Development of Force Field Parameters for the Simulation of Single- and Double-Stranded DNA Molecules and DNA-Protein Complexes

Supporting Information

Maxwell R. Tucker,[1] Stefano Piana,[1] Dazhi Tan,[1] Michael V. LeVine,[1] and David E. Shaw[1,2]

[1] D. E. Shaw Research, New York, NY 10036, USA.

[2] Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA.

Sequence	**T (K)**	**RMSE (Hz)**					
		Amber-bsc1	**DES-Amber SF1.0**	**DES-Amber**	**DES-Amber T**	**DES-Amber 3.20**	
AA[1]	278	1.245	0.449	0.481	1.141	0.68	*0.53*
	295	1.501	0.547	0.586	0.936	0.72	*0.61*
	339	2.157	0.709	0.717	0.495	0.84	*0.62*
AAA[1]	278	0.866	0.586	0.604	0.686	0.77	*0.69*
	284	0.846	0.57	0.566	0.671	0.73	*0.65*
	295	0.91	0.587	0.562	0.686	0.72	*0.65*
	316	1.058	0.595	0.576	0.543	0.71	*0.58*
	350	1.331	0.502	0.519	0.360	0.68	*0.53*
TATA[2]	281	1.441	0.577	0.539	0.317	0.72	*0.52*
	295	1.514	0.592	0.552	0.381	0.74	*0.55*
	317	1.765	0.602	0.609	0.367	0.77	*0.56*
TTA[3]	339	2.362	0.860	0.852	0.488	0.86	*0.59*
CCA[4]	298	3.019	1.208	1.180	0.846	1.08	*1.01*
	348	2.612	1.002	0.92	0.762	0.99	*0.94*
CCAA[4]	298	1.715	1.051	1.021	0.876	1.01	*0.94*

	348	1.962	0.806	0.763	0.620	0.86	*0.82*
TAAT[5,6]	273	1.445	0.794	0.783	0.594	0.94	*0.84*
	280	1.357	0.802	0.758	0.591	0.89	*0.78*
	287	1.360	0.786	0.783	0.582	0.88	*0.73*
	300	1.347	0.726	0.745	0.525	0.84	*0.70*
	316	1.514	0.797	0.787	0.558	0.88	*0.72*
	343	1.65	0.789	0.712	0.517	0.85	*0.68*
ATAT[2]	261	2.233	1.222	1.26	0.806	1.07	*0.85*
	308	2.690	1.729	1.677	1.375	1.31	*1.15*
ATGT[7]	291	1.317	0.798	0.765	0.666	0.88	*0.77*
	339	1.937	0.968	0.924	0.692	0.97	*0.79*
ACATGT[8]	336	1.932	0.990	0.932	0.736	0.98	*0.84*
GCG[9]	296	2.408	0.824	0.794	0.62	0.90	*0.75*
CGCG[3]	282	1.601	0.65	0.643	2.090	0.81	*0.91*
CGCGCG[3]	298	1.205	0.546	0.578	1.272	0.76	*0.81*
CACGTG[10]	338	1.609	0.605	0.615	0.646	0.85	*0.78*
CCAAG[4]	293	1.591	0.666	0.673	0.641	0.82	*0.77*

TGG[4]	298	1.661	0.755	0.718	0.476	0.85	*0.71*
	348	1.773	0.640	0.638	0.478	0.80	*0.67*
TTGG[4]	298	1.861	0.884	0.82	0.66	0.90	*0.76*
	348	1.946	0.837	0.799	0.634	0.86	*0.73*
CTTGG[4]	293	2.019	0.979	0.938	0.631	0.94	*0.88*
Average RMSE (Hz)		1.30	0.89	0.88	0.84	0.87	*0.76*

Table S1. Simulations of short DNA oligomers and comparison to NMR J-couplings. J-couplings were calculated for each system and each temperature using the Karplus relation from Hasnoot et al.[11,12] and compared to the experimentally determined J-couplings.[1–3,5–10] DES-Amber T is the variant of the DES-Amber force field where all of the negative charge of the 3'-terminal residues is localized on this residue, rather than split between the 5'- and 3'-terminal residues as is customary in the Amber force fields. For DES-Amber 3.20 RMS errors are reported before and after (italics) the refit of backbone torsion parameters.

		Exp	Amber-bsc1		DES-Amber		DES-Amber SF1.0		DES-Amber 3.20	
Sequence	**Salt**	**ΔG_{310}**	**#**	**ΔG_{310}**	**#**	**ΔG_{310}**	**#**	**ΔG_{310}**	**#**	**ΔG_{310}**
d(CACAG)	NaCl	-2.7	32	-1.8(2)	12	-0.8(4)	4	0.0(6)	12	-1.8(4)
d(CGCGG)	NaCl	-6.5	18	-6.4(5)	32	-7.5(4)	30	-5.6(4)	27	-7.4(5)
d(CGTACG)	KCl	-4.6	7	-1.3(1.2)	11	-2.7(5)	8	-1.9(9)	4	-3(1)
d(GAGTGAG)	NaCl	-4.7	19	-4.3(4)	2	-1.5(2.0)	6	-3.0(6)	10	-4.6(8)
d(GCGC)	KCl	-3.9	84	-2.8(2)	26	-4.1(4)	28	-5.5(6)	30	-4.9(3)
d(GGATCC)	KCl	-4.0	8	-3.0(5)	3	-3.0(6)	7	-3.2(1.1)	4	-3.2(5)
d(TCATGA)	KCl	-3.1	4	0.7(7)	3	0.8(1.0)	4	0.2(1.0)	5	-1.3(1.0)
RMSD (kcal mol^{-1})			2.0		2.1		2.0		1.1	

Table S2. Simulations of dsDNA reversible duplex formation. For each sequence and each force field, the number of melting events and the calculated thermodynamic stability (kcal mol^{-1}) at 310 K and 1 M oligomer concentration are reported.

	Amber-bsc1		DES-Amber		DES-Amber SF1.0		DES-Amber 3.20	
Sequence	**%Diss**	**<#C>**	**%Diss**	**<#C>**	**%Diss**	**<#C>**	**%Diss**	**<#C>**
d(CACAG)	8	34	65	7	65	7	58	10
d(CGCGG)	3	51	69	7	62	8	59	10
d(CGTACG)	3	45	67	8	58	9	59	11
d(GAGTGAG)	5	49	71	7	68	6	61	10
d(GCGC)	11	30	65	8	65	8	63	9
d(GGATCC)	12	35	77	4	71	5	69	7
d(TCATGA)	4	64	69	7	72	5	65	8

Table S3. Simulations of dsDNA reversible duplex formation. For each sequence and each force field the fraction of structures where the two DNA strands are non-interacting %Diss (defined as frames where there are no contacts between the strands) and the average number of contacts <#C> (defined as two heavy atoms with a distance < 5 Å) between DNA strands in non B-DNA structures is reported.

Sequence	**PDB ID**	**RMSD (Å) Amber-bsc1**		**RMSD (Å) DES-Amber**	**RMSD (Å) DES-Amber SF1.0**	**RMSD (Å) DES-Amber 3.20**
		This work	**Ref. 13**			
d(CGCGAATTCGCG)	1bna	2.16	2.08	2.26	2.42	2.42
	1naj	1.86	1.84	1.86	2.09	1.95
	1duf	1.76		1.69	1.84	1.83
d(GGATATATCC)	2lwg	1.83	1.92	1.66	*18.32*	1.64
d(CTTTTAAAAG)	1sk5	1.99	1.98	1.82	4.46	1.93
d(GTCCαAGGACG)	2lib	1.97		1.94	2.02	2.40
d(GGCAAAACGG)	1nev	2.74		2.12	2.15	2.04
d(CGCGAAAAAACG)	1d89	2.50	2.18	2.35	2.46	2.44
d(GGCAAAAAACGG)	1fzx	1.79	2.46	1.66	1.84	2.85
d(GGCAAGAAACGG)	1g14	1.87		1.74	1.84	1.79

d(GCGAGATCTGCG)	1opq	2.29		2.40	2.61	2.94
d(CTCGGCGCCATC)	2hkb	2.06	2.59	2.21	2.05	2.04
d(CCTCTGGTCTCC)	2k0v	2.33	2.04	2.52	2.60	2.66
d(CGCATGCTACGC)	2l8q	1.82	2.60	1.70	1.82	1.75
d(GCAAAATTTTGC)	1rvh	1.96	2.28	1.60	1.77	1.65
d(GCATCGATTGGC)	5uzd	1.70		1.80	1.84	1.91
d(GCGCATGCTACGCG)	2m2c	2.44	2.18	2.87	2.89	2.99
d(CGAGGTTTAAACCTCG)	1ss7	2.78		2.77	2.86	2.90
d(GGAAAATCTCTAGCAGT)	1tqr	3.92		3.53	3.69	3.74
Average RMSD		**2.20**	**2.24**	**2.13**	**2.40**	**2.30**

Table S4. Long DNA oligomers simulations. The average heavy atoms RMSD (Å) with respect to the starting PDB structure was calculated for the last 49 μs of each 50 μs simulation. For the Drew-Dickerson dodecamer, three simulations were performed starting from the X-ray (1bna) or NMR (1d89, 1fxz) structures. RMSD data are also reported for the Amber-bsc1 simulations performed by other authors and deposited in the BIGNASim database.[13] In the DES-Amber SF1.0 simulation of 2lwg the double helix is unstable and the two strands dissociate, as indicated by the high value of the RMSD of 18.32 Å. This data point was not used to compute the RMSD average over all simulations for the DES-Amber SF1.0 force field. For the DES-

Amber 3.20 force field 10 μs of simulation was performed and the last 9.9 μs was used for data analysis.

σ_{OW} =	**DES-A**	**3.18**	***3.18 refit***	**3.20**	***3.20 refit***	***3.20 NOE***	**3.23**	***3.23 refit***	**3.25**	***3.25 refit***
AAAA	1.01	1.58	*0.89*	1.72	*1.37*	*0.93*	1.50	*0.85*	1.65	*0.87*
CCCC	0.77	1.07	*0.58*	1.16	*0.59*	*0.58*	1.34	*0.60*	1.40	*0.63*
CAAU	1.07	1.36	*0.79*	1.38	*0.73*	*0.69*	1.51	*0.68*	1.50	*0.67*
GACC	0.71	0.96	*0.67*	0.95	*0.68*	*0.68*	1.09	*0.66*	1.04	*0.78*
UUUU	1.34	1.58	*1.48*	1.32	*1.03*	*1.27*	1.27	*1.13*	1.26	*0.86*
Average	0.98	1.31	*0.88*	1.42	*0.88*	*0.83*	1.34	*0.78*	1.37	*0.76*

Table S5. Deviations from experimental J-couplings for simulations of ssRNA tetranucleotides. RMS error from the experimental data (Hz) for J-couplings calculated for each system using the Karplus relation from refs. 12, 14, and 15. Experimental data was taken from refs. 16, 17, and 18. For each value of σ_{OW} tested, deviations from experiment are reported before and after the refit. For σ_{OW} = 3.20 we also report scalar coupling deviations for a torsion optimization where NOEs and nNOEs data was also used (3.20 NOE).

σ_{OW} =	**3.18**	**3.20**	**3.23**	**3.25**
α	0.39	0.40	0.65	0.97
β	0.09	0.10	0.36	0.29
γ	0.11	0.19	0.32	0.34
δ	0.04	0.08	0.10	0.24
ε	0.62	0.56	0.73	0.66
ζ	0.34	0.38	0.56	0.49

Table S6. Magnitude of torsion corrections from refitting based on experimental NMR J-coupling data. Sum of coefficients of torsion corrections (kcal mol^{-1}) for each dihedral angle obtained from refitting based on NMR data.

	DES-Amber 3.20 DNA						DES-Amber 3.20 RNA					
Torsion Potentials												
	k_1	φ_1	k_2	φ_2	k_3	φ_3	k_1	φ_1	k_2	φ_2	k_3	φ_3
α	0.0311	20.916	0.0401	-74.048	-0.0579	54.007	0.0465	33.139	-0.3458	-24.291	0.1910	-24.931
β	0.0040	83.050	0.0047	-75.487	-0.0084	-49.626	-0.2103	-80.789	0.0941	-14.932	-0.1329	87.223
γ	0.0138	73.641	-0.0182	-16.793	-0.0193	65.042	0.2073	-5.375	-0.1825	-59.290	0.0997	9.298
δ	-0.0673	-0.949	0.1476	-85.449	0.2277	16.127	-0.0400	-59.584	-0.1333	43.836	0.2207	-29.652
ε	0.0097	-36.802	-0.0146	28.646	0.0112	71.663	-0.2033	-47.991	0.2285	-6.411	-0.0310	-31.882
ζ	-0.0048	85.879	-0.0018	21.835	-0.0087	-86.289	-0.3159	-83.919	0.2163	59.267	0.0871	-40.126
Lennard-Jones pairs												
	σ	ε					σ	ε				
OP/OW	3.2000	0 21681					3.2000	0.21681				
CA/OW	3.2825	0 10000					3.2825	0.10000				
MG/OW	2.9000	0.025914					2.9000	0.025914				

Atomic Charges		
C2'		0.1670
O2'		-0.6839
HO2'		0.3886

Table S7. Modified force field parameters for DES-Amber 3.20. List of parameters that were modified with respect to standard DES-Amber during the development of DES-Amber 3.20. Torsion corrections for the α, β, γ, δ, ε, and ζ torsion angles were obtained from refitting based on experimental NMR J-coupling data as $V(\varphi) = \sum_{n=1}^{3} a_n sin(\varphi) + b_n \cos(\varphi)$ and transformed into the cosine expansion $V(\varphi) = \sum_{n=1}^{3} k_n cos(n\varphi - \varphi_n)$ that is commonly used in the Amber force field (and those transformed values are reported here).

		Exp[19,20]	DES-Amber 3.20		
Sequence	**[NaCl]**	**ΔG_{310}**	**t (μs)**	**#**	**ΔG_{310}**
CACAG	0 M	-3.6	500	18	-4.2(3)
CACAG	1 M	-4.7	500	12	-4.1(3)
CGCGG	0 Ml	-6.2	500	21	-6.8(3)
CGCGG	1 Ml	-7.3	500	16	-6.6(5)
UAAGGUA	0 M	-5.4	180	5	-5.2(7)
UAAGGUA	1 M	-6.95	150	7	-5.7(5)

Table S8. Simulations of dsRNA reversible duplex formation. For each sequence, the total simulation time (μs), number of melting events, and the calculated thermodynamic stability (kcal mol^{-1}) at 310 K are reported and compared against polymer model estimates at two different salt concentrations. The free energies in the presence of counterions only (0 M) were estimated from the polymer model at 0.1 M salt concentration.[19]

	Shear (Å)		**Stretch (Å)**		**Stagger (Å)**		**Buckle (deg)**		**Propeller (deg)**		**Opening (deg)**	
	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD
G-C	-0.02	0.36	0.20	0.25	0.09	0.44	0.78	9.54	0.68	9.19	0.62	2.75
T-A	0.02	0.27	0.15	0.19	-0.04	0.27	-0.70	6.55	-0.39	8.08	1.10	3.90
	Shift (Å)		**Slide (Å)**		**Rise (Å)**		**Tilt (deg)**		**Roll (deg)**		**Twist (deg)**	
AT/TA	0.12	0.35	0.00	0.32	-0.05	0.30	-0.04	1.74	3.95	5.73	-0.40	2.91
CG/GC	0.01	0.38	0.17	0.80	-0.19	0.39	-0.81	2.89	-0.24	8.16	0.85	3.51
CG/TA	0.02	0.43	-0.26	0.72	0.02	0.37	0.11	4.39	2.73	6.85	-0.24	4.48
GC/CG	-0.10	0.45	-0.68	0.85	0.10	0.33	-0.22	2.98	2.34	4.82	-6.66	8.14
GC/GC	0.13	0.56	-1.15	1.35	0.26	0.35	1.31	3.60	1.83	6.42	-2.05	5.28
GC/TA	0.12	0.53	-0.31	0.58	0.15	0.27	0.50	2.78	0.23	3.74	-3.10	4.23
TA/AT	0.04	0.18	0.17	0.81	0.07	0.15	0.28	1.18	-1.15	3.35	0.81	4.53
TA/CG	-0.14	0.49	-0.51	0.66	0.31	0.45	-0.90	3.63	0.27	3.08	-2.99	4.44

TA/GC	-0.02	0.49	-0.07	1.13	0.08	0.23	0.25	3.94	2.79	5.59	-1.47	6.95
TA/TA	-0.06	0.38	-0.22	0.51	0.13	0.20	-0.11	2.41	0.53	3.78	-1.14	2.98

Table S9. DES-Amber simulations of DNA duplexes. Base-pair and base-step parameters deviations from the reference X-ray structure. For each simulation, the average base-step and base-pair parameters were calculated. The RMSD deviation and mean signed deviation of the average MD values from the reference X-ray structure was computed for all base-airs and all base steps of each sequence. Here we report the average RMSD and MSD aggregated across all simulations by base-pair type and base-step type.

	Shear (Å)		**Stretch (Å)**		**Stagger (Å)**		**Buckle (deg)**		**Propeller (deg)**		**Opening (deg)**	
	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD
G-C	-0.02	0.36	0.20	0.25	0.09	0.44	0.73	9.56	0.60	9.17	0.63	2.74
T-A	0.02	0.27	0.15	0.19	-0.07	0.28	-0.64	6.42	-0.61	8.11	1.08	3.86
	Shift (Å)		**Slide (Å)**		**Rise (Å)**		**Tilt (deg)**		**Roll (deg)**		**Twist (deg)**	
AT/TA	0.11	0.35	0.00	0.33	-0.06	0.31	-0.09	1.73	4.10	6.08	-0.52	2.90
CG/GC	0.03	0.39	0.16	0.79	-0.17	0.37	-0.64	2.59	0.16	7.94	0.88	3.47
CG/TA	0.02	0.43	-0.25	0.71	0.01	0.37	0.15	4.43	3.15	7.10	-0.34	4.51
GC/CG	-0.11	0.46	-0.68	0.85	0.09	0.33	-0.24	3.04	2.42	4.88	-6.66	8.18
GC/GC	0.12	0.58	-1.18	1.37	0.26	0.35	1.40	3.61	2.00	6.58	-2.41	5.60
GC/TA	0.12	0.55	-0.31	0.58	0.14	0.27	0.48	2.78	0.62	3.72	-3.36	4.41
TA/AT	0.08	0.21	0.27	0.78	0.07	0.14	0.20	1.11	-0.58	3.06	1.06	4.53
TA/CG	-0.13	0.49	-0.53	0.68	0.31	0.45	-0.92	3.61	0.71	3.20	-3.33	4.65
TA/GC	-0.01	0.49	-0.06	1.13	0.08	0.23	0.30	3.97	3.13	5.77	-1.48	6.94

TA/TA	-0.06	0.38	-0.21	0.50	0.14	0.21	-0.12	2.44	1.26	3.94	-1.44	3.16

Table S10. DES-Amber 1.0 simulations of DNA duplexes. Base-pair and base-step parameters deviations from the reference X-ray structure. For each simulation, the average base-step and base-pair parameters were calculated. The RMSD deviation and mean signed deviation of the average MD values from the reference X-ray structure was computed for all base-airs and all base steps of each sequence. Here we report the average RMSD and MSD aggregated across all simulations by base-pair type and base-step type.

	Shear (Å)		Stretch (Å)		Stagger (Å)		Buckle (deg)		Propeller (deg)		Opening (deg)	
	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD
G-C	-0.02	0.36	0.20	0.25	0.08	0.44	0.57	9.38	0.29	9.18	0.71	2.80
T-A	0.02	0.26	0.15	0.19	-0.07	0.28	-0.42	6.19	-0.79	8.01	1.29	3.90
	Shift (Å)		Slide (Å)		Rise (Å)		Tilt (deg)		Roll (deg)		Twist (deg)	
AT/TA	0.13	0.36	0.01	0.33	-0.06	0.30	-0.12	1.65	4.61	6.21	-0.70	2.92
CG/GC	0.01	0.37	0.23	0.82	-0.19	0.40	-0.83	3.02	0.16	8.14	0.76	3.45
CG/TA	0.01	0.44	-0.23	0.70	0.01	0.36	0.12	4.46	3.51	7.10	-0.58	4.52
GC/CG	-0.11	0.47	-0.71	0.88	0.10	0.33	-0.19	2.97	2.37	4.86	-6.92	8.37
GC/GC	0.15	0.57	-1.20	1.38	0.27	0.35	1.25	3.46	2.33	6.60	-2.76	5.66
GC/TA	0.07	0.52	-0.32	0.57	0.14	0.26	0.48	2.81	0.70	3.66	-3.46	4.49
TA/AT	0.06	0.20	0.21	0.71	0.09	0.15	0.27	1.14	0.11	3.06	0.19	4.36
TA/CG	-0.13	0.49	-0.60	0.74	0.32	0.45	-0.86	3.58	0.79	3.21	-3.85	5.09
TA/GC	-0.02	0.49	0.01	1.17	0.07	0.23	0.27	4.00	3.30	5.92	-1.53	7.02

TA/TA	-0.06	0.38	-0.18	0.49	0.14	0.21	-0.14	2.38	1.33	3.96	-1.42	3.15

Table S11. DES-Amber 3.20 simulations of DNA duplexes. Base-pair and base-step parameters deviations from the reference X-ray structure. For each simulation, the average base-step and base-pair parameters were calculated. The RMSD deviation and mean signed deviation of the average MD values from the reference X-ray structure was computed for all base-airs and all base steps of each sequence. Here we report the average RMSD and MSD aggregated across all simulations by base-pair type and base-step type.

	Shear (Å)		**Stretch (Å)**		**Stagger (Å)**		**Buckle (deg)**		**Propeller (deg)**		**Opening (deg)**	
	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD
G-C	-0.03	0.38	0.21	0.25	0.07	0.42	1.17	9.58	2.41	9.54	-0.24	2.75
T-A	0.02	0.27	0.18	0.21	0.01	0.27	-0.03	7.44	1.14	8.87	0.90	3.77
	Shift (Å)		**Slide (Å)**		**Rise (Å)**		**Tilt (deg)**		**Roll (deg)**		**Twist (deg)**	
AT/TA	0.11	0.30	-0.10	0.34	-0.02	0.29	-0.04	1.70	2.75	5.33	-0.20	3.19
CG/GC	0.06	0.44	0.06	0.80	-0.07	0.36	-0.49	2.95	-3.05	8.88	0.49	4.17
CG/TA	0.10	0.51	-0.22	0.75	0.07	0.40	0.02	4.29	0.81	6.21	0.03	4.46
GC/CG	-0.08	0.44	-0.08	0.51	0.01	0.32	-0.28	2.84	-0.88	4.36	-2.85	5.40
GC/GC	0.02	0.47	-0.75	1.09	0.30	0.38	1.28	3.69	0.63	6.22	0.71	5.01
GC/TA	0.25	0.54	-0.25	0.63	0.15	0.32	0.71	2.39	-1.34	3.59	-2.18	3.89
TA/AT	-0.11	0.34	0.25	0.81	0.00	0.11	0.00	0.71	-2.03	3.79	0.26	5.74
TA/CG	-0.17	0.58	0.13	0.45	0.26	0.42	-0.72	2.85	-1.31	3.40	0.42	3.31
TA/GC	-0.05	0.54	-0.27	1.14	0.09	0.26	-0.10	3.54	1.30	4.95	-3.92	7.97
TA/TA	-0.11	0.45	-0.22	0.53	0.15	0.22	-0.54	2.65	1.20	4.00	-1.52	3.49

Table S12. Amber-bsc1 simulations of DNA duplexes. Base-pair and base-step parameters deviations from the reference X-ray structure. For each simulation, the average base-step and base-pair parameters were calculated. The RMSD deviation and mean signed deviation of the average MD values from the reference X-ray structure was computed for all base-airs and all base steps of each sequence. Here we report the average RMSD and MSD aggregated across all simulations by base-pair type and base-step type.

	Shear (Å)		**Stretch (Å)**		**Stagger (Å)**		**Buckle (deg)**		**Propeller (deg)**		**Opening (deg)**	
	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD
G-C	-0.01	0.37	0.21	0.26	0.04	0.41	0.67	9.10	0.77	9.39	-0.10	2.68
T-A	0.02	0.27	0.18	0.21	-0.03	0.28	0.45	8.19	-1.66	8.80	1.68	3.99
	Shift (Å)		**Slide (Å)**		**Rise (Å)**		**Tilt (deg)**		**Roll (deg)**		**Twist (deg)**	
AT/TA	0.10	0.30	0.05	0.33	-0.02	0.29	-0.06	1.63	3.21	5.61	-0.30	3.28
CG/GC	0.08	0.49	0.46	0.89	-0.00	0.35	-0.43	2.79	-2.65	8.90	2.47	4.76
CG/TA	0.15	0.56	0.08	0.70	0.09	0.40	0.05	4.38	2.13	6.66	1.03	4.89
GC/CG	-0.06	0.41	0.20	0.57	-0.09	0.33	-0.06	3.08	-0.68	4.43	-4.16	6.21
GC/GC	-0.05	0.54	-0.10	0.85	0.21	0.28	0.80	2.76	0.87	6.54	0.87	5.34
GC/TA	0.17	0.51	-0.03	0.51	0.12	0.29	0.66	2.72	-1.03	3.41	-2.47	4.14
TA/AT	-0.11	0.22	0.67	0.93	-0.02	0.11	-0.09	0.94	-1.36	3.37	0.19	5.67
TA/CG	-0.16	0.54	0.44	0.63	0.19	0.38	-0.49	2.77	-0.38	3.26	-0.22	3.24
TA/GC	-0.04	0.45	-0.01	1.12	0.12	0.29	-0.02	3.42	1.78	5.15	-3.19	8.18
TA/TA	-0.10	0.45	0.09	0.47	0.15	0.22	-0.42	2.63	2.39	4.45	-1.16	3.39

Table S13. Amber-OL15 simulations in TIP3P water of DNA duplexes. Base-pair and base-step parameters deviations from the reference X-ray structure. For each simulation, the average base-step and base-pair parameters were calculated. The RMSD deviation and mean signed deviation of the average MD values from the reference X-ray structure was computed for all base-airs and all base steps of each sequence. Here we report the average RMSD and MSD aggregated across all simulations by base-pair type and base-step type.

	Shear (Å)		**Stretch (Å)**		**Stagger (Å)**		**Buckle (deg)**		**Propeller (deg)**		**Opening (deg)**	
	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD	MSD	RMSD
G-C	-0.02	0.37	0.21	0.26	0.04	0.41	0.67	9.11	0.92	9.42	-0.15	2.66
T-A	0.02	0.26	0.18	0.21	0.02	0.28	-0.23	7.82	-1.12	8.64	1.59	4.02
	Shift (Å)		**Slide (Å)**		**Rise (Å)**		**Tilt (deg)**		**Roll (deg)**		**Twist (deg)**	
AT/TA	0.10	0.32	0.06	0.31	-0.03	0.29	-0.02	1.69	2.82	5.22	-0.89	3.37
CG/GC	0.09	0.53	0.45	0.87	-0.01	0.37	-0.32	2.84	-2.36	8.96	2.71	4.99
CG/TA	0.14	0.57	0.07	0.69	0.07	0.41	0.04	4.41	1.96	6.43	1.01	4.80
GC/CG	-0.06	0.42	0.14	0.56	-0.10	0.32	-0.10	3.11	-0.29	4.23	-4.52	6.40
GC/GC	-0.07	0.57	-0.06	0.87	0.19	0.27	0.75	2.69	1.00	6.51	1.01	5.28
GC/TA	0.16	0.47	-0.02	0.52	0.10	0.27	0.74	2.73	-1.42	3.79	-2.43	4.12
TA/AT	-0.10	0.25	0.54	0.85	-0.05	0.11	-0.03	1.03	-1.96	3.87	0.07	5.60
TA/CG	-0.17	0.53	0.34	0.58	0.18	0.38	-0.31	2.88	-0.90	3.39	0.01	3.36
TA/GC	-0.06	0.46	-0.01	1.08	0.09	0.27	-0.11	3.46	1.51	4.99	-2.75	7.80
TA/TA	-0.08	0.45	-0.00	0.46	0.13	0.21	-0.42	2.65	0.99	3.95	-0.74	3.10

Table S14. Amber-OL15 simulations in OPC water of DNA duplexes. Base-pair and base-step parameters deviations from the reference X-ray structure. For each simulation, the average base-step and base-pair parameters were calculated. The RMSD deviation and mean signed deviation of the average MD values from the reference X-ray structure was computed for all base-airs and all base steps of each sequence. Here we report the average RMSD and MSD aggregated across all simulations by base-pair type and base-step type.

	X-ray structure		Amber-bsc1		Amber-OL15		DES-Amber	
	Rise	**Twist**	**Rise**	**Twist**	**Rise**	**Twist**	**Rise**	**Twist**
1d89	3.303	36.130	3.354	34.528	3.355	35.025	3.328	34.326
1duf	3.413	34.679	3.341	34.662	3.339	35.071	3.325	33.742
1fzx	3.001	34.543	3.366	33.973	3.338	34.188	3.340	33.580
1g14	3.046	34.211	3.370	33.970	3.342	34.269	3.342	33.189
1naj	3.177	35.856	3.338	34.719	3.335	35.195	3.326	33.726
1nev	3.298	36.403	3.372	33.958	3.341	34.251	3.354	33.269
1opq	3.511	34.537	3.346	34.471	3.333	34.723	3.335	33.092
1qc1	3.374	35.274	3.349	33.958	3.333	34.787	3.348	32.498
1rvh	3.388	35.299	3.340	33.831	3.317	33.900	3.340	33.785
1sk5	3.229	37.509	3.334	34.367	3.320	34.671	3.324	34.827
1ss7	3.193	34.866	3.356	34.002	3.344	34.379	3.321	33.463
1tqr	3.111	37.537	3.359	34.204	3.339	34.369	3.337	33.350
2hkb	3.041	34.186	3.350	34.415	3.324	34.629	3.334	32.722
2k0v	3.271	34.703	3.388	33.945	3.362	34.385	3.364	32.595
2l8q	3.260	33.111	3.339	34.366	3.334	34.582	3.346	33.063
2lib	3.223	33.416	3.370	33.303	3.297	33.163	3.390	32.107
2lwg	3.250	32.008	3.397	34.358	3.364	34.512	3.383	32.799
2m2c	3.279	37.711	3.339	34.391	3.330	34.498	3.351	33.084
3ggb	3.293	35.584	3.350	33.493	3.326	34.146	3.349	32.459
5uzd	3.416	33.195	3.359	34.089	3.332	34.204	3.363	32.476
avg	**3.254**	**35.038**	**3.356**	**34.150**	**3.335**	**34.447**	**3.345**	**33.208**

Table S15. Average helical Rise and Twist in simulations of DNA duplexes. For each simulation, the average Rise (Å) and Twist (degrees) parameters were calculated using DSSR. The X-ray values are reported for reference, with the corresponding PDB ID. Simulations performed with Amber-OL15 and the OPC water model gave very similar results as Amber-OL15, and simulations performed with DES-Amber 1.0 and DES-Amber 3.20 gave very similar results as DES-Amber, and are not reported in this table.

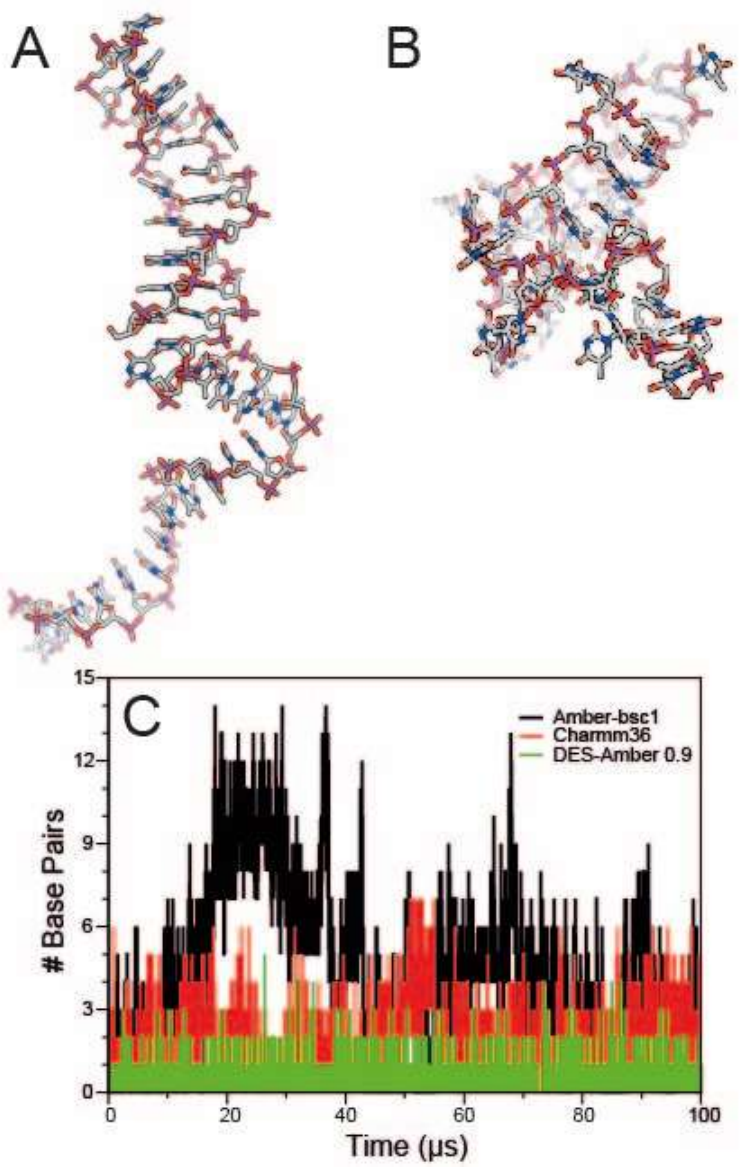

Figure S1. dT40 simulations. Last frame from the 100 μs simulations of dT40 in 0.4 M NaCl performed with the Amber-bsc1 (A) and CHARMM36 (B) force fields. (C) Number of base-pairs formed as a function of time in the 1M NaCl simulations as calculated using DSSR.

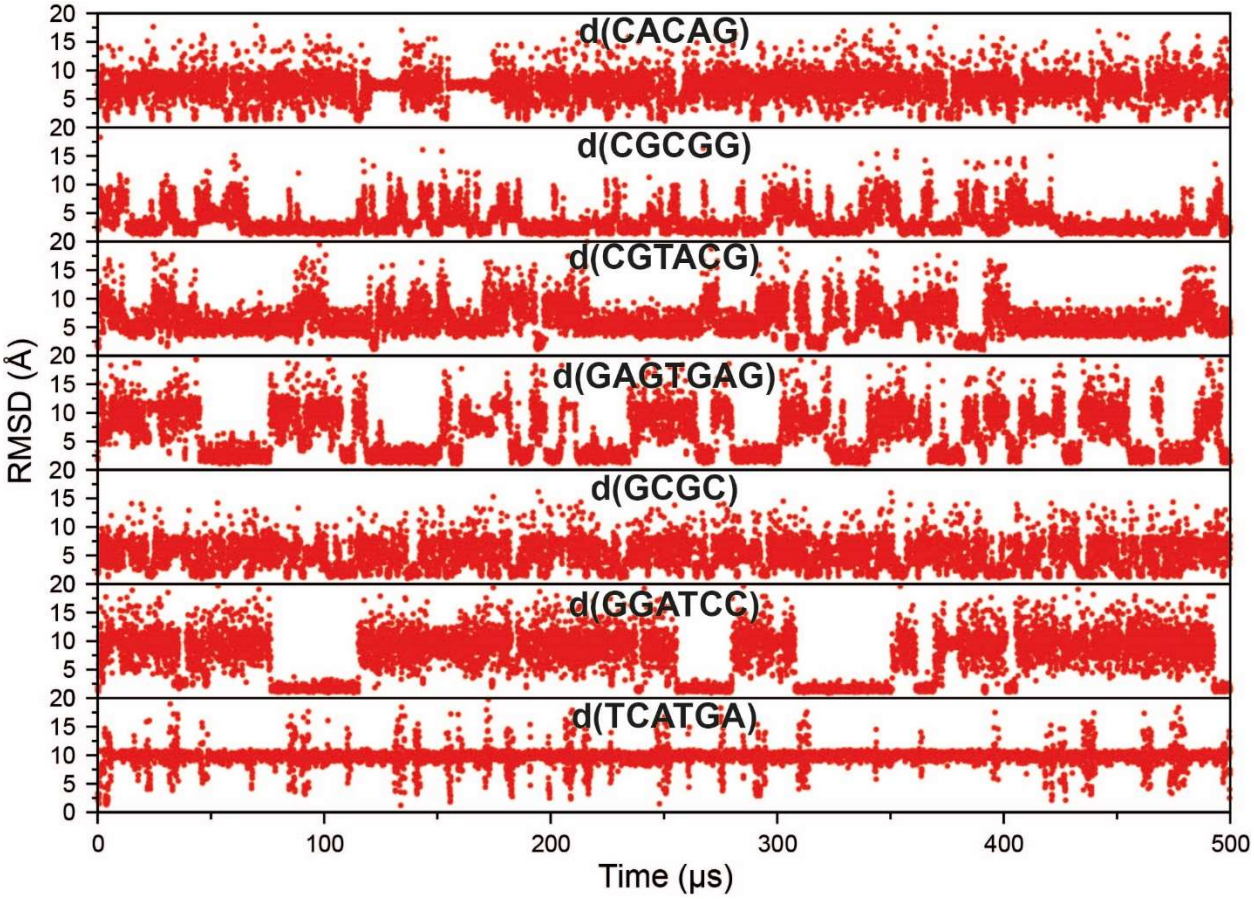

Figure S2. Simulated tempering simulations of DNA duplex formation performed with the Amber-bsc1 force field. RMSD traces with respect to the ideal B-DNA structure as a function of simulation time for the 7 systems investigated. The oligonucleotide sequences are reported for each system.

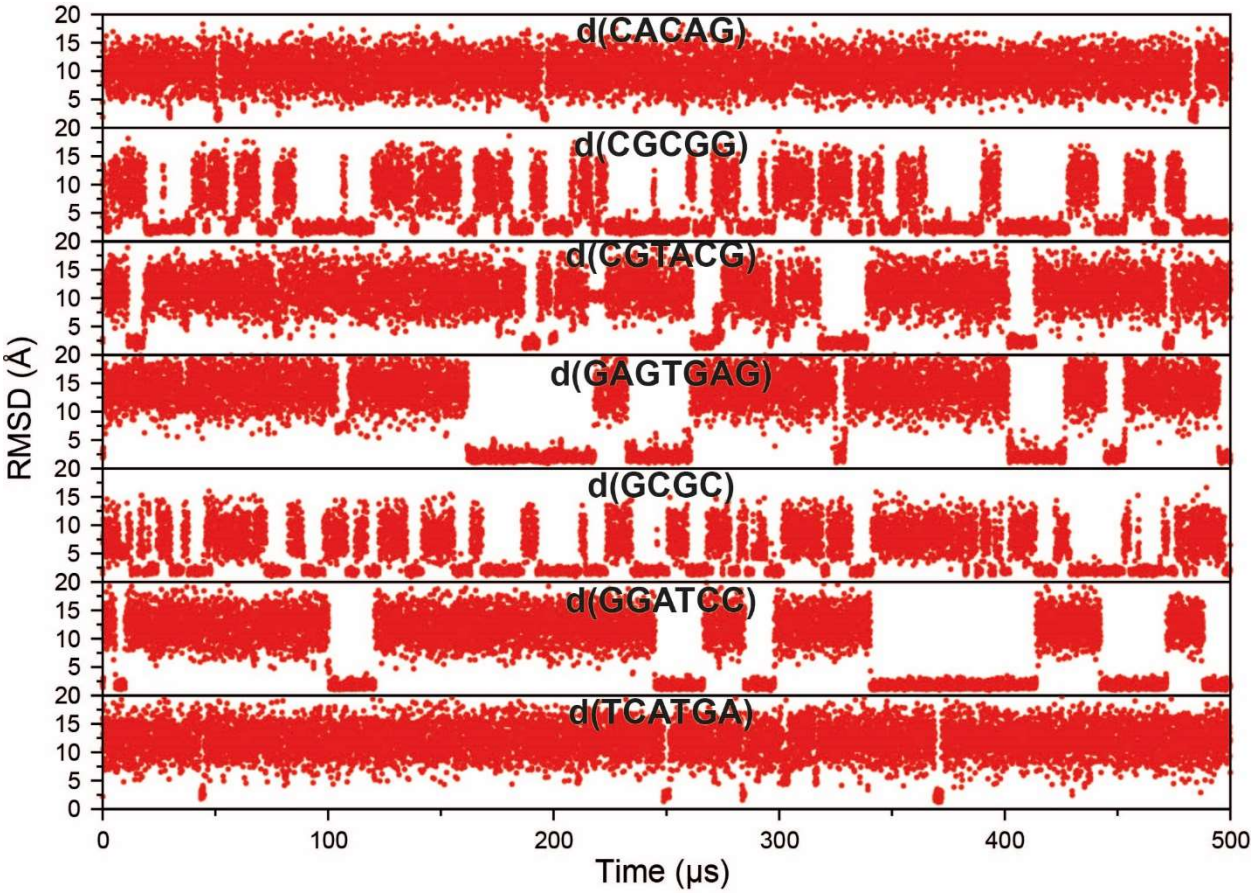

Figure S3. Simulated tempering simulations of DNA duplex formation performed with the DES-Amber SF1.0 force field. RMSD traces with respect to the ideal B-DNA structure as a function of simulation time for the 7 systems investigated. The oligonucleotide sequences are reported for each system.

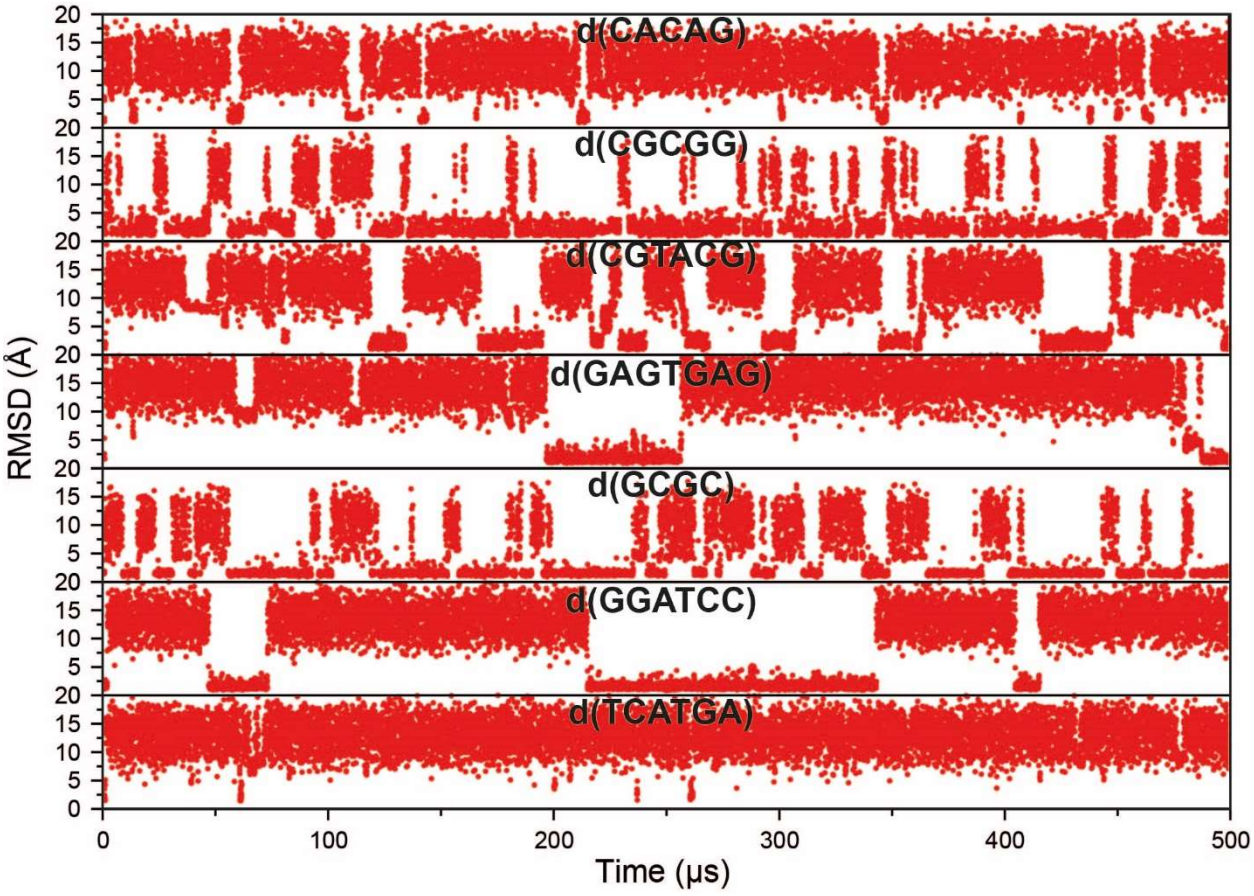

Figure S4. Simulated tempering simulations of DNA duplex formation performed with the DES-Amber force field. RMSD traces with respect to the ideal B-DNA structure as a function of simulation time for the 7 systems investigated. The oligonucleotide sequences are reported for each system.

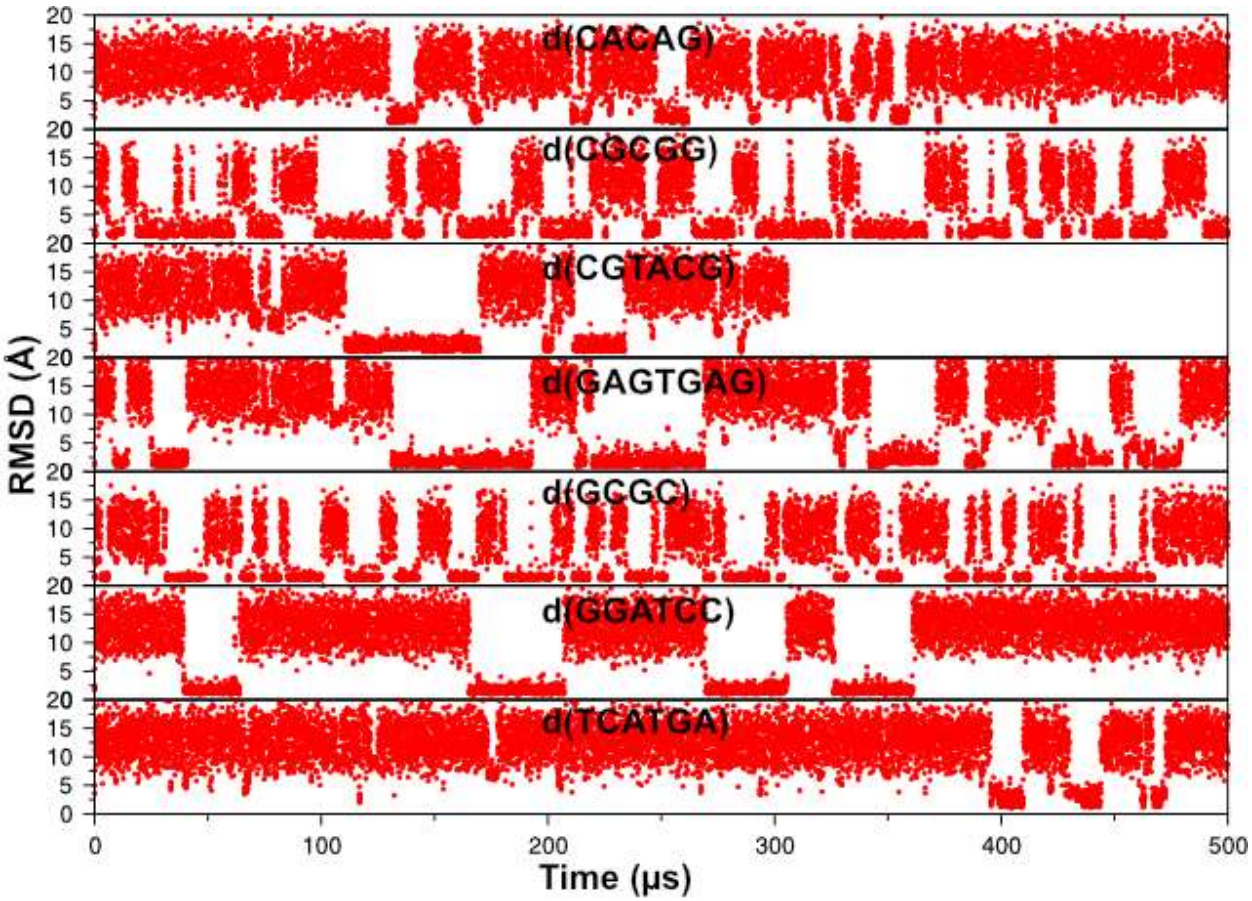

Figure S5. Simulated tempering simulations of DNA duplex formation performed with the DES-Amber 3.20 force field. RMSD traces with respect to the ideal B-DNA structure as a function of simulation time for the 7 systems investigated. The oligonucleotide sequences are reported for each system.

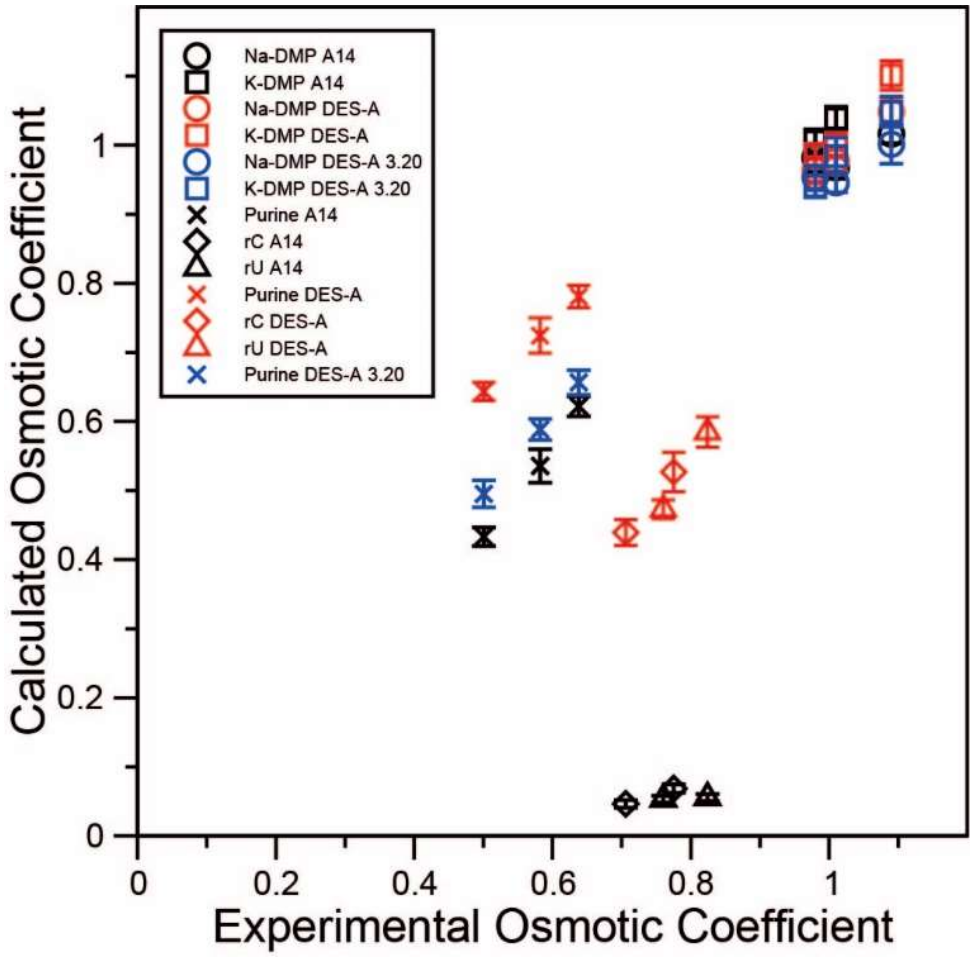

Figure S6. Comparison of calculated and experimental[21,22] osmotic coefficients for DMP, Purine, Cytidine and Uridine. Osmotic coefficients were calculated for solutions of Na^+-DMP (circles), K^+-DMP (squares), and Purine (crosses) using the Amber (black), DES-Amber (red), and DES-Amber 3.20 (blue) force fields. For solutions of Cytidine (diamonds) and Uridine (triangles), calculations were performed using the Amber (black) and DES-Amber (red) force fields only as, for these two systems, DES-Amber 3.20 is identical to DES-Amber.

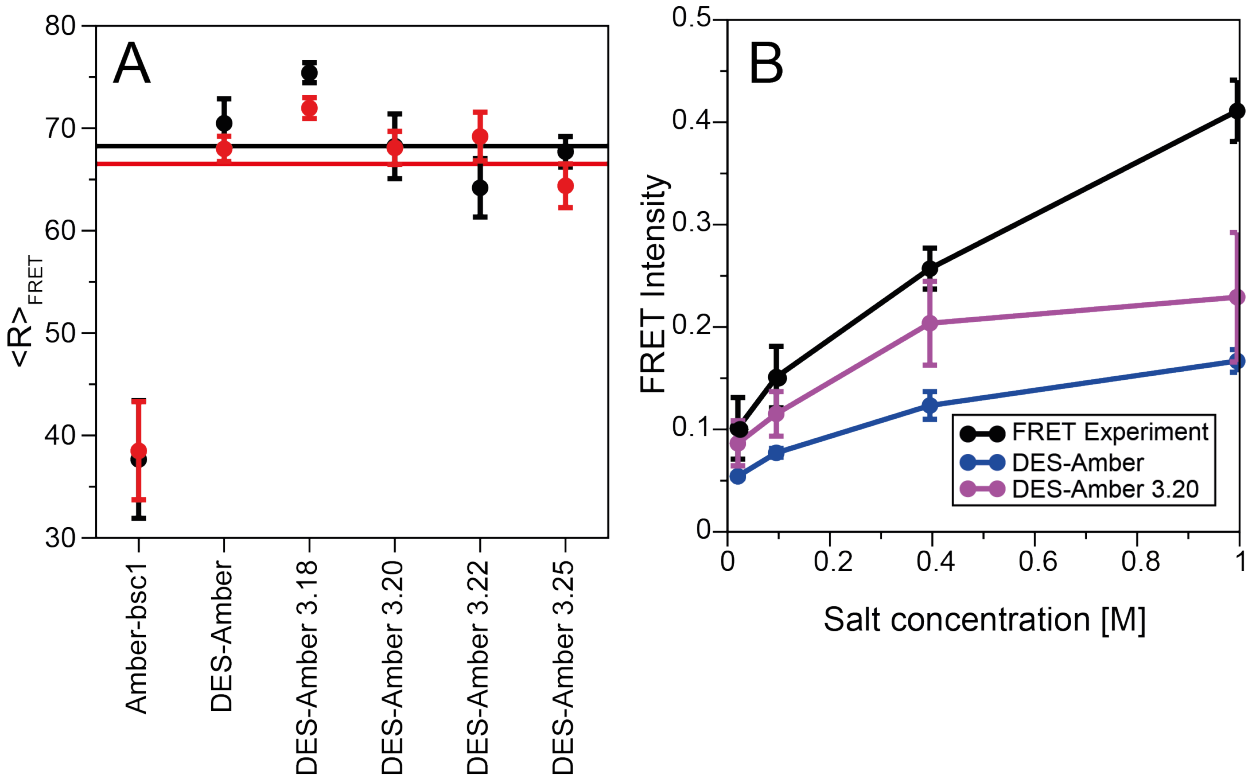

Figure S7. Comparison of the disordered states of rU40 (A) and dT40 (B) simulated with different force fields with FRET data.[23] (A) The average FRET end-to-end distance ($\langle R \rangle_{FRET}$) was calculated at 0.05 M (black) and 0.1 M (red) NaCl concentration using simulations with values of the sigma parameter for the LJ interaction between water and the phosphate oxygen ranging from 3.18 to 3.25. Results from simulations performed using Amber-bsc1 in OPC water and DES-Amber are also shown. The experimental value is shown as a solid line. (B) FRET intensity for dT40 was calculated from simulations at NaCl concentrations ranging from 0.05 M to 1.0 M using the DES-Amber and the DES-Amber 3.20 force fields, compared to the experimental data. The Förster radius was assumed to be 55 Å for rU40 and 56.4 Å for dT40.[23]

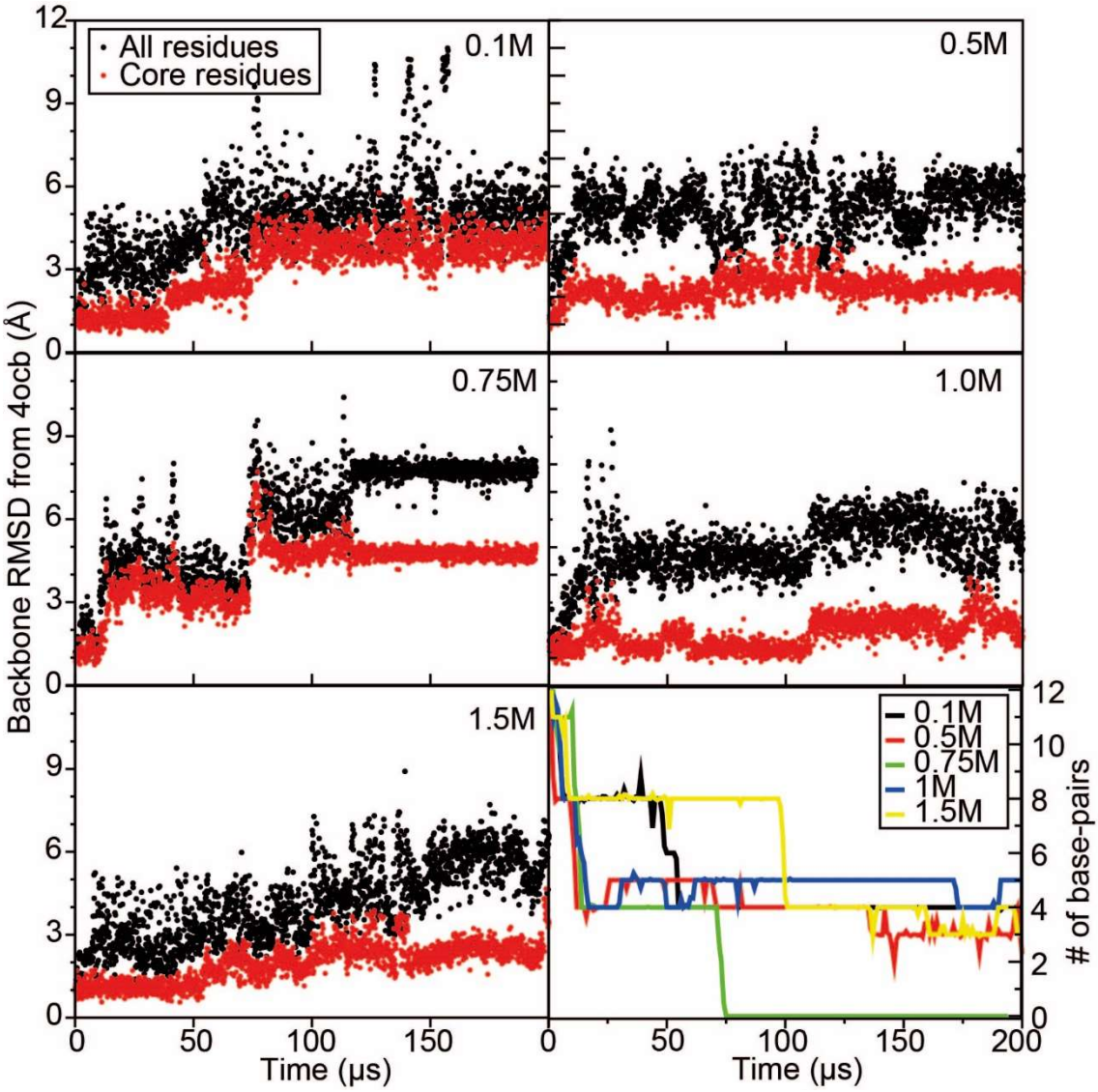

Figure S8. MD simulations of the d(CGCGCGCGCGCG) Z-DNA dodecamer at KCl concentrations ranging between 0.1 and 1.5M. For each concentration 200 μs of simulation were performed. Backbone RMSD traces with respect to the starting X-ray structure (PDB entry 4ocb) are reported for simulations performed using the DES-Amber force field. Simulations performed using DES-Amber SF1.0 gave similar results. For the calculation of the number of Z-DNA base-pairs, a dual cutoff methods was used where a base pair was considered to be in a Z-DNA conformation when the RMSD with respect to the starting structure, calculated on the

heavy atoms of the nucleotide pair, was < 1 Å, while the base pair was considered to be broken when the RMSD was > 3 Å.

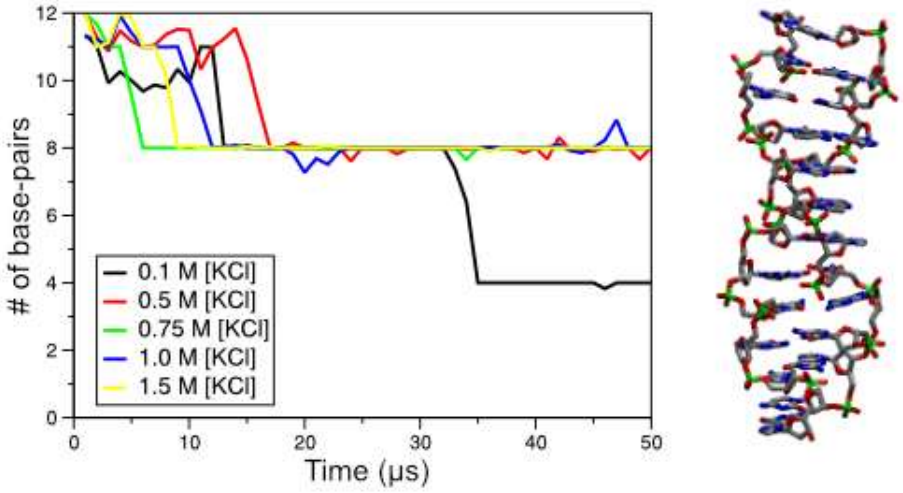

Figure S9. MD simulations of a d(CGCGCGCGCGCG) Z-DNA dodecamer performed with the DES-Amber 3.20 force field. Number of Z-DNA base pairs as a function of time during 50 μs MD simulations performed at KCl concentrations between 0.1 and 1.5 M.

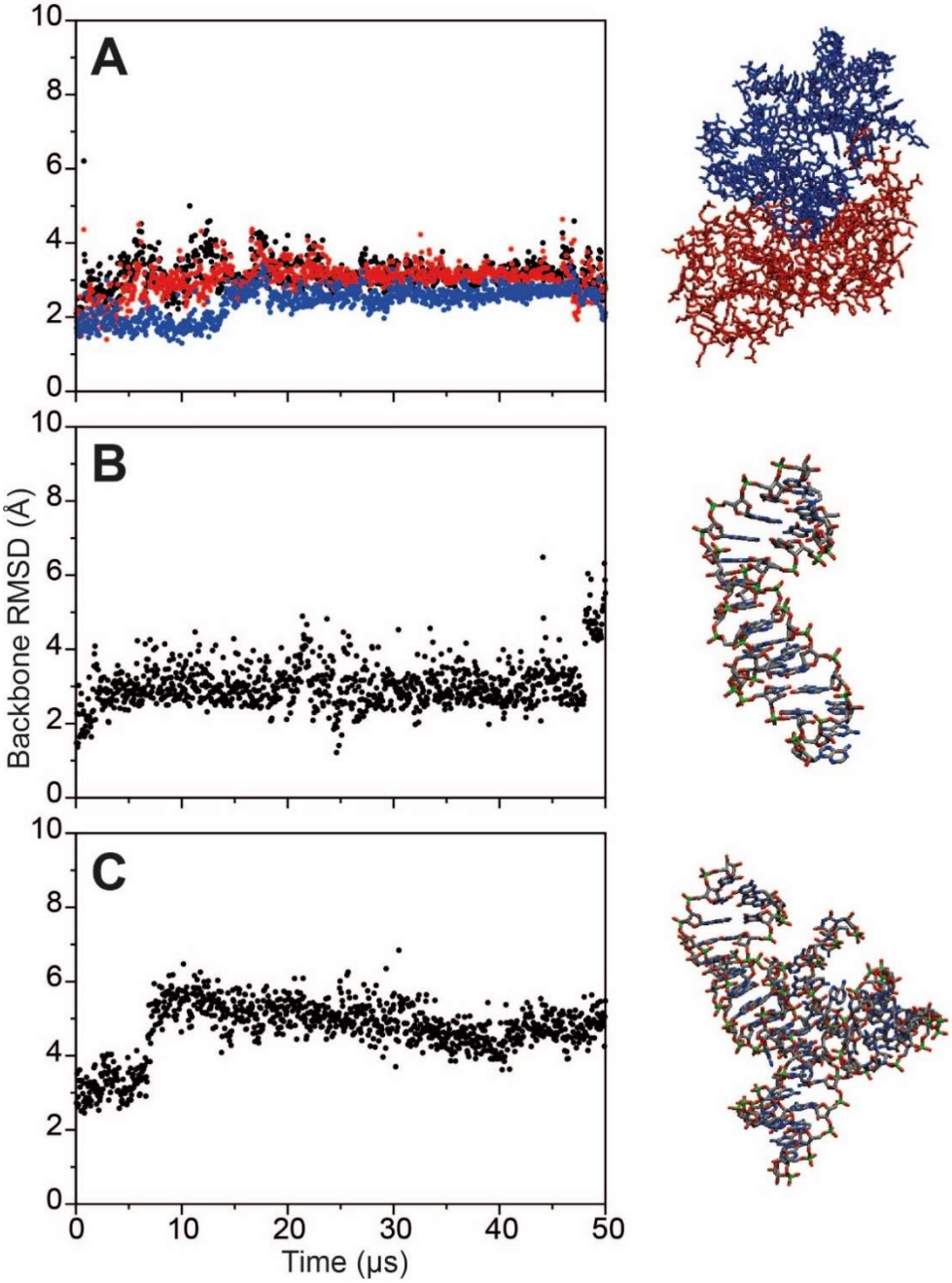

Figure S10. MD simulations of three complex RNA folds. Backbone RMSD deviations from the starting X-ray structure as function of time for three 50 μs simulations of three complex RNA folds performed with the DES-Amber 3.20 force field: (A) the ribosomal L1 stalk from Thermus

Thermophilus in complex with the ribosomal protein tthl1 (PDB entry 3u4m;[24] complex RMSD is in black, protein RMSD in red and RNA RMSD in blue), (B) the Sarcin/Ricin domain from ecoli 23S rRNA (PDB entry 3dw4[25]) and (C) an all-RNA "G8" hairpin ribozyme (PDB entry 2oue[26]).

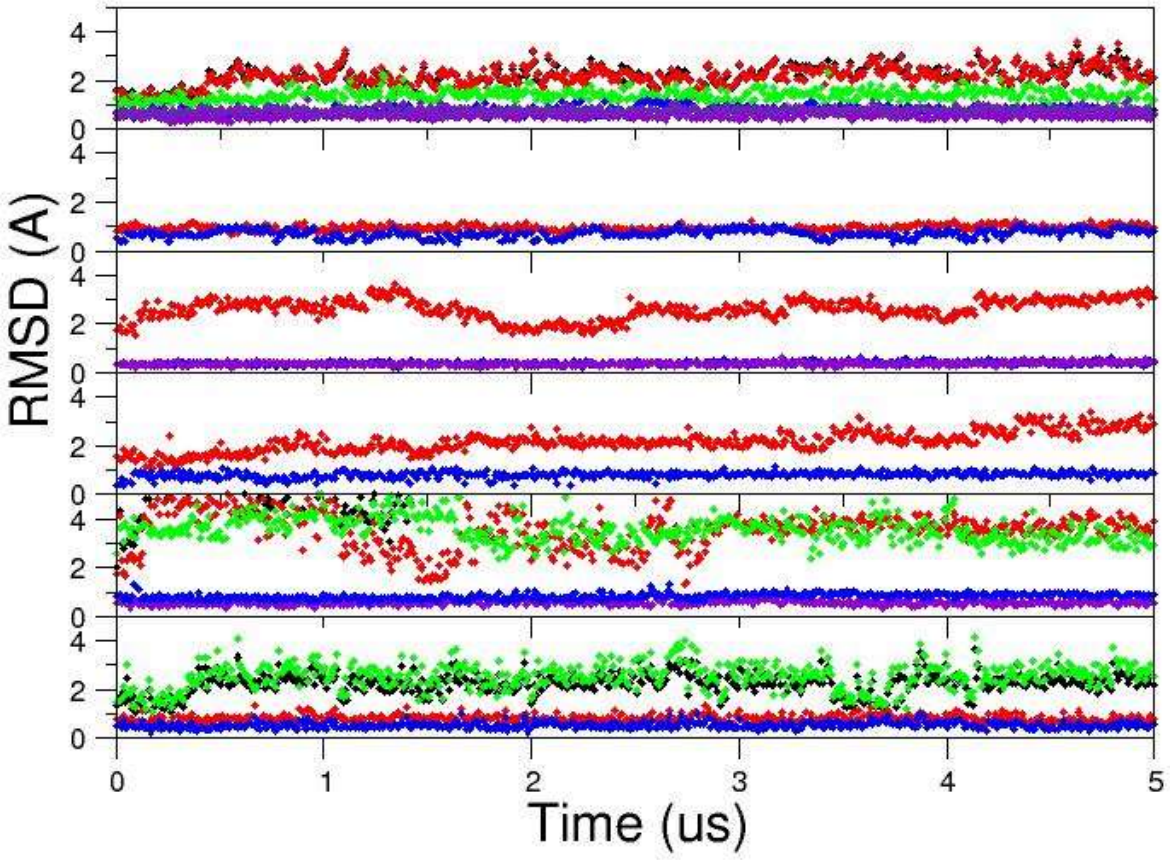

Figure S11. MD simulations of six proteins containing structural zinc ions. For each 5 μs simulation, the protein RMSD (red) and the RMSD of the heavy atoms of the residues coordinating zinc (blue, violet and indigo) are reported with respect to the experimentally determined structures (from top to bottom: 1AAY,[27] 1F57,[28] 1PZW,[29] 2AP1, 2EXF,[30] and 5W9S[31]). In cases where the protein structure has been determined in complex with DNA, the DNA RMSD (green) and the complex RMSD (black) are also reported. The coordination of zinc is CCCC for 1PZW and 5W9S, CCCH for 2AP1 and 2EXF, CHHE for 1F57, and CCHH for 1AAY.

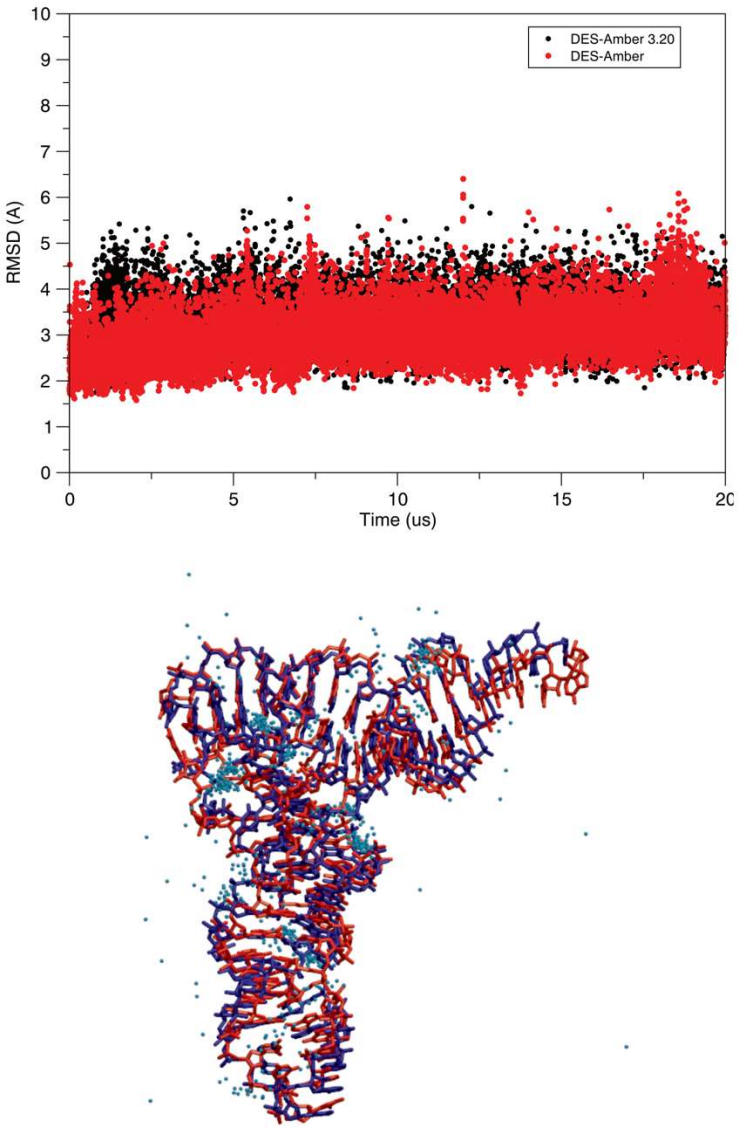

Figure S12. MD simulations of tRNA. The position of the Mg2+ ions was equilibrated by performing 36 μs of simulated tempering with position restraints where the tempered parameter (the Mg2+ LJ σ), allowed to vary between 2.61 and 3.045 Å. The last structure of the equilibration was used to perform 20 μs of unrestrained simulation using the DES-Amber and DES-Amber 3.20 force fields. The structure was found to be remarkably stable during the simulation, as indicated by the low backbone RMSD. A comparison of the initial structure

(blue) to the average structure of the last microsecond of simulation (red) is also shown. The position of the Mg^{2+} ions in 50 snapshots of the simulation is shown as cyan balls.

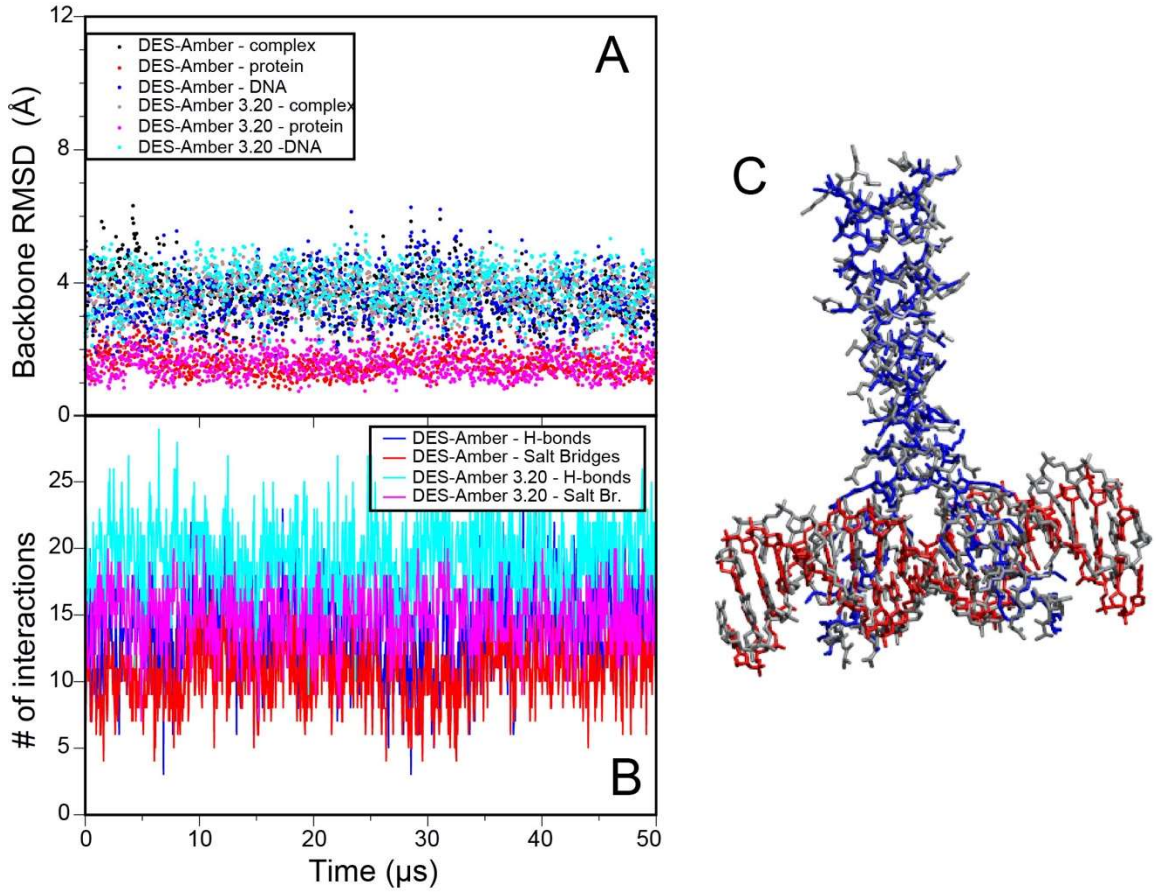

Figure S13. MD simulations of a leucine zipper/DNA complex. (A) Backbone RMSD from the X-ray structure (PDB ID 2DGC)[32] (B) The number of H-bonds and the number of salt-bridges between the protein and DNA is reported for simulations performed with DES-Amber (blue and red) and DES-Amber 3.20 (cyan and magenta). (C) Superimposition of the average structure of the last 2.5 μs of simulation (red -DNA, blue – protein) with the X-ray structure (gray)

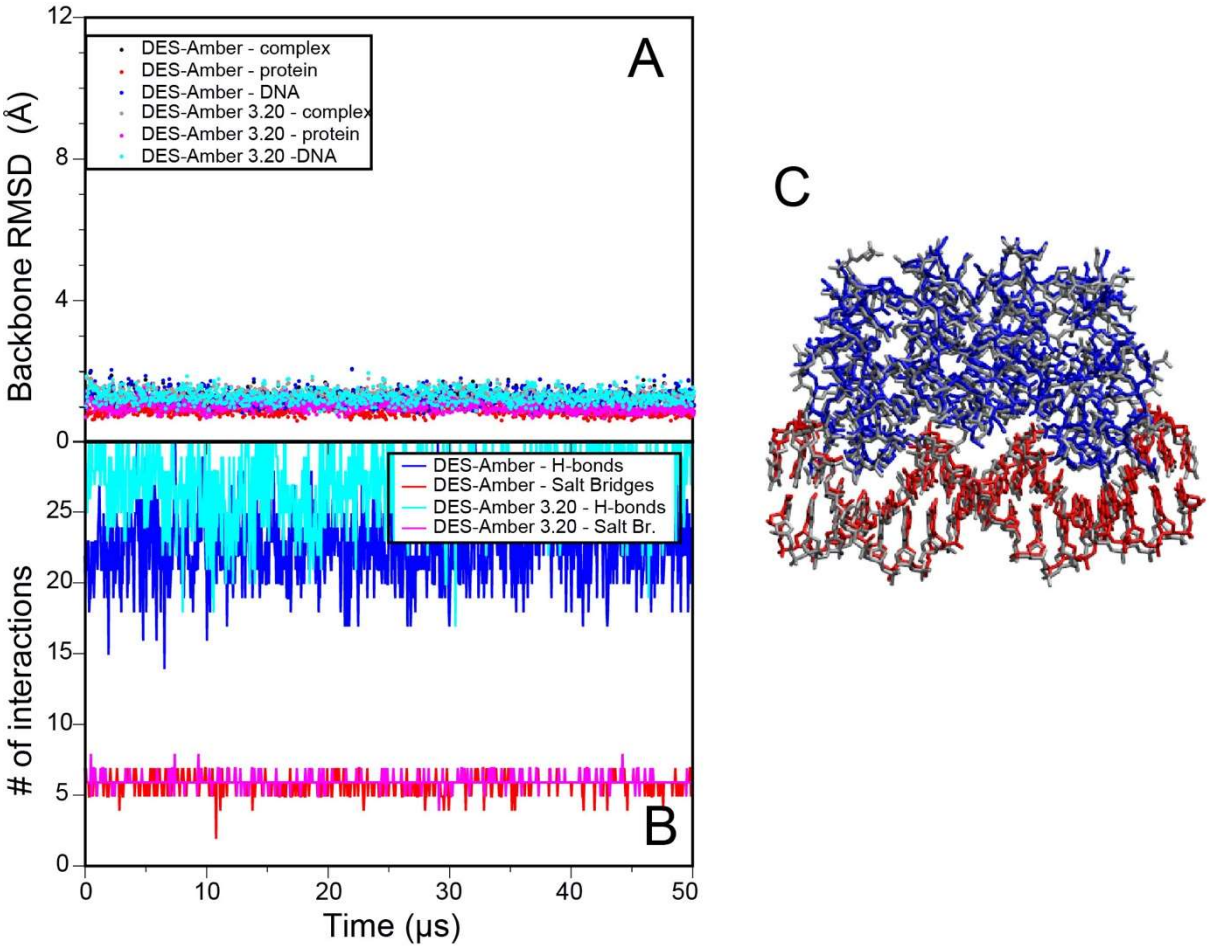

Figure S14. MD simulations of the Trp-repressor/operator complex. (A) Backbone RMSD from the X-ray structure (PDB ID 1TRO).[33] (B) The number of H-bonds and the number of salt-bridges between the protein and DNA is reported for simulations performed with DES-Amber (blue and red) and DES-Amber 3.20 (cyan and magenta). (C) Superimposition of the average structure of the last 2.5 μs of simulation (red -DNA, blue – protein) with the X-ray structure (gray)

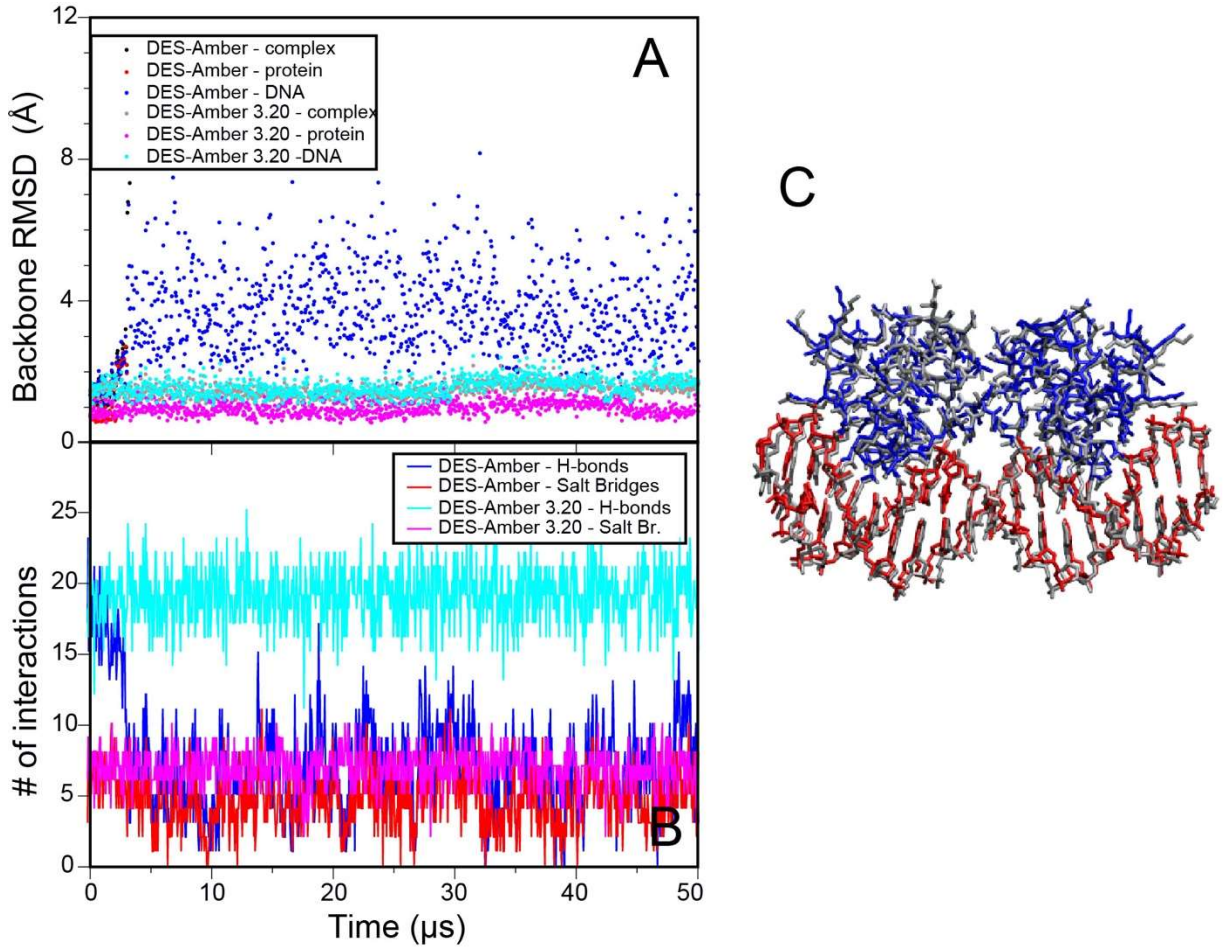

Figure S15. MD simulations of the P22 c2 repressor protein operator complex. (A) Backbone RMSD from the X-ray structure (PDB ID 3JXC).[34] (B) The number of H-bonds and the number of salt-bridges between the protein and DNA is reported for simulations performed with DES-Amber (blue and red) and DES-Amber 3.20 (cyan and magenta). (C) Superimposition of the average structure of the last 2.5 μs of simulation (red -DNA, blue – protein) with the X-ray structure (gray)

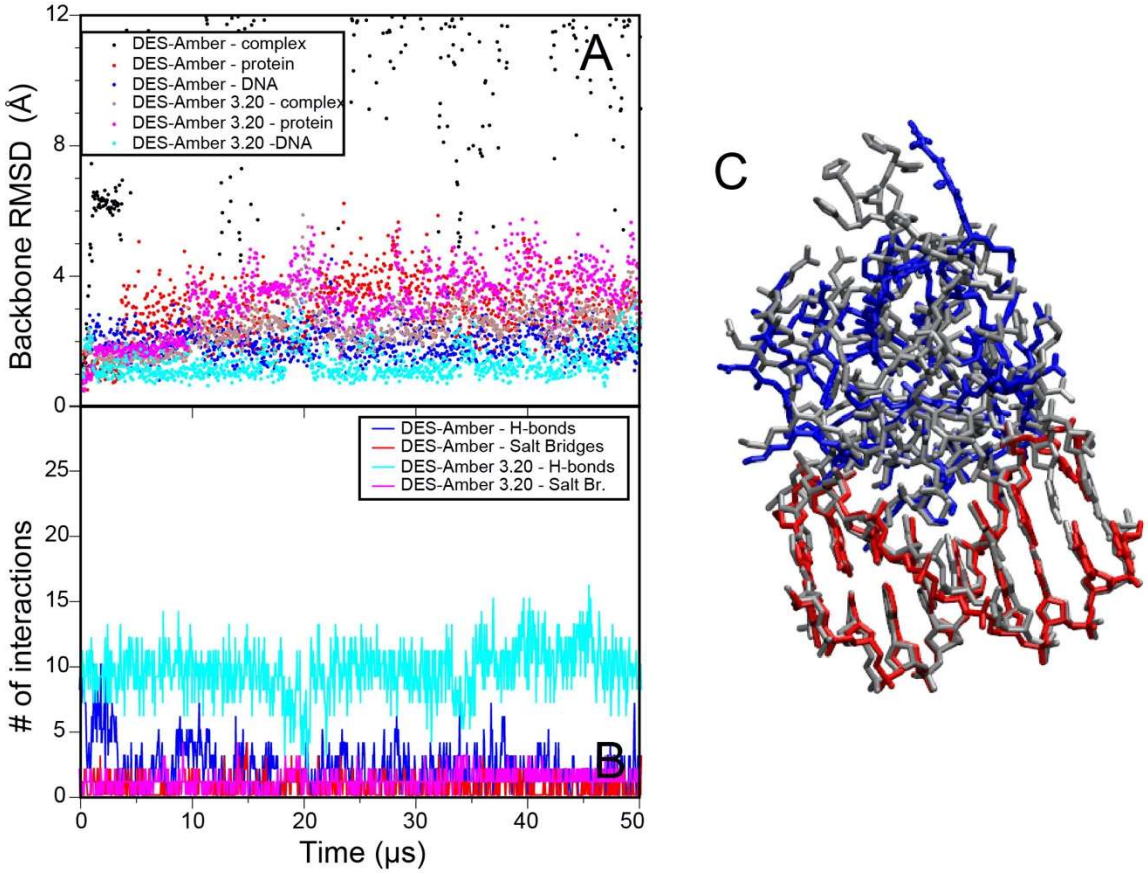

Figure S16. MD simulations of CI repressor in complex with the OL2 operator half-site. (A) Backbone RMSD from the X-ray structure (PDB ID 3ZHM).[35] The protein flexible tail residues were omitted from the RMSD calculation. (B) The number of H-bonds and the number of salt-bridges between the protein and DNA is reported for simulations performed with DES-Amber (blue and red) and DES-Amber 3.20 (cyan and magenta). (C) Superimposition of the average structure of the last 2.5 μs of simulation (red -DNA, blue – protein) with the X-ray structure (gray)

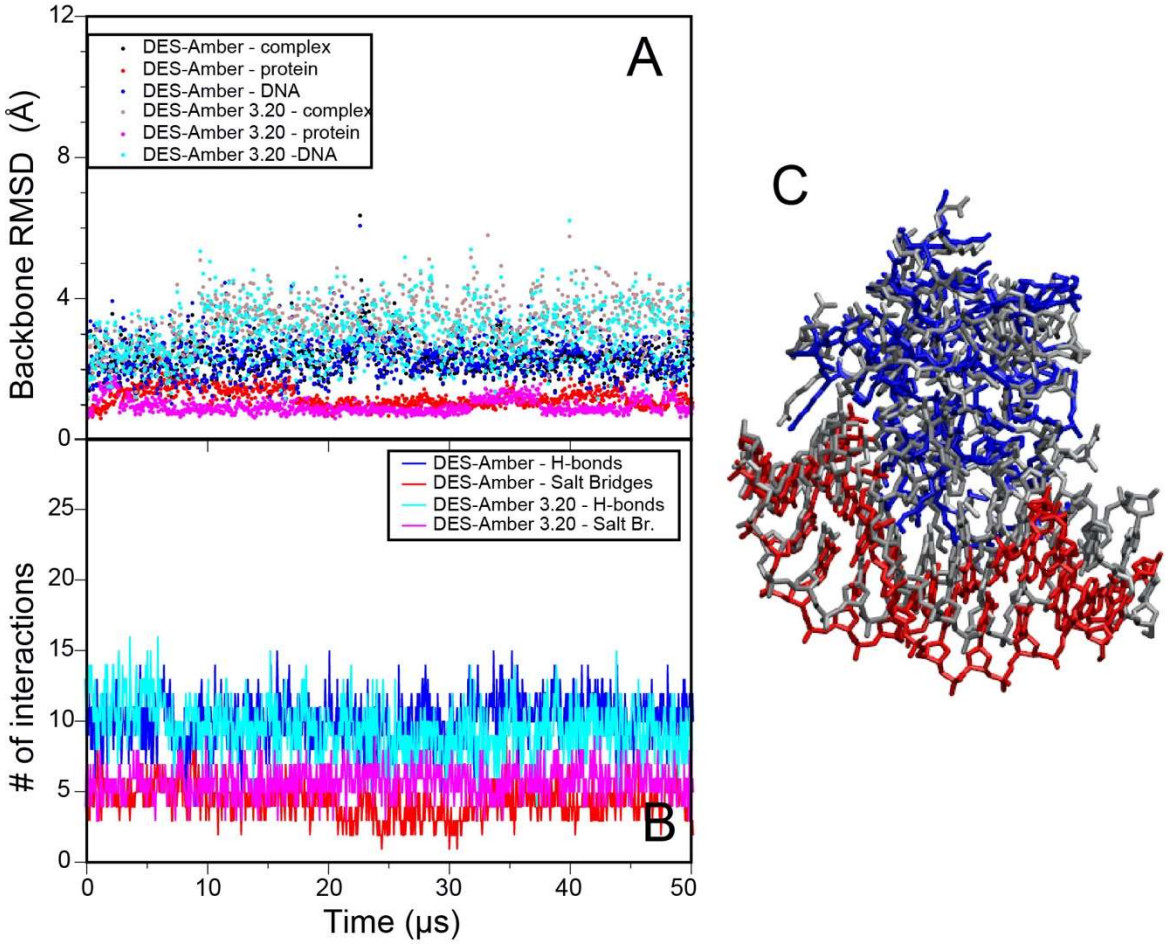

Figure S17. MD simulations of ETV6 bound to DNA. (A) Backbone RMSD from the X-ray structure (PDB ID 4MHG).[36] (B) The number of H-bonds and the number of salt-bridges between the protein and DNA is reported for simulations performed with DES-Amber (blue and red) and DES-Amber 3.20 (cyan and magenta). (C) Superimposition of the average structure of the last 2.5 μs of simulation (red -DNA, blue – protein) with the X-ray structure (gray)

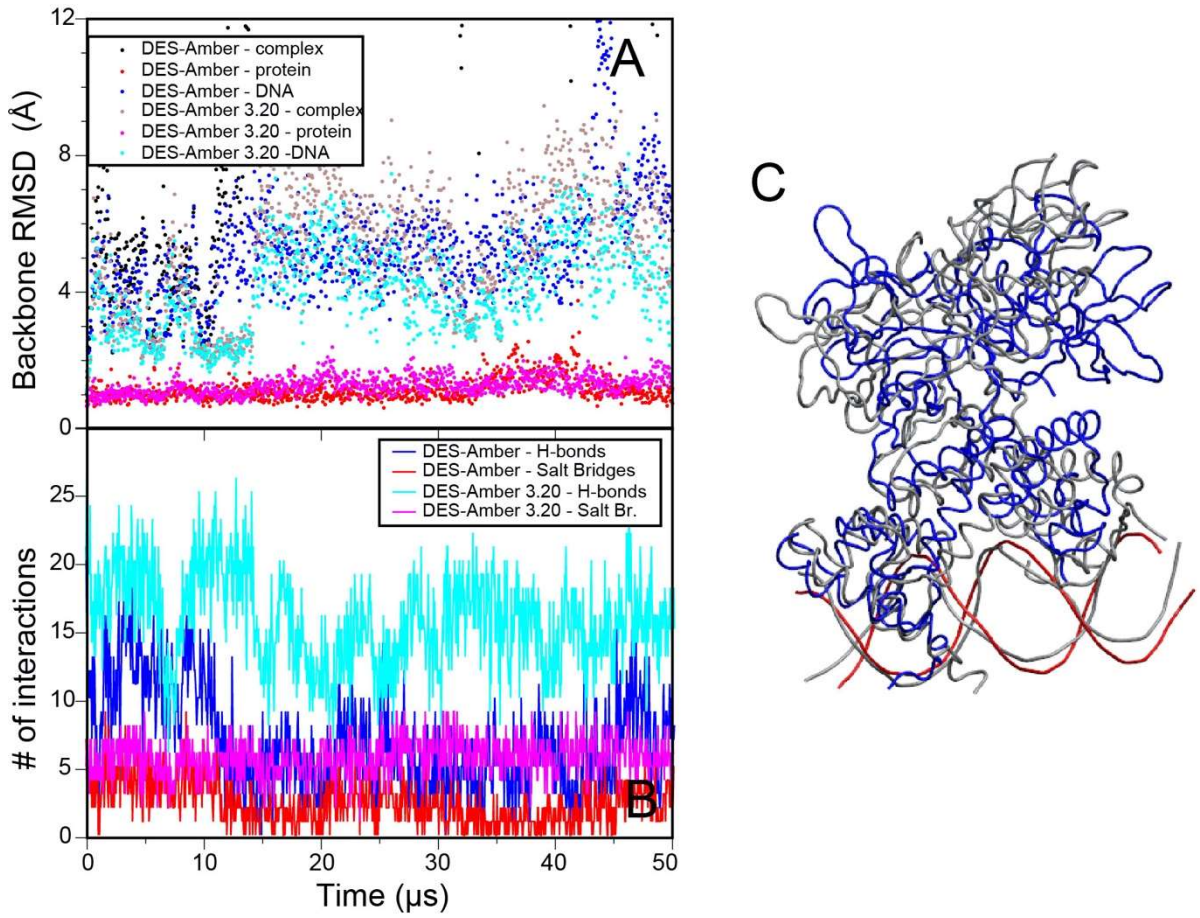

Figure S18. MD simulations of the λ-repressor. (A) Backbone RMSD from the X-ray structure (PDB ID 3BDN).[37] The large dimerization domain was omitted from the RMSD calculation as it is loosely attached to the DNA binding domain and displays substantial hinge motion. (B) The number of H-bonds and the number of salt-bridges between the protein and DNA is reported for simulations performed with DES-Amber (blue and red) and DES-Amber 3.20 (cyan and magenta). (C) Superimposition of the average structure of the last 2.5 μs of simulation (red -DNA, blue – protein) with the X-ray structure (gray)

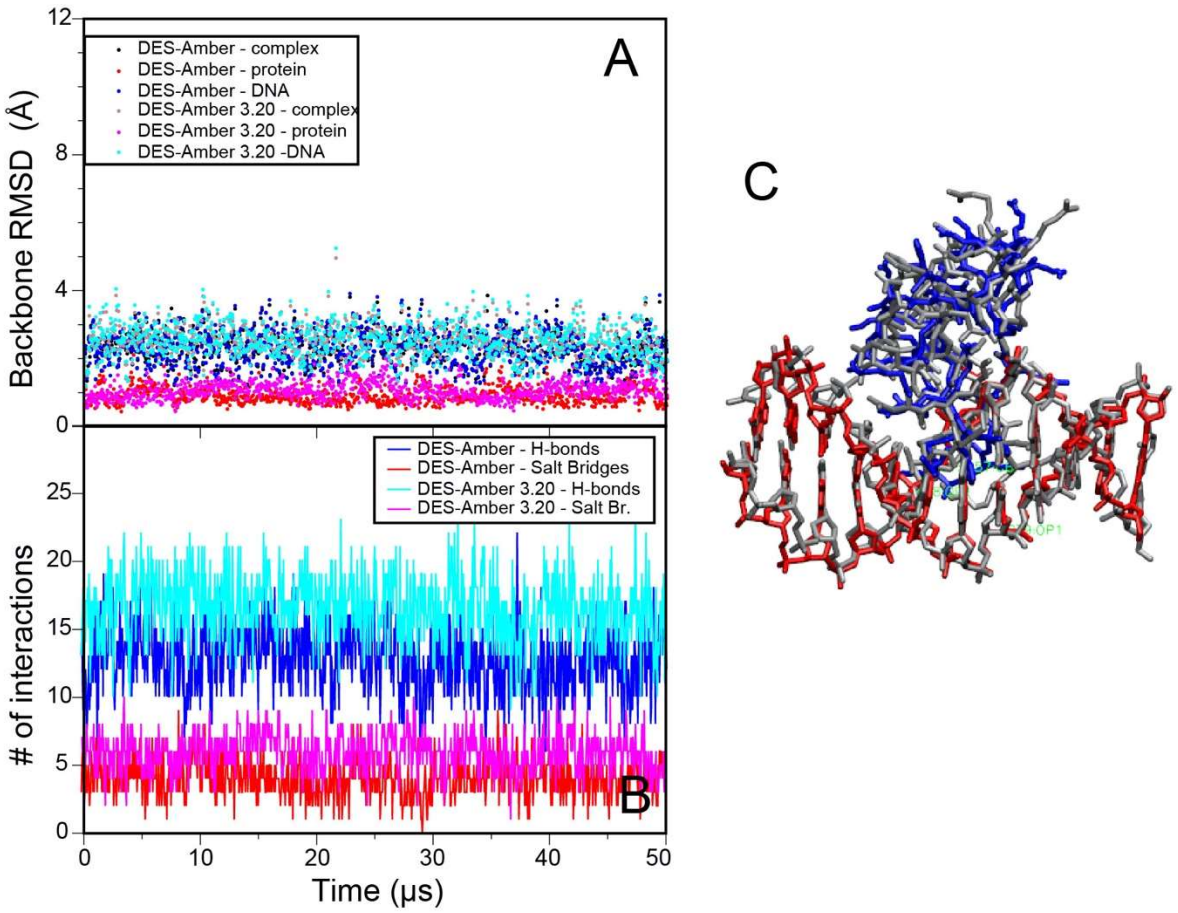

Figure S19. MD simulations of the zinc finger of human CXXC5 in complex with CpG DNA. (A) Backbone RMSD from the X-ray structure (PDB ID 5W9S).[38] The protein flexible tail residues were omitted from the RMSD calculation. (B) The number of H-bonds and the number of salt-bridges between the protein and DNA is reported for simulations performed with DES-Amber (blue and red) and DES-Amber 3.20 (cyan and magenta). (C) Superimposition of the average structure of the last 2.5 μs of simulation (red -DNA, blue – protein) with the X-ray structure (gray).

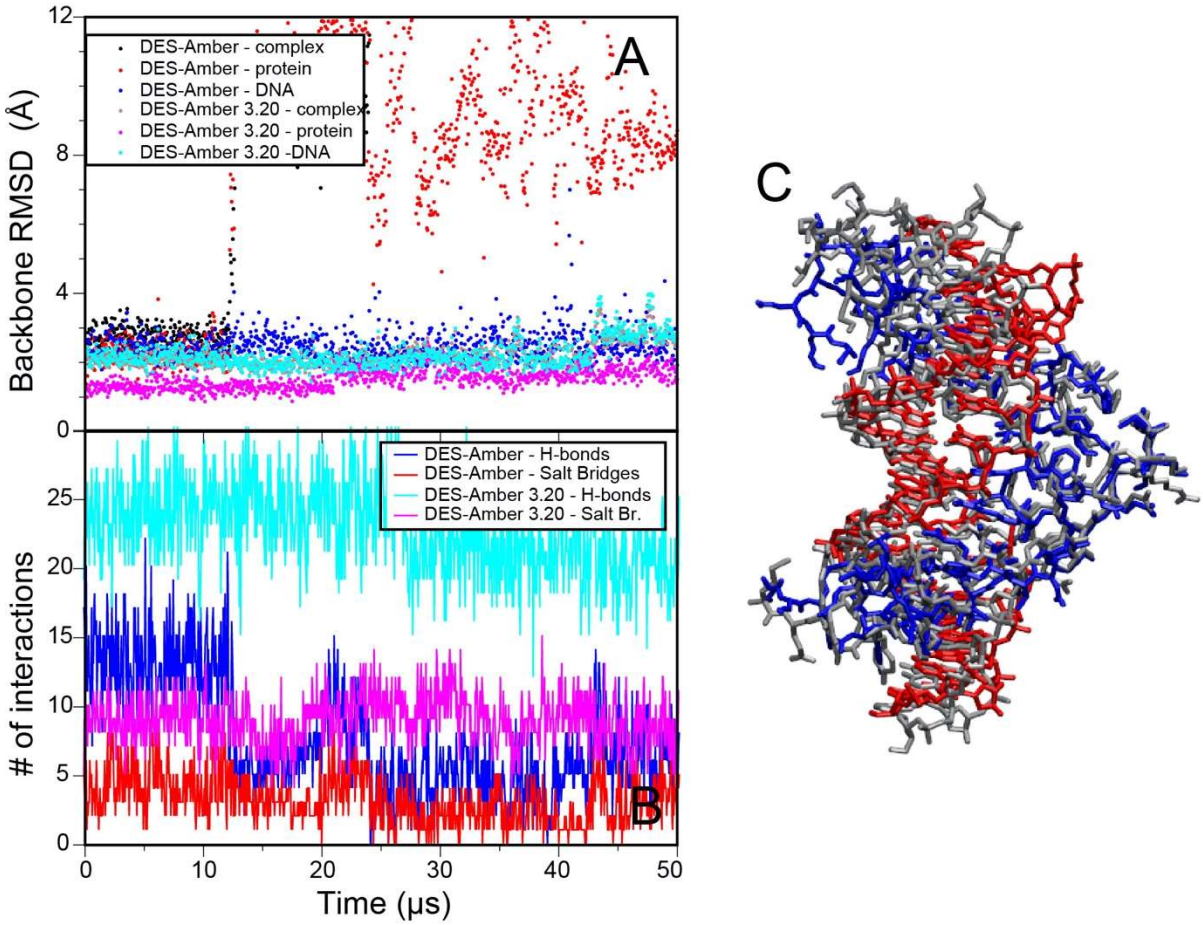

Figure S20. MD simulations of the ZIF268 zinc finger-DNA complex. (A) Backbone RMSD from the X-ray structure (PDB ID 1AAY).[39] The protein flexible tail residues were omitted from the RMSD calculation. (B) The number of H-bonds and the number of salt-bridges between the protein and DNA is reported for simulations performed with DES-Amber (blue and red) and DES-Amber 3.20 (cyan and magenta). (C) Superimposition of the average structure of the last 2.5 μs of simulation (red -DNA, blue – protein) with the X-ray structure (gray)

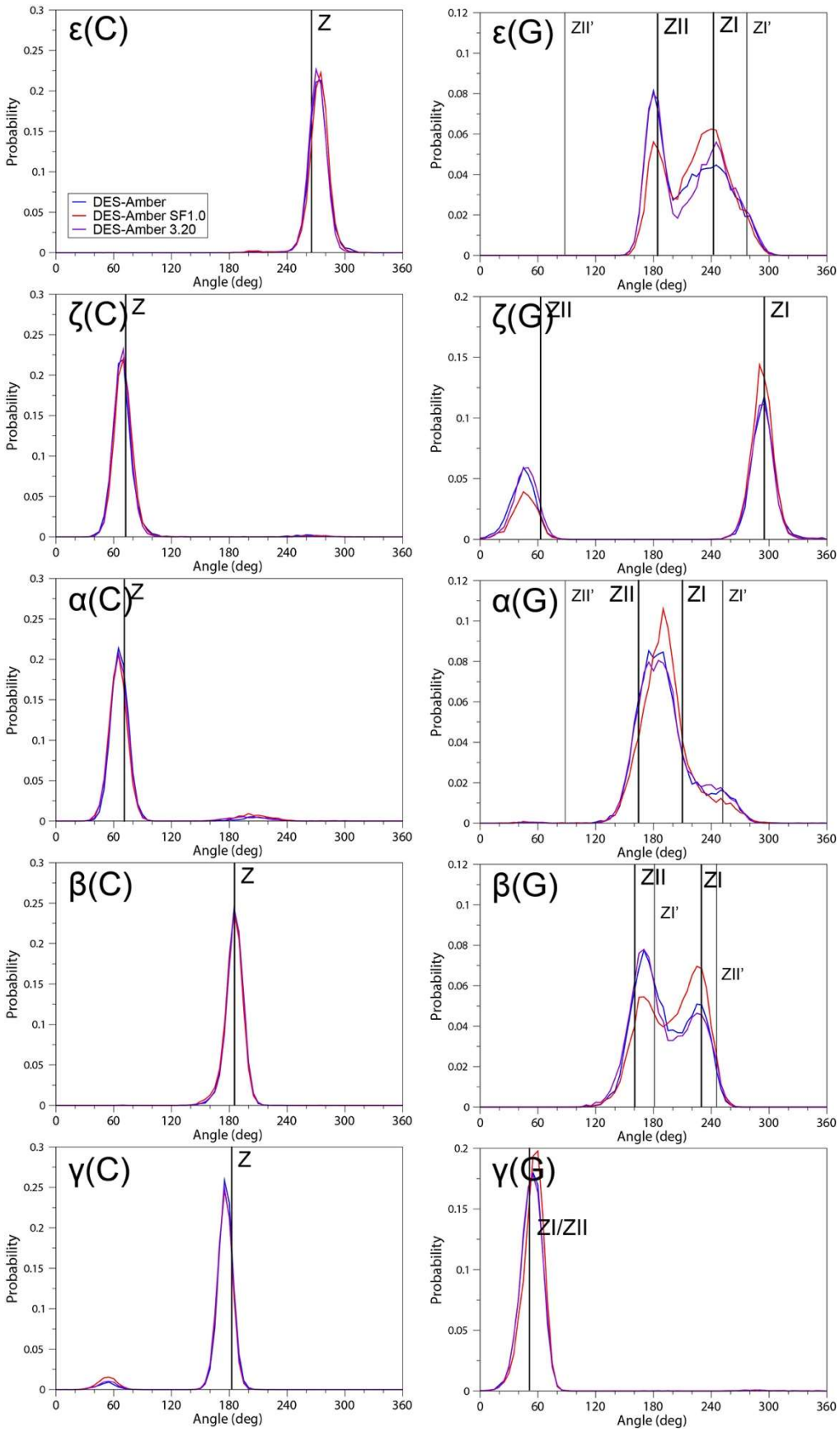
ε(C)
Z
DES-Amber
DES-Amber SF1.0
DES-Amber 3.20
ε(G)
ZII'
ZII
ZI
ZI'
ζ(C)
Z
ζ(G)
ZII
ZI
α(C)
Z
α(G)
ZII'
ZII
ZI
ZI'
β(C)
Z
β(G)
ZII
ZI
ZI'
ZII'
γ(C)
Z
γ(G)
ZI/ZII
Probability
Angle (deg)

Figure S21. MD simulations of Z-DNA in 1 M NaCl. (A) Probability distribution for the backbone torsion angles of cytosine (C) and guanosine (G). The position of the backbone conformers (ZI, ZI', ZII, ZII")[40] is indicated.

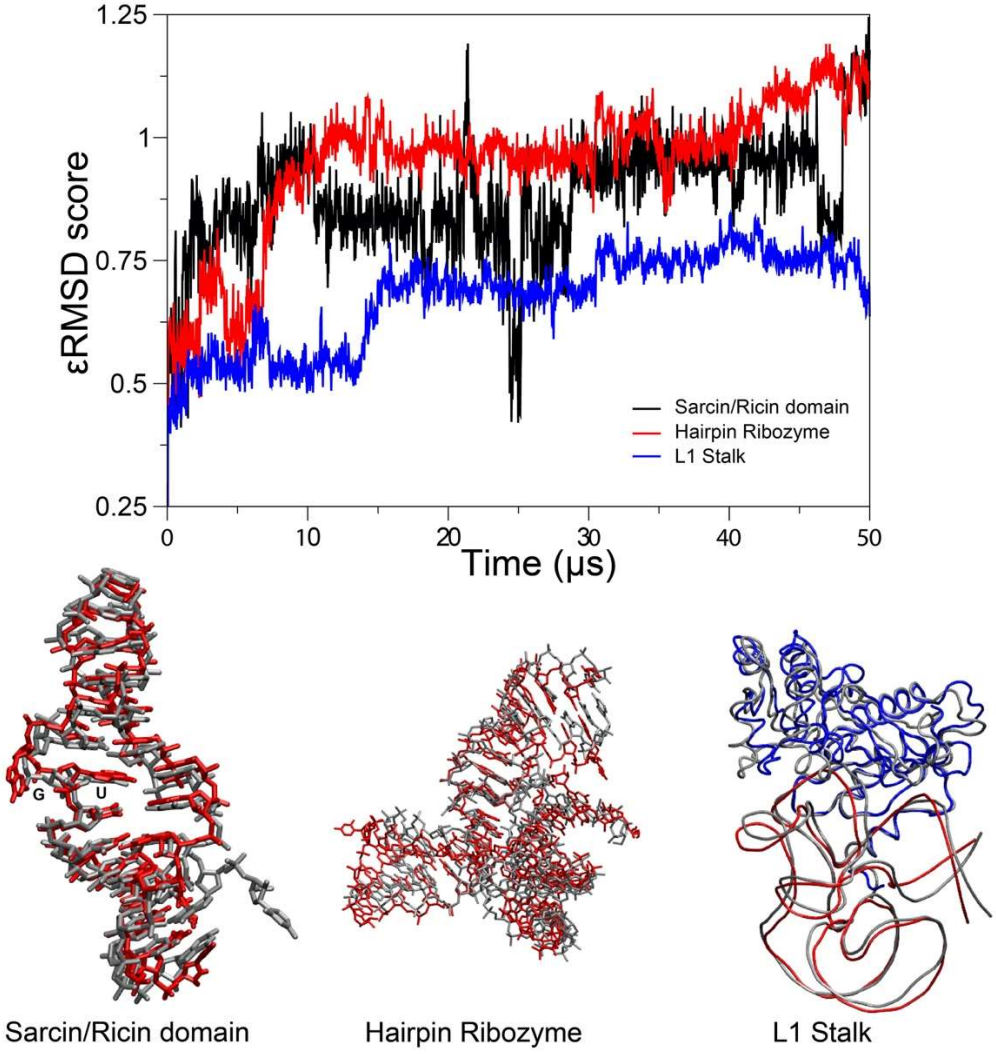

Figure S22. MD simulations of complex RNA folds. 50 μs of simulation of a Sarcin/Ricin domain, of a Hairpin ribozyme and of the L1 stalk were performed with the DES-Amber 3.20 force-field. The εRMSD score[41] was calculated with respect to the X-ray structure. In the bottom panel, the average structure of the last 2.5 μs of simulation is superimposed to the starting X-ray structures (PDB IDs 3DW4,[25] 2OUE[26] and 3U4M[24]).

Development of a generalized force field for the description of structural zinc ions compatible with DES-Amber

Structural zinc ions are a rather common feature of proteins that bind nucleic acids. To simulate these proteins and their interactions with DNA and RNA, we have created a set of parameters for structural zinc ions by adapting and applying to DES-Amber the EZAFF protocol developed by the group of Kenneth Merz[42] to cover the very common cases of tetracoordinated zinc ions. In the EZAFF force field, a preliminary QM calculation is performed to compute the partial charges on the system as well as the corresponding bond angles and bond lengths. Such an approach, while highly accurate, is computationally expensive and impractical in the context of the DES-Amber workflow and makes calculation harder to reproduce across different platforms.

In zinc complexes, a substantial fraction of the positive charge of zinc is usually delocalized on the coordination sphere rather than localized on the zinc ion and in the original EZAFF approach the preliminary QM calculation is used to accurately quantify this redistribution of charge. To approximately reproduce this characteristic feature without resorting to QM calculations, we defined special residue types for the amino acids coordinating the zinc ions and assigned 0.3 units of charge of the zinc ion on each of these residues. To preserve the total charge of the system, we assigned a charge of 0.6 to tetracoordinated zinc ion (note that in DES-Amber charges are scaled by 0.9, so the total charge on zinc is +1.8 rather than +2). We also created parameters for an exacoordinated zinc ion of charge 0, but these were not extensively tested. Bond lengths were set to the average values observed for each atom type in high resolution structures taken from the Protein Data Bank[43] and bond angles were set to 109.5 degrees for tetracoordinated zing and 90 degrees for exacoordinated zinc, corresponding to a tetrahedral and octahedral geometry, respectively. Force constants for bonds and angles were taken from EZAFF; force-constants for dihedral angles were set to 0, as in EZAFF. Lennard-Jones

parameters for zinc were also those of EZAFF. Files containing the full set of parameters are distributed with the Supporting Information.

We validated the force field by performing simulations of 5 μs each for six proteins containing a total of 11 tetra coordinated structural zinc ions in different coordination spheres (namely CCCC, CCCH, CCHH and CHHE). In all simulations the zinc binding site is stable and its geometry is well maintained throughout the simulation, as evidenced by the RMSD of the zinc and surrounding residues heavy atoms being < 1 Å for all simulations investigated (Figure S11). This result is comparable to the level of accuracy obtained for the EZAFF force field[42] and suggests that this approximate parametrization is adequate to perform stable simulations of proteins containing tetracoordinated structural zinc ions.

Parameterization strategy for non-standard residues

In some nucleic acid structures, nucleotides with a chemical structure slightly different the canonical one can be present due to changes in pH (protonation) or post-transcriptional modifications (methylation, oxidation, etc.). To develop parameters for these non-standard nucleotides, we started from the parameters for the most similar nucleotide in DES-Amber (the "parent" nucleotide) and followed the principle of introducing the smallest possible perturbation to these parameters. To this aim, we retain the bonded and non-bonded parameters for the atoms that are unchanged in the two nucleotides (the parent and the non-standard) and then assign GAFF bonded and non-bonded parameters[44,45] for the non-standard components and fit torsion parameters to torsion scans at the MP2 level. Charges for the non-standard components and the atoms bound to them are assigned according to the GAFF protocol with the constraints of preserving the total charge of the nucleobase.

Optimization of nucleic acid backbone non-bonded parameters against experimental data

The study described in the main text is mostly focused on the optimization of *nucleobase* non-bonded parameters and torsion potentials. Other parameters, however, may also influence the accuracy of nucleic acid simulations. *Backbone* non-bonded parameters are arguably the most important parameters not considered so far that may influence the outcome of simulations. Here we describe our attempts at tuning some of the backbone non-bonded parameters, namely the phosphate-water interactions and the 2' hydroxyl-group charge distribution in RNA. We found that tuning of these parameters resulted in a force field that improved the description of some of the systems investigated; in particular, notable improvements were observed in the description of protein-nucleic acid complexes.

It is worth pointing out that we are here addressing rather small quantitative deviations from experimental measurements in the context of an otherwise fairly accurate force field. Although somewhat minor, it is possible that some of the discrepancies observed in the simpler test systems (for which it is possible to get quantitative results) may be significant enough to limit our ability to describe some of the more complex nucleic acid systems. This type of fine-tuning of backbone non-bonded parameters would be difficult to perform in the context of force fields like Amber-bsc1, as the parameters of that force field provide a much less accurate description of many of the simpler systems considered in this study, like Na-DMP, K-DMP, Uracyl, Thymine, and rU40.

Reparametrization of the 2' alcohol group charges.

The 2' hydroxyl group in RNA plays an important role in determining the local conformational preference of the RNA backbone.[46] In a previous publication,[47] we have shown that the charge distribution used for the hydroxyl groups in the Amber force fields results in a poor description of the properties of water-alcohol mixtures (as assessed by KBIs) and the wrong order of stability for donor and acceptor hydrogen bonds to the hydroxyl group compared to QM calculations. These deficiencies can be remediated by rebalancing the charges of the alcohol group. As we did previously for the Ser and Thr side chains in the context of the DES-Amber protein force field, we therefore changed the charge distribution on the 2' atoms by decreasing the charge of O2' by 0.07 (from -0.6139 to -0.6839) and of HO2' by 0.03 (from 0.4186 to 0.3886) and increasing the charge of C2' by 0.1 (from 0.0670 to 0.1670). These changes allow us to quantitatively reproduce QM hydrogen bonding energies and KBIs of alcohol/water mixtures.[47]

Assessment of nucleic acid non-bonded parameters based on osmotic coefficients and aggregation propensities

The osmotic coefficient of Purine calculated using standard Amber and TIP3P water is in excellent agreement with experiment (RMS error 0.05) but it is slightly overestimated in DES-Amber and its variants (RMS error 0.14). We address this imbalance by optimization of the LJ interaction between the aromatic carbon atoms and water. We therefore changed the σ of this LJ interaction from 3.2325 Å to 3.2825 Å and ε from 0.12 kcal mol^{-1} to 0.1 kcal mol^{-1}. Using these parameters, the experimentally measured osmotic coefficients could be reproduced within experimental error (RMS error of 0.01). Note that this change only affects Adenine.

Osmotic coefficient calculations for sodium dimethyl-phosphate (Na-DMP) and potassium dimethyl-phosphate (K-DMP) solutions obtained from calculations performed with the DES-Amber force field are in very good agreement with the experimental measurements[22] (Figure S6) suggesting that phosphate LJ parameters developed by Steinbrecher and coworkers[48] and adopted in this force field are already close to optimal for these systems. Osmotic coefficient calculations for solutions of Uridine (rU) and Cytidine (rC) ribonucleotides result in osmotic coefficients that are slightly smaller than experiment[21] (Figure S6). This finding could be explained by assuming that the sugar ring or the Uridine and Cytidine nucleosides may be too hydrophobic in DES-Amber. Based on this observation, if the force field were fully transferable from small to larger molecules, we would have expected that a simulation of poly-U or poly-C would result in a structural ensemble that is too compact. Simulations of rU40 and dT40, two polynucleotides that are mostly disordered in solution, however, result in disordered ensembles that are slightly more *expanded* that the FRET data suggests (Figure S7), a result that is at odd with the expectations from the osmotic coefficient of the individual components. These results suggest that parameters obtained from small molecule calculations, while providing a good starting point, may not be fully transferable and some degree of additional fine-tuning may be required when applying these parameters to larger, more complex systems.

In an attempt to fine-tune the hydration properties of the phosphate backbone in the context of a large RNA or DNA macromolecule, we added to the Steinbrecher-Case parameters[48] that describe the phosphate group an additional set of LJ parameters to specifically modulate the interaction between the phosphate O2 oxygen atoms and the water oxygen atoms. Similar NBfix corrections have been recently used by several authors to fine-tune the properties of charged groups and hydrogen bonds in the context of non-polarizable force-fields.[49–55] In particular, we tested a number of increases in the sigma for the interaction (σ_{OW}) between the original value of 3.14 and 3.25 Å. This change introduces the smallest perturbation that would influence the

hydration properties of this group without changing the backbone torsion potentials or other LJ interactions, which would further complicate the interpretation of the results. This set of changes results in osmotic coefficients for Na-DMP and K-DMP that are still in good agreement with experiment (Figure S6). Simulations of rU40 and dT40 are only weakly affected by these changes, yielding disordered ensembles that are slightly more compact than DES-Amber and in good agreement with the experimental data (Figure S7). These results indicated that small changes in the LJ interaction of the phosphate group still result in force fields that exhibit reasonable global hydration properties and we decided to perform further testing of this force field to further quantify the effect of these parameter changes on the structure or stability of small ssRNA molecules.

Influence of phosphate non-bonded parameters on the structural properties of small ssRNA oligomers

We performed simulation of the five tetranucleotides described in the main text, namely AAAA, UUUU, CAAG, CCCC, and CGAA using values for σ_{OW} of 3.18, 3.20, 3.23, and 3.25 Å and measured the deviation from NMR scalar couplings and NOEs experimentally measured for these systems. We find that even small changes of σ_{OW} result in deviations from the experiment due to the formation of substantial amounts of intercalated structures (previously observed in simulations performed with a number of Amber force fields[17,56]) that are inconsistent with the experimental data, as shown by average error on the J-couplings that grows from 0.98 to > 1.3 Hz (Table S5). The amount of intercalated structures appears to increase with the increase of the value of σ_{OW} and the intercalated structures become dominant for values larger than 3.18 Å, a result consistent with previous observation of the importance of hydration for determining the structural properties of these tetranucleoties.[57]

It has been shown that reweighting with experimental data can be an effective approach to refining protein and nucleic acid force fields.[18,58,59,60] Here we used a curated dataset of J-couplings[18,60] to refine the RNA backbone torsion potentials by reweighting of the simulation ensemble.[60] Adding NOEs and not-observed NOEs (nNOEs) data to the fit only slightly affected the results (Table S5), as torsion potentials optimized only to J-couplings also reproduced well the NOE data. We find that, for each value of σ_{OW} between 3.18 and 3.25 Å, it is possible to obtain a force field that is in good agreement with the experimental data for the tetranucleotides (Average error on J-coupling < 0.9 Hz, Table S5); and, for some of these force fields, we verified that they can fold the UUCG and GGAC tetraloops to within 1 Å of the experimentally determined structure and also fold the CGCGG and CAGAG RNA duplexes to the correct A-form. In general, the resulting torsion correction appears to favor formation of the A-form of RNA with the largest changes observed in the α torsion term. In other words, there is a substantial degree of correlation in this dataset and care should be taken not to overfit to it. To help preventing overfitting, here we use an L2 regularization term restraining the coefficients of the fit to 0 with a force constant of 20 kcal mol^{-1} $Å^{-1}$. We find that the strength of the correction terms generally increases as the value of σ_{OW} increases (Table S6) and some torsion corrections can become as large as 1 kcal mol^{-1} for σ_{OW} = 3.25 Å.

This result indicates that the conformation of small ssRNA molecules is strongly influenced by both torsions and non-bonded potentials and so force fields that satisfactorily describe such systems can be obtained using a range of different parameters. We decided to pick the force field obtained using a value for σ_{OW} of 3.20 Å for further, more extensive testing on RNA and DNA systems as this force field provides a description of the solution properties of DMP that is still in good agreement with the experimental data while improving the description of rU40 and dT40 and requiring rather minor corrections to the torsions potentials to accurately describe the structural properties of the RNA tetranucleotides. To make this variant compatible with Mg^{2+}

we used osmotic coefficient calculations of $MgCl_2$,[61] $MgAce_2$,[62] and $MgDMP_2$ to calibrate the Mg^{2+}-OW LJ σ, changing it from 3.004 to 2.90 Å. We name this DES-Amber variant DES-Amber 3.20.

Assessment of the accuracy of DES-Amber 3.20 for DNA simulations

Refinement of backbone torsions based on structural properties of small DNA oligomers

As a first step, we initially performed a refinement of the backbone torsion parameters based on a total of 713 experimental NMR scalar couplings that probe the structural properties of 18 DNA oligomers ranging from 2 to 6 nucleotides (Table S1). In stark contrast to the RNA results, the change of the σ_{OW} for the water-phosphate LJ interaction does not affect the scalar couplings of the oligonucleotides. The RMS error from the experimental J-coupling is 0.88 Hz for DES-Amber and 0.87 Hz for DES-Amber 3.20. As the agreement with the experiment is already good in the starting force field, torsion refitting only reduces the error to 0.76 Hz; only the δ torsion corrections are larger than 0.1 kcal mol^{-1}, while corrections of less than 0.06 kcal mol^{-1} are applied to the other torsions (Table S7).

Thermodynamic stability of DNA double helix

We determined the thermodynamic stability of the DNA double helix for 7 sequences ranging from 4 to 7 nucleotides by performing ST simulations of 500 μs for each system (Figure S5) and computing the fraction of helix at each temperature. The helix stability at 310 K for each system

is reported in Table S2. The RMS error in the free energy of helix formation at 310 K is 2.0 kcal mol^{-1} for Amber-bsc1, 2.2 kcal mol^{-1} for DES-Amber and 1.1 kcal mol^{-1} for DES-Amber 3.20.

Thermodynamic stability of a DNA hairpin

We determined the thermodynamic stability of a DNA hairpin of sequence d(GCGAAGC) by running 295 μs of ST simulation and computing the fraction of helix at each temperature. The hairpin folds reversibly hundreds of times in simulation to within less than 1 Å RMSD from the experimentally-determined solution structure (PDB entry 1kr8).[63] The fraction of helix formed at 310 K is 13.5%, corresponding to a ΔG_{fold} at 310K of 1.1 kcal mol^{-1} for DES-Amber 3.20, which is comparable to the values of 1.3 and 1.5 kcal mol^{-1} obtained for DES-Amber and DES-Amber SF1.0, respectively.

Structural properties of DNA double helices

To verify that he force field is able to recapitulate the structure of double-stranded DNA, we performed simulations of 10 μs at 300 K for 17 DNA duplexes between 10 and 17 nucleotides long and compared the heavy-atom RMSD to the structures determined experimentally by X-ray or NMR. The heavy-atom RMSD deviations observed for DES-Amber 3.20 were not only comparable to those of DES-Amber but also to those of Amber-bsc1 (Table S4), a force field that is able to very accurately reproduce the structural features of double-stranded B-DNA.[64,65]

Structural properties of non-canonical DNA structures

We performed simulations for a number of non-canonical DNA structures to verify that the force field can correctly recapitulate the stability of non-Watson-Crick base pairs. In particular, we performed 20 μs simulations of two G-quadruplex structures[66,67] and a triplex structure[68] as well as 50 μs simulations of a Z-DNA helix[69] in salt concentrations ranging between 0.1 M and 1.5 M. The triplex and quadruplex structures were stable for the 20 μs of simulation (Figure S8). As observed for the DES-Amber force field, in simulations of the G-quadruplexes, the core quadruplex structure remained very close to the experimentally determined structure for the duration of the simulations, while the loops were more flexible. The Z-DNA structure is expected to be unstable in monovalent salt solution, although its stability is expected to increase at higher salt concentration. Within the first 10-20 μs, we observed fraying of the terminal base pairs in all salt conditions, but at the lowest salt concentration of 0.1 M [NaCl] the Z-DNA duplex melted almost completely, while core residues remained stable in the simulations run at higher salt concentration (Figure S9).

Assessment of the accuracy of DES-Amber 3.20 for RNA simulations.

We described previously the performance of this force field for rU40 and small ssRNA tetranucleotides. Not surprisingly, simulations of small ssRNA tetranucleotides are in excellent agreement with the NMR experimental data (Table S5) as the backbone torsion parameters of the force field were fitted to reproduce this data.

Thermodynamic stability of the RNA double-helix

We verified that this force field can describe the stability of A-RNA double-helical structures by performing ST simulations of three RNA A-helices of sequences CACAG, CGCGG, and UAAGGUA. The duplexes were simulated in 0 M (counterions only) and 1 M NaCl concentration. In each simulation, we observed between 5 and 21 helix formation events (defined as transitions between structures with greater than 10 Å RMSD and structures within 2 Å RMSD from the canonical A-helix structure), allowing us to estimate the stability of the duplexes with a statistical error of ~0.3-0.7 kcal mol^{-1} (Table S8). The RMS error in the free energies at 310 K with respect to the experimental data is 0.7 kcal mol^{-1}, indicating that the force field can reproduce the helix stability reasonably well, although the ionic strength dependence on stability might be weaker than what is observed in experiment.

Thermodynamic stability of RNA tetraloops

We verified that the force field is also able to recapitulate the structure of two RNA tetraloops, namely the ggcGCAAgcc and gccUUCGggc tetraloops. To this aim, we performed 500 μs of ST simulation for each of the tetraloops. In the two simulations, we observed multiple independent reversible folding events to structures that were within 1 Å RMSD from the experimentally determined structure. The calculated thermodynamic stability of the two tetraloops at 310 K is 0.0(3), -0.5(3) kcal mol^{-1} if only structures with the correct tetraloop configuration are considered as folded, or -0.8(5) and -1.1(4) kcal mol^{-1} if all hairpin structures are considered folded. These values can be compared to the experimentally determined stabilities[70] of -3.4 and -5.44 kcal mol^{-1} suggesting that, although the correct structure is recapitulated by the force field, the thermodynamic stability of the tetraloops is underestimated by 3–5 kcal mol^{-1}; this

underestimation is a problem that is commonly observed for the Amber RNA force fields[71,72] and may be related to the limited ability of point-charge force fields to properly describe hydrogen bonding.[71]

Structure and stability of complex RNA folds

Finally, we performed 50 μs simulations of three complex RNA folds, namely the ribosomial L1 stalk from Thermus Thermophilus in complex with the ribosomal protein tthl1 (PDB entry 3u4m[24]), the Sarcin/Ricin domain from Ecoli 23S rRNA (PDB entry 3dw4[25]), and an all-RNA "G8" hairpin ribozyme (PDB entry 2oue[26]). The structural Mg^{2+} and Na^{+} ions and the nucleobase modifications present in the structures were retained during the simulations as they were found to be important for structural stability.

We find that in all three cases, although some fluctuations are observed, overall the RNA fold is stable on the simulation timescale (Figure S10), suggesting that the force field provides a reasonable representation of the overall structure of non-canonical RNA structures. We also calculated the eRMSD score for each structure (Figure S22) which has values of ~1, indicating that some rearrangement does occur at the level of the base-pairs. In both the hairpin ribozyme and the Sarcin/Ricin domain simulations fraying occurs at the terminal residues; furthermore, in the ribozyme simulation the relative orientation of the two duplexes can fluctuate, further contributing to the overall RMSD from the X-ray structure. The Sarcin/Ricin domain features a peculiar S-loop where a bulged Guanine forms a characteristic GpU dinucleotide platform. In the 50 μs of simulation, the bulged Guanine residue is found to flip away from the GpU platform and becomes solvent exposed. This motion induces only minor perturbations in the remainder of the structure that closely maintains the X-ray conformation (Figure S22).

Supporting References

1. C. S. M. Olsthoorn, L. J. Bostelaar, J. H. Boom, C. Altona, Conformational Characteristics of the Trinucleoside Diphosphate dApdApdA and Its Constituents from Nuclear Magnetic Resonance and Circular Dichroism Studies. Extrapolation to the Stacked Conformers. *Eur J Biochem*. **112**, 95–110 (1980).

2. J.-R. Mellema, J. M. L. Pieters, G. A. Marel, J. H. Boom, C. A. G. Haasnoot, C. Altona, Sequence-dependent structural variation in single-helical DNA Proton NMR studies of d(T-A-T-A) and d(A-T-A-T) in aqueous solution. *Eur J Biochem*. **143**, 285–301 (1984).

3. Ts'o, P.O.P., Cheng, D.M., Frechet, D. and Kan, L. S. (1983) in Nucleic Acids: The Vectors of Life (B. Pullman and J. Jortner eds) pp. 201-206, D. Reidel Publ. Co., Dordrecht.

4. D. M. Cheng, L. S. Kan, E. E. Leutzinger, K. Jayaraman, P. S. Miller, P. O. P. Ts'o, Conformational Study Of Two Short Pentadeoxyribonucleotides, CpCpApApG and CpTpTpGpG, and their Fragments By Proton NMR. *Biochemistry*. **21**, 621–630 (1982).

5. J.-R. Mellema, C. A. G. Haasnoot, J. H. Van Boom, C. Altona, Complete assignment and conformational analysis of a deoxyribotetranucleotide, d(TAAT). A 360 and 500 MHz NMR study. *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis*. **655**, 256–264 (1981). https://doi.org/10.1016/0005-2787(81)90016-2

6. J.-R. Mellema, A. K. Jellema, C. A. G. Haasnoot, J. H. Boom, C. Altona, Conformational analysis of the single-helical DNA fragment d(T-A-A-T) in aqueous solution. The combined use of NMR proton chemical shifts and coupling constants obtained at 300 MHz and 500 MHz. *Eur J Biochem*. **141**, 165–175 (1984). https://doi.org/10.1111/j.1432-1033.1984.tb08171.x

7. J.-M. Neumann, T. Huynh-Dinh, S. K. Kan, B. Genissel, J. Igolen, S. Tran-Dinh, DNA Fragment Conformations: 1H-NMR and Relaxation Studies of 2′-Deoxyadenylyl(3′-

5′)thymidylyl- (3′-5′)deoxyguanosylyl(3′-5′)thymidine, d(A-T-G-T), in Neutra1 Aqueous Solution. *European Journal of Biochemistry*. **121**, 317–323 (1982). https://doi.org/10.1111/j.1432-1033.1982.tb05788.x

8. S. Tran-Dinh, J.-M. Neumann, T. Huynh-Dinh, B. Genissel, J. Igolen, G. Simonnot, DNA Fragment Conformations: 1H-NMR Studies of Helix-Coil Transition, Conformations and Dynamic Structures of the Self- Complementary Deoxyhexanucleotide d(A-C-A-T-G-T) in Aqueous. *European Journal of Biochemistry*. **124**, 415–425 (2005). https://doi.org/10.1111/j.1432-1033.1982.tb06609.x

9. J. H. J. Hartog, C. Altona, J. H. Boom, A. T. M. Marcelis, G. A. Marel, L. J. Rinkel, G. Wille-Hazeleger, J. Reedijk, Cis-Platinum Induced Distortions in DNA. Conformational Analysis of d(GpCpG) and cis-Pt(NH3)2{d(GpCpG)}, Studied by 500-MHz NMR. *Eur J Biochem*. **134**, 485–495 (1983). https://doi.org/10.1111/j.1432-1033.1983.tb07593.x

10. S. Tran-Dinh, J.-M. Neumann, J. Taboury, T. Huynh-Dinh, S. Renous, B. Genissel, J. Igolen, DNA Fragment Conformations. 1H-NMR Comparative Studies of Helix-Coil Transition and Conformation of d(C-A-C-G-T-G) and d(G-T-G-A-C). Influence of Helix Formation on Proton Chemical Shifts. *Eur J Biochem*. **133**, 579–589 (1983). https://doi.org/10.1111/j.1432-1033.1983.tb07502.x

11. C. A. G. Haasnoot, F. A. A. M. de Leeuw, H. P. M. de Leeuw, C. Altona, The relationship between proton-proton NMR coupling constants and substituent electronegativities. II—conformational analysis of the sugar ring in nucleosides and nucleotides in solution using a generalized Karplus equation. *Org. Magn. Reson.* **15**, 43–52 (1981). https://doi.org/10.1002/mrc.1270150111

12. C. A. G. Haasnoot, F. A. A. M. de Leeuw, C. Altona, The relationship between proton-proton NMR coupling constants and substituent electronegativities—I. *Tetrahedron*. **36**, 2783–2792 (1980). https://doi.org/10.1016/0040-4020(80)80155-4

13. A. Hospital, P. Andrio, C. Cugnasco, L. Codo, Y. Becerra, P. D. Dans, F. Battistini, J. Torres, R. Goñi, M. Orozco, J. Ll. Gelpí, BIGNASim: a NoSQL database structure and analysis portal for nucleic acids simulation data. *Nucleic Acids Res*. **44**, D272–D278 (2016). https://doi.org/10.1093/nar/gkv1301

14. P. P. Lankhorst, C. A. G. Haasnoot, C. Erkelens, C. Altona, Carbon-13 NMR in Conformational Analysis of Nucleic Acid Fragments 2. A Reparametrization of the Karplus Equation for Vicinal NMR Coupling Constants in CCOP and HCOP Fragments. *Journal of Biomolecular Structure and Dynamics*. **1**, 1387–1405 (1984). https://doi.org/10.1080/07391102.1984.10507527

15. M. M. W. Mooren, S. S. Wijmenga, G. A. van der Marel, J. H. van Boom, C. W. Hilbers, The solution structure of the circular trinucleotide cr(GpGpGp) determined by NMR and molecular mechanics calculation. *Nucl Acids Res*. **22**, 2658–2666 (1994). https://doi.org/10.1093/nar/22.13.2658

16. J. D. Tubbs, D. E. Condon, S. D. Kennedy, M. Hauser, P. C. Bevilacqua, D. H. Turner, The Nuclear Magnetic Resonance of CCCC RNA Reveals a Right-Handed Helix, and Revised Parameters for AMBER Force Field Torsions Improve Structural Predictions from Molecular Dynamics. *Biochemistry*. **52**, 996–1010 (2013). https://doi.org/10.1021/bi3010347

17. D. E. Condon, S. D. Kennedy, B. C. Mort, R. Kierzek, I. Yildirim, D. H. Turner, Stacking in RNA: NMR of Four Tetramers Benchmark Molecular Dynamics. *J. Chem. Theory Comput.* **11**, 2729–2742 (2015). https://doi.org/10.1021/ct501025q

18. S. Bottaro, G. Bussi, S. D. Kennedy, D. H. Turner, K. Lindorff-Larsen, Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. *Sci. Adv.* **4**, eaar8521 (2018). https://doi.org/10.1126/sciadv.aar8521

19. J. SantaLucia, A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proceedings of the National Academy of Sciences*. **95**, 1460–1465 (1998). https://doi.org/10.1073/pnas.95.4.1460

20. T. Xia, J. SantaLucia, M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, D. H. Turner, Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson−Crick Base Pairs †. *Biochemistry*. **37**, 14719–14735 (1998). https://doi.org/10.1021/bi9809425

21. P. O. P. Ts'o, I. S. Melvin, A. C. Olson, Interaction and Association of Bases and Nucleosides in Aqueous Solutions. *J. Am. Chem. Soc.* **85**, 1289–1296 (1963). https://doi.org/10.1021/ja00892a016

22. K. Tamaki, K. Suga, E. Tanihara, Solution Properties of Dialkyl Phosphate Salts. Apparent Molar Volumes, Viscosity *B* Coefficients, Heats of Solution, and Osmotic Coefficients. *BCSJ*. **60**, 1225–1229 (1987). https://doi.org/10.1246/bcsj.60.1225

23. H. Chen, S. P. Meisburger, S. A. Pabit, J. L. Sutton, W. W. Webb, L. Pollack, Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proceedings of the National Academy of Sciences*. **109**, 799–804 (2012). https://doi.org/10.1073/pnas.1119057109

24. S. Tishchenko, A. Gabdulkhakov, N. Nevskaya, A. Sarskikh, O. Kostareva, E. Nikonova, A. Sycheva, S. Moshkovskii, M. Garber, S. Nikonov, High-resolution crystal structure of the isolated ribosomal L1 stalk. *Acta Crystallogr D Biol Crystallogr*. **68**, 1051–1057 (2012). https://doi.org/10.1107/S0907444912020136

25. V. Olieric, U. Rieder, K. Lang, A. Serganov, C. Schulze-Briese, R. Micura, P. Dumas, E. Ennifar, A fast selenium derivatization strategy for crystallization and phasing of RNA structures. *RNA*. **15**, 707–715 (2009). https://doi.org/10.1261/rna.1499309

26. J. Salter, J. Krucinska, S. Alam, V. Grum-Tokars, J. E. Wedekind, Water in the Active Site of an All-RNA Hairpin Ribozyme and Effects of Gua8 Base Variants on the Geometry of Phosphoryl Transfer. *Biochemistry*. **45**, 686–700 (2006). https://doi.org/10.1021/bi051887k

27. M. Elrod-Erickson, M. A. Rould, L. Nekludova, C. O. Pabo, Zif268 protein–DNA complex refined at 1.6å: a model system for understanding zinc finger–DNA interactions. *Structure*. **4**, 1171–1180 (1996). https://doi.org/10.1016/S0969-2126(96)00125-6

28. D. M. F. van Aalten, C. R. Chong, L. Joshua-Tor, Crystal Structure of Carboxypeptidase A Complexed with d -Cysteine at 1.75 Å − Inhibitor-Induced Conformational Changes † , ‡. *Biochemistry*. **39**, 10082–10089 (2000). https://doi.org/10.1021/bi000952h

29. R. Jauch, G. P. Bourenkov, H.-R. Chung, H. Urlaub, U. Reidt, H. Jäckle, M. C. Wahl, The Zinc Finger-Associated Domain of the Drosophila Transcription Factor Grauzone Is a Novel Zinc-Coordinating Protein-Protein Interaction Module. *Structure*. **11**, 1393–1402 (2003). https://doi.org/10.1016/j.str.2003.09.015

30. S. Bourbigot, N. Ramalanjaona, C. Boudier, G. F. J. Salgado, B. P. Roques, Y. Mély, S. Bouaziz, N. Morellet, How the HIV-1 Nucleocapsid Protein Binds and Destabilises the (−)Primer Binding Site During Reverse Transcription. *Journal of Molecular Biology*. **383**, 1112–1128 (2008). https://doi.org/10.1016/j.jmb.2008.08.046

31. C. Xu, K. Liu, M. Lei, A. Yang, Y. Li, T. R. Hughes, J. Min, DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. *Structure*. **26**, 85-95.e3 (2018). https://doi.org/10.1016/j.str.2017.11.022

32. W. Keller, P. König, T. J. Richmond, Crystal Structure of a bZIP/DNA Complex at 2.2 Å: Determinants of DNA Specific Recognition. *Journal of Molecular Biology*. **254**, 657–667 (1995).

33. Z. Otwinowski, R. W. Schevitz, R.-G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, P. B. Sigler, Crystal structure of trp represser/operator complex at atomic resolution. *Nature*. **335**, 321–329 (1988).

34. D. Watkins, S. Mohan, G. B. Koudelka, L. D. Williams, Sequence Recognition of DNA by Protein-Induced Conformational Transitions. *Journal of Molecular Biology*. **396**, 1145–1164 (2010).

35. K. H. Frandsen, K. K. Rasmussen, M. R. Jensen, K. Hammer, M. Pedersen, J.-C. N. Poulsen, L. Arleth, L. Lo Leggio, Binding of the N-Terminal Domain of the Lactococcal Bacteriophage TP901-1 CI Repressor to Its Target DNA: A Crystallography, Small Angle Scattering, and Nuclear Magnetic Resonance Study. *Biochemistry*. **52**, 6892–6904 (2013).

36. S. De, A. C. K. Chan, H. J. Coyne, N. Bhachech, U. Hermsdorf, M. Okon, M. E. P. Murphy, B. J. Graves, L. P. McIntosh, Steric Mechanism of Auto-Inhibitory Regulation of Specific and Non-Specific DNA Binding by the ETS Transcriptional Repressor ETV6. *Journal of Molecular Biology*. **426**, 1390–1406 (2014).

37. S. Stayrook, P. Jaru-Ampornpan, J. Ni, A. Hochschild, M. Lewis, Crystal structure of the λ repressor and a model for pairwise cooperative operator binding. *Nature*. **452**, 1022–1025 (2008).

38. C. Xu, K. Liu, M. Lei, A. Yang, Y. Li, T. R. Hughes, J. Min, DNA Sequence Recognition of Human CXXC Domains and Their Structural Determinants. *Structure*. **26**, 85-95.e3 (2018).

39. M. Elrod-Erickson, M. A. Rould, L. Nekludova, C. O. Pabo, Zif268 protein–DNA complex refined at 1.6å: a model system for understanding zinc finger–DNA interactions. *Structure*. **4**, 1171–1180 (1996).

40. M. Zgarbová, J. Šponer, M. Otyepka, T. E. Cheatham, R. Galindo-Murillo, P. Jurečka, Refinement of the Sugar–Phosphate Backbone Torsion Beta for AMBER Force Fields Improves the Description of Z- and B-DNA. *J. Chem. Theory Comput.* **11**, 5723–5736 (2015).

41. S. Bottaro, F. Di Palma, G. Bussi, The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Research*. **42**, 13306–13314 (2014).

42. Z. Yu, P. Li, K. M. Merz, Extended Zinc AMBER Force Field (EZAFF). *J. Chem. Theory Comput.* **14**, 242–254 (2018). https://doi.org/10.1021/acs.jctc.7b00773

43. M. Laitaoja, J. Valjakka, J. Jänis, Zinc Coordination Spheres in Protein Structures. *Inorg. Chem.* **52**, 10983–10991 (2013). https://doi.org/10.1021/ic401072d

44. J. Wang, W. Wang, P. A. Kollman, D. A. Case, Automatic atom type and bond type perception in molecular mechanical calculations. *Journal of Molecular Graphics and Modelling*. **25**, 247–260 (2006).

45. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman, D. A. Case, Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).

46. L. Darré, I. Ivani, P. D. Dans, H. Gómez, A. Hospital, M. Orozco, Small Details Matter: The 2′-Hydroxyl as a Conformational Switch in RNA. *J. Am. Chem. Soc.* **138**, 16355–16363 (2016). https://doi.org/10.1021/jacs.6b09471

47. S. Piana, P. Robustelli, D. Tan, S. Chen, D. E. Shaw, Development of a Force Field for the Simulation of Single-Chain Proteins and Protein–Protein Complexes. *J. Chem. Theory Comput.* **16**, 2494–2507 (2020). https://doi.org/10.1021/acs.jctc.9b00251

48. T. Steinbrecher, J. Latzer, D. A. Case, Revised AMBER Parameters for Bioorganic Phosphates. *J. Chem. Theory Comput.* **8**, 4405–4412 (2012). https://doi.org/10.1021/ct300613v

49. J. Yoo, A. Aksimentiev, Improved Parameterization of Amine–Carboxylate and Amine–Phosphate Interactions for Molecular Dynamics Simulations Using the CHARMM and AMBER Force Fields. *J. Chem. Theory Comput.* **12**, 430–443 (2016).

50. J. Yoo, A. Aksimentiev, New tricks for old dogs: improving the accuracy of biomolecular force fields by pair-specific corrections to non-bonded interactions. *Phys. Chem. Chem. Phys.* **20**, 8432–8449 (2018).

51. J. Yoo, A. Aksimentiev, Improved Parametrization of Li^+, Na^+, K^+, and Mg^{2+} Ions for All-Atom Molecular Dynamics Simulations of Nucleic Acid Systems. *J. Phys. Chem. Lett.* **3**, 45–50 (2012).

52. V. Mlýnský, P. Kührová, T. Kühr, M. Otyepka, G. Bussi, P. Banáš, J. Šponer, Fine-Tuning of the AMBER RNA Force Field with a New Term Adjusting Interactions of Terminal Nucleotides. *J. Chem. Theory Comput.* **16**, 3936–3946 (2020).

53. P. Kührová, V. Mlýnský, M. Zgarbová, M. Krepl, G. Bussi, R. B. Best, M. Otyepka, J. Šponer, P. Banáš, Improving the Performance of the Amber RNA Force Field by Tuning the Hydrogen-Bonding Interactions. *J. Chem. Theory Comput.* **15**, 3288–3305 (2019).

54. R. M. Venable, Y. Luo, K. Gawrisch, B. Roux, R. W. Pastor, Simulations of Anionic Lipid Membranes: Development of Interaction-Specific Ion Parameters and Validation Using NMR Data. *J. Phys. Chem. B*. **117**, 10183–10192 (2013).

55. Y. Luo, B. Roux, Simulation of Osmotic Pressure in Concentrated Aqueous Salt Solutions. *J. Phys. Chem. Lett.* **1**, 183–189 (2010).

56. C. Bergonzo, N. M. Henriksen, D. R. Roe, T. E. Cheatham, Highly sampled tetranucleotide and tetraloop motifs enable evaluation of common RNA force fields. *RNA*. **21**, 1578–1590 (2015). https://doi.org/10.1261/rna.051102.115

57. C. Bergonzo, T. E. Cheatham, Improved Force Field Parameters Lead to a Better Description of RNA Structure. *J. Chem. Theory Comput.* **11**, 3969–3972 (2015). https://doi.org/10.1021/acs.jctc.5b00444

58. A. Cesari, A. Gil-Ley, G. Bussi, Combining Simulations and Solution Experiments as a Paradigm for RNA Force Field Refinement. *J. Chem. Theory Comput.* **12**, 6192–6200 (2016). https://doi.org/10.1021/acs.jctc.6b00944

59. R. B. Best, G. Hummer, Optimized Molecular Dynamics Force Fields Applied to the Helix−Coil Transition of Polypeptides. *J. Phys. Chem. B*. **113**, 9004–9015 (2009). https://doi.org/10.1021/jp901540t

60. A. Cesari, S. Bottaro, K. Lindorff-Larsen, P. Banáš, J. Šponer, G. Bussi, Fitting Corrections to an RNA Force Field Using Experimental Data. *J. Chem. Theory Comput.* **15**, 3425–3431 (2019). https://doi.org/10.1021/acs.jctc.9b00206

61. J. A. Rard, D. G. Miller, Isopiestic determination of the osmotic and activity coefficients of aqueous magnesium chloride solutions at 25 .degree.C. *J. Chem. Eng. Data*. **26**, 38–43 (1981). https://doi.org/10.1021/je00023a014

62. A. Apelblat, E. Korin, The vapour pressures of saturated aqueous solutions of magnesium, calcium, nickel and zinc acetates and molar enthalpies of solution of magnesium, calcium, zinc and lead acetates. *The Journal of Chemical Thermodynamics*. **33**, 113–120 (2001). https://doi.org/10.1006/jcht.2000.0731

63. P. Padrta, R. Štefl, L. Králík, L. Žídek, V. Sklenář, Refinement of d(GCGAAGC) hairpin structure using one- and two-bond residual dipolar couplings. *Journal of Biomolecular NMR*. **24**, 1–14 (2002). https://doi.org/10.1023/A:1020632900961

64. R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Šponer, M. Otyepka, P. Jurečka, T. E. Cheatham, Assessing the Current State of Amber Force Field Modifications for

DNA. *J. Chem. Theory Comput.* **12**, 4114–4127 (2016). https://doi.org/10.1021/acs.jctc.6b00186

65. P. D. Dans, I. Ivani, A. Hospital, G. Portella, C. González, M. Orozco, How accurate are accurate force-fields for B-DNA? *Nucleic Acids Res*, gkw1355 (2017). https://doi.org/10.1093/nar/gkw1355

66. S. Amrane, M. Adrian, B. Heddi, A. Serero, A. Nicolas, J.-L. Mergny, A. T. Phan, Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 Minisatellite. *J. Am. Chem. Soc.* **134**, 5807–5816 (2012). https://doi.org/10.1021/ja208993r

67. S. Haider, G. N. Parkinson, S. Neidle, Crystal Structure of the Potassium Form of an Oxytricha nova G-quadruplex. *Journal of Molecular Biology*. **320**, 189–200 (2002). https://doi.org/10.1016/S0022-2836(02)00428-X

68. J. L. Asensio, T. Brown, A. N. Lane, Solution conformation of a parallel DNA triple helix with 5′ and 3′ triplex–duplex junctions. *Structure*. **7**, 1–11 (1999). https://doi.org/10.1016/S0969-2126(99)80004-5

69. Z. Luo, M. Dauter, Z. Dauter, Phosphates in the Z-DNA dodecamer are flexible, but their P-SAD signal is sufficient for structure solution. *Acta Crystallogr D Biol Crystallogr*. **70**, 1790–1800 (2014). https://doi.org/10.1107/S1399004714004684

70. J. P. Sheehy, A. R. Davis, B. M. Znosko, Thermodynamic characterization of naturally occurring RNA tetraloops. *RNA*. **16**, 417–429 (2010). https://doi.org/10.1261/rna.1773110

71. P. Kührová, R. B. Best, S. Bottaro, G. Bussi, J. Šponer, M. Otyepka, P. Banáš, Computer Folding of RNA Tetraloops: Identification of Key Force Field Deficiencies. *J. Chem. Theory Comput.* **12**, 4534–4548 (2016). https://doi.org/10.1021/acs.jctc.6b00300

72. S. Bottaro, P. Banáš, J. Šponer, G. Bussi, Free Energy Landscape of GAGA and UUCG RNA Tetraloops. *J. Phys. Chem. Lett.* **7**, 4032–4038 (2016). https://doi.org/10.1021/acs.jpclett.6b01905