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

Parin Sripakdeevong, Mirko Cevec, Andrew T Chang, Michèle C Erat, Melanie Ziegeler, Qin Zhao, George E Fox, Xiaolian Gao, Scott D Kennedy, Ryszard Kierzek, Edward P Nikonowicz, Harald Schwalbe, Roland K O Sigel, Douglas H Turner & Rhiju Das

Noncoding RNA molecules form complex three-dimensional structures that have key roles in a multitude of cellular processes, but determining their three-dimensional structures remains challenging. We demonstrate that integrating $^1$H NMR chemical shift data with Rosetta de novo modeling can be used to consistently determine 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 experimental structures with high accuracy (0.6–2.0 Å all-heavy-atom r.m.s. deviation) in 18 cases.

Noncoding RNA molecules form complex three-dimensional structures that have key roles in a multitude of cellular processes from gene regulation to viral pathogenesis. These RNAs are typically composed of canonical helices interconnected by motifs with intricate noncanonical structures critical for catalysis, binding proteins and higher-order folding. Often comprising a few dozen nucleotides or less, these motifs are compelling targets for solution NMR spectroscopy approaches. Nevertheless, NMR spectroscopy–based characterization of RNA motifs does not always generate sufficient nuclear Overhauser effect (NOE) or other restraints to produce reliable atomic-resolution three-dimensional (3D) models.

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 torsions, and to refine 3D models. More recently, chemical shift data have been leveraged for de novo determination of protein structure. Similar tools for RNA are less developed. Chemical shift assignments through NOE spectroscopy and through bond-correlation spectroscopy experiments are standard first steps in NMR spectroscopy of RNA, but chemical shift values are generally not used at the structure-determination stage. Algorithms have been developed to ‘back-calculate’ non-exchangeable $^1$H chemical shifts from RNA 3D structure. In particular, the Nuchemicals program has been used to refine models generated from conventional NMR spectroscopy measurements (NOE, J-couplings, residual dipolar couplings) and to determine de novo structures of simple helical forms of nucleic acids. A recent study also demonstrated that chemical shift data could be used to stringently constrain RNA molecular dynamics simulations starting from a known structure. 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 yet been demonstrated, to our knowledge.

Here we show that assigned $^1$H chemical shift data provide sufficient information to determine the structures of noncanonical RNA motifs at high resolution, by integrating these data with recent advances in high-resolution RNA de novo structure prediction. We named the method CS-Rosetta-RNA and extensively benchmarked it on 23 RNA motifs, including 11 motifs for which conventional NMR structural models were unreleased to the public and were kept hidden from the modelers (here referred to as ‘blind’ targets). CS-Rosetta-RNA is freely available through a web server at http://rosie.rosettacommons.org/rna_denovo.

Methods for prediction of RNA structure by fragment assembly of RNA with full-atom refinement (FARFAR) and stepwise assembly (SWA) have permitted the modeling of RNA motifs that give atomic-resolution agreement to experimentally determined structures in favorable cases. However, as in protein studies, inaccuracies in available energy functions preclude high-resolution modeling in many cases. Fortunately, in such cases correct structures are still sampled, and even quite sparse experimental data can be used to identify these models with high confidence. We illustrate the use of CS-Rosetta-RNA with a complex RNA test motif that was challenging for prior Rosetta approaches, a conserved UUAAGU hexaloop from 16S ribosomal RNA (Fig. 1a). Standard Rosetta modeling without the use of chemical shift information generated models with atomic-resolution agreement to this hexaloop’s crystallographic

---

1Biophysics Program, Stanford University, Stanford, California, USA. 2Center for Biomolecular Magnetic Resonance, Institute for Organic Chemistry and Chemical Biology, Johann Wolfgang Goethe University Frankfurt, Frankfurt, Germany. 3Department of Biochemistry and Cell Biology, Rice University, Houston, Texas, USA. 4Department of Biochemistry, University of Oxford, Oxford, UK. 5Institute of Inorganic Chemistry, University of Zurich, Zurich, Switzerland. 6Feinberg School of Medicine, Northwestern University, Chicago, Illinois, USA. 7Department of Biology and Biochemistry, University of Houston, Houston, Texas, USA. 8Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, New York, USA. 9Institute of Bioorganic Chemistry, Polish Academy of Sciences, Poznan, Poland. 10Department of Chemistry, University of Rochester, Rochester, New York, USA. 11Department of Biochemistry, Stanford University, Stanford, California, USA. 12Department of Physics, Stanford University, Stanford, California, USA. Correspondence should be addressed to R.D. (rhiju@stanford.edu).
structure (0.52 Å all-heavy-atom r.m.s. deviation; Fig. 1a,b), but these models were ranked worse in computed Rosetta energy than non-native models (>5.0 Å r.m.s. deviation; Fig. 1c). Nevertheless, the experimentally measured chemical shifts of the non-exchangeable \(^1\)H atoms were 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 (\(E_{\text{shift}}\); Online Methods) then permitted confident discrimination of the near-native models (Fig. 1f; 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 of 23 RNA motifs (Table 1 and Online Methods). 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 data and, in some cases, crystallography data, were available (Supplementary Table 1). These RNA motifs included hairpins, internal loops, a three-way junction and a tetraloop–receptor interaction. On average, each data set included 6.0 non-exchangeable \(^1\)H chemical shifts per nucleotide (out of 7–8 possible), including both ribose and base protons (Supplementary Table 1). We tested CS-Rosetta-RNA on 11 blind RNA targets that were concurrently under investigation in five NMR spectroscopy laboratories. Sequences and assigned chemical shifts for these targets, but no other information, were provided by researchers in these laboratories to the authors of this work carrying out chemical shift–guided modeling. Subsequent comparison of CS-Rosetta-RNA models with structures derived from conventional NMR spectroscopy approaches thus served as rigorous evaluation of blind targets.

Over the entire benchmark set of 23 RNA motifs, CS-Rosetta-RNA returned 18 ‘success’ cases, defined here as cases in which at least one of the five lowest-energy cluster centers achieved better than 2.0 Å all-heavy-atom r.m.s. deviation (r.m.s. deviation values and cluster ranks are provided in Table 1 and Supplementary Tables 2 and 3; energy versus r.m.s. deviation 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 spectroscopy structures and the CS-Rosetta-RNA models (Supplementary Results, and Supplementary Figs. 4 and 5). CS-Rosetta-RNA performed well on both the test set of known structures (10/12 success cases) and the blind targets (8/11 success cases). 11 of the 23 cases satisfied a more stringent success criterion: the lowest-energy (top-ranked) model was within 1.5 Å all-heavy-atom r.m.s. deviation of the experimental structure (Table 1). Lastly, incorporating even sparse data (~1 chemical shift per nucleotide) 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 (r.m.s. deviation, 0.81 Å; Fig. 2a), a GAAA tetraloop–receptor interaction
Table 1 | The CS-Rosetta-RNA method benchmarked on 23 RNA motifs

| Motif name                        | PDB identifiera | Motif sizeb | r.m.s. deviation, lowest-energy modelc,d (Å) | r.m.s. deviation, top five lowest-energy modelsc,e (Å) |
|-----------------------------------|-----------------|-------------|---------------------------------------------|-----------------------------------------------------|
| **Known structures**              |                 |             |                                             |                                                      |
| Single G-G mismatch               | 1F5G            | 6           | 0.71 (ranked first)                         | 0.71 (ranked first)                                  |
| UUUC tetraloop                    | 2KOC            | 6           | 0.84 (first)                                | 0.84 (first)                                         |
| Tandem GA-GG mismatch             | 1MIS            | 8           | 1.10 (first)                                | 1.10 (first)                                         |
| Tandem GU-UA mismatch             | 2JSE            | 8           | 3.02 (first)                                | 2.52 (first)                                         |
| 16S rRNA UUAAGU loop              | 1ANR            | 8           | 5.86 (first)                                | 5.86 (first)                                         |
| tRNA^Met ASL                      | 1SZY            | 9           | 3.89 (first)                                | 1.35 (first)                                         |
| Conserved SRP internal loop       | 1LNT            | 12          | 0.81 (first)                                | 0.81 (first)                                         |
| R2 retrotransposon 4 × 4 loop     | 2LBF            | 12          | 1.17 (first)                                | 1.17 (first)                                         |
| Hepatitis C virus IRES IIa        | 2PN4            | 13          | 3.21 (first)                                | 1.48 (first)                                         |
| GAAA tetraloop receptor           | 2RBS            | 15          | 0.68 (first)                                | 0.68 (first)                                         |
| Sc.ai5 three-way junction         | 2LU0            | 16          | 3.66 (first)                                | 1.74 (first)                                         |
| **Blind targets**                 |                 |             |                                             |                                                      |
| UAAC tetraloopf                   | 4A4R            | 6           | 0.94 (first)                                | 0.94 (first)                                         |
| UCAC tetraloopf                   | 4A4S            | 6           | 1.00 (first)                                | 1.00 (first)                                         |
| UGAC tetraloopf                   | 4A4U            | 6           | 3.60 (first)                                | 1.67 (first)                                         |
| UUAC tetraloopf                   | 4A4T            | 6           | 1.72 (first)                                | 1.72 (first)                                         |
| Chimp HAR1 GAA loop               | 2LHP            | 7           | 2.88 (first)                                | 2.88 (first)                                         |
| Human HAR1 GAA loop               | 2LUB            | 7           | 2.26 (first)                                | 2.03 (first)                                         |
| GU-UAA internal loop              | 9               | 9           | 1.37 (first)                                | 1.37 (first)                                         |
| tRNA^Val ASL (cuUCca)h            | 2LBBL           | 9           | 3.28 (first)                                | 1.41 (first)                                         |
| tRNA^Met ASL (cuUCccg)h           | 2LBK            | 9           | 3.42 (first)                                | 1.94 (first)                                         |
| tRNA^Met ASL (cuUCCca)h           | 2LBJ            | 9           | 3.08 (first)                                | 2.93 (first)                                         |
| 5′-GAGU–3′-UGAG loop              | 2LX1            | 12          | 1.10 (first)                                | 1.10 (first)                                         |

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

aPDB identifier of reference experimental structure. bNumber 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 r.m.s. deviation over all nucleotides, excluding the boundary canonical base pairs after alignment over all nucleotides. Nucleotides found to be extrahelical bulges (both unpaired and unstacked) in the reference experimental structure were excluded from both the alignment and the r.m.s. deviation calculation. dAll-heavy-atom r.m.s. deviation of the first-ranked (lowest energy) model to the experimental structure. eLowest all-heavy-atom r.m.s. deviation to the experimental structure among the five lowest-energy cluster centers. fThe four UNAC tetraloops were treated as separate motifs despite adopting similar conformations owing to being blind targets. gExperimental structure (unpublished data; M.C. Erat and R.K.O. Sigel) has not yet been deposited into PDB. hSequence of the 7-nucleotide anticodon loop is given in parentheses with the anticodon triplet in upper case.

Figure 2 | Comparison of experimental and CS-Rosetta-RNA models for diverse RNA motifs. (a–g) CS-Rosetta-RNA models (in color) overlaid on the experimental structures (in white) for conserved internal loop from the SRP RNA (a; PDB: 1LNT), GAAA tetraloop-receptor tertiary interaction motif (b; PDB: 2RBS), three-way junction from yeast mitochondrial group II intron Sc.ai5 (c; PDB: 2LU0), 5′-GAGU–3′-UGAG-5′ self-complementary internal loop (d; PDB: 2LX1), UCAC tetraloop (e; PDB: 4A45), 5′-GU–3′-UAU-5′ internal loop from a group II intron (f), glycine tRNA(UCC) anticodon stem-loop from *Bacillus subtilis* (g; PDB: 2LBL). The r.m.s. deviation values between CS-Rosetta-RNA models (rank of model by energy given in parentheses) and the experimental structure are (a) 0.81 Å (ranked first), (b) 0.68 Å (first), (c) 1.74 Å (fourth), (d) 1.10 Å (first), (e) 1.00 Å (first), (f) 1.37 (first) and (g) 1.41 Å (third). The two-dimensional schematics are annotated based on the experimental structure and follow nomenclature in ref. 20.
(r.m.s. deviation, 0.68 Å; Fig. 2b), a three-way junction from yeast mitochondrial group II intron Sc.ai57 (r.m.s. deviation, 1.74 Å; Fig. 2c), and both the major and minor conformations of a G:G mismatch (Supplementary Fig. 7). Successful blind target cases included predictions for a highly irregular 5′-GAGU-3′-UG AG-5′ self-complementary internal loop that required synthesizing and probing additional constructs to solve by conventional NMR spectroscopy (r.m.s. deviation, 1.10 Å; Fig. 2d), all four UNAC tetraloops (where N refers to any nucleotide; Fig. 2e), a 5′-GU-3′-UAU-5′ internal loop from a group II intron (r.m.s. deviation, 1.37 Å; Fig. 2f) and a CUUCAA anticondon stem-loop of Bacillus subtilis RNA34 (r.m.s. deviation, 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 Rosetta units (approximately equal to kT, where k is the Boltzmann constant and T is temperature, 37 °C) to the next-lowest energy cluster and, in all of these cases, the model achieved better than 1.5 Å r.m.s. deviation to experimental structure (Supplementary Fig. 8). This energy gap thus appears to be a hallmark of CS-Rosetta-RNA accuracy (Supplementary Results). In one apparent exception, the SRP conserved internal loop, a large energy gap (5.5 Rosetta units) strongly suggested that the CS-Rosetta-RNA prediction should be accurate, but the lowest-energy CS-Rosetta-RNA model disagreed with the experimental NMR spectroscopy models3 (>2.0 Å r.m.s. deviation; Supplementary Fig. 9a,b). Additional analysis revealed that the experimental NMR spectroscopy models poorly explained the 1H chemical shift data published in the same study3 (r.m.s. deviationshift = 0.50 p.p.m.) and poorly agreed with subsequently solved crystallographic structures4,19 (r.m.s. deviation of 2.30 Å to Protein Data Bank (PDB) identifier 1LNT19). In contrast, the CS-Rosetta-RNA model gave excellent agreement with the chemical shift data (r.m.s. deviationshift = 0.18 p.p.m.) and closely matched the crystallographic structures (r.m.s. deviation of 0.81 Å to PDB identifier 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 spectroscopy–derived structures.

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. The standard approach2 of determining NOEs, J-couplings and, in some cases, residual dipolar couplings, does not always yield sufficient information to determine an RNA’s 3D structure by conventional means, as illustrated by the 5′-GAGU-3′-3′-UGAG-5′ case (Fig. 2d; see Supplementary Note, and Supplementary Figs. 10 and 11 for further modeling details of this highly irregular motif). Further integration of de novo modeling and NMR methodologies, including the incorporation of 13C, 15N and exchangeable 1H chemical shift data (Supplementary Results), may help accelerate determination of RNA structure and eventually help solve currently intractable 3D RNA structures.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank H. Al-Hashimi for suggesting the problem and the Rosetta community for sharing code, and G. Varani and J.D. Puglisi for providing experimental 1H chemical shift data (for HIV-1 TAR apical loop and hepatitis C virus IRES subdomain IIα, respectively). Calculations were carried out on the BioX2 cluster (US National Science Foundation award CNS-0619926) and Extreme Science and Engineering Discovery Environment (XSEDE) (allocation MCB090153). We acknowledge financial support from a Burroughs Wellcome Career Award at the Scientific Interface (to R.D.), US National Institutes of Health (NIH) grant R21 GM102716 (to R.D.), a C.V. Starr Asia/Pacific Stanford Graduate Fellowship (to P.S.), NIH grant GM73969 (to E.P.N.), NIH 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.E., M.Z. and H.S.).

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 NMR spectroscopy data for the 11 blind targets and participated in evaluating the blinded trials. All authors discussed the results and commented on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.
ONLINE METHODS

Generation of Rosetta models. Two complementary structure modeling methods, FARFAR\(^\text{18}\) and SWA\(^\text{17}\), 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 previously\(^\text{17}\). Each step involved enumerating several million conformations for each nucleotide, and all step-by-step build-up paths were covered in \(N\) building steps, where \(N\) is the number of nucleotides in the motif. At the final building steps, all models were 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 with respect 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 FARFAR in the Rosetta framework, as described previously\(^\text{16}\); the fragment source was the large ribosomal subunit of \(H.\) marismortui (PDB: 1JJ2). For each motif, 250,000 FARFAR SWA 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, which led to ~10,000–20,000 final models. The SWA and FARFAR build-up paths were covered in ~1 million conformations for each nucleotide, and all step-by-step generated models were finely clustered and a maximum of 10,000 low-energy FARFAR models were retained. The SWA approach is effective at generating models that are highly optimized with respect 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 FARFAR in the Rosetta framework, as described previously\(^\text{16}\); the fragment source was the large ribosomal subunit of \(H.\) marismortui (PDB: 1JJ2). For each motif, 250,000 FARFAR SWA 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, which led 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 models were then combined, which led 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 motif. The majority of the computations for this work were performed on Stanford University’s Bio-X 2 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 d (of wall-clock time) to perform 50,000 CPU hours of computation.

Incorporation of non-exchangeable \(^1\)H chemical shifts into structure modeling with CS-Rosetta-RNA. Information from the experimental non-exchangeable \(^1\)H chemical shifts was incorporated into the modeling process through the chemical shift pseudo-energy term:

\[
E_{\text{shift}} = c \sum_i (\delta_i^{\text{exp}} - \delta_i^{\text{calc}})^2
\]

(1)

where \(\delta_i^{\text{exp}}\) and \(\delta_i^{\text{calc}}\) are, respectively, the experimental and back-calculated chemical shift in p.p.m. units (the index \(i\) sums over all experimentally assigned non-exchangeable \(^1\)H chemical shifts in the RNA motif), and \(c\) is a weighting factor set to 4.0 \(k_B T/\text{p.p.m.}^2\) based on test runs with different motifs. The Nuchemcs program\(^\text{12}\) was used to back-calculate non-exchangeable \(^1\)H chemical shifts. In the 23-RNA-motif benchmark set, only three chemical shift data sets (UUCG tetraloop, chimp human accelerated region 1 (HAR1) GAA loop and human HAR1 GAA loop) included stereospecific assignments of the diastereotopic 1H5′ and 2H5′ protons pair. For the remaining 20 chemical shift data sets, the assignment of 1H5′ and 2H5′ 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 rescoped under the hybrid all-atom energy:

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

where \(E_{\text{Rosetta}}\) is the standard Rosetta all-atom energy function for RNA\(^\text{16}\), and \(E_{\text{shift}}\) is the chemical shift pseudo-energy term. Refinement of the models under the \(E_{\text{hybrid}}\) 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 Nuchemcs\(^\text{12}\) algorithm was rewritten into the Rosetta code base, http://www.rosetta-commons.org. After refinement, the models were rescoped and reranked under the \(E_{\text{hybrid}}\) all-atom energy function. Finally, all models were clustered, such that models with pairwise all-heavy-atom r.m.s. deviation below 1.5 Å 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.

Sources of experimental PDB structures used in this study. The sources of the experimental PDB structures used in this study were: (1) single G:G mismatch (PDB: 1FSG\(^\text{21}\), PDB: 1F5H\(^\text{21}\)); (2) UUCG tetraloop (PDB: 2KOC\(^\text{22}\), PDB: 1F7Y\(^\text{23}\)); (3) tandem GA:AG mismatch (PDB: 1MIS\(^\text{24}\)); (4) tandem UG:UA mismatch (PDB: 2JSE\(^\text{25}\)); (5) 16S rRNA UUAAGU (PDB: 1FG\(^\text{26}\), PDB: 1HS2; ref. 27); (6) HIV-1 TAR apical loop (PDB: 1ANR\(^\text{28}\)); (7) tRNA\(_{\text{Met}}\) (PDB: 1DSY\(^\text{27}\)); (8) conserved SRP internal loop (PDB: 1LNT\(^\text{19}\), PDB: 28SR\(^\text{26}\), PDB: 28SP\(^\text{3}\)); (9) R2 retrotransposon 4x4 loop (PDB: 2L8E\(^\text{30}\)); (10) hepatitis C virus IRES IIa (PDB: 2PN4; ref. 31, PDB: 1P5M\(^\text{32}\)); (11) GAAA tetraloop-receptor (PDB: 2R8S\(^\text{33}\), PDB: 2ADT\(^\text{34}\)); (12) Sc.ai5 three-way junction (PDB: 2LU0; ref. 35); (13) 16S rRNA UUAAGU (PDB: 4A4R\(^\text{36}\)); (14) UCAC tetraloop (PDB: 4A4S\(^\text{36}\)); (15) UGAC tetraloop (PDB: 4A4U\(^\text{36}\)); (16) UCAC tetraloop (PDB: 4A4T\(^\text{36}\)); (17) chimp HAR1 GAA loop (PDB: 2LHP\(^\text{37}\)); (18) human HAR1 GAA loop (PDB: 2LU0\(^\text{37}\)); (19) GU:UAU internal loop (unpublished data; M.C. Erat and R.K.O. Sigel; not yet deposited into PDB); (20) tRNA\(_{\text{Gly}}\) ASL (cuUCCaa) (PDB: 2LBI\(^\text{38}\)); (21) tRNA\(_{\text{Gly}}\) ASL (cuUCCcg) (PDB: 2LBK\(^\text{39}\)); (22) tRNA\(_{\text{Gly}}\) ASL (uUCCGa) (PDB: 2LB\(^\text{38}\)); and (23) 5′-GAGU-′3′-UGAG loop (PDB: 2LX1; ref. 39).

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

21. Burkard, M.E. & Turner, D.H. *Biochemistry* **39**, 11748–11762 (2000).
22. Nozinovic, S., Furtig, B., Jonker, H.R., Richter, C. & Schwalbe, H. *Nucleic Acids Res.* **38**, 683–694 (2010).
23. Ennifar, E. et al. *J. Mol. Biol.* **304**, 35–42 (2000).
24. Wu, M. & Turner, D.H. *Biochemistry* **35**, 9677–9689 (1996).
25. Shankar, N. et al. *Biochemistry* **46**, 12665–12678 (2007).
26. Carter, A.P. et al. *Nature* **407**, 340–348 (2000).
27. Zhang, H., Fountain, M.A. & Krugh, T.R. *Biochemistry* **40**, 9879–9886 (2001).
28. Aboul-ela, F., Karn, J. & Varani, G. *Nucleic Acids Res.* **24**, 3974–3981 (1996).
29. Schweisguth, D.C. & Moore, P.B. *J. Mol. Biol.* **267**, 505–519 (1997).
30. Lerman, Y.V. et al. *RNA* **17**, 1664–1677 (2011).
31. Zhao, Q., Han, Q., Kissing, C.R., Hermann, T. & Thompson, P.A. *Acta Crystallogr. D Biol. Crystallogr.* **64**, 436–443 (2008).
32. Lukavsky, P.J., Kim, I., Otto, G.A. & Puglisi, J.D. *Nat. Struct. Biol.* **10**, 1033–1038 (2003).
33. Ye, J.D. et al. *Proc. Natl. Acad. Sci. USA* **105**, 82–87 (2008).
34. Davis, J.H. et al. *J. Mol. Biol.* **351**, 371–382 (2005).
35. Donghi, D., Pechlaner, M., Finazzo, C., Knobloch, B. & Sigel, R.K. *Nucleic Acids Res.* **41**, 2489–2504 (2013).
36. Zhao, Q. et al. *Biopolymers* **97**, 617–628 (2012).
37. Ziegeler, M., Cevec, M., Richter, C. & Schwalbe, H. *ChemBioChem* **13**, 2100–2112 (2012).
38. Chang, A.T. & Nikonowicz, E.P. *Biochemistry* **51**, 3662–3674 (2012).
39. Kennedy, S.D., Kierzek, R. & Turner, D.H. *Biochemistry* **51**, 9257–9259 (2012).