0.5 µM	10 µL
Subtotal	-	-	22 µL

NOTE: Batch samples can be kept on ice after HEPES addition.

5. Denature RNA by adding DMSO:

- Add DMSO to each sample.
- Briefly vortex.

Reagents	[Reagent]	[Final]	Amount
RNA containing soln.	-	-	22 µL
DMSO	-	-	25 µL
Subtotal	-	-	47 µL

6. Start click reaction:

- Add ascorbic acid and Cu(II)/TBTA mix to samples.

Reagents	[Reagent]	[Final]	Amount
RNA containing soln.	-	-	47.0 µL
Ascorbic Acid – 20X	10 mM	0.5 mM	2.5 µL
Cu(II): TBTA – 41.7X	10 mM	240 uM	1.2 µL
Total	-	-	50.7 µL

NOTE: Initiate reaction within 5 minutes of adding DMSO (RNA stays denatured for ~30 minutes).

7. Incubate reaction.

- Cover reaction.
- Incubate at 40 °C for ~20 minutes.

NOTE: No further conversion after 1 h and more RNA degradation occurs.

8. Quench reaction.

- Add 1.0 µL of 0.5 M EDTA to each sample (50 µL).

RNA-Alkyne enrichment

Enrich crosslinked RNA via azide-magbead pull-down.

Materials:

- 0.5M Tris pH 7.5
- 2M NaCl
- 10% Tween-20
- 0.5M EDTA

Protocol:

1. Prepare Binding/Wash Buffer:

Component	[reagent]	[Final]	Amount (μ L)
NaCl	2 M	1 M	12,500
Tris pH 7.5	1 M	10mM	250
EDTA	0.5 M	1 mM	50
TWEEN-20	10%	0.01%	25
NFW	-	-	12,175
TOTAL	-	-	25,000

2. Pull-down beads:

- Immobilize beads with a magnetic rack.
- Remove and discard supernatant.

3. Wash beads:

- Resuspend beads in 200 μ L of Wash Buffer (1X).
- Immobilize beads with a magnetic rack.
- Remove and discard supernatant.
- Repeat (x3)

4. Resuspend beads:

- Resuspend beads in 7.5 μ L of NFW

Mutational profiling RT (DMS-optimized)

This protocol is based on Smola et.al., 2015. It has been optimized for the synthesis of long cDNA from DMS-modified RNA.

Materials:

- 0.5 M Tris pH 8.0
- 0.75 M KCl
- M DTT
- RNA (1 ng to 5 μ g of total RNA)
- μ M RT primer (50 to 250 ng random primer)
- 10 mM dNTP mix

- 5 M betaine
- 40 mM MnCl₂
- SuperScript II

Protocol:

1. Create fresh 10X NTP minus:

Reagent	Amount
1 M Tris ph 8.0	20 uL
2 M KCl	15 uL
1 M DTT	4 uL
NF Water	1 uL
TOTAL	40 uL

2. For each RNA sample combine:

Reagent	Amount
RNA (~10 ng)	7.5 uL
nonamer (~100 ng)	0.5 uL
10 mM dNTPs	2 uL
TOTAL	10 µL

3. Denature and anneal the ribosome by incubating at 90°C for 5 min, then cool on ice.

4. Create the following master mix:

- (*multiply by number of samples+1*)

Reagent	Amount
10X NTP minus	2 uL
5 M Betaine	4 uL
40 mM MnCl ₂	3 uL
TOTAL	9 uL

5. Add 9 µL of master mix to each sample.

6. Incubate at 25°C for 2 min.

7. Add 1 µL of SuperScript II.

8. Set Thermocycler to this program:

STEP	TEMP	TIME	CYCLES
Incubate	25 °C	10 (min)	1
Incubate	40 °C	90 (min)	1
Extension cycle:			
	50 °C	2 (min)	10
	42 °C	2 (min)	10
Deactivate	70 °C	10 (min)	1
HOLD	4 °C	Hold	1

Notes:

More Information:

[First-Strand cDNA Synthesis Using SuperScript II RT](#)

Second Strand Synthesis

From first-strand synthesis product, a mix of enzymes polymerizes a second cDNA from the first cDNA of a DNA-RNA hybrid, then digests away the RNA.

Materials:

- NEBNext second-strand synthesis enzyme mix
- NEBNext second strand synthesis 10x buffer
- Fresh 80% EtOH

Protocol:

NOTE: Always keep 2nd Strand Enzyme cold (below 16°C)

1. Pull-down cDNA product on beads:
 - Immobilize beads with a magnetic rack.
 - Remove and discard supernatant.
2. Prepare reaction:
 - Resuspend cDNA on ice.
 - Pipette mix.

Reagent	Amount
sscDNA	(on-bead)
nucleotide-free water	32 µL
10X 2nd Strand Reaction Buffer	4 µL
2nd Strand Enzyme Mix	2 µL
TOTAL	40 µL

3. Incubate at 16°C for 60 min.

cDNA purification - AMPure beads

Material:

- AMPure beads (equilibrate to RT)
- Fresh 80% ethanol
- Magnetic holder

Bead Ratio:

1.8X

Protocol:

1. Beads added to reaction mixture (1.8X)
2. Incubate at room temp. for 5 min.
3. Place in magnetic holder for 5 min.
4. Remove most of the supernatant
5. Wash beads with 80% ethanol (30 seconds) twice.
6. Add 20 µL of water, resuspend the pellet, and incubate for 5 minutes.
7. Remove and keep supernatant.

More Information:

[NGS Beads for DNA Size Selection](#)

Qubit

Quantify the dsDNA product from second-strand synthesis via Qubit HS DNA assay.

Materials:

- Qubit tubes
- Working solution
- Standards 1 & 2

Protocol:

1. Create standards:
 - Add 190 µL of working solution into a Qubit tube
 - Add 10 µL of standard
 - Equilibrate to RT
2. Create samples:
 - Add 198 µL of working solution into a Qubit tube
 - Add 2 µL of cDNA

- Equilibrate to RT
3. Incubate for 2 minutes.
 4. Reading concentrations using Qubit fluorimeter HS dsDNA assay.

Results:

- Record: Sample ID, [dsDNA] ng/ μ L

DNA Library Prep (Adapter Ligation) - NEBNEXT Ultra II

Materials:

- dsDNA (500 pg - 1 μ g)
- (green) NEBNext Ultra II End Prep Enzyme Mix
- (green) NEBNext Ultra II End Prep Reaction Buffer
- (red) NEBNext Ultra II Ligation Master Mix
- (red) NEBNext Ligation Enhancer
- (blue) NEBNext Ultra II Q5 Master Mix
- NEBNext Oligo Kit
- 80% Ethanol

Protocol:

1. Prepare Ends.
 - Add the following components into a PCR tube:

Components	Volume
dsDNA	25 μ L
End Prep Reaction Buffer	3.5 μ L
End Prep Enzyme Mix	1.5 μ L
TOTAL	30 μL

- Mix thoroughly and spin down.
- Incubate in thermal cycler with lid set to > 75 °C:

Step	Time (min.)	Temperature (°C)
Prep ends	30	20
Denature	30	65
End	HOLD	4

2. Adapter Ligation. (Adds ~60 nt)
 - Dilute adaptors based on input DNA quantity (from NEBNext Oligo Kit):

Input DNA	Adaptor Dilution	Working [Adapter]
	No Dilution	15 M
5- 100 ng	1:10	1.5 μM
0.5 - 5 ng	1:25 (recommended)	0.6 μM

- Add the following components directly into the End Prep Reaction Mixture:
NOTE: Do NOT premix Adapter with Master Mix or Ligation Enhancer.

COMPONENT	VOLUME
End Prep Reaction Mixture	30 μl
(red) NEBNext Adaptor (diluted)	1.25 μl
(red) NEBNext Ultra II Ligation Master Mix	15 μl
(red) NEBNext Ligation Enhancer	0.5 μl
TOTAL	47.75 μL

- Mix thoroughly and spin down.
- Incubate at 20 °C for 15 minutes with the lid heater off.
- Add USER enzyme.

COMPONENT	VOLUME
Ligation reaction mixture	47.75 μL
(red) USER enzyme	1.50 μl
TOTAL	49.25 μL

- Mix thoroughly.
- Incubate at 37 °C for 15 minutes with the lid heater set to > 47 °C

3. Cleanup of Adaptor-ligated DNA - SPRI beads

Bead Ratio: 1.4x (or lower)

- Beads added to reaction mixture (1.4X)
- Incubate at room temp. for 5 min.
- Place in magnetic holder for 5 min.
- Remove most of the supernatant
- Wash beads with 80% ethanol (30 seconds) twice.
- Add 17 μL of water, resuspend the pellet, and incubate for 5 minutes.
- Remove and keep supernatant.

4. PCR Amplification (Adds ~60 nt)

- Add the following components to a PCR strip:

COMPONENT	VOLUME
Adaptor Ligated DNA Fragments	15 µl
(blue) NEBNext Ultra II Q5 Master Mix	25 µl
Index Primer/i7 Primer	5 µl
Universal PCR Primer/i5 Primer	5 µl
TOTAL	50 µl

- Mix thoroughly and spin down.
- Choose number of PCR cycles using the table below:

INPUT DNA	# OF CYCLES (YIELD ~100 ng)
100 ng	3
50 ng	3–4
10 ng	6–7
5 ng	7–8
1 ng	9–10
0.5 ng	10–11+

- Perform PCR amplification using the following PCR cycling conditions:

STEP	PCR Step	TEMP	TIME	CYCLES
Denature		98 °C	30 (s)	1
PCR cycles:				
	Denaturation	98 °C	10 (s)	3-15 (15 rec.)
	Annealing/	65 °C	75 (s)	3-15 (15 rec.)
	Extension			
Final Extension		65 °C	300 (s)	1
HOLD		4 °C	Hold	1

5. Cleanup of PCR Reaction - SPRI beads

Bead Ratio: 0.95x

- Beads added to reaction mixture (0.95X)
- Incubate at room temp. for 5 min.
- Place in magnetic holder for 5 min.

- Remove most of the supernatant
 - Wash beads with 80% ethanol (30 seconds) twice.
 - Add 20 μ L of water, resuspend pellet, and incubate for 5 minutes.
 - Remove and keep supernatant.
6. Repeat step 5

Bead Ratio: 0.95x

Notes:

- Length of DNA = Insert (~150) + adaptor (60 nt) + primers (60 nt) = 270
- Frag-MaP usually requires 15 PCR cycles

Qubit

Quantify the dscDNA product from second strand synthesis via Qubit HS DNA assay.

Materials:

- Qubit tubes
- Working solution
- Standards 1 & 2

Protocol:

5. Create standards:
 - Add 190 μ L of working solution into Qubit tube
 - Add 10 μ L of standard
 - Equilibrate to RT
6. Create samples:
 - Add 198 μ L of working solution into Qubit tube
 - Add 2 μ L of cDNA
 - Equilibrate to RT
7. Incubate for 2 minutes.
8. Reading concentrations using Qubit fluorimeter HS dsDNA assay.

Results:

- Record: Sample ID, [dscDNA] ng/ μ L

Agilent Tape Station

Determine the average length of sequencing library.

Materials:

- High sensitivity D1000 DNA sample buffer
- High sensitivity D1000 DNA ladder
- High sensitivity D1000 DNA screentape
- Tape station tubes and caps

Specifications:

Analytical specifications	HS D1000 ScreenTape assay
Sizing range	35–1000bp
Sizing accuracy	±10 %
Quantitative range	10–1000 pg/µL

Protocol:

1. Allow High sensitivity D1000 Reagents to equilibrate to room temperature.
2. Prepare ladder (position: A1)

COMPONENT	VOLUME
D1000 Sample buffer	2 µl
Ladder	2 µl
TOTAL	4 µl

3. Prepare samples (position: A2-B8)
 - Dilute samples to < 1 ng/µL in NFW

COMPONENT	VOLUME
HS D1000 Sample buffer	2 µl
Sample	2 µl
TOTAL	4 µl

4. Apply caps to tube strips and/or foil seals to 96-well sample plates.
5. Vortex sample and spin down.
6. Load Screetape and tubes/plate into the tape station instrument.

Results:

- Record: Sample ID, Avg. bp length, % of total

Pool and dilute library

Dilute sample to the same molarity and pool for sequencing.

Materials:

- Illumina-mix-calculator.xlsx
- Sample_Sheet.xlsx

Target library concentration:

300 pM

Protocol:

1. Input data into [illumina-mix-calculator.xlsx](#).
 - Avg. bp length (Bioanalyzer)
 - ng/ μ L (Qubit)
 - % Non-dimer (Bioanalyzer)
 - Target library concentration (nM)

NOTE: Recommended minimum volume: 3 μ L
2. Pool sample library using the volumes calculated by [illumina-mix-calculator.xlsx](#).
 - Record: Sample ID, [non-dimer] (nM), Amount to pool (μ L)
3. Dilute the sequencing library to 2 nM using Qubit measurements.
 - Best practice: 10 μ L sample + 190 μ L Qubit working solution.
 - Store sequencing libraries at 2 nM.
 - Record storage concentration of pooled library: ng/ μ L, pM
4. Dilute 2 nM sequencing library to target sequencing concentration using Qubit measurements.
 - Best practice: 10 μ L sample + 190 μ L Qubit working solution.
 - Prepare dilution fresh before each sequencing run.
 - Record diluted concentration of pooled library: ng/ μ L, pM

Results:

- Record: Library Name, ng/ μ L, pM

Sequence Frag-MaP library (NextSeq)

Sequence Frag-MaP library using Illumina NextSeq.

Materials:

- low-bind/low-retention microcentrifuge tubes
- NextSeq1000/2000 kit (P1 or P2):

Parameters:

1. Adaptor reads:

Adapter	Sequence
AdapterRead1	AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
AdapterRead2	AGATCGGAAGAGCGTCGTAGGGAAAGAGTGT

2. Dilute pooled library:

- Use RSB TWEEN 20 to dilute pooled library to 2 nM in a low-bind tube.
- Measure concentration using [Qubit HS dsDNA](#) (recommend sample volume $\geq 4 \mu\text{L}$)

3. Dilute to loading concentration:

- Concentrations for Onboard Denature and Dilute

Library	concentration	Loading 2 nM Library	2 nM PhiX	RSB with TWEEN-20
Frag-MaP	250-300 pM	9.5 μL	0.5 μL	62.7 μL

NOTE: 1-5 % PhiX is recommended to improve run metrics and troubleshooting.

More information:

[Illumina: library dilution protocol](#)

APPENDIX D: SUPPLEMENTAL INFORMATION FOR CHAPTER 4

Supplemental Tables

Table D.1. PDB codes for non-redundant Hariboss RNA-ligand complex library.

1aju	2au4	2tra	3ski	4q9r	5v3f	6e84	6up0	7k00	8d2b
1am0	2be0	2yie	3skz	4qk9	5v9z	6e8t	6v9b	7kuk	8d5l
1arj	2bee	3b4a	3slm	4qln	5vcf	6e8u	6v9d	7kum	8d5o
1eht	2esi	3b4b	3slq	4qvi	5vci	6fz0	6va2	7kun	8eyu
1ei2	2esj	3bnq	3suh	4ts2	5vj9	6gxr	6va3	7kuo	8eyv
1evv	2et3	3d2g	3sux	4w90	5vjb	6hag	6va4	7kup	8eyw
1f1t	2et4	3d2v	3td1	4xwf	5xi1	6hbt	6vui	7kvu	8f4o
1f27	2et8	3d2x	3tzr	4y1j	5z1h	6hc5	6wzr	7kvv	8fza
1fmn	2f4t	3dig	3v7e	4y1m	5z1i	6hmo	6wzs	7lne	8gxb
1fuf	2f4u	3dil	3wru	4yaz	5z71	6jbf	6xb7	7lnf	8gxc
1fyp	2g5k	3diq	4e8k	4yb0	5zei	6jbg	6xkn	7lng	8hb3
1i2y	2gdi	3dir	4e8q	4yb1	5zej	6iji	6xrq	7mkt	8hb8
1i9v	2ho7	3ds7	4e8v	4zc7	6az4	6lau	6yl5	7oa3	8hba
1j7t	2hom	3e5c	4f8u	4znp	6fbf	6lax	6ylb	7oaw	8i3z
1j8g	2hoo	3e5e	4far	5bjo	6c64	6laz	6ymi	7oax	8i43
1koc	2hop	3e5f	4fau	5bjp	6c65	6n5o	6ymj	7rwr	8i45
1kod	2juk	3egz	4faw	5btp	6c8e	6p2h	6ymk	7szu	8i46
1lc4	2kd4	3f2q	4gpx	5c45	6c8k	6pq7	6yml	7td7	8i7n
1lvj	2kgp	3f4g	4gpy	5dhb	6c8l	6q57	6ymm	7tdc	8r62
1mwl	2ktz	3f4h	4jf2	5fjc	6c8m	6qiv	7a3y	7tzr	8r63
1n7a	2ku0	3fo4	4k31	5fk1	6c8o	6t3n	7d12	7tzt	8r8p
1n7b	2kx8	3fo6	4k32	5fk4	6cb3	6tb7	7d7v	7tzu	8swg
1nta	2kxm	3gca	4kqy	5fk6	6cc1	6tf0	7d7w	7u0y	8swo
1ntb	2l1v	3ger	4kzd	5hbw	6cc3	6tf1	7d81	7u87	8sx5
1o15	2l94	3gx2	4l81	5kpy	6ck4	6tf2	7dwh	7u88	8sx6
1q8n	2lwk	3gx3	4lvv	5krg	6ck5	6tf3	7e9e	7u89	8sxl
1raw	2m4q	3iwn	4lvx	5kx9	6db8	6tfe	7edl	7u8a	8sy1
1tn2	2miy	3k0j	4lvy	5lwj	6dlr	6tfg	7edt	7u8b	8u5k
1uts	2mxs	3k1v	4lx5	5ob3	6dmc	6u6j	7elp	7wib	8u5p
1uud	2n0j	3mij	4lx6	5t83	6dn1	6u89	7eog	7wif	8u5t
1uui	2o3v	3muv	4nfo	5ued	6dn2	6u8f	7eok	7wii	8u5z
1xpf	2o3w	3mxh	4nya	5uee	6dn3	6u8u	7eol	7ych	8vaw
1ykv	2o3x	3npq	4nyb	5ueg	6e1s	6uc7	7eom	7yci	
1yls	2o3y	3q3z	4p20	5ux3	6e1u	6uc8	7eon	7zj4	
1yrj	2oe5	3rkf	4p5j	5v0h	6e1v	6uc9	7eo0	8cf2	
1zz5	2pwt	3s4p	4pdq	5v0o	6e1w	6uej	7eop	8d28	
2a04	2qwy	3sd3	4q9q	5v1l	6e81	6uet	7fhi	8d2a	

Table D.2. Description of all pocket characteristics generated by fpocketR.

Characteristic	Description
Parameters	fpocket parameters used to find pockets.
Name	Structure name generated from input file.
PDB	First four characters of the inputted tertiary structure (typically the PDB accession code).
State	Conformational state from the tertiary structure file.
Type	Known: pocket overlaps with ligand. Novel: pocket does not significantly overlap with ligand.
Filter	Pass: pocket score above the quality filter threshold. Fail: pocket is not visualized in figures.
Pocket	Pocket ID number.
Score	† Pocket score from fpocket.
Drug_score	† Pocket drug score from fpocket.
a-sphere	† Number of alpha spheres in the pocket.
SASA	† Solvent accessible surface area.
Volume	† Volume of pocket (\AA^3).
Hydrophobic_density	† Sum of all apolar alpha sphere neighbors divided by the # of apolar alpha spheres in the pocket.
Apolar_a-sphere_proportion	† Percentage, reflects the proportion of apolar alpha spheres among all alpha spheres in the pocket.
Hydrophobicity_score	† Residue-based hydrophobicity scale reported in Monera & al. (1995) <i>Prot Sci</i> , 1, 319-329.
Polarity_score	† Measure the hydrophilicity character of the pocket.
PocketNT	List of all RNA residues that contact alpha spheres from the pocket.
Pocket_NPR1	First normalized principal moment of inertia ratio (I_1/I_3) for the pocket.
Pocket_NPR2	Second normalized principal moment of inertia ratio (I_2/I_3) for the pocket.
Pocket_shape	Description of pocket shape based on NPR: rod-like , disc-like , sphere-like , or balanced .
Ligand_ID	Chemical ID of the ligand assigned from the PDB Chemical Component reference Dictionary.
Pocket_overlap	Proportion of alpha sphere in the pocket that are within 3 \AA of an atom in the ligand.
Ligand_overlap	Proportion of atoms in the ligand that are within 3 \AA of an alpha spheres from the pocket.
Center_criteria	Minimum Euclidean distance between the geometric centers of the pocket and ligand.
QED_score	Ligand drug-likeness score reported in Bickerton,G.R. (2012) <i>Nat Chem</i> , 4, 90-98.
Ligand_NPR1	NPR1 (I_1/I_3) for the ligand calculated using the conformation from the PDB (not optimized!).
Ligand_NPR2	NPR2 (I_2/I_3) for the ligand calculated using the conformation from the PDB (not optimized!).
Ligand_shape	Description of ligand shape based on NPR: rod-like , disc-like , sphere-like , or balanced .

† Characteristics calculated and documented by fpocket v4.0.3.

Table D.3. Output files used and generated by fpocketR.

Filename	Description
<PDB>clean_out/<ligand_id>_model.sdf	Analyzed ligand structure from the PDB.
<PDB>clean_out/pockets/pocket<#>_atm.pdb	RNA atoms that contact alpha spheres used to determine pocketNT.
<PDB>clean_out/pockets/pocket<#>_surf.obj	Pocket surface used to calculate NPR values.
<PDB>clean_out/pockets/pocket<#>_vert.pqr	Alpha spheres in pocket used to generate * _real_sphere.pdb.
<PDB>clean_out/<PDB>_out_pocket_characteristics.csv	Geometric and physicochemical properties of pockets.
<PDB>clean_out/<PDB>_2D.png	Figure of pockets in RNA secondary structure.
<PDB>clean_out/<PDB>_2D.svg	Editable figure of pockets in RNA secondary structure.
<PDB>clean_out/<PDB>_out_real_sphere.pdb	RNA structure with pockets. Alpha sphere radius in B factor column.
<PDB>clean_out/<PDB>_out_real_sphere.pse	RNA structure with properly scaled alpha spheres in pockets.
<PDB>clean_out/<PDB>_3D_<dpi>.png	Figure of pockets in RNA tertiary structure.

Table D.4. Output files used and generated by fpocketR multistate analysis.

Filename	Description
state_tracker.txt	Stores last state analyzed used to resume interrupted analysis.
<PDB>_all_states_out_pocket_characteristics.csv	Geometric and physicochemical properties of pockets in all states.
<PDB>_2D_pocket_density.png	Figure of pockets density from all states in RNA secondary structure.
<PDB>_2D_pocket_density.svg	Editable figure of pockets density from all states in RNA secondary structure.
<PDB>_all_states_out_real_sphere.pse	RNA structure of all states with pockets depicted as transparent surfaces.
<PDB>_all_states_3D_<dpi>.png	Figure of pockets in all states of RNA tertiary structure.

REFERENCES

1. P. A. Sharp, The Centrality of RNA. *Cell* [Preprint] (2009). [Accessed 17 October 2020].
2. T. R. Cech, J. A. Steitz, The noncoding RNA revolution - Trashing old rules to forge new ones. *Cell* **157**, 77–94 (2014).
3. I. Dunham, *et al.*, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
4. E. A. Doherty, J. A. Doudna, Ribozyme structures and mechanisms. *Annu Rev Biochem* **69**, 597–615 (2000).
5. C. A. McHugh, *et al.*, The Xist lncRNA interacts directly with SHARP to silence transcription through HDAC3. *Nature* **521**, 232–236 (2015).
6. C. Chu, *et al.*, Systematic discovery of Xist RNA binding proteins. *Cell* **161**, 404–416 (2015).
7. A. Serganov, D. J. Patel, Ribozymes, riboswitches and beyond: regulation of gene expression without proteins. *Nat Rev Genet* **8**, 776–790 (2007).
8. I. A. Roundtree, M. E. Evans, T. Pan, C. He, Dynamic RNA Modifications in Gene Expression Regulation. *Cell* [Preprint] (2017). [Accessed 30 January 2021].
9. K. D. Warner, C. E. Hajdin, K. M. Weeks, Principles for targeting RNA with drug-like small molecules. *Nat Rev Drug Discov* **17**, 547–558 (2018).
10. J. D. Watson, F. H. C. Crick, Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **171**, 737–738 (1953).
11. I. Tinoco, C. Bustamante, How RNA folds. *J Mol Biol* **293**, 271–281 (1999).
12. Q. Vicens, J. S. Kieft, Thoughts on how to think (and talk) about RNA structure. *Proc Natl Acad Sci U S A* **119**, e2112677119 (2022).
13. T. R. Cech, The ribosome is a ribozyme. *Science* (1979) **289**, 878–879 (2000).
14. J. P. Staley, J. L. Woolford, Assembly of ribosomes and spliceosomes: complex ribonucleoprotein machines. *Curr Opin Cell Biol* **21**, 109–118 (2009).
15. P. J. McCown, K. A. Corbino, S. Stav, M. E. Sherlock, R. R. Breaker, Riboswitch diversity and distribution. *Rna* **23**, 995–1011 (2017).
16. A. Haller, M. F. Soulière, R. Micura, The dynamic nature of RNA as key to understanding riboswitch mechanisms. *Acc Chem Res* **44**, 1339–1348 (2011).
17. A. C. Tan, *et al.*, Fingerprinting Tertiary Structure in Complex RNAs Using Single-Molecule Correlated Chemical Probing. *Biochemistry* **63**, 2648–2657 (2024).
18. D. N. Wilson, Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nat Rev Microbiol* **12**, 35–48 (2014).
19. R. Santos, *et al.*, A comprehensive map of molecular drug targets. *Nat Rev Drug Discov* **16**, 19–34 (2016).
20. J. Harrow, *et al.*, GENCODE: The reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760–1774 (2012).

21. I. Ezkurdia, *et al.*, Multiple evidence strands suggest that there may be as few as 19 000 human protein-coding genes. *Hum Mol Genet* **23**, 5866–5878 (2014).
22. S. J. Dixon, B. R. Stockwell, Identifying druggable disease-modifying gene products. *Curr Opin Chem Biol* **13**, 549–555 (2009).
23. M. Matsui, D. R. Corey, Non-coding RNAs as drug targets. *Nat Rev Drug Discov* **16**, 167–179 (2016).
24. J. Zhang, M. W. Lau, A. R. Ferré-D'Amaré, Ribozymes and Riboswitches: Modulation of RNA Function by Small Molecules. *Biochemistry* **49**, 9123 (2010).
25. W. M. Hewitt, D. R. Calabrese, J. S. Schneekloth, Evidence for ligandable sites in structured RNA throughout the Protein Data Bank. *Bioorg Med Chem* **27**, 2253–2260 (2019).
26. T. Zhou, H. Wang, C. Zeng, Y. Zhao, RPocket: an intuitive database of RNA pocket topology information with RNA-ligand data resources. *BMC Bioinformatics* **22**, 428 (2021).
27. S. Karthikeyan, *et al.*, Crystal structure of human riboflavin kinase reveals a β barrel fold and a novel active site arch. *Structure* **11**, 265–273 (2003).
28. A. Serganov, L. Huang, D. J. Patel, Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* **458**, 233–237 (2009).
29. J. R. Thomas, P. J. Hergenrother, Targeting RNA with small molecules. *Chem Rev* **108**, 1171–1224 (2008).
30. J. T. Koehn, S. Felder, K. M. Weeks, Innovations in targeting RNA by fragment-based ligand discovery. *Curr Opin Struct Biol* **79**, 102550 (2023).
31. J. L. Childs-Disney, *et al.*, Targeting RNA structures with small molecules. *Nat Rev Drug Discov* **21**, 736–762 (2022).
32. R. C. Moellering, Linezolid: The first oxazolidinone antimicrobial. *Ann Intern Med* [Preprint] (2003). Available at: www.annals.org [Accessed 3 January 2021].
33. H. Ratni, R. S. Scalco, A. H. Stephan, Risdiplam, the First Approved Small Molecule Splicing Modifier Drug as a Blueprint for Future Transformative Medicines. *ACS Med Chem Lett* **12**, 874–877 (2021).
34. H. Ma, X. Jia, K. Zhang, Z. Su, Cryo-EM advances in RNA structure determination. *Sig Transduct Target Ther* **7**, 1–6 (2022).
35. D. Tan, S. Piana, R. M. Dirks, D. E. Shaw, RNA force field with accuracy comparable to state-of-the-art protein force fields. *Proc Natl Acad Sci U S A* **115**, E1346–E1355 (2018).
36. J. W. Arney, K. M. Weeks, RNA-Ligand Interactions Quantified by Surface Plasmon Resonance with Reference Subtraction. *Biochemistry* **61**, 1625–1632 (2022).
37. M. J. Smola, G. M. Rice, S. Busan, N. A. Siegfried, K. M. Weeks, Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile, and accurate RNA structure analysis HHS Public Access. *Nat Protoc* **10**, 1643–1669 (2015).
38. M. J. Smola, K. M. Weeks, In-cell RNA structure probing with SHAPE-MaP. *Nat Protoc* **13**, 1181–1195 (2018).
39. A. Ke, J. A. Doudna, Crystallization of RNA and RNA–protein complexes. *Methods* **34**, 408–414 (2004).

40. D. B. Haack, *et al.*, Scaffold-enabled high-resolution cryo-EM structure determination of RNA. *Nat Commun* **16**, 1–17 (2025).
41. W. Wang, *et al.*, trRosettaRNA: automated prediction of RNA 3D structure with transformer network. *Nat Commun* **14**, 7266 (2023).
42. T. Shen, *et al.*, Accurate RNA 3D structure prediction using a language model-based deep learning approach. *Nat Methods* **21**, 2287–2298 (2024).
43. K. Kappel, *et al.*, Accelerated cryo-EM-guided determination of three-dimensional RNA-only structures. *Nat Methods* **17**, 699–707 (2020).
44. K. Yazdani, *et al.*, Machine Learning Informs RNA-Binding Chemical Space. *Angew Chem Int Ed Engl.* **62**, e202211358 (2023).
45. F. P. Panei, P. Gkeka, M. Bonomi, Identifying small-molecules binding sites in RNA conformational ensembles with SHAMAN. *Nat Commun* **15** (2024).
46. F. Walter, Q. Vicens, E. Westhof, Aminoglycoside–RNA interactions. *Curr Opin Chem Biol* **3**, 694–704 (1999).
47. A. L. Feig, Applications of isothermal titration calorimetry in RNA biochemistry and biophysics. *Biopolymers* **87**, 293–301 (2007).
48. H. Mukherjee, *et al.*, PEARL-seq: A Photoaffinity Platform for the Analysis of Small Molecule-RNA Interactions. *ACS Chem Biol* **15**, 2374–2381 (2020).
49. B. M. Suresh, *et al.*, A general fragment-based approach to identify and optimize bioactive ligands targeting RNA. *Proc Natl Acad Sci U S A* **117**, 33197–33203 (2021).
50. P. J. Homan, *et al.*, Single-molecule correlated chemical probing of RNA. *Proc Natl Acad Sci U S A* **111**, 13858–13863 (2014).
51. K. M. Weeks, SHAPE Directed Discovery of New Functions in Large RNAs. *Acc Chem Res* **54**, 2502–2517 (2021).
52. A. L. Hopkins, C. R. Groom, The druggable genome. *Nat Rev Drug Discov* **1**, 727–730 (2002).
53. S. Djebali, *et al.*, Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
54. K. Kavita, R. R. Breaker, Discovering riboswitches: the past and the future. *Trends Biochem Sci* **48**, 119–141 (2023).
55. M. Felletti, J. S. Hartig, Ligand-dependent ribozymes. *WIREs RNA* **8** (2017).
56. N. R. Voss, M. Gerstein, 3V: cavity, channel and cleft volume calculator and extractor. *Nucleic Acids Res* **38**, W555–W562 (2010).
57. D. Monet, N. Desdouits, M. Nilges, A. Blondel, MkgridXf: Consistent Identification of Plausible Binding Sites despite the Elusive Nature of Cavities and Grooves in Protein Dynamics. *J Chem Inf Model* **59**, 3506–3518 (2019).
58. F. P. Panei, R. Torchet, H. Mé Nager, P. Gkeka, M. Bonomi, HARIBOSS: a curated database of RNA-small molecules structures to aid rational drug design. *Bioinformatics* **38**, 4185–4193 (2022).
59. J. An, M. Totrov, R. Abagyan, Pocketome via comprehensive identification and classification of ligand binding envelopes. *Mol Cell Proteomics* **4**, 752–761 (2005).

60. A. Donlic, *et al.*, R-BIND 2.0: An Updated Database of Bioactive RNA-Targeting Small Molecules and Associated RNA Secondary Structures. *ACS Chem Biol* **17**, 1556–1566 (2022).
61. R. P. Sheridan, V. N. Maiorov, M. K. Holloway, W. D. Cornell, Y. D. Gao, Drug-like density: A method of quantifying the “bindability” of a protein target based on a very large set of pockets and drug-like ligands from the protein data bank. *J Chem Inf Model* **50**, 2029–2040 (2010).
62. G. R. Bickerton, G. V. Paolini, J. Besnard, S. Muresan, A. L. Hopkins, Quantifying the chemical beauty of drugs. *Nat Chem* **4**, 90–98 (2012).
63. V. Le Guilloux, P. Schmidtko, P. Tuffery, Fpocket: An open source platform for ligand pocket detection. *BMC Bioinformatics* **10**, 168 (2009).
64. J. J. Clark, Z. J. Orban, H. A. Carlson, Predicting binding sites from unbound versus bound protein structures. *Sci Rep* **10**, 15856 (2020).
65. J. Nix, D. Sussman, C. Wilson, The 1.3 Å crystal structure of a biotin-binding pseudoknot and the basis for RNA molecular recognition. *J Mol Biol* **296**, 1235–1244 (2000).
66. J. Flinders, *et al.*, Recognition of Planar and Nonplanar Ligands in the Malachite Green–RNA Aptamer Complex. *ChemBioChem* **5**, 62–72 (2004).
67. A. Serganov, A. Polonskaia, A. T. Phan, R. R. Breaker, D. J. Patel, Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **441**, 1167–1171 (2006).
68. J. A. Howe, *et al.*, Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the *E. coli* FMN riboswitch. *RNA Biol* **13**, 946–954 (2016).
69. G. Landrum, *et al.*, RDKit: Open-source cheminformatics. [Preprint] (2023). Available at: <https://www.rdkit.org> [Accessed 5 September 2023].
70. Schrödinger LLC, “The PyMOL Molecular Graphics System, Version 2.5.0” (2015).
71. T. Odland, K. Subbarao, C. Polster, paretoset. <https://github.com/tommyod/paretoset> (2023).
72. P. S. Irving, K. M. Weeks, RNavigate: efficient exploration of RNA chemical probing datasets. *Nucleic Acids Res* **52**, 2231–2241 (2024).
73. T. Zok, *et al.*, RNApdbee 2.0: multifunctional tool for RNA structure annotation. *Nucleic Acids Res* **46**, W30–W35 (2018).
74. J. S. Reuter, D. H. Mathews, RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
75. A. Umuhire Juru, A. E. Hargrove, Frameworks for targeting RNA with small molecules. *J Biol Chem* **296**, 100191 (2021).
76. S. Kovachka, *et al.*, Small molecule approaches to targeting RNA. *Nat Rev Chem* **8**, 120–135 (2024).
77. Z. L. Watson, *et al.*, Structure of the bacterial ribosome at 2 Å resolution. *Elife* **9**, 1–62 (2020).
78. C. Crowe-McAuliffe, *et al.*, Structural Basis for Bacterial Ribosome-Associated Quality Control by RqcH and RqcP. *Mol Cell* **81**, 115–126 (2021).
79. C. G. Parker, *et al.*, Ligand and Target Discovery by Fragment-Based Screening in Human Cells. *Cell* **168**, 527–541 (2017).

80. S. Balaratnam, *et al.*, A chemical probe based on the PreQ1 metabolite enables transcriptome-wide mapping of binding sites. *Nat Commun* **12**, 5856 (2021).
81. D. N. Wilson, *et al.*, The oxazolidinone antibiotics perturb the ribosomal peptidyl-transferase center and effect tRNA positioning. *Proc Natl Acad Sci U S A* **105**, 13339–13344 (2008).
82. J. A. Ippolito, *et al.*, Crystal structure of the oxazolidinone antibiotic linezolid bound to the 50S ribosomal subunit. *J Med Chem* **51**, 3353–3356 (2008).
83. L. R. Ganser, M. L. Kelly, D. Herschlag, H. M. Al-Hashimi, The roles of structural dynamics in the cellular functions of RNAs. *Nat Rev Mol Cell Biol* **20**, 474–489 (2019).
84. H. Shi, *et al.*, Rapid and accurate determination of atomistic RNA dynamic ensemble models using NMR and structure prediction. *Nat Commun* **11**, 5531 (2020).
85. K. Zhang, *et al.*, Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat Commun* **10**, 5511 (2019).
86. A. D. Garst, A. Héroux, R. P. Rambo, R. T. Batey, Crystal Structure of the Lysine Riboswitch Regulatory mRNA Element. *Journal of Biological Chemistry* **283**, 22347–22351 (2008).
87. J. L. Jenkins, J. Krucinska, R. M. McCarty, V. Bandarian, J. E. Wedekind, Comparison of a PreQ1 Riboswitch Aptamer in Metabolite-bound and Free States with Implications for Gene Regulation. *Journal of Biological Chemistry* **286**, 24626–24637 (2011).
88. J. R. Stagno, *et al.*, Structures of riboswitch RNA reaction states by mix-and-inject XFEL serial crystallography. *Nature* **541**, 242–246 (2017).
89. C. Lu, *et al.*, Crystal structures of the SAM-III/SMK riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nat Struct Mol Biol* **15**, 1076–1083 (2008).
90. G. Qu, *et al.*, Structure of a group II intron in complex with its reverse transcriptase. *Nat Struct Mol Biol* **23**, 549–557 (2016).
91. Y. Tong, *et al.*, Transcriptome-Wide Mapping of Small-Molecule RNA-Binding Sites in Cells Informs an Isoform-Specific Degrader of QSOX1 mRNA. *J Am Chem Soc* **144**, 11620–11625 (2022).
92. L. Fang, *et al.*, Pervasive transcriptome interactions of protein-targeted drugs. *Nat Chem* **15**, 1374–1383 (2023).
93. N. A. Siegfried, S. Busan, G. M. Rice, J. A. E. Nelson, K. M. Weeks, RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* **11**, 959–964 (2014).
94. A. Gabdulkhakov, *et al.*, Crystal Structure of the 23S rRNA Fragment Specific to r-Protein L1 and Designed Model of the Ribosomal L1 Stalk from Haloarcula marismortui. *Crystals (Basel)* **7**, 37 (2017).
95. P. A. Sharp, The Centrality of RNA. *Cell* **136**, 577–580 (2009).
96. T. R. Cech, J. A. Steitz, The Noncoding RNA Revolution—Trashing Old Rules to Forge New Ones. *Cell* **157**, 77–94 (2014).
97. C. R. Fullenkamp, X. Liang, M. Pettersson, J. Schneekloth, “Outlook” in *RNA as a Drug Target*, J. Schneekloth, M. Pettersson, R. Mannhold, H. Buschmann, J. Holenz, Eds. (John Wiley & Sons, Ltd, 2024), pp. 355–384.
98. C. M. Connelly, M. H. Moon, J. S. Schneekloth, The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol* **23**, 1077–1090 (2016).

99. S. D. Veenbaas, J. T. Koehn, P. S. Irving, N. N. Lama, K. M. Weeks, Ligand-binding pockets in RNA, and where to find them. *bioRxiv* 2025.03.13.643147 (2025). <https://doi.org/10.1101/2025.03.13.643147>.
100. A. M. Mustoe, C. L. Brooks, H. M. Al-Hashimi, Hierarchy of RNA functional dynamics. *Annu Rev Biochem* **83**, 441–466 (2014).
101. S. Bosio, M. Bernetti, W. Rocchia, M. Masetti, Similarities and Differences in Ligand Binding to Protein and RNA Targets: The Case of Riboflavin. *J Chem Inf Model* **64**, 4570–4586 (2024).
102. P. Schmidtke, V. Le Guilloux, J. Maupetit, P. Tufféry, fpocket: Online tools for protein ensemble pocket detection and tracking. *Nucleic Acids Res* **38**, W582–W589 (2010).
103. J. S. Reuter, D. H. Mathews, RNAsstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
104. C. Zhang, Y. Zhang, A. M. Pyle, rMSA: A Sequence Search and Alignment Algorithm to Improve RNA Structure Modeling. *J Mol Biol* **435**, 167904 (2023).
105. B. A. Sweeney, *et al.*, RNACentral: a hub of information for non-coding RNA sequences. *Nucleic Acids Res* **47**, D221–D229 (2019).
106. B. Zdrazil, *et al.*, The ChEMBL Database in 2023: a drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Res* **52**, D1180–D1192 (2024).
107. R. Flamary, *et al.*, POT Python Optimal Transport. [Preprint] (2025). Available at: <https://github.com/PythonOT/POT>.
108. K. Darty, A. Denise, Y. Ponty, VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25**, 1974–1975 (2009).
109. B. A. Sweeney, *et al.*, R2DT is a framework for predicting and visualising RNA secondary structure using templates. *Nat. Commun.* **12**, 1–12 (2021).
110. E. K. S. Mcrae, *et al.*, Structure, folding and flexibility of co-transcriptional RNA origami. *Nat Nanotechnol* **18**, 808–817 (2023).
111. N. Sampedro Vallina, E. K. S. McRae, B. K. Hansen, A. Boussebayle, E. S. Andersen, RNA origami scaffolds facilitate cryo-EM characterization of a Broccoli–Pepper aptamer FRET pair. *Nucleic Acids Res* **51**, 4613–4624 (2023).
112. R. Das, *et al.*, Assessment of three-dimensional RNA structure prediction in CASP15. *Proteins:Struct., Funct., Bioinf.* **91**, 1747–1770 (2023).
113. M. Magnus, Z. Miao, RNA 3D Structure Comparison Using RNA-Puzzles Toolkit. *Methods Mol. Biol.* **2586**, 263–285 (2023).
114. G. M. Schroeder, D. Kiliushik, J. L. Jenkins, J. E. Wedekind, 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 (2023).
115. C. Bernard, G. Postic, S. Ghannay, F. Tahí, State-of-the-RNArt: benchmarking current methods for RNA 3D structure prediction. *NAR Genom Bioinform* **6** (2024).
116. S. Mohan, J. P. Donohue, H. F. Noller, Molecular mechanics of 30S subunit head rotation. *Proc Natl Acad Sci U S A* **111**, 13325–13330 (2014).

117. E. J. Rundlet, *et al.*, Structural basis of early translocation events on the ribosome. *Nature* **595**, 741–745 (2021).
118. H. Paternoga, *et al.*, Structural conservation of antibiotic interaction with ribosomes. *Nat Struct Mol Biol* **30**, 1380–1392 (2023).
119. D. Bulkley, F. Johnson, T. A. Steitz, The Antibiotic Thermorubin Inhibits Protein Synthesis by Binding to Inter-Subunit Bridge B2a of the Ribosome. *J Mol Biol* **416**, 571–578 (2012).
120. L. Zhang, *et al.*, The structural basis for inhibition of ribosomal translocation by viomycin. *Proc Natl Acad Sci U S A* **117**, 10271–10277 (2020).
121. R. E. Stanley, G. Blaha, R. L. Grodzicki, M. D. Strickler, T. A. Steitz, The structures of the anti-tuberculosis antibiotics viomycin and capreomycin bound to the 70S ribosome. *Nat Struct Mol Biol* **17**, 289–293 (2010).
122. W. H. B. Sauer, M. K. Schwarz, Molecular shape diversity of combinatorial libraries: A prerequisite for broad bioactivity. *J Chem Inf Comput Sci* **43**, 987–1003 (2003).
123. C. Fan, X. Wang, T. Ling, Y. Yang, H. Zhao, Characterizing RNA-binding ligands on structures, chemical information, binding affinity and drug-likeness. *RNA Biol* **20**, 431–443 (2023).
124. E. Menichelli, *et al.*, Discovery of small molecules that target a tertiary-structured RNA. *Proc Natl Acad Sci U S A* **119**, 102550 (2022).
125. X. Xu, *et al.*, Insights into xanthine riboswitch structure and metal ion-mediated ligand recognition. *Nucleic Acids Res* **49**, 7139–7153 (2021).
126. L. Huang, D. M. J. Lilley, Structure and ligand binding of the SAM-V riboswitch. *Nucleic Acids Res* **46**, 6869–6879 (2018).
127. A. L. Edwards, F. E. Reyes, A. Héroux, R. T. Batey, Structural basis for recognition of S-adenosylhomocysteine by riboswitches. *RNA* **16**, 2144–2155 (2010).
128. M. M. Matyjasik, S. D. Hall, R. T. Batey, High Affinity Binding of N2-Modified Guanine Derivatives Significantly Disrupts the Ligand Binding Pocket of the Guanine Riboswitch. *Molecules* **25**, 2295 (2020).
129. A. Sun, *et al.*, SAM-VI riboswitch structure and signature for ligand discrimination. *Nat Commun* **10**, 5728 (2019).
130. P. Daldrop, D. M. J. Lilley, The plasticity of a structural motif in RNA: Structural polymorphism of a kink turn as a function of its environment. *RNA* **19**, 357–364 (2013).
131. J. A. Liberman, *et al.*, Structural analysis of a class III preQ1 riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. *Proc Natl Acad Sci U S A* **112**, E3485–E3494 (2015).
132. K. Nagano, T. Kamimura, G. Kawai, Interaction between a fluoroquinolone derivative and RNAs with a single bulge. *The Journal of Biochemistry* **171**, 239–244 (2022).
133. J. L. Chen, *et al.*, Design, Optimization, and Study of Small Molecules That Target Tau Pre-mRNA and Affect Splicing. *J Am Chem Soc* **142**, 8706–8727 (2020).
134. Z. Du, K. E. Lind, T. L. James, Structure of TAR RNA Complexed with a Tat-TAR Interaction Nanomolar Inhibitor that Was Identified by Computational Screening. *Chem Biol* **9**, 707–712 (2002).

135. L. Huang, T.-W. Liao, J. Wang, T. Ha, D. M. J. Lilley, Crystal structure and ligand-induced folding of the SAM/SAH riboswitch. *Nucleic Acids Res* **3**, 7545–7556 (2020).
136. R. B. Paulsen, *et al.*, Inhibitor-induced structural change in the HCV IRES domain IIa RNA. *Proc Natl Acad Sci U S A* **107**, 7263–7268 (2010).
137. S. C. Y. Jeng, *et al.*, Fluorogenic aptamers resolve the flexibility of RNA junctions using orientation-dependent FRET. *RNA* **27**, 433–444 (2021).
138. J. Davila-Calderon, *et al.*, IRES-targeting small molecule inhibits enterovirus 71 replication via allosteric stabilization of a ternary complex. *Nature Communications* **2020** *11*:1 **11**, 4775 (2020).
139. E. Duchardt-Ferner, *et al.*, Structure of an RNA aptamer in complex with the fluorophore tetramethylrhodamine. *Nucleic Acids Res* **48**, 949–961 (2020).
140. A. Serganov, *et al.*, Structural basis for Diels-Alder ribozyme-catalyzed carbon-carbon bond formation. *Nat Struct Mol Biol* **12**, 218–224 (2005).
141. K. Huang, *et al.*, Structure-based investigation of fluorogenic Pepper aptamer. *Nat Chem Biol* **17**, 1289–1295 (2021).
142. L. F. M. Passalacqua, *et al.*, Co-crystal structures of the fluorogenic aptamer Beetroot show that close homology may not predict similar RNA architecture. *Nat Commun* **14**, 2969 (2023).
143. S. D. Gilbert, R. P. Rambo, D. Van Tyne, R. T. Batey, Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat Struct Mol Biol* **15**, 177–182 (2008).
144. M. Mieczkowski, *et al.*, Large Stokes shift fluorescence activation in an RNA aptamer by intermolecular proton transfer to guanine. *Nat Commun* **12**, 3549 (2021).
145. X. Peng, W. Liao, X. Lin, D. M. J. Lilley, L. Huang, Crystal structures of the NAD⁺-II riboswitch reveal two distinct ligand-binding pockets. *Nucleic Acids Res* **51**, 2904–2914 (2023).
146. D. Krochmal, *et al.*, Structural basis for substrate binding and catalysis by a self-alkylating ribozyme. *Nat Chem Biol* **18**, 376–384 (2022).
147. O. Pikovskaya, A. Polonskaia, D. J. Patel, A. Serganov, Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nat Chem Biol* **7**, 748–755 (2011).
148. P. Fernandez-Millan, A. Autour, E. Ennifar, E. Westhof, M. Ryckelynck, Crystal structure and fluorescence properties of the iSpinach aptamer in complex with DFHBI. *RNA* **23**, 1788–1795 (2017).
149. E. B. Porter, J. T. Polaski, M. M. Morck, R. T. Batey, Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat Chem Biol* **13**, 295–301 (2017).
150. C. J. Robinson, *et al.*, Modular riboswitch toolsets for synthetic genetic control in diverse bacterial species. *J Am Chem Soc* **136**, 10615–10624 (2014).
151. C. P. Jones, A. R. Ferré-D'Amaré, Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains. *Nat Struct Mol Biol* **22**, 679–685 (2015).
152. J. A. Liberman, M. Salim, J. Krucinska, J. E. Wedekind, Structure of a class II preQ1 riboswitch reveals ligand recognition by a new fold. *Nat Chem Biol* **9**, 353–355 (2013).
153. L. Truong, *et al.*, The fluorescent aptamer Squash extensively repurposes the adenine riboswitch fold. *Nat Chem Biol* **18**, 191–198 (2021).

154. H. Jiang, *et al.*, The identification and characterization of a selected SAM-dependent methyltransferase ribozyme that is present in natural sequences. *Nat Catal* **4**, 872–881 (2021).
155. P. Schmidtke, C. Souaille, F. Estienne, N. Baurin, R. T. Kroemer, Large-scale comparison of four binding site detection algorithms. *J Chem Inf Model* **50**, 2191–2200 (2010).
156. H. M. Berman, *et al.*, The Protein Data Bank. *Nucleic Acids Res* **28**, 235–242 (2000).
157. M. S. Svetlov, *et al.*, High-resolution crystal structures of ribosome-bound chloramphenicol and erythromycin provide the ultimate basis for their competition. *RNA* **25**, 600–606 (2019).
158. J. A. Dunkle, L. Xiong, A. S. Mankin, J. H. D. Cate, Structures of the Escherichia coli ribosome with antibiotics bound near the peptidyl transferase center explain spectra of drug action. *Proc Natl Acad Sci U S A* **107**, 17152–17157 (2010).
159. J. L. Hansen, T. M. Schmeing, P. B. Moore, T. A. Steitz, Structural insights into peptide bond formation. *Proc Natl Acad Sci U S A* **99**, 11670–11675 (2002).
160. Z. D. Aron, *et al.*, trans-Translation inhibitors bind to a novel site on the ribosome and clear *Neisseria gonorrhoeae* in vivo. *Nat Commun* **12**, 1799 (2021).
161. H. M. Wilt, P. Yu, K. Tan, Y. X. Wang, J. R. Stagno, FMN riboswitch aptamer symmetry facilitates conformational switching through mutually exclusive coaxial stacking configurations. *J Struct Biol X* **4**, 100035 (2020).
162. G. M. Schroeder, *et al.*, Analysis of a preQ1-I riboswitch in effector-free and bound states reveals a metabolite-programmed nucleobase-stacking spine that controls gene regulation. *Nucleic Acids Res* **48**, 8146–8164 (2020).
163. H. K. Lee, *et al.*, Crystal structure of Escherichia coli thiamine pyrophosphate-sensing riboswitch in the apo state. *Structure* **31**, 848-859.e3 (2023).
164. C. D. Stoddard, *et al.*, Free State Conformational Sampling of the SAM-I Riboswitch Aptamer Domain. *Structure* **18**, 787–797 (2010).
165. D. J. Klein, A. R. Ferré-D'Amaré, Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science (1979)* **313**, 1752–1756 (2006).
166. D. J. Klein, S. R. Wilkinson, M. D. Been, A. R. Ferré-D'Amaré, Requirement of Helix P2.2 and Nucleotide G1 for Positioning the Cleavage Site and Cofactor of the glmS Ribozyme. *J Mol Biol* **373**, 178–189 (2007).
167. A. Peselis, A. Serganov, ykkC riboswitches employ an add-on helix to adjust specificity for polyanionic ligands. *Nature Chemical Biology* **2018 14:9** **14**, 887–894 (2018).
168. A. J. Knappenberger, C. W. Reiss, S. A. Strobel, Structures of two aptamers with differing ligand specificity reveal ruggedness in the functional landscape of RNA. *Elife* **7**, e36381 (2018).