Supplemental Information for

AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids

Changsheng Zhang,[1]# Chao Lu,[2]# Zhifeng Jing,[1] Chuanjie Wu,[2] Jean-Philip Piquemal,[1,3] Jay W. Ponder,[2]* Pengyu Ren[1]*

[1] Department of Biomedical Engineering, The University of Texas at Austin,

Austin, Texas 78712, United States.

[2] Department of Chemistry, Washington University in Saint Louis, Saint Louis, Missouri 63130,

United States

3. Laboratoire de Chimie Théorique, Sorbonne Universités, UPMC, UMR7616 CNRS, Paris,

France

Table of Contents

Supplementary methods:

I. Stacking calculation in tetramer simulation analysis

The same as Turner [reference 111, D. E. Condon et al., J. Chem. Theory Comput. 11, 2729–2742 (2015)] suggested, the base-base stacking score contains three terms, center distance (d, Dscore), overlap (ω angle, Oscore), and parallel (Ξ angle, Pscore).

Stacking score = (Dscore + Oscore)×Pscore;

Dscore= 1, if (d<3.5A); 0, if (d>5.0); $(1.0/d^3-1.0/125)*65.2588$, else

Oscore=1, if (ω<25); 0, if (ω>50); 2-ω/25.0, else

Pscore=1 if (Ξ<45); -1 else

Percentage of base stacking observed in MD simulations:

II. Karplus equations for ^3J coupling with β, γ, ε, ν torsions

β vs J^3_{H5'-P} & J^3_{H5''-P}:
$$J^3_{H5'-P} = \frac{1}{N}\sum_{i=1}^{N}15.3\cos^2(\beta_i-120°)-6.1\cos(\beta_i-120°)+1.6;$$

$$J^3_{H5''-P} = \frac{1}{N}\sum_{i=1}^{N}15.3\cos^2(\beta_i+120°)-6.1\cos(\beta_i+120°)+1.6$$

ε vs J^3_{H3'-P}: $J^3_{H3'-P} = \frac{1}{N}\sum_{i=1}^{N}15.3\cos^2(\varepsilon_i+120°)-6.1\cos(\varepsilon_i+120°)+1.6$

γ vs J^3_{H4'-H5'} & J^3_{H4'-H5''}:
$$J^3_{H4'-H5'} = \frac{1}{N}\sum_{i=1}^{N}9.7\cos^2(\gamma_i-120°)-1.8\cos(\gamma_i-120°);$$

$$J^3_{H4'-H5''} = \frac{1}{N}\sum_{i=1}^{N}9.7\cos^2(\gamma_i)-1.8\cos(\gamma_i°)$$

ν vs J^3_ν $J^3_\nu = \frac{1}{N}\sum_{i=1}^{N}9.67\cos^2(\nu)-2.03\cos(\nu)$

Part I: DNA/RNA AMOEBA force field parameterization

Figure S1. DNA/RNA intra-molecular polarization group definition.

The multipole and torsional parameters for DNA/RNA C5'-O5'-P-O3' and C3'-O3'-P-O5' were directly transferred from DMP (dimethyl phosphate) and these atoms are kept in the same group.

R = H, for DNA

R = OH, for RNA
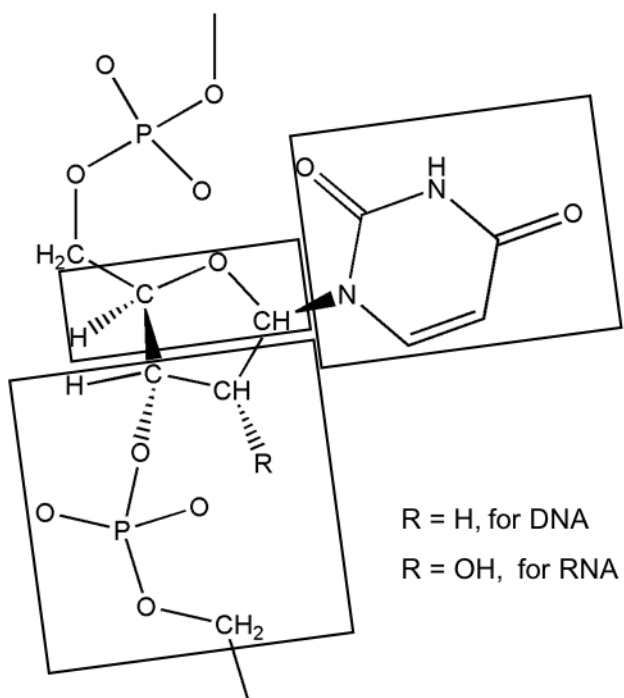
Figure S2. Potential energy surfaces of sugar with respect to v_0 and v_4.
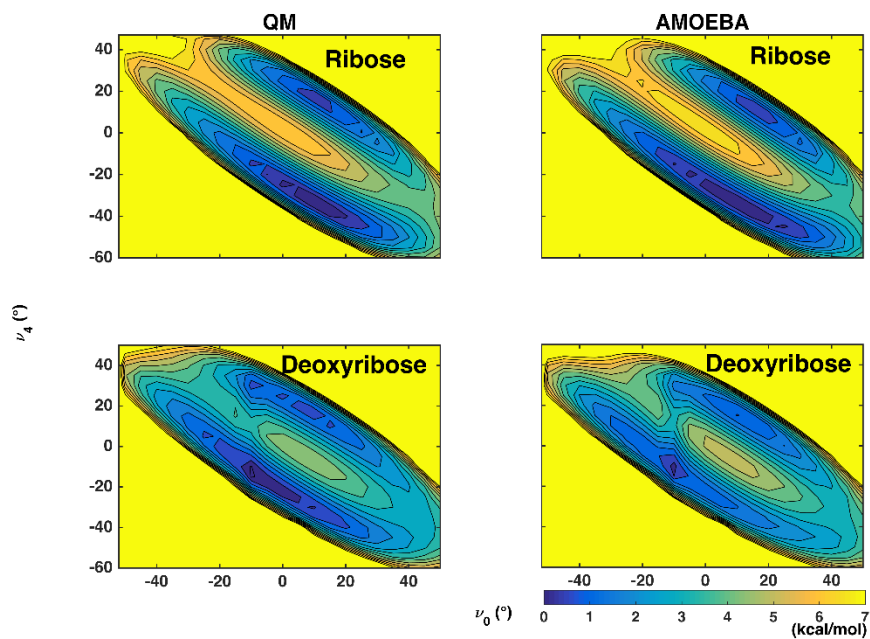
Figure S3. Comparison of AMOEBA and QM sugar puckering energy maps for dG and dT.
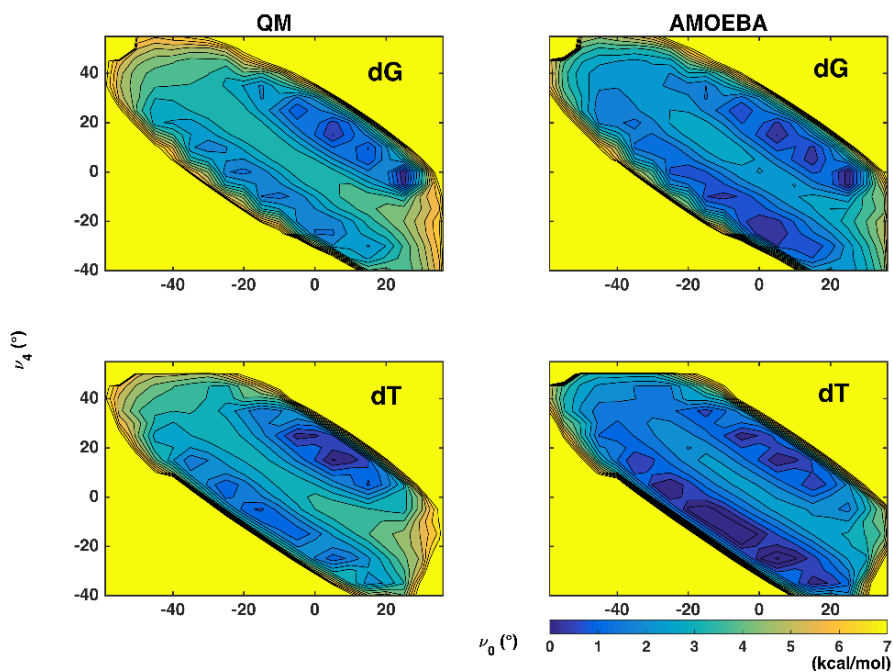
Figure S4. Comparison of AMOEBA and QM sugar puckering energy maps for rG and rU.
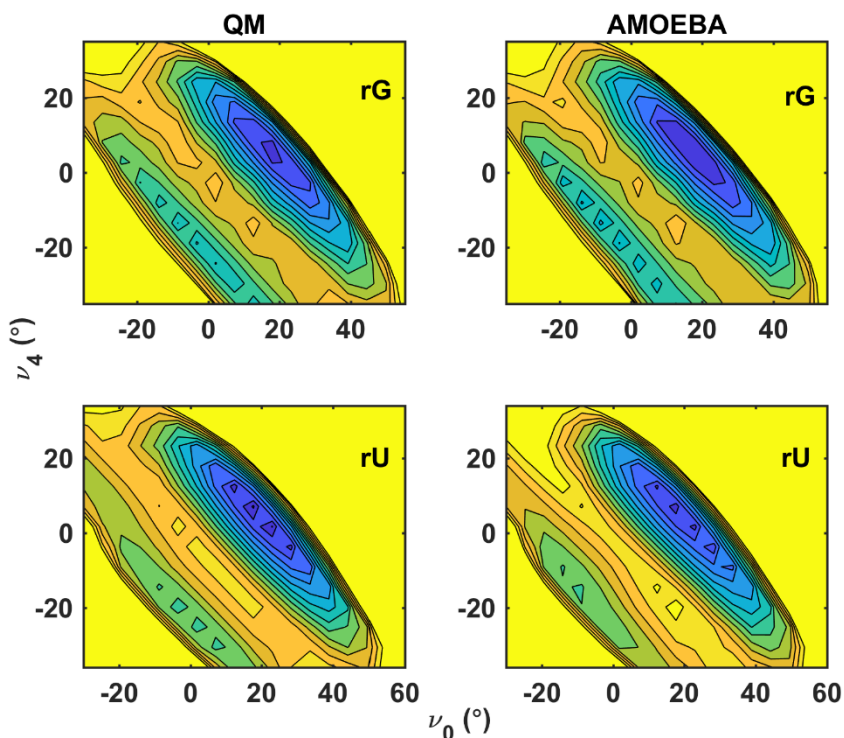
Figure S5. Nucleobases conformational energy profiles with the rotation of χ torsion.

**(A) deoxyribonucleosides (B) ribonucleosides. The (deoxy)ribose sugar was fixed at (C3')
C2' endo conformation in both QM (MP2/6-311G(1d,1p), black plus)) and AMOEBA (red
circle) calculations. The χ torsion is defined by O4'-C1'-N1-C2 for pyrimidines and O4'-
C1'-N9-C4 for purines.**

(A)
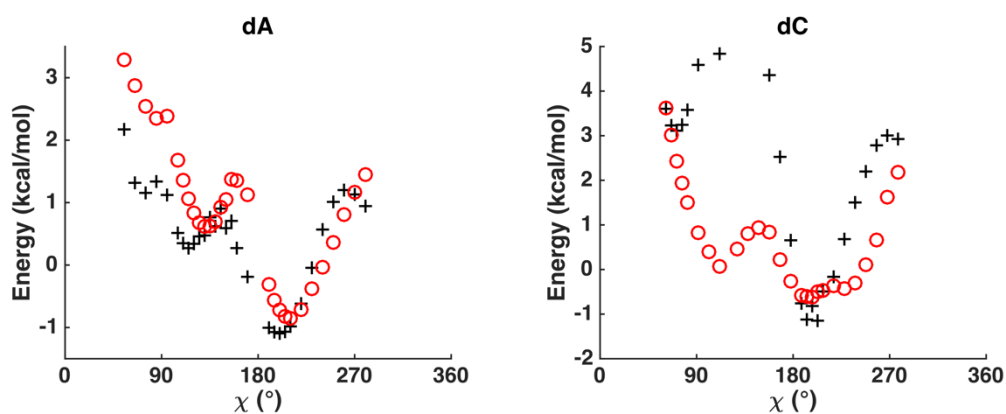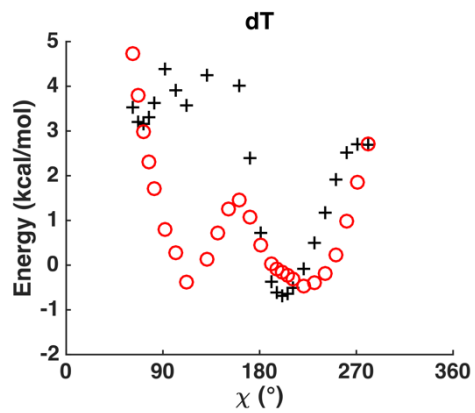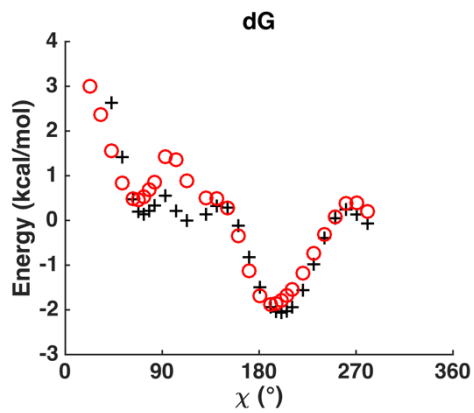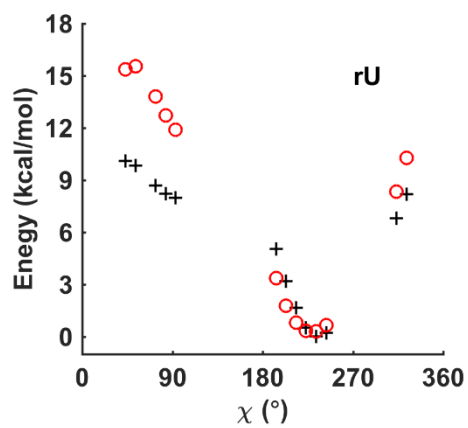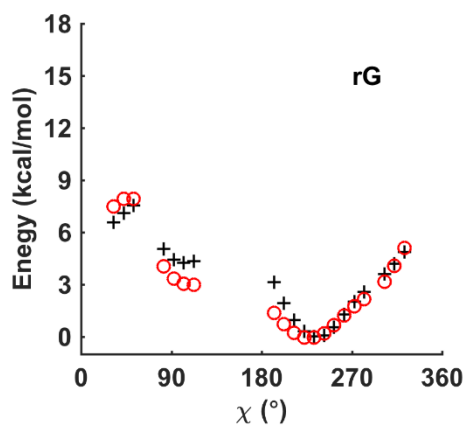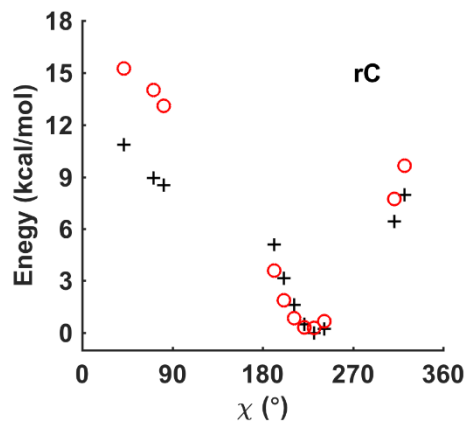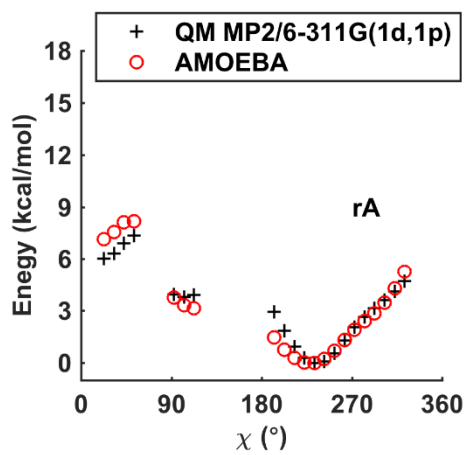
Figure S6. Nucleosides -C2'-O2'- torsional energy surfaces.

The ribose sugar was fixed at C3' endo conformation for both QM (MP2/6-311G(1d,1p), black plus) and AMOEBA calculation (red circle).
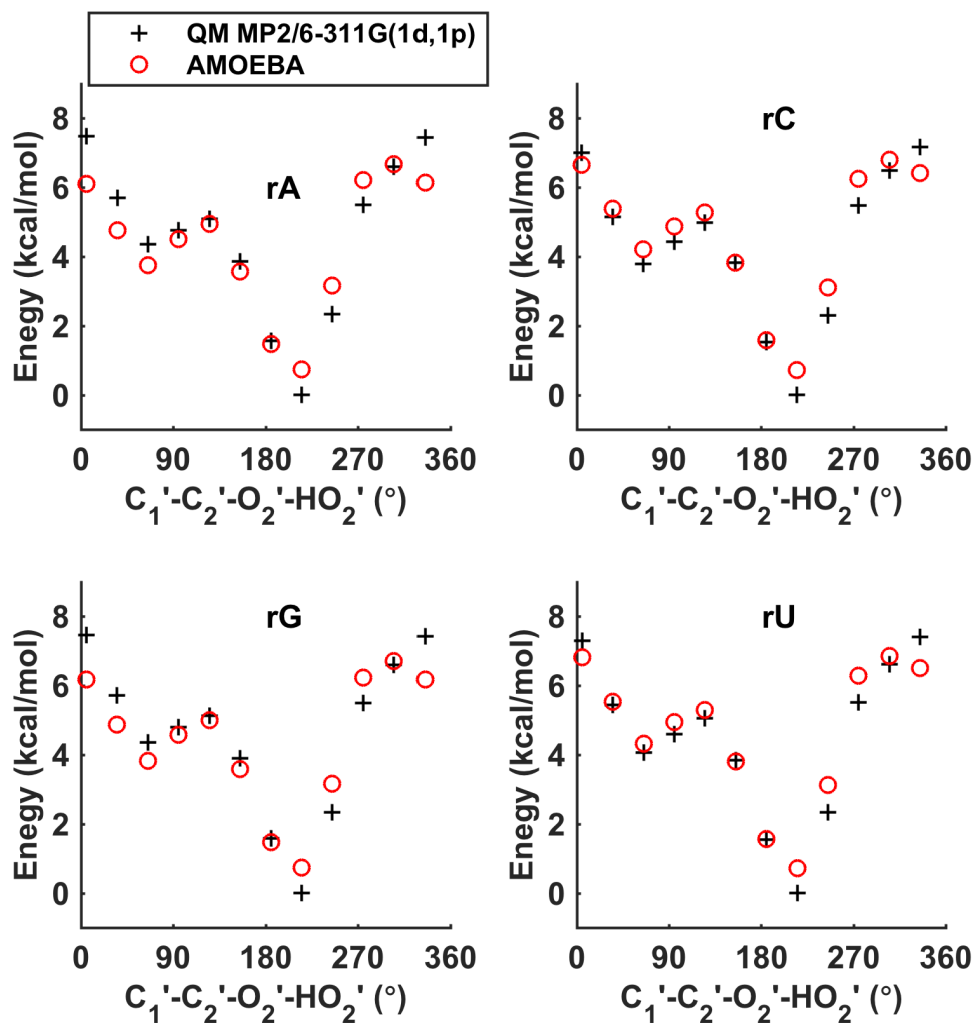
Figure S7. Nucleosides terminal C4'-C3'-O3'-HO3' and C4'-C5'-O5'-HO5' torsional energy surfaces.
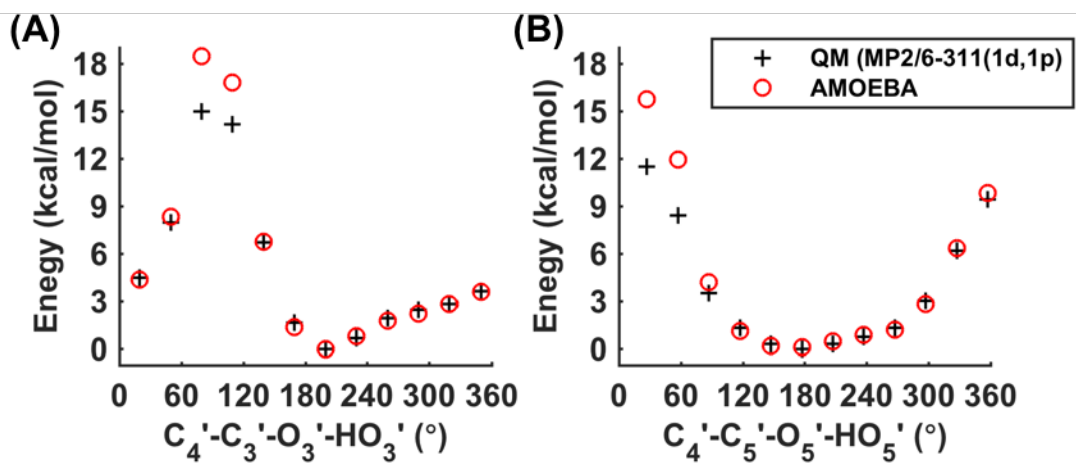
Table S1. Stretch-torsion (A) and angle-torsion coupling (B) parameters for O4'-C1'-N-C6/C8 torsion.

The coupling energy equations are:

$$U_{angtor}(\chi, a_1, a_2) = \sum_{m=1}^{2}\sum_{n=1}^{3} k_{angtor,mn}(a_m - a_{m0})[1 + \cos(n\chi + \phi_n)] \qquad \phi_1 = \phi_3 = 0, \phi_2 = \pi$$

$$U_{strtor}(\chi, b_1, b_2, b_3) = \sum_{m=1}^{3}\sum_{n=1}^{3} k_{strtor,mn}(b_m - b_{m0})[1 + \cos(n\chi + \phi_n)] \qquad \phi_1 = \phi_3 = 0, \phi_2 = \pi$$

(A)

n \ m	1 (C6/C8-N-C1')	2 (N-C1'-O4')
1	0.014 kcal/mol/degee	0
2	0	0
3	-0.058 kcal/mol/degee	-0.0110 kcal/mol/degee

(B)

n \ m	1 (C6/C8-N)	2 (N-C1')	3 (C1'-O4')
1	0	0	0
2	0	0	0
3	1.5 kcal/mol/Å	-4.0 kcal/mol/Å	5.2 kcal/mol/Å

Figure S8. Model compound used for DNA/RNA backbone torsion parameterization.

(A) Deoxyribose 3,5-bis (methyl phosphate) for DNA β, γ, and ε torsions. (B) MEP (Methyl ethyl phosphate) for RNA β torsion. (C) MHFP (2-Methyltetrahydrofuran phosphate) for RNA γ torsion. (D) Ribose 3,5-bis (methyl phosphate) for RNA ε torsion.
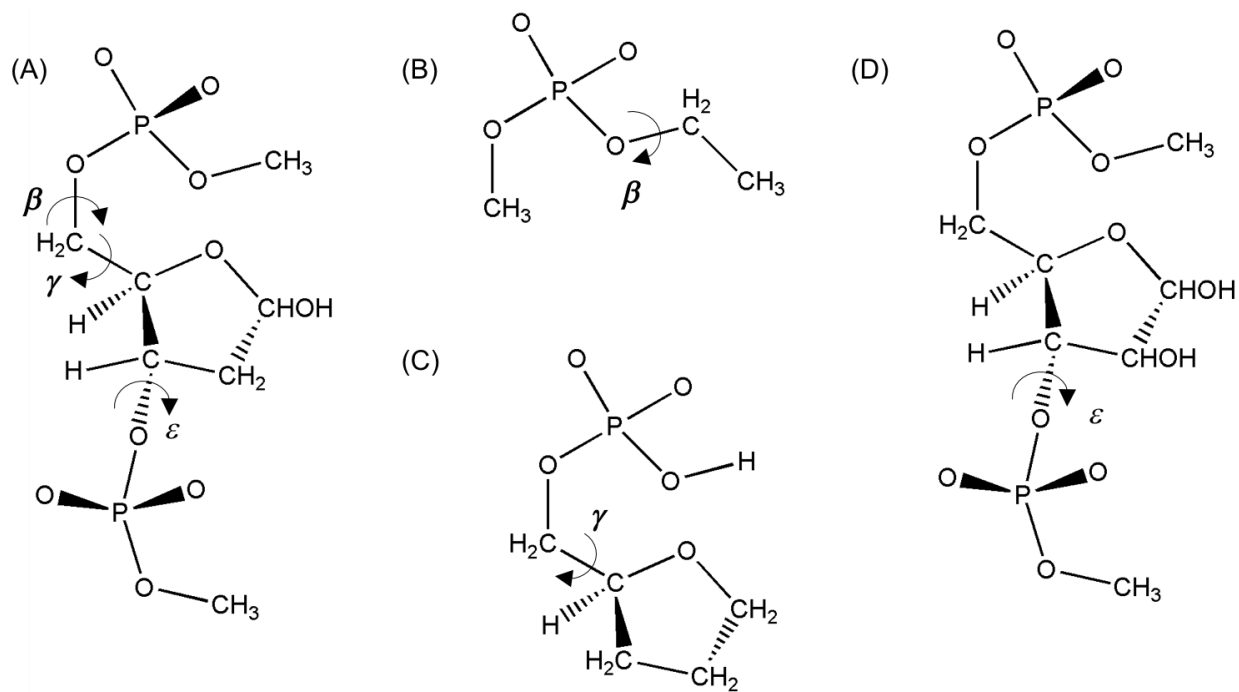
Figure S9. The γ (O5-C5-C4-C3) torsional energy profile using 2-Methyltetrahydrofuran phosphate as the model compound.
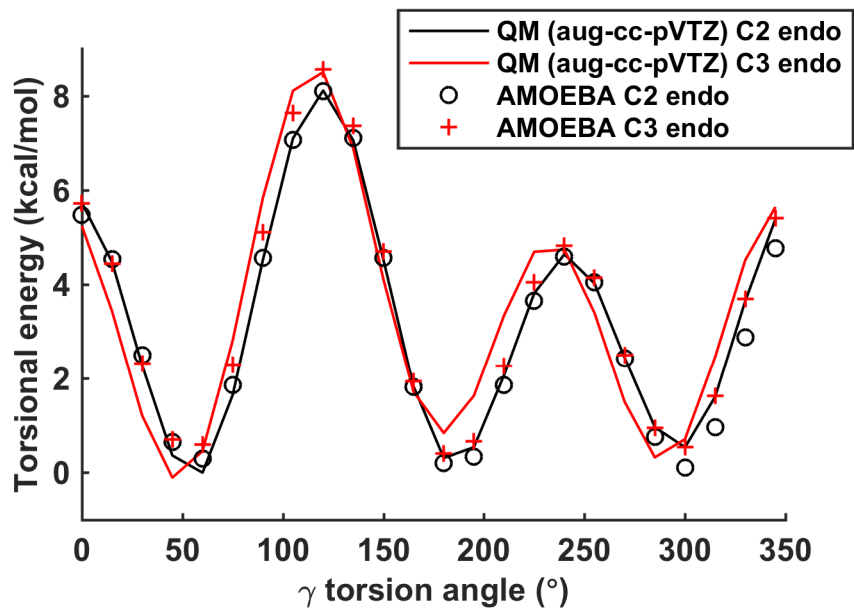
Table S2. β torsion determination using methyl ethyl phosphate as model compound.

(A) QM, statistical potential energy and the AMOEBA energy with different force constants. RNA PDB structures (no protein binding) were used for the statistical energy calculation. A torsional histogram with a 10° bin from 0-360° were calculated, and then statistical potential energy was calculated by taking the logarithm of the frequency (ρ) and times with kT. (B) β ^{3}J coupling on tetramers using soft (torsion force constant =0) and rigid (torsion force constant =1.905) torsion parameter ($^3J_{H5'-P}$ & $^3J_{H5''-P}$). A 1.0-μs simulation for each tetramer was used for the testing.

(A)

Torsion angles (P-O-C-C)	MEP (QM)	Statistic (kT×lnρ)	β torsion constant k_2= 0 kcal/mol	β torsion constant k_2= 1.905 kcal/mol
180	0	0	0	**0**
150,210	0.355	1.142	0.435	**0.910**
120,240	0.752	1.847	0.746	**2.175**
90,270	0.907	2.348	0.894	**2.799**
60,300	2.574	3.262	3.361	**4.790**

(B)

	CAAU (1.905)	CAAU (0)	CAAU (NMR)	AAAA (cal)	AAAA (0)	AAAA (NMR)	GACC (cal)	GACC (0)	GACC (NMR)
β_2	3.1, 3.0	7.4,5.9	3.7, 2,2	3.4, 2.1	3.5,2.1	3.8, 1.0	3.9, 1.8	4.1,1.8	3.7, 0.9
β_3	3.1, 2.7	4.7,2.2	3.5, <1	2.7, 2.5	2.8,2.5	3.0, 1.0	3.1, 2.2	3.4,2.1	4.0, 2.0
β_4	5.4, 3.6	10.5,7.8	3.8-4.3, 3.3	4.4, 2.9	6.0,5.9	3.2, 1.0	8.1, 6.2	10.1,8.7	4.4, 2.0

Part II: DNA/RNA simulations

Table S3. Simulation box content and size for DNA and RNA simulations.

(A) Simulations for DNA/RNA in water solution

Type	PDB	Water/Na+/Cl- number	Cubic box side length for NVT simulation (Å)
RNA double chain (Duplex)	2JXQ	8323/34/16	63.577
	1MIS	6178/26/12	57.649
	1F5G	9216/36/18	65.534
	2L8F	11220/42/22	69.807
	1RNA	10030/46/20	72.900
RNA single chain (Hairpin)	2KOC	6327/25/12	57.956
	1ZIH	5758/22/11	56.389
	1SZY	10804/41/21	69.360
DNA double helices in water	1D42	9219/16/2 (A-form) 6984/16/2 (B-form)	66.950 (A-form) 61.408 (B-form)
	1D20	8054/36/18 (A-form) 8040/36/18 (B-form)	62.700 (A-form) 62.700 (B-form)

	1NAJ	8423/40/18 (A-form)	63.712 (A-form)
		8234/40/18 (B-form)	63.250 (B-form)
	2HKB	7966/40/18 (A-form)	62.550 (A-form)
		8001/40/18 (B-form)	62.650 (B-form)
	1D23	7209/32/14(B-form)	65.493 (B-form)

(B) DNA Duplex in Ethanol/Water Mixture

PDB	Ethanol/Water/Na+/ Cl-number	Cubic box side length for NVT simulation (Å)
1D42	2053/654/16/2 (A-form)	60.400 (A-form)
	2053/654/14 (B-form)	60.400 (B-form)
1D20	2409/781/18 (A-form)	63.720 (A-form)
	2415/783/18 (B-form)	62.700 (B-form)
1NAJ	2464/810/22 (A-form)	64.345 (A-form)
	2459/807/22 (B-form)	64.296 (B-form)
2HKB	2366/770/22 (A-form)	63.400 (A-form)
	2359/785/22 (B-form)	63.400 (B-form)

(C) Crystal simulations

Type	PDB	Cell	NA/Water/ions	Box size (Å)
Z-form DNA	1LJX	unit	1 DNA/359 water /24 Na+	21.182×28.363×44.440
	292D	unit	1DNA/339 water /16 Na+	17.940×31.230×44.550
B-form DNA	1D23	unit	4 DNA/1072 water/36 Mg2+	38.930×39.630×33.300
	1D23	2×2×2	32 DNA/8578 water/288 Mg2+	77.860×39.630×33.300
RNA	1RNA	unit	4 RNA/1537 water/104 Na+	34.110×44.610×49.110
	1RNA	2×2×2	32 RNA/12300 water/832 Na+	68.220×89.220×98.220

Table S4. Simulation efficiency data of DNA and RNA systems on GPU.

All simulations were performed with a 3-fs time step and heavy hydrogen, polarization convergence of 10^{-4} Debye/atom, using NVIDIA GTX1070. For comparison, the simulation speed for the DHFR system (23558 atoms) with 3-fs time step and AMBER force field on Nvidia GTX 1070 is ~300 ns/day.

DNA/RNA Systems	Number of atoms	Simulation speed (ns/day)
rCAAU	11381	21.9
2KOC	19465	13.8
2JXQ	25661	10.6
TAR	63128	4.8
1NAJ	25518	10.9
1D42	21478	12.6
1D20	24807	10.9
2HKB	24818	10.5
1D23 (unit cell)	5780	30
1RNA (super cell)	65892	3.5

Figure S10. RMSD of RNA simulation structures was calculated by comparing with the first NMR structure.

Terminal residues are not included.
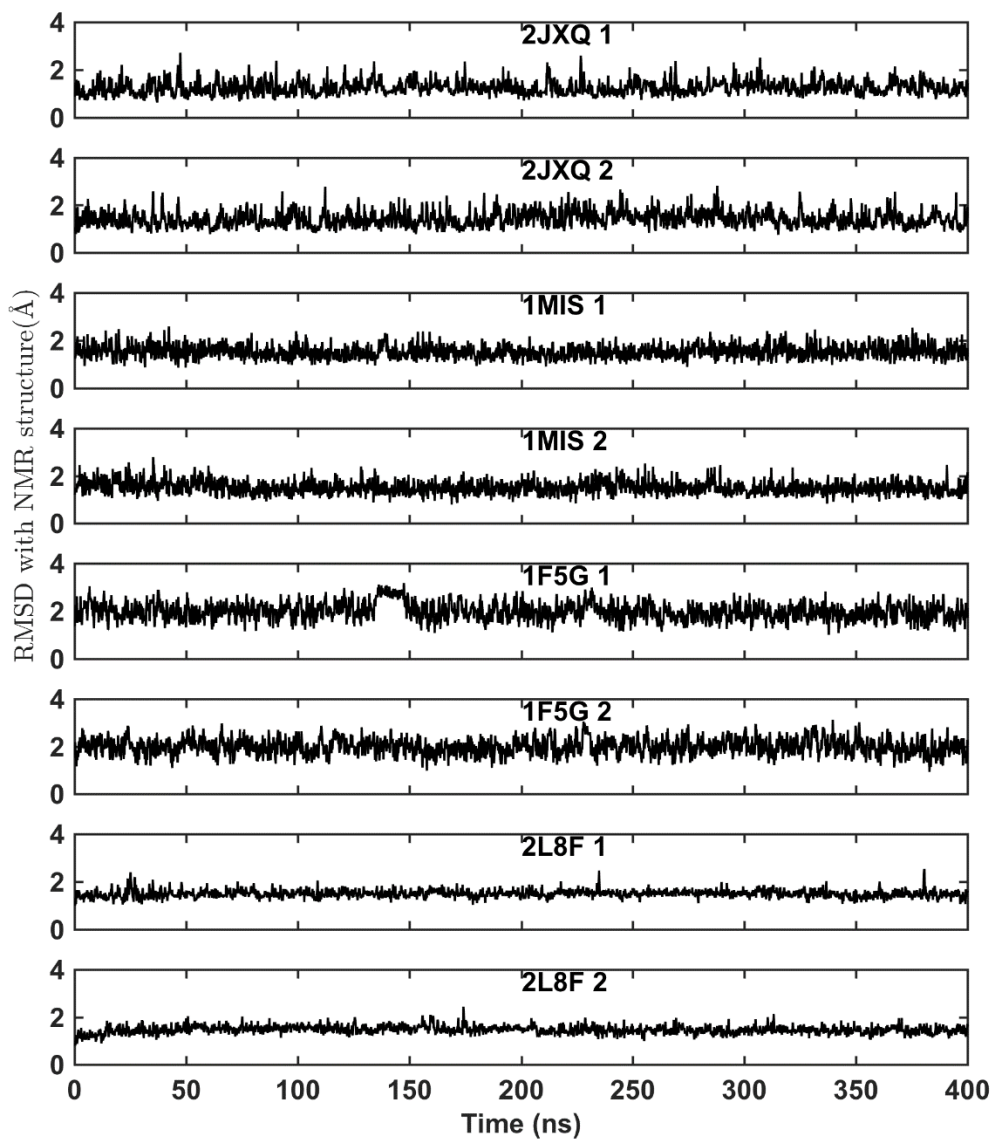
(A) RNA Duplex
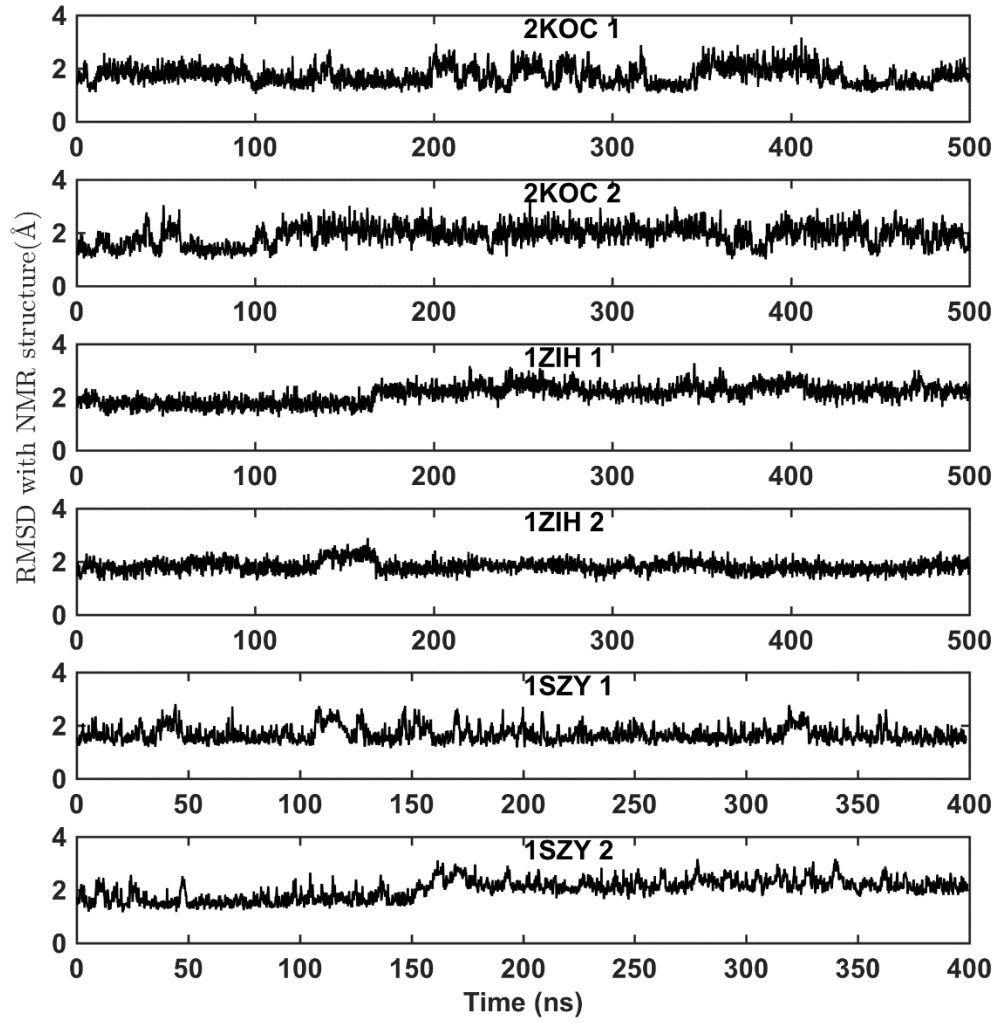
(B) RNA hairpin loops

Figure S11. Non-terminal heavy atom RMSD of DNA duplexes with respect to typical A-form structure (blue) and B-form structure (red) in aqueous solution.

(A) The simulation started with A-form structure. (B) The simulation started with B-form structure.

(A)

(B)

1D42 (B-form)

1NAJ (B-form)

1D20 (B-form)
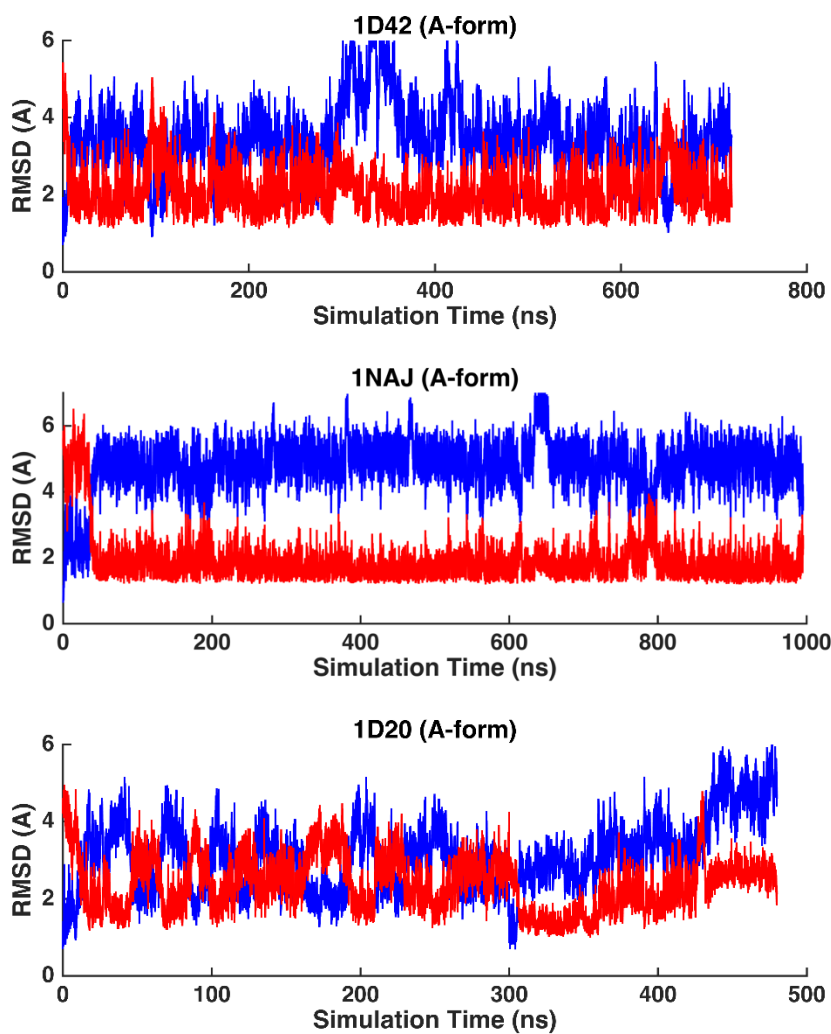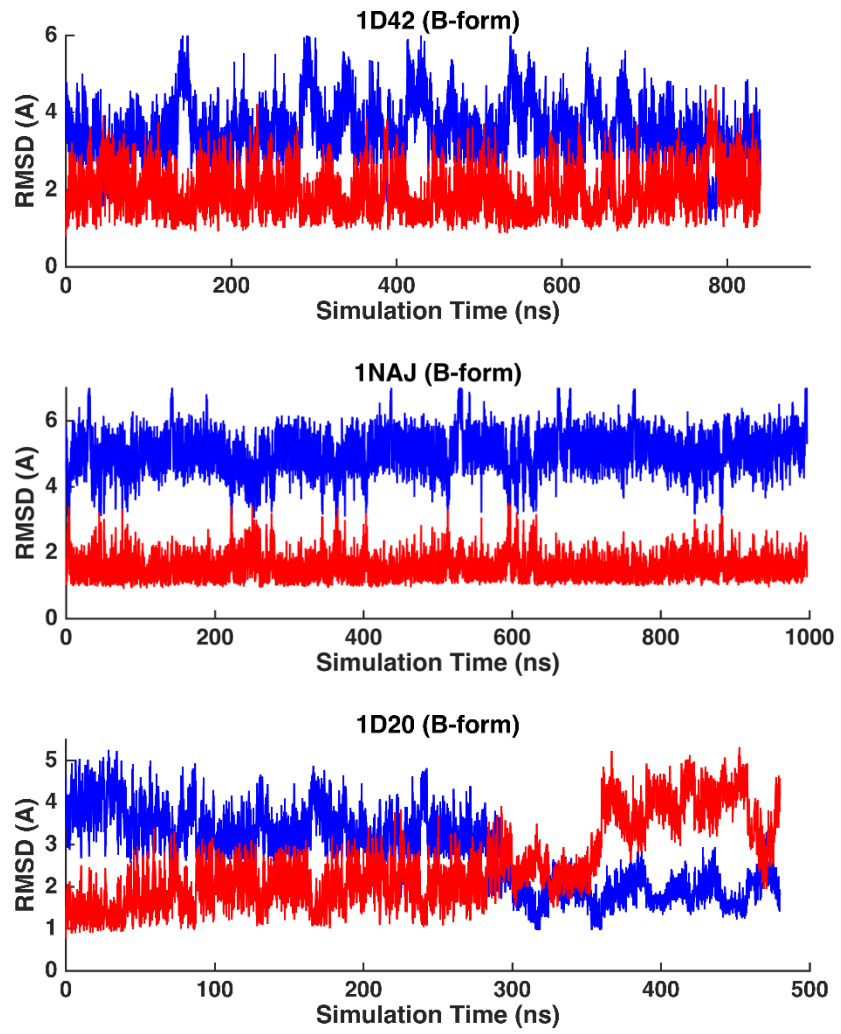
Figure S12. Non-terminal heavy atom RMSD of simulated DNA duplexes with respect to typical A-form structure (blue) and B-form structure (red) in ethanol-water mixture.

1D20 (A-form/328K)

1D20 (B-form/328K)

1D20 (A-form/298K)

1D20 (B-form/298K)
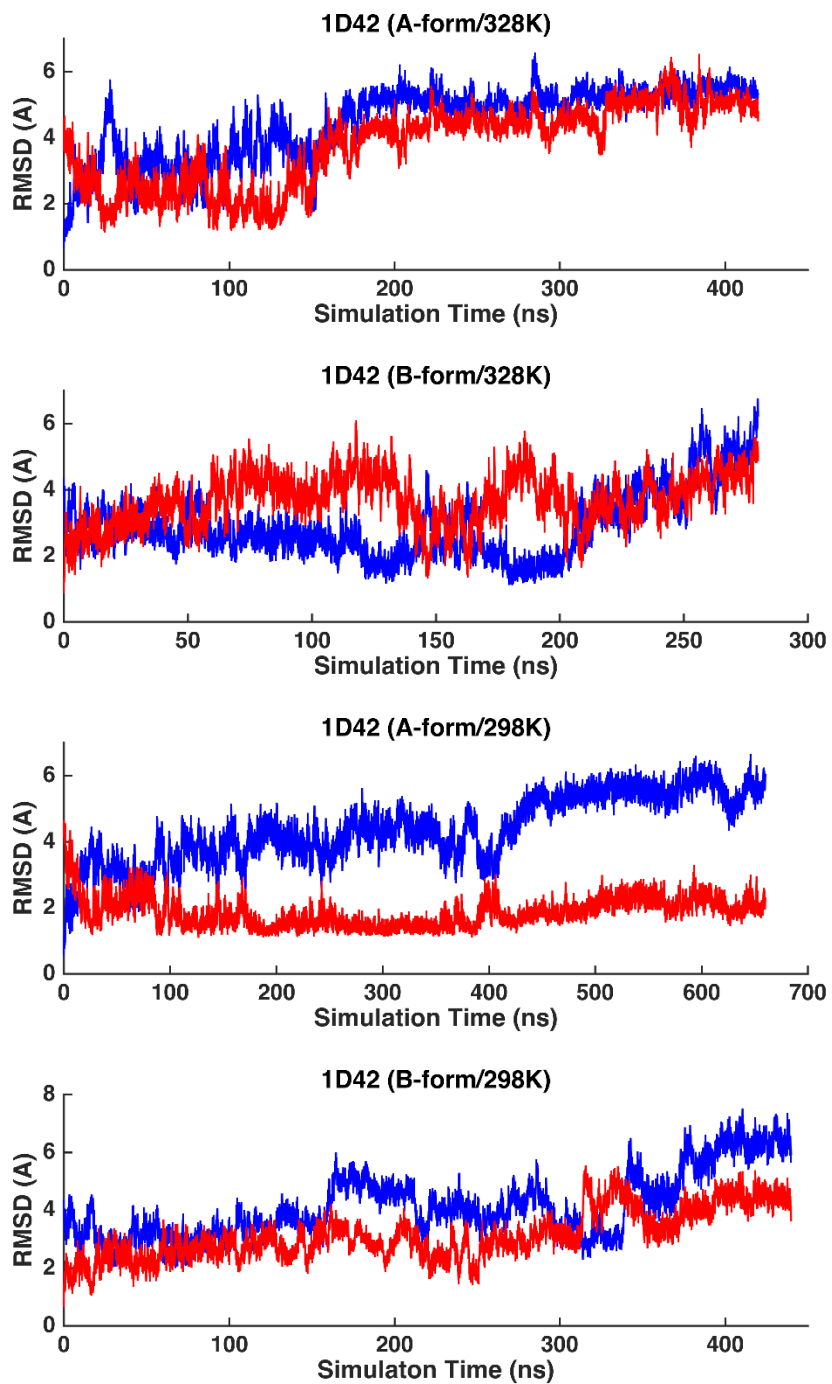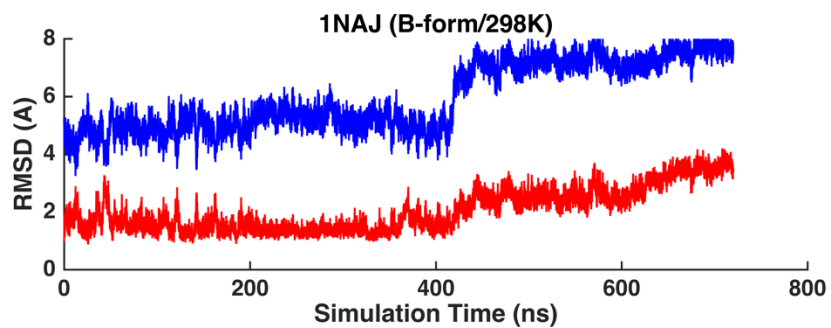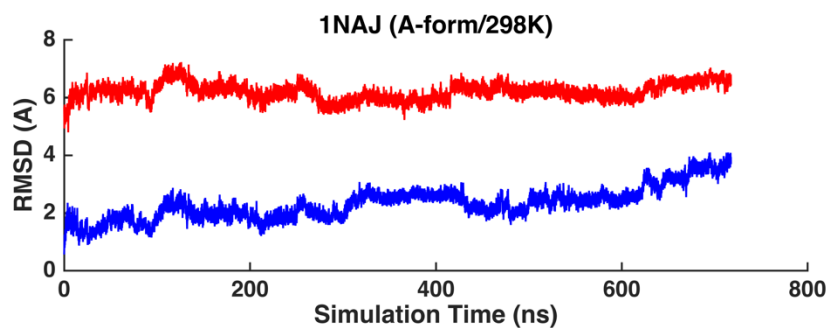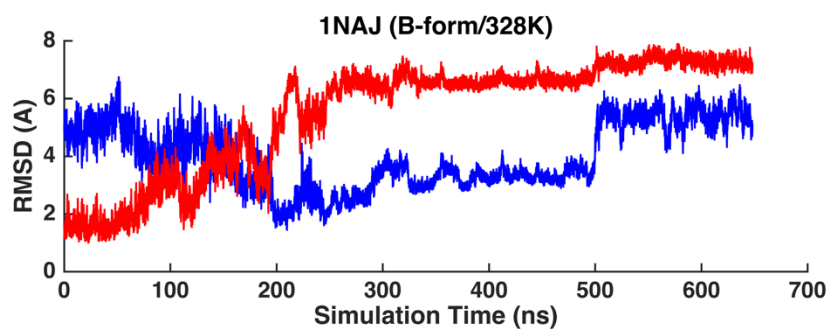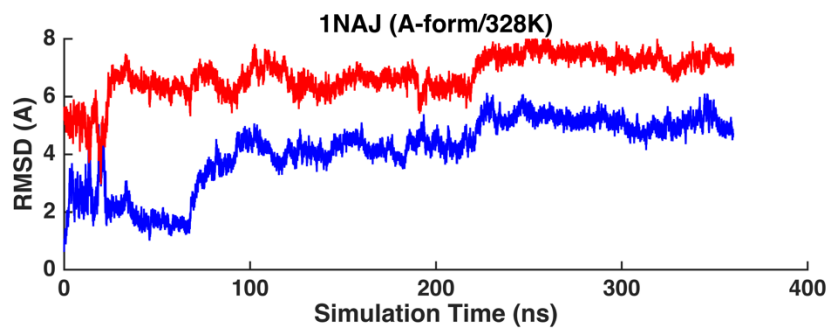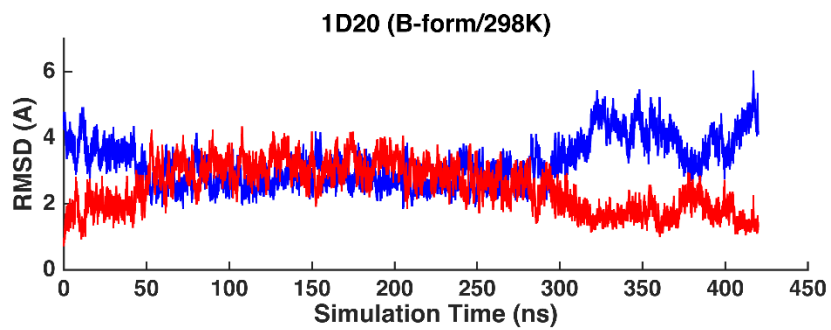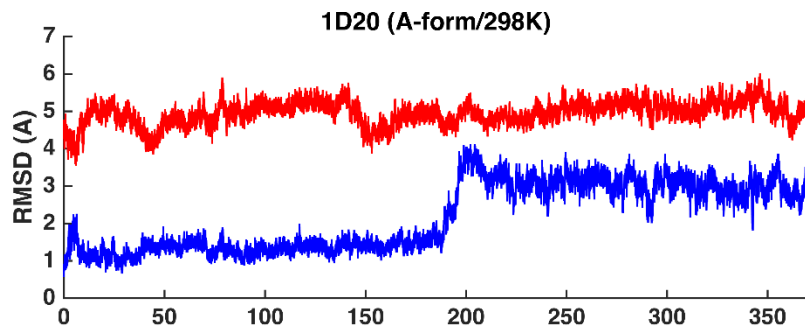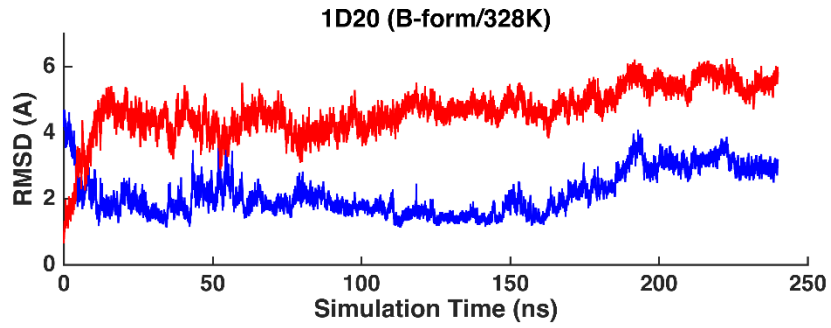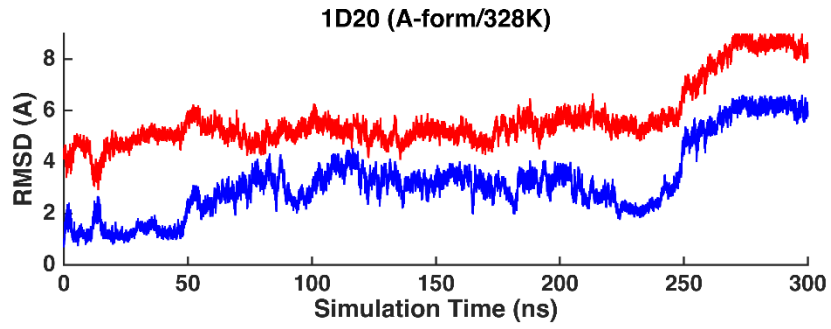
Figure S13 Comparison of Curves+ helicoidal parameters along the nucleotide position for DNA (1NAJ, 2HKB, 1D42, 1D20) and RNA (2JXQ) duplexes.

The parameters were computed using structures from solution-phase simulations (black) and NMR structures (red). See the method for Curves+ calculations in reference 109.

(A)

(B)

(C)

(D)

(E)
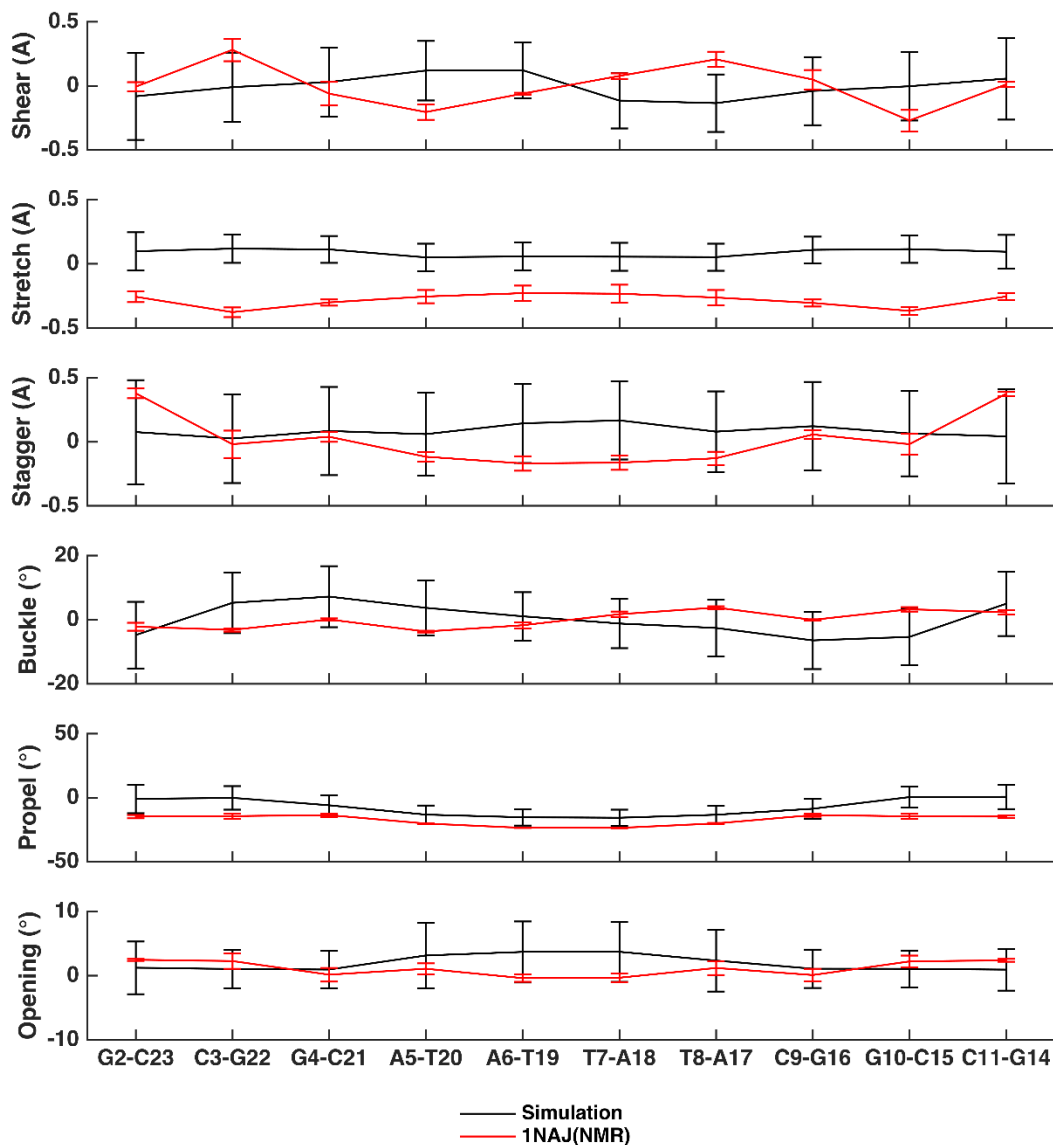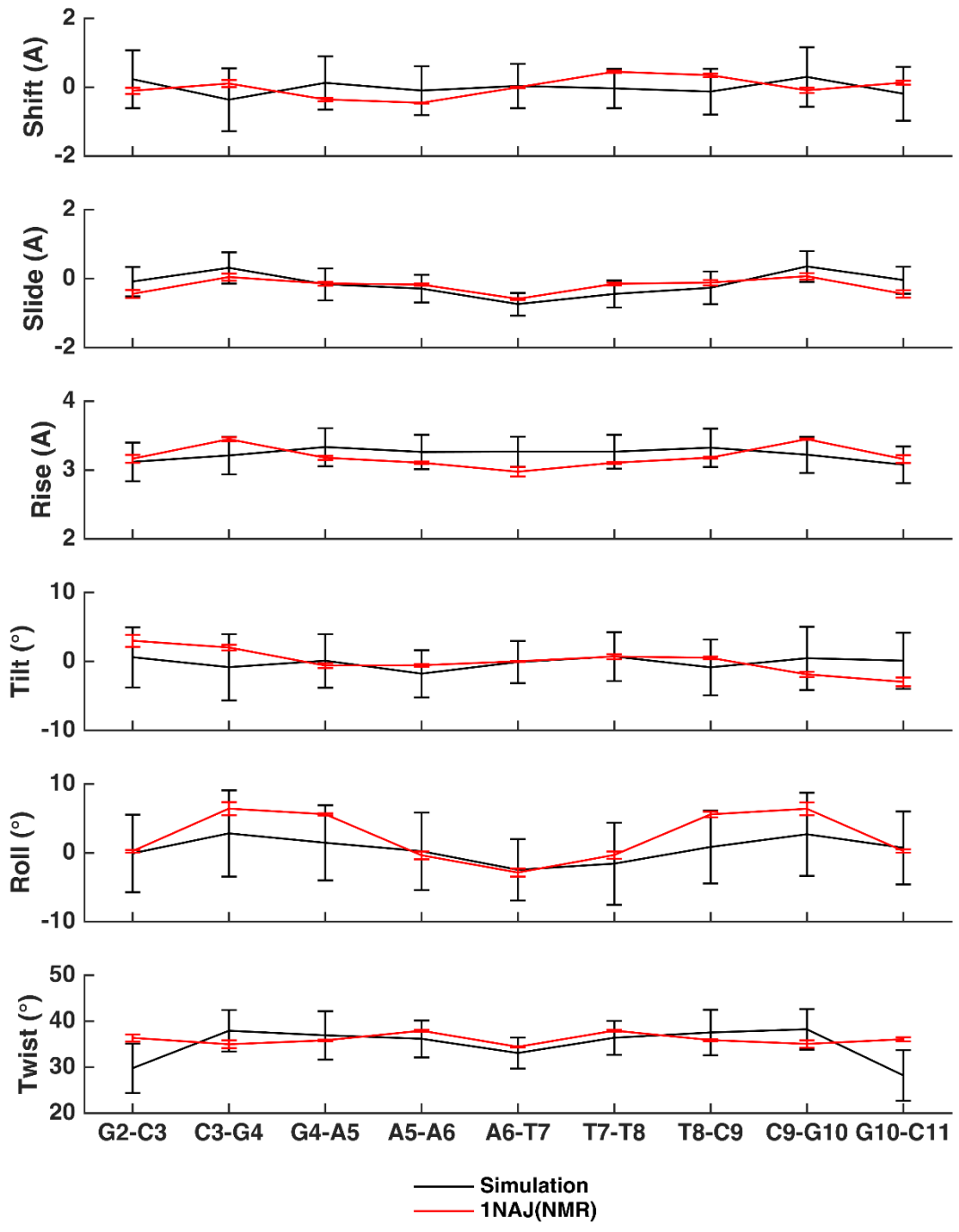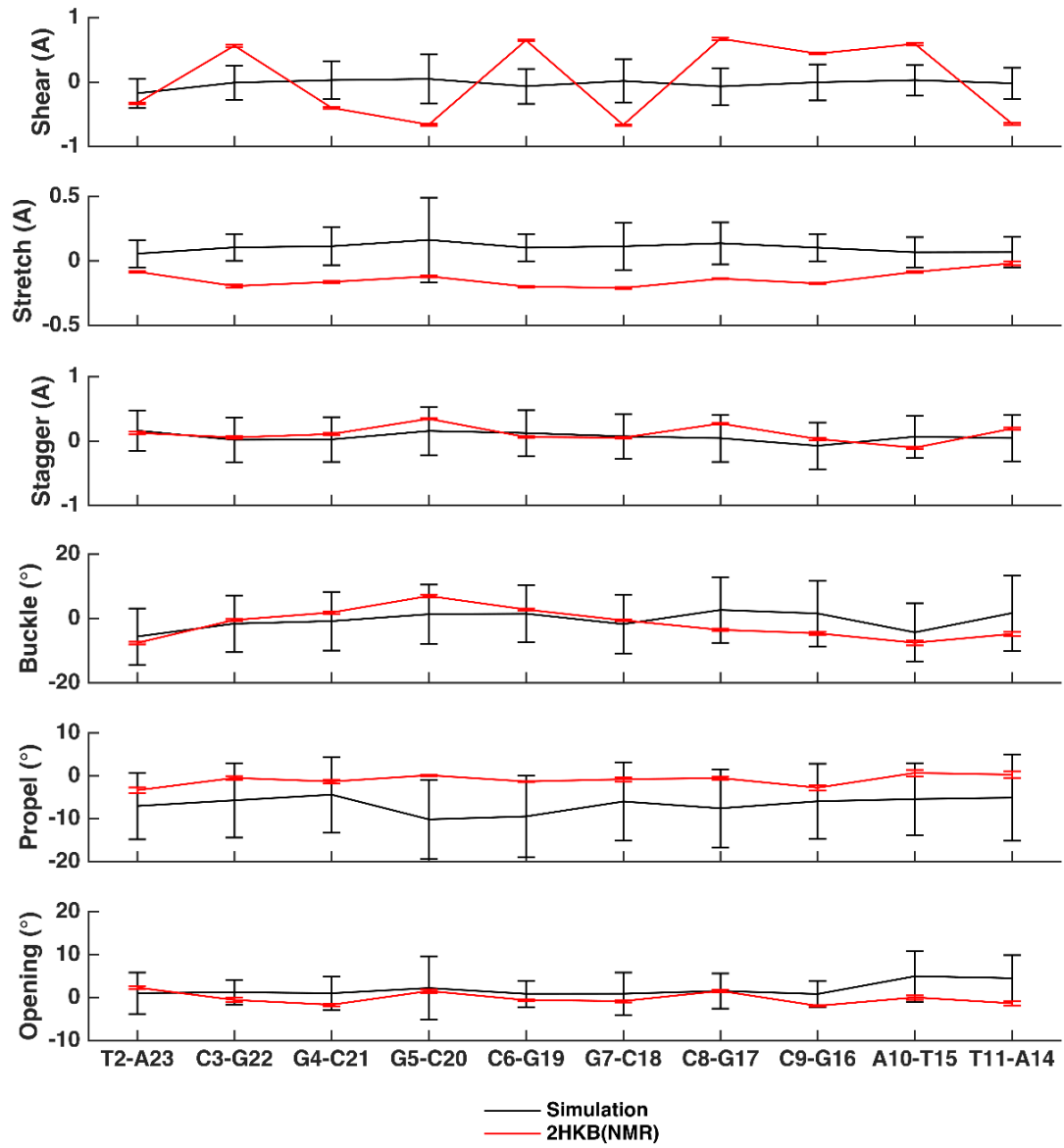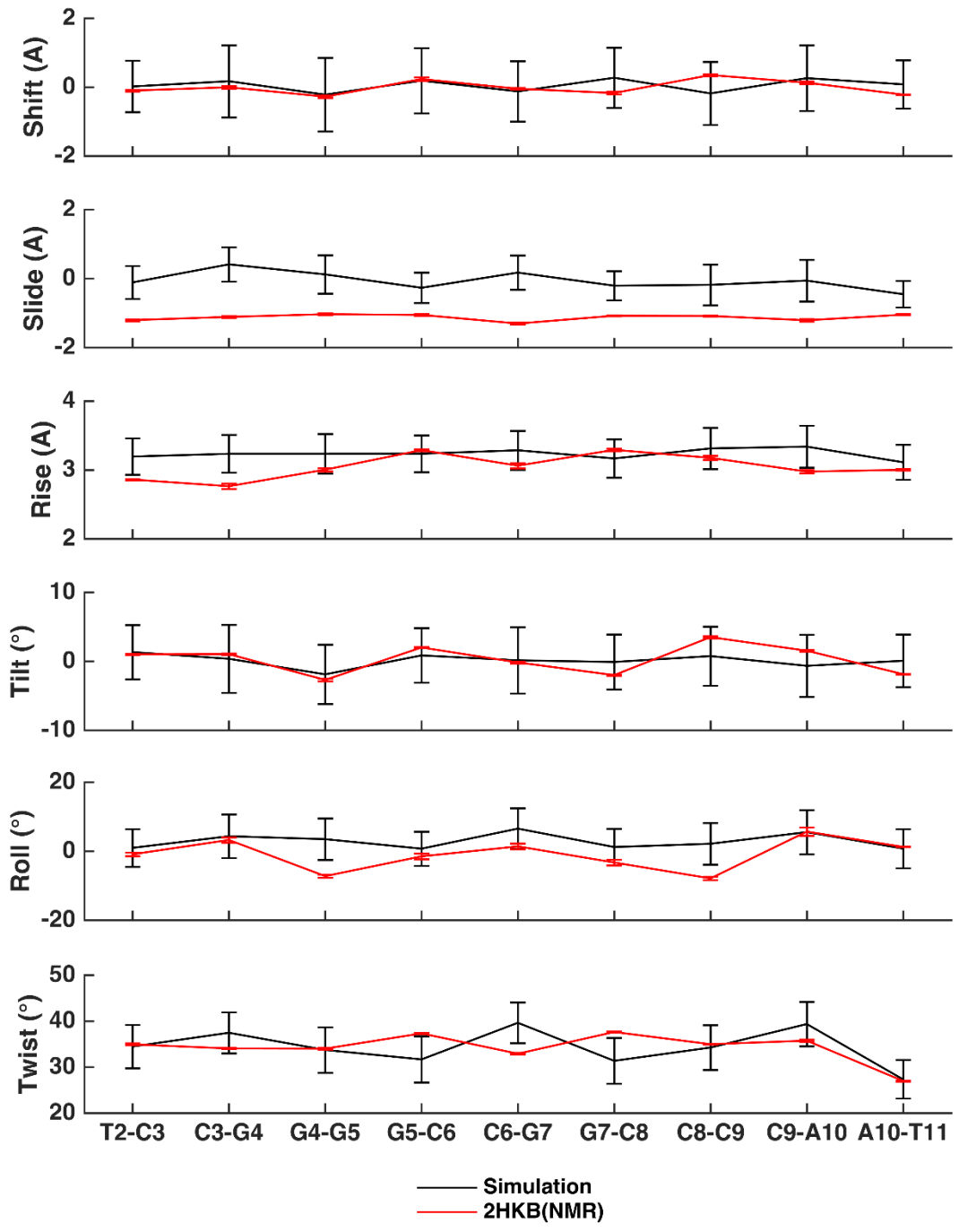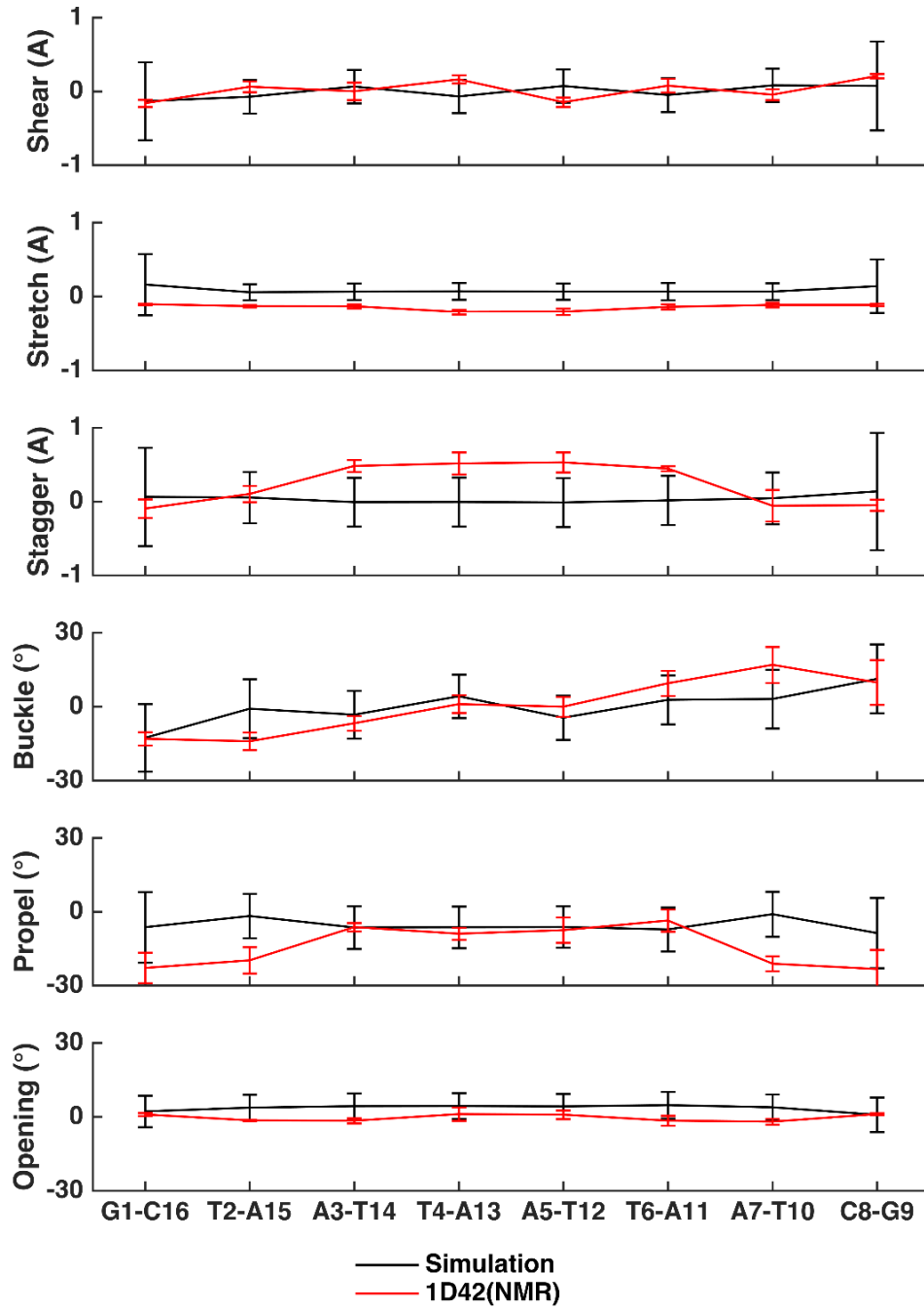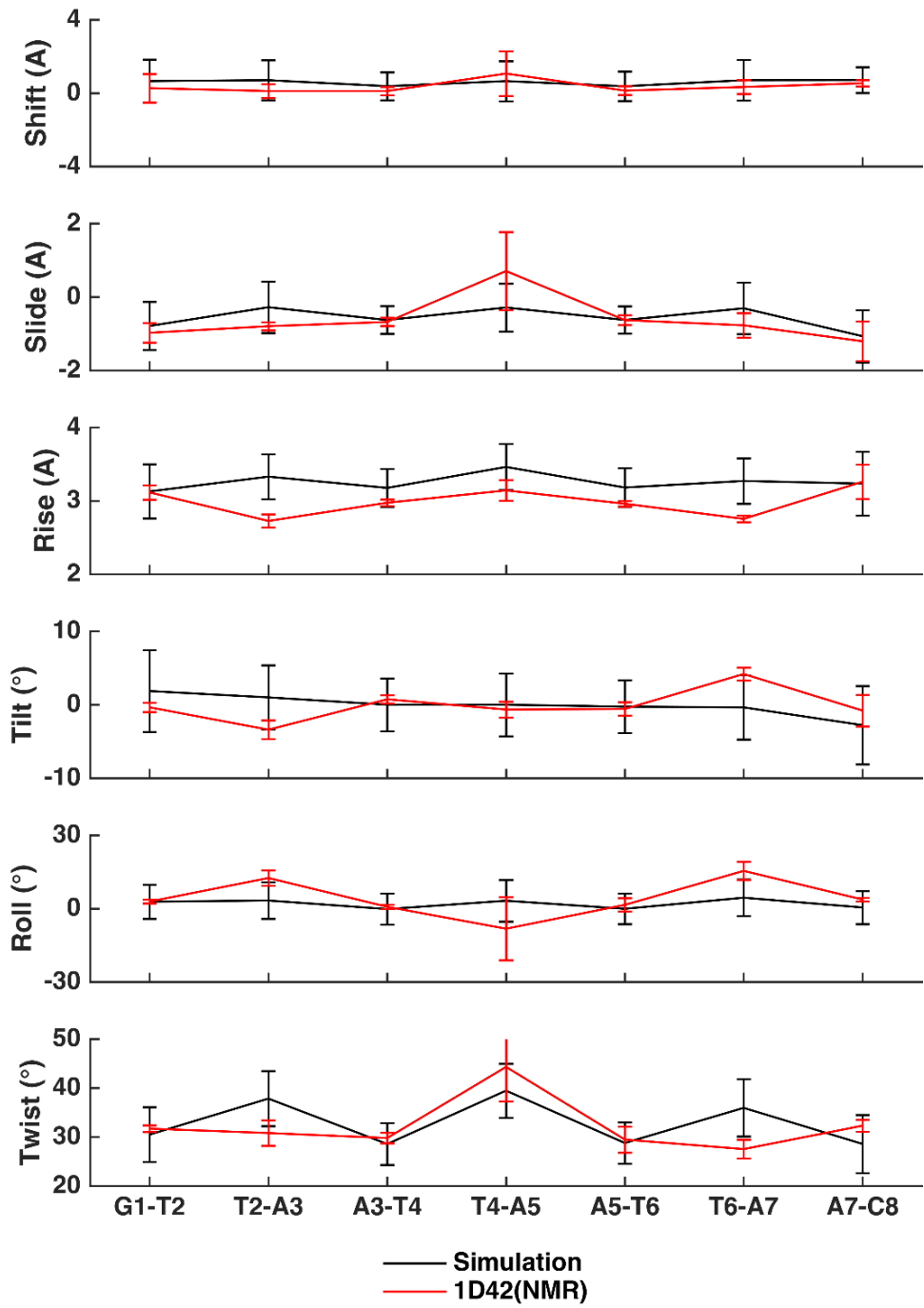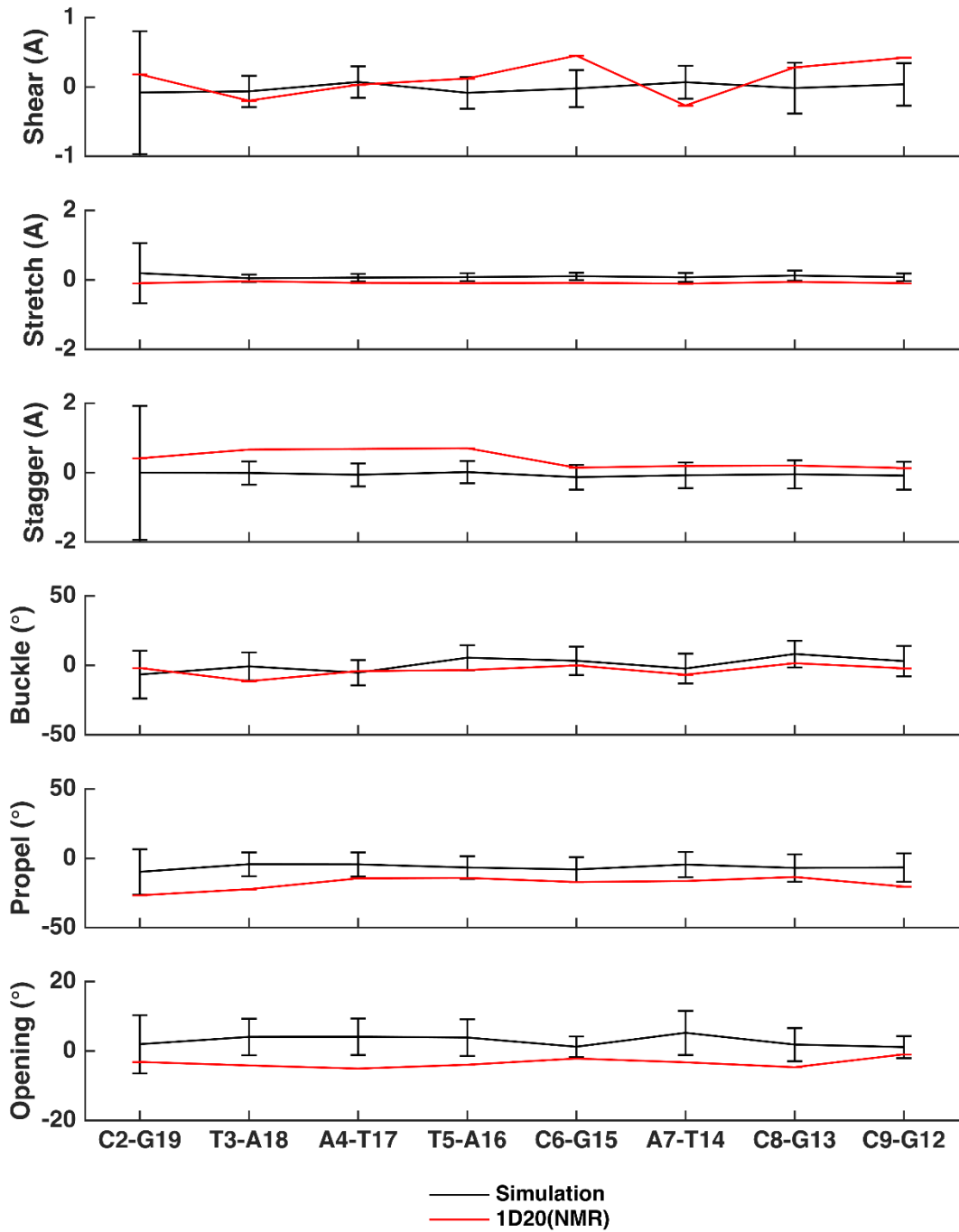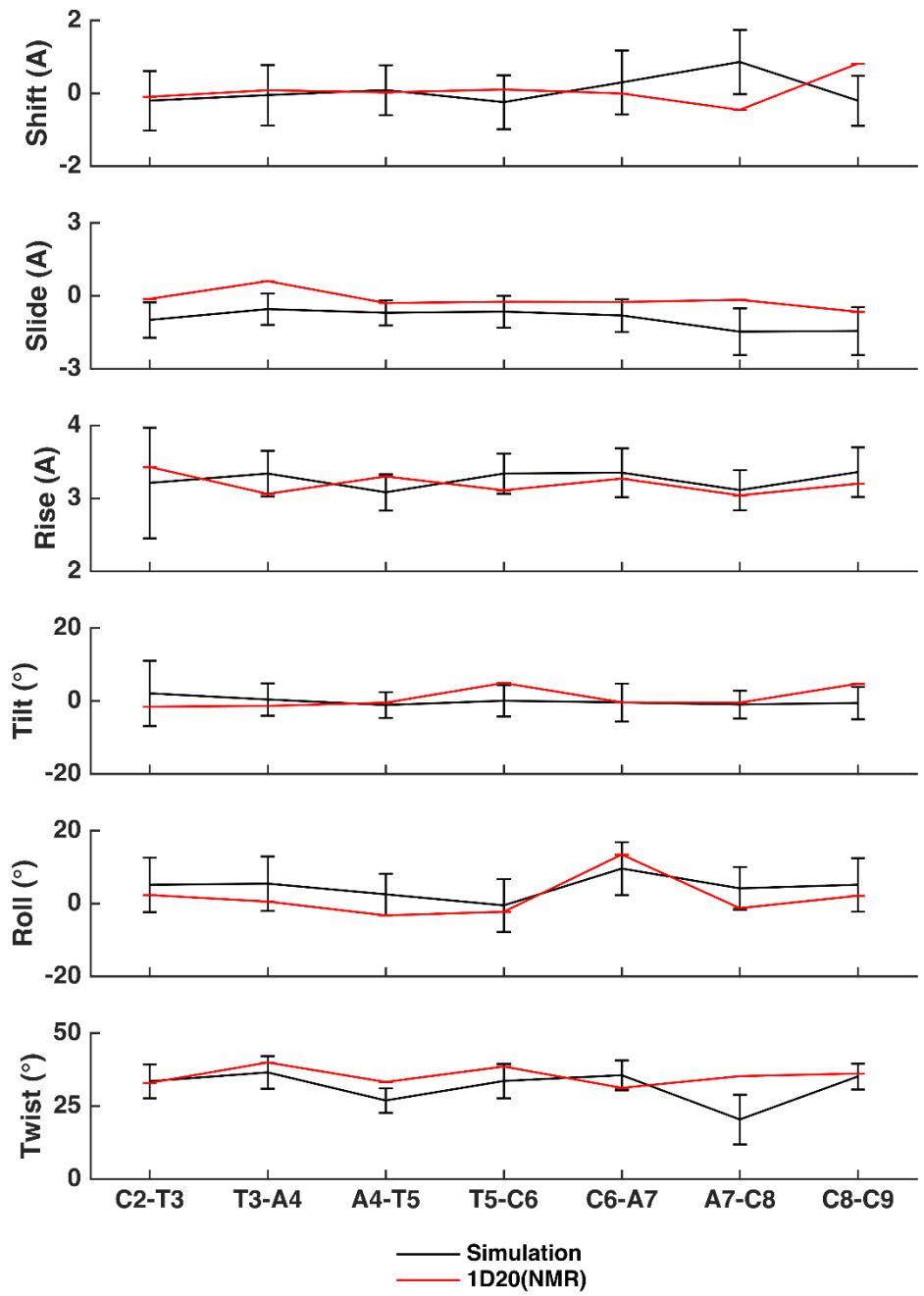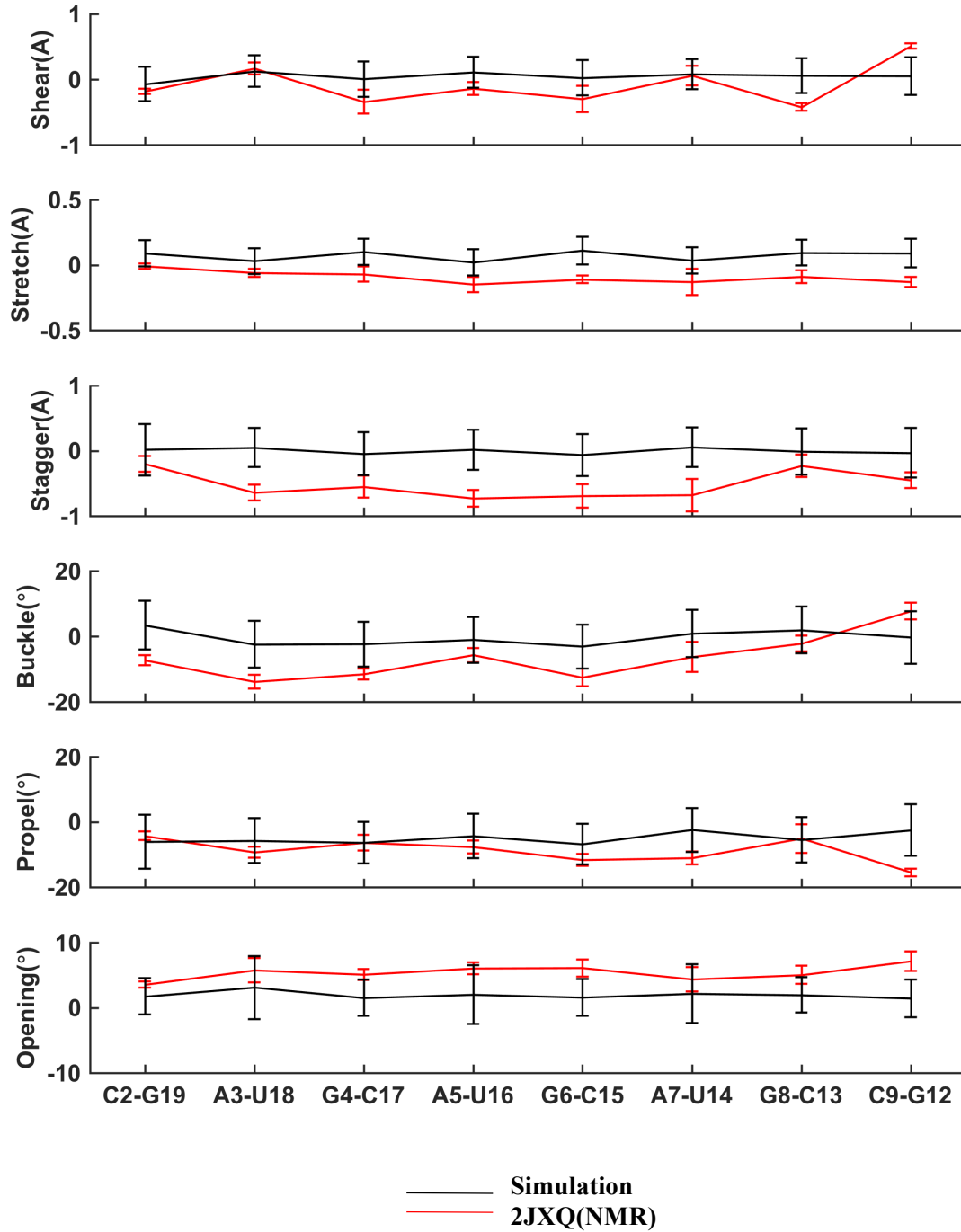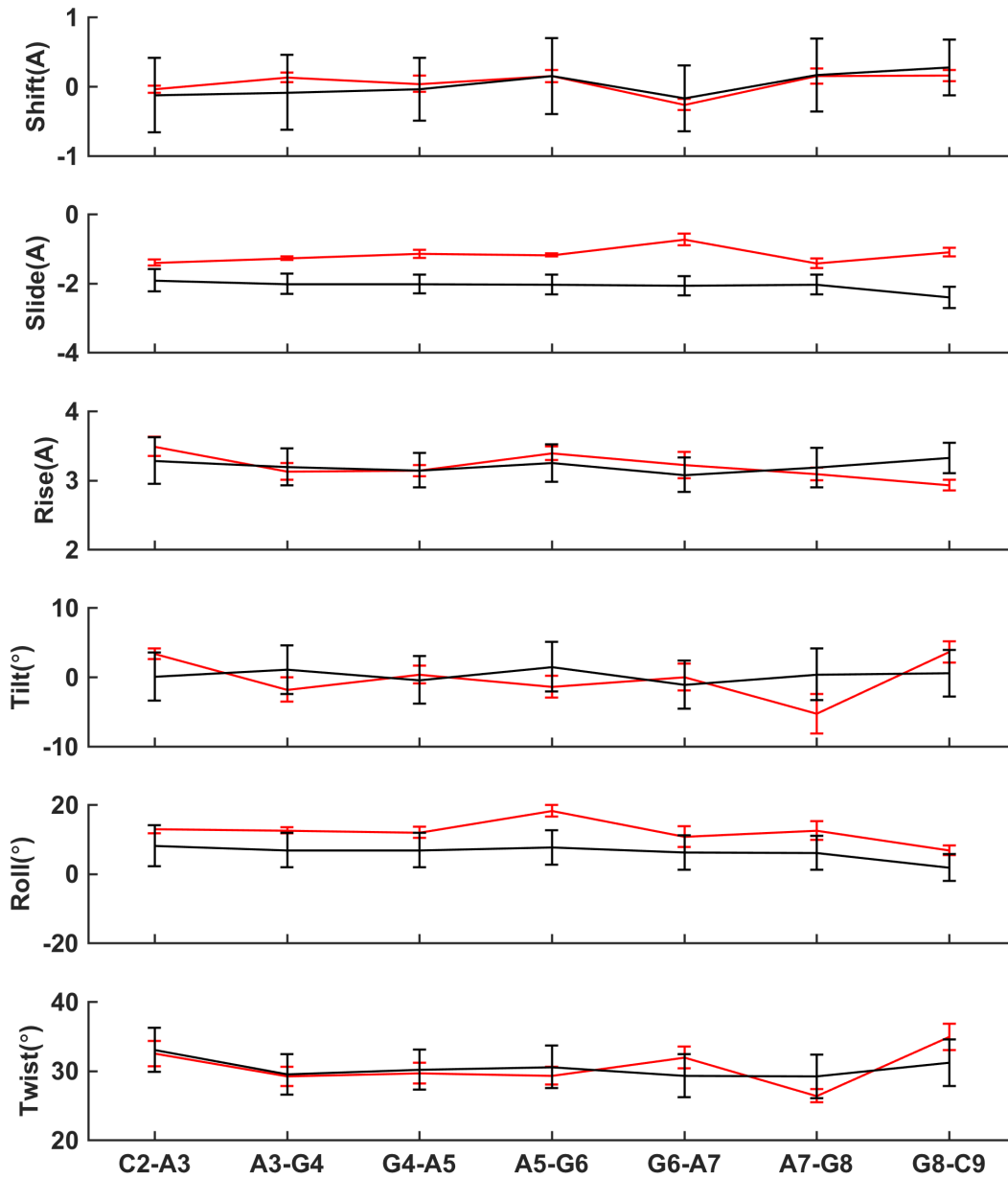
Figure S14 Comparison of Curves+ groove widths of 1NAJ, 2HKB and 2JXQ along nucleotide position between solution-phase simulation and NMR.

(A)
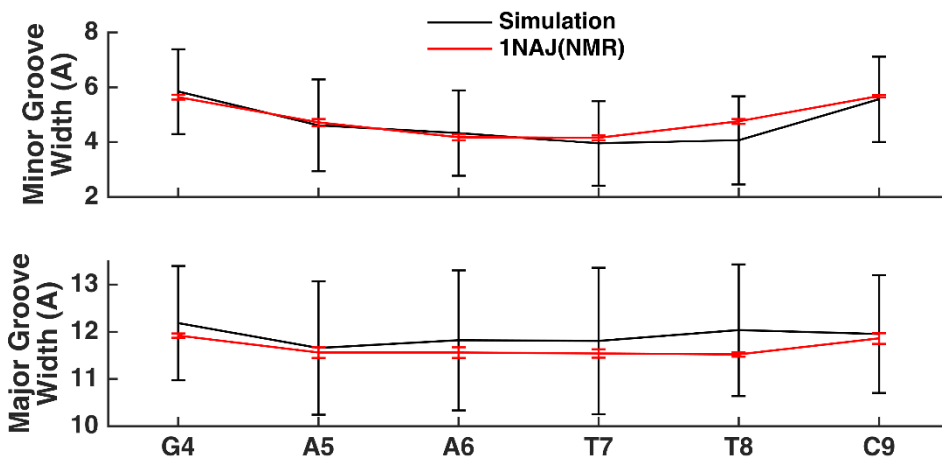
(B)
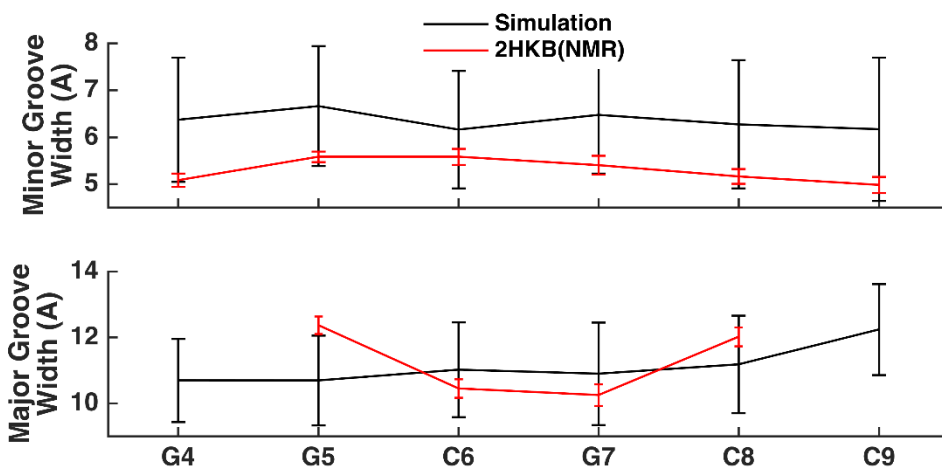
(C)
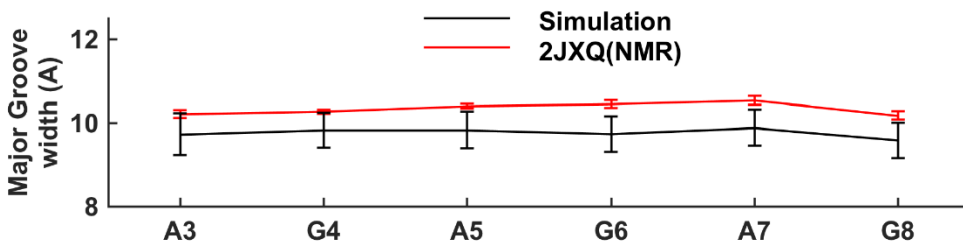
Figure S15 Comparison of Curves+ helicoidal parameters along the nucleotide position between crystal simulation, solution simulation and X-ray structures for DNA (1D23) and RNA (1RNA)

(A) 1D23
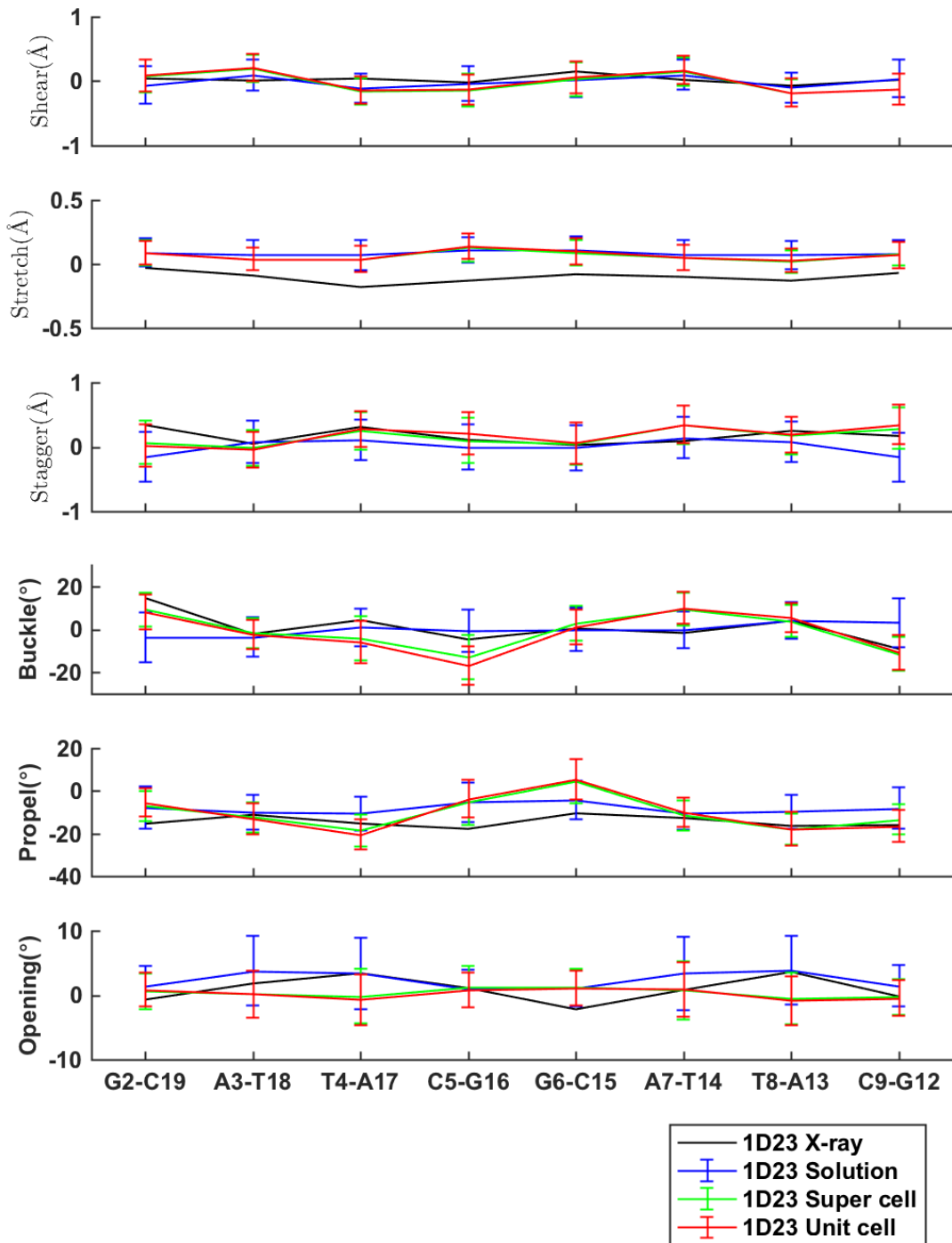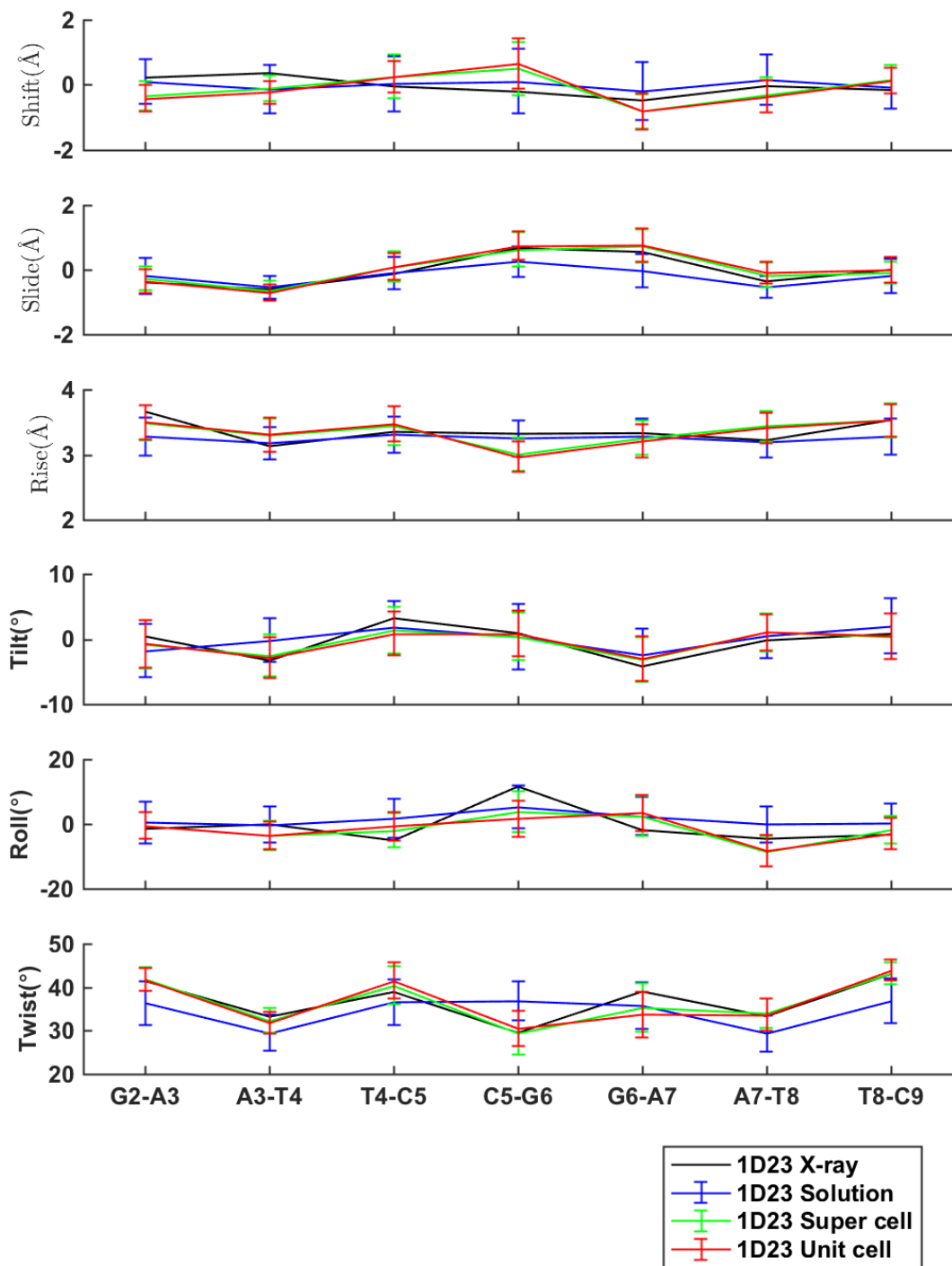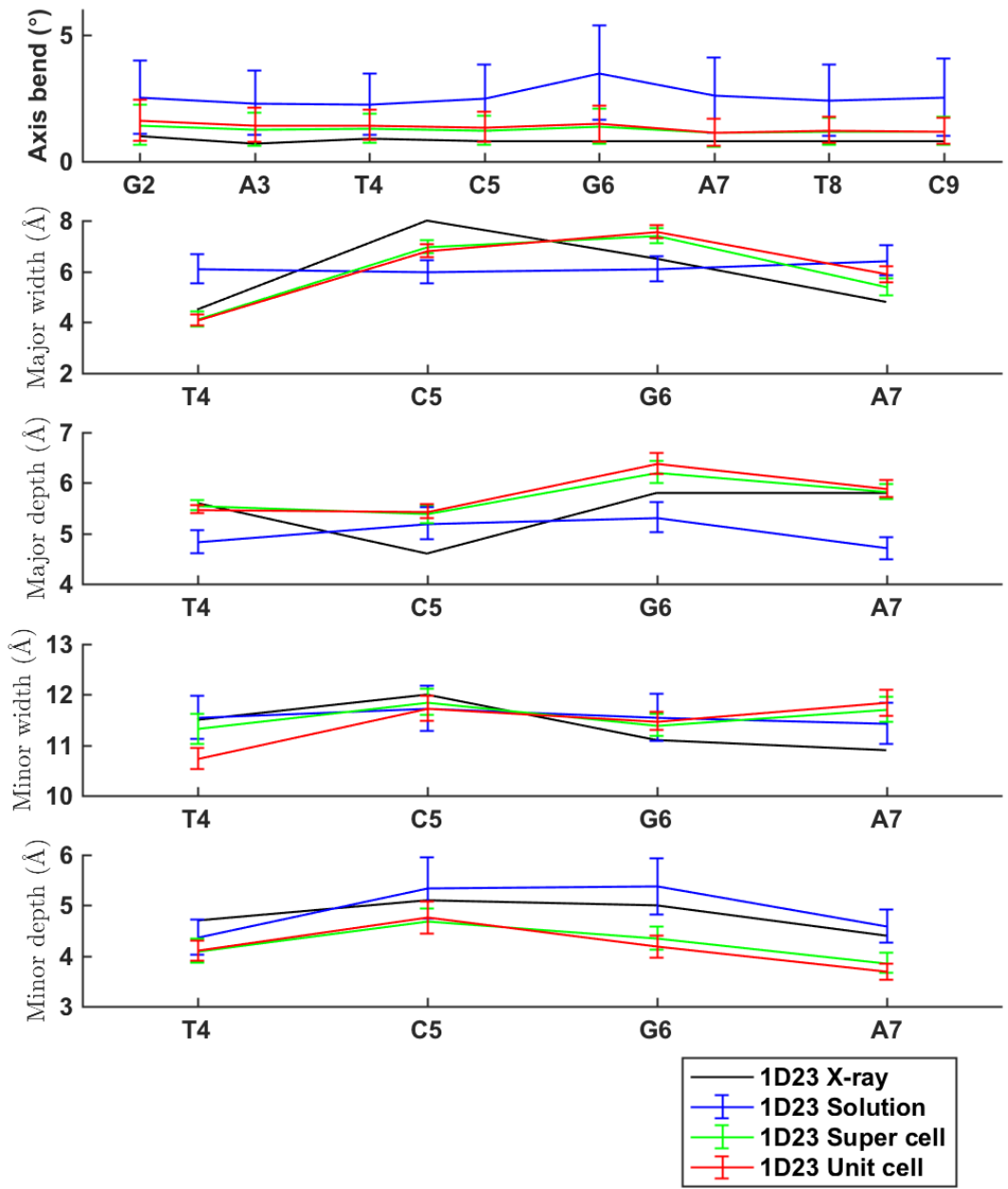
(B) 1RNA
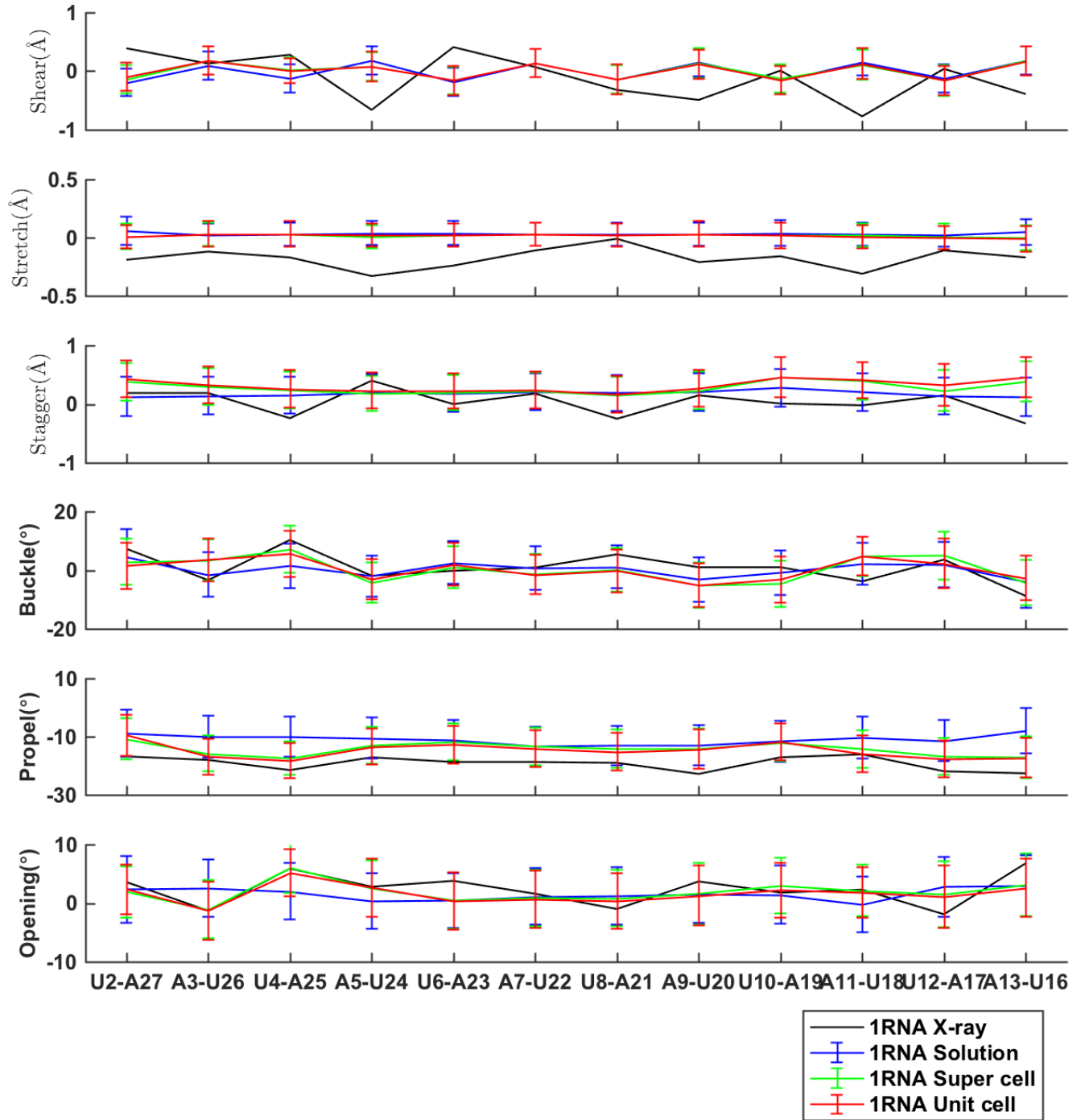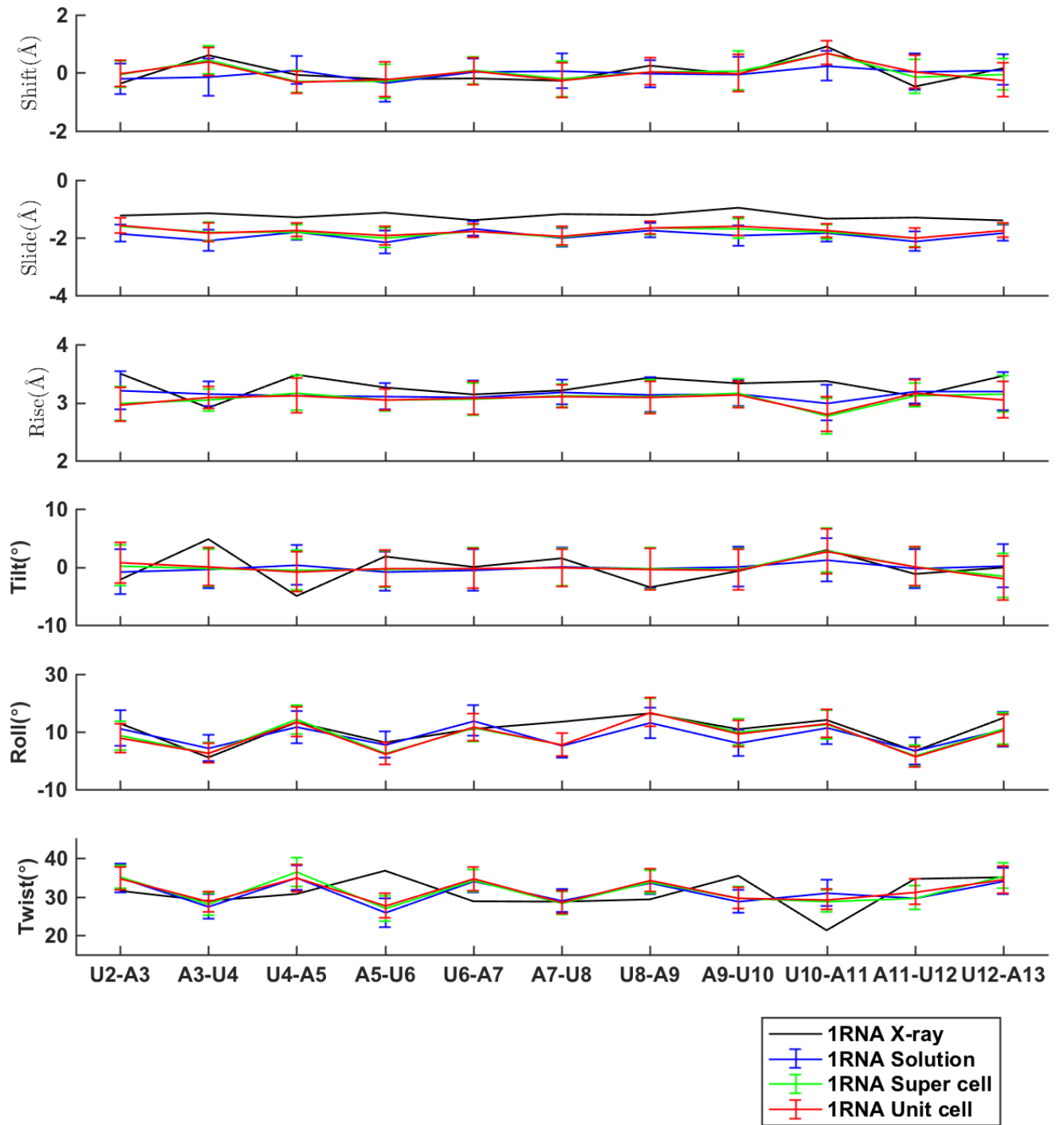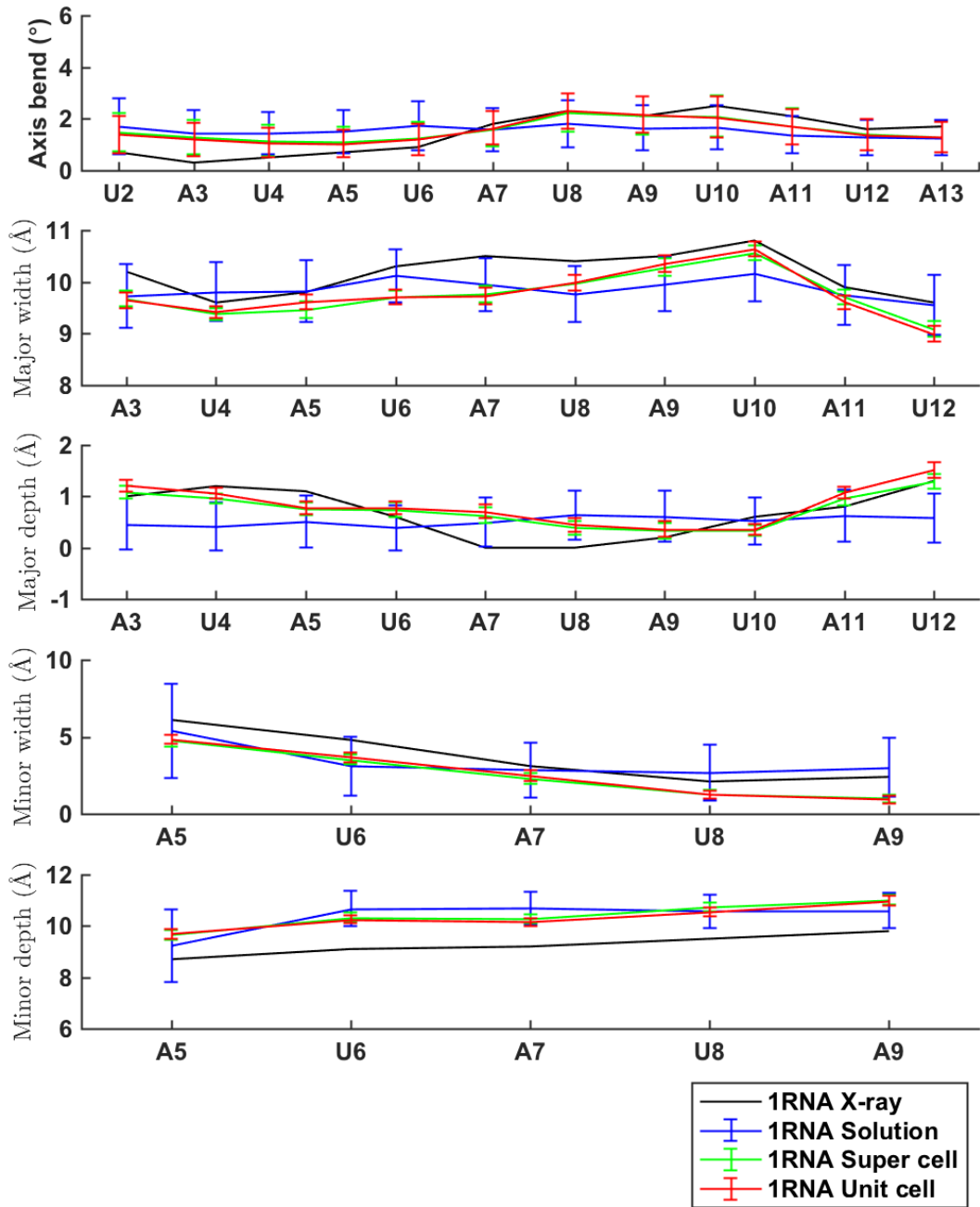
Figure S16. The stability of the RNA terminal base-pairs or the capping base pair of hairpin stem.

Most of these pairs are GC pair, and the distance of GN1 and CN3 atoms were used for detecting the stability of the pairs. If there is only one peak at ~3 Å, that shows the pair is quite stable, and never break up in the simulation.
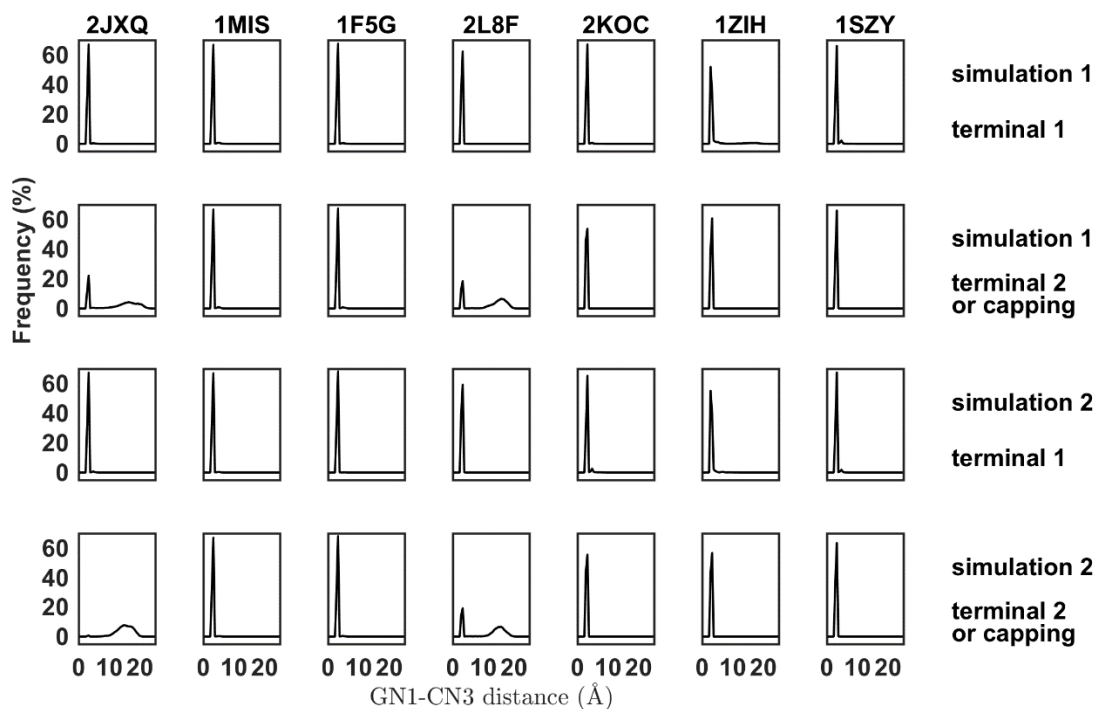
Figure S17 Terminal base-pair breakup and reforming in 2JXQ and 2L8F simulation.
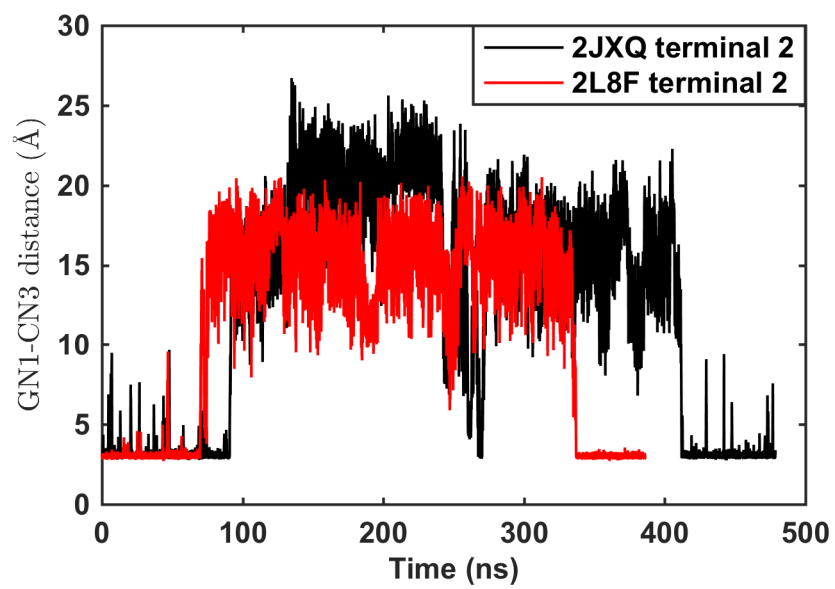
Figure S18. Comparison of the average values of RNA backbone torsions and χ torsion between simulated and NMR structures.

NMR data show in black circle. Simulation data of the two independent trajectories are in red and blue lines with standard deviation. The average unsigned errors are included for each torsion.

(A) 2JXQ
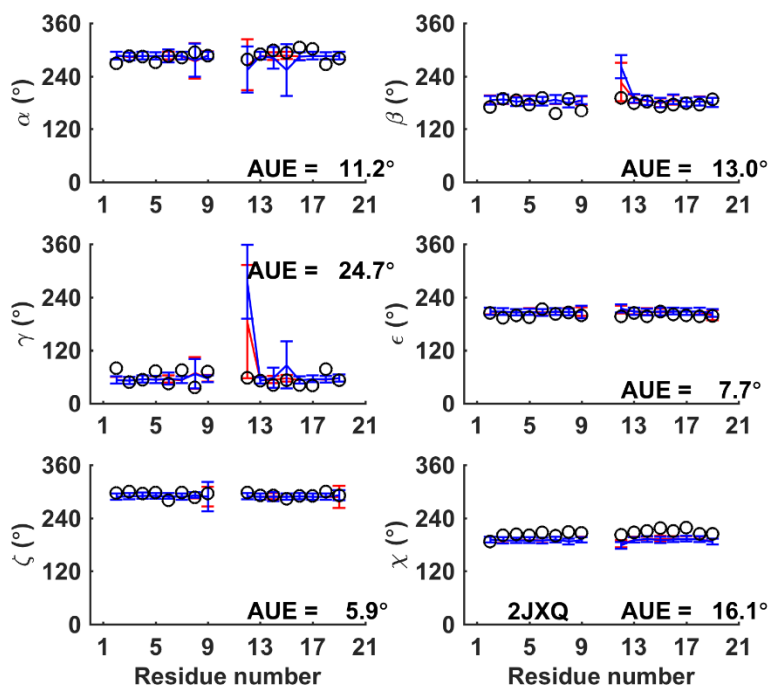
(B) 1MIS

(C) 1F5G

(D) 2L8F

(E) 2KOC

46
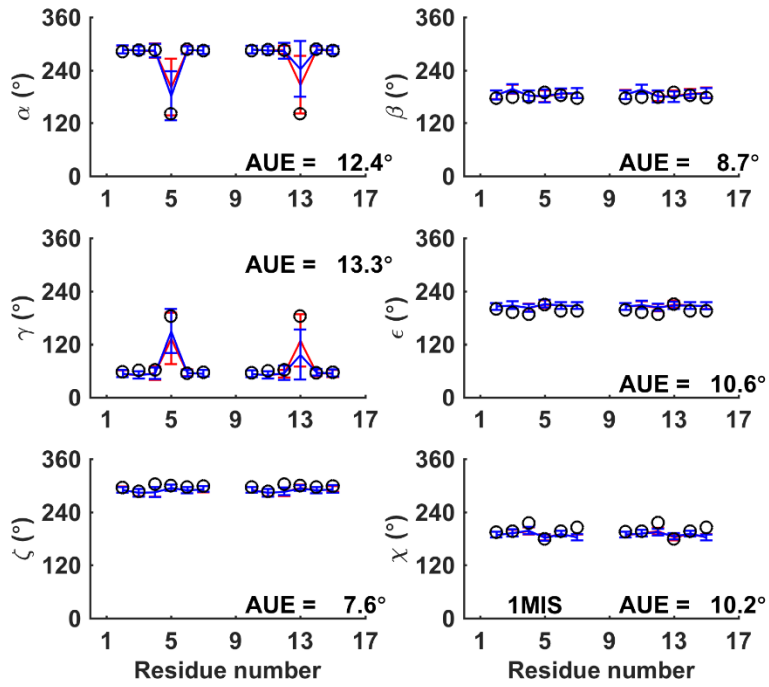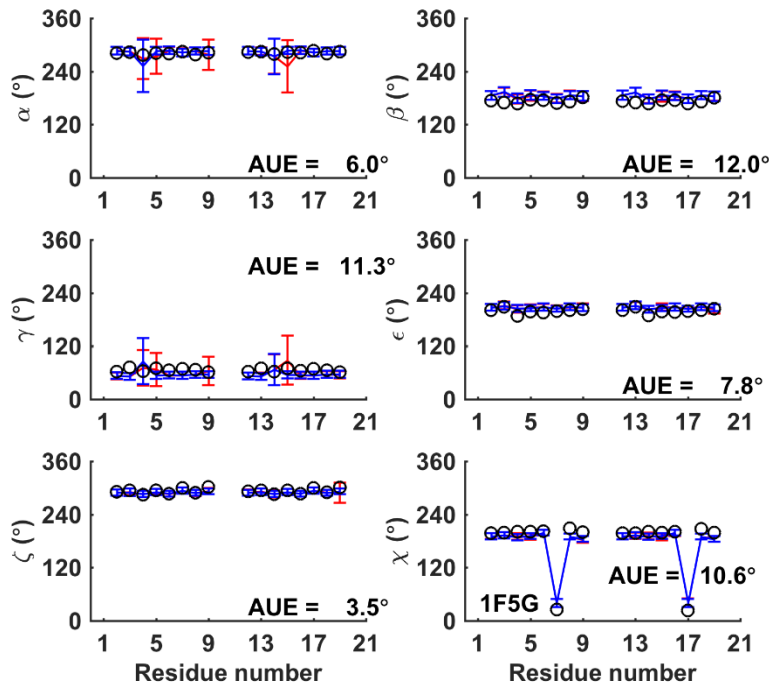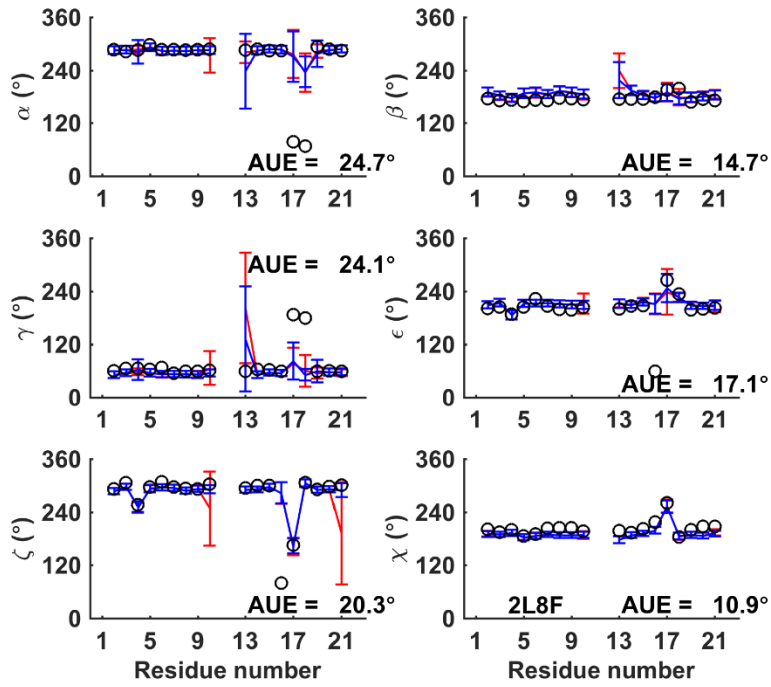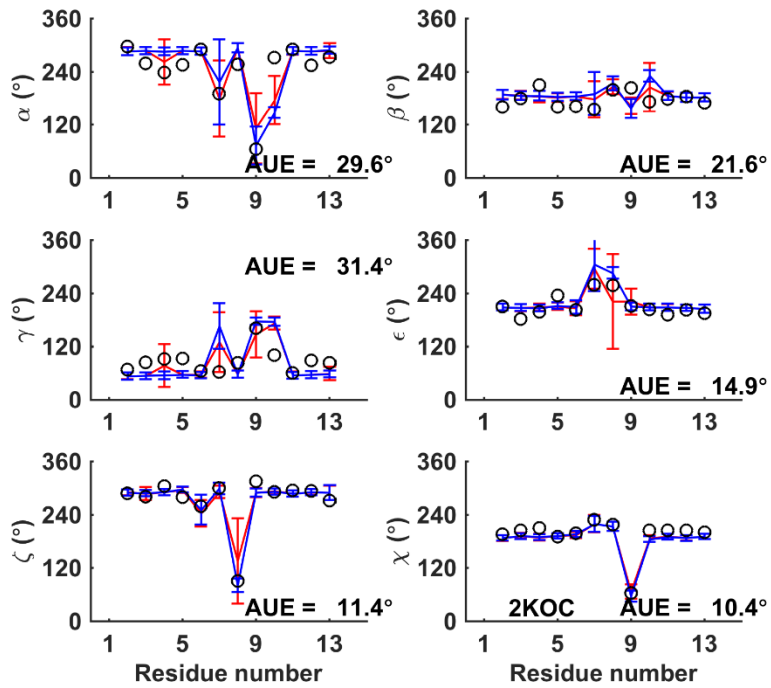
(F) 1ZIH
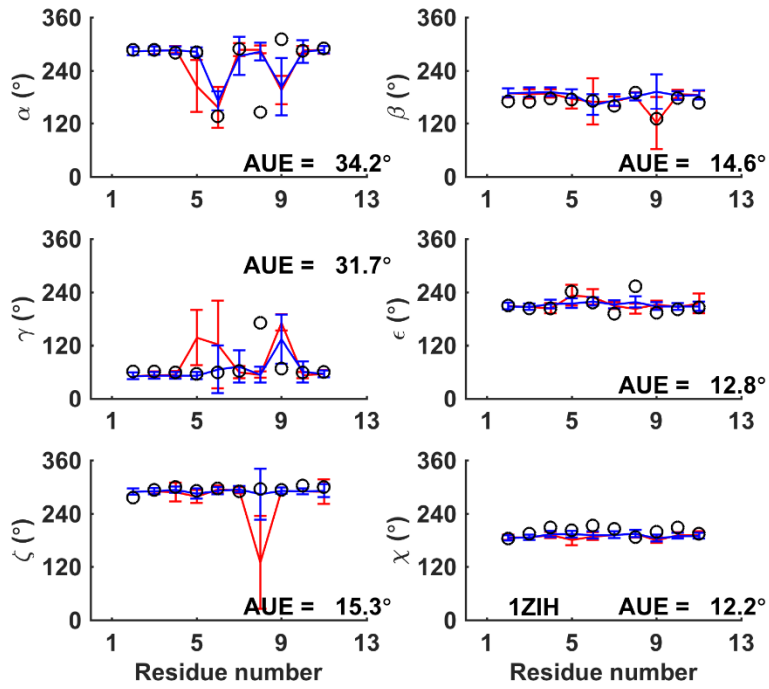
(G) 1SZY
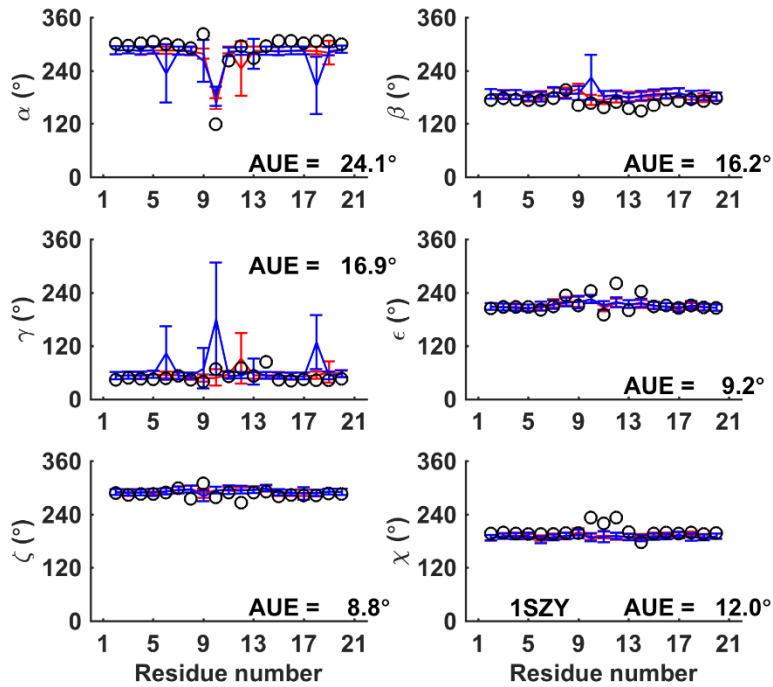
Table S5. Axis bend and groove parameters for the 4 RNA double helices calculated from simulated and NMR structures by using Curves+ program.

(A) Axis bend angle

Axis bend (°)	Number of Base pair	NMR	simulation1	simulation2
2JXQ (all WC)	10	10.2 ± 1.2	18.6 ± 10.9	19.5 ± 8.6
1F5G	10	46.9 ± 3.1	15.5 ± 8.5	15.7 ± 8.6
1MIS	8	14.0	9.6 ± 5.0	9.4 ± 4.9
2L8F	11	22.8 ± 1.7	32.2 ± 13.8	28.1 ± 11.4

(B) Width and depth of the major groove

Groove 1 (Å)	NMR width/depth	simulation1 width/depth	simulation2 width/depth
2JXQ	10.3 ± 0.2; 0.7 ± 0.2	9.0 ± 2.9; 0.4 ± 2.4	9.0 ± 2.8; 0.4 ± 1.9
1F5G	9.8 ± 0.4; 1.7 ± 0.4	10.2 ± 0.5; 0.0 ± 0.6	10.1 ± 1.2; 0.0 ± 1.3
1MIS	9.8 ± 0.2; 2.0 ± 0.3	10.6 ± 1.6; 0.2 ± 2.3	10.6 ± 2.0; 0.3 ± 3.5
2L8F	9.0 ± 0.9; −0.2 ± 0.6	9.4 ± 1.1; −0.8 ± 0.7	9.4 ± 1.2; −0.8 ± 0.7

(C) Width and depth of the minor groove

Groove 2 (Å)	NMR width/depth	simulation1 width/depth	simulation2 width/depth
2JXQ	2.9 ± 0.4; 9.8 ± 0.1	1.9 ± 3.4; 2.9 ± 4.6	2.2 ± 4.2; 2.0 ± 3.9
1F5G	11.4 ± 0.7; 8.0 ± 1.0	6.6 ± 3.8; 9.1 ± 2.6	6.2 ± 3.8; 8.3 ± 3.7
2L8F	7.5 ± 1.8; 4.9 ± 3.3	7.5 ± 1.5; 6.8 ± 3.3	7.4 ± 1.6; 6.7 ± 3.3

Figure S19. α and γ torsion transitions in RNA nucleotide residues.

(A) 2KOC U7 and G9, and (B) 1MIS A5.

(A)
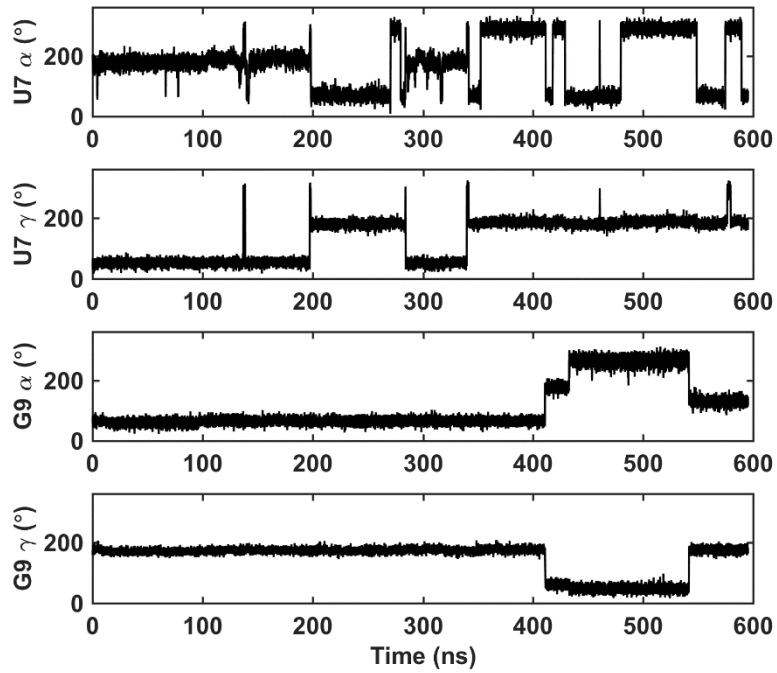
(B)
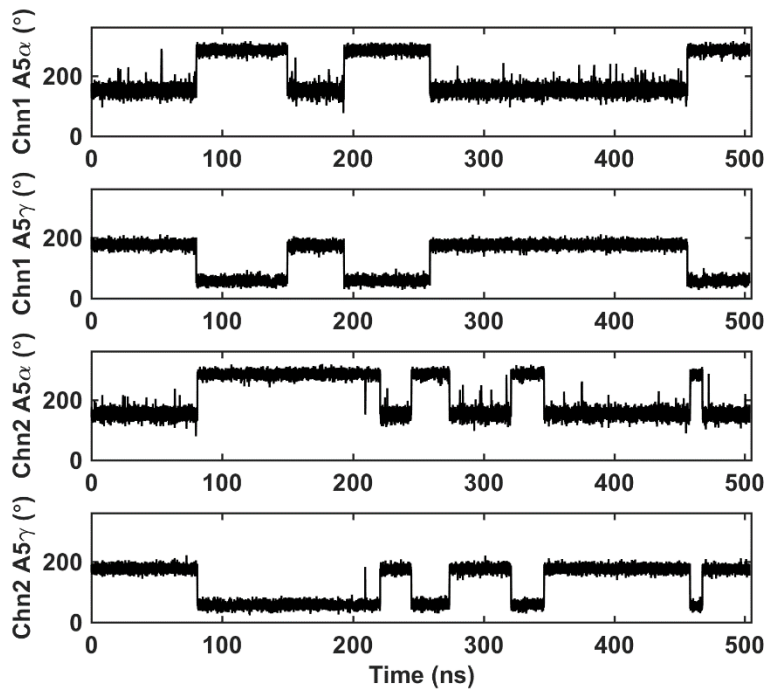
Figure S20. UUCG loop RMSDs (all heavy atoms including backbone, sugar and base) were calculated by comparing with the first NMR (2KOC) structure.

The colors show different clusters in simulation, which are the same as Figure 17B. Clusters 1-5 are in black, blue, magentas, red, and green, respectively.
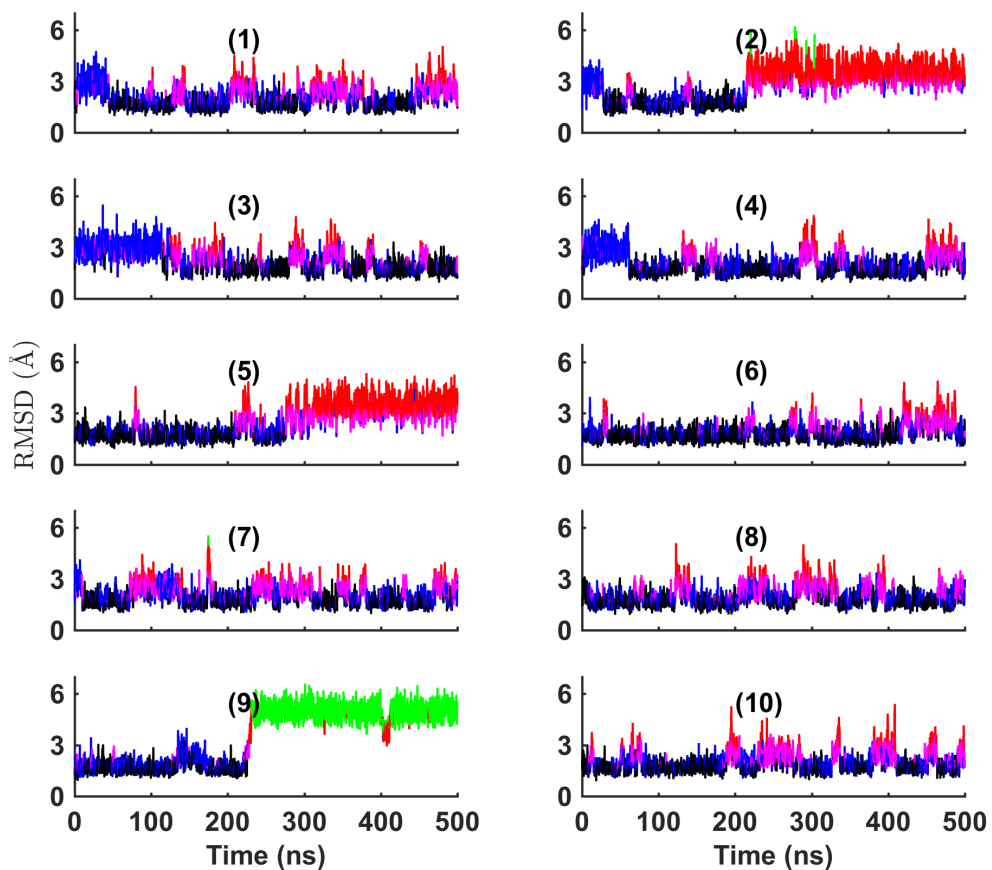
Figure S21. **Statistical population density map for ribose puckering from MD simulations of 14 RNA molecules.**

ν0 and ν4 as the two dimensions for the map, which is the same as QM and MM energy map (see Figures 2B and S4). Contour lines show the values of –ln(frequency)+3.51, which are correlated to the free energy.
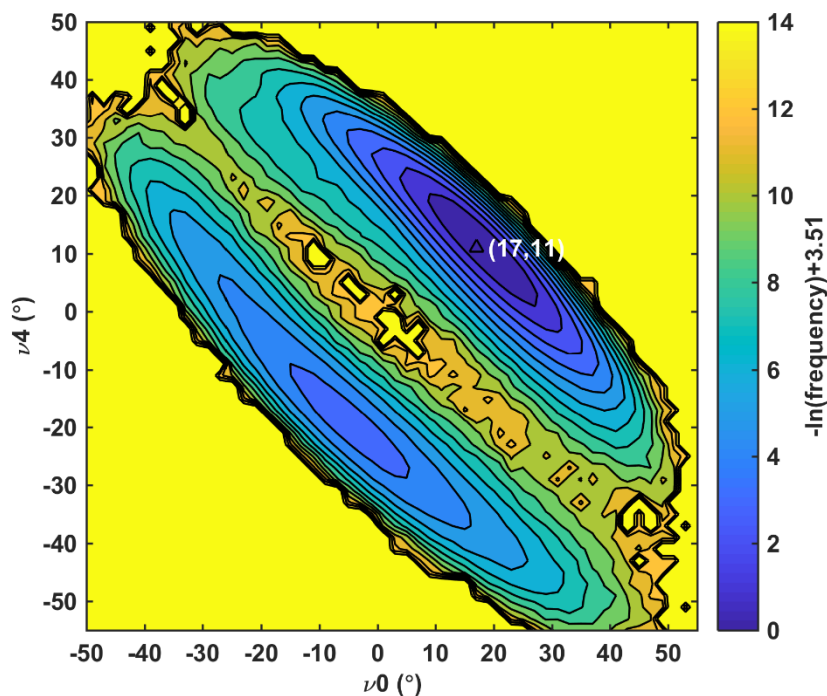
Table S6. UUCG loop torsion angles in simulation.

The same data used as Figure 17D in main text. The torsions with big error (> 30°) were shown in red.

torsions	Peaks (if two peaks, populations are showed)		NMR (°)	Error (°)
ç0	298 ± 5		291 ± 7	7
ç1	248 ± 19		262 ± 4	14
ç2	288 ± 11 (94.5%)	88 ± 8 (5.5%)	302 ± 8	14
ç3	78 ± 7		84 ± 12	6
ç4	288 ± 8		325 ± 6	37

α1	288 ± 7		277 ± 13	11
α2	183 ± 13 (88.0%)	293 ± 8 (12.0%)	195 ± 5	12
α3	293 ± 8 (92.5%)	163 ± 8 (5.0%)	268 ± 13	25
α4	68 ± 7		67 ± 16	1
α5	148 ± 7 (94.0%)	288 ± 6 (6.0%)	266 ± 8	Different conformation
β1	183 ± 7		170 ± 9	13
β2	178 ± 17		150 ± 6	28
β3	198 ± 12		198 ± 7	0
β4	158 ± 12		196 ± 9	38
γ1	58 ± 6		69 ± 10	11
γ2	53 ± 7 (83.5%)	188 ± 7 (15.5%)	62 ± 4	9
γ3	58 ± 6 (91.5%)	178 ± 7 (9.0%)	70 ± 10	12
γ4	178 ± 6		163 ± 8	15
δ1	83 ± 5		81 ± 2	2
δ2	148 ± 8		134 ± 3	14
δ3	153 ± 6		144 ± 3	9
δ4	93 ± 7		82 ± 2	11
ε1	203 ± 10		198 ± 6	5
ε2	288 ± 28		262 ± 10	26
ε3	288 ± 13		264 ± 4	24
ε4	208 ± 7 (94.5%)	283 ± 9 (5.5%)	205 ± 6	3
χ1	193 ± 6		199 ± 4	4
χ2	218 ± 23		231 ± 3	13
χ3	213 ± 9		224 ± 4	11
χ4	53 ± 11		58 ± 3	5

Table S7. Basic structural characteristic of UUCG loop and H-bond populations in simulations using AMOEBA NA and AMBER force fields.

The simulation data of AMBER force fields with best performance were selected from Table 2 of paper by Banas et al. (reference 122).

Simulations	$G_{L4}(N1)\cdots U_{L1}(O2)$	$C_{L3}(N4)\cdots U_{L2}(pro\text{-}R_p)$	$U_{L2}(O2')\cdots G_{L4}(N7)$	$U_{L1}(O2')\cdots G_{L4}(O6)$	$U_{L1}(O2')\cdots U_{L2}(O5')$	$G_{L4}\,\chi$ (°)	tSW U_{L1}/G_{L4} propeller (°)
NMR	2.7 ± 0.1	2.9 ± 0.1	2.9 ± 0.1	2.6 ± 0.1	3.4 ± 0.1	58 ± 4	-4.3 ± 4.5
AMOEBA	99%	66%	15%	80%	0%	59 ± 10	-17 ± 10
Amber 99χODE	90%	72%	54%	94%	4%	83 ± 13	-22 ± 11
Amber 99χYIL	95%	76%	41%	87%	6%	65 ± 15	1 ± 11
Amber bsc0χOL-DFT	93%	70%	50%	88%	9%	76 ± 14	1 ± 11

Table S8. Single strand RNA tetramer simulation structure clustering results.

Including RMSD, torsions, and the populations. (A) CAAU. (B) AAAA. (C) GACC

(A)

Cluster No.	NAME	RMSD vs standard A form (Å)	Torsions of cluster center (ζ3-α4-γ4)	population
CAAU 1	Mainly A form	1.0-2.5	G-G-G+	79.8%
CAAU 2	NMR minor	2.4-3.5	G-TG+	5.6%
CAAU 3	3-4 unstacking major	3.5-5.5	TG-G+	7.6%
CAAU 4	3-4 unstacking minor	4.5-5.5	TG+G+	4.4%
CAAU5	2-3 unstacking (unfolded)	5.5-7	___	2.6%

(B)

Cluster No.	NAME	RMSD vs standard A form (Å)	Torsions of cluster center (ζ3-α4-γ4)	population
AAAA 1	Mainly A form	1.0-2.5 (Green)	G-G-G+	88.8%
AAAA 2	NMR minor	2.5-3.5 (Cyan)	G-TG+	7.0%
AAAA 3	3-4 unstacking major	3.5-5.5 (Magenta)	TG-G+	3.2%
AAAA 4	3-4 unstacking minor	5-6.5 (Yellow)	G+G+G+	1.0%

(C)

Cluster No.	NAME	RMSD vs standard A form (Å)	Torsions of cluster center (ζ3-α4-γ4)	population
GACC 1	Mainly A form	1.0-2.6 (Green)	G-G-G+	42.8%
GACC 2	NMR minor	2.5-3.5 (Cyan)	G+TG+	17.2%
GACC 3	3-4 unstacking major	4.0-5.5 (Magenta)	G+G+T	26.4%
GACC 4	3-4 unstacking minor	3.5-4.5 (Yellow)	TG-G+	13.5%

Figure S22. Supposition of different RNA tetramer clusters.

(A) AAAA. (B) CAAU. (C) GACC. Configurational clusters 1 to 4 are in color black, blue, magentas, red, respectively.

rAAAA rCAAU rGACC
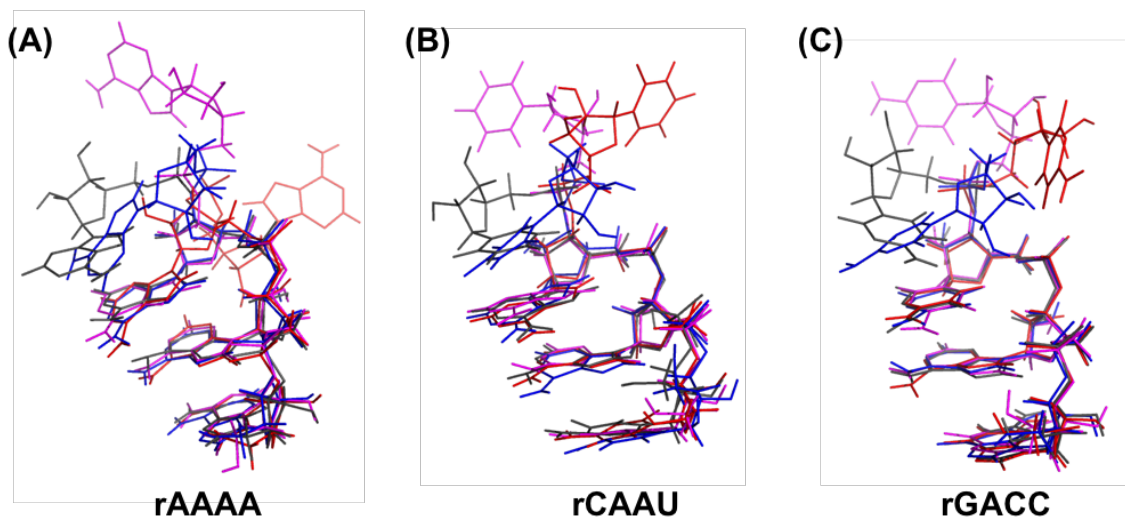
Figure S23. Single strand tetramer rCAAU simulation analysis.

(A) RMSD distribution (see RMSD calculation method in main text). (B) Torsion population compared with A-form values. (C) ζ3-α4 and α4 -γ4 statistical population maps. The contour colors show the value of the negative logarithm of the population density. The central structures in each of the 4 clusters are identified in the map.
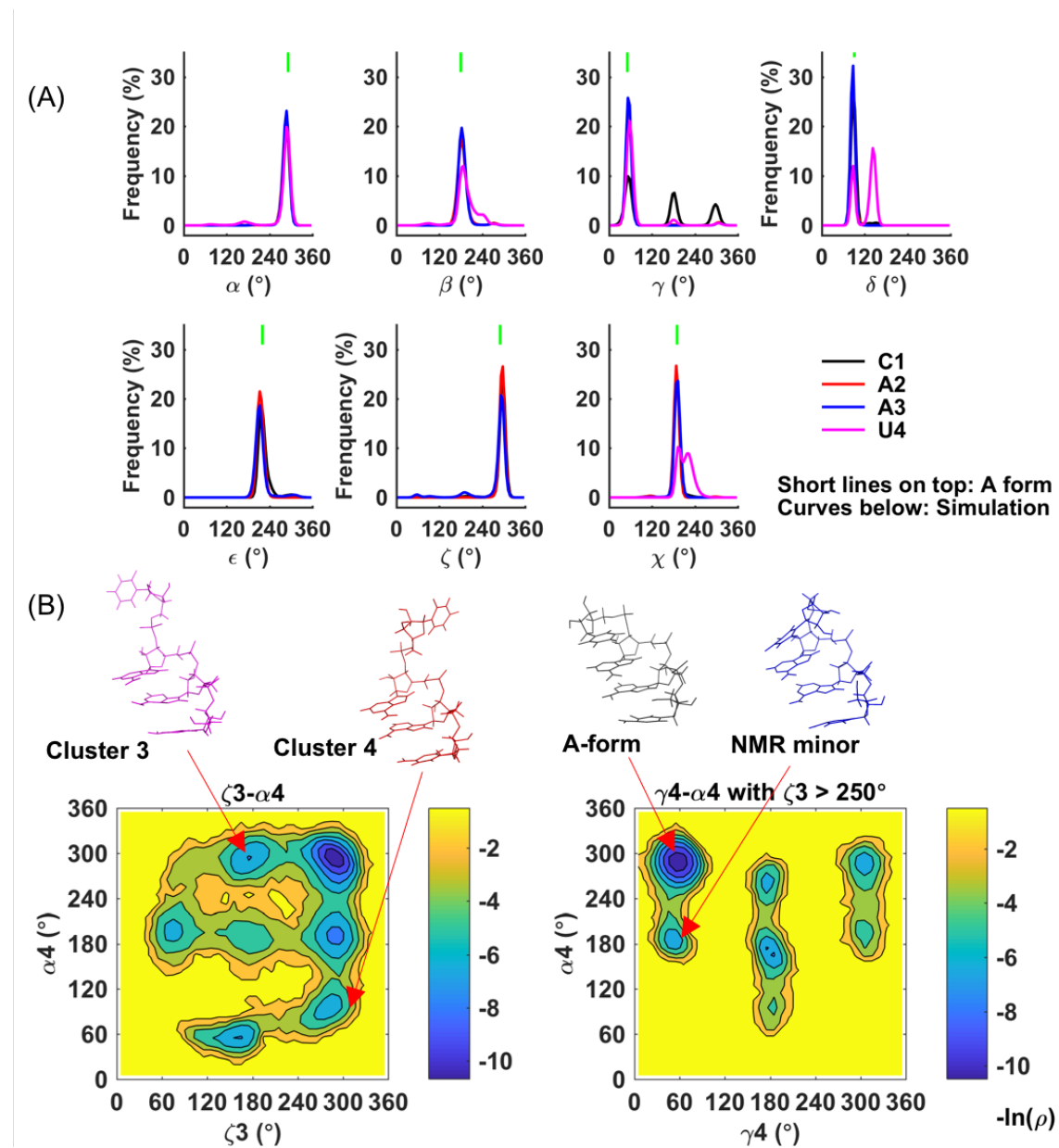
Figure S24. Single strand tetramer rAAAA simulation analysis.

(A) Torsion population compared with A-form values. (B) ζ3-α4 and α4 -γ4 torsion statistical population maps. The contour colors show the value of the negative logarithm of the population density. The central structures in each of the 4 clusters are identified in the map.
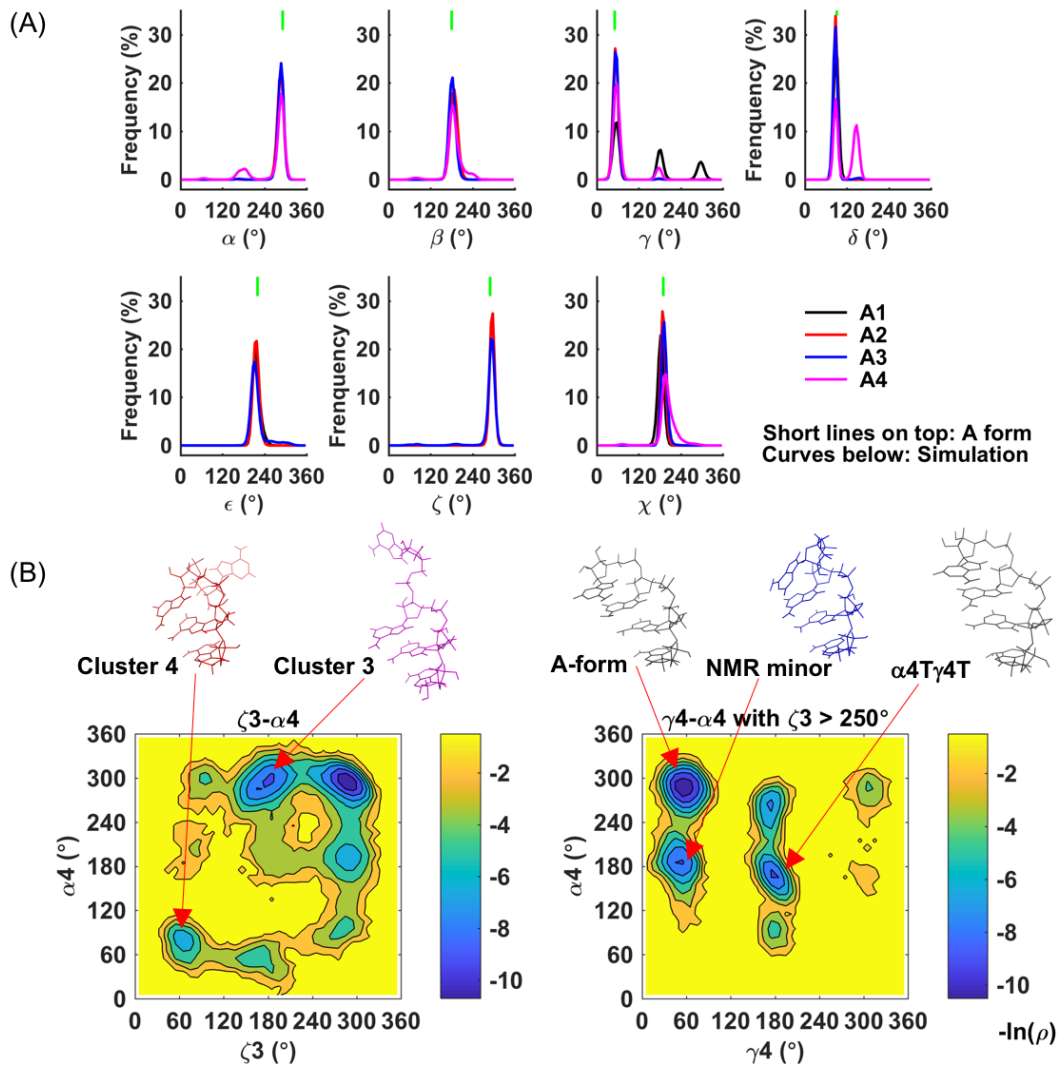
Figure S25. Single strand RNA tetramer RMSD was calculated by comparing the simulated structures with the standard A-form structure.

(A) rAAAA. (B) rCAAU. (C) rGACC. The colors show different clusters in simulation: clusters 1-5 are in black, blue, magentas, red, green, respectively.

(A)
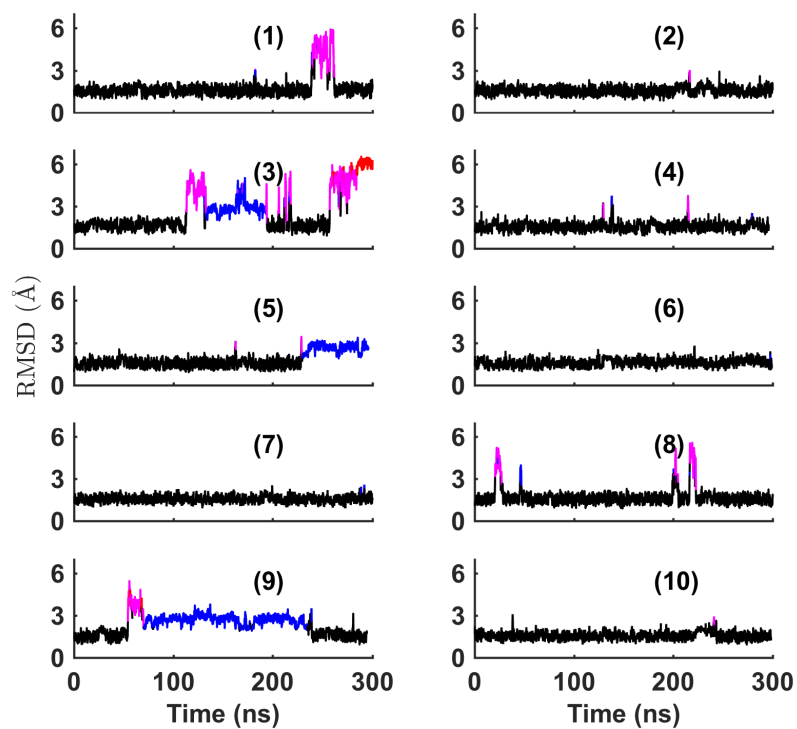
(B)
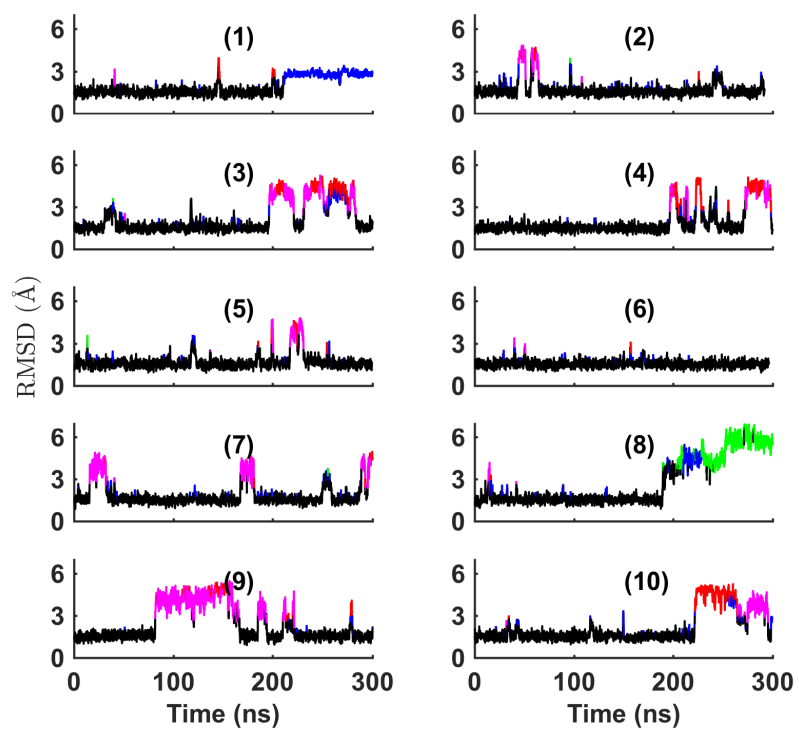
(C)
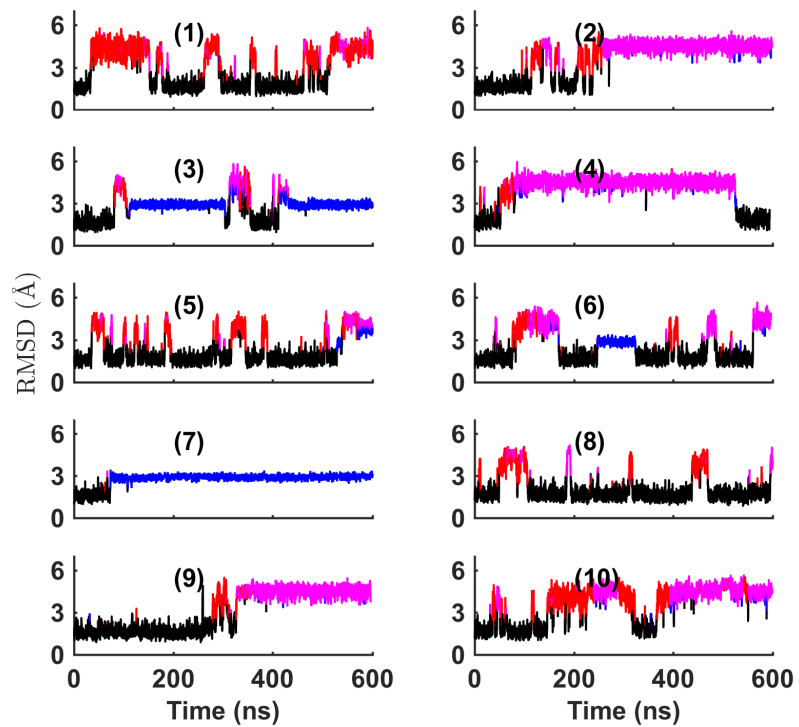
Figure S26. Extension of the rAAAA simulation trajectory 3 and UUCG loop simulation 5. The colors used for indicating configurational clusters are the same as Figure S25 or S20

(A)

(B)

.
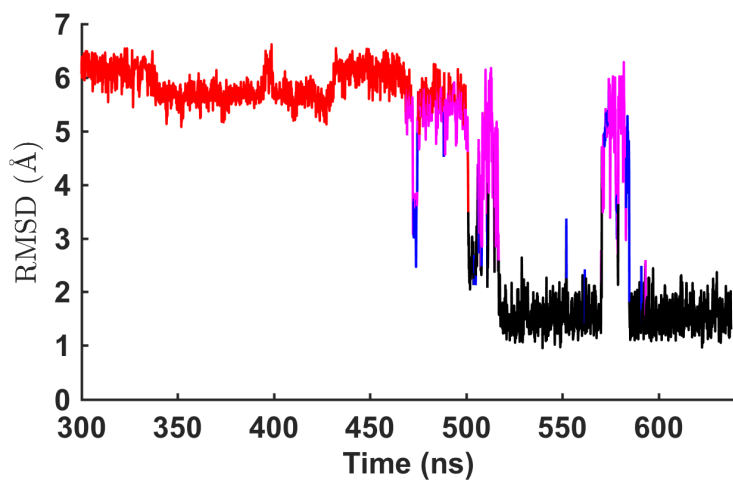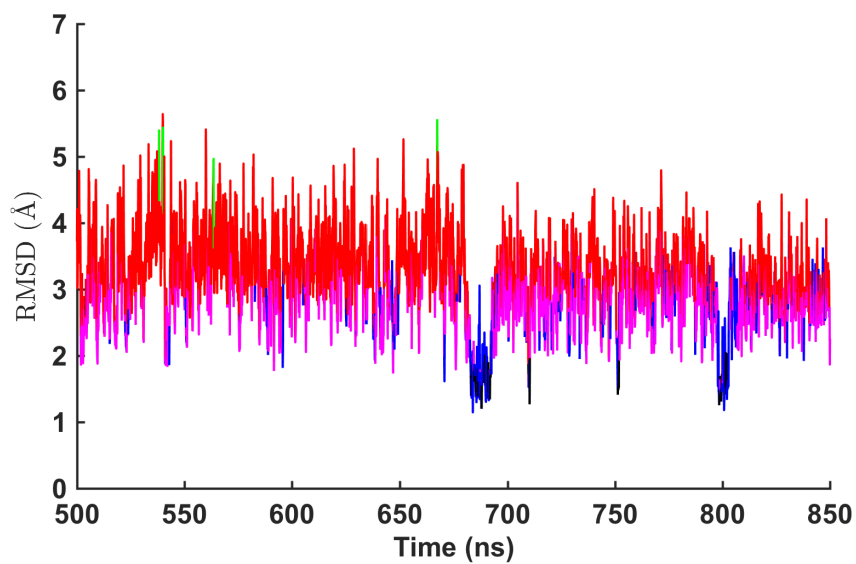
Table S9. β, γ Major conformation ratios in RNA tetramer simulation

(T: 150-210; G+: 30-90) Underlined data: missed (relative error > 50%). The NMR data were taken from the SI of reference 111.

	CAAU (simulation)	CAAU (NMR)	AAAA (simulation)	AAAA (NMR)	GACC (simulation)	GACC (NMR)
β_2 (T)	94.7%	~96%	98.2%	~100%	97.2%	~100%
β_3 (T)	96.5%	~100%	99.4%	~100%	98.7%	~100%
β_4 (T)	71.6%	~92%	84.2%	~100%	<u>41.4%</u>	<u>~95%</u>
γ_1 (G+)	<u>47.4%</u>	<u>~82%</u>	53.9%	~77%	45.1%	~66%
γ_2 (G+)	96.5%	~87%	99.2%	~100%	99.9%	~96%
γ_3 (G+)	97.5%	~93%	99.1%	~96%	99.9%	~100%
γ_4 (G+)	92.7%	~70%	89.4%	~96%	<u>68.8%</u>	<u>~100%</u>

Table S10. Calculated NOE distance data for the RNA single strand tetramers.

Calculation method: $r_{NOE} = \left(\sum_{i=1}^{N} \frac{r_i^{-6}}{N} \right)^{-1/6}$. The NMR data were taken from the SI of reference 111.

(A) CAAU: 6 of 84 missed. Missed ones are colored in red.

No.	Peaks	Calculation	NMR
1	C1H4'-C1H6	>6.0	4.8
2	C1H4'-A2H8	5.25	4.9
3	C1H3'-C1H5	4.60	No
4	C1H3'-C1H6	2.86	3.0
5	C1H3'-A2H4'	5.53	No
6	C1H3'-A2H3'	4.86	No
7	C1H3'-A2H1'	5.59	overlap
8	C1H3'-A2H8	2.39	overlap
9	C1H2'-C1H5	5.49	No
10	C1H2'-C1H6	3.71	3.3
11	C1H2'-A2H4'	3.95	overlap
12	C1H2'-A2H3'	4.54	No
13	C1H2'-A2H2'	5.69	Spin-Diffusion
14	C1H2'-A2H1'	4.02	4.3

15	C1H2'-A2H8	2.42	3.2
16	C1H1'-C1H5	5.24	No
17	C1H1'-C1H6	3.45	3.6
18	C1H1'-A2H1'	5.85	>5.0
19	C1H1'-A2H8	4.76	5.5
20	C1H5 -A2H8	4.80	No
21	C1H6 -A2H8	4.24	4.7
22	A2H4'-A2H8	4.24	4.4
23	A2H4'-A3H8	5.34	No
24	A2H3'-A2H8	2.88	3.0
25	A2H3'-A2H2	5.95	No
26	A2H3'-A3H4'	5.73	No
27	A2H3'-A3H3'	4.92	Spin-Diffusion
28	A2H3'-A3H1'	5.84	No
29	A2H3'-A3H8	2.56	2.7
30	A2H3'-U4H5	5.85	No
31	A2H2'-A2H8	4.03	3.3
32	A2H2'-A2H2	4.51	5.2
33	A2H2'-A3H4'	3.98	3.8
34	A2H2'-A3H3'	4.35	Spin-Diffusion
35	A2H2'-A3H2'	5.54	Spin-Diffusion
36	A2H2'-A3H1'	3.97	3.8
37	A2H2'-A3H8	2.35	2.7
38	A2H1'-A2H8	3.55	3.7
39	A2H1'-A2H2	4.82	4.5
40	A2H1'-A3H1'	5.77	No
41	A2H1'-A3H8	4.68	4.9
42	A2H8 -A3H2'	>6.0	>5.0
43	A2H8 -A3H8	4.43	No
44	A2H2 -A3H4'	5.75	No
45	A2H2 -A3H3'	4.92	No
46	A2H2 -A3H2'	5.18	No
47	A2H2 -A3H1'	2.96	3.9
48	A2H2 -A3H8	4.89	No
49	A2H2 -A3H2	5.13	No
50	A3H4'-A3H8	4.25	4.6
51	A3H4'-U4H2'	>6.0	4.8
52	A3H4'-U4H6	5.70	No
53	A3H3'-A3H8	2.84	2.7

54	A3H3'-U4H4'	5.69	No
55	A3H3'-U4H3'	5.08	No
56	A3H3'-U4H2'	3.73	3.0
57	A3H3'-U4H5	3.46	3.9
58	A3H3'-U4H6	2.96	3.1
59	A3H2'-A3H8	4.13	3.2
60	A3H2'-A3H2	4.55	No
61	A3H2'-U4H4'	4.14	3.3
62	A3H2'-U4H3'	4.91	No
63	A3H2'-U4H2'	4.76	No
64	A3H2'-U4H1'	4.48	4.6
65	A3H2'-U4H5	3.81	3.4
66	A3H2'-U4H6	2.54	3.0
67	A3H1'-A3H8	3.61	3.7
68	A3H1'-A3H2	4.77	5.4
69	A3H1'-U4H5	5.55	No
70	A3H1'-U4H6	5.10	No
71	A3H8 -U4H2'	4.93	3.9
72	A3H8 -U4H5	4.00	4.3
73	A3H8 -U4H6	4.87	4.9
74	A3H2 -U4H2'	5.80	No
75	A3H2 -U4H1'	3.54	4.5
76	A3H2 -U4H5	4.67	No
77	A3H2 -U4H6	4.97	No
78	U4H4'-U4H6	4.18	No
79	U4H3'-U4H5	5.25	No
80	U4H3'-U4H6	3.23	overlap
81	U4H2'-U4H5	4.71	overlap
82	U4H2'-U4H6	2.65	No
83	U4H1'-U4H5	5.35	No
84	U4H1'-U4H6	3.56	3.6

(B) AAAA: 16 of 80 missed. Missed ones are colored in red.

No.	Peaks	Calculation	NMR
1	A1H4'-A2H8	5.29	No
2	A1H3'-A1H8	3.07	3.4
3	A1H3'-A2H4'	5.73	No
4	A1H3'-A2H3'	4.94	No
5	A1H3'-A2H1'	5.83	No
6	A1H3'-A2H8	2.48	No

7	A1H2'-A1H8	4.33	3.6
8	A1H2'-A1H2	4.34	No
9	A1H2'-A2H4'	4.07	4.6
10	A1H2'-A2H3'	4.51	No
11	A1H2'-A2H2'	5.68	No
12	A1H2'-A2H1'	4.08	4.5
13	A1H2'-A2H8	2.33	3.0
14	A1H1'-A1H8	3.56	3.3
15	A1H1'-A1H2	4.87	No
16	A1H1'-A2H1'	5.94	No
17	A1H1'-A2H8	4.69	4.8
18	A1H8 -A2H8	4.60	No
19	A1H2 -A2H4'	5.55	No
20	A1H2 -A2H2'	5.47	No
21	A1H2 -A2H1'	2.83	3.8
22	A1H2 -A2H8	4.51	No
23	A1H2 -A2H2	5.41	No
24	A2H4'-A2H8	4.23	5.0
25	A2H4'-A3H8	5.29	No
26	A2H3'-A2H8	2.92	3.0
27	A2H3'-A3H4'	5.68	No
28	A2H3'-A3H3'	4.96	No
29	A2H3'-A3H1'	5.75	No
30	A2H3'-A3H8	2.50	2.9
31	A2H2'-A2H8	4.21	No
32	A2H2'-A2H2	4.54	No
33	A2H2'-A3H4'	3.93	No
34	A2H2'-A3H3'	4.40	No
35	A2H2'-A3H2'	5.48	No
36	A2H2'-A3H1'	3.90	No
37	A2H2'-A3H8	2.35	No
38	A2H2'-A4H3'	5.62	No
39	A2H1'-A2H8	3.62	3.7
40	A2H1'-A2H2	4.79	No
41	A2H1'-A3H1'	5.70	No
42	A2H1'-A3H8	4.73	4.7
43	A2H8 -A3H8	4.40	No
44	A2H2 -A3H4'	5.75	No
45	A2H2 -A3H2'	5.60	No
46	A2H2 -A3H1'	2.94	4.1
47	A2H2 -A3H8	4.97	No

No.	Peaks	Calculation	NMR
48	A2H2 -A3H2	5.16	No
49	A3H4'-A3H8	4.24	No
50	A3H4'-A4H3'	5.66	No
51	A3H4'-A4H2'	5.81	No
52	A3H4'-A4H8	5.46	No
53	A3H3'-A3H8	2.83	No
54	A3H3'-A4H4'	5.54	No
55	A3H3'-A4H3'	4.49	No
56	A3H3'-A4H2'	3.23	3.2
57	A3H3'-A4H1'	5.80	No
58	A3H3'-A4H8	2.69	3.1
59	A3H3'-A4H2	5.41	No
60	A3H2'-A3H8	4.01	No
61	A3H2'-A3H2	4.63	No
62	A3H2'-A4H4'	3.98	No
63	A3H2'-A4H3'	4.61	No
64	A3H2'-A4H2'	4.23	No
65	A3H2'-A4H1'	4.12	No
66	A3H2'-A4H8	2.48	No
67	A3H1'-A3H8	3.66	3.8
68	A3H1'-A3H2	4.75	No
69	A3H1'-A4H8	4.97	5.6
70	A3H8 -A4H2'	4.54	4.4
71	A3H8 -A4H8	4.58	No
72	A3H2 -A4H1'	3.38	4.8
73	A3H2 -A4H8	5.17	No
74	A3H2 -A4H2	4.93	No
75	A4H4'-A4H8	4.31	4.7
76	A4H3'-A4H8	3.10	No
77	A4H2'-A4H8	3.03	3.0
78	A4H2'-A4H2	5.03	No
79	A4H1'-A4H8	3.61	3.8
80	A4H1'-A4H2	4.68	No

(C) GACC: 8 of 82 missed. Missed ones are colored in red.

No.	Peaks	Calculation	NMR
1	G1H4'-A2H8	5.33	>5.0
2	G1H3'-G1H8	3.17	3.7
3	G1H3'-A2H4'	5.84	No
4	G1H3'-A2H3'	5.01	No

5	G1H3'-A2H8	2.58	overlap
6	G1H3'-C3H5	5.25	No
7	G1H2'-G1H8	4.40	3.4
8	G1H2'-A2H4'	4.18	3.2
9	G1H2'-A2H3'	4.55	>5.0
10	G1H2'-A2H2'	5.78	>5.0
11	G1H2'-A2H1'	4.24	>5.0
12	G1H2'-A2H8	2.28	2.7
13	G1H1'-G1H8	3.52	3.4
14	G1H1'-A2H1'	>6.0	>5.0
15	G1H1'-A2H8	4.66	4.2
16	G1H8 -A2H8	4.83	>5.0
17	A2H4'-A2H8	4.23	3.5
18	A2H4'-C3H6	5.41	4.2
19	A2H3'-A2H8	2.97	overlap
20	A2H3'-C3H4'	5.72	No
21	A2H3'-C3H3'	4.95	>5.0
22	A2H3'-C3H1'	5.89	No
23	A2H3'-C3H5	3.11	No
24	A2H3'-C3H6	2.67	2.6
25	A2H3'-C4H5	5.68	No
26	A2H2'-A2H8	4.24	overlap
27	A2H2'-A2H2	4.48	No
28	A2H2'-C3H4'	4.01	>5.0
29	A2H2'-C3H3'	4.45	3.7
30	A2H2'-C3H2'	5.54	No
31	A2H2'-C3H1'	4.09	3.6
32	A2H2'-C3H5	4.11	No
33	A2H2'-C3H6	2.32	2.5
34	A2H1'-A2H8	3.61	4.1
35	A2H1'-A2H2	4.80	>5.0
36	A2H1'-C3H1'	5.94	No
37	A2H1'-C3H5	5.64	No
38	A2H1'-C3H6	4.74	No
39	A2H8 -C3H2'	>6.0	>5.0
40	A2H8 -C3H5	3.78	4.1
41	A2H8 -C3H6	4.63	No
42	A2H2 -C3H4'	5.72	No
43	A2H2 -C3H2'	5.70	No
44	A2H2 -C3H1'	3.02	3.7
45	A2H2 -C3H5	5.65	No

46	A2H2 -C3H6	4.73	No
47	C3H4'-C3H6	4.06	No
48	C3H4'-C4H6	5.81	No
49	C3H4'-C4H2'	>6.0	overlap
50	C3H3'-C3H5	4.53	No
51	C3H3'-C3H6	2.74	2.4
52	C3H3'-C4H4'	5.36	No
53	C3H3'-C4H3'	4.43	No
54	C3H3'-C4H2'	3.00	overlap
55	C3H3'-C4H5	3.58	No
56	C3H3'-C4H6	2.97	2.8
57	C3H2'-C3H5	5.40	No
58	C3H2'-C3H6	3.67	3.0
59	C3H2'-C4H4'	4.61	overlap
60	C3H2'-C4H3'	5.34	No
61	C3H2'-C4H2'	4.79	No
62	C3H2'-C4H1'	5.08	Line Noise
63	C3H2'-C4H5	3.53	overlap
64	C3H2'-C4H6	2.84	2.7
65	C3H1'-C3H5	5.28	No
66	C3H1'-C3H6	3.49	3.3
67	C3H1'-C4H5	5.66	No
68	C3H1'-C4H6	5.53	No
69	C3H5 -C4H2'	4.66	No
70	C3H5 -C4H5	3.92	No
71	C3H5 -C4H6	5.47	No
72	C3H6 -C4H3'	5.98	>5.0
73	C3H6 -C4H2'	4.03	3.8
74	C3H6 -C4H5	4.18	No
75	C3H6 -C4H6	4.92	No
76	C4H4'-C4H6	4.18	No
77	C4H3'-C4H5	5.67	No
78	C4H3'-C4H6	3.51	2.6
79	C4H2'-C4H5	4.70	No
80	C4H2'-C4H6	2.60	2.8
81	C4H1'-C4H5	5.35	No
82	C4H1'-C4H6	3.57	2.8

Table S11. RDC data studied for HIV TAR.

Residue	H-C bond	Experimental RDC
18G	H8-C8	-1.538
19C	H5-C5	-0.755
20A	H8-C8	-1.162
20A	H1'-C1'	1.265
21G	H8-C8	-1.567
22A	H8-C8	-1.624
22A	H2-C2	-0.518
22A	H1'-C1'	0.314
26G	H8-C8	-1.32
27A	H8-C8	-1.447
27A	H2-C2	-0.979
27A	H1'-C1'	2.101
28G	H8-C8	-1.509
29C	H5-C5	-1.162
29C	H6-C6	-0.926
32U	H1'-C1'	-0.118
34G	H1'-C1'	-0.655
36G	H8-C8	-1.433
37C	H5-C5	-1.194
38U	H5-C5	0.197
39C	H5-C5	0.374
40U	H1'-C1'	1.37
42U	H5-C5	-0.98
43G	H8-C8	-0.929
44C	H5-C5	0.047

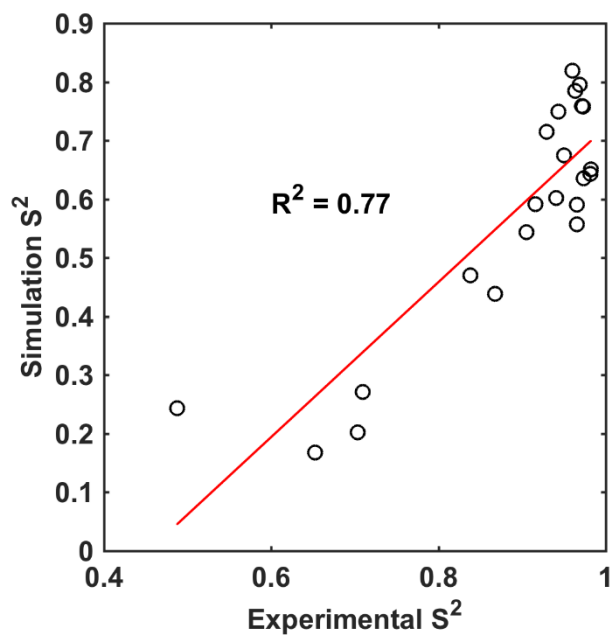**Figure S27. Correlation between calculated order parameters and the experimental values.**

Figure S28. Correlation coefficients between RDC calculated from the first 4 TAR PDB structures and the experimental values.

The PDB (NMR) structure of HIV TAR, 1ANR, was deposited on PDB in 1996 and the residual dipolar coupling data was collected latter, in 2003 (reference 123). The different sources of the 1ANR structures and the residual dipolar coupling data may be responsible for the poor correlation.
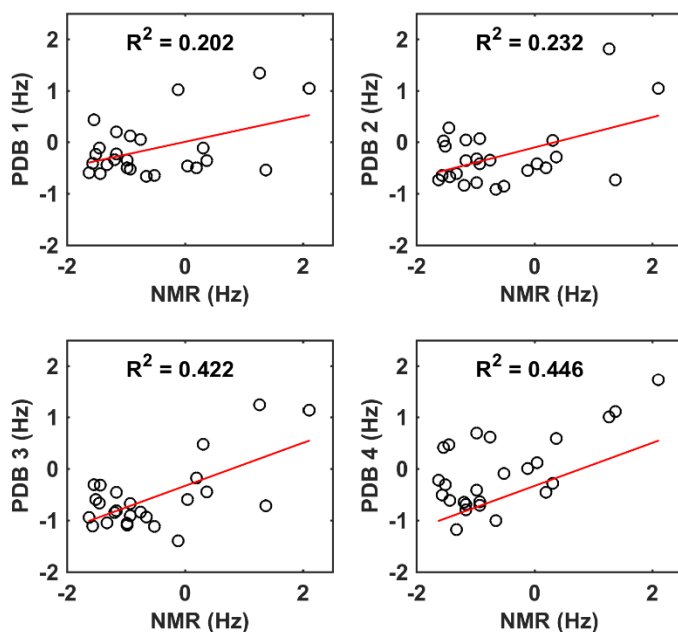
Figure S29. Correlation coefficients of calculated and experimental TAR RDC values.

The first 8 panel shows the calculated result for each of the 8 trajectories. Each of the first 4 NMR structures was selected as the starting structure of 2 trajectories. For example, the first two trajectories were starting from the first NMR structure.
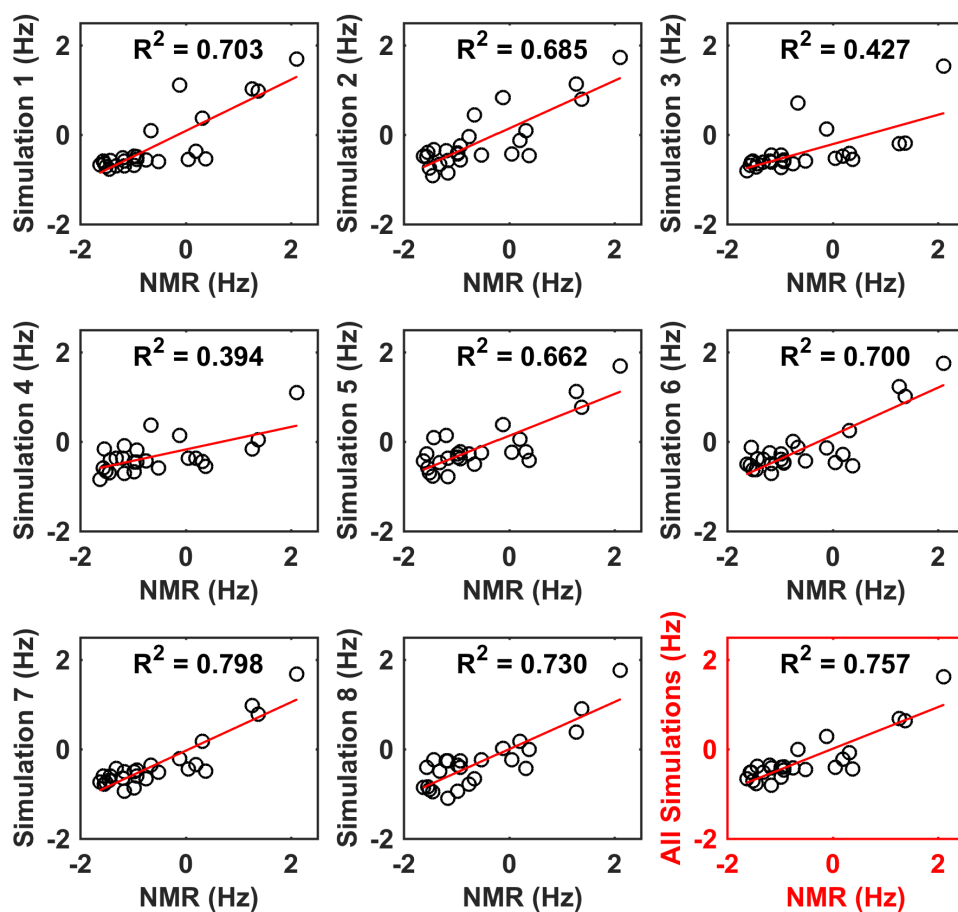
Figure S30. Polarization energy of adenine stacking in single strand A form RNA.

The standard A-form RNA –(A)n – was first built, and then the backbone and sugars were trimmed off, with only stacking adenines left. The energy here refers to interaction energy (total energy minus the sum of monomer energy). The average induced dipole on each adenine was ~ 0.47 Debye.