Supplementary material for: A systematically coarse-grained model for DNA, and its predictions for persistence length, stacking, twist, and chirality

Alex Morriss-Andrews, Joerg Rottler and Steven S. Plotkin
Department of Physics and Astronomy, University of British Columbia
6224 Agricultural Road, Vancouver, BC V6T1Z1, Canada.

S1 Supplementary material

S1.1 Tables

TABLE SI: RE^2 Parameters for base-base interactions. Units: $\sigma_{x,y,z}$ (Å), $\epsilon_{x,y,z}$ (none) and A_{12} (kcal/mol). The row indicates which base the radii $\sigma_{x,y,z}$ and inverse relative well depths $\epsilon_{x,y,z}$ belong to. The column indicates which base the row-base is interacting with. Note that all the properties of each base depend on what base it is interacting with.

		A	C	G	T
A	σ_x	2.4652	2.2860	2.4370	2.6909
	σ_y	2.1472	1.4051	1.5951	1.9398
	σ_z	0.5578	0.1262	0.5821	0.6384
	ϵ_x	3.6229	4.3800	3.6178	3.7656
	ϵ_y	5.0023l	2.6585	2.8420	2.1395
	ϵ_z	1.4057l	2.4634	1.5088	1.6393
	A_{12}	3455	17115	4188	6195
	σ_c	3.5969	3.6915	3.8501	3.7700
C	σ_x	2.4997	2.3127	2.1181	2.1897
	σ_y	2.0252	1.7950	1.6871	1.9022
	σ_z	0.7082	0.3903	0.4637	0.3069
	ϵ_x	2.9957	2.5848	1.7108	1.1875
	ϵ_y	7.8481	3.8092	3.6023	6.3381
	ϵ_z	1.0379	1.3424	1.163	1.2818
	A_{12}	17115	7826	3720	4839
	σ_c	3.6915	3.9060	3.961	3.8553
G	σ_x	3.1278	3.1633	3.3321	3.3842
	σ_y	2.5891	2.6258	1.9640	1.0974
	σ_z	0.4201	0.5201	0.4342	0.0290
	ϵ_x	3.5849	3.3655	4.3347	5.4026
	ϵ_y	4.4283	2.4695	2.3717	0.8356
	ϵ_z	0.6581	0.8089	0.7752	1.1117
	A_{12}	4188	3720	4274	51144
	σ_c	3.8501	3.961	4.0250	3.8664
T	σ_x	2.7194	3.1334	3.2489	2.9091
	σ_y	1.6628	2.0441	2.2016	2.2128
	σ_z	0.2994	0.5824	0.7588	0.5062
	ϵ_x	4.8654	4.5408	5.5485	3.6127
	ϵ_y	5.5447	3.3117	4.9745	3.4088
	ϵ_z	0.8189	0.9092	0.9895	0.9666
	A_{12}	6195	4839	51144	2437
	σ_c	3.7700	3.8553	3.8664	3.8096

TABLE SII: Equipotential radii (Å) and minimum energies (kT at 310K). The phosphate interaction is purely a screened Coulomb interaction and thus has its minimum at infinity.

		Adenine	Cytosine	Guanine	Thymine	Sugar	Phosphate
$1kT$	r_x	3.5002	3.4702	4.5140	3.9947	0.9418	2.7039
	r_y	3.1963	2.9713	3.1189	3.2961	0.9418	2.7039
	r_z	1.5479	1.5145	1.5391	1.5386	0.9418	2.7039
$10kT$	r_x	3.3836	3.3457	4.3526	3.8454	0.8315	0.3348
	r_y	3.0654	2.8266	2.9891	3.1497	0.8315	0.3348
	r_z	1.4768	1.4263	1.4707	1.4586	0.8315	0.3348
$V_{\min}$	$V_{\min,x}$	1.1790	1.1075	0.6649	0.6902	0.1540	0
	$V_{\min,y}$	0.9433	0.8102	1.0689	0.7212	0.1540	0
	$V_{\min,z}$	2.7053	2.0562	3.1969	2.3001	0.1540	0

TABLE SIII: Equipotential separation distances (Å) and minimum energies (kT at 310K) between Bases and S or P residues. Here distance rather than radius is used between unlike partners.

		A-P	C-P	G-P	T-P	A-S	C-S	G-S	T-S
$1kT$	d_x	6.1314	5.6869	7.7365	6.2227	4.5239	4.2680	5.7290	5.1273
	d_y	4.9390	4.9517	4.6075	5.5883	4.0709	3.5264	3.8037	4.1450
	d_z	3.4256	3.4359	3.3415	3.4049	1.8533	1.5695	1.6753	1.7636
$10kT$	d_x	5.7067	5.2886	7.3162	5.8186	4.3228	4.0728	5.5260	4.9296
	d_y	4.5398	4.5829	4.2459	5.1922	3.8663	3.3322	3.6158	3.9445
	d_z	3.1302	3.1330	3.0542	3.1168	1.6528	1.3725	1.4730	1.5676
$V_{\min}$	ϵ_x	0.2561	0.2644	0.2361	0.2419	0.0054	0.0039	0.0037	0.0047
	ϵ_y	0.3152	0.3417	0.3872	0.2600	0.0043	0.0030	0.0056	0.0049
	ϵ_z	0.7223	0.5997	0.7177	0.6692	0.0102	0.0060	0.0108	0.0109

TABLE SIV: RE^2 Parameters for base-phosphate sphere-asphere interactions. The $\sigma_{x,y,z}$ and $\epsilon_{x,y,z}$ parameters are properties of the base. Spherical particles under the RE^2 potential are considered to have $\epsilon_{x,y,z} = 1$, $\sigma_{x,y,z} = 0$, and nonzer σ_c. Units: $\sigma_{x,y,z}$ (Å), $\epsilon_{x,y,z}$ (none) and A_{12} (kcal/mol).

	A	C	G	T
σ_x	1.9844	1.7323	73.294	1.9983
σ_y	1.1571	1.2331	1.1070	1.5560
σ_z	0.1533	0.2070	0.2470	0.0966
ϵ_x	13.443	12.917	8.7902	15.0740
ϵ_y	9.0257	8.0570	4.1412	12.8730
ϵ_z	2.6714	3.5227	1.7311	2.9361
A_{12}	729.67	534.84	244.45	934.7900
σ_c	4.2489	4.1348	4.1409	4.2940

TABLE SV: RE^2 Parameters for base-sugar sphere-asphere interactions. The $\sigma_{x,y,z}$ and $\epsilon_{x,y,z}$ parameters are properties of the base. Spherical particles under the RE^2 potential are considered to have $\epsilon_{x,y,z} = 1$, $\sigma_{x,y,z} = 0$, and nonzero σ_c. Units: σ (Å), ϵ (none) and A_{12} (kcal/mol).

	A	C	G	T
σ_x	2.4652	2.3127	3.3321	2.9091
σ_y	2.1472	1.7950	1.9640	2.2128
σ_z	0.5578	0.3903	0.4342	0.5062
ϵ_x	3.6229	2.5848	4.3347	3.6127
ϵ_y	5.0023	3.8092	2.3717	3.4088
ϵ_z	1.4057	1.3424	0.7752	0.9666
$\tilde{A}_{12}$	0.25	0.25	0.25	0.25
σ_c	2	2	2	2

TABLE SVI: Translational and rotational friction coefficients of the bases found from atomistic simulations and the radii of ellipsoids which best fit the friction coefficients according to the continuum hydrodynamic theory (Eqs. (14a-14b)). These are referred to as the hydrodynamic radii of the bases. The hydrodynamic radii are plotted against the equipotential energetic radii in Fig. A17.

	ζ^{trans} $(10^{-11} kg/s)$				ζ^{rot} $(10^{-30} kg\ m^2/s)$			
	A	C	G	T	A	C	G	T
ζ_x	0.3961	0.2975	0.4730	0.3344	0.1467	0.0779	0.1616	0.1038
ζ_y	0.3972	0.3771	0.4974	0.3994	0.2183	0.1371	0.2910	0.1840
ζ_z	0.4835	0.5163	0.5524	0.4988	0.4046	0.1988	0.4091	0.2607

	$\sigma^{(H)}$ (Å)			
	A	C	G	T
σ_x	3.0809	2.6255	3.3773	2.889
σ_y	2.1573	1.5675	1.9315	1.698
σ_z	1.1348	0.9648	1.2677	1.071

TABLE SVII: Bond potential parameters: spring constants and equilibrium separations. The first four bonds are only listed for reference and were not used in the simulations. Because of the magnitude of the effective spring constant, the base-sugar system was treated as a rigid body.

Bond	k_r $(\mathrm{kcal/mol \cdot Å^2})$	r_0 (Å)
r_{CS}	330.8 ± 1.2	3.057 ± 0.00074
r_{TS}	323.4 ± 0.84	3.169 ± 0.00018
r_{AS}	229.2 ± 2.9	3.501 ± 0.00065
r_{GS}	186.5 ± 0.74	3.572 ± 0.0014
$r_{\mathrm{SP_{3'}}}$	40.45 ± 2.3	4.581 ± 0.0037
$r_{\mathrm{P_{5'}S}}$	5.596 ± 0.75	5.194 ± 0.032

TABLE SVIII: Bond angle potential parameters. Four triples are well fit to a harmonic potential, however six showed a clear secondary potential well (see Fig. S2B). For these, their statistics were fit to the sum of two Gaussian distributions, giving a potential as in Eq. (5).

	Harmonic	
	k_θ (kcal/mol)	θ_0 (rad)
θ_{PSP}	19.51 ± 0.69	1.406 ± 0.02
$\theta_{\mathrm{P_{5'}SG}}$	11.29 ± 1.7	1.926 ± 0.0089
$\theta_{\mathrm{P_{5'}SA}}$	7.852 ± 0.78	1.829 ± 0.051
θ_{SPS}	3.453 ± 0.39	1.922 ± 0.034

	Bi-harmonic				
	A	k_1(kcal/mol)	k_2(kcal/mol)	θ_1	θ_2
$\theta_{\mathrm{ASP_{3'}}}$	0.3229 ± 0.087	24.19 ± 5.8	34.57 ± 21	2.83 ± 0.00016	2.274 ± 0.026
$\theta_{\mathrm{CSP_{3'}}}$	0.2169 ± 0.0065	20.69 ± 3.2	29.75 ± 2.4	2.734 ± 0.02	2.015 ± 0.0036
$\theta_{\mathrm{TSP_{3'}}}$	0.168 ± 0.015	28.12 ± 3.2	52.8 ± 3.1	2.713 ± 0.025	1.982 ± 0.0055
$\theta_{\mathrm{GSP_{3'}}}$	0.3702 ± 0.096	32.12 ± 9.2	16.34 ± 9.2	2.81 ± 0.017	2.467 ± 0.11
$\theta_{\mathrm{P_{5'}SC}}$	0.7811 ± 0.018	32.57 ± 3	29.55 ± 0.97	1.652 ± 0.0084	2.01 ± 0.0069
$\theta_{\mathrm{P_{5'}ST}}$	0.5165 ± 0.0041	65.58 ± 2.4	22.18 ± 3.4	1.498 ± 0.0076	1.887 ± 0.0077

TABLE SIX: Dihedral potential parameters. The functional form of this potential is 3 terms of a 2π periodic series, shown in Eq. (6). Representative plots of the potentials are seen in Fig. S3. The form of potentials such as $P_{5'}SP_{3'}S_{3'}$ that did not involve the base did not show much variation base to base. The ± numbers are the variation base to base for these numbers. The ± numbers for the potentials involving the base represent half the 95% confidence interval of the fit of the statistics to the desired potential.

	K_1(kcal/mol)	K_2(kcal/mol)	K_3(kcal/mol)	ϕ_1	ϕ_2	ϕ_3
$\phi_{\mathrm{ASP_{3'}S_{3'}}}$	-0.09 ± 0.01	0.03 ± 0.01	-0.05 ± 0.01	-1.41 ± 0.09	0.8 ± 0.2	-4.0 ± 0.2
$\phi_{\mathrm{CSP_{3'}S_{3'}}}$	-0.10 ± 0.01	0.10 ± 0.01	-0.06 ± 0.01	-1.26 ± 0.07	-0.02 ± 0.07	-8.3 ± 0.1
$\phi_{\mathrm{GSP_{3'}S_{3'}}}$	-0.10 ± 0.01	-0.04 ± 0.01	-0.08 ± 0.01	-0.79 ± 0.08	-1.5 ± 0.2	-3.9 ± 0.1
$\phi_{\mathrm{TSP_{3'}S_{3'}}}$	-0.11 ± 0.01	-0.17 ± 0.01	0.07 ± 0.01	-0.70 ± 0.08	1.85 ± 0.04	-6.7 ± 0.3
$\phi_{\mathrm{S_{5'}P_{5'}SA}}$	-0.13 ± 0.01	-0.26 ± 0.01	-0.05 ± 0.01	-1.62 ± 0.06	-1.39 ± 0.03	-0.7 ± 0.2
$\phi_{\mathrm{S_{5'}P_{5'}SC}}$	-0.07 ± 0.02	-0.22 ± 0.02	-0.05 ± 0.02	-0.9 ± 0.1	-0.72 ± 0.04	0.4 ± 0.2
$\phi_{\mathrm{S_{5'}P_{5'}SG}}$	-0.12 ± 0.02	-0.23 ± 0.02	-0.02 ± 0.02	-14.71 ± 0.08	-1.24 ± 0.04	-0.2 ± 0.3
$\phi_{\mathrm{S_{5'}P_{5'}ST}}$	-0.18 ± 0.03	-0.22 ± 0.03	-0.14 ± 0.03	0.2 ± 0.1	-0.48 ± 0.09	-0.1 ± 0.1
$\phi_{\mathrm{P_{5'}SP_{3'}S_{3'}}}$	0.10 ± 0.02	0.15 ± 0.02	0.21 ± 0.02	0.0 ± 0.1	2.44 ± 0.02	2.17 ± 0.05
$\phi_{\mathrm{S_{5'}P_{5'}SP_{3'}}}$	-0.08 ± 0.01	-0.20 ± 0.01	0.044 ± 0.01	8.0 ± 0.1	-1.50 ± 0.05	-0.9 ± 0.2

TABLE SX: Improper potential parameters. The improper angle is calculated as the angle between the $P_{3'}SX$ and $P_{5'}SX$ plane, and the potential is fit to the same functional form as the dihedral potentials: Eq. (6).

	Class 2					
	K_1(kcal/mol)	K_2(kcal/mol)	K_3(kcal/mol)	ϕ_1	ϕ_2	ϕ_3
$\phi_{\mathrm{P_{5'}SAP_{3'}}}$	-1.4 ± 0.2	-0.3 ± 0.2	-0.3 ± 0.1	-0.77 ± 0.04	-0.9 ± 0.2	-0.6 ± 0.2
$\phi_{\mathrm{P_{5'}SGP_{3'}}}$	-2.0 ± 0.3	-0.7 ± 0.2	-0.4 ± 0.1	-0.51 ± 0.02	-1.04 ± 0.09	-1.0 ± 0.1
$\phi_{\mathrm{P_{5'}SCP_{3'}}}$	2.36 ± 0.06	0	0	1.97 ± 0.02	0	0
$\phi_{\mathrm{P_{5'}STP_{3'}}}$	2.66 ± 0.09	0	0	1.62 ± 0.03	0	0

S1.2 Parameterization of potentials

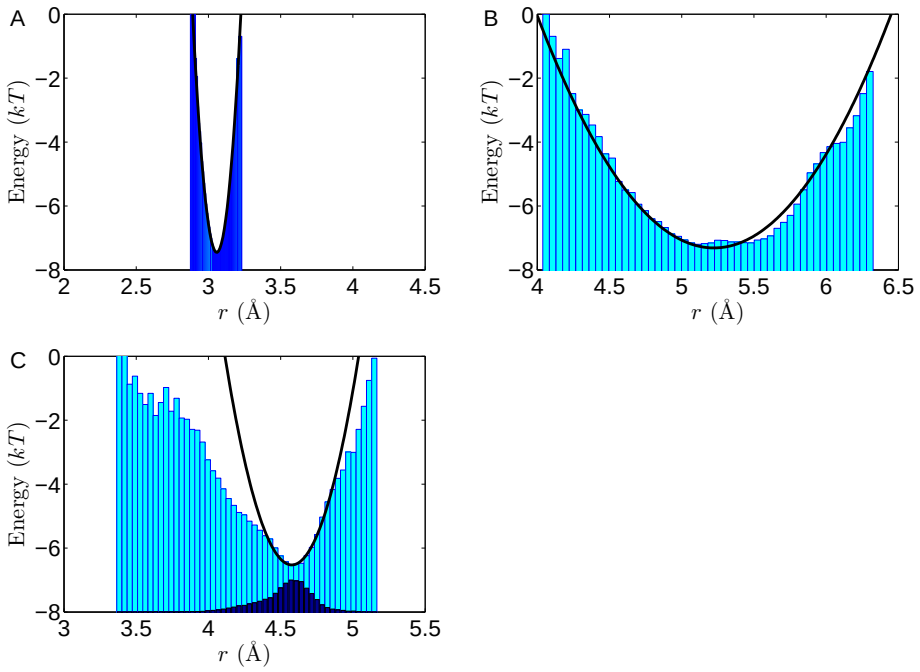

FIG. S1: Three effective bond potentials in Cytosine (a) r_{S-CM} (b) $r_{P_{5'}S}$ (c) The bond between the 3′ phosphate residue and the sugar, $r_{SP_{3'}}$, shows an anharmonic shoulder that may be accounted for by adding another Gaussian (see discussion of angle potentials where this is done), however the amount of time the bond spends at these distances is small: to see this the scaled probability distribution is shown. The harmonic approximation is thus reasonably good for this bond.

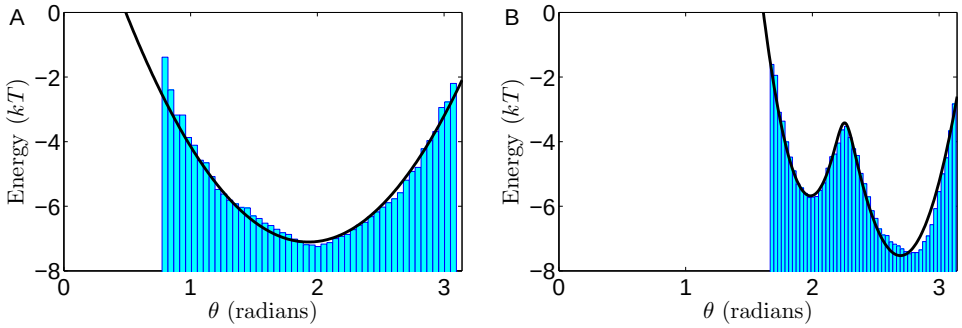

FIG. S2: Two representative examples of angle potentials (A) Harmonic potential between $S-P_{3'}-S_{3'}$, and (B) Bi-harmonic potential between $T-S-P_{3'}$.

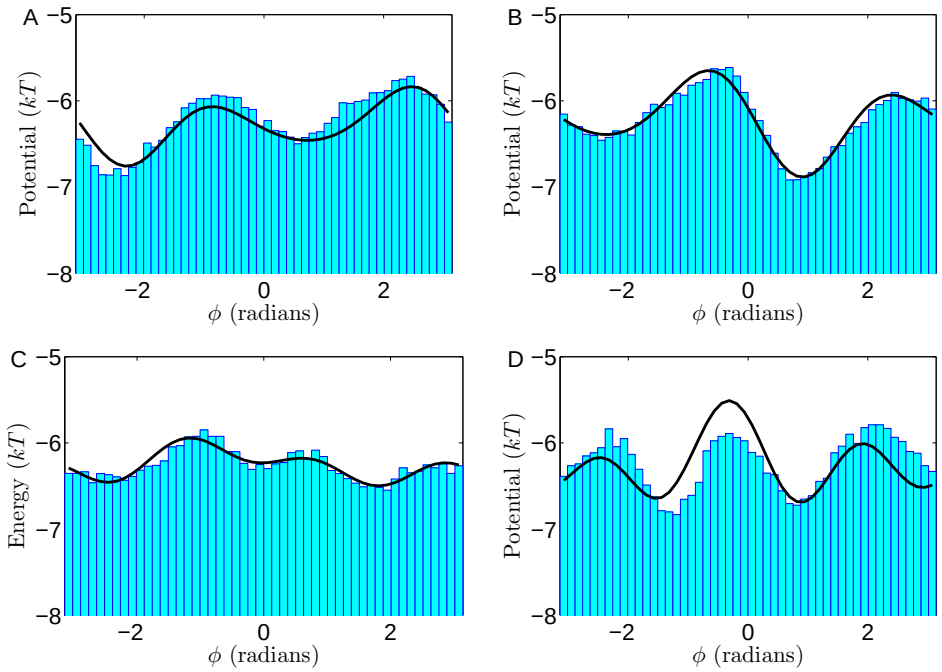

FIG. S3: Dihedral potentials for the angles (a) $S_{5'}P_{5'}SP_{3'}$, (b) $S_{5'}P_{5'}SA$, (c) $GSP_{3'}S_{3'}$, and (d) $P_{5'}SP_{3'}S_{3'}$. The potential fits for $S_{5'}P_{5'}SP_{3'}$ and $P_{5'}SP_{3'}S_{3'}$ result from taking the parameters obtained by averaging over all bases and given in Table SIX. We do not take these potentials to be base dependent.

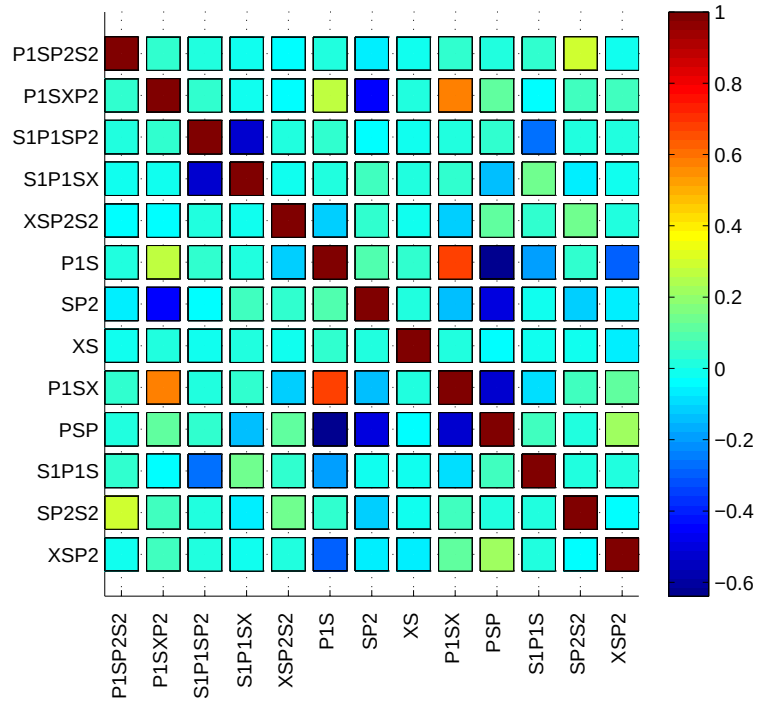

FIG. S4: The extent of coupling between degrees of freedom in our model can be investigated by calculating the correlation coefficient between coordinates from the all-atom simulations that were used to define the coarse-grained potentials. Here for notational simplicity X represents a base, $1 \equiv 5'$ coordinate and $2 \equiv 3'$ coordinate. For example the correlation between r_{P_1S} and r_{XS} is calculated as:

$$C = \frac{\left\langle \left(r_{P_1S} - \langle r_{P_1S}\rangle\right)\left(r_{XS} - \langle r_{XS}\rangle\right)\right\rangle}{\sqrt{\left\langle \left(r_{P_1S} - \langle r_{P_1S}\rangle\right)^2\right\rangle \left\langle \left(r_{XS} - \langle r_{XS}\rangle\right)^2\right\rangle}}$$

Above is the resulting color-coded plot the 13x13 table of correlation coefficients. Diagonal terms are unity; off-diagonal terms representing cross correlations are generally much smaller, indicating modest (but nonzero) effects due to coupling between the degrees of freedom. Larger values of cross-correlations are typically between overlapping degrees of freedom, e.g. between bonds that formed part of angle: $r(P_1S)$ and $\theta(P_1SX)$, $r(P_1S)$ or $r(P_2S)$ and $\theta(P_1SP_2)$; or between angles/dihedrals and angles/dihedrals that had overlapping coordinates: $\theta(P_1SP_2)$ and $\theta(P_1SX)$, $\theta(P_1SX)$ and $\phi(P_1SXP_2)$, $\phi(S_1P_1SP_2)$ and $\phi(S_1P_1SX)$.

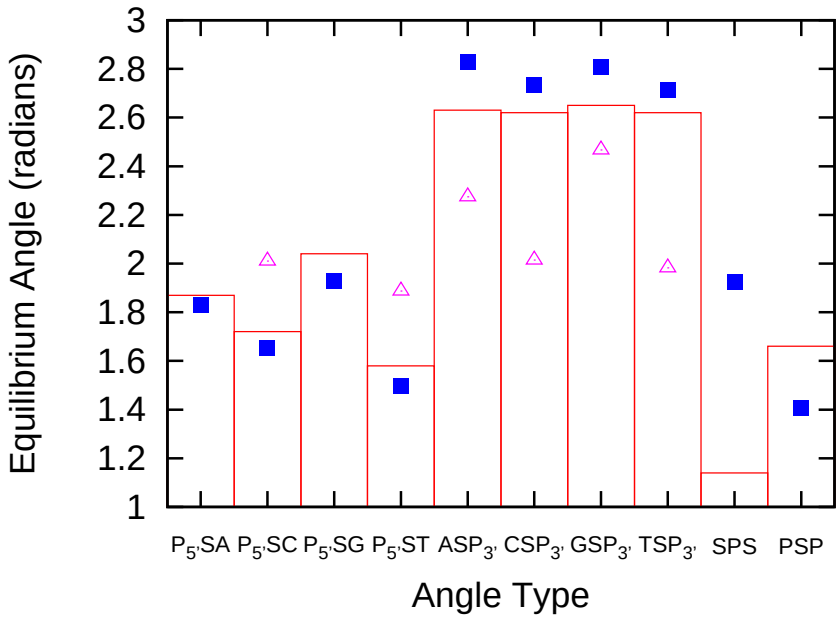

FIG. S5: Angle equilibrium values for the crystal structure (histogram), primary parametrized minima (solid squares) and secondary wells (open triangles, where applicable). The angles that are in the PDB crystal structure (as per the CRC *Handbook of Biochemistry and Molecular Biology 3e*) are shown as a histogram. The data points are the minima of the coarse-grained potentials, as parameterized from a small segment of ssDNA to avoid double counting of interactions. In cases where the potential is a double well, the locations of both the primary and secondary minima are shown. For most types of angle potentials, the minimum of the potential well lies fairly close to the observed value in the crystal structure, and in cases with double-welled angle potentials, the PDB angles lie between the two parameterized wells, closer to the deeper one. Thus our parameterization taken from a modified trimer of ssDNA in fact biases the angle potentials to a that in a double helix quite well. As can be readily seen, there are two exceptions corresponding to angles along the backbone. This is not a problem, however, as there are many factors which will influence what angles the coarse-grained will take in relation to each other. The most significant factor will be adjacent bases along either strand trying to stack with each other and pulling the backbone into the needed angles. This has been noted by Tepper *et al* in J. Chem. Phys. 122, 124906 (2005), who set the equilibrium of the backbone angle potentials to π, which lies above all the ordinates in the graph.

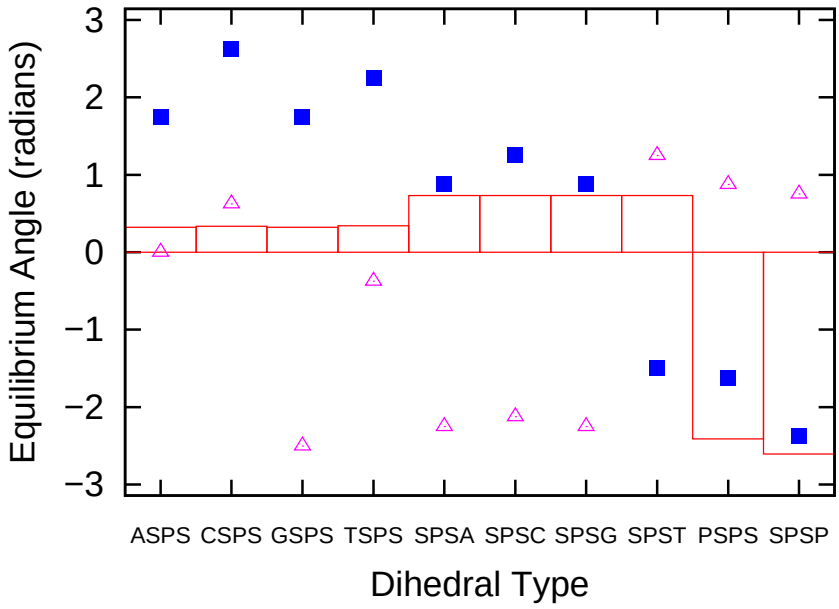

FIG. S6: Dihedral angle equilibrium values for the crystal structure (histogram), primary parametrized minima (solid squares) and secondary parametrized minima (open triangles).

S1.3 Supplementary figures

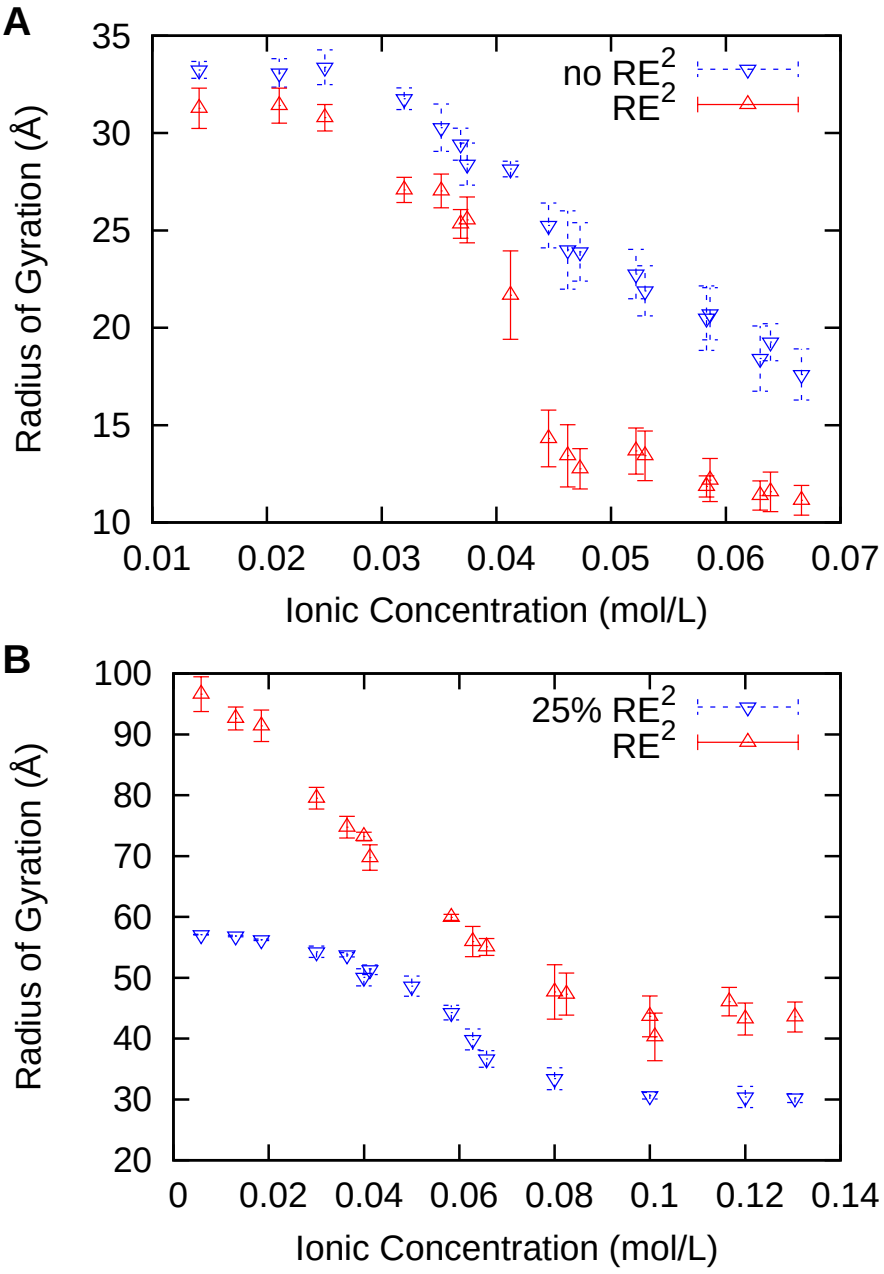

FIG. S7: Dependence of the radius of gyration on the ionic concentration. (A) Shows the results for a single strand of DNA and (B) for a double helix. The results are taken from the same simulations as shown in Fig. 4 A and B respectively.

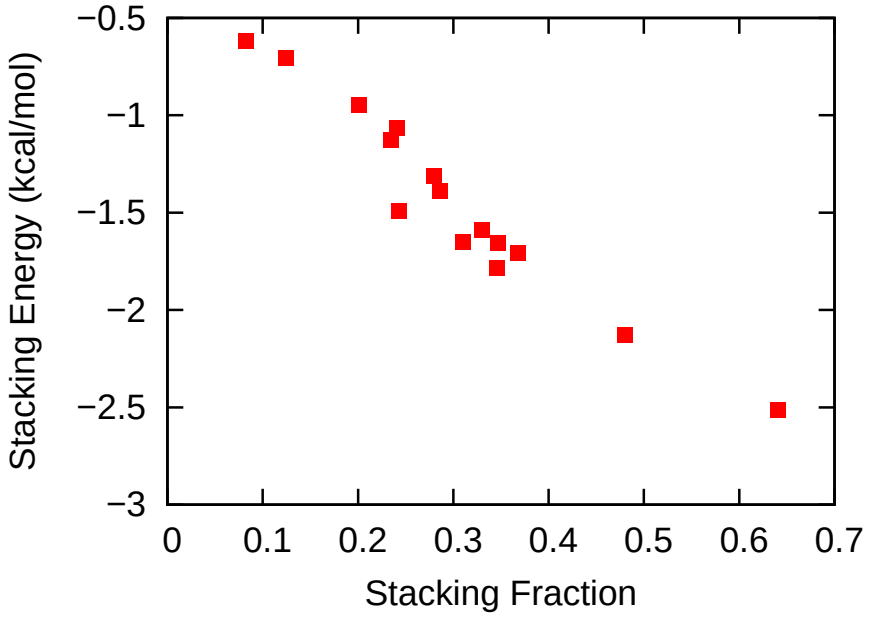

FIG. S8: Nearest neighbor stacking energy plotted as a function of stacking fraction, a geometrical order parameter. Results were obtained from a number of simulations run at temperatures varying from 0 to 0.34ϵ. These results were taken from the same simulations as the heterogeneous sequence from Fig. 14.

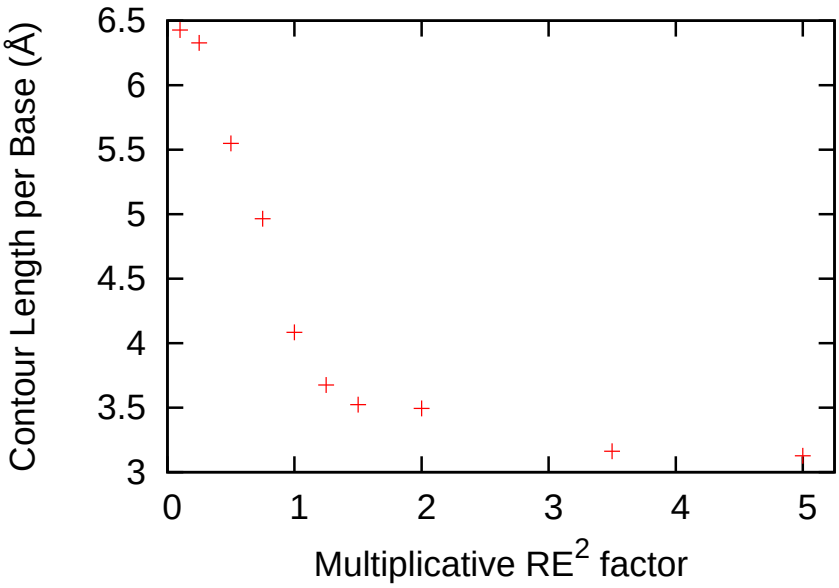

FIG. S9: Effect of scaling the RE^2 energies on the contour length of double stranded DNA. Contour length shows an initial rapid convergence to about 3.5Å per stacked base-pair at about 1.4 times the RE^2 energy obtained from the all-atom parametrization, then a slow convergence to its asymptotic value of about 3Å per stack. The RE^2 attractive potential lowers the system's energy by compressing the polymer. Simulation parameters are the same as in Fig. 4B.