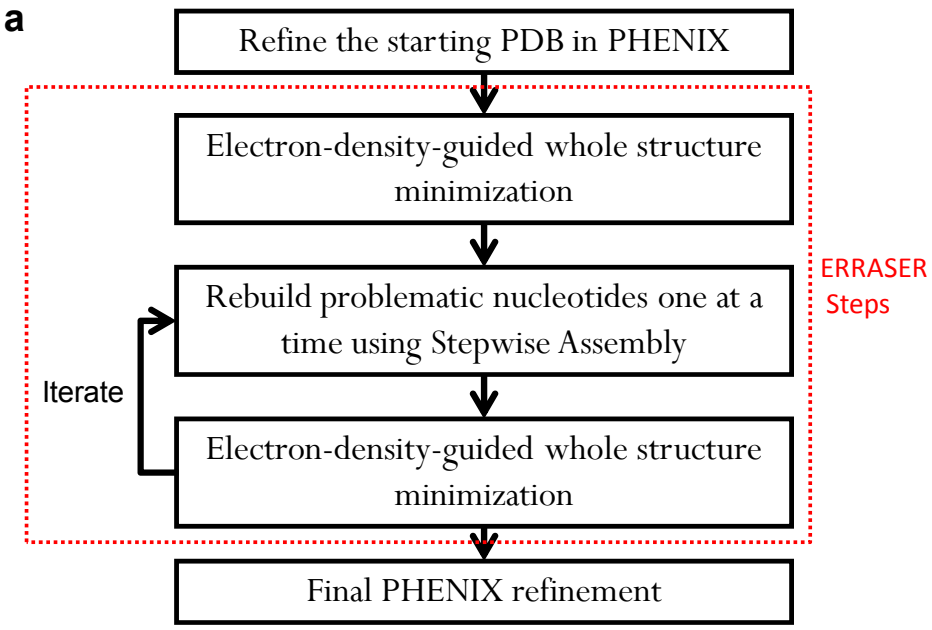
Supplementary Information

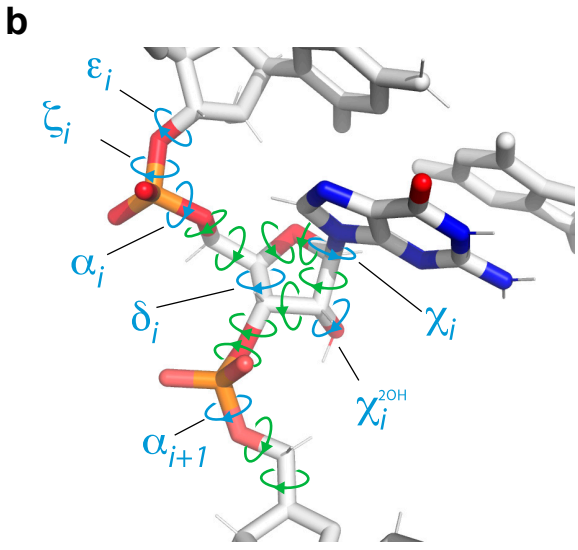
Correcting pervasive errors in RNA crystallography through enumerative structure prediction

Fang-Chieh Chou, Parin Sripakdeevong, Sergey M. Dibrov, Thomas Hermann & Rhiju Das*

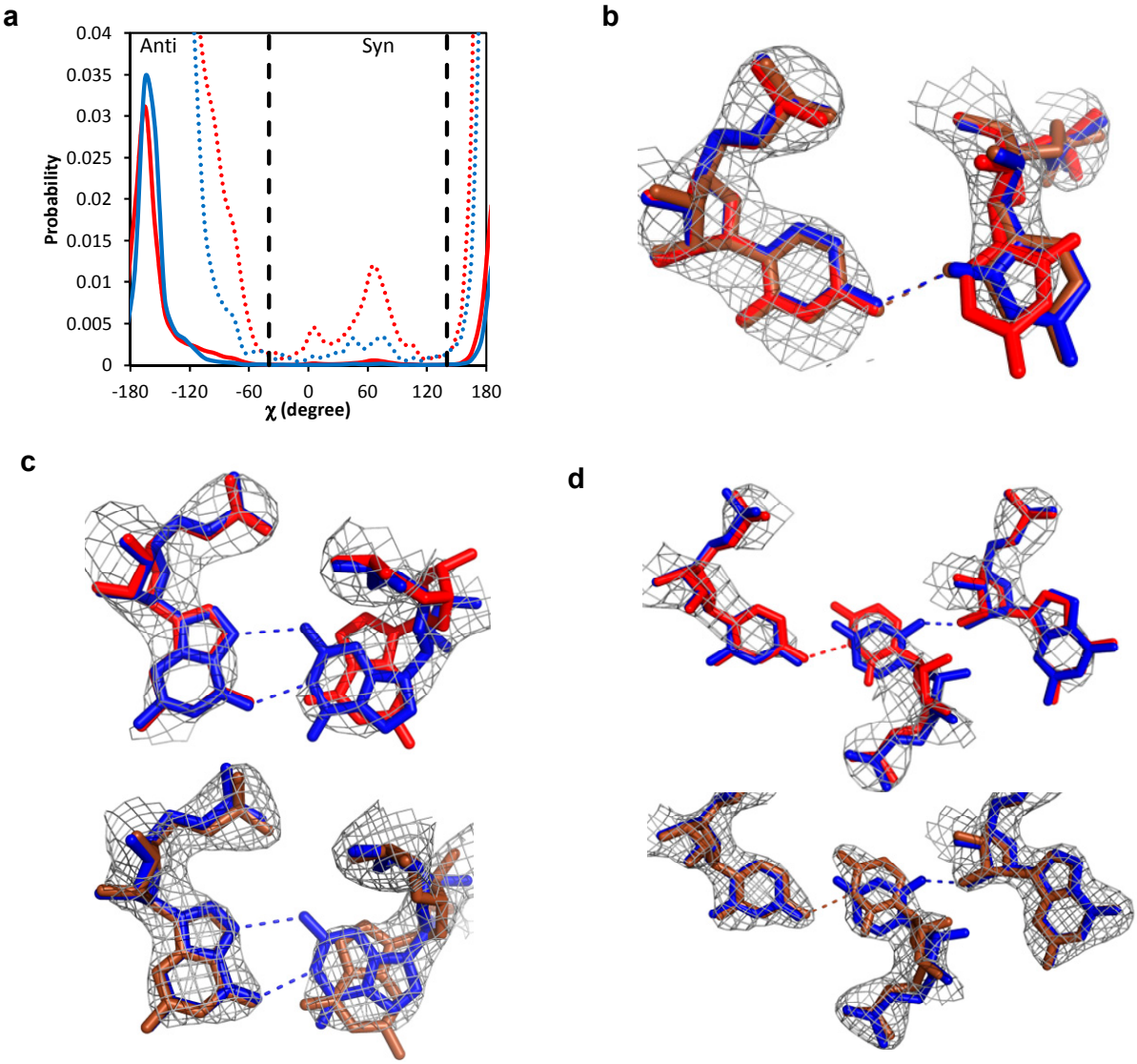
Supplementary Item	**Title**
Supplementary Figure 1	ERRASER-PHENIX algorithm.
Supplementary Figure 2	Base orientation improvements.
Supplementary Table 1	Benchmark test set of 24 structural models, sorted by resolution.
Supplementary Table 2	Outlier bond lengths and angles of the benchmark assessed by phenix.rna_validate.
Supplementary Table 3	Clashscore of the benchmark assessed by phenix.clashscore.
Supplementary Table 4	Outlier backbone rotamers and potentially incorrect sugar puckers of the benchmark assessed by phenix.rna_validate.
Supplementary Table 5	Number of base-pairs identified by MC-Annotate.
Supplementary Table 6	Summary of notable changes of base orientations (χ angle).
Supplementary Table 7	R factors of the benchmark.
Supplementary Table 8	R_{free} factors of the benchmark.
Supplementary Table 9	Similarity analysis between high-resolution and low-resolution models.
Supplementary Table 10	Torsional RMSDs (in degrees) between high-resolution and low-resolution models.
Supplementary Table 11	Similarity analysis for model pairs of the same or similar sequences.
Supplementary Table 12	Torsional RMSDs (in degrees) for model pairs of the same or similar sequences.
Supplementary Table 13	P-values of Wilcoxon signed-rank test between each method and the starting PDB dataset for all geometric features, R and R_{free}.
Supplementary Results	
Supplementary Notes	

Supplementary Figure 1. ERRASER-PHENIX algorithm. (**a**) Flow chart of the whole pipeline, the ERRASER steps are enclosed in red. (**b**) Enumerative RNA modeling in Rosetta steps. The blue torsions are explicitly sampled by enumeration and the green torsions are determined by automatic loop closure.





Supplementary Figure 2. Base orientation improvements. (**a**) Probability distribution of the χ angle in RNA09 (http://kinemage.biochem.duke.edu/databases/rnadb.php). The vertical lines show the range of syn vs. anti. Red: Purine; Blue: Pyrimidine; Dotted lines: 20X zoom-in of the distributions. (**b)** Base orientation changes in 2CKY, chain A, residue U35 Red: PDB model; Blue: ERRASER-PHENIX model; Brown: Reference model. **(c, d)** Base orientation changes in the ribosomal subunit 3OTO for residue **(c)** G266 and **(d)** U365. Red: PDB model; Blue: ERRASER-PHENIX model; Brown: Reference model (2VQE). Upper panel: Plot with electron density from 3OTO. Lower panel: Plot with electron density from 2VQE; the ERRASER model is aligned to the 2VQE model.



Supplementary Table 1. Benchmark test set of 24 structural models, sorted by resolution.

PDB ID	Name	RNA length[a]	Resolution (Å)
2A43	RNA Pseudoknot	26	1.34
3DIL	Lysine Riboswitch	174	1.9
1U8D	Guanine Riboswitch	68	1.95
3D2V	Eukaryotic TPP Riboswitch	77 × 2	2
2GDI	E. coli. TPP Riboswitch	80 × 2	2.05
3TZR	HCV IRES Subdomain IIa	36	2.21
3MXH	c-di-GMP Riboswitch	92	2.3
2PN4	HCV IRES Subdomain IIa	44 × 2	2.32
2QUS	Hammerhead Ribozyme	69 × 2	2.4
1Y27	Guanine Riboswitch	68	2.4
2OIU	L1 Ribozyme Ligase	71 × 2	2.6
2YGH	SAM-I Riboswitch	95	2.6
3DIZ	Lysine Riboswitch	174	2.85
3E5E	SAM-III Riboswitch	53	2.9
2GIS	SAM-I Riboswitch	94	2.9
2CKY	Eukaryotic TPP Riboswitch	77 × 2	2.9
2PN3	HCV IRES Subdomain IIa	44	2.9
3F2Q	FMN Riboswitch	112	2.95
3IWN	c-di-GMP Riboswitch	93 × 2	3.2
3BO3	*Azoarcus* Group I Ribozyme	212	3.4
3R4F	Prohead RNA	66	3.5
3P49	Glycine Riboswitch	169	3.55
1Y0Q	Twort Group I Ribozyme	233	3.6
3OTO	30S Ribosomal Subunit	1,522	3.69

[a] The length of RNA component of the molecule. "× 2" indicates 2 copies in the asymmetric unit.

Supplementary Table 2. Outlier bond lengths and angles of the benchmark assessed by phenix.rna_validate.

PDB ID	Outlier bonds(%)					Outlier angles(%)				
	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
2A43	0	0	0	0	0	0	0	0	0	0
3DIL	0	0	0	0	0	0	0	0	0	0
1U8D	0	0	0	0	0	2.99	0	0	0	0
3D2V	0	0	0	0	0	0	0	0	0	0
2GDI	0	0	0	0	0	0	0	0	0	0
3TZR	0	0	0	0	0	0	0	0	2.78	0
3MXH	0	0	0	0	0	0	0	0	0	0
2PN4	2.38	0	0	0	0	2.38	0	0	0	0
2QUS	1.47	0	0	0	0	3.68	0.74	0	0	0
1Y27	0	0	0	0	0	4.48	0	0	0	0
2OIU	2.11	0	0	0	0	0.7	0	0	0	0
2YGH	3.19	0	0	0	0	2.13	0	0	0	0
3DIZ	0	0	0	0	0	0	0	0	0	0
3E5E	1.92	0	0	0	0	0	0	0	0	0
2GIS	0	0	0	0	0	6.38	0	0	0	0
2CKY	1.3	0	0	0	0	0.65	0	0	0	0
2PN3	0	0	0	0	0	0	0	0	0	0
3F2Q	0	0	0	0	0	0	0	0	0	0
3IWN	0	0	0	0	0	4.3	0	0	0	0
3BO3	0.45	0	0	0	0	0	0	0	0	0
3R4F	0	0	0	0	0	0	0	0	0	0
3P49	0	0	0	0	0	0.59	0	0	0	0
1Y0Q	0	0	0	0	0	0	0	0	0	0
3OTO	0	0.2	0.27	0.07	0	0	0.07	0	0.13	0
Average	0.53	0.01	0.01	0.003	0	1.18	0.03	0	0.12	0
Avg. high-res[a]	0	0	0	0	0	0.60	0	0	0.56	0
Avg. low-res[b]	0.08	0.03	0.05	0.01	0	0.82	0.01	0	0.02	0
Equal to or better than PDB		23 / 24	23 / 24	23 / 24	24 / 24		23 / 24	24 / 24	22 / 24	24 / 24

Bond lengths and angles that have a deviation > 4 σ compared to PHENIX ideal geometry are counted as outliers.
[a] Average value for the five high resolution models 3DIL, 1U8D, 3D2V, 2GDI and 3TZR. Ultra-high resolution dataset 2A43 was excluded.
[b] Average value for the six lowest resolution models 3IWN, 3BO3, 3R4F, 3P49, 1Y0Q and 3OTO.

Supplementary Table 3. Clashscore of the benchmark assessed by phenix.clashscore.

PDB ID	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
2A43	1.19	1.19	2.38	2.38	1.19
3DIL	1.40	6.29	5.94	5.94	5.94
1U8D	14.02	16.22	13.59	12.71	10.08
3D2V	11.43	8.67	8.47	10.84	8.47
2GDI	5.52	8.41	8.98	8.60	6.50
3TZR	14.24	13.40	14.24	14.24	9.21
3MXH	8.48	5.58	6.69	8.93	6.92
2PN4	2.48	5.31	8.85	9.21	6.02
2QUS	12.97	13.43	10.52	12.75	6.71
1Y27	6.82	10.00	6.36	5.91	7.27
2OIU	8.48	7.61	7.61	9.13	5.00
2YGH	6.78	4.53	7.11	5.16	4.85
3DIZ	6.17	7.40	7.05	12.34	7.23
3E5E	4.51	4.53	2.83	9.07	5.09
2GIS	43.14	23.39	22.75	9.29	8.01
2CKY	20.47	12.99	11.22	10.83	7.68
2PN3	9.94	7.81	7.10	7.81	8.52
3F2Q	9.74	7.80	8.91	6.40	5.85
3IWN	55.65	23.31	19.63	13.94	12.71
3BO3	11.05	13.35	15.61	12.56	16.23
3R4F	49.62	16.07	15.12	16.54	10.85
3P49	17.56	5.43	6.57	7.43	3.00
1Y0Q	69.53	20.44	7.96	16.19	2.39
3OTO	41.42	15.75	15.30	14.69	3.17
Average	18.03	10.79	10.03	10.12	7.04
Avg. high-res[a]	9.32	10.60	10.24	10.47	8.04
Avg. low-res[b]	40.81	15.73	13.37	13.56	8.06
Equal to or better than PDB		15 / 24	17 / 24	15 / 24	17 / 24

Clashscore is the number of serious clashes (atom pairs that have more 0.4 Å van der waals overlap) per 1000 atoms.
[a] Average value for the five high resolution models 3DIL, 1U8D, 3D2V, 2GDI and 3TZR. Ultra-high resolution dataset 2A43 was excluded.
[b] Average value for the six lowest resolution models 3IWN, 3BO3, 3R4F, 3P49, 1Y0Q and 3OTO.

Supplementary Table 4. Outlier backbone rotamers and potentially incorrect sugar puckers of the benchmark assessed by phenix.rna_validate.

PDB ID	Outlier backbone rotamers					Potentially incorrect puckers				
	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
2A43	1	1	1	3	2	0	0	0	0	0
3DIL	17	13	14	11	11	4	2	1	0	0
1U8D	4	5	7	3	2	3	2	2	0	0
3D2V	23	25	26	18	23	5	2	2	2	3
2GDI	22	17	16	14	11	8	2	2	3	0
3TZR	5	5	5	3	3	1	1	1	0	0
3MXH	8	9	10	11	11	2	1	1	0	0
2PN4	5	2	1	1	0	1	0	0	0	0
2QUS	35	26	24	18	17	9	6	6	2	0
1Y27	13	10	8	2	2	4	2	2	0	0
2OIU	29	17	17	8	6	6	0	0	0	0
2YGH	8	7	8	5	3	1	0	0	0	0
3DIZ	26	19	17	11	12	5	1	1	0	0
3E5E	6	6	5	4	4	2	1	1	0	0
2GIS	21	16	20	14	6	8	7	7	1	0
2CKY	64	44	37	21	18	18	5	5	2	2
2PN3	7	2	2	2	3	1	0	0	0	0
3F2Q	21	15	15	9	3	3	2	2	2	0
3IWN	52	42	42	29	27	23	12	11	4	2
3BO3	78	70	63	22	19	15	6	6	1	0
3R4F	16	13	17	10	6	5	1	1	1	0
3P49	73	57	60	31	22	19	11	11	8	1
1Y0Q	52	59	58	55	21	20	14	11	8	0
3OTO	279	301	313	253	145	34	35	37	27	0
Average (%)[a]	18.8	15.2	15.3	10.3	7.9	5.0	2.4	2.4	1.0	0.2
Avg. high-res[b]	11.7	11.2	11.9	8.0	7.9	3.6	1.9	1.8	0.6	0.4
Avg. low-res[c]	28.6	25.6	26.4	16.6	10.7	8.1	4.3	4.0	2.3	0.3
Equal to or better than PDB		19 / 24	18 / 24	21 / 24	22 / 24		23 / 24	23 / 24	24 / 24	24 / 24

The backbone rotamer assignment is an effort of RNA Ontology Consortium[1]. Sugar pucker errors are determined using a geometric criterion based on the distance between the glycosidic bond vector (C1′–N1/9) and the following (3′) phosphate[2].

[a] Average error rate as percentage. Obtained by dividing the number of outliers with the total number of nucleotides.

[b] Average value for the five high resolution models 3DIL, 1U8D, 3D2V, 2GDI and 3TZR. Ultra-high resolution dataset 2A43 was excluded. The values are normalized by the numbers of nucleotides in the models.

[c] Average value (normalized) for the six lowest resolution models 3IWN, 3BO3, 3R4F, 3P49, 1Y0Q and 3OTO.

Supplementary Table 5. Number of base-pairs identified by MC-Annotate.

PDB ID	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
2A43	12	12	12	12	12
3DIL	88	89	89	90	91
1U8D	31	32	32	32	32
3D2V	69	68	68	69	68
2GDI	70	70	70	71	70
3TZR	15	15	15	15	15
3MXH	38	40	40	38	40
2PN4	32	32	32	32	32
2QUS	52	52	51	52	52
1Y27	36	34	35	34	36
2OIU	60	63	62	62	64
2YGH	37	37	37	39	43
3DIZ	88	89	89	89	88
3E5E	20	20	20	20	21
2GIS	40	36	38	37	40
2CKY	70	67	67	66	69
2PN3	16	17	16	16	16
3F2Q	43	44	44	44	44
3IWN	64	64	65	65	77
3BO3	86	85	81	85	92
3R4F	25	21	21	21	23
3P49	44	43	46	44	60
1Y0Q	97	89	87	92	104
3OTO	561	593	596	587	692
Average (%)[a]	83.4	82.7	82.7	82.8	87.0
Avg. high-res[b]	91.4	92.0	92.0	92.7	92.4
Avg. low-res[c]	72.2	69.4	69.3	70.1	81.5
Equal to or better than PDB		16 / 24	16 / 24	18 / 24	21 / 24

[a] Average value of (# of base-pairs) / (# of residues) × 2. The normalization is based on that the # of base-pairs in an ideal RNA duplex is half of the # of residues.
[b] Average value calculated for the five high resolution models 3DIL, 1U8D, 3D2V, 2GDI and 3TZR. Ultra-high resolution dataset 2A43 was excluded.
[c] Average value for the six lowest resolution models 3IWN, 3BO3, 3R4F, 3P49, 1Y0Q and 3OTO.

Supplementary Table 6. Summary of notable changes of base orientations (χ angle).

PDB ID	Resolution	Chain-Residue	Base Type	PDB		ERRASER-PHENIX		Reference PDB	Resolution	Chain-Residue	PDB	
				syn/anti	χ	syn/anti	χ				syn/anti	χ
3DIL[a]	1.9	A-110	G	syn	75	anti	-88	NA				
3D2V	2	B-35	U	syn	44	anti	-123	3D2V	2	A-35	anti	-110
2GDI	2.05	X-55	C	anti	-65	anti	-160	NA				
		Y-55	C	syn	52	anti	-135					
2QUS[b]	2.4	A-22	G	syn	97	anti	-78	NA				
		B-23	C	syn	56	anti	-119					
2OIU	2.6	P-19	U	syn	58	anti	-136	2OIU	2.6	Q-19	anti	-112
2YGH	2.6	A-14	A	syn	41	anti	-148	3GX5	2.4	A-9	anti	-85
3DIZ[a]	2.85	A-110	G	syn	80	anti	-85	NA				
2CKY	2.9	A-1	G	syn	-7	anti	174	3D2V	2	A-1	anti	-178
		A-35	U	syn	51	anti	-116			A-35	anti	-110
		A-74	U	syn	36	anti	-167			A-74	anti	-113
		B-67	C	syn	-39	anti	-129			B-67	anti	-100
		B-74	U	anti	-50	anti	-140			B-74	anti	-156
3IWN	3.2	A-7	C	anti	-51	anti	-162	3MXH	2.3	R-17	anti	-155
		A-33	A	anti	-100	syn	82			R-33	syn	67
3BO3	3.4	B-21	C	syn	91	anti	-150	NA				
3P49	3.55	A-731	C	anti	-58	anti	-152	NA				
		A-732	C	syn	-33	anti	-129					
1Y0Q[c]	3.6	A-35	A	anti	150	anti	-77	3BO3	3.4	NA		
		A-113	A	syn	21	anti	-176			B-121	anti	177
		A-158	A	syn	51	anti	-152			B-134	anti	-167
3OTO	3.69	A-91	C	anti	-58	anti	-164	2VQE	2.5	A-91	anti	-161
		A-108	G	syn	-11	anti	180			A-108	syn	3
		A-266	G	anti	-59	syn	54			A-266	anti	-83
		A-328	C	syn	116	anti	-85			A-328	syn	110
		A-346	G	syn	46	anti	-143			A-346	syn	61
		A-365	U	syn	43	anti	-157			A-365	syn	53
		A-421	U	syn	33	anti	-134			A-421	syn	6
		A-839	U	syn	72	anti	-164			A-839	syn	117
		A-1004	A	anti	-87	syn	67			A-1004	anti	-67
		A-1054	C	syn	81	anti	-139			A-1054	syn	77
		A-1279	A	syn	43	anti	-101			A-1279	syn	83

The table shows all the residues with $\Delta\chi > 90^{o}$ upon ERRASER-PHENIX refinement. The definition of syn and anti conformation is: syn: $-40 < \chi \leq 140$; otherwise is anti. See supplementary results for discussion.

[a] 3DIL and 3DIZ are structures of the same sequence with different resolution. However, ERRASER flipped residue 110 in both structure from syn to anti. Therefore we did not perform the high-resolution vs. low-resolution comparison in the table.

[b] For 2QUS, the backbone conformations for residue 22-23 for chain A and chain B, as well as the structures of nearby region, are quite different. Therefore we did not use the different chains as reference models in our analysis.

[c] 1Y0Q was compared to a homologous structure 3BO3. The two structures are group I ribozymes from different species. Residue 35 has no homologous partner in 3BO3 so was not compared.

Supplementary Table 7. *R* factors of the benchmark.

PDB ID	PDB (Deposited)	PDB (PHENIX calculated)	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
2A43	0.114	0.128	0.140	0.140	0.145	0.137
3DIL	0.192	0.183	0.172	0.169	0.171	0.170
1U8D	0.178	0.177	0.166	0.167	0.165	0.162
3D2V	0.207	0.202	0.212	0.210	0.224	0.211
2GDI	0.208	0.199	0.188	0.189	0.196	0.187
3TZR	0.182	0.182	0.184	0.185	0.188	0.180
3MXH	0.202	0.222	0.196	0.196	0.202	0.199
2PN4	0.261	0.258	0.264	0.270	0.268	0.264
2QUS	0.184	0.184	0.190	0.188	0.189	0.186
1Y27	0.232	0.224	0.204	0.205	0.207	0.203
2OIU	0.203	0.196	0.190	0.193	0.196	0.194
2YGH	0.200	0.186	0.197	0.206	0.199	0.205
3DIZ	0.193	0.200	0.198	0.201	0.203	0.202
3E5E	0.222	0.224	0.201	0.199	0.270	0.196
2GIS	0.266	0.249	0.216	0.219	0.218	0.216
2CKY	0.183	0.194	0.180	0.179	0.175	0.176
2PN3	0.229	0.214	0.213	0.212	0.239	0.210
3F2Q	0.200	0.197	0.199	0.202	0.202	0.200
3IWN	0.222	0.218	0.224	0.224	0.239	0.215
3BO3	0.282	0.266	0.244	0.240	0.274	0.239
3R4F	0.239	0.221	0.184	0.182	0.187	0.179
3P49	0.282	0.279	0.227	0.233	0.218	0.233
1Y0Q[a]	0.277	0.265	0.227	0.221	0.225	0.226
3OTO	0.173	0.177	0.163	0.164	0.163	0.186
Average	0.214	0.210	0.199	0.200	0.207	0.199
Avg. high-res[a]	0.193	0.188	0.185	0.184	0.189	0.182
Avg. low-res[b]	0.246	0.238	0.211	0.211	0.218	0.213
Equal to or better than PDB			16 / 24	15 / 24	12 / 24	16 / 24

[a] Average value for the five high resolution models 3DIL, 1U8D, 3D2V, 2GDI and 3TZR. Ultra-high resolution dataset 2A43 was excluded.
[b] Average value for the six lowest resolution models 3IWN, 3BO3, 3R4F, 3P49, 1Y0Q and 3OTO.

Supplementary Table 8. R_{free} factors of the benchmark.

PDB ID	PDB (Deposited)	PDB (PHENIX calculated)	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
2A43	0.190	0.180	0.166	0.165	0.176	0.162
3DIL	0.229	0.213	0.194	0.193	0.199	0.199
1U8D	0.228	0.218	0.198	0.202	0.202	0.198
3D2V	0.251	0.244	0.233	0.235	0.249	0.239
2GDI	0.241	0.229	0.216	0.219	0.219	0.217
3TZR	0.245	0.242	0.239	0.239	0.246	0.240
3MXH	0.239	0.270	0.231	0.233	0.239	0.235
2PN4	0.320	0.323	0.320	0.319	0.323	0.320
2QUS	0.253	0.253	0.235	0.231	0.234	0.225
1Y27	0.264	0.255	0.235	0.238	0.238	0.235
2OIU	0.238	0.233	0.224	0.225	0.225	0.224
2YGH	0.259	0.244	0.234	0.234	0.237	0.236
3DIZ	0.244	0.246	0.239	0.239	0.240	0.235
3E5E	0.259	0.258	0.254	0.256	0.348	0.261
2GIS	0.289	0.270	0.252	0.252	0.256	0.246
2CKY	0.250	0.253	0.236	0.235	0.231	0.234
2PN3	0.283	0.274	0.272	0.273	0.282	0.267
3F2Q	0.243	0.254	0.249	0.255	0.256	0.260
3IWN	0.292	0.287	0.282	0.281	0.297	0.270
3BO3	0.325	0.312	0.295	0.293	0.316	0.295
3R4F	0.271	0.252	0.245	0.241	0.243	0.239
3P49	0.310	0.299	0.290	0.279	0.280	0.295
1Y0Q	0.310	0.307	0.294	0.289	0.292	0.291
3OTO	0.231	0.232	0.225	0.225	0.223	0.233
Average	0.261	0.256	0.244	0.244	0.252	0.244
Avg. high-res[a]	0.239	0.229	0.216	0.218	0.223	0.219
Avg. low-res[b]	0.290	0.281	0.272	0.268	0.275	0.270
Equal to or better than PDB			24 / 24	24 / 24	17 / 24	21 / 24

[a] Average value for the five high resolution models 3DIL, 1U8D, 3D2V, 2GDI and 3TZR. Ultra-high resolution dataset 2A43 was excluded.
[b] Average value for the six lowest resolution models 3IWN, 3BO3, 3R4F, 3P49, 1Y0Q and 3OTO.

Supplementary Table 9. Similarity analysis between high-resolution and low-resolution models.

High res. models	Low res. models	Similar residues (%)[a]					Similar puckers (%)[b]				
		PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
4FE5[c]	1U8D	87.5	87.5	87.5	89.1	92.2	98.4	96.9	96.9	98.4	100.0
4FE5[c]	1Y27	67.2	78.1	78.1	82.8	85.9	93.8	100.0	100.0	100.0	100.0
1U8D	1Y27	68.8	76.6	75.0	84.4	82.8	95.3	96.9	96.9	98.4	96.9
2YGH	2GIS	65.6	66.7	61.3	68.8	81.7	88.2	92.5	92.5	95.7	97.9
3DIL	3DIZ	94.3	93.7	92.5	85.1	94.8	98.9	98.9	98.9	98.9	98.3
3MXH	3IWN_1	53.3	54.6	55.8	61.0	71.4	85.7	92.2	92.2	89.6	94.8
3MXH	3IWN_2	45.5	49.4	48.1	50.7	63.6	81.8	90.9	90.9	89.6	90.9
3D2V_1	2CKY_1	57.1	71.4	72.7	66.2	79.2	85.7	96.1	96.1	94.8	97.4
3D2V_1	2CKY_2	42.9	64.9	67.5	74.0	75.3	89.6	97.4	97.4	97.4	94.8
3D2V_2	2CKY_1	54.6	61.0	62.3	63.6	77.9	92.2	100.0	100.0	98.7	100.0
3D2V_2	2CKY_2	45.5	59.7	64.9	68.8	71.4	90.9	98.7	97.4	97.4	97.4
2PN4_1	2PN3	77.3	84.1	81.8	86.4	86.4	95.5	97.7	97.7	93.2	97.7
2PN4_2	2PN3	84.1	84.1	86.4	81.8	84.1	93.2	95.5	95.5	93.2	95.5
Average		64.9	71.7	71.9	74.1	80.5	91.5	96.4	96.3	95.8	97.0
Equal to or better than PDB			12 / 13	11 / 13	11 / 13	13 / 13		12 / 13	12 / 13	12 / 13	12 / 13

[a] Nucleotide pair in which the differences between all torsion angles are smaller than 40°.
[b] Nucleotide pair in which the difference between δ angle is smaller than 20°.
[c] 4FE5 is a ultra-high resolution (1.32 Å) guanine riboswitch structure not in the benchmark set.

Supplementary Table 10. Torsional RMSDs (in degrees) between high-resolution and low-resolution models.

High res. models	Low res. models	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
4FE5	1U8D	25.1	24.5	25.8	21.5	15.3
4FE5	1Y27	39.3	36.7	33.5	33.2	30.5
1U8D	1Y27	37.3	35.1	32.0	30.3	31.4
2YGH	2GIS	38.4	37.7	38.2	32.2	29.0
3DIL	3DIZ	13.5	12.8	13.4	26.2	16.4
3MXH	3IWN_1	48.0	45.6	44.2	41.3	37.3
3MXH	3IWN_2	51.9	49.5	48.8	48.2	42.2
3D2V_1	2CKY_1	41.0	38.7	37.5	38.7	33.0
3D2V_1	2CKY_2	40.9	37.4	35.1	33.8	35.2
3D2V_2	2CKY_1	42.6	39.5	38.6	38.8	31.5
3D2V_2	2CKY_2	42.7	40.2	39.7	37.2	36.9
2PN4_1	2PN3	25.0	21.7	21.8	23.6	23.8
2PN4_2	2PN3	24.5	22.9	22.7	22.8	23.9
Average		36.2	34.0	33.2	32.9	29.7
Equal to or better than PDB			13 / 13	12 / 13	12 / 13	12 / 13

RMSD is calculated between all the torsion angles in the model pair.

Supplementary Table 11. Similarity analysis for model pairs of the same or similar sequences.

Chain1	Chain 2	Similar residues (%)[a]					Similar puckers (%)[b]				
		PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
2GDI_1	2GDI_2	78.8	86.1	86.1	86.1	86.1	97.5	100.0	100.0	98.7	98.7
2OIU_1	2OIU_2	42.3	53.5	57.8	63.4	69.0	87.3	94.4	94.4	94.4	95.8
2QUS_1	2QUS_2	60.9	75.4	75.4	75.4	81.2	92.8	94.2	94.2	94.2	94.2
3P49_1	3P49_2	50.7	50.7	53.3	58.4	68.8	94.8	97.4	97.4	97.4	92.2
1U8D	1Y27	77.9	77.9	80.5	88.3	87.0	97.4	98.7	98.7	97.4	97.4
2YGH	2GIS	59.7	71.4	72.7	84.4	84.4	90.9	98.7	98.7	98.7	98.7
3DIL	3DIZ	86.4	90.9	88.6	90.9	90.9	97.7	97.7	97.7	97.7	97.7
3MXH	3IWN_1	34.2	39.5	34.2	68.4	73.7	84.2	94.7	92.1	94.7	100.0
3MXH	3IWN_2	68.8	78.1	82.8	85.9	84.4	95.3	100.0	100.0	100.0	100.0
3IWN_1	3IWN_2	65.6	71.0	66.7	71.0	89.3	88.2	93.6	93.6	95.7	97.9
3D2V_1	3D2V_2	94.3	95.4	94.8	92.5	97.7	98.9	100.0	100.0	100.0	100.0
3D2V_1	2CKY_1	53.3	55.8	54.6	64.9	72.7	85.7	92.2	92.2	92.2	96.1
3D2V_1	2CKY_2	45.5	48.1	49.4	55.8	62.3	81.8	92.2	92.2	93.5	90.9
3D2V_2	2CKY_1	57.1	70.1	70.1	77.9	85.7	85.7	98.7	98.7	98.7	98.7
3D2V_2	2CKY_2	42.9	64.9	70.1	79.2	77.9	89.6	100.0	100.0	100.0	97.4
2CKY_1	2CKY_2	54.6	63.6	64.9	79.2	83.1	92.2	100.0	100.0	98.7	98.7
2PN4_1	2PN4_2	45.5	62.3	64.9	77.9	74.0	90.9	98.7	98.7	97.4	100.0
2PN4_1	2PN3	77.3	81.8	79.6	84.1	81.8	95.5	97.7	97.7	93.2	97.7
2PN4_2	2PN3	84.1	84.1	86.4	93.2	86.4	93.2	95.5	95.5	95.5	95.5
Average		62.1	69.5	70.2	77.7	80.9	91.6	97.1	96.9	96.7	97.2
Equal to or better than PDB			19 / 19	19 / 19	18 / 19	19 / 19		19 / 19	19 / 19	18 / 19	18 / 19

[a] Nucleotide pair in which the differences between all torsion angles are smaller than 40°.
[b] Nucleotide pair in which the difference between δ angle is smaller than 20°.

Supplementary Table 12. Torsional RMSDs (in degrees) for model pairs of the same or similar sequences.

Chain 1	Chain 2	PDB	PHENIX	RNABC-PHENIX	RCrane-PHENIX	ERRASER-PHENIX
2GDI_1	2GDI_2	25.0	23.4	23.4	24.3	22.8
2OIU_1	2OIU_2	44.5	41.7	39.8	39.9	34.6
2QUS_1	2QUS_2	46.6	42.3	40.4	32.5	29.0
3P49_1	3P49_2	42.5	41.6	40.8	36.1	38.5
1U8D	1Y27	31.6	31.4	25.3	19.4	22.2
2YGH	2GIS	32.1	31.9	32.0	24.5	23.2
3DIL	3DIZ	19.2	17.5	22.4	17.7	16.4
3MXH	3IWN_1	50.9	49.7	51.7	37.0	32.9
3MXH	3IWN_2	37.3	34.4	26.0	28.2	28.8
3IWN_1	3IWN_2	38.4	37.4	37.5	30.5	22.1
3D2V_1	3D2V_2	13.5	12.0	11.7	19.5	11.7
3D2V_1	2CKY_1	48.0	45.6	44.0	38.0	37.3
3D2V_1	2CKY_2	51.9	49.4	46.7	42.1	42.5
3D2V_2	2CKY_1	41.0	38.5	35.1	31.4	27.6
3D2V_2	2CKY_2	40.9	36.9	30.8	27.3	31.9
2CKY_1	2CKY_2	42.6	39.6	37.8	27.5	24.9
2PN4_1	2PN4_2	42.7	40.3	37.8	27.5	31.0
2PN4_1	2PN3	25.0	21.2	26.7	24.1	22.2
2PN4_2	2PN3	24.5	21.3	20.0	16.5	20.2
Average		36.7	34.5	33.2	28.6	27.4
Equal to or better than PDB			19 / 19	16 / 19	18 / 19	19 / 19

RMSD is calculated between all the torsion angles in the model pair.

Supplementary Table 13. P-values of Wilcoxon signed-rank test between each method and the starting PDB dataset for all geometric features, R and R_{free}.

	Outlier bond	Outlier angle	Clashscore	Outlier backbone rotamer	Potentially incorrect pucker	Number of base-pairs	R	R_{free}
PHENIX	0.017	0.004	0.089	0.001	< 0.001	0.521	0.009	< 0.001
RNABC -PHENIX	0.017	0.005	0.045	0.007	< 0.001	0.394	0.024	< 0.001
RCrane -PHENIX	0.017	0.015	0.136	< 0.001	< 0.001	0.733	0.450	0.014
ERRASER -PHENIX	0.018	0.005	0.006	< 0.001	< 0.001	0.009	0.011	< 0.001

The Wilcocon signed-rank test is a non-parametric statistical test for pairs of related samples. The test can tell whether two datasets are significantly different from each other. Therefore it is suitable for the comparison between the PDB-deposited values and values after improvement methods. Here the two-sided P-value for each comparison is given in the table. All data in the benchmark (n = 24) are used for the analysis. Calculations are performed with the python library SciPy.

Supplementary Results

ERRASER-PHENIX improves RNA base-pairing geometry

ERRASER-PHENIX visually improved the base pairing patterns of the RNA models, enhancing the co-planarity of interacting bases. For example, Figure 1f shows a helical region in 3P49, a structure solved at 3.55 Å resolution. At this resolution, accurate positioning of base planes into the electron density map was difficult. Manual fits gave base pairs that were buckled or twisted compared to geometries seen in higher-resolution crystallographic models[3]. RNABC and RCrane held the base positions fixed during rebuilding and were thus unable to improve the base-pair planarity. On the other hand, ERRASER-PHENIX improved the planarity of the base-pairs, likely due to the hydrogen bonding potential included in the Rosetta energy function. Independent base-pair validation tools – which, like MolProbity, would permit unbiased assessment of improvement – are not currently available. However, we applied the base-pair assignment method MC-Annotate[4] and noted that the refined structures gave a higher number of automatically assigned base-pairs than the starting PDB models in 21 out of 24 cases (Supplementary Table 5). For the 3P49 case, ERRASER-PHENIX increased the number of base-pairs from 44 in the PDB model to 60. Other methods (RNABC-PHENIX and RCrane-PHENIX) lead to smaller improvements in this case, giving 46 and 44 base-pairs respectively.

ERRASER-PHENIX improves the base orientation

In addition to the base-pairing geometry improvement described above, ERRASER also improved the orientation of bases in the models. The glycosidic torsions in a RNA structure predominantly adopt two distinct conformations: syn and anti[5]. These two conformations can be distinguished by the value of χ torsion angle of the glycosidic bond. In the discussion below, the

syn conformation is defined as $-40^{o} < \chi \leq 140^{o}$, and the remaining angle ranges are defined as anti, based on the distribution of χ angles in RNA09 dataset (Supplementary Fig. 2a, http://kinemage.biochem.duke.edu/databases/rnadb.php). It is also evident that the anti conformation is much more probable than syn conformation, and syn pyrimidine conformers are especially rare. Therefore in the current single-nucleotide rebuilding scheme, syn pyrimidine conformers were sampled only if the starting model adopts syn conformation.

Supplementary Table 6 summarized all the notable base orientation changes ($\Delta\chi > 90^{o}$) in the benchmark. While most of the changes are syn-to-anti flips, in agreement with the higher frequency of anti conformations, there are still a few anti-to-syn flips, confirming that ERRASER did not blindly flip the base orientation from syn to anti. To evaluate the confidence of these remodeled bases, these changes were compared to reference models of similar or higher resolution with same or similar sequence. If such models were not available, and there were two different copies of structures in one asymmetric unit, the different copy (to which the target nucleotide did not belong) was used as reference model. For cases where reference models are available, the table shows that for all test cases except the ribosome, all 12 base orientation changes agreed well with the reference. In some cases these orientation changes even introduced extra hydrogen bonds, further supporting these fixes (Supplementary Fig. 2b). For the lowest-resolution ribosome test case (3OTO, 3.69 Å), most of the conformational changes did not match the higher-resolution reference model (2VGE, 2.5 Å). However, since both structures were solved using molecular replacement, it is possible that the two deposited structures inherited the same base orientations from an earlier model[6-7], therefore might share the same erroneous conformations. By detailed inspection of the conformational changes, we found that the ERRASER changes in 3OTO gave the same or more hydrogen bonds as the starting coordinates

and agreed well with the electron density. For example, Supplementary Figure 2c shows an example of such an orientation change where a guanosine flipped from anti to syn and forming a Watson-Crick vs. Hoogsten base-pair with two extra hydrogen bonds. The density map of higher-resolution model did not falsify the possibility of this new, energetically more favorable conformation. On the other hand, Supplementary Figure 2d demonstrates a case where the flipping is ambiguous, where a uridine flipped from syn to anti, and both conformations have one hydrogen bond with nearby nucleotide. Although the starting conformation fits slightly better in the higher-resolution density map, it is possible that the alternative conformation also existed in the crystal structure with a smaller occupancy. Across the benchmark, ERRASER-PHENIX gave improved, or at least alternatively possible, conformations for the base orientation in RNA.

Pairwise comparison for models with similar sequences

Analogous to the independent comparison between remodeled low-resolution and original high-resolution models (see the main text), we reasoned that pairs of models with the same or similar RNA sequences should give similar conformations at corresponding nucleotides, and an accurate refinement procedure should maintain or improve this similarity. We drew 19 such structural pairs from three categories: models of the same sequences determined independently at different resolutions (11 cases, same pairs as those used in high-res vs. low-res comparisons); two copies present in the asymmetric unit related by non-crystallographic symmetry (7 cases); and the conserved regions of two aptamer domains in glycine riboswitch (1 case). Supplementary Table 10-11 summarizes the results of the similarity comparison of the torsion angles of each nucleotide pair, sugar pucker assignment and the RMSD in torsional space for these structure pairs. In nearly all cases, ERRASER-PHENIX improved these metrics compared to the PDB models, and gave better average values than RNABC-PHENIX and RCrane-PHENIX.

Supplementary Notes

Example Rosetta command lines used in ERRASER-PHENIX

(1) Full structure minimization.

```
erraser_minimizer.<exe> -native <input pdb> -out_pdb <output
pdb> -score::weights rna/rna_hires_elec_dens -
score:rna_torsion_potential RNA11_based_new –constrain_P true -
vary_geometry true -fixed_res <fixed residue list> -
edensity:mapfile <map file> -edensity:mapreso 2.0 -
edensity:realign no
```

(2) Single nucleotide rebuilding with analytical chain closure.

```
swa_rna_analytical_closure.<exe> -algorithm rna_resample_test -s
<input pdb> -native <native pdb> -out:file:silent blah.out -
sampler_extra_syn_chi_rotamer true -
sampler_extra_anti_chi_rotamer true -constraint_chi true -
sampler_cluster_rmsd 0.1 -sampler_native_rmsd_screen true -
sampler_native_screen_rmsd_cutoff 3.0 -sampler_num_pose_kept 100
-PBP_clustering_at_chain_closure true -
allow_chain_boundary_jump_partner_right_at_fixed_BP true -
add_virt_root true - rm_virt_phosphate true -sample_res 2 -
cutpoint_closed 2  -fasta fasta -input_res 1 3-4 -fixed_res 1 3-
4 -jump_point_pairs NOT_ASSERT_IN_FIXED_RES 1-4 -alignment_res
1-4 -rmsd_res 4 -score:weights rna/rna_hires_elec_dens -
edensity:mapfile <map file> -edensity:mapreso 2.0 -
edensity:realign no -score:rna_torsion_potential RNA11_based_new
```

(3) Single nucleotide rebuild at terminal nucleotides.

```
swa_rna_main.<exe> -algorithm rna_resample_test -s <input pdb> -
native <native pdb> -out:file:silent blah.out -
sampler_extra_syn_chi_rotamer true -
sampler_extra_anti_chi_rotamer true -constraint_chi true -
sampler_cluster_rmsd 0.1 -sampler_native_rmsd_screen true -
sampler_native_screen_rmsd_cutoff 3.0 -sampler_num_pose_kept 100
-PBP_clustering_at_chain_closure true -
allow_chain_boundary_jump_partner_right_at_fixed_BP true -
add_virt_root true - rm_virt_phosphate true -sample_res 2 -
cutpoint_closed 2 -fasta fasta -input_res 1-4 -fixed_res 2-4 -
jump_point_pairs NOT_ASSERT_IN_FIXED_RES 1-4 -alignment_res 1-4
```

```
-rmsd_res 4 -score:weights rna/rna_hires_elec_dens -
edensity:mapfile <map file> -edensity:mapreso 2.0 -
edensity:realign no -score:rna_torsion_potential RNA11_based_new
```

Example PHENIX command lines used in ERRASER-PHENIX

(1) phenix.ready_set.

```
phenix.ready_set 3E5E.pdb
```

(2) One-cycle TLS refinement.

```
phenix.refine 2QUS-sf.mtz 2QUS.updated.pdb GTP.cif
2QUS.metal.edits 2QUS.link.edits tls.params
main.number_of_macro_cycles=1 strategy=tls
```

(3) Three-cycle refinement with manual parameter set.

```
phenix.refine 2GIS-sf.mtz  2GIS.pdb  2GDI.metal.edits
ordered_solvent=true  optimize_adp_weight=true
strategy=individual_sites+individual_adp+occupancies
main.number_of_macro_cycles=3 wxc_scale=0.1
```

(4) Three-cycle refinement with automatically optimized target weight.

```
phenix.refine 2GIS-sf.mtz
2GIS_phenix_erraser_refine_001_refine_001.pdb
2GIS_phenix_erraser_refine_001_refine_001.metal.edits
ordered_solvent=true optimize_adp_weight=true
strategy=individual_sites+individual_adp+occupancies
main.number_of_macro_cycles=3 optimize_xyz_weight=true
```

(5) Density map creation.

```
phenix.maps maps.params

maps.params:
...
   map {
     map_type = 2mFo-DFc
     format = xplor *ccp4
     file_name = 2GIS_cell.ccp4
```

```
      kicked = True
      fill_missing_f_obs = True
      grid_resolution_factor = 1/4.
      region = selection *cell
      atom_selection = None
      atom_selection_buffer = 3
      sharpening = False
      sharpening_b_factor = None
      exclude_free_r_reflections = True
      isotropize = True
   }
...
```


1. Richardson, J.S. et al. RNA backbone: Consensus all-angle conformers and modular string nomenclature (an RNA Ontology Consortium contribution). *RNA* **14**, 465-481 (2008).
2. Chen, V.B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst. D* **66**, 12-21 (2010).
3. Stombaugh, J., Zirbel, C.L., Westhof, E. & Leontis, N.B. Frequency and isostericity of RNA base pairs. *Nucleic Acids Res.* **37**, 2294-2312 (2009).
4. Gendron, P., Lemieux, S. & Major, F. Quantitative analysis of nucleic acid three-dimensional structures. *J. Mol. Biol.* **308**, 919-936 (2001).
5. Bloomfield, V.A. et al. Nucleic Acids: Structures, Properties, and Functions. (University Science Books, 2000).
6. Demirci, H. et al. Modification of 16S ribosomal RNA by the KsgA methyltransferase restructures the 30S subunit to optimize ribosome function. *RNA* **16**, 2319-2324 (2010).
7. Kurata, S. et al. Modified Uridines with C5-methylene Substituents at the First Position of the tRNA Anticodon Stabilize U·G Wobble Pairing during Decoding. *J. Biol. Chem.* **283**, 18801-18811 (2008).