

Supporting Information for:

SHAPE-enabled fragment-based ligand discovery for RNA

Meredith J. Zeller, Oleg Favorov, Kelin Li, Ashok Nuthanakanti, Dina Hussein, Auréliane Michaud, Daniel A. Lafontaine, Steven Busan, Alexander Serganov, Jeffrey Aubé, and Kevin M. Weeks*

* correspondence, weeks@unc.edu

Details of screening analysis methods, three tables, and four figures.

Supporting Ligand Screening Methods

Statistical test for a difference in modification rates of a given nucleotide

The screening analysis requires statistical comparison of the modification rate of a given nucleotide in the presence of a fragment as compared to its absence. For each nucleotide the number of modifications in a given reaction is a Poisson process with a known variance; the statistical significance of the observed difference in modification rates between two samples can therefore be ascertained by performing the Comparison of Two Poisson Counts test¹. That is, if m_1 modifications of a tested nucleotide were counted among n_1 reads in sample 1 and m_2 modifications were counted among n_2 reads in sample 2, the tested null hypothesis predicts that among all the counted modifications ($m_1 + m_2$), the proportion of modifications in sample 1 will be $p_1 = n_1/(n_1 + n_2)$. The Z-test of this hypothesis is:

$$Z_p = \frac{m_1 - p_1(m_1 + m_2) + 0.5}{\sqrt{p_1(1 - p_1)(m_1 + m_2)}}$$

$$Z_n = \frac{m_1 - p_1(m_1 + m_2) - 0.5}{\sqrt{p_1(1 - p_1)(m_1 + m_2)}}$$

$$Z = \min (|Z_p|, |Z_n|)$$

If the Z value exceeds a specified significance threshold, the tested nucleotide is taken to be statistically significantly affected by the presence of the test fragment.

Minimizing the multiple testing problem

For each fragment, the Z-test has to be performed on a large number of nucleotides comprising the RNA sequence, increasing the probability of false positives. While the numbers of false positive assignments of SHAPE reactivity per nucleotide can be minimized by raising Z significance threshold, this approach would reduce the sensitivity of the screen (meaning it would reduce the ability to detect weaker binding ligands). To reduce the number of Z-tests performed, such tests were applied only to nucleotides in the region of interest, rather than to all nucleotides in the RNA screening construct. For the dengue motif of the RNA, the region of interest was positions 59-110; for the TPP motif, the region of interest was positions 100-199. The number of Z-tests was reduced further by omitting nucleotides with low modification rates in both samples. The threshold for considering a nucleotide to have a low modification rate was

set at 25% of the plate-average modification rate, which was computed over all nucleotides in all 96 wells of a given plate. Z-tests were performed only on those nucleotides that, in at least one of the two compared samples, had the modification rate exceeding this 25% threshold.

Minimizing effects of uncontrolled factors on nucleotide reactivity

Ideally, the only difference between conditions in two compared samples would be the presence of a fragment in one sample but not in the other. Testing negative-control samples against each other can be used to gauge the prevalence of uncontrolled factors that might introduce across-sample variability in nucleotide modification rates. For example, if the Z significance threshold is set at 2.7, in the absence any such factors, the Z-test applied to pairs of negative-control (no fragment) samples should, theoretically, identify differentially reactive nucleotides with a probability $P = 0.0035$. However, when the Z-test was applied to pairs of negative-control samples selected at random from the 587 negative-control samples tested in the primary screen, the actual probability was 90 times higher with $P = 0.32$. Thus, there was statistically significant variability in SHAPE reactivities at individual nucleotides in the absence of fragments.

Although the majority of replicates shared essentially the same profiles, there were a substantial number of replicates with dissimilar profiles; some coefficients of determination were as low as 0.85. Applying the Z-test to dissimilar negative-control samples generated large numbers of cases where nucleotides were falsely classified as differentially reactive. To avoid this outcome, each sample was compared to the five most highly correlated negative-control samples. Z-tests applied to such selective pairs of negative controls with a Z significance threshold of 2.7, resulted in identification of differentially reactive nucleotides with a probability $P = 0.067$.

This probability is about 20 times higher than the theoretical $P = 0.0035$ indicating that there is variability in sample processing. Some of this variability scales equally across the reactivities of all the nucleotides of all RNAs in a sample. This variability can be removed by scaling down the overall reactivity in the more reactive sample so as to match the overall reactivity in the less reactive sample. Such scaling was performed by (i) computing for each nucleotide in the RNA sequence the ratio of its modification rate in the more reactive sample to that in the less reactive sample and (ii) dividing the modification rates of all the nucleotides in the more reactive sample by the median of the ratios obtained in step (i). Such scaling of correlation-maximized pairs of negative-control wells reduced the probability of finding nucleotide hits to $P = 0.030$, 9-fold higher than the theoretical probability. Thus, false-positive identification of fragments will occur,

as indeed occurs in all high-throughput screening assays, and we distinguished actual fragment hits from non-ligand variations by replicate SHAPE validation and by direct ligand binding measurement using ITC.

Maximizing the ratio of true-to-false ligand detections

Since an effective ligand is expected to affect modification rates of multiple nucleotides in the target RNA, a fragment was recognized as a hit only if the number of nucleotides with reactivity different from that in the negative control exceeded a defined threshold, which was set to 2. Second, since we are looking for relatively robust effects of fragments on the RNA, small relative differences in reactivity of a nucleotide, even if statistically significant, were excluded from the total count of differentially reactive nucleotides. In practice, the minimal accepted difference was set to 20% of the average:

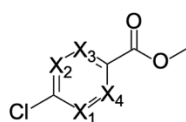
$$|r_1 - r_2| / (r_1 + r_2) / 2 = 0.2,$$

where r_1 and r_2 are the nucleotide modification rates in two samples. Third, a given sample was tested against the five negative-control samples with which it was most highly correlated. All five tests were required to find the test sample altered relative to the negative-control sample.

Finally, the sensitivity and specificity of the screen were controlled by the choice of Z significance threshold. Evaluation of samples containing fragments and all negative-control samples was performed at multiple Z significance threshold settings. For each such setting, the false-positive fraction (FPF) was computed as a fraction of the negative-control samples that were found to be altered, and the ligand fraction (LF) was estimated by subtracting FPF from the fraction of altered samples containing a fragment. The balance between LF and FPF was quantified by their ratio, LF/FPF. The best balance (LF/FPF \approx 1.3) for the TPP riboswitch RNA was achieved with Z significance threshold in the range between 2.5 and 2.7, at which $0.022 > \text{FPF} > 0.014$. For the dengue pseudoknot, the best balance (LF/FPF \approx 4) was achieved with Z significance threshold in the range between 2.5 and 2.65, at which $0.007 > \text{FPF} > 0.005$.

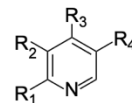
Table S1. Structure-activity relationships for analogs of fragment **5** binding to the TPP riboswitch RNA. Modifications to the (A) pyridine core and (B) pendant groups. Dissociation constants obtained by ITC.

A



Molecule	X ₁ , X ₂ , X ₃ , X ₄	K _d (μM)
5	N, C, C, C	265
S1	C, C, C, N	490
S2	N, N, C, C	420
S3	N, C, N, C	1200
S4	C, C, C, C	no binding

B



Molecule	R ₁	R ₂	R ₃	R ₄	K _d (μM)
S5		H	H		440
S6		H	H		390
S7		H	H		1800
S8		H	H		1100
S9		H	H		no binding
S10		H	H		no binding
S11			H		820
S12		H			93
S13		H	H		600
S14		H	H		1300
S15		H	H		1800
S16		H	H		1300
S17		H	H		no binding
S18		H	H		no binding

Table S2. X-ray crystallography data collection and refinement statistics for thiamine pyrophosphate (TPP) riboswitch co-crystallized with fragment and drug-like ligands.

Data collection				
Co-crystallized compounds	16	17	37	38
Wavelength (Å)	0.9252	0.9184	0.9791	0.9793
Space group	<i>C2</i>	<i>C2</i>	<i>P3212</i>	<i>P3212</i>
Cell dimensions				
a, b, c (Å)	150.16, 29.61, 95.51	148.73, 30.40, 95.57	61.02, 61.02, 102.85	61.70, 61.70, 102.79
α, β, γ (°)	90.00 94.32 90.00	90.00 93.51 90.00	90.00 120.00 90.00	90.00 120.00 90.00
Resolution (Å)	30.00–2.70 (2.80–2.70) ^a	29.78–2.21 (2.27–2.21) ^a	50.00–2.93 (2.98–2.93) ^a	30.00–2.87 (2.92–2.87) ^a
R _{merge} ^b	0.152 (0.581)	0.080 (0.729)	0.130 (3.612)	0.136 (1.774)
R _{pim} ^c	0.101 (0.407)	0.069 (0.623)	0.044 (1.140)	0.042 (0.583)
CC _{1/2}	0.981 (0.656)	0.997 (0.765)	0.995 (0.275)	0.944 (0.637)
$\ \sigma(I) \ $	9.9 (1.3)	10.4 (1.6)	304.8 (5.5)	32.0 (0.9)
Completeness (%)	95.2 (92.3)	99.2 (91.4)	99.9 (100)	99.8 (100)
Redundancy	2.9 (2.5)	4.0 (3.7)	10.0 (10.6)	11.8 (9.7)
No. unique reflections	11,400 (1,061)	21,835 (1,447)	4,715 (238)	5,295 (267)
Refinement				
Resolution (Å)	29.05 – 2.70	29.78 – 2.21	47.00 – 2.96	29.55 – 2.87
R _{work} /R _{free} (%)	20.3/26.1	20.4/24.4	24.3/25.7	22.9/26.9
No. of atoms				
RNA	3335	3361	1655	1655
Lead	48	24	11	25
Other ligands ^d	3	11	3	6
Water	34	126	-	2
Average B-factors (Å ²)				
RNA	43.47	53.50	115.97	95.67

Lead	35.13	39.14	68.55	79.90
Ligand	46.24	45.54	119.21	94.97
Water	28.91	36.79	-	79.83
RMS deviations				
Bond lengths (Å)	0.003	0.005	0.002	0.004
Bond angles (°)	0.848	1.115	0.605	0.748
PDB code	7TZR	7TZS	7TZT	7TZU



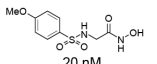


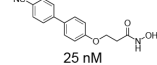


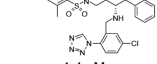


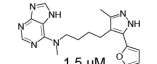


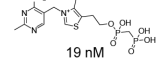


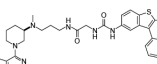


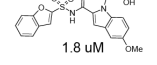


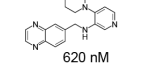


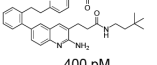


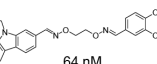


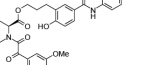


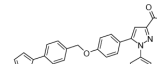


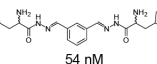


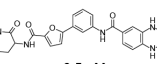
^aThe highest-resolution shell values are shown in parentheses.

^b $R_{\text{merge}} = \frac{\sum_h \sum_i |I(h)_i - \langle I(h) \rangle|}{\sum_h \sum_i I(h)_i}$, where $I(h)$ is the intensity for reflection h , \sum_h is the sum for all reflections, and \sum_i is the sum for i measurements of reflection h .

^c $R_{\text{rim}} = \frac{\sum_{hkl} \sqrt{(1/(n-1)) \sum_i |I(hkl)_i - \langle I(hkl) \rangle|}}{\sum_{hkl} \sum_i I(hkl)_i}$

^dLigand indicates components of the crystallization solution (buffer, cations, etc) except lead or drug molecules.

Table S3. Comparison of representative protein and RNA ligands developed by fragment-based methods. RNA examples are noted with an asterisk. Entries detail the two component fragments and their individual K_d values, the linked compound and its corresponding K_d value, and the ligand efficiency (LE) and linking coefficient (E) for the linked compound²⁻¹³

Fragment 1	Fragment 2	Linked compound	LE	E
 6.2 mM	 1.5 mM	 20 nM	0.62	0.0021
 17 mM	 20 μ M	 25 nM	0.49	0.06
 330 μ M	 12 μ M	 1.4 nM	0.30	0.35
 1.5 mM	 1 mM	 1.5 μ M	0.31	1.0
 11 μ M	 >1 mM	 19 nM	0.40	1.4 *
 240 μ M	 9.0 μ M	 3 nM	0.40	1.4
 1 mM	 1.1 mM	 1.8 μ M	0.26	1.6
 25 μ M	 >10 mM	 620 nM	0.34	2.5 *
 114 μ M	 140 nM	 400 μ M	0.28	25
 40 μ M	 41 μ M	 64 nM	0.32	39
 2 μ M	 100 μ M	 49 nM	0.22	250
 150 μ M	 580 μ M	 26 μ M	0.17	300
 13 μ M	 13 μ M	 54 nM	0.25	330
 >100 μ M	 >100 μ M	 6.5 μ M	0.17	650 *

A

GGUCGCGAGUAAUCGCGACCGCUGCAAAGAGAUUGUAGCGUGGGCACUUCGGUGUCCACACGCGA**AGG**AAAC
CGCGUGUCAACUGUGCAACAGCUGACA**AAGAGAUCCU**AAAACUCAGUACUCGGGGUGCCUUCUGCGUGA
AGGCUGAGAAAUACCCGUAUCACCUGAUCUGGAUAAUGCCAGCGUAGGGAAGU**GCUG**GAUCCGGUUCGCCG
GAUCAUCCGGCUUCGGUCCGGUUC

Green = Structure cassette

Orange = RNA barcode (barcode NT underlined)

Red = DENV pseudoknot (mutations **bold**)

Black = linker

Blue = TPP riboswitch (mutations **bold**)

B

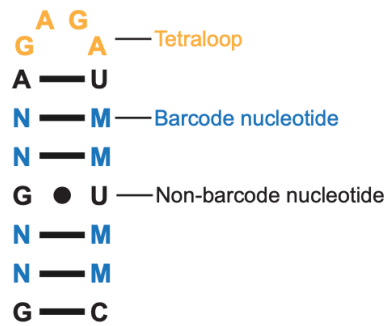


Figure S1. Screening construct design. (A) RNA sequence, colored by individual components. (B) Secondary structure of the RNA-sequence barcode in the context of its self-folding hairpin.

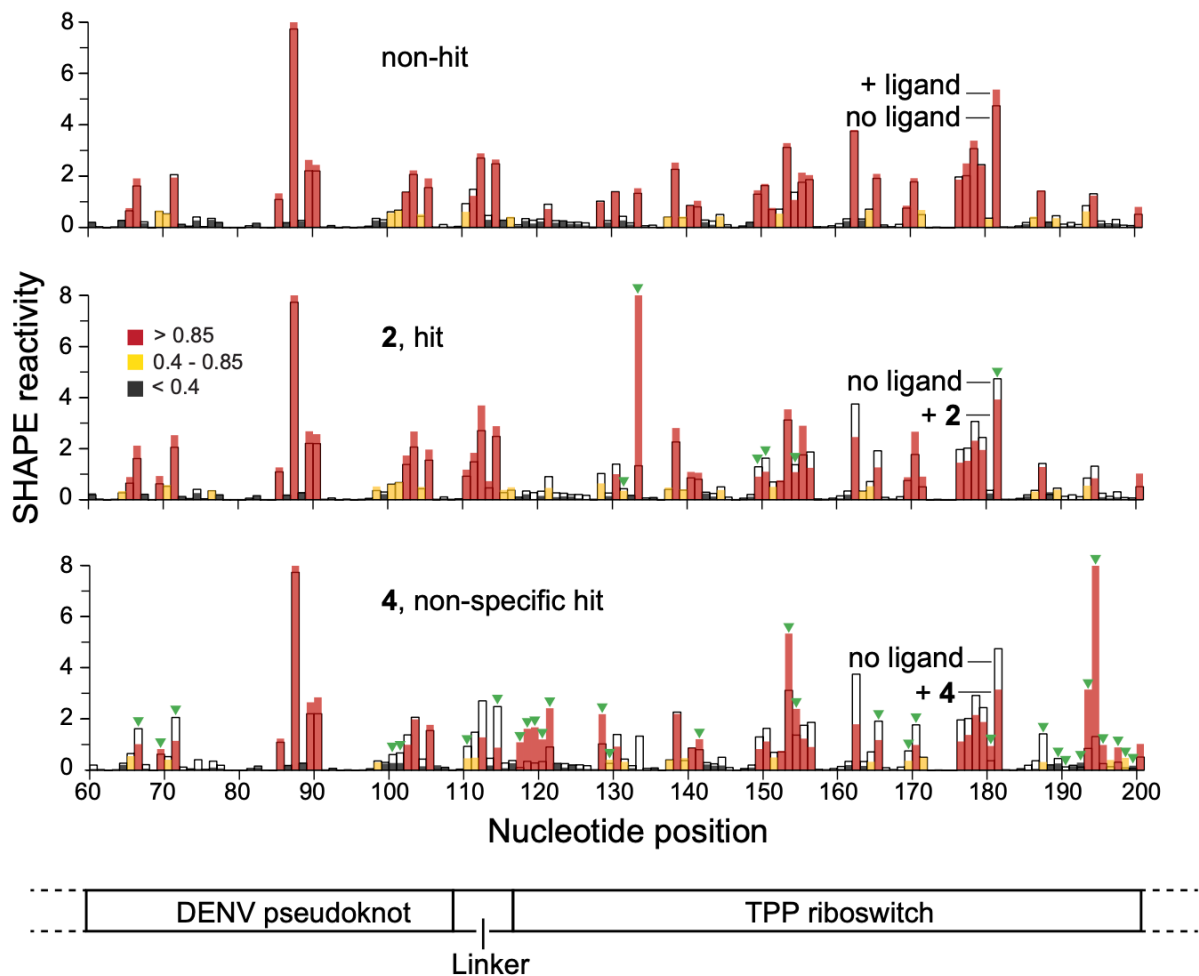


Figure S2. SHAPE profiles for non-hit, hit, and nonspecific hit fragments. Mutation rate traces corresponding to fragment-exposed and no-ligand control traces are in solid colors (red, yellow and black) and in black outline, respectively. Nucleotides determined to be statistically significantly different in fragment versus no fragment samples are denoted by green triangles. Mutation rate traces for the same fragments are shown schematically in Fig. 2.

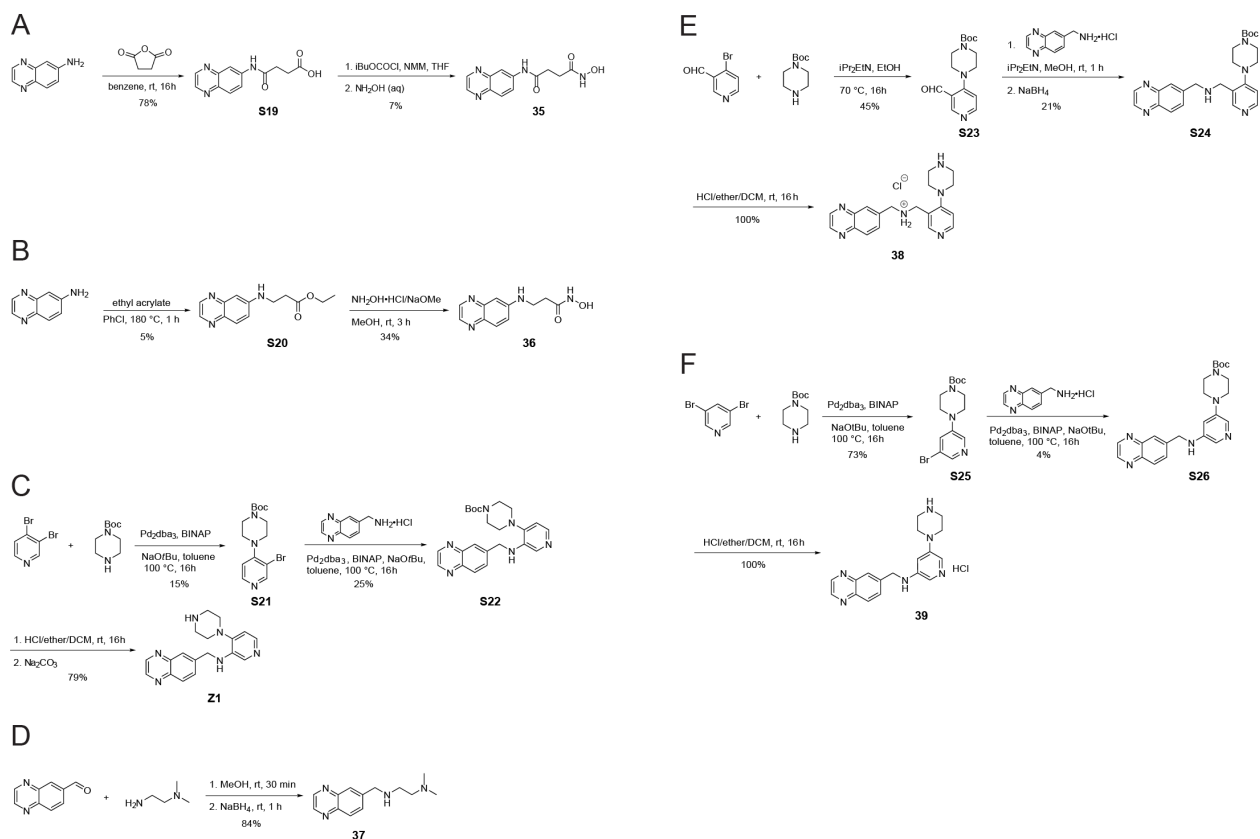


Figure S3. Synthetic schemes for ligands newly synthesized in this work.

(A) Compound **35**: 3-C linked hydroxamic acid **35** was prepared from carboxylic acid **S19** via a mixed anhydride intermediate by reacting with aqueous hydroxylamine. The acid **S19** was accessed by treating quinoxalin-6-amine with cyclized anhydride dihydrofuran-2,5-dione.

(B) Compound **36**: The 2-C linked analog **36** was obtained from the corresponding ester **S20** by reacting with hydroxylamine formed in situ. Ester **S20** was made via Michael addition of quinoxalin-6-amine with ethyl acrylate.

(C) Compound **Z1**: The Buchwald-Hartwig reaction was used for the synthesis of intermediate **S21** and **S22**. Protecting group (Boc) removal was achieved with HCl in ether, followed by further treatment with Na_2CO_3 to give **Z1**.

(D, E) Compounds **37**, **38**: **37** was prepared through imine formation and sodium borohydride reduction using quinoxalin-6-ylmethanamine hydrochloride and aldehyde **S23**, prepared via $\text{S}_{\text{N}}\text{Ar}$ reaction. Further (Boc) deprotection of **S24** with HCl gave **37**. Similar reaction conditions were used to generate **38**.

(F) Compound **39**: The less constrained analog **39** was made with two Buchwald-Hartwig reactions with 3,5-dibromopyridine, followed by (Boc) deprotection with HCl.

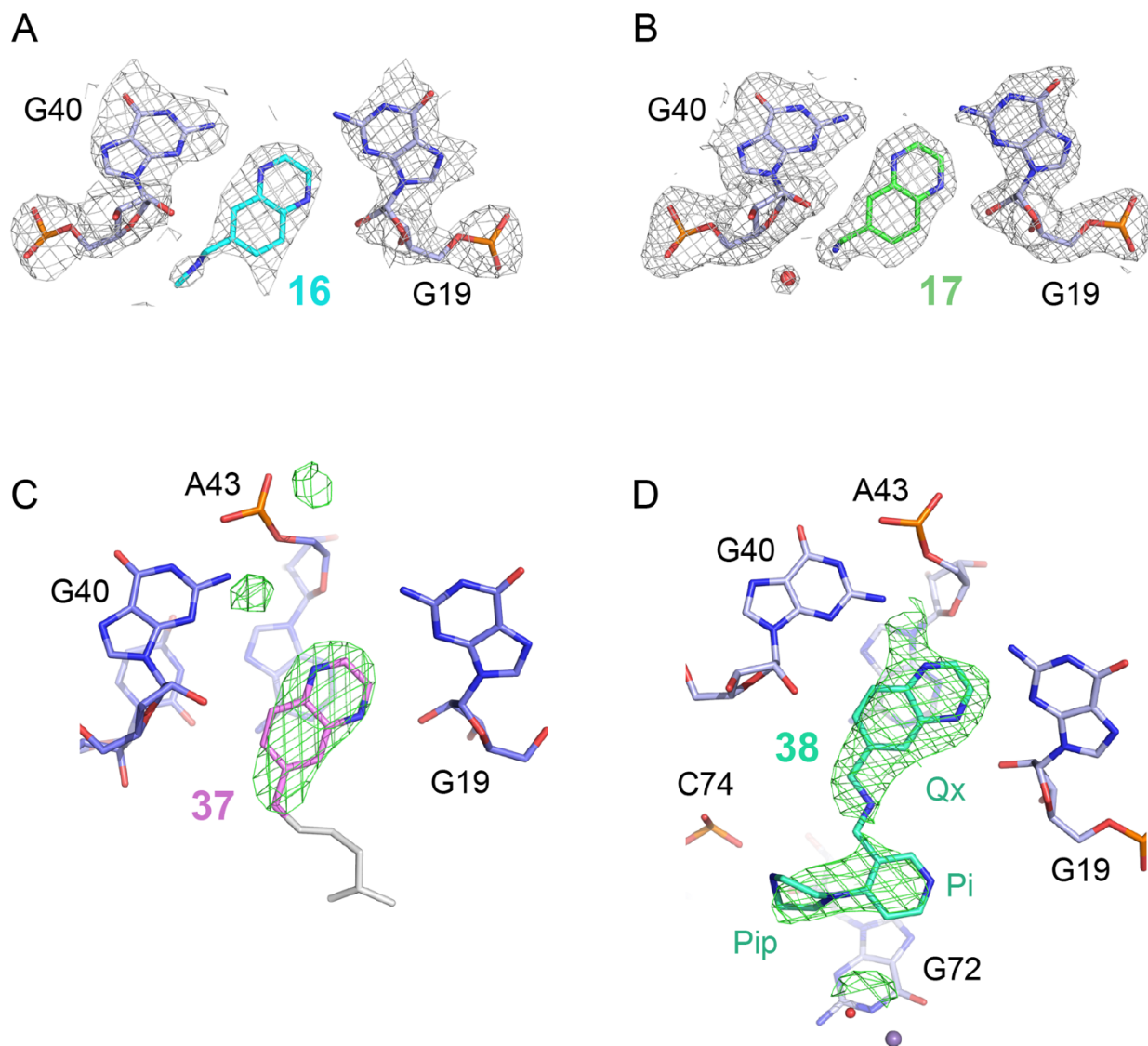


Figure S4. Electron density maps for riboswitch-ligand complexes. (A) and (B) $2F_o-F_c$ simulated annealing omit map contoured at 0.8σ level (gray mesh) for refined structures of RNA-bound fragments (A) **16** and (B) **17**, respectively. (C) F_o-F_c omit map contoured at 2.5σ level (green mesh) for the refined structure of the RNA-bound **37**. Flexible linker of **37**, not visible in the final structure, is modeled in gray. (D) F_o-F_c omit map contoured at 2.0σ level (green mesh) for the refined structure of RNA-bound compound **38**. Ox, quinoxaline; Pi, pyridine; Pip, piperazine.

Supporting References

1. Woolson, R. *Statistical Methods for the Analysis of Biomedical Data*. (John Wiley & Sons, 1987).
2. Shuker, S. B., Hajduk, P. J., Meadows, R. P. & Fesik, S. W. Discovering High-Affinity Ligands for Proteins : SAR by NMR. *Science (80-.)*. **274**, 1531–1534 (1996).
3. Hajduk, P. J. *et al.* Discovery of potent nonpeptide inhibitors of stromelysin using SAR by NMR. *J. Am. Chem. Soc.* **119**, 5818–5827 (1997).
4. Maly, D. J., Choong, I. C. & Ellman, J. A. Combinatorial target-guided ligand assembly: Identification of potent subtype-selective c-Src inhibitors. *Proc. Natl. Acad. Sci.* **97**, 2419–2424 (2000).
5. Swayze, E. E. *et al.* SAR by MS: A ligand based technique for drug lead discovery against structured RNA targets. *J. Med. Chem.* **45**, 3816–3819 (2002).
6. Howard, N. *et al.* Application of fragment screening and fragment linking to the discovery of novel thrombin inhibitors. *J. Med. Chem.* **49**, 1346–1355 (2006).
7. Barker, J. J. *et al.* Discovery of a novel Hsp90 inhibitor by fragment linking. *ChemMedChem* **5**, 1697–1700 (2010).
8. Borsi, V., Calderone, V., Fragai, M., Luchinat, C. & Sarti, N. Entropic contribution to the linking coefficient in fragment based drug design: A case study. *J. Med. Chem.* **53**, 4285–4289 (2010).
9. Ichihara, O., Barker, J., Law, R. J. & Whittaker, M. Compound design by fragment-linking. *Mol. Inform.* **30**, 298–306 (2011).
10. Jordan, J. B. *et al.* Fragment-Linking Approach Using ¹⁹F NMR Spectroscopy to Obtain Highly Potent and Selective Inhibitors of β -Secretase. *J. Med. Chem.* **59**, 3732–3749 (2016).
11. Mondal, M. *et al.* Fragment Linking and Optimization of Inhibitors of the Aspartic Protease Endothiapepsin: Fragment-Based Drug Design Facilitated by Dynamic Combinatorial Chemistry. *Angew. Chemie - Int. Ed.* **55**, 9422–9426 (2016).
12. Möbitz, H. *et al.* Discovery of Potent, Selective, and Structurally Novel Dot1L Inhibitors by a Fragment Linking Approach. *ACS Med. Chem. Lett.* **8**, 338–343 (2017).
13. Zeller, M.J., Nuthanakanti, A., Li, K., Aubé, J., Serganov, A. & Weeks, K.M., Multisite ligand recognition and cooperativity in the TPP riboswitch: Implications for fragment-linking in RNA ligand discovery, *ACS Chem. Biol.* **17**, 438-448 (2022).