Blind prediction of noncanonical RNA structure at atomic accuracy

Andrew Watkins[1], Caleb Geniesse[1,2], Wipapat Kladwang[1], Paul Zakrevsky[3,4], Luc Jaeger[4], Rhiju Das[1,2,5,*]

[1]Department of Biochemistry, Stanford University School of Medicine, Stanford, CA 94305.
[2]Biophysics Program, Stanford University, Stanford, CA 94305.
[3]Department of Chemistry, University of California, Santa Barbara, Santa Barbara, CA 93106.
[4]Current Affiliation: RNA Biology Laboratory, Center for Cancer Research, National Cancer Institute, Frederick, MD 21702
[5]Department of Physics, Stanford University, Stanford, CA 94305.

[*]Corresponding author. Phone: (650) 723-5976. E-mail: rhiju@stanford.edu

Abstract

Prediction of RNA structure from nucleotide sequence remains an unsolved grand challenge of biochemistry and requires distinct concepts from protein structure prediction. We report a stepwise Monte Carlo method that has enabled the first blind prediction recovering all noncanonical base pairs of a complex RNA structure, a double pseudoknot from the Zika virus genome posed as a community-wide RNA Puzzle. A benchmark of 82 diverse motif structure challenges and prospective experimental tests for three previously unsolved tetraloop/receptors support the method's general ability to recover noncanonical pairs *ab initio*, with remaining problems traced to limitations of current macromolecule free energy functions.

Main text

Significant success in protein modeling has been achieved by assuming that the native conformations of a macromolecule have the lowest free energy and that the free energy function can be approximated by a sum of hydrogen bonding, van der Waals, electrostatic, and solvation terms that extend over Angstrom-scale distances. Computational methods that subject large pools of low-resolution protein models to all-atom Monte Carlo minimization guided by these free energy functions have achieved near-atomic-accuracy predictions in the CASP community-wide blind trials (*1*). When adapted to RNA structure modeling, analogous methods have consistently achieved nucleotide resolution in the RNA-Puzzle blind trials but have not yet reached atomic accuracy, aside from previously solved motifs that happen to recur in new targets (*2*). A disappointing theme in recent RNA-Puzzle assessments is that the rate of accurate prediction of noncanonical base pairs is typically 20% or lower, even for models with correct global folds (*2*). Without recovery of such noncanonical pairs, RNA computational modeling will not be able to explain evolutionary data, predict molecular partners, or be prospectively tested by compensatory mutagenesis for the myriad biological RNAs that are being discovered at an accelerating pace.

The lag between the protein and RNA modeling fields is partly explained by differences in how protein and RNA molecules fold. Protein structures are largely defined by how α-

helices and β-sheets pack together. As abundant data exist on these regular protein elements and their side-chain interactions, protein models with reasonable accuracy can often be assembled from fragments of previously solved structures. Less regular loops interconnecting α- and β-elements are less critical for defining protein folds. Indeed, those loops are typically not recovered at high accuracy, even in the most exceptional blind predictions (*3, 4*). In contrast, the predictable and geometrically regular elements of RNA folding are Watson-Crick helices that sequester their side chains and therefore cannot be positioned by direct side-chain interactions. Instead, the RNA loops interconnecting those helices form intricate noncanonical base pairs that define an RNA's global helix arrangement. The RNA structure prediction problem, more so than the protein problem, depends on discovering these irregular loop conformations and their associated noncanonical base pairs *ab initio*. Unfortunately, discovering the lowest energy conformations of these loop motifs has not generally been tractable due to the vast number of deep, local minima in the folding free energy landscape of even small noncanonical RNA motifs. Most RNA modeling methods use coarse-grained modeling stages that allow for smoother conformational search but generally return conformations too inaccurate to be refined to high accuracy by Monte Carlo minimization or molecular dynamics refinement.

To address this challenge, we have developed a Rosetta method called stepwise Monte Carlo (SWM). This method removes barriers in conformational search through addition and deletion of residues rather than through low resolution coarse-graining or conventional high resolution moves that make smaller perturbations. We previously described how step-by-step buildup of an RNA structure, enforcing low-energy conformations for each added nucleotide, could lead to atomic accuracy models of single stranded RNA loops (*5*). Conceptually, each addition was proposed to simulate the stepwise formation of well-defined structure from "random coil"-like ensembles. The calculation, instantiated in the Rosetta modeling framework, involved a deterministic enumeration over build paths, which guaranteed a unique solution for the final conformational ensemble but necessitated significant expenditure of computational power. Only small loops could be modeled. We recently discovered that such

computationally intensive enumeration is not required for high accuracy modeling. Instead, an add-and-delete Monte Carlo minimization scheme stochastically produces conformations that give computed free energies as low as those achieved by enumerative stepwise assembly (SI Results). Through its increased speed, SWM allowed us to confirm that recent updates to the Rosetta free energy function (*6*) and estimation of conformational entropy of unstructured segments improve modeling accuracy for single-stranded loops (SI Results; SI Tables S1-S2; and SI Fig. S1).

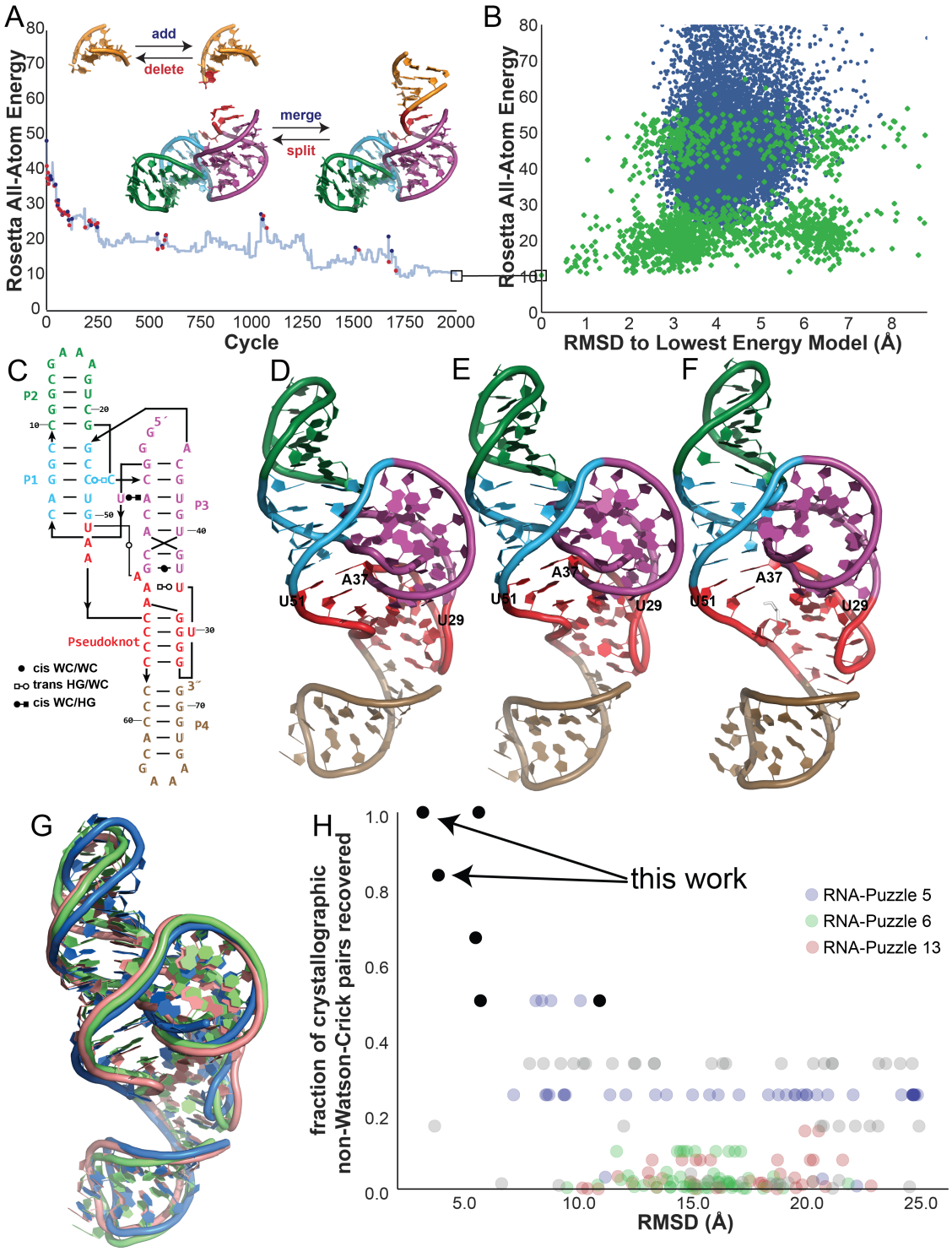

Figure 1. Searching the complex energy landscape of a noncanonical RNA tertiary fold. (A) Example 2000-cycle trajectory of the stepwise Monte Carlo method in which additions (blue) and deletions (red) are interspersed with standard resampling steps. 3D structures illustrate addition/deletion of terminal residues (top) and separate pre-defined segments (bottom). (B) The lowest energy structures discovered by stepwise Monte Carlo

(SWM, green) are not found with conventional fragment assembly (FARFAR, blue). (C) Two-dimensional diagram of the Zika xrRNA. Single dashes mark base pairs in secondary structure; other pairings, given in Leontis-Westhof annotation, were predicted by SWM. Amongst SWM models submitted for the RNA-Puzzles competition (D-E), one model produced the lowest overall RMSD and recovered all noncanonical pairs (E) in the subsequently released crystal structure (F), while others illustrated modeling uncertainties (D). (G) Overlay of crystal structure (marine) with SWM predictions (salmon and lime, respectively, for models in D and E). (H) Illustration of the fraction of noncanonical base pairs recovered and RMSD to native model obtained by Rosetta modeling (black) and other labs (gray) for RNA-Puzzle 18. As points of reference, several previous puzzles of similar complexity are plotted. Points recovering zero noncanonical pairs are given a small vertical perturbation to appear visually distinct. The models in (D) and (E) are the indicated points in the upper-left corner.

In late 2016, the blind modeling challenge RNA-Puzzle 18 provided an opportunity to test SWM rigorously (Fig. 1). On one hand, the 71-nucleotide target sequence was readily identified via PDB-BLAST (*7*) to be a Zika virus RNA homologous to a molecule with a previously solved X-ray structure, an Xrn1 endonuclease-resistant (xrRNA) fragment of Murray Valley Encephalitis virus (PDB ID: 4PQV) (*8*). However, the crystallographic environment of the prior structure disrupted a pseudoknot (between L3 and J1/4, Fig. 1) expected from sequence alignments so that nearly half of the prior structure could not be trusted as a template for homology modeling. Further complicating the modeling, two Watson-Crick pairs within stem P3 changed to or from G•U wobble pairs. Moreover, prior literature analysis (*8*) suggested extension of this helix by two further Watson-Crick pairs (U29-A37; U30-A36), albeit without direct evidence from phylogenetic covariation and in partial conflict with dimethyl sulfate probing (SI Results and SI Fig. S4). *Ab initio* modeling was therefore necessary for modeling the RNA, and we carried out stepwise Monte Carlo runs (Figs. 1A and 1B; SI Methods). Lowest free energy models converged to a tight ensemble of intricate structures illustrated in Figs. 1D and 1E. The Watson-Crick pairs U29-A37 and U30-A36 predicted in the literature did not occur in the models. Instead, several other features were consistently observed across the SWM models (red, Figs. 1D-E): co-axial stacking of the pseudoknot helix on P3, a noncanonical base pair between A37 and U51 interconnecting the P3 and P1 coaxial helical stacks, a UA-handle (*9*) formed by U29-A36, and lack of pairing by U30, A35, A52, and A53. These features were not uniformly present – or not predicted at all – in models created by our prior state-

of-the-art method for RNA-puzzle modeling (Fragment Assembly of RNA with Full Atom Refinement, FARFAR) or, as it later turned out, in models submitted by other RNA-Puzzle participants (see SI Fig. S2 and below).

The subsequent release of the crystal structure (Fig. 1F-G) (*10*) confirmed all base pairs predicted by SWM modeling (100% non-Watson-Crick recovery, compared to <50% in prior RNA-puzzles; Fig. 1H). Indeed, the only structural deviation involved A53, which was predicted in SWM models to be stacked on the pseudoknot helix (Figs. 1D-E). In the crystal, this nucleotide was indeed unpaired but bulged out of the core to form a contact with a crystallographic neighbor, while a 1,6-hexanediol molecule from the crystallization buffer took its place (white sticks, Fig. 1F); this arrangement was noted independently to be a likely crystallographic artifact (*10*). As illustrated in the overlay (Fig. 1G), there is striking overall fold agreement (3.0 Å RMSD over all 71 residues and 2.3 Å over just the most difficult noncanonical region, nucleotides 5-6, 27-39, and 49-57), much better than the ~10 Å agreement seen in previous RNA-puzzles of comparable difficulty (Fig. 1H) (*2*). Furthermore, SWM predicted all noncanonical base pairs accurately (F_{NWC}= 1, Fig. 1H).

To evaluate SWM more broadly, we tested whether it could recover noncanonical base pairs on 82 complex RNA motifs that we encountered in previous RNA-puzzles and other modeling challenges (SI Table S3 and SI Fig. S3). Due to the relative efficiency of SWM modeling and growing computational power, we could test a benchmark that was nearly three times larger than our most extensive previous efforts (*11*). Overall, SWM achieved a median RMSD accuracy (over the top five cluster centers) of 1.45 Å (Table 1 and SI Fig. S4) and mean recovery of non-Watson-Crick pairs of 76%. We observed numerous cases in which the SWM model and experimental structure were nearly indistinguishable by eye (Fig. 2). Examples included two-stranded motifs solved at much higher computational expense with the prior enumerative stepwise assembly method (*12*), such as the most conserved domain of the signal recognition particle (Fig. 2A; 1.26 Å RMSD, 5 of 5 noncanonical pairs recovered) and the first RNA-Puzzle challenge, a human thymidylate synthetase mRNA segment (Fig. 2B; 0.96 Å, 1 noncanonical pair and

1 extrahelical bulge recovered) (*2*). For several test cases, there was experimental evidence that formation of stereotyped atomic structures required flanking helices to be positioned by the broader tertiary context, so we expected these motifs might be particularly difficult. Nevertheless, if the immediately flanking helix context was provided, the median RMSD accuracy and non-Watson-Crick base pair recovery remained excellent (1.19 Å and 100%; Table 1, SI Fig. S4 and SI Table S3), as illustrated by the J5/5a hinge from the P4-P6 domain of the *Tetrahymena* group I intron (*13*) (Fig. 2C; 0.55 Å RMSD, all 4 noncanonical pairs and all 3 extrahelical bulges recovered).

Perhaps the most striking models were recovered for multi-helix junctions and tertiary contacts, which have largely eluded RNA modeling efforts seeking high resolution (*11, 14*). SWM achieves high accuracy models for the P2-P3-P6 three-way junction from the Varkud satellite ribozyme, previously missed by all modelers in the RNA-puzzle 7 challenge (Fig. 2D; 1.13 Å RMSD, 3 of 3 noncanonical pairs recovered); a highly irregular tertiary contact in a hammerhead ribozyme (Fig. 2E; 1.16 Å RMSD, 2 of 3 noncanonical pairs and 1 extrahelical bulge recovered); a complex between a GAAA tetraloop and its 11-nt receptor (0.64 Å RMSD, all 4 noncanonical pairs recovered, Fig. 2F); and the $tRNA^{phe}$ T-loop, a loop-loop tertiary contact stabilized by chemical modifications at 5-methyl-uridine, pseudouridine, and N1-methyl adenosine (Fig. 2G; 1.33 Å accuracy). Motifs without any flanking A-form helices offer particularly stringent tests for *ab initio* modeling but could also be recovered at high accuracy by SWM, as illustrated by the inosine-tetrad-containing quadruplex (Fig. 2H; 2.87 Å RMSD overall, 0.46 Å RMSD if the terminal uracils, which make crystal contacts, are excluded). For comparison, modeling with FARFAR gave worse recovery of noncanonical pairs and worse all-atom energies than SWM for each of these cases and, more broadly, across each motif category (Table 1 and SI Tables S3-S4).

Table 1. Benchmark of stepwise Monte Carlo (SWM) compared to prior Rosetta fragment assembly method (FARFAR).

	Motif Properties	Best of 5 Cluster Centers		Lowest RMSD	Lowest Energy
		RMSD (Å)	F_{NWC} [a]	RMSD (Å)	E-Gap (REU) [a]

Category	# Motifs	Length	Strands	SWM	FARFAR	SWM	FARFAR	SWM	FARFAR	SWM	FARFAR
Single helix or crystallographic context enforced											
trans-helix loop	15	6	1	**0.74**	3.30	**1.00**	0.77	**0.54**	1.91	**-0.28**	5.96
apical loop	4	5	1	**1.14**	2.96	**1.00**	**1.00**	**0.85**	1.17	**1.48**	3.19
two-way junction,	14	7.5	2	**0.74**	1.15	**1.00**	**1.00**	**0.55**	0.81	**-4.14**	4.09
multi-helix junction,	5	11	3	**1.91**	1.93	**0.80**	0.33	**1.41**	1.75	**-0.49**	10.26
tertiary contact	10	9	2	**1.25**	1.78	**0.74**	0.50	**1.18**	1.32	**-2.48**	9.62
Multiple helices without crystallographic context enforced											
two-way junction	15	7	2	1.59	**1.40**	**0.88**	0.55	**1.00**	1.18	**-1.21**	7.90
multi-helix junction	5	10	3	**2.60**	3.45	**0.40**	0.20	**2.13**	2.38	**-2.92**	13.06
tertiary contact	8	9	2	2.89	**2.27**	**0.36**	0.26	**1.98**	2.03	**-1.25**	23.50
non-helix embedded	5	10	4	**2.81**	4.30	**0.80**	0.71	**2.14**	2.48	**-5.12**	2.10
Overall	82	7	2	**1.45**	1.93	**0.89**	0.67	**1.09**	1.34	**-1.21**	7.85

[a] Fraction of non-Watson-Crick pairs from experimental structure observed in computational model.

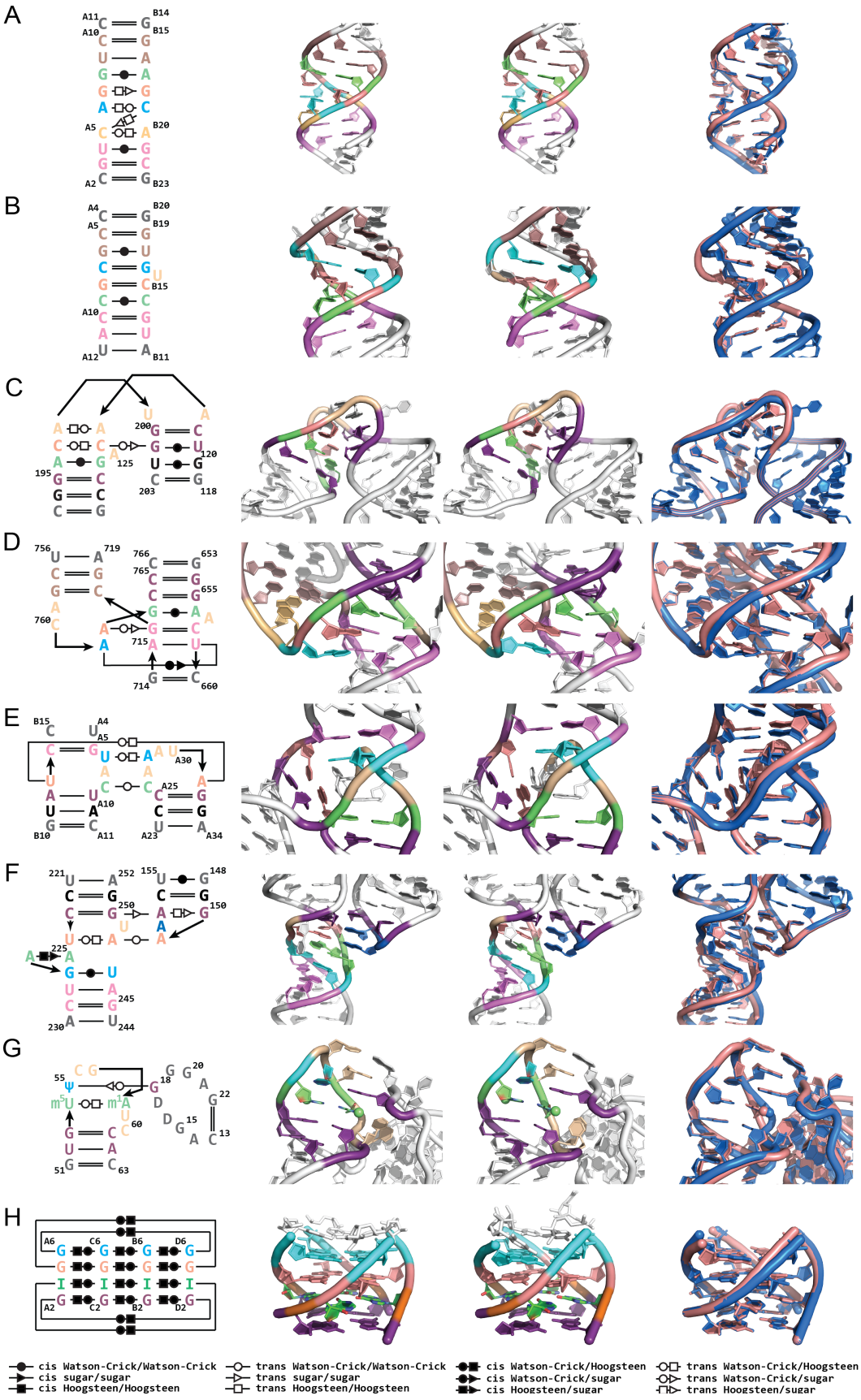
A
B
C
D
E
F
G
H
cis Watson-Crick/Watson-Crick
cis sugar/sugar
cis Hoogsteen/Hoogsteen
trans Watson-Crick/Watson-Crick
trans sugar/sugar
trans Hoogsteen/Hoogsteen
cis Watson-Crick/Hoogsteen
cis Watson-Crick/sugar
cis Hoogsteen/sugar
trans Watson-Crick/Hoogsteen
trans Watson-Crick/sugar
trans Hoogsteen/sugar

Figure 2. Stepwise Monte Carlo modeling recovers most or all noncanonical base pairs ab initio for complex RNA motifs (PDB ID in parentheses). From left to right in A-H, secondary structure with noncanonical base pairs, native, optimized model, and an overlay (native in marine; model in salmon). (A) Conserved domain of the human signal-recognition particle (1LNT); (B) noncanonical junction from human thymidylate synthase regulatory motif, RNA-puzzle 1 (3MEI); (C) irregular J5/5a hinge from the P4-P6 domain of the *Tetrahymena* group I self-splicing intron (2R8S); (D) P2-P3-P6 three-way A-minor junction from the Varkud Satellite nucleolytic ribozyme, RNA-puzzle 7 (4R4V); (E) highly irregular tertiary contact stabilizing the *Schistosoma* hammerhead nucleolytic ribozyme (2OEU); (F) recurrent tetraloop/receptor tertiary contact from the P4-P6 domain of the *Tetrahymena* group I self-splicing intron (2R8S); (G) T-loop/purine interaction from yeast $tRNA^{phe}$ involving three chemically modified nucleotides (1EHZ) (H) RNA quadruplex including an inosine tetrad (2GRB). In the secondary structure diagrams, and reflected in the model, violets indicate flanking helices built *de novo*; dark violet indicates a portion of the native structure used for modeling but allowed to minimize, while pastel colors indicate recovered noncanonical features and recovered extrahelical bulges are in wheat with white side-chains. In the secondary structure diagram, black indicate fixed context from the crystal structure, while gray indicates additional helical context not included in modeling; these context nucleotides are shown in white in the 3D models.

In some cases in the benchmark, SWM did not exhibit near-atomic-accuracy recovery and illuminated challenges remaining for computational RNA modeling (Fig. 2H and SI Fig. S5). While a few discrepancies between SWM models and X-ray structures could be explained by crystallographic interactions (e.g., edge nucleotides making crystal contacts, Fig. 2H), most problems were better explained by errors in the energy function. For 14 of the 19 cases in which the SWM modeling RMSD was worse than 3.0 Å (and thus definitively not achieving atomic accuracy), the energy of the lowest energy SWM model was lower than that of the optimized experimental structure, often by several units of free energy (calibrated here to correspond to k_BT (*6*)). In some cases, the RMSD achieved by FARFAR was better than by SWM, but not the fraction of base pairs recovered (Table 1). suggesting that conformational preferences encoded in database fragments in FARFAR needed to be captured by SWM, perhaps in its torsional potential. Results on the hepatitis C virus internal ribosome entry site, the sarcin ricin loop, and other test cases suggest that a more accurate torsional potential, as well as inclusion of metal ions, may eventually address these residual problems (Supplemental Fig. S5).

Despite these current limitations, the overall accuracy of SWM in the 82-motif benchmark suggested that it could predict noncanonical base pairs in motifs that have been refractory to NMR and crystallographic analysis and that the resulting models could be stringently validated or falsified by prospective biochemical tests. Success in the 11-nt tetraloop/receptor motif (Fig. 2F), a classic model system and ubiquitous tertiary contact in natural RNAs, encouraged us to model alternative tetraloop/receptor complexes selected for use in RNA engineering but not yet solved experimentally (*2, 15*). The resulting SWM models for the C7.2, C7.10, and R(1) tetraloop/receptors (Fig. 3) suggested striking structural homologies to the natural 11-nt receptors but also noncanonical features (extrahelical bulges, pairs) different from prior manual modeling efforts (SI Results). We tested these features using prospective experiments. CMCT (*N*-cyclohexyl-*N′*-(2-morpholinoethyl) carbodiimide tosylate) mapping on the receptors installed into the P4-P6 domain of the *Tetrahymena* ribozyme verified extrahelical bulging of single-nucleotide uridines predicted at different positions in the different receptors (Fig. 3C and SI Results). The R(1) receptor model included numerous unexpected noncanonical features, especially a base triple involving a new Watson-Crick singlet base pair G4-C9 and a dinucleotide platform at G4-U5. These features were stringently evaluated via compensatory mutagenesis. Chemical mapping on the P4-P6 domain confirmed the G4-C9 base pair but was not sensitive enough to test other compensatory mutants (SI Fig. S6). We therefore carried out native gel assembly measurements in a different system, the tectoRNA dimer, which allows precise energetic measurements spanning 5 kcal/mol (Fig. 3E). Observation of energetic disruption by individual mutations and then rescue by compensatory mutants confirmed the predicted interactions of G4-C9, the base triple G4-U5-C9, and noncanonical pair G6-A7 ($p < 0.01$ in all cases; SI Fig. 3D) as well as other features of the model (Fig. S7). Such mutagenesis-based validation of noncanonical pairs would have been intractable absent a predicted model, due to the large number of possible mutant combinations that would have to be tested.

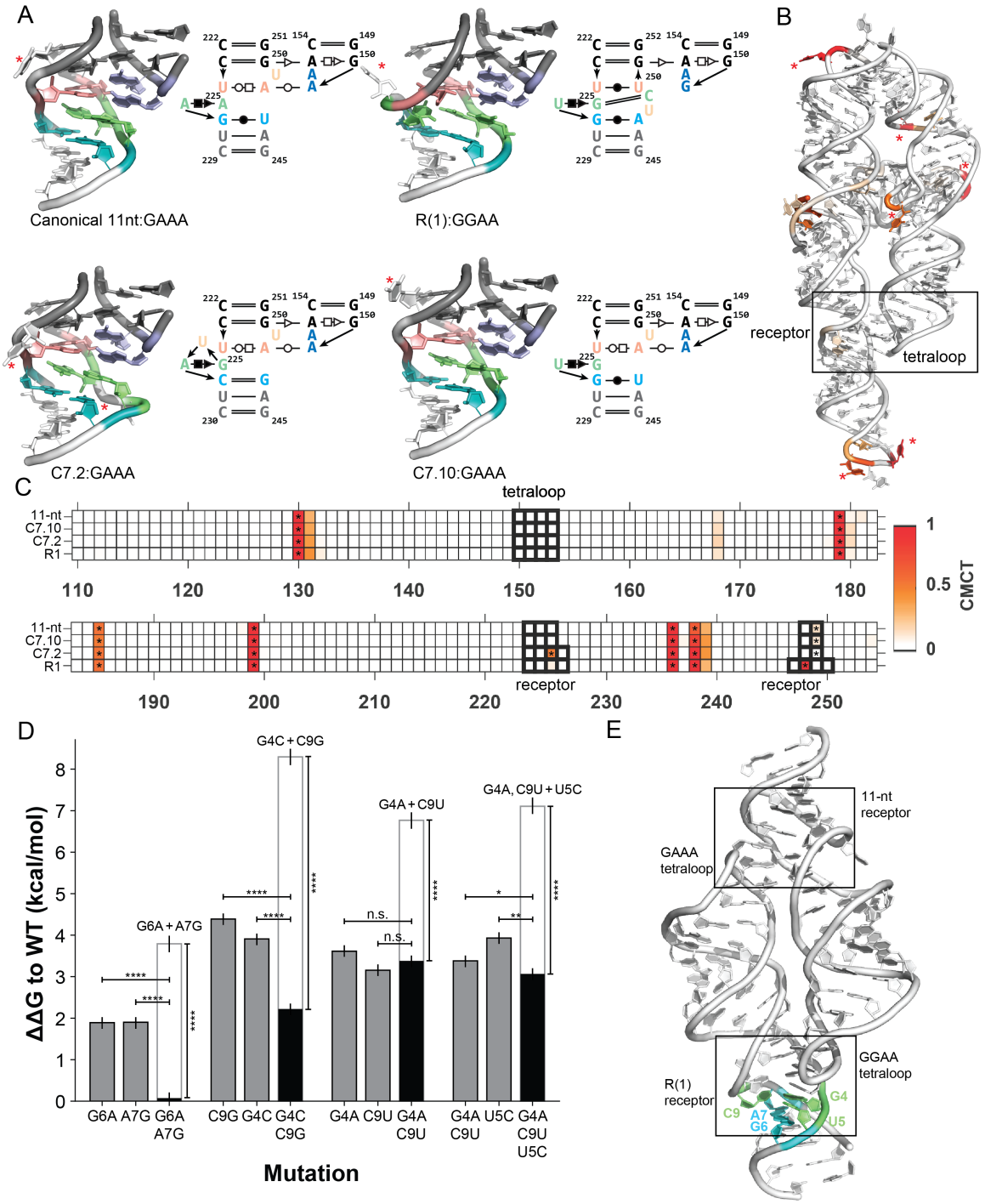

Figure 3. Stepwise Monte Carlo modeling of previously unsolved tetraloop-tetraloop receptor (TL/TR) motifs. (A) Homologies revealed between the canonical 11-nucleotide TL/TR motif and alternative motifs discovered through *in vitro* selection that have resisted crystallization. Structural features preserved across the receptors include: two purines from the GNRA tetraloop (lavender); a Watson-Crick/Hoogsteen U-A handle at the receptor docking site (salmon); a same-strand dinucleotide platform stacked under the tetraloop (lime); and a noncanonical pair which forms the transition between the platform

region and the lower canonical stem of the receptor (teal). The bottom flanking helix (white) was allowed to move relative to the top helices of the receptor and tetraloop (gray), which were held fixed. All other residues were built *ab initio* during SWM modeling. The modeling revealed bulged uracil residues at different positions in the different models (red asterisks). (B) The canonical 11-nt tetraloop receptor module from the P4-P6 domain of the *Tetrahymena* group I self-splicing intron (strands: 3, PDB: 2R8S), with structurally bulged uracil residues indicated by red asterisks. (C) CMCT mapping of the receptors installed into the P4-P6 domain of the *Tetrahymena* ribozyme (tetraloop and receptor indicated by black boxes) supports the bulged uracils in the predicted models (black asterisks). (D) Native gel assembly of a tectoRNA dimer (E) with the R(1) receptor allows quantitative tests by compensatory mutagenesis. For the base pairs and triples tested, rescue by double and triple mutants (black bars) was compared to energetic perturbations predicted based on sum of effects (white bars) of component mutations or, more conservatively, to the single mutants. One, two, three, and four asterisks represent $p < 0.05$, $p < 0.001$, $p < 5 \times 10^{-5}$, and $p < 1 \times 10^{-6}$,respectively; n.s., not significant; p values computed by Student's t-test for difference of means.

We have reported blind prediction of all noncanonical pairs of a complex RNA-puzzle, quantitative recovery of noncanonical pairs through the majority of an extensive benchmark that includes prior RNA-puzzle motifs, and prospective experimental tests of noncanonical features in three previously unsolved tetraloop-receptors. These results support stepwise nucleotide structure formation as a missing algorithmic principle for high resolution RNA structure modeling. Towards modeling and design of large multi-motif RNA structures, determining whether improved RNA torsional potentials and treatment of ionic effects can correct residual free energy function problems becomes a critical open question.

Acknowledgments

We thank Stanford Research Computing for expert administration of the BioX^3 clusters (supported by NIH 1S10RR02664701) and Sherlock clusters. We acknowledge financial support from the Burroughs-Wellcome Fund (CASI to R. D.), NIH R01 GM102519 and R21 GM102716 (to R. D.), and a RosettaCommons grant.

Materials and Methods

Detailed methods are available in the Supplementary Materials.

References

1. P. Bradley, K. M. Misura, D. Baker, Toward high-resolution de novo structure prediction for small proteins. *Science* **309**, 1868-1871 (2005).
2. Z. Miao *et al*., RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*, (2017).
3. P. S. Huang *et al*., De novo design of a four-fold symmetric TIM-barrel protein with atomic-level accuracy. *Nat Chem Biol* **12**, 29-34 (2016).
4. S. Ovchinnikov *et al*., Improved de novo structure prediction in CASP11 by incorporating coevolution information into Rosetta. *Proteins* **84 Suppl 1**, 67-75 (2016).
5. P. Sripakdeevong, W. Kladwang, R. Das, An enumerative stepwise ansatz enables atomic-accuracy RNA loop modeling. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 20573-20578 (2011).
6. R. F. Alford *et al*., The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. *J Chem Theory Comput*, (2017).
7. P. W. Rose *et al*., The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res* **45**, D271-D281 (2017).
8. J. S. Kieft, J. L. Rabe, E. G. Chapman, New hypotheses derived from the structure of a flaviviral Xrn1-resistant RNA: Conservation, folding, and host adaptation. *RNA Biol* **12**, 1169-1177 (2015).
9. L. Jaeger, E. J. Verzemnieks, C. Geary, The UA_handle: a versatile submotif in stable RNA architectures. *Nucleic Acids Res* **37**, 215-230 (2009).
10. B. M. Akiyama *et al*., Zika virus produces noncoding RNAs using a multi-pseudoknot structure that confounds a cellular exonuclease. *Science* **354**, 1148-1152 (2016).
11. R. Das, J. Karanicolas, D. Baker, Atomic accuracy in predicting and designing noncanonical RNA structure. *Nat Methods* **7**, 291-294 (2010).
12. P. Sripakdeevong *et al*., Structure determination of noncanonical RNA motifs guided by 1H NMR chemical shifts. *Nat Methods* **11**, 413-416 (2014).
13. A. A. Szewczak, T. R. Cech, An RNA internal loop acts as a hinge to facilitate ribozyme folding and catalysis. *RNA* **3**, 838-849 (1997).
14. M. Parisien, F. Major, The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* **452**, 51-55 (2008).
15. C. Geary, S. Baudrey, L. Jaeger, Comprehensive features of natural and in vitro selected GNRA tetraloop-binding receptors. *Nucleic Acids Res* **36**, 1138-1152 (2008).