Supporting Information: Chemically Accurate Coarse-Graining of Double-Stranded DNA

Alexey Savelyev[†] and Garegin A. Papoian[†]

[†]Department of Chemistry, University of North Carolina, Chapel Hill

1 Preparation, Parametrization and Molecular Dynamics Simulation Protocol of the All-Atom DNA System

We started from the canonical B-form of a 16-base-pair DNA oligomer $[d(CGAGGTTTAAACCTCG)]_2$ (1), which was built with the *nucleic acid generator* ('nucgen'), a part of the of the AMBER 10.0 suit of programs (2). The refined Amber Parmbsc0 force field for nucleic acids (3) was used to parametrize our DNA segment. The initial DNA structure was first neutralized by 15 Na^+ ions. An extra ~ 0.12 M of NaCl buffer (14 additional Na^+ ions and 14 Cl^- ions), corresponding to the physiological salt concentration, was then added to the system. We used a recently developed force-field for alkali and halide monovalent ions to parametrize inter-ionic and ion-water interactions, as well as the interactions between ions and the DNA (4). It is important to note that the earlier default force-field for monovalent ions in AMBER led to some overestimation of ionic pairing, as discussed in our recent works (5–7) as well as the works of others (8–10). The initial positions of the ions were determined from the computed electrostatic potential using LEaP module of the AMBER 10.0 (2). The system was further solvated in more than 6500 TIP3P (11) water molecules in a cubic box, having dimensions of $60 \times 60 \times 60$ Å. As a result, two DNA segments from neighboring periodic images were at least 35 Å apart. The overall number of atoms in the system was ~ 20000 in the periodic box. To equilibrate this starting structure we used a multistage equilibration process, reported by Orozco and coworkers (12). The subsequent production run was carried out at constant temperature (300 K) and pressure (1 bar) using the Langevin temperature equilibration scheme (see AMBER 10.0 manual), the "weak-coupling" pressure equilibration scheme (13), and periodic boundary conditions. The translational center-of-mass motion was removed every 2 ps. We used the SHAKE algorithm (14) to constrain all bonds involving hydrogens, which allows all MD simulations to use an increased time step of 2 fs without any instability. Particle Mesh Ewald method (15) was used to treat long-range interactions with a 9 Å nonbonded cutoff. The production run was carried out for 60 ns to ensure the equilibration of ions.

It was shown in prior works (16, 17) that 50 ns MD was enough to equilibrate the Na^+ atmosphere around DNA in a smaller system comprised of ~ 16000 atoms. Given a slightly larger size of our systems (~ 20000 atoms), we used extra 10 ns of MD to ensure equilibration.

2 Preparation and Molecular Dynamics Simulation Protocol for the Coarse-Grained Systems

We used the Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) (18) to carry out MD simulations of various coarse-grained (CG) systems. We simulated two linear DNA chains of different number of base-pairs and also one torsionally stressed circular DNA chain. The macromolecule used for the derivation of the CG Hamiltonian parameter set with the MRG-CG technique was the linear DNA segment of the same size as the all-atom system and comprised of 32 beads (16 base-pairs), whose initial coordinates were the geometric centers of the corresponding all-atomistic base-pair nucleotides. We call this system **1**. Next, we have carried out 7 MD simulations of much longer linear DNA chain of 200 base-pairs (400 beads), which was immersed in the NaCl salt buffer of various concentrations in a range $\sim [0.1 - 100]$ mM, to study the dependence of the DNA persistence length on the solution ionic strength. We refer to all those systems as system **2**. Finally, we have built and simulated a 90-base-pair DNA nanocircle with imposed torsional stress to predict the structural phase transition to a buckled state upon varying NaCl concentration in a range of $C \sim [1 - 500]$ mM. Details on the geometry and preparation of this nanocircle, which we call system **3**, are provided in the Methods Section of the main text. The Biochemical Algorithms Library (BALL) (19) was used to build all the models.

All CG DNA molecules were first neutralized by the necessary amount of the Na^+ ions and additional NaCl molecules were added to reproduce the physiological, or other conditions. Note there was no water in CG systems. The size of the periodically repeating cubic cell in the system **1** was the same as in the corresponding all-atom system, and the Particle Mesh Ewald technique was

used to treat the long-range (Coulomb) interactions. To be consistent when going from atomistic to coarse-grained simulations, we kept the sizes of simulation boxes equal to those of AA system, such that both the cutoff for direct Coulomb interactions and, importantly, the size of periodic cell along with the precision of the Ewald summation procedure (which defines the number of k-space points) remain unchanged. Our experience with varying the periodic box sizes suggests that the latter requirement is necessary, since different conditions for Ewald procedure may influence electrostatics. Following these requirements, the number of ions and the size of periodically repeating cubic cell in system **1** were the same as in the corresponding AA system.

Long DNA molecule in a set of systems **2** was immersed in a large periodic box having dimensions $\sim 800 \times 800 \times 800$ Å, and circular 90-base-pair DNA in a set of systems **3** was placed in a periodic cube of the linear dimension of ~ 120 Å. The number of mobile Na^+ and Cl^- ions in systems **2** and **3** varied to reproduce the desired ionic concentration.

Initially all systems were minimized according to the standard steepest descent algorithm. Then they were heated up to 300 K during the 5 ns and subsequently equilibrated for another 10 ns. We used the canonical NVT integration scheme (Nose/Hoover temperature thermostat) to update particle's positions and velocities at each timestep (20). To determine the biggest timestep that can be used to simulate the CG system without instabilities, we used the criteria of the total energy conservation, the latter being the energy of the CG system complemented by the contribution from the Nose-Hoover Hamiltonian (21). It turned out that it was safe to use the timesteps of up to 10 fs, so we used this upper limit in our MD simulations. The production run for each optimization iteration was 20 ns to ensure the convergence of the covariance matrix in Eq. (1), see the main text. We verified the convergence at each iteration by comparing the data generated by two halves of the MD trajectory.

3 Parametrization of the Coarse-Grained Models

Our 2-bead model of the double-stranded DNA chain is structurally analogous to the DNA model developed in our prior work (22). However, due to the presence of mobile Na^+ and Cl^- ions in the current model, Hamiltonian parameters for the DNA chain needed to be re-optimized with the MRG-CG technique (see the Table 1). As for the ionic interaction parameters between the mobile ions, we have taken those from our recent work on the coarse-graining the physiological NaCl solution (23), see the Table 2. As mentioned in the main text, we did not optimize inter-ionic potentials as they generated, after optimizing the rest of the Hamiltonian parameter set, very plausible inter-ionic radial distribution functions, as shown in Figure 1.

In contrast, the derivation of the functional forms and the corresponding parameters describing ion-DNA interactions was a nontrivial aspect of the present study. In particular, trial interaction potentials among the beads of DNA and Na^+ and Cl^- ions extracted from the all-atom system could not be directly used in our CG model. Indeed, conventional Boltzmann inversion procedure used for derivation of other types of structural potentials resulted in significantly overestimated attraction/repulsion among the bead of DNA and Na^+ /Cl^- ion. Apparently, this was caused by correlation effects from the neighboring DNA beads (due to chain connectivity). To cut off these correlation effects and single out a "typical" interaction between DNA bead and mobile ions we obtained an MD trajectory of a separate system comprised of a number of unconnected DNA "monomers", dymethylphosphate (DMPH) ions, and Na+ and Cl ions. Those DMPH ions were constructed by "clipping out" the phosphate group from the backbone strand of the all-atom DNA system, enveloped by 2 methyl groups. Each DMPH ion carries the net charge -1, immersed in the NaCl salt buffer. Further details for atomistic DMPH parameterization and MD simulations can be found in our prior work (7). The CG Hamiltonian terms obtained using this approach is introduced in the main text, Eq.(6), while all the defining parameters are given in the Table 3.

Extracted effective interaction potentials among DMPH ions and Na^+ and Cl^- ions appeared

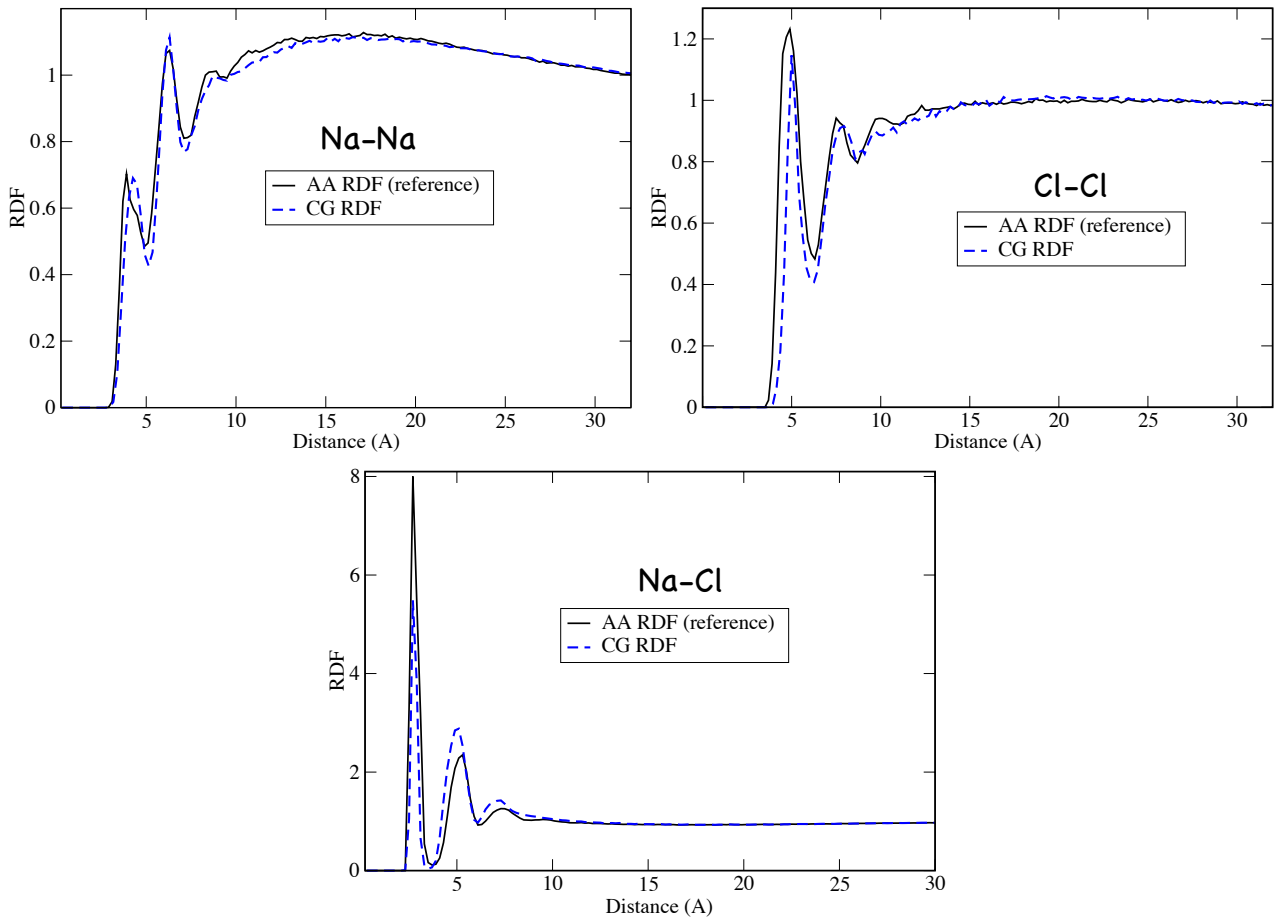

Figure 1: The comparison between inter-ionic radial distribution functions computed from the reference all-atom system (solid black lines) and the corresponding coarse-grained system (blue dashed lines).

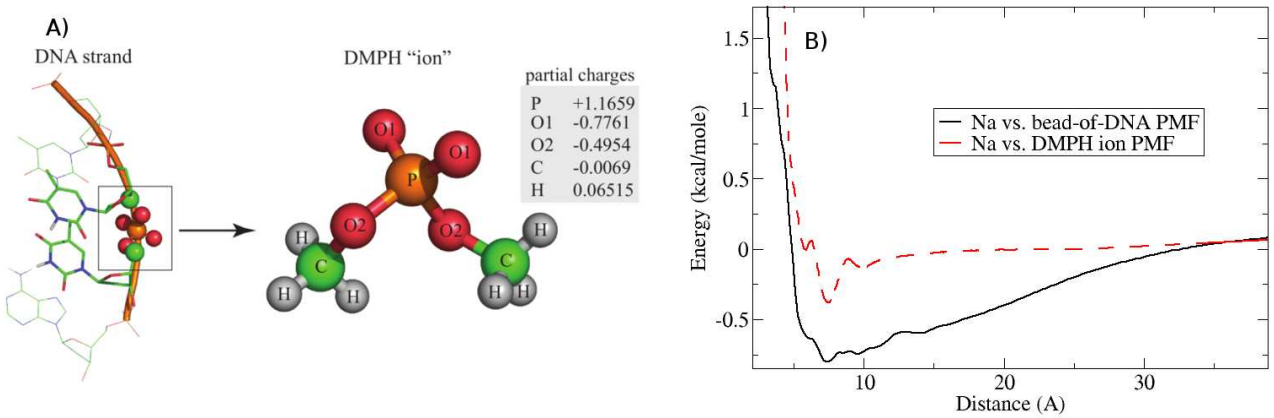

Figure 2: (A) The dimethyl phosphate (DMPH) ion, $(CH_3)_2PO_4^-$ (on the right), was obtained by adding hydrogens, to complete the valence of carbon atoms, to the $(-C-PO_4^--C-)$ functional group,"clipped out" from the DNA strand backbone (shown on the left). Partial charges of the DMPH anion atoms are provided in units of the electron charge. (B) Trial effective interacion potential between single DNA bead and Na^+ ion extracted from the actual system of the DNA chain in the aqueous NaCl solution (black curve) and the analogous system where DNA polymer is substituted with the unconnected DMPH ions (red dashed curve). It is seen that DNA connectivity (correlation effects caused by neighbouring DNA beads) results in a significantly overestimated ion–bead-of-DNA trial potential (black curve).

to be an excellent starting point for describing ion-DNA interactions in our CG model. Indeed, as shown in the Figure 3 of the main text those potentials generated both 1D and 2D distributions of mobile ions around DNA which were already very much consistent with the atomistic results. Notably, only 2 iterations of the following MRG-CG procedure were needed to improve those distributions even more. On the other hand, had we started from the trial ion–bead-of-DNA potentials extracted from the actual system – DNA *polymer* and mobile Na^+ and Cl^- ions – the number of MRG-CG iterations would be significantly larger. In addition, the convergence of the Hamiltonian parameter set always may be an issue when starting from unfavorable initial conditions. This point is illustrated in the Figure 2b.

4 Physical and Mathematical Background of the MRG-CG Technique

In the Methods Section we introduced the main relations of the MRG-CG method. Although the technique is based on the RG Monte Carlo method by R. Swendsen which was used as long as 30 years ago to study criticality in spin systems (24), we recently applied it to complex molecular systems, providing a new interpretation in light of the Field Theory. The latter, in particular, invites for an immediate method generalization by means of reproducing the moments of the order 2 and more for the physical observables entering the Hamiltonian. This is briefly mentioned in the Methods Section. Below we discuss other possibilities to enhance the method accuracy and also provide some important mathematical details on solving set of linear equations, Eq.(1) of the main text. Additional technical discussion can be found in prior works (22, 23).

4.1 Field-theoretical interpretation.

In the present study and our earlier works (22, 23) we have limited ourselves with matching only the first moments of the physical observables, which are the first derivatives of the partition function over the corresponding conjugate fields, see Eq.(2). According to the general field-theoretical scheme (25), we call this approach *mean-field*, although this terminology is ambiguous. To avoid misunderstanding, we emphasize that the "mean-field" here refers to the conjugate fields $\{K_\alpha\}$ (Hamiltonian parameters) coupled to the *collective* observables, $\{S_\alpha\}$, which may be complex functions of the "low-level" structural system characteristics, such as inter-particle separation, bending angles, and the like. For example, the role of $\{S_\alpha^{\rm angle}\}$ is played by various powers of the "primary" order parameter, $(\theta - \theta_0)$, which is the deviation from the equilibrium bending angle. The same is true for the bond and all of the fan interactions where the role of different observables is played by various monomials of the deviation from the equilibrium interparticle separation (see the Methods Section of the main text). The variable $(r - r_0)$ itself is the characteristic of primary

importance for, say, for chemists investigating ionic association/dissociation, where the inter-ionic distance *is* the plausible reaction coordinate. Because observables $\{S_\alpha\}$ represent monomials of $(\theta - \theta_0)$ and $(r - r_0)$ of the order 2 *and* more, matching the first moments in distributions of $\{S_\alpha\}$ really captures the partition function details well beyond from what can be expected from a simple mean-field picture.

This argument applies with lesser transparecy to the collective modes describing structure of the ionic "shells", cumulative Gaussian functions, $S_\alpha^{\rm Gauss} = \sum_{\rm allpairs} {\rm e}^{-C^{(\alpha)}\left[r-R^{(\alpha)}\right]^2}$, due to more complex functional form. However, the concept remains the same – reproducing averages of S_αs effectively captures finer details that are neglected in the mean-field regime.

To summarize, there are two ways to enhance the accuracy of the MRG-CG method. First, as mentioned in the beginning, one can try to reproduce higher-order correlators of physical observables entering linearly the CG Hamiltonian, $\mathcal{H} = \sum_{\alpha=1}^{N} K_\alpha S_\alpha$, and leave the number N of observables fixed. On the other hand, one can introduce additional variables, such as higher-order monomials of $(r-r_0)$, or more complicated functions of r, and remain in the regime, which we call *mean-field* over S_αs, for the lack of a better term. The former approach will necessarily demand dealing with over-determined systems of linear equations, as the initial system of N equations [see Eq.(1), main text] will be supplemented by $\sim N^n$ additional equations for $\Delta\langle S_1 ... S_n\rangle$ (see the Methods Section). The latter approach is conjugate with the search for additional physically motivated observables which would form a more complete basis set. At the same time, such an extended basis set has to be compact enough in order not to generate high degeneracy of the obtained solutions. For example, even in the present CG model of DNA and ions characterized by the total of 47 observables (39 of the DNA and 8 of the ion-DNA structural characteristics included into optimization) we faced the degeneracy problem caused by singularity of the covariance matrix entering Eq.(1) (see the main text). A standard way to reduce the degeneracy is to utilize the Singular Value Decomposition (SVD) technique to eliminate those covariance matrix eigenvectors which superfluously affect Hamiltonian parameters. We elaborate next on the nontrivial *inverse problem*

which we had to overcome before direct use of the SVD.

4.2 Compact Basis Set and Solving the Inverse Problem.

Eigenvalues of the covariance matrix, $\langle S_\alpha S_\gamma \rangle - \langle S_\alpha \rangle \langle S_\gamma \rangle$, indicate how changes in various dynamical modes affect different effective potentials. For the DNA problem, it turns out that the covariance matrix is nearly singular, resulting in degeneracy of solutions that represent various sets of parameters. Apparently, this problem is caused by some redundancy of interaction potential functions as well as by the noise which is normally present in the input data obtained from MD simulations (26, 27). When too many observables are used to describe the CG system, larger uncertainty in the covariance matrix inversion results, and, thus, the stronger the degeneracy of the resulting set of CG Hamiltonian parameters. This implies, in particular, a significant advantage of using our compact set of 47 basis functions associated with DNA and ionic motions, as opposed, for example, to a very large set of ~ 600 positional Dirac delta functions used to describe solely ionic behavior in work by Lyubartsev et al (28). Indeed, each type of interactions in Ref. (28) was defined by order of 200 observables (instead of 3 S_αs associated with DNA bond, bending angle and fan interactions in our case), defined by the corresponding set of positional Dirac delta functions. Since our model of DNA and ions is described by more than 15 interaction potentials, such representation would require to deal with order of 4000 variables, requiring inversion of a matrix of $\sim 10^7$ elements to solve the set of linear equations, Eq.(1) of the main text. Representing bending angle potentials, which are *3-body* interactions, is even more problematic in the positional basis, resulting in immense computational difficulty because of necessity to deal with very large arrays. Note also that had we included the *4-body* dihedral potential in the consideration, the corresponding matrices would be even larger. On the other hand, within our approach this computational difficulty is bypassed by projecting such a large many-dimensional array into a very compact 2-dimensional array defined in a set of basis functions of *different* dimensions (our S_αs).

Because the elements of the covariance matrix, $\langle S_\alpha S_\gamma \rangle - \langle S_\alpha \rangle \langle S_\gamma \rangle$, have dissimilar physical

units, SVD technique can not be directly used to eliminating those matrix eigenvectors which superfluously affect Hamiltonian parameters (to reduce the degeneracy). For example, the matrix element $\langle S_2^{bond} \cdot S_3^{angle} \rangle - \langle S_2^{bond} \rangle \langle S_3^{angle} \rangle$ has a dimension of [$\text{Å}^3 \cdot \text{Rad}^4$], while the diagonal element $\left\langle \left(S_2^{bond}\right)^2 \right\rangle - \langle S_2^{bond} \rangle^2$ is measured in units of [Å^6]. Therefore, to use SVD at each iteration, we reduced the corresponding covariance matrix to a dimensionless form by appropriately rescaling vectors ΔK_α and $\Delta\langle S_\alpha \rangle$. Then, in matrix notation the rescaled Eq.(1) of the main text reads,

$$\sum_j \frac{M_{ij}}{\sqrt{q_i \cdot q_j^T}} \cdot \left[X_j \cdot \sqrt{q_j} \right] = \frac{B_i}{\sqrt{q_i}}, \qquad q_i \equiv M_{ii}, \tag{1}$$

with M, X and B standing for the covariance matrix, vector of the corrections ΔK_α and the vector of deviations $\Delta\langle S_\alpha \rangle$, respectively. As follows from the second equation, vector q is composed from the diagonal elements of the original matrix M. Hence, the latter is reduced to a dimensionless form (with unit elements on the diagonal) after its element-by-element division by the tensor $\sqrt{q_i \cdot q_j^T}$. After filtering out near zero eigenvalues and performing a subsequent matrix inversion, the original units of the elements ΔK_α were obtained by reverse transformation.

In summary, the compact basis set of the basis functions, or observables $\{S_\alpha\}$ possessing various physical units, allowed the inclusion of many-body interactions into optimization scheme, which was one of the principal novelties of our approach.

	Intra-strand		*Inter*-strand (*fan*)										
	Bond	**Angle**	**5p**	**4p**	**3p**	**2p**	**1p**	**bp**	**1m**	**2m**	**3m**	**4m**	**5m**
K_1	2.625	9.22	0.3626	0.1386	0.1378	0.0955	0.115	2.92	0.040	0.123	0.085	1.324e-06	0.0467
K_2	-0.226	4.16	-0.077	-0.0568	-0.0527	-0.0459	-0.041	0.41	-0.010	-0.040	-0.0444	-0.0122	0.0021
K_3	0.0149	1.078	0.005	0.005	0.005	0.00502	0.0058	0.072	0.0008	0.0037	0.005	0.00185	0.000146
l_0, θ_0	4.96	156	16.9	12.5	10.2	9.2	9.9	11.3	12.3	13.45	14.7	16.35	17.1

Table 1: Final parameters for the coarse-grained DNA Hamiltonian defined by polynomials in Eqs.(4) are shown. Intra-starnd bond and bending angle potentials describe the polymeric interactions within each of the DNA strands. Inter-strand interactions are described by the *fan* potential representing the set of constraints among two polynucleotides to maintain the desired DNA structure (see Fig. 1 in the main text). These include, in particular, base pair interactions (denoted as "bp"), as well as the interactions among the given bead located on one strand and a series of beads located on the other strand (denoted as "1..5p" and "1..5m" and corresponding to beads $[(N \pm 1..5) - i]$ in notations of Fig. 1). Parameters K_1 and K_2 and K_3 for all types of interactions, aside from the bending angle potential, are given in units of [kcal/mol·Å^{-2}], [kcal/mol·Å^{-3}] and [kcal/mol·Å^{-4}], respectively. Analogous parameters for bending angle potential are measured in units of [kcal/mol·rad^{-2}], [kcal/mol·rad^{-3}] and [kcal/mol·rad^{-4}], respectively. l_0's are equilibrium interparticle separations for *bond* and *fan* interaction (in Å), and θ_0 is the equlibrium bending angle (in degrees).

	Na^+ –Na^+	Na^+ –Cl^-	Cl–Cl^-
A	$8.4 \cdot 10^5$	$1.14 \cdot 10^5$	$2.7 \cdot 10^7$
B_1	-0.013	-1.06	-0.59
B_2	0.61	2.61	0.33
B_3	-0.2	-0.29	-0.053
B_4	0.086	0.276	0.078
B_5	-0.023	-0.072	-0.022
C_1	5	8.4	5.8
C_2	2.46	3.5	3.12
C_3	4.38	3.52	3.5
C_4	3.73	5.66	4
C_5	7.4	5	9.9
R_1	3.8	2.75	4.9
R_2	4.75	3.52	6.1
R_3	6.2	5.1	7.5
R_4	7.2	6	8.5
R_5	8.7	7.3	9.7

Table 2: Inter-ionic parameters defining the Hamiltonian, Eq.(5), were taken from our prior work on coarse-graining NaCl solutions at various ionic concentrations (23). Parameters A and $B_1...B_5$ are given in units of [kcal/mol·Å^{12}] and [kcal/mol], respectively. Gaussian variances $C_1...C_5$ have units of [Å^{-2}], while positions of Gaussian functions, $R_1...R_5$, are measured in Å.

	Na^+ –DNA-bead	Cl^- –DNA-bead
A	$1.82 \cdot 10^4$	$1.48 \cdot 10^3$
B_1	0.11	-0.06
B_2	0.173	-0.19
B_3	0.035	-0.06
C_1	7.8	9
C_2	1.2	0.5
C_3	0.31	0.11
R_1	6.5	5.1
R_2	8.6	7.5
R_3	11.5	11.0

Table 3: Optimized parameters defining interactions among ions and beads of DNA, Eq.(6). Parameters A and $B_1...B_5$ are given in units of [kcal/mol·Å^6] and [kcal/mol], respectively. Gaussian variances $C_1...C_5$ have units of [Å^{-2}], while positions of Gaussian functions, $R_1...R_5$, are measured in Å.

References

1. McAteer, K, Aceves-Gaona, A, Michalczyk, R, Buchko, G. W, Isern, N. G, Silks, L. A. P, Miller, J. H, & Kennedy, M. A. (2004) Compensating bends in a 16-base-pair DNA oligomer containing a T(3)A(3) segment: A NMR study of global DNA curvature. *Biopolymers* **75**, 497–511.

2. Case, D, Cheatham, T, Darden, T, Gohlke, H, Luo, R, Merz, K, Onufriev, A, Simmerling, C, Wang, B, & Woods, R. (2005) The amber biomolecular simulation programs. *J Comput Chem* **26**, 1668–1688.

3. Prez, A, Marchn, I, Svozil, D, Sponer, J, Cheatham, T. E, Laughton, C. A, & Orozco, M. (2007) Refinement of the amber force field for nucleic acids: improving the description of alpha/gamma conformers. *Biophys J* **92**, 3817–3829.

4. Joung, I. S & Cheatham, T. E. (2008) Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J Phys Chem B* **112**, 9020–9041.

5. Savelyev, A & Papoian, G. A. (2007) Inter-dna electrostatics from explicit solvent molecular dynamics simulations. *J Am Chem Soc* **129**, 6060–6061.

6. Savelyev, A & Papoian, G. A. (2007) Free energy calculations of counterion partitioning between dna and chloride solutions *Mendeleev Commun* **17**, 97–99.

7. Savelyev, A & Papoian, G. A. (2008) Polyionic charge density plays a key role in differential recognition of mobile ions by biopolymers. *J Phys Chem B* **112**, 9135–9145.

8. Chen, A. A & Pappu, R. V. (2007) Parameters of monovalent ions in the amber-99 forcefield: assessment of inaccuracies and proposed improvements. *J Phys Chem B* **111**, 11884–11887.

9. Chen, A. A & Pappu, R. V. (2007) Quantitative characterization of ion pairing and cluster formation in strong 1:1 electrolytes. *J Phys Chem B* **111**, 6469–6478.

10. Auffinger, P, Cheatham, T. E, & Vaiana, A. C. (2007) Spontaneous formation of kcl aggregates in biomolecular simulations: A force field issue? *J Chem Theory Comput* **3(5)**, 1851–1859.

11. Miyamoto, S & Kollman, P. A. (1992) Settle: An analytical version of the shake and rattle algorithm for rigid water models. *J Comput Chem* **13**, 952–962.

12. Shields, G. C, Laughton, C. A, & Orozco, M. (1998) Molecular dynamics simulation of a pna·dna·pna triple helix in aqeous solution *J Am Chem Soc* **120**, 5895–5904.

13. Berendsen, H. J, Postma, J. P, van Gunsteren, W. F, DiNola, A, & Haak, J. R. (1984) Molecular dynamics with coupling to an external bath. *J Chem Phys* **81**, 3684–3690.

14. Ryckaert, J.-P, Ciccotti, G, & Berendsen, H. J. (1977) *J Comput Phys* **23**, 327–341.

15. Darden, T, York, D, & Pedersen, L. (1993) Sequence-specific binding of counterions to B-DNA. *J Chem Phys* **98**, 10089–10092.

16. Ponomarev, S. Y, Thayer, K. M, & Beveridge, D. L. (2004) Ion motions in molecular dynamics simulations on DNA. *Proc Natl Acad Sci U S A* **101**, 14771–14775.

17. Varnai, P & Zakrzewska, K. (2004) Dna and its counterions: a molecular dynamics study. *Nucleic Acids Res* **32**, 4269–4280.

18. Plimpton, S. (1995) Fast parallel algorithms for short-range molecular dynamics *Journal of Computational Physics* **117**, 1–19.

19. Kohlbacher, O & Lenhof, H. P. (2000) BALL–rapid software prototyping in computational molecular biology. Biochemicals Algorithms Library. *Bioinformatics* **16**, 815–824.

20. Hoover, W. G. (1985) Canonical dynamics: Equilibrium phase-space distributions *Phys Rev A* **31**, 1695–1697.

21. Knotts, T. A, Rathore, N, Schwartz, D, & de Pablo, J. J. (2007) A coarse grain model fro dna *J Chem Phys* **126**, 084901.

22. Savelyev, A & Papoian, G. A. (2009) Molecular renormalization group coarse-graining of polymer chains: application to double-stranded dna. *Biophys J* **96**, 4044–4052.

23. Savelyev, A & Papoian, G. A. (2009) Molecular renormalization group coarse-graining of electrolyte solutions: application to aqueous nacl and kcl. *J Phys Chem B* **113**, 7785–7793.

24. Swendsen, R. H. (1979) Monte carlo renormalization group *Phys Rev Lett* **42**, 859–861.

25. Zinn-Justin, J. (2002) *Quantum Field Theory and Critical Phenomena*. (Clarendon press, Oxford).

26. Lyubartsev, A. P & Laaksonen, A. (2000) Determination of effective pair potentials from ab initio simulations: application to liquid water *Chem Phys Lett* **325**, 15–21.

27. Lyubartsev, A. P. (2005) Multiscale modeling of lipids and lipid bilayers. *Eur Biophys J* **35**, 53–61.

28. Lyubartsev, A. P & Laaksonen, A. (1995) Calculation of effective interaction potentials from radial distribution functions: A reverse monte carlo approach. *Phys Rev E* **52**, 3730–3737.