CHAPTER 1

Inverse Monte Carlo Methods

Alexander P. Lyubartsev

CONTENTS

1.1 INTRODUCTION

Modeling of many important biomolecular and soft matter systems requires consideration of length and time scales not reachable by atomistic simulations. An evident solution of this problem is introducing simplified models with lower spacial resolution, which have received a common name: *coarse-grained* (CG) models. In CG models, atoms of (macro)molecules are united into CG sites and solvent atoms are often not considered explicitly. This reduces greatly the number of degrees of freedom of the studied system and allows simulations of much larger systems which are not feasible to

simulate at the atomistic level. Studies of models which can be characterized as "coarse-grained" started at the earlier stages of molecular modeling in the 1960s and 1970s (when the term "coarse-grained" was not used at all). For example, a primitive model of electrolytes represented hydrated ions as charged spheres in a dielectric media (Vorontsov-Velyaminov and Elyashevich, 1966; Card and Valleau, 1970), and a simple freely jointed model of a polymer chain (Gottlieb and Bird, 1976) was used to model polymers in solution. Simple rod-like particles with two or three interaction sites were used to describe lipids in lipid bilayers and other self-assembled structures (Noguchi and Takasu, 2001; Farago, 2003; Brannigan and Brown, 2004). Such models were designed to illustrate general physical behavior of the studied systems. In order to relate such models to real physical systems, one needs to find model parameters, such as effective (hydrated) ion radius in the primitive electrolyte model, which can be done empirically by fitting to known experimental data.

Development of more advanced CG models for studies of specific molecular structures including lipids, proteins, DNA, polymers, etc., sets higher requirements for the choice of interaction potentials describing interactions in such systems. In the so-called *top-down* methodology, one is trying to parametrize the model to reproduce experimentally measurable macroscopic properties of the system. One of the most popular modern CG models of this kind is described in terms of the MARTINI force field (Marrink et al., 2004, 2007), which represents groups of about four heavy atoms by CG sites. The MARTINI force field, originally developed for lipids (Marrink et al., 2007), was later extended to proteins (Monticelli et al., 2008; de Jong et al., 2012), carbohydrates (Lopez et al., 2009), and some other types of molecules (de Jong et al., 2012; Marrink and Tieleman, 2013). Within the MARTINI force field model, CG sites interact by the electrostatic and Lennard-Jones potentials with parameters fitted to reproduce experimental partitioning data between polar and apolar media. This functional form of the force field is convenient as it coincides with that for the atomistic simulations and is implemented in all major simulation packages, but it may be also a source of problems with overstructuring of molecular coordination and with consistent description of multicomponent systems, which can in principle be solved by using softer (than Lennard-Jones) CG potentials (Marrink and Tieleman, 2013). A complicating circumstance in this respect is that (differently from atomistic models) even a functional form of the effective interaction potentials is in many cases not known *a priori*.

In the alternative *bottom-up* methodology, effective CG potentials are derived from atomistic simulations. Atomistic force fields reflect the real chemical structure of the studied system and they are generally more established than CG force fields. Furthermore, the *bottom-up* methodology can use as a starting point an even deeper, quantum-chemical level of modeling. Within the *bottom-up* methodology, a CG force field is parameterized to fit some important physical properties that result from a high-resolution (atomistic) simulation. Several *bottom-up* approaches to parametrize CG force fields have been formulated recently. Within the force-matching approach (Ercolessi and Adams, 1994; Izvekov et al., 2004), also called multiscale coarse-graining (Izvekov and Voth, 2005; Ayton et al., 2010)), the CG potential is built in a way to provide the best possible fit to the forces acting on CG sites in the atomistic simulation. Within the inverse Monte Carlo (IMC) technique (Lyubartsev and Laaksonen, 1995, 2004) and similar renormalization group coarse-graining (Savelyev and Papoian, 2009b), as well as in the related iterative Boltzmann inversion (IBI) method (Soper, 1996; Reith et al., 2003), the target property is the radial distribution functions (RDFs) as well as internal structural properties of molecules such as distributions of bond lengths, covalent angles, and torsion angles. For this reason, methods based on IMC or IBI techniques are called structure-based coarse-graining. In the relative entropy minimization method (Shell, 2008; Chaimovich and Shell, 2011), the CG potential is defined by a condition to provide minimum entropy change between the atomistic and CG system, which is also equivalent to minimizing information loss in the coarse-graining process. In the conditional work approach (Brini et al., 2011), the effective potentials between CG sites are obtained as free energy (potentials of mean force) between the corresponding atom groups determined in a thermodynamic cycle. More details of different *bottom-up* multiscale methodologies and their analysis can be found in a recent review (Brini et al., 2013).

This chapter is devoted to the systematic coarse-graining methodology based on the IMC method. Here, we discuss the term "inverse Monte Carlo" only in application to multiscale coarse-graining, while there exists a more general definition of "IMC" as a solution of any inverse problem by any type of Monte Carlo (MC) technique (Dunn and Shultis, 2012). In Section 1.2, the theoretical background, practical algorithms, problems, and possible limitations of the approach are considered. Section 1.3 considers various applications of the IMC methodology to the ionic systems, lipid assemblies, DNA, and a few other systems. Finally, perspectives and possible limitations of the methodology are discussed.

1.2 MULTISCALE SIMULATIONS USING IMC

1.2.1 Theoretical Background

Generally, a problem of going from a high-resolution (atomistic) description to a CG one can be formulated as follows. Assume that at the high-resolution level, the system is described by a Hamiltonian (potential energy) $H(\{r_i\})$, where $\{r_i\}, i = 1, ..., n$ are the coordinates of atoms. The potential energy function represents typically the atomistic force field, but in the case of *ab-initio* modeling it can represent the energy surface obtained within any quantum-chemical computation method. A coarse-graining is described in terms of mapping of atomistic coordinates (degrees of freedom) $\{r_i\}_{i=1,...,n}$ to CG coordinates $\{R_j\}_{j=1,...,N}$, which is mathematically expressed in terms of mapping functions $R_j = R_{\text{map}(j)}(\{r_i\})$. Often CG coordinates R_j are center of masses of groups of atoms united in CG site j, but they may just coincide with coordinates of some selected atoms, or other choices can be made. The original Hamiltonian $H(\{r_i\})$ defines all properties of the high-resolution system, and through the mapping functions, all properties of the CG system. The task is to define effective interaction potential for CG sites, which provide the same properties for the CG system as the properties which follows from the CG mapping of the high-resolution system.

If only structural properties are a matter of interest, an exact formal solution of the above formulated problem can be written in terms of N-body potential of mean force. The N-body potential of mean force is obtained by inclusion of the CG degrees of freedom into the partition function of the original high-resolution system and subsequent integration over atomistic coordinates:

$$\begin{aligned} Z &= \int \prod_{i=1}^{n} \mathrm{d}r_i \exp(-\beta H(\{r_i\}) = \\ &= \int \prod_{i=1}^{n} \mathrm{d}r_i \int \prod_{j=1}^{N} \mathrm{d}R_j \delta(R_j - R_{\text{map}(j)}(\{r_i\})) \exp(-\beta H(\{r_i\}) \\ &= \int \prod_{j=1}^{N} \mathrm{d}R_j \exp(-\beta H_{\text{CG}}(\{R_j\}) \end{aligned} \tag{1.1}$$

with $\beta = 1/k_B T$ and N-body potential of mean force (CG Hamiltonian) $H_{\text{CG}}(\{R_j\})$ defined by

$$H_{\text{CG}}(\{R_j\}) = -\frac{1}{\beta} \ln \int \prod_{i=1}^{n} \mathrm{d}r_i \delta(R_j - R_{\text{map}(j)}(\{r_i\})) \exp(-\beta H(\{r_i\}) \tag{1.2}$$

A CG system with a Hamiltonian (Equation 1.2) has the same structural properties (canonical averages) for the CG degrees of freedom as the underlying high-resolution system with the original Hamiltonian. Since the CG and high-resolution systems have the same partition function, the thermodynamic properties (average energy, free energies, pressure) can also be reconstructed. The important point is, however, that the CG Hamiltonian depends on the thermodynamic conditions (temperature, concentration) and these dependences need to be taken into account while obtaining thermodynamic properties by derivation of the partition functions by the thermodynamics parameters. Reconstruction of the correct dynamics in the CG system is a more challenging task. In addition to renormalization of the Hamiltonian according to Equation 1.2, the dynamic equations of motion need to be changed to the generalized Langevin equation with a memory function (Romiszowski and Yaris, 1991). Some practical approaches to dynamic coarse-graining within dissipative particle dynamics can be found in the literature (Eriksson et al., 2008; Hijon et al., 2010).

As discussed above, the N-body potential of mean force (Equation 1.2) provides an exact solution of the coarse-graining problem; however, in practical terms simulations involving an N-body potential are infeasible. A common way to proceed is to approximate it by a more convenient expression, for example, by a sum of distance-dependent pair potentials:

$$H_{\mathrm{CG}}(R_1, ..., R_N) \approx \sum_{i<j} V_{ij}^{\mathrm{eff}}(R_{ij}) \tag{1.3}$$

with $R_{ij} = |R_i - R_j|$. It is, however, not necessary to be limited by pair potentials; any practically usable expression can be given in Equation 1.3, for example, angle or torsion potential terms for macromolecular CG models, or some other simple forms expressing three or four body interactions. From the computational point of view, we are first interested in pair-wise approximations: the very aim of coarse-graining is computational speed-up, and extensive use of many-body potentials would greatly hamper this goal.

The task of building a CG force field can be thus reformulated to find an "as best as possible" approximation according to Equation 1.3. Definitions of what "the best approximation" can be, however, differ. Usually one determines a set of target properties that one wish the CG model to keep. These properties can be either of a microscopical character, as forces or instantaneous energies, or canonical averages as RDFs, average energies, or pressure. For example, minimizing the force difference coming from both

sides of Equation 1.3 (weighted with the Boltzmann factor) is equivalent to the force-matching method (Ercolessi and Adams, 1994; Izvekov et al., 2004; Izvekov and Voth, 2005). Within the structure-based coarse-graining approach, the target is various structural properties, such as RDFs as well as distributions of internal degrees of freedom in complex molecules. In principle, other properties of interest or any combination of them can be used for parameterization of effective potentials.

1.2.2 IMC: Newton Inversion

The task of determining computationally feasible CG potentials can thus be formulated in the following way (Lyubartsev et al., 2010). Assume our effective potentials $H(\{r_i\})$ (we now remove index "CG" from the notations) are determined by a (finite) set of parameters $\{\lambda_\gamma\}$, and the set of target properties (which we know from the atomistic simulations) is $\{A_\alpha^{\text{ref}}\}$. We assume here that the number of potential parameters $\gamma = 1, ..., M$ is equal to the number of target properties $\alpha = 1, ..., M$. If we know the set of $\{\lambda_\gamma\}$, we can always compute average properties $\{\langle A_\alpha\rangle\}$ in direct molecular dynamics (MD) or MC simulations. The inverse problem, finding parameters $\{\lambda_\gamma\}$ from averages $\{\langle A_\alpha\rangle\}$, is less trivial. We can consider the relationship between $\{\lambda_\gamma\}$ and $\{\langle A_\alpha\rangle\}$ as a nonlinear multidimensional equation, and use the Newton inversion method (known also as the Newton–Raphson method) to solve it iteratively.

The method is based on the expression relating small changes of the potential parameters λ_γ and changes of the canonical energies $\langle A_\alpha\rangle$ caused by these changes of the potential parameters:

$$\Delta\langle A_\alpha\rangle = \sum_\gamma \frac{\partial\langle A_\alpha\rangle}{\partial\lambda_\gamma}\Delta\lambda_\gamma + O(\Delta\lambda^2) \tag{1.4}$$

The matrix of derivatives $\frac{\partial\langle A_\alpha\rangle}{\partial\lambda_\gamma}$ (Jacobian) can, by the use of statistical mechanics expressions for averages in canonical ensemble, be presented in the following form (Lyubartsev et al., 2010; Wang et al., 2013a):

$$\begin{aligned}\frac{\partial\langle A_\alpha\rangle}{\partial\lambda_\gamma} &= \frac{\partial}{\partial\lambda_\gamma}\frac{\int\prod_{i=1}^N \mathrm{d}r_i A_\alpha(\{\lambda_\gamma\},\{r_i\})\exp(-\beta H(\{r_i\}))}{\int\prod_{j=1}^N \mathrm{d}r_j \exp(-\beta H(\{r_j\}))} = \\ &= \left\langle\frac{\partial A_\alpha}{\partial\lambda_\gamma}\right\rangle - \beta\left(\left\langle\frac{\partial H}{\partial\lambda_\gamma}A_\alpha\right\rangle - \left\langle\frac{\partial H}{\partial\lambda_\gamma}\right\rangle\langle A_\alpha\rangle\right)\end{aligned} \tag{1.5}$$

Now the Jacobian is expressed in terms of canonical averages which all can be evaluated by running direct simulation with a Hamiltonian defined by a given set of parameters $\{\lambda_\gamma\}$. Equations 1.4 and 1.5 can be used to solve the inverse problem iteratively. One starts from some initial potential determined by a trial set of parameters $\{\lambda_\gamma^{(0)}\}$, runs a simulation, and computes the deviation of computed values $\{\langle A_\alpha^{(0)}\rangle\}$ from the target values A_α^{ref}:

$$\Delta\langle A_\alpha\rangle = A_\alpha^{\text{ref}} - \{\langle A_\alpha^{(0)}\rangle\} \tag{1.6}$$

We also compute a matrix of derivatives (Equation 1.5). Then, the system of linear equations (Equation 1.4) is solved neglecting second-order corrections, resulting in corrected values of parameters λ_γ:

$$\lambda_\gamma^{(n+1)} = \lambda_\gamma^{(n)} + \Delta\lambda_\gamma \tag{1.7}$$

The procedure is repeated until convergence is reached. If initial approximation $\{\lambda_\gamma^{(0)}\}$ is poor, some regularization of the iterative procedure might be necessary, in which the difference in Equation 1.6 is multiplied by some factor between 0 and 1.

While the Newton inversion procedure was depicted above for the transition from the atomistic to the CG level, it can work in the same way for the connection between *ab initio* and atomistic levels, with the only difference that the quantum-mechanical energy surface is used instead on the N-body potential of mean force in Equation 1.3.

The Newton inversion algorithm can be straightforwardly implemented if the number of potential parameters is equal to the number of target properties. If the number of properties exceeds the number of potential parameters, the problem can be solved in the variational sense, by finding a set of $\{\lambda_\gamma\}$ which provides the least possible deviation of the computed properties from the target values. Optimization using the same equations (Equation 1.5) can in this case be carried out according to the Gauss–Newton algorithm. This approach, under the name the force balance method, has been recently implemented for parametrization of a water model from ab-initio and experimental data (Wang et al., 2013a).

1.2.3 IMC: Reconstruction of Pair Potentials from RDFs

An important case of the general approach described above is when parameters $\{\lambda_\gamma\}$ are values of the pair potential in a regular set of points covering the whole range of distances (i.e., the potential is given in a table form), and

the target properties are values of the RDF in the same set of points. Then, the inverse problem is reformulated as finding the pair interaction potential which reconstructs the given RDF. For a multicomponent case, a set of pair interaction potentials is reconstructed from a set of RDFs between the same CG site types. Even intramolecular interactions such as bond, angular, and torsion potentials can be linked to the distribution of the corresponding bond lengths, angles, and torsions and included in the inversion procedure. The equations given in the previous section become equivalent in this case to the IMC algorithm introduced previously (Lyubartsev and Laaksonen, 1995, 2004) and described below.

We consider a class of systems which is described by a Hamiltonian (potential energy) in the following form:

$$H = \sum_{\alpha} V_{\alpha} S_{\alpha} \tag{1.8}$$

The Hamiltonian of any system with pair interaction can be presented in such form when the pair potential is given by a set of tabulated values as a step-wise function of distance:

$$\tilde{V}(r) = V(r_{\alpha}) \equiv V_{\alpha}$$

for

$$r_{\alpha} - \frac{1}{2M} < r < r_{\alpha} + \frac{1}{2M}; \quad r_{\alpha} = (\alpha - 0.5) r_{\text{cut}}/M; \;\; \alpha = 1, ..., M \tag{1.9}$$

where r_{cut} is some cutoff distance and M is the number of grid points within the interval $[0, r_{\text{cut}}]$. The S_{α} values represent the number of particles pairs with the distances found inside α-slice. Evidently, S_{α} is an estimator of the RDF: $\langle S_{\alpha} \rangle = 4\pi r^2 \rho(r) N^2/(2V)$. Thus, the inverse problem is formulated now as finding the values of the interaction potential in grid points V_{α} from RDFs expressed in averages $\langle S_{\alpha} \rangle$.

With these notations, the Jacobian matrix of Newton inversion (Equation 1.5) becomes

$$\frac{\partial \langle S_{\alpha} \rangle}{\partial V_{\gamma}} = -\beta \left(\langle S_{\alpha} S_{\gamma} \rangle - \langle S_{\alpha} \rangle \langle S_{\gamma} \rangle \right) \tag{1.10}$$

Here, indexes α, γ are running over all interaction types (nonbonded, bonded, angular), within each interaction type over all CG types of sites, and for each set of CG types, over the relevant range of distances (or angles). In all cases, average values of $\langle S_{\alpha} \rangle$ and cross-correlation terms $\langle S_{\alpha} S_{\gamma} \rangle$ can

be acquired from a simulation of the CG system. In the inverse procedure, one starts from a trial set of potentials (in practical simulations, one can start either from zero or from mean force potentials), runs an MC simulation, computes RDFs expressed in terms of average values of $\langle S_\alpha \rangle$ as well as cross-correlation terms according to Equation 1.10, solves a system of linear equations (Equation 1.4) (with substitution $\lambda_\alpha \rightarrow V_\alpha$ and $A_\alpha \rightarrow S_\alpha$), obtains new corrections to the interaction potential ΔV_α, and repeats the procedure until convergence.

Another approach to invert RDFs as well as bond or angle distributions was described earlier by Schommers (1983), reintroduced by Soper under the name empirical potential structure refinement method (Soper, 1996), and become most known as the iterative Boltzmann inversion Reith et al. (2003). As in the IMC method, one starts a simulation with some trial values of the potential $V_\alpha^{(0)}$, and corrects at each iteration the potential according to

$$V_\alpha^{(i+1)} = V_\alpha^{(i)} + k_B T \ln \frac{\langle S_\alpha^{(i)} \rangle}{S_\alpha^{\text{ref}}} \tag{1.11}$$

Correction of potential according to Equation 1.11 is straightforward to implement, and such an approach was used in a number of studies (Shelley et al., 2001; Reith et al., 2003; Harmandaris et al., 2006; Carbone et al., 2008; Wang and Deserno, 2010). In the IBI approach, correction to the potential is determined only by the value of the same distribution function at the same distance point, which is why the IBI approach faces convergence problem in the multicomponent case (Hess et al., 2006a) where different RDFs can be strongly interconnected. In practical calculation of CG potentials by RDF inversion, it might be instructive to start the iterative process using an IBI approach, which brings the system RDFs closer to the reference values, and then switch to IMC, which takes into account correlations between different distribution functions and provides better convergence when the RDFs become close to the reference functions.

It is known that for the relationship between pair potential and RDF, solution of the inverse problem is unique with the precision of an additive constant to the potential. This was proven previously by Henderson for the monocomponent case (Henderson, 1974), and generalized later for a multicomponent case and intramolecular interactions (Rudzinski and Noid, 2011). On the other hand, the inversion problem for the RDF–pair potential relation is often ill-defined; different potentials may in some cases produce RDFs which are very close to each other (Soper, 1996). For this

reason, the IMC and IBI approaches can produce different results for effective potentials while having RDFs undistinguishable by eye on a graph (Rühle et al., 2009; Wang et al., 2013b). Still, even a small difference in RDF, especially at large distances, may be of importance in correct reproduction of the Kirkwood–Buff integral (which is determined by expression $\int 4\pi r^2(g(r) - 1)\mathrm{d}r$), which is important for consistent description of thermodynamics of mixtures (Mukherji et al., 2012). Within the IMC approach, the target function is proportional to $r^2 g(r)$, which is why the Kirkwood–Buff integrals are well reproduced in the CG simulations, while the IBI method is less sensitive to the behavior of RDF at large distances, and may need corrections in order to reproduce Kirkwood–Buff integrals accurately (Ganguli et al., 2012).

1.2.4 Software

The IMC method is currently implemented in two open source software packages: versatile object-oriented toolkit for coarse-graining applications (VOTCA) (Rühle et al., 2009) and MagiC (Mirzoev and Lyubartsev, 2013). The VOTCA package (which also implements force-matching and IBI methods) uses GROMACS (Lindahl et al., 2001) as a sampling engine for computations of necessary canonical averages at each iteration of the inverse procedure. It analyzes the trajectory obtained by GROMACS with trial potentials, computes necessary averages, and computes the updated tabulated potentials to be used at the next iteration. A problem related to the use of MD to sample system configurations during the IMC procedure is that standard MD software such as GROMACS has certain requirements to the smoothness of the used tabulated potentials while in the IMC procedure the potentials may change unpredictably during the optimization procedure and thus be a source of instability.

The MagiC package was developed specially to implement systematic structure-based coarse-graining of arbitrary molecular models using the IMC or IBI methodology. As input, MagiC uses atomistic trajectories generated by any simulation software, from which it computes necessary reference RDFs between CG sites as well as bond and angle distributions. MagiC has its own MC multithread sampling engine which can run simultaneously many copies of the simulated system on each available processor/core. Simulations are run for a trial set of CG potentials, which may include nonbonded interactions between all CG site types, as well as bonded and angle intramolecular interactions. The electrostatic interactions are taken out of the inversion scheme and are treated by the conventional Ewald summation

method (Allen and Tildesley, 1987). The program evaluates the canonical averages necessary for RDF inversion (Equation 1.10) by averaging results generated from all the threads and computes updated potentials which are distributed back to all the threads. The use of multithread methodology improves the quality of sampling, resulting in better stability and faster convergence of the iteration procedure. Recently, MagiC has been used for computations of effective potentials for CG models of lipids (Mirzoev and Lyubartsev, 2014), CG DNA and solvent-mediated DNA–ion interactions (Korolev et al., 2014; Naomé et al., 2014), and a CG model of an ionic liquid (Wang et al., 2013b).

1.3 APPLICATIONS OF THE IMC

1.3.1 Simple Electrolytes

As a first example of application of the IMC methodology, we consider computations of effective solvent-mediated potentials of ions in aqueous solution (Lyubartsev and Laaksonen, 1995, 1997). We recapitulate this study here because of its simplicity and instructive character. One of the typical approximations used in the description of macromolecules and particularly polyelectrolytes is substituting of solvent molecules by a continuum media. For example, in the continuum (called also primitive) electrolyte model, ions in water are substituted by charged spheres moving in dielectric media with a proper dielectric constant. Evidently, this is a serious simplification at a small (a few Å) distance between the ions where it is impossible to define a dielectric constant in a consistent manner. Moreover, an ion radius (in terms of hard sphere, or softer repulsive r^{-12} or r^{-9} potential) within the continuum electrolyte model is an adjustable parameter without clear physical meaning.

A better model of effective ion–ion interactions in aqueous solution must take into account the solvation structure of water around the ions. Practically, effective solvent-mediated ion–ion potentials may be constructed by the IMC method from ion–ion RDFs, generated in high-quality atomistic MD simulations of ions in water. This approach has been already implemented in the first paper describing the IMC technique (Lyubartsev and Laaksonen, 1995). In subsequent publications (Lyubartsev and Laaksonen, 1997; Lyubartsev and Marcelja, 2002; Mirzoev and Lyubartsev, 2011), the effective-solvent mediated potentials for NaCl aqueous solution have been calculated with greater precision as well as for a number of concentrations and temperatures.

An example of computation of solvent-mediated ion–ion potentials for NaCl aqueous solution is illustrated in Figure 1.1. The underlying MD simulations have been performed for the flexible simple point charge (SPC)

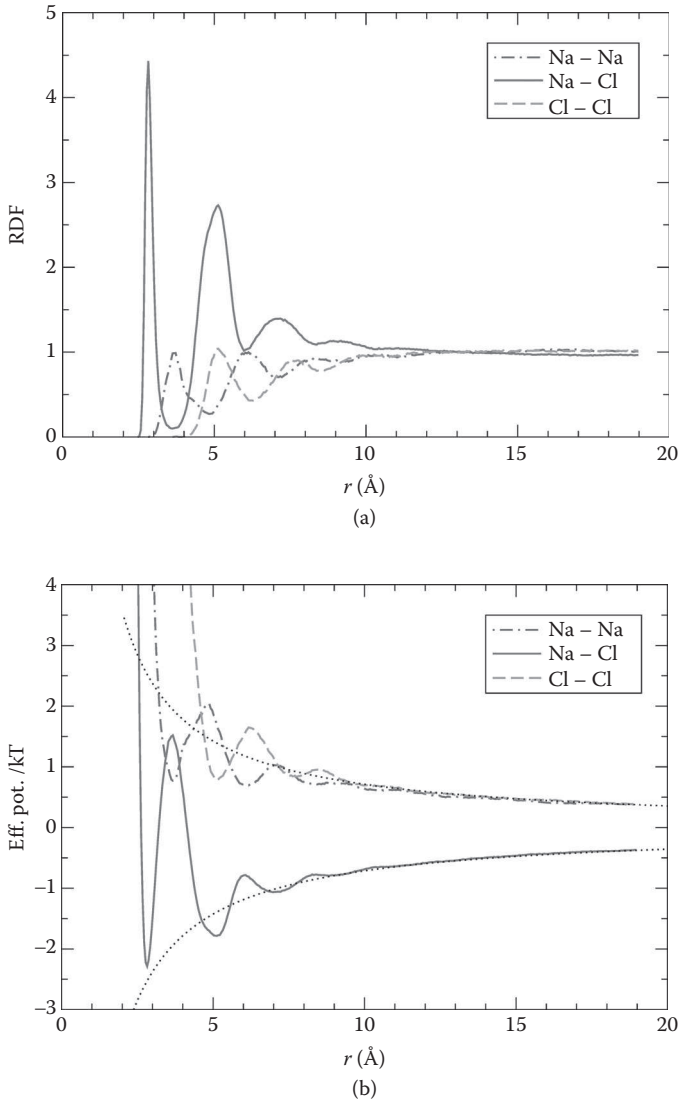

FIGURE 1.1 (a) and (b) Radial distribution functions (RDFs) between Na^+ and Cl^- ions in water computed from atomistic simulations at temperature 298 K and ion concentration 0.5 M (Lyubartsev and Laaksonen, 1997), and corresponding effective potentials derived from these RDFs by the inverse Monte Carlo (IMC) method. Dotted lines show Coulombic potential at dielectric permittivity 80. (Compiled from data in Lyubartsev, A. P. and A. Laaksonen, *Phys. Rev. E*, 55, 5689–5696, 1997.)

water model (Toukan and Rahman, 1985) and Smith–Dang parameters for Na^+ and Cl^- ions (Smith and Dang, 1994) resulting in the ion–ion RDFs shown in Figure 1.1a. They were fed into an IMC procedure resulting in effective potentials between the ions shown in Figure 1.1b. The effective potentials make one to two oscillations, thereby reflecting the molecular nature of the solvent, and then finally approach the primitive model potential with dielectric constant close to 80. With distances more than 10 Å, the effective potentials almost perfectly coincide with the Coulombic potential. These characteristic features of ion–ion solvent-mediated potentials were also observed in other works (Hess et al., 2006a,b; Savelyev and Papoian, 2009a).

An important issue of CG effective potentials is their state-point dependence, which originates in integration over "nonimportant" degrees of freedom (see Equation 1.1). Studies of solvent-mediated ion–ion potentials showed that they do depend both on the ion concentration (Lyubartsev and Laaksonen, 1997) and on the temperature (Mirzoev and Lyubartsev, 2011). It was, however, shown that most of this dependence can be taken care of by introducing concentration (Hess et al., 2006b) and temperature-dependent (Mirzoev and Lyubartsev, 2011) dielectric permittivity. This can be done by considering the short-range part of the effective potential,

$$V_{\mathrm{sh}}(r) = V_{\mathrm{tot}}(r) - \frac{q_i q_j}{4\pi\epsilon_0 \epsilon r} \tag{1.12}$$

and optimizing dielectric permittivity ϵ by the requirement that the short-range part of all three ion–ion potentials be most close to zero at distances outside 10 Å (or a similar cutoff distance). Besides definition of the transferable short-range part of the effective potential, this approach provides also an alternative definition of effective dielectric permittivity of a solvent. It was demonstrated in a paper (Mirzoev and Lyubartsev, 2011) that such a definition of the dielectric constant is consistent with conventional computations of dielectric permittivity from the dipole moment fluctuations, as well as with experimental data.

1.3.2 CG Lipid Model

Simulations of lipid membranes and other self-assembled structures have attracted much attention during the last decade due to the fact that lipid membranes form the outer shells of living cells (Lyubartsev and Rabinovich, 2011). However, atomistic simulation of even a relatively small piece of

membrane consisting of a few hundred lipids and surrounding water is a computational challenge, while many actual biophysical problems, such as studies of inhomogeneous membrane mixtures, membrane mechanical properties, association and partitioning of polypeptides, nanoparticles, other membrane bound compounds, etc., require consideration of substantially larger membrane fragments. For investigation of all these phenomena in molecular simulations, a coarse-grain level of modeling provides practically the only possible choice.

A large variety of various CG models of lipids differing by the level of detail and the way of defining CG potentials have been reported in the last two decades (Drouffe et al., 1991; Goetz and Lipowsky, 1998; Shelley et al., 2001; Murtola et al., 2004; Marrink et al., 2004; Lyubartsev, 2005; Cooke and Deserno, 2005; Izvekov and Voth, 2005; see also the reviews of Pandit and Scott, 2009; Shinoda et al., 2012). In several cases, interaction potentials were determined within a systematic *bottom-up* approach from atomistic simulations, using IBI (Shelley et al., 2001), force matching (Izvekov and Voth, 2005), and IMC (Lyubartsev, 2005). In one work (Lyubartsev, 2005), a dimyristoylphosphatidylcholine (DMPC) lipid was CG to a 10-site model (see Figure 1.2). The CG site–site RDFs as well as CG bond-length distributions were computed from atomistic MD simulations of 16

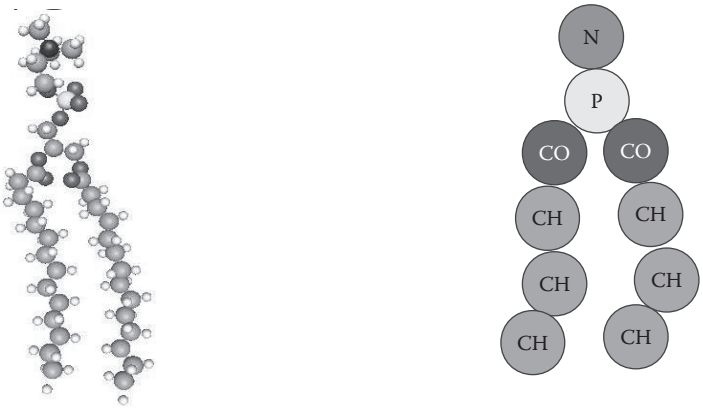

FIGURE 1.2 Atomistic (a) and coarse-grained (b) lipid models. (With kind permission from Springer Science+Business Media: *Eur. Biophys. J.,* Multiscale modeling of lipids and lipid bilayers, 35, 53, 2005, Lyubartsev, A. P.)

lipids dissolved in water and described by the CHARMM27 force field, and were inverted simultaneously within the IMC procedure, resulting in 10 nonbonded potentials between four different types of CG sites and four bonded potentials. The resulting CG model has showed the ability to reproduce a bilayer structure consistent with atomistic simulations, and to describe self-assembly of lipids into bicell as well as the formation of spherical vesicle structures (Figure 1.3).

In subsequent development (Mirzoev and Lyubartsev, 2014), the effective potentials derived in work (Lyubartsev, 2005) were reparameterized after recomputation of CG site–site RDFs according to recent modification of the CHARMM27 force field described in Högberg et al. (2008). This modification of the CHARMM force field has been done with a primary aim to improve agreement with experiments for atomistic simulations of the lipid bilayer, and to reproduce correctly average area per lipid

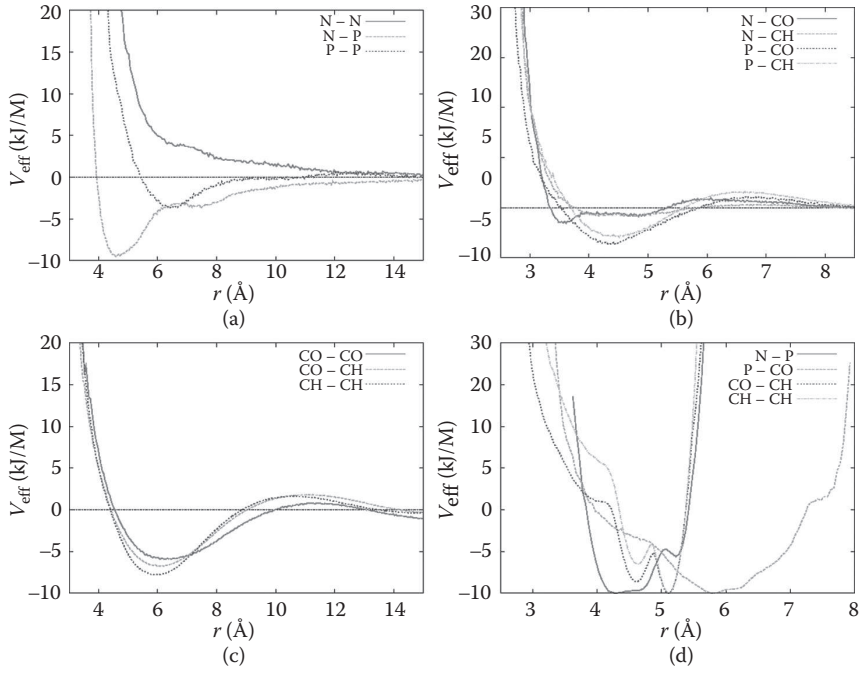

FIGURE 1.3 Potentials for various atom pairs are shown in panels (a)-(d). Effective potentials for the coarse-grained (CG) lipid model displayed in Figure 1.2 computed by the IMC method from RDF and bond-length distributions determined in atomistic simulations. (With kind permission from Springer Science+Business Media: *Eur. Biophys. J.,* Multiscale modeling of lipids and lipid bilayers, 35, 53, 2005, Lyubartsev, A. P.)

in particular. Atomistic simulations of DMPC lipid mixtures with water at different molar ratios were run up to 400 ns in order to ensure reliable converged RDFs. The set of intramolecular potentials of previous work (Lyubartsev, 2005) was complemented by angular potentials determined from distribution of relevant angles between CG sites. It was found that effective potentials obtained from atomistic simulations carried out at different concentrations, and properties of lipid bilayers simulated using these potentials, show nonnegligible concentration dependence. Thus, potentials based on low lipid concentration overestimate the effective hydrophobic attraction of the lipid tails, which favors a more gel-like and more ordered structure of the bilayer. The potentials based on higher lipid concentration in the atomistic simulations provide more fluid-like structure with larger area per lipid. The best agreement with reference data as well with experiment was achieved with a set of potentials derived from atomistic simulations at 1:30 lipid:water molar ratio, which also provides full saturating hydration of the DMPC headgroup in the bilayer. Comparison of some characteristics of the DMPC lipid bilayer obtained in atomistic and CG lipid models is given in Table 1.1.

Despite a certain degree of the state-point dependency of the effective potentials, all the derived potentials obtained in conditions of unordered lipid–water mixture provided a stable bilayer structure with correct partitioning of different lipid groups across the bilayer as well as with acceptable values of the average lipid area, orientational tail ordering, and

TABLE 1.1 Comparison of Some Properties of Lipid Bilayer Obtained in Atomistic and CG Simulations for Four CG Models Derived by IMC from Atomistic Simulations at Different Concentrations.

Model	A (Å^2)	K_A (mN/m)	D (Å)	$S1$	$S2$
CG 1:100	49.1	910	38.8	0.73	0.70
CG 1:50	50.1	910	40.2	0.77	0.77
CG 1:30	59.7	370	38.0	0.57	0.52
CG 1:20	65.7	190	37.6	0.52	0.45
Atomistic	60.0	250	39.6	0.58	0.52

Source: Mirzoev, A. and A. P. Lyubartsev, *J. Comput. Chem.*, 35, 1208–1218, 2014.

Note: A, average area per lipid; K_A, compressibility; D, membrane thickness determined from the distribution maxima of N-sites; $S1$ and $S2$, order parameters defined by vectors connecting CO and the second CH sites, and the first and the third CH sites of lipid tails, respectively. For details, see Mirzoev and Lyubartsev (2014).

compressibility. They also demonstrated the ability of CG lipids to self-assemble in bilayer, bicell, or vesicle structures depending on simulation condition as is, e.g., displayed in Figure 1.4. This behavior of the CG lipid model, derived using an IMC approach, was reached without use of any additional information except that which was available from atomistic simulations.

1.3.3 CG DNA Models

The DNA molecule was studied by computer simulations methods for many decades. DNA is a strongly charged polyelectrolyte, and understanding of many properties, including DNA packing in the cell nuclei, requires proper description of both long-range electrostatic interactions

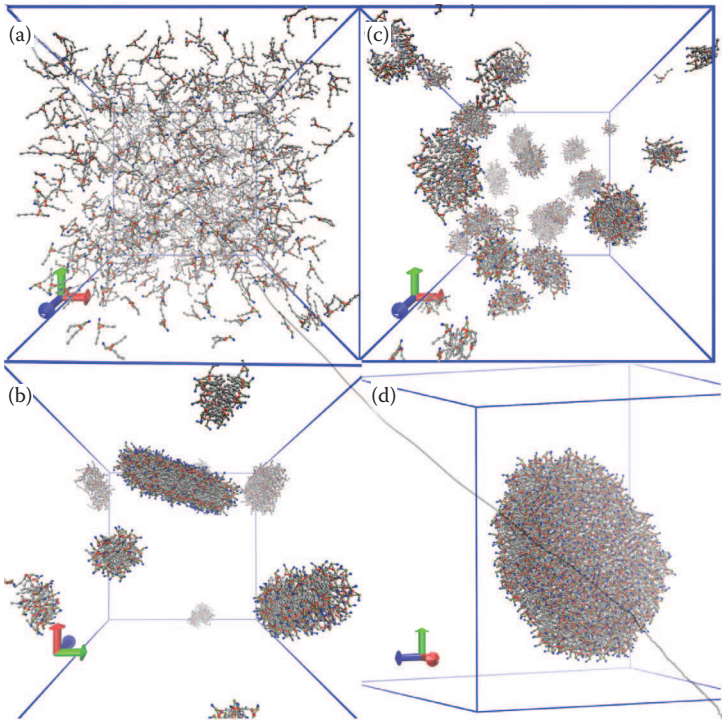

FIGURE 1.4 Trajectory snapshots are shown in panels (a)-(d), as the CG simulation is evolved in time. Formation of a bicell from initially random distribution of CG lipids observed in Langevin molecular dynamics simulations of work. (From Mirzoev, A. and A. P. Lyubartsev. Systematic implicit solvent coarse graining of dimyristoylphosphatidylcholine lipids. *J. Comput. Chem.* 2014. 35. 1208–1218. Copyright Wiley-VCH Verlag GmbH & Co. KGaA. Reproduced with permission.)

and chemical (atomistic) details of DNA interactions with surrounding molecules including ions and histones (Korolev et al., 2012). This necessitates the use of multiscale approaches when continuum solvent CG models are robust enough to catch the atomistic chemically specific details and effect of water hydration.

In one paper (Lyubartsev and Laaksonen, 1999), the IMC approach was used to derive effective solvent-mediated potentials between alkali ions and sites of CG DNA model where each nucleotide was presented as three sites representing phosphate, base, and sugar. The RDFs between ions and these sites used as an input to the IMC procedure were computed in atomistic simulations of a DNA fragment in ionic solution. The DNA periodic fragment used in this study was fixed, and the obtained ion–DNA solvent-mediated potentials were used to model binding affinity of different alkali ions to DNA.

A flexible CG DNA model, presented as two connected helical chains representing DNA phosphate groups, was considered in a series of works from the Papoian group (Savelyev and Papoian, 2009b, 2010; Savelyev et al., 2011). Both solvent-mediated ion–DNA and internal DNA interactions were determined from the structural information obtained in atomistic simulations of DNA in aqueous ion solution using the molecular renormalization group approach, which is in most details equivalent to the IMC method. This model was used primarily to study DNA persistence length at different ion conditions. A similar DNA–ion model has been considered in a recent paper (Naomé et al., 2014) with ion–CG DNA and internal DNA interactions computed by the combined iterative Boltzmann/IMC method from atomistic simulations of a DNA oligonucleotide.

A CG DNA model considered in a recent study (Korolev et al., 2014) included, along with CG sites representing phosphate groups, space filling sites located along DNA axis with each such site representing two base pairs (see Figure 8.5a). The internal structure of DNA was maintained by three bond and three angular potentials between neighboring CG sites. The reference atomistic simulations were carried out for four DNA nucleotides which gave the possibility to extract effective DNA–DNA site–site interactions which are intended to be used in perspective studies of nucleosome folding. It was also found in a paper (Korolev et al., 2014) that the internal DNA potentials can be well fitted to harmonic potentials, which gives the possibility to describe the internal CG DNA interactions in terms of a standard molecular mechanics force field implemented in most MD software.

In another recent work (Biase et al., 2014), the IMC approach was used to parametrize effective potentials for the CG model of a single-strand DNA with CG sites representing phosphate, sugars, and four distinguishable types of DNA bases. The obtained model was used to describe longtime dynamics of a single-strand DNA in a nanopore.

1.3.4 Other Systems

Besides computations of effective potentials for CG models, there were a few attempts to use the IMC approach to derive atomistic potentials from *ab-initio* (Car–Parrinello type) simulations. In on paper (Lyubartsev and Laaksonen, 2000), the RDFs between the O and H atoms of water, obtained in Car–Parrinello molecular dynamics (CPMD) (Car and Parrinello, 1985) simulations of 32 water molecules, were inverted to produce atomistic interaction potentials which turned out to be very similar to potentials of conventional SPC and TIP3P water models. Later unpublished studies have showed, however, that a "ab-initio" derived water model could not compete with the traditional empirically parametrized water models mainly because of deficiencies in the density functional theory (DFT) functionals used in CPMD simulations. In another work (Lyubartsev et al., 2001), CPMD simulations of a Li^+ ion in water were used to extract the interaction potential between the Li^+ ion and water oxygen. The nonelectrostatic part of the Li^+–O potential was found to be well approximated by an exponential function like the one used in the Buckingham potential (see Figure 1.5). The *ab-initio* derived model of Li^+ ion interaction was further used in MD simulations of a LiCl solution (Egorov et al., 2003).

The IMC method in two dimensions was used in a work (Murtola et al., 2007) to study domain formation in a lipid bilayer containing cholesterol. Two-dimensional RDFs of lipids and cholesterol center-of-masses projections to *XY* plane were determined in atomistic simulations, and were then used to determine effective potentials for a two-dimensional model of a lipid–cholesterol mixture. This two-dimensional model was further used to study formation of cholesterol-rich and cholesterol-poor domains in a mixed lipid–cholesterol bilayer.

In Wang et al. (2013b), the IMC method was used to build potentials for a CG model of $Bmim^+$-PF_6^- ionic liquid. The $Bmim^+$ cation of ionic liquid was presented as three beads and the PF_6^- anion as a single bead as depicted in Figure 1.6. As in other cases, RDFs for the IMC procedure were obtained in atomistic simulations of this system. The obtained effective potentials were used in large-scale simulations of $Bmim^+$-PF_6^- ionic liquid in

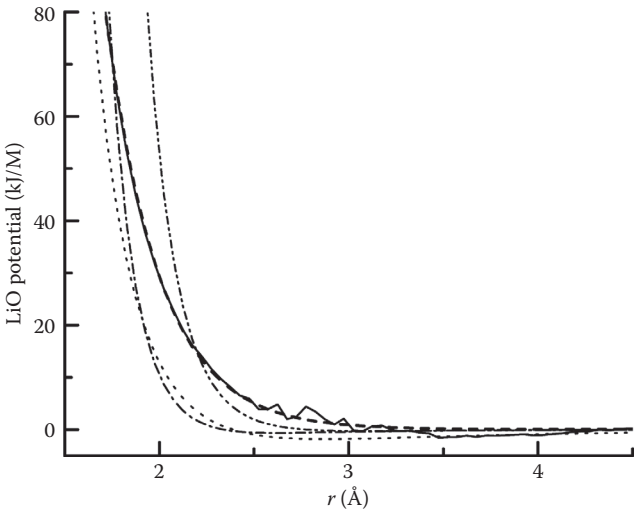

FIGURE 1.5 Short-range part of the effective LiO potential. Solid line—derived from *ab initio* simulation using the inverse Monte Carlo; bold dashed line—exponential fit; and other lines—potentials from other works, see Lyubartsev et al. (2001) for details. (Reprinted with permission from Lyubartsev, A. P. et al., *J. Chem. Phys.,* 114, 3120. Copyright 2001, American Institute of Physics.)

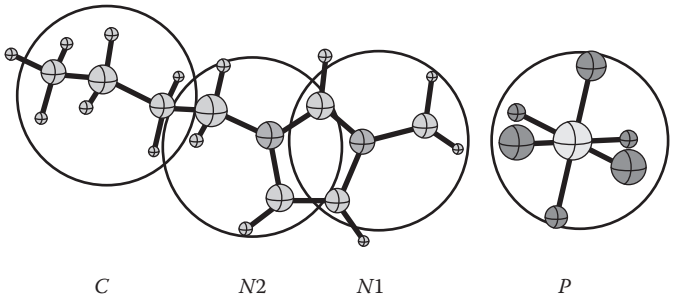

FIGURE 1.6 Coarse-graining scheme of $Bmim^{+}$-PF_6^{-} ionic liquid. (From Wang, Y.-L. et al., *Phys. Chem. Chem. Phys.,* 15, 7701, 2013. Reproduced by permission of The Royal Society of Chemistry.)

which, among other properties, experimental x-ray scattering factors were reproduced.

A few other examples of using IMC for reconstruction of interaction potentials from RDF can be mentioned: computation of effective potentials between charged colloids (Lobaskin et al., 2001), single-site water model (Eriksson et al., 2008), proline molecules in dimethyl sulfoxide (DMSO)

solvent (Lyubartsev et al., 2010). In Zhang and Berkowitz (2009), atomistic interaction potentials in Ag–Rh were determined by IMC from RDFs obtained from experimental scattering data.

1.4 FINAL REMARKS

In this chapter, the IMC methodology to extract interaction potentials from the structural properties of a molecular system (RDFs, bond and angles distributions) has been presented and its role in multiscale modeling is discussed. Examples given in this chapter show that the method is very general and can find applications for a very wide variety of molecular and macromolecular systems. In principle, the IMC method can serve as a key element in the systematic hierarchical multiscale modeling approach, which starts from *ab-initio* level, produce interaction potentials for classical atomistic simulations, and then proceed to different levels of coarse-graining. It might be, however, unrealistic to believe that the approach can produce CG models suitable for large-scale macromolecular simulations completely in an *ab-initio* manner, so experimental input as well as experimental validation will be always necessary. There are several reasons for this: the approximate nature of DFT functionals or atomistic force fields used at the starting stage; limited sampling of the phase space during RDF calculations; and inherent limitations of the coarse-graining procedure itself which in a typical case neglects high-order terms in the N-body potential of mean force. Transferability of the CG potentials, that is, their dependence on the temperature, concentration, and composition of the system, may be a serious issue. In some cases, like in ionic solutions, temperature and concentration dependence of the effective potentials can be effectively included into effective dielectric permittivity (Hess et al., 2006b; Mirzoev and Lyubartsev, 2011); in other cases, such as in a CG lipid model (Mirzoev and Lyubartsev, 2014), transferability studies need to be carried out. A reasonable strategy to construct effective potentials for CG models might be to determine the general form of the effective potential using IMC or other type of *bottom-up* coarse-graining, and then fine-tune parameters of the potential to fit available experimental data. The IMC methodology presented here would help to increase the fraction of "*ab initio*" derived features in CG molecular models at the expense of "empirically fitted" or "*ad hoc*" ones, which would enhance the predictive character and reliability of large-scale molecular simulations and advance them further to problems not yet covered within today's molecular simulation techniques.

ACKNOWLEDGMENTS

This work has been supported by the Swedish Research Council (Vetenskapsrådet).

REFERENCES

Allen, M. P., and D. J. Tildesley (1987). *Computer Simulations of Liquids* (2nd ed.). Oxford: Clarendon.

Ayton, G. S., E. Lyman, and G. A. Voth (2010). Hierarchical coarse-grained strategy for protein-membrane systems to access mesoscopic scales. *Faraday Discuss. 144*, 347–358.

Biase, P. M. D., S. Markosyan, and S. Noskov (2014). Microsecond simulations of DNA and ion transport in nanopores with novel ion–ion and ion–nucleotides effective potentials. *J. Comput. Chem. 35*, 711–721.

Brannigan, G., and F. L. H. Brown (2004). Solvent-free simulations of fluid membrane bilayers. *J. Chem. Phys. 120* (2), 1059–1071.

Brini, E., E. A. Algaer, P. Ganguly, C. Li, F. Rodriguez-Ropero, and N. F. A. van der Vegt (2013). Systematic coarse-graining methods for soft matter simulations a review. *Soft Matter 9*, 2108–2119.

Brini, E., V. Marcon, and N. F. A. van der Vegt (2011). Conditional reversible work method for molecular coarse graining applications. *Phys. Chem. Chem. Phys. 13*, 10468–10474.

Car, R., and M. Parrinello (1985). Unified approach for molecular dynamics and density functional theory. *Phys. Rev. Lett. 55* (22), 2471–2474.

Carbone, P., H. A. K. Varzaneh, X. Chen, and F. Müller-Plathe (2008). Transferability of coarse-grained force fields: The polymer case. *J. Chem. Phys. 128*, 064904.

Card, D. N., and J. P. Valleau (1970). Monte Carlo study of the thermodynamics of electrolyte solutions. *J. Chem. Phys. 52*, 6232–6240.

Chaimovich, A., and M. S. Shell (2011). Coarse-graining errors and numerical optimization using a relative entropy framework. *J. Chem. Phys. 134*, 094112.

Cooke, I. R., and M. Deserno (2005). Solvent-free model for self-assembling fluid bilayer membranes: Stabilization of the lipid phase based on broad attractive tail potentials. *J. Chem. Phys. 123*, 224710.

de Jong, D. H., G. Singh, W. F. D. Bennett, C. Arnarez, T. A. Wassenaar, L. V. Schäfer, X. Periole, D. P. Tieleman, and S. J. Marrink (2012). Improved parameters for the martini coarse-grained protein force field. *J. Chem. Theory Comput. 9* (1), 687–697.

Drouffe, J. M., A. C. Maggs, and S. Leibler (1991). Computer simulations of self-assembled membranes. *Science 254* (5036), 1353–1356.

Dunn, W. L., and J. K. Shultis (2012). *Exploring Monte Carlo Methods*. Boston, MA: Elsevier.

Egorov, A. V., A. V. Komolkin, V. I. Chizhik, P. V. Yushmanov, A. P. Lyubartsev, and A. Laaksonen (2003). Temperature and concentration effects on Li^+-ion

hydration. A molecular dynamics simulation study. *J. Phys. Chem. B 107*, 3234–3242.

Ercolessi, F., and J. B. Adams (1994). Interatomic potentials from first-principles calculations: The force-matching method. *Europhys. Lett. 26*, 583–588.

Eriksson, A., M. N. Jacobi, J. Nyström, and K. Tunstrøm (2008). Using force covariance to derive effective stochastic interactions in dissipative particle dynamics. *Phys. Rev. E 77*, 016707.

Farago, O. (2003). "Water-free" computer model for fluid bilayer membranes. *J. Chem. Phys. 119* (1), 596–605.

Ganguli, P., D. Mukherji, C. Junghans, and N. F. A. van der Vegt (2012). Kirkwood–Buff coarse-grained force fields for aqueous solutions. *J. Chem. Theory Comput. 8*, 1802–1807.

Goetz, R., and R. Lipowsky (1998). Computer simulations of bilayer membranes: Self-assembly and interfacial tension. *J. Chem. Phys. 108* (17), 7397–7409.

Gottlieb, M., and R. B. Bird (1976). A molecular dynamics calculation to confirm the incorrectness of the random-walk distribution for describing the Kramers freely jointed bead-rod chain. *J. Chem. Phys. 65* (8), 2467–2468.

Harmandaris, V. A., N. P. Adhikari, N. F. A. van der Vegt, and K. Kremer (2006). Hierarchical modeling of polystyrene: From atomistic to coarse-grained simulations. *Macromolecules 39*, 6708–6719.

Henderson, R. L. (1974). A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A 49A* (3), 197–198.

Hess, B., C. Holm, and N. van der Vegt (2006a). Modeling multibody effects in ionic solutions with a concentration dependent dielectric permittivity. *Phys. Rev. Lett. 96* (14), 147801.

Hess, B., C. Holm, and N. van der Vegt (2006b). Osmotic coefficients of atomistic NaCl (aq) force fields. *J. Chem. Phys. 124* (16), 164509.

Hijon, C., P. Espanol, E. Vanden-Eijnden, and R. Delgado-Buscalioni (2010). Mori–Zwanzig formalism as a practical computational tool. *Faraday Discuss. 144*, 301–322.

Högberg, C.-J., A. M. Nikitin, and A. P. Lyubartsev (2008). Modification of the CHARMM force field for DMPC lipid bilayer. *J. Comp. Chem. 29*, 2359–2369.

Izvekov, S., M. Parrinello, C. J. Burnham, and G. A. Voth (2004). Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys. 120*, 10896–10913.

Izvekov, S., and G. A. Voth (2005). A multiscale coarse-graining method for biomolecular systems. *J. Phys. Chem. B 109*, 2469–2473.

Korolev, N., Y. Fan, A. P. Lyubartsev, and L. Nordenskiöld (2012). Modelling chromatin structure and dynamics: Status and prospects. *Curr. Opin. Struct. Biol. 22* (2), 151–159.

Korolev, N., D. Luo, A. P. Lyubartsev, and L. Nordenskiöld (2014). A coarse-grained DNA model parameterized from atomistic simulations by inverse Monte Carlo. *Polymers 6*, 1655–1675.

Lindahl, E., B. Hess, and D. van der Spoel (2001). Gromacs 3.0: A package for molecular simulations and trajectory analysis. *J. Mol. Model. 7*, 306–317.

Lobaskin, V., A. P. Lyubartsev, and P. Linse (2001). Effective macroion-macroion potentials in asymmetric electrolytes. *Phys. Rev. E 63*, 020401.

Lopez, C. A., A. J. Rzepiela, A. H. de Vries, L. Dijkhuizen, P. H. Hunenberger, and S. J. Marrink (2009). Martini coarse-grained force field: Extension to carbohydrates. *J. Chem. Theory Comput. 5*, 3195–3210.

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

Lyubartsev, A. P., and A. Laaksonen (1995). Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E 52* (4), 3730–3737.

Lyubartsev, A. P., and A. Laaksonen (1997). Osmotic and activity coefficients from effective potentials for hydrated ions. *Phys. Rev. E 55* (5), 5689–5696.

Lyubartsev, A. P., and A. Laaksonen (1999). Effective potentials for ion–DNA interactions. *J. Chem. Phys. 111* (24), 11207–11215.

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

Lyubartsev, A. P., and A. Laaksonen (2004). On the reduction of molecular degrees of freedom in computer simulations. *Lect. Notes Phys. 640*, 219–244.

Lyubartsev, A. P., K. Laasonen, and A. Laaksonen (2001). Hydration of Li^+ ion. An *ab initio* molecular dynamics simulation. *J. Chem. Phys. 114* (7), 3120–3126.

Lyubartsev, A. P., and S. Marcelja (2002). Evaluation of effective ion-ion potentials in aqueous electrolytes. *Phys. Rev. E 65*, 041202.

Lyubartsev, A., A. Mirzoev, L. J. Chen, and A. Laaksonen (2010). Systematic coarse-graining molecular models by the Newton inversion method. *Faraday Discuss. 144*, 43–56.

Lyubartsev, A. P., and A. L. Rabinovich (2011). Recent development in computer simulations of lipid bilayers. *Soft Matter 7*, 25–39.

Marrink, S. J., A. H. de Vries, and A. E. Mark (2004). Coarse grained model for semiquantitative lipid simulations. *J. Phys. Chem. B 108*, 750–760.

Marrink, S. J., H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. de Vries (2007). The MARTINI force field: Coarse grained model for biomolecular simulations. *J. Phys. Chem. B 111* (27), 7812–7824.

Marrink, S. J., and D. P. Tieleman (2013). Perspective on the Martini model. *Chem. Soc. Rev. 42*, 6801–6822.

Mirzoev, A., and A. P. Lyubartsev (2011). Effective solvent-mediated interaction potentials of Na^+ and Cl^- in aqueous solution: Temperature dependence. *Phys. Chem. Chem. Phys. 13*, 5722–5727.

Mirzoev, A., and A. P. Lyubartsev (2013). MagiC: Software package for multiscale modeling. *J. Chem. Theory Comput. 9*, 1512–1520.

Mirzoev, A., and A. P. Lyubartsev (2014). Systematic implicit solvent coarse graining of dimyristoylphosphatidylcholine lipids. *J. Comput. Chem. 35*, 1208–1218.

Monticelli, L., S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, and S. J. Marrink (2008). The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput. 112*, 819–834.

Mukherji, D., N. F. A. van der Vegt, K. Kremer, and L. D. Site (2012). Kirkwood–Buff analysis of liquid mixtures in an open boundary simulation. *J. Chem. Theory Comput. 8*, 375–379.

Murtola, T., E. Falck, M. Karttunen, and I. Vattulainen (2007). Coarse-grained model for phospholipid/cholesterol bilayer employing inverse Monte Carlo with thermodynamic constraints. *J. Chem. Phys. 126* (7), 075101.

Murtola, T., E. Falck, M. Patra, M. Karttunen, and I. Vattulainen (2004). Coarse-grained model for phospholipid/cholesterol bilayer. *J. Chem. Phys. 121* (18), 9156–9165.

Naomé, A., A. Laaksonen, and D. P. Vercauteren (2014). A solvent-mediated coarse-grained model of DNA derived with the systematic Newton inversion method. *J. Chem. Theory Comput. 10* (8), 3541–3549.

Noguchi, H., and M. Takasu (2001). Self-assembly of amphiphiles into vesicles: A Brownian dynamics simulation. *Phys. Rev. E 64*, 041913.

Pandit, S. A., and H. L. Scott (2009). Multiscale simulations of heterogeneous model membranes. *Biochim. Biophys. Acta 1788*, 136–148.

Reith, D., M. Pütz, and F. Müller-Plathe (2003). Deriving effective mesoscale potentials from atomistic simulations. *J. Comp. Chem. 24*, 1624–1636.

Romiszowski, P., and R. Yaris (1991). A dynamic simulation method suppressing uninteresting degrees of freedom. *J. Chem. Phys. 94*, 6751–6761.

Rudzinski, J. F., and W. G. Noid (2011). Coarse-graining entropy, forces, and structures. *J. Chem. Phys. 135*, 214101.

Rühle, V., C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko (2009). Versatile object-oriented toolkit for coarse-graining applications. *J. Chem. Theory Comput. 5* (12), 3211–3223.

Savelyev, A., C. K. Materese, and G. A. Papoian (2011). Is DNA's rigidity dominated by electrostatic or nonelectrostatic interactions? *J. Am. Chem. Soc. 133*, 19290–19293.

Savelyev, A., and G. A. Papoian (2009a). Molecular renormalization group coarse-graining of electrolyte solutions: Applications to aqueous NaCl and KCl. *J. Phys. Chem. B 113*, 7785–7793.

Savelyev, A., and G. A. Papoian (2009b). Molecular renormalization group coarse-graining of polymer chains: Application to double-stranded DNA. *Biophys. J. 96*, 4044–4052.

Savelyev, A., and G. A. Papoian (2010). Chemically accurate coarse graining of double-stranded DNA. *Proc. Natl. Acad. Sci. USA 107*, 20340–20345.

Schommers, W. (1983). Pair potentials in disordered many-particle systems: A study for liquid gallium. *Phys. Rev. A 28*, 3599–3605.

Shell, M. S. (2008). The relative entropy is fundamental to multiscale and inverse thermodynamic problems. *J. Chem. Phys. 129*, 144108.

Shelley, J. C., M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein (2001). A coarse grained model for phospholipid simulations. *J. Phys. Chem. B 105*, 4464–4470.