AIP | The Journal of Chemical Physics

C–IBI: Targeting cumulative coordination within an iterative protocol to derive coarse-grained models of (multi-component) complex fluids

Tiago E. de Oliveira, Paulo A. Netz, Kurt Kremer, Christoph Junghans, and Debashish Mukherji

Citation: The Journal of Chemical Physics **144**, 174106 (2016); doi: 10.1063/1.4947253
View online: http://dx.doi.org/10.1063/1.4947253
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/144/17?ver=pdfcov
Published by the AIP Publishing

C–IBI: Targeting cumulative coordination within an iterative protocol to derive coarse-grained models of (multi-component) complex fluids

Tiago E. de Oliveira,[1,2] Paulo A. Netz,[1,2] Kurt Kremer,[1] Christoph Junghans,[3,a)] and Debashish Mukherji[1,a)]
[1]*Max-Planck Institut für Polymerforschung, Ackermannweg 10, 55128 Mainz, Germany*
[2]*Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil*
[3]*Computer, Computational, and Statistical Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA*

(Received 21 January 2016; accepted 10 April 2016; published online 3 May 2016)

We present a coarse-graining strategy that we test for aqueous mixtures. The method uses pair-wise cumulative coordination as a target function within an iterative Boltzmann inversion (IBI) like protocol. We name this method coordination iterative Boltzmann inversion (C–IBI). While the underlying coarse-grained model is still structure based and, thus, preserves pair-wise solution structure, our method also reproduces solvation thermodynamics of binary and/or ternary mixtures. Additionally, we observe much faster convergence within C–IBI compared to IBI. To validate the robustness, we apply C–IBI to study test cases of solvation thermodynamics of aqueous urea and a triglycine solvation in aqueous urea. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4947253]

I. INTRODUCTION

Systematic structural coarse-graining, or systematically reducing degrees of freedom of a complex (macro)molecular system, is a paramount challenge of multiscale modeling.[1–3] Deriving coarse-grained (CG) models has several advantages — (1) When a multi-atom molecule and/or segments of a macromolecule are represented by a single site bead, the molecular dynamics (MD) simulation setups result in a smaller number of particles and thus give a significant computational gain. (2) The non-bonded interactions between CG beads are usually smooth. Therefore, large simulation time steps can be chosen. (3) The smooth interaction potentials lead to faster dynamics, which results in faster equilibration of the reference system. In this context, there are several possible CG techniques of deriving CG potentials, such as force matching,[4,5] inverse Monte Carlo,[6,7] Boltzmann inversion (BI)[8,9] and its extension to (iterative) Boltzmann inversion (IBI),[10] relative entropy,[11] and/or potential of mean force.[12–14] Additionally there are also well known CG models, examples include the Molinero water model[15] and the free energy based MARTINI model.[16] All these methods aim to target (or reproduce) a certain property of the underlying all-atom reference systems. Therefore, it is often difficult to map every property of a physical system within a unified CG model, posing grand challenge in the representability and transferability of CG models.[1,2] For example, in the case of liquid water, an IBI derived CG model usually presents a pressure of about 6000 bars,[17] which can be readjusted to 1 atm using a pressure correction.[10,17] However, this pressure correction compromises the fluid compressibility and thus results in unphysical fluctuations. In this context, a more recent work, employing a pressure correction at barostat level, could preserve both pressure and compressibility within a unified CG model.[18] The complexity of deriving CG models grows even further when dealing with macromolecular solvation in solution mixtures, where thermodynamic properties are intimately linked to delicate intermolecular interactions and local concentration and/or conformational fluctuations.[19]

A widely used structure based CG method is the well known BI[8,9] and the IBI,[10] where the pair-wise non-bonded potential is obtained by inverting $g(r)$ within an iterative procedure. In this context, being a simplified method, IBI works exceedingly well for several systems, including polymer melt,[8–10] single component fluids[20] and, also, to some extent for multicomponent fluids, to name a few. However, IBI does not guarantee that the derived CG model reproduces the same solvation thermodynamic state point as that of the reference all-atom system, especially for multi-component fluids. This is particularly because IBI targets to fit $g(r)$ and, for binary mixtures, the convergence of pair-wise $g(r)$ (unity at large distances) often suffers from the very nature of CG protocol. Therefore, a small absolute deviation in $g(r)$ can lead to a significant error in the cumulated coordination numbers. For example, estimation of coordination, given by

$$C_{ij}(r) = 4\pi \int_0^r g_{ij}(r')r'^2 dr' \tag{1}$$

with the indices i and j standing for every set of pairs, uses a volume integral of $g(r)$. This requires $g(r)$ to be multiplied by a factor of $4\pi r^2$ and a small error in $g(r)$ are weighted by a factor of $4\pi r^2$. Therefore, it is important to obtain a precise estimate of $g(r)$ for all r values and thus presents a need to devise a better, yet simple, CG protocol, which is the motivation behind this work.

Additionally, an accurate, yet simple, CG model is highly important for hybrid simulations, such as the adaptive

a)Authors to whom correspondence should be addressed. Electronic addresses: junghans@lanl.gov and mukherji@mpip-mainz.mpg.de

resolution scheme (AdResS).[21,22] In AdResS a small all-atom region is coupled to a CG reservoir. Correct thermodynamic conditions within the all-atom region are strongly related to the particle fluctuations and thus requiring a CG region that presents precise measure of the fluctuations compared to the all-atom region. This is even more important for the multicomponent fluids.[19,23]

The above mentioned reasoning poses grand challenges to the derivation of CG models to study solvation properties of solvent mixtures, especially because solvation thermodynamics is dictated by the following: (1) the energy density within the solvation volume of the macromolecule, (2) the local concentration fluctuation of the two solvent components, and (3) the entropic contributions, especially near the transition region of macromolecules where a delicate balance between entropy and energy plays a key role. In this context, the energy density is not only related to the (co)solvent-macromolecule interaction strengths but also to the solution composition within the solvation volume[19] and thus is related to the first shell coordination number. However, fluctuations are related to the convergence of the tails of pair-wise radial distribution functions g(r).[24] This presents a need for a protocol that can get both the above scenarios correct within a simplified CG strategy. Therefore, in this work we devise a method that aims to use $C(r)$, as a target function within an IBI-like iterative protocol. Our method not only gives a precise estimate of the coordination number in comparison to the reference all-atom system but also the precise estimate of the solvation properties. Added advantage of this protocol is that it presents a much faster convergence in comparison to the conventional IBI protocol.

The remainder of the paper is organized as follows: in Section II we sketch the method followed by the result and discussion in Section III. Finally in Section IV we draw our conclusions.

II. METHOD AND MODEL

A. All-atom simulations

The CG model is derived from an underlying all-atom reference system. We use test cases of aqueous urea mixtures and the solvation of a single triglycine in aqueous urea mixtures, which was studied by two of us in an earlier work.[25] The reference all-atom simulations are performed using GROMACS.[26] We use the Kirkwood-Buff (KB) derived force field for urea[27] and the SPC/E water model.[28] A combination of these two force-fields for the aqueous urea mixtures are known to reproduce correct solution thermodynamics. We consider four different urea molar concentrations c_u, ranging between 2.0M and 8.0M. We restrict the concentration to below 8.0M because urea is known to denature proteins at around 6M solutions.[29] System sizes are chosen to be consisting of ∼16 000 molecules, where we consider four different mole fractions 2.0M, 4.0M, 6.0M, and 8.0M. The specific choice of these system sizes give reasonable convergence in the thermodynamic properties, which usually suffer from severe system size effects within small systems.[19,23] The force field parameters for triglycine are taken from Gromos43a1.[30] The all-atom simulations are performed for 25 ns within an NpT ensemble, where the pressure is controlled with a Berendsen barostat at 1 atm pressure with a coupling time of 0.5 ps.[31] The initial configurations for the all-atom simulations are taken from a 50 ns long equilibrated sample from our earlier study.[25] The temperature is set to 300 K using a Berendsen thermostat with a coupling time of 0.1 ps. The integration time step is 1 fs. The interaction cutoff is chosen as 1.4 nm. Electrostatics is treated using particle mesh ewald.[32] The bond lengths of the urea molecules and triglycine is constrained using the LINCS algorithm.[33]

B. Coarse-grained simulations

The IBI and $C-$IBI derived CG potentials are used to simulate full blown CG configurations. The temperature is set to 300 K using a Langevin thermostat with a damping constant of 0.2 ps. Simulation time step is chosen as 4 fs and the cutoff distance is 1.4 nm. Simulations are conducted for 50 ns. We use the last 25 ns of a trajectory from 50 ns to calculate observables, such as g(r), urea activity coefficients γ_{uu}, and the shift in solvation free energy $\Delta\mathcal{G}_t$ of triglycine. CG simulations are also performed using GROMACS.[26]

III. RESULTS AND DISCUSSIONS

A. $C-$IBI: coordination iterative Boltzmann inversion

Before describing our $C-$IBI method, we first briefly comment on the conventional IBI method. The procedure starts from an initial guess for the potential of the CG model using $g_{ij}(r)$ obtained from the all atom simulation,

$$V_0(r) = -k_B T \ln\left[g_{ij}(r)\right]. \tag{2}$$

Then the potential is updated over several iterations using the protocol,

$$V_n^{\mathrm{IBI}}(r) = V_{n-1}^{\mathrm{IBI}}(r) + k_B T \ln\left[\frac{g_{ij}^{n-1}(r)}{g_{ij}^{\mathrm{target}}(r)}\right]. \tag{3}$$

During every iteration, a 1 ns long MD run of the CG system is performed using the potential obtained at the end of the preceding iteration. In Fig. 1(a) we present a comparison between fitted g(r) after 25 IBI iterations (symbols) and the reference all-atom data (solid lines). At a first look, it appears to be in reasonably good agreement. Moreover, the first shell coordination $C_{ij}^f = 4\pi \int_0^{r_0} g_{ij}(r') r'^2 dr'$ shows a deviation of roughly ∼2%–4%. Note that C_{ij}^f requires integration over the first peak of g(r), thus we have chosen $r_0 = 0.32$ nm for water-water, 0.48 nm for urea-water and 0.58 nm for urea-urea distributions, respectively. For example, a small error within the first few solvation shells (as observed in g(r)) cumulatively adds up to a large error at the tail and thus severely disturbs particle fluctuations. In this context, this small error is not recognized within the IBI protocol, where the corrections are weighted with a factor of $1/r^2$ – when looking into the coordination numbers. This leads to a position dependent error, which is most severe for large r values, and also added cumulative error from the earlier r values. Therefore, there is a need of a protocol, especially for binary mixtures, that

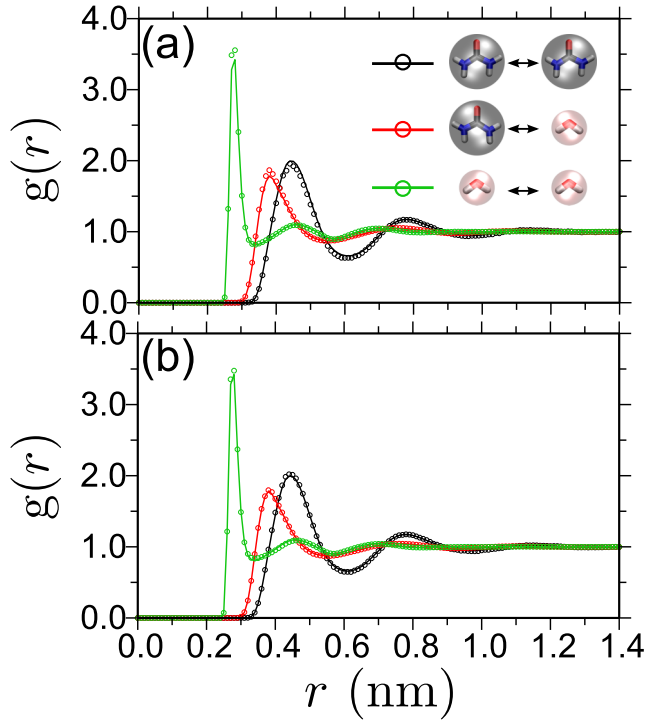

FIG. 1. Pair distribution function g(r) between different solvent pairs for a 6M aqueous urea mixture. The solid lines present the reference all-atom data and the symbols are fitted coarse-grained g(r) after 25 iterations. (a) is the data corresponding to the iterative Boltzmann inversion (IBI) and (b) corresponds to the coordination iterative Boltzmann inversion ($C-$IBI).

gives precise solvation properties. A theory that can serve as an excellent guide to achieve this purpose is the fluctuation theorem of Kirkwood and Buff (KB).[24] KB theory connects the pair-wise coordination with particle fluctuations and, thus, with the solution thermodynamics. KB theory makes use of the "so called" Kirkwood-Buff integrals (KBIs) G_{ij} defined as

$$\begin{aligned} \mathrm{G}_{ij} &= V\left[\frac{\langle N_i N_j\rangle - \langle N_i\rangle\langle N_j\rangle}{\langle N_i\rangle\langle N_j\rangle} - \frac{\delta_{ij}}{\langle N_j\rangle}\right] \\ &= 4\pi\int_0^\infty \left[\mathrm{g}_{ij}^{\mu\mathrm{VT}}(r) - 1\right] r^2 dr \end{aligned} \tag{4}$$

where averages in the grand canonical ensemble (μVT) are denoted by brackets $\langle\cdot\rangle$, V is the volume, N_i the number of particles of species i, δ_{ij} is the Kronecker delta, and $\mathrm{g}_{ij}^{\mu\mathrm{VT}}(r)$ is the pair distribution function in the μVT ensemble. For finite systems, however, a reasonable approximation leads to $\mathrm{g}_{ij}^{\mu\mathrm{VT}}(r) \approx \mathrm{g}_{ij}^{\mathrm{NVT}}(r)$ with $\mathrm{g}_{ij}^{\mathrm{NVT}}(r)$ being the pair distribution function in the canonical (NVT) ensemble. For big system sizes this is nearly almost (always) a safe approximation and thus leading to

$$\begin{aligned} \mathrm{G}_{ij}(r) &= 4\pi\int_0^r \left[\mathrm{g}_{ij}^{\mathrm{NVT}}(r') - 1\right] r'^2 dr' \\ &= C_{ij}(r) - \frac{4}{3}\pi r^3. \end{aligned} \tag{5}$$

Here the second term in the last line is a volume correction to $C_{ij}(r)$. Therefore, the quantity $G_{ij}(r)$ is also referred to as the excess-coordination, which could be connected to solvation properties of multi-component mixtures.[19,27,34] Therefore, we not only need the precise estimate of g(r) but also correct G_{ij}. This presents a need for an improved protocol that can correctly reproduce pair-wise coordination and the solvation properties. Thus we propose coordination iterative Boltzmann inversion ($C-$IBI). Here also, the initial guess is the same as in Eq. (2). However, the iterative protocol is modified to target $C_{ij}(r)$ given by

$$\mathrm{V}_n^{C-\mathrm{IBI}}(r) = \mathrm{V}_{n-1}^{C-\mathrm{IBI}}(r) + k_\mathrm{B}T\ \ln\left[\frac{C_{ij}^{n-1}(r)}{C_{ij}^{\mathrm{target}}(r)}\right]. \tag{6}$$

A cutoff distance for $C_{ij}(r)$ is chosen to be 1.5 nm, which is typically of the order of the correlation length of water-based molecular fluids. The advantage of using Eq. (6), unlike the IBI protocol, is that it presents equal weightage at every r value and, therefore, corrects $C_{ij}(r)$ at every r points precisely. Furthermore, because $C(r)$ is exactly reproduced using Eq. (6), it also exactly reproduces g(r). In Fig. 1(b) we present g(r) obtained using the $C-$IBI protocol. While there is hardly any visible distinction between g(r) obtained from $C-$IBI and the reference all-atom simulations, we find a much improved first shell coordination that shows ~0.5% deviation and also an improved tail convergence. A comparison of CG potentials derived from both methods, IBI and $C-$IBI, is shown in Fig. 2. It can be appreciated that the potentials derived from the two methods are distinctly different even when they show very similar g(r) (see Fig. 1), suggesting that a mere 25 IBI iterations may not be sufficient to get the correct coordination and, hence, the solvation properties.

It should be noted that the IBI protocol is the simplest form of CG method that works exceedingly well for several systems.[8–10,17] The initial guess of $V(r)$ in IBI is deduced from the Boltzmann distribution and the subsequent corrections in Eq. (3) are based on the difference in the distribution function while ignoring the higher order correlations. Furthermore, IBI can also be considered as IMC without cross correlation. In this context, IMC[6] can be derived from a thermodynamic argument. In our $C-$IBI method, we choose the same initial guess as the IBI (see Eq. (6)) and subsequent corrections

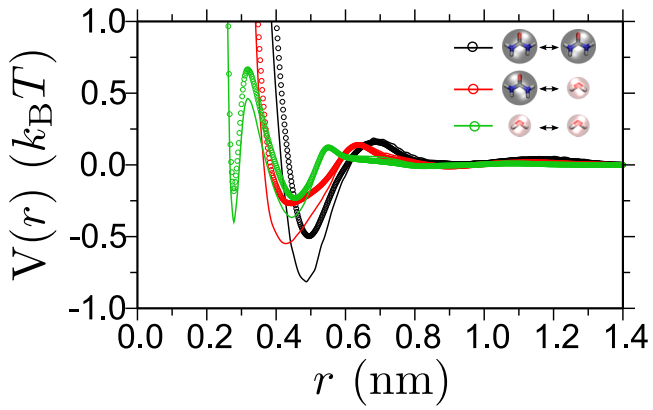

FIG. 2. Pair-wise coarse-grained (CG) potentials V(r) between different solvent pairs for a 6M aqueous urea derived using two different CG methods. The solid lines present CG potentials derived from IBI and the symbols are for the $C-$IBI protocol. Data is shown after 25 iterations within both protocols.

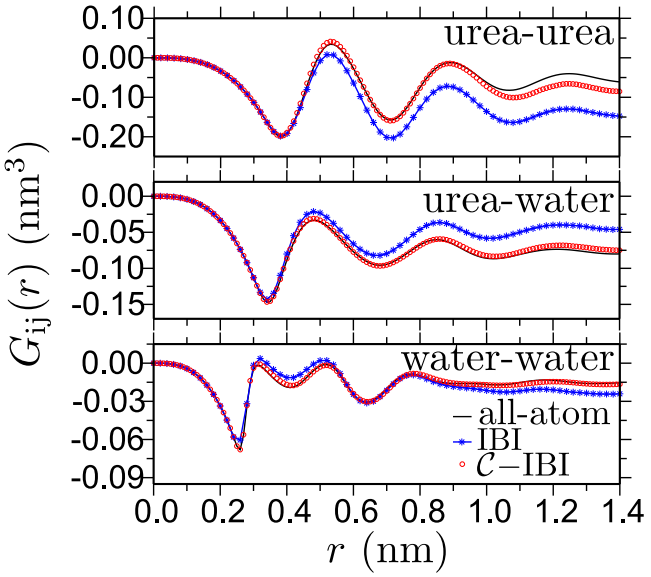

FIG. 3. Kirkwood-Buff integrals $G_{ij}(r)$ between different solvent pairs for a 6M aqueous urea mixture. The all-atom data are compared to the IBI and $\mathcal{C}-$IBI methods. The data is shown for a 25 ns long MD trajectory. The CG simulations are performed with the potentials obtained in Fig. 2.

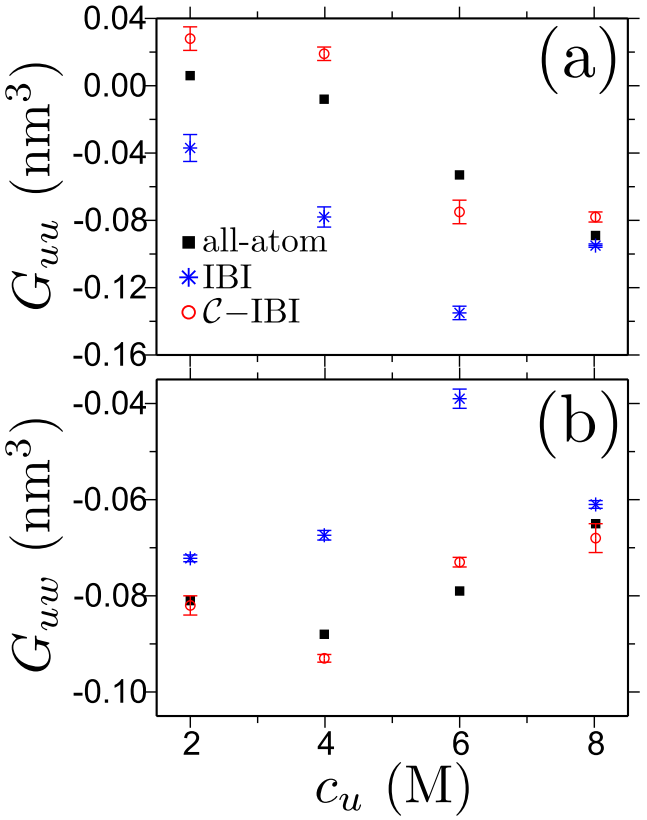

FIG. 4. Kirkwood-Buff integrals between urea-urea G_{uu} (a) and urea-water G_{uw} (b) as a function of molar concentration of urea c_u. We present a comparative plot of all-atom simulation, IBI and $\mathcal{C}-$IBI methods. The data is shown for a 25 ns long MD trajectory. The error bars are standard deviations obtained from four simulation trajectories. Note that we only show G_{ij} between the minor components of urea-urea and urea-water pairs that are most effected by the CG protocol.

are based on the difference in $C(r)$. Because of the nature of $\mathcal{C}-$IBI protocol, which aims to reproduce $C(r)$, this also tunes any irregularities that may cumulatively add up to an error at large r values. Therefore, reproducing $C(r)$ automatically guarantees the reproduction of underlying g(r). However, just targeting g(r) in an iterative procedure may not give a precise estimate of $C(r)$ and thus may lead to unrealistic fluctuation, especially for the multi-component fluids.

In Fig. 3 we present a comparative plot of $G_{ij}(r)$ between different solvent pairs. It can be seen that $\mathcal{C}-$IBI shows a reasonably satisfactory convergence to the reference all-atom data, while IBI data show significant deviation, especially between urea–urea and urea–water. Note that the values of G_{ij} are calculated by taking the averages of $G_{ij}(r)$ between 1 nm and 1.4 nm. In Fig. 4, we show G_{ij} between urea-urea G_{uu} and urea-water G_{uw}. It can be appreciated that the data from $\mathcal{C}-$IBI CG model can closely reproduce G_{ij} obtained from all-atom simulations. Note that we only show the data for urea-urea and urea-water pairs, where urea is the minor species. For water-water KBI, both models give reasonable agreement. Here it is important to mention that a slight deviation of G_{ij} can result in a large deviation in the particle fluctuation and thus leading to wrong solvation thermodynamics. Therefore, in Sec. III B, we will show that our method also gives a correct estimate of the solvation free energy.

The summary of the first shell excess coordination and the G_{ij} is presented in Tables I and II obtained from IBI and $\mathcal{C}-$IBI CG simulations and their comparison to the reference all-atom

TABLE I. A table showing comparative detail of the first shell excess coordination, obtained from all-atom (AA) simulations, iterative Boltzmann inversion (IBI) and coordination iterative Boltzmann inversion ($\mathcal{C}-$IBI). Results for IBI and $\mathcal{C}-$IBI are shown for $N_{\text{iterations}}$ iterations. For 2M and 4M due to the lower urea concentrations we perform a set of 10 IBI iterations before a set of $\mathcal{C}-$IBI iterations. This specific choice is adequate to obtain a reasonably good first estimate of the potential, before starting $\mathcal{C}-$IBI. We also include data for a set of simulations where we use IBI derived CG potential after 125 iterations.

		First shell excess coordination $= 4\pi \int_0^{r_0} \text{g}(r') r'^2 dr' - 4\pi r_0^3/3$											
		Urea-urea ($r_0 = 0.58$ nm)				Urea-water ($r_0 = 0.48$ nm)				Water-water ($r_0 = 0.32$ nm)			
c_u (M)	$N_{\text{iterations}}$	AA	$\mathcal{C}$-IBI	IBI	IBI-125	AA	$\mathcal{C}$-IBI	IBI	IBI-125	AA	$\mathcal{C}$-IBI	IBI	IBI-125
2	10 + 64	0.045	0.048	0.032	0.098	−0.036	−0.035	−0.034	−0.038	−0.011	−0.023	−0.026	−0.024
4	10 + 30	0.011	0.023	−0.009	−0.010	−0.033	−0.030	−0.030	−0.030	−0.005	−0.005	−0.004	−0.005
6	25	−0.013	−0.009	−0.033	−0.021	−0.034	−0.032	−0.021	−0.026	−0.002	−0.001	0.004	0.002
8	15	−0.039	−0.033	−0.029	−0.030	−0.018	−0.017	−0.015	−0.015	0.011	0.012	0.012	0.010

TABLE II. Same as Table I, but for Kirkwood-Buff integrals (or excess coordination) G_{ij}.

		G_{ij}											
		Urea-urea				Urea-water				Water-water			
c_u (M)	$N_{\text{iterations}}$	AA	C-IBI	IBI	IBI-125	AA	C-IBI	IBI	IBI-125	AA	C-IBI	IBI	IBI-125
2	10 + 64	0.006	0.028	−0.037	0.288	−0.081	−0.082	−0.072	−0.105	−0.023	−0.023	−0.024	−0.020
4	10 + 30	−0.008	0.019	−0.078	−0.092	−0.088	−0.093	−0.067	−0.067	−0.015	−0.015	−0.021	−0.020
6	25	−0.053	−0.075	−0.135	−0.076	−0.079	−0.073	−0.039	−0.073	−0.015	−0.016	−0.022	0.012
8	15	−0.089	−0.078	−0.095	−0.099	−0.065	−0.068	−0.061	−0.057	−0.009	−0.007	−0.012	−0.012

data. It can be seen that for the same number of iterations of both protocols, C−IBI gives much better estimates of the quantities than the standard IBI. It should be noted that the G_{ij} and C_{ij}^f are related to the volume around a given molecule. Therefore, smaller molecules also lead to smaller G_{ij} and C_{ij}^f values, making them highly sensitive to simulation protocol. Considering this, our C−IBI method seems to be working exceedingly well for the fluid mixtures. Furthermore, C−IBI also shows faster convergence than the IBI protocol. For the 6M aqueous urea mixture (see Fig. 3), we get a reasonable convergence within 25 iterations of C−IBI, which otherwise is not possible even after 125 iterations of IBI. In Tables I and II we also include IBI data after 125 iterations. A careful look on the tables also shows that neither G_{ij} nor the first shell excess coordination is correctly reproduced within the IBI protocol irrespective of the number of iterations, certainly not both quantities at the same time. However, C−IBI almost, always reproduces both these quantities within reasonable accuracy.

Furthermore, it should also be noted that for the smaller concentrations of urea, namely for 2M and 4M, we first run a set of 10 iterations of IBI, with 1 ns each step, followed by a certain number of C−IBI iterations. This procedure was performed to obtain a reasonable guess for the initial potential in Eq. (2), especially for the urea-urea pairs. Note that the convergence of g(r) for large r values are highly sensitive for multi-component systems, especially when one of the solvent components present at low concentrations.[19,23] Additionally, we also want to point out that for the smallest urea concentrations, namely 2M and 4M, IBI almost never gives any reasonable estimate of G_{ij} and C_{ij}^f. For example, in Table II and in Fig. 4, it can be appreciated that the urea-urea and urea-water KBIs using IBI CG models show large deviations from their all-atom data. Thus suggesting that IBI, despite giving after some iterations a reasonable starting potential guess, may still not be a suitable scheme to obtain reasonable fluctuations, especially when one of the solvent components is in low concentrations.

We would also like to point out that the C−IBI CG potentials are obtained without incorporating any adjustable pre-factors in the second term of Eq. (6). There are related methods that aim to reproduce KBI of binary[35] and ternary mixtures.[36] This method makes use of a pressure-like[10] KBI-based ramp correction. The advantage of ramp correction protocol is that it can be used to tune any thermodynamic property within a simplified protocol, such as the pressure, KBI and/or surface tension. However, a ramp correction usually requires a careful tuning of the pre-factor. Furthermore, while the ramp corrections can be used to tune a particular property of interest, it often sacrifices other properties. For example, when pressure corrections are applied to a system, it sacrifices fluid compressibility.[10] Therefore, the parameter free C−IBI is a protocol that, by construction, reproduces coordination, excess coordination, pair-wise solution structure, and thus the solvation free energies. Furthermore, because of the structure based nature of the C−IBI method, transferability is almost impossible over a wide range of concentrations. This is because when CG potentials are derived at two concentrations of urea, these two potentials only give precise thermodynamic properties on those two concentration state points. The use of these potentials in between concentrations often lead to inconsistent results. Therefore, the structure based CG protocols (such as C−IBI method) is thermodynamically consistent, but presents no concentration transferability. Moreover, when dealing with phase transition by changing temperature, one can use the

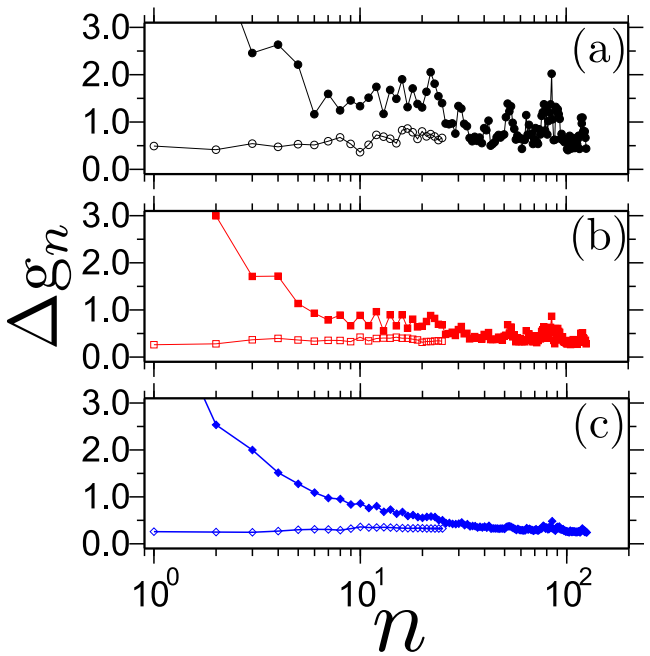

FIG. 5. Relative error Δg_n for a 6M aqueous urea solution obtained over coarse-graining iterations n. Δg_n measures the difference between the target radial distribution function $g^{\text{target}}(r)$ and the pair distribution per iterations $g_n(r)$, obtained over coarse-graining iterations. Solid symbol are obtained from IBI and the open symbols represent C−IBI. Data are shown for urea-urea in (a), urea-water in (b), and water-water in (c).

method proposed in Ref. 37 in conjunction with $C-$IBI method and thus presenting a possibility of obtaining temperature transferable CG model with $C-$IBI protocol. Furthermore, the pressure of the CG model derived using $C-$IBI remains around 5000 bars, a typical shortcoming of the almost all CG models. In Sec. III B, we will show how a slight change in G_{ij}, as reported in the Table II, can lead to a large, unphysical, deviations in the reference thermodynamic properties.

Lastly in this section we also want to comment on the convergence of g(r) in the $C-$IBI scheme and the IBI scheme. For this purpose, we calculate the relative error Δg_n between $g^{\text{target}}(r)$ and $g_n(r)$ after every iterations n, given by

$$\Delta g_n = \frac{\sqrt{\int [g^{\text{target}}(r) - g_n(r)]^2 dr}}{\int g^{\text{target}}(r) dr}. \quad (7)$$

In Fig. 5 we present Δg_n. It can be appreciated that the $C-$IBI converges much faster than the IBI. Furthermore, in both, IBI and $C-$IBI corrections, the convergence of one pair always disturbs the convergence of others. However, we not only find that $C-$IBI converges faster but also that they present much less structural fluctuations (see Fig. 5). More interestingly, we find that a reasonable structure can be obtained from almost very beginning of the $C-$IBI protocol, and any further iterations are performed to get a reasonable convergence of the tail of g(r) so that the model can reproduce correct fluctuations.

B. Solvation thermodynamics

1. Activity coefficient of aqueous urea

Solvation of a urea (u) molecule in the mixtures of water (w) and urea can be calculated using the expression[27]

$$\gamma_{uu} = 1 + \left(\frac{\partial \ln \gamma_u}{\partial \ln \rho_u}\right)_{p,T} = \frac{1}{1 + \rho_u (C_{uu} - C_{uw})}, \quad (8)$$

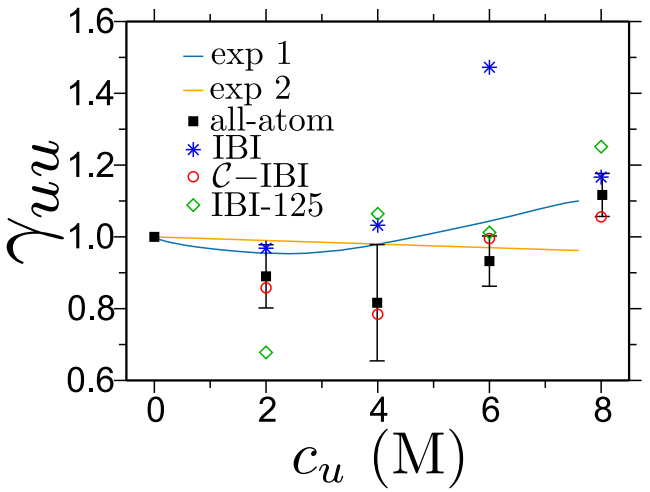

FIG. 6. γ_{uu} as a function of urea molar concentration c_u (see Eq. (8)). We present comparative data obtained using different coarse-grained method, all-atom reference system, and experiments. The data set corresponding to experiment 1 is taken from Ref. 38 and experiment 2 is taken from the Ref. 39, respectively.

where γ_u is the molar cosolvent activity coefficient, $\mu_u = k_B T \ln \gamma_u$ is the cosolvent chemical potential, the urea number density is ρ_u, C_{uu} is the urea-urea coordination number and C_{uw} is the urea-water coordination. In Fig. 6 we present γ_{uu} as a function of c_u. The data corresponding to $C-$IBI matches nicely with the all-atom reference system[25] and both data sets follow a similar trend as the experimental data set 1.[38] Furthermore, the IBI derived CG models (irrespective of the number of iterations) show a rather random variation in γ_{uu}. Fig. 6 also shows that $C-$IBI is a particularly powerful method over the full range of c_u, while the standard IBI CG models only give a slightly better estimate for large c_u and for 125 iterations of IBI. Note that while the convergence of the tail of g(r) is a grand challenge within an iterative procedure, $C-$IBI appears to be a much better alternative within a reasonable number of iterations.

2. Solvation free energy of triglycine in aqueous urea

So far we have presented results for the aqueous urea mixtures. In this section, we focus on reproducing the solvation properties of a triglycine in aqueous urea within our $C-$IBI protocol. For this purpose, we simulate one triglycine in a box containing water and urea with varying c_u as described earlier in the method section. For the CG model, we map the full triglycine molecule onto one CG bead. Furthermore, as in the cases of 2M and 4M aqueous urea mixtures, we first perform an initial set of 25 IBI iterations, followed by 30 $C-$IBI iterations. This is again motivated by the fact that we want to have a reasonable initial guess for the potential (see Eq. (2)). Here, however, every iteration consists of a 10 ns MD trajectory.

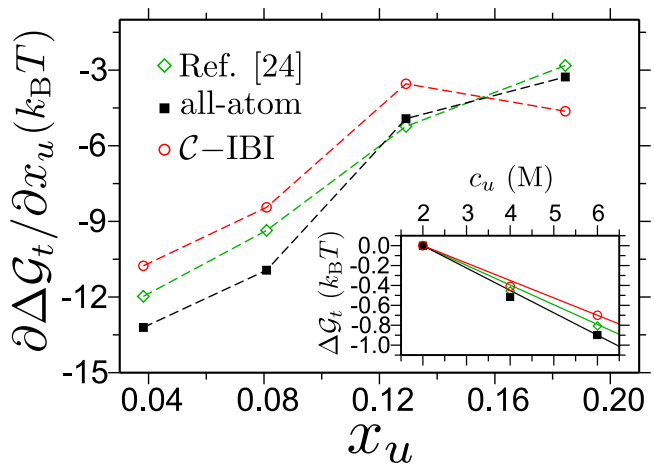

FIG. 7. Derivative of triglycine solvation free energy $\partial \Delta \mathcal{G}_t / \partial x_u$, as shown in Eq. (9), as a function of urea mole fraction x_u. Note here we use urea mole fraction, instead of urea molar concentration c_u, in the abscissa to be consistent with the Eq. (9). We present data for the all-atom simulations and from $C-$IBI models. For comparison we also include data from Ref. 25, which were obtained using a hybrid multiscale method. In the inset we present the variation of shift in solvation energy of a triglycine $\Delta \mathcal{G}_t$ with c_u. Note that we only restrict our data in the inset till $c_u = 6.0$M concentration of urea, because the experimental data is only available at around this c_u value. Straight lines are the linear fits to the data with the slopes listed in Table III.

TABLE III. A comparative table showing $m-\text{value} = \partial\Delta\mathcal{G}_t/\partial c_u$ obtained from the linear fits to the data in the inset of Fig. 7. The results are shown for all-atom simulations, $C-$IBI model, previous simulations,[25] and experimental data.[40] Note that while the experimental data are usually presented in kJ mol^{-2} L, the energy unit in our simulations is k_BT. Therefore, for better representability we provide the $m-$value in both these units.

	$m-$value	
	k_BT mol^{-1} L	kJ mol^{-2} L
All-atom simulation	−0.225	−0.557
$C-$IBI simulation	−0.175	−0.433
Simulation Ref. 25	−0.198	−0.492
Experiment Ref. 40	−0.197	−0.489

Note that we deliberately chose single triglycine molecules, to test the robustness of our method under extreme CG simulation conditions.

When a triglycine t at infinite dilution ($\rho_t \to 0$) is dissolved in an aqueous urea solution, the shift in the solvation free energy of triglycine ($\Delta\mathcal{G}_t$) is given by[34]

$$\lim_{\rho_t\to 0}\left(\frac{\partial\Delta\mathcal{G}_t}{\partial x_u}\right)_{p,T} = \frac{k_BT(\rho_w+\rho_u)^2}{\eta}\left(C_{tw}-C_{tu}\right), \quad (9)$$

where x_u is the urea mole fraction, k_B is the Boltzmann constant, $\eta = \rho_w + \rho_u + \rho_w\rho_u\left(C_{ww}+C_{uu}-2C_{uw}\right)$ is the preferential solvation parameter, and ρ_i is the number density of the ith component of the aqueous solutions. In Fig. 7 we present $\partial\Delta\mathcal{G}_t/\partial x_u$ as a function of x_u. $C-$IBI gives a reasonably good agreement with the all-atom data, while the data corresponding to the IBI CG model after 60 iterations did not show any visible convergence of C_{tw} and C_{tu} that could be used to obtain a reasonable estimate of the solvation energy. Furthermore, we do not only get a reasonable estimate of $\partial\Delta\mathcal{G}_t/\partial x_u$ but also for different G_{ij} components in Eq. (9), i.e., C_{tw}, C_{tu}, and η.

Integration of Eq. (9) gives the direct measure of the shift in solvation energy $\Delta\mathcal{G}_t$ with urea concentration. In the inset of Fig. 7 we show the variation of $\Delta\mathcal{G}_t$ with c_u. The slope of the linear fit to the data in the inset of Fig. 7 gives the direct measure of the "so called" $m-$value, which is defined as $\partial\Delta\mathcal{G}_t/\partial c_u$. Additionally, the $m-$value can be efficiently used to make a reasonable comparison between simulation and experimental observations. In Table III, we present $m-$values of a triglycine obtained from different methods. A reasonably good agreement between $C-$IBI, all-atom simulations and experiments suggest that the method can be used for any multi-component complex fluids.

IV. CONCLUSION

We have presented a parameter free coarse-graining (CG) strategy for complex mixtures. Our method uses cumulative coordination as a target function within an iterative protocol. We name our method $C-$IBI. $C-$IBI method not only gives a correct estimate of the pair-wise coordination but also by construction gives a good estimate of the solvation thermodynamics. More specifically, our CG method correctly reproduces both—energy density within the solvation volume and the local concentration fluctuations. Additionally, $C-$IBI shows much faster convergence with respect to the standard iterative Boltzmann inversion (IBI). We have used $C-$IBI derived CG potentials to study aqueous urea mixtures and the solvation of a small peptide in aqueous urea. The method presents a new, simplified, CG protocol and thus can be further used to study more complex (bio-)macromolecular systems, especially in mixed solvent environment.

ACKNOWLEDGMENTS

We thank Christine Peter and Nico van der Vegt for stimulating discussions. T.E.O. and P.A.N. acknowledges financial support from CNPq and CAPES from Brazilian Government and hospitality at the Max-Planck Institut für Polymerforschung, where this work was performed and generous allocation of computational facilities at the supercomputing center of CESUP-UFRGS. C.J. thanks LANL for a Director's fellowship. Assigned: LA-UR-15-28326. LANL is operated by Los Alamos National Security, LLC, for the National Nuclear Security Administration of the U.S. DOE under Contract No. DE-AC52-06NA25396. We thank Robinson Cortes-Huerto, Tanja Kling, Tristan Bereau, and Torsten Stühn for critical reading of the manuscript. Simulation snapshots in this manuscript are rendered using VMD.[41]

APPENDIX: $C-$IBI AS AN EXTENSION IN VOTCA

$C-$IBI method is implemented as an extension of the VOTCA package[20] that requires certain additional lines, presented in Fig. 8, to be included within the settings file to perform $C-$IBI iterations.

```
<cg>
 <non-bonded>
  <name>SOL</name>
  ...
  <inverse>
   <post_update>cibi</post_update>
   <post_update_options>
     <cibi>
      <do>1</do>
     </cibi>
     ...
   </post_update_options>
   ...
  </inverse>
 </non-bonded>
 ...
</cg>
```

FIG. 8. A schematic showing part of the script that is required within the settings file for the $C-$IBI iterations.

[1]C. Peter and K. Kremer, Soft Matter **5**, 4357 (2009).
[2]C. Peter and K. Kremer, Faraday Discuss. **144**, 9 (2010).
[3]W. G. Noid, J. Chem. Phys. **139**, 090901 (2013).
[4]F. Ercolessi and J. B. Adams, Europhys. Lett. **26**, 583 (1994).
[5]S. Izvekov and J. G. Voth, J. Chem. Phys. **123**, 134105 (2005).
[6]A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E. **52**, 3730 (1995).
[7]A. P. Lyubartsev, A. Naome, D. P. Vercauteren, and A. Laaksonen, J. Chem. Phys. **143**, 243120 (2015).
[8]W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn, Acta Polym. **49**, 61 (1998).
[9]W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn, Acta Polym. **49**, 75 (1998).
[10]D. Reith, M. Pütz, and F. Müller-Plathe, J. Comput. Chem. **24**, 1624 (2003).
[11]M. S. Shell, J. Chem. Phys. **129**, 144108 (2008).
[12]V. Harmandaris and K. Kremer, Macromolecules **42**, 791 (2009).
[13]J.-W. Shen, C. Li, N. F. A. van der Vegt, and C. Peter, J. Chem. Theory Comput. **7**, 1916 (2011).
[14]C. Dalgicdir, O. Sensoy, C. Peter, and M. Sayar, J. Chem. Phys. **139**, 234115 (2013).
[15]V. Molinero and W. B. Moore, J. Phys. Chem. B **113**, 4008 (2009).
[16]S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman, and A. H. De Vries, J. Phys. Chem. B **111**, 7812 (2007).
[17]H. Wang, C. Junghans, and K. Kremer, Euro. Phys. J. E **28**, 221 (2009).
[18]N. Dunn and W. G. Noid, J. Chem. Phys. **143**, 243148 (2015).
[19]D. Mukherji and K. Kremer, Macromolecules **46**, 9158 (2013).
[20]S. Y. Mashayak, M. N. Jochum, K. Koschke, N. R. Aluru, V. Rühle, and C. Junghans, PLoS one **10**, e131754 (2015).
[21]M. Praprotnik, L. Delle Site, and K. Kremer, J. Chem. Phys. **123**, 224106 (2005).
[22]S. Fritsch, S. Poblete, C. Junghans, G. Ciccotti, L. Delle Site, and K. Kremer, Phys. Rev. Lett. **108**, 170602 (2012).
[23]D. Mukherji, N. F. A. van der Vegt, K. Kremer, and L. Delle Site, J. Chem. Theory Comput. **8**, 375 (2012).
[24]J. G. Kirkwood and F. P. Buff, J. Chem. Phys. **19**, 774 (1951).
[25]D. Mukherji, N. F. A. van der Vegt, and K. Kremer, J. Chem. Theory Comput. **8**, 3536 (2012).
[26]S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess, and E. Lindahl, Bioinformatics **29**, 845 (2013).
[27]S. Weerasinghe and P. E. Smith, J. Phys. Chem. B **107**, 3891 (2003).
[28]H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, J. Phys. Chem. **91**, 6269 (1987).
[29]C. B. Afinsen, Science **181**, 223 (1973).
[30]W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, Gromos43a1 Hochschulverlag AG an der ETH Zürich, Zürich Switzerland, 1996.
[31]H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).
[32]U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. A. Pedersen, J. Chem. Phys. **103**, 8577 (1995).
[33]B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije, J. Comput. Chem. **18**, 1463 (1997).
[34]A. Ben-Naim, *Molecular Theory of Solutions* (Oxford University Press, New York, 2006).
[35]P. Ganguly, D. Mukherji, C. Junghans, and N. F. A. van der Vegt, J. Chem. Theory Comput. **8**, 1802 (2012).
[36]P. Ganguly and N. F. A. van der Vegt, J. Chem. Theory and Comput. **9**, 5247 (2013).
[37]B. Mukherjee, L. Delle Site, K. Kremer, and C. Peter, J. Phys. Chem. B **116**, 8474 (2012).
[38]R. H. Stokes, Aust. J. Chem. **8**, 2087 (1967).
[39]O. Miyawaki, A. Saito, T. Matsuo, and K. Nakamura, Biosci., Biotechnol., Biochem. **61**, 466 (1997).
[40]M. Auton and D. Wayne Bolen, Proc. Natl. Acad. Sci. U. S. A. **102**, 15065 (2005).
[41]W. Humphrey, A. Dalke, and K. Schulten, J. Mol. Graphics **14**, 33 (1996).