JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

A Solvent-Mediated Coarse-Grained Model of DNA Derived with the Systematic Newton Inversion Method

Aymeric Naômé,*[,†,‡,∥] Aatto Laaksonen,[‡,§,∥] and Daniel P. Vercauteren[†]

[†]Laboratoire de Physico-Chimie Informatique, Unité de Chimie Physique Théorique et Structurale, University of Namur, 5000 Namur, Belgium

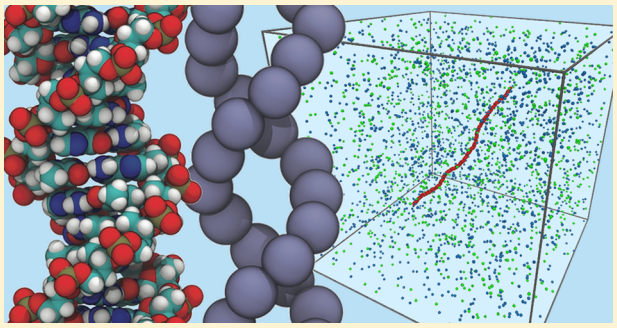
[‡]Division of Physical Chemistry, Department of Materials and Environmental Chemistry, Arrhenius Laboratory, Stockholm University, 10691 Stockholm, Sweden

[∥]Science for Life Laboratory, 17121 Solna, Sweden

[§]Stellenbosch Institute of Advanced Studies (STIAS), Wallenberg Research Centre, Stellenbosch University, 7600 Stellenbosch, South Africa

Supporting Information

ABSTRACT: We present a new class of coarse-grained (CG) force fields (FFs) for B-DNA with explicit ions suited for large-scale mesoscale simulations at microsecond−micrometer scale using a wide spectrum of particle simulation methods from molecular dynamics to dissipative particle dynamics. The effective solvent-mediated pairwise interactions making up the FFs are obtained by inverting radial distribution functions and other particle−particle distributions obtained from all-atom simulations of numbers of octadecamer DNA fragments from the Ascona B-DNA library. The inverse Monte Carlo (IMC) method, later known as Newton inversion (NI) (Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E*, **1995**, *52*, 3730−3737), was used together with the iterative Boltzmann inversion (IBI) scheme to compute the effective CG potentials. We show that this systematic structure-based approach is capable of providing converged potentials that accurately reproduce the structural features of the underlying atomistic system within a few percents of relative difference. We also show that a simple one-site-per-nucleotide model with 10 intramolecular pair interaction potentials is able to reproduce key features of DNA, for example, the persistence length and its dependence on the ionic concentration, experimentally determined around 50 nm at physiological salt concentration.



1. INTRODUCTION

The development of coarse-grained (CG) models of large biological molecules for efficient and accurate large scale molecular simulations is rapidly gaining increasing interest. Such simplified models allow studies across length and time scales not amenable at a detailed atomistic resolution. CG models not only reduce the amount of computing time but also the vast amount of data produced by atomistic simulations.

Although the "coarse-grained" approach is not new, its recent revival and the development of modern CG models differs in motivation from the origins of simplistic models proposed in the 1970s. In the beginning, simple models were mostly designed as the first steps toward more detailed ones, developing in complexity as the computer capabilities were growing. Today, in the era of ever growing computer power, palliating limited resources is not the only driving force for CG models to be developed. Bridging the phenomenological scales by deliberatively dropping details that are not necessary or relevant at a given spatial/temporal resolution to leave a simpler yet physically sound picture of the system of interest can often prove to be more beneficial than fully detailed brute-force approaches. Applying CG procedures to study phenomena as an alternative to all-atom (AA) models is thus a matter of cleverly getting rid of noise coming from too many details and to focus on the interesting behavior. For an overview of the development and usage of CG models in biophysical studies, we refer the reader to a recent review by Kamerlin and Warshel.[1]

Approaches to CG modeling fall, in general, in two distinct classes differing by the source of the information it incorporates. The "top-down" models, which are the most numerous, are parametrized following a trial-and-error refinement process targeting the reproduction of key experimental structural or thermodynamical features. On the other hand, the more recent "bottom-up" models are created by systematically reconstructing a CG Hamiltonian based on underlying more detailed simulations. Two approaches of the latter class are (i) the structure-based method, grounded on the renormalization group theory,[2] developed by Lyubartsev and Laaksonen[3]

Received: March 16, 2014
Published: May 27, 2014

(IMC/NI, inverse Monte Carlo/Newton inversion) and Savelyev et al.[4,5] (MRG-CG, molecular renormalization group coarse graining) but in different formalisms and (ii) the force-matching method, coined multiscale coarse-graining scheme (MS-CG) due to Voth et al.[6−8] For further information about systematic CG models, the reader is referred to the following review by Brini et al.[9]

The past decade particularly has seen a blooming in the development of CG models for biomolecules, including lipids,[10] proteins,[11−13] or nucleic acids.[14,15] Despite the plethora of published models, either derived following an empirical or a rigorous way, only a handful of them are tackling nucleic acids. And among those, only a few have led to repeated publications and applicative use. Let us consider more closely the empirical three-site model of de Pablo and co-workers,[16−22] of Pantano and co-workers,[23−26] of Ouldridge and co-workers,[27] and the systematic MRG-CG one-site model due to Savelyev et al.[5,28−31]

The original solvent-free model of de Pablo with an implicit representation of ions was parametrized to reproduce melting curves of double-stranded DNA (dsDNA). It has consecutively evolved throughout the years, incorporating implicit solvent effects,[17] explicit ions,[18] and explicit water.[19] The latest refined model[22] addresses and fixes most of the limitations of previous versions. This family of models is capable of reproducing the thermodynamics of melting, the bubble formation dynamics, and the salt dependence of the persistence length of dsDNA and single-stranded DNA (ssDNA). The first application[21] studying the hybridization of ssDNA on surface-tethered probes in DNA microarrays was performed with the original model. Pantano and colleagues devised a similar model with six interaction sites per nucleotide tuned to match the structure of canonical B-DNA.[23] Hydration and ionic strength effects were included by using a generalized Born implicit model. Their model manages to reproduce melting curves and their ionic strength dependence as well as the breathing dynamics of the fraying ends. Notably, they also implemented the model into a direct hybrid multiresolution AA/CG scheme.[24] They studied the curvature of DNA induced by kinks, bubble formation, bending and breathings along multimicrosecond trajectories and suggested that curvatures present in most structures of protein−DNA complexes can arise from thermally induced deformation.[25] In their last contribution to date,[26] they assess the accuracy of their FF by extensive comparison to microsecond long AA MD trajectory. The model of Ouldridge et al. differs mainly inasmuch as the backbone geometry has no imposed helicity through angle and dihedral terms, instead the use of base stacking potentials alone succeeds to render realistic representations of both ss- and dsDNA.[27] The focus on accurate description of DNA hybridization thermodynamics and kinetics makes this model very powerful for DNA nanotechnology applications.[32,33] The model of Savelyev et al. differs from the others in that it lies in the systematic CG category. Their methodology will be discussed later as it is almost identical to the approach of Lyubartsev and Laaksonen.[3] Their model of dsDNA with explicit ions and effective solvent-mediated potentials reproduces the persistence length of DNA at various ionic concentrations. Interestingly, their simulations of backbone neutralization[29] and at high salt concentration (0.1−1 M)[31] suggested that theories of polyelectrolyte chains due to Manning[34] and Odijk−Skolnick−Fixman[35] (OSF) overestimate and underestimate DNA stiffness, respectively. Two applications of their model showed salt-induced buckling of DNA minicircles[28] and DNA wrapping around a charged nanosphere.[30]

We should also mention the original models of Morriss-Andrews et al.,[36] Sayar et al.,[37] Linak et al.,[38] Edens et al.,[39] He et al.,[40] and Cragnolini et al.[41]

The IMC method developed in 1995[3] provides an elegant iterative way of solving the inverse problem of statistical mechanics, viz. reconstructing the interaction potentials from a distribution of particles, here from a set of radial distribution functions (RDFs) calculated in atomistic simulations. This iterative scheme corrects a guessed set of interaction potentials, typically the potential of mean force (PMF) obtained by direct Boltzmann inversion of the distributions, until the distributions generated by a molecular mechanics simulations method converge to their target values. We use Metropolis Monte Carlo simulations to generate ensemble averages, while Savelyev et al.[4,5] applied MD. The IMC process differs from the iterative Boltzmann inversion scheme (IBI)[42] in that in IMC the correlations between the distributions are accounted for: the value of the potential at some value of a given structural variable will not only influence the distribution of this variable but also all other distributions. It follows that the set of potentials generated is self-consistent. Since the term IMC is misleading as it only refers to the way the phase space is sampled, we rather coin the method NI, for Newton inversion. This is because the whole procedure resembles the iterative solution of a multidimensional nonlinear equation using the Newton−Raphson method.

The applicability of the NI method was already illustrated with the construction of effective potentials for water molecules and water−ion interactions based on ab initio molecular dynamics (MD),[43,44] solvent-mediated ion−ion[45−47] and ion−DNA,[48] one-site model of water, L- and D-proline in DMSO,[49] for lipid models[50−53] and for ionic liquids[54] all based on classical MD simulations. These applications to relatively simple systems have proven successful, but the capability of the method to yield CG potentials suited for performing simulations with large complex biomolecules has not been tested yet and was so far subject to certain reservation.[3−5] In the present paper, we particularly address the important issue concerning the method's convergence. Provided that the NI procedure setup is carefully prepared the convergence does not need to become an issue. We also present a class of FFs for DNA and ions suitable for simulations on the microsecond−micrometer scale and beyond. We have validated our CG models by reproducing the AA structural information we incorporated into the inversion procedure and by calculating other independent properties. The quality of a simple one-site-per-nucleotide DNA CG model is assessed by comparing the persistence length under varying ionic concentration.

2. METHODS

2.1. The Inverse Monte Carlo/Newton Inversion Procedure. The Hamiltonian (the potential energy contribution) of a system of N particles interacting via n different pair potentials is of the form:

$$H(\mathbf{r}^N) = V(\mathbf{r}^N) = \sum_{i<j}^{N} V_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) = \sum_{p}^{n} \sum_{i<j,\in p}^{N} V_p(|\mathbf{r}_i - \mathbf{r}_j|) \quad (1)$$

where the sum runs over all pair types n and all pairs of interaction sites i and j with coordinates $\mathbf{r}_i$ and $\mathbf{r}_j$. Giving the potential $V(\mathbf{r}^N)$ as a function spanning the whole Hamiltonian space Γ, the concise eq 1 can be rewritten as

$$\begin{aligned} H(\mathbf{r}^N) &= \int_\Gamma V(\mathbf{r}^N) \sum_{i<j}^{N} \delta(\mathbf{r}^N - \mathbf{r}_i)\delta(\mathbf{r}^N - \mathbf{r}_j)\mathrm{d}\mathbf{r}^N \\ &= \int_0^\infty V_{ij}(r_{ij}) \sum_{i<j}^{N} \delta(r_{ij} - |\mathbf{r}_i - \mathbf{r}_j|)\mathrm{d}r_{ij} \\ &= \sum_{p}^{n} \int_0^\infty V_p(r_p) \sum_{i<j,\in p}^{N} \delta(r_p - |\mathbf{r}_i - \mathbf{r}_j|)\mathrm{d}r_p \end{aligned} \tag{2}$$

which is the product of some continuous potential function $V_p(r_p)$ depending on particle pairs separation by a set of positional Dirac functions $\delta(r_p - |\mathbf{r}_i - \mathbf{r}_j|)$ (r is not restricted to distance separations, it can be any collective variable). To proceed to a practically amenable Hamiltonian, the continuous Hamiltonian in eq 2 can be further expressed in a truncated stepwise grid approximation version as

$$H(r_\alpha^{N_\alpha}) = \sum_{\alpha}^{N_\alpha} K_\alpha S_\alpha \tag{3}$$

where K_α denotes the value of $V_p(r_p)$ in the slice $r_p - \Delta r_\alpha/2 < r_\alpha < r_p + \Delta r_\alpha/2$, S_α is the number of particle pairs whose interdistance r_p lies within the r_α slice, and N_α is the total number of bins for the n pair types. S_α is thus an estimator of the RDF, $\langle S_\alpha \rangle = 4\pi r_\alpha^2 \Delta r_\alpha g_\alpha m_p / V_{\mathrm{box}}$ (m_p being the total number of pairs of a given type: $m_p|_{\mathrm{A-A}} = N_\mathrm{A}(N_\mathrm{A} - 1)/2$, $m_p|_{\mathrm{A-B}} = N_\mathrm{A}N_\mathrm{B}$). If we know a certain set of RDFs $\{g_\alpha^*\} \equiv \{S_\alpha^*\}$ describing the equilibrium distribution of particles of a system, it is then possible to reconstruct a set of pairwise effective interaction potentials that will approach the exact continuous, untruncated, complete Hamiltonian of eqs 1 and 2. If we had the knowledge of the full distance distribution of every particle pair, we would be able, in theory, to recreate the exact AA Hamiltonian.

We are of course not interested in this objective as the functional form of the AA FF is already known. We rather integrate out unnecessary degrees of freedom while keeping the physical significance of the underlying Hamiltonian. As to reconstruct such an effective Hamiltonian, we need a set of RDFs between CG sites mapped onto the detailed structure of the system.

The iterative procedure to obtain the desired effective CG Hamiltonian proceeds as follows. Starting from a set of trial potentials $\{K_\alpha^{(0)}\}$, a Monte Carlo (MC) simulation of the CG molecular system is performed, the RDFs $\{S_\alpha^{(0)}\}$ are calculated, compared to the reference ones, and a correction $\{\Delta K_\alpha^{(0)}\}$ is brought to the potentials. This sequence is repeated with the new potentials $K_\alpha^{(1)} = K_\alpha^{(0)} + \Delta K_\alpha^{(0)}$ until a satisfactory convergence threshold of the RDFs is reached, i.e., $\langle S_\alpha \rangle \approx g_\alpha^*$. Practically, this iterative procedure is carried out in two stages, differing in the way the correction is evaluated, namely through IBI and NI. IBI is used to first rough-hew the set of PMFs $K_\alpha^{(0)} = -kT \ln S_\alpha^*$ used as the starting trial potentials. This method first suggested by Soper[42] is to be preferred over NI since a poor initial approximation of the potentials, like the PMF, converges faster (less sampling) to values reasonably close to the target distributions. The new potentials are calculated as a refinement of the initial potentials according to the mean field approximation:

$$K_\alpha^{(1)} = K_\alpha^{(0)} + \frac{1}{\beta}\ln[S_\alpha^{(0)} - S_\alpha^*]\lambda \tag{4}$$

where λ is a regularization factor between 0 and 1 making sure the RDFs converge to the reference without oscillating back and forth around the target values. When no further improvement in the RDFs is observed using IBI, the second stage (NI) is initiated to further refine the RDFs. Since S_α in eq 3 are some functions of the potentials K_α, a truncated series expansion can be written as

$$\Delta\langle S_\alpha \rangle = \sum_\gamma \frac{\partial \langle S_\alpha \rangle}{\partial K_\gamma} \Delta K_\gamma + O(\Delta K_\gamma^2) \tag{5}$$

and from exact statistical mechanics relationships[3] and taking into account the linearity in $\{K_\alpha\}$ of the Hamiltonian eq 3:

$$\begin{aligned} \frac{\partial \langle S_\alpha \rangle}{\partial K_\gamma} &= \frac{\partial}{\partial K_\gamma} \frac{\int S_\alpha(\mathbf{r}^N) \exp[-\beta \sum_\lambda^{N_\alpha} K_\lambda S_\lambda(\mathbf{r}^N)]\mathrm{d}\mathbf{r}^N}{\int \exp[-\beta \sum_\lambda^{N_\alpha} K_\lambda S_\lambda(\mathbf{r}^N)]\mathrm{d}\mathbf{r}^N} \\ &= -\beta\left[\left\langle S_\alpha \frac{\partial H}{\partial K_\gamma}\right\rangle - \langle S_\alpha \rangle \left\langle \frac{\partial H}{\partial K_\gamma}\right\rangle\right] \\ &= -\beta[\langle S_\alpha S_\gamma \rangle - \langle S_\alpha \rangle \langle S_\gamma \rangle] \end{aligned} \tag{6}$$

By solving the system of linear equations (eq 5) with the covariance matrix elements defined in eq 6, taking $\Delta\langle S_\alpha \rangle^{(0)} = \langle S_\alpha \rangle^{(0)} - S_\alpha^*$ and omitting the quadratic terms $O(\Delta K_\gamma^2)$, we obtain the correction $\Delta K_\alpha^{(0)}$. Since NI takes into account the cross-correlation between the distributions, a higher degree of convergence can be achieved compared to IBI. Again, a regularization factor λ scales the correction to ensure proper convergence. As the intramolecular pair distributions converge faster than RDFs, we also reduce the size of the covariance matrix by freezing the related potentials in the last stage of the procedure.

The MC simulation protocol follows the customary Metropolis scheme. A first stage of equilibration of a few million MC steps is carried out prior to the averaging of the RDFs over a further $5-20 \times 10^6$ steps with maximum displacement $\mathrm{d}r = 0.5$ Å. Electrostatics is explicitly accounted for between charged CG sites involved in intermolecular potentials. As a result, the interaction potentials between such pairs subject to optimization do not contain the contribution from electrostatics which needs to be also considered explicitly in the CG-MD simulations. Electrostatics is effectively accounted for in the intramolecular potentials involving charged CG pairs and do not need to be added. For the MC simulations, the Ewald summation treatment is used with a direct space cutoff of 2.4 nm, the highest magnitude of wave vectors in reciprocal space set to 7, and $\varepsilon_r = 80$. The size, composition, and temperature of the solvent-free CG systems for IBI/NI optimization are the same as for the AA system. The whole procedure is carried out using the software package MagiC,[52] recently released and available at http://code.google/com/p/magic/. The IBI and NI methods are also implemented in the VOTCA toolkit for coarse-grain modeling.[55] Another software performing IBI is IBIsCO.[56]

In the context of this work, we find interesting to briefly describe and discuss the similar MRG-CG method. In their

implementation, Savelyev and co-workers adopt a dimensional reduction of the CG Hamiltonian (eq 3) by taking $\{K_\alpha\}$ in the sense of the parameters of polynomial fitting functions for bond, angular, and torsional intramolecular interaction potentials and a linear combination of Gaussians with excluded volume (r^{-6} or r^{-12}) for intermolecular terms:

$$V_{\text{bond}} = \sum_{n=2}^{4} K_{\text{bond}}(l - l_0)^n,\ V_{\text{angle}} = \sum_{n=2}^{4} K_{\text{angle}}(\theta - \theta_0)^n$$

$$V_{\text{ion-DNA;ion-ion}} = \frac{A}{r^{6;12}} + \sum_{k=1}^{3;5} B_k e^{-C_k(r-R_k)^2} \tag{7}$$

The set of observables $\{S_\alpha\}$ thus becomes the corresponding structural observables: $S_\alpha^{\text{bond/angle}} = (x - x_0)^n$, $S_\alpha^{\text{excl.vol.}} = r^{-6,-12}$, and $S_\alpha^{\text{Gauss}} = e^{-C_\alpha(r-R_\alpha)^2}$, where $\{C_\alpha\}$ and $\{R_\alpha\}$ are determined from the initial (trial) PMFs and fixed throughout the iterative procedure. This choice of basis functions has the obvious advantage of reducing drastically the number of covariance matrix elements but could in turn put too much constrain on the shape of the free energy surface and impede convergence. Pair PMFs between spherically symmetrical species show a small number of oscillations and can be easily fitted with a few Gaussian functions localized at peaks and valleys. However, when coarse-graining groups of atoms with internal degrees of freedom, the resulting RDFs are often more rugged and fitting the corresponding potentials requires many Gaussians not especially centered on extrema. Moreover, unlike for RDFs reported in ref 4, nothing ensures that more complex shapes of distributions can be reproduced with the same variances and heights of the Gaussians that fit the direct Boltzmann inverted PMFs. These parameters can unfortunately not be subject to optimization since the assumed linearity of the Hamiltonian eq 3.

2.2. All-Atom Molecular Dynamics Simulations. To build the CG models of DNA described in the following section, we have taken advantage of about 2 μs of AA MD trajectories from a work due to members of the Ascona B-DNA Consortium.[57] Those consist of a set of 39 50 ns-long MD simulations of octadecamer B-DNA fragments containing the 136 unique tetranucleotide sequences. DNA is modeled with the modified parmbsc0 FF,[58] ions with the Dang parameters,[59] and water with SPC/E[60] (see ref 57 and http://gbio-pbil.ibcp.fr/ABC for further details). Those trajectories provide us with enough conformational sampling to confidently construct CG models containing all needed molecular details.

2.3. Coarse-Grain Mapping and Topology Definitions. The topology of the CG representations is defined as a two-level construct. The first level is the geometric mapping of the atomistic structure. In this step, nucleotides are reduced to a number of interaction sites (beads) which are located at the center of mass (c.o.m.) of the atoms they represent and bear the sum of their individual masses. Ions are considered as CG sites on their own. The second level is the definition of the interaction matrix, namely, listing CG sites pairs for calculating the RDFs that finally yield the effective potentials. In the present work, we have chosen to map each of the four nucleotides to a generic interaction site (the mass is taken as the average of the four different nucleotides masses). The definition of the interaction network is arbitrary. However, a minimum number and range of interactions are necessary to maintain a well-behaved helical structure. Evidently, the maximum range of interactions is determined by the size of the original AA simulations. It is also obvious that more numerous and long-ranged the pairwise interactions are, the more tedious it is to achieve convergence in the IBI/NI procedure. Intermolecular interactions are subject to long-ranged electrostatics interactions in addition to the short-ranged nonbonded effective terms. Pairs of particles which are involved in an intramolecular interaction do not have electrostatic interactions in our model. Similarly in AA FFs, atoms involved in bonded interactions are excluded from nonbonded interactions (except for 1−4 interactions where nonbonded interactions are scaled down). The term "bond", in that regard, means nothing more than a pairwise interaction for which nonbonded interactions are excluded.

In this work, the interactions between CG sites are described by either 6, 8, 10, 13, or 16 pairwise intramolecular and 5 intermolecular potentials. Intramolecular potentials are used in the sense of bonds only (inherent to the choice of positional Dirac functions, eq 2), and angular as well as torsional interactions are indirectly represented by the network of bond-type interactions linking the different CG centers as it is depicted in Figure 1. Let us note that angular and torsion

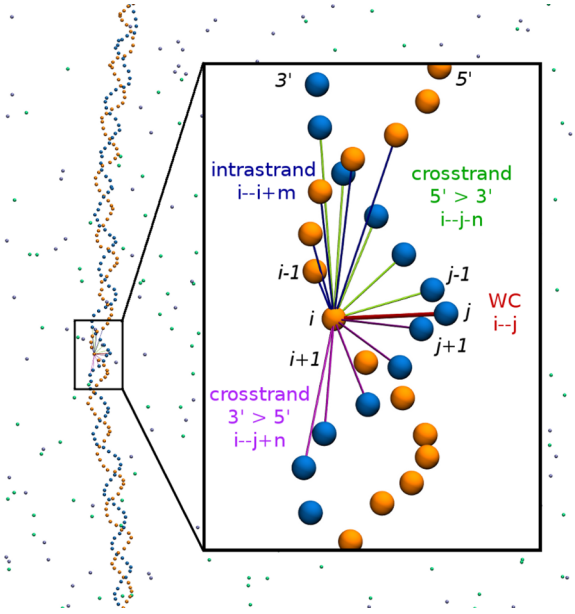


Figure 1. The one site per nucleotide model. The beads represent the interaction centers placed at the c.o.m. of a nucleotide. Every site interacts with other nucleotides through intrastrand interactions and interstrand interactions. For *16pot*, *13pot*, and *10pot*, $m = n = 5$, 4, and 3, respectively. *8pot* and *6pot* have $m = 3$ and $n = 2$ or 1, respectively. Symmetrical interaction matrices (i.e., n,m are the same in the 5′ → 3′ and 3′ → 5′ directions) up to $m = n = 6$ can be extracted from the trimmed reference octadecamer.

potentials can be straightforwardly included into the IBI/NI scheme by introducing $\langle S_\alpha \rangle$ as histograms over distributions of the corresponding angles or torsions. These many-body potentials can be dealt with in the MagiC software.[52]

The 2 μs of trajectories[57] were mapped with our reduction scheme, and distance distributions were extracted discarding the two capping residues subject to fraying. The intramolecular distributions are trimmed for population below 0.3% of the particle average density as to remove both low probability

dx.doi.org/10.1021/ct500222s | J. Chem. Theory Comput. 2014, 10, 3541−3549

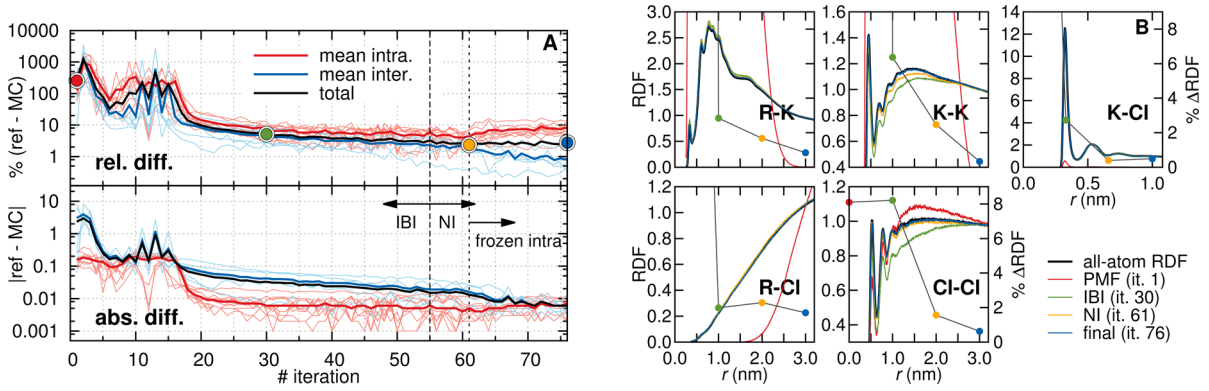


Figure 2. (A) Relative and absolute differences between target and MC generated distributions (*10pot*). Mean values are averaged over bin values. (B) Convergence of the RDFs for selected iterations. "R" stands for nucleotide beads. The lines' colors correspond to their position in A (filled circles). The colored circles in B are the RDF's individual relative differences.

configurations and avoid noise in the potentials. Since the AA MD simulations were performed using the Amber simulation package[61] with truncated octahedral PBC,[57] we calculated the RDFs and performed the MC simulations using the same conditions. As Amber defines the box with the distance between hexagonal faces and MagiC[52] between square faces, box dimensions in the original AA trajectories have to be multiplied by a factor $2/\sqrt{3}$. The system subject to the IBI/NI procedure is the following: 70 K^+, 36 Cl^-, 18 nucleotides for a total of 142 CG sites in a 8.95 nm truncated octahedron box.

2.4. Coarse-Grained Molecular Dynamics. The CG MD simulations were performed using the MD program package GROMACS v4.[62] The potentials were smoothed and interpolated (500 points/nm) with a Gaussian smoothing scheme are stored in look-up tables. The forces are calculated by GROMACS as minus the numerical derivative of the tabulated values. Like in the MC simulations, the interaction potential between charged CG sites comprises an electrostatic contribution, here either in the form of a Coulomb $1/r$ potential or particle mesh Ewald (PME). The short-ranged tabulated potentials were modified to avoid cutoff effects. In order to avoid discontinuity of the force at the right side border of the short ranged potential r_{SR}. The potential energy jump from $V_{SR}(r_{SR})$ to zero was removed by shifting every intermolecular potential by $-V_{SR}(r_{SR})$. For plain-cutoff simulations, the total force should ideally be zero at the electrostatic cutoff radius r_{cut}, what is ensured by applying a switching function from $r_{cut}/2$ to r_{cut} (see the Supporting Information for more details about the postprocessing of the output potentials). The relative dielectric constant is taken as ε_r = 80 to account for the screening effect of water. Several plain cutoff's have been tried for the Coulomb interactions: r_{cut} = 3.2(= r_{SR}), 8.0, 12.0 nm. For a DNA double-helix in a straight conformation, those distances roughly correspond to interactions between phosphates as far as 10, 24, and 36 base pairs (bp) apart, respectively. For the PME simulations, a direct space cutoff of 8 nm with a grid spacing of 1.6 nm is used (~885 000 grid points for the largest systems considered in this work). MD simulations are performed in the NVT ensemble with Nosé–Hoover thermostat. A stable integration time step of 5 fs is employed in all simulations, with a neighbor list update frequency of 2 time steps.

2.5. Validation and Characterization. We validated our model by first checking the agreement between the target and CG MD bond distributions and RDFs, and if other distributions, not explicitly incorporated into the model, are satisfactorily reproduced. The fluctuation of the CG site positions was also compared to the CG mapped AA trajectories as well as the spatial density distributions of cations around DNA. We also determined the acceleration of the dynamical scale due to the absence of friction and to a lesser extent to the smoothening of the potential energy surface inherent to the coarse-graining process by inspecting the diffusion behavior of the ions and the relaxation times of the DNA twist angle and end-to-end distance.

A convenient measure of the bending ability of polymer chains is the persistence length l_P that quantifies its stiffness as the distance after which the orientation of the chain becomes uncorrelated.[63] We thus tested the applicability and quality of our model by assessing its ability to reproduce this mechanical property of DNA and its dependence on the ionic concentration. The persistence length l_P is also the unique mechanical parameter of a theoretical model known as the worm-like chain (WLC) depicting DNA as a fluctuating chain of inextensible joined links.[64,65] This model gives remarkably good agreement with results from single molecule experiments.[65,66] We evaluated l_P using the exponent decay approximation:

$$\langle \hat{\mathbf{l}}_i \cdot \hat{\mathbf{l}}_j \rangle_{|i-j|=n} = \exp(-n\|\mathbf{l}\|/l_P) \tag{8}$$

where $\hat{\mathbf{l}}_i = \mathbf{l}/\|\mathbf{l}_i\|$ is a unit vector along segment i direction and $\|\mathbf{l}\|$ is the average segment length.

3. RESULTS AND DISCUSSION

3.1. Convergence of the IBI/NI Procedure. In this section, we assess the convergence of the IBI/NI procedure. The cutoff for the 5 intermolecular RDFs is 32 Å, and the bin size for all distributions is 0.05 Å. We thus have a set of 5 × (32/0.05) = 3200 observables $\langle S_\alpha \rangle$ that enter eqs 5 and 6. Solving for ΔK_α thus necessitates the inversion of the 3200 × 3200 matrix of the ~5 million unique covariance elements $\langle S_\alpha S_\gamma \rangle - \langle S_\alpha \rangle \langle S_\gamma \rangle$. Practically though, $\langle S_\alpha \rangle$, which are zero-valued, are excluded from the system of linear equations (and the corresponding potential values are set to very high values). Therefore, in the present case, the actual number of equations in eqs 5 and 6 is not more than 2900 (max. 3800 including intramolecular terms for *10pot*). The starting configuration is taken as a random system containing a 18 bp DNA fragment and appropriate number of ions. Starting from the PMFs, we performed the IBI/NI procedure until convergence of every

dx.doi.org/10.1021/ct500222s | J. Chem. Theory Comput. 2014, 10, 3541–3549

pair potential is reached. The statistics of observables are collected from eight independent MC simulations running in parallel. The convergence profile for the *10pot* force field is illustrated in Figure 2A. The intramolecular terms typically converge faster, during the IBI phase (iteration 1−55). When no further significant improvement is observed with IBI, the NI scheme is turned on. As already reported by Rühle et al.,[55] if NI allows better convergence, it also requires significantly more statistics. We also use a lower λ to dampen the effect of subsequent too abrupt potential updates (Table 1). To further speed-up and ease the slower convergence of intermolecular terms, we freeze the intramolecular potentials during the last NI iterations (iterations 62−77). After convergence, the average intermolecular and intramolecular relative differences are reduced to <1% and <9%, respectively (Figure 2). RMSDs are not plotted but almost quantitatively correspond to the mean absolute RDF differences and reach a value of 0.01 for both inter- and intramolecular distributions. Note that the convergence profile shown in Figure 2A corresponds to the first exploratory IBI/NI and that similar convergence can be achieved in a dozen of iterations with the optimal number of MC steps and regularization factor λ. The individual relative RDF differences are shown in Figure 2B at selected stages of the procedure. It is obvious from the red curves that the set of direct Boltzmann inverted PMFs (the trial potentials), i.e., every distribution considered as an independent degree of freedom, cannot be reliably used as a CG FF *per se*. Since the different FFs were derived independently (i.e., *13pot* interaction potentials is not a subset of the converged *16pot* FF, and so forth), homologue pair potentials do not necessarily have the same shape as they are correlated with different number of other pair potentials (a figure showing the final converged interaction potentials and their target distributions is available in the Supporting Information).

Table 1. IBI/NI Procedure for *10pot* FF

no. iterations	method	MC steps[a] (× 10^6)	λ
15	IBI	10(1)	0.5
30	IBI	10(1)	0.2
10	IBI	11(1)	0.2
4	INI	14(2)	0.05
2	INI	14(2)	0.08
15	INIfrozen	14(2)	0.1

[a]In parentheses, equilibration steps not included in the averaging.

3.2. Consistency of CG-MD Simulations. The intramolecular distributions are virtually unchanged compared to the IBI/NI converged distributions, independently of the time step, pair list update frequency (nstlist), and cutoff (data not shown). Longer-ranged intramolecular pair distributions which are not explicitly included in the tested FFs, e.g., the helical turn length $i - (i + 10)$, are satisfactorily reproduced for *16···10pot* FFs. The backbone angle and torsion angle distributions which are not incorporated in the FFs as such are also well matched (Figure 3). Three intrastrand and five interstrand interactions ($m = 3$, $n = 2$) are sufficient to conserve a helical structure. The *6pot* FF was discarded since the structural deviations are too severe. The RDFs are more affected by the choice of time step in CG simulations, especially the ion−ion contact peak for the pair K−Cl (Figure 3). We assessed the reproduction of the distributions for both PME and plain cutoff descriptions of electrostatics. If only PME gives RDFs virtually the same as the original ones (Figure 3), the plain cutoff scheme still provides high accuracy at a cheaper computational cost.

It is also informative to display the ionic atmosphere around DNA in three dimensions. For the role of ions around DNA the reader is referred to a recent review by Mocci and Laaksonen.[67] Figure 4B shows the isosurfaces of constant local K^+

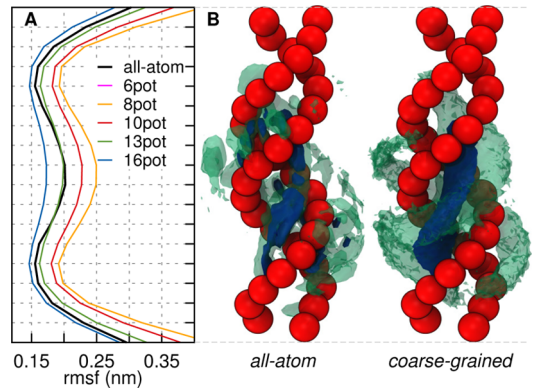


Figure 4. (A) Position fluctuation profile of the CG sites along the sequence. The fluctuations for *6pot* are in the range 0.57−0.83 nm. (B) Distribution of K^+ around DNA calculated from the AA trajectories and a 1000 ns CG MD simulation. The isosurfaces correspond to AA 20 and 10 times, CG 7.5 and 3.5 times the bulk ion density.

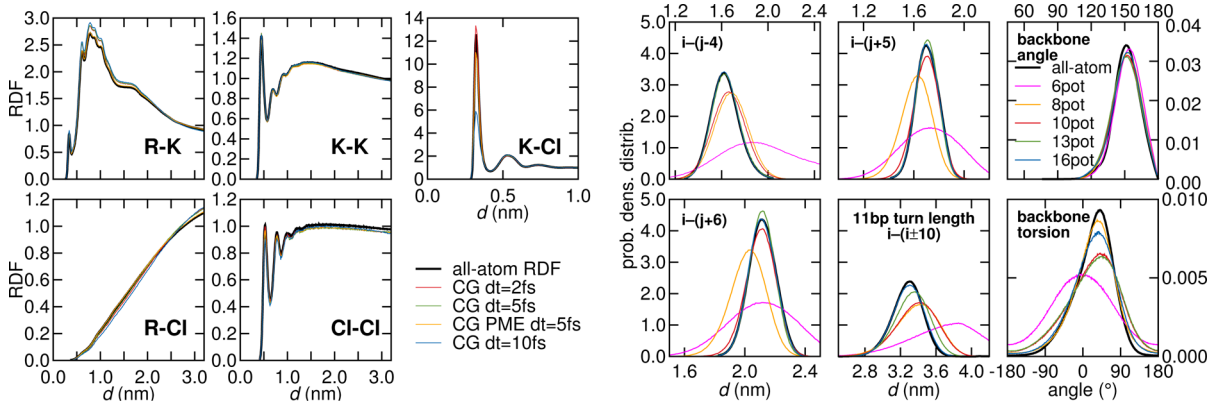


Figure 3. (left) RDFs for ions and DNA site with different time steps and nstlist = 2 (*16pot*). (right) Distribution of intramolecular terms not incorporated into the models. Explicit interaction potentials exist for distributions with $n = 4$ and $n = 4,5$ in the *13pot* and *16pot* FFs, respectively.

dx.doi.org/10.1021/ct500222s | J. Chem. Theory Comput. 2014, 10, 3541−3549

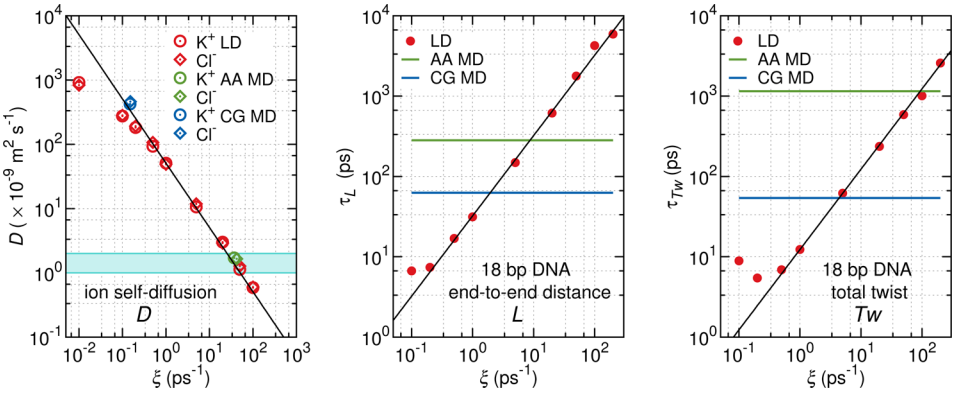


Figure 5. (left) Self-diffusion coefficient D of ions, (middle) relaxation time τ_{Tw} of DNA total twist, and (right) τ_L of DNA end-to-end distance with respect to the friction coefficient ξ for LD and MD simulations. The blue area delineates the diffusion coefficient range for most inorganic ions in aqueous solution.

concentration for both the AA and the CG models. Note that the ionic atmosphere is smeared-out in the CG model as the specificity of the ion coordination sites on DNA nucleobases is averaged out to a single spherical interaction site for all four different nucleotides. The nucleotides c.o.m. position fluctuations (rmsf) profile along the sequence is another property which is well reproduced by the CG models, independently of the number of intramolecular potentials used (Figure 4A), despite the implicit treatment of the solvent. DNA twist is equally well reproduced by the 8pernucl to 16pernucl FFs and is in the range 33.8−32.8°. The average twist calculated for the mapped AA trajectories is 33.2° (canonical B-DNA structure: 34.3°) (see Supporting Information for how the twist is determined).

3.3. Scaling of the Dynamics. In order to determine the acceleration of the dynamical scale for *in vacuo* CG MD due to the absence of a source of friction, i.e., the solvent, we calculated the mean squared displacement (MSD) of K^+ and Cl^- and extracted the self-diffusion coefficient D according to the Einstein equation for diffusion ($D = \lim_{t\to\infty}\langle|x_i(t_0 + t) - x_i(t_0)|^2\rangle_{i,t_0}/6t$). We obtain $D_{K^+} = 425 \times 10^{-9}$ and $D_{Cl^-} = 462 \times 10^{-9}$ m^2 s^{-1}, about 220× higher than the AA MD ($D_{K^+} = 1.71 \times 10^{-9}$ and $D_{Cl^-} = 2.032 \times 10^{-9}$ m^2 s^{-1}) and experimental values ($D_{K^+} = 1.957 \times 10^{-9}$ and $D_{Cl^-} = 2.032 \times 10^{-9}$ m^2 s^{-1}).[68] For those diffusion coefficients, the Stokes' law gives a friction coefficient of $\xi = 0.15$ ps^{-1} ($\xi = 37$ ps^{-1} for AA MD). We also inspected the diffusional behavior for a series of Langevin dynamics (LD) simulations. Figure 5 shows that D scales linearly with ξ up to the friction for regular MD simulations. (For $\xi = 0.01$ ps^{-1}, the LD diffusion is faster than for MD, and this is because of a higher temperature since the friction term is too low to ensure proper thermostatting.) Since the linear dependence of the rate of dynamical processes upon the friction coefficient in the limit of large friction ($\xi \geq 10$) predicted by Kramers' theory[69,70] also applies here for lower ξ, meaningful difusional rates could be calculated from frictionless CG MD simulations thus saving a factor ∼220 of simulation time. Various sources mention a friction coefficient in the range $\xi = 50-90$ ps^{-1} to approximate the collision frequency due to liquid water.[70−73] In qualitative agreement, a friction coefficient of $\xi \sim 35$ ps^{-1} returns the experimental diffusion coefficients. A value of ξ between 25 and 55 ps^{-1} leads to diffusion coefficients in the range $1-2 \times 10^{-9}$ m^2 s^{-1}, like for most inorganic ions. Since the dynamics of DNA itself is most probably not affected to the same extent, we inspected the time relaxation of the twist angle and of the end-to-end distance for the original AA MD trajectories, for CG MD, and LD simulations. The relaxation times are obtained from the fitting of the time correlation functions $C(t)$ with a two-exponential function $C(t) = a\exp(-t/\tau_1) + b\exp(-t/\tau_1)$ as $\tau = a\tau_1 + b\tau_2$. The acceleration factor is determined as 4.5× for the end-to-end distance and 21× for the total twist (Figure 5). The lower extent of acceleration of the dynamical scale for these observables is due to the lesser importance of the solvent degrees of freedom on the given dynamical properties. Expressed differently, the degrees of freedom defining those observables are less coarse-grained. The total twist and the end-to-end distance extracted from the undampened CG MD simulations display a time correlation function with an additional oscillatory component with frequencies determined from the Fourier transform of $C(t)$. The fitting function used in that case is $C(t) = [a\exp(-t/\tau_1) + b\exp(-t/\tau_1)]\cdot[c\cos(2\pi t/T_1) + d\cos(2\pi t/T_2)]$. Those frequencies are 1.0 cm^{-1} ($T_1 = 33$ ps) for the end-to-end distance and 1.3 cm^{-1}; 3.4 cm^{-1} ($T_1 = 25$ ps; $T_2 = 10$ ps) for the total twist. The highest frequency can be attributed to the vibration of the individual (base pair step) twist angles, whereas the lower corresponds to the collective in-phase twisting giving rise to the breathing of DNA which, in turn, for short DNA fragments, produces most of the end-to-end vector fluctuations.

3.4. Persistence Length. To test the ability of our model to reproduce the mechanical properties of long dsDNA fragments and how they are affected by phosphate−phosphate electrostatic repulsion and ion−DNA interactions, we calculated the persistence length l_P (eq 8) of a long 300 bp DNA double-helix for three Coulomb cutoff ($r_{cut} = 3.2$, 8.0, 12.0 nm) and K^+ concentrations in the range 0−38 mM (see Supporting Information for details about the calculation of the persistence length). The smallest cutoff fails to reproduce the reciprocal dependence of the rigidity of DNA to the ionic concentration (data not shown). A cutoff of 8.0 nm performs better and seems to be a safe choice as the persistence lengths obtained do not change much further with $r_{cut} = 12.0$ nm and *16pot* (Figure 6). In view of the similar asymptotic values of l_P for *10pot* with cutoff and PME electrostatics, the choice of using a faster plain cutoff scheme instead of a more accurate Ewald method also appears to be valid for bigger systems (Figure 6), despite a less steep decay for low concentrations. The persistence lengths averaged over 1 μs trajectories and extrapolated to physiological ionic concentration give an asymptotic value of 93 nm with *16pot*. We can directly compare this value to estimations

dx.doi.org/10.1021/ct500222s | J. Chem. Theory Comput. 2014, 10, 3541−3549

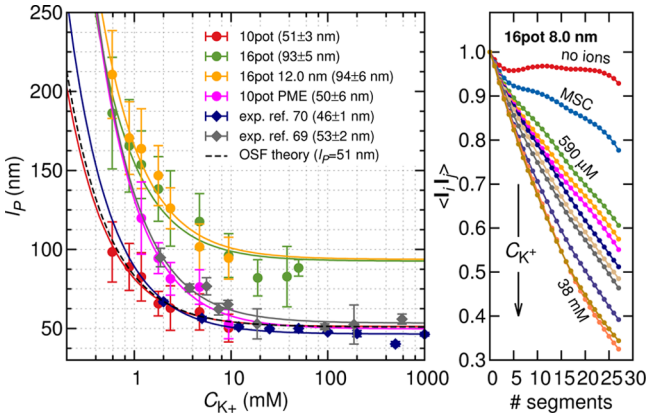


Figure 6. (left) Persistence lengths as a function of ionic concentration for the different FFs and from experiment.[78,79] The legend gives the asymptotic values of the persistence length fitted with a reciprocal law. The error bars of simulation results represent the standard deviation calculated from 100 ns subtrajectories. (right) Segment direction autocorrelations for *16pot* with concentrations up to 38 mM.

calculated from AA MD simulations since our CG FFs are extracted from simulations using the parmbsc0 AA FF. Bending persistence lengths calculated from AA MD simulations of short DNA fragments (3−11 and 25 bp) treated with the earlier parm98/parm94 FFs have been determined around 80−90 nm.[74,75] More recent studies with the parmbsc0 FF and longer fragments (12, 36, and 56 bp) give values of 43[76] or 44 nm.[77] Obviously, removing the longer-ranged intramolecular potential terms has the effect of reducing l_P down to 76 nm with *13pot* and 51 nm with *10pot*, in better agreement with the experimental value of ∼50 nm (corresponding to ∼150bp with C_{Na^+} = 150 mM).[78] In the range of concentrations considered, the reciprocal function fit of the *10pot* series shows a perfect overlap with the OSF theory of polyelectrolytes flexibility ($l_P(C) = l_{P,0} + l_{P,el} = l_{P,0} + 1/4\kappa^2 l_B = l_{P,0} + 32.4C^{-1}$ nm; κ^{-1}: Debye−Hückel screening length, l_B: Bjerrum length).[35] (see the Supporting Information for a version of Figure 6 for all considered FFs)

4. CONCLUSION AND PERSPECTIVES

We have shown that the Newton inversion (NI) method, originally known as inverse Monte Carlo (IMC), is able to provide effective solvent-mediated potentials suited for coarse-grained (CG) molecular dynamics (MD) simulations of DNA with explicit ions. The NI scheme brings distribution functions closer to their target values after that successive initial rounds of IBI do not lead to further improvement, especially for the intermolecular terms. Reducing the dimensionality of the inversion matrix by freezing the intramolecular potentials also proved to effectively improve convergence in the very last stage when full NI has reached its limit. We have constructed five CG DNA models reducing one nucleotide to a single site interacting with either 6, 8, 10, 13, or 16 intramolecular interaction potentials (*6···16pot*). All models (except *6pot*) reproduce the reciprocal behavior of the persistence length with increasing salt concentration. The *10pot* force field (FF) gives the asymptotic persistence length l_P = 51 nm closest to the experimental value of about 50 nm. In a future work, we will apply our FFs to the study of DNA circles structure and dynamics. A separate ongoing work on CG model developments is focused on constructing a peptide-DNA FF using the same methodology.

■ ASSOCIATED CONTENT

Supporting Information

Description of the postprocessing of the IBI/NI potentials before they are used in CG-MD, a figure showing the entire sets of effective potentials (*8, 10, 13, 16pot*) together with the target distributions, and extra information about how the persistence lengths are determined. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: aymeric.naome@unamur.be.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

A.N. thanks the Belgian National Fund for Research (F.R.S.-FNRS) for his F.R.I.A. doctoral scholarship, the European Science Foundation (ESF) SimBioMa networking programme for an exchange grant. A.N. is also grateful to Dr Laurence Leherte, Dr Alexander Mirzoev, Prof. Alexander Lyubartsev, Dr Francesca Mocci, and Mr Mathieu Fossépré for fruitful discussions. A.L. thanks the Swedish Science Council (Vetenskapsrådet) for continued support. A.N. and D.P.V. thank the IAP 7/05 for financial support. All authors acknowledge the Swedish National Infrastructure for Computing (SNIC, Sweden) and the Plateforme Technologique de Calcul Intensif (PTCI, supported by the F.R.S.-FNRS, Belgium) for computing resources. Special thanks are addressed to Prof. Richard Lavery for kindly sharing the all-atom molecular dynamics trajectories from which our coarse-grained models are derived.

■ REFERENCES

(1) Kamerlin, S. C. L.; Warshel, A. *Phys. Chem. Chem. Phys.* **2011**, *13*, 10401−11.
(2) Swendsen, R. H. *Phys. Rev. Lett.* **1979**, *14*, 859−61.
(3) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1995**, *52*, 3730−7.
(4) Savelyev, A.; Papoian, G. A. *Biophys. J.* **2009**, *96*, 4044−52.
(5) Savelyev, A.; Papoian, G. A. *J. Phys. Chem. B* **2009**, *113*, 7785−93.
(6) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.
(7) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.
(8) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.
(9) Brini, E.; Algaer, E. A.; Ganguly, P.; Li, C.; Rodríguez-Ropero, F.; van der Vegt, N. F. A. *Soft Matter* **2013**, *9*, 2108−19.
(10) Bennun, S. V.; Hoopes, M. I.; Xing, C.; Faller, R. *Chem. Phys. Lipids* **2009**, *159*, 59−66.
(11) Neri, M.; Anselmi, C.; Cascella, M.; Maritan, A.; Carloni, P. *Phys. Rev. Lett.* **2005**, *95*, 218102.
(12) Tozzini, V. *Acc. Chem. Res.* **2010**, *43*, 220−30.
(13) Saunders, M. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2012**, *22*, 144−50.
(14) de Pablo, J. J. *Annu. Rev. Phys. Chem.* **2011**, *62*, 555−74.
(15) Potoyan, D. A.; Savelyev, A.; Papoian, G. A. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 69−83.
(16) Knotts, T. A., IV; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. *J. Chem. Phys.* **2007**, *126*, 084901.

(17) Sambriski, E. J.; Schwartz, D. C.; de Pablo, J. J. *Biophys. J.* **2009**, *96*, 1675−90.
(18) Freeman, G. S.; Hinckley, D. M.; de Pablo, J. J. *J. Chem. Phys.* **2011**, *135*, 165104.
(19) DeMille, R. C.; Cheatham, T. E., III; Molinero, V. *J. Phys. Chem. B* **2011**, *115*, 132−42.
(20) Florescu, A.-M.; Joyeux, M. *J. Chem. Phys.* **2011**, *135*, 085105.
(21) Schmitt, T. J.; Rogers, J. B.; Knotts, T. A., IV *J. Chem. Phys.* **2013**, *138*, 035102.
(22) Hinckley, D. M.; Freeman, G. S.; Whitmer, J. K.; de Pablo, J. J. *J. Chem. Phys.* **2013**, *139*, −.
(23) Dans, P. D.; Zeida, A.; Machado, M. R.; Pantano, S. *J. Chem. Theory Comput.* **2010**, *6*, 1711−25.
(24) Machado, M. R.; Dans, P. D.; Pantano, S. *Phys. Chem. Chem. Phys.* **2011**, *II*, 1−2.
(25) Zeida, A.; Machado, M. R.; Dans, P. D.; Pantano, S. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2012**, *86*, 021903.
(26) Dans, P. D.; Darré, L.; Machado, M. R.; Zeida, A.; Brandner, A. F.; Pantano, S. Advances in Bioinformatics and Computational Biology, Lecture Notes in Computer Science; Springer International Publishing: Switzerland, 2013; Vol. *8213*; pp 71−81.
(27) Ouldridge, T. E.; Louis, A. A.; Doye, J. P. K. *J. Chem. Phys.* **2011**, *134*, 085101.
(28) Savelyev, A.; Papoian, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 20340−5.
(29) Savelyev, A.; Materese, C. K.; Papoian, G. A. *J. Am. Chem. Soc.* **2011**, *133*, 19290−3.
(30) Cao, Q.; Zuo, C.; Ma, Y.; Li, L.; Zhang, Z. *Soft Matter* **2011**, *7*, 506−14.
(31) Savelyev, A. *Phys. Chem. Chem. Phys.* **2012**, *14*, 2250−4.
(32) Doye, J. P. K.; Ouldridge, T. E.; Louis, A. A.; Roman, F.; Šulc, P.; Matek, C.; Snodin, B. E. K.; Rovigatti, L.; Schreck, J. S.; Harrison, R. M.; Smith, W. P. J. *Phys. Chem. Chem. Phys.* **2013**, *15*, 20395−20414.
(33) Ouldridge, T. E.; Hoare, R. L.; Louis, A. A.; Doye, J. P. K.; Bath, J.; Turberfield, A. J. *ACS Nano* **2013**, *7*, 2479−2490.
(34) Manning, G. S. *Biophys. J.* **2006**, *91*, 3607−16.
(35) Skolnick, J.; Fixman, M. *Macromolecules* **1977**, *10*, 944−8.
(36) Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. *J. Chem. Phys.* **2010**, *132*, 035105.
(37) Sayar, M.; Avşaroğlu, B.; Kabakçoğlu, A. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **2010**, *81*, 041916.
(38) Linak, M. C.; Tourdot, R.; Dorfman, K. D. *J. Chem. Phys.* **2011**, *135*, 205102.
(39) Edens, L. E.; Brozik, J. A.; Keller, D. J. *J. Phys. Chem. B* **2012**, *116*, 14735−43.
(40) He, Y.; Maciejczyk, M.; Ołdziej, S.; Scheraga, H. A.; Liwo, A. *Phys. Rev. Lett.* **2013**, *110*, 098101.
(41) Cragnolini, T.; Derreumaux, P.; Pasquali, S. *J. Phys. Chem. B* **2013**, *117*, 8047−60.
(42) Soper, A. K. *Chem. Phys.* **1996**, *202*, 295−306.
(43) Lyubartsev, A. P.; Laaksonen, A. *Comput. Phys. Commun.* **2000**, *128*, 565−89.
(44) Lyubartsev, A. P.; Laasonen, K.; Laaksonen, A. *J. Chem. Phys.* **2001**, *114*, 3120−6.
(45) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1997**, *55*, 5689−96.
(46) Lyubartsev, A. P.; Karttunen, M.; Vattulainen, I.; Laaksonen, A. *Soft Mater.* **2002**, *1*, 121−37.
(47) Mirzoev, A.; Lyubartsev, A. P. *Phys. Chem. Chem. Phys.* **2011**, *13*, 5722−7.
(48) Lyubartsev, A. P.; Laaksonen, A. *Comput. Phys. Commun.* **1999**, *121−122*, 57−9.
(49) Lyubartsev, A. P.; Mirzoev, A.; Chen, L.; Laaksonen, A. *Faraday Discuss.* **2010**, *144*, 43−56.
(50) Lyubartsev, A. P. *Eur. Biophys. J.* **2005**, *35*, 53−61.
(51) Murtola, T.; Karttunen, M.; Vattulainen, I. *J. Chem. Phys.* **2009**, *131*, 055101.
(52) Mirzoev, A.; Lyubartsev, A. P. *J. Chem. Theory Comput.* **2013**, *9*, 1512−20.
(53) Mirzoev, A.; Lyubartsev, A. P. *J. Comput. Chem.* **2014**, *35*, 1208−18.
(54) Wang, Y.-L.; Lyubartsev, A. P.; Lu, Z.-Y.; Laaksonen, A. *Phys. Chem. Chem. Phys.* **2013**, *15*, 7701−12.
(55) Rühle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theory Comput.* **2009**, *5*, 3211−23.
(56) Karimi-Varzaneh, H. A.; Qian, H.-j.; Chen, X.; Carbone, P.; Müller-Plathe, F. *J. Comput. Chem.* **2011**, *32*, 1475−87.
(57) Lavery, R.; Zakrzewska, K.; Beveridge, D.; Bishop, T. C.; Case, D. A.; Cheatham, T. E., III; Dixit, S. B.; Jayaram, B.; Lankas, F.; Laughton, C.; Maddocks, J. H.; Michon, A.; Osman, R.; Orozco, M.; Perez, A.; Singh, T.; Spackova, N.; Sponer, J. *Nucleic Acids Res.* **2010**, *38*, 299−313.
(58) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. E., III; Laughton, C. A.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817−29.
(59) Dang, L. X. *J. Am. Chem. Soc.* **1995**, *117*, 6954−60.
(60) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. *J. Phys. Chem.* **1987**, *91*, 6269−71.
(61) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2013**, *3*, 198−210.
(62) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−47.
(63) Hagerman, P. J. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 265−86.
(64) Kratky, O.; Porod, G. *Recl. Trav. Chim. Pays-Bas* **1949**, *68*, 1106−22.
(65) Peters, J. P.; Maher, L. J., III *Q. Rev. Biophys.* **2010**, *43*, 23−63.
(66) Bustamante, C.; Smith, S. B.; Liphardt, J.; Smith, D. *Curr. Opin. Struct. Biol.* **2000**, *10*, 279−85.
(67) Mocci, F.; Laaksonen, A. *Soft Matter* **2012**, *8*, 9268−84.
(68) *CRC Handbook of Chemistry and Physics*, 86th ed.; Lide, D. R., Ed.; CRC Press: Boca Raton, FL, 2005.
(69) Hänggi, P.; Talkner, P.; Borkovec, M. *Rev. Mod. Phys.* **1990**, *62*, 251−341.
(70) Feige, M. J.; Paci, E. *J. Mol. Biol.* **2008**, *382*, 556−65.
(71) Brünger, A.; Brooks, C. L.; Karplus, M. *Chem. Phys. Lett.* **1984**, *105*, 495−500.
(72) Pastor, R. W.; Brooks, B. R.; Szabo, A. *Mol. Phys.* **1988**, *65*, 1409−19.
(73) Ponder, J. W. *Tinker v4.2 User Guide*; Jay Ponder Lab, Washington University: St. Louis, MO, 2004; p 46.
(74) Lankas, F.; Sponer, J.; Hobza, P.; Langowski, J. *J. Mol. Biol.* **2000**, *299*, 695−709.
(75) Mazur, A. K. *Biophys. J.* **2006**, *91*, 4507−18.
(76) Noy, A.; Golestanian, R. *Phys. Rev. Lett.* **2012**, *109*, 228101.
(77) Spiriti, J.; Kamberaj, H.; de Graff, A. M. R.; Thorpe, M. F.; van der Vaart, A. *J. Chem. Theory Comput.* **2012**, *8*, 2145−56.
(78) Baumann, C. G.; Smith, S. B.; Bloomfield, V. A.; Bustamante, C. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94*, 6185−90.
(79) Rizzo, V.; Schellman, J. *Biopolymers* **1981**, *20*, 2143−63.