**Structure, Volume *24***

## Supplemental Information

# Improving NMR Structures of RNA

**Guillermo A. Bermejo, G. Marius Clore, and Charles D. Schwieters**

**SUPPLEMENTAL INFORMATION**

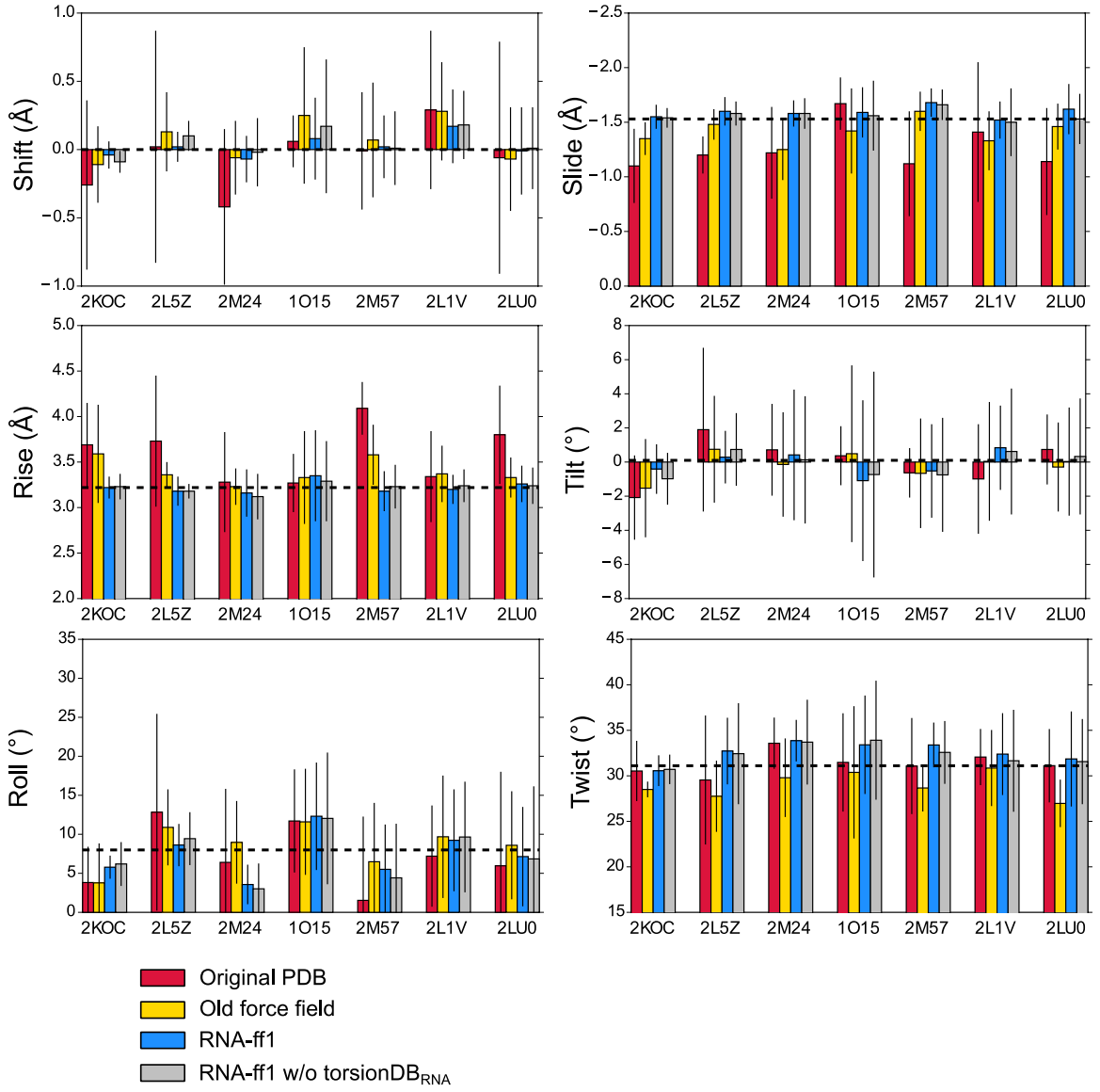


**Figure S1. Related to Figure 4. Base Pair Step Parameters in Helical Stems of the Original PDB and Xplor-NIH NMR Models**

Statistics on original PDB models (red) and Xplor-NIH structures calculated with the old force field (yellow), RNA-ff1 (blue), and RNA-ff1 without the torsionDB$_{RNA}$ potential (grey) are grouped by the PDB code of the corresponding original NMR bundle (*y*-axes). Each statistic represents the average over the structure bundle (standard deviation indicated as error bar). A dashed line indicates the A-form parameter value observed in high-resolution X-ray structures (Olson et al., 2001). For the definition of the base pair step parameters see Figure 1 in Lu and Olson, 2003. The slide and rise plots appear in Figure 4.

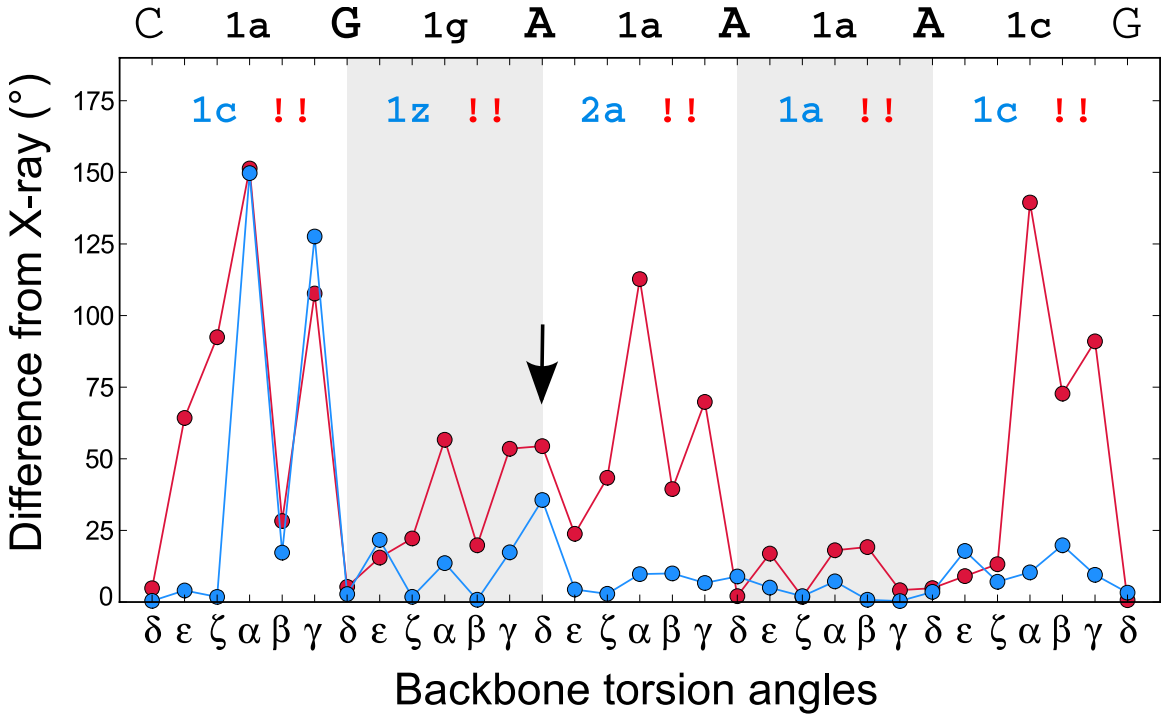


**Figure S2. Related to Figure 7. Improvement of the Backbone Conformation of the GAAA Tetraloop in PDB Structure 2LU0**

The sequence of the tetraloop, including two flanking residues, is indicated on top. The conformation assigned by the program Suitename to each suite between the corresponding bases is indicated for the GAAA tetraloop of PDB X-ray structure 1HQ1 (black), the model 1 of the original NMR bundle 2LU0 (red), and the structure with the lowest experimental energy calculated with the RNA-ff1 force field (blue). "**!!**" denotes an outlier suite; all other labels represent known rotamers (Richardson et al., 2008). For each backbone torsion angle, the plot shows the difference from the X-ray structure for the 2LU0 structure (red) and the RNA-ff1 structure (blue). Background shading delineates the different suites.

The first suite of the RNA-ff1 structure has **1c** conformation, only subtly different from the **1a** of the reference X-ray model (see main text for details). The RNA-ff1 structure also differs from the X-ray reference in the conformation of the second and third suites, which share a common δ (indicated with an arrow in the figure), forced to adopt a C2`-endo value both directly and indirectly (via $\nu_1$ and $\nu_2$) by torsion angle restraints (Donghi et al., 2013). The C2`-endo sugar ring pucker conformation is unusual for GAAA tetraloops (Richardson et al., 2008), consistent with this being the δ that shows the largest difference with the X-ray reference; changing its value to one associated with the C3`-endo pucker would render the conformation of the involved suites identical to that of the reference X-ray structure (i.e., **1z** → **1g** and **2a** → **1a**).

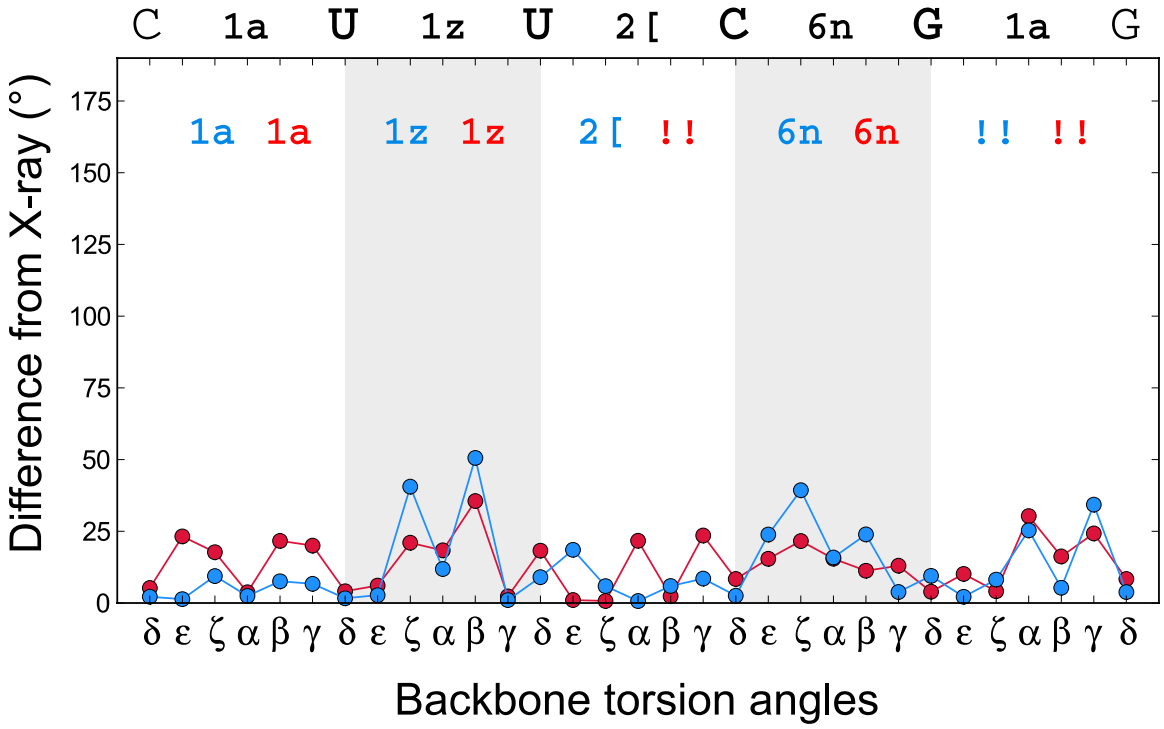


**Figure S3. Related to Figure 7. Improvement of the Backbone Conformation of the UUCG Tetraloop in PDB Structure 2KOC**

The sequence of the tetraloop, including two flanking residues, is indicated on top. The conformation assigned by the program Suitename to each suite between the corresponding bases is indicated for the TL1 loop of PDB X-ray structure 1F7Y (black), the model 1 of the original NMR bundle 2KOC (red), and the structure with the lowest experimental energy calculated with the RNA-ff1 force field (blue). "**!!**" denotes an outlier suite; all other labels represent known rotamers (Richardson et al., 2008). For each backbone torsion angle, the plot shows the difference from the X-ray structure for the 2KOC structure (red) and the RNA-ff1 structure (blue). Background shading delineates the different suites.

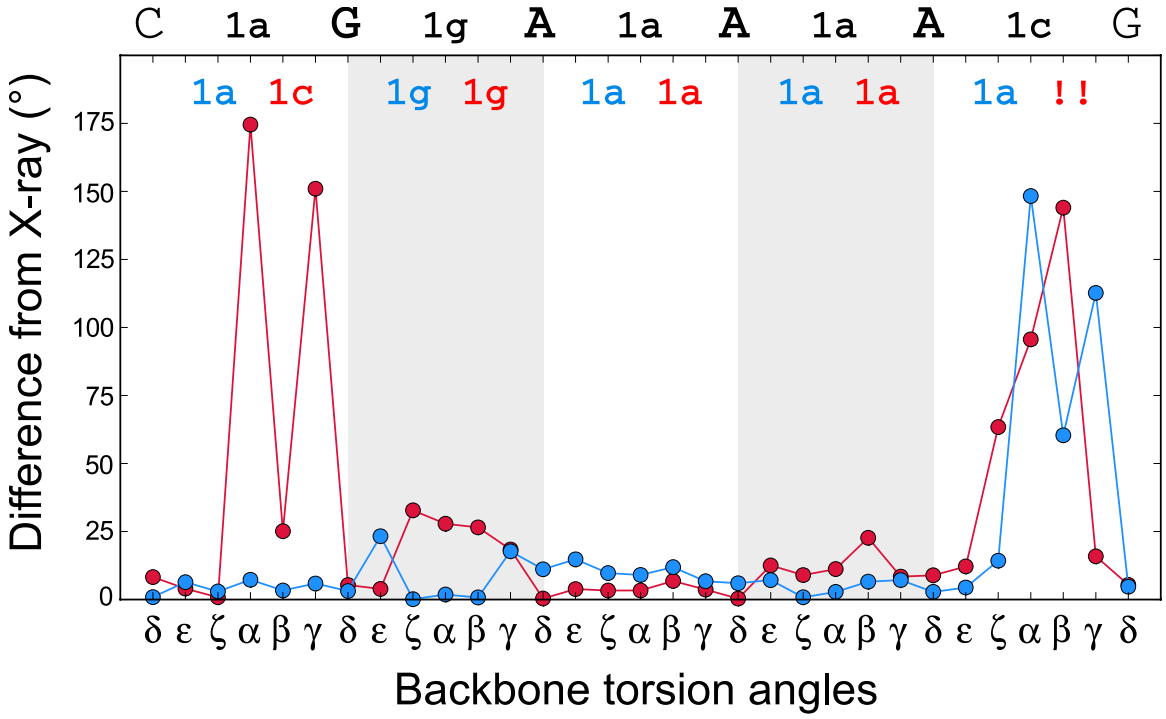


**Figure S4. Related to Figure 7. Improvement of the Backbone Conformation of the GAAA Tetraloop in PDB Structure 2L5Z**

The sequence of the tetraloop, including two flanking residues, is indicated on top. The conformation assigned by the program Suitename to each suite between the corresponding bases is indicated for the GAAA tetraloop of PDB X-ray structure 1HQ1 (black), the model 1 of the original NMR bundle 2L5Z (red), and the structure with the lowest experimental energy calculated with the RNA-ff1 force field (blue). "**!!**" denotes an outlier suite; all other labels represent known rotamers (Richardson et al., 2008). For each backbone torsion angle, the plot shows the difference from the X-ray structure for the 2L5Z structure (red) and the RNA-ff1 structure (blue). Background shading delineates the different suites.

The last suite of the RNA-ff1 structure has a **1a** conformation, subtly different from the **1c** of the reference X-ray model (see main text for details).

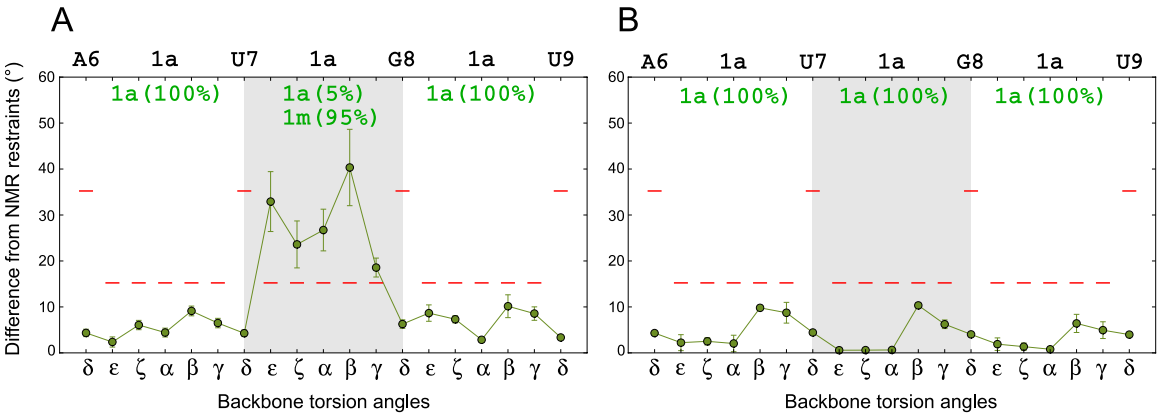


**Figure S5. Related to Figure 3. Example of Bad Convergence with the Old Xplor-NIH Statistical Torsional Potential**

Preliminary calculations with the 2M24 data, using the old statistical torsional potential in Xplor-NIH (Clore and Kuszewski, 2003) suggest that a suite becomes trapped in an incorrect local energy minimum. The sequence of a segment (residues 6−9) that includes such suite is shown on top of each plot, along with the **1a** (i.e., A-form) conformation for the backbone of all suites involved. The **1a** conformation was expected by the authors of the NMR dataset, and reflected in the associated torsion angle restraints (Kruschel et al., 2014). Calculations were performed with a protocol based on the RNA-ff1 force field, identical to the final protocol (see Methods), except that it lacked the extra sugar ring flexibility (i.e., only bond angles involving C4` and O4` were flexible, as opposed to all endocyclic angles). The protocol included either torsionDB$_{RNA}$ or the old statistical torsional potential. Each plot shows the difference between the average torsions in the structure bundle and those of the corresponding restraint targets (dots; standard deviations indicated as error bars). Red dashes denote the threshold beyond which the restraints are considered violated (5° above published tolerances (Kruschel et al., 2014)). The conformational population of each suite in the structure bundle is indicated. Background shading delineates the different suites.

(A) Calculations with the old statistical torsional potential.

(B) Calculations with torsionDB$_{RNA}$.

The U7−G8 suite adopts a **1m** backbone conformation in 95% (19 out of 20) of the structures computed with the old statistical potential, which correlates with consistent restraint violations for its ε−ζ−α−β−γ torsions (A). In contrast, this suite is in the expected **1a** conformation in all structures computed with torsionDB$_{RNA}$, in agreement with the torsion angle restraints (B). These results suggest that, due to its roughness, the old statistical potential frustrates the suite in the **1m** minimum.

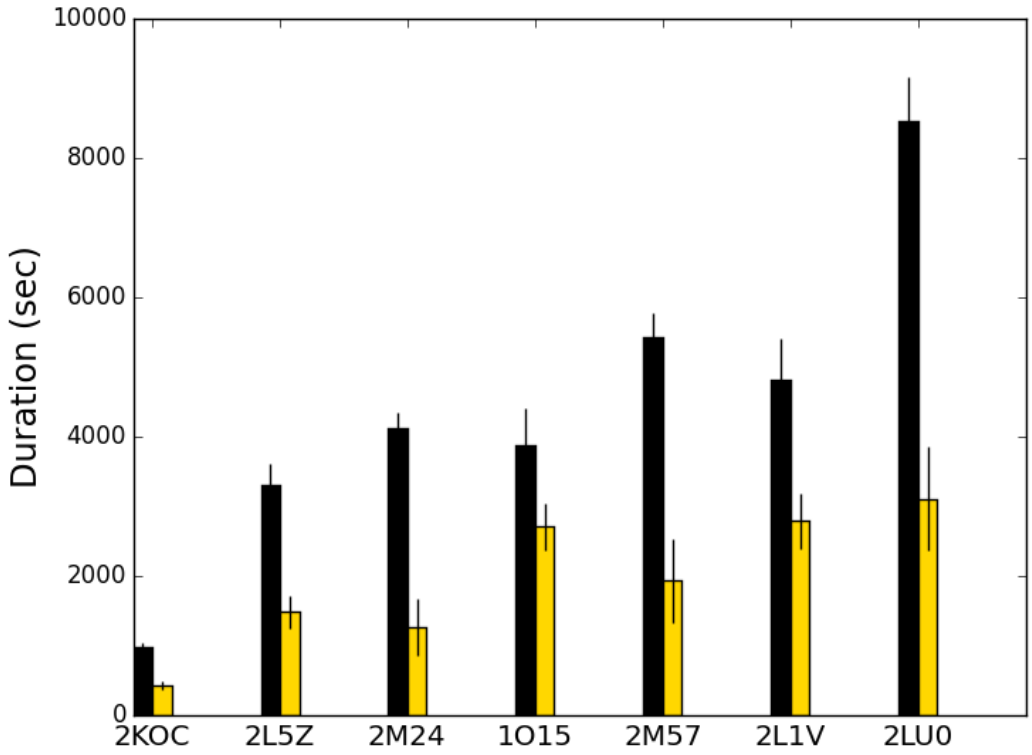


**Figure S6. Related to Figure 3. Average Duration for the Calculation of a Structure in the Refinement Stage of the Structure Calculation Protocol**

Average time taken to calculate a single structure with the refinement protocol of the RNA-ff1 calculation (see Methods) (yellow) and a similar protocol, where torsion$DB_{RNA}$ is replaced by the old statistical torsional potential in Xplor-NIH (Clore and Kuszewski, 2003) (black) (standard deviations indicated as error bars).

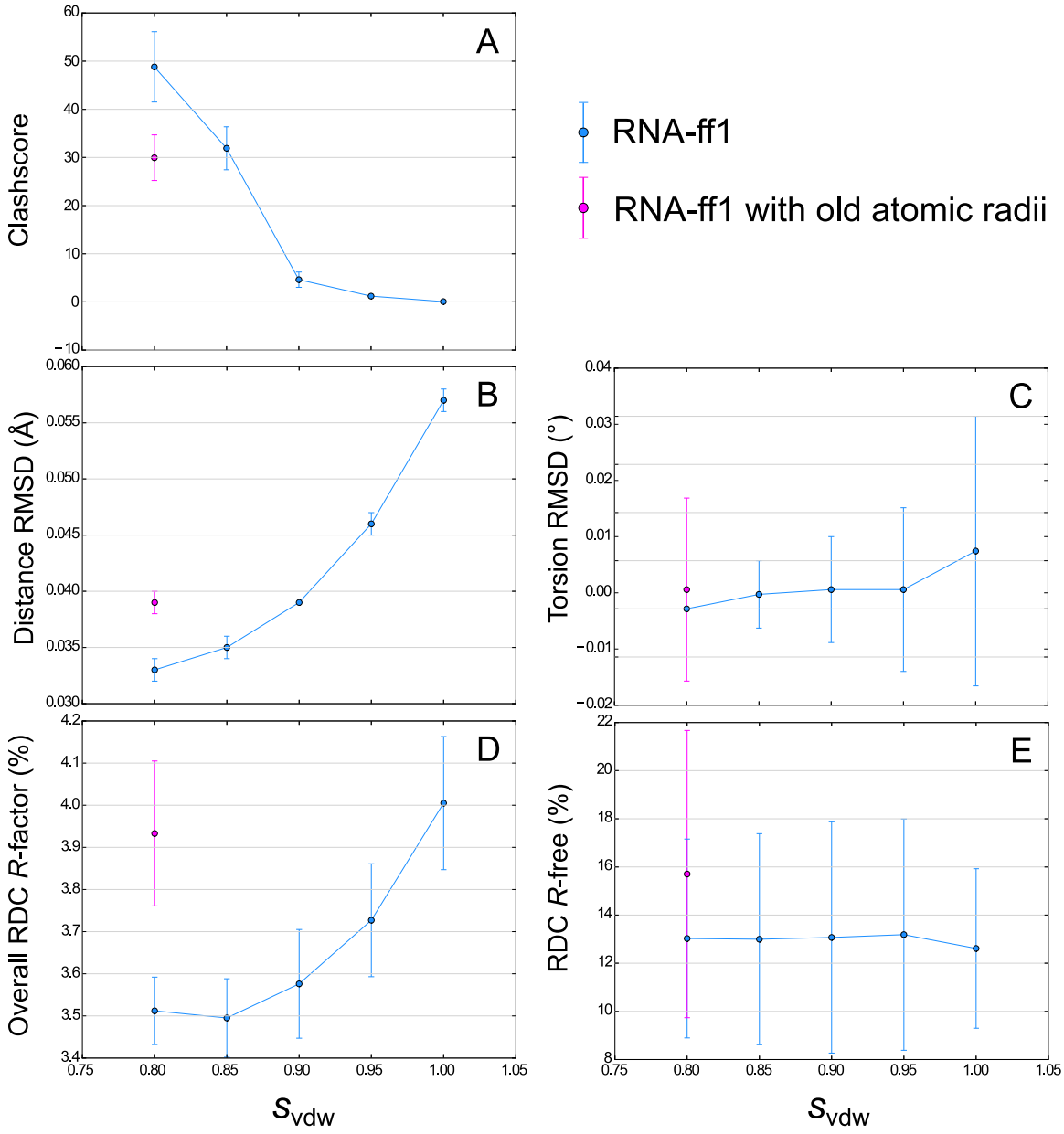


**Figure S7. Related to Methods. Search for the Optimal Atomic Radius Scale Factor, $s_{vdw}$, in the RNA-ff1 Force Field with the 2L5Z NMR Data**

The repulsive-only van der Waals-like energy term employed by the RNA-ff1 force field (Equation 1) requires the optimization of the atomic radius scale factor, $s_{vdw}$. As an example, results are shown for structure calculations performed with the 2L5Z NMR data, using the RNA-ff1 force field with various values of $s_{vdw}$ (blue). For comparison, the result of a similar calculation carried out with RNA-ff1 modified with the nonbonded setup of the old Xplor-NIH force field (i.e., that in topology/parameter files nucleic-1.1.top/nucleic-1.1.par) is shown with its optimal $s_{vdw}$ of 0.80 (magenta). Calculations include the full NMR dataset, except for panel E, associated with the RDC cross-validation scheme. Standard deviations are indicated as error bars.

Relative to the old nonbonded setup, at $s_{vdw}$ = 0.90 RNA-ff1 yields a significantly lower clashscore (A), while achieving similar or better fit to the distance restraints (B), torsion angle restraints (C), the full set of RDCs (D), and the cross-validated set of RDCs (E).

## SUPPLEMENTAL EXPERIMENTAL PROCEDURES

### Structure Calculation Procedure

The overall molecular dynamics/simulated-annealing procedure used with the RNA-ff1 force field (an example of which is provided in the eginput/rna directory within Xplor-NIH's distribution package) consists of two separate protocols for (a) folding an initial extended conformation with satisfied covalent geometry, and (b) refining a folded model, selected on the basis of experimental energy. All degrees of freedom are in torsion angle space, with certain exceptions within sugar rings (see Methods).

The folding protocol comprises the following sequential stages (where $k_n$ represents the force constant or scale factor of energy term $n$):

(i) High-temperature dynamics at 3,500 K for the smaller of 15 ps or 15000 timesteps, subject to restraints for experimental torsion angles ($k_{ta}$ = 10 kcal mol$^{-1}$ rad$^{-2}$), experimental interatomic distances (NOEs and hydrogen bonds; $k_{dist}$ = 2 kcal mol$^{-1}$ Å$^{-2}$), base pair planarity (see Methods), bond lengths ($k_{bond}$ = 1000 kcal mol$^{-1}$ Å$^{-2}$), bond angles ($k_{angle}$ = 200 kcal mol$^{-1}$ rad$^{-2}$), improper dihedrals ($k_{impr}$ = 50 kcal mol$^{-1}$ rad$^{-2}$), the statistical torsional potential torsionDB$_{RNA}$ ($k_{tdb}$ = 0.5), and van der Waals-like repulsions ($k_{vdw}$ = 0.004 kcal mol$^{-1}$ Å$^{-4}$; only C1`–C1` interactions active, with atomic radius scale factor $s_{vdw}$ = 1.2; see Equation 1);

(ii) Dynamics with simulated annealing, where temperature is reduced from the initial value (3,500 K) to 25 K in steps of 12.5 K for the smaller of 0.2 ps/step or 200 timesteps/step, with $k_{ta}$ = 200 kcal mol$^{-1}$ rad$^{-2}$, $k_{bond}$ = 1000 kcal mol$^{-1}$ Å$^{-2}$ (same as above), $s_{vdw}$ = 0.9 (optimal value found by grid search; see Methods and Figure S7), and $k_{dist}$, $k_{angle}$, $k_{impr}$, $k_{tdb}$, and $k_{vdw}$ geometrically increased from the initial values to 30 kcal mol$^{-1}$ Å$^{-2}$, 500 kcal mol$^{-1}$ rad$^{-2}$, 500 kcal mol$^{-1}$ rad$^{-2}$, 4, and 4 kcal mol$^{-1}$ Å$^{-4}$, respectively (all van der Waals repulsions active in this stage, except for atoms three or fewer bonds apart from each other);

(iii) 500 steps of Powell minimization using the final state (force constants, etc.) of the previous stage.

The refinement protocol is the same as that used for folding, with a few exceptions/additions. First, the high-temperature stage is performed at 3,000 K with the addition of RDC restraints ($k_{rdc}$ = 0.05 kcal mol$^{-1}$ Hz$^{-2}$) and the statistical base−base positional potential (Clore and Kuszewski, 2003) ($k_{pos}$ = 0.002). Second, the simulated annealing stage starts at 3,000 K, the temperature reduced to 25 K in steps of 12.5 K, for the smaller of 0.63 ps/step or 630 timesteps/step, with $k_{rdc}$ and $k_{pos}$ geometrically increased to 1 kcal mol$^{-1}$ Hz$^{-2}$ and 0.3, respectively. The molecular alignment tensor is estimated from the initial structure by singular value decomposition (Losonczi et al., 1999), its orientation, magnitude and rhombicity subsequently optimized along with the atomic coordinates.

Structure calculations performed with Xplor-NIH's old force field followed a procedure identical to that used with RNA-ff1 (described above), with a few exceptions. First, the old force field relies on atomic radii different from those in RNA-ff1, therefore requiring a different optimal $s_{vdw}$ of 0.8 (Nilges et al., 1988). Finally, the force field's torsion angle information is supplied by the older statistical potential (Clore and Kuszewski, 2003), implemented with initial and final scale factors of 0.5 and 1, respectively.

### RDC Fit

The agreement of a structure with the RDC data was assessed via singular value decomposition (Losonczi et al., 1999) using the calcTensor helper program provided in the Xplor-NIH distribution package. The full RDC dataset was used to find the optimal alignment tensor, both for the standard and cross-validation schemes. The RDC fit was assessed by the $R$-factor figure of merit, $R_{dip}$, given by (Clore and Garrett, 1999):

$$R_{dip} = 100\,[\langle (D_{obs} - D_{calc})^2 \rangle / (2\langle D_{obs}^2 \rangle)]^{1/2}, \qquad \text{(S1)}$$

where $D_{obs}$ and $D_{calc}$ are observed and calculated RDC values, respectively, and angular brackets denote averaging over all RDCs, normalized relative to a common nuclear pair type (if applicable).

### Software Details

All structure calculations were carried out with Xplor-NIH (Schwieters et al., 2006; Schwieters et al., 2003). A locally run copy of MolProbity (version 4.01a-450) (Chen et al., 2010; Davis et al., 2007) was used throughout. The reported fraction of suite backbone outliers by MolProbity was found to be in terms

of the total number of residues ($n$), instead of the number of suites ($n - 1$); the latter was used here. Suitename version 0.3.070628 (Richardson et al., 2008) was used to classify backbone conformations, and calculate the average suiteness score of non-outlier suites in a given structure, based on the score of individual suites. DSSR version 1.1.8-2014oct09 (Lu et al., 2015) was used to analyze base pair steps in helical stems. Two-dimensional plots (box, bar, and dot plots; Figures 1, 4 and 7) were generated with Matplotlib (Hunter, 2007), and three-dimensional surfaces (Figure 3) with Mayavi (Ramachandran and Varoquaux, 2011).

**SUPPLEMENTAL REFERENCES**

Hunter, J.D. (2007). Matplotlib: A 2D graphics environment. Comput. Sci. Eng. *9*, 90-95.

Losonczi, J.A., Andrec, M., Fischer, M.W., and Prestegard, J.H. (1999). Order matrix analysis of residual dipolar couplings using singular value decomposition. J. Magn. Reson. *138*, 334-342.

Ramachandran, P., and Varoquaux, G. (2011). Mayavi: 3D Visualization of Scientific Data. Comput. Sci. Eng. *13*, 40-50.