On the Reduction of Molecular Degrees of Freedom in Computer Simulations

Alexander P. Lyubartsev and Aatto Laaksonen

Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University,
106 91, Stockholm, Sweden

Abstract. Molecular simulations, based on atomistic force fields are a standard theoretical tool in materials, polymers and biosciences. While various methods, with quantum chemistry incorporated, have been developed for condensed phase simulations during the last decade, there is another line of development with the purpose to bridge the time and length scales based on coarse-graining. This is expected to lead to some very interesting breakthroughs in the near future. In this lecture we will first give some background to common atomistic force fields. After that, we review a few common simple techniques for reducing the number of motional degrees of freedom to speed up the simulations. Finally, we present a powerful method for reducing uninteresting degrees of freedom. This is done by solving the "Inverse Problem" to obtain the interaction potentials. More precisely, we make use of the radial distribution functions, and by using the method of Inverse Monte Carlo [Lyubartsev & Laaksonen, Phys. Rev. E. **52**, 3730 (1995)], we can construct effective potentials which are consistent with the original RDFs. This makes it possible to simulate much larger system than would have been possible by using atomistic force fields. We present many examples: How to simulate aqueous electrolyte solutions without any water molecules but still having the hydration structure around the ions – at the speed of a primitive electrolyte model calculation. We demonstrate how a coarse-grained model can be constructed for a double-helix DNA and how it can be used. It is accurate enough to reproduce the experimental results for ion condensation around DNA for several different counterions. We also show how we can construct site-site potentials for large-scale atomistic classical simulations of arbitrary liquids from smaller scale *ab initio* simulations. This methodology allows us to start from a simulation with the electrons and atomic nuclei, to construct a set of atomistic effective interaction potentials, and to use them in classical simulations. As a next step we can construct a new set of potentials beyond the atomistic description and carry out mesoscopic simulations, for example by using Dissipative Particle Dynamics. In this way we can tie together three different levels of description. The Dissipative Particle Dynamics method appears as a very promising tool to use with our coarse-grained potentials.

1 Introduction

Atoms and molecules, the building blocks of matter from a chemist's point of view, interact with each other. They are attracted at long distances, whereas at short distances the interactions become strongly repulsive. In fact, there is no need to seek for a proof from the underlying physics to confirm that this is really the case as we can observe it indirectly on a macroscopic level with our own eyes every day. For example, due to the attraction water molecules nucleate to form drops, which in turn rain down to fill rivers. Thanks to the strong repulsion, on the other hand, one cubic meter of that very same water weighs basically always about 1000 kg, regardless of how much pressure we use in trying to compress it.

A.P. Lyubartsev and A. Laaksonen, On the Reduction of Molecular Degrees of Freedom in Computer Simulations, Lect. Notes Phys. **640**, 219–244 (2004)
http://www.springerlink.com/

Today, molecular modelling is carried out by following and analysing interacting structural models in computers and visualizing the results using computer graphics. Historically, modelling of molecules started long before computers were invented. Already in the mid 1800's structural models were suggested for molecules as an aid to explain their chemical and physical properties. Similarly, the first attempts to model interacting molecules go back to the pioneering work of van der Waals [1]. With his simple equation of state where the molecules are approximated as spheres attracting each other, he was able to predict a first-order gas-liquid phase transition. Many of the conceptual ideas from over a century ago are still used in modern methods of molecular modelling and simulations.

The interactions involving molecules can be explained based on quantum mechanics. Close to a contact distance between molecules, there is a complex mixture of forces of quantum mechanical origin [2]. Although the physical background is known, the mathematical complexity effectively prohibits any analytical attempts to describe condensed matter systems on the basis of equations for interacting particles. Fortunately, thanks to computers, there are now approximate but powerful methods to treat many-body systems and extract macroscopic properties from the data. In their simplicity these methods and models work amazingly well, based on man-made force fields, and applied on atomistic molecular models together with a marriage of classical physics and statistical thermodynamics.

1.1 Atomistic Force Fields

The methods of atomistic force-field based computer simulations of condensed matter have become a standard tool to investigate practically everything from metallic solids to complex biological systems in their physiologically natural environments, i.e. aqueous solutions of ions and organic molecules. Quite surprisingly though, these tools are still basically the same as they were when they first were introduced. It may have been the phenomenally rapid development of computer technology during the last three decades hampering the development of molecular simulation methods, or that these methods were very robust already from the very beginning when presented by Alder and Wainwright [3] and Rahman [4] for simple liquids, and developed further by Rahman and Stillinger [5] for molecular liquid water and McCammon and coworkers for proteins [6]. Most likely it is the combination of both. Both the time scales covered by routine molecular dynamics simulations and the number of particles in the simulation cell have increased roughly by one order of magnitude each decade from the days of the pioneering paper by Rahman and Stillinger [5].

In molecular computer simulations, the internal energy of the whole system, at any moment during the simulation, can be calculated by summing up the contributions from a number of separate potential energy functions with the actual spatial and geometrical coordinates as parameters. This collection of potential functions is called the *force field* because it provides a mathematical tool to guide each atom in a system towards a new direction under the influence of all the other atoms in its surrounding. Conceptually these terms are divided to so-called bonded and non-bonded, or alternatively, intra- and intermolecular interactions. For small rigid molecules the intermolecular potentials are often enough.

Bonded Interactions. In the most commonly used force fields [7–10] the intra-molecular interactions are divided to terms involving two, three and four connected atoms, respectively. For example, there are the harmonic terms describing the distortions from equilibrium positions in bond-stretching and angle-bending. The "energies" calculated from these terms, however, are simply just penalties due to deviations from natural bond lengths and angles rather than any real binding energies. Therefore these are only relative quantities. Their use in the context of a force field was introduced by Hill [11] to describe molecular deformations due to steric interactions. Additionally, there is a term (constructed from four connected atoms) to describe the periodic torsional motion of dihedral angles. This torsional model goes back to the investigations of Kemp and Pitzer [12] on the ethane molecule to describe the rotation around the sp^3–sp^3 bond. The periodic nature is implemented naturally in trigonometric cosine expansions with the amplitudes related to corresponding rotational barrier heights to hinder the rotation around a covalent bond due to steric interactions and/or the internal structure of the molecular orbitals forming the bond.

The above three terms can be considered as a minimum description of the intramolecular interactions. Although it could be in principle possible to omit the torsional terms and treat them with non-bonded interactions between the outermost atoms describing the dihedral angle (see below), an additional term to describe the out-of-plane motion and inversion should be mentioned. In some cases it is important to keep part of the molecule planar (for example involving an sp^2 carbon). In other cases an umbrella type of motion may take place (e.g. an amine group). As this term is also constructed from four atoms (not in a sequence, but three of the atoms connected to the fourth central atom) it is normally known as the improper torsion term.

Non-Bonded Interactions. While considering what would make the non-bonded interactions we could simply imagine for a moment that there are no bonds between the atoms of the molecules, only a collection of moving atoms. The atoms have specific volumes and can exert interactions on each other. While the non-ideal origin of the interactions was recognized by van der Waals already, it can also be rigorously derived using quantum mechanics. These van der Waals interactions are basically a mixture of repulsion due to overlapping valence electrons and attraction due to induction and dispersion forces [2]. These interactions are commonly approximated all together in the 80 years old 12–6 potential [13], where the distance dependence of the repulsion term is proportional to r^{-12} (mimicing the exponential soft-wall behaviour) while it is proportional to r^{-6} for the attraction, r being the inter-atomic distance.

Certain atoms can be considered as more electronegative [14] than others. They gain electrons, while their neighbouring, less electronegative atoms will consequently suffer from a corresponding lack of electrons. Although the whole molecule is always electroneutral (unless it is an ion), the individual atoms in molecules mostly are not. Common force fields normally represent this redistribution of the total electron density by collecting the local net charges to single sites. Normally the charge is attached to the site of an atom. Sometimes it may also be placed along the bonds, or outside an atom to mimic a lone pair. Atomic charges are practical as they allow the use of Coulomb's law to describe their mutual interactions at the same time as they provide multipoles for

molecules or individual functional groups (charge groups). The electrostatic interactions, however, decay as r^{-1} with respect to the charge-charge distance, making them very long-ranged compared to the van der Waals interactions. Correct treatment of Coulombic interactions is, however, crucial in almost all categories of molecular simulations because they play an important role at both short and long distances. Hill [11] used the Lennard–Jones 12–6 potential [13] for non-bonded interactions, but did not consider the electrostatic interactions. The use of partial charges in force fields was later introduced by Kitaigorodskii [15].

Normally the short-range van der Waals and long-range Coulombic interactions are computed together because they both require calculations of the same site-site distances. Distance calculation is the most computer-intensive part of every simulation as it requires computation of a square root and an inverse ($1/\sqrt{r^2}$) for $N(N-1)/2$ pairs, where N is the number of sites in the system to be simulated. In practice it is not quite an $\mathcal{O}(N^2)$ problem as it is possible to restrict the interactions within a cut-off sphere measured from each atom. This is obvious for the fast decaying Lennard–Jones terms but is applied even for the Coulombic interactions when methods based on Ewald summation techniques are used. Also, modern computers often have the inverse and square root computations implemented in the hardware.

At this point we can put all the covalent bonds back again. Based on the underlying philosophy behind the atomistic force fields, a number of closest "non-bonded" pairs will be excluded because they have already been treated as "bonded" interactions, although this is only a very small fraction of all the $N(N-1)/2$ pairs. Now, there is one specific term in most force fields that should be given some more attention as it turns out to be an intermediate case: a pair of atoms in the outermost positions of a dihedral angle consisting of four atoms and three bonds. In many cases a compromise is made to treat this particular pair partially as a bonded and partially as a non-bonded interaction, and be called the "1–4" interaction. The reason is that steric interactions affect rotations around a bond. However, they are rather rough and in most cases lack further details of local internal conformational degrees of freedom. However, due to the short distance between the 1–4 atoms, both the Lennard–Jones and the Coulombic interactions are normally scaled down substantially. The 1–2 or 1–3 non-bonded interactions are, although somewhat arbitrarily, omitted completely. It is assumed that they are properly described with bond and angle potentials, most often within the harmonic approximation.

It should be mentioned that hydrogen bonds are sometimes enhanced by terms containing a narrow well to adjust its spatial length and directional terms to adjust the angle of the hydrogen bond. Both are parameterized to fit into the geometric definition of a hydrogen bond. In most cases hydrogen bonds are naturally formed due to the Coulombic interactions.

There are many suggestions to improve the terms discussed above; simply too many to review them all here. To mention just a few, the harmonic bond-stretching term is sometimes replaced by the anharmonic Morse potential [16]. This is a realistic model containing the anharmonicity and dissociation behaviour. It is required in calculations of spectroscopic properties as harmonic potentials do not give, for example, overtones. Morse potential is shown to describe the important O–H stretching fairly well. The angle in the angle-bending term can sometimes be replaced by the distance and described as a

bond thereby giving a somewhat more realistic behaviour. A nice example of these both features is the flexible water model by Toukan and Rahman [17] where the water molecule is made to act like a triangle by an additional H–H "bond", while O–H bonds are described by a Morse potential. The Lennard–Jones 12–6 term has occasionally been replaced by the physically more correct exponential term (e^{-6}) [18]. The above improvements become important if the molecules are simulated in more extreme conditions away from equilibrium geometries or configurations.

In general the very simple functional forms of the potential functions discussed above are adopted as they approximately describe the specific isolated physical interactions in situations close to equilibrium conformations and configurations. It is also assumed that the various terms do not interfere each other too much under normal conditions. If they do, cross terms can be introduced. Higher order terms are normally not used in force fields as they can easily lead to unexpected behaviour outside equilibrium conditions.

An additional problem is the treatment of non-bonded interactions between unlike atoms. Most often so-called simple (Lorentz–Berthelot) combination rules are applied for the Lennard–Jones potentials characterized by ε and σ for energy and length scales, respectively. Then, given ε_i to describe the potential energy scale between particles of type "i" (and ε_j accordingly), the parameter for unlike atoms is obtained by geometric mean: $\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j}$. For σ, one applies arithmetic mean, i.e. $\sigma_{ij} = (\sigma_i + \sigma_j)/2$. Strictly speaking, the Lennard–Jones parameters apply only for pure substances. Any type of non-ideal behaviour between molecular mixtures is difficult to incorporate into force fields. Simple combination rules normally give too attractive cross-interaction. Other empirical combination rules than the above mentioned Lorentz–Berthelot description exist in the literature.

In summary, the whole force-field can be visualized as a collection of soft balls and springs to build the molecules, with the size of the balls determined by the Lennard–Jones 12-term, an elastic collision parameter. The strength of the springs (for covalent bonds and angles) is set by force constants, while rotations around bonds are not completely free but hindered by periodic frictions (barriers). The balls attract or repel each other pairwisely depending on their mutual distances and/or the charges they carry. Force fields, consisting of continuous functions, are simple to use in modelling software to search for the energy minimum conformations, or in simulation programs to move molecules to sample the phase space, either randomly or deterministically by solving the equations of motions for each mass point (atom). In the latter case, derivatives of force fields are needed. A typical force field as discussed above with a minimum number of terms, used in computer simulations of flexible molecules, looks then like the following:

$$
\begin{aligned}
V = & \sum_{\text{covalent bonds}} K_{\mathrm{b}}(r - r_{\mathrm{eq}})^2 + \sum_{\text{covalent angles}} K_{\mathrm{a}}(\theta - \theta_{\mathrm{eq}})^2 \\
& + \sum_{\text{torsional angles}} \frac{1}{2} K_{\mathrm{t}} \left(1 + \cos(m_{\mathrm{t}}\phi - \gamma_{\mathrm{t}})\right) \\
& + \sum_{i<j} \left\{ 4\,\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^{6} \right] \frac{q_i q_j}{r_{ij}} \right\}.
\end{aligned}
$$

1.2 Parameterization of Force Fields

Although the functional forms of atomistic force fields have been the same for several decades, force fields do differ from each other. The real value of a force field lies on its parameterization. What makes the parameterization a non-trivial matter is the demand for a force field to be transferable from one molecular system to another. Otherwise it would not have much practical use. To be more realistic, this requirement has been softened slightly so that the parameterization should be transferable from one system to another similar system. Historically, such general force fields were first constructed for organic molecules, thereafter for proteins, nucleic acids, membranes and carbohydrates. All of them require some special care to be taken to work well, and to give results in good or reasonable agreement with experiments. New revisions appear frequently in the literature.

There is no simple way to derive force field parameters from the underlying physics (quantum mechanics). They are therefore deduced empirically using a variety of sources by combining data from suitable experiments with *ab initio* electronic calculations. Functionality normally requires large sets of atomic types with their parameters. This means that instead of one carbon atom with a mass of 12 atomic mass units, most force fields have several of them depending on their hybridization. Same is true for oxygen depending on its position (–O–, =O or –O–H). There are no reliable potentials for multivalent metal ions due to a substantial three-body interactions and strong polarizing effects. It should be mentioned that there are several main philosophies to assign the atomic types and carry out the parameterization [19].

Among the sources to parameterize force fields are the experimental information about the molecular structure and geometry coming from X-ray, electron and neutron diffraction, NMR spectroscopy as well as optical spectroscopy for small molecules. Other types of experimental data are vibrational frequencies, internal rotational barriers, conformational differences and thermodynamic data such as heat of formation and solvation free energies. Although it seems that plenty of experimental information can be used, the problem is to find it. It is also possible to obtain structures, conformations, and vibrational frequencies from electronic structure calculations. Particularly the modern density functional theory (DFT) methods have proven to be reliable. Besides the two main lines of methodologies, there are some empirical rules one can apply as well. Also, computer simulations can be used to optimise the potential parameters. For example, thermodynamic properties such as density or pressure can be corrected by adjusting van der Waals interactions. Clearly the information is gathered from all possible sources. Only with the final results from a simulation at hand, it is possible to judge the quality of the force field.

Besides the important question of transferability from one system to another there is a question of applicability at different physical conditions like temperature or pressure. Also it is important to remember that there is no single universal force field applicable at all the three states of matter, not even for the smallest molecules. An illustrating example of this point is a work where an interaction potential was constructed for $Na^+ Cl^-$ at a very high level based on electronic structure calculations (multi-reference CI) [20]. It was fitted accurately to a very flexible function and applied to calculate gas, liquid and solid phase properties of sodium chloride. Because the potential function was con-

structed from $Na^+ - Na^+$, $Cl^- - Cl^-$ and $Na^+ - Cl^-$ interaction energies from gas state (vacuum) conditions it reproduced all spectroscopic (cross- and anharmonic) constants very accurately in a Dunham analysis. It worked fairly well for phonon dispersion curves and reproduced the polymorphic FCC $\rightarrow$ BCC phase transition. However, it failed to produce the interaction energies for molten salts of NaCl by 30 %. Clearly, because it was a pure pair-potential, it was lacking the many-body character needed in liquid state simulations. The empirically parameterized force fields, however, are effective pair potentials as they get some of the many-body interactions into the parameters when fitted on experimental data obtained from macroscopic systems. Another critical test for a potential function is to use it in simulations of ordered systems or close to interfaces. For example, water models doing well in simulations of liquid water sometimes fail when used in clusters (droplets) of water.

Last but certainly not least there is the problem how to assign the atomic fractional charges. As there is no such a thing as the atomic charge (it is neither an experimental nor a theoretical observable) there is no unique way to determine it. Normally when a force field is constructed there remains the problem where to put the charges. For small molecules they can be obtained by fitting the charges to reproduce a set of molecular multipole moments (dipole, quadrupole, octopole, hexadecapole, etc.). For example, in the case of HCl [21] the charges were fitted to reproduce the first four multipole moments of HCl. However, even in this case the procedure was nothing but trivial: accurate enough fitting required one to add a new site behind the chlorine atom, and to vary the distance from the chlorine atom to the virtual charged site. This illustrates that by assigning the charges entirely on atomic sites is not always sufficient. For large molecules charges are a major headache. Various schemes are discussed for example in [19].

2 Reduction of Molecular Degrees of Freedom

Dynamical molecular systems made up of interacting particles are characterized by numerous degrees of freedom. All of them are important to include in a simulation model, but not always at the same time. It is neither practical nor even possible today to simulate, for example, a protein in water solution over nanosecond time scales using *ab initio* MD methods. Therefore the time scale and/or the length scale associated with phenomena one aims to study using computer simulations are decisive in the consideration of a suitable level of approximations. In other words, which degrees of freedom should be kept while still hoping to get reliable results.

The thermal motion of molecules can be divided between translational, rotational and vibrational degrees of freedom based on the classical equipartition principle. Quantum mechanically we also have electronic and nuclear degrees of freedom. While internal nuclear degrees of freedom fall outside the regime of condensed matter simulations, the much slower motion of the nuclei themselves compared to that of the surrounding electrons is the essence of the Born–Oppenheimer approximation. It is also the basis of molecular dynamics simulations with quantum mechanics incorporated. Examples of other degrees of freedom are the description of the environment, such as solvent molecules. In all, this issue gives rise to a number of general questions. Do the solvent molecules interact with the solute in some specific way like water molecules or ions

with biomolecules? Alternatively, is it sufficient to describe the solvent as a viscous or dielectric medium? Are there external fields applied on the system under investigation? Are some regions of conformational space much more probable than others?

2.1 Eliminating Fast Fluctuations

In molecular dynamics simulations the input parameter of main importance is the time step. This is the key parameter in solving the equations of motion numerically based on finite difference algorithms. Clearly this discrete unit of time should be considerably shorter than the period of the fastest dynamical fluctuations to accurately determine the trajectories for all the particles in the simulated system. On the other hand, considering the whole simulation and the time scale covered by the simulation, the time step is the main limiting factor. In simulations of excited state properties of a chromophore the time scales of interest may be within a few hundreds of femtoseconds, while the time step is a fraction of a femtosecond. Conformational changes in large biomolecules take place during a few hundreds of nanoseconds, however. Thus, one wants to use as long time steps as possible to reach these long time scales. To increase the size of the time step one has to get rid of the fast (but interesting) motional degrees of freedom. Below we give a short overview of techniques to do it and thereafter present our own hierarchical method.

In real quantum mechanical simulations one should preferably treat all particles as quantum particles, even the nuclei. Unfortunately no rigorous methods exist able to go beyond the Born–Oppenheimer approximation due to the great complexity of the underlying theory. Currently there are semi-classical wave packet methods and path integral approaches, while all so-called quantum simulation methods are based on the Born–Oppenheimer scheme.

In classical atomistic computer simulations the electronic degrees of freedom have been replaced with point charges as discussed above. This leaves the vibrational fluctuations of bond stretching as the fastest fluctuations. For all-atom simulations with all the hydrogen atoms included it would require a time step of about 0.1 – 0.2 femtoseconds – a painfully short time step for simulations of biological systems. Normally, these bond vibrations are not of interest, especially when harmonic potentials are used. This idea has lead to approaches where bond stretching has been frozen. This was done first by Ryckaert et al. [22] who introduced the so-called SHAKE algorithm to constrain the bond lengths iteratively while keeping the center-of-mass fixed. In constrained dynamics the time step can be increased up to one order-of-magnitude. Another approach to deal with the same problem is to treat the fast and slow motions separately. This can be done by multiple time step methods, like the one introduced by Tuckerman et al. [23]. Then the bond stretching and angle bending, for example, can be treated using a short enough time step, while the remaining degrees of freedom are described with a long time step (comparable to what is used in constrained dynamics). In the computing time this method is competitive with constrained dynamics, while at the same time one can still describe the vibrational fluctuations and keep all the internal degrees of freedom (in constrained dynamics there are no vibrational contributions from bond stretching). Sometimes some degrees of freedom are restrained to certain intervals, such as bond stretching or angle bending but also other non-bonded interatomic distances such as the

proton-proton distances from NMR–NOE (nuclear magnetic resonance nuclear Overhauser enhancement) measurements. This is done by adding penalty terms to the force field. Note that there are still contributions to the thermal energy from these fluctuations.

Next step would be to keep the molecules completely rigid. This was done already by Rahman and Stillinger in their pioneering MD simulation of water [5]. In fact, it is a standard method to treat small and effectively rigid molecules. In a rigid molecule MD, the equations of translational motion are solved for the molecular center-of-masses. Simultaneously, the Eulerian equations are solved for the rotational motion, calculated using the torques. At the beginning of a new time step the rotated molecules are placed on their center-of-mass positions to provide an updated configuration. Obviously the use of this method cannot be extended for large and flexible molecules.

2.2 Simplifying Molecular Models

Another way to reduce the complexity of a system is to use so-called united atoms. For example, it is common to describe CH_2 and CH_3 groups as united atoms having the masses of 14 and 15 atomic mass units, respectively. In doing so, an effective, larger Lennard–Jones sphere is put either on the center-of-mass of the whole group or on top of the heavy atom. This has several benefits. It eliminates the light hydrogen atoms and thereby allows one to increase the time step. Further, it also reduces the number of interaction sites and therefore increases the speed of the simulation. This is also a reasonable approximation as the CH_2 and CH_3 groups are inert. Consequently hydrocarbon chains, like the ones in lipid membranes are often described using united atoms. Also there are popular three-site models for methanol, thus eliminating the internal rotation of the methyl group. In some cases the united atom model may work even better than the existing all-atom model. It is also possible to extend the united atom model to larger groups like what was done in the simulation of t-butyl alcohol $(CH_3)_3OH$ by reducing a 15-site model to a 3-site model [24].

In principle the simplification procedure can be taken much further by constructing intermediate and low-resolution models made up of beads or clusters of beads connected with springs, representing functional groups or building blocks like amino acids, etc. These types of models have been used for a long time, and particularly in simulations of protein folding. An illustrative selection of these models is presented by Hall and co-workers [25]. We will return to bead models later when we present our own coarse-grained DNA model (see Sect. 4.2).

2.3 Elimination of Explicit Solvent Molecules

Computer simulations of large biomolecules under physiologically relevant conditions and environment require plenty of solvent water molecules around the macromolecule. First, there should be several layers of water forming the hydration shell structure. In addition, a generous amount of bulk water should be added to fill the whole simulation cell allowing exchange of molecules between hydration and bulk phases. Obviously, water in large amounts is important in most biological systems to stabilize their three-dimensional structure, but also as a medium for ions and other molecules coming close to them.

Single water molecules are also vital for the key functions of biomolecules. However, in these simulations the CPU time goes mainly to compute water-water and water-solute interactions. In the first simulations of proteins [26] and many others reported thereafter the water molecules were simply omitted. The solvent was represented as a dielectric continuum. So was done in the first simulation of DNA [27], where even the Coulombic interactions were switched off to prevent the breakdown of the double-helix structure due the repulsion of the anionic phosphate groups. Now, replacing the aqueous solvent of explicit water molecules by a dielectric constant for liquid water is a very crude approximation, particularly in the regions close to the macromolecule where solute-solvent interactions can be very specific on atomic scale. However, various forms of polarizable continuum models can sometimes do the trick. For an interesting method to link together the continuum and microscopic regimes, see [28]. Again, we return to this problem to eliminate the solvent molecules as we present the method of Inverse Monte Carlo in Sect. 3.

3 Effective Potentials and the Inverse Monte Carlo Method

3.1 Inverse Problem in Statistical Mechanics

In statistical mechanical modelling of condensed matter systems, one needs a set of interaction potentials or the Hamiltonian to describe the model system. After the Hamiltonian is specified, analytical or numerical methods within the pairwise approximation can be applied, including computer simulation techniques. The aim is to compute canonical averages of physical quantities and compare them with experiments if these are available. In particular, we can calculate various thermodynamic and structural properties in terms of distribution functions. In case of dynamical simulations we can calculate time-dependent transport properties. Realistic simulations of large enough systems with long time scales are straightforward but computationally expensive.

An interesting and fundamental problem in statistical physics is the "Inverse Problem". Namely, how to deduce the interaction potential (or reconstruct the Hamiltonian) from canonical averages. These averages can be known, for example, from experiments. One such property is the radial distribution function obtained through a Fourier transform of the structure factor from X-ray or neutron scattering experiments of condensed phase samples [29].

In 1974, Henderson presented a theorem [30] of a unique, point-to-point correspondence between pair potentials and radial distribution functions. For a given system under given conditions of temperature and density, two pair potentials, giving rise to the same RDF $g(r)$, do not differ from each other by more than a single constant. This constant can further be defined from the condition that the interaction potential vanishes at the infinite distance, making the potential uniquely defined.

Not surprisingly, the solution of the inverse problem is not as straightforward in practice as a calculation of the RDF from a known interaction potential. It is not possible to write a formal expression for how to compute the interaction potential from the RDF. Some special techniques are required to solve the inverse problem as we will show later after discussing a few related ideas.

In 1988, McGreevy and Pusztai suggested the so-called Reverse Monte Carlo (RMC) method where the starting point for a simulation was the radial distribution function [31]. Within this technique, a MC simulation can be carried out in a consistent way to retain the input RDF. However, this task is accomplished without any knowledge of the interaction potential. The set of configurations produced by the RMC can be used for computation of further structural information, e.g. calculation of 3D spatial or orientational correlations. However, it should be stressed that because the interaction potential is not reconstructed, the inverse problem is not solved. It is not possible to calculate thermodynamical or dynamical properties of the system using this approach. Naturally, it cannot be used for any type of coarse-grained simulations.

In a related approach, Reatto and coworkers [32] have used the Hypernetted Chain (HNC) approximation to solve the inverse problem. This or any similar approach, based on some closure of the Ornstein–Zernike equation, were also applied in several works during the last decade to produce interaction potentials from radial distribution functions [33,34]. Again, it should be stressed that computations within the HNC theory are only feasible for relatively simple models. Although it gives sometimes accurate results [35], the HNC theory is not an exact mathematical solution to a statistical-mechanical problem. Therefore its accuracy should be validated in every specific case. We should also mention a few other works devoted to the inverse problem [36–38].

A practical way to solve the inverse problem completely is to reconstruct the interaction potential from the distribution function along the lines suggested in our paper [39]. This method together with its applications to compute various effective potentials on different levels of description of condensed matter, will be discussed in more details in the following sections.

3.2 The Method

Consider first a simple system of identical particles interacting through a pair potential. The corresponding interaction Hamiltonian is given as

$$H = \sum_{i,j} V(r_{ij}), \tag{1}$$

where $V(r_{ij})$ is a pair potential, and r_{ij} the distance between the particles i and j. Assume further that the radial distribution function $g(r)$ is known so we can construct the corresponding interaction potential $V(r)$.

Let us first apply a grid approximation to digitalize the Hamiltonian:

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

for

$$\begin{aligned} &r_\alpha - \frac{1}{2M} < r < r_\alpha + \frac{1}{2M}; \\ &r_\alpha = (\alpha - 0.5) r_{\mathrm{cut}}/M; \\ &\alpha \ = 1, \ldots, M\,, \end{aligned}$$

where r_{cut} is a cutoff distance and M is the number of grid points within the interval $[0, r_{\mathrm{cut}}]$. Then we can rewrite the Hamiltonian (1) as:

$$H = \sum_{\alpha} V_{\alpha} S_{\alpha} \,, \tag{2}$$

where S_{α} is the number of pairs of particles whose mutual distance is inside the α-slice, and serves as an estimator of the radial distribution function:

$$\langle S_{\alpha} \rangle = \frac{4\pi r^2 \, g(r) N^2}{2V}.$$

The average values of S_{α} are now some functions of the potential V_{α}, and one can write the expansion

$$\Delta\langle S_{\alpha} \rangle = \sum_{\gamma} \frac{\partial \langle S_{\alpha} \rangle}{\partial V_{\gamma}} \Delta V_{\gamma} + \mathcal{O}(\Delta V^2) \,, \tag{3}$$

where the derivatives $\partial\langle S_{\alpha}\rangle / \partial V_{\gamma}$ can be expressed using exact relationships based on statistical mechanics [39]:

$$\begin{aligned} \frac{\partial \langle S_{\alpha} \rangle}{\partial V_{\gamma}} &= \frac{\partial}{\partial V_{\gamma}} \frac{\int \mathrm{d}q S_{\alpha}(q) \exp(-\beta \sum_{\lambda} V_{\lambda} S_{\lambda}(q))}{\int \mathrm{d}q \exp(-\beta \sum_{\lambda} V_{\lambda} S_{\lambda}(q))} \\ &= -\frac{\langle S_{\alpha} S_{\gamma} \rangle - \langle S_{\alpha} \rangle \langle S_{\gamma} \rangle}{k_{\mathrm{B}} T}. \end{aligned} \tag{4}$$

Equations (3) and (4) give us the interaction potential V_{α} from radial distribution functions $\langle S_{\alpha} \rangle$ iteratively. Let $V_{\alpha}^{(0)}$ be some trial potential such as the corresponding potential of mean force,

$$V_{\alpha}^{(0)} = -k_{\mathrm{B}} T \, \ln g(r_{\alpha}).$$

Then, using standard Monte Carlo simulations, one can evaluate the averages $\langle S_{\alpha} \rangle$ and their deviations from the reference values S_{α}^{*}, defined from the given RDF:

$$\Delta\langle S_{\alpha} \rangle^{(0)} = \langle S_{\alpha} \rangle^{(0)} - S_{\alpha}^{*}.$$

By solving the system of linear equations (3) with the coefficients defined by (4), and omitting the terms $\mathcal{O}(\Delta V^2)$, we can obtain corrections to the potential $\Delta V_{\alpha}^{(0)}$. Then the procedure is repeated with the new potential

$$V_{\alpha}^{(1)} = V_{\alpha}^{(0)} + \Delta V_{\alpha}^{(0)},$$

until convergence is achieved. The whole procedure is similar to a solution of a multidimensional non-linear equation using the Newton–Rhapson method.

If the initial approximation of the potential is poor, some type of regularization of the iteration procedure is needed. In this case, we multiply the required change of the RDF by a factor between 0 and 1. By doing so, the term $\mathcal{O}(\Delta V^2)$ in (3) can be made small enough to guarantee convergence of the whole procedure at the cost that the number of iterations will increase in this case.

The above algorithm also gives us a possibility to evaluate the uncertainty of the inverse procedure. The input RDF (it can be either an experimental RDF or a RDF obtained in an accurate computer simulation) contains normally some uncertainty. An analysis of the eigen values and eigen vectors of the matrix (4) gives us information about how changes in a certain part of the RDF will change the effective potential. For example, eigen vectors with eigen values close to zero correspond to changes in the potential hardly affecting the RDF. The presence of eigen values close to zero makes the inverse problem less well-defined, which in some cases (for example, liquid water) may cause some serious problems in the RDF inversion [40].

It should be mentioned that our approach is related to the renormalization group Monte Carlo method [41,42], used previously to study phase transitions in lattice models (e.g., Ising models for ferromagnets, polymer models, etc.) as well as in the quantum field theory. The renormalization procedure represents, in fact, a change of scale in representating the given system, during which one consecutively moves towards more and more coarse-grained descriptions of the system. The set of (3)–(4) has been used in references [41,42] to describe how the parameters of the Hamiltonian change during the renormalization procedure. The applications of this method are clearly more general than the lattice systems near the phase transition point, and cover even soft-matter problems, allowing us to "renormalize" the Hamiltonian of a molecular system in such a way that only those degrees of freedom we are interested in are kept.

3.3 How to Implement the Method for More General Systems

The method described above for a mono-component system can easily be extended for more general systems, including mixtures of molecules, presence of long-range electrostatic forces, etc.

In the case of a multicomponent system (assume K different species), we have a set of $K(K+1)/2$ pair potential functions and their corresponding RDFs. Consequently, we will have $M \times K(K+1)/2$ terms in the sum (2) describing the Hamiltonian in the grid approximation, and the same number of lines in the linear system equation (3). In other respects, the algorithm is a complete analogue of the one-component case. Similarly, the method can be generalized to provide effective potentials between molecules, being presented in this case as a sum of site-site interactions.

In calculations of potentials between mobile particles (ions) and more static sites of a large (fixed) macromolecule, we have a set of $K(K+1)/2$ potential functions, describing the interactions between K species of mobile particles, and $K \times K_{\mathrm{M}}$ potential functions, describing interactions between the mobile particles and K_{M} different sites on the macromolecule. Naturally, in order to calculate this set of pair potentials, one needs the corresponding set of radial distribution functions.

In a typical case the RDF between two atoms is zero at short distances. In the numerical treatment, we can put the potential at the corresponding distances to $+\infty$ (hard core). This again leads to zero RDF intensities at these distances and zero coefficients in corresponding lines of the linear equation system (3). These zero lines can simply be omitted from the group of equations by assuming that if $g(r)$ is zero for some inter-particle distance, then the probability should be zero for any pair of particles approaching

the distance, and the corresponding potential can be therefore set to infinity. The true potential can still assume a finite value, but the absolute value is not important.

In the case of charged particles (ions) special care should be taken of the long-range part of the electrostatic interactions. The inverse procedure yields the effective potential for the range of distances for which the input RDF is defined. Therefore we need some assumptions about the behaviour of the effective potential at distances outside the cut-off range. A reasonable assumption is that outside the cut-off distance the effective potential is made equal to the Coulombic potential with a proper dielectric constant. Our previous studies of ionic solutions [43] show that this assumption is valid already at distances of 10 Å and beyond.

In practical calculations for systems involving charged particles, we can divide the effective pair potential in the whole range $[0, \infty]$ into two parts: the short-range part up to the cut-off radius and the Coulombic potential with some dielectric constant. As the Coulombic part is not changed during the simulations it can be calculated using the Ewald method, while the short-range part changes according to (3) and (4) as it is fitted to the input RDF.

The inverse Monte Carlo algorithm may also easily be generalized to other, non-pairwise types of interactions. For example, intramolecular covalent angles and torsional angle potentials can be obtained, in a similar manner, from the population distributions of these angles. Also, three-body potentials can, in principle, be calculated from three-body correlation functions, however such computations would require very large computer memories to store the matrix (4) for a three-dimensional grid.

4 Some Examples and Discussion

4.1 *Ab–Initio* Effective Potentials

Liquid Water. As discussed in the Introduction, *ab-initio* parameterization of force field parameters from a large number of dimer configurations is very complicated and behind a tedious work. In our recent work [40] we proposed a novel approach to construct site-site interaction potentials from *ab-initio* simulations both accurately and efficiently as well as without any pain. The main idea behind this approach is to use detailed high quality *ab initio* liquid simulations to calculate a set of RDFs between all the different pairs of atoms (or arbitrary sites) in the system. The RDFs will then be used to construct a set of effective pair potentials between the same sites using the Inverse Monte Carlo procedure described in Sect. 3.2.

The suggested method provides us with an automatic procedure to construct a complete set of pair interaction potentials from *ab-initio* (for example, Car–Parrinello) simulations. Unlike, while doing the tedious work manually, we do not need to worry about the sampling of the configurational space because all the relevant parts of the phase space become properly weighted as the *ab-initio* simulation proceeds at ambient temperatures. Most importantly and in contrast to previous *ab initio* potentials like the MCY model [44], which is by definition a pure pair potential, our pair potential is a true effective potential since it is calculated at the proper liquid state conditions at finite temperatures rather than in vacuum at 0 K. An effective pair potential includes all high-order interactions averaged.

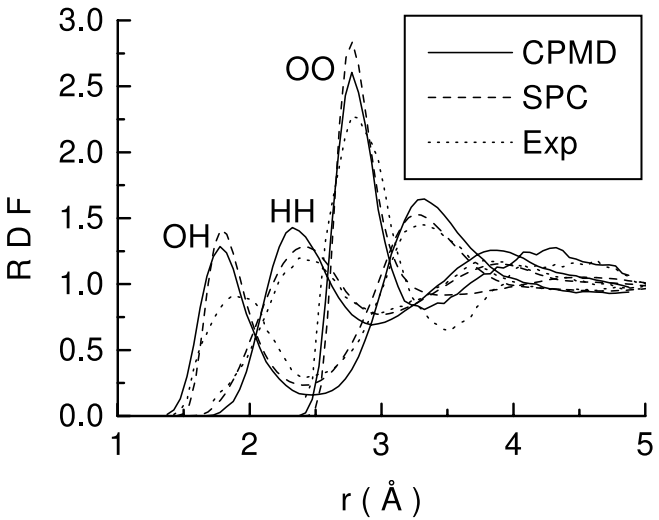

Fig. 1. Radial distribution functions of water obtained in the Car–Parrinello simulations. Results of classical MD for the SPC model and experimental neutron scattering data [46] are also shown

The first application of this approach was the reconstruction of pair potential for water at ambient conditions [40]. Liquid water is an obvious choice to test new theoretical methods. Water is the most studied system using computer simulations, and for water there exist both relatively reliable potential models and experimental data. Most water models are empirically parameterized to reproduce some set of available experimental data.

The Car–Parrinello molecular dynamics simulations [45] were carried out for 32 water molecules in a cubic box of linear size 9.855 Å (corresponding to a density of $1\,\mathrm{g/cm^3}$) with periodic boundary conditions. The time step was set to 0.15 fs. The initial configuration was taken from the molecular dynamics simulations of the SPC water model after a proper equilibration at a temperature of 300 K. The system was then further equilibrated for 4 ps using the Car–Parrinello MD method, followed by 10 ps of averaging. The temperature was controlled by scaling the velocity if the kinetic temperature deviated from the target value (300 K) more than 50 K. As a result, the average temperature during the production run was 297 K. More detailed simulation details can be found in [40].

The resulting RDFs for the different pairs of atoms are shown in Fig. 1, and the effective interaction potentials obtained from the RDFs are shown in Fig. 2. These effective potentials reproduce the very same RDFs as were obtained from Car–Parrinello MD simulations, within a precision of at least 0.01 for all distances. For comparison, the potential curves of a popular, empirically parameterized three-site SPC water model [47] are also shown in Fig. 2. It is interesting to observe that the *ab initio* effective potentials are almost superimposed on the SPC model potentials at distances exceeding 3.5 Å. At shorter distances (2.8–3.5 Å for O–O and 1.6–3 Å for O–H pairs) the *ab initio* O–O and O–H potentials rise somewhat less steeply, while at very short distances they are noticeably steeper than the corresponding SPC potential curves.

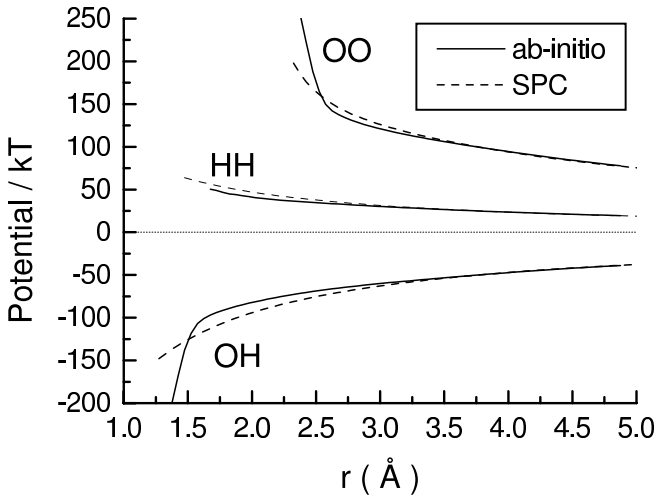

Fig. 2. The effective potentials for water, calculated by the inverse MC method from *ab initio* RDFs. The SPC model potentials are given for comparison

We would like to stress that the main purpose of this example was not to suggest yet another water model, but rather to present and test a pure "*ab initio*" way to derive a set of effective atom-atom potentials for classical MD simulations. We have constructed them from the structure of liquid water through Car–Parrinello simulations. At the moment the effective interactions yield, in a few minutes, the very same set of RDFs as the Car–Parrinello calculations that require a few months of computing time on the same computer. Of course, the Car–Parrinello method does not represent the ultimate high-end simulation tool. Other, more accurate, quantum chemical approaches could be used to calculate the forces during the *ab-initio* molecular dynamics run. It can also be mentioned that *ab initio* simulations, such as the Car–Parrinello technique, are carried out by solving the classical equations of motion on the Born–Oppenheimer surface. In truly quantum mechanical simulations, even the motion of the light atoms should be treated quantum mechanically. The now obtained raw "*ab initio*" effective potential has to be tuned further to give a variety of liquid state properties, both thermodynamical and dynamical, in agreement with the experiment, making it finally a new water model.

Li^+ – Water Potential. In another application, the pair interaction potential between a Li^+ ion and water molecules has been derived in a similar way. The problem to make a right choice among all the existing empirical ion-water potentials is rather severe: for example, for a Li^+ ion, described by the additive pair-potential of Lennard–Jones type, the effective ion diameter σ varies as much as from 0.9 to 2.8 Å in different proposed potential models (see [48–51] and references therein). Such a big change in the effective diameter leads, for example, to a change of the hydration number from 4 to 6 (see [50]), and correspondingly from tetrahedral water coordination to octahedral coordination in the first hydration shell. Other types of ion-water potentials, with an exponential short-range repulsion [52], or the so-called 4–7 potential (in terms of powers of r) [53] and

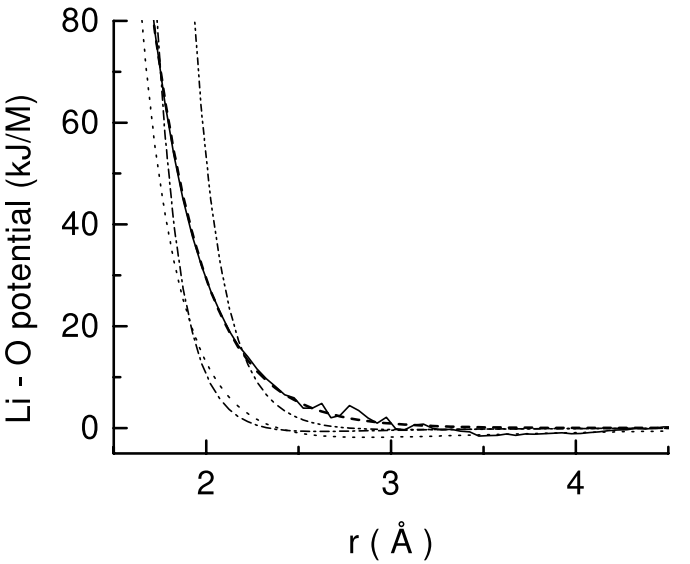

Fig. 3. Short-range part of the effective Li–O potential. The potential derived from *ab initio* simulation using the Inverse Monte Carlo (*solid line*), an exponential fit (*bold dashed line*), two potentials of the Lennard–Jones type from [48] (*dash-dotted line with one dot*) and from [49] (*dash-dotted line with two dots*), and the potential of 4–7 type from [53] (*dot line*)

.

potentials with more complicated functional forms have been suggested [54,55]. Obviously, different properties of the Li^+ hydration shell as well as the whole electrolyte solution are very sensitive to the chosen potential model.

The underlying Car–Parrinello simulation has been carried out for a single Li^+ ion and 32 water molecules [51]. The simulation setup was mainly the same as in the case of the liquid water described above (for details see the original paper [51]). The simulation time was extended to 20 ps after 5 ps of equilibration.

In principle, it would be possible to invert the whole set of radial distribution functions, obtained in this Car–Parrinello simulation and reconstruct both Li^+-water and water-water potentials. However, for liquid water there already exist several simple models which describe liquid water more or less well under ambient conditions. Therefore we constructed only the Li-water potential, describing water-water interactions by the well-known SPC model. Moreover, since the hydrogens in the SPC model have only electrostatic potential, we kept the same feature in the Li^+-water interactions. Thus, we only vary the Li–O potential, while employing pure electrostatic interactions between water hydrogens and the Li^+ cation.

The result for the short-range part of the Li–O potential (electrostatic part subtracted) is shown in Fig. 3. This potential provides exactly the same Li–O RDF as the one generated in the original *ab initio* simulation. While this potential was fitted to a number of different functional forms, we found that the simple exponential form

$$V_{\text{eff}}(r) = A\,\exp(-Br)$$

with $A = 37\,380$ kJ / M and $B = 3.63\,\text{Å}^{-1}$ provided the best fit (see Fig. 3). In attempts to fit the potential to different power functions of $1/r$, we found that $1/r^6$ or $1/r^7$ repulsive potentials provided the best fit, but were less good than the above exponential form.

For comparison three other Li^+–O potentials are also shown: two of the Lennard–Jones type from [48] and [49] and a "4–7" type of potential recently suggested by Periole et al. [53]. Clearly, the Lennard–Jones cases have a much stronger repulsive core. It is also interesting that all potentials based on the *ab initio* computations [52,56,57] have a less stiff repulsive part – either exponential (which do not even approach infinity at a zero distance) or a power-law decay with a smaller exponent than in the Lennard–Jones potential. The longer-range repulsive part of the derived potential is responsible for the lower first minimum and correct position of the second maxima of the Li–O RDF, which are impossible to satisfy with a Lennard–Jones potential.

Reliability of the *Ab–Initio* Effective Potentials. The effective potentials calculated by the Inverse Monte Carlo method provide exactly the same structural properties as those obtained through the underlying *ab-initio* simulations. The feasibility of the effective potentials to describe thermodynamical and dynamical properties is still a matter of further research. For the liquid water potential, derived in [40], the internal energy was shown to be -31 kJ / M, compared with the experimental value of -41 kJ / M. The Li^+–water potential obtained in [51] yields the solvation free energy -419 kJ / M, while the experimental value is -475 kJ / M. While it is not possible to speak about a perfect agreement, the results may still be considered as good when we keep in mind that no adjustable parameters were used in the derivation of the effective potentials.

There are several approximations affecting the accuracy of the obtained effective potentials. The most significant of these are:

1. several approximations within the Car–Parrinello method itself and in the underlying density functional theory
2. statistical uncertainty of the RDFs, calculated in the Car–Parrinello simulations during a relatively short (10 – 20 ps) simulation time
3. small system size
4. limitations of the rigid water model
5. pair potential approximation

With these approximations in mind, the agreement with the experimental results can be considered as good. For the solvation free energy of a Li^+ ion, this method yields results of the same level of precision as those available from empirical models [51]. It is clear that further improvement of quantum-chemical simulation techniques (beyond the Born–Oppenheimer limit) and a substantial increase in computing power would make this method a powerful instrument to derive reliable atom-atom interaction potentials.

4.2 Effective Solvent-Mediated Potentials

Ion-Ion Potentials for Simple Electrolytes. The classical molecular dynamics computer simulations with site-site pair potentials can now be run for molecular systems

consisting of the order of 10^4 atoms. This corresponds to a linear system size of 50–60 Å. This size may be large enough for simulations of isotropic liquids of simple molecules. Other systems such as biological macromolecules, however, often require even larger simulation cells. The same can be said about simulations near phase transition points, or about studies of hydrodynamic phenomena. On the other hand, these large-scale simulations do not typically require atomistic resolution, and could be carried out within some simplified, or coarse-grained models.

One of the typical simplifications used in description of macromolecules is to substitute the solvent molecules by a continuum medium. For example, in the primitive electrolyte model the ions in the water solution are substituted by charged spheres moving in a dielectric medium with a proper dielectric constant. Evidently, this is a serious simplification at short distances (a few ångströms) around the ions where it is not possible to define a dielectric constant. Sometimes, good agreement between the experiment can be obtained using the primitive electrolyte model after adjusting the ionic radius. However, the adjusted parameter does not have any physical significance.

A reliable model for effective ion-ion interactions in aqueous solution must take into account the solvation structure around the ions. Such effective solvent-mediated ion-ion potentials can be constructed from the ion-ion radial distribution functions, generated in high-quality all-atomic molecular dynamics simulations, using the Inverse Monte Carlo method. This approach has been already suggested in our original paper describing the inverse Monte Carlo technique [39]. In subsequent papers [35,43], the effective solvent-mediated potentials for NaCl aqueous solution have been calculated with a greater precision and for a number of concentrations.

An example of solvent-mediated ion-ion potentials is presented in Fig 4. The underlying molecular dynamics simulations have been performed using the flexible SPC water model [17] and Smith–Dang parameters for Na^+ and Cl^- ions [58]. Simulation details are described in the original papers [39,43]. Observe that the effective potential makes 1–2 oscillations, thereby reflecting the molecular nature of the solvent, and then finally approaches the primitive model potential with a dielectric constant close to 80. At distances longer than 10 Å, the effective potential almost perfectly coincides with the Coulombic potential. This can be seen from Fig. 5, where the Coulombic part has been removed and only the short-range part of the effective potential is plotted.

Note also that the effective potentials differ from the potential of mean force (PMF) $\psi_{\mathrm{pmf}} = -k_{\mathrm{B}}T \ln g(r)$, which corresponds to the Kirkwood approximation for the N-particle correlation function. The potential of mean force is screened by the other ions in the system and decays as $(1/r)\exp(-r/r_D)$ rather than via a Coulombic potential. The potential of mean force is therefore not a suitable choice to present the ion-ion interactions within a continuum solvent model at finite ion concentrations.

Figure 5 shows the effective potentials obtained for 0.5 M and 1 M ion solution after 10 ns MD simulations in a box of linear size 39 Å. One can clearly see that the differences between the two concentrations are very small. This means that the effective potentials depend very little on the ion concentration, and one can use potentials calculated at one specific concentration to describe properties of ion solutions at other concentrations.

Recently, the ion-ion potentials have been successfully used in dynamical mesoscopic simulations through the Dissipative Particle Dynamics (DPD) method [59,60]. It

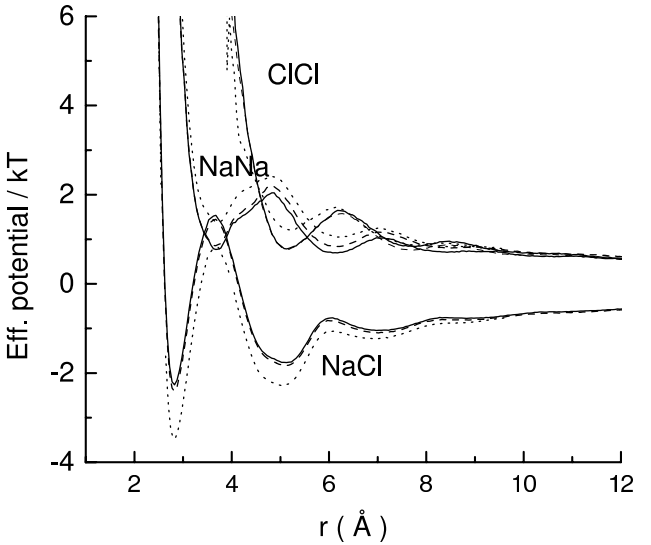

Fig. 4. Ion-ion effective potentials for NaCl ion solution. Ion concentrations: 0.5 (*solid line*), 1 (*dashed line*) and 4 M (*dotted line*)

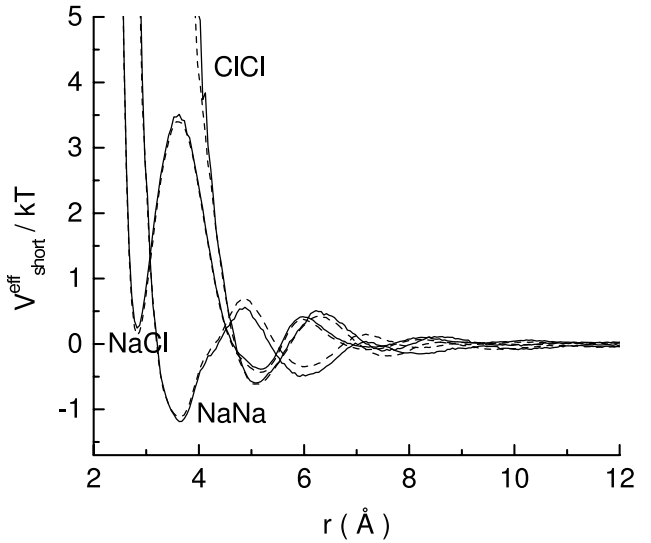

Fig. 5. Short-range part of the effective potential between ion-ion pairs in an aqueous NaCl electrolyte. The cases shown are 0.5 M (*continuous line*) and 1 M (*dashed line*). In subtracting the Coulombic potential we assumed a dielectric constant of 79 and 78 for 0.5 M and 1 M solutions, respectively

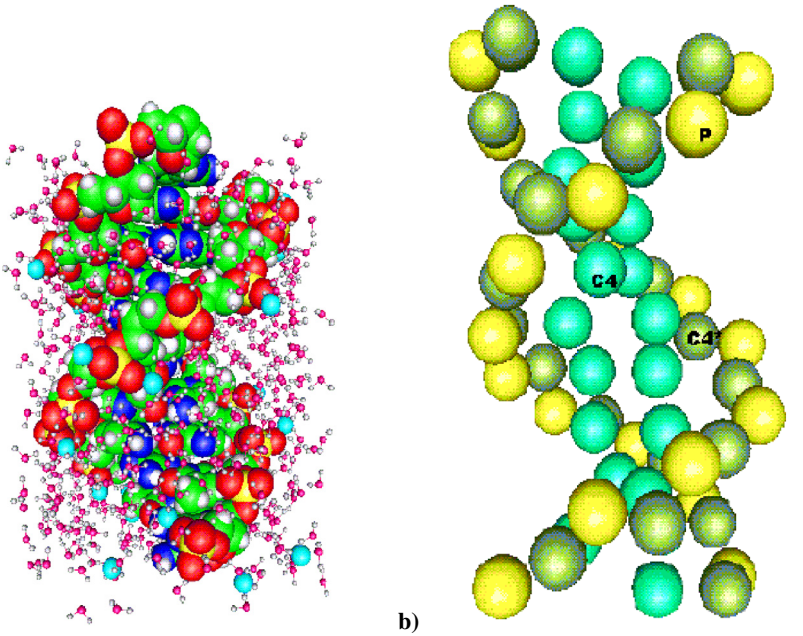

Fig. 6. (a) All-atom model of DNA. **(b)** Coarse-grained model of DNA in aqueous electrolyte solution

was found that the results for structural properties are reproduced accurately, and that transport properties like diffusion coefficients follow the correct temperature dependence very nicely. We are planning to extend the use of DPD simulations with effective potentials to more complex systems such as lipid bilayers. DPD is the obvious choice when effective potentials are used to calculate time-dependent properties.

Ion–DNA Effective Potentials. In the same way as for ionic solutions, one can introduce effective potentials between ions and a macromolecule, for example DNA. Such computations are described in our recent paper [61]. First, MD simulations of a (pseudo) infinitely long DNA model, presented as a periodically repeated fragment (10 base pairs $d(ATGCAGTCAG)_2$) of double-helix B–DNA, surrounded by water and different mono-valent ions in a rectangular box with periodic boundary conditions in all directions, have been carried out. Then, a simplified (coarse-grained) DNA model have been constructed in which three sites on each nucleotide monomer where selected (see Fig. 6) to describe ion-DNA interactions. These sites are the effective middle points of three building blocks of DNA: phosphorus atoms, $C4$ atoms of the bases and $C4'$ atoms of the sugars.

The phosphorus atom is a natural choice since the negative DNA charges are located on the phosphate groups. The choice of other sites is, however, somewhat arbitrary. It is expected that these sites should represent impenetrability of the DNA body for the ions. It is possible to choose additional sites on DNA to calculate RDFs and the effective potentials, in order to provide a more detailed description of the ion distribution within

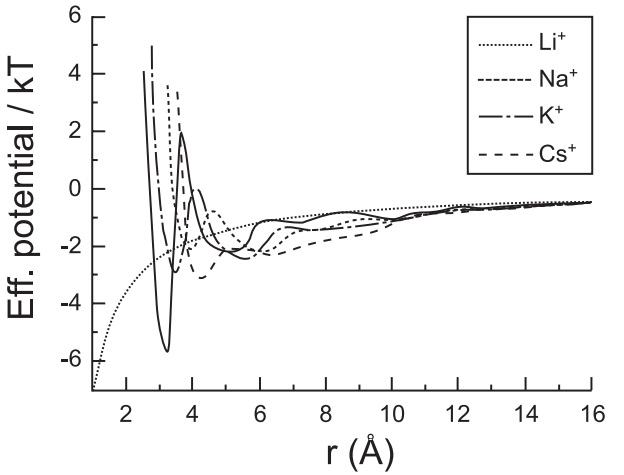

Fig. 7. Effective potentials between counter-ions and phosphorous atoms of DNA, including the Coulombic potential with $\varepsilon = 78$ (*thin dotted line*)

the simplified model. This specific choice of sites to construct the coarse-grained model can be considered as a minimal set of sites, still enough to reproduce the general features of ion binding properties and ion distribution around DNA.

A selected set of effective potentials for ion-DNA interactions are shown in Fig. 7. It can be seen that the effective potentials, as in the case of simple electrolyte solutions, have an oscillatory character representing the molecular nature of the solvent. We have seen that the effective potentials between the charged particles (or sites) approach the Coulombic potential at distances $r > 10\,\text{Å}$. The effective potentials between the uncharged DNA groups and ions tend to go to zero faster, reaching it already at distances around 8 Å. Also, the potentials have a less regular shape than the corresponding potentials between the ions and charged phosphates. Again, we want to stress that the displayed effective potentials should not be interpreted as potentials between the ions and specific DNA atoms, but rather as potentials between the ions and the whole functional group of DNA (phosphate group, sugar ring, nucleotide base), represented by the corresponding interaction center.

One important observation made in [61] was that the presence of the bulky DNA has only a weak influence on the effective potential between the ions. The influence was about the same order of magnitude as was caused by an increase of the salt concentration in pure electrolyte solutions. This observation shows a reasonable transferability among the effective potentials. The effective potentials, calculated at some conditions, can to a good approximation be used to describe other situations, e.g at different concentrations or in the presence of other components or solutes.

As an application of the ion-DNA effective potentials, we have calculated the relative binding affinities of different alkali ions to DNA. The problem of ion binding and ion association to DNA is biologically highly important because of the strong effect different ions exert on DNA. A problem of general interest could be formulated by asking how many ions of a given species can be found near the DNA surface, say, within 5 Å distance

(and be able to take part in specific ion-DNA interactions or reactions) at a given bulk ion concentration? All polyelectrolyte theories, calculating ion distributions and electrostatic potentials around polyions, are trying to provide the answer to this problem. To find the answer based on data from computer simulations, we must simulate both the region around the DNA and also the bulk solution. Moreover, the thermodynamic equilibrium between these two regions has to be reached. For physiological ion concentrations of the order of 0.1 – 0.2 M, the "bulk" conditions, i.e., a constant ion distribution and zero electrostatic potential are typically reached at distances of about 50 Å from the DNA; for lower salt concentrations this region should be even greater. In typical molecular dynamics simulations of DNA, the distance from the DNA surface to the border of the simulation box is on the scale of few ångströms. Therefore, such MD simulations allow one to obtain, for example, the answer to the question: where are the binding sites of different ions. However, the MD simulations would not allow one to address the question: what is the total amount of bound DNA ions, related to the ion concentration in the bulk. Currently, it would not be realistic to perform full-atomic MD simulations with a box length of the order of 100 Å, especially since just the equilibration of such large systems may well exceed the nanosecond time scale.

A coarse-grained model, based on our effective potentials could give us the answer to all the questions mentioned above. Monte Carlo simulations within the coarse-grained model have been carried out in a rectangular box of size $100 \times 100 \times 68\,\text{Å}^3$. Cylindrically averaged density profiles of different alkali ions around DNA were calculated. They are shown in Figs. 8(a) and (b). The binding affinities of alkali ions to DNA, as determined in the present work, follow the order: $\text{Cs}^+ > \text{Li}^+ > \text{Na}^+ \gtrsim \text{K}^+$. This order of binding affinities is in perfect agreement with experimental measurements performed by techniques of very different nature [61].

4.3 Effective Potentials for Macromolecules

As an additional example of coarse graining we can describe the computation of effective potentials between charged colloidal particles. This system is typically simulated as an "asymmetric electrolyte model" where the colloid particles are represented as highly charged and large spheres surrounded by small counterions and eventually coions, all interacting by a Coulombic potential with a suitable dielectric constant. As a step of coarse-graining, the ion component is removed and substituted by an effective interaction between colloid particles.

Such effective potentials for different concentrations of ions have been calculated in [62]. The radial distribution function between colloid particles were determined in Monte Carlo calculations of 80 spherical macroions of charge $-60\,\text{e}$ and radius 40 Å, and few thousands of small ions. Then the effective potentials were determined in the inverse MC procedure. It was found that in the case of monovalent ions the effective potentials could be nicely fitted to screened Coulomb (Yukawa) potentials, corresponding to the well-known DLVO (Derjaguin–Landau–Verley–Overbeek) theory of colloidal systems. However, in the case of divalent ions, a deep minimum in the effective potential appears at small distances between colloids (see Fig. 9). This minimum corresponds to an effective attraction between the macroions, which in some cases may cause a spontaneous aggregation of colloid particles.

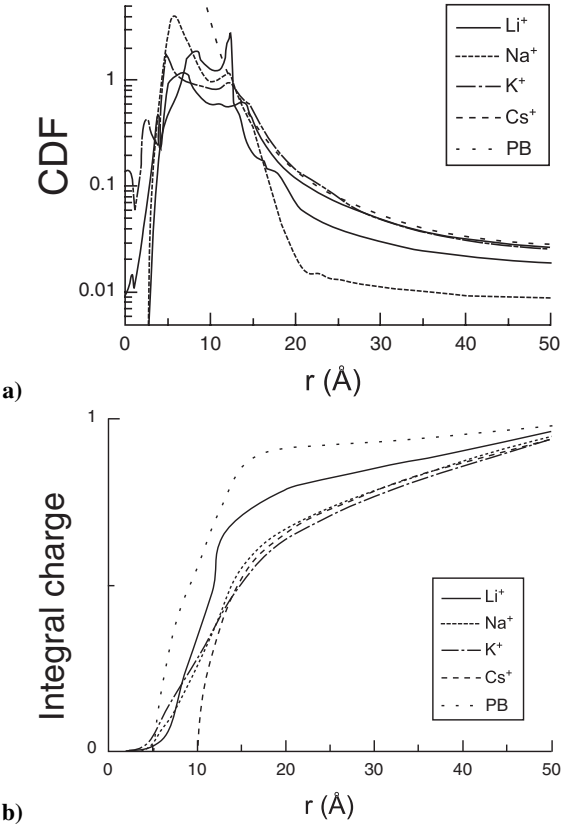

Fig. 8. (a) Cylindrically averaged density profiles of counter-ions around DNA for the case of salt-free solution, obtained in MC simulations with effective potentials. The solution of the Poisson–Boltzmann equation is included for comparison (*thin dotted line*). **(b)** The integrated charge around DNA, calculated from the cylindrical distribution functions

5 Summary

To summarize, the results from the above examples show how one can successively reduce molecular degrees of freedom in molecular computer simulations, going from the *ab-initio* description of electron structure to assemblies of macromolecules. Although we did not demonstrate it all the way for any of the examples, it could have been possible using the method of Inverse Monte Carlo. *Ab initio* MD simulations for DNA are still too expensive at the moment, although in principle they are possible. We are confident that with fine-tuning together with more powerful computers, it will be the ultimate tool for coarse-graining.

Acknowledgements

This work has been supported by the Swedish Science Council (VR).

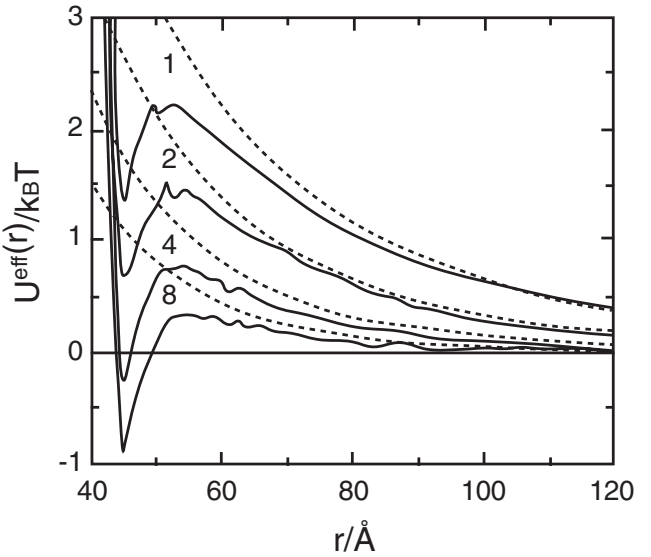

Fig. 9. Effective potentials between spherical macroions in the presence of divalent counterions at different polyion number densities in units of 10^{-6}Å^{-3}. For comparison, Yukawa fits to the long-range repulsive tail are also shown (*dotted lines*)

References

1. J.D. van der Waals: *Over de continuïteit van den Gas – en Vloieistoftoestand*. PhD thesis, University of Leiden, The Netherlands (1873)
2. M. Rigby, E.B. Smith, W.A. Wakeham, G.C. Maitland: *Intermolecular Forces, Their Origin and Determination* (Clarendon Press, Oxford 1981)
3. B.J. Alder, T.E. Wainwright: J. Chem. Phys. **27**, 1208 (1957)
4. A. Rahman: Phys. Rev. **136**A, 405 (1964)
5. A. Rahman, F.H. Stillinger: J. Chem. Phys. **55**, 3336 (1971)
6. J.A. McCammon, B.R. Gelin, M. Karplus: Nature **267**, 585 (1976)
7. P.K. Weiner, P.A. Kollman: J. Comp. Chem. **2**, 287 (1981)
8. B.R.B. Brooks, R.E. Bruccoleri, B.D. Olafson, D.J. States, S. Swaminathan, M. Karplus: J. Comp. Chem. **4**, 187 (1983)
9. W.F. van Gunsteren, H.J.C. Berendsen: *GROMOS Library Manual* (Biomos Nijenborgh 16, Groningen 1987)
10. W.L. Jorgensen, J. Tirado-Rives: J. Am. Chem. Soc. **110**, 1657 (1988)
11. T.L. Hill: J. Chem. Phys. **14**, 465 (1946)
12. J.D. Kemp, K.S. Pitzer: J. Chem. Phys. **4**, 749 (1936)
13. J.E. Lennard-Jones: Proc. R. Soc. London A **106**, 463 (1924)
14. L. Pauling: J. Am. Chem. Soc. **54**, 3570 (1932)
15. A.I. Kitaigorodskii: Tetrahedron **14**, 230 (1961)
16. P.M. Morse: Phys. Rev. **34**, 57 (1929)
17. K. Toukan, A. Rahman: Phys. Rev. B **31**, 2643 (1985)
18. R.A. Buckingham: Proc. R. Soc. London A **163**, 57 (1938)
19. A.D. Leach: *Molecular Modelling – Principles and Applications* (Addison–Wesley Longman, Edinburgh 1996)

20. A. Laaksonen, E. Clementi: Mol. Phys. **56**, 495 (1985)
21. A. Laaksonen, P.-O. Westlund: Mol. Phys. **73**, 663 (1991)
22. J.P. Ryckaert, G. Ciccotti, H.J.C. Berendsen: J. Comp. Phys. **23**, 327 (1977)
23. M. Tuckerman, B.J. Berne, G.J. Martyna: J. Chem. Phys. **97**, 1990 (1992)
24. P.G. Kusalik, A.P. Lyubartsev, D.L. Bergman, A. Laaksonen: J. Phys. Chem. B **104**, 9526 (2000)
25. A. Voegler-Smith, C.K. Hall: J. Chem. Phys. **133**, 9331 (2000)
26. J.A. McCammon, B.R. Gelin, M. Karplus: Nature **267**, 267 (1977)
27. M. Levitt: Cold Spring Harbor Symp. Quant. Biol. **47**, 251 (1983)
28. G.S.D. Ayton, S. Barderhagen, P. McMurtry, D. Sulsky, G.A. Voth: IBM J. Res. & Dev. **45**, 417 (2001)
29. J.P. Hansen, I.R. McDonald: *Theory of Simple Liquids* (Academic Press, London 1986)
30. R.L. Henderson: Phys. Lett. A**49**, 197 (1974)
31. R.L. McGreevy, L. Pusztai: Mol. Sim. **1**, 359 (1988)
32. L. Reatto, D. Levesque, J.J. Weis: Phys. Rev. A **33**, 3451 (1986)
33. Y. Rosenfeld, G. Kahl: J. Phys.: Condens. Matter **9**, L89 (1997)
34. P. Gonzalez-Mozuelos, M.D. Carbajal-Tinoco: J. Chem. Phys. **24**, 11074 (1998)
35. A.P. Lyubartsev, S. Marcelja: Phys. Rev. E **65**, 041202 (2002)
36. M. Ostheimer, H. Bertagnolli: Mol. Sim. **3**, 227 (1989)
37. A.K. Soper: Chem. Phys. **202**, 295 (1996)
38. G. Toth, A. Baranyai: J. Mol. Liquids **85**, 3 (2000)
39. A.P. Lyubartsev, A. Laaksonen: Phys. Rev. E **52**, 3730 (1995)
40. A.P. Lyubartsev, A. Laaksonen: Chem. Phys. Lett. **325**, 15 (2000)
41. R.H. Swendsen: Phys. Rev. Lett. **42**, 859 (1979)
42. G.S. Pawley, R.H. Swendsen, D.J. Wallace, K.G. Wilson: Phys. Rev. B **29**, 4030 (1984)
43. A.P. Lyubartsev, A. Laaksonen: Phys. Rev. E **55**, 5689 (1997)
44. O. Matsuoka, E. Clementi, M. Yoshimine: J. Chem. Phys. **64**, 1351 (1976)
45. R. Car, M. Parrinello: Phys. Rev. Lett. **55**, 2471 (1985)
46. A.K. Soper, F. Bruni, M.A. Ricci: J. Chem. Phys. **106**, 247 (1997)
47. H.J.C. Berendsen, J.P.M. Postma, W.F. van Gunsteren, J. Hermans: 'Interaction Models for Water in Relation to Protein Hydration'. In: *Jerusalem Symposia on Quantum Chemistry and Biochemistry*, Vol. 14, ed. by B. Pullman (Reidel, Dordrecht, Holland 1981) pp. 331–342
48. L.X. Dang: J. Chem. Phys. **96**, 6970 (1992)
49. K. Heinzinger: Physica B & C **131**, 196 (1985)
50. Y. Suwannachot, S. Hannongbua, B.M. Rode: J. Chem. Phys. **102**, 7602 (1995)
51. A.P. Lyubartsev, K. Laasonen, A. Laaksonen: J. Chem. Phys. **114**, 3120 (2001)
52. H. Kistenmasher, H. Porkie, E. Clementi: J. Chem. Phys. **59**, 5842 (1973)
53. X. Periole, D. Allouche, A. Ramirez-Solis, I. Ortega-Blake, J.P. Daudey, Y.H. Sanejouand: J. Phys. Chem. B **102**, 8579 (1998)
54. D.G. Bounds: Mol. Phys. **54**, 1335 (1985)
55. B.T. Gowda, S.W. Benson: J. Chem. Phys. **79**, 1235 (1983)
56. M. Migliore, G. Corongie, E. Clementi, G.C. Lee: J. Chem. Phys. **88**, 7766 (1988)
57. X. Periole, D. Allouche, J.P. Daudey, Y.H. Sanejouand: J. Phys. Chem. B **101**, 5018 (1997)
58. D.E. Smith, L.X. Dang: J. Chem. Phys. **100**, 3757 (1994)
59. A.P. Lyubartsev, M. Karttunen, I. Vattulainen, A. Laaksonen: Soft Materials **1**, 121, (2002)
60. M. Karttunen, A. Laaksonen, A.P. Lyubartsev, I. Vattulainen: unpublished
61. A.P. Lyubartsev, A. Laaksonen: J. Chem. Phys. **109** 11207 (1999)
62. V. Lobaskin, A.P. Lyubartsev, P. Linse: Phys. Rev. E **63**, 020401 (2001)