Supplementary Material

A coarse-grained approach for investigating the structure and dynamics of large nucleic acid species

Sweta Vangnaveti[1], *Rebecca J. D'Esposito*[2], *Jennifer L. Lippens*[3], *Daniele Fabris*[2,4]*, *Srivathsan Ranganathan*[1]*

1) The RNA Institute, University at Albany, 2016A Life Sciences Research Building, Albany, NY
2) Department of Chemistry, University at Albany, NY
3) Discovery Analytical Sciences, Amgen, Thousand Oaks, CA
4) Department of Biological Sciences, University at Albany, NY

* Corresponding authors: Daniele Fabris: fabris@albany.edu
Srivathsan Ranganathan: svranganathan@albany.edu

TABLES

Table 1S. Sequences of the oligonucleotides employed in the study. Species marked with ◆ are single stranded (ss), whereas those marked with * were combined with the respective complementary strand (not listed) to form a duplex construct of corresponding length.

Figure No.	Length	Sequence	Remarks
2,2S	3mer◆*	CCU	Used for Parameterization & Refinement (Structures are generated using Nucleic Acid Builder (NAB) package of Amber)
2,2S	5mer◆*	CCUAC	
2,2S	10mer◆*	CCUACUCGUU	
2,2S	16mer◆*	CCUACUCGUUACCUUC	
2,2S	32mer◆*	CCUACUCGUUACCCUCUUCUGAUACUGUUUAA	
2,2S	48mer◆*	CCUACUCGUUACCUUCUUCUGACUUCCCUCUUUCUUCCUACUGUUUAA	
2,2S	64mer◆*	CCUACUCGUUACCUUCUUCUGACUUCCCUCUUUCUUCCUAUCUGAAUUGUGUUCUCCUGUUUAA	
3	10,16,32,48,64mers◆	PolyA, polyC, polyG, polyU	Used for Testing (Structures are generated using Nucleic Acid Builder (NAB) package of Amber)
4	6mer◆*	CGCGCG	
4	10mer◆*	CGCGCGCGCG	

4	10mer◆*	CCUACUCGUU	Used for Testing. (Structures are snapshots from all atom MD simulations)
4	16mer◆*	CCUACUCGUUACCUUC	
4	18mer◆*	CGCGCGCGCGCGCGCGCG	
4	22mer◆*	GCGCGCGCGCGCGCGCGCGCG	
4	26mer◆*	CGCGCGCGCGCGCGCGCGCGCGCGCG	
4	30mer◆*	CUGUUGCACUAUGCCAGACAAUAAUUUUCU	
4	32mer◆*	CCUACUCGUUACCCUCUUCUGAUACUGUUUAA	
4	48mer◆*	CCUACUCGUUACCUUCUUCUGACUUCCCUCUUU CUUCCUACUGUUUAA	
4	64mer◆*	CCUACUCGUUACCUUCUUCUGACUUCCCUCUUU CUUCCUAUCUGAAUUGUGUUCUCCUGUUUAA	
4		(*All DNA sequences are the same as those of the RNA oligomers above with the Us replaced by Ts)*	Used for Testing. (Structures are snapshots from
5		*(All sequences are the same as those of the ones in figure 2,2S. The CG structures are generated using the oxRNA model)*	Used for Testing. (Structures are snapshots from all atom MD simulations)
6	20mer hairpin (SL 20 RNA)	GGACUAGCGGAGGCUAGUCC	Used for Testing. (Structures are snapshots from all atom and coarse-grained MD simulations)

6	34mer hairpin (SL 34 RNA)	GUCAGGGUCAGGAAAAAAAAAACCUGACCCUGAC	Used for Testing. (Structures are snapshots from coarse-grained MD simulations only)
6	16mer hairpin (SL 16 DNA)	TGCGATACTCATCGCA	Used for Testing. (Structures are snapshots from all atom and coarse-grained MD
6	28mer hairpin (SL 28 DNA)	GCGTTCATCAGAGTCATCTGATGAACGC	Used for Testing. (Structures are snapshots from all atom and coarse-grained MD

Table 2S. Hard-sphere radii for the five-bead representation of nucleotides.

Pseudo-atom	R_{HS0} Å (initial guess) (r_B= 3.00Å, f=0, s=0, s_T=0)	R_{HS1} Å (optimized for single nucleosides) (r_B=2.88Å, f=0, s=0.05, s_T=0.1)	R_{HS2} Å (optimized for DNA/RNA oligos) (r_B=2.88Å, f=0.125, s=0.05, s_T=0.1)
P	3.8*	3.8	3.32
R	4.2*	4.2	3.68
dR	3.9*	3.9	3.41
B_1^Y	3.00	3.02	3.17
B_1^R	3.00	2.74	2.60

B_{23}^{N}	3.00	2.88	3.12
B_{23}^{T}	3.00	3.17	3.43

*r_P =3.8 Å, r_R = 4.2 Å, r_{dR} = 3.9 Å are R_{HS} values for the phosphate, sugar and deoxy-sugar group, respectively, which are independently obtained as described in the parameterization section of the main text.

Table 3S. Comparison of the five-bead model (R_{HS1}) with the all-atom model. Nucleobases are indicated with the respective alphabets, and nucleotides with an R/D prefix to indicate RNA/DNA, respectively.

	CCS, EHSS (Å^2)			CCS, PA (Å^2)		
	Five-bead	All atom	% Δ	Five-bead	All atom	% Δ
A	59.55	60.09	-0.91	56.65	58.97	-3.94
C	55.40	55.59	-0.34	52.78	54.58	-3.30
G	66.09	64.90	1.84	62.02	63.50	-2.34
U	55.32	54.66	1.21	52.74	53.64	-1.67
T	61.68	61.20	0.48	59.31	59.67	-0.36
RA	113.00	114.52	-1.34	106.72	107.62	-0.84
RC	107.56	107.41	0.15	101.84	101.15	0.69
RG	119.23	118.80	0.36	111.93	111.42	0.46
RU	107.78	106.77	0.95	101.89	100.44	1.43
DT	106.76	106.16	0.60	101.14	100.13	1.01

Table 4S. Comparison of the five-bead model (R_{HS2}) with the all-atom model. Nucleobases are indicated with the respective alphabets, and nucleotides with an R/D prefix to indicate RNA/DNA, respectively.

	CCS, EHSS (Å^2)			CCS, PA (Å^2)		
	Five-bead	All atom	% Δ	Five-bead	All atom	% Δ
A	66.60	60.09	10.83	63.88	58.97	7.68
C	61.98	55.59	11.49	59.54	54.58	8.33
G	73.98	64.90	14.00	69.85	63.50	9.08
U	61.85	54.66	13.16	59.46	53.64	9.79
T	69.14	61.20	13.00	66.98	59.67	12.25
RA	104.35	114.52	-8.9	98.26	107.62	-9.53
RC	99.61	107.41	-7.3	93.62	101.15	-8.04
RG	111.46	118.80	-6.2	104.18	111.42	-6.95
RU	99.67	106.77	-6.7	93.62	100.44	-7.28
DT	101.98	106.16	-3.93	96.55	100.13	-3.58

FIGURES

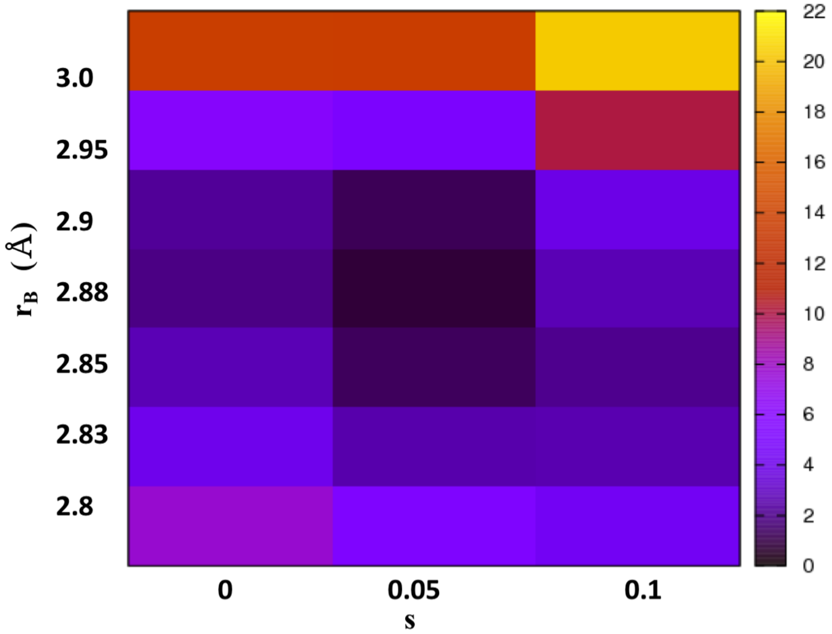

Figure 1S. The mean squared errors for the five-bead and all-atom CCS values averaged over the four nucleobases, plotted as a function of s and rB. Starting with an initial value of 3.0 (r_B) and 0.0 (s), the parameter space (r_B → (3.0,2.95,2.9,2.88,2.85,2.83,2.8 Å), s→ (0.0,0.05,0.1)) was explored. The values of r_B and s that minimized the mean square errors between the calculated and the reference all-atom CCS values were chosen for the R_{HS1} model.

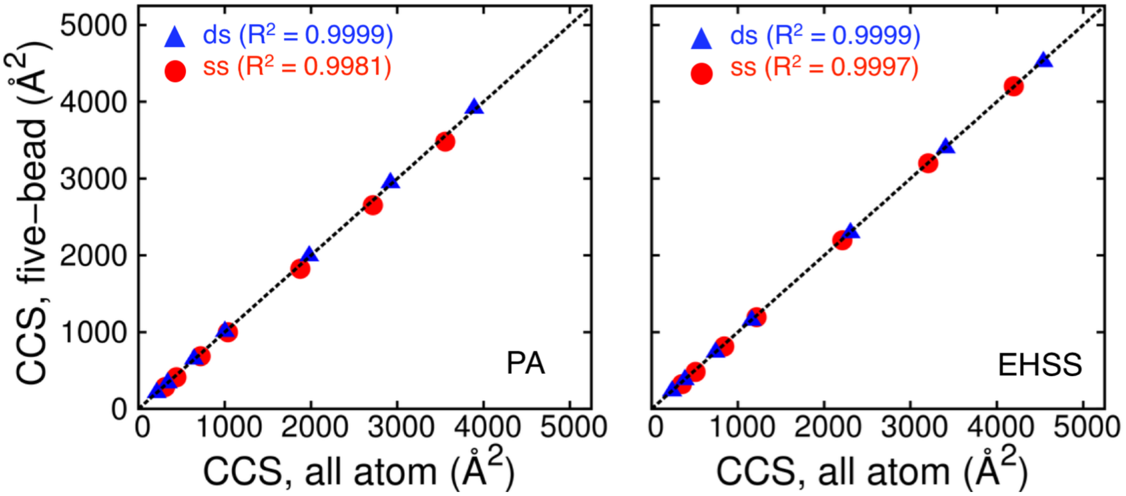

Figure 2S. Comparison of collision cross-section (CCS) values obtained from static all-atom and five-bead models by using the R_{HS2} parameter set. Single-stranded species are marked by blue triangles, while the double-stranded ones are red-circles. The CCS values were obtained by both the PA (left) and EHSS (right) algorithms.

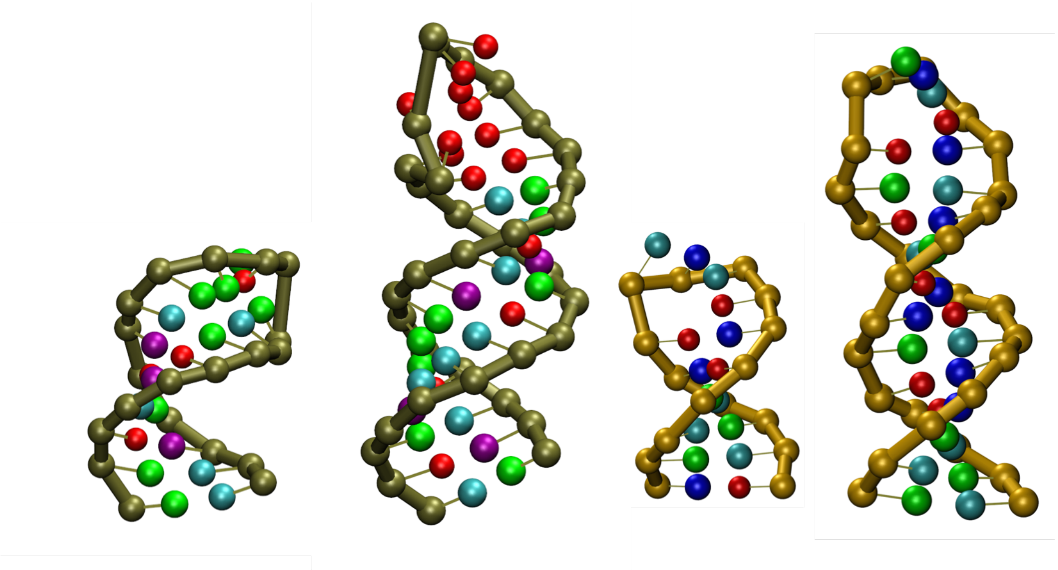

Figure 3S. The simulated hairpins (16mer & 28mer DNA hairpins and 20mer & 34mer RNA hairpins) in CG representation. The backbone is represented in dark (RNA) & light (DNA) tan and the different colored beads represent different nucleobases. (red – A, green – G, cyan – C, magenta – U, blue – T)