RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures

ZHICHAO MIAO,[1] RYSZARD W. ADAMIAK,[2] MARC-FRÉDÉRICK BLANCHET,[3] MICHAL BONIECKI,[4] JANUSZ M. BUJNICKI,[4,5] SHI-JIE CHEN,[6] CLARENCE CHENG,[7] GRZEGORZ CHOJNOWSKI,[4] FANG-CHIEH CHOU,[7] PABLO CORDERO,[7] JOSÉ ALMEIDA CRUZ,[1] ADRIAN R. FERRÉ-D'AMARÉ,[8] RHIJU DAS,[7] FENG DING,[9] NIKOLAY V. DOKHOLYAN,[10] STANISLAW DUNIN-HORKAWICZ,[4] WIPAPAT KLADWANG,[7] ANDREY KROKHOTIN,[10] GRZEGORZ LACH,[4] MARCIN MAGNUS,[4] FRANÇOIS MAJOR,[3] THOMAS H. MANN,[7] BENOÎT MASQUIDA,[11] DOROTA MATELSKA,[4] MÉLANIE MEYER,[12] ALLA PESELIS,[13] MARIUSZ POPENDA,[2] KATARZYNA J. PURZYCKA,[2] ALEXANDER SERGANOV,[13] JULIUSZ STASIEWICZ,[4] MARTA SZACHNIUK,[14] ARPIT TANDON,[10] SIQI TIAN,[7] JIAN WANG,[15] YI XIAO,[15] XIAOJUN XU,[6] JINWEI ZHANG,[8] PEINAN ZHAO,[6] TOMASZ ZOK,[14] and ERIC WESTHOF[1]

[1]Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de biologie moléculaire et cellulaire du CNRS, 67000 Strasbourg, France
[2]Department of Structural Chemistry and Biology of Nucleic Acids, Structural Chemistry of Nucleic Acids Laboratory, Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Poland
[3]Institute for Research in Immunology and Cancer (IRIC), Department of Computer Science and Operations Research, Université de Montréal, Montréal, Québec, Canada H3C 3J7
[4]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland
[5]Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland
[6]Department of Physics and Astronomy, Department of Biochemistry, and Informatics Institute, University of Missouri-Columbia, Columbia, Missouri 65211, USA
[7]Department of Physics, Stanford University, Stanford, California 94305, USA
[8]National Heart, Lung and Blood Institute, Bethesda, Maryland 20892-8012, USA
[9]Department of Physics and Astronomy, College of Engineering and Science, Clemson University, Clemson, South Carolina 29634, USA
[10]Department of Biochemistry and Biophysics, University of North Carolina, School of Medicine, Chapel Hill, North Carolina 27599, USA
[11]Génétique Moléculaire Génomique Microbiologie, Institut de physiologie et de la chimie biologique, 67084 Strasbourg, France
[12]Institut de génétique et de biologie moléculaire et cellulaire, 67400 Strasbourg, France
[13]Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York 10016, USA
[14]Poznan University of Technology, Institute of Computing Science, 60-965 Poznan, Poland
[15]Department of Physics, Huazhong University of Science and Technology, 430074 Wuhan, China

ABSTRACT

This paper is a report of a second round of RNA-Puzzles, a collective and blind experiment in three-dimensional (3D) RNA structure prediction. Three puzzles, Puzzles 5, 6, and 10, represented sequences of three large RNA structures with limited or no homology with previously solved RNA molecules. A lariat-capping ribozyme, as well as riboswitches complexed to adenosylcobalamin and tRNA, were predicted by seven groups using RNAComposer, ModeRNA/SimRNA, Vfold, Rosetta, DMD, MC-Fold, 3dRNA, and AMBER refinement. Some groups derived models using data from state-of-the-art chemical-mapping methods (SHAPE, DMS, CMCT, and mutate-and-map). The comparisons between the predictions and the three subsequently released crystallographic structures, solved at diffraction resolutions of 2.5–3.2 Å, were carried out automatically using various sets of quality indicators. The comparisons clearly demonstrate the state of present-day de novo prediction abilities as well as the limitations of these state-of-the-art methods. All of the best prediction models have similar topologies to the native structures, which suggests that computational methods for RNA structure prediction can already provide useful structural information for biological problems. However, the prediction accuracy for non-Watson–Crick interactions, key to proper folding of RNAs, is low and some predicted models had high Clash Scores. These two difficulties point to some of the continuing bottlenecks in RNA structure prediction. All submitted models are available for download at http://ahsoka.u-strasbg.fr/rnapuzzles/.

Keywords: 3D prediction; bioinformatics; force fields; X-ray structures; models; structure quality

Corresponding author: e.westhof@ibmc-cnrs.unistra.fr

Article published online ahead of print. Article and publication date are at http://www.rnajournal.org/cgi/doi/10.1261/rna.049502.114. Freely available online through the *RNA* Open Access option.

INTRODUCTION

More than 100,000 structures are currently available in the Protein Data Bank (PDB) (Berman et al. 2000); however, RNA-containing structures take up <6% of these depositions, including RNA structures complexed with other molecules. Although protein related structures constitute >90% of the structure database, <1/1000th of the proteins with known sequences have experimental structures available (Moult 2008). Given the vast number of noncoding RNA molecules being discovered in cells and viruses, it is likely that a very small part of the RNA conformational space has been structurally characterized. RNA structure determination efforts still have a long way to go, and computational modeling could play a major role in providing structural insights for various biological problem explorations.

"RNA-Puzzles" is a CASP-like (Moult et al. 2014) collective blind experiment for the evaluation of three-dimensional (3D) RNA structure prediction. The primary aims of RNA-Puzzles are (i) to determine the capabilities and limitations of current methods of 3D RNA structure prediction based on sequence, (ii) to find whether and how progress has been made, as well as what has yet to be done to achieve better solutions, (iii) to identify whether there are specific bottlenecks that hold back the field, (iv) to promote the available methods and guide potential users in the choice of suitable tools for real-world problems, and (v) to encourage the RNA structure prediction community in their efforts to improve the current tools and to make automated prediction tools available. Until now, 12 puzzles have been set up and assessments of three puzzles were previously published (Cruz et al. 2012).

We now report a second round focusing on the prediction of large RNA structures, a lariat-capping ribozyme (formerly named GIR1), an adenosylcobalamin-binding riboswitch, and a T-box–tRNA complex (Peselis and Serganov 2012; Zhang and Ferre-D'Amare 2013; Meyer et al. 2014). No closely homologous structures existed in structure databases at the time of the experiment, except for the Azoarcus group I intron (Adams et al. 2004) as a potential template for the catalytic core of the GIR1 ribozyme, templates for the tRNA, and crystallographic structure of a segment of a T-box RNA without the tRNA (Wang et al. 2010; Grigg et al. 2013). This round of prediction focuses on (i) the automatic assessment of de novo prediction of large RNA structures, especially structure topology, (ii) the evaluation of the contribution of simple and fast experimental data in structure prediction, such as chemical probing data, and (iii) the identification of bottlenecks in modeling 3D interactions. The ultimate aim is to derive force fields and programming systems allowing for automatic folding of RNA sequences in three-dimensional. However, at this stage, the assessment does not make a distinction between those groups deriving models based solely on ab initio predictions from those incorporating experimental data like chemical probing. As a matter of fact, RNA-Puzzles led to the development of automatic production and retrieval of solution data (see Kladwang et al. 2011a,b, 2014).

For the three puzzles, the best RMSDs range between 6.8 and 11.7 Å, and all display similar topologies to the native structures. Given the sizes of the RNAs (>160 nt), this is a very positive trend for de novo structure modeling. The best models always show much better prediction of non-Watson–Crick interactions but also, surprisingly, relatively high clash scores. This reemphasizes the importance of non-Watson–Crick interactions for RNA 3D structure modeling as well as the difficulty of predicting such interactions on the basis of RNA secondary structure even when complemented with chemical probing data. The observed atomic clashes, possibly due to the inclusions of experimental constraints for nucleotide contacts in the prediction without adequate optimization, have led to further experiments and insight toward better solutions, discussed below.

THE THREE RNA PUZZLES

Problem 5: the lariat-capping ribozyme

The lariat-capping ribozyme represents an individual family of ribozymes that has evolved specific architectural features from a group I intron ancestor (Meyer et al. 2014). The LC ribozyme catalyzes a distinct reaction involving formation of a 3-nt 2′,5′ lariat. The 188-nt long sequence is the following:

5′-GGUUGGGUUGGGAAGUAUCAUGGCUAAUCA
CCAUGAUGCAAUCGGGUUGAACACUUAAUU
GGGUUAAAACGGUGGGGGACGAUCCCGUAAC
AUCCGUCCUAACGGCGACAGACUGCACGGC
CCUGCCUCUUAGGUGUGUUCAAUGAACAGU
CGUUCCGAAAGGAAGCAUCCGGUAUCCCAAG
ACAAUC-3′

The crystal structure was resolved to 2.45 Å resolution. Two crystallographic models became available after modeling, with PDB ID's 4P95 and 4P9R.

Problem 6: the adenosylcobalamin riboswitch

An adenosylcobalamin riboswitch was crystallized (Peselis and Serganov 2012). The 168 nt adenosylcobalamin riboswitch consists of a ligand-bound structured core and a bent peripheral domain. The sequence is the following:

5′-CGGCAGGUGCUCCCGACCCUGCGGUCGGGA
GUUAAAAGGGAAGCCGGUGCAAGUCCGGCAC
GGUCCCGCCACUGUGACGGGGAGUCGCCCCU
CGGGAUGUGCCACUGGCCCGAAGGCCGGGAA
GGCGGAGGGGCGGCGAGGAUCCGGAGUCAGG
AAACCUGCCUGCCG-3′

The crystal structure (PDB 4GXY) has a resolution of 3.05 Å. An adenosylcobalamin molecule is given in the crystal structure but was not revealed at the start of the puzzle.

Problem 10: the T-box–tRNA complex structure

A T-box–tRNA complex structure was solved (Zhang and Ferré-D'Amaré 2013). The sequence of the 96 nt T-box is as follows:

5′-UGCGAUGAGAAGAAGAGUAUUAAGGAUUUAC
UAUGAUUAGCGACUCUAGGAUAGUGAAAGCU
AGAGGAUAGUAACCUUAAGAAGGCACUUCGAG
CA-3′

The sequence of tRNA is the following (75 nt):

5′-GCGGAAGUAGUUCAGUGGUAGAACACCACCUU
GCCAAGGUGGGGGUCGCGGGUUCGAAUCCCGU
CUUCCGCUCCA-3′

The structure of the complex was solved at a resolution of 3.20 Å (PDB 4LCK). The crystallized sequence was slightly different (the acceptor region was engineered in tRNA), but this detail of the crystal structure was not disclosed in the puzzle. Several RNA modules, including a K-turn, a G-bulge, a double T-loop and an anticodon loop, appeared in this complex structure.

Additional chemical-mapping data

The Das group provided chemical-mapping data on the three puzzles to all the modelers. One-dimensional chemical-mapping data and mutate-and-map (M^2) data were acquired, quantitated, and normalized as described in Kladwang et al. (2014) and Seetin et al. (2014), respectively. Three probes were used: 1M7 (a SHAPE reagent, 1-methyl-7-nitroisatoic anhydride, which acylates 2′-hydroxyls of flexible nucleotides); DMS (dimethyl sulfate, reacting with exposed N1/N3 of adenosine/cytosine; and CMCT (1-cyclohexyl(2-morpholinoethyl) carbodiimide metho-*p*-toluene sulfonate, reacting with exposed N1/N3 of guanosine/uracil) ([Mortimer and Weeks 2007; Cordero et al. 2012a] and references therein). Data were released to modelers on the RNA Mapping Database in standardized formats (Rocca-Serra et al. 2011; Cordero et al. 2012b) in accession codes RNAPZ5_STD_0000, RNAPZ5_1M7_0002, RNAPZ5_DMS_0002; RNAPZ6_STD_0001, RNAPZ6_1M7_0002; RNAPZ10_STD_0001, RNAPZ10_STD_0002. Each group was given the possibility to use those data and each group describes below at which stage and how these solution data were used during the modeling process.

OVERALL COMPARISON RESULTS

Assessment methods

The automatic model assessment methods were the same as previously used in RNA-Puzzles (Cruz et al. 2012). To geometrically compare predicted models with the experimental structures, we used the Root Mean Square Deviation (RMSD) measure, the Deformation Index (DI), and the complete Deformation Profile matrix (DP) which provides an evaluation of the predictive quality of a model at multiple scales (Parisien et al. 2009). The Clash Score as evaluated by MolProbity is also used as a control measurement for the quality of the geometric parameters of the models (Chen et al. 2010). Additionally, MCQ (Mean of Circular Quantities) score (Zok et al. 2014) was added as a reference to assess prediction in terms of torsion angle space. MCQ measures the dissimilarity between structures taking into account rotatable bonds and sugar pucker. Due to its sensitivity to local differences and independence from structural alignment, it may serve as a complement to methods based on atom coordinates. A single distortion, which can significantly increase global RMSD, influences only distorted residues in case of MCQ. On the other hand, numerous irregularities that sharpen the backbone may cancel out when RMSD is considered, but are revealed in torsion angle space. An implementation of MCQ score is publicly available for download under http://www.cs.put.poznan.pl/tzok/mcq. It allows for several usage scenarios, among which the *global* option was used to assess models in RNA-Puzzles Round II. For each pair consisting of the target and the predicted structure, MCQ-*global* provides a single distance score, representing their mean dissimilarity. Its value was computed upon the differences between the corresponding sugar pseudorotation angle (P) and seven dihedral angles defined for a residue (α, β, γ, δ, ε, ζ, and χ). The final rank was built to grasp the overall resemblance of models to the target structure in terms of their trigonometric representation. The MCQ score ranges between 0° and 180°. MCQ-*local* computes raw differences between particular dihedral angles, thus being sensitive even to the smallest discrepancy and allows the observation of high dissimilarity at the residue level. MCQ-*global* introduces an inevitable bias, since the information about single distortions gets lost during the averaging. Therefore an interpretation of the global score should take into account structure size. For large RNAs, *global* MCQ <15° indicates high similarity of structures, while *global* MCQ >45° indicates an overall dissimilarity.

Problem 5: the lariat-capping ribozyme

A total of 25 predicted models were submitted with RMSDs ranging from 9.15 to 36.5 Å (mean RMSD is 24.1 Å, see Table 1). The top two models are better than the others in terms of RMSD, Deformation Index, non-Watson–Crick interactions and stacking (Fig. 1). The top three models also have >90% Watson–Crick (WC) base pairs correctly predicted. Most of the groups have, however, submitted models with very low accuracy in non-WC interactions. The last three models ranked by RMSD have similarly poor levels of accuracy for non-WC interactions, with worse prediction of WC pairs as well as worse stacking predictions. Several models present

TABLE 1. Summary of the results for Puzzle 5

Group[a]	Number[b]	RMSD[c]	Rank[d]	DI all[e]	Rank[d]	INF all[f]	Rank[d]	INF wc[g]	Rank[d]	INF nwc[h]	Rank[d]	INF stack[i]	Rank[d]	Clash Score[j]	Rank[d]	MCQ[k]	Rank[d]
Das	2	9.15	1	12.02	1	0.76	1	0.91	3	0.33	1	0.75	2	6.79	12	10.57	2
Das	1	9.95	2	13.15	2	0.76	2	0.92	1	0.26	2	0.75	1	9.44	21	11.63	4
Adamiak	1	16.78	3	23.96	3	0.70	13	0.90	5	0.00	5	0.66	20	11.10	23	13.30	13
Dokholyan	2	19.90	4	29.14	6	0.68	19	0.84	13	*	*	0.66	22	7.62	14	13.87	17
Dokholyan	4	19.90	5	29.14	7	0.68	20	0.84	14	*	*	0.66	23	7.62	15	13.87	18
Bujnicki	2	20.77	6	28.09	4	0.74	3	0.86	9	0.00	11	0.74	5	1.66	8	12.14	6
Dokholyan	6	20.82	7	28.65	5	0.73	4	0.91	2	0.00	7	0.69	12	10.43	22	15.81	23
Dokholyan	7	21.44	8	30.61	8	0.70	14	0.91	4	0.00	8	0.66	19	8.77	19	14.63	22
Bujnicki	4	22.24	9	32.80	11	0.68	21	0.79	22	0.00	13	0.68	15	1.16	6	10.34	1
Bujnicki	3	22.37	10	36.99	16	0.61	25	0.72	24	0.00	12	0.62	25	5.30	10	12.75	12
Dokholyan	5	23.01	11	32.55	9	0.71	11	0.87	8	*	*	0.68	17	6.46	11	13.68	16
Dokholyan	3	23.69	12	32.64	10	0.73	5	0.88	7	0.00	6	0.71	10	8.61	18	14.61	21
Bujnicki	5	23.81	13	34.15	13	0.70	16	0.86	10	0.00	14	0.68	16	0.66	5	11.73	5
Dokholyan	8	24.01	14	33.84	12	0.71	10	0.89	6	0.00	9	0.67	18	9.11	20	14.40	20
Bujnicki	1	24.78	15	36.69	15	0.68	22	0.79	21	0.00	10	0.69	13	1.32	7	11.23	3
Chen	2	25.67	16	36.33	14	0.71	12	0.80	20	0.00	16	0.73	6	7.78	16	13.36	14
Chen	7	26.26	17	39.50	19	0.67	24	0.76	23	0.00	21	0.69	14	0.00	3	13.39	15
Dokholyan	1	27.20	18	38.03	17	0.72	8	0.84	12	*	*	0.70	11	7.45	13	14.33	19
Chen	1	27.24	19	38.90	18	0.70	15	0.81	18	0.00	15	0.71	9	4.64	9	12.19	7
Chen	3	28.71	20	41.69	20	0.69	18	0.71	25	0.00	17	0.74	4	11.42	24	12.35	10
Chen	5	31.39	21	44.06	22	0.71	9	0.81	19	0.00	19	0.73	7	13.41	25	12.19	8
Chen	4	31.52	22	43.50	21	0.73	6	0.83	15	0.00	18	0.74	3	8.28	17	12.25	9
Chen	6	31.55	23	44.07	23	0.72	7	0.83	16	0.00	20	0.73	8	0.00	2	12.47	11
Xiao	2	32.53	24	48.54	24	0.67	23	0.83	17	0.20	4	0.64	24	0.17	4	19.90	25
Xiao	1	36.54	25	52.90	25	0.69	17	0.86	11	0.21	3	0.66	21	0.00	1	19.71	24
Mean		24.05		34.48		0.70		0.84		0.05		0.69		5.97		13.47	
Standard deviation		4.91		7.11		0.03		0.05		0.06		0.03		4.30		2.30	
												X-ray model		5.86			

Values in each row correspond to a predicted model.
[a]Name of the research group that submitted the model.
[b]Number of the model among all models from the same group.
[c]RMSD of the model compared with the accepted structure (in Å).
[d]Columns indicate the rank of the model with respect to the left-hand column metric.
[e]DI all is the Deformation Index taking into account all interactions (stacking, Watson–Crick, and non-Watson–Crick).
[f]INF all is the Interaction Network Fidelity taking into account all interactions.
[g]INF wc is the Interaction Network Fidelity taking into account only Watson–Crick interactions.
[h]INF nwc is the Interaction Network Fidelity taking into account only non-Watson–Crick interactions.
[i]INF stack is the Interaction Network Fidelity taking into account only stacking interactions.
[j]Clash Score as computed by the MolProbity suite (Davis et al. 2007).
[k]MCQ Score is a comparison to native structure in torsion angle space (in degrees).
*Means the non-WC interaction is not available in the structure.

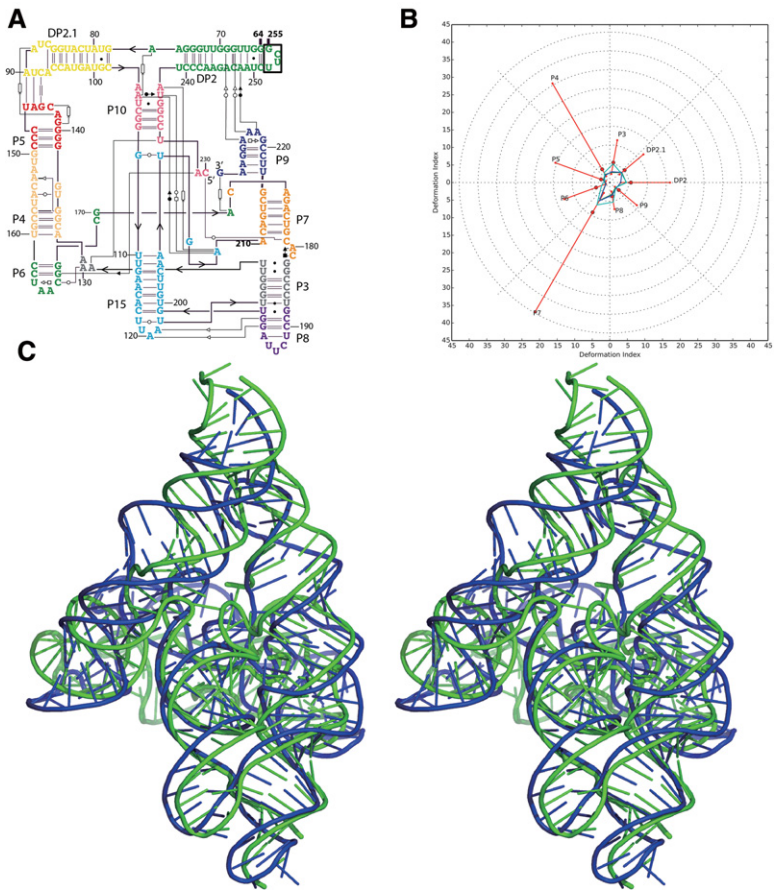


FIGURE 1. Problem 5: the lariat-capping ribozyme (*A*) secondary structure and (*B*) Deformation Profile values for the three predicted models with lowest RMSD: Das model 2 (green), Das model 1 (blue), and Adamiak model 1 (cyan). (Radial red lines) The minimum, maximum, and mean DP values for each domain. (*C*) Structure superimposition between native structure (green) and best predicted model (blue, Das model 2) with wall–eye stereo representation.

high values for the Clash Score (Chen et al. 2010), while the Clash Score for the crystal structure is low at 5.86. This implies a need for updated dictionaries of distances and angles or stronger constraints toward reasonable values both during crystal structure refinement and in structure modeling. The crystal structure of problem 5 shows an open ring structure around the center formed by a kissing interaction between two peripheral helices. This striking architecture with a clearly visible "hole" through the ring is not exactly predicted by any of the prediction models, although some groups correctly identified the overall topology (Fig. 2).

Problem 6: the adenosylcobalamin riboswitch

The RMSDs of the 34 submitted prediction models range from 11.4 to 37.0 Å with a mean value of 23.1 Å (Table 2). These are very large RMSDs but the non-inclusion of the ligand in the puzzle could be largely the cause of such high values. The Das, Major, and Chen groups rank at the top as they have relatively high accuracies in non-WC interaction prediction, while the other groups do not have the correct non-WC interactions. As a large riboswitch structure, the native structure has a Clash Score of 7.98. In such a situation, it is probably understandable that clashes appear in prediction models in order to maintain the same topology as the native structure. Models from the Das group show much better similarity to the native structure but with much higher Clash Scores due to the limited time available for refining models for this target.

Problem 10: T-box–tRNA complex

Twenty-six prediction models were submitted ranging from 6.8 to 16.9 Å RMSD with 11.5 Å as the mean value (Table 3). As this is a complex of two RNA molecules, we also compared the models of each molecule separately. The RMSDs of the T-box ranges from 5.96 to 17.9 Å, with a mean RMSD of 12.1 Å, exactly the same as the average RMSD of the molecular complex. As the structure topology of tRNA is well known, the modeling is more accurate, and the average RMSD achieved is 3.8 Å with a RMSD range between 2.49 and 6.9 Å. Therefore, the key comparisons between predictions are the T-box structure and the relative orientation/interaction between the T-box and the tRNA.

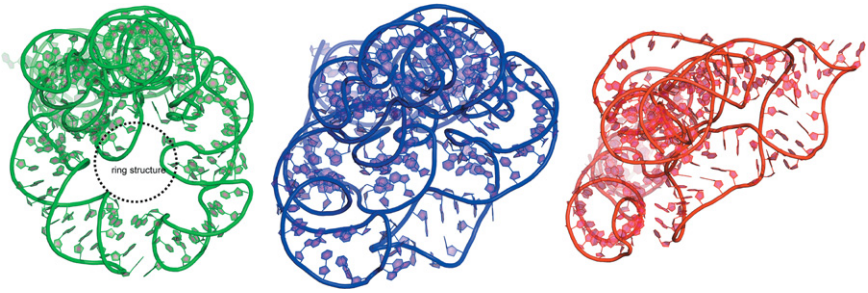


FIGURE 2. Illustration of the "ring" topology structure in Problem 5. Native structure with "ring" topology is shown in green; the best prediction model Das model 2 and the third best prediction Adamiak model 1 are shown in the same aspect in blue and red, respectively. Although the best model cannot totally capture the "ring" topology, it is more similar to native topology than others.

TABLE 2. Summary of the results for Puzzle 6

Group[a]	Number[b]	RMSD[c]	Rank[d]	DI all[e]	Rank[d]	INF all[f]	Rank[d]	INF wc[g]	Rank[d]	INF nwc[h]	Rank[d]	INF stack[i]	Rank[d]	Clash Score[j]	Rank[d]	MCQ[k]	Rank[d]
Das	4	11.39	1	15.72	1	0.72	11	0.89	5	0.32	8	0.70	11	23.48	32	14.20	8
Das	6	12.11	2	16.67	2	0.73	8	0.89	7	0.33	7	0.70	12	24.59	34	15.66	14
Das	2	13.27	3	18.29	3	0.73	7	0.89	4	0.35	2	0.71	8	17.24	29	14.95	11
Das	1	14.27	4	19.60	4	0.73	6	0.89	3	0.33	6	0.71	7	23.85	33	16.06	15
Das	9	14.70	5	20.46	5	0.72	14	0.89	8	0.35	3	0.69	16	17.98	30	14.06	7
Das	3	15.48	6	21.38	7	0.72	10	0.87	13	0.22	14	0.72	5	17.05	28	13.86	4
Das	5	15.57	7	20.80	6	0.75	1	0.89	6	0.36	1	0.74	2	13.39	26	13.19	1
Das	7	17.43	8	23.75	9	0.73	4	0.87	14	0.24	13	0.73	4	13.21	25	15.51	13
Das	8	17.57	9	23.54	8	0.75	2	0.88	11	0.33	5	0.74	1	13.02	24	14.21	9
Major	3	20.62	10	28.81	11	0.72	15	0.88	9	0.20	16	0.70	14	0.18	3	20.24	30
Major	2	20.96	11	28.96	12	0.72	9	0.86	16	0.20	15	0.72	6	0.18	2	19.94	27
Major	1	21.35	12	28.65	10	0.75	3	0.87	15	0.34	4	0.74	3	0.18	1	20.85	34
Major	7	21.44	13	29.74	13	0.72	13	0.88	10	0.29	10	0.70	15	0.55	10	20.02	28
Major	4	21.55	14	29.79	14	0.72	12	0.89	1	0.20	17	0.70	10	0.18	4	20.44	31
Chen	4	21.77	15	33.35	16	0.65	25	0.84	24	0.00	29	0.63	27	0.73	12	17.01	19
Chen	2	21.87	16	36.58	21	0.60	33	0.72	33	0.29	11	0.58	33	0.18	6	20.79	33
Dokholyan	5	22.27	17	33.36	17	0.67	24	0.85	19	0.00	25	0.64	26	7.52	20	17.42	20
Major	6	22.52	18	30.82	15	0.73	5	0.89	2	0.29	9	0.70	9	0.37	7	20.05	29
Chen	5	23.33	19	36.27	20	0.64	28	0.82	26	0.00	30	0.62	28	15.23	27	15.03	12
Chen	3	23.37	20	36.84	22	0.64	30	0.85	20	0.11	19	0.59	32	0.37	8	20.66	32
Major	5	23.63	21	33.96	18	0.70	17	0.86	17	0.10	20	0.68	18	0.55	9	18.60	25
Chen	1	23.88	22	39.21	26	0.61	32	0.80	30	0.18	18	0.57	34	0.18	5	18.93	26
Dokholyan	4	24.40	23	36.21	19	0.67	22	0.85	18	0.00	24	0.65	23	7.52	19	17.70	21
Dokholyan	1	25.51	24	37.89	25	0.67	23	0.84	23	0.00	21	0.66	22	7.34	18	16.70	18
Dokholyan	3	25.77	25	37.23	23	0.69	19	0.84	22	0.00	23	0.69	17	6.60	17	18.18	24
Dokholyan	2	26.10	26	37.67	24	0.69	18	0.83	25	0.00	22	0.68	19	7.89	21	17.86	22
Das	10	28.64	27	40.47	27	0.71	16	0.88	12	0.25	12	0.70	13	19.82	31	14.28	10
Bujnicki	2	30.55	28	44.96	28	0.68	21	0.81	27	*	*	0.67	21	1.28	14	13.58	3
Chen	7	30.74	29	48.99	30	0.63	31	0.80	29	0.00	32	0.61	30	9.72	23	16.41	16
Bujnicki	4	31.73	30	48.67	29	0.65	27	0.78	31	0.00	28	0.65	24	1.83	15	13.99	5
Bujnicki	3	31.87	31	53.39	31	0.60	34	0.70	34	*	*	0.59	31	0.92	13	13.34	2
Chen	6	35.73	32	55.76	33	0.64	29	0.81	28	0.00	31	0.62	29	2.02	16	16.48	17
Bujnicki	1	36.61	33	53.70	32	0.68	20	0.85	21	0.00	27	0.67	20	0.55	11	14.05	6
Dokholyan	6	36.96	34	56.67	34	0.65	26	0.78	32	0.00	26	0.65	25	9.54	22	17.92	23
Mean		23.09		34.06		0.69		0.84		0.17		0.67		7.80		16.83	
Standard deviation		6.87		11.54		0.04		0.05		0.14		0.05		8.14		2.54	
											X-ray model			7.98			

Values in each row correspond to a predicted model.
[a]Name of the research group that submitted the model.
[b]Number of the model among all models from the same group.
[c]RMSD of the model compared with the accepted structure (in Å).
[d]Columns indicate the rank of the model with respect to the left-hand column metric.
[e]DI all is the deformation index taking into account all interactions (stacking, Watson–Crick, and non-Watson–Crick).
[f]INF all is the Interaction Network Fidelity taking into account all interactions.
[g]INF wc is the Interaction Network Fidelity taking into account only Watson–Crick interactions.
[h]INF nwc is the Interaction Network Fidelity taking into account only non-Watson–Crick interactions.
[i]INF stack is the Interaction Network Fidelity taking into account only stacking interactions.
[j]Clash Score as computed by the MolProbity suite (Davis et al. 2007).
[k]MCQ Score is a comparison to native structure in torsion angle space (in degree).
*Means the non-WC interaction is not available in the structure.

TABLE 3. Summary of the results for Puzzle 10

Group[a]	Number[b]	RMSD[c]	Rank[d]	DI all[e]	Rank[d]	INF all[f]	Rank[d]	INF wc[g]	Rank[d]	INF nwc[h]	Rank[d]	INF stack[i]	Rank[d]	Clash Score[j]	Rank[d]	MCQ[k]	Rank[d]
Das	3	6.75	1	8.30	1	0.81	12	0.95	3	0.70	4	0.79	14	11.09	24	16.70	4
Das	4	6.98	2	8.44	2	0.83	4	0.95	1	0.68	7	0.81	8	10.73	21	15.91	1
Das	1	7.49	3	9.09	3	0.82	6	0.92	5	0.70	3	0.81	7	11.64	25	16.75	5
Bujnicki	4	9.24	4	12.33	7	0.75	15	0.87	17	0.40	16	0.74	15	0.91	2	21.30	15
Bujnicki	1	10.01	5	11.88	4	0.84	2	0.92	6	0.68	8	0.83	1	1.27	4	19.92	11
Bujnicki	8	10.05	6	12.15	6	0.83	5	0.90	11	0.64	10	0.82	3	1.45	7	19.11	6
Bujnicki	7	10.11	7	12.65	15	0.80	14	0.88	14	0.54	13	0.80	13	1.45	6	19.66	10
Bujnicki	5	10.15	8	12.39	8	0.82	9	0.87	19	0.56	12	0.83	2	1.81	10	19.44	7
Bujnicki	6	10.24	9	12.62	14	0.81	13	0.86	20	0.51	14	0.82	5	1.27	5	19.45	8
Bujnicki	9	10.25	10	12.57	12	0.82	11	0.88	15	0.69	6	0.81	10	0.91	3	19.94	12
Bujnicki	2	10.27	11	12.53	11	0.82	8	0.90	9	0.67	9	0.81	9	1.63	9	19.59	9
Bujnicki	3	10.27	12	12.48	10	0.82	7	0.91	7	0.56	11	0.82	6	0.91	1	21.23	14
Bujnicki	10	10.28	13	12.58	13	0.82	10	0.91	8	0.69	5	0.80	12	1.45	8	21.04	13
Das	5	10.28	14	12.14	5	0.85	1	0.95	2	0.78	2	0.82	4	10.91	22	16.39	3
Das	2	10.32	15	12.43	9	0.83	3	0.93	4	0.78	1	0.80	11	11.64	26	16.13	2
Chen	1	11.14	16	15.47	16	0.72	16	0.84	24	0.48	15	0.71	17	7.81	14	23.02	25
Dokholyan	3	12.00	17	16.78	17	0.72	17	0.88	16	0.10	23	0.71	16	7.80	13	22.22	20
Dokholyan	1	13.04	18	18.84	18	0.69	19	0.86	21	0.10	22	0.68	18	11.07	23	21.86	16
Dokholyan	10	13.05	19	19.11	20	0.68	20	0.89	13	0.00	26	0.65	21	10.16	20	22.56	22
Dokholyan	4	13.21	20	18.95	19	0.70	18	0.87	18	0.15	18	0.67	19	9.07	18	22.58	23
Dokholyan	2	13.58	21	20.20	21	0.67	21	0.89	12	0.12	21	0.63	24	8.89	16	22.60	24
Dokholyan	8	15.26	22	22.70	22	0.67	23	0.85	22	0.13	20	0.65	23	8.71	15	22.02	18
Dokholyan	7	15.91	23	23.68	23	0.67	22	0.83	25	0.20	17	0.66	20	6.35	11	21.95	17
Dokholyan	9	16.34	24	24.71	24	0.66	25	0.81	26	0.00	25	0.65	22	9.80	19	22.33	21
Dokholyan	5	16.42	25	26.00	26	0.63	26	0.84	23	0.00	24	0.59	26	8.89	17	23.28	26
Dokholyan	6	16.94	26	25.38	25	0.67	24	0.90	10	0.13	19	0.62	25	6.71	12	22.03	19
Mean		11.52		15.63		0.76		0.89		0.42		0.74		6.32		20.35	
Standard deviation		2.87		5.40		0.07		0.04		0.28		0.08		4.26		2.32	
												X-ray model		2.28			

Values in each row correspond to a predicted model.
[a]Name of the research group that submitted the model.
[b]Number of the model among all models from the same group.
[c]RMSD of the model compared with the accepted structure (in Å).
[d]Columns indicate the rank of the model with respect to the left-hand column metric.
[e]DI all is the Deformation Index taking into account all interactions (stacking, Watson–Crick, and non-Watson–Crick).
[f]INF all is the Interaction Network Fidelity taking into account all interactions.
[g]INF wc is the Interaction Network Fidelity taking into account only Watson–Crick interactions.
[h]INF nwc is the Interaction Network Fidelity taking into account only non-Watson–Crick interactions.
[i]INF stack is the Interaction Network Fidelity taking into account only stacking interactions.
[j]Clash Score as computed by the MolProbity suite (Davis et al. 2007).
[k]MCQ Score is a comparison to native structure in torsion angle space (in degree).

We find models of the tRNA segment from the Bujnicki group rank at the top, while the Das group does better for both the T-box alone and overall T-box/tRNA models. For the T-box, the Das group shows good predictions for both WC base pairs and non-WC interactions. However, their models, even the tRNA structures, involve more atomic clashes. In comparison, the Bujnicki group achieved better accuracy on tRNA and included fewer atomic clashes.

MODELING METHODS

Seven research groups pursuing the development of automatic modeling approaches participated in this round of RNA-Puzzles experiments. The following provides a brief description of the methodology and protocols used by the modeling groups (arranged alphabetically), together with comments and discussions.

Adamiak group

The RNA 3D structure from the Adamiak group was predicted using automated method RNAComposer (Popenda et al. 2012) in its batch mode. RNAComposer server (http://rnacomposer.cs.put.poznan.pl) uses sequence and secondary structure topology information in dot-bracket notation. Secondary structure (using the RNAStructure software) (Reuter and Mathews 2010) was adjusted using experimental data given for that RNA sequence by the Das group. Additional information about potential pseudoknots or tertiary contacts was obtained from manual analysis of the mutate-and-map data provided by the Das laboratory (RNA Mapping Database).

For Problem 5, two interactions were found: (i) 28UC29 with 93GA94 and (ii) 111GACUG115 with 148CAGUC152. Both were introduced as squared brackets into extended dot-bracket notation input:

```
GGUUGGGUUGGGAAGUAUCAUGGCUAAUCACC
  AUGAUGCAAUCGGGUUGAACACUUAAUUGGG
  UUAAAACGGUGGGGGACGAUCCCGUAACAUCC
  GUCCUAACGGCGACAGACUGCACGGCCCUGCC
  UCUUAGGUGUGUUCAAUGAACAGUCGUUCCG
  AAAGGAAGCAUCCGGUAUCCCAAGACAAUC
(((((..(((((..((((((((((....[[.)))))))))..(((((((((((((......((((....
  ((((((((((..]]))))).....)).))))((....)).....[[[[[....)))).(((.....)))
  )))))))).....]]]]].((((....)))))..)))))))...))))))..)))))
```

Before pressing "Compose" button in the batch mode, the option "Add atom distance restraints" was checked to introduce restraints concerning the interactions 111GACUG115 with 148CAGUC152. To do so, the RNA duplex with the same sequence was extracted from the X-ray structure (PDB 2Z75, resolution 1.7 Å) and selected using search engine of RNA FRABASE (Popenda et al. 2008). Subsequently, related 508 distance restraints were calculated (between atoms P, C1′, C2′, C5′, O3′, C2, C4, C6, and C8) and uploaded to RNAComposer. RNAComposer (64-bit Intel Xeon 2.33 GHz processor-based platform with scalable 8 GB memory) predicts 10 3D models within <10 min.

The resulting models were inspected for the total energy value calculated by RNAComposer and for the preservation of the 111GACUG115/148CAGUC152 and 28UC29/93GA94 interactions. Models showing lowest total energy were chosen for further analysis. Some fragment selections chosen by RNAComposer for the 3D structure assembly prohibited the formation of required contacts and such models were rejected at this stage. Subsequently, the selected models were investigated for the proximity of the 111GACUG115/148CAGUC152 pseudoknot region to the RNA termini. The mutate-and-map data (RNA Mapping Database) suggested that region hosting pseudoknot 111GACUG115/148CAGUC152 should be close to the molecule termini, namely to the internal loop 6GG7/182GA183. The model fulfilling this criterion and representing the lowest total energy estimated by RNAComposer was selected. Since RNAComposer automatically conducts two energy minimization steps prior to returning final RNA 3D structure this model did not require any further refinement. The model was validated using NUCheck (Feng et al. 1998).

Bujnicki group

The Bujnicki group used a hybrid strategy similar to the one used in the previous editions of the RNA-Puzzles experiment (Cruz et al. 2012), which comprised template-based (comparative) modeling, global folding with restraints using a coarse-grained method for template-free folding, and high-resolution refinement.

First, for all target sequences they attempted to identify homologous families in the Rfam database (Burge et al. 2013) and homologous RNAs with experimentally determined structures. For RNA sequences or sequence fragments that exhibited homology with RNAs with experimentally determined structures, initial models were constructed by template-based modeling and fragment assembly using ModeRNA (Rother et al. 2011). Target-template alignments were prepared manually, with the aid of secondary structure information extracted from Rfam and corrected if needed with the use of predictions made with RNA metaserver (http://genesilico.pl/rnametaserver/) developed as a part of the CompaRNA project (Puton et al. 2013). This stage was very similar to that which they used previously in RNA-Puzzles (Cruz et al. 2012). For Problem 5, a lariat-capping ribozyme related to group I self-cleaving introns (Rfam family RF01807), they used a group I intron structure (PDB 1ZZN) as the main template. For Problem 6, an adenosylcobalamin riboswitch (RF00174), they were unable to find a suitable template, therefore no template-based modeling was performed. For Problem 10, a tRNA bound to a T-box RNA, template-based modeling of the entire complex was based on the structure of a related complex (Grigg et al. 2013) built

manually by the authors based on crystal and NMR structures of fragments (PDB 4JRC and PDB 2KHY), with the use of cross-linking, mutagenesis, and SAXS data.

The aforementioned initial models of target structures (in the case of Problem 6—an artificial circular conformation of the target sequence with 5′ and 3′ ends close to each other) were used as starting points for global refinement, using the SimRNA method for RNA folding simulations, which uses a coarse-grained representation, relies on the Monte Carlo method for sampling the conformational space, and uses a statistical potential to approximate the energy and identify conformations that correspond to biologically relevant structures (MJ Boniecki, G Lach, K Tomala, W Dawson, P Lukasz, T Soltysinski, KM Rother, and JM Bujnicki, in prep.). Here, they used a novel version of SimRNA, which uses five (rather than three) atoms per residue: P of the phosphate group, C4′ of the ribose moiety, and in which base moieties are represented by triangles: N1–C2–C4 for pyrimidines and N9–C2–C6 for purines. This representation provides much improved description of base faces and edges compared with the previous version that used only one atom per base (Cruz et al. 2012; Rother et al. 2012) and therefore improves the modeling of stacking and base-pairing interactions, e.g., it discriminates much better between canonical and noncanonical base-pairing. Regions predicted to be confidently modeled in initial models were "frozen" while other regions were allowed to change conformation. For modeling of complex 3D structures, SimRNA can use additional restraints, derived from experimental or computational analyses, including information about secondary structure and/or long-range contacts. They have used such information depending on its availability. Typically, predictions were first made with restraints on predicted secondary structure and if additional data became available sufficiently long before the prediction deadline (e.g., results of experiments performed by the Das group and made available to all participants of RNA-Puzzles), additional simulations were conducted. Given the very tight deadline for Problem 6, they were unable to utilize additional data for this RNA, leading to poor results.

Predictions generated by SimRNA were converted to full-atom representation and ranked for submission using a combination of various criteria, including the results of clustering (the higher number of similar well-scored structures the better), agreement with experimental data not used in the process of modeling, manual inspection, and scoring with independent methods such as RASP (Capriotti et al. 2011). If time permitted, models selected for submission were subjected to high-resolution refinement whose aim was to reduce clashes, idealize geometries, and improve local interactions such as in standard and non-WC base pairs. Here, they used a different method than previously, namely an in-house software tool QRNAS (J Stasiewicz and JM Bujnicki, unpubl.) that extends the AMBER force field with energy terms explicitly modeling hydrogen bonds, idealizes base pair planarity and regularizes backbone conformation.

As in their previous (Cruz et al. 2012) modeling exercise, human intervention was relatively large. Most of the time was devoted to searching for additional information related to target RNA sequences and discussions within the group. Time used for alignment preparation and for selection of models for submission varied greatly depending on the difficulty of the Problem. Time used for template-based modeling was negligible. Time required for SimRNA modeling was typically a few days per target, and the final refinement was typically run overnight.

Chen group

The Chen group used a hierarchical approach to predict RNA 3D structure from the sequence (Xu et al. 2014). For a given RNA sequence, they first predict the secondary structure from the free energy landscape using the Vfold model (Cao and Chen 2005, 2006, 2009; Chen 2008). A unique feature of the Vfold model at secondary level is its ability to compute the RNA motif-based loop entropies. Using two virtual bonds per nucleotide to represent the backbone conformation, Vfold model samples fluctuations of loops/junction conformations in 3D space through conformational enumeration model (Cao and Chen 2005, 2006, 2009; Chen 2008). By calculating the probability of loop formation, the model can give the conformational entropy parameters for the formation of the different types of loops such as pseudoknot loops. Another notable feature of the Vfold model at secondary level is the modeling of RNA loop free energy. By enumerating all the possible (sequence-dependent) intra-loop mismatches, the Vfold model partially accounts for the sequence-dependence of the loop free energy.

Next, a 3D coarse-grained scaffold is constructed based on the predicted secondary structure (Cao and Chen 2011). To construct a 3D scaffold, the predicted helix stems are modeled as A-form helices. For the loops/junctions, 3D fragments from the known PDB database were used. Specifically, a structural template database (Xu et al. 2014) was built by classifying the structures according to the different motifs such as hairpin loops and internal/bulge loops, three-way junctions, four-way junctions, pseudoknots, etc. For each junction, the optimal (top 5) fragments were selected for the further structure assembly of the whole RNA. Any 3D structures generated by the structure assembly with structural clashes would be excluded. The final all-atom structures were built based on the coarse-grained model, followed by refinement using AMBER energy minimization. Two thousand steps of energy minimization were run, applying 500.0 kcal/mol constraints to all the residues, followed by another 2000 steps of minimization without constraints.

In order to increase the accuracy of RNA secondary structure prediction, they applied Rfam (Burge et al. 2013) to identify the possible conserved base pairs and used the most conserved base pair information as constraint to the Vfold algorithm to predict secondary structures. If available, the

SHAPE (selective 2′-hydroxyl acylation analyzed by primer extension) (Merino et al. 2005) experimental data were also used as constraint in the Vfold algorithm for secondary structure prediction. The SHAPE reactivity is strongly related to the nucleotide flexibility at single nucleotide resolution. Specifically, some nucleotides are restricted to be in loop regions (without forming base pairs with other nucleotides) because of their high SHAPE reactivity. The combination of SHAPE data and/or homologous sequence information from Rfam and the Vfold algorithm led to enhanced accuracy of RNA secondary structure prediction.

For the RNA/RNA complex of Problem 10, they built the 3D structures for each strand separately using the above hierarchical approach (Xu et al. 2014). The final complex structure was built manually based on the previously published SAXS-reconstructed envelope from DAMMIF (Fig. 5 in Grigg et al. 2013). Then, they ran a short-time MD simulation to stabilize the interactions between the two RNA molecules.

In summary, the computation involved two steps: (a) the prediction of the secondary structure and the construction of the coarse-grained 3D structure and (b) AMBER energy minimization. The computation time (Ta, Tb) for the two steps are (~2–3 h, <1 h), (<1 h, <1 h), and (<1 h, <3 h), for Problems 5, 6, and 10, respectively. The computations were performed on a desktop PC with Intel Core(TM) i7-2600 CPU at 3.40 GHz. They manually incorporated the constraints from the SHAPE experiments and the Rfam results into the Vfold model for secondary structure prediction. In addition to the construction of 3D structures, the RNA/RNA complex for Problem 10 involved human interference based on the SAXS data. All other steps were achieved automatically by computations.

Das group

The chemical probing data, obtained in by the Das group, were first used to model secondary structure with available automated algorithms and then further used for tertiary structure prediction. Automated secondary structure modeling with RNAstructure was carried out as described in Tian et al. (2014), with tools available by server (http://rmdb.stanford.edu/structureserver/) (Cordero et al. 2012b). In all three Problems, use of 1D SHAPE data improved RNAstructure 5.4 secondary structure predictions compared with modeling without data, leading to perfect recovery of helices in Problem 10 (Supplemental Table S1; Supplemental Fig. S6B). Nevertheless, approximately half of helices remained mispredicted in Problems 5 and 6. In Problem 5, use of DMS/CMCT with RNAstructure gave better models than SHAPE-guided modeling; but for Problems 6 and 10, use of DMS/CMCT made models worse compared even with modeling without data. On an encouraging note, a bootstrapping procedure that gives conservative estimates of modeling uncertainty (Kladwang et al. 2011b; Ramachandran et al. 2013) was able to highlight confident and nonconfident regions in all cases. For example, any helix modeled with >75% bootstrap values agreed with the subsequently released crystallographic model (Supplemental Figs. S1–S3) (Fig. 3).

More recent versions of RNAstructure have updated parameters for nearest-neighbor energies and for converting SHAPE values to pseudoenergies, and also have the ability to model pseudoknots (Hajdin et al. 2013). Although not a strictly blind test, the data above allowed a test of these advances. Use of RNAstructure 5.6 *Fold* for SHAPE-directed

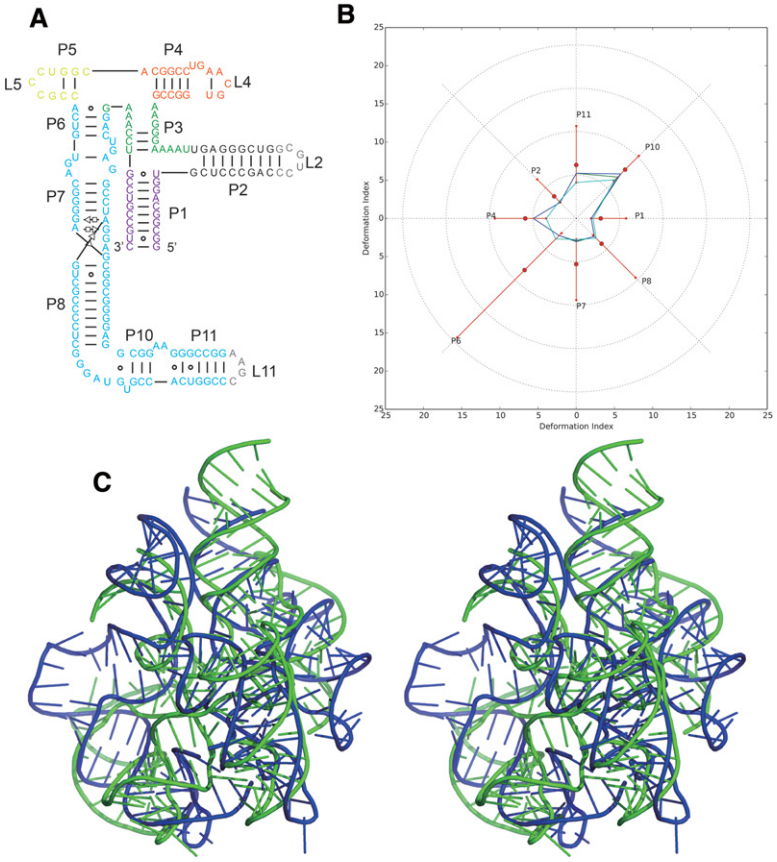


FIGURE 3. Problem 6: the adenosylcobalamin riboswitch (*A*) secondary structure and (*B*) Deformation Profile values for the three predicted models with lowest RMSD: Das model 4 (green), Das model 6 (blue) and Das model 2 (cyan). (Radial red lines) The minimum, maximum, and mean DP values for each domain. (*C*) Structure superimposition between native structure (green) and best predicted model (blue, Das model 4) with wall–eye stereo representation.

modeling did not improve modeling of Problem 5, and gave less accurate models of the other Problems 6 and 10, compared with RNAstructure 5.4 *Fold*; the difference appears to be due to a change in the parameters for SHAPE pseudoenergy. Use of the *ShapeKnots* executable produced a significant improvement but still imperfect model in SHAPE-directed modeling of Problem 5, which contains a pseudoknot in its catalytic core. In the other cases, *ShapeKnots* predictions did not improve upon pseudoknot-free *Fold* modeling. These results underscore the challenge of modeling RNA secondary structure using conventional 1D chemical-mapping data, even with continuing algorithmic advances. As has been discussed previously (Cordero et al. 2012a; Leonard et al. 2013; Rice et al. 2014), protection of nucleotides may signal non-Watson–Crick rather than Watson–Crick pairing in the structure, but current methods do not generally distinguish these possibilities (Figs. 4, 5).

For Problems 5 and 6, secondary chemical-mapping data were also acquired through the M^2 approach (Cordero et al. 2014). In this method, chemical-mapping profiles are measured not only for the sequence of interest but also for variants mutating each nucleotide in the RNA (Kladwang et al. 2011a). Increased reactivity of one nucleotide upon mutation of a sequence-distance nucleotide can signal their interaction in three dimensions, and these data can be leveraged for automatic secondary structure inference in RNAstructure. For Problem 5, RNAstructure 5.4 *Fold* guided by M^2-SHAPE data recovered all helices longer than 2 bp, except the catalytic core pseudoknot. Integrating M^2 data with the more recent RNAstructure 5.6 *ShapeKnots* recovered all of these helices, including the pseudoknot (Supplemental Table S1; Supplemental Fig. S1L). For Problem 6, all helices longer than 2 bp were recovered correctly with M^2-SHAPE data and by either RNAstructure executable. Errors in edge base pairs of several helices remain, as well as a register shift in Problem 5 (which was corrected in M^2-DMS analysis; Supplemental Fig. S1P). Overall, these comparisons confirm that secondary chemical mapping coupled to automated algorithms consistently achieves correct global secondary structures for complex RNA folds in terms of helix recovery, but resolving fine errors in edge base pairs will require methodological improvements.

Beyond basic secondary structure modeling, this round of puzzles also inspired development of computational methods for 3D modeling, primarily in three areas. First, simple automated tools in the Rosetta framework (Leaver-Fay et al. 2011) were created for threading structural templates into the desired sequence, such as the catalytic core of Problem 5, the lariat-capping ribozyme (see also below). Second, the Das group expanded fragment assembly of RNA with full-atom refinement (FARFAR) (Das et al. 2010; Kladwang et al. 2011a), whose interface for job setup was previously cumbersome, especially to solve subpieces of large RNAs (e.g., the three-way P15/P8/P3 junction of Problem 5). The RNA puzzles inspired them to write a single python script (rna_denovo_setup.py) for straightforward setup of FARFAR jobs, taking as input the full-model sequence and secondary structure, the residues of desired subdomain, and PDB models for any known subpieces of the subdomain. For Problem 6, the Das group also created a mode for setting up rigid-body "docking" of multiple RNA pieces including a placeholder sphere for the adenosylcobalamin ligand. Finally, an expansion of "stepwise" assembly was piloted, previously developed for enumerative high-resolution modeling of motifs (Sripakdeevong et al. 2011; Chou et al. 2013), to gene rate complex RNA folds by progressively closing "rings" of motifs through numerous tertiary buildup paths (e.g., to model the ring-like connection of the catalytic core with the P3/P8/P15 junction and the P2.1/P5 kissing loops in the GIR1 lariat-capping ribozyme). Compared with prior fragment assembly approaches from Das group (Kladwang et al. 2011a), this

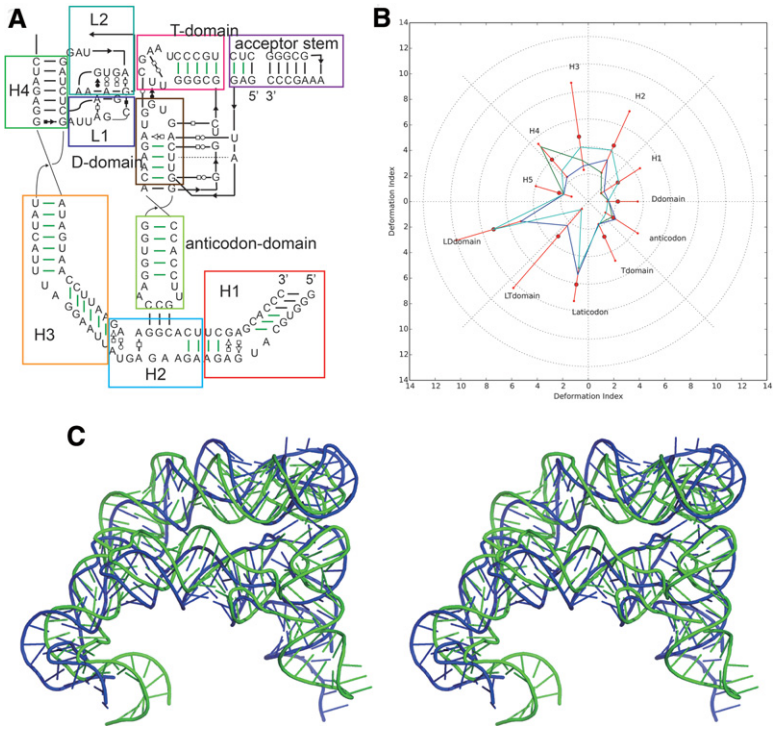


FIGURE 4. Problem10: the T-box–tRNA complex (*A*) secondary structure and (*B*) Deformation Profile values for the three predicted models with lowest RMSD: Das model 3 (green), Das model 4 (blue), and Das model 1 (cyan). (Radial red lines) The minimum, maximum, and mean DP values for each domain. (*C*) Structure superimposition between native structure (green) and best predicted model (blue, Das model 3).

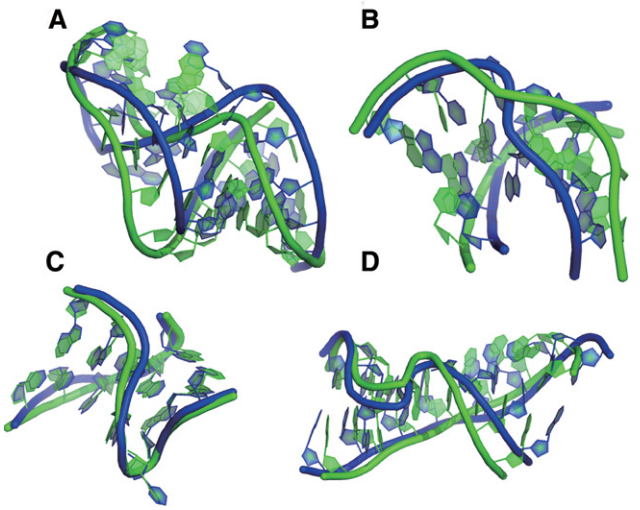


FIGURE 5. Modules in Problem 10. (*A*) Detailed structure of T-loop of Das model 4, (*B*) detailed structure of U30 of Das model 4, (*C*) detailed structure of K-turn of Das model 4, (*D*) detailed structure of Loop-E of Das model 4.

stepwise strategy was efficient in generating realistic, converged conformations of complex folds with multiple tertiary contacts at subhelical resolution. All Rosetta tools are freely available to academic researchers (Leaver-Fay et al. 2011) and documented at https://www.rosettacommons.org/docs/latest/rna-denovo-setup.html/.

The process was a mix of automatic and manual steps, as many of the tools were being developed "on-the-fly." On one hand, chemical-mapping-guided inference of secondary structure to double-check models from the literature was carried out automatically, but inference of some tertiary contacts from these data was guided by visual inspection of the M^2 data (see below). On the other hand, identification of potential templates for threading/homology modeling was not carried out automatically. Structural templates and alignments were instead derived from literature search (group I intron alignment to the lariat-capping ribozyme for Problem 5 (Beckert et al. 2008); mapping "half" of the FMN riboswitch to the adenosylcobalamin-binding core for Problem 6 (Barrick and Breaker 2007; Geary et al. 2011); and mapping double T-loop (Grigg et al. 2013; Lehmann et al. 2013), sarcin–ricin loop (Yang et al. 2001) and kink-turn motifs (Vidovic et al. 2000) to T-box, the tRNA (Fukai et al. 2000), and the T-Box/tRNA interface derived from ribosome (Dunkle et al. 2011) for Problem 10) and manual expert inspection (refining the core of the Problem 5 group I intron alignment; a previously unrecognized kink-turn in the Problem 6 adenosylcobalamin riboswitch; the ribosome-bound tRNA/mRNA-like interaction in the Problem 10 T-box/tRNA complex). Fully automating template recognition and tertiary contact inference-with guidance from readily available chemical-mapping data-appears to be an important challenge for the field (Table 4).

For all three targets, experimental data were critical for ruling out structural hypotheses that would have required substantial computational expense to explore, and in some cases, gave critical data that guided modeling, illustrated in Supplemental Figures S4–S6. For Problem 5, two peripheral tertiary contacts were not recognized in previous literature but were important for defining ~1/3 of the model. The contacts were apparent in M^2 data as changes in chemical mapping on one side of the contact in response to mutations on the other side. For Problem 6, several secondary structure models had been proposed in the literature (Ravnum and Andersson 2001; Nahvi et al. 2002, 2004; Vitreschak et al. 2003; Barrick and Breaker 2007), and the M^2 analysis was important for unambiguously confirming the correct model. Further, the M^2 data showed no evidence of extensive interaction between the helix P2 of the P1/P2/P3 junction and the long "arm" P7–P11, or for interactions within the "arm"; so runs with those contacts were not set up. For Problem 10, M^2 data was not acquired due to time constraints (three other Problems were being modeled concomitantly), but the available 1D chemical-mapping data helped rule out a potential fourth base pair neighboring the three base pair interaction of the tRNA anticodon and its T-box binding site; enforcing that interaction would have produced inaccurate distortions (Table 5).

As many of the computational methods were being developed at the same time as modeling, performance was not optimized. Thousands of CPU-hours were used (12 h for ~20–100 cores) for each 3D modeling step that involved fragment-based assembly and refinement of subpieces. For the case of Problem 5, the Das group ended up expending at least 30,000 CPU-hours. Nevertheless, since the prediction period, further automation and optimization has brought the computational expense of these procedures to under 10,000 CPU-hours per target, taking less than a week of wall clock time. It is noted that academic researchers interested in using these tools can make use of free "startup" allocations on the XSEDE supercomputers of up to 20,000 CPU-hours. As for Problem 6 (adenosylcobalamin riboswitch), both experiments and computational modeling were carried out in 1 wk.

Dokholyan group

The Dokholyan group at the University of North Carolina at Chapel Hill in collaboration with the Ding group from Clemson provided predictions for Problems 5, 6, and 10 using multiscale discrete molecular dynamics (DMD) method (Proctor et al. 2011; Shirvanyants et al. 2012). The structure modeling was performed with coarse-grained folding simulations followed by an all-atom reconstruction. In the coarse-grained folding simulations, the three-bead RNA model, where each nucleotide is represented by three pseudoatoms corresponding to base, sugar, and phosphate groups (Ding et al. 2008), was used. The interactions between the three beads are modeled based on information available from high-resolution RNA structure database. Bonded

TABLE 4. Summary of the results for Puzzle 10 t-box

Group[a]	Number[b]	RMSD[c]	Rank[d]	DI all[e]	Rank[d]	INF all[f]	Rank[d]	INF wc[g]	Rank[d]	INF nwc[h]	Rank[d]	INF stack[i]	Rank[d]	Clash Score[j]	Rank[d]	MCQ[k]	Rank[d]
Das	4	5.96	1	7.40	1	0.81	3	0.91	2	0.62	6	0.80	3	10.73	21	13.92	1
Das	3	6.51	2	8.18	2	0.80	5	0.90	4	0.65	4	0.79	9	11.09	24	15.39	5
Das	1	8.08	3	10.10	3	0.80	4	0.86	7	0.65	3	0.81	2	11.64	25	14.87	3
Das	5	9.32	4	11.24	4	0.83	1	0.91	3	0.74	1	0.82	1	10.91	22	15.08	4
Das	2	9.40	5	11.67	5	0.81	2	0.91	1	0.67	2	0.80	5	11.64	26	14.74	2
Bujnicki	4	10.62	6	16.64	15	0.64	20	0.81	20	0.00	25	0.64	18	0.91	2	21.77	21
Bujnicki	1	11.66	7	14.80	6	0.79	6	0.85	9	0.50	10	0.80	6	1.27	4	19.85	10
Bujnicki	8	11.69	8	14.86	7	0.79	7	0.85	10	0.56	8	0.79	7	1.45	7	18.32	6
Bujnicki	5	11.87	9	15.35	8	0.77	8	0.81	17	0.33	13	0.80	4	1.81	10	19.33	7
Bujnicki	7	11.88	10	16.14	14	0.74	14	0.81	18	0.30	14	0.75	14	1.45	6	20.07	11
Bujnicki	3	11.98	11	15.74	11	0.76	11	0.86	8	0.38	12	0.77	10	0.91	1	22.13	25
Bujnicki	6	12.00	12	15.95	13	0.75	13	0.79	24	0.17	15	0.79	8	1.27	5	19.46	8
Bujnicki	10	12.02	13	15.60	9	0.77	9	0.85	11	0.63	5	0.76	12	1.45	8	22.13	24
Bujnicki	9	12.02	14	15.83	12	0.76	12	0.81	19	0.59	7	0.76	13	0.91	3	20.43	12
Bujnicki	2	12.02	15	15.71	10	0.77	10	0.84	13	0.50	9	0.77	11	1.63	9	19.56	9
Dokholyan	2	12.29	16	19.44	18	0.63	23	0.80	22	0.00	17	0.62	20	8.89	16	20.90	14
Dokholyan	3	12.67	17	18.75	17	0.68	16	0.84	12	0.00	18	0.67	16	7.80	13	21.65	20
Chen	1	13.01	18	18.53	16	0.70	15	0.77	26	0.42	11	0.73	15	7.81	14	22.07	23
Dokholyan	8	13.22	19	21.39	20	0.62	24	0.83	14	0.00	23	0.59	25	8.71	15	21.63	19
Dokholyan	9	13.36	20	21.96	22	0.61	25	0.78	25	*	*	0.59	24	9.80	19	21.24	17
Dokholyan	10	13.47	21	20.90	19	0.65	19	0.88	6	0.00	24	0.61	22	10.16	20	21.98	22
Dokholyan	4	14.14	22	22.23	23	0.64	22	0.82	15	0.00	19	0.62	21	9.07	18	21.60	18
Dokholyan	1	14.42	23	21.93	21	0.66	17	0.80	21	0.00	16	0.66	17	11.07	23	20.92	15
Dokholyan	5	15.99	24	26.69	25	0.60	26	0.82	16	0.00	20	0.57	26	8.89	17	22.41	26
Dokholyan	7	16.87	25	26.47	24	0.64	21	0.79	23	0.00	22	0.64	19	6.35	11	20.64	13
Dokholyan	6	17.93	26	27.69	26	0.65	18	0.88	5	0.00	21	0.61	23	6.71	12	21.21	16
Mean		12.09		17.35		0.72		0.84		0.31		0.71		6.32		19.74	
Standard deviation		2.76		5.35		0.08		0.04		0.29		0.09		4.26		2.67	
											X-ray model			2.28			

Values in each row correspond to a predicted model.
[a]Name of the research group that submitted the model.
[b]Number of the model among all models from the same group.
[c]RMSD of the model compared with the accepted structure (in Å).
[d]Columns indicate the rank of the model with respect to the left-hand column metric.
[e]DI all is the Deformation Index taking into account all interactions (stacking, Watson–Crick, and non-Watson–Crick).
[f]INF all is the Interaction Network Fidelity taking into account all interactions.
[g]INF wc is the Interaction Network Fidelity taking into account only Watson–Crick interactions.
[h]INF nwc is the Interaction Network Fidelity taking into account only non-Watson–Crick interactions.
[i]INF stack is the Interaction Network Fidelity taking into account only stacking interactions.
[j]Clash Score as computed by the MolProbity suite (Davis et al. 2007).
[k]MCQ Score is a comparison to native structure in torsion angle space (in degrees).
*Means the non-WC interaction is not available in the structure.

TABLE 5. Summary of the results for Puzzle 10 tRNA

Group[a]	Number[b]	RMSD[c]	Rank[d]	DI all[e]	Rank[d]	INF all[f]	Rank[d]	INF wc[g]	Rank[d]	INF nwc[h]	Rank[d]	INF stack[i]	Rank[d]	Clash Score[j]	Rank[d]	MCQ[k]	Rank[d]
Bujnicki	7	2.49	1	2.75	1	0.91	8	0.97	10	0.91	7	0.88	8	1.45	6	19.03	5
Bujnicki	9	2.50	2	2.76	2	0.91	9	0.97	11	0.91	8	0.88	9	0.91	3	19.17	6
Bujnicki	4	2.57	3	2.80	3	0.92	3	0.97	8	1.00	2	0.89	7	0.91	2	20.56	15
Bujnicki	2	2.60	4	2.85	4	0.91	4	0.97	9	0.91	6	0.89	4	1.63	9	19.63	10
Bujnicki	6	2.63	5	2.89	5	0.91	5	0.94	13	1.00	4	0.89	6	1.27	5	19.42	8
Bujnicki	8	2.65	6	2.98	10	0.89	10	0.94	14	0.80	15	0.88	10	1.45	7	20.39	14
Bujnicki	10	2.67	7	2.93	6	0.91	6	0.97	12	0.85	14	0.90	2	1.45	8	19.39	7
Bujnicki	5	2.68	8	2.96	9	0.91	7	0.92	16	1.00	3	0.89	5	1.81	10	19.62	9
Bujnicki	3	2.69	9	2.93	7	0.92	2	0.97	7	0.89	9	0.90	1	0.91	1	19.82	12
Bujnicki	1	2.71	10	2.95	8	0.92	1	1.00	5	0.91	5	0.89	3	1.27	4	20.03	13
Das	3	2.92	11	3.39	11	0.86	15	1.00	2	0.85	11	0.81	15	11.09	24	18.78	4
Das	5	2.96	12	3.41	12	0.87	12	1.00	4	0.85	13	0.83	11	10.91	22	18.45	3
Das	1	3.01	13	3.50	13	0.86	14	1.00	1	0.85	10	0.81	14	11.64	25	19.82	11
Das	2	3.21	14	3.69	14	0.87	11	0.97	6	1.00	1	0.81	12	11.64	26	16.19	1
Das	4	3.32	15	3.85	15	0.86	13	1.00	3	0.85	12	0.81	13	10.73	21	17.12	2
Chen	1	3.66	16	4.59	16	0.80	16	0.91	19	0.60	16	0.78	17	7.81	14	24.33	26
Dokholyan	6	4.63	17	6.52	18	0.71	24	0.90	20	0.26	20	0.67	25	6.71	12	22.38	18
Dokholyan	5	4.87	18	7.30	22	0.67	26	0.84	24	*	*	0.63	26	8.89	17	23.74	25
Dokholyan	9	4.88	19	6.93	20	0.71	25	0.82	26	0.00	24	0.72	22	9.80	19	23.39	24
Dokholyan	3	5.04	20	6.44	17	0.78	17	0.92	18	0.20	23	0.80	16	7.80	13	22.96	21
Dokholyan	1	5.27	21	7.01	21	0.75	19	0.92	17	0.22	22	0.74	19	11.07	23	23.38	23
Dokholyan	7	5.27	22	7.32	23	0.72	22	0.84	25	0.45	18	0.71	23	6.35	11	22.05	16
Dokholyan	4	5.30	23	6.91	19	0.77	18	0.92	15	0.45	17	0.73	21	9.07	18	23.18	22
Dokholyan	10	5.52	24	7.47	24	0.74	21	0.89	22	0.00	25	0.74	20	10.16	20	22.89	20
Dokholyan	8	5.74	25	7.63	25	0.75	20	0.87	23	0.26	21	0.75	18	8.71	15	22.15	17
Dokholyan	2	6.90	26	9.62	26	0.72	23	0.89	21	0.26	19	0.69	24	8.89	16	22.79	19
Mean		3.80		4.78		0.83		0.94		0.65		0.81		6.32		20.79	
Standard deviation		1.33		2.15		0.08		0.05		0.34		0.08		4.26		2.18	
												X-ray model		2.28			

Values in each row correspond to a predicted model.
[a]Name of the research group that submitted the model.
[b]Number of the model among all models from the same group.
[c]RMSD of the model compared with the accepted structure (in Å).
[d]Columns indicate the rank of the model with respect to the left-hand column metric.
[e]DI all is the Deformation Index taking into account all interactions (stacking, Watson–Crick, and non-Watson–Crick).
[f]INF all is the Interaction Network Fidelity taking into account all interactions.
[g]INF wc is the Interaction Network Fidelity taking into account only Watson–Crick interactions.
[h]INF nwc is the Interaction Network Fidelity taking into account only non-Watson–Crick interactions.
[i]INF stack is the Interaction Network Fidelity taking into account only stacking interactions.
[j]Clash Score as computed by the MolProbity suite (Davis et al. 2007).
[k]MCQ Score is a comparison to native structure in torsion angle space (in degree).
*Means the non-WC interaction is not available in the structure.

interactions are based on parameters derived from covalent-bonding, bond angles and dihedral angles, while the non-bonded interactions are derived from base-pairing, base stacking, hydrophobic interactions, and phosphate–phosphate repulsions. Replica exchange DMD simulations were performed (Ding et al. 2008) followed by a selection protocol to select the lowest energy structures. Briefly, structures were selected from coarse-grained simulations based on energies obtained using the coarse-grained energy function. In the first filter, all the structures from every replica, which are the lowest ten percent of the energies and then perform hierarchical clustering for identifying the most dominant state among the lowest energy ensemble, were selected. The centroid of the most populated cluster was selected as the representative structure for the simulation. For the representative structure, the model was further refined by performing all-atom reconstruction. The all-atom DMD approach for RNA is similar to one used for all-atom protein modeling (Ding et al. 2008).

The CPU time for DMD-based RNA structure prediction depends on the length of the RNA. Previous benchmarks' showed linear dependence on RNA length. For Problems 6 and 10, the simulations were performed on UNC Killdevil computing cluster (each compute node consists of 12 core, 2.99 GHz Intel processors, with either 48 of 96 Gb memory), for Problem 6, a 169-nt length RNA, the total CPU time was ~21 h for eight compute nodes, which roughly translates to ~2.75 h of real time in simulation. The clustering algorithm run on similar compute node took <15 min to complete.

In the predictions they used base-pairing information, which was derived for each Problem using different methods. Problem 5 was mainly based on biochemical data provided by the Das group and sequence comparative analysis obtained by multiple sequence alignment from Rfam (Griffiths-Jones et al. 2003). The same was true for Problem 6 with additional data obtained from the Mfold server (Zuker 2003). Problem 10 used biochemical data from the Das group and the Mfold server (Zuker 2003).

Experimentally derived tertiary structure information was also used. In Puzzle 5, the specific long-range proximity constraint between nucleotides 78 and 170 was inferred from the cleavage sites between helices 5 (P5) and 10 (P10). For Problem 6, the two groups used the in-line probing data to approximate the solvent accessibility (Nahvi et al. 2004), which were added to DMD simulation as the hydroxyl radical probing approach (Ding et al. 2012). In the case of Problem 10, the two binding sites between tRNA and tBox were taken from experimental study (Grigg et al. 2013). The method is fully automated once the base-pairing information for the RNA has been provided.

Major group

The adenosylcobalamin riboswitch of Problem 6 was found by sequence similarity using BLAST. The secondary structure for this riboswitch was deduced by Barrick and Breaker (2007) and it was used as a primary template. Among the alternative structural elements, they kept the three-way and four-way junctions, as well as the T-loop-type interaction between the four-way junction and the lower part of P7.

The Major group fed the stems P2 and P4 as constraints to MC-Fold (Parisien and Major 2008). Various alternative secondary structural elements were indicated by Barrick and Breaker, especially between P7 and P11. The Major group identified a sequence with a potential to adopt a kink-turn, similar to the kink-turn predicted in the snoRNA U3 C′/D box (Rozhdestvensky et al. 2003). This kink-turn has a GA tandem and an asymmetric loop. We decided to assume its formation by adding it to the constraints using a mask in MC-Fold:

```
GGGAGU-GCGAGGAUC
((((((-))...))))
```

The 3D model was built by using the T-loop crystal structure of a tRNA (PDB 1EVV). The kink-turn area was modeled after Kt-7 (PDB ID 3CC2 of the 23S rRNA of *Haloarcula Marismortui*). The remaining parts and the final assembly were modeled using MC-Sym (major.iric.ca/Web/mctools). The models generated by MC-Sym were minimized up to the "brushup" level of the MC-Pipeline. The selection of the candidate models was based on "Score" values, a homemade all-atoms force-field which is part of the "Analysis" module of the MC-pipeline.

Xiao group

The Xiao group used 3dRNA web server (http://122.205.6.127/3dRNA/3dRNA.html) (Zhao et al. 2012) to complete the prediction for Problems 5 and 6. 3dRNA builds tertiary structure of an RNA molecule by assembling three-dimensional (3D) structural templates of its secondary structure elements, including helix, hairpin loop, internal loop, bulge loop, and multiway junction. The 3D templates are from a library extracted from experimental RNA structures. In addition, 3dRNA can build different models for an RNA molecule by using different templates for each of secondary structure elements.

In the prediction for Problems 5 and 6, they first predicted the secondary structures of the submitted sequences by using Mfold (Zuker 2003) and RNAfold (Hofacker et al. 1994) and picked out the optimal prediction. Then, the sequences and predicted secondary structures were submitted to 3dRNA web server and a set of structural models was generated. Finally, these structural models were ranked with a scoring function 3dRNAscore (data not shown) and the lowest energy structures were selected as the candidate structures.

For Problem 5, the Xiao group used Mfold server (Zuker 2003) and RNAfold (Hofacker et al. 1994) to predict the secondary structure. The Mfold web server gave 10 secondary structures consisting of an optimal and nine suboptimal

folds. The RNAfold web server gave an optimal secondary structure by minimum free energy and an optimal secondary structure by thermodynamic ensemble. They used both the Mfold and RNAfold optimal folds to predict the tertiary structure. One of the optimal secondary structure is a three-way-junction structure whose secondary structure is (..().().....()). This three-way junction had no 3D templates in the library but 3dRNA could automatically search the library and pick out a nearest three-way junction, whose secondary structure is (..()..()....()). In this case, the loop was extracted from a ribozyme fragment 1GID. 3dRNA then made deletion and insertion operations on it to match the secondary structure needed, i.e., (..().().....()). After that, the 3D templates of other secondary structural elements were automatically searched for by the searching module of 3dRNA and then assembled to a whole all-atom structure by the assembling module of 3dRNA. Finally, the structure was refined by AMBER energy minimization. All calculations were performed on an Intel S5500BC server (Intel(R) Xeon (R) CPU E5620 @ 2.40 GHz). The template searching and assembling process took ~2 min. The AMBER energy minimization took ~3 min.

As above, for Problem 6, the Xiao group used Mfold and RNAfold to predict the secondary structure. The optimal secondary structure predicted by Mfold is a four-way junction structure and the optimal secondary structure predicted by RNAfold is made of three-way junction structures. The barrier for 3dRNA is the lack of the 3D templates for the four-way junction: (()).....()....()........). As previously, 3dRNA searched the templates library for the nearest loop. A four-way junction with the secondary structure (..()........()..... ()..) was picked out. It was extracted from rRNA 1C2W. After deletion and insertion operations, a 3D template of four-way junction with the secondary structure (()).....()....()........) was created. After that, the whole tertiary structure was assembled smoothly. Ten models were predicted for each of the optimal predicted secondary structures and then scored by 3dRNAscore. The lowest energy model was selected as the candidate and further refined with AMBER energy minimization. All calculations were performed on an Intel S5500BC server (Intel(R) Xeon(R) CPU E5620 @ 2.40 GHz). The whole template searching and assembling process took <3 sec this time as the 3dRNA has been reimplemented. The AMBER energy minimization process took 38 sec.

DISCUSSION

Except in some cases (Levitt 1969; Michel and Westhof 1990), RNA 3D structure prediction has historically lagged behind protein structure prediction, although RNA Watson–Crick pairing (secondary structure) is simpler to predict than, for example, β−sheet pairings in proteins. Nevertheless, compared to protein structure, RNA has more degrees of freedom. In addition, despite the limited number of non-Watson–Crick base pairs that simplifies analysis and inspection of tertiary structure, these non-Watson–Crick base pairs are difficult to recognize but central to the three-dimensional architecture of folded RNA molecules. The backbone of a nucleotide has six rotatable bonds, while an amino acid includes three (and the ω dihedral angle is generally fixed ~180° in peptide plane). Therefore, the RNA conformational landscape is potentially much larger and the three-dimensional structure prediction of RNAs with 100 nt is comparable, in terms of the number of degrees of freedom and molecular weight, to the challenge of modeling proteins of 200–300 aa. Although the structures in this round of RNA-Puzzles are large, topologies of the best predictions are not extreme compared with the native structures. This is a positive signal for RNA structure modeling.

In the current stage, most predictions can achieve good accuracy on Watson–Crick base pairs, while non-Watson–Crick interactions remain an open challenge that constitute an important bottleneck in RNA structure modeling. To improve non-WC pair prediction, RNA module prediction should be emphasized, since RNA modules are stable in structure but difficult to predict. Programs such as RMDetect (Cruz and Westhof 2011) could help in predicting RNA modules and improve the non-WC interaction prediction, e.g., for the K-turn in the adenosylcobalamin riboswitch which was recognized by modelers but not automatically. Unlike the numerous structures available for proteins, the number of RNA structures solved by crystallization is still limited and the available conformational space of RNA folding is far from complete. The prediction of non-WC pair and stacking could also be improved with the increase of known RNA structures and a complete search for RNA modules.

Other than module prediction programs, easy and fast experiments can provide direct constraints in structure modeling. In the protein structure prediction trials CASP Round X (Moult et al. 2014), a new category of "contact-assisted" prediction was proposed. Experimental data such as NMR, chemical shift, cross-linking, and surface labeling have been proved to be instrumental. Previously, contacts inferred from evolutionary information also achieved success in protein structure modeling (Morcos et al. 2011) but, at the time of writing, they still have not had an impact in blind structure prediction tests (Moult et al. 2014). Nevertheless, these explorations have revealed a trend in structure modeling: With the help of simple experimental constraints, structure modeling could achieve the application level in providing structural information for biological problems, even if no homologous structures are available. According to the three large RNA structures in this round of RNA-Puzzles, the modeling of RNA topology structures are already close to native, and the relative orientation between T-box and tRNA structures are recovered at a resolution (6.8 Å) comparable to the spacing between nucleotides.

Although the best predictions are similar to the topology of native RNA structures, some beauties in native structure topologies cannot be captured. As an example, the dramatic hole

in the ring structure formed by two helices in Problem 5 was not described at even nucleotide resolution by any prediction model. Current fast experiments can only help in detecting local and detailed interactions rather than global architectures determined by long-range contacts. These methods are mainly sensitive to detailed contacts but are largely uninformative as to global information such as holes. For consistent refinement to higher resolution 3D models that predict these striking features, we still need deeper understanding of RNA structure and/or new fast experimental tools.

Currently, the major challenges in RNA structure prediction lie in (1) further improvement of algorithms that incorporate simple experimental data (contact-assisted data), (2) structure optimization to alleviate atomic clashes and improve accuracy, and (3) accumulation of comprehensive RNA structure knowledge with the help of database increases and automated structural bioinformatic tools. The surprising high values for the clash scores in several otherwise respectable models led to attempts to improve the clash score values by rerefinement. The Das group had run ERRASER (Chou et al. 2013), but ERRASER only works to find solutions with each nucleotide within ~2 Å RMSD of the starting solution. Many of the derived models used fragments of other crystal structures that did not fit well together. The relief of chain-breaks and clashes require bigger changes than ERRASER can currently handle. After an exchange of the previous versions of this article, the Bujnicki group ran their refinement method QRNAS on the Das models (all 17 models submitted for the three Problems). QRNAS is essentially a reimplementation of AMBER with additional regularization and it is used as the final element in the Bujnicki modeling pipeline. In all cases, a dramatic reduction of Clash Scores was obtained; in 10 models even down to zero. Only in three cases the Clash Scores remained larger than 4; however, these models had initial Clash Scores of nearly 30. Supplemental Table S2 shows the values of all the metrics used for comparisons and most of them display an improvement or at least no worsening. However, the bond angle deviations increase severely in all cases, a not so surprising result since that parameter was kept free during optimization. Thus, further work is required for resolving clashes in automatically derived models.

As attested by the number of coauthors involved in these three RNA Puzzles in most modeling groups, the automaticity of the 3D structure prediction process still requires a major investment in computer science and in the development of user-friendly and straightforward computer tools. Therefore, in order to make RNA 3D structure prediction available to the biological community in solving biological problems, we encourage web servers for automatic RNA 3D structure prediction. Such web servers should take query sequences, probably together with simple experimental data, and return possible RNA 3D coordinates. As described, several groups have already advanced in this direction. As inspiration, in recent years, servers have largely caught up with human expert groups in protein structure prediction (Moult et al. 2014), and it will be interesting to see if the RNA community can accomplish the same.

Finally, in the present comparisons, it is assumed that the crystal structure is the relevant and correct target. Crystallographic structures constitute highly relevant models representing with high precision and accuracy particular experiments and conditions. However, not all segments of crystallized structures are at the same level of accuracy, because of resolution issues, disorder, or high segmental mobilities (as represented by the thermal B-factors). For those segments, the uncertainty of the reference structure is a real question. A meaningful comparison would thus require that the prediction programs derive also a theoretical B-factor for the nucleotides representing some aspects of the uncertainty in the prediction. Preliminary results indicate that regions with high experimental B-factors correlate with regions in disagreement with the rest of the structure (regions in red color in the deformation profiles). Thoughtful weighted comparisons need to be developed to address these issues of molecular dynamics during comparisons between crystal structures and predicted models.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

This work (Z.M. and E.W.) is supported by the French Government grant ANR-10-BINF-02-02 <BACNET>. This work (E.W.) has been published under the framework of the LABEX: ANR-10-LABX-0036_NETRNA and benefits from a funding managed by the French National Research Agency as part of the Investments for the future program. N.V.D. acknowledges support of the National Institutes of Health (NIH) grant 2R01GM064803-09 (PI: K.M. Weeks). The 3D structure modeling was supported by the NIH (5 T32 GM007276 to C.C.; R01 GM102519 to R.D.), the Burroughs-Wellcome Foundation (CASI to R.D.), Bio-X and HHMI international fellowships (F.C.C.), a Stanford Graduate Fellowship (S.T.), a Conacyt fellowship (P.C.), European Research Council (grant RNA+P=123D to J.M.B.), Foundation for Polish Science (grant TEAM/2009-4/2 to J.M.B.), European Commission (EU structural funds, POIG.02.03.00-00-003/09 to J.M.B.), by the Polish National Science Centre (2012/06/A/ST6/00384 to R.W.A. and 2012/04/A/NZ2/00455 to J.M.B.) and Polish Ministry of Science and Higher Education (0492/IP1/2013/72 to K.J.P.). A.P. was supported by the NIH grant T32 GM088118. A.S. was supported by NYU Whitehead fellowship and the NIH grant R21MH103655. A.R.F. and J.Z. were supported by the Intramural Program of the National Heart, Lung and Blood Institute-NIH.

Received January 10, 2015; accepted February 12, 2015.

REFERENCES

Adams PL, Stahley MR, Kosek AB, Wang JM, Strobel SA. 2004. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430:** 45–50.

Barrick JE, Breaker RR. 2007. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* **8:** R239.

Beckert B, Nielsen H, Einvik C, Johansen SD, Westhof E, Masquida B. 2008. Molecular modelling of the GIR1 branching ribozyme gives new insight into evolution of structurally related ribozymes. *EMBO J* **27:** 667–678.

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. 2000. The protein data bank. *Nucleic Acids Res* **28:** 235–242.

Burge SW, Daub J, Eberhardt R, Tate J, Barquist L, Nawrocki EP, Eddy SR, Gardner PP, Bateman A. 2013. Rfam 11.0: 10 years of RNA families. *Nucleic Acids Res* **41:** D226–D232.

Cao S, Chen SJ. 2005. Predicting RNA folding thermodynamics with a reduced chain representation model. *RNA* **11:** 1884–1897.

Cao S, Chen SJ. 2006. Predicting RNA pseudoknot folding thermodynamics. *Nucleic Acids Res* **34:** 2634–2652.

Cao S, Chen SJ. 2009. Predicting structures and stabilities for H-type pseudoknots with interhelix loops. *RNA* **15:** 696–706.

Cao S, Chen SJ. 2011. Physics-based de novo prediction of RNA 3D structures. *J Phys Chem B* **115:** 4216–4226.

Capriotti E, Norambuena T, Marti-Renom MA, Melo F. 2011. All-atom knowledge-based potential for RNA structure prediction and assessment. *Bioinformatics* **27:** 1086–1093.

Chen SJ. 2008. RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu Rev Biophys* **37:** 197–214.

Chen VB, Arendall WB III, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC. 2010. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr* **66:** 12–21.

Chou FC, Sripakdeevong P, Dibrov SM, Hermann T, Das R. 2013. Correcting pervasive errors in RNA crystallography through enumerative structure prediction. *Nat Methods* **10:** 74–76.

Cordero P, Kladwang W, VanLang CC, Das R. 2012a. Quantitative dimethyl sulfate mapping for automated RNA secondary structure inference. *Biochemistry* **51:** 7037–7039.

Cordero P, Lucks JB, Das R. 2012b. An RNA mapping dataBase for curating RNA structure mapping experiments. *Bioinformatics* **28:** 3006–3008.

Cordero P, Kladwang W, VanLang CC, Das R. 2014. The mutate-and-map protocol for inferring base pairs in structured RNA. *Methods Mol Biol* **1086:** 53–77.

Cruz JA, Westhof E. 2011. Sequence-based identification of 3D structural modules in RNA with RMDetect. *Nat Methods* **8:** 513–521.

Cruz JA, Blanchet MF, Boniecki M, Bujnicki JM, Chen SJ, Cao S, Das R, Ding F, Dokholyan NV, Flores SC, et al. 2012. RNA-Puzzles: a CASP-like evaluation of RNA three-dimensional structure prediction. *RNA* **18:** 610–625.

Das R, Karanicolas J, Baker D. 2010. Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods* **7:** 291–294.

Davis IW, Leaver-Fay A, Chen VB, Block JN, Kapral GJ, Wang X, Murray LW, Arendall WB 3rd, Snoeyink J, Richardson JS, et al. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res* **35:** W375–W383.

Ding F, Sharma S, Chalasani P, Demidov VV, Broude NE, Dokholyan NV. 2008. Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA* **14:** 1164–1173.

Ding F, Lavender CA, Weeks KM, Dokholyan NV. 2012. Three-dimensional RNA structure refinement by hydroxyl radical probing. *Nat Methods* **9:** 603–608.

Dunkle JA, Wang L, Feldman MB, Pulk A, Chen VB, Kapral GJ, Noeske J, Richardson JS, Blanchard SC, Cate JH. 2011. Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332:** 981–984.

Feng Z, Westbrook J, Berman HM. 1998. *NUCheck*. NDB-407 Rutgers University, New Brunswick, NJ.

Fukai S, Nureki O, Sekine S, Shimada A, Tao J, Vassylyev DG, Yokoyama S. 2000. Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA(Val) and valyl-tRNA synthetase. *Cell* **103:** 793–803.

Geary C, Chworos A, Jaeger L. 2011. Promoting RNA helical stacking via A-minor junctions. *Nucleic Acids Res* **39:** 1066–1080.

Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. 2003. Rfam: an RNA family database. *Nucleic Acids Res* **31:** 439–441.

Grigg JC, Chen Y, Grundy FJ, Henkin TM, Pollack L, Ke A. 2013. T box RNA decodes both the information content and geometry of tRNA to affect gene expression. *Proc Natl Acad Sci* **110:** 7240–7245.

Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM. 2013. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc Natl Acad Sci* **110:** 5498–5503.

Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. 1994. Fast folding and comparison of RNA secondary structures. *Monatsh Chem* **125:** 167–188.

Kladwang W, VanLang CC, Cordero P, Das R. 2011a. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat Chem* **3:** 954–962.

Kladwang W, VanLang CC, Cordero P, Das R. 2011b. Understanding the errors of SHAPE-directed RNA structure modeling. *Biochemistry* **50:** 8049–8056.

Kladwang W, Mann TH, Becka A, Tian S, Kim H, Yoon S, Das R. 2014. Standardization of RNA chemical mapping experiments. *Biochemistry* **53:** 3063–3065.

Leaver-Fay A, Tyka M, Lewis SM, Lange OF, Thompson J, Jacak R, Kaufman K, Renfrew PD, Smith CA, Sheffler W, et al. 2011. ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. *Methods Enzymol* **487:** 545–574.

Lehmann J, Jossinet F, Gautheret D. 2013. A universal RNA structural motif docking the elbow of tRNA in the ribosome, RNAse P and T-box leaders. *Nucleic Acids Res* **41:** 5494–5502.

Leonard CW, Hajdin CE, Karabiber F, Mathews DH, Favorov OV, Dokholyan NV, Weeks KM. 2013. Principles for understanding the accuracy of SHAPE-directed RNA structure modeling. *Biochemistry* **52:** 588–595.

Levitt M. 1969. Detailed molecular model for transfer ribonucleic acid. *Nature* **224:** 759–763.

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127:** 4223–4231.

Meyer M, Nielsen H, Olieric V, Roblin P, Johansen SD, Westhof E, Masquida B. 2014. Speciation of a group I intron into a lariat capping ribozyme. *Proc Natl Acad Sci* **111:** 7659–7664.

Michel F, Westhof E. 1990. Modelling of the three-dimensional architecture of group I catalytic introns based on comparative sequence analysis. *J Mol Biol* **216:** 585–610.

Morcos F, Pagnani A, Lunt B, Bertolino A, Marks DS, Sander C, Zecchina R, Onuchic JN, Hwa T, Weigt M. 2011. Direct-coupling analysis of residue coevolution captures native contacts across many protein families. *Proc Natl Acad Sci* **108:** E1293–E1301.

Mortimer SA, Weeks KM. 2007. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* **129:** 4144–4145.

Moult J. 2008. Comparative modeling in structural genomics. *Structure* **16:** 14–16.

Moult J, Fidelis K, Kryshtafovych A, Schwede T, Tramontano A. 2014. Critical assessment of methods of protein structure prediction (CASP)—round x. *Proteins* **82:** 1–6.

Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR. 2002. Genetic control by a metabolite binding mRNA. *Chem Biol* **9:** 1043.

Nahvi A, Barrick JE, Breaker RR. 2004. Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res* **32:** 143–150.

Parisien M, Major F. 2008. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452:** 51–55.

Parisien M, Cruz JA, Westhof E, Major F. 2009. New metrics for comparing and assessing discrepancies between RNA 3D structures and models. *RNA* **15:** 1875–1885.

Peselis A, Serganov A. 2012. Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nat Struct Mol Biol* **19:** 1182–1184.

Popenda M, Blazewicz M, Szachniuk M, Adamiak RW. 2008. RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures. *Nucleic Acids Res* **36:** D386–D391.

Popenda M, Szachniuk M, Antczak M, Purzycka KJ, Lukasiak P, Bartol N, Blazewicz J, Adamiak RW. 2012. Automated 3D structure composition for large RNAs. *Nucleic Acids Res* **40:** e112.

Proctor EA, Ding F, Dokholyan NV. 2011. Discrete molecular dynamics. *Wiley Interdiscip Rev Comput Mol Sci* **1:** 80–92.

Puton T, Kozlowski LP, Rother KM, Bujnicki JM. 2013. CompaRNA: a server for continuous benchmarking of automated methods for RNA secondary structure prediction. *Nucleic Acids Res* **41:** 4307–4323.

Ramachandran S, Ding F, Weeks KM, Dokholyan NV. 2013. Statistical analysis of SHAPE-directed RNA secondary structure modeling. *Biochemistry* **52:** 596–599.

Ravnum S, Andersson DI. 2001. An adenosyl-cobalamin (coenzyme-B12)-repressed translational enhancer in the *cob* mRNA of *Salmonella typhimurium*. *Mol Microbiol* **39:** 1585–1594.

Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11:** 129.

Rice GM, Leonard CW, Weeks KM. 2014. RNA secondary structure modeling at consistent high accuracy using differential SHAPE. *RNA* **20:** 846–854.

Rocca-Serra P, Bellaousov S, Birmingham A, Chen C, Cordero P, Das R, Davis-Neulander L, Duncan CD, Halvorsen M, Knight R, et al. 2011. Sharing and archiving nucleic acid structure mapping data. *RNA* **17:** 1204–1212.

Rother M, Rother K, Puton T, Bujnicki JM. 2011. ModeRNA: a tool for comparative modeling of RNA 3D structure. *Nucleic Acids Res* **39:** 4007–4022.

Rother K, Rother M, Boniecki M, Puton T, Tomala K, Lukasz P, Bujnicki JM. 2012. Template-based and template-free modeling of RNA 3D structure: inspirations from protein structure modeling. In *RNA 3D structure analysis and prediction* (ed. Leontis NB, Westhof E). Springer, Berlin.

Rozhdestvensky TS, Tang TH, Tchirkova IV, Brosius J, Bachellerie JP, Hüttenhofer A. 2003. Binding of L7Ae protein to the K-turn of archaeal snoRNAs: a shared RNA binding motif for C/D and H/ACA box snoRNAs in Archaea. *Nucleic Acids Res* **31:** 869–877.

Seetin MG, Kladwang W, Bida JP, Das R. 2014. Massively parallel RNA chemical mapping with a reduced bias MAP-seq protocol. *Methods Mol Biol* **1086:** 95–117.

Shirvanyants D, Ding F, Tsao D, Ramachandran S, Dokholyan NV. 2012. Discrete molecular dynamics: an efficient and versatile simulation method for fine protein characterization. *J Phys Chem B* **116:** 8375–8382.

Sripakdeevong P, Kladwang W, Das R. 2011. An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling. *Proc Natl Acad Sci* **108:** 20573–20578.

Tian S, Cordero P, Kladwang W, Das R. 2014. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA* **20:** 1815–1826.

Vidovic I, Nottrott S, Hartmuth K, Lührmann R, Ficner R. 2000. Crystal structure of the spliceosomal 15.5kD protein bound to a U4 snRNA fragment. *Mol Cell* **6:** 1331–1342.

Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. 2003. Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA* **9:** 1084–1097.

Wang JC, Henkin TM, Nikonowicz EP. 2010. NMR structure and dynamics of the Specifier Loop domain from the *Bacillus subtilis* tyrS T box leader RNA. *Nucleic Acids Res* **38:** 3388–3398.

Xu XJ, Zhao PN, Chen SJ. 2014. Vfold: a web server for RNA structure and folding thermodynamics prediction. *PLoS One* **9:** e107504.

Yang X, Gerczei T, Glover LT, Correll CC. 2001. Crystal structures of restrictocin-inhibitor complexes with implications for RNA recognition and base flipping. *Nat Struct Biol* **8:** 968–973.

Zhang JW, Ferré-D'Amaré AR. 2013. Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA. *Nature* **500:** 363–366.

Zhao YJ, Huang YY, Gong Z, Wang YJ, Man JF, Xiao Y. 2012. Automated and fast building of three-dimensional RNA structures. *Sci Rep* **2:** 734.

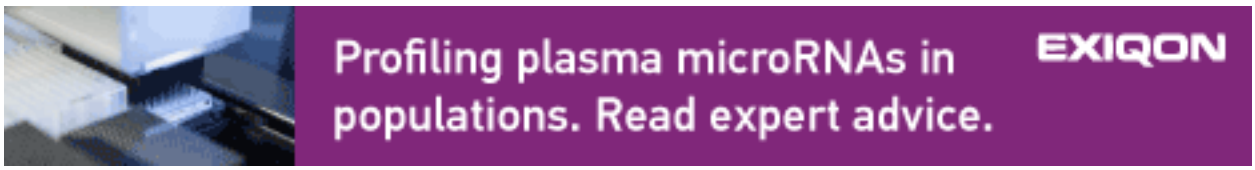
Zok T, Popenda M, Szachniuk M. 2014. MCQ4Structures to compute similarity of molecule structures. *Cent Eur J Oper Res* **22:** 457–473.

Zuker M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31:** 3406–3415.

RNA-Puzzles Round II: assessment of RNA structure prediction programs applied to three large RNA structures

Zhichao Miao, Ryszard W. Adamiak, Marc-Frédérick Blanchet, et al.

RNA published online April 16, 2015



RNA-Puzzles Round II: Assessment of RNA structure prediction programs applied to three large RNA structures

Zhichao Miao[1], Ryszard W. Adamiak[2], Marc-Frédérick Blanchet[3], Michal Boniecki[4], Janusz M. Bujnicki[4,5], Shi-Jie Chen[6], Clarence Cheng[7], Grzegorz Chojnowski[4], Fang-Chieh Chou[7], Pablo Cordero[7], José Almeida Cruz[1], Adrian Ferre-D'Amare[8], Rhiju Das[7], Feng Ding[9], Nikolay V. Dokholyan[10], Stanislaw Dunin-Horkawicz[4], Wipapat Kladwang[7], Andrey Krokhotin[10], Grzegorz Lach[4], Marcin Magnus[4], François Major[3], Thomas H. Mann[7], Benoît Masquida[11], Dorota Matelska[4], Mélanie Meyer[12], Alla Peselis[13], Mariusz Popenda[2], Katarzyna J. Purzycka[2], Alexander Serganov[13], Juliusz Stasiewicz[4], Marta Szachniuk[15], Arpit Tandon[10], Siqi Tian[7], Jian Wang[14], Yi Xiao[14], Xiaojun Xu[6], Jinwei Zhang[8], Peinan Zhao[6], Tomasz Zok[15] and Eric Westhof[1,*]

[1]Architecture et Réactivité de l'ARN, Université de Strasbourg, Institut de biologie moléculaire et cellulaire du CNRS, 67000 Strasbourg France; [2]Department of Structural Chemistry and Biology of Nucleic Acids, Structural Chemistry of Nucleic Acids Laboratory, Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland; [3]Institute for Research in Immunology and Cancer (IRIC), Department of Computer Science and Operations Research, Université de Montréal, Montréal, Québec H3C 3J7, Canada; [4]Laboratory of Bioinformatics and Protein Engineering, International Institute of Molecular and Cell Biology in Warsaw, 02-109 Warsaw, Poland; [5]Laboratory of Bioinformatics, Institute of Molecular Biology and Biotechnology, Faculty of Biology, Adam Mickiewicz University, 61-614 Poznan, Poland; [6]Department of Physics and Astronomy, Department of Biochemistry, University of Missouri Informatics Institute, University of Missouri-Columbia, MO 65211, U.S.A.; [7]Department of Physics, Stanford University, Stanford, California 94305, USA; [8]National Heart, Lung and Blood Institute, 50 South Drive, MSC 8012, Bethesda, Maryland 20892-8012, USA; [9]Department of Physics and Astronomy at Clemson University, College of Engineering and Science, USA; [10]Department of Biochemistry and Biophysics, University of North Carolina, School of Medicine, Chapel Hill, North Carolina, USA;[11]Génétique Moléculaire Génomique Microbiologie, Institut de physiologie et de la chimie biologique, 21 rue René Descartes 67084 Strasbourg, France;[12]Institut de génétique et de biologie moléculaire et cellulaire, 1 Rue Laurent Fries, 67400 Strasbourg, France; [13]Department of Biochemistry and Molecular Pharmacology, New York University School of Medicine, New York, New York, USA, [14]Department of Physics, Huazhong University of Science and Technology, Wuhan, China; [15]Poznan University of Technology, Institute of Computing Science, Poznan, Poland.

Supplementary information

TABLE S1.Summary of experimental data			base-pair sensitivity	ppv	helix sensitivity	ppv
Puzzle 5	No data	5.4 Fold	50.0%	48.4%	46.2%	42.9%
		5.6 Fold	50.0%	48.4%	46.2%	42.9%
		5.6 ShapeKnots	50.0%	50.0%	46.2%	42.9%
	1D SHAPE	5.4 Fold	45.0%	41.5%	38.5%	35.7%
		5.6 Fold	45.0%	41.5%	69.2%	35.7%
		5.6 ShapeKnots	78.3%	72.3%	46.2%	69.2%
	1D DMS/CMCT	5.4 Fold	58.3%	53.9%	46.2%	42.9%
		5.6 Fold	58.3%	53.9%	76.9%	42.9%
		5.6 ShapeKnots	81.7%	79.0%	61.5%	76.9%
	1D SHAPE/DMS/CMCT	5.6 Fold	66.7%	60.6%	61.5%	53.3%
		5.6 ShapeKnots	73.3%	67.7%	69.2%	66.7%
	2D SHAPE	5.4 Fold	76.7%	74.2%	69.2%	69.2%
		5.6 Fold	76.7%	74.2%	84.6%	69.2%
		5.6 ShapeKnots	90.0%	85.7%	92.3%	91.7%
	2D DMS	5.4 Fold	81.7%	77.8%	76.9%	71.4%
		5.6 Fold	81.7%	77.8%	76.9%	71.4%
		5.6 ShapeKnots	93.3%	86.2%	92.3%	85.7%
Puzzle 6	No data	5.4 Fold	73.7%	73.7%	54.6%	54.6%
		5.6 Fold	73.7%	73.7%	54.6%	54.6%
		5.6 ShapeKnots	73.7%	73.7%	54.6%	54.6%
	1D SHAPE	5.4 Fold	77.2%	81.5%	63.6%	77.8%
		5.6 Fold	73.7%	73.7%	54.6%	54.6%
		5.6 ShapeKnots	79.0%	81.8%	63.6%	77.8%
	1D DMS/CMCT	5.4 Fold	63.2%	64.3%	54.6%	50.0%
		5.6 Fold	63.2%	64.3%	54.6%	50.0%
		5.6 ShapeKnots	71.9%	70.7%	54.6%	54.6%
	1D SHAPE/DMS/CMCT	5.6 Fold	49.1%	57.1%	36.4%	33.3%
		5.6 ShapeKnots	63.2%	72.0%	63.6%	58.3%
	2D SHAPE	5.4 Fold	96.5%	96.5%	90.9%	90.9%
		5.6 Fold	96.5%	96.5%	90.9%	90.9%
		5.6 ShapeKnots	96.5%	96.5%	90.9%	90.9%
Puzzle 10	No data	5.4 Fold	83.0%	79.6%	88.9%	80.0%
		5.6 Fold	83.0%	79.6%	88.9%	80.0%
		5.6 ShapeKnots	83.0%	79.6%	88.9%	80.0%
	1D SHAPE	5.4 Fold	100.0%	97.9%	100.0%	100.0%
		5.6 Fold	87.2%	82.0%	88.9%	80.0%
		5.6 ShapeKnots	97.9%	92.0%	100.0%	90.0%
	1D DMS/CMCT	5.4 Fold	72.3%	66.7%	77.8%	70.0%
		5.6 Fold	72.3%	66.7%	77.8%	70.0%
		5.6 ShapeKnots	85.1%	71.4%	88.9%	72.7%

	1D SHAPE/DMS/CMCT	5.6 Fold	55.3%	55.3%	55.6%	50.0%
		5.6 ShapeKnots	63.8%	61.2%	66.7%	60.0%

Structure Refinement based on Das lab models done by Bujnicki lab.

Lab	Num	RMSD	P-value	DI	INF	INF_wc	INF_nwc	INF_stacking	clash	pct_badbonds	pct_resbadbonds	pct_badangles	pct_resbadar
DasRef*	1	9.96	0.00E+000	13.002	0.766	0.919	0.256	0.766	0	0.7	7.45	8.33	
DasRef	2	9.165	0.00E+000	11.94	0.768	0.908	0.344	0.757	0.17	0.25	1.6	7.07	
										0.475	4.525	7.7	
Das	1	9.948	0.00E+000	13.148	0.757	0.919	0.256	0.751	9.44	0.74	9.57	1.48	2
Das	2	9.152	0.00E+000	12.019	0.761	0.906	0.334	0.751	6.79	0.49	6.38	1.52	3
										0.615	7.975	1.5	29
Lab	Num	RMSD	P-value	DI	INF	INF_wc	INF_nwc	INF_stacking	clash	pct_badbonds	pct_resbadbonds	pct_badangles	pct_resbadar
DasRef	1	14.504	3.23E-009	19.022	0.762	0.897	0.433	0.746	3.85	0.41	2.38	7.88	
DasRef	2	13.683	1.92E-010	17.889	0.765	0.905	0.361	0.755	0	0	0	7.04	
DasRef	3	15.778	1.73E-007	21.019	0.751	0.889	0.334	0.744	0.18	0.05	0.6	7.5	
DasRef	4	11.714	9.59E-014	15.595	0.751	0.897	0.416	0.731	5.32	0.78	5.95	9.34	
DasRef	5	15.862	2.22E-007	20.509	0.773	0.897	0.433	0.763	0	0	0	6.8	
DasRef	6	12.421	1.68E-012	16.464	0.754	0.897	0.433	0.734	3.85	0.37	1.79	7.77	
DasRef	7	17.959	5.12E-005	24.221	0.741	0.877	0.316	0.732	0	0	0	7.15	
DasRef	8	17.97	5.26E-005	23.451	0.766	0.897	0.416	0.755	0	0.82	5.95	8.32	
DasRef	9	15.107	2.26E-008	19.785	0.764	0.905	0.316	0.755	0	0.05	0.6	6.72	
DasRef	10	29.226	9.91E-001	39.946	0.732	0.897	0.312	0.718	0.37	0.41	2.38	8.02	
										0.289	1.965	7.654	
Das	1	14.478	2.96E-009	19.893	0.728	0.885	0.333	0.705	23.85	0.64	8.33	0.87	1
Das	2	13.627	1.57E-010	18.774	0.726	0.885	0.347	0.705	17.24	0.6	7.14	0.84	1
Das	3	15.752	1.60E-007	21.759	0.724	0.874	0.217	0.72	17.05	0.55	7.14	0.62	
Das	4	11.699	9.02E-014	16.151	0.724	0.885	0.316	0.702	23.48	0.64	8.33	0.92	1
Das	5	15.834	2.04E-007	21.147	0.749	0.885	0.361	0.738	13.39	0.37	4.76	0.65	
Das	6	12.405	1.58E-012	17.08	0.726	0.885	0.333	0.702	24.59	0.64	8.33	0.92	
Das	7	17.875	4.22E-005	24.363	0.734	0.869	0.237	0.731	13.21	0.55	7.14	0.6	
Das	8	17.956	5.08E-005	24.06	0.746	0.877	0.334	0.743	13.02	0.41	5.36	0.62	
Das	9	15.048	1.88E-008	20.942	0.719	0.885	0.347	0.694	17.98	0.6	7.14	0.79	1
Das	10	29.182	9.91E-001	41.226	0.708	0.877	0.25	0.698	19.82	0.64	7.74	0.62	
										0.564	7.141	0.745	9
Lab	Num	RMSD	P-value	DI	INF	INF_wc	INF_nwc	INF_stacking	clash	pct_badbonds	pct_resbadbonds	pct_badangles	pct_resbadar
DasRef	1	7.64	0.00E+000	8.876	0.861	0.929	0.7	0.861	0	0	0	6.67	9
DasRef	2	10.539	1.50E-015	12.191	0.864	0.938	0.802	0.847	0	0	0	6.59	9
DasRef	3	6.837	0.00E+000	8.157	0.838	0.936	0.717	0.823	0	0	0	6.46	9
DasRef	4	7.077	0.00E+000	8.305	0.852	0.938	0.717	0.842	0	0	0	6.7	9
DasRef	5	10.482	1.17E-015	11.944	0.878	0.938	0.778	0.87	0	0	0	6.73	9
										0	0	6.63	9
Das	1	7.58	0.00E+000	9.199	0.824	0.92	0.7	0.811	11.64	0.36	4.09	0.56	
Das	2	10.447	9.99E-016	12.588	0.83	0.929	0.778	0.803	11.64	0.36	4.09	0.56	
Das	3	6.803	0.00E+000	8.365	0.813	0.946	0.7	0.786	11.09	0.41	4.68	0.64	
Das	4	7.062	0.00E+000	8.539	0.827	0.948	0.684	0.809	10.73	0.41	4.68	0.56	
Das	5	10.417	8.88E-016	12.295	0.847	0.948	0.778	0.823	10.91	0.27	2.92	0.45	
										0.362	4.092	0.554	8

Figure S1. Chemical mapping data and secondary structure predictions of RNA Puzzle 5.

(A) Secondary structure from crystallographic structures.

(B) Secondary structure prediction using no experimental data with RNAstructure 5.4 or 5.6 *Fold*. Nucleotides are colored according with SHAPE reactivities. Crystallographic pairings missing in this model and new non-crystallographic pairings are drawn as yellow and blue lines, respectively. Percentage labels give bootstrap support values.

(C) Secondary structure prediction using no data with RNAstructure 5.6 *ShapeKnots*.

(D) Secondary structure prediction using 1D SHAPE data with RNAstructure 5.6 *Fold*.

(E) Secondary structure prediction using 1D SHAPE data with RNAstructure 5.6 *ShapeKnots*.

(F) Secondary structure prediction using 1D DMS/CMCT data with RNAstructure 5.4 *Fold*.

(G) Secondary structure prediction using 1D DMS/CMCT data with RNAstructure 5.6 *Fold*.

(H) Secondary structure prediction using 1D DMS/CMCT data with RNAstructure 5.6 *ShapeKnots*.

(I) Secondary structure prediction using 1D SHAPE and DMS/CMCT data with RNAstructure 5.6 *Fold*.

(J) Secondary structure prediction using 1D SHAPE and DMS/CMCT data with RNAstructure 5.6 *ShapeKnots*.

(K) Secondary structure prediction using 2D SHAPE M^2 data with RNAstructure 5.6 *Fold*.

(L) Secondary structure prediction using 2D SHAPE M^2 data with RNAstructure 5.6 *ShapeKnots*.

(M) Mutate-and-map (M^2) dataset probed by the DMS.

(N) Secondary structure prediction using 2D SHAPE M^2 data with RNAstructure 5.4 *Fold*.

(O) Secondary structure prediction using 2D DMS M^2 data with RNAstructure 5.6 *Fold*.

(P) Secondary structure prediction using 2D DMS M^2 data with RNAstructure 5.6 *ShapeKnots*.

Figure S2. Chemical mapping data and secondary structure predictions of RNA Puzzle 6.

(A) Secondary structure from crystallographic structures.

(B) Secondary structure prediction without experimental data with RNAstructure 5.4 or 5.6 *Fold*. Nucleotides are colored with SHAPE reactivities. Crystallographic pairings missing in this model and new non-crystallographic pairings are drawn as yellow and blue lines, respectively. Percentage labels give bootstrap support values.

(C) Secondary structure prediction using no data with RNAstructure 5.6 *ShapeKnots*.

(D) Secondary structure prediction using 1D SHAPE data with RNAstructure 5.6 *Fold*.
(E) Secondary structure prediction using 1D SHAPE data with RNAstructure 5.6 *ShapeKnots*.
(F) Secondary structure prediction using 1D DMS/CMCT data with RNAstructure 5.4 *Fold*.
(G) Secondary structure prediction using 1D DMS/CMCT data with RNAstructure 5.6 *Fold*.
(H) Secondary structure prediction using 1D DMS/CMCT data with RNAstructure 5.6 *ShapeKnots*.
(I) Secondary structure prediction using 1D SHAPE and DMS/CMCT data with RNAstructure 5.6 *Fold*.
(J) Secondary structure prediction using 1D SHAPE and DMS/CMCT data with RNAstructure 5.6 *ShapeKnots*.
(K) Secondary structure prediction using 2D SHAPE M^2 data with RNAstructure 5.6 *Fold*.
(L) Secondary structure prediction using 2D SHAPE M^2 data with RNAstructure 5.6 *ShapeKnots*.

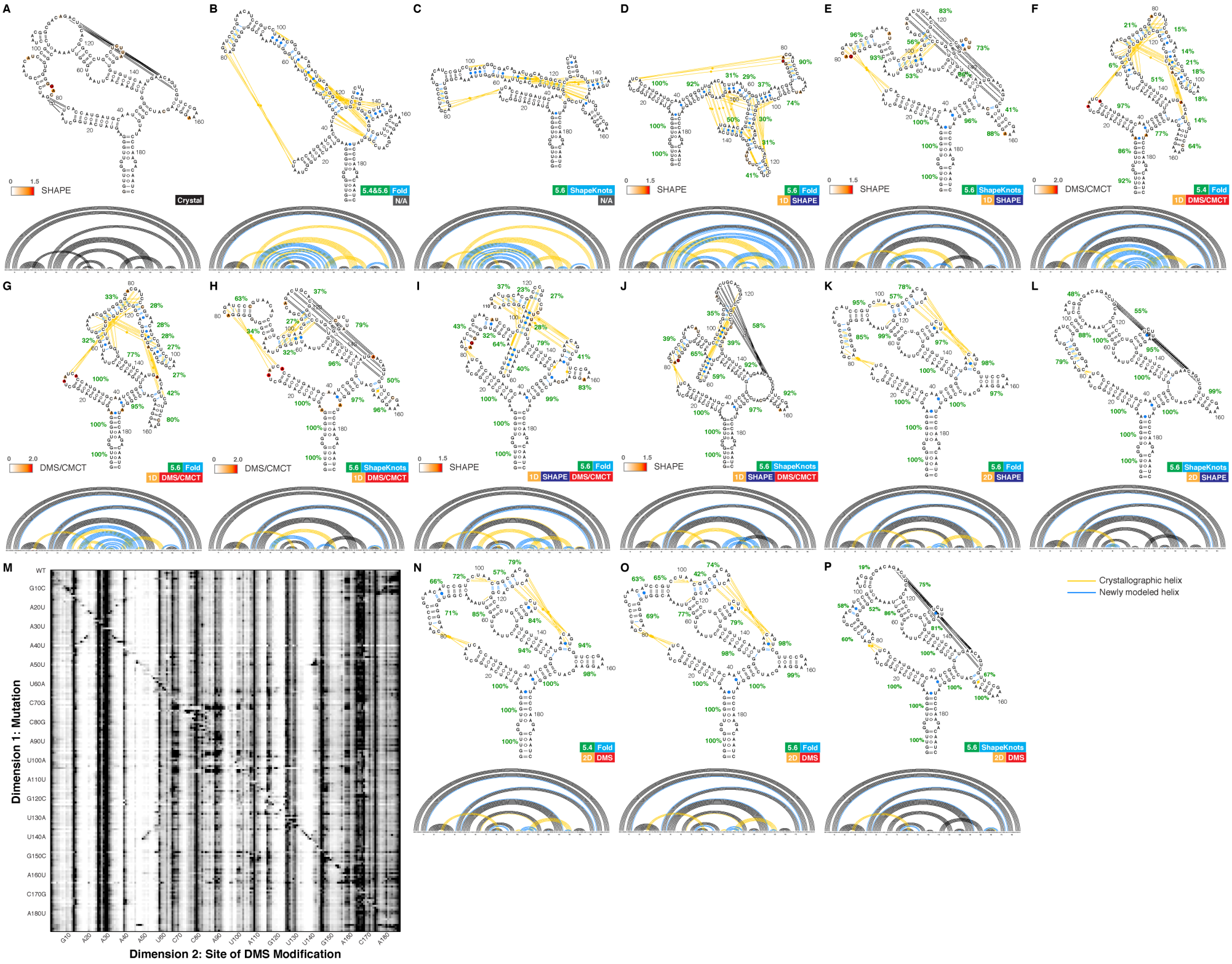
Figure S3. Chemical mapping data and secondary structure predictions of RNA Puzzle 10.
(A) Secondary structure from crystallographic structures.
(B) Secondary structure prediction without experimental data with RNAstructure 5.4 or 5.6 *Fold*. Nucleotides are colored with SHAPE reactivities. Crystallographic pairings missing in this model and new non-crystallographic pairings are drawn as yellow and blue lines, respectively. Percentage labels give bootstrap support values.
(C) Secondary structure prediction using no data with RNAstructure 5.6 *ShapeKnots*.
(D) Secondary structure prediction using 1D SHAPE data with RNAstructure 5.6 *Fold*.
(E) Secondary structure prediction using 1D SHAPE data with RNAstructure 5.6 *ShapeKnots*.
(F) Secondary structure prediction using 1D DMS/CMCT data with RNAstructure 5.4 *Fold*.
(G) Secondary structure prediction using 1D DMS/CMCT data with RNAstructure 5.6 *Fold*.
(H) Secondary structure prediction using 1D DMS/CMCT data with RNAstructure 5.6 *ShapeKnots*.
(I) Secondary structure prediction using 1D SHAPE and DMS/CMCT data with RNAstructure 5.6 *Fold*.
(J) Secondary structure prediction using 1D SHAPE and DMS/CMCT data with RNAstructure 5.6 *ShapeKnots*.

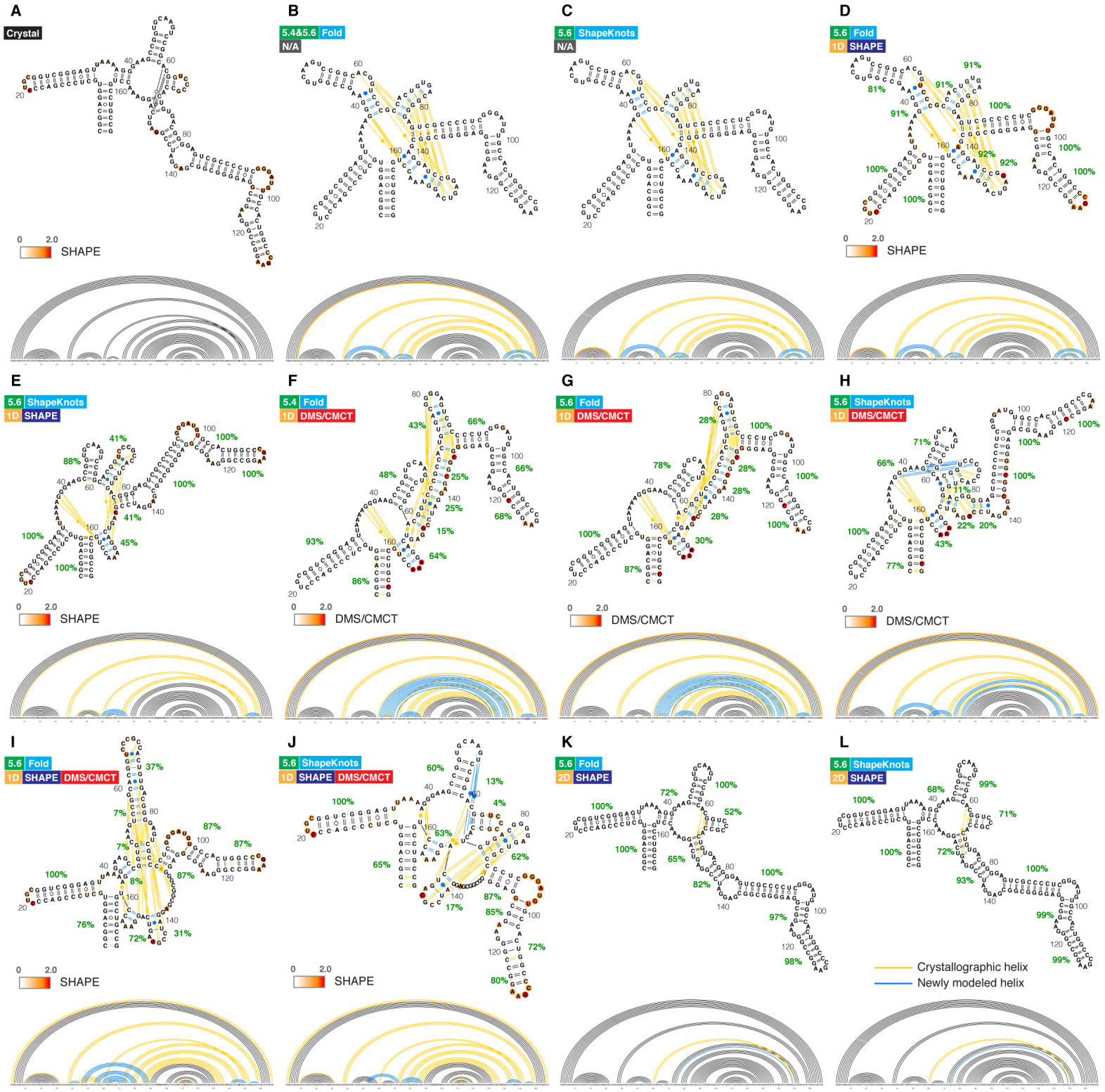
Figure S4. Chemical mapping data and secondary structure predictions of RNA Puzzle 5.

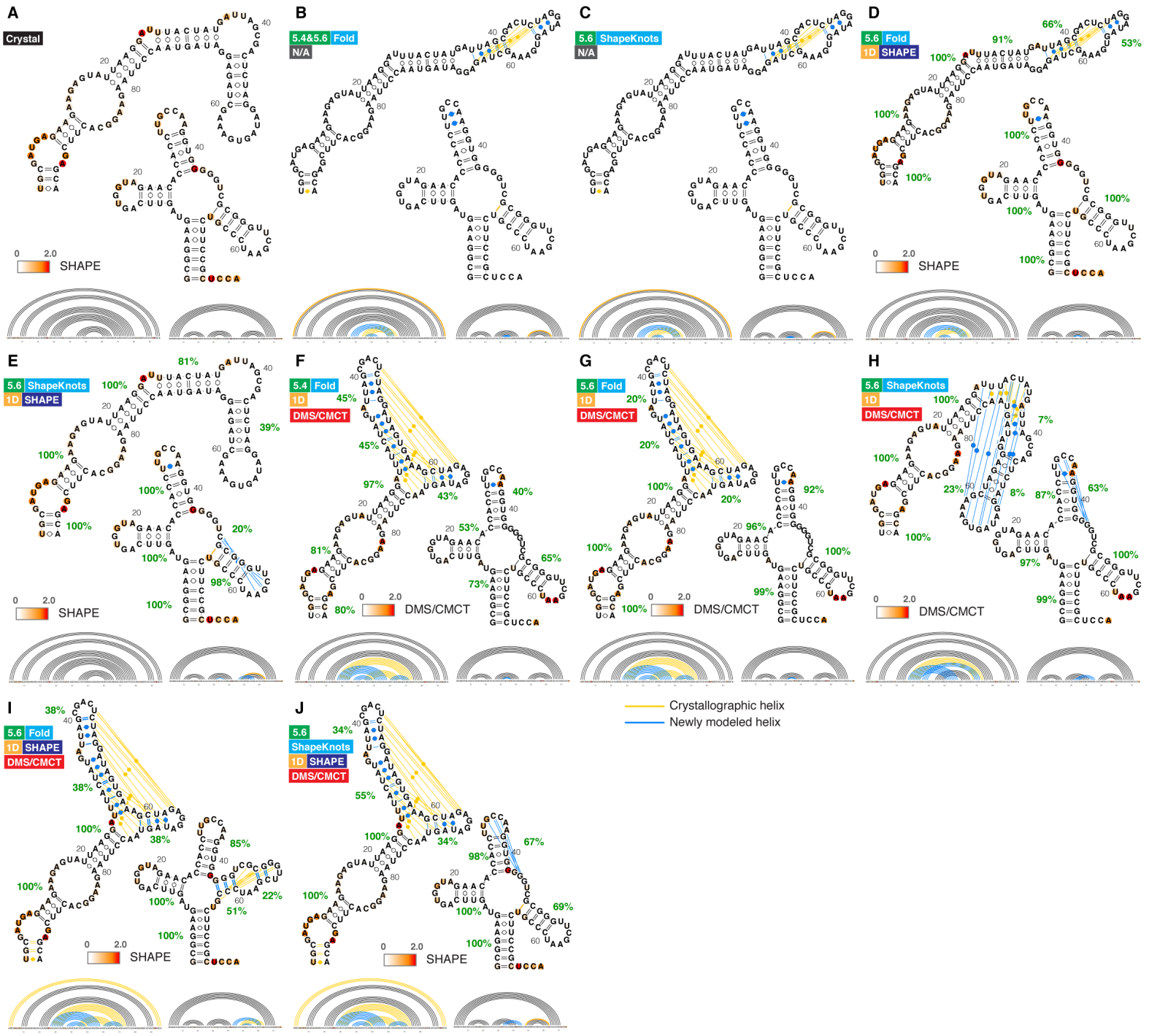
(A) Normalized reactivity of RNA Puzzle 5 RNA, using SHAPE (1M7), DMS and CMCT in 1-dimensional chemical mapping. Reactivities were normalized to GAGUA referencing hairpins (not shown).
(B) Secondary structure prediction using 1-dimensional SHAPE (1M7) data. Nucleotides are colored with SHAPE reactivities. Crystallographic pairings missing in this model and new non-crystallographic pairings are drawn as yellow and blue lines, respectively. Percentage labels give bootstrap support values.
(C) Mutate-and-map (M^2) dataset probed by the SHAPE reagent 1M7.
(D) Secondary structure prediction using 2D SHAPE M^2 data.

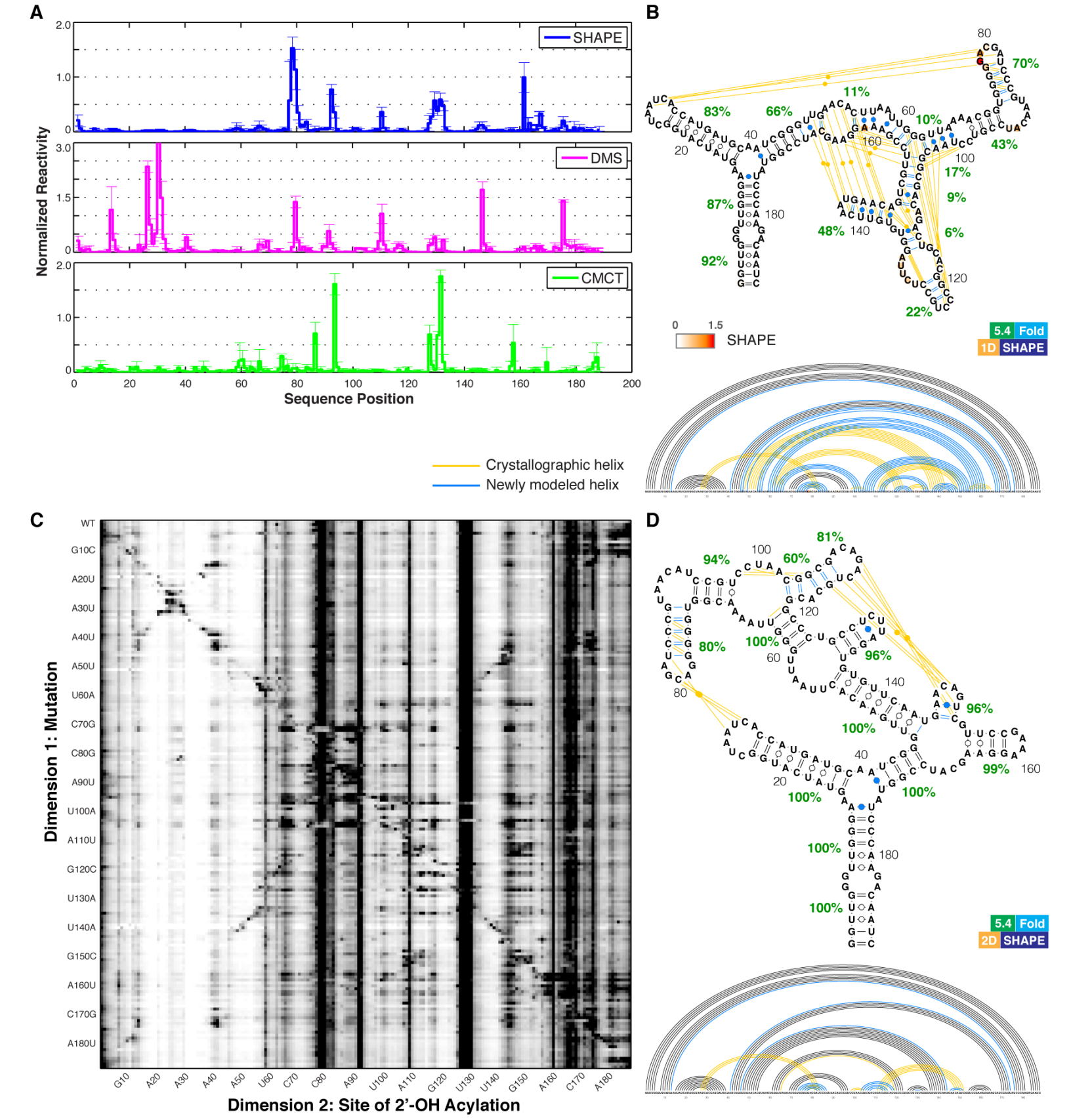
Figure S5. Chemical mapping data and secondary structure predictions of RNA Puzzle 6.
(A) Normalized reactivity of RNA Puzzle 6 RNA, using SHAPE (1M7), DMS and CMCT in 1-dimensional chemical mapping, in presence of 60 μM adenosylcobalamin. Reactivities were normalized to GAGUA referencing hairpins (not shown).
(B) Secondary structure prediction using 1-dimensional SHAPE (1M7) data. Nucleotides are colored with SHAPE reactivities. Crystallographic pairings missing in this model and new non-crystallographic pairings are drawn as yellow and blue lines, respectively. Percentage labels give bootstrap support values.
(C) Mutate-and-map (M^2) dataset probed by the SHAPE reagent 1M7, in presence of 60 μM adenosylcobalamin.
(D) Secondary structure prediction using 2D SHAPE M^2 data.

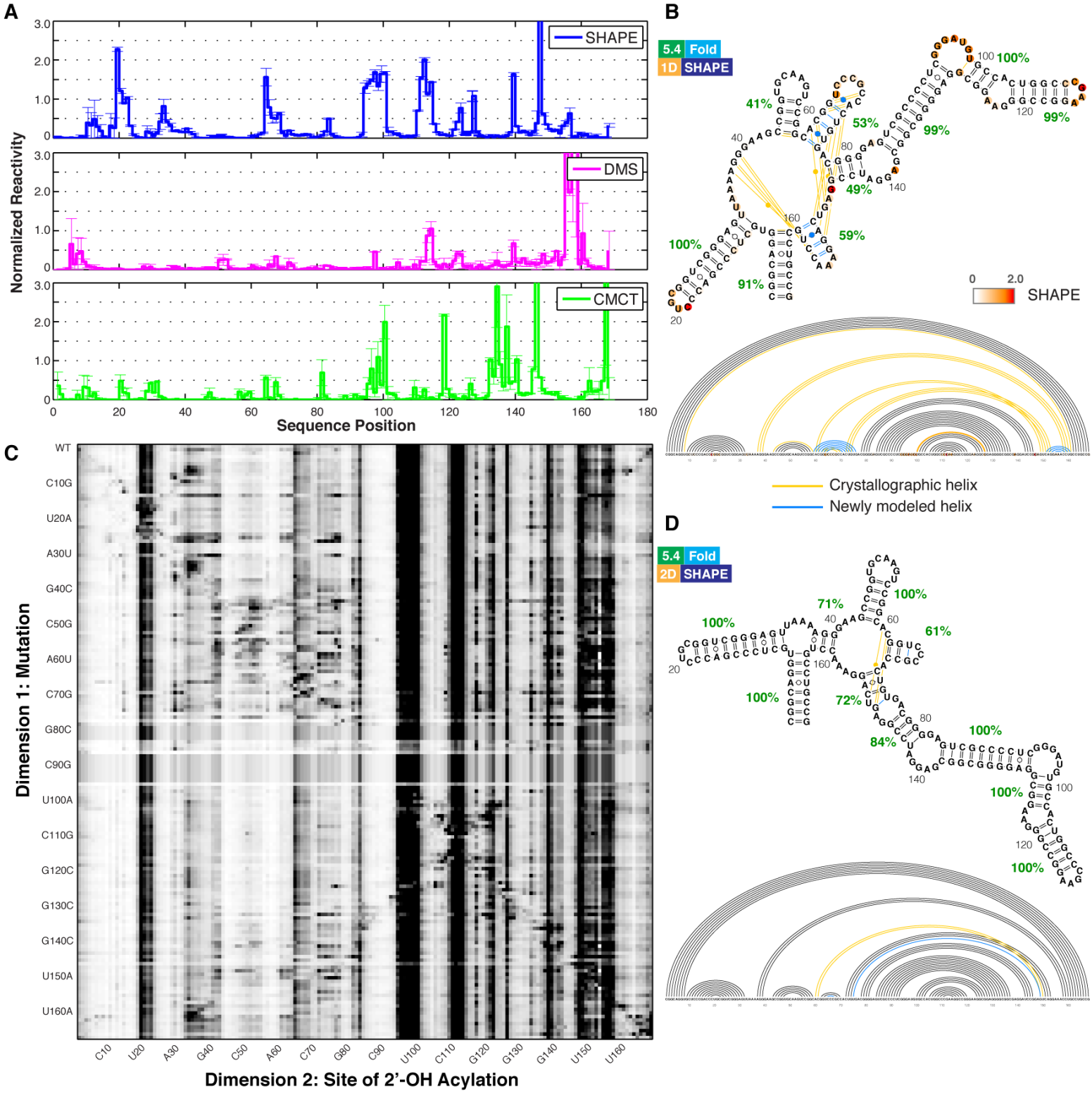
Figure S6. Chemical mapping data and secondary structure predictions of RNA Puzzle 10.
(A) Normalized reactivity of RNA Puzzle 10 RNA, using SHAPE (1M7), DMS and CMCT in 1-dimensional chemical mapping, in presence of 1 μM partner RNA strand. Reactivities were normalized to GAGUA referencing hairpins (not shown).
(B) Secondary structure prediction using 1-dimensional SHAPE (1M7) data. Nucleotides are colored with SHAPE reactivities. Crystallographic pairings missing in this model and new non-crystallographic pairings are drawn as yellow and blue lines, respectively. Percentage labels give bootstrap support values.

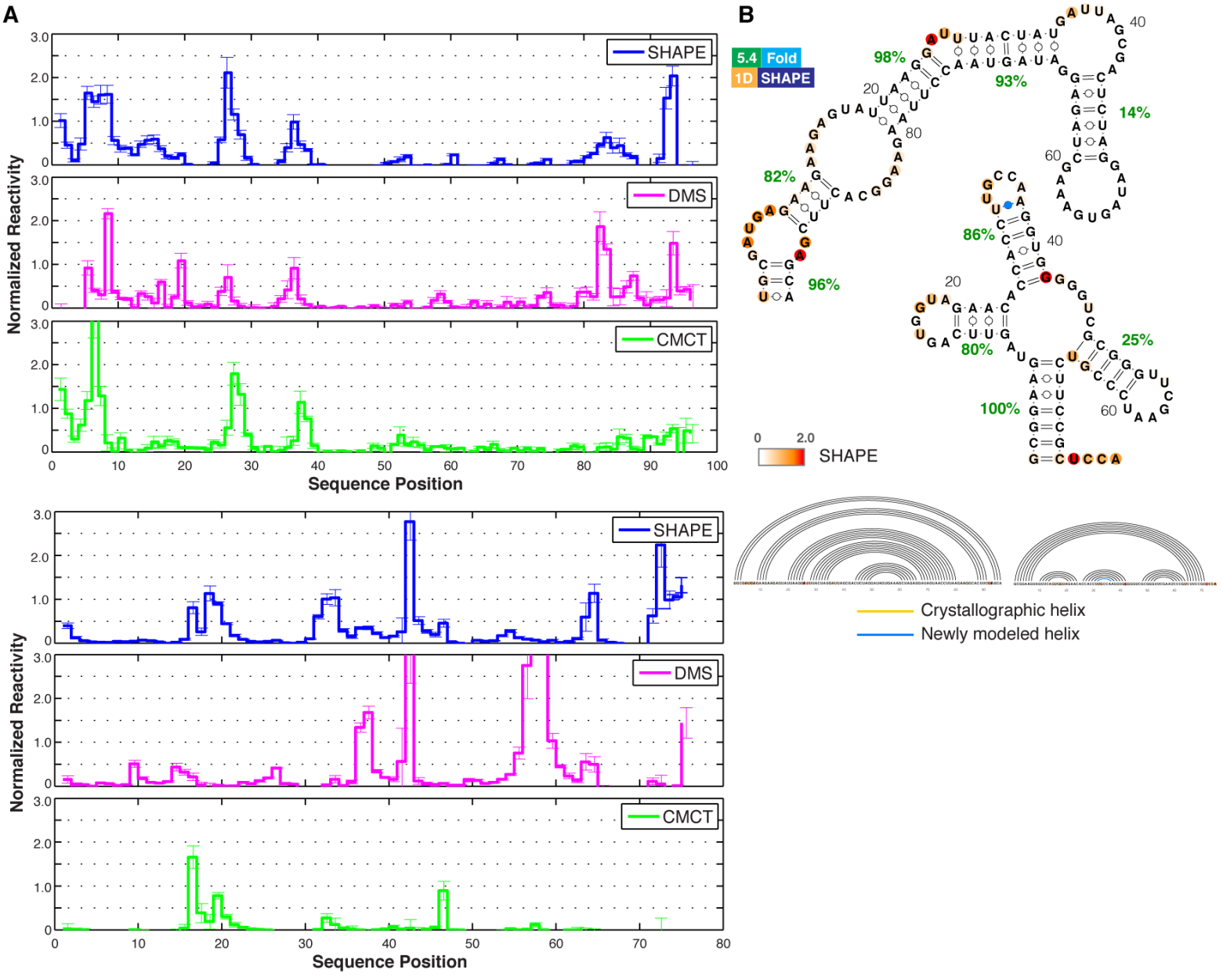

A
Crystal
SHAPE
B
5.4&5.6 Fold
N/A
C
5.6 ShapeKnots
N/A
D
5.6 Fold
1D SHAPE
E
5.6 ShapeKnots
1D SHAPE
F
5.4 Fold
1D DMS/CMCT
G
5.6 Fold
1D DMS/CMCT
H
5.6 ShapeKnots
1D DMS/CMCT
I
5.6 Fold
1D SHAPE DMS/CMCT
J
5.6 ShapeKnots
1D SHAPE DMS/CMCT
K
5.6 Fold
2D SHAPE
L
5.6 ShapeKnots
2D SHAPE
M
Dimension 1: Mutation
Dimension 2: Site of DMS Modification
N
5.4 Fold
2D DMS
O
5.6 Fold
2D DMS
P
5.6 ShapeKnots
2D DMS
Crystallographic helix
Newly modeled helix


A
Crystal
0 2.0
SHAPE
B
5.4&5.6 Fold
N/A
C
5.6 ShapeKnots
N/A
D
5.6 Fold
1D SHAPE
0 2.0
SHAPE
E
5.6 ShapeKnots
1D SHAPE
0 2.0
SHAPE
F
5.4 Fold
1D DMS/CMCT
0 2.0
DMS/CMCT
G
5.6 Fold
1D DMS/CMCT
0 2.0
DMS/CMCT
H
5.6 ShapeKnots
1D DMS/CMCT
0 2.0
DMS/CMCT
I
5.6 Fold
1D SHAPE DMS/CMCT
0 2.0
SHAPE
J
5.6 ShapeKnots
1D SHAPE DMS/CMCT
0 2.0
SHAPE
K
5.6 Fold
2D SHAPE
L
5.6 ShapeKnots
2D SHAPE
Crystallographic helix
Newly modeled helix


A
Crystal
B
5.4&5.6 Fold
N/A
C
5.6 ShapeKnots
N/A
D
5.6 Fold
1D SHAPE
E
5.6 ShapeKnots
1D SHAPE
F
5.4 Fold
1D
DMS/CMCT
G
5.6 Fold
1D
DMS/CMCT
H
5.6 ShapeKnots
1D
DMS/CMCT
I
5.6 Fold
1D SHAPE
DMS/CMCT
J
5.6
ShapeKnots
1D SHAPE
DMS/CMCT
0 2.0
SHAPE
DMS/CMCT
Crystallographic helix
Newly modeled helix


A
SHAPE
DMS
CMCT
Normalized Reactivity
Sequence Position
Crystallographic helix
Newly modeled helix
B
SHAPE
5.4 Fold
1D SHAPE
C
Dimension 1: Mutation
Dimension 2: Site of 2'-OH Acylation
D
5.4 Fold
2D SHAPE


A
SHAPE
DMS
CMCT
Normalized Reactivity
Sequence Position
B
5.4 Fold
1D SHAPE
SHAPE
Crystallographic helix
Newly modeled helix
C
Dimension 1: Mutation
Dimension 2: Site of 2'-OH Acylation
D
5.4 Fold
2D SHAPE


A
SHAPE
DMS
CMCT
Normalized Reactivity
Sequence Position
B
5.4 Fold
1D SHAPE
98%
93%
14%
82%
86%
96%
80%
25%
100%
SHAPE
Crystallographic helix
Newly modeled helix