EPJ manuscript No.
(will be inserted by the editor)

Comparison of iterative inverse coarse-graining methods

David Rosenberger[1], Martin Hanke[2] and Nico F.A.van der Vegt[1,a]

[1] Eduard-Zintl-Institut für Anorganische und Physikalische Chemie and Center of Smart Interfaces, Technische Universität Darmstadt, Alarich-Weiss-Strasse 10, 64287 Darmstadt, Germany
[2] Institut für Mathematik, Johannes Gutenberg Universität Mainz, 55099 Mainz, Germany

Abstract. Deriving potentials for coarse-grained Molecular Dynamics (MD) simulations is frequently done by solving an inverse problem. Methods like Iterative-Boltzmann-Inversion (IBI) or Inverse Monte-Carlo (IMC) have been widely used to solve this problem. The solution obtained by application of these methods guarantees a match in the radial distribution function (RDF) between the underlying fine-grained system and the derived coarse-grained system. However, these methods often fail in reproducing thermodynamic properties. To overcome this deficiency, additional thermodynamic constraints such as pressure or Kirkwood-Buff-Integrals (KBI) may be added to these methods. In this communication we test the ability of these methods to converge to a known solution of the inverse problem. With this goal in mind we have studied a binary mixture of two simple Lennard-Jones (LJ) fluids, in which no actual coarse-graining is performed. We will further discuss whether full convergence is actually needed to achieve thermodynamic representability.

1 Introduction

Molecular Dynamics (MD) simulations can be used to follow the microscopic dynamics of atomic-scale processes up to several hundreds of nanoseconds. While this may be sufficient for many condensed phase systems, many properties of soft condensed matter are determined by processes on time and length scales much longer than that.[1–5] Owing to high computational cost of sampling a vast number of atomistic degrees of freedom, the relevant time and length scales in soft matter are often inaccessible with classical, atomistic MD models. To make these scales accessible with current computational power the number of degrees of freedom (DOF) of the models have to be reduced to a degree where a correct physical description of the system is still guaranteed. This process is commonly referred to as coarse-graining.

Coarse-graining can be achieved with bottom-up procedures in which degrees of freedom of an detailed-atomistic, or fine-grained (FG), system are systematically integrated out.[6,7] A major challenge in systematic coarse-graining is the representability

[a] e-mail: vandervegt@csi.tu-darmstadt.de

of the coarse-grained (CG) model, i.e. observables computed with the CG model and its parent FG model must be the same. In practice, however, full representability cannot be achieved because the multibody potential of mean force of the CG degrees of freedom is usually approximated by effective pair potentials. [8–16]

The focus of this work is on iterative inverse coarse-graining methods. Methods of this class derive effective pair potentials which aim to reproduce certain target properties of the FG system. Common target properties are the radial distribution functions (RDFs). The derivation of an effective pair potential reproducing the RDF constitutes an inverse problem. According to the Henderson theorem [17] this inverse problem has a unique solution, and the two most common methods to solve this problem are the Iterative Boltzmann Inversion (IBI)[11] and Inverse Monte Carlo[12] (IMC) methods. Both methods use an iterative scheme in which an initial guess for the effective pair potential is updated until convergence between the RDF of the CG system and the RDF of the FG system is achieved. The representability of the CG model however crucially depends on how well the effective pair potentials approximate the multibody potential of mean force associated with the coarse variables and has been discussed by Noid et al.[18,19] and others.[20–23] While some approaches use thermodynamic constraints, selected to represent quantities like the pressure in the iterative optimization procedure [11,24–26], other coarse graining approaches have emphasized properties of the pair potentials in terms of the underlying, effective, atomistic interactions.[21,27, 28] While application of thermodynamic constraints in iterative inversion methods may interfere with the Henderson theorem, in practice it is often considered more important to recover thermodynamic quantities than to satisfy the Henderson theorem. IBI-derived models have been derived over the last years for various soft matter systems, including polymers [11,29–31] and liquid crystals.[32]

A systematic study of the method itself was previously reported by Jain et al. and Fu et al.[33,34] They tested the capability of IBI to converge to a known solution for the inverse problem. In the two studies, a simple one-component Lennard Jones (LJ) fluid was taken as a reference or FG system. Both investigated the possibility to recover the potentials that were applied to generate the target RDFs without loosing any DOF. The study of Jain et al. shows that even after 1000 iterations full convergence in the potential has not been accomplished, whereas the structures have converged already after 10 iterations. This is due to the fact that the structure of such a simple fluid is mainly determined by short ranged repulsive interactions. This part of the interaction potential converges much faster than the attractive long range part. To account for the attractive part constraints can be added to the standard IBI scheme. Application of these constraints leads to a faster convergence towards the correct potential.[33]. Fu et al. extended the study.[34] They tested the influence of the cut-off and the capability of the effective pair potential to reproduce the isothermal compressibility. The authors suggest to choose the cut-off such that the oscillating long range part of the RDF is removed, since the method is insensitive to this part. Moreover, it can be shown that effective pair potentials based on IBI lead to lower isothermal compressibility compared to the FG system. With an additional constraint added to the method the compressibility increases compared to the FG system, but is closer to the target value.[34]

In this paper we extend these studies for a binary mixture of LJ fluids without loosing any DOF. We systematically show the strengths and weaknesses of the IBI method with and without additional constraints and compare them with the IMC method. We furthermore want to address to what degree convergence is needed to achieve

appropriate thermodynamic representability with the generated effective pair potentials.

2 Methods

2.1 Iterative Boltzmann Inversion

Generation of effective pair potentials for non-bonded interactions which guarantee structural agreement at a pair level between the FG and CG system can be achieved by application of the IBI technique [11]. Here, an effective pair potential is derived from the potential of mean force (PMF)

$$U_{ij}^{0}(r) = -k_B T \ln g_{ij}^{ref}(r) \tag{1}$$

between molecules of two species i and j separated by the distance r, where g_{ij}^{ref} is the reference RDF between the two species i and j, k_B is the Boltzmann constant and T is the temperature of the system. The initial guess is usually a poor choice and lacks the ability to reproduce the target RDF. For that U_{ij}^{0} gets updated (see Eq. 2) n-times until convergence between the target RDF and the RDF generated from the effective pair potential is reached.

$$U_{ij}^{n}(r) = U_{ij}^{n-1}(r) + k_B T \ln \left[\frac{g_{ij}^{n-1}(r)}{g_{ij}^{ref}(r)} \right] = U_{ij}^{n-1}(r) + \Delta U_{ij}(r) \tag{2}$$

Recent work by Ivanizki revealed that IBI is a modified Newton method [35] based on an approximation of the Jacobian

$$\frac{\partial \log g_{ij}}{\partial U_{ij}} = \frac{1}{g_{ij}} \frac{\partial g_{ij}}{\partial U_{ij}} \approx -\frac{1}{k_B T} I, \tag{3}$$

I denoting the identity matrix, which becomes exact in the low density limit where $g_{ij} = \exp(-U_{ij}/k_B T)$.

RDFs are predominately determined by short ranged repulsive interactions, i.e. a match in the RDF guarantees a match in the repulsive part of the potential between the FG and CG system. On the other hand, thermodynamic properties like pressure are mainly determined by long range attractive interactions, which are not well represented by IBI potentials. As a result, IBI potentials often overestimate the virial pressure of the system.

It should be emphasized that, in general, the incorporation of thermodynamical constraints violates the Henderson theorem, because there is only one effective pair potential that reproduces the RDF, and the thermodynamical constraints are either satisfied automatically, or they simply do not hold true for the Henderson potential.[17] Therefore, thermodynamical constraints can only be incorporated by relaxing the accuracy with which the RDF is being matched.

For IBI such a relaxation is difficult to realize because the whole iteration process is designed to target for a precise match of the RDF. Accordingly, it has been suggested to intertwine the IBI iteration with somewhat artificial correction steps to enforce

the validity of thermodynamical constraints. For example, to account for the correct pressure and to improve on the attractive interactions, an additional update

$$\Delta V_{ij}(r) = \alpha \left(1 - \frac{r}{r_{cut}}\right) \tag{4}$$

can be added to the standard IBI procedure [11].This correction is a simple linear ramp correction that becomes 0 at the cut-off r_{cut} and α at $r = 0$, where the exact form of α is discussed in section 3.2. From now this method will be referred to as P-IBI. Besides this simple linear polynomial there are more complex functions that can be used for ΔV_{ij} as well. These functions take the density of the system into account or try to ensure correct pressure via a variational principle.[24–26] Inclusion of the pressure correction changes Eq. 2 to:

$$U_{ij}^{n}(r) = U_{ij}^{n-1}(r) + \Delta U_{ij}(r) + \Delta V_{ij}(r). \tag{5}$$

It should be pointed out that every gain comes with a cost. The cost in the case of having matched pressure as well as the RDF is the loss of exact representation of the compressibility at a coarse-grained resolution.[24]

Besides the RDF another property can be introduced as a target property namely the Kirkwood-Buff integrals (KBI). These integrals can be related to a variety of other thermodynamic properties of multicomponent systems like activity coefficients, solvation free energies, partial molar volumes and the isothermal compressibility.[36] KBIs are integrals of the RDF over a finite volume [37,38], namely

$$G_{ij}^{V} = \frac{1}{V} \int_V \int_V [g_{ij}^{\mu VT}(r_{12}) - 1] d\mathbf{r}_1 d\mathbf{r}_2 \,. \tag{6}$$

In Eq. 6, $g_{ij}^{\mu VT}$ is the RDF in the grand-canonical ensemble. In the limit $V \to \infty$, the integral in Eq. 6 changes to [38]

$$G_{ij}^{\infty} = 4\pi \int_0^{\infty} [g_{ij}(r) - 1] r^2 dr \tag{7}$$

where G_{ij}^{∞} is the KBI in the thermodynamic limit. The KBIs (Eq. 7) should be formally calculated in grand-canonical simulations, but it was shown by several research groups that the calculation of KBIs is also possible under NpT or NVT conditions.[39–48] To this end, the running Kirkwood-Buff integral (RKBI), defined as

$$G_{ij}^{V}(r) = 4\pi \int_0^{r} [g_{ij}(s) - 1] s^2 ds, \tag{8}$$

is calculated and G_{ij}^{∞} is approximated by averaging the RKBI in a suitable range of r, typically between 1.0 nm and 2.0 nm in aqueous solution mixtures, where the RKBI reaches a limiting plateau value. The justification for this lies first within that density fluctuations are only local far away from the critical point and second within that correlation distances between the particles are small enough compared to the box size.[48]

The idea proposed by Ganguly et al. is to use RKBIs within the framework of IBI. This method is called Kirkwood-Buff Iterative Boltzmann Inversion (KB-IBI) [47]. The main objective of this method is to guarantee not only structural, but also thermodynamic representability at a coarse-grained level, for systems with multiple components. In case of KB-IBI the update is performed according to a ramp correction as [47]

$$\Delta KBI_{ij}(r) = \alpha'(G_{ij}^{n-1} - G_{ij}^{ref}) \left(1 - \frac{r}{r_{cut}}\right) \tag{9}$$

where G_{ij}^{ref} is the plateau value of the RKBI calculated from the reference simulations and G_{ij}^{n-1} is the plateau value of the RKBI after the $n-1$th step. The values for G_{ij} are obtained by averaging the RKBI in a range where the oscillations around its plateau value are small. The parameter α' is a system specific prefactor, which, as a good estimation, lies in the range between 0.01 and 0.1 kJ nm^{-3} mol^{-1}. [47] The KB-IBI ramp correction (Eq. 9) can be combined with the regular IBI update $\Delta U_{ij}(r)$, in a way similar to the pressure ramp correction in Eq. 5.

Finally, the running integral of the RDF can be considered to compute the update

$$\Delta C_{ij}(r) = k_B T \ln \left[\frac{\int_0^r g_{ij}^{n-1}(s) s^2 ds}{\int_0^r g_{ij}^{ref}(s) s^2 ds} \right]. \tag{10}$$

This method is called C-IBI according to Oliveira et al.[49] The potential gets updated n times until convergence between the running integrals of the FG and CG system is reached.

$$U_{ij}^n(r) = U_{ij}^{n-1}(r) + \Delta C_{ij}(r)\,. \tag{11}$$

2.2 Inverse Monte Carlo

The IMC method developed by Lyubartsev and Laaksonen provides another way to derive effective pair potentials which reproduce a target RDF. [12] In contrast to IBI, IMC is an exact Newton method [50] and aims to directly fit g_{ij} rather than $\log g_{ij}$ as compared to Eq. 2, i.e.,

$$U_{ij}^n = U_{ij}^{n-1} - A^{-1}(g_{ij}^{n-1} - g_{ij}^{ref}), \tag{12}$$

where

$$A = \frac{\partial g_{ij}}{\partial U_{ij}} \tag{13}$$

is the corresponding Jacobi matrix. As shown in [12,51,52] this matrix consists of cross correlation quantities; it can therefore be assembled during the forward simulation, but this requires longer simulation times to achieve sufficiently good statistics.

Instead of solving Eq. 12 exactly, one can reformulate this as a variational problem and add some penalty term to enhance stability: [53]

$$\Delta U_{ij} = \arg\min \|A\Delta U_{ij} - (g_{ij}^{n-1} - g_{ij}^{ref})\|^2 + \lambda \|R\Delta U_{ij}\|^2, \tag{14}$$

because A is often very close to being singular. Eq. 14 is known as Tikhonov regularization [54]; $\lambda > 0$ is the regularization parameter and R is the regularization operator, which, for example, can be the identity matrix. With this choice of R Eq. 14 can be rewritten as a linear system

$$(A^T A + \lambda I)\Delta U_{ij} = A^T (g_{ij}^{n-1} - g_{ij}^{ref}). \tag{15}$$

As mentioned before, the major advantage of IMC compared to IBI is the exact update scheme, in which the elements of the Jacobian matrix can be expressed in terms of fluctuation quantities following from statistical thermodynamics, whereas IBI uses an empirical update scheme. Another difference is that for a multi-component system every individual IMC potential update depends on the match of all target RDFs

simultaneously. This means that in case of a mismatch of just one of the RDFs, all interaction potentials that are present in the system are affected. In IBI this interdependence is not given for multi-component systems. Finally, thermodynamic constraints are fairly easy to incorporate into the IMC scheme by augmenting the scheme in Eq. 14 to the effect that the minimization process is restricted over ΔU_{ij} candidates that maintain a preassigned thermodynamic quantity, as it was previously done for the surface tension. [53] To enforce the correct virial pressure of a single component molecular fluid one could restrict the potential update ΔU of Eq. 14 to satisfy the orthogonality constraint

$$\int_0^\infty \Delta U'(r) r g^{ref}(r) r^2 dr = 0. \tag{16}$$

But these advantages come with some disadvantages. IMC needs more statistics to lead to reasonable results and therefore it requires higher computational cost per iteration. Furthermore, regions of poor sampling might perturb the stability of the algorithm and thus have to be removed.[55]

3 Simulation Setup

All simulations were performed with Gromacs 4.6.5.[56] For the generation of the effective pair potentials the Versatile Object-Oriented Toolkit for Coarse-Graining Applications (VOTCA version 1.3_rc1) [55,57,58] with Gromacs 4.6.5 [56] was applied.

3.1 Reference Simulation

The target RDFs were calculated from a binary mixture of Lennard-Jones particles (interaction parameters see Table 1). There were 1000 particles per type in a cubic box with an initial side length of 5 nm. The simulation time was 50 ns with a time step of 2 fs. For integration of the equations of motion the stochastic dynamics (sd) integrator of Gromacs 4.6.5 was used with a time constant of 0.5 ps^{-1}.[56] LJ interactions were treated with the cut-off method and were truncated at a distance of 1.2 nm. The simulation was run in a NpT ensemble at 1 bar and 85 K with the Parrinello-Rahman barostat ($\tau_p = 0.5$ ps).[59,60] Periodic boundary conditions were applied in x, y and z direction. The analysis of the RDFs was performed over the last 25 ns.

Table 1. Interaction parameters for the reference binary Lennard-Jones system

Interaction	σ [nm]	ϵ [kJ/mol]
LJ1-LJ1	0.340	0.98
LJ2-LJ2	0.296	0.84
LJ1-LJ2	0.317	0.91

3.2 Generation of Effective Pair Potentials for Non-Bonded Interactions via Iterative Inverse Methods

The target RDFs were Boltzmann inverted and used as an input for the methods described in section 2. It is crucial at this point to remind again that in the study

presented here no real coarse-graining is performed. The major objective is to test the ability of the methods to reproduce the generating potentials of the target RDFs. All simulations were run under NVT conditions taking the average box size from the reference simulation and its final configuration as the starting configuration. IBI simulations without any additional constraints were run for 1 ns per iteration for 300 iteration steps. The time step for the integration was 2 fs. As an integrator the sd integrator of Gromacs 4.6.5 was used.[56] There were three different cut-off values tested for the LJ interactions: 1.2 nm, 0.85 nm and 0.60 nm. All three interactions were updated according to Eq. 2 in each iteration step. The same setup was used for P-IBI. The pressure was corrected by application of Eq. 4 every third iteration, where an additional scaling factor, f, of 0.0003 is included in the pre-factor α (see Eq. 4). The pressure correction is applied in VOTCA version 1.3 according to [61]

$$\alpha = -\text{sign}(\Delta p) 0.1 k_B T \min(1, f|\Delta p|) \tag{17}$$

where Δp is the pressure difference between the actual value and the target value. For the KB-IBI update we followed the recipe of Ganguly et. al. presented in the supporting information of [47]. 30 preliminary iterations of P-IBI were followed by 50 iterations of KB-IBI update according to Eq. 9. The scaling factor α' was gradually decreased from 0.05 for the first 10 iterations to 0.03 for the next ten iterations and finally to 0.01 for the final 30 iterations. RKBIs were averaged between 0.5 nm and 0.85 nm. During the C-IBI steps the potential was updated according to Eq. 11. C-IBI was run for 100 steps and the integrals were evaluated between 0.2 and 0.85 nm.

The IMC simulations were performed without having any thermodynamic constraint included. The run time per iteration has been increased up to 5 ns. To guarantee stability of the IMC scheme regions of insufficient sampling were removed from each RDF (interval for RDF/interaction evaluation see table 2). Parameter like time-step and integrator were kept the same as in the IBI steps. The potential updates were performed in cycles, i.e only one interaction got updated per step. Further, we tested the influence of 30 preliminary steps of P-IBI followed by 4 IMC updates per interaction. The update scheme was additionally stabilized by the application of Eq. 14 with regularization parameter λ of different magnitudes (100, 300, 1000).

Table 2. Interval for RDF/interaction evaluation in the IMC scheme

Interaction	r(0) [nm]	r_{cut} [nm]
LJ1-LJ1	0.32	0.85
LJ2-LJ2	0.27	0.85
LJ1-LJ2	0.29	0.85

3.3 Calculation of Thermodynamic properties

The generated effective pair potentials have been applied to calculate the following thermodynamic properties: pressure, KBIs, isothermal compressibility, the partial molar volume and the derivative of the activity coefficient. To this end, 100 ns of MD simulations were performed in the NVT ensemble. For integration of the equations of motion the sd integrator was applied.[56] The time step for the integration was 2 fs. The reference temperature was 85 K. The cut-off for the LJ interactions was chosen

according to the cut-off within the generation procedure. This means for the derived potentials a cut-off of 0.85 nm was used. The analysis of the data has been performed over the last 90 ns.

4 Results and Discussion

In figure 1 the target RDFs for the three interactions (dashed lines) present in the system are shown together with their underlying LJ potentials (solid lines, U_{ref}).

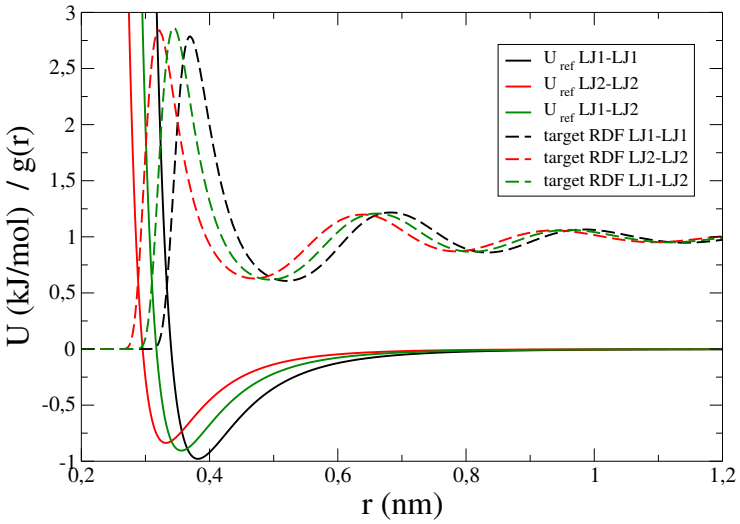

Fig. 1. Reference LJ potentials (solid lines) and target RDFs (dashed lines) for the three interactions present in the system: LJ1-LJ1 (black), LJ2-LJ2 (red) and LJ1-LJ2 (green).

4.1 Effective pair potentials via different variations of the Iterative Boltzmann Inversion method

The target RDFs were Boltzmann inverted and the different variations of the IBI method were executed as described in Section 3.2. For convenience the results are discussed based on the LJ1-LJ1 interaction. Since the remaining interactions in the system (LJ2-LJ2, LJ1-LJ2) show similar behaviour, the following discussion can be assigned to them accordingly.

In figures 2 and 3 the influence of the pressure correction on the convergence of the IBI method with a cut-off of 0.85 nm is shown. One can see clearly that convergence in the RDF has been reached after 50 steps for IBI (figure 2) as well as for P-IBI (figure 3). But in both cases the potential has not converged to the reference potential. Without pressure correction one obtains less attractive potentials after the IBI process in comparison to the reference potential. This is in agreement with the fact that RDFs are mainly determined by the repulsive part of the LJ potential. By approaching a match between the RDFs the short ranged repulsive part of the potential converges faster than the long ranged attractive part. As a consequence the virial pressure of the system is overestimated by the IBI potentials. As one can see

further, even after 300 steps the potential well is not matched, although an agreement in the RDF has been accomplished after 50 steps. Further iterations would lead to a closer approximation towards the reference potential, but full convergence cannot be reached within a reasonable number of iterations.

To account for the large virial pressure and the lack of attraction in the resulting potential the linear pressure correction (see Eq. 4) is applied. This correction in P-IBI leads to a final potential which is slightly more attractive than the reference potential, but it is in better agreement with it as the potential obtained from IBI. Moreover, the method seems to have converged to a stable solution after 80 steps, i.e. further iteration does not significantly improve the obtained potentials. What we have shown up to this point is in accordance to what is known from literature for single component LJ systems.[11,33,34] Since IBI treats each interaction independently this is not a surprise. Independent treatment means that an update of one interaction does not affect the other interactions in the system.[55]

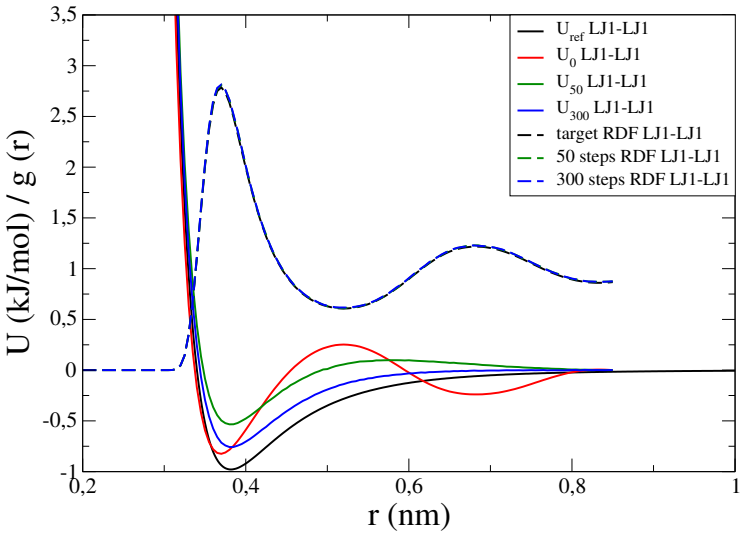

Fig. 2. IBI: RDFs (dashed lines) and potentials (solid lines) for the LJ1-LJ1 interaction: reference or target (black), the PMF (red), after 50 iterations (green) after 300 iterations (blue)

In the next step we investigate the influence of different cut-off values on the convergence of P-IBI. The results are shown in figure 4. In all three cases the RDF is well-matched after 50 iterations. If the cut-off is short (0.60 nm), the depth of the potential well closely matches with the reference potential, but the potential decays too rapidly beyond the minimum. If the cut-off is chosen to be as long as in the reference simulations (1.20 nm) a secondary minimum has evolved. As mentioned before, RDFs are mainly determined by short range repulsive interactions. Therefore, inclusion of long range correlations in IBI iterations leads to a poor convergence in that part of the potential and this is exactly what we observe for a cut-off of 1.20 nm. If the cut-off is chosen at the intermediate value (0.85 nm), the final potential is a bit more attractive compared to the shorter cut-off, but converges faster in the long range part compared to the longer cut-off.

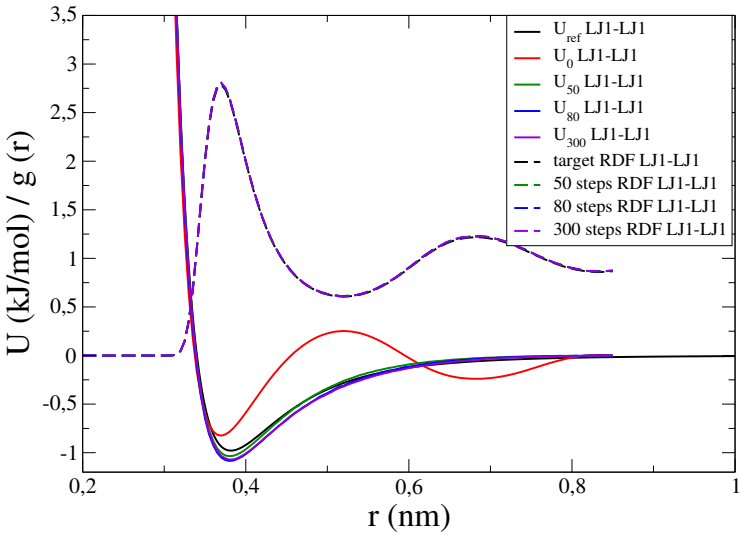

Fig. 3. P-IBI: RDFs (dashed lines) and potentials (solid lines) for the LJ1-LJ1 interaction: reference or target (black), the PMF (red), after 50 steps (green), after 80 steps (blue), after 300 steps (violet)

To determine the optimal cut-off distance for the LJ interaction throughout the IBI process, one should choose it long enough to get good representation of the potential beyond the minimum, but also short enough to exclude fluctuations on the very long distances, which lead to poor convergence of the method. Again this is in accordance to what is known from literature for a single component system.[34]

In the last step, KB-IBI and C-IBI updates have been investigated and compared to the previous results. As we see in figure 5, the RDF gets well reproduced by all IBI methods tested. By having a look at figure 6 one sees that P-IBI and IBI lead to larger RKBIs compared to the FG reference. The blue (IBI) and the violet (P-IBI) RKBIs are always above the black target RKBI. KB-IBI updates do not provide a significant improvement (red curve). If the update gets performed according to the C-IBI method, the RKBI is matched the best, especially in the longer range ($r > 0.60$ nm), but in this case the potential (solid green line in figure 5) becomes too repulsive compared to the reference. The reason why the KB-IBI method (solid red line in figure 5) does a better job in the reproduction of the potential is because the correction term shifts the potential directly by adding a ramp potential on top of it (Eq. 9). In C-IBI the correction has not direct relation to the potential, since the update is based on matching the integrated RDF over the whole range.

To summarize it, the C-IBI scheme reproduces the target RKBI best by construction. This is achieved, however, with a potential which deviates significantly from the reference atomistic potential. Potentials obtained from KB-IBI, P-IBI and IBI do not match the target RKBI in the long distance range but approximate the reference atomistic potential better. Agreement of, both, the pair potential and the RKBI with the atomistic references cannot be achieved with any of the IBI-related methods for the binary Lennard-Jones system studied in this work. Neither the KB-IBI method nor the C-IBI method provide an improvement over the P-IBI method in terms of the reproduction of the reference potential. We note that for the present Lennard-Jones system a complication arises from the fact that the RKBIs do not converge to a lim-

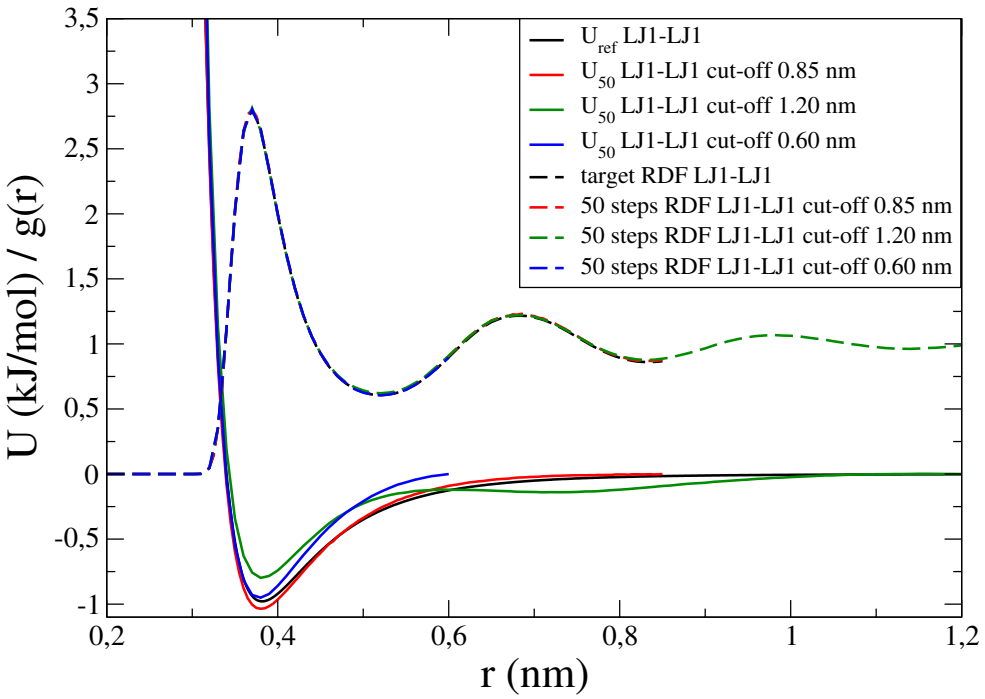

Fig. 4. Influence of different cut offs for P-IBI for the LJ1-LJ1 interaction: potentials (solid lines) and RDFs (dashed lines) target or reference (black), cut-off 0.85 nm (red), cut-off 1.20 nm (green), cut-off 0.6 nm (blue)

iting plateau value but keep on oscillating at long range. This makes it difficult to apply the KB-IBI update (see Eq. 9), which relies on a limiting plateau value being found in the RKBI.

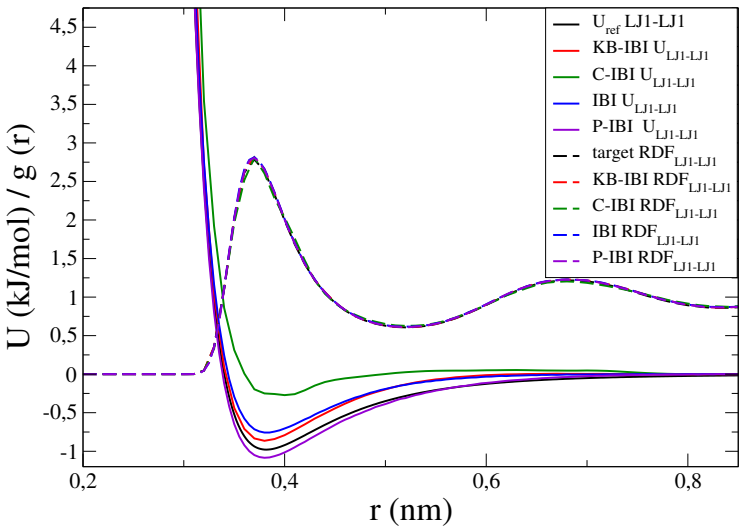

Fig. 5. Comparison of the different IBI methods with the atomistic reference (black) for the LJ1-LJ1 interaction, RDFs (dashed lines) and final potentials (solid lines) for KB-IBI(red), C-IBI (green), IBI (blue) and P-IBI (violet)

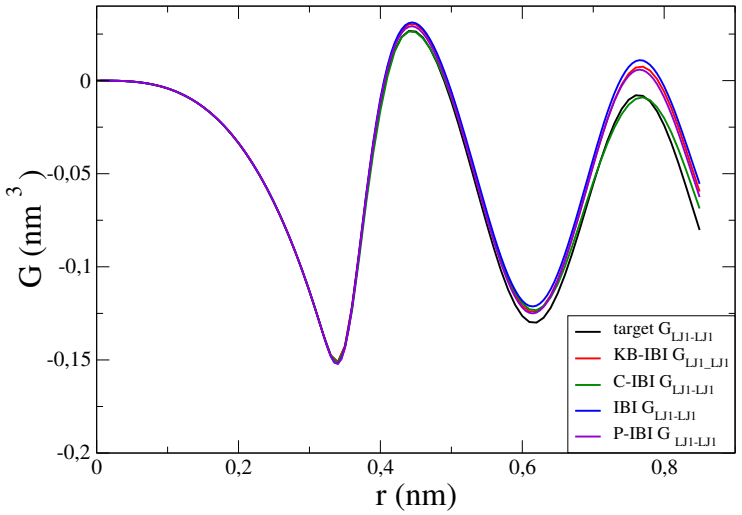

Fig. 6. Comparison of the RKBIs for the LJ1-LJ1 interaction: reference (black), KB-IBI (red), C-IBI (green), IBI (blue) and P-IBI (violet)

4.2 Effective pair potentials via the Inverse Monte-Carlo Method

Since the goal of IMC is the same as for IBI, i. e. generation of effective pair potentials by achieving a match in the RDF between the CG and FG system, comparable results are expected.

As we see from the red curve in figure 7, simple application of IMC leads to a complete mismatch in the RDF and a non-physical shape of the potential (figure 8). This results in a crash of the simulation after a few iterations. It is known that this is likely to happen if the initial guess, i.e. the PMF, is too far away from the solution

of the problem. To provide a better starting point 30 preliminary steps of P-IBI are performed. But from the green curve in figures 7 and 8 we do not see any further improvement. The algorithm cannot converge to a stable solution. For further stabilization we add a regularization term to the IMC update scheme according to Eq. 14.

A singular value decomposition (SVD) of the matrix A (see Eq. 13) allows an analy-

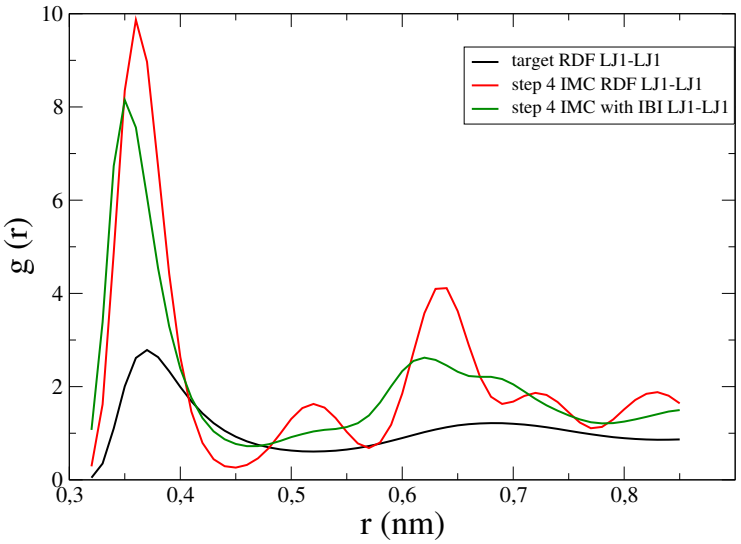

Fig. 7. Target RDF between LJ1-LJ1 (black line), RDF after 4 IMC steps without previous IBI (red line), RDF after 4 steps with previous IBI (green line).

sis on how large the regularization parameter λ has to be. The SVD decomposes the matrix A to

$$A = U\Sigma V^T = \sum_{k=1}^{n} \mathbf{u_k}\sigma_k \mathbf{v_k^T} \tag{18}$$

where Σ is a diagonal matrix with the singular values $\sigma_1, ..., \sigma_n$ of A and $V(\mathbf{v_1}, ...\mathbf{v_n})$ and $U(\mathbf{u_1}, ...\mathbf{u_n})$ are both unitary $n \times n$ matrices. Now substitution of Eq. 18 in Eq. 15 leads to

$$\Delta U_{ij} = \sum_{k=1}^{n} \frac{\sigma_k}{\sigma_k^2 + \lambda} \left(\mathbf{u_k^T}(g_{ij}^{n-1} - g_{ij}^{ref})\right)\mathbf{v_k}. \tag{19}$$

A good regularization parameter λ should be such that it dominates the smallest singular values (squared) but is itself small compared to the larger ones. The optimal λ parameter can only be determined if the exact solution is known, which unfortunately is not the case in practice. Finding good heuristic choices of λ is still a problem that is intensively studied in the mathematical community.

In our case the SVD analysis shows the smallest singular values of a magnitude of 100 or below. These are the values we considered to be regularized. So the IMC method is run with a regularization parameter λ of 100. The red curves in figure 9 show the final result of 4 IMC updates per interaction with a regularization value of 100. As one can see with additional regularization a match in the RDF could be achieved and a physical correct looking shape of the potential was obtained for all

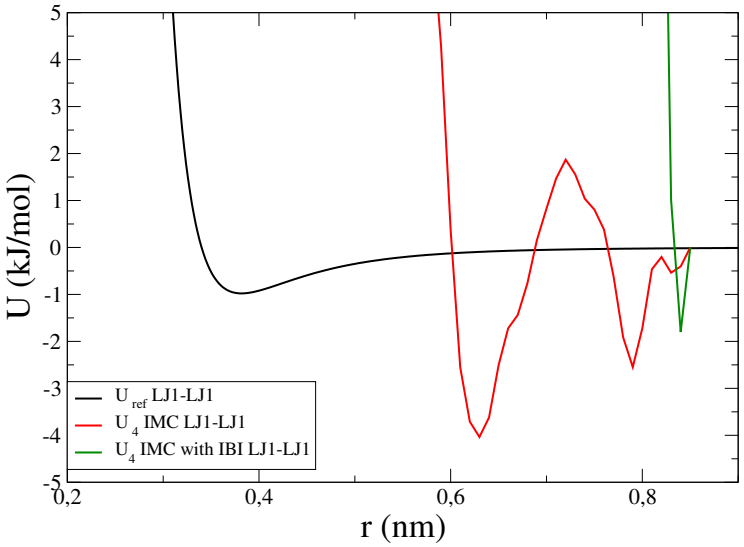

Fig. 8. Reference LJ potential for the LJ1-LJ1 interaction (black line), $U_{LJ1-LJ1}$ after 4 IMC steps without previous IBI (red line), $U_{LJ1-LJ1}$ after 4 steps with previous IBI (green line).

three interactions LJ1-LJ1 (upper panel), LJ2-LJ2 (middle panel) and LJ1-LJ2 (lower panel). Nevertheless the generated potentials are a bit too attractive compared to the reference potentials (black lines in all the panels of figure 9). That is why larger regularization parameter of 300 and 1000 are tested. These larger values lead to a better approximation towards the reference potential (green lines $\lambda = 300$ and blue lines $\lambda = 1000$ in figure 9). Especially with $\lambda = 1000$ the potentials for the LJ1-LJ1 and LJ1-LJ2 interaction could almost be fully recovered. Both potentials show only minor divergence in the long-range part of the potential. Beyond that the generated LJ1-LJ2 potential has a slightly deeper potential well. The generated LJ2-LJ2 potential stays a bit more attractive for both larger λ values.

For further calculations we take the potentials generated with a regularization parameter of 1000. Although this value might lead to a too strong dominance of the regularization term over the data fit term in Eq. 14, it leads to the best reproduction of the underlying potentials for all three interactions.

In this section we could show that the run of previous IBI steps and addition of a regularization term stabilizes the IMC method. Without these auxiliary steps the algorithm did not converge to a stable solution. But with them, comparable results to the P-IBI method are obtained as shown in figure 10. In this graph the results of P-IBI after 80 steps and the regularized IMC run ($\lambda = 1000$) after 4 steps (with preliminary 30 P-IBI steps) for the LJ1-LJ1 interaction are compared. One sees clearly that the IMC method results in a better reproduction of the reference Lennard-Jones potential than the P-IBI method does. This also holds true for the LJ1-LJ2 interaction. The final LJ2-LJ2 interaction is for both methods nearly the same, but in both cases slightly too attractive. Overall it seems that the interdependence of the potential update in IMC leads to an improvement in the reproduction of all potentials present in a binary system. Compared to P-IBI also less iterations are needed to obtain reasonable well matched RDFs and potentials. But as there is no free lunch, the fewer amount of iterations comes with the cost of a longer run time per iteration.

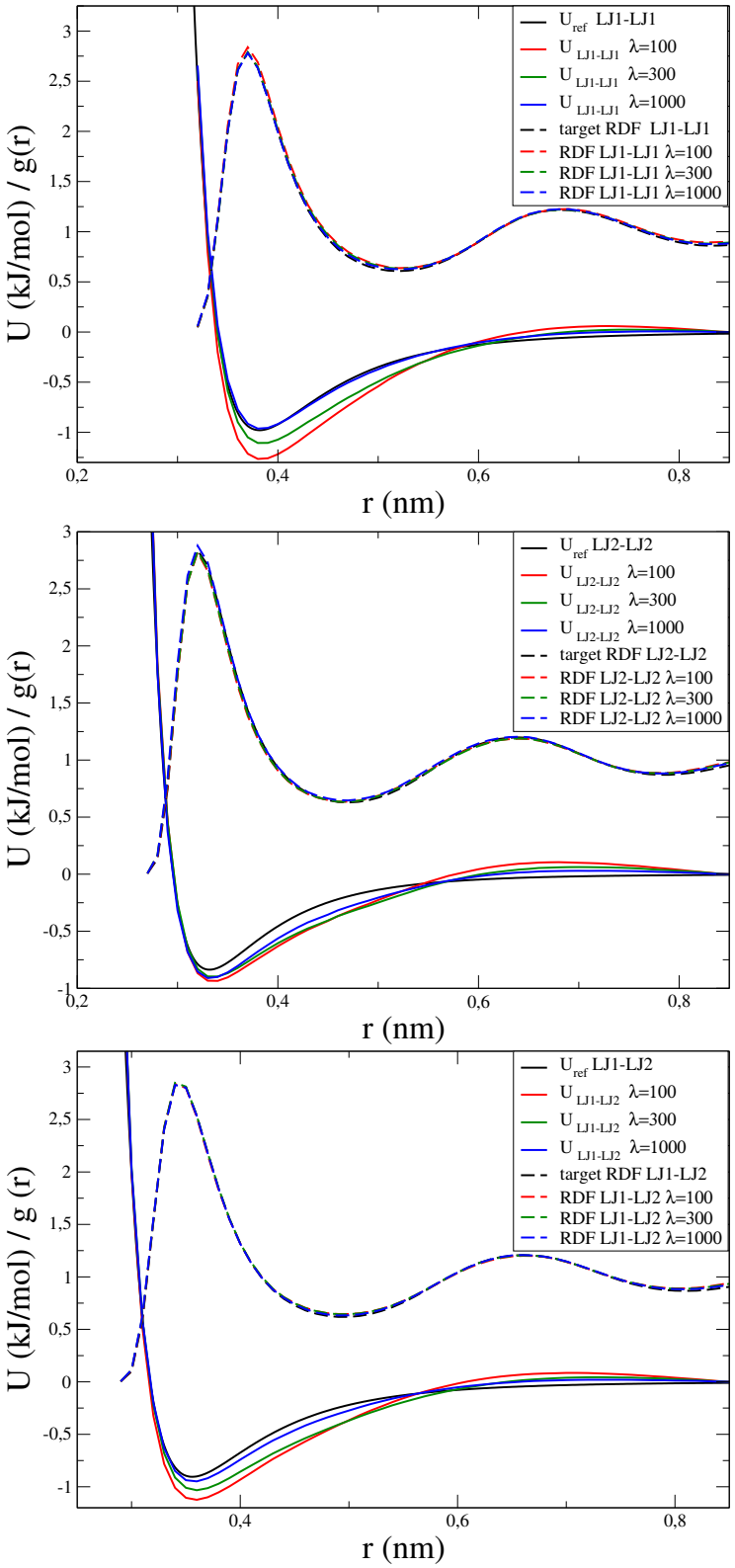

Fig. 9. Influence of different regularization values on the IMC results shown for the LJ1-LJ1 (upper panel), LJ2-LJ2 (middle panel) and LJ1-LJ2 (lower panel) interaction: Potentials (solid lines) and RDFs (dashed lines) for the FG system (black), for $\lambda = 100$ (red), $\lambda = 300$ (green) and $\lambda = 1000$ (blue).

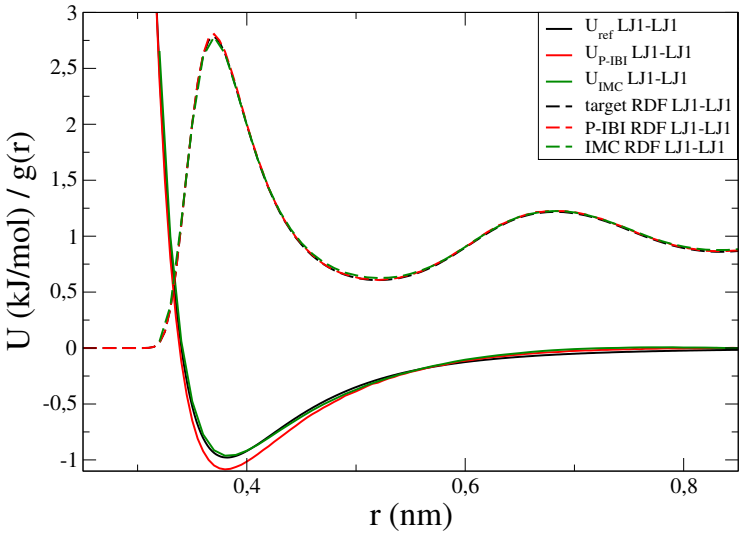

Fig. 10. Final potentials (solid lines) and RDFs (dashed lines) of the LJ1-LJ1 interaction obtained from P-IBI (red) and IMC with a regularization parameter of 1000 (green) in comparison with the FG reference (black).

4.3 Thermodynamic quantities

The generated effective pair potentials are now further evaluated as described in Section 3.3. Before we start with the KB analysis, we first investigate how well the RDFs are reproduced for the different potentials. In figure 11 we see the RDFs for the LJ1-LJ1 interaction in good agreement with the FG reference (target RDF) even beyond the cut-off for mostly all of the applied methods. Only the KB-IBI derived potentials lead to a small shift in the RDF towards smaller distances. The LJ2-LJ2 RDFs (not shown here) obtained with IMC and KB-IBI potentials show a slightly larger first peak, whereas slightly smaller first and second maxima are obtained with the C-IBI derived potential. For the LJ1-LJ2 interaction the RDF (not shown here) obtained with the KB-IBI potential shows a first and second maximum which are too large, whereas the other methods are in good agreement with the reference. So in general one can say that an update based on the KB-IBI method comes with the cost of less structural accuracy in the production run. This structural inaccuracy might be related to the fact, that KB-IBI targets to match a quantity which is mainly determined by the long-range part of the potential, whereas the RDF is mainly determined by the short-range part of the potential. Thus to get accurate structural representation, it may be preferable not to match the reference potential exactly. It may be more important that the update of the generating method focuses on the short-range part of the potential, since this is the part which mainly is determining the RDF.

In order to calculate thermodynamic quantities using KBIs obtained from molecular simulations, finite size artefacts should be considered. To this end, Ganguly et al. applied a tail correction to the RDF and studied the dependence of the RKBI on system size in NpT simulations.[48] They could show that for sufficiently large systems, the RKBI approaches a limiting plateau value, which corresponds to the correct thermodynamic limit (Eq. 7). In particular, this could be demonstrated by comparing the KBIs obtained by means of this approach with an alternative method

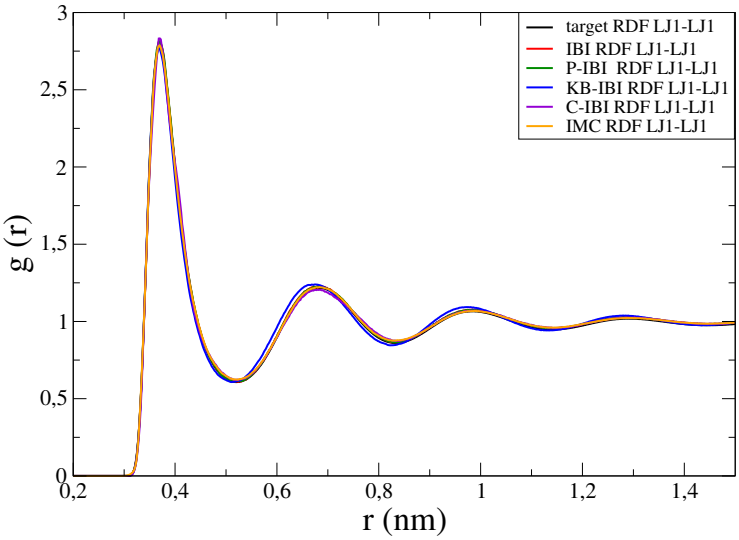

Fig. 11. RDF for the LJ1-LJ1 interaction obtained from different generating potentials: reference LJ potential (black), IBI (red), P-IBI (green), KB-IBI (blue), C-IBI (violet) and IMC (orange).

introduced by Schnell et al. [62,63] These authors showed that KBI values in the thermodynamic limit can be obtained from an analysis of particle number fluctuations in small sub-volumes (sub-boxes) $V_s \subset V$ in simulations of closed NVT (or NpT) systems. By following the formalism of Hill,[64] they could show that SKBIs ($G_{ij}^{V_s}$, defined below) scale linearly with the inverse of system size $L_s = V_s^{1/3}$ of these small subboxes:[62]

$$G_{ij}^{V_s} = G_{ij}^{\infty} + \frac{a}{L_s} \tag{20}$$

The small boxes can be considered as grand-canonical ensembles embedded in a larger box, which acts as particle bath with which an exchange of particles and energy takes place. Assuming our simulation box (NVT or NpT) with size L is large enough compared to the small box with size L_s, this method can be applied for our system as well. The SKBIs (see Eq. 20) are calculated from particle number fluctuations in these small sub-boxes according to: [63]

$$G_{ij}^{V_s} = V_s \left[\frac{\langle N_i N_j \rangle - \langle N_i \rangle \langle N_j \rangle}{\langle N_i \rangle \langle N_j \rangle} - \frac{\delta_{ij}}{\langle N_i \rangle} \right] \tag{21}$$

where N_i is the particle number of type i in the sub-box with the volume V_s and δ_{ij} is the Kronecker delta. The averages are taken over sub-boxes randomly inserted along the trajectory of the NVT system. Following this procedure, SKBIs are calculated for various sub-box sizes (L_s). From these calculations the KBIs in the thermodynamic limit can be obtained via Eq. 20. To make use of this equation, first the regime where the SKBIs (Eq. 21) scale linear with the inverse box size has to be identified and second this regime has to be extrapolated for $L_s \to \infty$ or $1/L_s \to 0$. In table 3 the limiting KBI values (G_{ij}^{∞}) for all interactions with different generating potentials are presented. In figure 12 the plots from which the limiting KBI values are obtained are shown. Since we choose a spherical shape for the sub-boxes, the radii of these spheres is used as size L_s.

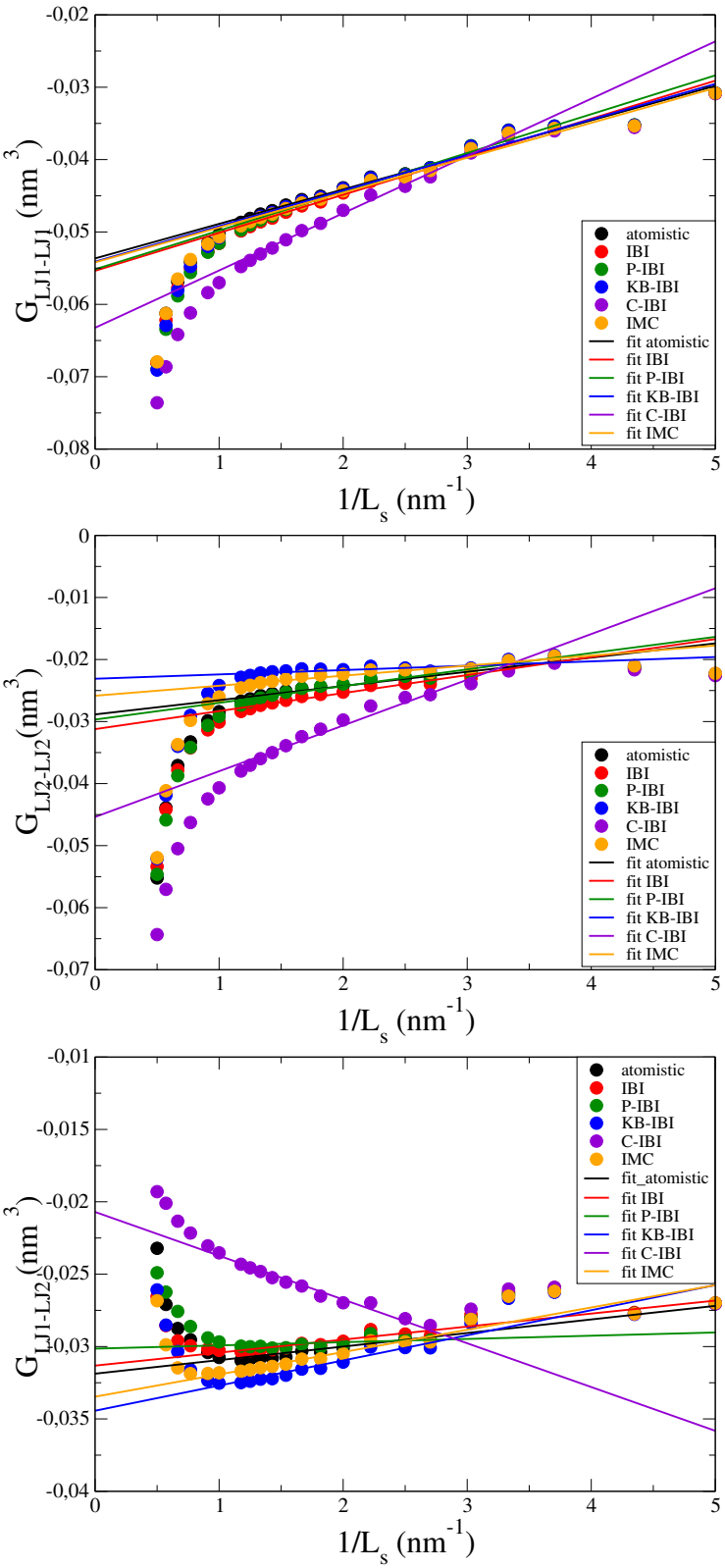

Fig. 12. LJ1-LJ1 (upper panel), LJ2-lJ2 (middle panel) and LJ1-LJ2 (lower panel) SKBIs calculated from number fluctuations of spherical sub-boxes as a function of the inverse sub-box radii for the differently generated potentials with different system size (solid dots). Solid straight lines are fitted to the linear regimes of the plots.

Table 3. Limiting KBI values for the interactions LJ1-LJ1, LJ2-LJ2,LJ1-LJ2 for different generating potentials

Method	G_{11}^{∞} (nm^3)	G_{22}^{∞} (nm^3)	G_{12}^{∞} (nm^3)
atomistic	-0.054	-0.029	-0.032
IBI	-0.055	-0.031	-0.031
P-IBI	-0.055	-0.030	-0.030
KB-IBI	-0.054	-0.023	-0.034
C-IBI	-0.063	-0.045	-0.021
IMC	-0.054	-0.025	-0.033

In table 3 the limiting KBIs for the different generating potentials are listed. The KBIs obtained from potentials derived via IBI and P-IBI slightly underestimate the atomistic reference for the LJ1-LJ1 and LJ2-LJ2 interaction. They overestimate the LJ1-LJ2 KBI. The potentials derived via KB-IBI and IMC show the best reproduction of the KBI for the LJ1-LJ1 interaction. They underestimate the KBI of the LJ2-LJ2 interaction and overestimate the KBI for the LJ1-LJ2 interaction. It is interesting to see that the C-IBI method shows the poorest performance in reproducing the KBIs for all interactions, although this method shows the best fit in case of the RKBIs. A reason for this mismatch might be that finite size effects have a too strong influence on the potentials derived via C-IBI. This also applies for the KB-IBI method. But in case of KB-IBI the underlying potential is closer to the reference one than it is the case for C-IBI. This might lead to the better representability of the KBI by potentials derived via KB-IBI compared to potentials generated via C-IBI.

Having obtained the limiting KBI values, we calculate the isothermal compressibility, κ_T, of the system and the partial molar volume for each of the particles, $\overline{v_1}$ and $\overline{v_2}$. Following the KB theory the compressibility can be calculated as [36]

$$\kappa_\tau = \frac{\zeta}{\eta\ k_B T} \tag{22}$$

and the partial molar volume for particle type LJ1 as [36]

$$\overline{v_1} = \frac{1 + \rho_2(G_{22}^{\infty} - G_{12}^{\infty})}{\eta} \tag{23}$$

and for type LJ2 respectively as [36]

$$\overline{v_2} = \frac{1 + \rho_1(G_{11}^{\infty} - G_{12}^{\infty})}{\eta} \tag{24}$$

where

$$\zeta = 1 + \rho_1 G_{11}^{\infty} + \rho_2 G_{22}^{\infty} + \rho_1\rho_2(G_{11}^{\infty}G_{22}^{\infty} - G_{12}^{\infty 2})\ , \tag{25}$$

and

$$\eta = \rho_1 + \rho_2 + \rho_1\rho_2(G_{11}^{\infty} + G_{22}^{\infty} - 2G_{12}^{\infty})\ , \tag{26}$$

ρ_1 is the number density for particle type LJ1, G_{11}^{∞} is the limiting KBI for the LJ1-LJ1 interaction, ρ_2 is the number density for particles type LJ2, G_{22}^{∞} is the limiting KBI for the LJ2-LJ2 interaction and G_{12}^{∞} is the limiting KBI for the LJ1-LJ2 interaction. [36]

To complete the KB analysis, the derivative of the activity coefficient, γ_1, with respect to the mole fraction of the mixture, x_1, for particle type LJ1 is calculated.[65]

$$\left(\frac{\partial \ln \gamma_1}{\partial \ln x_1}\right)_{p,T} = - \frac{\rho_2 x_1 (G_{11}^{\infty} + G_{22}^{\infty} - 2G_{12}^{\infty})}{1 + \rho_2 x_1 (G_{11}^{\infty} + G_{22}^{\infty} - 2G_{12}^{\infty})} \,. \tag{27}$$

The results for the KB analysis together with the virial pressure and the average energy density of the system are presented in table 4.

Table 4. Virial pressure p in [bar], the isothermal compressibility κ_τ in $[10^{-9}\mathrm{m}^2/\mathrm{N}]$, the partial molar volumes $\overline{v_1}$ and $\overline{v_2}$ in $[\mathrm{nm}^3]$, the derivative of the activity coefficient $\left(\frac{\partial ln \gamma_i}{\partial ln x_i}\right)_{p,T}$ and the total energy density in $[10^{-22}\mathrm{kJ/nm}^3]$ for the different generating potentials

Method	p	κ_τ	$\overline{v_1}$	$\overline{v_2}$	$\left(\frac{\partial ln \gamma_1}{\partial ln x_1}\right)_{p,T}$	Energy density
atomistic	1.0	1.07	0.0298	0.0208	0.133	-1.762
IBI	942.4	0.83	0.0296	0.0207	0.176	-0.836
P-IBI	16.7	1.31	0.0298	0.0204	0.182	-1.739
KB-IBI	609.8	1.02	0.0315	0.0208	0.055	-1.135
C-IBI	2317.9	0.71	0.0256	0.0172	0.726	0.321
IMC	413.1	0.99	0.0308	0.0208	0.089	-1.442

As one sees from table 4 the reference pressure of 1 bar is not matched with any of the derived potentials. The P-IBI potentials provide the best approximation of the reference value, but in this case the compressibility is slightly overestimated compared to the atomistic reference. C-IBI potentials generate the highest virial pressure compared to all other potentials which is due to a mismatch of the C-IBI potentials in the attractive and repulsive regions (see figure 5). In case of the IBI method, the potential well is too high which also leads to more repulsion between the particles and as a consequence to a larger virial pressure. The potentials generated via IMC also overestimate the virial pressure despite that all three IMC potentials are closest to the atomistic reference potentials. This may be explained because IMC potentials are slightly more repulsive in the tail compared to reference potential (see figure 9). This shows that even small differences in the potential have a big impact on the pressure of the system. Moreover it shows how well the long range part of the potential has to be matched to achieve the correct pressure. The KB-IBI potentials also overestimate the virial pressure of the system, but reproduce the compressibility of the system quite well compared to the atomistic reference. In general, the compressibility is underestimated if the potential is too repulsive compared to the atomistic reference. The partial molar volumes of the two different components agree well with the reference values for nearly all potentials, except the C-IBI generated potentials. This results from the poor agreement of the limiting KBIs with the atomistic reference as we can see from figure 12. The derivatives of the activity coefficient overestimate the reference values for mostly all IBI potentials. The IMC and KB-IBI potentials however underestimate this quantity.

All the results of the KB analysis can be related to how well each method could reproduce the limiting KBIs compared to the target system. A weakness of the small-system method applied to calculate the limiting KBIs is the reliability of the linear fit, especially if the differences have a small magnitude. But from the thermodynamic

analysis we see that properties that are insensitive to the long range part of the potential can be reproduced by all applied methods in good agreement with the atomistic reference. To achieve a better match in a quantity, e.g. the derivative of the activity coefficient, that is mainly determined by the tails of the potential, a better reconstruction of this part for all interactions in the system has to be guaranteed.

Finally the average energy densities of the systems have been compared. This quantity reflects how well the different methods could reproduce the potentials for all interactions present in the system. P-IBI potentials reproduce the energy density most accurately. Here all the generated potentials have a slightly deeper minimum compared to the atomistic reference potentials. KB-IBI and IMC potentials provide less negative energy densities. A reason for this can be the more repulsive tails of the potentials derived via IMC. KB-IBI potentials are also always slightly more repulsive compared to the atomistic reference, but less repulsive than potentials generated via IBI. That is why in this case the energy density is even less negative. The energy density predicted by the C-IBI potentials is positive, which can again be related to the overly repulsive potentials generated by this method.

5 Conclusion

In this paper we have addressed the question whether it is possible to recover the generating potentials of target RDFs via different variations of the IBI method and the IMC method. The point of interest was to see if these methods can converge to a known solution of the inverse problem of finding effective pair potentials for non-bonding interactions derived from a given set of RDFs. The system under study was a binary mixture of two LJ particles.

For none of the methods it was possible to recover the potentials for all interactions, although good structural match could be achieved for all methods that have been considered. We could show that for IBI-type methods implementation of thermodynamic constraints improves the convergence of the method with respect to the potential and the RDF, if used in combination with a well chosen cut-off. The cut-off should be chosen such that it is large enough to include the relevant parts of the potential, but not too large, since this affects the convergence negatively. As additional thermodynamic constraints the pressure (P-IBI), KBIs (KB-IBI) and the integrated RDF (C-IBI) have been investigated. An update based on C-IBI does not improve the convergence towards the reference potential most probably because of finite size effects. The KB-IBI method leads to results that are in between those obtained with P-IBI and C-IBI. The linear corrections used in P-IBI and KB-IBI lead to reproducing the reference potential better than the alternative update based on the integrated RDF used in C-IBI. The C-IBI-based update however provides the best representation of the RKBIs. In terms of convergence of the method towards the reference potential, inclusion of the pressure as a thermodynamic constraint in the standard IBI procedure is preferred over the alternative quantities (KBIs or integrated RDFs), although this may interfere with the Henderson theorem.[17]

For IMC we could show how sensitive this method reacts to several factors. First, the initial guess should be close to the actual potential. Hence, the PMF is not a good choice. Second, one should provide enough statistics to achieve convergence of the algorithm. In our case we could not achieve convergence within $5 \cdot 10^6$ MD steps (compare IBI: $1 \cdot 10^6$ MD steps). But we could show that it is possible to decrease the statistical demand by adding a regularization term, which stabilizes the algorithm. A

SVD analysis of the matrix helps to get a good educated guess for the magnitude of the regularization parameter. The regularization allows the generation of potentials which have the best agreement with the reference potentials for all three interactions present in the binary system.

We further addressed the thermodynamic representability of the derived potentials in comparison with the atomistic reference. To this end, the isothermal compressibility, the partial molar volumes of the binary mixture components, and the derivative of the activity coefficient with respect to the mole fraction (for component 1) have been calculated from the KBIs obtained with the derived potentials. The IBI potentials do not reproduce the isothermal compressibility of the atomistic reference model. The KB-IBI potentials best reproduce this quantity. The partial molar volumes of the two mixture components are represented well with the majority of the potentials. The pressure of the system is overestimated with all potentials. While the P-IBI potential shows the best agreement with the reference pressure, it overestimates the isothermal compressibility. All IBI potentials overestimate the derivative of the activity coefficient computed for one of the components. Hence, these potentials do not accurately represent the dependence of the chemical potential on composition which is mainly determined by the tails of the LJ potentials. The KB-IBI potentials underestimate the derivative of the activity coefficient. The reason that KB-IBI and C-IBI do not well represent the quantities they are designed for is related to finite-size system effects, which are stronger in Lennard-Jones mixtures as compared to aqueous solution mixtures previously studied.[47, 49]

Comparison with the results obtained with IMC potentials further showed that even a better match of the potential around the attractive minimum does not improve the thermodynamic representability of the model. The non-sensitivity of the RDFs on long-range tails of the potentials clearly poses a challenge in representing thermodynamic properties with models derived by inverse coarse-graining methods. The derived potentials that best reproduce the attractive tail of the reference potential also represent the average energy density of the system better, but already small deviations in the long range part do not allow an accurate representation of this quantity.

The authors would like to thank Christine Peter, Christoph Junghans, Denis Andrienko, Christoph Scherer and Pritam Ganguly for fruitful discussions and useful remarks on technical aspects of the applied methods. Funding has been granted by the German Research Foundation (DFG) within the Collaborative Research Center "Multiscale Simulation Methods for Soft-Matter Systems"(SFB-TRR 146). Computational time provided on the Lichtenberg High Performance Computer by the TU Darmstadt is acknowledged.

References

1. W. Tschöpp, K. Kremer, J. Batoulis, T. Bürger, O. Hahn, Acta Polym. **49**, 61 (1998)
2. M. Murat, K. Kremer, J. Chem. Phys. **108**, 4340 (1998)
3. F. Müller-Plathe, ChemPhysChem **3** (9), 754 (2002)
4. M. Praprotnik, L. Delle Site, K. Kremer, Annu. Rev. Phys. Chem. **59**, 545 (2008)
5. C. Peter, K. Kremer, Soft Matter **5**, 4357 (2009)
6. E. Brini, E. A. Algaer, P. Ganguly, C. Li, F. Rodriguez-Ropero, N. F. A. van der Vegt, Soft Matter **9**, 2108 (2013)
7. T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, M. Karttunen, Phys. Chem. Chem. Phys. **11**, 1869 (2009)

8. B. Hess, C. Holm, N. F. A. van der Vegt, J. Chem. Phys. **124**, 164509 (2006)
9. Y. Wang, W. G. Noid, P. Liu, G. A. Voth, Phys. Chem. Chem. Phys. **11**, 2002 (2009)
10. E. Brini, V. Marcon, N. F. A. van der Vegt, Phys. Chem. Chem. Phys. **13**, 10468 (2011)
11. D. Reith, M. Puetz, F. Müller-Plathe, J.Comput. Chem. **24**, 1624 (2003)
12. A. P. Lyubartsev, A. Laaksonen, Phy. Rev. E **52** (4), 3730 (1995)
13. F. Ercolessi, J. B. Adams, Europhys. Lett. **26**, 583 (1994)
14. S. Izvekov, G. A. Voth, J. Phys. Chem. B **109**, 2469 (2005)
15. J. W. Mullinax, W. G. Noid, J. Phys. Chem. C **114**, 5661 (2010)
16. M. S. Shell, J. Chem. Phys. **129**, 144108 (2008)
17. R.L. Henderson, Physics Letters **49A**, 197 (1974)
18. W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, H. C. Andersen, J. Chem. Phys. **128**, 244114 (2008)
19. W. G. Noid, P. Liu, Y. Wang, J.-W. Chu, G. S. Ayton, S. Izvekov, H. C. Andersen, G. A. Voth, J. Chem. Phys. **128**, 244115 (2008)
20. M. E. Johnson, T. Head-Gordon, A. A. Louis, J. Chem. Phys. **126**, 144509 (2007)
21. E. Brini, N. F. A. van der Vegt, J. Chem. Phys. **137**, 154113 (2012)
22. P. Ganguly, N. F. A. van der Vegt, J. Chem. Theory. Comput. **9**, 5247 (2013)
23. L. C. Jacobson, R. M. Kirby, V. Molineo, J. Phys. Chem. B **118**, 8190 (2014)
24. H. Wang, C. Junghans, K. Kremer, Eur. Phys. J. E. **28**, 221 (2009)
25. A. Das, H. C. Andersen, J. Chem. Phys. **132**, 164106 (2010)
26. N. J. Dunn, W. G. Noid, J. Chem. Phys. **143**, 243148 (2015)
27. B. Hess, C. Holm, N. F. A. van der Vegt, J. Chem. Phys. **124**, 164509 (2006)
28. J. W. Shen, C. Li, N. F. A. van der Vegt, C. Peter, J. Chem. Theory. Comput. **7**, 1916 (2011)
29. H. J. Qian, P. Carbone, C. Xiaoyu, H. A. Karimi-Varzaneh, C. C. Liew, F. Müller-Plathe, Macromolecules **41**, 9919 (2008)
30. H. Eslami, H. A. Karimi-Varzaneh, F. Müller-Plathe, Macromolecules **44**, 3117 (2011)
31. M. Langeloth, T. Sugii, M. C. Boehm, F. Müller-Plathe, J. Chem. Phys. **143**, 243158 (2015)
32. C. Peter, L. Delle Site, K. Kremer, Soft Matter **4**, 859 (2004)
33. S. Jain, S. Garde, S. K. Kumar, Ind. Eng. Chem. Res. **45**, 5614 (2006)
34. C. Fu, P. M. Kulkarni, M. S. Shell, L. G. Leal, J. Chem. Phys. **137**, 164106 (2012)
35. D. Ivanizki, *Numerical Analysis of the relation between interactions and structures in a molecular fluid* (Ph.D Thesis University of Mainz, 2015).
36. A. Ben-Naim, *Molecular Theory of Solutions* (Oxford University Press New York, 2006).
37. J. G. Kirkwood, F. P. Buff, J. Chem. Phys. **19**, 774 (1951)
38. P. Krüger, S. K. Schnell, D. Bedeaux, S. Kjelstrup, T. J. H. Vlugt, J.-M. Simon, J. Phys. Chem. Lett. **4**, 235 (2013)
39. S. Weerashinge, P. E. Smith, J. Chem. Phys. **118**, 10663 (2003)
40. S. Weerashinge, P. E. Smith, J. Phys. Chem. B **107**, 3891 (2003)
41. S. Weerashinge, P. E. Smith, J. Chem. Phys. **121**, 2180 (2004)
42. S. Weerashinge, P. E. Smith, J. Phys. Chem. B **109**, 15080 (2005)
43. M. E. Lee, N. F. A. van der Vegt, J. Chem. Phys. **122**, 114509 (2005)
44. M. B. Gee, N. R. Cox, Y. F. Jiao, N. Bentenitis S. Weerashinge, P. E. Smith, J. Chem. Theory Comput. **8**, 1802 (2012)
45. M. Kang, P. E. Smith, J. Comput. Chem. **27**, 1477 (2006)
46. M. Fyta, R. R. Netz, J. Chem. Phys. **136**, 124103 (2012)
47. P. Ganguly, D. Mukherji, C. Junghans, N. F. A van der Vegt, J.Chem. Theor. and Comp. **8**, 1802 (2012)
48. P. Ganguly, N. F. A. van der Vegt, J. Chem. Theory. Comput. **9**, 1347 (2013)
49. T. E. Oliveira, P. A. Netz, K. Kremer, C. Junghans, D. Mukherji, J. Chem. Phys. **144**, 174106 (2016)
50. A. Lyubartsev, A. Mirzoev, L-J. Chen, A. Laaksonen Faraday Discuss. **144** 43 (2010)
51. M. Hanke, arXiv:1603.03899 [math-ph] (2016)
52. M. Hanke, arXiv:1603.03900 [math-ph] (2016)
53. T. Murtola, E. Falck, M. Karttunen, I. Vattulainen , J.Chem. Phys. **126**, 075101 (2007)

54. H. W. Engl, M. Hanke, A. Neubauer *Regularization of Inverse Problems* (Kluver Academic Publishers, 2000).
55. V. Ruehle, C. Junghans, A. Lukyanov, K. Kremer, D. Andrienko, J. Chem. Theo. Comp. **5** (12), 3211 (2009)
56. B. Hess, C. Kutzner, D. van der Spoel, E. Lindahl, J. Chem. Theo. and Comp. **4**(2), 435 (2008)
57. V. Ruehle, C. Junghans, Macromol. Theory Simul. **20**, 472 (2011)
58. S. Y. Mashayak, M. Jochum, K. Koschke, N. R. Aluru, V. Ruehle, C. Junghans, PLoS one **10**, e131754 (2015)
59. M. Parrinello, J. Applied Physics **52**(12), 7182 (1981)
60. S. Nosė. M. L. Klein, Molecular Physics, **50**(5), 1055 (1983)
61. https://github.com/votca/downloads/blobmaster/votca-csg-manual-1.3_rc1.pdf
62. S. K. Schnell, T. J. H. Vlugt, J.-M. Simon, D. Bedeaux, S. Kjelstrup, Chem. Phys. Lett. **504**, 199 (2011)
63. S. K. Schnell, X. Liu, J.-M. Simon, A. Bardow, D. Bedeaux, T. J. H. Vlugt, S. Kjelstrup, J. Phys. Chem. B **115**, 10911 (2011)
64. T. L. Hill, *Thermodynamics of Small Systems, Part 1* (W. A. Benjamin: New York, 1963).
65. A. Villa, C. Peter, N. F. A. van der Vegt, J. Chem. Theory. Comput. **6**, 2434 (2010)