University of Groningen

Martini Coarse-Grained Force Field

Uusitalo, Jaakko J.; Ingolfsson, Helgi I.; Akhshi, Parisa; Tieleman, D. Peter; Marrink, Siewert J.

Published in:
Journal of Chemical Theory and Computation

DOI:

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

# Martini Coarse-Grained Force Field: Extension to DNA

Jaakko J. Uusitalo,[†] Helgi I. Ingólfsson,[†] Parisa Akhshi,[‡] D. Peter Tieleman,[‡] and Siewert J. Marrink[*,†]

[†]Groningen Biomolecular Sciences and Biotechnology Institute and Zernike Institute for Advanced Materials, University of Groningen, Nijenborgh 7, 9747 AG Groningen, The Netherlands

[‡]Department of Biological Sciences and Centre for Molecular Simulation, University of Calgary, 2500 University Drive NW, Calgary, Alberta, Canada T2N 1N4

Ⓢ Supporting Information

ABSTRACT: We systematically parameterized a coarse-grained (CG) model for DNA that is compatible with the Martini force field. The model maps each nucleotide into six to seven CG beads and is parameterized following the Martini philosophy. The CG nonbonded interactions are based on partitioning of the nucleobases between polar and nonpolar solvents as well as base–base potential of mean force calculations. The bonded interactions are fit to single-stranded DNA (ssDNA) atomistic simulations and an elastic network is used to retain double-stranded DNA (dsDNA) and other specific DNA conformations. We present the implementation of the Martini DNA model and demonstrate the properties of individual bases, ssDNA as well as dsDNA, and DNA–protein complexes. The model opens up large-scale simulations of DNA interacting with a wide range of other (bio)molecules that are available within the Martini framework.
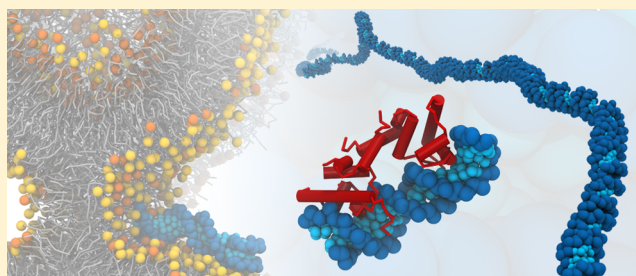
INTRODUCTION

Coarse-graining in biomolecular simulations reduces the complexity of an atomistic simulation system by averaging over nonessential degrees of freedom. In practice, the reduction is achieved by replacing groups of atoms by larger units that approximate the interactions of the underlying atoms. Such structural coarse-graining speeds up simulations by reducing the number of computations performed per timestep and enables the use of larger timesteps due to smoothening of the energy landscape. If done carefully, CG models offer a powerful tool to study systems at increased length and time scales with only limited loss of accuracy.[1,2]

DNA seems like an ideal molecule for CG molecular dynamics (MD) simulations, given the problematic nature of sampling DNA conformations at the all-atom level.[3−5] In double-stranded (dsDNA) form DNA molecules span sizes from nanometers to centimeters and important structural features arise at widely different length scales, which can be addressed with different levels of coarse-graining. Initially, CG models for DNA were developed mainly for large length scales, where individual base pairs are largely ignored and the main focus is on elastic and packing properties of dsDNA.[6−9] More recently, an increasing number of models closer to atomistic resolution have been developed. For example, the 3SPN (three sites per nucleotide) models[10,11] have enough detail to include specific hydrogen bonding and base stacking potentials and derive their bonded interactions from the canonical B-DNA structure. These models describe elastic properties of DNA as well as DNA melting fairly well. The OxDNA model[12,13] also reproduces both thermodynamical and structural properties of DNA successfully by parameterizing the CG interactions

against experimental stacking propensity and dsDNA melting temperatures as well as the ideal B-DNA structure. Another CG model with three beads per nucleotide by Linak et al.[14] is also able to describe some non-Watson–Crick type base pairs using directional potentials. A similar model by Maciejczyk et al.[15] uses potentials of mean force calculations to derive the CG potentials. The model by Maffeo et al.[16] models ssDNA with two beads per nucleotide but does not describe sequence specificity. The SIRAH force field has a CG DNA model[17] that differs from the others by using more CG beads per nucleotide (six beads) with spherically uniform interactions. Another example of a CG DNA model with higher resolution is the HiRE model[18] that describes each nucleotide with seven beads. Traditionally, CG DNA models parameterize the solvent implicitly, and to our knowledge, only the SIRAH model has an explicit solvent version available. Most of these models are focused on the structural or thermodynamic properties of DNA. There have not been many CG models that can describe DNA interactions with other biomolecules. Two models that do include DNA interactions with other biomolecules are the PRIMO(NA)[19] model and the model by Poulain et al.[20] In addition, Liwo et al.[21] have announced that they are in the final stages of combining their protein and DNA force fields (UNRES and NARES-2P, respectively).

Here, we present a Martini[22] CG DNA model systematically parameterized according to the Martini philosophy and, therefore, compatible with other Martini models for biomolecules and solvents. The CG Martini force field[22−24]

combines top-down and bottom-up approaches for parameterization. The main targets are experimental values like densities of liquids and partitioning free energies of small solutes between polar and nonpolar solvents which are used to determine nonbonded interaction parameters as well as atomistic reference simulations which are used mainly to extract bonded interaction parameters but also to fine-tune the nonbonded interactions. Martini maps roughly four non-hydrogen atoms to one CG bead and has a restricted number of bead and interaction types. Each bead type describes one or more chemical building blocks and mimics their properties. This approach makes the CG Martini models transferable to different systems and compatible with each other. The Martini force field has parameters available for an increasing number of molecules, including lipids,[22,23] proteins,[25,26] carbohydrates,[27] and polymers,[28,29] and includes aqueous solvent explicitly at two different available levels of detail.[23,30] Since transferability and compatibility are intrinsic properties of models built using the Martini force field, a CG Martini DNA model opens up possibilities for CG DNA simulations of complex biological systems that other approaches have not been able to address. Previous work on nucleotides using the Martini force field include an *ad hoc* version of a small piece of Martini DNA, parameterized by Khalid and co-workers[31] and used to study complexes of DNA and lipids.[31,32] A modified version of the SIRAH DNA model has also been adapted to work with Martini solvent to simulate DNA hairpin translocation and unzipping in nanopores.[33] A preliminary version has also been used in a model of a lipid nanoparticle containing RNA.[34,35] A separate model was made by Kim et al.[36] to study RNA–peptide complexes.

The parameterization approach we take here follows the general strategy for Martini, combining top-down information from experiment (notably, partitioning free energies) with bottom-up information derived from reference atomistic simulations. In particular, we first selected bead types for the nucleobases based on partition free energies from water to chloroform or hydrated octanol. The bonded interactions were fitted to reproduce bond, angle, and dihedral distributions obtained from atomistic simulations of short ssDNAs. Finally, for dsDNA, an elastic network was optimized to retain the double helical structure and the persistence length of dsDNA. Our DNA model can be used to simulate both ssDNA and dsDNA and is compatible with all other Martini models. The main limitation of the model is in the base-pairing, which is not specific enough and requires the elastic network to keep dsDNA in its canonical form. Reproducing melting curves or spontaneous hybridization is therefore not within the scope of our model. Furthermore, the elastic network used for the dsDNA will keep the DNA structure close to its initial form. This allows for simulations of various DNA structures with or without other biomolecules but limits local reorganization which makes our model less suitable to study structural properties of dsDNA.

■ MODEL

In Martini DNA, each nucleotide is mapped to six or seven CG beads. The backbone is modeled with three beads by mapping the phosphate to one and the sugar to two beads. The pyrimidines (cytosine and thymine) are modeled as three-bead rings and the purines (adenine and guanine) as four-bead rings. The mapping of each base is illustrated in Figure 1, and the exact mappings are listed in Table S1 of the Supporting
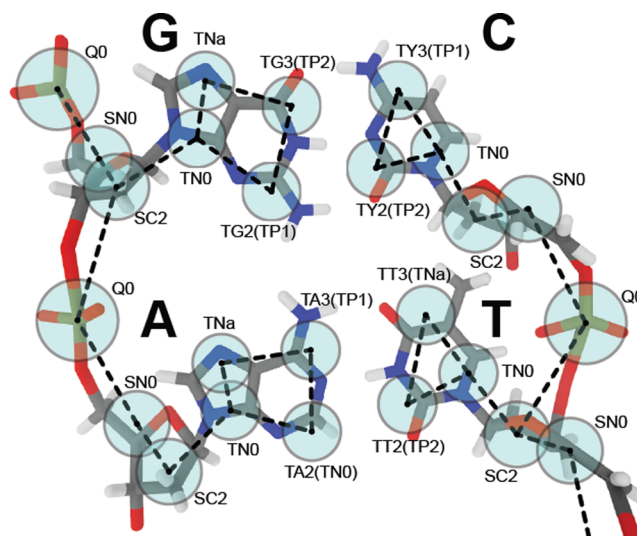
Figure 1. CG mapping of Martini DNA. The DNA backbone is modeled with one bead describing the phosphate and two beads describing the sugar. The pyrimidines are modeled with three beads and the purines with four beads. The Martini bead type of each bead is shown; the T-prefix marks the beads that use the new tiny bead type. For hydrogen bonding beads, the new special bead types are shown together with the bead type describing their interactions with all beads except the special hydrogen bonding beads (in parentheses).

Information. For each residue, the beads are divided into backbone beads (BB1, BB2, and BB3) and side chain beads (SC1, SC2, SC3, and SC4 for the purines). The first backbone bead (BB1) is the phosphate and last one (BB3) the 3′ end of the sugar. For the side chains, the beads are defined in cyclical order so that SC1 is attached to backbone, and in dsDNA, the SC2 and SC3 beads would be base-pairs with the opposing strand.

The bases in dsDNA are stacked very close to each other, with a distance between bases of only 0.34 nm. These stacked rings lead to problems in a CG model that describes the underlying atomistic structure with large isotropic beads. In Martini, approximately four heavy atoms are described with a single bead. The spherical shape is necessary partly due to the simplicity of the model but is commonly justified by the fairly uniform distribution of underlying atoms inside the spheres. While this approximation is good enough for most atomistic structures, it fails for thin, planar molecules like the nucleobases since the beads are too large in the dimension perpendicular to the plane of the base. Standard Martini beads use a LJ parameter $\sigma = 0.47$ nm, and even the smaller ring type beads (S type) use $\sigma = 0.43$ nm, which is too large to model base stacking. In order to fit the bases at the correct distance in the CG DNA structure, we therefore created an even smaller bead type. We selected $\sigma = 0.32$ nm for this new bead type and denote these beads with T (for tiny). T particles interact with each other using the reduced σ but behave as S type particles toward standard and S type particles ($\sigma = 0.43$ nm and epsilon reduced to 75% with S particles and $\sigma = 0.47$ nm with regular particles).

Another addition to standard Martini was made to describe hydrogen bonding between the bases. Martini does not have any special interactions to model directional hydrogen bonds, although they are crucial for the formation of dsDNA and the specificity of base pairing. We therefore specifically tuned the interactions between the hydrogen bonding beads. Since

Martini explicitly defines interactions pairwise for each bead type, we used special bead types for the hydrogen bonding particles, adding eight special beads that are meant solely for this purpose. These beads interact with all other bead types based on their underlying chemical group (e.g., TA2, based on TN0, interacts like TN0 with all other bead types except for the eight special hydrogen bonding bead types), but the energy levels are modified for interactions between these special eight beads. The interaction matrix of these beads is presented in Table S2 of the Supporting Information.

■ METHODS

This section covers the methods used to parameterize and test the DNA model. First, the simulation parameters are detailed for the four different types of CG simulations used: free energy calculations of nucleobases in solution, equilibrium simulations of DNA in solution, equilibrium simulations of DNA–protein systems, and finally, free energy simulations of nucleobases in membranes. Second, simulation parameters for the corresponding atomistic simulations are detailed. Third, the implementation of alchemical free energy calculations and potential of mean force calculations are explained in detail. The section concludes with an explanation of how the CG-bonded interactions were obtained from atomistic simulations and how CG structures after simulation were backmapped to atomistic resolution.

CG Simulations. The molecular dynamics (MD) simulations were performed using versions 4.5.x and 4.6.x of the GROMACS simulation package.[37] The default Martini parameters[22,23] were used for nonbonded interactions: Lennard–Jones (LJ) interactions were shifted between 0.9 and 1.2 nm; Coulomb interactions were shifted between 0.0 and 1.2 nm and screened by a uniform relative electric permittivity (ε_r = 15). In simulations with the polarizable water model,[30] the relative permittivity was adjusted to ε_r = 2.5 and long-range electrostatics were included using PME.[38,39] The pair list radius was set to 1.4 nm and updated every 10th step. The timestep varied depending on the system; the systems including only the nucleobases were simulated with a 20 fs timestep, complete DNA molecules with a 10 fs timestep.

The CG free energy simulations in solution were performed in periodic rhombic dodecahedron boxes with the temperature maintained using the velocity rescaling thermostat[40] at a reference temperature of 298 K and time constant τ_T = 0.5 ps. The Parrinello–Rahman barostat[41] was used for pressure coupling with a reference pressure of 1 bar, 4.5×10^{-5} bar^{-1} compressibility, and a time constant τ_p = 4.0 ps. Equilibrium simulations of CG systems used the velocity rescaling thermostat[40] with a reference temperature of 298 K and time constant τ_T = 0.5 ps and the Berendsen barostat[42] with a pressure of 1 bar, 3×10^{-4} bar^{-1} compressibility, and τ_p = 3.0 ps. For the DNA and DNA–protein stability test simulations, the temperature was maintained at 310 K using the Berendsen temperature coupling algorithm[42] with a time constant τ_T = 2.0 ps. The Berendsen barostat[42] was used to maintain a reference pressure of 1.0 bar using a time constant τ_p = 3.0 ps and a compressibility of 3.0×10^{-5} bar^{-1}. For membrane partitioning free energy simulations, the temperature was maintained at 303 K using the Berendsen temperature coupling algorithm[42] with a time constant τ_T = 0.3 ps and semi-isotropic pressure coupling was applied using the Berendsen barostat[42] with a reference pressure of 1.0 bar using a time constant τ_p = 3.0 ps and a compressibility of 3.0×10^{-5} bar^{-1}. Note the use of different

coupling schemes illustrates the robustness of our model. Additional tests showed our results to be independent of the specific choice of thermostat or barostat (data not shown). The recommended parameters for Martini DNA simulations can be found at http://cgmartini.nl.

AA Simulation Setup. Atomistic reference simulations were run with a timestep of 2 fs with all bonds constrained. The coupling schemes were identical to the CG simulations, except for membrane partitioning free energy simulations in which a time constant τ_T = 0.1 ps was used for the temperature coupling and a time constant τ_p = 2.0 ps and a compressibility of 4.5×10^{-5} bar^{-1} for the pressure coupling. Parameters used for nonbonded interactions were based on Lavery et al.[43] for AMBER, except for the addition of a short switching region to avoid cutoff artifacts. A neighbor list of 1.2 nm was updated every 10th step, while both LJ and Coulomb interactions were switched off between 0.8 and 0.9 nm. Long-range electrostatics were handled with PME,[38,39] and a dispersion correction was used for both energy and pressure. For CHARMM, parameters from Bjelkmar et al.[44] were used, i.e., a neighbor list of 1.2 nm was updated every 10th step, Coulomb interactions were cutoff at 1.2 nm, and PME was used for the long-range part whereas LJ interactions were switched off between 1.0 and 1.2 nm and dispersion correction was employed for the long-range part. For GROMOS, parameters from Oostenbrink et al.[45] were used. A neighbor list of 0.8 nm was updated every fifth step together with a twin-range setup where Coulomb and LJ interactions were calculated until 1.4 nm. Reaction-field and dispersion correction were used for the long-range interactions.

Together with the CHARMM and AMBER force fields for the bases, the TIP3P[46] water model was used, whereas SPC[47] was used with GROMOS. The membrane partitioning free energy simulations used TIPS3P[48] with CHARMM36.[49] The octanol and chloroform topologies used with the AMBER bases were GAFF[50] taken from Caleman et al.[51] The octanol for CHARMM is based on the CGenFF[52] parameters for ethanol (adapted to octanol by Casteblanco[53]), whereas the chloroform used is the same as the GAFF topology used for AMBER. Although this chloroform model has been previously shown to be compatible with CHARMM27,[54,55] mixing force fields is always a reason for caution. Therefore, we also included a comparison to previously published results of Wolf et al.[56] who studied chloroform water partitioning using different chloroform parameters for CHARMM and AMBER (with CHARMM and OPLS parameters, respectively). We denote the results from Wolf et al. as CHARMM-Wolf and AMBER-Wolf throughout the manuscript. For the GROMOS bases, we used the GROMOS53A6[45] parameters for both octanol and chloroform. For membrane partitioning free energy calculations with AMBER,[57] lipid parameters were taken from refs 58 and 59.

Free Energy Calculations. The partitioning free energies of small building blocks are the main benchmark used to parameterize nonbonded interactions in Martini.[22,23] We determined the bead types for the nucleobases by calculating their partitioning free energies from water to octanol (hydrated with a 0.26 mole fraction of water to match experimental conditions[60]) and from water to chloroform. We used experimental values as well as results from all-atom simulations to benchmark the CG partitioning behavior.

All partitioning free energies are calculated by separately calculating the solvation free energies in both solvents and

using a simple thermodynamic cycle, which yields the partitioning free energy as

$$\Delta G_{s_1 \to s_2} = \Delta G_{s_1 \to \varnothing} - \Delta G_{s_2 \to \varnothing}$$

where s_1 and s_2 denote the two solvents, $\varnothing$ denotes vacuum, and ΔG is the change in free energy related to the process described in the subscript. The free energies of solvation were obtained by simulating the reverse process. The solute was first equilibrated in the solvent and then decoupled from it in discrete steps. The completely decoupled state corresponds to having the solute in vacuum (which approximates its gas phase) and thus the change in free energy from the fully coupled state to the uncoupled state is (the negative of) the solvation free energy of the solute in that solvent. The decoupling was performed in separate steps that makes sampling from equilibrium possible and also allows running the simulations in parallel. The degree of coupling was described using a λ parameter that ranges from 0 (fully coupled) to 1 (fully uncoupled). In all simulations that use the λ parameter the intramolecular interactions were not dependent on λ and were thus fully present regardless of the λ value. Separate simulations of each λ value between 0 and 1 were performed and during each simulation the energy of the system was calculated using the two neighboring λ values at each neighbor search step. These energy differences between the native and foreign λ values were translated into free energy difference using the Bennett's acceptance ratio method (BAR)[61] as implemented in the GROMACS tool g_bar.

The decoupling of the CG solute from the solvent could be performed in one stage as the CG solutes have no charges. To calculate each CG free energy value, 11 simulations with uniformly distributed λ parameters were performed (a test with 10 additional windows showed similar results). Each simulation was started from a pre-equilibrated system where the solute was fully interacting with the solvent. Each simulation was then energy minimized (steepest descent, 2000 steps) and equilibrated in NVT for 1 ns and in NpT for 0.5 ns with the appropriate λ parameter. Production runs of 100 ns were then performed. The simulations were run in rhombic dodecahedron boxes, where the distance between periodic images was 2.6−3.4 nm depending on the system. Soft-core interactions were employed to avoid problems with small λ values, using GROMACS soft-core parameters sc_alpha = 0.5 and sc_power = 1.

The decoupling of the solute from the solvent in the AA simulations was performed in two stages. First, the Coulomb interactions were decoupled, while the LJ interactions were kept fully coupled. Second, the LJ interactions were decoupled when no charges were present in the solute. Soft-core interactions were used in the second stage with the same parameters as in CG simulations to avoid singularities.[62] The first stage was performed with six separate simulations and the second one with 11 simulations, each 20 ns long. For exact comparison with the available experimental data on unmethylated bases in water and octanol and methylated bases in water and chloroform, we created both types of bases using parameters from CHARMM27,[54,55] AMBER94,[63] and a modified version of GROMOS with charges scaled to improve the partitioning of bases to chloroform (kindly provided by Jozica Dolenc). For the chemical structures of the unmethylated and methylated nucleobases, see Figure S1 of the Supporting Information. The topologies were taken from the DNA building blocks, and if necessary, the bases were

neutralized: for AMBER, a hydrogen with the correct charge was added, for CHARMM a hydrogen with the default (0.07) charge was added, and the charge of the attached nitrogen was modified to neutralize the molecule.

PMF Calculations. A number of potentials of mean force (PMFs) were calculated, namely, PMFs of base stacking, base pairing, and base partitioning into a membrane environment. In each case, the PMF profiles were obtained from umbrella sampling simulations.

To obtain the PMFs related to the base stacking, two bases were placed in a box with about 500 water beads. This is large enough so that even at the longest restraint distance the periodic images of the bases were further away than the distance between the two bases in the box. For the CG systems, windows were spaced 0.05 nm apart from 0.25 to 2.20 nm between the centers of mass of the bases. The distance between the bases was restrained with a harmonic umbrella potential with a force constant of 2000 kJ mol^{-1} nm^{-2}. Each window was simulated for 100 ns. The atomistic simulations followed the same methodology with similar box sizes, but the increased number of particles was limited simulation length to 10 ns per window.

The base pairing PMF profiles were obtained by constraining the bases to the same plane using position restraints in the directions perpendicular to the pulling direction. The bases were aligned so that the X-axis passed through the N1 atom of the purines and the N3 atom of the pyrimidines, and the base plane was set as the XY plane. Position restraints with a force constant of 400 kJ mol^{-1} nm^{-2} were used in the Y direction and a 4000 kJ mol^{-1} nm^{-2} force constant in the Z direction. Windows were again spaced 0.05 nm apart between 0.45 and 2.00 nm. For CG systems, each window was simulated for 100 ns, whereas the atomistic systems were simulated for 10 ns.

PMFs of base partitioning into lipid bilayers was explored in a series of over 50 simulations, transporting the different bases along various phospholipid bilayers (1,2-dioleoyl-sn-glycero-3-phosphocholine, DOPC; 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, POPC; 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine, DOPE; and 1,2-dioleoyl-sn-glycero-3-phospho-L-serine, DOPS). For each system, both CG and AA simulations were performed. For the CG simulations, both polarizable and standard Martini water models were tested, and for the AA simulations, both AMBER99sb_parmbsc0[57] and CHARMM36[49] were used with the simulation parameters detailed above. Each system consisted of two nucleobases, a lipid bilayer (64 lipids in each leaflet) and ~3500 CG water particles or ~10 000 AA water molecules. Additionally, all systems were neutralized by adding counterions. The two independent nucleobases were separated by 4 nm, using a harmonic potential force constant of 200 kJ mol^{-1} nm^{-2} on each nucleobase, and simulated in 20 windows, spaced 0.2 nm apart. Each CG window was then simulated for 300 ns, corresponding to an overall simulation time of 6 μs for each system and each AA window for 100−150 ns, for a total of 2−3 μs per PMF profile.

Parameterization of Bonded Terms. Bonded parameters in Martini are usually based on reference AA simulations. The aim is to match the conformations available in Martini as closely as possible to the conformational space of the reference AA model. This means that in Martini the bonded parameters are frequently used to balance out some inaccuracies in CG nonbonded interactions. For DNA, we decided to use CHARMM27[54,55] as the reference AA force field. In our

Table 1. Martini DNA Parameters[a] (bonds, angles, and dihedrals)

beads	type	position of minimum (nm or degrees)	Force constant (kJ mol⁻¹ nm⁻², kJ mol⁻¹, or kJ mol⁻¹ rad⁻²)	beads	type	position of minimum (nm or degrees)	Force constant (kJ mol⁻¹ nm⁻², kJ mol⁻¹, or kJ mol⁻¹ rad⁻²)
BB1-BB2	1	0.360	20 000	BB1-BB2-BB3-CSC1	2	−78.0	25
BB2-BB3	1	0.198	80 000	BB2-BB3-CSC1-CSC2	2	−90.0	20
BB3-BB1	1	0.353	10 000	BB2-BB3-CSC1-CSC3	2	−142.0	50
BB1-BB2-BB3	2	110.0	200	BB3-GSC1	1	0.300	30 000
BB2-BB3-BB1	2	102.0	150	GSC1-GSC2	1	0.295	constraint
BB3-BB1-BB2	2	106.0	75	GSC2-GSC3	1	0.295	constraint
BB1-BB2-BB3-BB1	2	95.0	25	GSC2-GSC4	1	0.389	20 000
BB2-BB3-BB1-BB2	1	180^3	2	GSC3-GSC4	1	0.285	constraint
BB3-BB1-BB2-BB3	9	85.0[b]	2	GSC4-GSC1	1	0.161	constraint
BB3-BB1-BB2-BB3	9	160.0[c]	2	BB2-BB3-GSC1	2	94.5	250
BB3-ASC1	1	0.300	30 000	BB3-GSC1-GSC2	2	137.0	300
ASC1-ASC2	1	0.229	constraint	BB3-GSC1-GSC4	2	130.0	250
ASC2-ASC3	1	0.266	constraint	GSC1-GSC2-GSC3	1	69.5	200
ASC2-ASC4	1	0.326	20 000	GSC1-BB3-BB1	2	157.0	150
ASC3-ASC4	1	0.288	constraint	GSC2-GSC1-GSC4	1	125.0	200
ASC4-ASC1	1	0.162	constraint	GSC2-GSC3-GSC4	1	84.0	200
BB2-BB3-ASC1	2	94.0	250	GSC3-GSC4-GSC1	1	94.0	200
BB3-ASC1-ASC2	2	160.0	200	BB1-BB2-BB3-GSC1	2	−90.0	20
BB3-ASC1-ASC4	2	140.0	200	BB2-BB3-GSC1-GSC2	2	−117.0	1
ASC1-ASC2-ASC3	1	85.0	200	BB2-BB3-GSC1-GSC4	2	92.0	15
ASC1-BB3-BB1	2	158.0	200	BB3-TSC1	1	0.270	30 000
ASC2-ASC1-ASC4	1	125.0	200	TSC1-TSC2	1	0.217	constraint
ASC2-ASC3-ASC4	1	74.0	200	TSC2-TSC3	1	0.322	constraint
ASC3-ASC4-ASC1	1	98.0	200	TSC3-TSC1	1	0.265	constraint
BB1-BB2-BB3-ASC1	2	−90.0	20	BB2-BB3-TSC1	2	92.0	220
BB2-BB3-ASC1-ASC2	2	−116.0	0.5	BB3-TSC1-TSC2	2	107.0	300
BB2-BB3-ASC1-ASC4	2	98.0	15	BB2-TSC1-TSC3	1	145.0	400
BB3-CSC1	1	0.270	30 000	TSC1-BB3-BB1	1	180.0	30
CSC1-CSC2	1	0.220	constraint	TSC1-TSC2-TSC3	1	55.0	100
CSC2-CSC3	1	0.285	constraint	TSC2-TSC1-TSC3	1	83.0	100
CSC3-CSC1	1	0.268	constraint	TSC2-TSC3-TSC1	1	42.0	100
BB2-BB3-CSC1	2	95.0	210	BB1-BB2-BB3-TSC1	2	−75.0	40
BB3-CSC1-CSC2	2	95.0	300	BB2-BB3-TSC1-TSC2	2	−110.0	15
BB2-CSC1-CSC3	1	150.0	500	BB2-BB3-TSC1-TSC3	2	−145.0	65
CSC1-BB3-BB1	1	180.0	30				
CSC1-CSC2-CSC3	1	61.0	200				
CSC2-CSC1-CSC3	1	71.0	200				
CSC2-CSC3-CSC1	1	47.0	200				

[a]In addition to the standard exclusions of bonded neighbors, the second neighbors are also excluded in the backbone, and the base beads are excluded from the backbone beads of the same residue. [b]Multiplicity of the dihedral is 2. [c]Multiplicity of the dihedral is 3.

tests, it gave a more flexible ssDNA than AMBER94.[63] These findings align well with the results of Guy et al.[64] who found CHARMM to sample ssDNA configurations more efficiently than the AMBER force fields. The ssDNA was chosen as the reference state as we did not want to bias our model toward a more ordered DNA state such as dsDNA. We selected 10 different ssDNA sequences, each four bases long, to use as our primary test systems. The sequences used are shown in Table S4 of the Supporting Information. The short length of the strands avoids complications in parameterization that hairpins or other interactions between the ends of the strands would cause. The set of different sequences was used to test whether the distributions of the bonds, angles, and dihedrals of the backbone were sequence dependent at the CG level.

Backmapping. Backmapping of DNA and protein−DNA complexes was performed from CG Martini to CHARMM36[49] all atom representation using *backward*.[65] The backmapping procedure takes an input CG structure and converts it to a target atomistic structure using the provided Martini to CHARMM mapping for each molecule (available at the Martini portal, http://cgmartini.nl/) and the atomistic topology. Using the *backward* scripts, the following procedure was used to relax the atomistic structure, after the initial backmapping step. First, an energy minimization for 500 steps was performed, followed by MD simulations with increasing timesteps of 0.1, 0.2, 0.5, 1.0, 1.5, and 2.0 fs. The LJ and Coulomb cutoffs were set at 0.9 nm. The neighbor list cutoff was set at 0.9 nm. Temperature was scaled using the velocity rescale algorithm and held at 200 K.[40] The resulting structures were further equilibrated for 50 ns using the CHARMM36[49] force field and the same parameters as the CHARMM36 atomistic simulations described above.

■ **RESULTS**

This section describes the parameterization and performance of our Martini DNA model in detail. The mapping of the model was described above in the Model section. The model was parameterized in three steps. First, the nonbonded interactions

were selected based on partitioning behavior of the bases and base−base interactions in water. Then, the bonded interactions of ssDNA were fitted to atomistic simulation data. Last, the elastic network for dsDNA was parameterized. In addition to the parameterization of the model, we describe the behavior of Martini DNA in different environments and compare it to experimental and atomistic simulation data.

Bonded Parameters. The bonded parameters were optimized based on all-atom (AA) reference simulations of short single-stranded DNA (ssDNA). We tested a number of different sequences and determined that the backbone terms were not sequence-specific at the Martini level, and thus, only one set of parameters was developed using the averaged data from all systems. Examples of the differences in distributions from different sequences can be found in Figure S2 of the Supporting Information. The bonded terms were manually adjusted so that the CG distributions cover most of the AA distributions without extending to values not observed in AA simulations. The full list of all bonded terms in the model is given in Table 1, while examples of the corresponding AA and CG distributions are shown in Figure 2. The CG distributions
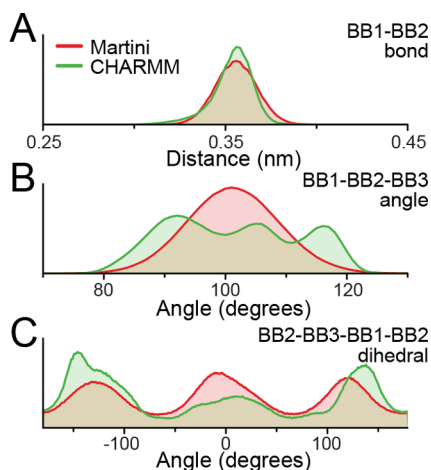
Figure 2. Example distributions of bonded terms from both AA and CG simulations. Distributions from atomistic CHARMM simulations are in green, and CG Martini distributions are in red. The backbone beads are numbered starting from the phosphate (BB1) and then the beads representing the sugar (BB2 and BB3). For a comparison of AA vs CG distributions for all bonded terms, see Figure S3 of the Supporting Information.

for each bonded term are shown together with the target AA distribution in Figure S3 of the Supporting Information. Overall, the distributions are well reproduced, although CG distributions are smoother across the board, as expected.

Partitioning of Bases. The most important test for bead types in Martini is how the building blocks partition between polar and nonpolar environments. To test our models of the nucleotide bases, we calculated the free energies of partitioning from water to hydrated octanol and from water to chloroform. In both cases, we compared the CG free energy values to experimental values as well as results from atomistic simulations, both previously published simulations and additional simulations were performed for this comparison (see Methods section). The partitioning free energies are shown in Figure 3 as well as in Table S3 of the Supporting Information. Figure 3 shows how the CG values have an absolute error no larger than the atomistic force fields and furthermore follow the
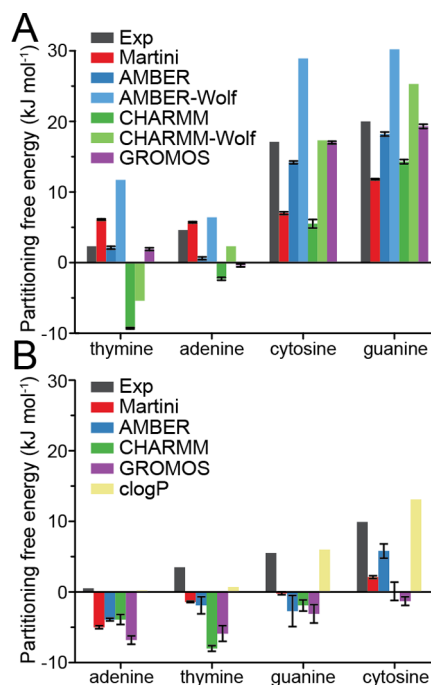
Figure 3. Partitioning free energies of the nucleobases. (A) Shows the water−chloroform partitioning values for each tested force field together with the experimental values. (B) Shows the water−octanol partitioning free energies. In each figure, the bases are sorted according to the experimental partitioning free energies. AMBER-Wolf and CHARMM-Wolf results are from Wolf et al.[56] and use different solvent models than we did. Experimental results are from refs 76 and 77, while the clogP data is from the ACD/Labs algorithm in Chemspider (www.chemspider.com, retrieved 15 January 2015). The raw data is presented in Table S3 of the Supporting Information.

trend of the experimental results well. Here, it is worth noting the large differences between the experimental results and all the computational results as well as the large differences between values obtained from different atomistic force fields. We also show calculated logP (clogP) values obtained from the ACD/Labs logP prediction algorithm.[66] These correlate well with the experimental values; this is consistent with the values being based on experimental logP values of similar molecular fragments. In the case of chloroform, the differences between our results for AMBER and CHARMM and the published results (AMBER-Wolf and CHARMM-Wolf) are due to different parameters used for chloroform (see Methods section). Overall, in comparison to the other computational models, Martini DNA fares well by following the experimental order of partitioning free energies both in water−octanol and water−chloroform. Guanine and cytosine are more hydrophilic than adenine and thymine in both cases.

Membrane Partitioning. To further explore the partitioning of the nucleotide bases between polar and nonpolar environments, we calculated the free energy profiles of moving the nucleobases into lipid bilayers. Figure 4 shows a snapshot and the PMF profiles of moving the nucleobases into a DOPC bilayer along an axis perpendicular to the bilayer, in both AA and CG simulations. PMFs of the nucleobases partitioning into POPC, DOPE, and DOPS bilayers are shown in Figure S4 of the Supporting Information. The general features of the AA and CG PMF profiles match very well for all systems. In both sets of simulations, each PMF profile shows a clear free energy barrier at the bilayer center, which results from pulling polar
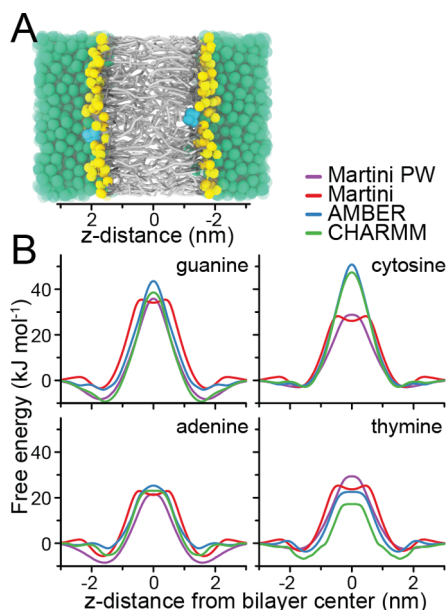
Figure 4. Partitioning of the nucleobases into bilayers. (A) Simulation snapshot showing two nucleobases pulled through a lipid bilayer in CG simulations. The nucleobases are shown in blue, water in green, lipid headgroups in yellow, and lipid tails in gray. (B) PMF profiles of the nucleobases partitioning through a DOPC bilayer, shown for CHARMM, AMBER, and Martini with both standard and polarizable water (green, blue, red, and purple lines, respectively). The results of nucleobases partitioning into POPC, DOPE, and DOPS bilayers are presented in Figure S4 of the Supporting Information.

nucleobases into the hydrophobic tail region. The PMF value decreases rapidly when moving away from the bilayer center, with a similar slope in both AA and CG simulations. A minimum in the free energy profile is observed where the hydrophobic lipid tails end and the hydrophilic lipid polar headgroups begin, consistent with the amphipathic nature of nucleobases. At this location, the apolar aromatic rings of nucleobases interact with hydrophobic lipid tails, while nucleobase polar groups interact with lipid headgroups. Further out, a small barrier is sometimes observed when the nucleobases move past the polar/charged lipid headgroups and then the profile flattens out when nucleobases reach bulk water.

The AMBER and CHARMM AA force fields produce similar profiles in general, although the exact values for the energy barriers differ somewhat. In AA simulations, guanine and cytosine generally show the highest and adenine and thymine the lowest energy barriers at the center of the bilayer, in general agreement with the water−octanol and water−chloroform simulations, as well as previous computational and experimental results (see above). The CG simulations follow the same trend, although for cytosine the central barrier is lower than in the AA simulations. In the AA cytosine and guanine simulations, water defects form and keep the nucleobases solvated, but these are not observed at the CG level. Due to the large granularity of CG simulations, there is no water structure in these kinds of defects, and the corresponding missing entropic contribution makes the defects either disproportionately expensive or nonexistent,[67,68] which can contribute to some of the observed differences in the PMFs. In the bilayer interface region, AA simulations prefer guanine and adenine and then thymine to interact with the phospholipid, which is well reproduced in CG

simulations. The CG PMF profiles using the standard (nonpolarizable) water model show a reasonable agreement with both AA and the CG polarizable water. They reproduce well the main energetic features for nucleobase transport along the lipid bilayer as discussed above (see Figure 4 and Figure S4 of the Supporting Information). Overall, considering that atomistic solvation free energy calculations as well as partitioning free energies between water and cyclohexane[69] have associated errors of up to 8 kJ mol^{-1}, our CG free energy calculations reproduce the AA results quite reasonably.

Base Stacking. The partitioning simulations describe how the bases interact with their environment. Next, we tested how they interact with each other in water. A PMF profile of a free cytosine and a guanine in water is shown in Figure 5A; PMF
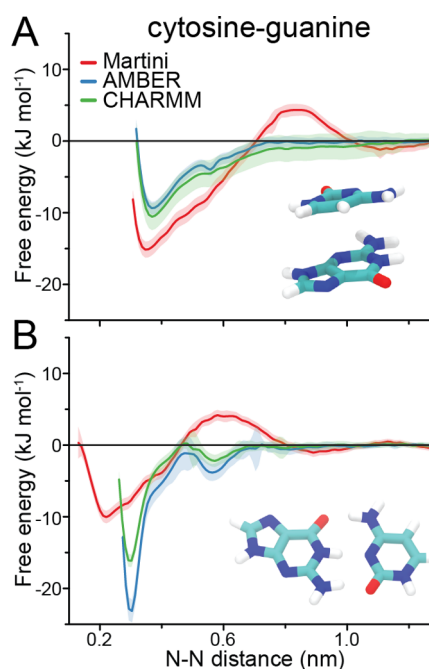
Figure 5. DNA base−base interactions. (A) PMF of two bases (cytosine and guanine) interacting without additional constraints in water (only the distance between the bases in each pulling window is constrained). This probes mostly base−base stacking (see insert). (B) PMF of cytosine and guanine hydrogen bonding while restricted onto a plane (the bases are constrained in the plane and then pulled, see insert). PMFs are shown for CHARMM, AMBER, and Martini (green, blue, and red, respectively). The data for other base pairs are presented in Figure S5 (free/stacking simulations) and Figure S6 (hydrogen bonding simulations) of the Supporting Information.

profiles for all other nucleobase combinations are provided in Figure S5 of the Supporting Information. The AA and CG profiles match fairly well, showing the free energy minimum to be at a distance of 0.34 nm where the bases are stacked for all three force fields (Martini, CHARMM, and AMBER). The main difference is the higher free energy barrier in the CG profile around 0.8 nm, arising from the desolvation of the base−base interaction by the relatively large size of CG water beads compared to atomistic water molecules.

Base−Base Hydrogen Bonding. The PMF profiles of two free bases in water show mainly the effect of stacking. The orientation where two bases can form hydrogen bonds between each other is fairly restricted and thus does not show up clearly in these PMF profiles. Since this is one of the most important

features of DNA hybridization, we study it by constricting the bases on a plane and calculating the PMF of moving them along the axis toward each other. Figure 5B shows an example profile, while profiles for all the other nucleobase combinations are shown in Figure S6 of the Supporting Information. The parameterization of the special class of hydrogen bonding beads (see above) focused on getting the relative preferences between different base pairs correct. However, the spherically symmetric nature of the potentials that Martini uses means that any strong interaction between the bases in hydrogen bonding would also affect their stacking. Thus, the shallower minima observed in the hydrogen-bonding PMFs (Figure 5B and Figure S6, Supporting Information) are necessary to better match the base stacking PMFs discussed above. The free energy minima of the CG PMF profiles are shifted relative to the AA profiles in Figure 5B. This is inevitable since the CG bead size is selected based on the distance between two stacking bases and the increased hydrogen bonding distance cannot be compensated for in the size of the beads. The CG beads are mapped to the centers of mass of the underlying atoms and thus the beads of one nucleobase are further away from the other base than the hydrogen bonding atoms are. Table 2 summarizes the Martini DNA base−base pairing strengths. Importantly, Martini reproduces the order of corresponding atomistic results.

Table 2. Nucleobase Hydrogen Bonding Strength

base pair	AMBER ΔG (kJ mol^{-1})	CHARMM ΔG (kJ mol^{-1})	Martini ΔG (kJ mol^{-1})
C-G	-22.6 ± 1.6	-16.5 ± 1.3	-10.0 ± 1.0
A-T	-15.8 ± 0.8	-13.6 ± 0.2	-9.8 ± 1.5
C-T	-11.3 ± 1.2	-11.8 ± 0.3	-9.2 ± 0.8
T-T	-9.4 ± 0.6	-6.5 ± 0.2	-8.7 ± 1.0
A-G	-8.8 ± 1.0	-9.6 ± 0.7	-8.3 ± 0.8
G-T	-7.8 ± 1.0	-5.2 ± 0.2	-6.2 ± 1.0
A-A	-4.6 ± 0.6	-5.8 ± 0.7	-5.6 ± 1.0
A-C	-3.2 ± 0.8	-2.6 ± 0.8	-5.1 ± 1.0
C-C	-1.6 ± 1.5	-1.0 ± 1.2	-5.0 ± 1.0
G-G	3.7 ± 1.7	5.4 ± 0.3	-4.8 ± 0.5

Single-Stranded DNA. We tested a number of ssDNAs, differing in sequence, length, and the ionic strength of the solution. Figure 7 shows the radius of gyration (R_g) of ssDNA with three different ion concentrations together with experimental results as well as atomistic simulations. The Martini DNA agrees well with the CHARMM force field that was used in parameterization of the bonded terms, especially at higher ion concentrations. The AMBER force field seems stiffer than CHARMM, but the differences are fairly minor and there are questions whether its stiffness is realistic.[64] The recent experimental results,[70] also shown in Figure 7, predict larger R_g values and much straighter ssDNA strands than any of the computational results show, even at very short strand lengths. We think the general discrepancy is due to different assumptions that are made in measuring radii of gyrations experimentally and thus should not be directly compared. Computationally, R_g values similar to the experimental values for the shortest strands would only be obtained for a single strand that is about as extended/rigid as DNA in double helical form.

Correspondence between both atomistic force fields and Martini is good, and with higher ion concentrations, they are very similar, but at low ion concentrations, Martini is more
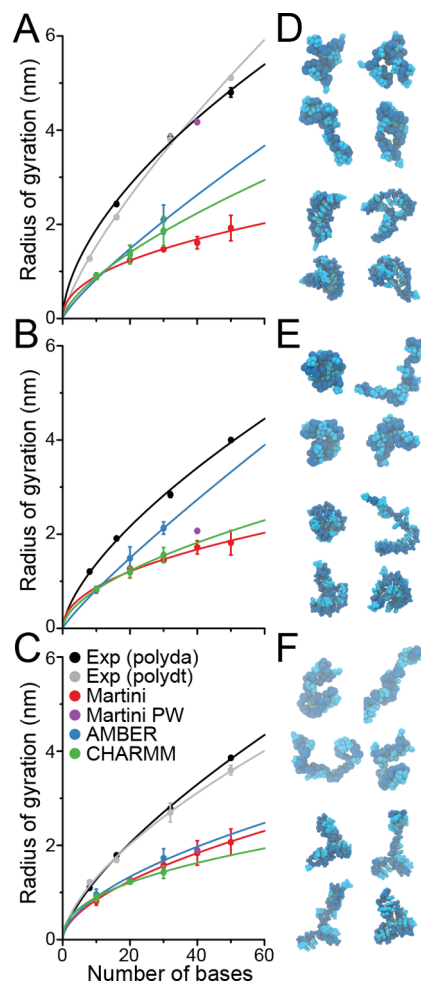
Figure 6. Flexibility of ssDNA. The radius of gyration of ssDNA segments as a function of strand length in a solution. Results are shown for CHARMM, AMBER, and Martini with both standard and polarizable water (green, blue, red, and purple, respectively). All simulations were run for a random sequence of ssDNA. The lines are a fit of the measurements to $R_g = AN^\nu$, where N is the number of monomers in the DNA, and the rest are fitting parameters. (A) Shows results in solution with counterions only, (B) with 100 mM NaCl ,and (C) with 1 M NaCl solution. Experimental data (black and gray) is from Sim et al.[70] with a salt concentration of 12.5 mM in (A), 125 mM in (B), and 1025 mM in (C). (D−F) Snapshots of the 20 base long ssDNAs from CG (top) and AA (bottom) simulations in a solution with only counterions (D), 100 mM NaCl (E), and 1 M NaCl (F), demonstrating the versatile structures found in all the simulations. The DNA backbone is colored in blue and nucleobases in cyan.

compact than the atomistic force fields. We think this is due to the implicit screening of the charges in Martini. To illustrate this, we simulated the 40 base long ssDNA with each ion concentration using the polarizable Martini water (PW) model. With the polarizable water model, the effect of ion concentration is clearly visible. This suggests that Martini ssDNA with the standard water model best describes systems with moderate to high salt concentrations.

Double-Stranded DNA. Opposing ssDNAs do not hybridize spontaneously, and prehybridized DNA structures are not stable when simulated using Martini DNA. This is likely due to the limited strength of interactions between the DNA strands as well as the large size of the solvent beads. For simulations of dsDNA, we resorted to a similar solution as with Martini
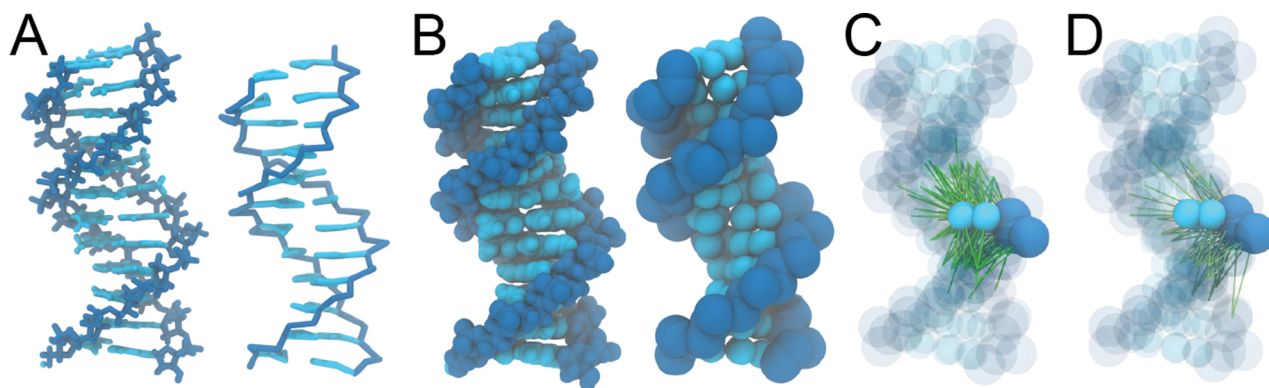
Figure 7. dsDNA mapping and elastic network for a dsDNA strand (PDB ID: 1BNA). (A) Wire-frame structure of AA (left) and CG (right) models. (B) Presents the same models using a space-filling representation. (C) Shows all the elastic bonds connected to a single residue in the stiff elastic network model. (D) Shows the elastic bonds for the soft model of dsDNA. The backbone is colored in blue, the nucleobases in cyan, and the elastic bonds in green.

proteins where an elastic network is used to preserve the secondary structure.[71] For stable and accurate simulation of dsDNA, we developed an elastic network to be used with the bonded terms that are based on ssDNA simulations. We parameterized two separate elastic network schemes for dsDNA. The first restrains the whole structure, is stable with larger simulation timesteps, and retains the initial structure closely. The second is more flexible and does not hinder rotations of the bases, enabling more realistic movement of and interactions with the bases. We distinguish these two in this manuscript by calling them stiff and soft elastic networks. Both networks can be built using a modified version of the *martinize* script[26] called *martinize-dna*. The different kinds of elastic networks are illustrated in Figure 7. The stiff elastic network was designed to closely retain initial structures, including unusual ones (such as DNA in a protein complex, G-quadruplexes, or large-scale structures like those arising in DNA-origami), whereas the flexible elastic network was a compromise between realistic flexibility of the DNA and numerical stability.

To calculate the persistence length, we used the same definition as Ouldridge et al.[13] The persistence length is determined by calculating the average angle between vectors that connect consecutive bases in dsDNA and fitting an exponential decay to the data:

$$\langle n_i \cdot n_0 \rangle = e^{-i \langle l_0 \rangle / L}$$

where n_i is the ith vector, $\langle l_0 \rangle$ is the time averaged length of the first vector (used here as an estimate for the distance between bases), and L is the persistence length. The estimate for persistence length is obtained by placing the n_0 in 10 separate positions and fitting the equation to the averaged data. The errors are estimated as the standard deviation after doing block averaging with five blocks on the data. Persistence length was determined from simulations of 100 bp random sequence dsDNA, where the elastic network was created on the ideal B-DNA starting conformation. Simulations with counterions, 100 mM, and 1 M NaCl solution were each run for 500 ns for the stiff model and 3000 ns for the soft model, with the first 100 ns treated as equilibration and excluded from analysis. Figure 8 shows the fit for 100 mM NaCl simulations. The reason for the shortest distances not fully reaching a correlation of one are small errors in defining the centers of the dsDNA base pairs corresponding to the vector end points. The difference between
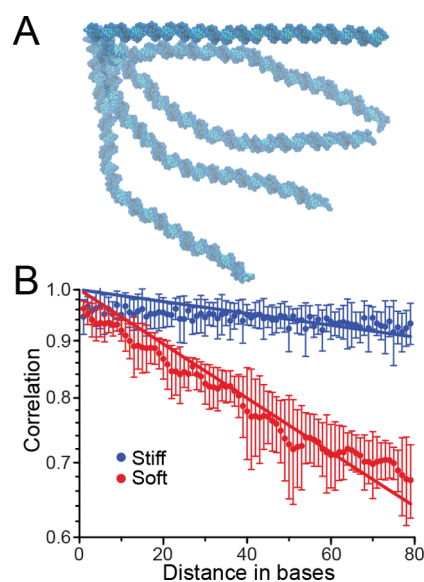

Figure 8. Persistence length of dsDNA. (A) Snapshots from a simulation with the soft elastic network showing the bending range of a 100 bp long dsDNA. (B) Correlation plot used to determine the persistence length. The average angle between vectors $\langle n_x \cdot n_0 \rangle$, each of which connects centers of two consecutive bases, is plotted on the y-axis. The distance x (in bases) of these two vectors is plotted on the x-axis. Fit is shown for both elastic network models.

the soft and stiff elastic networks is evident in the calculated persistence lengths of these two models; the former has a persistence length of 66 ± 8 nm, whereas the latter has a persistence length of 280 ± 60 nm when only counterions are present. The values do not depend significantly on the ion concentration; in 100 mM NaCl solution, the persistence lengths are 68 ± 14 nm and 270 ± 60 nm, respectively, and in 1 M NaCl, 78 ± 11 nm and 280 ± 70 nm. In comparison, the generally accepted experimental value of dsDNA persistence length is about 50 nm[72] in high salt concentrations. Additional softening of the elastic network to decrease the persistence length resulted in numerically unstable simulations.

To further examine the properties of the dsDNA model, we calculated the ion distribution around a dsDNA molecule and compared that to the ion distribution obtained from atomistic simulations. These distributions were calculated in solutions with only counterions as well as with ion concentrations of 100

mM and 1 M NaCl. These tests were performed with standard Martini water as well as polarizable Martini water with PME. The ion distributions around dsDNA with a 100 mM NaCl solution are presented in Figure 9. Results for the other
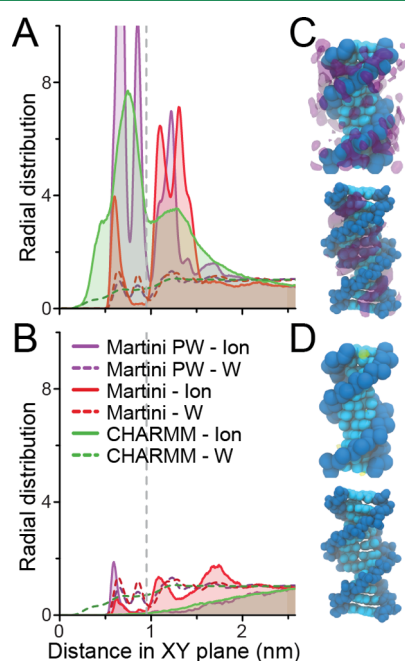
Figure 9. dsDNA ion distribution. (A) Shows the distributions of sodium as a function of distance from the DNA center in the *XY* plane for CHARMM (green), Martini (red), and Martini with polarizable water and PME (purple). (B) Presents the distributions of chloride with these force fields. In each case, the corresponding water distribution is shown for comparison (dotted lines). (C) Shows the volume maps of locations most commonly occupied by sodium in Martini (top) and CHARMM (bottom) simulations. (D) Shows the most commonly occupied locations of chloride. In the case of CHARMM (bottom), chloride does not localize around DNA but is present in Martini in low amounts (top). The figures for Martini are from simulations using the standard Martini water model, and all simulations have a 100 mM NaCl solution. Ion distribution from simulations with counterions only and 1 M NaCl are shown in Figure S7 of the Supporting Information.

conditions tested are shown in Figure S7 of the Supporting Information. In atomistic simulations, the sodium ions condensate on the grooves of the DNA, and chloride is completely excluded from the DNA. In the CG simulations with standard Martini water, sodium condenses both in the grooves and near the phosphates, likely due to the size of CG beads preventing more ions packing into the grooves. In the CG simulations, chloride is also present in small amounts near the DNA, which could be due to the implicit screening and short-ranged cutoff of electrostatics. Using the polarizable water model, the ions are somewhat more ordered around the DNA, and thus, the peaks in the radial distribution function are sharper.

We also tested four 10−12 base pairs long dsDNA oligomers (PDB ID: 1BNA, 424D, 109D, 158D). In Figure 10A, 1BNA is shown atomistically, coarse-grained, and after mapping the CG structure back to atomistic resolution. The soft elastic DNA network mostly affects only the DNA backbone and is flexible enough to allow for a reasonable persistence length. The overall structure of these oligomers is very stable, with a root-mean-
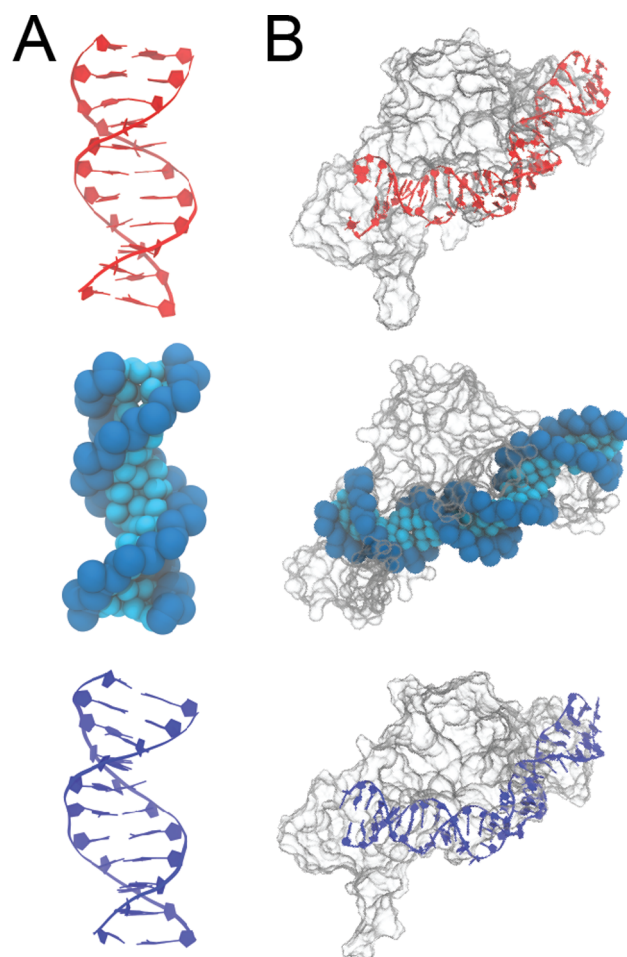
Figure 10. DNA stability. (A) Snapshots of a dsDNA (PDB ID: 1BNA) are shown at the atomistic resolution (top), after 100 ns of CG Martini simulation (middle), and after backmapping the final frame of the CG simulation to atomistic and simulating for 50 ns (bottom). (B) Snapshots of the DNA−protein complex (PDB ID: 1MNM) are shown at the atomistic resolution (top), after 100 ns of CG Martini simulation (middle), and after backmapping the final frame of the CG simulation to atomistic and simulating for 50 ns (bottom).

square deviation (RMSD) of 0.2−0.3 nm with the greatest flexibility in the terminal nucleobases (Table S5, Supporting Information).

DNA−Protein Complexes. In order to assess the compatibility of the Martini DNA force field to other biomolecules, we simulated several DNA−protein complexes. For each system, the initial crystal structure was coarse-grained using the *martinize* script.[26] The soft elastic network was used for the DNA and the ElNeDyn network for the proteins.[71] The systems were then equilibrated and simulated for 100 ns. We tested three DNA−protein complexes: the lambda repressor−operator complex, the sporulation-specific transcription factor Ndt80, and the yeast MATalpha2/MCM1/DNA ternary complex (PDB IDs: 1LMB, 1MNN, and 1MNM, respectively). All the proteins remained firmly bound to their respective DNA and the DNA−protein complexes showed low RMSD of around 0.1−0.2 nm (Table S5, Supporting Information). Figure 10B demonstrates this for the 1MNM structure, showing the initial atomistic and final CG structures.

Backmapping to Atomistic Resolution. CG simulations can reach time and length scales inaccessible to AA simulations

but are not appropriate for all problems. When higher resolution/atomistic insight is needed, it is convenient to be able to map CG configurations onto AA coordinates. We demonstrate the feasibility of backmapping Martini DNA considering the final snapshots of the CG Martini DNA (1BNA) and DNA−protein complex (1MNM) simulations, as shown in Figure 10. The DNA and DNA−protein complexes were mapped to atomistic coordinates using *backward*[65] (see Methods) and simulated with CHARMM for 50 ns.

DISCUSSION

We have described the parameterization and performance of a coarse-grained Martini DNA model. With respect to the bases, the behavior of Martini is comparable, and in some cases better, than most current atomistic DNA force fields in reproducing the experimental partitioning free energies. Free bases in solution have similar stacking free energies at the CG level as at atomistic level. The relative order of the strength of hydrogen bonding between bases is retained compared to atomistic force fields even though the absolute values do not match. The partitioning profiles of the bases across lipid membranes are also comparable to those obtained at the AA level.

The Martini parameterization strategy that is used to obtain the described DNA model is of course dependent on the target data used in the parameterization. The selection of the atomistic force field used in conjunction with experimental data affects the parameters and performance of the model. In this manuscript, we have shown that the nonbonded interactions derived from atomistic simulation data are not sensitive to the choice of force field, at least between the tested CHARMM and AMBER force fields. The bonded parameters are much more sensitive to this choice, as shown from the ssDNA radius of gyration results and the partitioning free energies. Properties of dsDNA, however, are less sensitive to the reference AA force field selection since they are largely determined by the elastic network. In principle, the parameterization method described in the paper is fully transferable to another reference force field, and future versions of the Martini DNA model could be reparameterized using improved atomistic models.[57,73−75]

Concerning DNA strands, our model reproduces the radius of gyration of ssDNA from atomistic simulations well, in particular at moderate to high ion concentrations. At low ionic strength, the behavior is improved using the polarizable Martini water model. For the dsDNA model, two elastic network levels are supported: a stiff network that maintains an arbitrary initial structure closely and a soft network that is more flexible and allows for movement of individual bases and shows a realistic persistence length. The best choice of network will depend on the particular need in the application. For dsDNA, the ion distribution of Na^+ is reproduced reasonably well, but in contrast with the AA simulations, the CG model shows a small accumulation of Cl^- in the DNA grooves.

The main current limitation of the model is the reduced strength of the base-pairing interactions (Figure 5B) due to the inability of Martini to model directional hydrogen bonds. Consequently, Martini DNA cannot be used to study DNA hybridization, melting, hairpin formation, or intercalation. Furthermore, the explicit solvent limits some applications that require length and time scales that can be achieved only with even faster implicit solvent models that have drastically fewer degrees of freedom. We should stress that the limitations of the model should also direct users into what kind of applications

the model is designed for. We do not recommend using Martini DNA for structural studies of DNA; the elastic network on the dsDNA and the limited accuracy of the model make it an imperfect choice for such applications. Instead, we foresee the model to be used in intermolecular applications where DNA plays its own role in a complex environment composed of other biomolecules such as DNA−DNA interactions (e.g., DNA winding/unwinding in the nucleosome or packing of DNA in viral capsids), DNA−protein interactions (e.g., protein−DNA transcription and repair complexes or histone-mediated DNA packing), and DNA−lipid interactions (e.g., gene delivery vehicles such as lipoplexes or controlled vesicle fusion mediated by DNA strands) or in nanotechnology applications (e.g., DNA transport through nanopores, DNA wrapping around nanotubes, or large DNA origami and pyramid structures).

In summary, we have developed a new coarse-grain DNA model based on the Martini force field. The speed and compatibility of the model opens the way to perform large-scale modeling of complex biomolecular systems involving DNA. The DNA model will furthermore serve as a basis for the development of a Martini RNA model. All the parameters needed to use the Martini DNA model, as well as a tutorial on how to build and simulate a DNA system using Martini, are available at the Martini portal http://cgmartini.nl.

ASSOCIATED CONTENT

Ⓢ Supporting Information

Figures for systems not shown in the manuscript, values for partitioning free energies that are presented as a figure in the main manuscript, the interaction matrix for the special hydrogen bonding beads, DNA bead mapping, and RMSD values for DNA and DNA-protein stability. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00286.

AUTHOR INFORMATION

Corresponding Author
*E-mail: s.j.marrink@rug.nl.

Notes
The authors declare no competing financial interest.

■ ABBREVIATIONS

AA, all-atom; CG, coarse-grained; dsDNA, double-stranded DNA; ssDNA, single-stranded DNA; PMF, potential of mean force

■ REFERENCES

(1) Ingólfsson, H. I.; López, C. A.; Uusitalo, J. J.; De Jong, D. H.; Gopal, S. M.; Periole, X.; Marrink, S.-J. The Power of Coarse Graining in Biomolecular Simulations. *WIREs Comput. Mol. Sci.* **2014**, *4*, 225−248.

(2) Noid, W. G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, *139*, 090901.

(3) Hyeon, C.; Thirumalai, D. Capturing the Essence of Folding and Functions of Biomolecules Using Coarse-Grained Models. *Nat. Commun.* **2011**, *2*, 487.

(4) Maffeo, C.; Yoo, J.; Comer, J.; Wells, D. B.; Luan, B.; Aksimentiev, A. Close Encounters with DNA. *J. Phys.: Condens. Matter* **2014**, *26*, 413101.

(5) Kara, M.; Zacharias, M. Theoretical Studies of Nucleic Acids Folding. *WIREs Comput. Mol. Sci.* **2014**, *4*, 116−126.

(6) Tan, R. K. Z.; Harvey, S. C. Molecular Mechanics Model of Supercoiled DNA. *J. Mol. Biol.* **1989**, *205*, 573−591.

(7) Schlick, T.; Olson, W. K. Supercoiled DNA Energetics and Dynamics by Computer Simulation. *J. Mol. Biol.* **1992**, *223*, 1089−1119.

(8) Drukker, K.; Wu, G.; Schatz, G. C. Model Simulations of DNA Denaturation Dynamics. *J. Chem. Phys.* **2001**, *114*, 579.

(9) Savelyev, A.; Papoian, G. A. Chemically Accurate Coarse Graining of Double-Stranded DNA. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20340−20345.

(10) Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. A Coarse Grain Model for DNA. *J. Chem. Phys.* **2007**, *126*, 084901.

(11) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J. An Experimentally-Informed Coarse-Grained 3-Site-Per-Nucleotide Model of DNA: Structure, Thermodynamics, and Dynamics of Hybridization. *J. Chem. Phys.* **2013**, *139*, 144903.

(12) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. DNA Nanotweezers Studied with a Coarse-Grained Model of DNA. *Phys. Rev. Lett.* **2010**, *104*, 178101.

(13) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. Structural, Mechanical, and Thermodynamic Properties of a Coarse-Grained DNA Model. *J. Chem. Phys.* **2011**, *134*, 085101.

(14) Linak, M. C.; Dorfman, K. D. Analysis of a DNA Simulation Model Through Hairpin Melting Experiments. *J. Chem. Phys.* **2010**, *133*, 125101.

(15) Maciejczyk, M.; Spasic, A.; Liwo, A.; Scheraga, H. A. DNA Duplex Formation with a Coarse-Grained Model. *J. Chem. Theory Comput.* **2014**, *10*, 5020−5035.

(16) Maffeo, C.; Ngo, T. T.; Ha, T.; Aksimentiev, A. A Coarse-Grained Model of Unstretched Single-Stranded DNA Derived From Atomistic Simulation and Single-Molecule Experiment. *J. Chem. Theory Comput.* **2014**, *10*, 2891−2896.

(17) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. A Coarse Grained Model for Atomic-Detailed DNA Simulations with Explicit Electrostatics. *J. Chem. Theory Comput.* **2010**, *6*, 1711−1725.

(18) Cragnolini, T.; Derreumaux, P.; Pasquali, S. Coarse-Grained Simulations of RNA and DNA Duplexes. *J. Phys. Chem. B* **2013**, *117*, 8047−8060.

(19) Gopal, S. M.; Mukherjee, S.; Cheng, Y.-M.; Feig, M. PRIMO/PRIMONA: a Coarse-Grained Model for Proteins and Nucleic Acids That Preserves Near-Atomistic Accuracy. *Proteins: Struct., Funct., Genet.* **2010**, *78*, 1266−1281.

(20) Poulain, P.; Saladin, A.; Hartmann, B.; Prévost, C. Insights on Protein-DNA Recognition by Coarse Grain Modelling. *J. Comput. Chem.* **2008**, *29*, 2582−2592.

(21) Liwo, A.; Baranowski, M.; Czaplewski, C.; Gołaś, E.; He, Y.; Jagieła, D.; Krupa, P.; Maciejczyk, M.; Makowski, M.; Mozolewska, M. A.; Niadzvedtski, A.; Ołdziej, S.; Scheraga, H. A.; Sieradzan, A. K.; Slusarz, R.; Wirecki, T.; Yin, Y.; Zaborowski, B. A Unified Coarse-Grained Model of Biological Macromolecules Based on Mean-Field Multipole-Multipole Interactions. *J. Mol. Model.* **2014**, *20*, 2306.

(22) Marrink, S.-J.; de Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108*, 750−760.

(23) Marrink, S.-J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111*, 7812−7824.

(24) Marrink, S.-J.; Tieleman, D. P. Perspective on the Martini Model. *Chem. Soc. Rev.* **2013**, *42*, 6801−6822.

(25) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4*, 819−834.

(26) De Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S.-J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9*, 687−697.

(27) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Hünenberger, P. H.; Marrink, S.-J. Martini Coarse-Grained Force Field: Extension to Carbohydrates. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210.

(28) Lee, H.; de Vries, A. H.; Marrink, S.-J.; Pastor, R. W. A Coarse-Grained Model for Polyethylene Oxide and Polyethylene Glycol: Conformation and Hydrodynamics. *J. Phys. Chem. B* **2009**, *113*, 13186−13194.

(29) Rossi, G.; Monticelli, L.; Puisto, S. R.; Vattulainen, I.; Ala-Nissila, T. Coarse-Graining Polymers with the MARTINI Force-Field: Polystyrene as a Benchmark Case. *Soft Matter* **2011**, *7*, 698−708.

(30) Yesylevskyy, S. O.; Schäfer, L. V.; Sengupta, D.; Marrink, S.-J. Polarizable Water Model for the Coarse-Grained MARTINI Force Field. *PLoS Comput. Biol.* **2010**, *6*, e1000810.

(31) Khalid, S.; Bond, P. J.; Holyoake, J.; Hawtin, R. W.; Sansom, M. S. P. DNA and Lipid Bilayers: Self-Assembly and Insertion. *J. R. Soc., Interface* **2008**, *5*, 241−250.

(32) Corsi, J.; Hawtin, R. W.; Ces, O.; Attard, G. S.; Khalid, S. DNA Lipoplexes: Formation of the Inverse Hexagonal Phase Observed by Coarse-Grained Molecular Dynamics Simulation. *Langmuir* **2010**, *26*, 12119−12125.

(33) Stachiewicz, A.; Molski, A. A Coarse-Grained MARTINI-Like Force Field for DNA Unzipping in Nanopores. *J. Comput. Chem.* **2015**, *36*, 947−956.

(34) Leung, A. K. K.; Hafez, I. M.; Baoukina, S.; Belliveau, N. M.; Zhigaltsev, I. V.; Afshinmanesh, E.; Tieleman, D. P.; Hansen, C. L.; Hope, M. J.; Cullis, P. R. Lipid Nanoparticles Containing siRNA Synthesized by Microfluidic Mixing Exhibit an Electron-Dense Nanostructured Core. *J. Phys. Chem. C* **2012**, *116*, 18440−18450.

(35) Rozmanov, D.; Baoukina, S.; Tieleman, D. P. Density Based Visualization for Molecular Simulation. *Faraday Discuss.* **2014**, *169*, 225−243.

(36) Kim, M.; Kim, H. R.; Chae, S. Y.; Larson, R. G.; Lee, H.; Park, J. C. Effect of Arginine-Rich Peptide Length on the Structure and Binding Strength of siRNA−Peptide Complexes. *J. Phys. Chem. B* **2013**, *117*, 6917−6926.

(37) Pronk, S.; Pall, S.; Schulz, R.; Larsson, P.; Bjelkmar, P.; Apostolov, R.; Shirts, M. R.; Smith, J. C.; Kasson, P. M.; van der Spoel, D.; Hess, B.; Lindahl, E. GROMACS 4.5: a High-Throughput and Highly Parallel Open Source Molecular Simulation Toolkit. *Bioinformatics* **2013**, *29*, 845−854.

(38) Darden, T. A.; York, D.; Pedersen, L. Particle Mesh Ewald: an N· Log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089.

(39) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T. A.; Lee, H.; Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *J. Chem. Phys.* **1995**, *103*, 8577.

(40) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.

(41) Parrinello, M.; Rahman, A. Polymorphic Transitions in Single Crystals: a New Molecular Dynamics Method. *J. Appl. Phys.* **1981**, *52*, 7182−7190.

(42) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. Molecular Dynamics with Coupling to an External Bath. *J. Chem. Phys.* **1984**, *81*, 3684.

(43) Lavery, R.; Zakrzewska, K.; Beveridge, D. L.; Bishop, T. C.; Case, D. A.; Cheatham, T. E., III; Dixit, S.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Pérez, A.; Singh, T.; Spackova, N.; Sponer, J. A Systematic Molecular Dynamics Study of Nearest-Neighbor Effects on Base Pair and Base Pair Step Conformations and Fluctuations in B-DNA. *Nucleic Acids Res.* **2010**, *38*, 299−313.

(44) Bjelkmar, P.; Larsson, P.; Cuendet, M. A.; Hess, B.; Lindahl, E. Implementation of the CHARMM Force Field in GROMACS: Analysis of Protein Stability Effects From Correction Maps, Virtual Interaction Sites, and Water Models. *J. Chem. Theory Comput.* **2010**, *6*, 459−466.

(45) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. A Biomolecular Force Field Based on the Free Enthalpy of Hydration and Solvation: the GROMOS Force-Field Parameter Sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, *25*, 1656−1676.

(46) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926.

(47) Berendsen, H. J. C.; Postma, J.; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands, 1981; Vol. *14*, pp 331−342.

(48) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(49) Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell, A. D.; Pastor, R. W. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *J. Phys. Chem. B* **2010**, *114*, 7830−7843.

(50) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(51) Caleman, C.; van Maaren, P. J.; Hong, M.; Hub, J. S.; Costa, L. T.; van der Spoel, D. Force Field Benchmark of Organic Liquids: Density, Enthalpy of Vaporization, Heat Capacities, Surface Tension, Isothermal Compressibility, Volumetric Expansion Coefficient, and Dielectric Constant. *J. Chem. Theory Comput.* **2012**, *8*, 61−74.

(52) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, A. D. CHARMM General Force Field: a Force Field for Drug-Like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2009**, *31*, 671−690.

(53) Casteblanco, F. F. *Modelling and Predicting the Free Energy and Solubility Behavior of Lovastatin and Simvastatin Using Molecular Simulations*; Rutgers University, 2013; pp 1−82.

(54) Foloppe, N.; MacKerell, A. D., Jr. All-Atom Empirical Force Field for Nucleic Acids: I. Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data. *J. Comput. Chem.* **2000**, *21*, 86−104.

(55) MacKerell, A. D.; Banavali, N. All-Atom Empirical Force Field for Nucleic Acids: II. Application to Molecular Dynamics Simulations of DNA and RNA in Solution. *J. Comput. Chem.* **2000**, *21*, 105−120.

(56) Wolf, M. G.; Groenhof, G. Evaluating Nonpolarizable Nucleic Acid Force Fields: a Systematic Comparison of the Nucleobases Hydration Free Energies and Chloroform-to-Water Partition Coefficients. *J. Comput. Chem.* **2012**, *33*, 2225−2232.

(57) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of Conformers. *Biophys. J.* **2007**, *92*, 3817−3829.

(58) Jämbeck, J. P. M.; Lyubartsev, A. P. Derivation and Systematic Validation of a Refined All-Atom Force Field for Phosphatidylcholine Lipids. *J. Phys. Chem. B* **2012**, *116*, 3164−3179.

(59) Jämbeck, J. P. M.; Lyubartsev, A. P. An Extension and Further Validation of an All-Atomistic Force Field for Biological Membranes. *J. Chem. Theory Comput.* **2012**, *8*, 2938−2948.

(60) Bernazzani, L.; Cabani, S.; Conti, G.; Mollica, V. Thermodynamic Study of the Partitioning of Organic Compounds Between Water and Octan-1-Ol. Effects of Water as Cosolvent in the Organic Phase. *J. Chem. Soc., Faraday Trans.* **1995**, *91*, 649−655.

(61) Bennett, C. H. Efficient Estimation of Free Energy Differences From Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245−268.

(62) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic Ingredients of Free Energy Calculations: a Review. *J. Comput. Chem.* **2010**, *31*, 1569−1582.

(63) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.

(64) Guy, A. T.; Piggot, T. J.; Khalid, S. Single-Stranded DNA Within Nanopores: Conformational Dynamics and Implications for Sequencing; a Molecular Dynamics Simulation Study. *Biophys. J.* **2012**, *103*, 1028−1036.

(65) Wassenaar, T. A.; Pluhackova, K.; Böckmann, R. A.; Marrink, S.-J.; Tieleman, D. P. Going Backward: a Flexible Geometric Approach to Reverse Transformation From Coarse Grained to Atomistic Models. *J. Chem. Theory Comput.* **2014**, *10*, 676−690.

(66) Petrauskas, A. A.; Kolovanov, E. A. ACD/Log P Method Description. *Perspect. Drug Discovery Des.* **2000**, *19*, 99−116.

(67) Bennett, W. F. D.; Tieleman, D. P. Water Defect and Pore Formation in Atomistic and Coarse-Grained Lipid Membranes: Pushing the Limits of Coarse Graining. *J. Chem. Theory Comput.* **2011**, *7*, 2981−2988.

(68) Bennett, W. F. D.; Tieleman, D. P. The Importance of Membrane Defects—Lessons From Simulations. *Acc. Chem. Res.* **2014**, *47*, 2244−2251.

(69) MacCallum, J. L.; Tieleman, D. P. Calculation of the Water-Cyclohexane Transfer Free Energies of Neutral Amino Acid Side-Chain Analogs Using the OPLS All-Atom Force Field. *J. Comput. Chem.* **2003**, *24*, 1930−1935.

(70) Sim, A. Y. L.; Lipfert, J.; Herschlag, D.; Doniach, S. Salt Dependence of the Radius of Gyration and Flexibility of Single-Stranded DNA in Solution Probed by Small-Angle X-Ray Scattering. *Phys. Rev. E* **2012**, *86*, 021901.

(71) Periole, X.; Cavalli, M.; Marrink, S.-J.; Ceruso, M. A. Combining an Elastic Network with a Coarse-Grained Molecular Force Field: Structure, Dynamics, and Intermolecular Recognition. *J. Chem. Theory Comput.* **2009**, *5*, 2531−2543.

(72) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. Ionic Effects on the Elasticity of Single DNA Molecules. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6185−6190.

(73) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *J. Chem. Theory Comput.* **2012**, *8*, 348−362.

(74) Chen, A. A.; García, A. E. High-Resolution Reversible Folding of Hyperstable RNA Tetraloops Using Molecular Dynamics Simulations. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 16820−16825.

(75) Savelyev, A.; MacKerell, A. D. All-Atom Polarizable Force Field for DNA Based on the Classical Drude Oscillator Model. *J. Comput. Chem.* **2014**, *35*, 1219−1239.

(76) Hansch, C.; Leo, A.; Hoekman, D.; Heller, S. R. *Exploring QSAR - Hydrophobic, Electronic, and Steric Constants*; American Chemical Society: Washington, DC, 1995.

(77) Cullis, P. M.; Wolfenden, R. Affinities of Nucleic Acid Bases for Solvent Water. *Biochemistry* **1981**, *20*, 3024−3028.