# Structure modeling of RNA using sparse NMR constraints

Benfeard Williams II[1,2,†], Bo Zhao[2,3,†], Arpit Tandon[1,2], Feng Ding[4], Kevin M Weeks[3], Qi Zhang[1,2,*], Nikolay V Dokholyan[1,2,*]

[1] Department of Biochemistry and Biophysics, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
[2] Molecular and Cellular Biophysics Program, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
[3] Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA.
[4] Department of Physics and Astronomy, Clemson University, Clemson, SC 29634, USA.

* To whom correspondence should be addressed. Tel: +1 919 843 2513; Fax: +1 919 966 2852; Email: dokh@unc.edu Correspondence may also be addressed to: Tel: +1 919 966 5770; Fax: +1 919 966 2852; Email: zhangqi@unc.edu

† The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

**This document includes:**

**Supplementary Text**

Brief descriptions of each RNA structure

The guanosine binding site of Group I intron from *Tetrahymena thermophila* is a 22-nt RNA that mimics the binding of the 3' terminus of the intron with the guanosine-binding site during the second transesterification reaction. In the structure reported by Kitumura *et al.* (PDB ID: 1K2G (1)), the conserved 3' guanosine forms a base triple interaction with a G·C base pair by binding in the minor groove. The base triple forms stacking interactions with the neighboring A·U base pair. These interactions are revealed through the inter-proton connectivity between G3, G22, and A4. The minimum energy structure from iFoldNMR has a backbone RMSD of 3.1 Å and all-atom RMSD of 3.7 Å relative to the published structure. Notably, although the overall orientation of the base triple is correct, the terminal G is slightly off plane in the triple in the iFoldNMR structure. The low RMSD between the two structures upon exclusion of loops and bulges suggests that the reason for the difference between the iFoldNMR structure and the published structure is due to the lack of experimental constraints on the loop conformation.

The mouse mammary tumor virus pseudoknot is required for ribosomal frameshifting during translation of the viral RNA. The NMR structure of a mutant of this pseudoknot, which does not efficiently frameshift, was presented by Kang *et al.* (PDB ID: 1KAJ (2)). The RNA element is formed by two helices, bent relative to each other but containing no other complex elements beyond the pseudoknot. Overall the backbone RMSD between the iFoldNMR structure and the published structure is 5.4 Å, and the all-atom RMSD is 5.1 Å. The high RMSD is mainly due to the different conformation of the eight-membered loop, which makes no contacts with the rest of the structure. Comparison of the iFoldNMR and published structures excluding this loop reveals the predicted structure is much closer than the RMSDs would suggest.

The *Aquifex aeolicus* tmRNA pseudoknot PK1 is a highly compact pseudoknot required for a bacterial ribosomal rescue mechanism (PDB ID: 2G1W (3)). The structure contains a series of non-canonical hydrogen bonding interactions involving no direct base pairing interactions. The presence of a non-canonical interaction involving G4, C11, and A16 results in a tightly packed structure. The backbone RMSD for the iFoldNMR vs. published structure is 4.3 Å and the all-atom RMSD is 5.0 Å. The key differences between our structure and the published structure are the result of minor differences in the backbone and in non-constrained bases; this explains why the all-atom RMSD is worse than the backbone RMSD. The only available sparse constraints consist of amino-amino distances, which, like non-exchangeable protons, are more difficult to assign than the iminos. When the loops and bulges are ignored, the overall RMSD is 4.6 Å, highlighting the ability of iFoldNMR to determine the overall topology but its inability to predict the exact conformation of flexible regions.

The HIV-2 TAR hairpin kissing dimer consists of three coaxially stacked helices that form between the HIV-2 TAR hairpin loop and a complimentary helix (PDB ID: 1KIS (4)). The topology was reproduced accurately by iFoldNMR with an RMSD compared to the published structure of 2.4 Å.

The murine leukemia virus pseudoknot is an RNA switch that induces ribosomal read-through. The structure presented by Houck-Loomis and co-workers is of the inactive pseudoknot conformation (PDB ID: 2LC8 (5)). The pseudoknot consists of two coaxially stacked helices connected by a large loop (L2) and a single nucleotide bulge (L1), which forms a key base-triple in the active conformation. As was the case for the mouse mammary tumor virus pseudoknot (PDB ID: 1KAJ), the high backbone RMSD of 4.0 Å and all-atom RMSD of 6.7 Å when the iFoldNMR structure of the leukemia virus pseudoknot is compared to the published structure are due to the relatively unconstrained loop. The backbone and all-atom RMSDs without this unconstrained loop are 3.8 Å and 4.3 Å, respectively

The pea enation mosaic virus P1-P2 pseudoknot is one of two elements required for efficient ribosomal frameshifting in this virus. In the published structure of this pseudoknot loop nucleotides make non-canonical contacts with the short helical stem (PDB ID: 2RP0 (6)). Inter-proton NOE distances and $J_{NN}$-COSY reveal the presence of a base-triple involving a protonated cytosine along the Hoogsteen edge of guanosine and the formation of a Hoogsteen base pair between A27 and U09. In our simulations, we included distance imino-imino distance constraints but ignored the presence of protonation on the cytosine. The backbone topology was reproduced by iFoldNMR with a backbone RMSD of 4.6 Å. The largest differences between our structure and the published structure involve loop residues, which have few contacts with the helix. Heterogeneity possible in the loop results in differences in the relative alignment of the two helices. The all-atom RMSD of 5.3 Å was improved to an RMSD of 4.6 Å when loop and bulge residues were excluded.

The sugarcane yellow leaf virus mRNA pseudoknot promotes programmed ribosomal frameshifting. It has a topology similar to that of the pea enation mosaic virus P1-P2 pseudoknot (PDB ID: 2RP0). The sugarcane virus pseudoknot consists of two helices connected by two loops with some interactions between the loop nucleotides and helices and a protonated cytosine base triple (PDB ID: 1YG3 (7)). iFoldNMR reproduced the overall topology of the published structure with lack of constraints in loops. The backbone RMSD relative to the published structure is 3.4 Å, and the all-atom RMSD is 4.8 Å. Some of the differences occur due to non-canonical hydrogen bonding involving the 2' OHs of several residues that are not accurately reproduced in our predictions as no constraints involving this position were used in the iFoldNMR pipeline.

The *Bacillus subtilis* $PreQ_1$ riboswitch class I aptamer regulates queuosine biosynthesis. The riboswitch bound to $PreQ_1$, a biosynthetic precursor to queuosine, induces transcription termination of genes that regulate queuosine biosynthesis. The riboswitch forms a compact pseudoknot with the formation of two base triples to create a binding pocket that recognizes $PreQ_1$ through two residues, C19 and A32 (PDB ID: 2LIV (8)). In the presence of $PreQ_1$, iFoldNMR reproduces the overall topology of the published structure, although loop 3 adopts a different conformation in our structure. Instead of folding against the minor groove of P1, it is looped out and away from P1 in our structure. As all nucleotides in the loop adenosine, no imino signals inform the relative orientation of this loop. A relatively unconstrained bulge, G02, also contributes to the relatively large backbone RMSD of 6.6 Å. The all-atom RMSD of 5.5 Å was improved to 4.9 Å when the bulge and loop were ignored.

The *Kluyveromyces lactis* telomerase RNA pseudoknot structure consists of two coaxially stacked helices with two loops that interact with the major groove through a series of base triples;

the formation of these triples is critical for telomerase function (PDB ID: 2M8K (9)). The iFoldNMR structure closely replicates both the overall topology (with a backbone RMSD of 4.1 Å) and fine structures of the base triples (all-atom RMSD 4.5 Å). The difference between all-atom and backbone RMSDs is due to loop residues that were unconstrained in the DMD simulation. Non-hydrogen bonded proton constraints guided development of the published model; these constraints pulled the loop into the groove of the helix.

The *Streptococcus pneumonia* $PreQ_1$ class II riboswitch (PDB ID: 2MIY (10)) belongs to a class of $PreQ_1$ binding riboswitches that are structurally distinct but functionally similar to the $PreQ_1$ riboswitch described above. Like the class I riboswitch, the consensus class II riboswitch forms a H-type pseudoknot fold. Although the overall architecture of the published riboswitch structure was reproduced by iFoldNMR modeling, there are key differences between individual helices as well as the relative orientations of the helices. The most notable difference is in relative orientation of the J2-4 loop with respect to stem 1 of the H-type pseudoknot domain. This loop is on the outside of the helix in the NMR structure. In our structural model the J4-3 helix is almost parallel to stem 1, whereas in the published structure it is perpendicular. The difference between the J2-4 helix orientation and difference in the J4-3 helix orientations accounts for the observed RMSD of 6.5 Å between the two structures.

The *Neurospora* Varkud satellite ribozyme is an RNA enzyme that catalyzes phosphodiester bond self-cleavage and self-ligation critical for replication of VS RNA. The ribozyme is a large multi-domain RNA consisting of six helical domains (I-VI). Although the large size of this RNA (>800 nt) makes it difficult to study by NMR, the modular nature of RNA enables studies of isolated domains. Legault *et al*. have reported structures of three VS ribozyme domains (11-13). Stem I, the substrate, forms a kissing interaction with stem V, an interaction important in substrate recognition and preparation for catalysis. The remaining stems form the catalytic core of the ribozyme. Stems II-III-VI and III-IV-V each form three-way junctions.

The Varkud satellite ribozyme stem I-V kissing-loop interaction is a magnesium-dependent interaction that forms between the substrate (stem-loop I) and the catalytic domain (stem-loop V); this domain has been characterized by high-resolution NMR (PDB ID: 2MI0 (11)). The iFoldNMR model accurately reproduces the overall topology of the published NMR structure with a backbone RMSD of 4.4 Å and an all-atom RMSD of 4.5 Å. The differences in the models result from the orientation of the two helices relative to each other as alignments within individual helices agree quite well. The minor difference in the upper helix results lack of constraints on the base pairing of C07 and G14.

The Varkud satellite ribozyme II-III-VI three-way junction serves to orient the single nucleotide bulge of stem loop VI with the active site of stem loop I (PDB ID: 2N3R (13)). A difficulty in modeling of three-way junctions is the lack of long-range orientation constraints that confine helices relative to each other. Our iFoldNMR model accurately predicted the stacking of helices of stem loops II and III, with an RMSD of 4.8 Å. The largest difference between the structures is due to the orientation of stem loop VI relative to the other two helices. This major difference results in backbone and all-atom RMSDs of 15.9 Å and 13.4 Å, respectively. There are no non-solvent-exchangeable proton NOEs that would have guided positioning of A06 between the A32-A46 and the G34-C45 base pairs the iFoldNMR simulation.

The Varkud satellite ribozyme III-IV-V three-way junction ensures proper orientation of stem loop V with stem loop I (PDB ID: 2MTJ (12)). As described above, modeling of the orientations of stems in a three-way junction presents difficulties due to lack of NOE-based constraints. Although the region of the coaxially stacked helices of stems III and IV have an RMSD of 3.3 Å when the iFoldNMR model is compared to the published structure, the relative orientation of stem V is not recapitulated, leading to an all-atom RMSD of 7.6 Å. This is likely due to the failure to reproduce the U-turn structure for residues U37, G38, and A39, which serves to align helix V in the published structure. This motif was positioned in the published structure by a series of non-solvent-exchangeable constraints on oxygen- and amino-mediated hydrogen bonds; these cannot be detected or assigned using 11-echo-NOESY and $J_{NN}$-COSY type experiments.

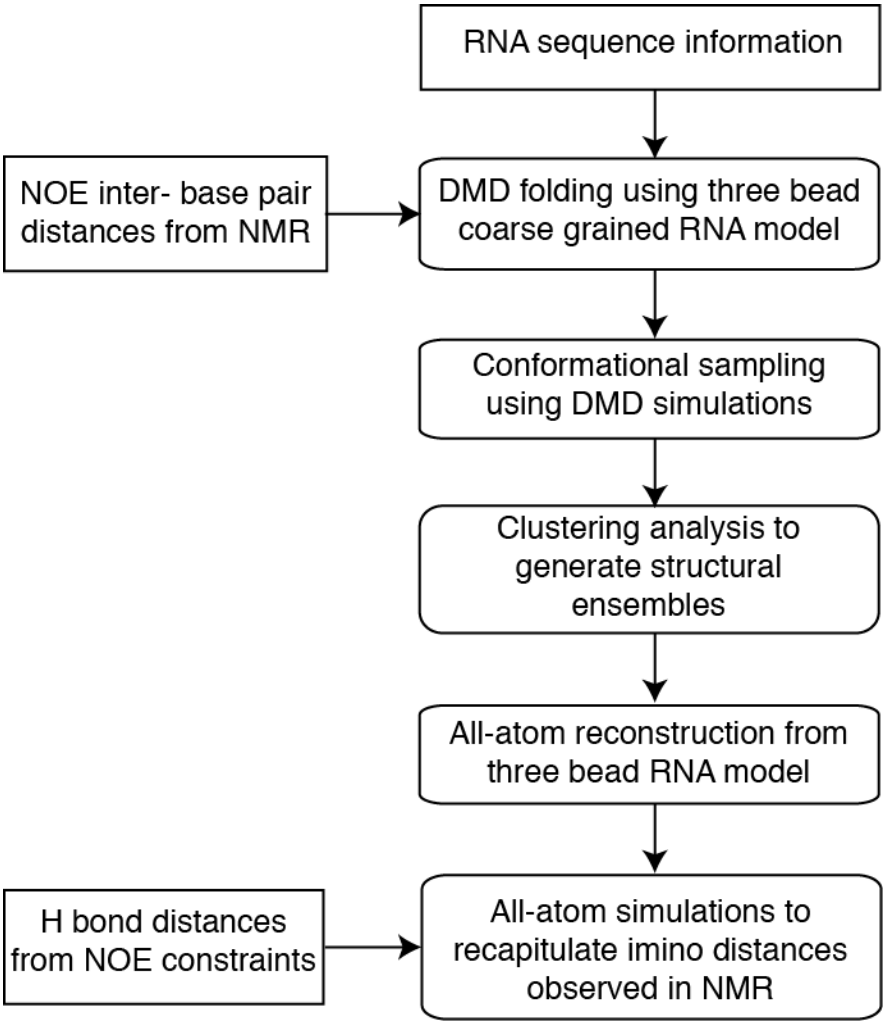


**Figure S1.** iFoldNMR: an NMR guided DMD workflow.

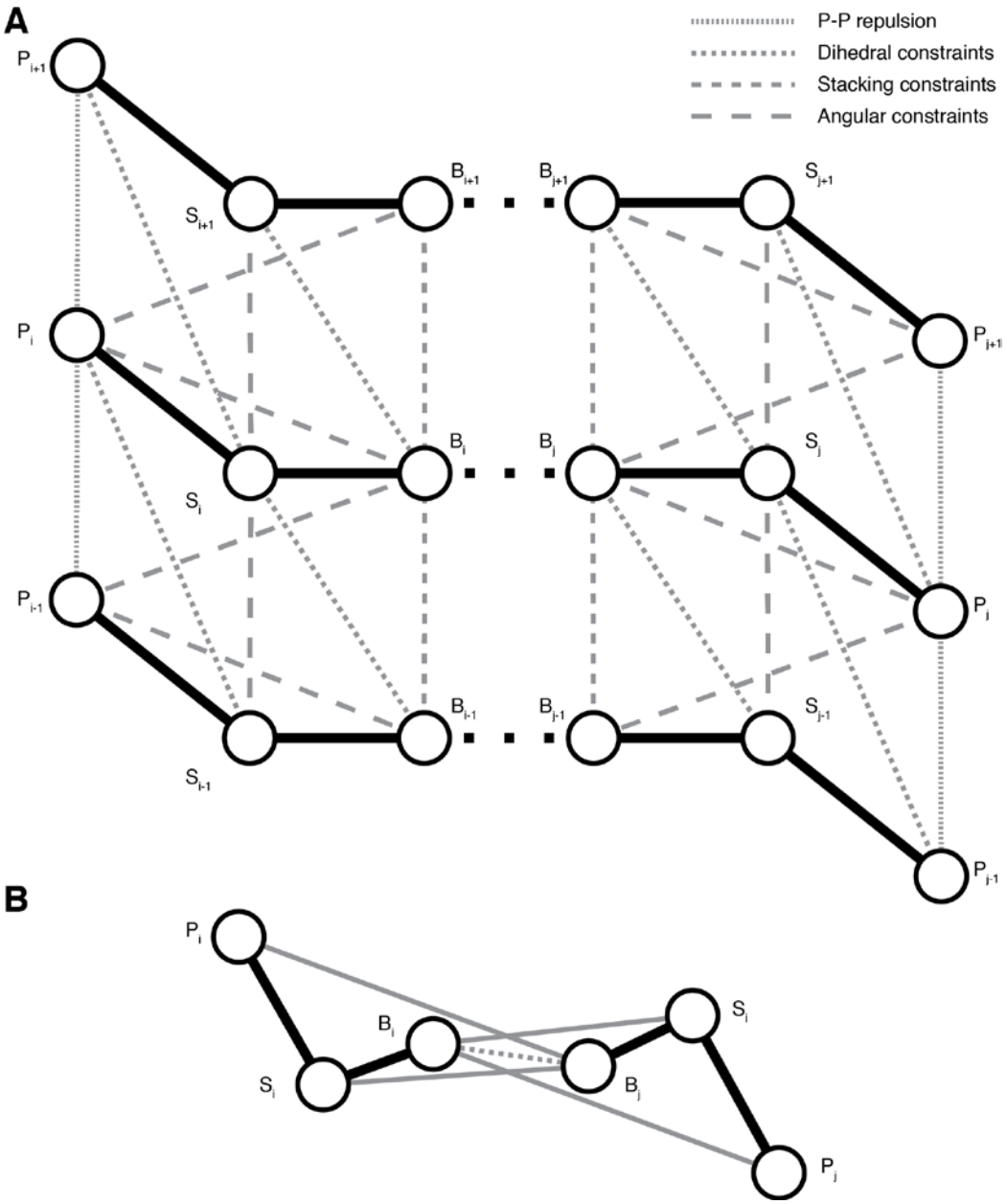


**Figure S2.** Coarse-grained structural model of RNA used in DMD simulations. (A) The coarse-grained RNA representation consists of phosphate (P), base (B), and sugar (S) beads. The thick lines correspond to the covalent interactions, dashed lines correspond to angular constraints, dashed dotted lines correspond to dihedral constraints. Additional steric constraints are used to model base stacking. (B) Hydrogen bonds in RNA base pairing are modeled using our reaction algorithm **(14)**. Base pairing between bases are shown using dashed lines.

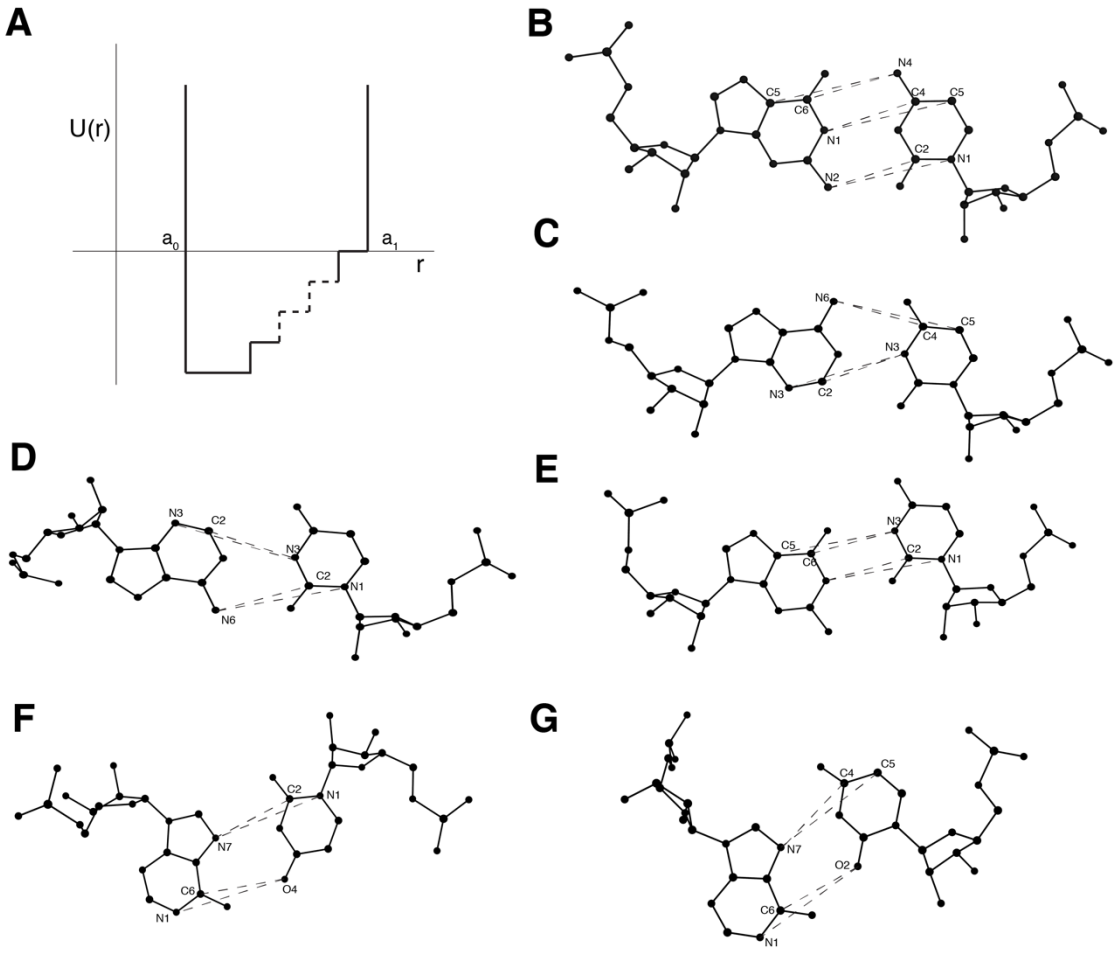


**Figure S3.** Constraints and base pair interactions in DMD. (A) Non-bonded interactions between the atoms are modeled using a stepwise potential function to constrain the distances between atoms. Atom pair constraints used for iFoldNMR are shown for (B) G·C and (C) A·U Watson-Crick base pairs, (D) A·U reverse Watson-Crick base pairs, (E) G·U wobble base pairs, (F) A·U Hoogsteen base pairs and (G) A·U reverse Hoogsteen base pairs.

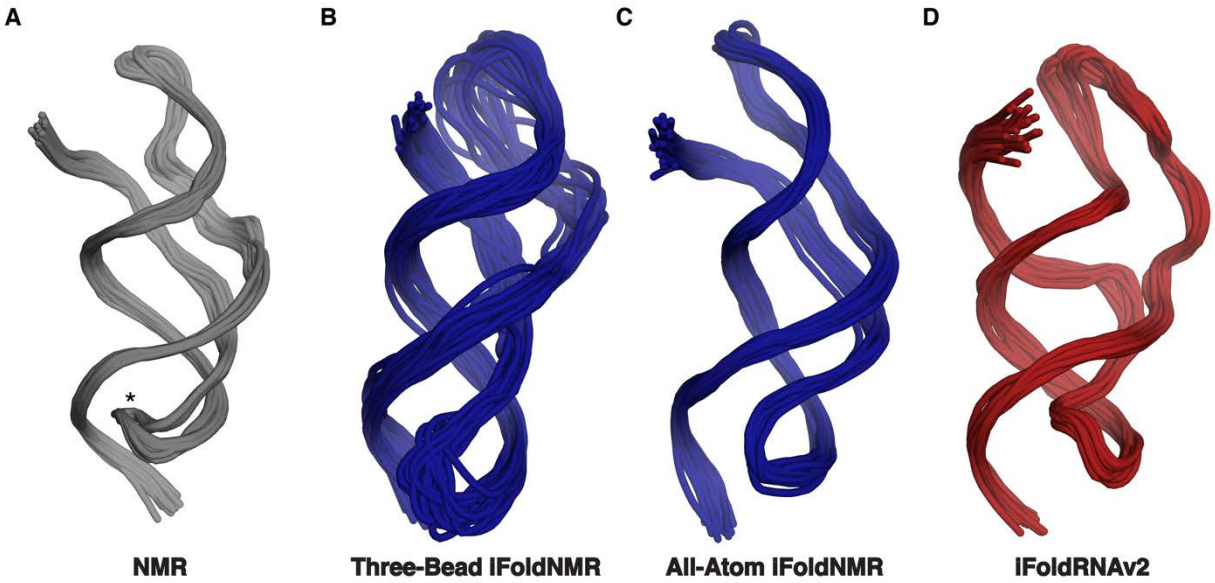


**Figure S4.** Human telomerase pseudoknot backbone traces for the 20 lowest energy structures from (A) the published NMR structure (PDB ID: 2K96)**(15)** with a characteristic kink(*), (B) the three-bead iFoldNMR structural models after clustering analysis, (C) the all-atom iFoldNMR models, and (D) the iFoldRNAv2 models without sparse NMR constraints.

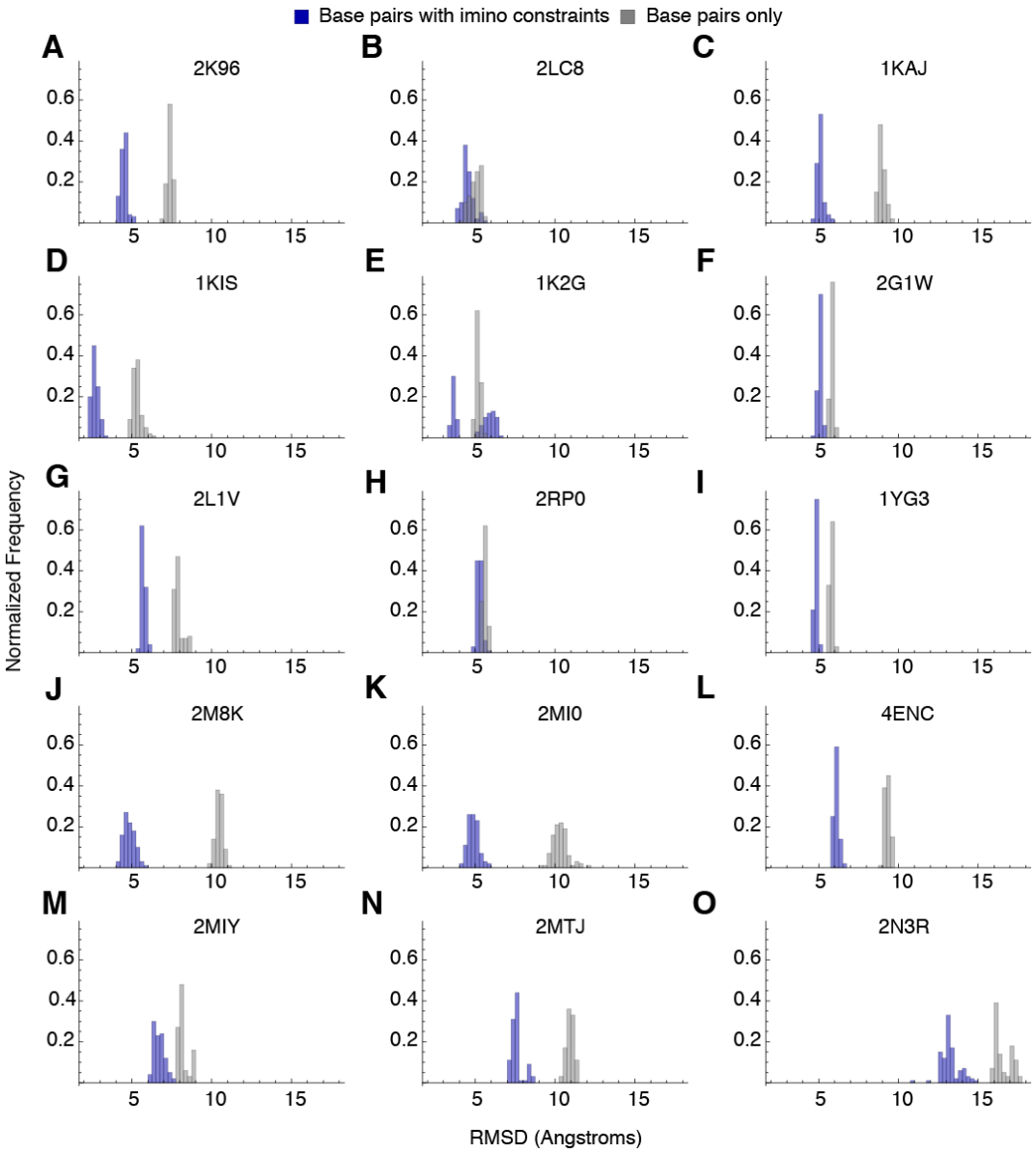


**Figure S5.** RMSD distributions of predicted structures. Histograms for the RMSD values of modeled RNA structures with only base pair constraints are in gray and the modeled RNA structures with all imino proton NMR constraints are in blue. All RMSD values were calculated using all heavy atoms for the 100 lowest energy structures with the corresponding accepted NMR solution structure for (A) human telomerase pseudoknot (PDB ID: 2K96), (B) murine leukemia virus pseudoknot (PDB ID: 2LC8) ignoring the unconstrained loop, (C) mouse mammary tumor virus pseudoknot (PDB ID: 1KAJ), (D) HIV-2 TAR hairpin kissing dimer (PDB ID: 1KIS), (E) guanosine binding site of Group I intron from *Tetrahymena thermophile* (PDB ID: 1K2G), (F) *Aquifex aeolicus* tmRNA pseudoknot PK1 (PDB ID: 2G1W), (G) *Bacillus subtilis* $PreQ_1$ riboswitch class I aptamer (PDB ID: 2L1V), (H) pea enation mosaic virus P1-P2 pseudoknot (PDB ID: 2RP0), (I) sugarcane yellow leaf virus mRNA pseudoknot (PDB ID: 1YG3), (J) *Kluyveromyces lactis* telomerase RNA pseudoknot (PDB ID: 2M8K), (K) *Neurospora* Varkud satellite ribozyme stem I-V kissing-loop interaction (PDB ID: 2MI0), (L) *Streptococcus*

*pneumonia* $PreQ_1$ class II riboswitch (PDB ID: 4ENC), (M) *Neurospora* VS ribozyme II-III-VI three-way junction (PDB ID: 2MIY), (N) *Neurospora* VS ribozyme III-IV-V three-way junction (PDB ID: 2MTJ), (O) *Bacillus cereus* fluoride riboswitch aptamer (PDB ID: 2N3R).

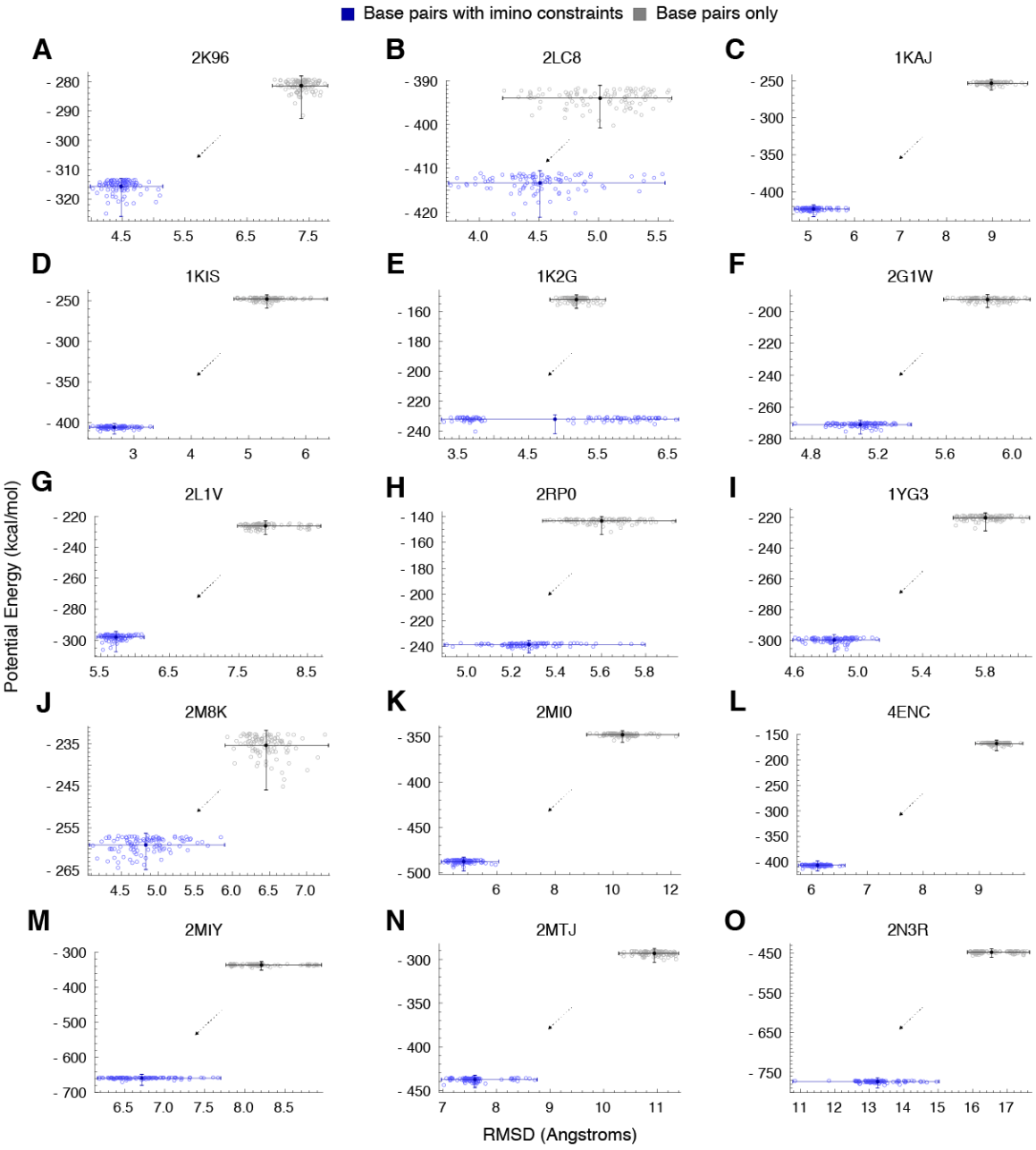


**Figure S6.** RMSD versus potential energy plots for predicted structures reveal an improvement in the RMSD of the ensemble of iFoldNMR structures in blue relative to modeled RNA structures with only base pair constraints in gray. All energy values were calculated by the Medusa force field and RMSD values were calculated using all heavy atoms for the 100 lowest energy structures with the corresponding accepted structure for (A) human telomerase pseudoknot (PDB ID: 2K96), (B) murine leukemia virus pseudoknot (2LC8) ignoring the unconstrained loop, (C) mouse

mammary tumor virus pseudoknot (1KAJ), (D) HIV-2 TAR hairpin kissing dimer (1KIS), (E) guanosine binding site of Group I intron from *Tetrahymena thermophile* (1K2G), (F) *Aquifex aeolicus* tmRNA pseudoknot PK1 (2G1W), (G) *Bacillus subtilis* $PreQ_1$ riboswitch class I aptamer (2L1V), (H) pea enation mosaic virus P1-P2 pseudoknot (2RP0), (I) sugarcane yellow leaf virus mRNA pseudoknot (1YG3), (J) *Kluyveromyces lactis* telomerase RNA pseudoknot (2M8K), (K) *Neurospora* Varkud satellite ribozyme stem I-V kissing-loop interaction (2MI0), (L) *Streptococcus pneumonia* $PreQ_1$ class II riboswitch (4ENC), (M) *Neurospora* VS ribozyme II-III-VI three-way junction (2MIY), (N) *Neurospora* VS ribozyme III-IV-V three-way junction (2MTJ), (O) *Bacillus cereus* fluoride riboswitch aptamer (2N3R).

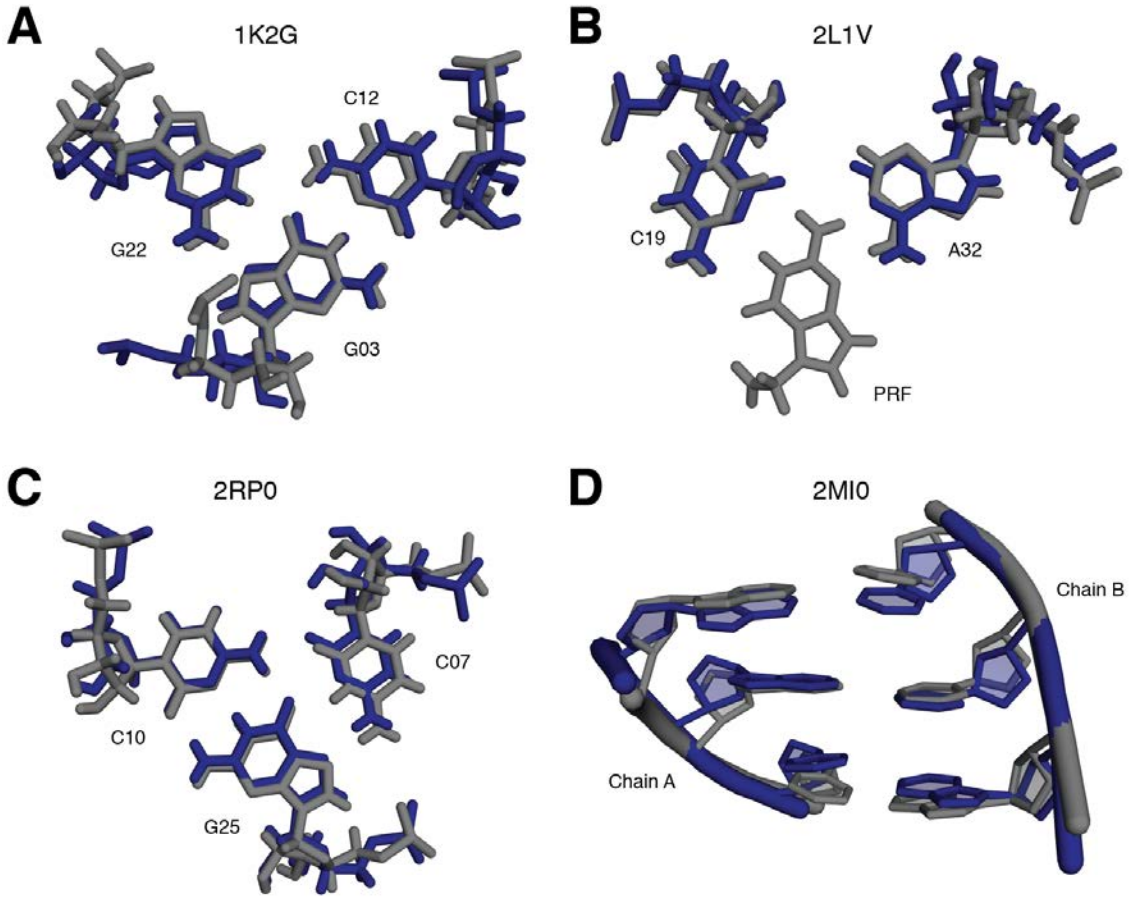


**Figure S7.** Non-canonical RNA features predicted by iFoldNMR. (A) GCG base triple in the guanosine binding site of Group I intron from *T. thermophile* (PDB ID: 1K2G). (B) Interaction with small ligand, PRF (7-deaza-7-aminomethyl-guanine), in *B. subtilis* preQ1 riboswitch class I aptamer (PDB ID: 2L1V). (C) Base triple with protonated cytosine C07 in pea enation mosaic virus P1-P2 pseudoknot (PDB ID: 2RP0). (D) Base pairs involved the *Neurospora* Varkud satellite ribozyme stem I-V kissing-loop interaction (PDB ID: 2MI0).

**Table S1.** Detailed RMSD values for RNA systems. Backbone only and all-atom RMSD values are presented for each RNA system. Columns labeled 'with imino' include values calculated using structural models produced from DMD simulations with imino-proton constraints. Columns labeled 'SS only' include values calculated using structure models produced from DMD simulations with only secondary structure information. No loop or bulge RMSD values excluded any nucleotide that lack base pairing and only sequential NMR constraints are reported. The RMSD of the cluster reports the largest RMSD value between the lowest energy structure and any other member of the cluster reported in the Protein Data Bank. All RMSD values reported are in Å. Interaction Network Fidelity (INF) values (scale 0.00 to 1.00), calculated using equations from Parisien *et al.* **(16)**, are a classification of the base stacking and base pairing between a model and a reference structure. Any entry with "n/a" in the "RMSD of cluster" column means that only one NMR structure has been deposited in the Protein Data Bank, and therefore, we were unable to compare the ensemble. For 4ENC, any entry with "n/a" means that we were unable to make calculations about the INF and loop regions due to sequence differences between the crystal structure and our NMR-derived structural model.

| PDB ID | Backbone RMSD (P atom) | All-atom RMSD | | INF (all atoms) | | INF (Watson-Crick) | | No loop or bulges (# of residues excluded) | RMSD of cluster |
|---|---|---|---|---|---|---|---|---|---|
| | | with imino | SS only | with imino | SS only | with imino | SS only | | |
| 2K96 | 4.200 | 4.245 | 7.705 | 0.775 | 0.744 | 1.00 | 1.00 | 3.854 (2) | 1.437 |
| 2LC8 | 3.796 | 4.287 | 5.124 | 0.755 | 0.802 | 1.00 | 1.00 | 3.383 (18) | 1.480 |
| 1KAJ | 5.366 | 5.139 | 9.475 | 0.776 | 0.644 | 1.00 | 1.00 | 2.905 (11) | n/a |
| 1KIS | 2.532 | 2.394 | 5.554 | 0.803 | 0.688 | 1.00 | 0.935 | 2.394 (0) | n/a |
| 1K2G | 3.079 | 3.734 | 5.355 | 0.740 | 0.685 | 1.00 | 1.00 | 3.757 (1) | n/a |
| 2G1W | 4.261 | 5.000 | 6.075 | 0.591 | 0.700 | 1.00 | 1.00 | 4.572 (3) | 1.545 |
| 2L1V | 6.643 | 5.534 | 8.431 | 0.487 | 0.515 | 0.943 | 0.943 | 4.914 (1) | 1.239 |
| 2RP0 | 4.572 | 5.275 | 5.495 | 0.657 | 0.613 | 1.00 | 0.943 | 4.587 (3) | 2.150 |
| 1YG3 | 3.431 | 4.846 | 5.575 | 0.709 | 0.712 | 1.00 | 1.00 | 3.402 (3) | 2.993 |
| 2M8K | 4.080 | 4.458 | 11.553 | 0.814 | 0.754 | 1.00 | 0.97 | 4.471 (2) | 1.524 |
| 2MI0 | 4.440 | 4.508 | 10.294 | 0.711 | 0.725 | 1.00 | 1.00 | 4.500 (2) | 3.916 |
| 2MIY | 6.940 | 6.472 | 9.586 | 0.824 | 0.760 | 0.973 | 0.944 | 6.031 (9) | 3.725 |
| 2N3R | 15.919 | 13.440 | 16.930 | 0.567 | 0.644 | 1.00 | 1.00 | 13.456 (2) | 5.413 |
| 2MTJ | 7.471 | 7.601 | 11.293 | 0.688 | 0.592 | 1.00 | 0.968 | 6.558 (3) | 7.822 |
| 4ENC | 5.281 | 5.836 | 9.258 | n/a | n/a | n/a | n/a | n/a | n/a |

**Table S2.** Breakdown of imino NMR constraints. Summary of the total number of imino constraints for each RNA system and the subset that corresponds to base pairing.

| PDB ID | Total number of imino constraints | Number of constraints involved in base pairing | Length (nt) |
|---|---|---|---|
| 2K96 | 74 | 59 | 47 |
| 2LC8 | 64 | 64 | 56 |
| 1KAJ | 43 | 36 | 32 |
| 1KIS | 64 | 53 | 16+16 |
| 1K2G | 36 | 22 | 22 |
| 2G1W | 29 | 28 | 22 |
| 2L1V | 45 | 25 | 36 |
| 2RP0 | 32 | 30 | 27 |
| 1YG3 | 47 | 30 | 28 |
| 2M8K | 84 | 63 | 48 |
| 2MI0 | 91 | 77 | 22+21 |
| 2MIY | 100 | 67 | 59 |
| 2N3R | 126 | 101 | 62 |
| 2MTJ | 96 | 73 | 47 |
| 4ENC | 73 | 56 | 47 |

**Table S3.** Summary of RMSD results from different structure modeling programs. Results are listed in Angstroms (Å) for RMSD values. Table cells in the FARFAR column with a value of "n/a" represent RNAs that exceed the 32nt limit for modeling structures in FARFAR. Table cells in the 3dRNA-2.0 and RNAComposer columns with a value of "n/a" represent RNAs with multiple chains, which are not supported by these modeling servers.

| PDB ID | Length (nt) | iFoldNMR | FARFAR | 3dRNA-2.0 | RNAComposer |
|---|---|---|---|---|---|
| 2K96 | 47 | 4.25 | n/a | 11.90 | 9.11 |
| 2LC8 | 56 | 4.29 | n/a | 29.40 | 13.47 |
| 1KAJ | 32 | 5.13 | 6.22 | 13.98 | 8.81 |
| 1KIS | 16+16 | 2.39 | 2.90 | n/a | n/a |
| 1K2G | 22 | 3.73 | 3.00 | 9.22 | 2.45 |
| 2G1W | 22 | 5.00 | 4.34 | 16.92 | 7.79 |
| 2L1V | 36 | 5.53 | n/a | 19.24 | 11.38 |
| 2RP0 | 27 | 5.28 | 3.18 | 20.41 | 8.90 |
| 1YG3 | 28 | 4.85 | 4.72 | 13.44 | 4.95 |
| 2M8K | 48 | 4.46 | n/a | 14.93 | 8.79 |
| 2MI0 | 22+21 | 4.51 | n/a | n/a | n/a |
| 2MIY | 59 | 6.47 | n/a | 21.51 | 13.02 |
| 2N3R | 62 | 13.44 | n/a | 6.27 | 17.85 |
| 2MTJ | 47 | 7.60 | n/a | 7.09 | 8.29 |
| 4ENC | 47 | 5.84 | n/a | 15.55 | 0.76* |

*Note: the 4ENC structure was used as a fragment for RNAComposer, which results in an extremely low RMSD between the experimental and predicted structures.

**Supplemental References**


1. Kitamura,A., Muto,Y., Watanabe,S., Kim,I., Ito,T., Nishiya,Y., Sakamoto,K., Ohtsuki,T., Kawai,G., Watanabe,K., *et al.* (2002) Solution structure of an RNA fragment with the P7/P9.0 region and the 3'-terminal guanosine of the tetrahymena group I intron. *RNA*, **8**, 440–451.

2. Kang,H., Hines,J.V. and Tinoco,I. (1996) Conformation of a non-frameshifting RNA pseudoknot from mouse mammary tumor virus. *J. Mol. Biol.*, **259**, 135–147.

3. Nonin-Lecomte,S., Felden,B. and Dardel,F. (2006) NMR structure of the Aquifex aeolicus tmRNA pseudoknot PK1: new insights into the recoding event of the ribosomal trans-translation. *Nucleic Acids Res.*, **34**, 1847–1853.

4. Chang,K.Y. and Tinoco,I. (1997) The structure of an RNA 'kissing' hairpin complex of the HIV TAR hairpin loop and its complement. *J. Mol. Biol.*, **269**, 52–66.

5. Houck-Loomis,B., Durney,M.A., Salguero,C., Shankar,N., Nagle,J.M., Goff,S.P. and D'Souza,V.M. (2011) An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature*, **480**, 561–564.

6. Giedroc,D.P. and Cornish,P.V. (2009) Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res.*, **139**, 193–208.

7. Cornish,P.V., Hennig,M. and Giedroc,D.P. (2005) A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated -1 ribosomal frameshifting. *Proceedings of the National Academy of Sciences*, **102**, 12694–12699.

8. Kang,M., Peterson,R. and Feigon,J. (2009) Structural Insights into riboswitch control of the biosynthesis of queuosine, a modified nucleotide found in the anticodon of tRNA. *Mol. Cell*, **33**, 784–790.

9. Cash,D.D., Cohen-Zontag,O., Kim,N.-K., Shefer,K., Brown,Y., Ulyanov,N.B., Tzfati,Y. and Feigon,J. (2013) Pyrimidine motif triple helix in the Kluyveromyces lactis telomerase RNA pseudoknot is essential for function in vivo. **110**, 10970–10975.

10. Kang,M., Eichhorn,C.D. and Feigon,J. (2014) Structural determinants for ligand capture by a class II preQ1 riboswitch. **111**, E663–71.

11. Bouchard,P. and Legault,P. (2014) Structural insights into substrate recognition by the Neurospora Varkud satellite ribozyme: importance of U-turns at the kissing-loop junction. *Biochemistry*, **53**, 258–269.

12. Bonneau,E. and Legault,P. (2014) Nuclear magnetic resonance structure of the III-IV-V three-way junction from the Varkud satellite ribozyme and identification of magnesium-binding sites using paramagnetic relaxation enhancement. *Biochemistry*, **53**, 6264–6275.

13. Bonneau,E., Girard,N., Lemieux,S. and Legault,P. (2015) The NMR structure of the II-III-VI three-way junction from the Neurospora VS ribozyme reveals a critical tertiary interaction and provides new insights into the global ribozyme structure. *RNA*, 10.1261/rna.052076.115.

14. Ding,F., Sharma,S., Chalasani,P., Demidov,V.V., Broude,N.E. and Dokholyan,N.V. (2008) Ab initio RNA folding by discrete molecular dynamics: from structure prediction to folding mechanisms. *RNA*, **14**, 1164–1173.

15. Kim,N.-K., Zhang,Q., Zhou,J., Theimer,C.A., Peterson,R.D. and Feigon,J. (2008) Solution structure and dynamics of the wild-type pseudoknot of human telomerase RNA. *J. Mol. Biol.*, **384**, 1249–1261.

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