Kirkwood–Buff Coarse-Grained Force Fields for Aqueous Solutions

Pritam Ganguly,† Debashish Mukherji,‡ Christoph Junghans,§,∥ and Nico F. A. van der Vegt*† † Center of Smart Interfaces, Technische Universität Darmstadt, Petersenstrasse 32, 64287 Darmstadt, Germany
‡ Max-Planck Institut für Polymerforschung, Ackermannweg 10, 55128 Mainz, Germany
§ Theory Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States
∥ Center of Smart Interfaces, Technische Universität Darmstadt, Petersenstrasse 32, 64287 Darmstadt, Germany

Supporting Information

ABSTRACT: We present an approach to systematically coarse-grain liquid mixtures using the fluctuation solution theory of Kirkwood and Buff in conjunction with the iterative Boltzmann inversion method. The approach preserves both the liquid structure at pair level and the dependence of solvation free energies on solvent composition within a unified coarse-graining framework. To test the robustness of our approach, we simulated urea—water and benzene—water systems at different concentrations. For urea—water, three different coarse-grained potentials were developed at different urea concentrations, in order to extend the simulations of urea—water mixtures up to 8 molar urea concentration. In spite of their inherent state point dependence, we find that the single-site models for urea and water are transferable in concentration windows of approximately 2 M. We discuss the development and application of these solvent models in coarse-grained biomolecular simulations.

1. INTRODUCTION

Biomolecules in water can be salted-in, salted-out, or chemically denatured by the presence of cosolvents, such as alcohols, inorganic salts, guanidinium chloride, and urea, to name a few.¹,² Herein, we are interested in developing systematically coarse-grained (CG) single-site models for water and chemical denaturants such as urea, which may find application in coarse-grained biomolecular simulations.³ Urea is a well-known salting-in agent: preferential interaction of urea (over water) with nonpolar molecules⁴,⁵ as well as with nonpolar and polar groups on peptides⁶,⁷ including the peptide backbone of proteins,⁸ favors the solvation of these groups and leads to a decrease of their solvation free energies. A systematic molecular coarse-graining method, which provides solvent models that reproduce the solvation free energies while keeping the required structural information, is presently not available and will be proposed in this work.

Effective nonbonded pair potentials for CG models have successfully been developed for polymers⁹—¹² and nonpolar molecular liquids¹³—¹⁶ by reversible work techniques in which averages are taken over degrees of freedom no longer represented by the CG model, such as angular orientations. Owing to multibody correlations, this type of approach will however fail for hydrogen bonded liquids, and alternative approaches are needed. Herein, we propose an approach that is based on the thermodynamic theory of Kirkwood and Buff introduced in the early 1950s.¹⁷ Instead of relating the thermodynamic properties to the intermolecular potentials, this theory relates the thermodynamic properties to integrals of radial distribution functions (RDF) over the volume. For solution components i and j, these so-called Kirkwood–Buff integrals (KBI) are defined as¹⁷

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

where \( g_{ij}(r) \) is the RDF and \( G_{ij} \) is the KBI. Away from the critical point where density fluctuations become long-ranged, contributions to this integral are local and are determined by fluctuations on length scales \( R < 1 \) nm. Physically, \( \rho_{ij} G_{ij} \) can be interpreted as the change in the number of \( j \) molecules in a spherical region of radius \( R \) in the solution before and after placing a molecule \( i \) at the origin of that region (\( \rho_j \) is the number density of component \( j \)).¹⁸ We thus see that \( G_{ij} \) is a local quantity which can be used as a measure of the affinity between solution components \( i \) and \( j \). In the binary system of cosolvent (c) and water (w), the link to the solvation thermodynamics is given by¹⁸,¹⁹

\[ \left( \frac{\partial \ln \gamma_c}{\partial \ln \rho_c} \right)_{p,T} = -\frac{\rho_c (G_{cc} - G_{cw})}{1 + \rho_c (G_{cc} - G_{cw})} \]  

where \( kT \ln \gamma_c \) is the cosolvent solvation free energy (at pressure \( p \), temperature \( T \), and cosolvent number density \( \rho_c \)) and \( \gamma_c \) is the cosolvent molar scale activity coefficient. Similar expressions have been derived for systems that have a solute (s) at infinite dilution (\( \rho_s \to 0 \)) in a cosolvent—water solution. In this case, the solvation free energy of the solute (\( \Delta G_s \)) varies with the solution composition according to¹⁸

\[ \left( \frac{\partial \Delta G_s}{\partial x_c} \right)_{p,T} = \lim_{\rho_s \to 0} \frac{RT(\rho_s + \rho_c)}{\eta} (G_{sw} - G_{cw}) \]  

where \( R \) is the gas constant, \( \eta = \rho_w + \rho_c + \rho_s \rho_c (G_{sw} + G_{cw} - 2G_{cw}) \), and \( \rho \) is the number density of individual components of the aqueous solutions. Preferential solvation of the solute by cosolvent molecules (\( G_{sw} - G_{cw} < 0 \)) results in a decrease of \( \Delta G_s \) upon increasing the cosolvent mole fraction \( x_c \) (salting-in).

Received: February 2, 2012
Published: March 21, 2012
In this paper, we pursue the idea that a CG model provides a good representation of the realistic system if it reproduces the solvent composition dependence of the solvation free energies as expressed by eq 2 and eq 3 in a range of nearby concentrations. A conformational transition of a biomolecule driven by changes of the solvation shell composition provides just one illustrative example where this is important. The coarse-grained view provided by Kirkwood–Buff theory tells us that this requirement can be met with models that reproduce the \( G_{ij} \)'s. Although the RDFs need not necessarily be reproduced to realistically model salting-in and salting-out processes, CG models that represent both the RDFs and KBIs of the real system significantly extend the scope and applicability of CG biomolecular simulations.

A CG water–cosolvent model that represents the pairwise liquid structure without sacrificing the required thermodynamic accuracy has previously been reported for benzene in water.\(^20\) The approach reported there however relies on pairwise additivity of hydrophobic interactions between small molecules at low concentration and cannot readily be generalized to hydrophilic compounds. Alternatively, the MARTINI model\(^11,22\) is instead parametrized to reproduce experimental transfer free energies. Although the MARTINI model is very useful in studies of, among others, self-assembly processes, it is not sufficiently accurate to reproduce the dependence of solvation free energies on solvent composition and the corresponding changes in liquid structure. As illustrated by Figure S1 in the Supporting Information, the MARTINI model predicts a typical Lennard-Jones-type fluid structure for an aqueous solution with a polar cosolvent. This structure is however not representative of aqueous systems in which the RDFs show significantly less pronounced long-range oscillations.

2. SIMULATION DETAILS

Atomistic simulations were performed with the GROMACS molecular dynamics package.\(^25\) The force field parameters for urea were taken from the Kirkwood–Buff–derived force field,\(^19\) for benzene, the Gromos 43A1 parameters were used.\(^24\) Water was modeled with the SPC/E potential.\(^3\) The all-atom simulations were performed in the NpT ensemble. The pressure was controlled with a Parrinello–Rahman barostat\(^26\) at 1 atm pressure with a coupling time of 3 ps. The temperature was set to 300 K in all simulations using a Nose–Hoover thermostat\(^27,28\) with a relaxation time of 0.5 ps. The integration time step was set to 2 fs, and 100 ns trajectories were accumulated. Electrostatic interactions were calculated with the particle mesh Ewald (PME) method.\(^29\) The nonbonded interaction cutoff was chosen as 1 nm. The simulations of aqueous urea were performed with approximately 11 000 water molecules; for aqueous benzene, the number of water molecules varied between 10 000 to 56 000. The number of cosolvent molecules was varied according to the concentration. The urea concentrations were taken between 2.6 and 7.7 M and for benzene between 0.1 and 0.5 M.

The coarse-grained simulations were performed in an NVT ensemble with the GROMACS simulation package at the average NpT volume of the corresponding atomistic simulation. The equations of motion were integrated using the leapfrog stochastic algorithm. The inverse friction constant was set to 0.2 ps for urea/water and 1.0 ps for the benzene/water systems. The integration time step was set to 4 fs, and the cutoff was set to 1.4 nm.

3. RESULTS AND DISCUSSIONS

3.1. Method and Implementation. CG solvent models that represent the RDFs in principle also represent the thermodynamic solvation properties in eqs 2 and 3. CG methods that optimize effective pair potentials in order to reproduce the RDFs have previously been reported in the literature and include the inverse Monte Carlo\(^30,31\) and iterative Boltzmann inversion (IBI)\(^32\) methods. These methods provide, at least in principle, the required balance of structural and thermodynamic properties but, as we will show here, need to be further refined, since in practice small variations in the RDFs lead to large variations in the corresponding KBIs, owing to the volume integration in eq 1.

We have simulated all-atom and coarse-grained systems of urea in water and benzene in water. In this work, we use the IBI method implemented in the VOTCA package.\(^33\) The procedure starts from an initial guess for the coarse-grained pair potential, \( U_{ij}^{(0)}(r) \), which is obtained from a reference distribution, in this case the RDF, \( g_{ij}^{(ref)}(r) \), between the molecular centers of mass sampled in an all-atom simulation:

\[
U_{ij}^{(0)}(r) = -k_B T \ln g_{ij}^{(ref)}(r)
\]

The coarse-grained pair potential is iteratively refined until consistency is achieved between the coarse-grained and the reference distributions:

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

In every iteration, a 10-ns- (for benzene) and a 1-ns- (for urea) long MD simulation is performed. The final, converged CG potential is then used to run a MD simulation that generates a 25-ns-long trajectory. Here, we first start by discussing the results for aqueous urea. Figure 1 shows \( g_{ij}(r) \) and \( G_{ij}(r) \) for urea–urea, urea–water, and water–water pairs obtained with the all-atom and IBI coarse-grained models. The KBIs (eq 1) are obtained from the \( G_{ij}(r) \) functions by taking the limit for large \( r \). Typically, a limiting plateau value is observed in these functions for distances greater than 1 nm, provided that the box dimension is chosen to be large enough.\(^34\) In this work, the KBIs are obtained by averaging \( G_{ij}(r) \) in the interval between 1 and 1.4 nm. While the RDFs are reproduced within the line thickness, the limiting \( G_{ij} \) values are shifted in comparison to the target all-atom \( G_{ij}s \) (see black and red curves in Figure 1). In particular, in mixtures, any small error in the fitted \( g_{ij}(r) \) at short-range can propagate to the tail of \( G_{ij}(r) \), giving rise to the discrepancy observed in Figure 1.

In order to reproduce the exact KBIs, we add a correction term into the coarse-grained potential:

\[
\Delta U_{ij}^{(n)}(r) = A(G_{ij}^{(n)} - G_{ij}^{(ref)}) \left( 1 - \frac{r}{r_{cut}} \right)
\]

where \( G_{ij}^{(ref)} \) is the KBI calculated from the reference all-atom simulation and \( G_{ij}^{(n)} \) is the KBI after the \( n \)th iteration. The idea behind using the specific ramp can be rationalized as follows: If the KBIs of the CG model are larger than the all-atom KBIs (as in the urea–urea and water–water KBIs in Figure 1), this infers an unphysical local excess coordination of molecules. Therefore, some repulsion needs to be added to the potential in order to weaken the aggregation. Similarly, if the local aggregation is underestimated (as in urea–water KBI in Figure 1), then some
attraction is needed in the potential. In principle, we can choose any appropriate functional form. However, we chose to use the simplest function that has been shown to work perfectly well in the case of linear pressure correction. The prefactor \( A \) is system-specific and can be tuned on the basis of convergence. With our system and specific simulation protocol, a good estimate of \( A \) is in the range between 0.01 and 0.10 kJ nm\(^{-3}\) mol\(^{-1}\). The prefactors are summarized in the Supporting Information. The KBIs are significantly improved by using the correction in eq 6 as a ramp (see the blue curve in the Figure 1). The models obtained in this way are referred to as KB-IBI. The modifications to the VOTCA package used for the KB-IBI method will be a part of VOTCA release 1.3, and we will also include an example from this work. In Figure 2, we show a comparative plot of the pairwise coarse-grained potentials obtained from the two separate approaches.

3.2. Applications. Having validated the approach for one concentration of urea, we now want to test if the same approach can also be used for a wider range of cosolvent concentrations. Figure 3 shows \( G_{ij} \) as a function of urea molar concentration \( c_u \). The unmodified IBI model shows significant deviations from the all-atom data. The KB-IBI model does a much better job and reproduces the KBIs and, therefore, the solution thermodynamic properties of the parent atomistic model. We point out that procedures which identically match the liquid structure of the CG and all-atom models necessarily provide identical KBIs, but not vice versa. Therefore, it is imperative to investigate to what extent the KB-IBI procedure preserves the liquid structure. In Figure 4, we show the urea–urea RDF for three concentrations corresponding to the middle panel of Figure 3. These results confirm that the RDF is reproduced very well. The urea–urea RDF converges slower owing to the smaller number of urea molecules compared to water molecules present in the system. The urea–water and water–water RDFs (not shown) show equally good agreement.

---

**Figure 1.** Radial distribution functions \( g_{ij}(r) \) and running integrals \( G_{ij}(r) = 4\pi \int_0^r [g_{ij}(s) - 1] s^2 \, ds \) for aqueous urea mixture at 4.7 M urea. Comparative data are shown for all-atom and the IBI and Kirkwood–Buff IBI (KB-IBI) coarse-graining methods. We present data for all three pairs, urea–urea (uu), urea–water (uw), and water–water (ww).

**Figure 2.** Coarse-grained potentials for a urea–water mixture obtained at 4.7 M urea solution. Solid lines denote IBI potentials; the dashed lines denote KB-IBI potentials. The inset shows an enlarged view of the potential minima.

**Figure 3.** Kirkwood–Buff integrals \( G_{ij} \) for aqueous urea solutions as a function of the molar urea concentration \( c_u \). Results are shown for all three pairs; urea–urea (uu), urea–water (uw), and water–water (ww). The CG pair potentials (IBI and KB-IBI) were developed for the solution systems (at urea concentration \( c_u \)) indicated by the arrows and were subsequently used in the concentration windows bounded by the vertical dashed lines. Solid lines are fits to the all-atom data.
The urea and water CG potentials are state-point-dependent. Their transferability to systems with varying urea concentration is therefore not guaranteed. The CG urea–water systems were however simulated with the same CG potential in a finite window of urea concentrations delineated by the vertical dashed lines in Figure 3. We thus developed the CG potential at only three urea concentrations, namely, 2.6, 4.7, and 6.8 M, indicated by the arrows in the Figure 3. The potentials are shown in the Supporting Information. The data in Figure 3 show that the KB-IBI models are transferable in concentration ranges of approximately 2 M. This is partially due to the invariance of the pair structure over such a small concentration range.

Next, we show the results for aqueous benzene solutions. The pairwise CG potential for the benzene–water solution is shown in the Figure 5. Figure 6 presents the \( G_{ij} \)'s as a function of benzene molar concentration. In this case, the results are significantly better for the KB-IBI model compared to the normal IBI model, which shows too strong benzene–benzene aggregation at all concentrations. The CG potential, developed at \( c_b = 0.2 \) M, shows good transferability up to a concentration slightly above 0.5 M, where the system becomes unstable (phase separates) in the all-atom simulation. In CG simulations with the IBI, KB-IBI, and earlier developed models,\(^{20}\) the benzene–water system however remains stable above 0.5 M. This observation indicates that KB-IBI CG models indeed realistically describe thermodynamically stable solutions but fail to describe systems outside equilibrium and processes including lipid self-assembly for which the MARTINI model\(^ {21} \) provides a better choice. Figure 7 presents benzene–benzene RDFs. The data clearly support that the structure is reasonably well reproduced. A closer inspection of the plot reveals that the first peak at 0.6 nm is better reproduced for the lower molar concentrations. However, at larger concentrations, the agreement is relatively poor. Furthermore, longer ranged correlations (beyond 0.75 nm) are always better obtained by the KB-IBI model. We show only the benzene–benzene RDF, which is mostly affected because of poor statistics. The benzene–water and water–water RDFs are in closer agreement with the reference atomistic model (not shown). Interestingly, KBI-IBI provides a better converged benzene–benzene RDF with significantly fewer iterations (in total 50 iterations were performed) compared to the standard IBI method (100 iterations).
Finally, we consider the solvation free energies of the aqueous mixture components. The dependence of the urea and benzene solvation free energies on their molar concentration in solution is described by the quantity $f_{cc} \equiv \left( \frac{\partial \ln \gamma}{\partial \ln \rho_s} \right)_{p,T}$ and is presented in Figure 8 as a function of the cosolvent (urea or benzene) concentration $c_c$. The KB-IBI model shows significantly better agreement with the all-atom data and with experiments in comparison to the IBI model.

Practical application of the models discussed in this paper requires further development of pair potentials that describe urea and water interactions with chemical groups of dissolved solutes (peptides, proteins, etc.). The KB-IBI method can readily be extended to systems with additional solutes and therefore may provide a useful route to construct nonbonded potentials that can be used in CG simulations of salting-in and salting-out phenomena, protein denaturation, or stabilization by chemical denaturants and/or osmolytes. In these applications, the phenomena of interest are driven by fluctuations in solvent composition, and CG models must therefore reproduce the changes in solvation free energies with quantitative accuracy; i.e., the CG models must reproduce the $G_f$’s in a suitable range of concentrations.

### 3.3. Conclusions

We have proposed a systematic molecular coarse-graining approach, which, inspired by Kirkwood–Buff (KB) solution theory, provides a new route to developing CG models that reproduce the liquid structure at the pair level and the solvation free energies of the mixture components. The approach is based on the iterative Boltzmann inversion (IBI) method and is denoted KB-IBI. We developed three different coarse-grained KB-IBI potentials for urea/water systems at different urea concentrations, allowing the simulation of urea–water mixtures up to 8 molar urea concentration. In spite of their inherent state point dependence, we find that the single-site models for urea and water are transferable in concentration windows of approximately 2 M. We furthermore find that the KB-IBI method provides converged potentials for the liquid mixtures studied here with significantly fewer iterations compared to standard IBI. The KB-IBI method can be easily generalized to multicomponent systems, offering opportunities to parametrize CG nonbonded potentials for interactions between solvent components and chemical groups on biomolecules. A next step in this direction would be to consider single solutes in urea–water mixtures and optimize only the KB-IBI solute–solvent potentials without further optimizing the KB-IBI solvent–solvent potentials, which are then taken from the binary urea–water mixtures studied in this work. Since the salting-in (or salting-out) behavior in dilute solute/urea/water systems is determined by the solute–solvent Kirkwood–Buff integrals only (see eq 3), while the solvent–solvent KBIs can be assumed to remain unaffected in the limit of very low solute concentration, KB-IBI solute–solvent potentials can be developed for a variety of solutes in combination with a fixed set of potentials for the solvent–solvent interactions.

### ASSOCIATED CONTENT

#### Supporting Information

Tables listing $A$ values in eq 6, simulation results obtained using the MARTINI force field, and KB-IBI potentials obtained at three different urea concentrations (2.6, 4.7, and 6.8 M). This information is available free of charge via the Internet at http://pubs.acs.org.

### AUTHOR INFORMATION

**Corresponding Author**

*E-mail: vandervegt@csi.tu-darmstadt.de.*

**Notes**

The authors declare no competing financial interest.

### ACKNOWLEDGMENTS

The authors are grateful to Christine Peter for stimulating discussions. We thank Kurt Kremer and Emiliano Brini for useful comments on the manuscript.

### REFERENCES


**NOTE ADDED AFTER ASAP PUBLICATION**

This article was published ASAP on April 2, 2012. The Supporting Information file has been replaced. The correct version was published on April 25, 2012.