

# Hierarchical Multiscale Modelling Scheme from First Principles to Mesoscale

Alexander Lyubartsev<sup>1</sup>, Yaoquan Tu<sup>2</sup>, and Aatto Laaksonen<sup>1,\*</sup>

<sup>1</sup>Division of Physical Chemistry, Arrhenius Laboratory, Stockholm University, SE-106 91 Stockholm, Sweden

<sup>2</sup>Department of Natural Sciences, Örebro University, SE-701 82 Örebro, Sweden

We present a straight-forward implementation of a practical hierarchical multiscale modelling scheme which enables us to start from first-principles atomistic computer simulation and successively coarse-grain the model by leaving out uninteresting degrees of freedom. Using the Car-Parrinello method or our recently developed highly efficient tight-binding-like approximate density-functional quantum mechanical method, we first perform *ab initio* simulations. From these first-principles simulations we obtain a set of atomistic pair-wise effective interaction potentials to be used as a force field with *no* empirical data for subsequent classical all-atom simulations while scaling up the system size 2–3 orders of magnitude. The atomistic simulations similarly provide a new set of effective potentials but at a chosen coarse-grain level suitable for large-scale mesoscopic or soft-matter simulations beyond the atomic resolution. Show several examples are shown of how this scheme is done based on effective interaction potentials to tie together the various scales of modelling.

**Keywords:** DFT, Tight-Binding, Electronic Structure Calculations, *Ab Initio*, Inverse Monte Carlo, Coarse-Graining, Multi-Scale Modeling.

## CONTENTS

1. Simulations and the Real World . . . . .	1
2. Hierarchical Multiscale Modelling Approach . . . . .	3
3. Electronic Degrees of Freedom: <i>Ab Initio</i> Tight-Binding Like DFT Method . . . . .	5
4. Examples . . . . .	6
4.1. Derivation of Site–Site Atomistic Interaction Potentials from <i>Ab Initio</i> Simulations . . . . .	6
4.2. Effective Solvent-Mediated Ion–Ion and Ion–DNA Potentials . . . . .	7
4.3. Coarse-Grained Phospholipid Model for Bilayers and Vesicles . . . . .	8
5. Conclusions . . . . .	8
Acknowledgments . . . . .	9
References . . . . .	9

## 1. SIMULATIONS AND THE REAL WORLD

From the very early days of computer simulations three practical parameters have been, and still are, crucial in performing computer experiments. All of them are equally important and are assigned their values collectively based on the available computing resources at the moment. For the purpose of discussion we can illustrate them as they

span an operational space with three “orthogonal” axes as bases: (1) System *size* (2) Motional *time scale*<sup>a</sup> (3) The *accuracy* of the model

In moving inside this space of computer simulations it is necessary to choose an optimal region to keep the computations feasible while producing reliable results. “Compromise” is the key word. During the early decades of molecular computer simulations little effort was put to improve the methods and the models. This may have mainly been because of the phenomenally rapid development in the computer technology (steadily following the Moore’s law) helping the users to move the operational points further out along the three axes thereby making the simulations more and more realistic each year. Although limited in today’s standards, quite a few of the early simulation works are in many ways ground breaking, in particular how the analysis of the trajectories was based on statistical mechanics, thermodynamics and response theory in order to compare simulations with many real experiments. Most early molecular dynamics (MD) computer simulations of biomolecular systems, however, starting from a crystal structure can now be considered as rather meaningless simply because the *time* covered in

\*Author to whom correspondence should be addressed.

<sup>a</sup>In Monte Carlo simulations this corresponds to the number of moves.

the simulation was far too short to observe anything of value during the few hundred picoseconds MD simulations which were considered as the “state-of-the-art” during the late 80’s. In addition, too small simulation cells (*size*) and too short cut-off radii (*accuracy*) caused many artifacts and led to many wrong conclusions.

Today’s multi-core desk-top computers allow conventional MD simulations of systems containing order of

100 000 atoms to cover times to hundreds of nano seconds using particle mesh Ewald summation to treat the long-ranged Coulombic interactions. Although this is a tremendous improvement it is still not enough to study complex biological processes. Or to design new materials with desired properties which would require the simulations to be connected all the way from molecules to industry scale. During the last two decades several new



**Alexander Lyubartsev** Born in 1962 in Leningrad (now St. Petersburg) Russia. Graduated from Leningrad State University, Faculty of Physics in 1985, with specialization in statistical physics. In 1985–88 he was carrying out Ph.D. research on Monte Carlo simulation of polyelectrolyte models at the department of Molecular Biophysics, Leningrad University. He defended Ph.D. thesis in June 1988. In 1988–93 A. Lyubartsev was working as a junior researcher at the same department. In 1993 he moved to Sweden and began to work as a researcher at the Physical Chemistry division of Stockholm University. In 2001 he got a prestigious 6-year senior research position from the Swedish Science council, and from 2007 he is a professor at Stockholm University. The scientific work of A. Lyubartsev is in development of new advanced molecular modeling techniques and methodologies, such as free energy calculations using expanded ensembles, path integral Monte Carlo and molecular

dynamics for simulations of quantum particles, multiscale modeling techniques, with the application area in simulations of biological macromolecules and supramolecular structures, DNA and lipid membranes.



**Yaoquan Tu** was born in Shanghai, China, in 1962. He received a Bachelor’s degree (1984) and a Master’s degree (1987) in Chemical Physics from East China University of Science and Technology. In 2002, he received a Ph.D. degree in Physical Chemistry from Stockholm University, Sweden. After the postdoctoral work in the Division of Physical Chemistry at Stockholm University and the Theoretical Chemistry Department at Royal Institute of Technology, Sweden, he remained as a research associate in the Theoretical Chemistry Department at the Royal Institute of Technology in 2006. He is now an assistant professor in the School of Science and Technology at Örebro University, Sweden. His main research interests have covered the development of approximate quantum chemistry methods for large molecular systems, molecular properties in the interaction systems, theoretical modeling of nonlinear optical properties of organic molecular materials, and multi-scale modeling of biological systems.



**Aatto Laaksonen** (born in 1949 in Helsinki Finland), is Professor in Physical Chemistry at Arrhenius Laboratory of Stockholm University. He obtained Ph.C. and degree in Mathematics at Stockholm University 1976 and Ph.D. in Physical Chemistry with a thesis in Quantum Chemistry in 1981. In 1982 he worked at Daresbury Laboratory (UK) in the area of Quantum Chemistry in the group of Dr. Victor Saunders. In 1983–1985 He worked at the Computational Biophysics group of Dr. Enrico Clementi at IBM research laboratories in Poughkeepsie and in Kingston (USA) moving to the area of molecular simulations. In 1993/94 he spent a sabbatical year at Dalhousie University, Halifax (Canada), working in the groups of Professor Rod Wasylishen (NMR), Professor Peter Kusalik (StatMech) and Professor Russell Boyd (Quantum Chemistry) developing QM/MM simulation schemes. In 2002 and 2005 he has been working in Japan (at Japan Atomic Energy Research Institute)

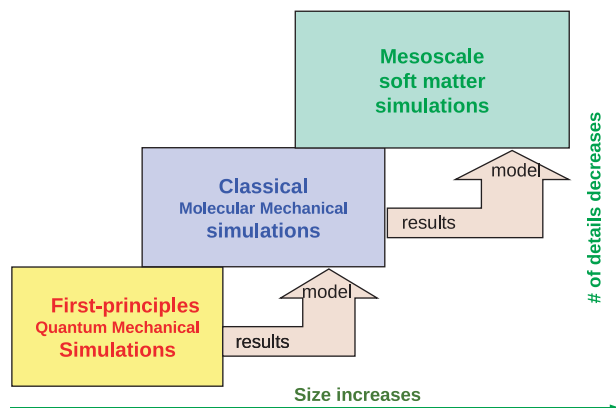
in Tokai, to carry out computer modelling of damaged DNA and corresponding repair enzymatic mechanism. Now he is moving towards materials science, being responsible of modeling work to design new nano and meso porous (including chiral) materials for separation and storage of gases and for heterogeneous catalysis processes inside the pores within the newly established center-of-excellence at Stockholm University where modern inorganic and organic syntheses are combined. Laaksonen’s research interests include simulation and modelling of many types of molecular systems from molecular nano clusters at atmospheric conditions, solutions and liquid mixtures (mixed solvents) and basically all types of biomolecular systems. A truly hierarchical multiscale simulation methodology from first-principles quantum chemistry to meso-scale soft-matter has been developed in Laaksonen’s group including a DFT method for large molecular systems based on *ab initio* tight-binding methodology. Laaksonen has published more than 200 scientific papers in the literature.

simulation methods have been introduced to stretch the *time* and *length* scales very much further. Also schemes were proposed to increase the *accuracy* by bring molecular dynamics (MD) to the domains of quantum mechanics (QM) and thereby to the first principles of physics. This allows MD simulations free of any empirical parameters (such as potential functions or molecular mechanical (MM) force fields) as input. The most widespread technique of this kind of methods is the Car-Parrinello molecular dynamics.<sup>1</sup> Hybrid methods mixing QM and MM based schemes<sup>2,3</sup> are also common tools today. The most interesting development may still be the schemes beyond atomistic resolution to model meso- and nano-scale systems and soft matters.<sup>4,5</sup> In other words there are now reliable simulation methods available to treat a system at three levels of physical description (QM, MM and mesoscopic soft matter), where the accuracy is successively decreasing while allowing the system length and time-scales to be increased. Examples of these are Car-Parrinello molecular dynamics, classical atomistic MD based on MM force fields, and dissipative particle dynamics (DPD). In common terminology, models beyond the atomistic resolution are result of coarse-graining (CG). There is no unique way to do coarse-graining within off-lattice framework. For heterogeneous systems like biological molecules some ways from *ad hoc* to parameterize CG potentials have been used<sup>6,7</sup> while for homogeneous systems, like in materials design, finite element and grid-based models are commonly employed. In the case of biological systems a coarse-grained description of water molecules surrounding biomolecules represents a great challenge. Such simplifications may include implicit description of the solvent with the help of solvent-mediated potentials<sup>8</sup> or coarse-grained representation of solvent molecules.<sup>7</sup> The problem is however how we can specify interaction potentials for such coarse-grained models.

In this paper, we first discuss a straight-forward hierarchical multiscale modeling approach which enables us to link together different levels (physical models) of simulations, ranging from electronic structure level to those of mesoscale. In the next section, an efficient *ab initio* electronic structure computation scheme is presented which provides substantial time saving comparing to the conventional DFT calculations. Thereafter several applications are presented: derivation of molecular-mechanical (MM) potentials from *ab initio* simulations, solvent-mediated potentials between ions and polyions in aqueous solutions, and coarse-grained modeling of lipids and lipid assemblies.

## 2. HIERARCHICAL MULTISCALE MODELLING APPROACH

In this section we describe a general approach how to use results of a more fundamental, more accurate theory, to construct simplified, coarse-grained models to be



**Fig. 1.** Hierarchical multiscale approach: Results of simulations of a more detailed model are used to build a model for simulation on a larger scale. Within the process, the size of the system increases while the level of details decreases.

used in large-scale simulations<sup>9</sup> according to the scheme depicted in Figure 1. The idea of the approach is the following: First, we carry out detailed simulations on a more fundamental, *ab-initio* level. Such simulations provide full information about the system, although on a limited length scale for reasons pointed out above. From these detailed simulations we can determine a set of radial distribution functions (RDFs) between sites representing interesting degrees of freedom. Then we ask: which interaction potential within a coarse-grained model would reproduce the very same set of RDFs as we obtained in the detailed model? To obtain the answer to this question, we need to solve an *inverse problem*: reconstruct interaction potentials from the previously obtained RDFs. If the inverse problem can be solved, we have in hand effective interaction potentials, which for the coarse-grained model reproduce the same structural properties of the system as detailed, *ab initio* model. This allows us to increase the length scale of the problem and thus to use the computed effective potentials for simulations on considerably larger scale.

It is clear that the key to this approach is the solution of the inverse problem. It is known from the theory,<sup>10</sup> that the solution of the inverse problem is unique in terms of pairwise potentials and RDFs. If we consider all possible potentials (not only pairwise), the solution of the inverse problem is no longer unique. However, from the computational point of view, we are interested just in pairwise solutions: the very aim of coarse-graining is the computational speed-up, and use of many body potentials would greatly hamper this goal. The pairwise potentials, obtained from the RDFs within the inverse approach, can be considered in some sense as “the best possible” pairwise approximation to the true many body mean force potentials.

There exist also alternative schemes to build effective coarse-grained potentials by targeting other than RDF properties of the detailed system. One of them is the “force matching” approach.<sup>11,12</sup> The idea is to fit pairwise potentials by minimizing the square deviation of the

force acting on each coarse grained site from its “exact” value obtained from the detailed model. Similarly, one can also think about fitting the potential energy.<sup>13</sup> In the case of linking *ab initio* and “atomistic molecular mechanical” levels of description, “exact” forces acting on individual atoms or energies are directly available from the *ab-initio* level of simulations, and the force- or energy- matching approaches can be used straight-forwardly. However on the next level of multiscale modeling, which binds the atomistic and coarse-grained levels of description, the exact potentials of mean force for coarse-grained units is typically unavailable, and some additional assumptions have to be made.<sup>14</sup> In this paper we concentrate on the multiscale modeling approach where the target is radial distribution functions.

The effective coarse-grained potentials can be derived from RDFs with the help of the Inverse Monte Carlo (IMC) method originally suggested by us in Ref. [8]. Briefly, the IMC method consists in the following. Assume that the Hamiltonian (potential energy) of the studied system can be represented as a sum of pairwise interactions:

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

where  $V(r_{ij})$  is a pairwise potential, and  $r_{ij}$  the distance between particles  $i$  and  $j$ . Let us first apply a grid approximation to digitalize the Hamiltonian:

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

for

$$r_\alpha - \frac{1}{2M} < r < r_\alpha + \frac{1}{2M};$$

$$r_\alpha = (\alpha - 0.5)r_{\text{cut}}/M; \alpha = 1, \dots, M \quad (2)$$

where  $r_{\text{cut}}$  is a cutoff distance and  $M$  is the number of grid points within the interval  $[0, r_{\text{cut}}]$ .

Then we can rewrite the Hamiltonian (Eq. 1) as:

$$H = \sum_{\alpha} V_{\alpha} S_{\alpha} \quad (3)$$

where  $S_{\alpha}$  is the number of pairs between the particles with their mutual distances inside  $\alpha$ -slice.  $S_{\alpha}$  is an estimator of the radial distribution function:

$$\langle S_{\alpha} \rangle = 4\pi r^2 \rho(r) N(N-1)/(2V) \quad (4)$$

Equation (3) can be readily generalized for the case of different interaction sites, where index  $\alpha$  runs both over all pairs of sites and distance intervals and corresponding  $\langle S_{\alpha} \rangle$  represent a complete set of pair distribution functions. Even intramolecular bond, angle and torsion potentials can be represented in form (3). Then index  $\alpha$  runs even over possible (discretized) values of the bond lengths, angles and torsions, with  $V_{\alpha}$  values representing the corresponding bond, angle or torsion potentials and  $\langle S_{\alpha} \rangle$  corresponding to distributions over bond lengths, angles or torsions.

In all cases, average values of  $\langle S_{\alpha} \rangle$  can be acquired from a detailed simulation of a small system.

The average values of  $S_{\alpha}$  are also some functions of the potential  $V_{\alpha}$ , which is initially unknown. Anyway, one can write down expansion:

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

where the derivatives  $\partial \langle S_{\alpha} \rangle / \partial V_{\gamma}$  can be expressed using the exact statistical mechanics relationships:<sup>8</sup>

$$\frac{\partial \langle S_{\alpha} \rangle}{\partial V_{\gamma}} = -(\langle S_{\alpha} S_{\gamma} \rangle - \langle S_{\alpha} \rangle \langle S_{\gamma} \rangle) / (kT) \quad (6)$$

Equations (5) and (6) allow us to define interaction potentials  $V_{\alpha}$  from radial distribution functions  $\langle S_{\alpha} \rangle$  iteratively, starting from some trial potentials (for which the potentials of mean force can always be chosen). Conventional MC simulations with a trial potential are carried out in which averages of  $S_{\alpha}$  and their cross correlations  $\langle S_{\alpha} S_{\gamma} \rangle$  are determined. Then a set of linear Eq. (5) without second-order term is solved which yield corrections to the trial potential. After each interaction, the potential is corrected according to:

$$V^{(i+1)} = V^{(i)} + \alpha \Delta V^{(i)} \quad (7)$$

where  $\Delta V^{(i)}$  is determined from (5) at each iteration ( $i$ ), and  $\alpha$  is a regularization parameter between 0 and 1. The procedure is repeated until convergence. More details on the inverse Monte Carlo procedure can be found in the earlier papers.<sup>8,9</sup>

There exist a few other approaches which can be used to invert RDFs. Soper introduced an “empirical potential structure refinement” method (also known as “iterative Boltzmann inversion”) in which pair potential is corrected at each iteration according to the mean field approximation:

$$V^{(i+1)} = V^{(i)} + kT \ln \frac{g^{(i)}(r)}{g_{\text{ref}}(r)} \quad (8)$$

Correction of potential according to (8) is straightforwardly to implement, and such an approach was used in a number of studies.<sup>15,16</sup> In cases when several different types of coarse-grained sites, and correspondingly several different potentials are involved, cross-correlations between RDFs according to Eqs. (6)–(7) need to be taken into account in order to provide convergence.

In some cases it is possible to solve the inverse problem using a numerical solutions of the liquid theory equations, for example Hypernetted-Chain (HNC) approximation.<sup>17</sup> In the case of solvent-mediated potentials between ions, HNC solution was found to provide very accurate solutions of the inverse problem, coinciding with the results obtained by the inverse MC simulations.<sup>18</sup> We should also mention a few other works devoted to the inverse problem.<sup>19–22</sup>

### 3. ELECTRONIC DEGREES OF FREEDOM: AB INITIO TIGHT-BINDING LIKE DFT METHOD

In this section we describe an approach which provides us with simulation data on the most detailed, *ab-initio* level. Electronic structure calculations are essential in the study of the properties of molecules and materials. In the past decades, many electronic structure calculation methods have been developed. Of the methods, those developed from the density functional theory (DFT)<sup>24–26</sup> have become very popular. Currently, DFT methods can be used to study a system of up to about 100 atoms with considerable accuracy. Despite such great advance, DFT calculations are still very time-demanding. For systems of several hundred atoms or more, such calculations are far too slow to be applied in practice. In recent years, tremendous progresses have been made in developing efficient and reliable approximate electronic structure calculation methods that can be used for a wide variety of purposes, for example, for modeling the forces on atoms in atomistic molecular dynamics computer simulations. One of such advances is the development of *ab initio* tight-binding-like (AITB) electronic structure methods which can be hopefully used as a general tool for electronic structure calculations.<sup>27–29</sup> We give short outline of our recent work on the development of such an accurate yet highly efficient AITB method.<sup>23</sup>

AITB methods usually start from the Harris-Foulkes functional<sup>30, 31</sup> which is equivalent to expanding the electron–electron interactions in the Kohn-Sham energy functional with respect to a reference density  $\tilde{\rho}$ , keeping only the zero-th order approximation and the first order correction, and neglecting the second order and higher order corrections. In order to make the calculations simpler and faster, we made a further simplification of the Harris-Foulkes functional. The simplified functional is given as

$$E_{\text{HF}} = \sum_i^{N_{\text{occ}}} f_i \varepsilon_i - \frac{1}{2} \iint \frac{\tilde{\rho}(\mathbf{r}_1) \tilde{\rho}(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_1 d\mathbf{r}_2 + E_{\text{xc}}[\rho^{(0)}] - \int \rho^{(0)}(\mathbf{r}) V_{\text{xc}}(\rho^{(0)}(\mathbf{r})) d\mathbf{r} + V_{I-I} \quad (9)$$

with

$$V_{\text{xc}}(\mathbf{r}) = \frac{\delta E_{\text{xc}}[\rho(\mathbf{r})]}{\delta \rho(\mathbf{r})} \quad (10)$$

where  $f_i$  is the occupation number on orbital  $i$ .  $E_{\text{xc}}$  is the exchange-correlation energy functional and  $V_{\text{xc}}$  is the so-called exchange-correlation potential.  $\rho$  is the electron density.  $V_{I-I}$  is the ion–ion interaction term. The orbital energy  $\varepsilon_i$  is found from the following one-electron orbital equation:

$$\left[ -\frac{1}{2} \nabla^2 + V_{\text{ext}}(\mathbf{r}_1) + \int \frac{\tilde{\rho}(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 + V_{\text{xc}}(\rho^{(0)}(\mathbf{r}_1)) \right] |\psi_i(\mathbf{r}_1)\rangle = \varepsilon_i |\psi_i(\mathbf{r}_1)\rangle \quad (11)$$

where  $V_{\text{ext}}$  is the external potential. Compared to the original Harris-Foulkes functional,<sup>30, 31</sup> our simplification is in the exchange-correlation part. Equation (9) is equivalent to expanding the Coulomb and exchange-correlation interaction terms in the Kohn-Sham energy functional with respect to  $\tilde{\rho}$  and  $\rho^{(0)}$ , respectively, and neglecting all the second order and higher order corrections. Thus, the error in the total energy caused by such simplification is also only in the second order and would have only a minor effect on the calculation results.

The reference electron densities  $\tilde{\rho}$  and  $\rho^{(0)}$  we used are super-positions of spherically distributed atomic-like densities having the forms

$$\tilde{\rho}(\mathbf{r}) = \sum_I \tilde{\rho}_I(r_I) \quad (12)$$

and

$$\rho^{(0)}(\mathbf{r}) = \sum_I \rho_I^{(0)}(r_I) \quad (13)$$

respectively, with

$$\tilde{\rho}_I(r_I) = \rho_I^{(0)}(r_I) + \Delta n_I f_I(r_I) \quad (14)$$

where  $\rho_I^{(0)}$  is the valence electron density of neutral atom  $I$  and  $f_I(r_I)$  corresponds to the density of a single electron in the highest occupied atomic orbital.  $r_I = |\mathbf{r} - \mathbf{R}_I|$  is the distance between  $\mathbf{r}$  and atomic site  $\mathbf{R}_I$ .  $\Delta n_I$  can be considered as the net number of electrons that atom  $I$  obtains in a molecular system and is determined by

$$\frac{\partial E_{\text{HF}}}{\partial \Delta n_I} = 0 \quad (15)$$

Equation (15) leads to a set of  $\{\Delta n_I\}$  required in solving Eq. (11). Therefore, in practical calculation,  $\{\Delta n_I\}$  and Eq. (11) are solved self-consistently.

Under the LCAO-MO approximation, molecular orbitals are linear combinations of atomic orbitals, that is,

$$\psi_i = \sum_{\mu} C_{\mu i} \phi_{\mu} \quad (16)$$

where  $\{\phi_{\mu}\}$  represent atomic orbitals. The coefficients  $\{C_{\mu i}\}$  and the orbital energies  $\varepsilon_i$  can be obtained by solving the following equation self-consistently:

$$\mathbf{FC} = \mathbf{SC}\varepsilon \quad (17)$$

with

$$S_{\mu\nu} = \langle \phi_{\mu} | \phi_{\nu} \rangle \quad (18)$$

and

$$F_{\mu\nu} = \left\langle \phi_{\mu} \left| -\frac{1}{2} \nabla^2 \right| \phi_{\nu} \right\rangle + \langle \phi_{\mu} | V_{\text{ext}} | \phi_{\nu} \rangle + \left\langle \phi_{\mu}(\mathbf{r}_1) \left| \int \frac{\tilde{\rho}(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 \right| \phi_{\nu}(\mathbf{r}_1) \right\rangle$$

$$\begin{aligned}
 & + \langle \phi_\mu(\mathbf{r}) | V_{xc}(\rho^{(0)}(\mathbf{r})) | \phi_v(\mathbf{r}) \rangle \\
 & = \left\langle \phi_\mu \left| -\frac{1}{2} \nabla^2 \right| \phi_v \right\rangle + \left\langle \phi_\mu \left| \sum_I V_I^{(PP)} \right| \phi_v \right\rangle \\
 & + \left\langle \phi_\mu(\mathbf{r}_1) \left| \sum_I \int \frac{\rho_I^{(0)}(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 \right| \phi_v(\mathbf{r}_1) \right\rangle \\
 & + \sum_I \Delta n_I \left\langle \phi_\mu(\mathbf{r}_1) \left| \int \frac{f_I(\mathbf{r}_2)}{r_{12}} d\mathbf{r}_2 \right| \phi_v(\mathbf{r}_1) \right\rangle \\
 & + \langle \phi_\mu(\mathbf{r}) | V_{xc}(\rho^{(0)}(\mathbf{r})) | \phi_v(\mathbf{r}) \rangle \quad (19)
 \end{aligned}$$

where  $V_I^{(PP)}$  is the pseudo-potential due to the nucleus and core electrons of atom  $I$ .

In our approach,  $\rho_I^{(0)}$  and  $f_I(\mathbf{r})$  are all expanded as linear combinations of 1S type gaussians by least-square fittings.  $V_I^{(PP)}$  corresponds to the norm-conserving separable dual-space pseudo-potential devised by Goedecker, Teter, and Hutter<sup>32</sup> for DFT calculations. The atomic orbitals  $\{\phi_\mu\}$  are also expressed as linear combinations of primitive gaussians. Therefore, it is clear that all the integrals, except for those related to the exchange-correlation potential and energy functional, can be expressed as closed forms and calculated analytically.

For the calculation of integrals related to the exchange-correlation potential or energy functional, the approximation of many-center expansion is adopted. For the off-center integrals, the same expansion formula as that given by Horsfield<sup>28</sup> is used, that is

$$\begin{aligned}
 \langle \phi_{I\alpha} | V_{xc}(\rho^{(0)}) | \phi_{J\beta} \rangle & = \langle \phi_{I\alpha} | V_{xc}(\rho_I^{(0)} + \rho_J^{(0)}) | \phi_{J\beta} \rangle \\
 & + \sum_{K(\neq I, J)} \langle \phi_{I\alpha} | V_{xc}(\rho_I^{(0)} + \rho_J^{(0)} + \rho_K^{(0)}) \\
 & - V_{xc}(\rho_I^{(0)} + \rho_J^{(0)}) | \phi_{J\beta} \rangle \quad (20)
 \end{aligned}$$

For the on-site integrals, we developed an improved many-center expansion scheme by including higher order terms in the original expressions of Horsfield. The improved integrals are given as

$$\begin{aligned}
 \langle \phi_{I\alpha} | V_{xc}(\rho^{(0)}) | \phi_{I\beta} \rangle & = \langle \phi_{I\alpha} | V_{xc}(\rho_I^{(0)}) | \phi_{I\beta} \rangle \\
 & + \sum_{J(\neq I)} \langle \phi_{I\alpha} | V_{xc}(\rho_I^{(0)} + \rho_J^{(0)}) \\
 & - V_{xc}(\rho_I^{(0)}) | \phi_{I\beta} \rangle + \delta V_{I\alpha, I\beta} \quad (21)
 \end{aligned}$$

and

$$\begin{aligned}
 & \int \rho_I^{(0)}(\mathbf{r}) \varepsilon_{xc}(\rho^{(0)}(\mathbf{r})) d\mathbf{r} \\
 & = \int \rho_I^{(0)}(\mathbf{r}) \varepsilon_{xc}(\rho_I^{(0)}(\mathbf{r})) d\mathbf{r} \\
 & + \sum_{J(\neq I)} \int \rho_I^{(0)}(\mathbf{r}) [\varepsilon_{xc}(\rho_I^{(0)}(\mathbf{r}) + \rho_J^{(0)}(\mathbf{r})) - \varepsilon_{xc}(\rho_I^{(0)}(\mathbf{r}))] d\mathbf{r} \\
 & + \delta E_{xc, I} \quad (22)
 \end{aligned}$$

with

$$\begin{aligned}
 \delta V_{I\alpha, I\beta} & \approx \frac{1}{2} \sum_{\substack{J(\neq I) \\ K(\neq I, J)}} \langle \phi_{I\alpha} | V_{xc}(\rho_I^{(0)} + \rho_J^{(0)} + \rho_K^{(0)}) + V_{xc}(\rho_I^{(0)}) \\
 & - V_{xc}(\rho_I^{(0)} + \rho_J^{(0)}) - V_{xc}(\rho_I^{(0)} + \rho_K^{(0)}) | \phi_{I\beta} \rangle \quad (23)
 \end{aligned}$$

and

$$\begin{aligned}
 \delta E_{xc, I} & \approx \frac{1}{2} \sum_{\substack{J(\neq I) \\ K(\neq I, J)}} \int \rho_I^{(0)}(\mathbf{r}) [\varepsilon_{xc}(\rho_I^{(0)}(\mathbf{r}) + \rho_J^{(0)}(\mathbf{r}) + \rho_K^{(0)}(\mathbf{r})) \\
 & + \varepsilon_{xc}(\rho_I^{(0)}(\mathbf{r})) - \varepsilon_{xc}(\rho_I^{(0)}(\mathbf{r}) + \rho_J^{(0)}(\mathbf{r})) \\
 & - \varepsilon_{xc}(\rho_I^{(0)}(\mathbf{r}) + \rho_K^{(0)}(\mathbf{r}))] d\mathbf{r} \quad (24)
 \end{aligned}$$

In the above equations,  $\alpha$  and  $\beta$  denote atomic orbitals and  $I$ ,  $J$ , and  $K$  denote atoms.  $\varepsilon_{xc}$  is the exchange-correlation energy density.  $\delta V_{I\alpha, I\beta}$  is the correction for the on-site potential integral and  $\delta E_{xc, I}$  is the correction for the energy integral.

In practical AITB calculations, the integrals required are often obtained by finding the corresponding integrals in local coordinate systems through the interpolation of the look-up tables and then by transforming the integrals from the local coordinate systems to the molecular coordinate system. Except in those for the one-center integrals, in look-up tables are integrals calculated on a pre-defined mesh of inter-atomic distances in the local coordinate systems. For more details of the theoretical background the reader is referred to Ref. [23].

## 4. EXAMPLES

### 4.1. Derivation of Site-Site Atomistic Interaction Potentials from *Ab Initio* Simulations

Pairwise atom-atom potentials for classical molecular dynamics simulations can be derived from RDFs computed in the above described AITB method, or in more conventional *ab initio*, for example Car-Parrinello molecular dynamics simulations. At today's level of computer power, it is possible to carry out Car-Parrinello *ab-initio* simulations for a few hundred of atoms on a several tens of picoseconds. As a first test of the suggested methodology, Car-Parrinello simulations of water (32 and 64 molecules) have been carried out, in which oxygen-oxygen, oxygen-hydrogen and hydrogen-hydrogen RDFs were computed. Then these RDFs used sent as input to our IMC procedure and the effective potentials for water were thus obtained. They came out rather similar to those of the conventional SPC water model, with some differences at short distances between atoms, see details in Ref. [34]. While the obtained water potential was not superior to the "standard" empirical water models like SPC or TIPnP (the above potentials were very thoroughly fitted to reproduce the properties of real water), this example demonstrated that it is possible



to obtain realistic interatomic potentials solely on the basis of *ab-initio* simulations.

In another example, the effective potential between  $\text{Li}^+$  ion and water was computed from *ab initio* RDFs. It turned out that non-electrostatic part of this potential has a simple exponential form, without any specific attractive interaction term:<sup>35</sup>

$$V_{\text{eff}}(r) = A \exp(-Br) \quad (25)$$

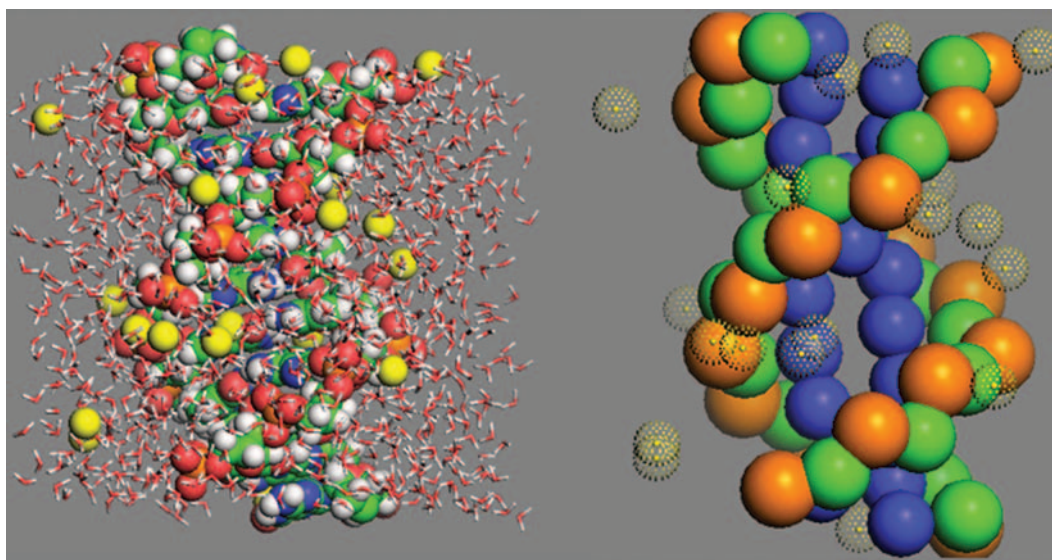
with  $A = 37380 \text{ kJ/M}$  and  $B = 3.63 \text{ \AA}^{-1}$ . This potential produces a hydration shell for the  $\text{Li}^+$  ion in a good agreement with available experimental neutral scattering data. Later this potential was used in classical molecular dynamics simulations of a Li-salt ion solution in which diffusional behavior of  $\text{Li}^+$  ions was studied.<sup>36</sup> These examples showed that such completely *ab initio* parameterization of interaction potentials can be used in cases where necessary experimental data to parameterize empirical potentials are not available.

#### 4.2. Effective Solvent-Mediated Ion–Ion and Ion–DNA Potentials

At the next level of coarse-graining, the explicit description of water molecules can be removed and substituted by effective solvent-mediated potentials between solute molecules, while atomistic molecular mechanical models of macromolecules can be substituted by their coarse-grained representation. Figure 2 shows an example of such coarse-graining in the case of ionic environment of DNA, where all the water molecules in the atomistic model, as well as atomistic details of the DNA structure, are removed, resulting in the coarse grained DNA model in the presence of solvated ions (right panel). The number of degrees of freedom is reduced by factor 50. The

effective solvent-mediated potentials between the solutes of the coarse grained model are computed by the inverse Monte Carlo method from the RDFs obtained in atomistic simulations.

The effective potentials were obtained first for  $\text{Na}^+$  and  $\text{Cl}^-$  ions in water solutions.<sup>8,37</sup> It turned out that at large distances (more than  $10 \text{ \AA}$ ), the effective Na–Cl potential is very close to the Coulombic potential in a media with dielectric permittivity of water ( $\epsilon = 80$ ), but at shorter distances it shows a few oscillations, reflecting the effects from the structure of water molecules. In further development, effective solvent mediated potentials between different alkali ions and DNA were computed from atomistic simulations of ions in a water layer around a DNA fragment.<sup>38</sup> These ion–DNA effective potentials were used in coarse-grained simulations of ion atmosphere around DNA. It was found that in these simulations, the relative binding affinities of different alkali ions to DNA follow the order:  $\text{Cs}^+ > \text{Li}^+ > \text{K}^+ \approx \text{Na}^+$ .<sup>38</sup> This order was just found previously in experimental studies of very different nature (Donnan equilibrium, ion exchange, NMR, circular dichroism) which represented a puzzle because of the location of  $\text{Li}^+$  ion between  $\text{Cs}^+$  and  $\text{K}^+$ . It is impossible to clarify such ordering with continuum solvent models only (such as the primitive electrolyte model), while atomistic simulations of a fully equilibrated ionic atmosphere around DNA are even now not feasible. However, our multiscale approach with effective solvent mediated potentials derived exclusively from atomistic simulations produced the experimental order of binding affinity automatically. It becomes also clear that the stronger affinity of  $\text{Li}^+$  ion (in comparison with  $\text{Na}^+$  and  $\text{K}^+$ ) is explained by specific hydration shells of DNA phosphate groups creating a favourable “pocket” for  $\text{Li}^+$  ions. This favourable



**Fig. 2.** Atomistic and coarse-grained representations of ionic environment of DNA. The number of degrees of freedom in the coarse grained model is reduced by factor 50.

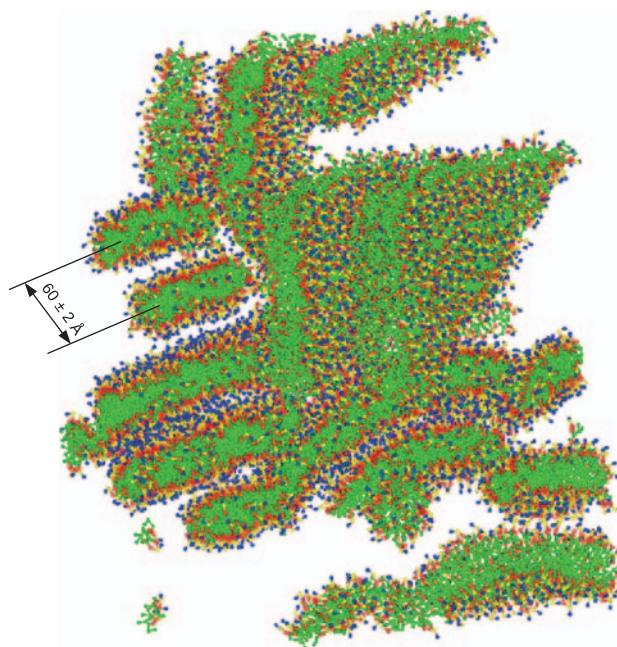
binding site was reflected in RDF-s computed in atomistic simulations, and then led to a more attractive effective solvent-mediated potential for  $\text{Li}^+$  ions.

#### 4.3. Coarse-Grained Phospholipid Model for Bilayers and Vesicles

Simulations of lipid membranes have attracted much attention during the last decade due to the fact that such membranes form outer shells of living cells. However, atomistic simulation of even a small piece of membrane consisting of about 100 lipids and surrounding water is a computational challenge, while many actual biophysical problems, such as studies of membrane mechanical properties, fusion, morphology, rafts formation, etc, require consideration of substantially larger membrane fragments. For investigation of all these phenomena in molecular simulations, coarse-grain level of modeling provides practically the only possible choice.

As another example of our multiscale modeling scheme, the effective solvent-mediated potentials for 10-sites CG model of DMPC lipid molecule have been constructed.<sup>39</sup> The starting point was a system of 16 lipid molecules dissolved in 1600 waters which has been simulated at the atomistic scale for 30 ns. The CHARMM27 force field with partial charges recalculated by fitting the electrostatic potential (ESP) from *ab-initio* Hartree-Fock calculations was used. The radial distribution functions as well as distributions of intramolecular distances between the coarse-grained sites were determined from these atomistic simulations. They were used as an input to the IMC procedure which yielded the interaction potentials, both inter- and intra-molecular, between all coarse-grained sites.

The coarse-grained model, derived in the way described above, was used in subsequent simulations of lipid systems on longer length- and time-scales, both within MC and MD simulations. First, it was demonstrated that the coarse-grained model provides the same structure of a plane bilayer as the atomistic model.<sup>39</sup> Then a number of other simulations, with the number of lipids in the range of 400–5000 and the system size of 200–500 Å was performed. It was shown that, depending on conditions, lipids organize themselves in different structures. If the number of lipids is small (less than 1000) the prevailing resulting structure is a bicell (a piece of bilayer of discoid shape). A larger bilayer fragment (with 3500 lipids) was found to spontaneously form a spherical vesicle. At larger lipid concentration, a tendency to form multi-lamellar structures was observed. An example of the formation of such a multi-lamellar structure is shown in Figure 3, which contains a snapshot of a system consisting of 5000 coarse-grained DMPC lipids in a periodic box of 400 Å, obtained after 50 ns Langevin molecular dynamics simulation started from a randomly initiated condition. The distance between the layers in the formed multi-lamellar stack was determined as 60 Å, in a perfect agreement with experimental



**Fig. 3.** Formation of multi-lamellar structures in Langevin dynamics simulation of 5000 coarse-grained DMPC lipids in a periodic cubic box of 400 Å. A snapshot after 50 ns of simulation started from random initial condition is shown. The choline groups of lipids are shown in blue, phosphates yellow, ester groups red and hydrocarbon tails green.

value for DMPC lipids. Such a distance means that there is a space of about 20 Å between the surfaces of neighboring layers, filled by water. It is remarkable that this distance is reproduced even in our coarse-grained model without explicit water. Thus the coarse-grained model, developed exclusively from all-atomic simulation data, reproduces well all the basic features of lipids in water solution.

## 5. CONCLUSIONS

A hierarchical true multiscale modelling approach presented here links together three levels of molecular modelling: *ab initio* molecular dynamics, classical molecular dynamics and meso-scale simulations. The method providing the link between these levels is the Inverse Monte Carlo approach. For the *ab initio* molecular dynamics, a highly efficient and accurate AITB scheme was recently developed. Thus the described methodology provides a consistent scheme to build molecular models for different scales without the need of empirical fitting of parameters. Some elements of this scheme were also demonstrated. Of course, there is still a very long way to go to define properties of molecules or materials exclusively *in silico*, and some tuning of the models against available experimental data is always an option. Also, transferability of the coarse-grained potentials need to be checked in every case. Nevertheless the suggested approach would increase the fraction of “*ab-initio*” derived features in molecular models in expense of “*ad hoc*” or “*empirically fitted*” ones,



which would enhance reliability of molecular simulations, increase their predictive power and open possibilities to address to new “large-scale” problems which are not yet considered in the “molecular simulation” domain.

**Acknowledgments:** This work has been supported by grants from the Swedish Science Council (VR) and computing grants from the Swedish National Infrastructure for Computing (SNIC).

## References

1. R. Car and M. Parrinello, *Phys. Rev. Lett.* 55, 2471 (1985).
2. A. Warshel and M. Levitt, *J. Mol. Biol.* 103, 227 (1976).
3. M. J. Field, P. A. Bash, and M. Karplus, *J. Comp. Chem.* 11, 700 (1990).
4. P. Espanol, *Phys. Rev. E* 52, 1734 (1995).
5. R. Benzi, S. Succi, and M. Vergassola, *Phys. Rep.* 222, 145 (1992).
6. R. D. Groot, *Langmuir* 16, 7493 (2000).
7. S. J. Marrink, A. H. de Vries, and A. E. Mark, *J. Phys. Chem. B* 108, 750 (2004).
8. A. P. Lyubartsev and A. Laaksonen, *Phys. Rev. E* 52, 3730 (1995).
9. A. P. Lyubartsev and A. Laaksonen, *Novel Methods in Soft Matter Simulations*, edited by M. Karttunen, I. Vattulainen, and A. Lukkari-nen, Springer, Berlin (2004), pp. 219–244.
10. R. L. Henderson, *Phys. Lett. A* 49, 197 (1974).
11. F. Ercolessi and J. Adams, *Europhys. Lett.* 26, 583 (1994).
12. S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth, *J. Chem. Phys.* 120, 10898 (2004).
13. G. Toth, *J. Phys.: Condens. Matter* 19, 335222 (2007).
14. S. Izvekov and G. A. Voth, *J. Phys. Chem. B* 109, 2469 (2005).
15. J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, and M. L. Klein, *J. Phys. Chem. B* 105, 4464 (2001).
16. D. Reith, M. Putz, and F. Muller-Plathe, *J. Phys. Chem. B* 109, 2469 (2005).
17. L. Reatto, D. Levesque, and J. J. Weis, *Phys. Rev. A* 33, 3451 (1986).
18. A. P. Lyubartsev and S. Marcelia, *Phys. Rev. E* 65, 041202 (2002).
19. M. Ostheimer and H. Bertagnolli, *Molecular Simulations* 3, 227 (1989).
20. Y. Rosenfeld and G. Kahl, *J. Phys.: Condens. Matter* 9, L89 (1997).
21. G. Toth and A. Baranyai, *J. Mol. Liquids* 85, 3 (2000).
22. N. G. Almaraz and E. Lomba, *Phys. Rev. E* 68, 011202 (2003).
23. Y. Tu, S. P. Jacobsson, and A. Laaksonen, *Phys. Rev. B* 74, 205104 (2006).
24. P. Hohenberg and W. Kohn, *Phys. Rev. B* 136, 864 (1964).
25. W. Kohn and L. J. Sham, *Phys. Rev. A* 140, 1133 (1965).
26. R. G. Parr and W. Yang, Oxford University Press, Oxford (1989).
27. O. F. Sankey and D. J. Niklewski, *Phys. Rev. B* 40, 3979 (1989).
28. A. P. Horsfield, *Phys. Rev. B* 56, 6594 (1997).
29. J. P. Lewis, K. R. Glaesemann, G. A. Voth, J. Fritsch, A. A. Demkov, J. Ortega, and O. F. Sankey, *Phys. Rev. B* 64, 195103 (2001).
30. J. Harris, *Phys. Rev. B* 31, 1770 (1985).
31. W. M. C. Foulkes and R. Haydock, *Phys. Rev. B* 39, 12520 (1989).
32. S. Goedecker, M. Teter, and J. Hutter, *Phys. Rev. B* 54, 1703 (1996).
33. A. A. Demkov, J. Ortega, O. F. Sankey, and M. P. Grumbach, *Phys. Rev. B* 52, 1618 (1995).
34. A. P. Lyubartsev and A. Laaksonen, *Chem. Phys. Lett* 325, 15 (2000).
35. A. P. Lyubartsev, K. Laasonen, and A. Laaksonen *J. Chem. Phys.* 114, 3120 (2001).
36. A. V. Egorov, A. V. Komolkin, V. I. Chizhik, P. V. Yushmanov, A. P. Lyubartsev, and A. Laaksonen, *J. Phys. Chem. B* 107, 3234 (2003).
37. A. P. Lyubartsev and Laaksonen, *Phys. Rev. E* 55, 5689 (1997).
38. A. P. Lyubartsev and A. Laaksonen, *J. Chem. Phys.* 111, 11207 (1999).
39. A. P. Lyubartsev, *Eur. Biophys. J.* 35, 53 (2005).

Received: 12 July 2008. Accepted: 7 September 2008.