Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

SUPPLEMENTARY INFORMATION

Simultaneous NMR characterisation of multiple minima in the free energy landscape of an RNA UUCG tetraloop

Aditi N. Borkar[1], Pramodh Vallurupalli[2], Carlo Camilloni[1],
Lewis E. Kay[2], and Michele Vendruscolo[1]

[1]Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK
[2]Departments of Molecular Genetics, Biochemistry, and Chemistry, University of Toronto, Toronto, Canada M5S 1A8

Residual dipolar coupling measurements

RNA sample preparation. T7 RNA polymerase was used to transcribe the RNA *in vitro* from a partially double stranded DNA template that contained the T7 RNA polymerase promoter and the appropriate DNA sequence and the RNA was purified by standard denaturing electrophoresis[1]. The last two nucleotides of the DNA template were C2' methoxy modified[2]. The two RNA samples used in this study consisted of ~2mM 14mer dissolved in 20 mM potassium phosphate, and 0.4 mM EDTA, pH 6.4, 100% $^{2}H_2O$. Additionally one of the samples contained ~25 mg/ml Pf1 phage[3] to align the RNA. We note that in the comparison with the results of Ref.[4] one should note that in that case RNA was phosphorylated from 5' and obtained synthetically.

NMR Spectroscopy. All the one bond ^{13}C-^{1}H $^{1}J/(^{1}J+RDC)$ couplings were measured at natural abundance ^{13}C on a 600 MHz Varian Inova spectrometer equipped with a cold probe using the IPAP method[5-7]. Data was processed using NMRPipe[8] and visualized using SPARKY[9]. Errors were estimated on the basis of repeat measurements.

Molecular dynamics simulations

Preparation of the starting conformations. A high-resolution NMR structure (PDB 2KOC[4]) of a 14-nucleotide RNA hairpin containing the UUCG tetraloop was used as the starting conformation in the simulations (**Figure 2**). All molecular dynamics simulations reported in this work were performed using GROMACS 4.5[10] with the AMBER99bsc0 force field[11] with the χ parameterization[12]. The first model from the 2KOC PDB file along with 13 neutralising K^+ ions and 100 mM $MgCl_2$ were placed in an octahedral box with sides 1.2 nm away from the initial structure and solvated with TIP3P water molecules[13]. After an energy minimization, first with the steepest descent method and then with the low memory Broyden-Fletcher-Goldfarb-Shanno quasi-Newtonian method[14], the system was simulated for 50 ps with position restraints while the temperature was raised to 200 K. Next, these position restraints were removed and the system was simulated under NVT conditions for a further 50 ps. Finally, the system was heated to the final temperature of 298.15 K under isothermal-isobaric (NPT) conditions. A total of eight NPT runs were performed by using different random seed velocities and monitored until the pairwise root mean square distance (RMSD) correlation of the eight trajectories dropped to zero

$$r_{i,j} = \frac{\sum_{k=1}^{N} \left(X_{i,k} - \bar{X}_i\right)\left(X_{j,k} - \bar{X}_j\right)}{\sqrt{\sum_{k=1}^{N} \left(X_{i,k} - \bar{X}_i\right)^2 \sum_{k=1}^{N} \left(X_{j,k} - \bar{X}_j\right)^2}} \sim 0 \quad (1)$$

where, $i = 1...R$ and $j = i + 1...R$, R being the total number of trajectories (8 in this case). In this way the final conformations of the eight trajectories were significantly different from each other (**Figure 2**). All further simulations were continued using these eight structures as starting points for the different replicas.

Replica-averaged metadynamics (RAM) simulations. In this work, we carried out molecular dynamics simulations using the replica-averaged metadynamics (RAM) method[15,16] (**Figure 2**), which implements at the same time the NMR-derived structural restraints as a correction to the force-field[17] and the enhancement of the sampling due to the metadynamics approach[18,19] in the bias-exchange mode[20]. In the bias-exchange mode, the dynamics of each replica is biased in a direction that changes stochastically with time[20]. The sum of Gaussians is then exploited for reconstructing iteratively an estimator of the free energy[18,19]. This approach is highly effective in forcing the system to escape from local minima and explore a complex free energy landscape. For the UUCG tetraloop, we used six collective variables (CVs) corresponding to specific backbone torsion angles (β, δ, ε and ζ of UL2 and α of CL3) and a network of signature tetraloop hydrogen bonds (**Figure 1**). In the simulations two of the eight replicas were not subjected to a bias.

Metadynamics trajectories were post-processed using METAGUI [21]. The sampled conformations were first clustered into substates and the free energies of each substate were computed by a weighted-histogram procedure after allowing for a suitable equilibration period in the simulation[15,16]. All the conformations from the converged part of the trajectory were extracted to build a conformational ensemble[15,16].

Convergence of the metadynamics simulations. The convergence of the metadynamics simulations was verified by comparing the free energy landscapes for the six different CVs at different stages of the simulation (**Figure S1**).

Generation of the restrained conformational ensemble ('RAM ensemble'). In the RAM simulations, RDCs were back-calculated from the structures using the recently introduced 'ϑ method'[22] (see below) for each replica at each time step, and an additional term was incorporated in the force field to penalize differences between the experimental and simulated values of the RDCs[15,22]. These simulations generated a restrained conformational ensemble (the 'RAM ensemble'). The algorithm for calculating the RDCs and deriving the energy restraint term is implemented in the PLUMED 2 suite[23].

Generation of the unrestrained conformational ensemble ('MD ensemble'). In another set of metadynamics simulations using the setup described above, no restraints were imposed to generate an unrestrained molecular dynamics (MD) ensemble (the 'MD ensemble') of the UUCG tetraloop.

The ϑ method for calculating residual dipolar couplings. The RDC between two nuclear spins can be written as[24]

$$D = D_{\max} \left\langle \left(3\cos^2 \vartheta - 1\right)/2 \right\rangle \qquad (2)$$

where γ_1 and γ_2 are the gyromagnetic ratios of the two spins, r is their distance, ϑ is the angle between the inter-nuclear vector and the external magnetic field, $D_{\max} = -\mu_0 \gamma_1 \gamma_2 h / 8\pi^3 r^3$ is the maximal value of the dipolar coupling for the two nuclear spins, μ_0 is the magnetic constant and h is the Planck constant. The angular brackets describe the thermal averaging over the orientation of the inter-nuclear vector with respect to the external magnetic field. In isotropic solutions the RDCs average to zero because all directions are equivalent. By contrast, if the solution is anisotropic, as in the case of the addition of an alignment medium, the rotational symmetry is broken, and non-zero values of the RDCs may appear[24,25].

Eq. (2) provides the RDC of a given inter-nuclear bond vector as a function of the angle ϑ between the vector and the magnetic field, whose direction is usually taken as that of the z-axis. One can thus use the information about the ϑ angles provided by the RDCs to refine the structures of proteins[22,26,27]. In this approach one asks if there is a structure that satisfies at the same time all the

inter-nuclear vector orientations specified from the ϑ angles with respect to the z-axis. In order to implement this strategy for structural refinement, we first maximized the correlation, ρ, between the calculated, D^{calc}, and the experimental, D^{exp}, RDCs

$$V_\theta = -K_\theta\left[\rho\left(D^{calc},D^{exp}\right) - 1\right] \quad (3)$$

Once a high correlation is obtained it is possible to find the scaling factor for the RDCs as the slope of the line that fits D^{exp} as a function of D^{calc}. Having found the scaling factor, it becomes possible to apply a more stringent restraining potential of the form

$$E_\theta = K_\theta\left[\sum_i^N \left(D_i^{calc} - D_i^{exp}\right)^2\right] \quad (4)$$

where D_i^{calc} is calculated as an average of the RDCs of instantaneous conformations of all the replicas. In the implementation presented in Eqs. (3,4), the ϑ method can be applied to multiple bonds measured in a single alignment medium, although it is possible to extend its use to multiple alignment media[28]. In order to extract the information about dynamics provided by RDCs, we incorporated them as replica-averaged structural restraints in molecular dynamics simulations[22]. To this effect in Eq. (3) we averaged the calculated RDCs over 8 replicas of the RNA molecule.

Construction of the sketch-map. Sketch-map is a dimensionality reduction algorithm that works by preserving the distance connectivity information between a set of high dimension points in a low dimension space[29]. Thus, it can be used to visualize the conformational variability of biomolecules on a free energy landscape[16]. To implement this approach, we first calculated the six backbone torsion angles (α, β, γ, δ, ε, ζ) and the glycosidic χ torsion angle of the tetraloop residues (UL1, UL2, CL3 and GL4) and its closing base pair (CL-1 and GL+1) for the RAM and MD ensembles. Next, we calculated the sketch-map of this 42-dimensional torsion angle hyperspace and used the resulting 2-dimensional projections (which will be referred to as sketch-map CVs). The resulting free energy landscape as a function of the sketch-map CVs represents the projection of all the high-dimensional conformational variability inherent in the RAM or MD ensembles onto a low dimensional surface without biasing the analysis towards any one particular ground state or excited state conformation.

Q factor. The Q factor is a normalised metric for agreement between the experimental RDCs (D^{exp}) and the RDCs calculated from the RAM or MD ensembles (D^{calc})

$$Q = \sqrt{\frac{\sum_{i=1}^{N}(D_i^{exp} - D_i^{calc})^2}{\sum_{i=1}^{N}(D_i^{exp})^2}}$$

Table S1: The 53 CH bond RDCs measured and used as restraints (Set A' in **Figure 3**) and free data (Set A'' in **Figure 3**) in the RAM simulations.

S. No.	Res#	Atom name	Res#	Atom name	RDC value (Hz)	Remarks
1	1	C4'	1	H4'	-5.7	Restrained
2	1	C8	1	H8	11.77	Restrained
3	2	C1'	2	H1'	-1.06	Restrained
4	2	C8	2	H8	4.62	Restrained
5	3	C5	3	H5	10.35	Restrained
6	3	C2'	3	H2'	-11.82	Restrained
7	4	C8	4	H8	10.35	Restrained
8	4	C2	4	H2	2.53	Restrained
9	4	C2'	4	H2'	-5.97	Restrained
10	5	C1'	5	H1'	8.78	Restrained
11	5	C6	5	H6	17.56	Restrained
12	5	C5	5	H5	7.23	Restrained
13	5	C2'	5	H2'	10.01	Restrained
14	6	C1'	6	H1'	8.35	Restrained
15	6	C6	6	H6	6.13	Restrained
16	6	C5	6	H5	15.44	Restrained
17	7	C4'	7	H4'	-11.64	Restrained
18	7	C1'	7	H1'	2.68	Restrained
19	7	C3'	7	H3'	2.77	Restrained
20	8	C4'	8	H4'	3.8	Restrained
21	8	C1'	8	H1'	-20.54	Restrained
22	8	C6	8	H6	-4.86	Restrained
23	8	C5	8	H5	-11.69	Restrained
24	8	C3'	8	H3'	-14.78	Restrained
25	9	C1'	9	H1'	-5.07	Restrained
26	9	C3'	9	H3'	10.2	Restrained
27	10	C1'	10	H1'	-19.61	Restrained
28	10	C2'	10	H2'	5.72	Restrained
29	11	C6	11	H6	12.43	Restrained
30	11	C5	11	H5	10.28	Restrained
31	12	C1'	12	H1'	0.92	Restrained
32	12	C8	12	H8	12.21	Restrained
33	13	C1'	13	H1'	9.66	Restrained
34	13	C6	13	H6	3.07	Restrained
35	13	C5	13	H5	14.68	Restrained
36	13	C2'	13	H2'	-16.32	Restrained
37	14	C4'	14	H4'	14.07	Restrained
38	14	C1'	14	H1'	12.23	Restrained
39	14	C6	14	H6	9.59	Restrained
40	1	C1'	1	H1'	2.98	Free data
41	1	C2'	1	H2'	-2.66	Free data
42	2	C4'	2	H4'	-0.46	Free data
43	4	C1'	4	H1'	7.22	Free data
44	6	C2'	6	H2'	12.82	Free data
45	7	C6	7	H6	7.3	Free data
46	7	C5	7	H5	4.46	Free data

47	9	C2'	9	H2'	2.64	Free data	51	6	C4'	6	H4'	15.09	Free data
48	10	C8	10	H8	14.76	Free data	52	10	C3'	10	H3'	-11.37	Free data
49	11	C3'	11	H3'	-4.7	Free data	53	7	C2'	7	H2'	-1.29	Free data
50	14	C5	14	H5	1.06	Free data							

Table S2: The 30 RDCs sourced from BMRB database (BMRB entry number 5705) and used as free data (Set B in **Figure 3**) for validation of the RAM and MD ensembles.

S. No.	Res#	Atom name	Res#	Atom name	RDC value (Hz)	Remarks
1	3	C6	3	H6	3.54	Free data
2	5	C6	5	H6	10.99	Free data
3	6	C6	6	H6	4	Free data
4	7	C6	7	H6	3.47	Free data
5	8	C6	8	H6	-3.66	Free data
6	11	C6	11	H6	9.09	Free data
7	13	C6	13	H6	2.83	Free data
8	14	C6	14	H6	6.43	Free data
9	1	C8	1	H8	7.29	Free data
10	2	C8	2	H8	1.63	Free data
11	4	C8	4	H8	6.02	Free data
12	9	C8	9	H8	0.34	Free data
13	10	C8	10	H8	11.1	Free data
14	12	C8	12	H8	5.36	Free data
15	1	C1'	1	H1'	-2.39	Free data
16	2	C1'	2	H1'	-0.32	Free data
17	3	C1'	3	H1'	6.36	Free data
18	4	C1'	4	H1'	6.44	Free data
19	5	C1'	5	H1'	4.95	Free data
20	6	C1'	6	H1'	6.89	Free data
21	7	C1'	7	H1'	-1.95	Free data
22	8	C1'	8	H1'	-14.66	Free data
23	9	C1'	9	H1'	-5.14	Free data
24	11	C1'	11	H1'	-0.43	Free data
25	12	C1'	12	H1'	6.77	Free data
26	13	C1'	13	H1'	6.53	Free data
27	14	C1'	14	H1'	7.28	Free data
28	2	N1	2	H1	0.3	Free data
29	12	N1	12	H1	-0.37	Free data
30	11	N3	11	H3	-1.75	Free data

	Set A			Set A"			Set 'B not A'			Set B			Set 'A not B'		
	M	Q	R	M	Q	R	M	Q	R	M	Q	R	M	Q	R
2KOC	3.58	0.34	0.94	4.17	0.50	0.84	1.13	0.31	0.95	2.20	0.36	0.92	4.68	0.46	0.89
RAM	1.17	0.11	1	3.12	0.37	0.92	1.17	0.32	0.95	2.46	0.39	0.92	2.62	0.26	0.97
MD	4.44	0.42	0.94	4.13	0.50	0.92	1.35	0.37	0.93	3.78	0.62	0.83	4.57	0.45	0.91

Table S3. Assessment of the quality of the 2KOC, RAM and MD ensembles analysed in this study. For RDCs comparisons we used the RMSD (M, in Hz), the Q factor and the Pearson's coefficient of correlation (R). The RDC sets are given as column headings and ensembles are denoted as row headings (**Figure 3**). Values in red denote the quality check of the restrained bonds and those in blue demote the refinement of the unrestrained bonds in the 2KOC and RAM ensembles. All values are for weighted-averages of the RDCs calculated by fitting a single alignment tensor to each substate in the RAM and MD ensembles.

Structure	RMSD (in nm)	Population (in %)
R1	0.07	57.0
R2	0.29	10.2
R3	0.28	4.8
R4	0.08	3.8
R5	0.33	2.1
R6	0.32	1.5
U1	0.10	64.2
U2	0.34	9.1
U3	0.38	10.9
U4	0.11	1.8
U5	0.13	2.8
U6	0.16	2.2

Table S4. List of RMSD values (in nm) from the 2KOC structure and populations of the R1-R6 and U1-U6 representative structures shown in **Figure 4**. The populations for each structure have been calculated as a sum of all the microstates lying close together within the respective minimum on the free energy landscape in **Figure 4.**

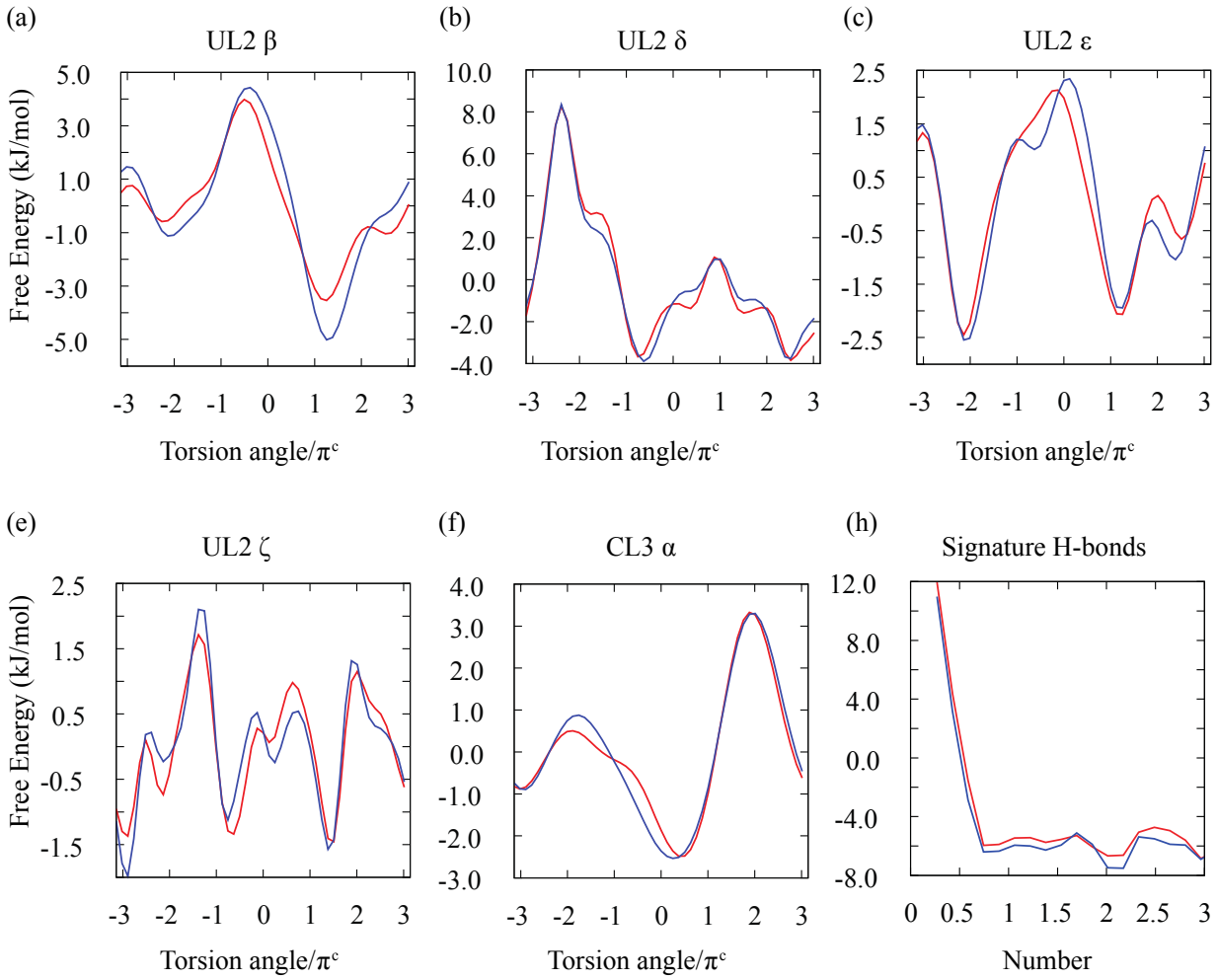

Figure S1. Analysis of the convergence of the RAM simulations. For a converged trajectory, the free energy surfaces for two halves of the simulation after a suitable equilibration time, overlap. For the RAM simulations of the RNA hairpin, after 70 x 8 ns of simulations, most of the CVs converge as seen by the free energy profiles calculated for the simulation run spanning 70 to 100 ns (red curve) and 100 to 130 ns (blue curve). These profiles were constructed in METAGUI by weighted histogram procedure [21]. Thus, the part of simulation spanning a total of 50 x 8ns (from 70 to 130 ns) after this equilibration time was used to construct the RAM ensemble.

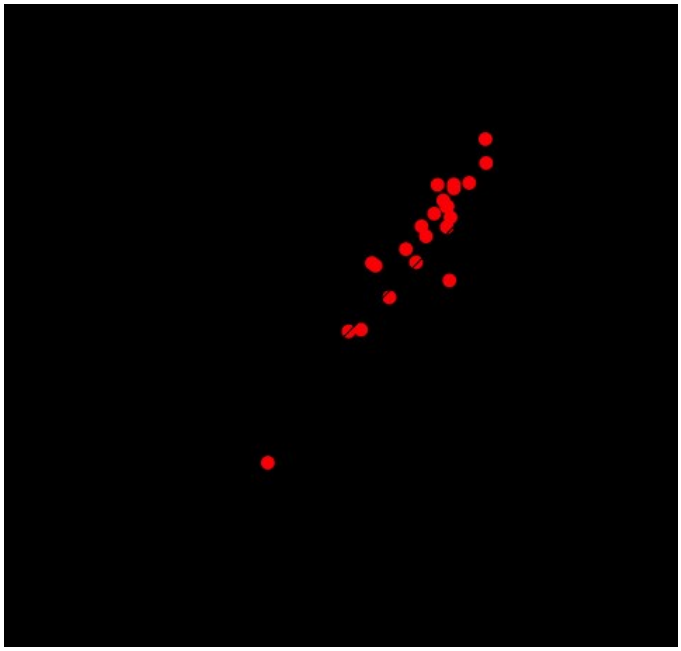

Figure S2. Correlation between the RDCs measured in Pf1 phage alignment in this work (**Figure 3a**, set A) for the 14-nucleotide RNA hairpin containing the UUCG tetraloop and those obtained from the BMRB entry number 5705[4] (**Figure 3a**, set B). Only the 23 RDCs present in both the A and B sets are shown (**Figure 3a**).

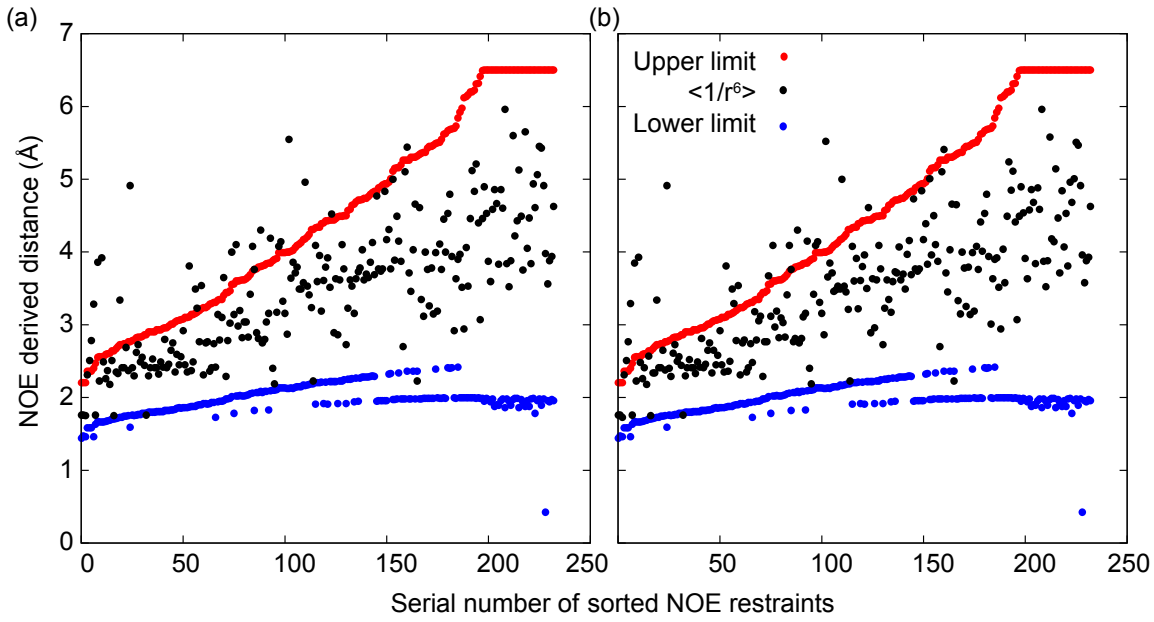

Figure S3. Validation of the RAM and MD ensembles using NOE-derived interatomic distances. The average $1/\mathbf{r}^6$ distances were calculated as ensemble averages for the RAM (a) and MD (b) ensembles. The upper (red) and lower (blue) limits derived from the experimental NOEs are shown for comparison. Only 2 violations > 1 Å are present in each of the two ensembles indicating that both the RAM and MD simulations produced ensembles in agreement with NMR observables not used as restraints in the simulations.

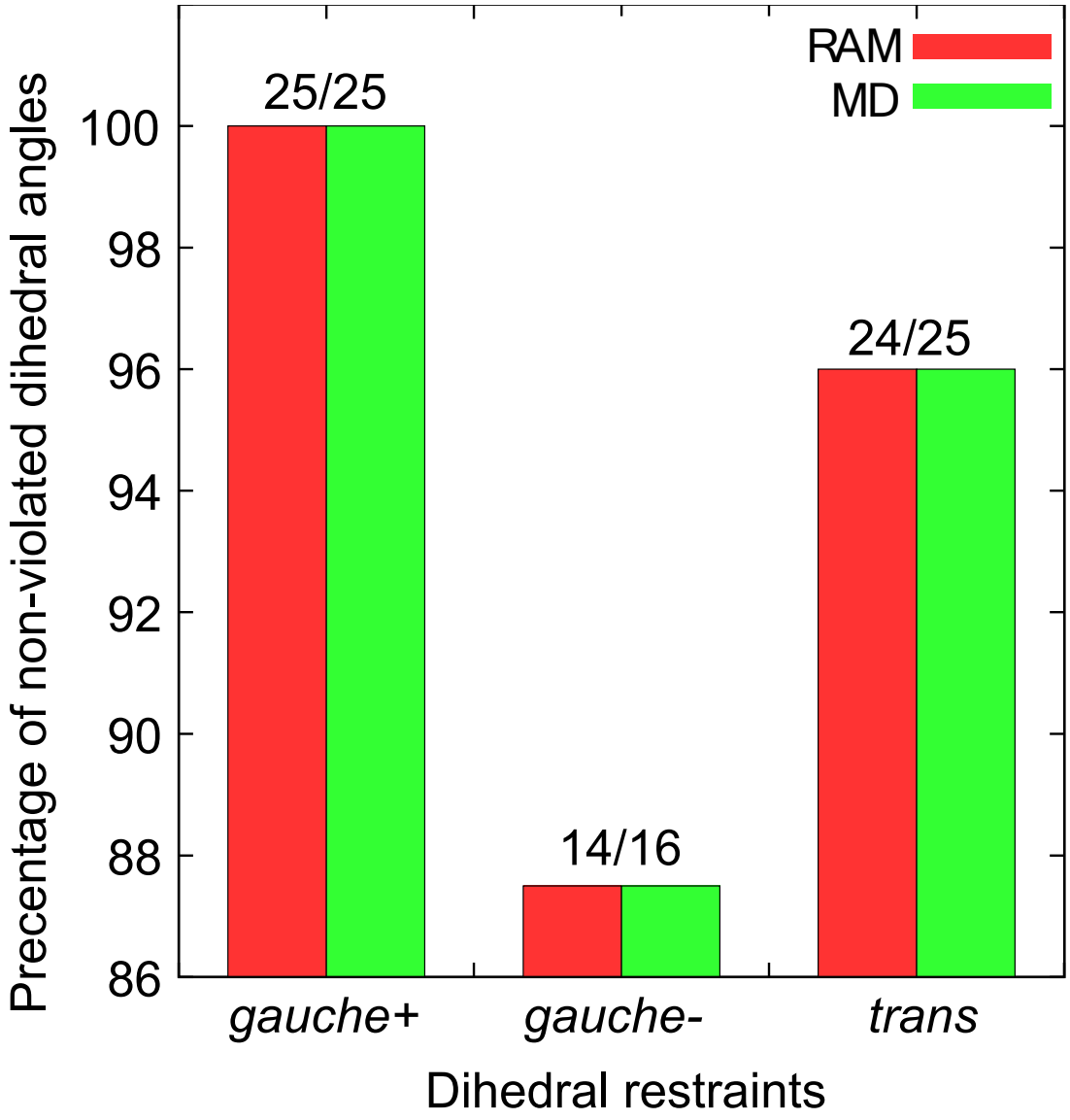

Figure S4. Validation of the RAM and MD ensembles using J-coupling-derived dihedral angles. The dihedral angles α, β, γ, δ, ε, ζ and χ were calculated as ensemble averages from the RAM and MD ensembles. The calculated values are grouped together based upon whether they had a *gauche* (+60°), *trans* (180°) or *gauche*- (-60°) value as determined from the corresponding experimental J-couplings. The two violations observed are for ζ of UL2 and GL4. Thus, both the RAM and MD simulations produced ensembles in agreement with NMR observables not used as structural restraints in the simulations.

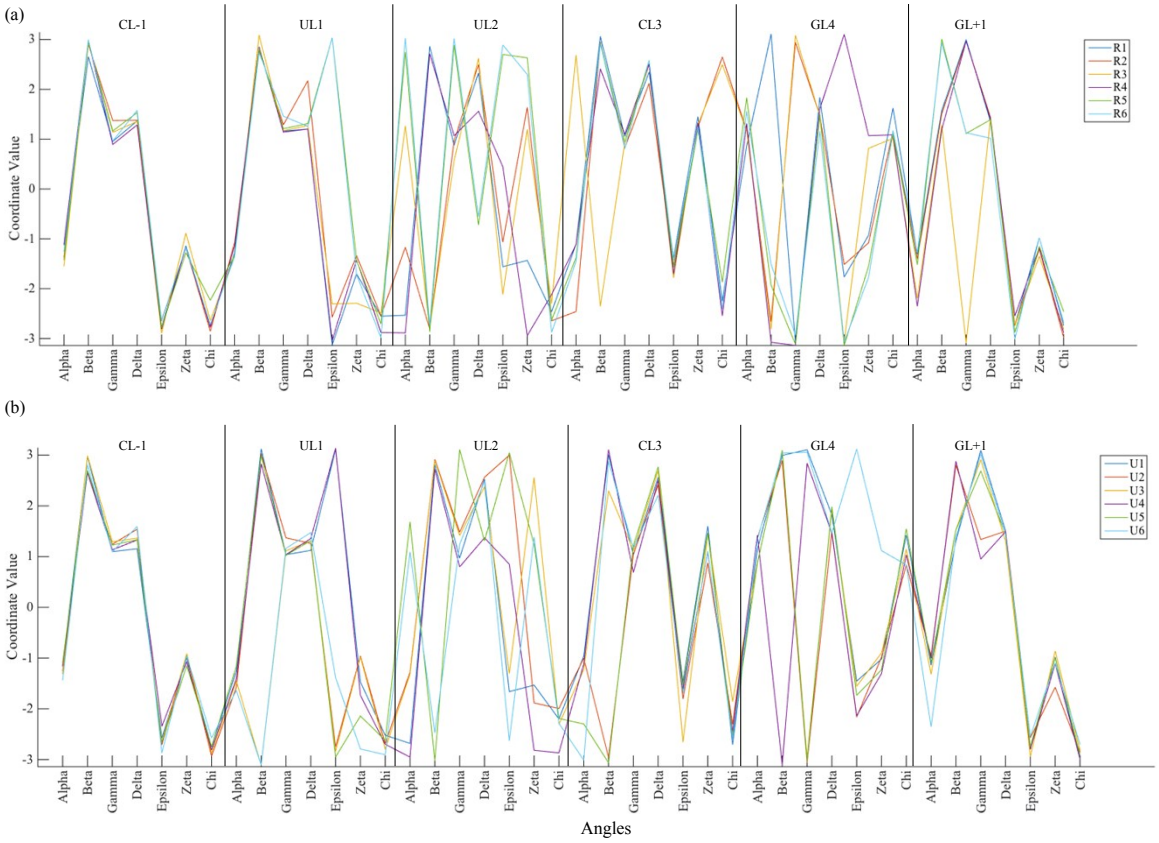

Figure S5. Parallel coordinate plot for comparison of the (a) R1-R6 and (b) U1-U6 structures of Figure 4. The plot illustrates dihedral angle values for the six backbone torsion angles (α, β, γ, δ, ε, ζ) and the glycosidic χ torsion angle of the tetraloop residues (UL1, UL2, CL3 and GL4) and its closing base pair (CL-1 and GL+1) in each of the R1-R6 and U1-U6 structures. The figure shows that the major difference between these structures of the UUCG tetraloop mainly arise due to the conformational flexibility of dihedral angles of UL2 and their immediate neighbours in UL1 and CL3. Additionally, the GL4 and GL+1 in the RAM structures appear to be much more dynamic than those in the MD ensemble. This plot also suggests much more conformational heterogeneity for the UUCG tetraloop than is apparent in the illustrations of Figure 4. Thus, different combinations of the dihedral angles in RNA can possibly give rise to similar relative orientations of its residues.

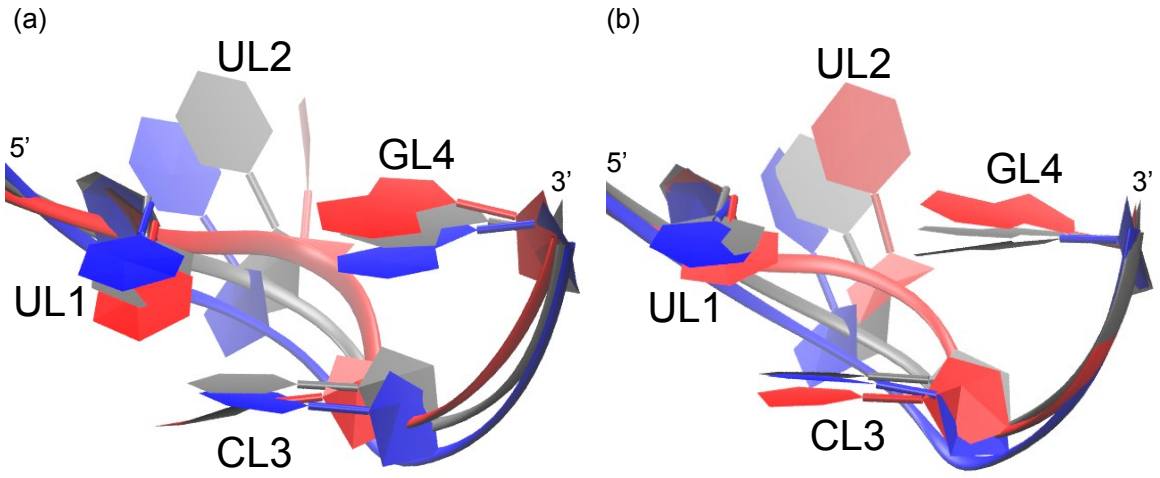

Figure S6. Superposition of (a) R1 (blue) and R4 (red) conformations and (b) U1 (blue) and U4 (red) conformations with the 2KOC structure (gray) of the UUCG tetraloop.

Supplementary References

(1) Puglisi, J. D.; Wyatt, J. R. *Methods Enzymol.* **1994**, *261*, 323.

(2) Kao, C.; Rüdisser, S.; Zheng, M. *Methods* **2001**, *23*, 201.

(3) Hansen, M. R.; Mueller, L.; Pardi, A. *Nat. Struct. Mol. Biol.* **1998**, *5*, 1065.

(4) Nozinovic, S.; Fürtig, B.; Jonker, H. R.; Richter, C.; Schwalbe, H. *Nucl. Acids Res.* **2010**, *38*, 683.

(5) Ottiger, M.; Delaglio, F.; Bax, A. *J. Mag. Res.* **1998**, *131*, 373.

(6) Yang, D.; Nagayama, K. *J. Mag. Res. A* **1996**, *118*, 117.

(7) Vallurupalli, P.; Hansen, D. F.; Stollar, E.; Meirovitch, E.; Kay, L. E. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 18473.

(8) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277.

(9) Goddard, T.; Kneller, D. **2004**.

(10) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D. *Bioinformatics* **2013**, *29*, 845.

(11) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T. E.; Laughton, C. A.; Orozco, M. *Bioph. J.* **2007**, *92*, 3817.

(12) Banas, P.; Hollas, D.; Zgarbová, M.; Jurecka, P.; Orozco, M.; Cheatham III, T. E.; Šponer, J. i.; Otyepka, M. *J. Chem. Theor. Comp.* **2010**, *6*, 3836.

(13) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.

(14) Head, J. D.; Zerner, M. C. *Chem. Phys. Lett.* **1985**, *122*, 264.

(15) Camilloni, C.; Cavalli, A.; Vendruscolo, M. *J. Chem. Theor. Comp.* **2013**, *9*, 5610.

(16) Camilloni, C.; Vendruscolo, M. *J. Am. Chem. Soc.* **2014**, *136*, 8982.

(17) Cavalli, A.; Camilloni, C.; Vendruscolo, M. *J. Chem. Phys.* **2013**, *138*, 094112.

(18) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562.

(19) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.

(20) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553.

(21) Biarnés, X.; Pietrucci, F.; Marinelli, F.; Laio, A. *Comp. Phys. Comm.* **2012**, *183*, 203.

(22) Camilloni, C.; Vendruscolo, M. *J. Phys. Chem. B* **2015**, *119*, 653.

(23) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comp. Phys.*

Comm. **2014**, *185*, 604.

(24) Bax, A. *Protein Sci.* **2003**, *12*, 1.

(25) Tolman, J.; Flanagan, J.; Kennedy, M. A.; Prestegard, J. *Proc. Natl. Acad. Sci. USA* **1995**, *92*, 9279.

(26) Wirz, L. N.; Allison, J. R. *J. Phys. Chem. B* **2015**, *119*, 8223.

(27) Camilloni, C.; Vendruscolo, M. *J. Phys. Chem. B* **2015**, *119*, 8225.

(28) Olsson, S.; Ekonomiuk, D.; Sgrignani, J.; Cavalli, A. *J. Am. Chem. Soc.* **2015**, *137*, 6270.

(29) Ceriotti, M.; Tribello, G. A.; Parrinello, M. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 13023.