Supporting Information

Figure S1. Guided fragment assembly for multi-loops
(A) Secondary structure of tRNA multi-loop with all helices independent and multi-loop closure assembled. Green line represents single-stranded path that will be folded at the midpoint to bring NT1 and NT2 into a base-paired conformation. (B) RSIM assembled tertiary structure for secondary structure illustrated in A. (C) An example of one fragment assembly path with distance constraints represented by zigzagged lines and green positions representing fragment replacement positions. (D) RSIM assembled tertiary structure of tRNA multi-loop.

Figure S2. Fragment representations
(A) Virtual bond model of fragments based on C1' atoms. Fragments are represented by four bond lengths, $\rho_{1\text{-}4}$, three bond angles, $\theta_{1\text{-}3}$, and two torsion angles, $\tau_{1\text{-}2}$. This representation is stored in RSIM frag_db file (B) The virtual bond model and base information are stored as nine (x,y,z) coordinates in the RSIM frag_nt file. (C) The all-atom representation records the (x,y,z) coordinates of all heavy atoms in the backbone and is stored in the RSIM frag_at file.

Figure S3. Composite score correlation with RMSD
Pearson correlation between composite score and C1' atom RMSD for each RNA benchmark. Plots show the score and C1' RMSD for the minimal scored conformations in secondary structure clusters for each benchmark. Only scores below 40 were included to insure that the highest resolution metrics had been reached.

Figure S4. Local RMSD for Benchmarks
Local RMSD measures are shown for the all-atom RMSD between two structures in the local region around a given nucleotide. All bases connected to a nucleotide by Voronoi polyhedra in the native structure define the local region at that particular nucleotide. Plots have nucleotide position along the x-axis and all-atom local RMSD along the y-axis. For over 50% of RNA benchmarks the local RMSD is consistently below 2 Å (red line) indicating native-like conformations were sampled in those regions.

NMR and X-ray structures used to construct the fragment library and evaluate the scoring metrics
PDB structure files with the following IDs where downloaded from the RCSB Protein Data Bank. For NMR structures with multiple models, the first model was used.

Ribosome Structure PDB IDs (47)
1i94, 1i95, 1i96, 1i97, 1j5a, 1jzx, 1jzy, 1jzz, 1k01, 1njn, 1ond, 1p9x, 1vs5, 1vs7, 1z58, 2aar, 2avy, 2aw7, 2d3o, 2hhh, 2o43, 2o44, 2o45, 2ogm, 2ogn, 2ogo, 2qal, 2qan, 2qb9, 2qbb, 2qbd, 2qbf, 2qbh, 2qbj, 2qou, 2qow, 2qoy, 2qp0, 2vho, 2vhp, 2z4k, 2z4m, 3bbn, 3df1, 3df3, 3i1m, 3i1o

Small RNA PDB IDs (348)

17ra, 1a51, 1a60, 1a9l, 1afx, 1ajf, 1aju, 1akx, 1am0, 1anr, 1aqo, 1arj, 1ato, 1atv, 1atw, 1b36, 1bgz, 1bn0, 1bvj, 1byj, 1bz2, 1bz3, 1bzt, 1bzu, 1c0o, 1c2x, 1cq5, 1cql, 1d0t, 1d0u, 1duh, 1e4p, 1e95, 1ebq, 1ebr, 1ebs, 1eht, 1ehz, 1ei2, 1esh, 1esy, 1evv, 1f1t, 1f6x, 1f6z, 1f78, 1f79, 1f7f, 1f7g, 1f7h, 1f7i, 1f84, 1f85, 1f9l, 1feq, 1fhk, 1fl8, 1fmn, 1fqz, 1fyo, 1fyp, 1hlx, 1hs1, 1hs2, 1hs3, 1hs4, 1hs8, 1hwq, 1i3x, 1i3y, 1i46, 1i4b, 1i4c, 1ie1, 1ie2, 1ik1, 1ikd, 1j4y, 1jo7, 1jox, 1jp0, 1jtj, 1jtw, 1ju7, 1jur, 1jwc, 1jzc, 1k2g, 1k4a, 1k4b, 1k5i, 1k6g, 1k6h, 1kaj, 1kfo, 1kka, 1kks, 1kos, 1kp7, 1kpd, 1kpy, 1kpz, 1kxk, 1l1w, 1lc6, 1ldz, 1lu3, 1luu, 1lux, 1lvj, 1m5l, 1m82, 1mfj, 1mfk, 1mfy, 1mt4, 1n66, 1n8x, 1na2, 1nbk, 1nbr, 1nc0, 1nem, 1nz1, 1o15, 1oq0, 1osw, 1ow9, 1p5m, 1p5n, 1p5o, 1p5p, 1pbr, 1pjy, 1q75, 1q8n, 1qc8, 1qd3, 1qwa, 1qwb, 1r2p, 1r7w, 1r7z, 1raw, 1rfr, 1rht, 1rmn, 1rng, 1rnk, 1roq, 1s2f, 1s34, 1s9s, 1scl, 1slo, 1slp, 1sy4, 1syz, 1szy, 1t28, 1tbk, 1tfn, 1tjz, 1tlr, 1tn2, 1tob, 1tra, 1txs, 1u2a, 1u3k, 1u8d, 1u9s, 1uts, 1uud, 1uui, 1uuu, 1vop, 1wks, 1wts, 1wtt, 1xhp, 1xjr, 1xsg, 1xsh, 1xst, 1xsu, 1xwp, 1xwu, 1y26, 1y27, 1yg3, 1yg4, 1ylg, 1ymo, 1yn1, 1yn2, 1ync, 1yne, 1yng, 1ysv, 1z2j, 1z30, 1z31, 1z43, 1zc5, 1zif, 1zig, 1zih, 255d, 283d, 28sp, 28sr, 2a43, 2a64, 2a9l, 2aht, 2ap0, 2ap5, 2au4, 2awq, 2b57, 2b7g, 2cd1, 2cd3, 2cd5, 2cd6, 2ees, 2eet, 2eeu, 2eev, 2eew, 2es5, 2euy, 2evy, 2f87, 2f88, 2fdt, 2fey, 2frl, 2g1g, 2g1w, 2g9c, 2gbh, 2gio, 2gip, 2gis, 2grw, 2gv3, 2gv4, 2gvo, 2h2x, 2hem, 2hns, 2hua, 2ixy, 2jr4, 2jrg, 2jrq, 2jse, 2jsg, 2jtp, 2juk, 2jwv, 2jxv, 2jym, 2k4c, 2k5z, 2k63, 2k66, 2k95, 2k96, 2kd8, 2ke6, 2kez, 2kf0, 2kfc, 2kgp, 2khy, 2kmj, 2koc, 2ldz, 2o33, 2pcv, 2q1r, 2qbz, 2qh2, 2qh3, 2qh4, 2rlu, 2ro2, 2rp0, 2rp1, 2rpk, 2rpt, 2tob, 2tpk, 2u2a, 2zy6, 387d, 3bbv, 3bnt, 3d0m, 3d0u, 3d0x, 3dhs, 3dig, 3dil, 3dim, 3dio, 3diq, 3dir, 3dis, 3dix, 3diy, 3diz, 3dj0, 3dj2, 3dw4, 3e5c, 3e5e, 3e5f, 3f2q, 3f2w, 3f2x, 3f2y, 3f30, 3fo4, 3fo6, 3g4m, 3gao, 3gca, 3ger, 3ges, 3gog, 3got, 3gx2, 3gx3, 3gx5, 3gx6, 3gx7, 3la5, 3php, 3tra, 430d, 480d, 4tna, 4tra, 6tna

Small RNA with Protein (142)

1a1t, 1a4t, 1aud, 1b23, 1biv, 1bmv, 1c0a, 1c9s, 1cx0, 1d6k, 1drz, 1dul, 1e8o, 1eiy, 1ekz, 1etf, 1etg, 1euq, 1euy, 1exd, 1exy, 1f6u, 1f8v, 1fje, 1g70, 1gtf, 1gtn, 1gtr, 1gts, 1hji, 1hq1, 1i9f, 1j1u, 1jid, 1k1g, 1k8w, 1l1c, 1l9a, 1lng, 1mfq, 1mnb, 1mzp, 1nyb, 1o0b, 1o0c, 1ob2, 1oln, 1qf6, 1qfq, 1qrs, 1qrt, 1qru, 1qtq, 1qu3, 1rkj, 1sj3, 1sj4, 1sjf, 1t4l, 1u0b, 1u6p, 1ull, 1vbx, 1vby, 1vbz, 1vc0, 1vc5, 1vc6, 1vc7, 1zbn, 1zl3, 1zo1, 1zse, 2a9x, 2ab4, 2ake, 2anr, 2asb, 2b63, 2b6g, 2bx2, 2dr2, 2dr5, 2dr7, 2dr8, 2dr9, 2dra, 2drb, 2du3, 2du4, 2du5, 2du6, 2dvi, 2ese, 2fk6, 2fy1, 2gic, 2hgh, 2hw8, 2ihx, 2ix1, 2iy5, 2jq7, 2kdq, 2nue, 2oih, 2oj3, 2pjp, 2pxb, 2pxd, 2pxe, 2pxf, 2pxk, 2pxl, 2pxp, 2pxq, 2pxt, 2pxu, 2pxv, 2q23, 2q25, 2q26, 2r8s, 2r93, 2rd2, 2re8, 2z2q, 2zh1, 2zh2, 2zh3, 2zh4, 2zh5, 2zh6, 2zh7, 2zh8, 2zh9, 2zha, 2zhb, 3hhz, 3iab, 3irw, 484d

FARNA Benchmarks (20)

157d, 1a4d, 1csl, 1dqf, 1esy, 1i9x, 1j6s, 1kd5, 1kka, 1l2x, 1mhk, 1q9a, 1qwa, 1xjr, 1zih, 255d, 283d, 28sp, 2a43, 2f88

Table S1. Fragment Library

All X-ray and NMR structures were decomposed into trinucleotide fragments of the same pyrimidine/purine sequence composition. For NMR structures the first model was used. Each fragment contained all heavy atoms between and including the 5' and 3' C1' atoms

of the nt before the first and after the third nt in the fragment. To generate the final fragment library, fragments that had all atom RMSD values within 0.5Å were considered duplicates and only one from such a pair was kept. To ensure that the library contained a diverse set of fragments the secondary structure of each fragment with respect to its parent RNA molecule was determined. For all 8 possible pyrimidine/purine sequence compositions 1,800 fragments from each class were randomly chosen such that 900 were from completely double-stranded regions and 900 from single-stranded or partially double-stranded regions. All fragments were aligned to a local coordinate system defined by the C1' atoms of the fragments. The position of the final C1' atom with respect to the local coordinate system origin was recorded. For each fragment class 300 clusters centers were identified with *k*-means clustering of the final C1' atom position. The radius of each cluster was set to half the distance to the nearest cluster center or the minimal distance needed to contain all cluster members. Any fragment that fell outside of a cluster was excluded. This resulted in a final number of fragments for each fragment class being 1660(RRR), 1674(RRY), 1664(RYR), 1664(RYY), 1644(YRR), 1670(YRY), 1656(YYR), and 1659(YYY).

Table S2. Generating a looping database

Given any RNA sequence of 6- or 7-nt, the probability of randomly identifying fragments that bring the terminal bases into a planar conformation (looped) is not uniform for all sequence compositions. This creates a bias that leads to preferential looping of certain sequences during a Monte Carlo simulation. The position of the loop ultimately determines the composition of the helix stemming from the loop base. Hence, any biases in the looped sequence composition could lead to false enhancements of tertiary structure and secondary structure predictions. To eliminate this potential bias we generated 20 million random conformations for each pyrimidine/purine sequence composition and identified conformations that brought the terminal bases into a planar conformation. This resulted in 2,857 loops of length 6 and 4,035 loops of length 7. We were then able to generate looped conformations for arbitrary sequences by threading the sequences through all looped conformations and identifying conformations with no steric clashes for the given sequence.

Table S3. Scoring Metrics

We evaluated nucleotide level metrics that were to be incorporated into a statistical potential on 348 small RNA structures (listed above), 47 ribosomal structures (listed above), the top 505 FARNA predictions (1), and the corresponding known structures for FARNA benchmarks (listed above). The total number of guanines was 2433, 29736, 181, and 91405 for small RNAs, ribosomes, FARNA benchmarks, and FARNA generated decoys, respectively. The total number of adenines was 1398, 23297, 111, and 56055 for small RNAs, ribosomes, FARNA benchmarks, and FARNA generated decoys, respectively. The total number of uracils was 1445, 17706, 85, and 42925 for small RNAs, ribosomes, FARNA benchmarks, and FARNA generated decoys, respectively. The total number of cytosines was 2130, 22091, 144, and 72720 for small RNAs, ribosomes, FARNA benchmarks, and FARNA generated decoys, respectively. Many of the metrics are derived from properties of the Voronoi polyhedra constructed around

atoms of the RNA molecules. These polyhedra were generated using algorithms from the Voro++ C++ library (2).

Graph Order
Each face of a polyhedron bisects a vector between two atoms. If the two atoms are from different nt, then the face is considered shared between the two nt. The graph order measures the number of unique nt sharing faces with the base atoms of a particular nucleotide. The mean and range is reported.

Base Volumes
The base volume is defined as the total volume of all packed Voronoi polyhedra in a particular nt. A packed atom is defined as one that does not share a face with the bounding box.

Atom Volumes
When analyzing at the volumes of atoms in particular bases, many were found to be highly correlated. We identified three atoms from each base that essentially defined the plane of the base and displayed independent volumes. These three atoms are shown for each type of base. The average volume and range are reported.

Atom Packing Probability
The packing probability gives the probability that a particular atom shares no faces with the bounding box. The packing probabilities for the three atoms with uncorrelated atom volumes are reported.

Backbone Volumes
The backbone volume is the sum of the volumes of all unpacked atoms in the backbone of a particular nucleotide. The mean and range is reported.

Unpacked Atoms in Backbone
The total number of unpacked atoms in the backbone is defined as the total number of atoms in the backbone of a particular nucleotide that share faces with the bounding box. The mean and range are reported.

Doublet Library
The doublet library is a Gaussian representation of the Skyes et al. 100-member library. Each member of the 100-member library is considered the center of a Gaussian distribution with 3Å variance for x,y,z and 0.2 variance for θ and ϕ. The score assigned to a pair of nt with parameters (x', y', z', θ', ϕ') for a particular cluster with center(x, y, z, θ, ϕ) is given by the formula

$$g(x',y',z',\theta',\phi') = e^{-\left(\frac{(x-x')^2}{3}+\frac{(y-y')^2}{3}+\frac{(z-z')^2}{3}+\frac{(\theta-\theta')^2}{0.2}+\frac{(\phi-\phi')^2}{0.2}\right)}$$

The mean and range for all scores are reported.

Hydrogen Bonds
Hydrogen bonds parameter is a probabilistic measure of hydrogen bonds for a given nt. The "Overall" parameter describes the average number of hydrogen bonds enjoyed by a given nt regardless of binding partner. The "Paired with" parameters report the average number of hydrogen bonds for a given nt-nt interaction.

Constraint types for guided fragment assembly
W = Watson face of base
H = Hoogsteen face of base
S = Sugar face of base
Base pairs are defined by the hydrogen bonds between each face. For example, W:W means hydrogen bonds are between the Watson faces of both nt in the pair. Planar interactions are defined as bases with an angle between the planes of the bases less than 20 degrees.

-2 No constraints
-1 C1' Distance < 14 A
0 Planar orientation but no hydrogen bonds

Base pairings in CIS
1 Planar with W:W hydrogen bonds
2 Planar with W:H hydrogen bonds
3 Planar with W:S hydrogen bonds
4 Planar with H:W hydrogen bonds
5 Planar with H:H hydrogen bonds
6 Planar with H:S hydrogen bonds
7 Planar with S:W hydrogen bonds
8 Planar with S:H hydrogen bonds
9 Planar with S:S hydrogen bonds

Base pairings in TRANS
10 Planar with W:W hydrogen bonds
11 Planar with W:H hydrogen bonds
12 Planar with W:S hydrogen bonds
13 Planar with H:W hydrogen bonds
14 Planar with H:H hydrogen bonds
15 Planar with H:S hydrogen bonds
16 Planar with S:W hydrogen bonds
17 Planar with S:H hydrogen bonds
18 Planar with S:S hydrogen bonds

Stacking
19 inward stacking
20 outward stacking
21 upward stacking

Availability
Source code and simple documentation have been deposited at:

http://www.github.com/jpbida

The documentation includes demonstrations illustrating the creation of fragment databases from PDB files, generating guided fragment assembly paths, assembling an RNA structure with secondary structure constraints, and using closed moves to minimize against a scoring function.

References

1. Das R & Baker D (2007) *Proc Natl Acad Sci U S A* **104,** 14664-14669.
2. Rycroft CH (2009) *Chaos* **19,** 041111.

Table S1. Fragment library composition descriptive statistics

Position*			Composition†							
1	2	3	RRR	RRY	RYR	RYY	YRR	YRY	YYR	YYY
U	U	U	4812	2976	2747	2035	3091	1843	2086	1941
U	U	P	639	433	190	237	235	227	164	96
U	P	U	166	209	169	69	117	74	206	60
U	P	P	355	414	343	485	216	210	243	232
P	U	U	489	253	273	239	492	296	311	211
P	U	P	92	156	95	304	157	232	167	108
P	P	U	410	300	387	341	284	316	419	327
P	P	P	1775	685	732	757	702	794	772	1743
Total			8738	5426	4936	4467	5294	3992	4368	4718

*Positions 1, 2, and 3 describe the location of the nt from the 5' end of the fragment. U indicates that the nt at that position is not hydrogen bonded to any other nt. P indicates that the nt at that position is hydrogen bonded to another nt.

†Composition describes the pyrimidine (Y) and purine (R) composition of 41,939 RNA trinucleotide fragments in the 5' to 3' direction. The table displays the frequency of fragments with particular sequence compositions and hydrogen bonding patterns (secondary structure).

Table S2. Looping frequencies observed for single-stranded hexanucleotide and heptanucleotide RNA sequences with different pyrimidine (Y) / purine (R) compositions

Frequency N6*

0 RRRYYR	0 YYYRYR	10 RYRRYY	35 RRRRYY
0 RRYYRY	0 YYYYRY	10 YYRRRY	39 RRYRRY
0 RRYYYR	0 YYYYYR	11 RYYRRR	39 YYRYRR
0 RRYYYY	0 YYYYYY	12 RRRRRR	41 RRYRRR
0 RYRRYR	1 RYRRRY	13 RRRYRY	41 YYRYRY
0 RYRYRY	2 RYRRRR	16 YRRYYY	45 RYYRYY
0 RYRYYY	3 RRRYRR	16 YYYYRR	46 RRRYYY
0 RYYYRR	4 YRYYRY	18 YRRYRR	97 YRRYRY
0 RYYYRY	4 YYRRYR	20 RRYRYY	98 YRRRYY
0 RYYYYR	5 YRYRYR	21 RRYYRR	131 YRRRRY
0 RYYYYY	5 YYYRRR	21 YYRRYY	150 YRYRYY
0 YRYYRR	6 YYYRRY	22 RYYRRY	179 YRYRRY
0 YRYYYR	7 RYRYYR	25 YYRRRR	228 YRRYYR
0 YRYYYY	8 RRYRYR	28 RYYRYR	391 RRRRRY
0 YYRYYR	8 RYRYRR	32 YRRRYR	455 YRRRRR
0 YYRYYY	10 RRRRYR	33 YRYRRR	471 YYYRYY

Frequency N7 †

1 RRRYRRY	2 YYRYRYY	7 RYYYRRR	21 YYRYYYY	66 YRYRYYY
1 RYRRRRR	2 YYYRYYR	8 RRRRYYY	22 YRRYRRR	69 YRRRYYR
1 RYRRYRR	3 RRRYYRR	9 RYYRRRY	23 YRRYYYR	75 YRRYRYY
1 RYRYRRR	3 RYYRYRY	10 RYYRYYR	24 YYYRYRR	83 YRRRRRR
1 RYYYRYY	3 YYYRRRY	10 RYYRYYY	25 RRRRRRR	94 YRRYRRY
1 RYYYYYR	3 YYYYYYR	10 YRYYRRY	29 RRRRRRY	99 YRYRYYR
1 YRYYYRR	4 RYRYRRY	11 RRYRYYY	30 YYRRRYR	100 YRYRRYR
1 YYYRRRR	4 RYYYRYR	13 RRYRRYR	32 RYRRYYY	104 YRYRRYY
1 YYYYRRY	4 RYYYYRR	13 YRRYYRY	34 YRYRYRR	109 YRRRRYY
1 YYYYRYR	4 YYRRYRY	13 YRYRYRY	35 YRYRRRY	125 YRRRRYR
1 YYYYYRY	4 YYRRYYY	13 YYYYYRR	40 YRYRRRR	142 YRRRYYY
2 RRRYRYY	5 RYRYYRR	15 YRRYYRR	41 RRRRYRY	157 YRRRYRY
2 RRYYYYY	5 RYYRRRR	16 RRRRYRR	42 RRRRYYR	175 RRYRRYY
2 RYRRRYY	5 RYYYYRY	16 RRRYYRY	43 YYRRRRR	192 RRYRRRY
2 RYYYYYY	5 YYYRYYY	16 YRRYRYR	44 RYRRRYR	218 RRYRRRR
2 YRYYYRY	6 RRYRYYR	17 YRRYYYY	48 YYRRRRY	260 YRRRRRY
2 YYRRRYY	6 RRYYYRY	19 YYRYRRY	57 RRRRRYR	339 RRRRRYY
2 YYRRYRR	6 RYRYYRY	20 RRRYYYR	59 YYRYYRR	551 YYRYYYR
2 YYRYRRR	7 RRYRYRR	20 RYRRRRY	63 YRRRYRR	

* Reports the number of conformations out of 20 million random conformations generated that brought the first and last nt of the indicated RNA hexanucleotide into a plane.

† Reports the number of conformations out of 20 million random conformations generated that brought the first and last nt of the indicated RNA heptanucleotide into a plane.

In both cases random conformations were generated using the 14,400-member fragment library with the only constraint on conformations being avoidance of steric clashes.

Table S3. Summary statistics for scoring metrics evaluated on ribosomes, small RNAs, native structures for FARNA benchmarks, and FARNA generated decoys.

		Nucleotide[¶]			
		G	**A**	**U**	**C**
Graph Order*					
	small RNAs	5.78 (0-21)	5.23 (0-20)	5.10 (0-13)	5.44 (0-11)
	ribosomes	6.31 (1-14)	5.93 (0-14)	6.05 (0-14)	5.96 (0-14)
	benchmarks	5.54 (2-11)	5.06 (0-11)	5.41 (0-9)	5.33 (1-10)
	decoys	5.28 (0-14)	5.31 (0-13)	5.42 (0-13)	5.38 (0-12)
Base Volumes[†]					
	small RNAs	138.5 (0-242.5)	127.0 (0-218.3)	94.28 (0-187.4)	106.7 (0-189.9)
	ribosomes	161 (0-259.2)	148.4 (0-254.0)	114.7 (0-222.5)	121.1 (0-218.7)
	benchmarks	135.2 (0-220.2)	125.1 (0-210)	93.8 (0-169.9)	106.5 (0-171.8)
	decoys	136.6 (0-273.5)	127.5 (0-259.8)	105.4 (0-211.2)	107.5 (0-216.5)
Atom Volumes[‡]					
Atom 1		**N2**	**N3**	**O2**	**O2**
	small RNAs	32.8 (14.319-47.812)	25.6 (13.069-43.155)	28.9 (11.275-46.165)	28.8 (14.732-45.667)
	ribosomes	29.2 (11.829-51.167)	23.5 (9.377-44.372)	27.8 (13.129-49.625)	27.4 (12.867-48.901)
	benchmarks	29.6 (17.444-46.333)	26.4 (15.743-40.417)	30.4 (14.062-44.355)	27.9 (16.564-41.149)
	decoys	26.9 (4.449-49.441)	27.4 (3.488-43.245)	31.6 (7.467-49.224)	30.7 (7.188-49.268)
Atom 2		**O6**	**N6**	**O4**	**N4**
	small RNAs	31.1 (14.858-49.311)	30.0 (14.332-47.561)	32.4 (15.967-47.769)	32.4 (17.929-48.67)
	ribosomes	30.8 (11.377-50.466)	29.8 (14.428-50.385)	30.8 (12.131-49.57)	32.1 (13.051-50.539)
	benchmarks	30.4 (17.457-44.334)	30.5 (14.831-44.594)	27.6 (15.967-39.253)	33.5 (19.591-45.258)
	decoys	28.6 (4.311-50.667)	28.7 (3.879-50.66)	30 (3.066-51.42)	31.3 (3.523-50.676)
Atom 3		**N7**	**N7**	**C5**	**C5**
	small RNAs	30.9 (11.568-45.339)	26.3 (9.865-45.518)	29.3 (8.735-42.466)	29.7 (11.162-43.737)
	ribosomes	29.6 (9.117-46.611)	24.5 (9.597-47.309)	28.1 (9.696-45.722)	28.0 (12.637-44.913)
	benchmarks	29.8 (14.414-43.946)	25.4 (11.971-42.992)	28.8 (15.682-39.817)	29.3 (18.053-41.58)
	decoys	30.8 (3.78-46.646)	27.3 (4.109-46.895)	29.4 (6.39-44.815)	29.5 (5.773-45.001)
Atom Packing Probability[§]					
Atom 1		**N2**	**N3**	**O2**	**O2**
	small RNAs	0.37	0.59	0.51	0.68
	ribosomes	0.65	0.79	0.72	0.77
	benchmarks	0.34	0.51	0.59	0.65
	decoys	0.15	0.29	0.3	0.34
Atom 2		**O6**	**N6**	**O4**	**N4**
	small RNAs	0.47	0.42	0.2	0.29
	ribosomes	0.58	0.64	0.4	0.49
	benchmarks	0.54	0.5	0.19	0.29
	decoys	0.52	0.56	0.4	0.4
Atom 3		**N7**	**N7**	**C5**	**C5**
	small RNAs	0.55	0.72	0.55	0.69
	ribosomes	0.72	0.83	0.74	0.86
	benchmarks	0.62	0.72	0.51	0.7
	decoys	0.64	0.78	0.72	0.74

Table S3 (cont.). Summary statistics for scoring metrics evaluated on ribosomes, small RNAs, native structures for FARNA benchmarks, and FARNA generated decoys.

		Nucleotide[¶]			
		G	**A**	**U**	**C**
Backbone Volumes[‖]					
	small RNAs	103.5 (22.08-235.9)	110.8(21.95-268.9)	111.2(13.25-266.0)	103.4(21.75-268.0)
	ribosomes	143.7 (13.33-306.2)	147.0(25.33-326.6)	141.4(34.96-301.9)	137.9(33.75-307.7)
	benchmarks	97.13 (22.34-197.3)	109.8(21.95-205.2)	93.51(21.82-197.2)	97.04(21.75-198.3)
	decoys	90.87 (22.26-295.9)	96.55(23.95-285)	85.35(13.07-282.7)	85.16(22.13-298.4)
Unpacked Atoms in Backbone[**]					
	small RNAs	5.81 (1-10)	5.51 (0-10)	5.4 (0-11)	5.81 (0-10)
	ribosomes	3.93 (0-11)	3.81 (0-10)	3.99 (0-9)	4.13 (0-9)
	benchmarks	5.93 (2-10)	5.51 (2-8)	5.99 (2-9)	5.93 (2-9)
	decoys	6.58 (0-10)	6.32 (0-10)	6.68 (0-11)	6.7 (0-10)
Doublet Library[††]					
	small RNAs	0.1369 (3.518-0)	0.1571 (2.360-0)	0.0793 (3.498-0)	0.0691 (3.496-0)
	ribosomes	0.2074 (3.640-0)	0.2610 (4.026-0)	0.0974 (3.840-0)	0.1067 (3.750-0)
	benchmarks	0.1024 (2.714-0)	0.1229 (2.360-0)	0.0449 (3.622-0)	0.0725 (2.810-0)
	decoys	0.0684 (3.876-0)	0.0663 (3.556-0)	0.0194 (2.892-0)	0.0161 (3.942-0)
Hydrogen Bonds[‡‡]					
Overall					
	small RNAs	0.717 (0-3.87)	0.445 (0-1.93)	0.523 (0-2.51)	1.31 (0-3.88)
	ribosomes	0.595 (0-3.48)	0.394 (0-1.99)	0.572 (0-2.79)	1.05 (0-3.95)
	benchmarks	0.83 (0-2)	0.532 (0-1.07)	0.832 (0-2.00)	1.49 (0-2.95)
	decoys	0.537 (0-4.42)	0.315 (0-2.86)	0.716 (0-3.99)	1.06 (0-4.05)
Paired with G					
	small RNAs	0.501 (0-3.87)	0.484 (0-1.93)	0.557 (0-2.51)	1.94 (0-3.88)
	ribosomes	0.233 (0-3.48)	0.315 (0-1.99)	0.529 (0-2.79)	1.57 (0-3.95)
	benchmarks	1.29 (0-2)	0.62 (0-1.07)	0.746 (0-1.42)	2.01 (0-2.95)
	decoys	0.288 (0-4.42)	0.452 (0-2.21)	0.828 (0-3.61)	1.55 (0-4.05)
Paired with A					
	small RNAs	0.43 (0-1.58)	0.367 (0-1.92)	1.04 (0-2.17)	0.403 (0-2.93)
	ribosomes	0.328 (0-2.00)	0.216 (0-1.83)	0.852 (0-2.52)	0.157 (0-2.75)
	benchmarks	0.689 (0-1.01)	0.385 (0-0.999)	1.15 (0-2.00)	0.804 (0.385-1.03)
	decoys	0.471 (0-3.07)	0.311 (0-2)	0.947 (0-3.99)	0.528 (0-3.44)
Paired with U					
	small RNAs	0.623 (0-1.92)	0.816 (0-1)	0.118 (0-1.87)	0.000708 (0-0.059)
	ribosomes	0.516 (0-1.77)	0.685 (0-1)	0.265 (0-1.74)	0.0304 (0-1.99)
	benchmarks	0.74 (0-1.00)	0.831 (0-1)	0 (0-0)	8e-04 (0-0.002)
	decoys	0.361 (0-1.96)	0.441 (0-1.15)	0.118 (0-1.94)	0.314 (0-1.99)
Paired with C					
	small RNAs	0.885 (0-1.82)	0.124 (0-1)	0.122 (0-1.96)	0.134 (0-2.21)
	ribosomes	0.79 (0-2.04)	0.102 (0-1.78)	0.0885 (0-1.99)	0.056 (0-2.86)
	benchmarks	0.94 (0-1.02)	0.138 (0-0.976)	0.32 (0-0.985)	0.317 (0-1.00)
	decoys	0.793 (0-3.57)	0.108 (0-2.86)	0.563 (0-2.76)	0.291 (0-4.01)

Summary statistics report the mean and range of the indicated parameters except where indicated below.

*Mean and range of bases that share a Voronoi polyhedron face with another base.

†Mean and range of the sum of volumes (units $Å^3$) of all atoms in a particular base.

‡Mean and range of the volume (units $Å^3$) of particular atoms in each base. Three atoms (Atom 1, Atom 2, Atom3) were chosen based on the independence of their volumes. These three atoms define the plane of each base relative to neighboring bases.

§Probability that a particular atom is contained in a Voronoi polyhedron that does not share a face with the solvent.

¶Summary statistics are reported for each type of nucleotide found in the RNA molecule. The total number of guanines is 2433, 29736, 181, and 91405 for small RNAs, ribosomes, FARNA benchmarks, and FARNA generated decoys respectively. The total number of adenines is 1398, 23297, 111, and 56055 for small RNAs, ribosomes, FARNA benchmarks, and FARNA generated decoys respectively. The total number of uracils is 1445, 17706, 85, and 42925 for small RNAs, ribosomes, FARNA benchmarks, and FARNA generated decoys respectively. The total number of cytosines is 2130, 22091, 144, and 72720 for small RNAs, ribosomes, FARNA benchmarks, and FARNA generated decoys respectively.

‖Mean and range of the sum of all backbone atom volumes (units $Å^3$) for a given nucleotide.

**Total number of backbone atoms that share a face of their Voronoi polyhedron with the bounding box.

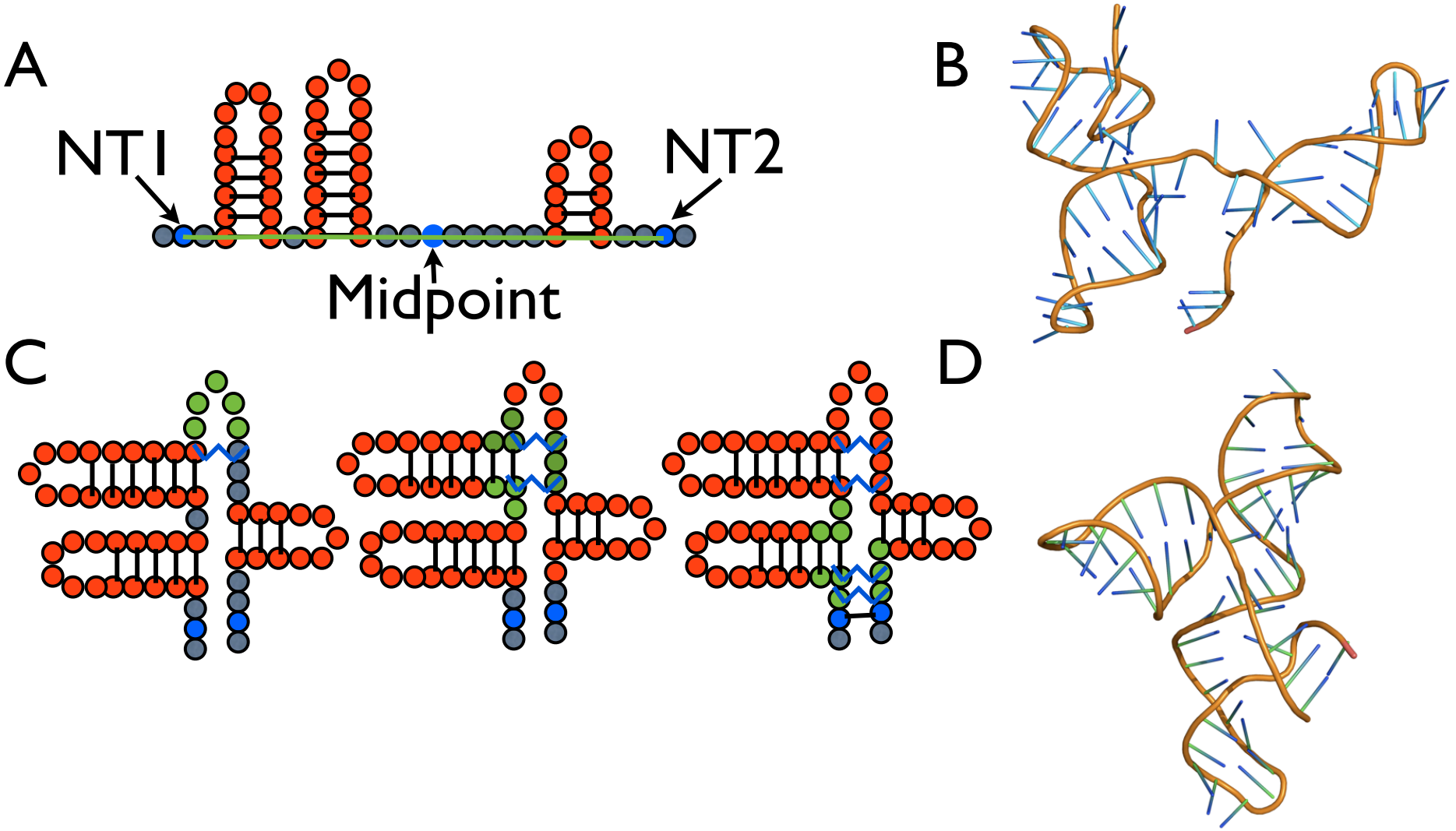
††Mean and range of the sum of all Gaussian distributions representing the 100-member doublet library.

‡‡Mean and range of the probabilistic measure of hydrogen bonds for a given nucleotide. The “Overall” parameter describes the average number of hydrogen bonds enjoyed by a given nucleotide regardless of binding partner. The “Paired with” parameters report the average number of hydrogen bonds for a given nucleotide-nucleotide interaction.

Table S4 Frequencies of Non-Canonical/Canonical (W-W:W-W:W-W) Fragments in PDB

"---"	604129	"W-H:H-W:H-S"	2	"S-H:H-W:W-W"	8	"W-H:W-H:H-W"	107
"S-H:H-S:W-S"	0	"W-H:S-W:W-W"	2	"W-W:S-W:H-S"	8	"S-H:H-H:W-H"	114
"W-S:W-S:W-S"	0	"W-H:W-H:S-H"	2	"H-S:W-W:W-H"	9	"S-H:H-W:H-S"	138
"W-S:W-W:W-S"	0	"W-H:W-W:W-S"	2	"W-H:H-H:W-W"	11	"H-W:H-W:W-W"	157
"H-S:W-H:W-W"	1	"W-W:H-W:H-S"	2	"H-W:W-W:W-H"	13	"H-W:H-S:W-W"	158
"H-S:W-S:W-W"	1	"W-W:H-W:W-H"	2	"W-H:H-W:W-W"	15	"W-W:W-H:W-W"	168
"H-W:H-W:W-S"	1	"W-W:H-W:W-S"	2	"W-W:S-H:W-S"	15	"W-W:W-H:S-H"	170
"H-W:S-H:H-S"	1	"W-W:S-H:S-S"	2	"W-W:S-W:W-W"	15	"W-W:W-H:W-H"	171
"H-W:S-W:W-S"	1	"W-W:W-H:H-H"	2	"W-W:W-S:S-W"	15	"W-S:W-W:W-W"	174
"S-H:S-H:H-W"	1	"W-W:W-H:H-S"	2	"S-H:S-H:H-H"	20	"H-H:W-W:W-W"	211
"S-H:S-H:W-S"	1	"H-W:W-H:W-H"	3	"W-H:S-H:H-S"	20	"W-W:S-H:W-H"	231
"S-W:H-H:H-W"	1	"H-W:W-S:W-W"	3	"W-W:W-S:W-W"	21	"H-S:W-W:S-H"	232
"S-W:H-H:W-W"	1	"S-H:H-S:H-W"	3	"W-W:S-H:W-W"	22	"S-H:W-H:H-S"	232
"S-W:S-W:W-W"	1	"S-W:H-S:H-S"	3	"W-W:H-S:W-W"	26	"S-H:H-S:H-S"	268
"S-W:W-W:S-H"	1	"W-H:W-H:W-W"	3	"S-W:W-W:W-W"	28	"W-H:W-W:W-W"	304
"S-W:W-W:S-W"	1	"W-H:W-W:S-H"	3	"H-W:H-W:H-W"	33	"H-S:H-S:W-W"	346
"S-W:W-W:W-H"	1	"W-W:H-W:S-H"	3	"W-W:W-W:S-S"	33	"S-H:S-H:H-S"	356
"W-H:W-H:W-H"	1	"W-W:S-W:W-H"	3	"H-W:H-S:H-W"	35	"W-W:W-W:S-W"	378
"W-H:W-W:H-W"	1	"W-W:W-S:S-S"	3	"S-W:H-S:W-W"	35	"W-W:S-H:S-H"	435
"W-H:W-W:S-W"	1	"H-S:W-W:H-W"	4	"H-S:H-W:W-W"	42	"W-W:S-H:H-H"	528
"W-S:W-W:S-H"	1	"H-W:W-H:H-S"	4	"W-W:H-W:W-W"	44	"H-H:H-S:W-W"	553
"W-S:W-W:W-H"	1	"S-H:H-W:W-H"	4	"W-H:W-H:H-S"	63	"H-W:W-W:W-W"	647
"W-W:H-H:W-W"	1	"W-H:S-H:H-H"	4	"S-H:H-H:W-W"	66	"S-H:H-S:W-W"	663
"W-W:H-W:H-W"	1	"W-H:W-S:W-W"	4	"S-H:H-W:H-W"	67	"W-W:S-H:H-S"	665
"W-W:S-W:S-H"	1	"H-W:H-W:W-H"	5	"W-W:W-W:H-W"	69	"W-W:S-H:H-W"	810
"W-W:W-H:W-S"	1	"H-W:W-W:S-H"	5	"W-H:H-W:H-W"	79	"W-W:W-W:W-H"	1174
"H-H:W-S:W-W"	2	"S-H:W-W:W-W"	5	"S-S:W-W:W-W"	80	"W-H:H-S:W-W"	1600
"S-H:W-S:W-W"	2	"W-W:S-W:H-W"	5	"W-W:W-W:W-S"	84	"H-S:W-W:W-W"	7094
"S-W:S-H:H-S"	2	"W-W:W-W:H-S"	5	"W-S:W-S:W-W"	88	"W-W:W-W:S-H"	7536
"W-H:H-H:H-W"	2	"W-S:H-S:W-W"	6	"W-W:W-W:H-H"	88	"W-W:W-W:W-W"	110115
		"W-W:W-H:H-W"	6	"W-H:S-H:S-H"	104		

* Reports the number of fragments in non-canonical and canonical helices with particular base pairing interactions (W=Watson, H=Hoogsteen, S=Sugar).



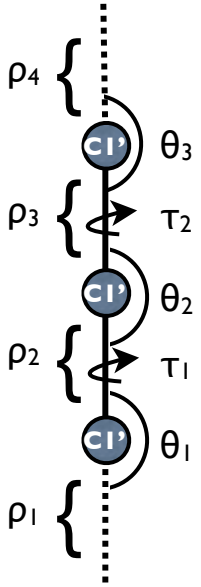
Bida Fig. S1

Virtual Bonds

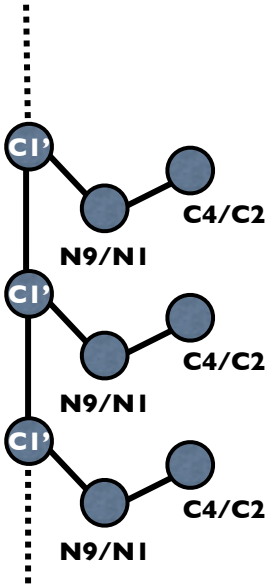
Virtual Bonds +
Base Orientation

All Atom

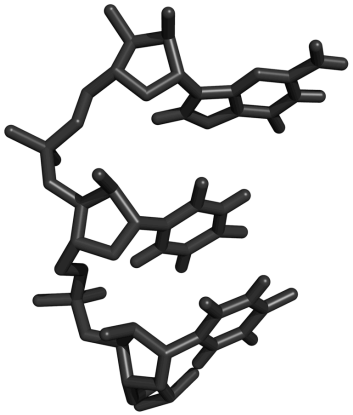
A



B



C

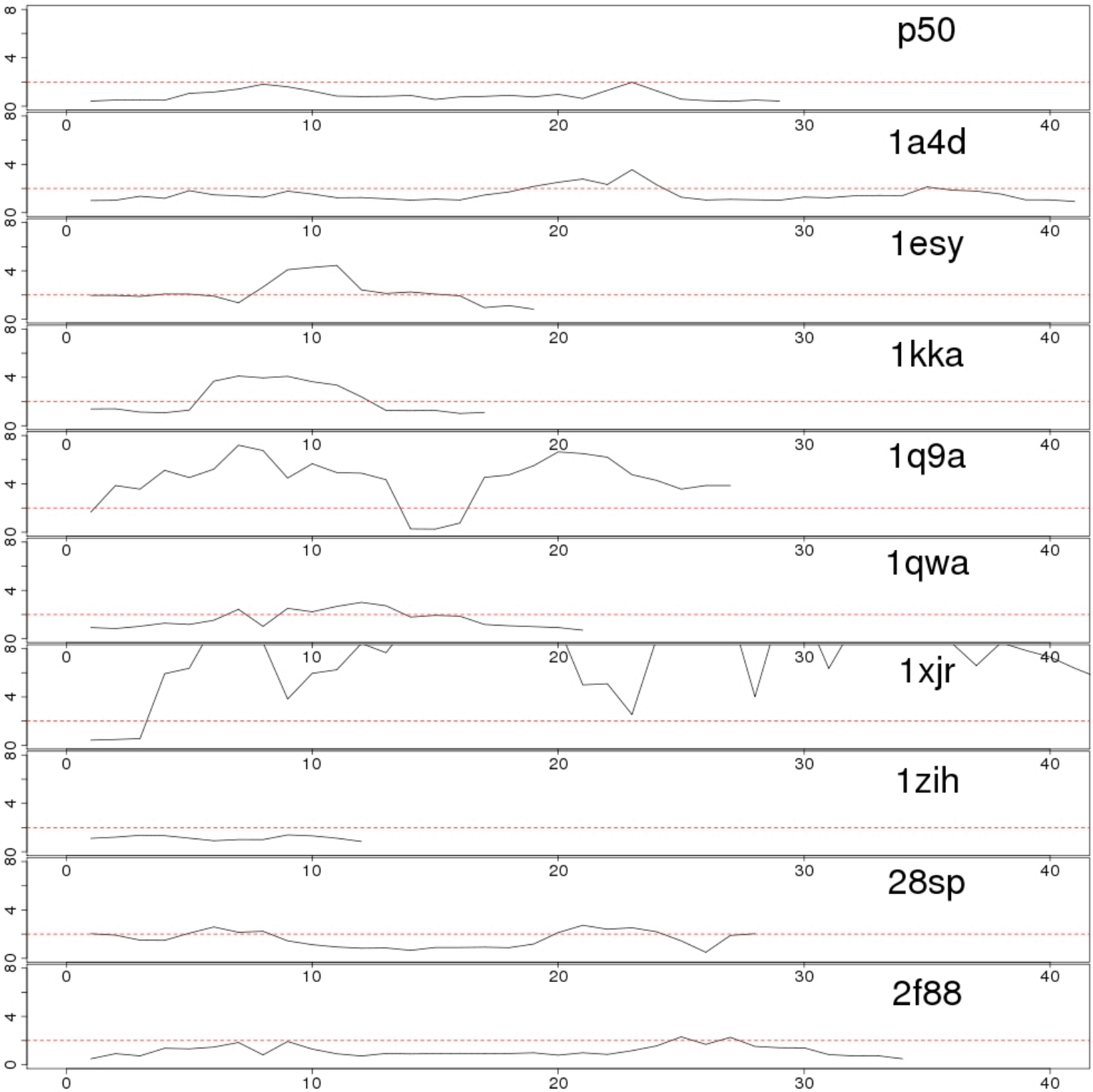


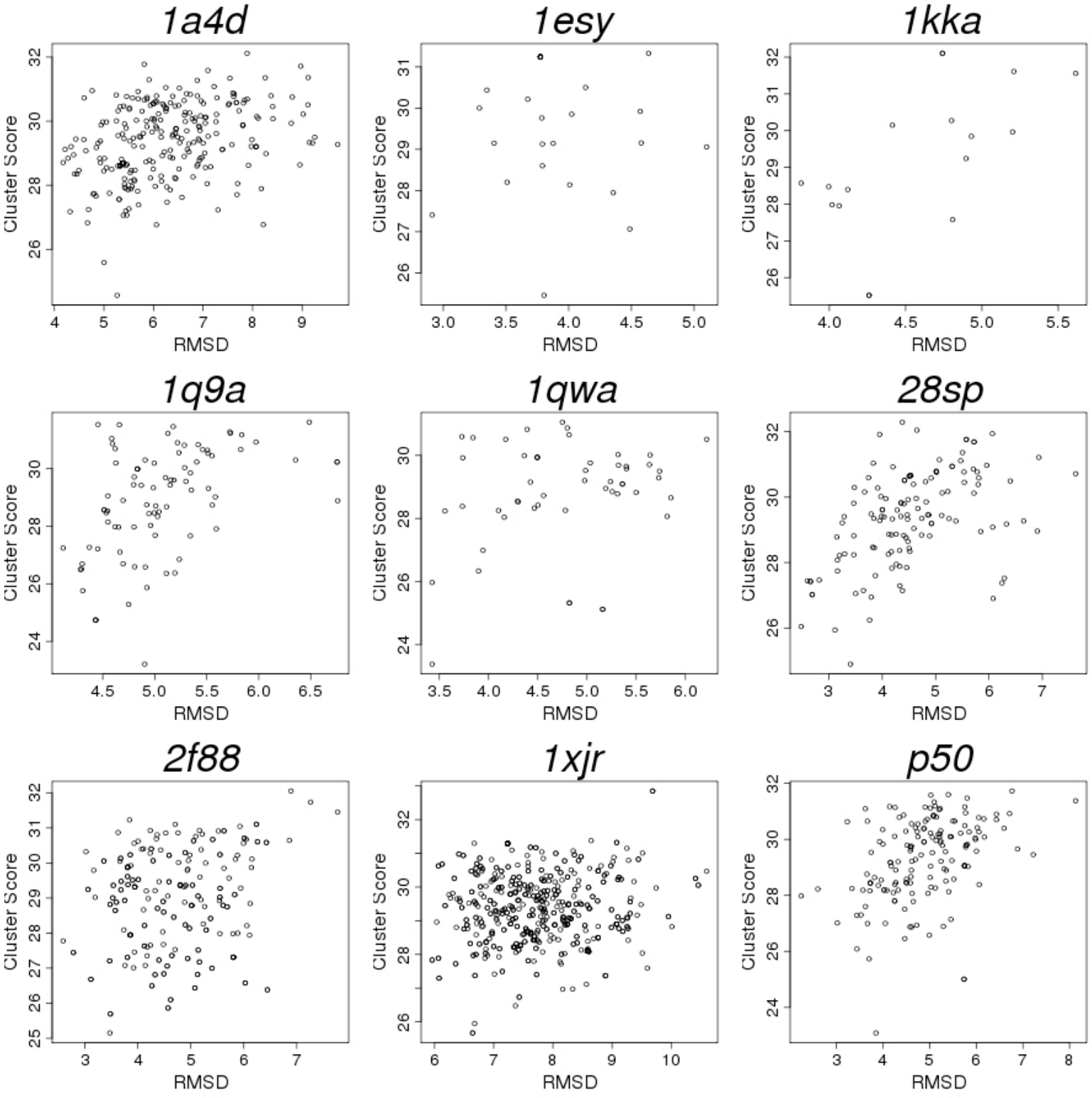
frag_db

frag_nt

frag_at

Bida Fig. S2


p50
1a4d
1esy
1kka
1q9a
1qwa
1xjr
1zih
28sp
2f88
0
10
20
30
40
80
4
8


1a4d
1esy
1kka
1q9a
1qwa
28sp
2f88
1xjr
p50
Cluster Score
RMSD