



Published in final edited form as:

Nat Methods. 2014 April ; 11(4): 413–416. doi:10.1038/nmeth.2876.

Structure determination of noncanonical RNA motifs guided by ¹H NMR chemical shifts

Parin Sripakdeevong¹, Mirko Cevec², Andrew T. Chang³, Michèle C. Erat^{4,5}, Melanie Ziegeler², Qin Zhao⁶, George E. Fox⁷, Xiaolian Gao⁷, Scott D. Kennedy⁸, Ryszard Kierzek⁹, Edward P. Nikonowicz³, Harald Schwalbe², Roland K. O. Sigel⁵, Douglas H. Turner¹⁰, and Rhiju Das^{1,11,12}

¹Biophysics Program, Stanford University, Stanford, CA 94305, USA ²Center for Biomolecular Magnetic Resonance, Institute for Organic Chemistry and Chemical Biology, Johann Wolfgang Goethe University Frankfurt, Max-von-Laue-Strasse 7, 60438 Frankfurt, Germany ³Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77251, USA ⁴Department of Biochemistry, University of Oxford, Oxford OX1 3QU, United Kingdom ⁵Institute of Inorganic Chemistry, University of Zurich, Winterthurerstrasse 190, CH-8057 Zurich, Switzerland ⁶Feinberg School of Medicine, Northwestern University, Chicago, IL 60611, USA ⁷Department of Biology and Biochemistry, University of Houston, Houston, TX 77004, USA ⁸Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, NY 14642, USA ⁹Institute of Bioorganic Chemistry, Polish Academy of Sciences, 60-714 Poznan, Noskowskiego 12/14, Poland ¹⁰Department of Chemistry, University of Rochester, Rochester, NY 14627, USA ¹¹Department of Biochemistry, Stanford University, Stanford, CA 94305, USA ¹²Department of Physics, Stanford University, Stanford, CA 94305, USA

Abstract

Structured non-coding RNAs underline fundamental cellular processes, but determining their 3D structures remains challenging. We demonstrate herein that integrating NMR ¹H chemical shift data with Rosetta *de novo* modeling can consistently return high-resolution RNA structures. On a benchmark set of 23 noncanonical RNA motifs, including 11 blind targets, Chemical-Shift-ROSETTA for RNA (CS-ROSETTA-RNA) recovered the experimental structures with high accuracy (0.6 to 2.0 Å all-heavy-atom rmsd) in 18 cases.

RNA molecules form complex three-dimensional structures that play key roles in a multitude of cellular processes from gene regulation to viral pathogenesis [see e.g., ref.¹].

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

To whom correspondence should be addressed: Phone: (650) 723-5976. Fax: (650) 723-6783. rhiju@stanford.edu.

Author Contributions

P.S. and R.D. designed the research. P.S. implemented the method, generated the data, analyzed the results, and wrote the paper. R.D. assisted in analyzing the data and writing the paper. M.C., A.T.C., M.C.E., M.Z., Q.Z., G.E.F., X.G., S.D.K., R.K., E.P.N., H.S., R.K.O.S. and D.H.T. provided experimental data for the 11 blind NMR targets and participated in the blind trials. All authors discussed the results and commented on the manuscript.

Competing Financial Interests

The authors declare no competing financial interests.

These RNAs are typically composed of canonical helices interconnected by motifs with intricate, noncanonical structures critical for RNA catalysis, binding, and higher-order folding. With sizes of a few dozen nucleotides or less, these motifs offer compelling targets for solution NMR approaches². Nevertheless, NMR characterization of RNA motifs does not always generate sufficient NOE or other restraints to produce reliable atomic-resolution 3D models^{3–6}.

NMR chemical shifts can be an important additional source of structural information for functional macromolecules. In protein studies, backbone chemical shifts are widely used to constrain protein secondary structures and backbone torsions⁷ and to refine three-dimensional models⁸. More recently, chemical shift data have been leveraged for *de novo* protein structure determination (see e.g., refs.^{9,10}). Similar tools for RNA are less developed. Chemical shift assignments through NOESY and through-bond correlation spectroscopy experiments are standard first steps in RNA NMR, but the resulting chemical shift values are generally not used at the structure determination stage². Algorithms have been developed to ‘back-calculate’ non-exchangeable ¹H chemical shifts from RNA 3D structure^{11,12}. In particular, the well-calibrated NUCHEMICS¹² program has been used to refine models¹³ generated from conventional NMR measurements (NOE, J-couplings, residual dipolar couplings) and to successfully determine *de novo* structures of simple helical forms of nucleic acids¹⁴. Recently, Frank et al. demonstrated the power of chemical shift data to stringently constrain RNA molecular dynamics simulations starting from known structure¹⁵. This study hypothesized that chemical-shift-based modeling without previous knowledge of the structure should be possible, but such *de novo* structure determination has not been demonstrated.

In this work, we show that assigned ¹H chemical shift data indeed provide sufficient information to determine the structures of noncanonical RNA motifs at high resolution, without other NMR measurements as structural inputs. The key innovation has been the integration of chemical shift data with recent advances in high-resolution RNA *de novo* structure prediction^{16,17}. This article presents the resulting method, Chemical-Shift-ROSETTA for RNA (CS-ROSETTA-RNA) and its extensive benchmark on 23 RNA motifs, including 11 blind targets. The method is also made freely available through a web server at http://rosie.rosettacommons.org/rna_denovo.

RNA structure prediction methods by fragment assembly of RNA with full-atom refinement (FARFAR)¹⁶ and stepwise assembly (SWA)¹⁷ have permitted the modeling of RNA motifs that give atomic-resolution agreement to experimentally determined structures in favorable cases^{16,17}. However, as in protein studies, inaccuracies in available energy functions preclude high-resolution modeling in many cases¹⁸. Fortunately, in such problem cases, correct structures are still sampled¹⁷, and even quite sparse experimental data can identify these models with high confidence^{10,18}. Figure 1 illustrates this approach on a complex RNA test motif that was challenging for prior Rosetta approaches, a conserved UUAAGU hexaloop from 16S ribosomal RNA (Fig. 1a). Rosetta modeling successfully generated models with atomic-resolution agreement to this hexaloop’s crystallographic structure (0.52 Å all-heavy-atom rmsd; Fig. 1a–b), but these were ranked worse than non-native models (>5.0 Å rmsd; Fig. 1c). The experimentally measured chemical shifts of the non-

exchangeable ^1H atoms are in strong agreement with the predicted chemical shifts from the near-native models but not from any of the non-native models (Fig. 1d–e and Supplementary Figs. 1 and 2). Supplementing the Rosetta energy function with a chemical shift-based pseudo-energy score (E_{shift} ; see Methods) then permits confident discrimination of the atomic-accuracy models (Fig. 1f; also see Supplementary Results for further discussions on the importance of base and ribose proton chemical shifts for recovering the native structure).

To evaluate the generality and accuracy of CS-ROSETTA-RNA, we carried out modeling on a benchmark set containing 23 RNA motifs (Table 1). First, we applied CS-ROSETTA-RNA to a test set of 12 noncanonical motifs for which published chemical shift data as well as structural models derived from NMR and, in some cases, crystallography were available (Supplementary Table 1). These RNA motifs included hairpins, internal loops, a 3-way junction, and a tetraloop-receptor interaction. On average, 6.0 non-exchangeable ^1H chemical shifts per nucleotide (out of 7–8 total) were assigned, including both ribose and base protons (Supplementary Table 1). In addition to these cases, we further tested CS-ROSETTA-RNA on 11 blind RNA targets that were concurrently under investigation in five NMR laboratories. Sequences and assigned chemical shifts for these targets, but no other information, were made available for chemical-shift-guided modeling. Subsequent comparison of CS-ROSETTA-RNA models with structures derived from conventional NMR approaches thus served as blind evaluations.

Over the entire benchmark of 23 RNA motifs, CS-ROSETTA-RNA returned 18 cases in which at least one of the five lowest energy cluster centers achieved better than 2.0 Å all-heavy-atom rmsd (rmsd values and cluster ranks are provided in Table 1 and Supplementary Tables 2 and 3; energy vs. rmsd plots are provided in Supplementary Fig. 3; PDB files of experimental structures and five lowest energy cluster centers are provided in Supplementary Data). In four of the remaining five cases, structural dynamics in solution precluded high-resolution agreement between the NMR structures and the CS-ROSETTA-RNA models (Supplementary Results and Supplementary Figs. 4 and 5). CS-ROSETTA-RNA also performed well on both the test set of known structures (10/12 success cases) and the blind targets (8/11 success cases). Furthermore, 11 of the 23 cases satisfied a more stringent success criterion: the lowest energy (top-ranked) model was within atomic-accuracy of the experimental structure (under 1.5 Å all-heavy-atom rmsd). Lastly, incorporating even sparse data (~1 chemical shift per nucleotide) gave improved accuracy (Supplementary Results and Supplementary Fig. 6).

CS-ROSETTA-RNA success cases included high-resolution models from diverse sources, such as the most conserved internal loop from the signal recognition particle (SRP) RNA (rmsd of 0.81 Å; Fig. 2a); a GAAA tetraloop-receptor interaction (rmsd 0.68 Å; Fig. 2b); a three-way junction from yeast mitochondrial group II intron Sc.ai5 γ (rmsd 1.74 Å; Fig. 2c); and both the major and minor conformations of a G:G mismatch (Supplementary Fig. 7). Successful blind predictions included a highly irregular 5'-GAGU-3'/3'-UGAG-5' self-complementary internal loop that required additional synthesis efforts to solve by conventional NMR means (rmsd 1.10 Å; Fig. 2d); all four UNAC tetraloops (Fig. 2e); a 5'-GU-3'/3'-UAU-5' internal loop from a group II intron (rmsd of 1.37 Å; Fig. 2f); and a cuUCCaa anticodon stem-loop of *Bacillus subtilis* tRNA^{Gly} (rmsd of 1.41 Å; Fig. 2g).

Several CS-ROSETTA-RNA predictions gave strong convergence, as defined by a distinct energy ‘funnel’: a single dominant conformation and geometrically similar models achieved better energy than all other conformations. In seven benchmark cases, the lowest energy model gave an energy gap of $>3.0 k_B T$ to the next-lowest energy cluster and, in all of these cases, the model achieved atomic-accuracy (under 1.5 \AA rmsd to experimental structure; Supplementary Fig. 8). This energy gap thus appears to be a hallmark of CS-ROSETTA-RNA accuracy (see also *A criterion for confidence prediction*, Supplementary Results). In one apparent exception, the SRP conserved internal loop, a large energy gap ($5.5 k_B T$) strongly suggested that the CS-ROSETTA-RNA prediction should be accurate, but the lowest energy CS-ROSETTA-RNA model disagreed with the experimental NMR models³ ($>2.0 \text{ \AA}$ rmsd; Supplementary Fig. 9a–b). Further analysis revealed that the experimental NMR models poorly explained the ^1H chemical shift data published in the same study³ ($\text{rmsd}_{\text{shift}} = 0.50 \text{ ppm}$) and poorly agreed with subsequently solved crystallographic structures^{4,19} (rmsd of 2.30 \AA to PDB: 1LNT¹⁹). In contrast, the CS-ROSETTA-RNA model gave excellent agreement with the chemical shift data ($\text{rmsd}_{\text{shift}} = 0.18 \text{ ppm}$) and closely matched the crystallographic structures (rmsd of 0.81 \AA to PDB: 1LNT; Fig. 2a and Supplementary Fig. 9c–d). The SRP motif case supports the use of CS-ROSETTA-RNA as a tool to independently cross-validate or remodel NMR-derived structures.

By integrating assigned NMR chemical shift data into a new generation of RNA *de novo* modeling algorithms, CS-ROSETTA-RNA enables confident determination of noncanonical RNA motif structures in a manner fundamentally distinct from prior methods, using independent and far less experimental information. While obtaining resonance assignments is a necessary step of NMR-based RNA characterization, the structural information contained in chemical shifts are typically left aside during the generation of RNA structural models². Furthermore, the standard operating procedure² of determining NOEs, J-couplings, and, in some cases, residual dipolar coupling, does not always yield sufficient information to determine an RNA’s three-dimensional structure by conventional means, as illustrated by the 5'-GAGU-3'/3'-UGAG-5' case (Fig. 2d; also see Supplementary Notes and Supplementary Figs. 10 and 11 for further modeling details of this highly irregular motif). Here, incorporating assigned ^1H chemical shift data into Rosetta-based RNA modeling gives high accuracy structures for the majority of cases in a 23-RNA benchmark, including 8 of 11 blind targets. Further integration of *de novo* modeling and NMR methodologies, including the incorporation of ^{13}C , ^{15}N , and exchangeable ^1H chemical shift data (see Supplementary Results), may allow not just the acceleration of structure determination but eventually lead to the solution of currently intractable three-dimensional RNA structures.

Online Methods

Generation of Rosetta Models

Two complementary structure-modeling methods, Fragment Assembly with Full-Atom Refinement (FARFAR)¹⁶ and Stepwise Assembly (SWA)¹⁷, were used in parallel to generate the Rosetta models for each motif. SWA models were constructed using a series of recursive building steps, as described previously¹⁷. Each step involved enumerating several million conformations for each nucleotide, and all step-by-step build-up paths were covered

in N^2 building steps where N is the number of nucleotides in the motif. At the final building steps, all models are finely clustered and a maximum of 10,000 low energy SWA models were retained. The SWA approach is effective at generating models that are highly optimized to the underlying all-atom energy function, but can produce primarily incorrect models when the assumed energy function is inaccurate. Therefore, models were also generated by fragment assembly followed by full-atom refinement (FARFAR) in the Rosetta framework, as described previously¹⁶; the fragment source was the large ribosomal subunit of *H. marismortui* (PDB: 1JJ2). For each motif, 250,000 FARFAR models were generated; these models were then finely clustered and a maximum of 10,000 low energy FARFAR models were retained. The SWA and FARFAR models were then combined, leading to ~10,000–20,000 final Rosetta models for each motif. The SWA method was used to model all 23 RNA motifs in the benchmark except for the GAAA tetraloop-receptor interaction and the Sc.ai5γ 3-way junction. The FARFAR method was used to model all 23 RNA motifs in the benchmark except for the 5'-GAGU-3'/3'-UGAG-5' RNA structural switch (see Supplementary Notes). Algorithms and complete documentation are incorporated into Rosetta release 3.5 (www.rosettacommons.org), freely available for academic use.

The total computational costs for the generation of SWA and FARFAR models in term of modern central processing unit (CPU) are as follow. For SWA runs, the computational cost ranged from ~5,000 CPU hours for a 6-nucleotide motif to ~50,000 CPU hours for the 13-nucleotide motif investigated in this work (using Intel Xeon E5345 2.33 GHz CPUs). For FARFAR runs, the computational cost ranged from ~3,000 CPU hours for the 6-nucleotide motif to ~8,000 CPU hours for the 13-nucleotide motif. The majority of the computations for this work were performed on Stanford University's Bio-X² cluster, a supercomputer with 2,208 CPUs (Intel Xeon E5345 2.33 GHz). When using 500 CPU (the maximum allocated to each user), it takes less than half a day (of wall-clock time) to perform 5,000 CPU hours of computation and less than 5 days (of wall-clock time) to perform 50,000 CPU hours of computation.

To further encourage usage of the CS-ROSETTA-RNA method by the general NMR RNA community, a public web server where users can access and submit CS-ROSETTA-RNA modeling jobs is made freely available at <http://rosie.rosettacommons.org/rnadenovo>. Documentations and tutorials on how to submit the modeling jobs are also provided at the website. Due to computational resource limitations and to ensure short queue time, the web server runs a slightly modified version of CS-ROSETTA-RNA where the models are generated using only the FARFAR method and the maximum number of models per job submission is limited to 50,000.

Incorporation of non-exchangeable ¹H chemical shifts into structure modeling

Information from the experimental non-exchangeable ¹H chemical shifts were incorporated into the modeling process through the chemical shift pseudo-energy term:

$$E_{\text{shift}} = c \times \sum_i \left(\delta_i^{\text{exp}} - \delta_i^{\text{calc}} \right)^2 \quad [1]$$

where δ_i^{exp} and δ_i^{calc} are, respectively, the experimental and back-calculated chemical shift in ppm units (the index i sums over all experimentally assigned non-exchangeable ^1H chemical shifts in the RNA motif), and c is a weighting factor set to $4.0 k_{\text{B}}T/\text{ppm}^2$ based on test runs with different motifs. The NUCHEMICS program¹² was used to back-calculate non-exchangeable ^1H chemical shifts. In the 23 RNA motifs benchmark set, only 3 chemical shift datasets (UUCG tetraloop, Chimp HAR1 GAA loop, and Human HAR1 GAA loop) included stereospecific assignments of the diastereotopic $1\text{H}5'$ and $2\text{H}5'$ protons pair. For the remaining 20 chemical shift datasets, the assignment of $1\text{H}5'$ and $2\text{H}5'$ was determined for each model based on which values gave better agreement between the experimental and back-calculated chemical shifts.

Each Rosetta model was refined and rescored under the hybrid all-atom energy:

$$E_{\text{hybrid}} = E_{\text{Rosetta}} + E_{\text{shift}} \quad [2]$$

where E_{Rosetta} is the standard Rosetta all-atom energy function for RNA¹⁶, and E_{shift} is the chemical shift pseudo-energy term. Refinement of the models under the E_{hybrid} all-atom energy function was carried out using continuous minimization in torsional space with the Davidson–Fletcher–Powell algorithm under the Rosetta framework [For this purpose, the NUCHEMICS algorithm was rewritten inside the Rosetta codebase (www.rosettacommons.org).] After refinement, the models were rescored and re-ranked under the E_{hybrid} all-atom energy function. Finally, all models were clustered, such that models with pairwise all-heavy-atom rmsd below 1.5 \AA were grouped. The lowest energy member of each cluster was designated as the cluster center and the five lowest energy cluster centers were designated the CS-ROSETTA-RNA predictions.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

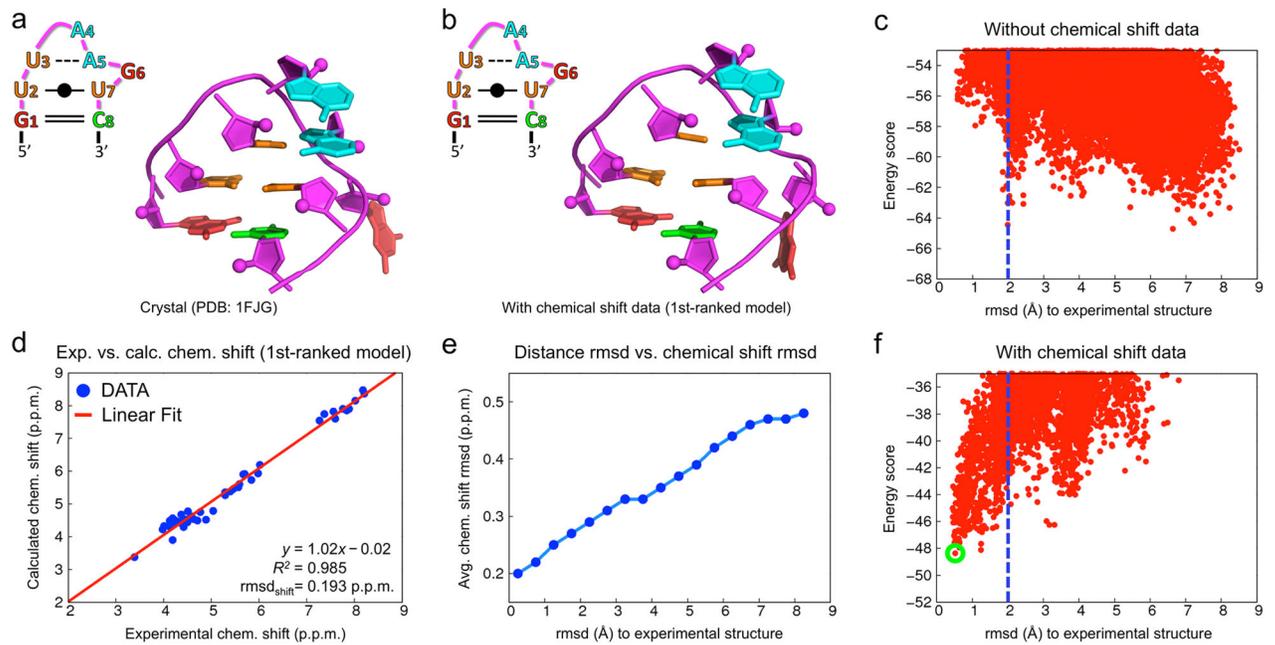
Acknowledgments

We thank H. Al-Hashimi (University of Michigan) for suggesting the problem and the Rosetta community for sharing code. We are grateful to G. Varani (University of Washington) and J. D. Puglisi (Stanford University) for providing experimental ^1H chemical shift data (for HIV-1 TAR apical loop and hepatitis C virus IRES subdomain IIa, respectively). Calculations were carried out on the BioX² cluster (NSF award CNS-0619926) and XSEDE resources (allocation MCB090153). We acknowledge financial support from a Burroughs Wellcome Career Award at the Scientific Interface (to R.D.), a C. V. Starr Asia/Pacific Stanford Graduate Fellowship (to P.S.), the US National Institutes of Health Grant GM73969 (to E.P.N.), the US National Institutes of Health Grant GM22939 (to D.H.T.), the Swiss National Science Foundation and the European Commission (BIO-NMR, ERC-Starting Grant, Marie Curie Fellowship) (to M.C.E. and R.K.O.S.), and the Deutsche Forschungsgemeinschaft (Cluster of Excellence) and the European Commission (Bio-NMR, WeNMR, Marie Curie Fellowship) (to M.C., M.Z. and H.S.).

References

1. Gesteland, RF.; Cech, T.; Atkins, JF. The RNA world: the nature of modern RNA suggests a prebiotic RNA world. Vol. 43. Cold Spring Harbor Lab Press; Cold Spring Harbor, NY: 2006.
2. Scott LG, Hennig M. Methods Mol Biol. 2008; 452:29–61. [PubMed: 18563368]
3. Schmitz U, James TL, Lukavsky P, Walter P. Nat Struct Biol. 1999; 6:634–638. [PubMed: 10404218]

4. Jovine L, et al. *Structure*. 2000; 8:527–540. [PubMed: 10801497]
5. Nabuurs SB, Spronk CAEM, Vuister GW, Vriend G. *PLoS Comput Biol*. 2006; 2:e9. [PubMed: 16462939]
6. Tolbert BS, et al. *J Biomol NMR*. 2010; 47:205–219. [PubMed: 20549304]
7. Cornilescu G, Delaglio F, Bax A. *J Biomol NMR*. 1999; 13:289–302. [PubMed: 10212987]
8. Clore GM, Gronenborn AM. *Proc Natl Acad Sci U S A*. 1998; 95:5891–5898. [PubMed: 9600889]
9. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M. *Proc Natl Acad Sci U S A*. 2007; 104:9615–9620. [PubMed: 17535901]
10. Shen Y, et al. *Proc Natl Acad Sci U S A*. 2008; 105:4685–4690. [PubMed: 18326625]
11. Case DA. *J Biomol NMR*. 1995; 6:341–346. [PubMed: 8563464]
12. Cromsig JA, Hilbers CW, Wijmenga SS. *J Biomol NMR*. 2001; 21:11–29. [PubMed: 11693565]
13. Girard FC, Ottink OM, Ampt KA, Tessari M, Wijmenga SS. *Nucleic Acids Res*. 2007; 35:2800–2811. [PubMed: 17430968]
14. van der Werf RM, Tessari M, Wijmenga SS. *J Biomol NMR*. 2013; 56:95–112. [PubMed: 23564038]
15. Frank AT, Horowitz S, Andricioaei I, Al-Hashimi HM. *J Phys Chem B*. 2013; 117:2045–2052. [PubMed: 23320790]
16. Das R, Karanicolas J, Baker D. *Nat Methods*. 2010; 7:291–294. [PubMed: 20190761]
17. Sripakdeevong P, Kladwang W, Das R. *Proc Natl Acad Sci U S A*. 2011; 108:20573–20578. [PubMed: 22143768]
18. Fleishman SJ, Baker D. *Cell*. 2012; 149:262–273. [PubMed: 22500796]
19. Deng J, Xiong Y, Pan B, Sundaralingam M. *Acta Crystallogr D Biol Crystallogr*. 2003; 59:1004–1011. [PubMed: 12777762]
20. Leontis NB, Westhof E. *RNA*. 2001; 7:499–512. [PubMed: 11345429]

**Figure 1.**

The CS-ROSETTA-RNA method illustrated on a UUAAGU hairpin. **(a)** The crystallographic structure (PDB: 1FJG). **(b)** Rosetta near-native model with a 0.52 Å all-heavy-atom rmsd to the crystallographic structure (rmsd calculated over the entire loop, excluding the flexible G6 extra-helical bulge). Two-dimensional schematics follow the Leontis and Westhof nomenclature²⁰. **(c)** Plot of the Rosetta energy vs. rmsd to the crystallographic structure for all Rosetta models before the inclusion of the chemical shift pseudo-energy term. **(d)** Plot of back-calculated chemical shifts from the Rosetta near-native model vs. experimental ¹H chemical shift values (rmsd_{shift} = 0.19 ppm). **(e)** Plot of the average rmsd_{shift} of all Rosetta models in 0.5-Å rmsd bins from the crystallographic structure. **(f)** Plot of the Rosetta energy vs. rmsd to the crystallographic structure for all Rosetta models after the inclusion of the chemical shift pseudo-energy term. With chemical shift data, the near-native model shown in **b** becomes the lowest energy model overall (green circle).

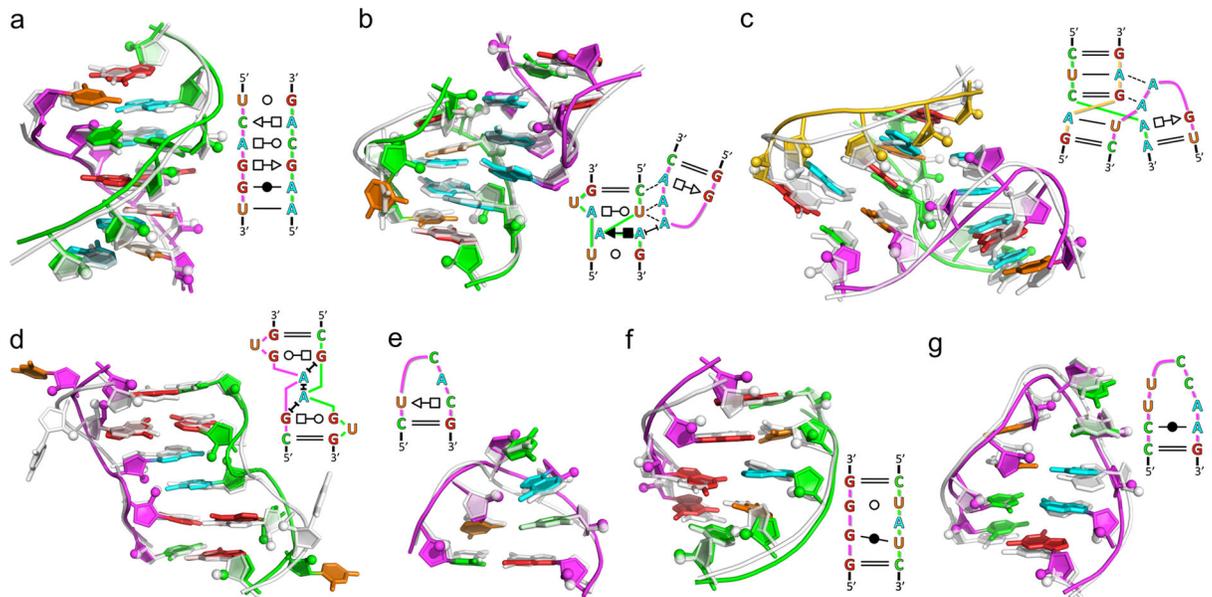


Figure 2.

Comparison of experimental and CS-ROSETTA-RNA models for diverse RNA motifs. **(a)** Conserved 4×4 internal loop from the SRP RNA (PDB: 1LNT). **(b)** GAAA tetraloop-receptor tertiary interaction motif (PDB: 2R8S). **(c)** 3-way junction from yeast mitochondrial group II intron *Sc.ai5 γ* (PDB: 2LU0). **(d)** 5'-GAGU-3'/3'-UGAG-5' self-complementary internal loop (PDB: 2LX1). **(e)** 5'-GU-3'/3'-UAU-5' internal loop from a group II intron. **(f)** Glycine tRNA(UCC) anticodon stem-loop from *Bacillus subtilis* (PDB: 2LBL). **(g)** UCAC tetraloop (PDB: 4A4S). The CS-ROSETTA-RNA models (shown in color) are overlaid on the experimental structures (shown in white). The rmsds between CS-ROSETTA-RNA models (energy cluster rank) and the experimental structure are **(a)** 0.81 Å (first), **(b)** 0.68 Å (first), **(c)** 1.74 Å (fourth), **(d)** 1.10 Å (first), **(e)** 1.37 Å (first), **(f)** 1.41 Å (third), and **(g)** 1.00 Å (first). The two-dimensional schematics are annotated based on the experimental structure and follow the Leontis and Westhof nomenclature²⁰.

Table 1

The CS-ROSETTA-RNA method benchmarked on 23 RNA motifs.

Motif name	PDB ^a	N _{nt} ^b	rmsd-top1 ^{c,d} (Å)	rmsd-top5 ^{c,e} (Å)
Known structures				
Single G:G mismatch	1F5G	6	0.71	0.71
UUCG tetraloop	2KOC	6	0.84	0.84
Tandem GA:AG mismatch	1MIS	8	1.10	1.10
Tandem UG:UA mismatch	2JSE	8	3.02	2.52
16S rRNA UUAAGU loop	1FJG	8	0.52	0.52
HIV-1 TAR apical loop	1ANR	8	5.86	5.86
tRNA _i ^{Met} ASL	1SZY	9	3.89	1.35
Conserved SRP internal loop	1LNT	12	0.81	0.81
R2 retrotransposon 4×4 loop	2L8F	12	1.17	1.17
Hepatitis C virus IRES IIa	2PN4	13	3.21	1.48
GAAA tetraloop-receptor	2R8S	15	0.68	0.68
Sc.ai5γ 3-way junction	2LU0	16	3.66	1.74
Blind targets				
UAAC tetraloop ^f	4A4R	6	0.94	0.94
UCAC tetraloop ^f	4A4S	6	1.00	1.00
UGAC tetraloop ^f	4A4U	6	3.60	1.67
UUAC tetraloop ^f	4A4T	6	1.72	1.72
Chimp HAR1 GAA loop	2LHP	7	2.88	2.88
Human HAR1 GAA loop	2LUB	7	2.26	2.03
GU:UAU internal loop	_g	9	1.37	1.37
tRNA ^{Gly} ASL (cuUCCaa) ^h	2LBL	9	3.28	1.41
tRNA ^{Gly} ASL (cuUCCcg) ^h	2LBK	9	3.42	1.94
tRNA ^{Gly} ASL (uuGCCaa) ^h	2LBJ	9	3.08	2.93
5'-GAGU/3'-UGAG loop	2LX1	12	1.10	1.10
rmsd < 1.50 Å	–	–	11/23	14/23
rmsd < 2.00 Å	–	–	12/23	18/23

Additional information and full motif names provided in Supplementary Tables 1 and 3.

^aPDB ID of reference experimental structure.^bMotif size, the number of nucleotides in the modeled RNA motif. Each motif consists of noncanonical core nucleotides closed by boundary canonical (W.C. or G:U wobble) base pairs.^cAll-heavy-atom rmsd over all nucleotides, excluding the boundary canonical base pairs after alignment over all nucleotides. Nucleotides found to be extra-helical bulges (both unpaired and unstacked) in the reference experimental structure were excluded from both the alignment and the rmsd calculation. Bold text indicates rmsd better than 2.0 Å.^dAll-heavy-atom rmsd of the first-ranked (lowest energy) model to the experimental structure.^eLowest all-heavy-atom rmsd to the experimental structure among the five lowest energy cluster centers.

^fThe 4 UNAC tetraloops were treated as separate motifs despite adopting similar conformations due to being blind targets.

^gThe experimental structure was solved by the Sigel group at University of Zurich and has not yet been deposited into the PDB database.

^hThe sequence of the 7-nt anticodon loop is given in parentheses with the anticodon triplet in uppercase.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript