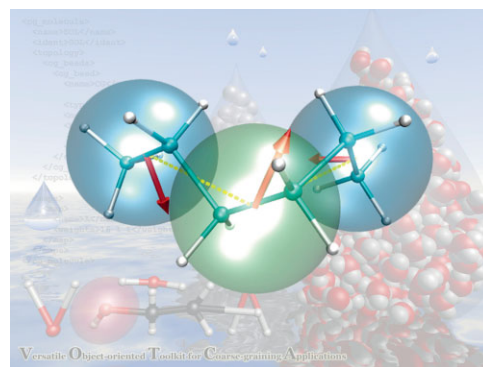


Hybrid Approaches to Coarse-Graining using the VOTCA Package: Liquid Hexane^a

Victor Rühle, Christoph Junghans

Several systematic coarse-graining techniques have been developed in recent years. The method of choice depends on the system of interest and the properties to be reproduced. We present three hybrid schemes to combine force-matching and Boltzmann inversion. The methods are tested on liquid hexane and the results are compared to iterative Boltzmann inversion. All approaches can easily be extended to mixtures of molecules and provide insight for parametrizing building blocks of bigger molecules. An implementation is provided in the VOTCA package which is available under an open source software license.



Introduction

In recent years, coarse-grained modelling has evolved into an important tool to extend the accessible range of time- and length-scales of particle based simulations of soft matter systems.^[1–10] A low resolution (coarse-grained) description is computationally more efficient due to three reasons. First, less interactions are involved which have to be evaluated. Second, interactions are usually smoother and thus simulations can be performed with a bigger timestep. Third, coarse-graining leads to an intrinsic speedup of the dynamics of the system.^[1,11,12]

Coarse-grained simulations range from generic models to bottom-up approaches. The latter provide a systematic link between two systems with different numbers of degrees of freedom. The process of deriving a low-resolution model based on a high resolution (reference) model is called *systematic coarse-graining*.

Various methods have been developed, depending on which properties need to be reproduced by the coarse-grained model. Structure based approaches aim to reproduce selected distribution functions (e.g. angle, bond or pair distribution functions) of the reference system. Examples are Boltzmann inversion,^[1] iterative Boltzmann inversion (IBI)^[4] and inverse Monte Carlo (IMC).^[3,13,14] However, these methods use only a subset of distribution functions and often neglect many-body correlations. The multiscale coarse-graining method (MS-CG)^[15,16] aims to match the full many-body potential of mean force. MS-CG is derived from liquid state theory using the variational principle. The resulting equations are identical to the force matching (FM) equations,^[17] but extensions are possible.^[18] In cases where, e.g. solvation free energies are studied, a parametrization based on free energies can be beneficial. One example is the MARTINI force-field for coarse-grained simulations of biomolecules such as lipids^[5] or proteins.^[19]

With the large variety of methods available, the versatile object-oriented toolkit for coarse-graining applications (VOTCA)^[20] was designed to offer a consistent implementation. This allows the evaluation of the different methods for new systems via a single tool. Currently, Boltzmann inversion, IBI, IMC and FM are implemented. In ref.^[20] we compared IBI, IMC and FM^[17] by coarse-graining four test systems, namely SPC/E water, methanol, liquid

Dr. V. Rühle, Dr. C. Junghans

Max Planck Institute for Polymer Research, Ackermannweg 10,
55128 Mainz, Germany

E-mail: {ruehle,junghans}@mpip-mainz.mpg.de

^a <http://www.votca.org>

propane and a single chain of hexane. Several problems were pointed out. For example, IMC can have convergence problems for small system sizes while force-matching may lead to unphysical results if the set of interaction which represent the potential energy surface of the coarse-grained force-field is incomplete. The latter was recently demonstrated for coarse-grained water, where a three-body potential is required to reproduce the pair distribution function when using the force-matching method.^[21]

For more complicated molecules, deriving all interactions at once can be complicated and it can be beneficial to treat specific interactions separately. One common example is the separation of bonded- and non-bonded interactions. Bonded interactions are stiffer and therefore often can be kept constant while parametrizing the non-bonded contributions.^[1,4] On a higher level, assembling molecules from building blocks which were derived for molecular fragments, is also possible.^[5,22,23]

In this work, we focus on the separation of bonded and non-bonded interactions using different hybrid approaches, specifically for cases where a single method fails. We combine FM with Boltzmann inversion and compare different approaches to separate bonded and non-bonded interactions in liquid hexane. As coarse-graining of a single chain of hexane in vacuum has already been performed in ref.^[20] to point out problems which can arise due to correlations and lack of interactions to represent the coarse-grained force-field, it is a good example to validate the separation ansatz, when different methods are used to develop interactions separately.

Methods

Methods for systematic coarse-graining, where Boltzmann inversion, IBI and force-matching are used in this work, have already been described in detail in various publications.^[1,4,7,15,16] Here, just a brief overview of the methods is given. In Hybrid Approaches section possible approaches to combine force-matching with potentials obtained via a different method are introduced.

Boltzmann Inversion

The simplest method is Boltzmann inversion.^[1] It is based on the fact that in a canonical ensemble, the distribution of independent degrees of freedom, $\{q\}$, satisfy a Boltzmann distribution,

$$P(q) \propto \exp[-\beta U(q)] \quad (1)$$

where $P(q)$ is a *normalized* distribution. These can be measured from a simulation and the potential can then be

determined by inverting the distribution function

$$U(q) = -k_B T \ln P(q) \quad (2)$$

As stated above, Equation (1) assumes that the degrees of freedom are independent. As can be seen from Equation (2), resulting potentials are in principle free energies. Using these for bonded interaction is often a reasonable assumption, however, non-bonded interactions need further refinement.

To compensate for many-body effects and entropic contributions, the potential can be corrected iteratively. A common correction is to use the difference in the potentials of mean force (PMF) of the desired (target) P^{target} and the current distribution $P^n(q)$ function,

$$\Delta U(q)^{n+1} = U_{\text{PMF}}^{\text{target}} - U_{\text{PMF}}^n = k_B T \ln \frac{P^n(q)}{P^{\text{target}}(q)} \quad (3)$$

For each iteration, a coarse-grained run is performed and the correction term evaluated. The method is repeated until convergence is reached.

Force Matching

The underlying idea of the force-matching method^[15,17] is to derive a coarse-grained force-field which reproduces the forces acting on the (mapped) coarse-graining beads. Technically, the fit is performed by minimizing the expression

$$\chi^2 = \sum_l^L \sum_i^M |F_{il}^{\text{ref}} - F_{il}^{\text{cg}}|^2 \quad (4)$$

where F_{il}^{ref} is the mapped reference force and F_{il}^{cg} is the coarse-grained force. In general, both have bonded- and non-bonded contributions. The sum L and M denote averaging over all snapshots and beads in the system, respectively.

In the context of coarse-graining, the method is often referred to as MS-CG. Furthermore, a derivation which is based on liquid state theory using the variational principle was given by Noid et. al.^[16] One important result of ref.^[16] is that special care has to be taken if not a centre of mass mapping is used or atoms are shared by several coarse-grained beads. In case of a centre of mass mapping, the total force on a coarse-grained bead is the sum of all forces of the associated atoms. Furthermore, additional terms to equation Equation (4) are possible, for example to match the pressure of the system.^[18]

An essential criteria to successfully apply the force-matching method is a complete set of interaction which

represent the coarse-grained force field. For water, where three-body correlations are important,^[24] a two-body potential is not sufficient to reproduce the atomistic radial distribution function,^[20] so an additional three-body potential is required.^[21]

Hybrid Approaches

In some cases, a single method does suffice to derive a coarse-grained model. For example, if one chooses an incomplete interaction space to represent the coarse-grained potential energy surface, unphysical potentials can be obtained when using force-matching.^[16,20] On the other hand, although success was reported for coarse-graining systems with an interface, IBI can be problematic for inhomogeneous systems or cases where the RDF is not well defined. In such cases, a combination of Boltzmann inversion for the bonded interactions and force-matching for non-bonded interactions can be useful, similarly as proposed in ref.^[25] These hybrid methods utilize the separation ansatz, which assumes that bonded and non-bonded interactions can be treated separately. Their extension to building blocks, i.e. separately parametrizing different molecules or fragments of molecules, is straight forward.

We now discuss three approaches for such a hybrid scheme, where the bonded interactions have already been derived by Boltzmann inverting the distribution functions.

(i) Full force-matching with *substitution*

In the *full force-matching method*, force-matching is performed on all interactions, bonded and non-bonded. Subsequently, bonded potentials are replaced by the Boltzmann inverted ones and only non-bonded force-matching potentials remain.

(ii) Force-matching with *subtraction*

The Boltzmann inverted bonded potentials are used to calculate forces of the mapped configuration, which are subtracted from the mapped reference forces. The remaining, non-bonded parts are then fitted via the force-matching method.

The benefit of the *subtraction method* is that it allows a decrease in the number of fitting parameters for the coarse-grained potentials. This can be favourable for large systems where solving the full force-matching equations Equation (4) is problematic. In addition, the method is consistent with the minimization condition, Equation (4), i.e. it tries to match the full forces which are present in the system.

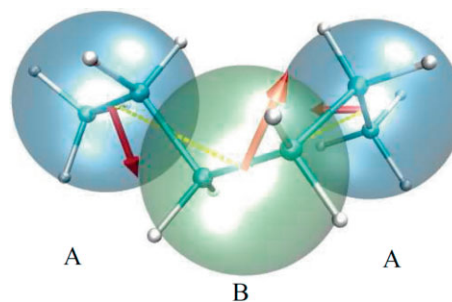


Figure 1. Mapping scheme for hexane. Each hexane molecule was coarse-grained into a three site model (A1–B1–A2), with two carbons per bead. Intramolecular interactions are given by bond- and angle potentials. The mapped intramolecular forces have an off plane component (red arrows) which leads to problems in the angle potential when using force-matching. The coarse-grained bond and angle forces both act within the plane of the molecule. Therefore, off plane contributions cannot be captured by this model. Taken from ref.^[20].

A problem of both of the aforementioned methods is that those force-contributions which cannot be captured by the coarse-grained force field (such as the off-plane components due to bonded interactions in hexane, see Figure 1 and Model System and Simulation Details section) persist and have to be compensated for by the fitted ones. This issue is overcome in the third approach,

(iii) Force-matching with *exclusions*

Only reference forces, that correspond to non-bonded interactions at coarse-grained level are recalculated for the reference trajectory. Then, force-matching is applied only to those recalculated forces. Bonded forces of the reference model and non-bonded forces, which contribute to bonded interactions at the coarse-grained level, are excluded.

Model System and Simulation Details

We use liquid hexane as a test system. Each (all-atom) hexane molecule is coarse-grained into a three bead chain, with two carbon atoms per bead (Figure 1). Different bead types are used for the inner (B) and outer beads (A). Each molecule has two AB-bonds, one ABA-angle and intermolecular non-bonded interactions, A–A, A–B and B–B.

Atomistic simulations were performed with 1 000 chains using an all-atom OPLS force field.^[26] The system was prepared at a density of $\rho = 0.65 \text{ g} \cdot \text{cm}^{-3}$ (cubic box of length $L = 6.042 \text{ nm}$) and equilibrated for 10 ns in a NVT simulation using the stochastic velocity rescaling thermostat.^[27] The reference data (forces + distributions) were generated in a

run of 20 ns, taking snapshots every 1.0 ps. No constraints were applied^b and a timestep of 1 fs was used.

The final configuration was mapped to coarse-grained level and used as the initial configuration for the coarse-grained runs. Bonded potentials were obtained by Boltzmann inverting the reference distribution functions of the liquid system (see Equation 3). It is important to note that the bonded reference distributions of liquid hexane are identical to those of a single molecule in vacuum given in ref.^[20]

Non-bonded potentials were obtained using IBI and the force-matching schemes described in Hybrid Approaches section. For IBI, the non-bonded potentials were refined for 500 iterations, each of length 0.5 ns, writing snapshots every 0.05 ps using a stochastic dynamics integrator. The cut-offs were chosen at $r_{AA} = 1.5$ nm, $r_{AB} = 1.35$ nm and $r_{BB} = 1.4$ nm. The force-matching schemes were performed on 4 000 snapshots in blocks of 20 frames. Forces were fitted via cubic splines with the intervals [0.35, 1.0] nm, [0.33, 1.0] nm and [0.37, 1.2] nm for A–A, A–B and B–B, respectively, and a grid spacing of 0.02 nm.

In the *full force-matching method*, bonded interactions were replaced by the IBI tables. For the *subtraction method*, the atomistic reference trajectory was mapped to coarse-grained level of detail and the bonded forces were calculated using the GROMACS rerun function with the IBI potentials. The force subtraction is done by VOTCA using the fully atomistic and rerun coarse-grained trajectories^c. For the *exclusion method*, a new GROMACS topology was created. All bonded interactions were removed and exclusions were added between those atoms, which correspond to bonded interactions on coarse-grained level (for this particular molecule, no intramolecular interactions are evaluated). Subsequently, the atomistic forces for the reference trajectory were recalculated using GROMACS rerun with the modified topology.

All methods discussed are implemented in the VOTCA package^[20] and will be available in release 1.1. Atomistic and coarse-grained runs were performed using GROMACS.^[28]

Results

To evaluate the different methods, we start with the discussion of the A–A potentials and associated distribution function shown in Figure 2 and 3, respectively. IBI perfectly reproduces the RDF. In addition, the

^b Constraints can lead to problems when using force-matching.^[16]

^c The implementation of the method was verified by plugging in the bonded potentials of the *full force-matching method*. The resulting non-bonded potentials were identical to the *full force-matching method*.

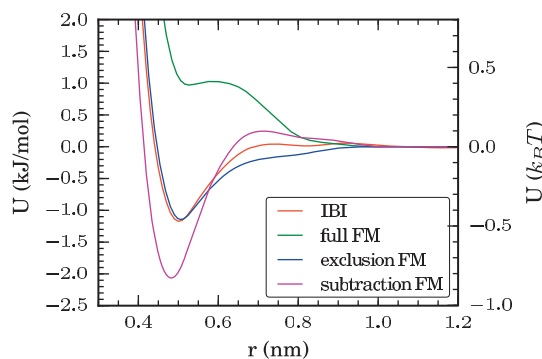


Figure 2. A–A potentials for the different methods. The interaction potentials between the A beads of IBI and the *exclusion method* are similar. The potential obtained by the *subtraction method* has a deeper minimum which results in overstructuring of the A–A RDF (see Figure 4). The potential obtained by the *full force-matching method* is purely repulsive and therefore cannot describe clustering of chains without imposing strong outer constraints on the simulation box.

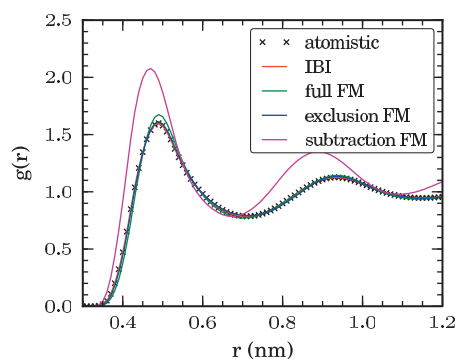


Figure 3. Radial distribution function of A–A beads. All methods except for the *subtraction method* lead to a good agreement of the A–A radial distribution function with the atomistic reference RDF. The deeper minimum of the *subtraction method* potential leads to overstructuring.

full force-matching method and the *exclusion method* show very good agreement, which indicates that pair potentials are sufficient for non-bonded interactions. The only method which does not reproduce the A–A RDF is the *subtraction method*, which exhibits overstructuring. Because of this, we omit this method from further discussions.

It is important to note that for IBI and the *exclusion method*, all three potentials, that is A–A, A–B and B–B, are very similar (see Figure 4). This reflects the chemical similarity of the A and B beads and suggests the possibility of a one bead type mapping scheme as discussed later. In contrast, the A–A potential from the *full force-matching method* is very different from the A–B and B–B potentials. Additionally, the purely repulsive shape (shown in Figure 2) is unphysical, since it cannot form structures without

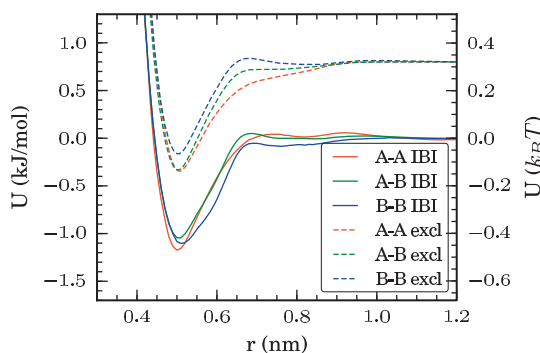


Figure 4. Non-bonded potentials of IBI and the *exclusion method*. The force-matching potentials are shifted by $0.8 \text{ kJ} \cdot \text{mol}^{-1}$ for better illustration. Both methods are in good agreement. Additionally, the different interaction potentials (A-A, A-B and B-B) are rather similar which reflects the close structural relation of the coarse-grained beads.

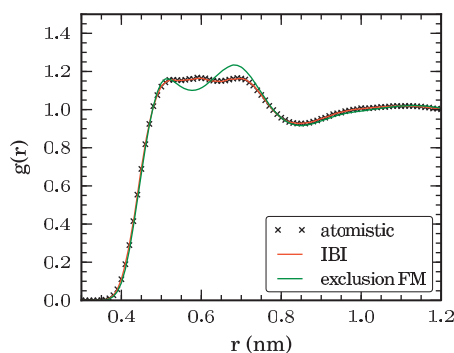


Figure 5. B-B radial distribution function. Both methods are in good agreement with the reference RDF. The exclusion method shows slight overstructuring of the B-B RDF.

imposing strong constraints on the system, i.e. the fixed volume of the NVT simulation. This discrepancy can be explained due to intramolecular contributions, which are not properly captured by the force-matching procedure.^[20] In force-matching, forces are optimized in a least square sense. Therefore, these contributions have to be compensated for by the non-bonded interactions. The bonded forces obtained from Boltzmann inversion lead to correct distributions, but are not necessarily compatible to the mapped atomistic ones (off-plane components). Therefore, in the *subtraction method*, a problem similar to the one in the *full force-matching method* appears. We conclude that neither the *subtraction method* nor the *full force-matching method* should be used if the set of bonded interactions of the coarse-grained force-field is incomplete.

From all hybrid methods studied in this work, the *exclusion method* seems to be the most appropriate. The resulting non-bonded potentials of all three interactions are shown in Figure 4. Their shape reflects the chemical similarity of respective beads and the structure of the system is well reproduced. In addition, the potentials are similar to the IBI potentials. The only distribution function which shows slight deviations is the B-B RDF as depicted in Figure 5. However, the difference is of the order of 0.1 and extends only over a small interval ($0.5 \text{ nm} < r < 0.7 \text{ nm}$).

The fact that all non-bonded interaction potentials are similar implies that one type of bead, in other words a single pair interaction type, is sufficient. Hence, we applied the *exclusion method* to a hexane molecule composed of three beads of the same kind (A-A-A). As can be seen in Figure 6, all distribution functions are in excellent agreement. Performing IBI with the A-A-A mapping scheme leads to similar results, hence only the potential is plotted in the top of Figure 6. We conclude that one bead type is sufficient to describe hexane, and most likely also longer alkyl chains

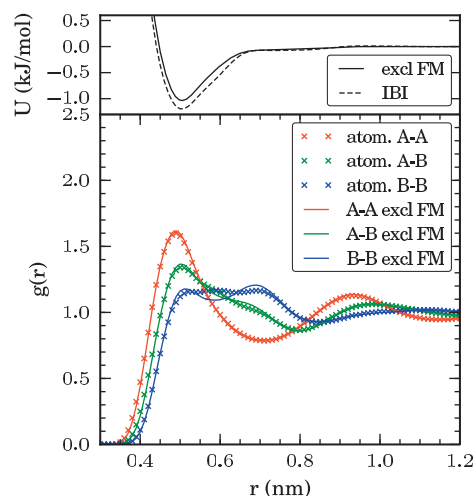


Figure 6. Potentials and RDFs for an A-A-A mapping scheme. The potentials of both methods, IBI and the *exclusion method*, are in perfect agreement. Only the force-matching RDFs are shown (bottom), both methods perfectly match the structure of the system.

which use such a 2:1 mapping. In this special case, the small differences in the potentials have no significant effect on the resulting distributions.

Finally, we may comment on the validity of the separation ansatz. This is intrinsically included in all hybrid approaches, since the bonded potentials (bond and angle) were kept constant when fitting the non-bonded interactions. Although the angle distributions agree well on coarse-grained and atomistic level, the coarse-grained bond distribution, depicted in Figure 7, differs from that of the atomistic reference liquid. Iterative corrections could be performed but are neglected here. In addition, the correlations of bonds and angles differ from that of the reference simulation. Therefore, cross correlation terms should be taken into account.

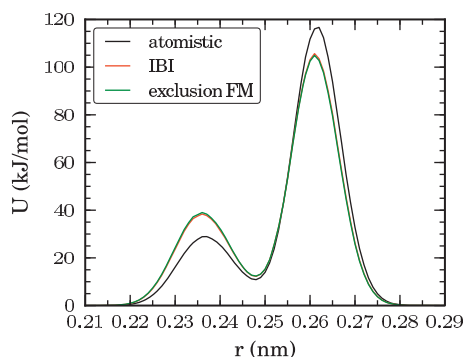


Figure 7. Bond distribution function. The bond distribution functions on the coarse-grained level differ from that of the reference system. Therefore, in a strong sense, the separation ansatz for this system not fully valid.

Conclusion and Outlook

Three hybrid methods to combine the force-matching method for non-bonded and IBI for bonded interactions were presented. All methods were tested on liquid hexane and resulting potentials were compared to IBI.

The best results were achieved when solely intermolecular contributions to forces were considered. This approach we referred to as the *exclusion method*. In contrast, the *full force-matching method* and the *subtraction method* either lead to an unphysical potential or could not reproduce the radial distribution function.

The potentials obtained using the *exclusion method* are similar to the ones from IBI. All pair distribution functions, A–A, A–B and B–B, are in excellent agreement with the reference RDF. Only the B–B distribution of the *exclusion method* shows some slight deviations. Furthermore, all three non-bonded potentials have similar shaped and a one bead type mapping scheme is sufficient to provide reasonable agreement of the distribution functions.

Applications for hybrid schemes go beyond the separation of bonded and non-bonded interactions. For example in more complicated molecules, parametrizing of all interactions simultaneously is impractical. Rather, interactions can be determined for fragments, so called building blocks, which reflect a chemical sub-structure for the full molecule.^[5,22,23]

All methods are implemented within the framework of the VOTCA package, release 1.1 and the input files of the simulations will be available as part of the tutorials.

Acknowledgements: We acknowledge Christine Peter for fruitful discussions during the entire project as well as Dominik Fritz and Mara Jochum for critical reading of the manuscript. CJ was financially supported by SFB 625 in the framework of the

multiscale modelling initiative of the Max-Planck Society (M³). Google Inc. is acknowledged for hosting the code projects.

Received: January 28, 2011; Revised: March 25, 2011; Published online: DOI: 10.1002/mats.201100011

Keywords: coarse-graining; computer modelling; molecular dynamics; molecular modelling; multiscale modelling; simulations

- [1] W. Tschöp, K. Kremer, J. Batoulis, T. Burger, O. Hahn, *Acta Polym.* **1998**, *49*, 61.
- [2] J. C. Shelley, M. Y. Shelley, R. C. Reeder, S. Bandyopadhyay, M. L. Klein, *J. Phys. Chem. B* **2001**, *105*, 4464.
- [3] A. P. Lyubartsev, M. Karttunen, I. Vattulainen, A. Laaksonen, *Soft Mater.* **2002**, *1*, 121.
- [4] D. Reith, M. Pütz, F. Müller-Plathe, *J. Comp. Chem.* **2003**, *24*, 1624.
- [5] S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, A. H. de Vries, *J. Phys. Chem. B* **2007**, *111*, 7812.
- [6] M. Praprotnik, L. Delle Site, K. Kremer, *Annu. Rev. Phys. Chem.* **2008**, *59*, 545.
- [7] G. A. Voth (Ed.), *Coarse-Graining of Condensed Phase and Biomolecular Systems*, Taylor & Francis, London **2008**.
- [8] T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, M. Karttunen, *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869.
- [9] C. Peter, K. Kremer, *Soft Matter* **2009**, *5*, 4357.
- [10] B. P. Lambeth, C. Junghans, K. Kremer, C. Clementi, L. Delle Site, *J. Chem. Phys.* **2010**, *133*, 221101.
- [11] V. A. Harmandaris, V. G. Mavrantzas, D. N. Theodorou, M. Kröger, J. Ramírez, H. C. Öttinger, D. Vlassopoulos, *Macromolecules* **2003**, *36*, 1376.
- [12] V. A. Harmandaris, K. Kremer, *Macromolecules* **2009**, *42*, 791.
- [13] A. P. Lyubartsev, A. Laaksonen, *Phys. Rev. E* **1995**, *52*, 3730.
- [14] A. K. Soper, *Chem. Phys.* **1996**, *202*, 295.
- [15] S. Izvekov, M. Parrinello, C. J. Burnham, G. A. Voth, *J. Chem. Phys.* **2004**, *120*, 10896.
- [16] W. G. Noid, J. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, H. C. Andersen, *J. Chem. Phys.* **2008**, *128*, 244114.
- [17] F. Ercolessi, J. B. Adams, *Europhys. Lett.* **1994**, *26*, 583.
- [18] A. Das, H. C. Andersen, *J. Chem. Phys.* **2010**, *132*, 164106.
- [19] L. Monticelli, S. K. Kandasamy, X. Periole, R. G. Larson, D. P. Tieleman, S. Marrink, *J. Chem. Theory Comput.* **2008**, *4*, 819.
- [20] V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, D. Andrienko, *J. Chem. Theory Comput.* **2009**, *5*, 3211.
- [21] L. Larini, L. Lu, G. A. Voth, *J. Chem. Phys.* **2010**, *132*, 164107.
- [22] A. Villa, N. F. A. van der Vegt, C. Peter, *Phys. Chem. Chem. Phys.* **2009**, *11*, 2068.
- [23] A. Villa, C. Peter, N. F. A. van der Vegt, *Phys. Chem. Chem. Phys.* **2009**, *11*, 2077.
- [24] V. Molinero, E. B. Moore, *J. Phys. Chem. B* **2009**, *113*, 4008.
- [25] Y. Wang, S. Izvekov, T. Yan, G. A. Voth, *J. Phys. Chem. B* **2006**, *110*, 3564.
- [26] W. L. Jorgensen, J. Tirado-Rives, *J. Chem. Soc., Abstr.* **1998**, *216*, U696.
- [27] G. Bussi, D. Donadio, M. Parrinello, *J. Chem. Phys.* **2007**, *126*, 014101.
- [28] B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4*, 435.