ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/24411423

Molecular Renormalization Group Coarse-Graining of Electrolyte Solutions: Application to Aqueous NaCl and KCl

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY B · JUNE 2009

Impact Factor: 3.3 · DOI: 10.1021/jp9005058 · Source: PubMed

CITATIONS
49

READS
29

2 AUTHORS, INCLUDING:

Garegin A Papoian
University of Maryland, College Park
83 PUBLICATIONS **1,979** CITATIONS

SEE PROFILE

Available from: Garegin A Papoian
Retrieved on: 26 January 2016

Molecular Renormalization Group Coarse-Graining of Electrolyte Solutions: Application to Aqueous NaCl and KCl

Alexey Savelyev and Garegin A. Papoian*

Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290

Received: January 17, 2009; Revised Manuscript Received: April 12, 2009

Simplified, yet accurate, coarse-grained models are needed to explore the behavior of complex biological systems by means of Molecular Dynamics (MD) simulations, because many interesting processes occur at long time scales and large length scales that are not amenable to studies by atomistic simulations. The aqueous salt buffer provides an important contribution to the structure and function of biological molecules. While in many simplified models both water and salt are treated as a continuous medium, it is often desirable to describe mobile ions in an explicit manner. For example, the discrete nature of ions was shown to play a very important role in their interaction with highly charged biomolecules, such as DNA. In this work, we have derived an effective interaction potential for monovalent ions by systematically coarse-graining the all-atom NaCl and KCl aqueous solutions at several different ionic concentrations. Our approach is based on explicitly accounting for cross-correlations among various observables that constitute the compact basis set of the coarse-grained Hamiltonian. Compactness of the Hamiltonian ensures computational efficiency of the optimization procedure. In addition, it allows us to accurately reproduce many-body effects, in contrast with many existing algorithms. The resulting Hamiltonian produced ionic distributions that are virtually identical to those obtained in atomistic simulations with explicit water, capturing short-range hydration effects. Our coarse-grained model of monovalent electrolyte solutions allows the incorporation of ions into complex coarse-grained biomolecular simulations, where both electrostatic and short-range hydration effects must be taken into account.

Introduction

Because aqueous salt environments are crucially important in determining the structure and function of all biological molecules and processes, they are often studied by means of all-atom Molecular Dynamics (MD) simulations. However, the computational cost of treating water and the mobile ions in an explicit way significantly limits the applicability of all-atom MD simulations to relatively small molecular systems ($\sim 10^5$ atoms) evolving over short time scales ($\sim$100 ns). In contrast, many biophysical phenomena occur at much larger length scales and longer time scales. When solvent and ions are included explicitly, simulation of many biomolecular structures, such as chromatin fibers, requires computation for hundreds of thousands to tens of millions of particles, which are mostly solvent. Hence, to efficiently study such systems, a simplified, coarse-grained (CG) representation is needed. The main difficulty in developing CG models is retaining sufficient fidelity to the original explicit solvent description, such that essential physics of processes of interest can still be reliably captured.

In the majority of works on coarse-graining various molecular systems, a great deal of attention is paid to the accuracy of simplified models for the solute molecules, while the surrounding aqueous salt environment is treated as a continuous medium. There are, however, numerous examples suggesting the importance of explicitly considering the mobile ions in CG models. For instance, it is desirable and may even be necessary from a physical standpoint to include explicit ions in systems whose behavior is primarily driven by the electrostatic interactions. Compaction of highly charged DNA molecules inside nuclei of eukaryotic cells, so-called chromatin folding,[1] is illustrative of phenomena driven by electrostatics, where the extent of chromatin compaction and stability of specific conformations strongly depend on the concentration of mobile ions in a solution.[2] This is because counterions, such as Na^+ or K^+, interact with the oppositely charged DNA phosphate groups, mitigating the electrostatic repulsion between distant parts of DNA molecule and facilitating chromatin folding. Counterions also play an important role in RNA folding.[3]

Because of the complex nature of interactions among DNA, water molecules, and ions, the commonly used continuous approximations for mobile ions ignore some of the important interactions that occur at short distances (see refs 4, 5), which are vital to chromatin organization in the compact state. For example, representing ions explicitly may significantly improve coarse-grained modeling of a nucleosome, a nucleoprotein complex comprised of the DNA chain wrapped around the protein histone core, which is the basic unit of the chromatin fiber.[2] Counterions crucially interact not only with negatively charged DNA, but also the protein part of the nucleosome.[2] When we compared counterion association obtained from all-atom simulations with the nonlinear Poisson−Boltzmann equation results, at the same salt concentration, we found significant differences.[4] This points to the importance of the discrete character of ions and water, which is manifested by nontrivial hydration effects. Consequently, one may expect ionic interactions and hydration effects to have a profound effect on the formation of the higher order chromatin structures, because a typical internucleosomal distance in the chromatin fiber is of the order of $\sim$10 Å,[6,7] a distance at which continuous electrostatics breaks down.[4,5]

* Corresponding author. E-mail: gpapoian@unc.edu.

10.1021/jp9005058 CCC: $40.75 © 2009 American Chemical Society
Published on Web 05/08/2009

The influence of mobile ions on the dynamics of charged macromolecule has been demonstrated by recent works on the coarse-graining double-stranded DNA chain.[8−10] For example, a good agreement with experiment was obtained for the values of DNA persistence lengths calculated at various NaCl concentrations in a CG model of DNA segment, where mobile Na^+ and Cl^- ions were treated explicitly.[9] In a different and more detailed CG DNA model, however, with no explicit mobile ions, the DNA persistence length appeared to be twice smaller than the experimental value.[8] Finally, our recent model of the DNA chain,[10] also lacking explicit ions, produced a DNA persistence length that is larger by ∼25% than the corresponding experimental value (unpublished data). However, because all of these models differ in resolution and no parameter optimization was performed in the two first of the studies, it is not possible to infer the extent of the DNA model improvement due to explicit treatment of mobile ions. The above discussion points to the lack of an accurate CG model of DNA in which the mobile ions are treated explicitly. It is important to have such a model because there obviously exists a noticeable difference between the dynamics of CG DNA chain modulated by explicit mobile ions and the corresponding dynamics observed with the continuous model of surrounding salt.

The discussion thus far suggests that to accurately represent electrostatic interactions in simplified models of complex charged systems it is necessary to consider mobile ions on equal footing with the charged object itself. Toward this goal, implicit-solvent interionic potentials are necessary to describe the bulk properties of monovalent electrolyte solutions by systematically integrating out the solvent degrees of freedom. This is the goal of the present study.

Our approach is based on constructing an effective Hamiltonian for the CG system, mobile Na^+ (or K^+) and Cl^- ions without explicit water, where the Hamiltonian is expressed as a linear combination of molecular interaction potentials, which are simultaneously thought of as relevant physical observables, followed by subsequent iterative parameter optimization (see Methods section). Our motivation is to make the Hamiltonian compact and physically plausible, which significantly reduces the choice of physical observables from the potentially infinite manifold. In this respect, our approach differs principally from the closely related optimization method due to Lyubartsev and Laaksonen, where all pairwise interactions were decomposed over the very large and not physically motivated set of positional Dirac delta-functions.[11−13] An extremely large number of the latter basis functions are needed for describing many-body interactions in polymer chains, such as bending and dihedral interactions, making the subsequent inversion problem impractical, as discussed in detail in our recent work.[10] On the other hand, the Lyubartsev−Laaksonen (LL) optimization technique seemed to have worked reasonably well for simple systems with small number of pairwise interactions, as demonstrated by coarse-graining of ionic solutions.[11] However, even for such systems, the optimization procedure may be quite problematic because of inherent redundancy of the large delta-function basis set, which is further elaborated below. As a result of this degeneracy, multiple unrelated parameter sets might result from LL optimizations, posing serious convergency and uniqueness issues, as recognized in the earlier works.[11]

Our CG Hamiltonian, on the other hand, is highly compact at the price of introducing basis set functions of different dimensionality. As explained in the Methods section, the Hamiltonian compaction may be viewed as "projecting" the dynamics of the original AA system onto the relevant set of dynamical modes, associated with the reduced CG Hamiltonian basis set. The advantage of our method rests on the linearity of the Hamiltonian with respect to optimizable parameters and also Hamiltonian's compactness, where these properties play a key role during parameter optimization. The latter, within the mean field approximation, relies on matching various pair (cross)-correlators for the dynamical variables in CG and the reference AA systems. At the same time, Hamiltonian compactness makes it computationally feasible to go far beyond matching solely pair correlators and permits reproduction of correlators of higher orders, thus significantly enhancing the accuracy of the model. Additionally, the Hamiltonian linearity over physical observables allows us to interpret the system's partition function as a generating functional that can be differentiated to obtain the corresponding correlation functions. Mathematically then, matching various order correlators of the physical observables that directly enter the CG Hamiltonian between CG and AA systems ensures a significant equivalence of the corresponding partition functions, making the coarse-graining in this work reminiscent of renormalization group (RG) theory,[14] as further elaborated below. From this perspective, integrating out the solvent degrees of freedom, as we pass from the AA to the simplified CG system (comprised of ions only), corresponds to one-step renormalization. In more complex molecular systems, multiple rounds of coarse-graining would be straightforward using this approach.

Notably, both LL and our methods are RG-based optimization techniques. In particular, the LL technique was adapted from the RG Monte Carlo method developed by Swendsen to study critical phenomena in three-dimensional Ising models.[15] However, the above-mentioned difficulty of incorporating the many-body effects into the optimization scheme is a drawback of LL approach, restricting its applicability to simple molecular systems. After Lyubartsev et al., we generalized Swendsen's RG method even further to effectively deal with many-body effects in complex molecular systems. For example, molecular renormalization group coarse-graining (MRG-CG), using molecular basis functions in the RG based optimization scheme, has been applied to developing a two-bead per base pair model of double-stranded DNA.[10] In the present work, we test the method on much simpler systems of mobile Na^+ (K^+) and Cl^- ions at various concentrations. Despite simplicity, however, such systems are characterized by long-range electrostatic interactions and complex dynamical correlations among the ions. Thus, care should be exercised in the choice of the relevant physical observables to guarantee the completeness of the Hamiltonian basis set, which, in turn, assures the convergence of the parameter optimization procedure, as discussed below. The interaction potentials for the monovalent ionic solutions derived in this work (see the Supporting Information for the parameter set tables) may be used to incorporate ions into complex coarse-grained biomolecular simulations, where both electrostatic and short-range hydration effects need to be taken into account.

Method Section

MD Simulations of All-Atom Systems. To build CG models of NaCl electrolyte at various ionic concentrations by integrating out water from the corresponding all-atom aqueous solutions, we prepared three systems representing NaCl solutions at 100, 300, and 500 mM ionic concentrations. The systems were comprised of 7, 21, and 35 NaCl molecules, respectively, placed evenly throughout the cubic box having dimensions 45 × 45 × 45 Å. We note that a larger simulation box could have been used to reduce the finite size effects. However, we are only trying to determine short-range interactions for the coarse-

grained Hamiltonian, on the order of 10 Å. The long-range interactions in the CG model are given by the Coulomb potential, without adjustable parameters. Because the periodic cell linear dimension is at least 4-fold greater as compared to the length scale of the short-range interactions, we expect that finite size effects do not appreciably influence the optimized parameter values for these interactions. All systems were further solvated in more than 2800 TIP3P water molecules.[16] We used the recently developed force field for alkali and halide monovalent ions to parametrize interionic and ion−water interactions[17] and AMBER 10.0 software suite to carry out all MD simulations.[18] It is important to note that the earlier default force field for monovalent ions in AMBER led to small overestimation of K^+Cl^- association, as discussed in our recent works[5,19,20] as well as the works of others.[21−23] All initial structures have been first minimized according to the standard steepest descent algorithm. Next, they were heated to 300 K over a period of 1 ns and subsequently equilibrated for another 5 ns in the canonical NVT ensemble. The subsequent production run for each system was carried out at constant temperature (300 K) and pressure (1 bar) using the Langevin temperature equilibration scheme, the "weak-coupling" pressure equilibration scheme,[24] and periodic boundary conditions. The translational center-of-mass motion was removed every 2 ps. We used the SHAKE algorithm[25] to constrain all bonds involving hydrogens, which allows all MD simulations to use an increased time step of 2 fs without any instability. Particle Mesh Ewald method[26] was used to treat long-range interactions with a 10 Å nonbonded cutoff. The production run was carried out for 60 ns to ensure the equilibration of ions. The above protocol was also used to simulate a 300 mM solution of K^+ and Cl^- ions, whose composition and geometry were identical to those of the 300 mM NaCl system.

MD Simulations of Coarse-Grained Systems. We used the large-scale atomic/molecular massively parallel simulator (LAMMPS)[27] to carry out MD simulations of CG ionic systems at 100, 300, and 500 mM concentrations. Number of Na^+ (K^+) and Cl^- ions as well as the size of the periodically repeating cubic cell in each system were the same as in the corresponding all-atom system. To be consistent when going from atomistic to coarse-grained simulations, we kept the sizes of simulation boxes equal to those of the AA system, such that both the cutoff for direct Coulomb interactions and, importantly, the size of periodic cell along with the precision of the Ewald summation procedure (which defines the number of k-space points) remain unchanged. Our experience with varying the periodic box sizes suggests that the latter requirement is necessary, because different conditions for Ewald procedure may influence electrostatics. Following these requirements, the number of ions and the size of periodically repeating cubic cell in each of the CG systems were the same as in the corresponding AA system. We used the canonical NVT integration scheme (Nose/Hoover temperature thermostat) to update particle's positions and velocities at each time step.[28] Similarly to all-atom systems, all CG systems were minimized, heated to 300 K, and equilibrated before the production run used for further analysis. Particle Mesh Ewald technique[26] was used to treat the long-range (Coulomb) interactions. To determine the largest time step that can be used to simulate the CG system without instabilities, we used the criteria of the total energy conservation, the latter being the energy of the CG system complemented by the contribution from the Nose−Hoover Hamiltonian.[8] It appeared that it was safe to use the time-steps of up to 10 fs, so we used this upper limit in the MD simulations. The production run for each optimization iteration was 40 ns to ensure the convergence of the covariance matrix for physical observables (see eq 5). We verified the convergence at each iteration by comparing the data generated from two halves of the MD trajectory.

Computing the Ionic Potential of Mean Force. As is customary, we use the potentials of mean force (PMF) as a starting approximation for the effective interactions in the CG system.[29] PMFs are obtained by the Boltzmann inversion of the corresponding all-atomistic distribution functions.[30] In the context of the present work, interionic PMF, U_{PMF}, is expressed through the ionic radial distribution function (RDF), $g(r)$, as follows:

$$U_{\mathrm{PMF}} = -k_{\mathrm{B}}T \ln g(r) \tag{1}$$

To calculate $g(r)$, we first compute Na−Na, Na−Cl, and Cl−Cl separations for all ionic pairs and construct the corresponding distance histograms from each snapshot of MD simulation. Next, these histograms are subject to normalization by the volume (spherical) Jacobian, $J(r)\,\mathrm{d}r = 4r\pi^2\,\mathrm{d}r$, to ensure that the number of neighbors within a distance r from a given ion is

$$n(r) = \rho \int_0^r g(r)J(r)\,\mathrm{d}r \tag{2}$$

where ρ is the average concentration of ions.

Optimizing Force-Field Parameters Using RG-Inspired Approach. The optimization scheme used in the present work closely follows the Monte Carlo RG method developed by Swendsen and co-workers to compute critical exponents in Ising models.[15] To proceed with mathematical formulation of the problem, we first introduce an effective CG Hamiltonian $\mathscr{H}(\{K_\alpha\})$, defined by a parameter set, $\{K_\alpha\}$, $\alpha = 1...N$, and a set of observables of interest, $\{S_\alpha(\{K_\alpha\})\}$, subject to canonical averaging over $\mathscr{H}(\{K_\alpha\})$. In this work, we are mostly interested in Hamiltonians, which are of the following form, $\mathscr{H} = \sum_\alpha K_\alpha S_\alpha$. Here, the term "observable" refers to the quantity collected in AA and CG systems, thus being "measured" in the context of numerical experiments. As explained in the section Hamiltonian as a Linear Combination of Physical Observables, for the present case of electrolyte systems, the set of observables is related to specific integral characteristics composed of interatomic interaction potentials. The difference, $\Delta\langle S_\alpha\rangle \equiv \langle S_\alpha\rangle_{\mathrm{CG}} - \langle S_\alpha\rangle_{\mathrm{AA}}$, between the expectation values of an observable, S_α, averaged over CG and AA systems may be expressed as

$$\Delta\langle S_\alpha\rangle = \sum_\gamma \frac{\partial\langle S_\alpha\rangle_{\mathrm{CG}}}{\partial K_\gamma}\Delta K_\gamma + O(\Delta K^2) \tag{3}$$

which is simply an expansion of $\langle S_\alpha\rangle_{\mathrm{CG}}$ around some point in space of the Hamiltonian $\{K_\alpha\}$. The analogous observables, $\langle S_\alpha\rangle_{\mathrm{AA}}$, can be defined and computed from atomistic simulations; however, they enter into eq 3 as specific numbers, through $\Delta\langle S_\alpha\rangle \equiv \langle S_\alpha\rangle_{\mathrm{CG}} - \langle S_\alpha\rangle_{\mathrm{AA}}$. The derivative in eq 3 is given by (CG subscripts are omitted):

$$\frac{\partial\langle S_\alpha\rangle}{\partial K_\gamma} = -\frac{1}{k_{\mathrm{B}}T}\left[\left\langle S_\alpha \cdot \frac{\partial\mathscr{H}}{\partial K_\gamma}\right\rangle - \langle S_\alpha\rangle\left\langle\frac{\partial\mathscr{H}}{\partial K_\gamma}\right\rangle\right] \tag{4}$$

and represents "susceptibility" of observable $\langle S_\alpha\rangle$ to the change of parameter K_γ (α and γ may be different). Hence, eq 3 may be viewed as the system's linear response to an external potential ΔK. This analogy is particularly beneficial in case of Hamiltonians linear in $\{K_\alpha\}$, having the form $\mathscr{H} = \sum_\alpha K_\alpha S_\alpha$. Next, eq 3 reduces to

$$\Delta\langle S_\alpha\rangle = -1/(k_{\mathrm{B}}T)\sum_\gamma [\langle S_\alpha S_\gamma\rangle - \langle S_\alpha\rangle\langle S_\gamma\rangle]\Delta K_\gamma \tag{5}$$

being expressed in terms of cross-correlators of various observables, as expected for susceptibilities. The following

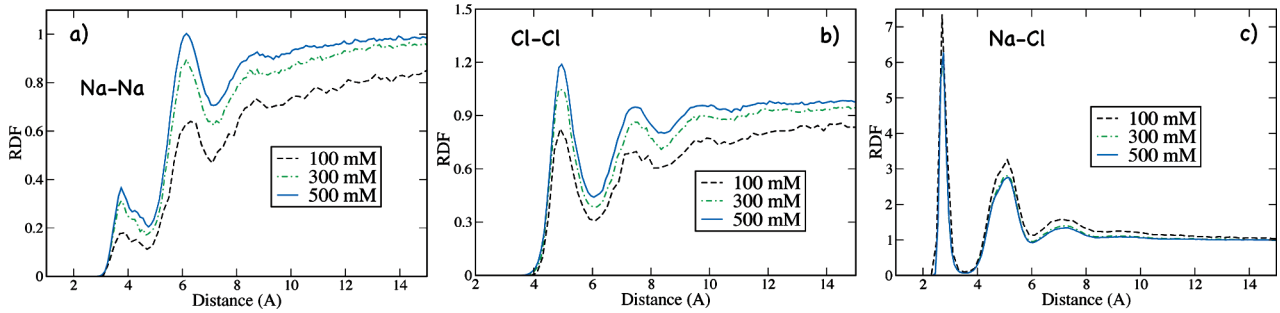

Figure 1. Radial distribution functions obtained from all-atom MD simulations of NaCl aqueous solution at three different ionic concentrations.

parameter optimization scheme may be used to decrease $\Delta\langle S_\alpha\rangle$. First, the $\langle S_\alpha S_\gamma\rangle_{CG}$ correlators are obtained from MD simulations of the CG system using some trial set of Hamiltonian parameters, $\{K_\alpha^{(0)}\}$, followed by the calculation of the deviations $\Delta\langle S_\alpha\rangle$ of each CG variable from their corresponding reference AA values. Subsequently, the system of linear eq 5 is solved to yield the corrections for the Hamiltonian parameters, $\Delta K_\alpha^{(0)}$, which define a new parameter set $K_\alpha^{(1)} = K_\alpha^{(0)} + \Delta K_\alpha^{(0)}$ for the next CG iteration. The procedure is repeated until the convergence of all CG variables is reached, that is, $\langle S_\alpha\rangle_{CG} \approx \langle S_\alpha\rangle_{AA}$.

Solving the Inverse Problem. Eigenvalues of covariance matrix (5) indicate how changes in various dynamical modes affect different effective potentials. For the present ionic problem, it turns out that the covariance matrix is nearly singular, resulting in degeneracy of solutions that represent various sets of parameters. Apparently, this problem is caused by the redundancy of interaction potential functions as well as by the noise that is normally present in the input data obtained from MD simulations.[13,31] When too many observables are used to describe the CG system, larger uncertainty in the covariance matrix inversion results, and thus the stronger is the degeneracy of the resulting set of CG Hamiltonian parameters. This implies, in particular, a significant advantage of using our compact set of 18 basis functions as compared to a very large set of ~600 positional Dirac delta-functions used in the LL approach. After a compact Hamiltonian basis set was introduced, further reduction in the degeneracy can be achieved by eliminating those matrix eigenvectors that superfluously affect Hamiltonian parameters. Singular value decomposition (SVD) could have been directly used to address this issue if the elements of the covariance matrix in eq 5 had identical physical units. However, the matrix element $\langle S_1^{ev} \cdot S_2^{Gauss}\rangle - \langle S_1^{ev}\rangle\langle S_2^{Gauss}\rangle$ involving cumulative Gaussian and the repulsive $(1/r)^{12}$ functions (see section Hamiltonian as a Linear Combination of Physical Observables) is measured in units of [Å^{-12}], while the diagonal element $\langle (S_2)^2\rangle - \langle S_2\rangle^2$ is dimensionless. Therefore, to use SVD at each iteration, we reduced the corresponding covariance matrix to a dimensionless form by appropriately rescaling vectors ΔK_α and $\Delta\langle S_\alpha\rangle$. Next, in matrix notation, the rescaled eq 5 takes a form,

$$\sum_j \frac{M_{ij}}{\sqrt{q_i \cdot q_j^T}} \cdot [X_j \cdot \sqrt{q_j}] = \frac{B_i}{\sqrt{q_i}},\ q_i \equiv M_{ii} \qquad (6)$$

with M, X, and B standing for the covariance matrix, vector of the corrections ΔK_α, and the vector of deviations $\Delta\langle S_\alpha\rangle$, respectively. As follows from the second equation, vector q is composed from the diagonal elements of the original matrix M. Hence, the latter is reduced to a dimensionless form (with unit elements on the diagonal) after its element-by-element division by the tensor $(q_i \cdot q_j^T)^{1/2}$. After near zero eigenvalues were filtered out and a subsequent matrix inversion was performed, the original units of the elements ΔK_α were obtained by reverse transformation.

Coarse-Grained Modeling of NaCl Solutions. We model an aqueous NaCl solution as a collection of particles, each carrying a charge of either +1 or −1, corresponding to Na^+ and Cl^- ions, respectively, which interact via potentials accounting for the ionic hydration and associated with the structural behavior of ions at small separations. The ion−ion RDFs computed from all-atom MD simulations of all three NaCl systems are shown in Figure 1. It is seen that there are several pronounced peaks in each RDF, indicating a formation of ionic "shells" around a given ion. In particular, the first peak corresponds to direct ionic contacts, while other peaks reflect the water-mediated interactions. Note, in more dilute solutions the repulsion between like charged ions is stronger because they are less screened electrostatically by ions of the opposite charge (see Figure 1a,b). This trend is consistent with the stronger attraction among Na^+ and Cl^- ions in dilute solutions, as indicated in Figure 1c. This structural analysis suggests that the effective CG Hamiltonian may be represented as a sum of (at least) three contributions,

$$\mathcal{H} = \mathcal{U}_{ex} + \mathcal{U}_{hyd} + \mathcal{U}_{el} \qquad (7)$$

In this expression, the first term indicates the energy due to excluded volume interactions, the second term is responsible for ionic hydration effects, and the last term represents the sum of electrostatic interactions.

As suggested in the previous section, the functional forms for individual energetic contributions can be inferred from the ionic PMFs obtained by the Boltzmann inversion of the corresponding RDFs (see eq 1). An example of the PMF characterizing Na−Cl interactions in the half-molar NaCl aqueous solution is demonstrated in Figure 2a.

First, we assume that particles in the CG system interact via long-range Coulomb potential. This, particularly, allowed us to set the absolute scale of the ionic PMF by equating the PMF value at the largest ionic separation to the interaction energy calculated from the analytical Coulomb potential. Aside from electrostatics, another obvious component of the PMF is a strongly repulsive short-range potential characterizing excluded volume interactions, whose functional form we approximate by the repulsive part of the Lennard-Jones potential, $(1/r)^{12}$. Finally, to capture structural ionic peaks and minima in the PMF, we propose to use Gaussian functions. Indeed, a certain PMF peak or minimum can be described by a function that is nonzero in a narrow range of interparticle separations, and a Gaussian function is ideally suited for that purpose. This is also a physically sound choice, because Gaussians corresponding to peaks of the PMF may be associated with the free energy penalty

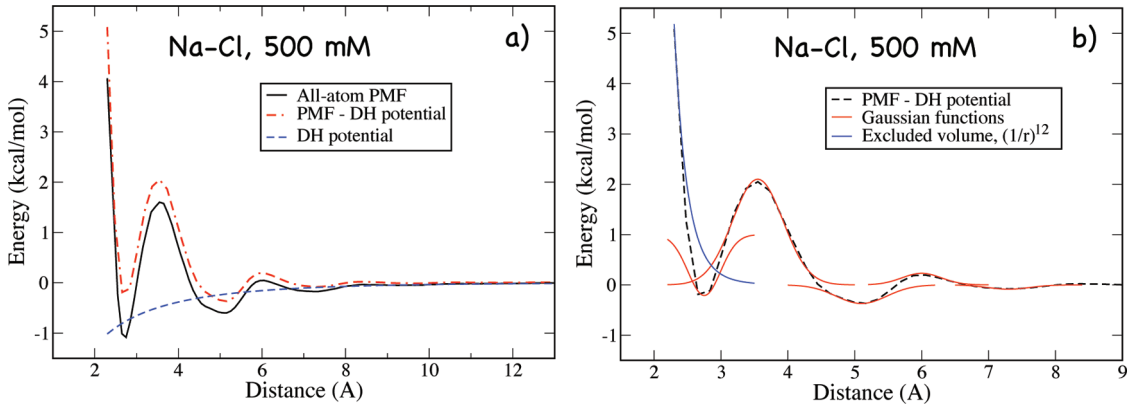

Figure 2. (a) The PMF (solid black line) representing Na−Cl interactions in a half-molar NaCl solution obtained by Boltzmann inversion of the corresponding all-atom RDF. For fitting purposes, it was decomposed into the Debye−Hückel (DH) potential and the remaining part characterizing structural ionic behavior. The latter was approximated by the repulsive $(1/r)^{12}$ potential and five Gaussian functions, as shown in panel (b). See the text for explanations.

for two ions to approach each other due to ionic dehydration. Similarly, Gaussians representing PMF minima describe a favorable range of ionic separations when ions are in a direct contact, separated by a certain amount of water molecules or fully hydrated. As a result, our effective Hamiltonian has the following functional form:

$$\mathscr{H} = \sum_{i>j}\left[\frac{A}{r_{ij}^{12}} + \sum_{k=1}^{5} B^{(k)} e^{-C^{(k)}[r_{ij} - R^{(k)}]^2} + \frac{q_i q_j}{4\pi\varepsilon_0 \varepsilon r_{ij}}\right] \tag{8}$$

defined by the set of parameters, $\{A,B^{(k)},C^{(k)}\}$, and the positions of Gaussian peaks and minima, $\{R^{(k)}\}$. The sum is over all pairs of CG particles. The Coulomb potential for interacting particles of charges q_i and q_j is scaled by the dielectric constant for water, $\varepsilon = 80$. We decided to include up to $k = 5$ Gaussian functions in the Hamiltonian to accurately capture the PMF structural properties represented by first three minima and first two maxima, as shown in Figure 2b.

Initial values for all Hamiltonian parameters can be obtained by fitting the above functional forms into ionic PMFs.[29] A remark regarding the technical side of doing this is in order. First, we approximate the electrostatic interactions by the Debye−Hückel (DH) potential, $(1/r)e^{-\kappa r}$, with the Debye length, κ^{-1}, corresponding to a given ionic concentration. The reason we substitute the real Coulomb interactions by DH potential follows from the fact that ionic PMF was obtained through averaging over all but two particles of interest. Indeed, on the level of mean field theory, the electrostatic interactions between any two ions in the electrolyte solution are represented by the screened Coulomb, or DH potential.[30,32] As a result, all of the initial parameters, $\{A,B^{(k)},C^{(k)},R^{(k)}\}$, follow from fitting analytical forms in eq 8, aside from the Coulomb potential, into the numerical difference between the all-atom PMF and the DH potential, as demonstrated in Figure 2a,b. Note, we used the DH potential exclusively for obtaining the initial guess of parameters for the short-range interactions. Because ions are explicitly present in the CG model, their bare interactions are Coulombic, and not of the DH form. Thus, we used the pure Coulomb potential, damped by the dielectric constant of water, in MD simulations of CG systems of many ions, as seen in eq 8. As discussed below, the trial values for Hamiltonian parameters obtained in a way just described generated ionic distributions already very plausible and close to the corresponding all-atom RDFs. In particular, relying upon the Debye−Hückel theory of electrolyte solutions resulted in a number of subsequent optimization steps significantly smaller than that had we started from a different and less physical approximation for electrostatic interactions during fitting (see section Optimizing the Hamiltonian Parameters and Comparing to All-Atom Results).

Hamiltonian as a Linear Combination of Physical Observables. As explained in the Methods section, the present optimization technique relies on representing the effective Hamiltonian in the following form, $\mathscr{H} = \sum_{\alpha=1}^{N} K_\alpha S_\alpha$, which is a linear decomposition over the N physical observables, $\{S_\alpha\}$, defined by a set of parameters $\{K_\alpha\}$. To elucidate what are physical observables in the current ionic problem, we analyze the structure of the effective Hamiltonian, eq 8. It follows that each type of effective ionic interactions is described by a very small number of observables, which are structure-based collective order parameters. For example, five collective modes characterizing structure of ionic "shells" are cumulative Gaussian functions, $S_\alpha^{\text{Gauss}} = \sum_{\text{allpairs}}[e^{-C_\alpha(r - R_\alpha)^2}]$, $\alpha = 1...5$, while the corresponding parameters $\{K_\alpha\}$ are given by the set of constants $\{B^{(k)}\}$; see eq 8. Note, the linearity of the Hamiltonian implies that only coefficients $\{K_\alpha\}$ in the expansion $\mathscr{H} = \sum_\alpha K_\alpha S_\alpha$ are subject to adjustment by optimization procedure (see Methods section). Therefore, while optimizing K_α's, we kept the Gaussian variances C_α's and the positions R_α's in observables $\{S_\alpha^{\text{Gauss}}\}$ to be equal to the values obtained from fitting the Hamiltonian (eq 8) into PMFs, as explained above. This is, indeed, a reasonable approximation, because both the width and the position of each Gaussian function remain unchanged in the course of optimization procedure (see Figures 3 and 4). It is worth noting, finally, that for observables $\{S_\alpha^{\text{Gauss}}\}$ to be physically meaningful, the variances of Gaussian functions should be extracted from their separate fitting into the corresponding peaks/minima of ionic PMFs and not from the best fit of the whole Hamiltonian into ionic PMF. This requirement reflects the locality of these observables and ensures the absence of spurious long-range correlations among $\{S_\alpha^{\text{Gauss}}\}$. The latter, in particular, significantly enhances the quality of the optimization procedure.

Analogously, the collective observable for excluded volume interactions is given by $S_\alpha^{\text{ev}} = \sum_{\text{allpairs}}(1/r^{12})$, and the role of the corresponding scaling paramter is played by the constant A. Finally, we do not associate the electrostatics with any observables because exact Coulomb interactions have no adjustable parameters. As a result, aside from electrostatics, $(5 \times 3) + 3 = 18$ constants $\{K_\alpha\}$ enter the effective Hamiltonian, because

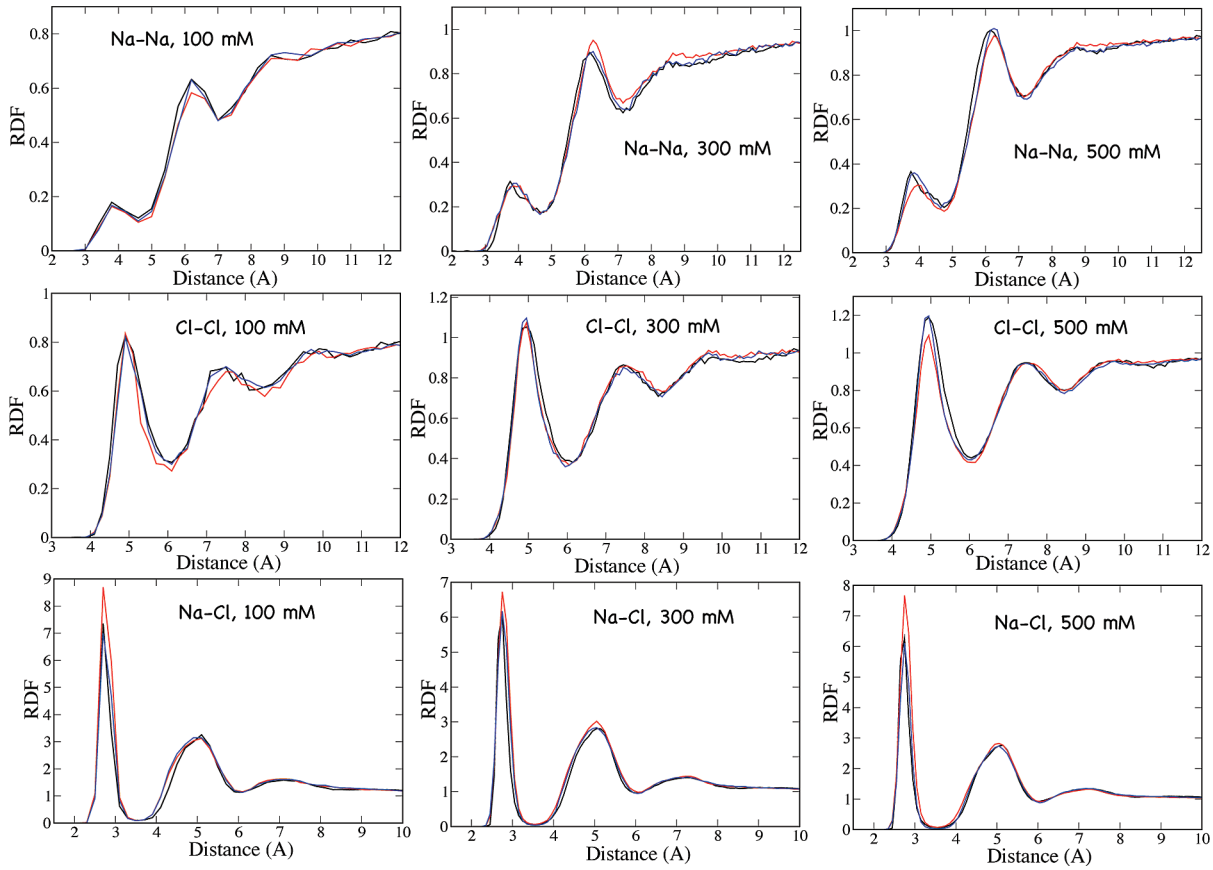

Figure 3. Radial distribution functions computed in three systems with 100, 300, and 500 mM ionic concentrations. Black, red, and blue lines represent the reference AA, initial CG, and the corrected by optimization final CG distributions, respectively. Initial CG distributions are those generated by potentials obtained by fitting to ionic PMFs using DH approximation for the electrostatics, as explained in the Coarse-Grained Modeling of NaCl Solutions section.

there are five $\{S^{\text{Gauss}}\}$ terms characterizing each of three types of ionic interactions (Na−Na, Na−Cl, and Cl−Cl), and also three corresponding excluded volume terms.

In the above discussion, K_α's may be understood as fields "conjugate" to collective observables S_α's. These fields are subject to optimization procedure aimed to reducing the difference between the expectation values of observables, $\Delta\langle S_\alpha\rangle = \langle S_\alpha\rangle_{\text{CG}} - \langle S_\alpha\rangle_{\text{AA}}$, averaged over CG and AA systems. From the other and equivalent perspective, S_α's can be viewed as a set of basis functions over which an effective Hamiltonian is spanned. A completeness of the given basis set is consistent with all $\Delta\langle S_\alpha\rangle$'s nearly vanishing after parameter optimization.

Parallels with prior related works can be drawn to illustrate various choices of physical observables. For example, in Swendsen's original work,[15] S_α's indicated various cumulative spin products, corresponding to interactions between nearest-neighbor and distant spins, as well as many-spin interactions (generated by RG). Lyubartsev et al. expressed[11] ionic radial distribution functions (RDF) in terms of S_α's, where the latter were positional Dirac delta-functions. As discussed in the Methods section, this renders the inverse problem significantly ill-posed, even for simple systems characterized by pairwise interactions, but especially when many-body effects are taken into account. In our recent work on DNA coarse-graining,[10] we related S_α's to various collective modes associated with different types of effective molecular interactions in a DNA chain, such as bond, bending angle interactions, etc. In the present work, we follow a similar strategy to decompose the ionic Hamiltonian over a compact set of physically motivated collective observables.

Optimizing the Hamiltonian Parameters and Comparing to All-Atom Results. The trial set of Hamiltonian parameters, $\{K_\alpha^{(0)}\}$, derived from fitting the functional forms in eq 8 to the PMFs (see section Coarse-Grained Modeling of NaCl Solutions) generated very plausible ionic distributions, as shown in Figure 3. Evidently, such a good agreement with all-atom results is a consequence of utilizing the DH potential to approximate the electrostatic part of the ionic PMF. To further improve the quality of ionic distributions, we optimized the Hamiltonian parameters by solving eq 5 according to the technique outlined in the Methods. MD simulations of all CG systems were carried out using the Los Alamos atomic/molecular massively parallel simulator (LAMMPS).[27] The details of the simulation protocol are provided in the Methods section.

It appeared that a very small number of optimization steps was required for all $\Delta\langle S_\alpha\rangle = \langle S_\alpha\rangle_{\text{CG}} - \langle S_\alpha\rangle_{\text{AA}}$ to nearly vanish. Because each $\Delta\langle S_\alpha\rangle$ is expressed in terms of cross-correlators for various CG degrees of freedom (see Methods section), this implies the absence of significant correlations between ions in a solution beyond DH correlations. For example, only one iteration was needed for the convergence of optimization procedure in a dilute 100 mM system of Na^+ and Cl^- ions. At the same time, two and three iterative steps led to the convergence in systems with ionic concentrations of 300 and 500 mM, respectively. This trend may be explained by less pronounced correlation effects in a more dilute solution. Hence, among studied systems, the PMF is the best starting approximation for 0.1 M NaCl solution, and, consequently, the least number of optimization steps were needed. Ionic distributions for all three simulated systems at the initial and final stages of

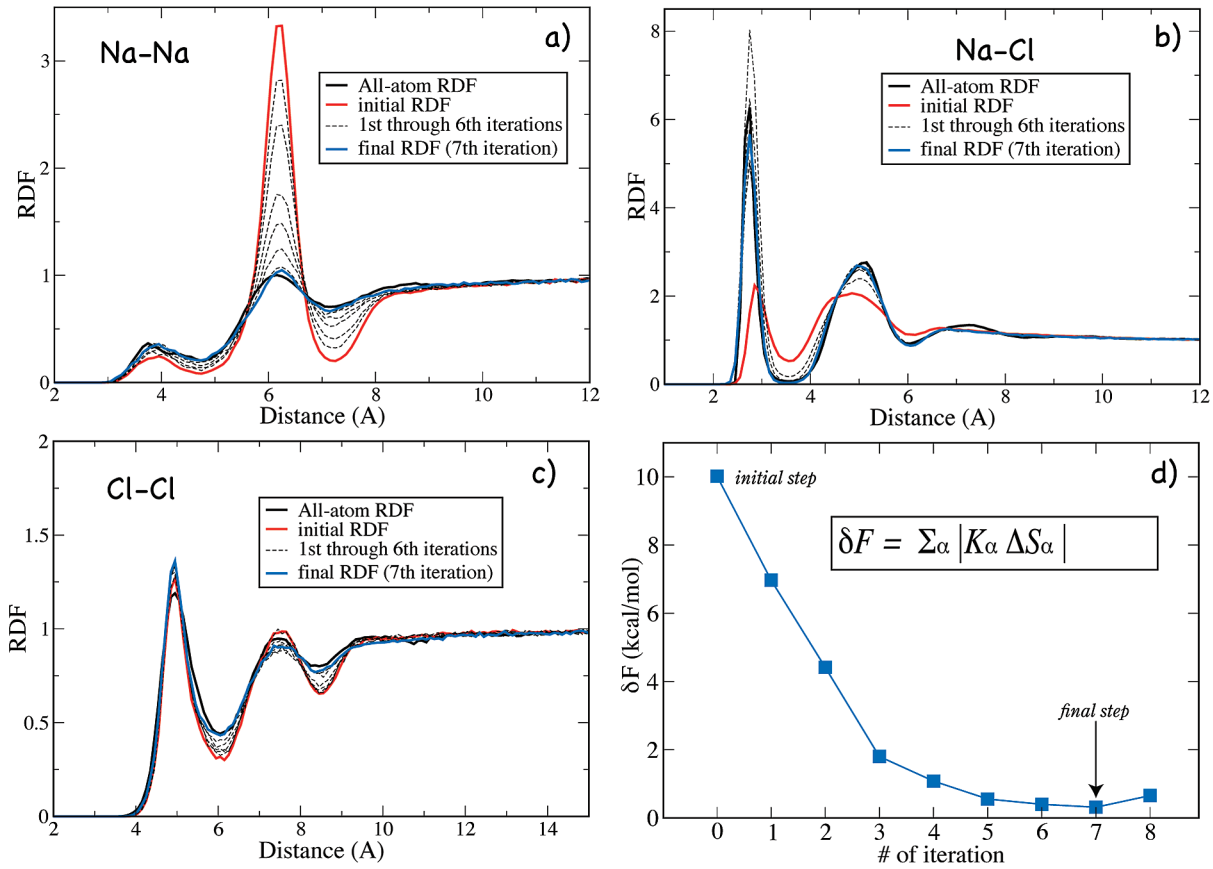

Figure 4. (a)−(c) Radial distribution functions for ions in a half-molar NaCl system obtained from optimization run starting from unfavorable initial conditions. Line colors are consistent with those in Figure 3. Distributions for intermediate steps are also indicated as dash lines. Last panel (f) demonstrates the reduction of the total free energy difference δF between AA and CG models with optimization iterations.

optimization procedure, as well as the corresponding all-atom RDFs, are shown in Figure 3.

Table 1 in the Supporting Information summarizes the final ionic parameters for NaCl solutions at 100, 300, and 500 mM concentrations. It is seen that parameters do not vary significantly as we alter the concentration of ions. Apparently, this is because the ionic RDFs are quite similar to each other, as indicated in Figure 1. As an additional test of the MRG-CG method, we derived an effective interaction potential for K^+ and Cl^- ions by coarse-graining the 300 mM KCl solution. Both AA and CG KCl systems were simulated according to the MD protocols outlined in the Methods section. As in the case of NaCl systems, we used the DH approximation for electrostatic interactions among ions to extract the trial set of Hamiltonian parameters. The optimized ionic parameters are given in Table 2 of the Supporting Information. The quality of distributions for ions in the CG KCl system (not shown) appeared to be comparable to the quality of ionic distributions in NaCl systems (Figure 3). Particularly, an excellent agreement was found with the all-atomistic results, and only two iterative steps were needed to reach the convergence of optimization procedure.

As discussed above, the parameters for ions do not change significantly as the concentrations are varied between 0.1 and 0.5 M. Therefore, we suggest, for example, that the parameters obtained for 0.1 and 0.5 M solutions may be used to simulate slightly more dilute and more dense electrolyte solutions, respectively, outside of the 0.1−0.5 M range. At the same time, it has to be kept in mind that parameter transferability depends strongly on the specific question being asked: if, for example, the observable of interest is very sensitive to the ionic concentration, then the coarse-grained model may need to be parametrized for a particular concentration of interest. Conversely, if one is interested, for example, in elucidating the dependence of the conformational dynamics of DNA on the concentration of the ionic buffer in a range of common interest, 0.1−1 M, the use of parameters obtained here is expected to be satisfactory.

Because the initial distributions generated by the trial set of Hamiltonian parameters are already very close to the reference all-atomistic ones, we also carried out an additional optimization run in a 500 mM NaCl system starting from much less favorable initial conditions. The ionic distributions obtained at different stages of such optimization are shown in Figure 4. It follows that the present RG-based optimization method is robust, demonstrating good agreement among atomistic and coarse-grained distribution functions for various structural observables.

We can estimate the change in the total free energy difference, $\delta F = \sum_\alpha K_\alpha \Delta S_\alpha$, between AA and CG systems in the course of optimization procedure. As illustrated in the last panel of Figure 4, only seven iterations are needed to reduce the (average) total free energy difference between AA and CG systems to a small value within the statistical error of the simulation ($\langle \delta F \rangle \approx 0.6 k_B T$). As noted earlier, when we utilized the DH approximation to infer the initial ionic parameters, the number of iterations were 1, 2, and 3, respectively, for 100, 300, and 500 mM NaCl solutions. Two optimization steps were needed for the 300 mM KCl solution. Finally, the discrepancies between the thermally averaged individual CG and AA terms, $||K_\alpha S_\alpha||$, were on the order of $0.01 k_B T$ per degree of freedom, indicating good correspondence between CG and AA Hamiltonians.

Because NaCl and KCl solutions are among the most commonly used salt buffers in both experimental studies (of,

e.g., DNA fibers and chromatin) and MD simulations, the obtained ionic parameters may be directly used in CG models of complex biological systems, where an explicit treatment of ions is desirable and necessary from physical standpoint. At the same time, the effective interactions between ions and the solute molecule itself have to be inferred from a separate analysis of the corresponding distribution functions. For example, in an ongoing work on incorporating an explicit NaCl salt into the recently developed CG DNA model,[10] we use the present optimization scheme to capture the coupling between the dynamics of a DNA chain and the surrounding Na^+ and Cl^- mobile ions, along with the ionic parameters derived in this work to represent the bulk properties of the NaCl solution.

Discussion

We suggest that the present coarse-graining scheme compares favorably with other commonly used optimization methods. Interestingly, prior works using this method for spin[15] and ionic systems[11] did not clearly elaborate on the specifics of its close relationship to the RG theory. Here, we point out these connections and demonstrate how to generalize the method to achieve an arbitrarily high accuracy. We start by noticing that representing Hamiltonian as a linear decomposition over observables S_α allows one to interpret the partition function, $\mathscr{Z}(\{K\}) \propto \sum \exp[-1/(k_B T)\sum_{\alpha=1}^{N} K_\alpha S_\alpha]$, as a generating function that can be differentiated to obtain all correlation functions,[14]

$$\langle S_1 \ldots S_n \rangle \propto \frac{\delta^n \ln \mathscr{Z}}{\delta K_1 \ldots \delta K_n} \tag{9}$$

Again, K_α's here may be viewed as the fields "conjugate" to observables S_α's. Thus, certain equivalence between the AA and CG partition functions may be achieved if their corresponding derivatives become equal. The equality of higher derivatives would suggest a more accurate correspondence. Hence, if two partition functions generate two identical sets of various auto- and cross-correlators of order n and less (hence, identical nth derivatives of the free energies), we can think of n as a degree of similarity between two generating functions. From this perspective, Swendsen's optimization method, which matches only first moments in distributions over observables S_α, corresponds to order $n = 1$ of equivalency between CG and AA systems. Within the present framework, it is straightforward to achieve higher accuracy in CG system description by demanding the coincidence of higher moments in S_α. This, in turn, would require computing (cross) correlators of order $n + 1$, to be used in equations equivalent to eq 5.

For example, we can use the condition $\Delta\langle S_\alpha S_\gamma \rangle \approx 0$ to match various second-order correlators, ensuring the equivalence of second-order derivatives of CG and AA free energies with respect to conjugate fields, $\{K_\alpha\}$. In that case, the system of N linear equations, eq 5, would be supplemented by $N(N-1)/2$ equations for $\Delta\langle S_\alpha S_\gamma \rangle$ expressed in terms of various correlators of the third order. Because simple electrolyte solutions are characterized by a relatively small number of observables, $N \lesssim 10^2$, it is computationally feasible to solve such an extended system of (still linear) equations. It is worth noting that discrepancy in higher order correlators is not expected in simple homogeneous systems, such as relatively dilute electrolyte solutions described in this work; however, this issue may become important in heterogeneous systems. In particular, we have found in a related study that to accurately capture the coupling between the dynamics of the DNA chain and the surrounding ionic atmosphere, the latter being strongly inhomogeneous along the macromolecule, it is necessary to ensure that second-order correlators are well reproduced. This will be elaborated elsewhere.

It is also interesting to qualitatively compare the MRG-CG technique, elaborated here, with other approaches to coarse-graining, in particular, where many-body correlations are taken into account. The multiscale coarse-graining (MS-CG) method developed by Voth and co-workers is based on the force matching between atomistic and coarse-grained systems to obtain CG model parameters.[33−35] This technique has been applied to the coarse-graining of mixed lipid bilayers, peptides, and ionic liquids.[36] The major difference between the present and the MS-CG techniques comes from different ways of practically implementing low-order truncations. In that regard, it will be interesting to compare these methods, when applied to various systems, in terms of obtaining unique, transferrable solutions and the accuracy of the CG force field, as compared to the reference atomistic simulations. Yet another physical coarse-graining approach, which takes into account many-body interactions, has been developed by Scheraga and co-workers in the context of developing a CG protein force field.[37] Their approach, however, is very different from our approach, based on a high-temperature cumulant expansion. It would be interesting to apply this technique to coarse-grain electrolyte solutions and compare the obtained accuracy with the MRG-CG results, presented in this work.

In summary, our generalization of Swendsen's method compares favorably with many other commonly used alternative schemes aimed at matching some ad-hoc structural characteristics (see ref 38 and references therein), but not partition functions. The MRG-CG method was recently applied[10] to develop a two-bead double-stranded DNA model, where complicated correlations among polymeric degrees of freedom were taken into account. In the current work, we showed that structural properties of monovalent ionic solutions are accurately captured by a small number of the collective dynamic modes, on which the whole system dynamics is projected to build up the effective Hamiltonian. The optimization scheme for finding Hamiltonian parameters is computationally efficient, as confirmed by ionic RDFs and quick reduction of the free energy difference between AA and CG systems. We estimate the error between AA and CG Hamiltonians to be less than $0.01 k_B T$ per degree of freedom. By showing the close relationship between our technique and the RG theory, we suggest this method might allow one to achieve high accuracy in coarse-graining of many molecular systems. In general, we expect that CG models of various complex biological molecules, including electrostatic and hydration effects, can be built with the present technique. However, atomistic simulations need to be equilibrated; thus, coarse-graining of proteins remains an interesting challenge.

Acknowledgment. This work was supported by the Beckman Young Investigator Award and Petroleum Research Fund Award 47593-G6.

Supporting Information Available: Optimized parameters for the coarse-grained Hamiltonians for NaCl and KCl solutions provided in Tables 1 and 2. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Luger, K.; Hansen, J. C. *Curr. Opin. Struct. Biol.* **2005**, *15*, 188–196.
(2) Schiessel, H. *J. Phys.: Condens. Matter* **2003**, *15*, R699−R774.
(3) Koculi, E.; Hyeon, C.; Thirumalai, D.; Woodson, S. A. *J. Am. Chem. Soc.* **2007**, *129*, 2676–2682.

(4) Savelyev, A.; Papoian, G. A. *J. Am. Chem. Soc.* **2006**, *128*, 14506–14518.
(5) Savelyev, A.; Papoian, G. A. *J. Am. Chem. Soc.* **2007**, *129*, 6060–6061.
(6) Schalch, T.; Duda, S.; Sargent, D. F.; Richmond, T. J. *Nature* **2005**, *436*, 138–141.
(7) Dorigo, B.; Schalch, T.; Kulangara, A.; Duda, S.; Schroeder, R. R.; Richmond, T. J. *Science* **2004**, *306*, 1571–1573.
(8) Knotts, T. A.; Rathore, N.; Schwartz, D.; de Pablo, J. J. *J. Chem. Phys.* **2007**, *126*, 084901.
(9) Mielke, S. P.; Gronbech-Jensen, N.; Benham, C. J. *Phys. Rev. E* **2008**, *77*, 031924.
(10) Savelyev, A.; Papoian, G. A. *Biophys. J.*, (2009), doi: 10.1016/j.bpj.2009.02.067.
(11) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730–3737.
(12) Lyubartsev, A. P.; Laaksonen, A. *J. Chem. Phys.* **1999**, *111*, 11207–11215.
(13) Lyubartsev, A. P. *Eur. Biophys. J.* **2005**, *35*, 53–61.
(14) Zinn-Justin, J. *Quantum Field Theory and Critical Phenomena*; Clarendon Press: Oxford, 2002.
(15) Swendsen, R. H. *Phys. Rev. Lett.* **1979**, *42*, 859–861.
(16) Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
(17) Joung, I. S.; Cheatham, T. E. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
(18) Wang, J.; Cieplak, P.; Kollman, P. *J. Comput. Chem.* **2000**, *21*, 1049–1074.
(19) Savelyev, A.; Papoian, G. A. *Mendeleev Commun.* **2007**, *17*, 97–99.
(20) Savelyev, A.; Papoian, G. A. *J. Phys. Chem. B* **2008**, *112*, 9135–9145.
(21) Chen, A. A.; Pappu, R. V. *J. Phys. Chem. B* **2007**, *111*, 11884–11887.
(22) Chen, A. A.; Pappu, R. V. *J. Phys. Chem. B* **2007**, *111*, 6469–6478.
(23) Auffinger, P.; Cheatham, T. E.; Vaiana, A. C. *J. Chem. Theory Comput.* **2007**, *3*, 1851–1859.
(24) Berendsen, H. J.; Postma, J. P.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.
(25) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327–341.
(26) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
(27) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1–19.
(28) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695–1697.
(29) Fukunaga, H.; Takimoto, J.; Doi, M. *J. Chem. Phys.* **2002**, *116*, 8183–8190.
(30) Barrat, J.-L.; Hansen, J.-P. *Basic Concepts for Simple and Complex Liquids*; Cambridge University Press: New York, 2003.
(31) Lyubartsev, A. P.; Laaksonen, A. *Chem. Phys. Lett.* **2000**, *325*, 15–21.
(32) Robinson, R. A.; Stokes, R. H. *Electrolyte Solutions*; Dover Publications, Inc.: New York, 2002.
(33) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469–2473.
(34) 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.
(35) 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.
(36) Noid, W. G.; Ayton, G. S.; Izvekov, S.; Voth, G. A. In *Coarse-Graining of Condensed Phase and Biomolecular Systems*; Voth, G. A., Ed.; CRC Press: Boca Raton, FL, 2008; Chapter 3, pp 21−40.
(37) Liwo, A.; Czaplewski, C.; Pillardy, J.; Scheraga, H. *J. Chem. Phys.* **2001**, *115*, 2323–2347.
(38) Nielsen, S. O.; Lopez, C. F.; Srinivas, G.; Klein, M. *J. Phys.: Condens. Matter* **2004**, *16*, R481–R512.

JP9005058