Biophysical Journal, Volume 98

Supporting Material

Free Energy Profile of RNA Hairpins: a Molecular Dynamics Simulation Study

Nanjie Deng and Piotr Cieplak

Biophysical Journal, Volume 98

Supporting Material

Free Energy Profile of RNA Hairpins: a Molecular Dynamics Simulation Study

Nanjie Deng and Piotr Cieplak

Supporting Information

Umbrella sampling

To calculate PMF using umbrella sampling, a series of MD simulations is performed on a RNA hairpin with the harmonic distance restraint between the C1'-C1' atoms of the two terminal nucleotides.. The biasing potential in the *i*-th simulation window is $W_i(r) = \frac{1}{2} k_i \left(r - r_i^0\right)^2$, where k_i is the force constant, r is the reaction coordinate and r_i^0 is the reference distance for the *i*-th sampling window. Starting from the folded conformation, the reference distance r_i^0 is incremented for the successive windows until it reaches approximately 35 Å. The largest value of r_i^0 corresponds to the equilibrium end-to-end distance of a single stranded RNA molecule, estimated from a 2-ns unrestrained MD simulation starting from the canonical A-form single strand conformation.

To choose an appropriate set of r_i^0 and k_i, a trial calculation of PMF is performed using 15 uniformly spaced sampling windows and a single force constant k = 1.2 kcal mol^{-1} Å^{-2}. For this trial calculation, 2-ns sampling per window are performed to obtain a rough estimate of PMF. Based on the shape of the trial PMF, force constants k_i are empirically adjusted to make the biasing potential offset the curvature of the trial PMF. To ensure significant overlap between adjacent windows in the production run, more sampling windows are inserted in the regions where the slope in the free energy is large. The full range of the reaction coordinate is covered using 24 windows centered at: r_i^0 = 10.7, 11.5, 12.0, 12.5, 13.0, 13.8, 14.8, 16.1, 17.5, 18.8, 20.2, 21.5, 22.9, 24.2, 25.6, 26.9, 27.9, 28.8, 29.6, 30.4, 31.3, 32.1, 33.5, 34.9 Å. For the UUCG tetraloop with AMBER ff99, the final set of k_i = 1.2, 8.0, 8.0, 6.0, 4.0, 1.6, 1.6, 1.6, 1.6, 1.6, 1.6, 1.6, 1.6, 1.6, 1.6, 1.6, 1.6, 2.0, 6.0, 6.5, 5.0, 4.0, 3.0, 2.0 kcal mol^{-1} Å^{-2}. In each sampling window, a 12-ns explicit solvent MD simulation is performed, starting from the last conformation of the previous sampling window. For the first window, the simulation is started from the folded structure. The first 2 ns are treated as equilibration, which allows the system to adjust to the current umbrella potential. The last 10 ns of sampling data are used for the calculation of PMF.

The biased probability distributions of the end-to-end distance $P'(r_i)$ accumulated in these sampling windows are unbiased and combined using the Weighted Histogram Analysis Method (WHAM) method to yield the unbiased distribution $P(r_i)$ and the

associated potential of mean force (1, 2). The WHAM program implemented by Grossfield (3) is used to calculate the PMF.

MD protocol

The MD simulations were performed using the CHARMM program (4) version 33b1. The all-atom AMBER ff99 (5) and CHARMM27 parameter set (6) were used to model RNA in aqueous solutions. Sodium chloride counterions were added to neutralize net charges on the RNA and to maintain an ionic concentration of 0.15 M. A truncated octahedral box containing TIP3P water molecules (7) previously equilibrated at 300 K and 1 atm pressure was used to solvate the RNA molecules. The dimension of the solvent box is 56.63 Å × 56.63 Å × 56.63 Å, which is chosen to ensure that the solute atoms are separated from nearest walls of the box by at least 9 Å when the RNA molecules are in extended conformations. To minimize the possible impact of solvent box on the calculated PMF, the same solvated box was used for all umbrella sampling windows. Water molecules within 2.8 Å of solute atoms or counterions were removed. The final solvated systems contained between 13,000 and 14,000 atoms.

The electrostatic interactions were computed using the particle-mesh Ewald (PME) method (8) with a real space cutoff of 10 Å and a grid spacing of 1.05 Å. The van der Waals interactions were smoothly switched off between 8 Å and 10 Å. The bond lengths involving hydrogen atoms were constrained by the SHAKE algorithm (9). The equation of motion was integrated with a time step of 2 fs. MD simulations were performed in the NPT ensemble, under the atmospheric pressure. The following protocol was used to prepare the initial solvated system for the production run: the solvent alone was first energy minimized for 500 steps of steepest descent and 500 steps of adopted basis Newton-Raphson (ABNR) method (4), with the solute atoms fixed in space. The whole system was then energy minimized for 500 steepest descent steps and 500 ABNR steps without constraints. Following these energy minimizations, the system was heated to 310 K from 50 K within 100 ps, with the solute atoms harmonically restrained. Before the production run, the system was equilibrated at the target temperature and pressure for 300 ps, during which the harmonic restraint on the solute atoms was gradually switched off. The production MD trajectories were saved every 1 ps for analysis.

References

1. Kumar S., D. Bouzida, R. Swendsen, P. Kollman, A. Rosenberg, The weighted histogram analysis method for free energy calculations on biomolecules. I. The method. *J. Comput. Chem*., **1992**, 13, 1011-1021.

2. Roux, B. The calculation of the potential of mean force using computer simulations. *Comp. Phys. Commun.*, **1995**, 91, 275-282.

3. Grossfield, A. An implementation of WHAM: the Weighted Histogram Analysis Method, http://membrane.urmc.rochester.edu/Software/WHAM/WHAM.html, **2008**, April, 14.

4. Brooks, B., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan, M. Karplus, CHARMM: A program for macromolecular energy, minmimization, and dynamics calculations. *J. Comp. Chem*., **1983**, 4, 187-217.

5. Cheatham, T., P. Cieplak, P. Kollman, modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat. *J. Biomol. Struct. Dyn.* **1999**, 16, 845-861.

6. Foloppe N., A. MacKerell Jr., All atom empirical force field for nucleic acids: 1) Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.*, **2000**, 21, 86-104.

7. Jorgensen, W., J. Chandrashekhar, J. Madura, R. Impey, and M. Klein, Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, 79, 926-935.

8. Essmann, U., L. Perera, M. Berkowitz, T. Darden, H. Lee, and L. Pedersen, A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, 103, 8577–8593.

9. Ryckaert J., G. Ciccotti, H. Berendsen, Numerical integration of cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes, *J. Comp. Phys*., **1977**, 23, 327–341.

Table S1 Free energy contributions to folding in UUCG, calculated with MM-PB/SA.[j]

Energy	Hairpin	Single strand	Δ: hairpin – single strand
E(internal)[a]	495.8 (0.6)	489.1 (0.7)	6.7
E(vdw)[b]	-74.0 (0.3)	-63.5 (0.4)	-10.5
E(elec)[c]	-885.0 (0.9)	-857.3 (1.4)	-27.7
G(solv_elec)[d]	-1659.4 (0.7)	-1675.5 (1.2)	16.1
G(total_elec)[e]	-2544.4 (0.4)	-2532.8 (0.4)	-11.6
G(solv_np)[f]	12.7 (<0.1)	14.1 (<0.1)	-1.4
G(gas+solv)[g]	-2109.9 (0.6)	-2093.1 (0.7)	-16.8
-TS(vib)[h]	-257.4 (0.1)	-261.6 (0.5)	4.2
G(total)[i]	-2367.3 (0.6)	-2354.8 (0.7)	-12.6

a. Solute internal energy: the sum of bonded terms, i.e. bond, angle and torsion energies.

b. Solute van der Waals energy.

c. Electrostatics energy of solute.

d. Electrostatics component of the solvation free energy, computed by solving Poisson-Boltzmann equation in solution and in vacuum.

e. G(total_elec) = E(elec) + G(total_elec).

f. Non-polar component of the solvation free energy.

g. The effective energy, G(gas+solv) = E(intra) + E(vdw) + E(elec) + G(solv_elec) + G(solv_np).

h. Solute vibrational entropy.

i. G(total) = G(gas+solv) – TS(vib).

j. Results are averaged over 500 coordinate frames taken from the last 5 ns of trajectory. The numbers in the parenthesis are standard error of the means.

Table S2 Free energy contributions to folding in UUUU, calculated with MM-PB/SA.[j]

Energy	Hairpin	Single strand	Δ: hairpin – single strand
E(internal)[a]	488.3 (0.6)	477.6 (0.6)	10.7
E(vdw)[b]	-62.9 (0.3)	-41.3 (0.4)	-21.6
E(elec)[c]	-724.7 (0.9)	-763.9 (1.5)	39.2
G(solv_elec)[d]	-1638.6 (0.7)	-1604.1 (1.3)	-34.6
G(total_elec)[e]	-2363.3 (0.4)	-2368.0 (0.4)	4.6
G(solv_np)[f]	12.9 (<0.1)	14.7 (<0.1)	-1.8
G(gas+solv)[g]	-1925.0 (0.6)	-1917.0 (0.6)	-8.0

-TS(vib)[h]	-258.5 (0.2)	-263.8 (0.2)	5.3
G(total)[i]	-2183.5 (0.6)	-2180.8 (0.6)	-2.8

a-i: See footnote of Table S1.

Table S3. Sugar puckering for the UUUU tetraloop.

	NMR[a]	CHARMM27	AMBER ff99
U4	C3'-endo	C3'-endo	C3'-endo
U5	C2'-endo	C3'-endo	C2'-endo
U6	37% C2'-endo 63% C3'-endo	C3'-endo	C2'-endo
U7	C2'-endo	C3'-endo	C2'-endo

a. Fürtig B., Richter C., Wöhnert J., and Schwalbe H., NMR spectroscopy of RNA. *ChemBioChem*, **2003**, 4, 936-962.

Figure S1. Snapshots of the UUCG hairpin observed in the umbrella sampling simulations using AMBER ff99. The end-to-end distances are indicated in each structural view.

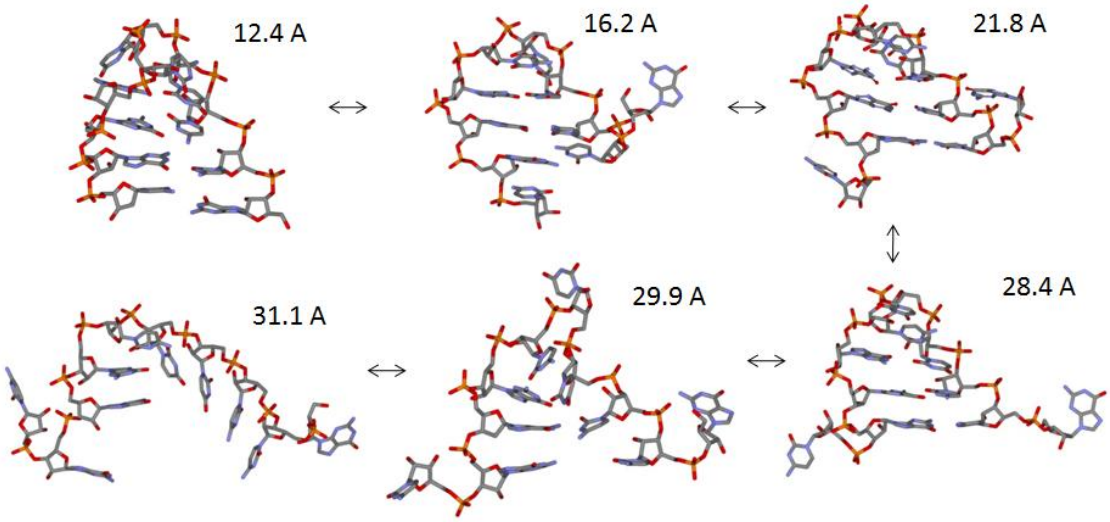


Figure S2. Same as Fig. S1, except for the UUUU hairpin.

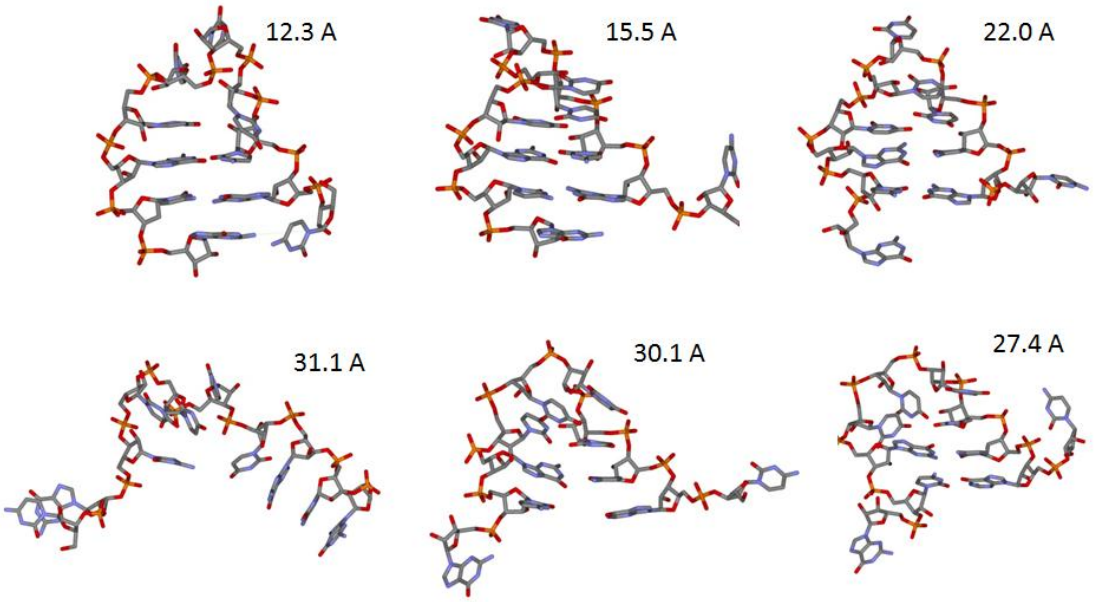


Figure S3. Snapshots of the UUCG hairpin observed during umbrella sampling simulations using CHARMM27.

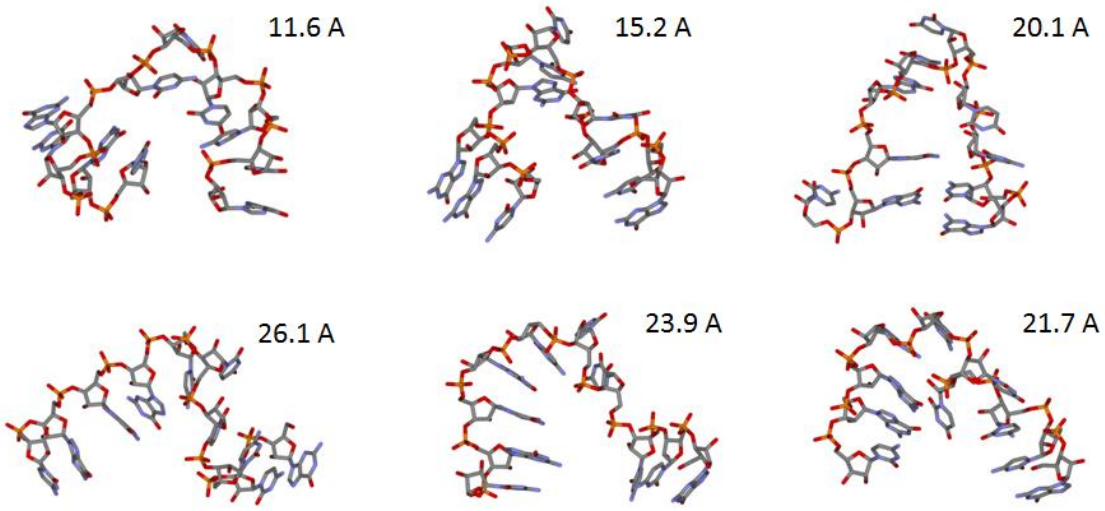
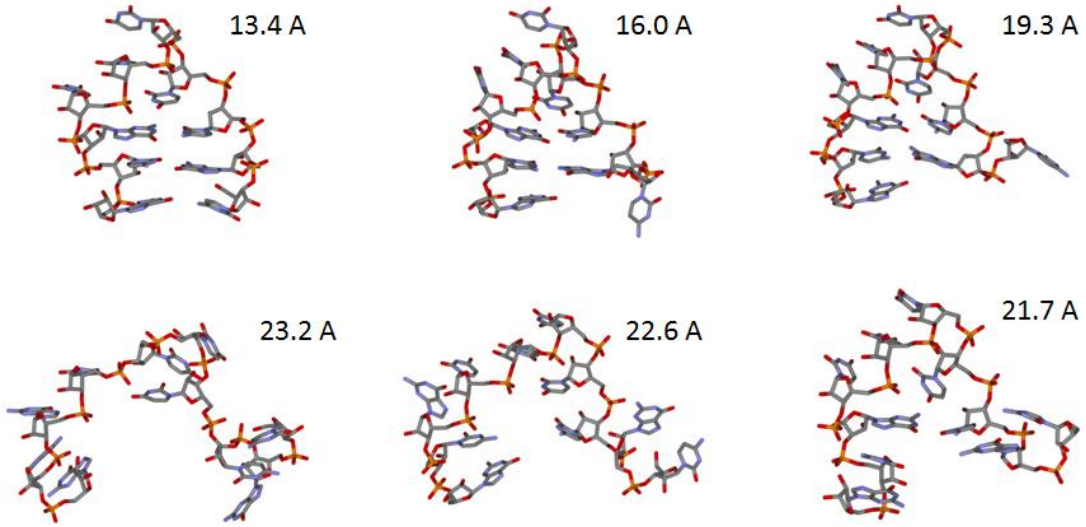


Figure S4. Same as Fig. S3, except for the UUUU hairpin.



.

Figure S5 Snapshots of the UUCG hairpin observed during the 7 ns CHARMM27 simulation in explicit solvent.

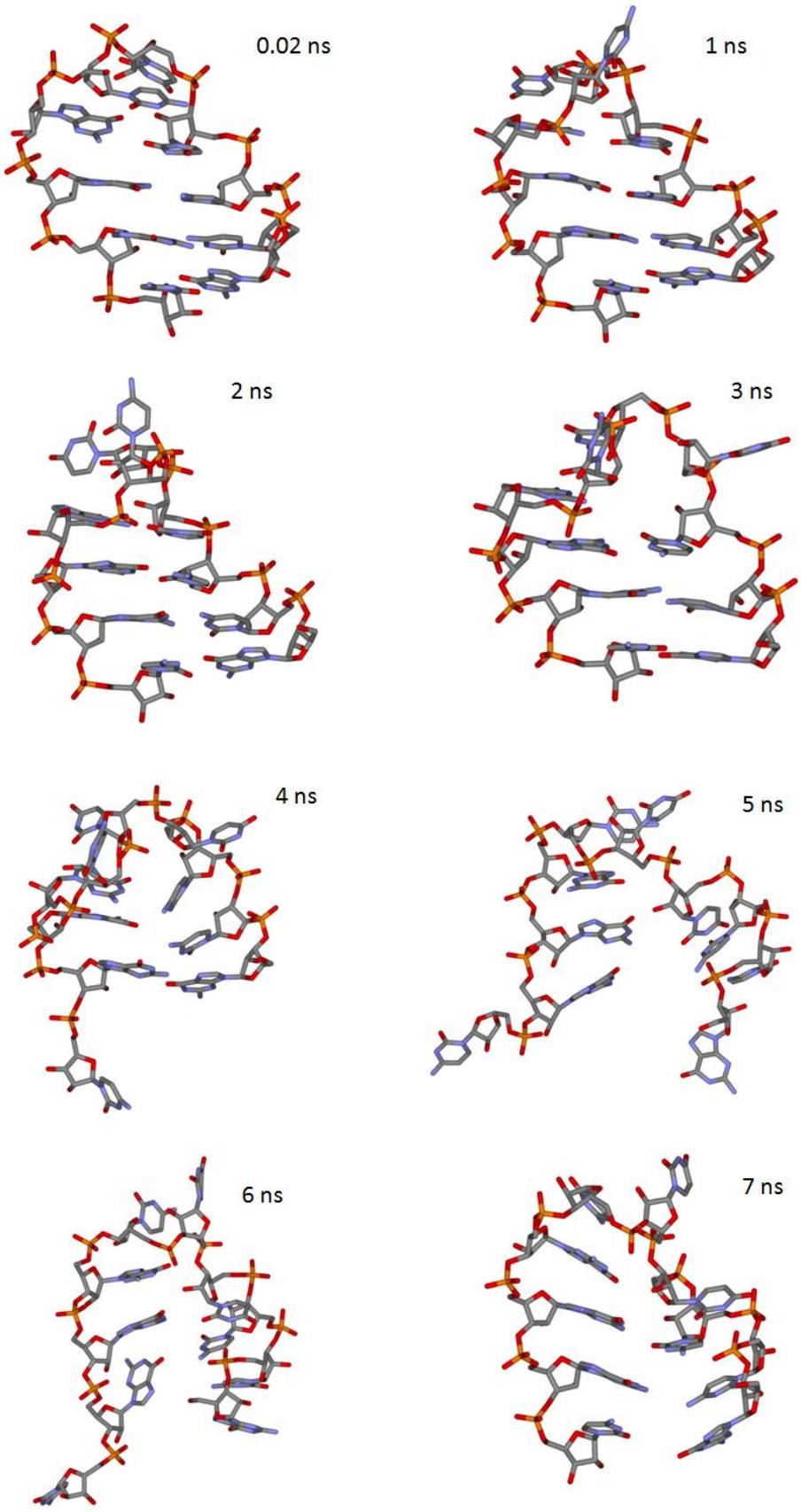
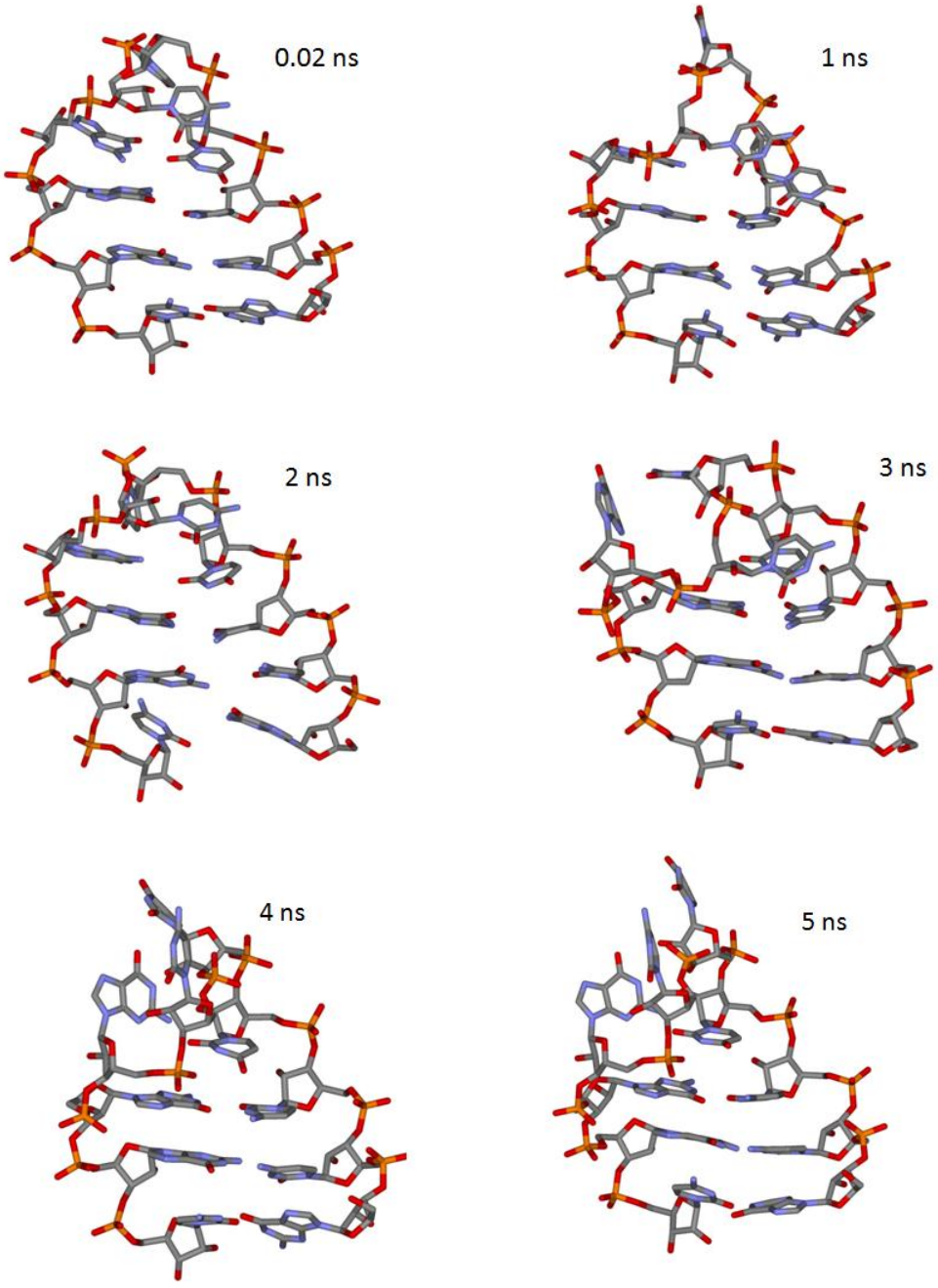
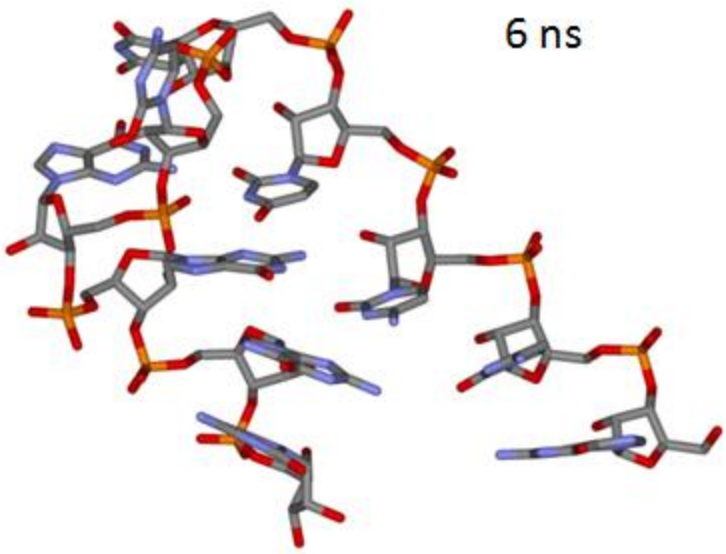

0.02 ns
1 ns
2 ns
3 ns
4 ns
5 ns
6 ns
7 ns

Figure S6 Snapshots of the UUCG hairpin observed during the 10 ns CHARMM27 simulation in explicit solvent.




6 ns

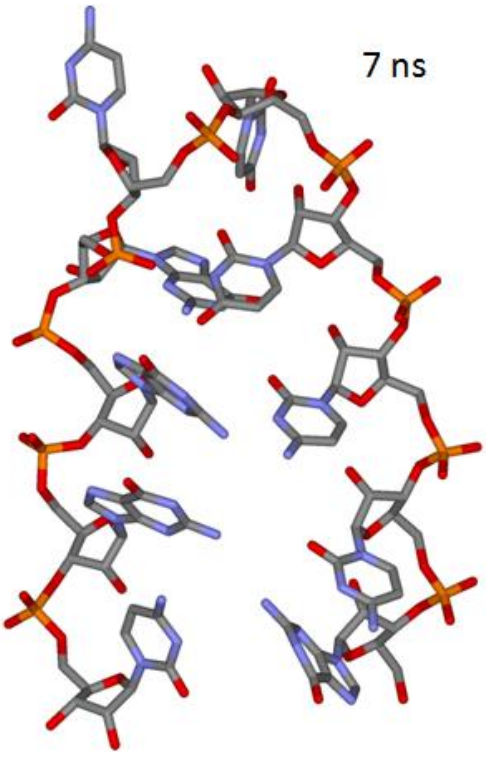
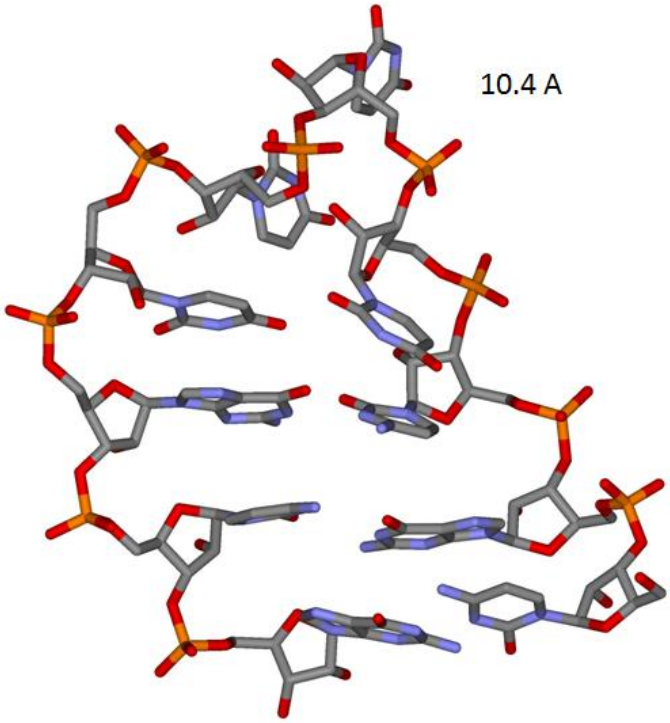

7 ns

Figure S7 Representative structures of the UUUU hairpin at the global free energy minima obtained with (a) AMBER ff99 and (b) CHARMM27. The dashed line in the CHARMM27 structure indicates a hydrogen bond between the U4 base and the phosphate linking U6 and U7.

a.



b.

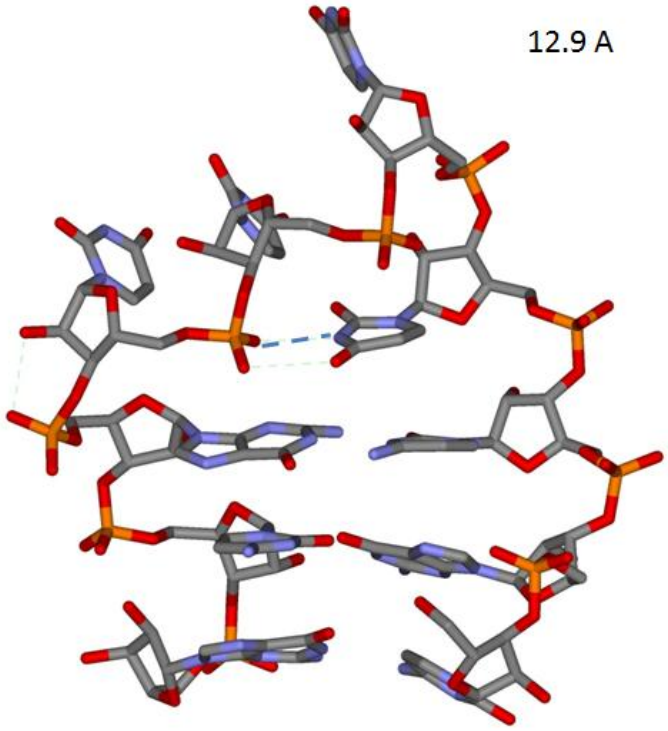
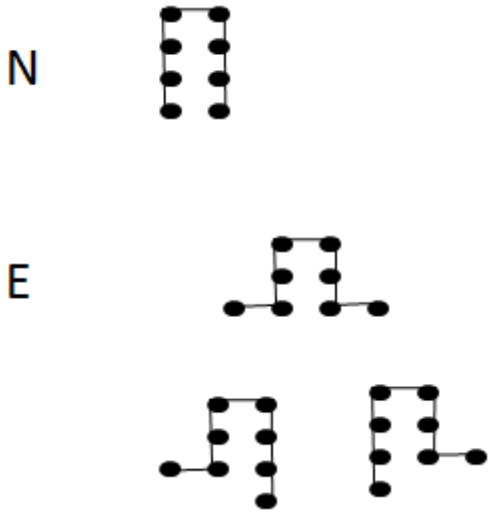


Figure S8 Configurations of the four-state model used by Ma. *et al.* for the 8-nt RNA hairpins. N: native state; E: a state with intact lop and frayed terminal base pair; U: unfolded state; S: a state with nonnative loop and native stem.



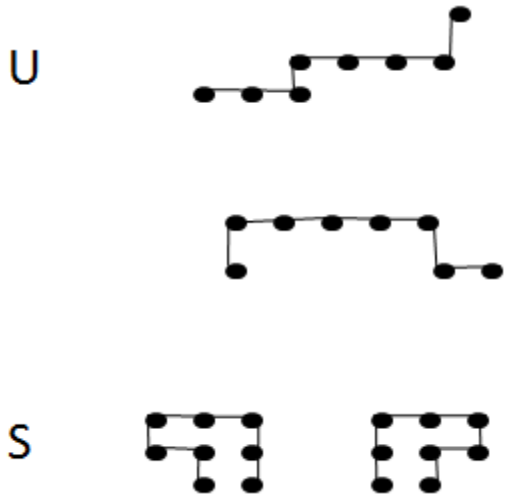


Figure S9 To determine the suitability of the AMBER force fields in modeling RNA hairpin at long times, two unrestrained MD simulations were performed with AMBER ff99 and AMBER parmbsc0 parameters set. Both trajectories were started from the experimental structure and lasted for 100 ns. The figure shows the RMSD relative to the crystal structure of UUCG for AMBER ff99 (blue) and AMBER parmbsc0 (red)

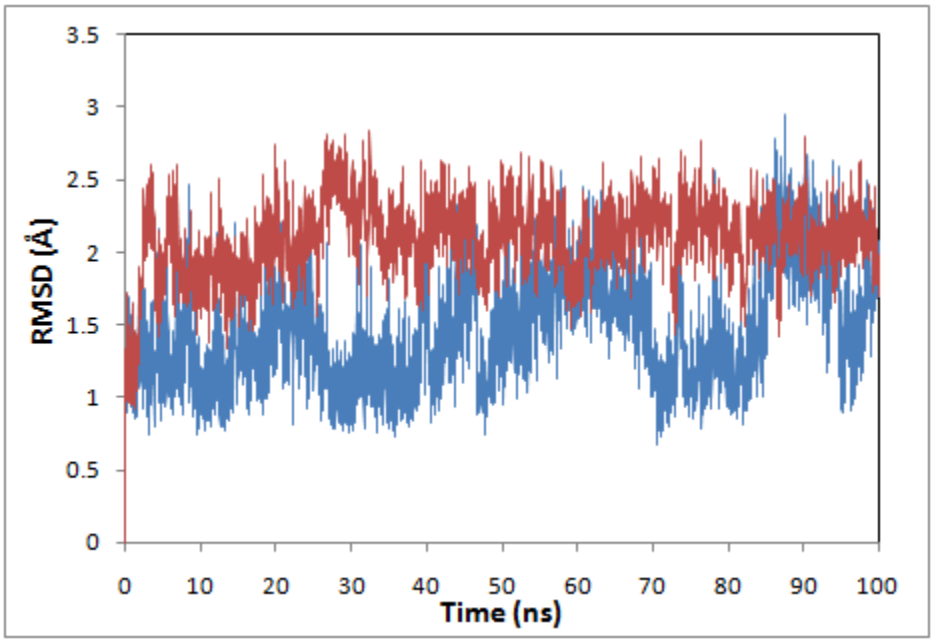


Figure S10 Results of the AMBER ff99 simulation (see Fig. S9): RMSD from X-ray structure for the loop (red) and stem (blue) regions.

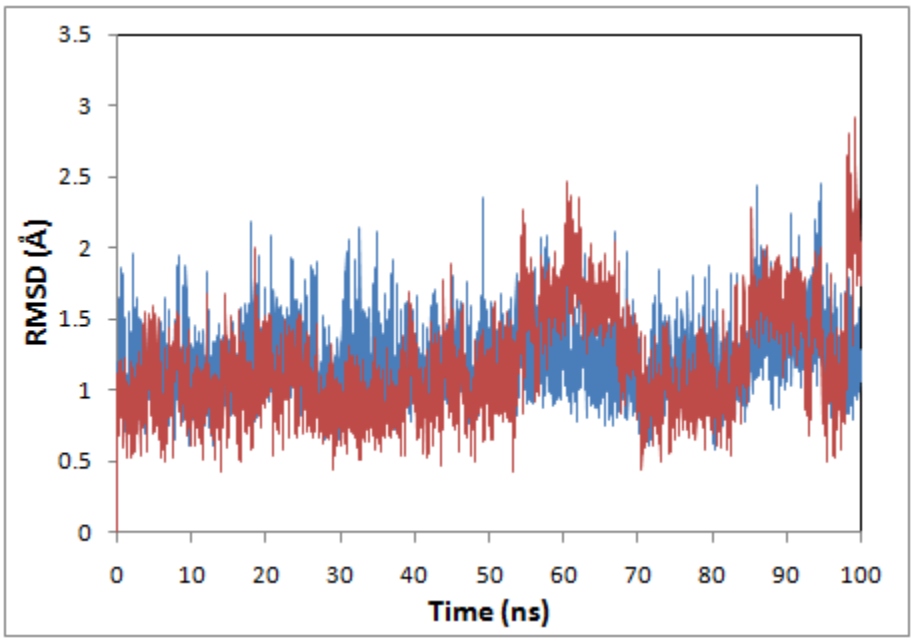


Figure S11 The same as Fig. S10, except using AMBER parmbsc0.

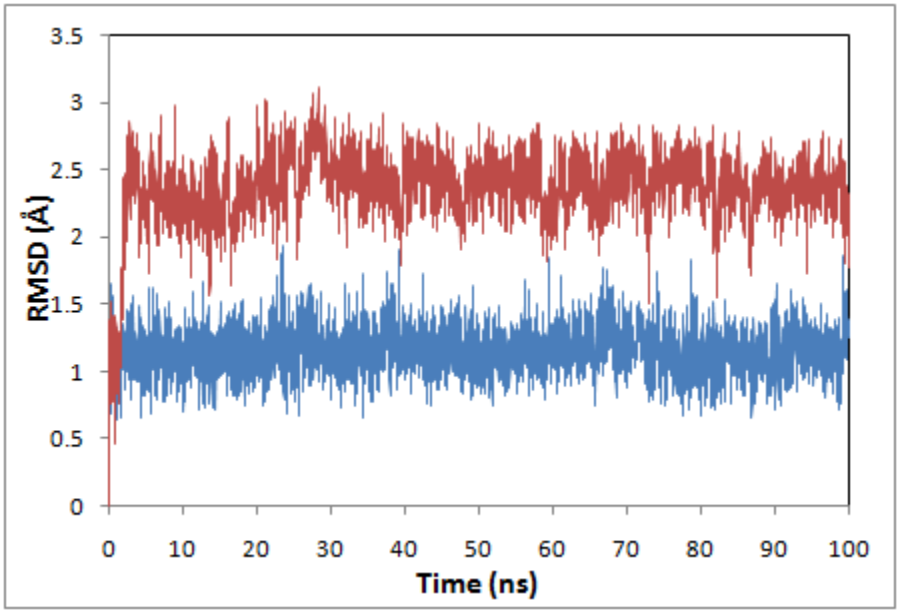


Figure S12 Results of the AMBER ff99 simulation (see Fig. S9): snapshot (red) taken at 100 ns, superimposed onto the X-ray structure (blue).

Figure S13 Results of the AMBER parmbsc0 simulation (see Fig. S9): snapshot (red) taken at 100 ns, superimposed onto the X-ray structure (blue).

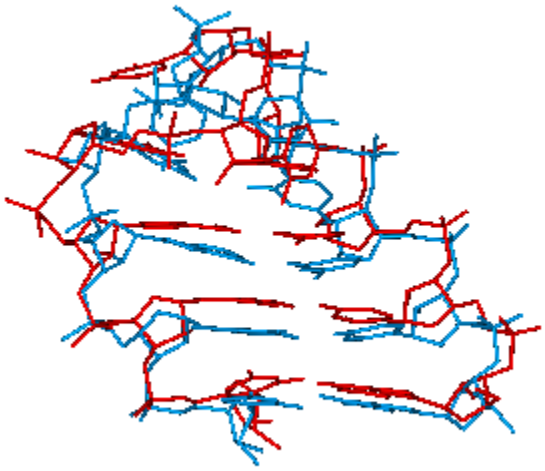


Figure S14 PMF of UUCG calculated using different trajectory lengths:
5 ns (blue), 7 ns (red), 8.5 ns (green) and 10 ns (black).

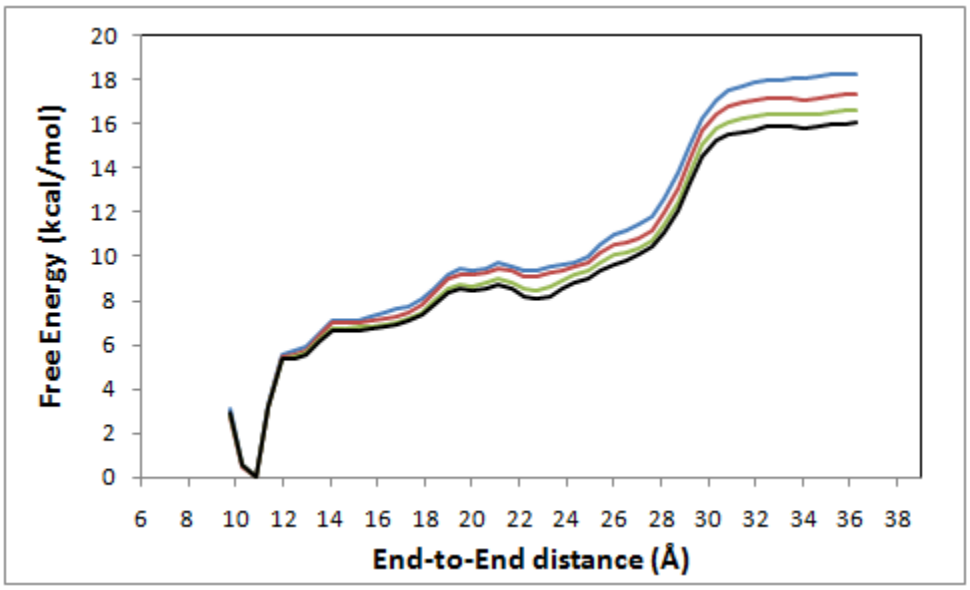


Figure S15 Same as Fig. S14, except for UUUU.

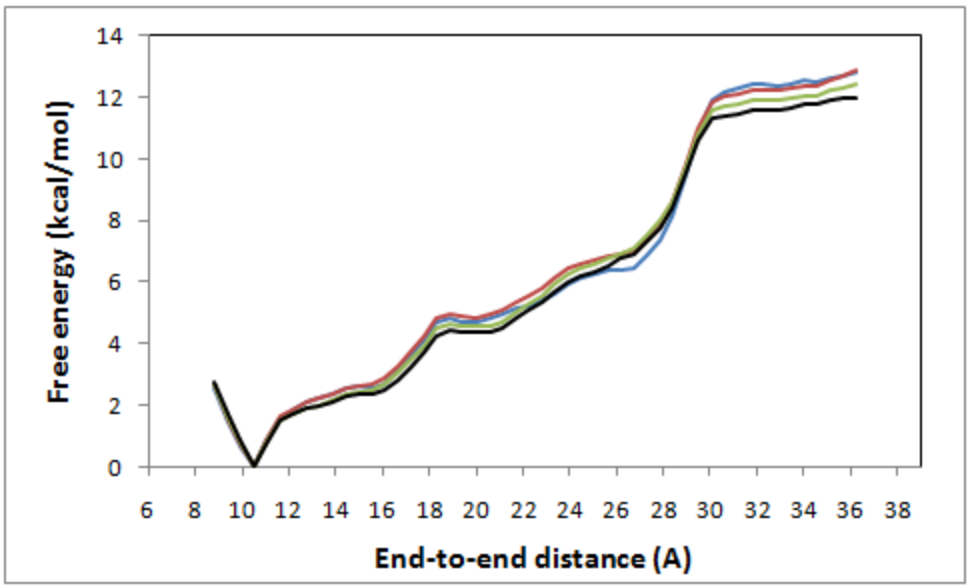
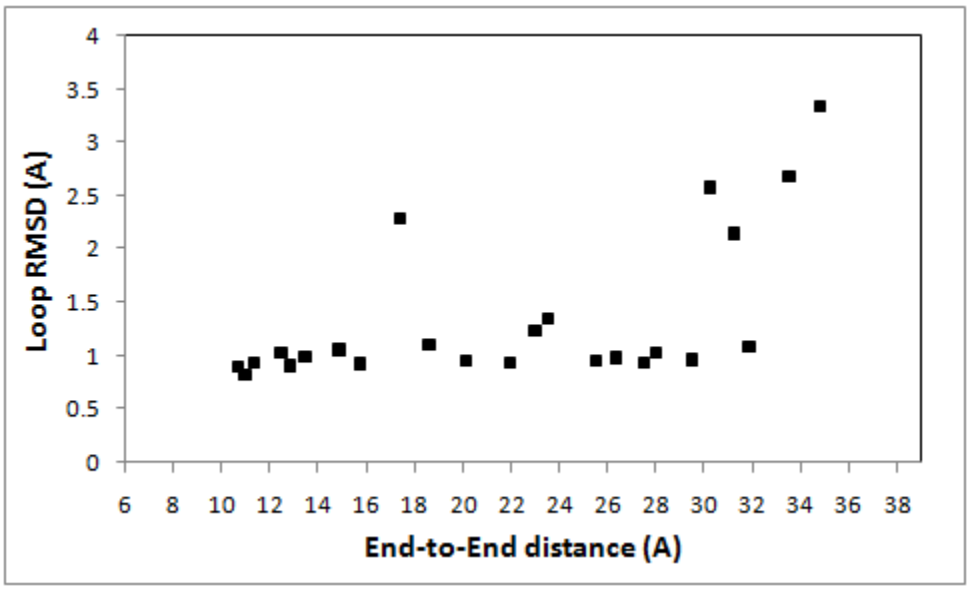


Figure S16 RMSD of the loop segment of UUCG relative to the crystal structure, as a function of end-to-end distance. Each RMSD value is obtained as the average over 10 ns umbrella sampling trajectory in each window.


Loop RMSD (A)
4
3.5
3
2.5
2
1.5
1
0.5
0
6
8
10
12
14
16
18
20
22
24
26
28
30
32
34
36
38
End-to-End distance (A)