Role of Short-Range Directional Interactions in Coarse-Graining of Protic/Aprotic Liquids

Piotr Kowalczyk,*,† Piotr A. Gauden,‡ and Alina Ciach§

Applied Physics, RMIT University, GPO Box 2476 V, Victoria 3001, Australia, Department of Chemistry, Physicochemistry of Carbon Materials Research Group, N. Copernicus University, Gagarin St. 7, 87-100 Torun, Poland, and Institute of Physical Chemistry, Polish Academy of Sciences, Kasprzaka Street 44/52, 01-224 Warsaw, Poland

Received: April 27, 2009; Revised Manuscript Received: August 19, 2009

We study the role of short-range directional interactions in coarse-graining (CG) of protic (i.e., acetamide, methanol, ethanol, and water) and aprotic (i.e., acetone, benzene, and toluene) liquids at normal conditions. For this purpose, we introduce a new CG method in which the average interactions between atomistic molecules and CG beads measured in an N,P,T ensemble are preserved. We show that the spherically symmetric effective CG potential constructed according to our scheme is able to reproduce structural/thermodynamic properties of aprotic liquids; the heat of vaporization and total bonding energy profile for monomer are reproduced with good accuracy, while the density and radial distribution function are reproduced with fair accuracy within the proposed method. In contrast, the isobaric heat capacity is underestimated in the CG simulation because some of the fluctuations have been washed out from atomistic aprotic liquids. For protic liquids, spherically symmetric effective CG potential produces more structure, enhanced packing of beads, and underestimated isobaric heat capacity of CG liquids. This fundamental difference between protic and aprotic liquids can be explained by the presence of short-range directional interactions in the former liquids. We conclude that some information during the CG into spherically symmetric interaction potentials of protic liquids has to be lost. However, understanding how short-range directional interactions influence the structural and thermodynamic properties of the CG liquids seems to be the key for improving the CG methods.

I. Introduction

Coarse-graining (CG) of liquids, soft materials, and biomolecular systems is a fundamental problem of modern statistical mechanics for which a general solution has not been found yet. Many important biomolecular and condensed phase processes evolve on length and time scales that expand far beyond the capabilities of the conventional atomic-level simulation (i.e., atomistic time scales on the order of 10 ns and typical system sizes of atomistic simulation on the order of 50−60 Å). Thus, a description of static and dynamic properties of complex systems by well-defined lower resolution models is one of the most important problems of modern statistical mechanics and molecular simulation. CG molecular models seem to be particularly promising for simulations of complex systems since they can be used to bridge between the microscopic and the mesoscopic level of description. As mentioned by Voth,[1] "CG promises to provide a revolutionary advance for the scientific community, especially in the field of computer simulation". However, more fundamental works connected with a CG description of matter need to be done to fully explore its applicability and limitations.

Common to all CG methods is that the groups of atoms are clustered into new CG beads.[2−25] Those CG beads then interact through more computationally efficient effective interactions. The combination of the reduction of the total number of degrees of freedom of the system with these simplified effective interactions allows for a significant jump in the accessible spatial/temporal scales.[1] Reduction of molecular scale information during the CG process results in lower resolution models of the systems of interest. Beside a required simplification of atomistic structures/interactions, a CG model should retain the key physical features of the system of interest.

In a majority of papers connected with CG, the authors have concentrated on the reproduction of a particular structural property of the atomistic system (i.e., pair correlation function, etc.) at the CG level. Moreover, the authors often focused on a single particular macromolecule[26] or complex system.[27] In a few papers, the authors investigated the reproduction of several structural/thermodynamic properties of atomistic systems in CG simulations. Marrink and co-workers[15] have developed transferable coarse-grained potentials for semiquantitative lipid and surfactant simulations. The authors reproduced the density, isothermal compressibility, and mutual solubility of water and *n*-alkanes at the CG level with reasonable accuracy (see Table 1 in ref 15). However, the authors stated that[15] "Our goal was to reproduce the experimental densities of pure water and alkane systems around room temperature, the mutual solubility of oil and water, and the relative diffusion rates". Thus, by construction of CG effective potentials, all these properties computed in atomistic and CG simulations should be the same. Baron et al.[28] compared other thermodynamic properties computed from atomistic and transferable CG simulations. For water and *n*-alkanes, Baron et al. found systematic deviations of thermodynamic properties computed in atomistic and CG simulations (see Table 1 in ref 28). For instance, for liquid water at 303 K, the enthalpy of vaporization computed from atomistic simulation of 42 kJ/mol was comparable with 43.9 kJ/mol measured experimentally. In contrast, two CG models predicted 29 and 30 kJ/mol, respectively. For *n*-alkanes a large deviation of the

* To whom correspondence should be addressed. Tel.: +61 (03) 99252571. Fax: +61 (03) 99255290. E-mail: piotr.kowalczyk@rmit.edu.au.
† RMIT University.
‡ N. Copernicus University.
§ Polish Academy of Sciences.

10.1021/jp9038368 CCC: $40.75
Published on Web 09/09/2009

enthalpy of vaporization from experiment (up to 36% for *n*-butane) was observed.[28] Further, as the author explained, the CG models evidenced stronger ordering than their atomistic counterparts, especially in the first shell. The radial distribution functions computed from CG simulations of liquid *n*-C16 at 323 K are characterized by much higher correlations at short and long distances (see Figure 1 in ref 28). Hydration free energy, solute−solvent interaction energy change upon hydration, and solute−solvent entropy change upon hydration were underestimated in CG simulations for all studied *n*-alkanes (see Figures 1−4 in ref 28). These results are surprising because all structural/thermodynamic properties of *n*-alkanes can be easily reproduced by the CG force field developed by Ungerer et al.[29,30] or the simple isotropic CG scheme introduced by Mognetti et al.[31−33] In a series of papers, Mognetti et al.[31−33] successfully reproduced thermodynamic properties of carbon dioxide, *n*-alkanes, krypton, nitrogen, xenon, and benzene. For linear carbon dioxide and disklike benzene molecules, the authors coarse-grained the quadruple interactions by an isotropic quadruple model, whereas the dispersion interactions were expressed by a (12,6) Lennard-Jones potential.[32,33] In contrast to the CG scheme due to Marrink et al.,[15,28] Mognetti et al. successfully reproduced the thermodynamic properties of *n*-alkanes (including long ones, such as $C_{16}H_{12}$) such as coexistence densities and interfacial tensions between the coexisting vapor and liquid phases (see Figures 6−8 in ref 32). It is important to realize that the CG scheme introduced by Mognetti et al. was built on the strong theoretical foundations (see the work of Stell et al.[34]). Interestingly, the authors were able to predict the mixture properties of the molecules mentioned above by using the standard Lorentz−Berthelot combining rule for cross interactions.[32] These results are remarkable achievement in the description of fluid mixtures by computer simulations. However, it is worth emphasizing that the CG method due to Mognetti et al. was applied only to small molecules that are characterized by weak electrostatic forces.[31−33] The open question is how this CG scheme works for simple, nearly spherical molecules characterized by strong anisotropy of molecular forces such as water, alcohols, or their mixtures. In fact, the description of intermolecular forces between water molecules in the liquid state by spherical potentials is a hard problem that has been studied by Johnson et al.,[35] Izvekov and Voth,[36] and others. Marrink et al.[15] stated that "the properties of bulk water can be reproduced by coarse graining four water molecules into one LJ bead of type P". However, the authors were only able to reproduce the density and isothermal compressibility of liquid water (see Table 2 in ref 15). In contrast, Johnson et al.[35] concluded that "Because at a given state point the pair potential that reproduces the pair structure is unique, we have therefore explicitly demonstrated that it is impossible to simultaneously represent the pair structure and several key equilibrium thermodynamic properties of water with state point dependent radially symmetric pair potentials". Indeed, Izvekov and Voth[36] successfully reproduced the radial distribution functions for liquid water by their CG method. However, all thermodynamic properties of liquid water, except density, computed from a spherically symmetric CG model of the water molecule were highly underestimated (see Table 2 in ref 36). For instance, the average potential energy of liquid water computed in atomistic simulation at 298 K of −41.5 kJ/mol was reduced up to −17.9 kJ/mol in one-site water CG simulation. Clearly, this disagreement cannot be explained by the packing effect because atomistic water molecules are nearly spherical. Thus, the following question arises: *Does the observed disagreement result from the intrinsic properties of water?* Allen and Rutledge presented a new density-dependent implicit solvent CG model.[37] Within their CG scheme, the authors successfully reproduced the radial distribution function and the excess chemical potential for a system of Lennard-Jones particles in an implicit solvent simulation (i.e., the solvent−solute and solvent−solvent interactions were embedded into effective potential between solute particles). Note, however that Allen and Rutledge described the system of spherical particles interacting via weak dispersion forces.[37] How this CG method works for particles interacting via highly anisotropic forces is an interesting question to address.

We believe that the fundamental problem is to connect the accuracy and limitations of the CG model with the molecular interactions between beads at an atomistic level. Finally, we argue that the ability of the CG model to predict a broad spectrum of properties of atomistic systems at the CG level is a more interesting and challenging issue then reproduction of just one particular feature. Here, the term "broad spectrum" is not limited to structural/thermodynamic properties that have been used for construction of the CG molecular forces. For example, it has been shown that many CG methods are able to reproduce the atomistic radial distribution functions at the CG level.[7,38,39] It is not surprising because at short distance the correlations in dense liquids are dominated by excluded volumes (i.e., the correct size of the cluster of atoms grouped into a bead is essential for reproduction of the short-range correlations).[40−43] However, from both theoretical and practical points of view, it is more interesting to show how the remaining structural/thermodynamic properties computed in atomistic simulation (i.e., total bonding energy for monomers, heat of vaporization, heat capacity, density, etc.) are reproduced at the CG level. If those properties are different in atomistic and CG simulation, the question to answer is the source of this disagreement. Clearly, any comparison between structural/thermodynamic properties that have been used for the adjustment of the CG forces is meaningless because those properties are equal in atomistic and CG simulations by construction of the CG model.

In fact, as will be intuitively obvious, many beads look very similar at finite temperatures (here, "beads" can be visualized as rotating dynamical objects). Continuing, one can expect that interactions between beads of quasi-spherical shape, which interact via weak isotropic interactions (for example, aprotic liquids) can be approximated by a simple spherically symmetric potential. In other words, for these particular types of beads, many structural/thermodynamic properties computed at the atomistic level can be easily reproduced by a carefully formulated CG scheme. On the other hand, highly directional short-range interactions between quasi-spherical beads (for example, protic liquids) cannot be accurately described by a spherically symmetric potential. Some atomistic information has to be lost because the interactions between those beads are highly anisotropic at finite temperatures (i.e., the assumption of isotropic symmetry of molecular forces leads in principle to incorrect description of protic beads interacting with highly directional forces).

In the current paper, we address the following question: *how does the molecular shape and directionality of interactions at the atomistic level influence the coarse-graining of protic/aprotic liquids?* To answer this question, we introduce a new CG method. For selected protic (i.e., methanol, acetamide, ethanol, and water) and aprotic (i.e., acetone, benzene, and toluene) liquids, we compare several properties measured in atomistic and CG simulations, including the following: heat of vaporization, density, isobaric heat capacity, radial distribution functions,

and distributions of total bonding energy for monomers. Let us note that the specific heat depends on the variance of the enthalpy and is determined by fluctuations of the energy and the volume in the N,P,T ensemble.[44,45] In the case of spherical beads interacting with spherically symmetric interaction potentials, the energy and volume fluctuations are associated with deviations of the interparticle separations from the average value. For particular relative orientations of protic molecules, hydrogen bonds (H-bonds) are formed and the energy is significantly lower compared to the orientations in which the H-bonds are absent. Because of thermal motion, the molecules rotate, and different relative orientations of the molecules, associated with different energies, yield additional contribution to the energy fluctuations. As a result, the energy distribution is much broader than for isotropic potentials. Similarly, the volume of the system depends on relative orientations of nonspherical molecules, and this dependence on orientations yields additional contribution to the volume fluctuations. Thus, we can expect that the quantities determined by fluctuations are underestimated in the CG simulations, and it is interesting to compare this effect for protic and aprotic molecules with different nearly spherical shapes. Indeed, the isobaric heat capacity is significantly underestimated in the CG simulations. Beside the isobaric heat capacity (the property that is computed from fluctuations) and density that deviates from the atomistic simulation result by 2−10%, our CG method successfully reproduces all the remaining properties listed above for aprotic liquids at 298 K and 1 bar. In contrast, we found that the spherically symmetric CG potential, optimized so as to reproduce the average potential energy, produces much more structure and enhanced packing for protic liquids at 298 K and 1 bar. We explain this observation by taking into account the role of the directionality of short-range interactions in atomistic simulations.

II. Theory

II.1. Coarse-Graining. Let us consider a system of identical atomistic rigid molecules, consisting of n atoms whose separations within the molecule are fixed. The intermolecular force between molecules A and B is determined by the potential $U_a(\mathbf{q})$, where $\mathbf{q} \equiv \{\mathbf{r}_i^A - \mathbf{r}_j^B\}_{i,j=1,2,...,n}$ is the set of separations between each atom in the molecule A and each atom in the molecule B ($\mathbf{r}_i^A$ and $\mathbf{r}_j^B$ denote the position of the ith atom in molecule A and jth atom in the molecule B, respectively). Because in the case of rigid molecules the intramolecular separations $||\mathbf{r}_i^A - \mathbf{r}_j^A||$ are fixed, the independent variables determining $U_a(\mathbf{q})$ can be chosen as $\bar{\mathbf{q}} = (r_{A,B},\boldsymbol{\omega}_A,\boldsymbol{\omega}_B)$, where $r_{A,B} = ||\mathbf{r}_A - \mathbf{r}_B||$ is the distance between the centers of mass of the molecules A and B and $\boldsymbol{\omega} = (\theta,\phi,\psi)$ represents Euler angles. Let us approximate such an atomistic rigid molecule of any shape by an effective spherical bead. Following this assumption, the anisotropic potential $U_a(\mathbf{q})$ is approximated by the spherically symmetric effective potential $U_c(r)$, which depends only on the distance $r = ||\mathbf{r}||$ between the centers of mass of the beads. Although the effective potential $U_c(r)$ describing the interactions between the spherical beads on the CG level is unknown, its asymptotic properties are,

$$\begin{aligned} U_c(r) &\to \infty, \quad \text{as} \quad r \to \sigma \\ U_c(r) &\to 0, \quad \text{as} \quad r \to \infty \end{aligned} \tag{1}$$

where σ denotes the diameter of the hard-core of the spherical bead.

To extract a trial effective potential $\Psi(r)$ from an atomistic molecular simulation, Ercolessi et al.[46,47] proposed a so-called "force matching method". Following this concept in the present context, we extract a trial spherically symmetric CG potential between the beads from the atomistic simulation. In the first step, we compute a pair potential of the orientation-averaged force between two molecules (say, A and B). More precisely, the force between the two molecules is averaged over their orientations and over positions and orientations of the remaining molecules in the N,P,T ensemble. The potential is given by

$$\Psi(r) = \int_{r_{\min}}^{r} f(\xi)\, d\xi \quad \text{for } r \le r_{\text{cut}} \tag{2}$$

where r is the distance between the centers of mass of the molecules A and B, r_{cut} is the cutoff distance used for calculation of intermolecular potentials in atomistic fluid, and $r_{\min}$ is the shortest distance between the molecules. The term $f = ||\mathbf{f}||$ is the magnitude of the average force $\mathbf{f}$ between the molecules A and B when the separation between their centers of mass is ξ, and $\mathbf{f}$ is given by

$$\mathbf{f}(\xi) = -\left\langle \sum_{i\in A}\sum_{j\in B} \nabla u_{ij}(r_{ij}^{A,B}) \right\rangle_{\xi} \tag{3}$$

where $\langle ... \rangle_{\xi}$ denotes the conditional ensemble average, with fixed separation ξ between the centers of mass of the molecules A and B, $r_{ij}^{A,B} = ||\mathbf{r}_i^A - \mathbf{r}_j^B||$ is the distance between the ith atom in molecule A and the jth atom in the molecule B, and $u_{ij}(x)$ denotes the distance dependent site−site intermolecular potential expressed by (12,6) Lennard-Jones or Columbic potential (see the next paragraph). The function $\mathbf{f}(\xi)$ corresponds to the net force between two molecules when their centers are separated by the distance ξ, derived by Voth and co-workers (see refs 1 and 10). Voth and co-workers demonstrated that this force function could be derived as a mean field approximation to the many-body mean force field.[1,10] The average force given by eq 3 is simply collected in atomistic simulation. For this purpose, the distance $0 \le r \le r_{\text{cut}}$ is divided into bins with a width of 0.02 Å. After an initial equilibration of the atomistic fluid, the equilibrium configurations are swapped with regular frequency. In these configurations, each pair of molecules separated by the distance ξ was used for the accumulation of the average force between the beads given by eq 3. The trial spherically symmetric potential between the beads is computed from eq 2 by numerical integration. We found that the trial effective CG potential computed from eq 2 is characterized by correct asymptotic properties and reproduces the pair distribution function reasonably well, as expected by construction of this potential (see Appendix A for more details). However, it underestimates the average potential energy for all studied fluids measured in CG simulations (see Figures 1, 2, 5, and 6). This is because the distribution function for positions of the centers of mass is obtained by integrating the Boltzmann factor over orientational degrees of freedom, while the average energy is obtained by integrating over orientational degrees of freedom the Boltzmann factor times energy (see Appendix B for more details).

In this work, we propose a simple method of improving the trial CG potential. Our method of optimizing the coarse-graining procedure is neither unique nor yielding the best possible form of the CG potential. The advantage of our method is its simplicity and possibility of comparing the performance of the effective potential optimized within our scheme for different classes of molecules, including protic and aprotic ones.

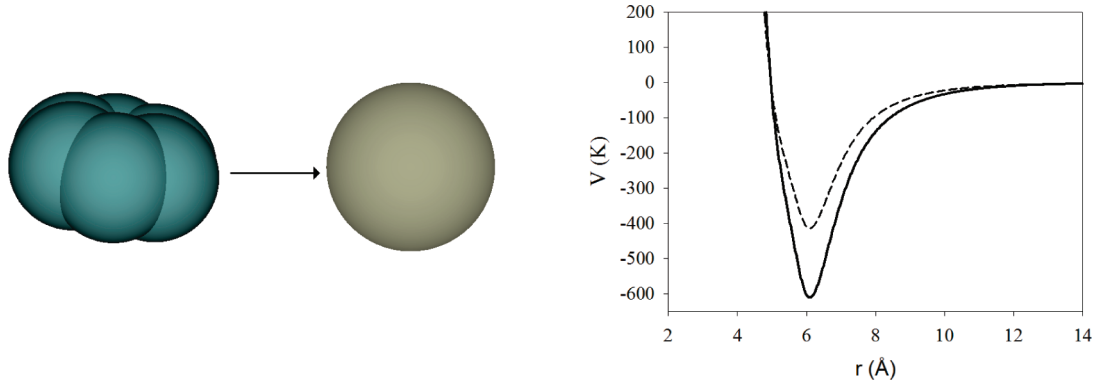

Figure 1. Coarse-graining of benzene molecule at 298 K and 1 bar to a spherical bead of $\sigma = 4.96$ Å (i.e., we define the collision diameter of the spherical bead as a value for which the effective potential is zero). The right panel presents the trial effective potential, $\Psi(r)$, (dashed line) and the final effective CG potential (solid line), $U_c(r) = C\Psi(r)$ where $C = 1.474\,976$.

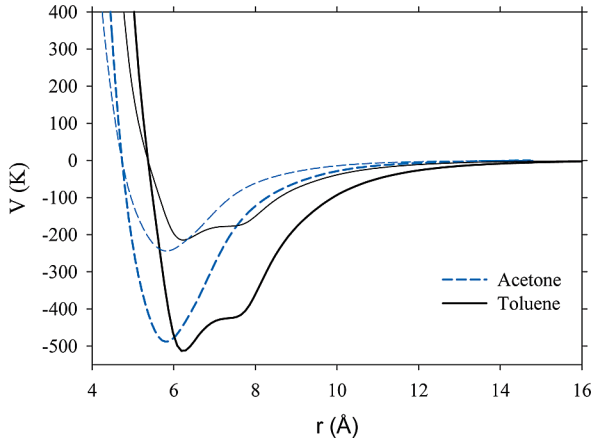

Figure 2. Effective CG (bold lines) and trial effective potential (thin lines) computed for acetone and toluene at 298 K and 1 bar. The effective sizes of the acetone and the toluene beads are 4.74 and 5.37 Å, respectively (i.e., we define the effective size of the spherical bead as a value for which the effective potential is zero).

In order to describe the radial distribution function with reasonable accuracy, let us postulate that the CG effective potential is just proportional to the trial effective CG potential calculated in the atomistic simulation

$$U_c(r) = C\Psi(r) \tag{4}$$

For the above form of the effective potential, we require that the average potential energy per bead in the atomistic and CG model is equal. The optimal CG potential can be found from the solution of the following variational functional:

$$\min[(\langle U_a\rangle - \langle U_c\rangle)^2] \tag{5}$$

where $\langle U_a\rangle$ and $\langle U_c\rangle$ are the average potential energy per bead computed from the atomistic and CG molecular simulation, respectively. In each case, the MC simulations are performed in an isobaric−isothermal ensemble. To solve the variational functional given by eq 5, we combined the Monte Carlo simulation in the N,P,T ensemble with the real-coded genetic algorithm.[48] However, we want to underline that any optimization method can be used for the solution of the variational functional given by eq 5.

For an unspecified shape of the effective potential $U_c(r)$, the result of the procedure given by eq 5 is not unique, but with the assumption given by eq 4, the unknown parameter C can be determined in a unique way. For the molecules studied in this work, the optimal values of C are given in Table 1. The average energy per bead in the atomistic and CG N,P,T simulation with the potentials $U_c(r)$ determined by eq 5 should be the same, but the radial distribution functions will be different for different shapes of $U_c(r)$. Since we reduce the possible shapes of the effective potential to the form given by eq 4, we obtain the effective potential that should yield reasonable, but not necessarily the best possible form of the radial distribution function under the condition that the average energy per bead is reproduced precisely. We emphasize that all our results are directly related to the approximation given by eq 4.

TABLE 1: Constants, C, Used for Rescaling of the Trial Effective CG Potential Given by Equation 4

molecule	C
acetone	2.001599
benzene	1.474976
toluene	2.397265
methanol	2.32518
acetamide	2.293975
ethanol	2.265495
water	2.24115

Apart from the assumption expressed by eq 4, there is only one requirement that controls our CG of an atomistic molecule, i.e. the average energy per bead in atomistic and CG N,P,T simulation should be the same. All other structural/thermodynamic properties measured on the CG level result from these two fundamental assumptions. *Note that since our CG method is the same for all types of molecules, the discrepancies between results obtained in the atomistic and CG simulations should shed light on the role of directionality of interactions and of the geometrical shape of the molecule for the considered quantity*.

II.2. Potential Models in Atomistic Simulations. In the atomistic simulations, we express the intermolecular potential between two molecules A and B as a sum of site−site terms,[49,50]

$$U^{A,B}(\mathbf{q}) = \sum_{i\in A}\sum_{j\in B} u_{ij}(r_{ij}^{A,B}) \tag{6}$$

where the sum is taken over all sites i of molecule A and sites j of molecule B, $r_{ij}^{A,B}$ is the distance between two LJ sites i and j on molecules A and B, respectively, and $u_{ij}(r)$ denotes distance dependent site−site interaction potential. Depending on the type of the site, $u_{ij}(r)$ is given by a pairwise dispersion or electrostatic

interaction energy. We assumed that the dispersion energy of interaction is given by the (12,6) Lennard-Jones equation[44,45,50]

$$u_{ij}^{A,B} = 4\varepsilon^{A,B}\left[\left(\frac{\sigma^{A,B}}{r_{ij}^{A,B}}\right)^{12} - \left(\frac{\sigma^{A,B}}{r_{ij}^{A,B}}\right)^{6}\right] \quad (7)$$

The parameters of the potential for all investigated molecules, i.e. $\sigma^{A,B}$ and $\varepsilon^{A,B}$, are taken from different force fields to ensure that the presented analysis does not depend on the particular force field but only on intrinsic properties of the studied liquids. We used the following force fields: benzene and toluene (Wick et al.[51] force field for arenes), water (TIP4P in the work of Jorgensen et al.[52]), acetone, acetamide, methanol, and ethanol (OPLS all atoms force field in the work of Jorgensen et al.[53]).

We modeled electrostatic force via the Coulomb law of electrostatic potential[44,45,50]

$$u_{i,j}^{A,B} = \frac{1}{4\pi\varepsilon_0}\frac{q_i^A q_j^B}{r_{ij}^{A,B}} \quad (8)$$

where ε_0 is the permittivity of free space ($\varepsilon_0 = 8.85419 \times 10^{-12}$ $C^2/(N\ m^2)$), q_i^A denotes the value of point charge i on molecule A, q_j^B denotes the value of point charge j on molecule B, $r_{ij}^{A,B}$ is the distance between two charges i and j on molecules A and B, respectively. The values of the point charges for all investigated molecules are taken from the force fields listed above.[51−53]

II.3. Simulation Methodology on the Atomistic and CG Level. We performed all atomistic and CG simulations in an isobaric−isothermal ensemble by the Monte Carlo method at 1 bar and 298 K. We modeled all investigated molecules as rigid bodies. In the N,P,T ensemble, we used two perturbation moves: displacement and rotation of a randomly selected molecule and the simulation cell volume change step to equilibrate the external pressure. All simulations were performed in cycles; with each cycle consisting of N number (N denotes the number of molecules in the simulation cell; here, we assume $N = 1000$) of attempted particle displacements and a single simulation cell volume change step. The center of mass of the selected molecule was randomly displaced, whereas the rotation of the molecule was performed in the system of the internal coordinates. The maximum displacement as well as maximum change of the simulation cell volume is adjusted to an ≈50% acceptance rate. The transition probabilities for molecule displacement, rotation, and simulation volume change step in the N,P,T ensemble are given elsewhere.[44,45] In the isothermal−isobaric ensemble, we generated 6×10^7 configurations. The first 1×10^7 configurations were discarded to guarantee equilibration, whereas the remaining 5×10^7 configurations were used to obtain the desired thermodynamic properties. For arbitrarily selected points, we stored the fluctuations in the total energy. Next, the variations of the internal energy were analyzed to ensure that thermodynamic equilibrium was achieved. The accuracy of our MC simulation in the N,P,T ensemble has been tested for several models of liquid water (see the Supporting Information).

II.4. Thermodynamic and Structural Properties Measured in the Atomistic and CG Simulations. We computed the average density of all investigated liquids from the following expression:[52]

$$\langle\rho\rangle = N/\langle V\rangle \quad (9)$$

where $\langle ... \rangle$ denotes an ensemble average, V is the volume of the simulation box, and N denotes a number of atomistic/CG beads. Additionally, we computed the heat capacity and heat of vaporization from the fluctuation formulas[52]

$$C_p = (\partial\langle H\rangle/\partial T)_{N,p} = (1/Nk_bT^2)(\langle H^2\rangle - \langle H\rangle^2) + 3R \quad (10)$$

$$\Delta H_{vap} = E(\text{gas}) - E(\text{liq}) + p(V(\text{gas}) - V(\text{liq})) \approx - \langle E(\text{liq})\rangle/N + RT \quad (11)$$

$$H = E(\text{fluid}) + pV = \sum_{A<B} E^{A,B} + pV \quad (12)$$

where $E^{A,B}$ is the total potential energy between particles A and B, T denotes temperature, p is the pressure, H is the enthalpy, R is the universal gas constant, and k_b denotes the Boltzmann constant. The last term in eq 10 is associated with translational and rotational kinetic energy of molecules. In the CG simulations, the beads have no rotational degrees of freedom. Since the MC simulations of the CG system are supposed to model real molecules, we assume that the kinetic energy should be the same as in the real system; therefore, we use the same formula (10) in the CG simulations.

We compared the structure of the atomistic and the corresponding CG liquids by computing the radial distribution function (mass center−mass center)[44,45]

$$g(r) = \rho_r/\rho = \frac{N_r V}{N(4\pi r^2)\,\mathrm{d}r} \quad (13)$$

were r denotes the intermolecular distance between the mass centers of two molecules, ρ_r is the number of atomistic/CG molecules (mass centers) in the volume element, ρ is the overall density, N_r is the number of atomistic/coarse-grained beads (mass centers) inside the shell of thickness dr, N is the total number of molecules in the system, and V is the defined volume.

Finally, we compared the total bonding profiles for monomers computed for atomistic liquids and their CG counterparts. The histograms of total bonding energies for monomers were constructed by the algorithm described in a series of papers by Jorgensen's group.[54]

III. Results and Discussion

At the beginning of this section, we want to underline that we will not consider the agreement between the average potential energy per bead computed from the atomistic and CG simulations for the optimized value of the scaling parameter C, because by construction of the CG effective potential the difference between these two energies should vanish (see eqs 4 and 5). The agreement between the average potential energy per bead computed from the atomistic and CG simulation is the assumption and says nothing about the limitations and accuracy of the presented CG method. For illustration of the dependence of the average potential energy and radial distribution function on the scaling parameter C, we will present the results for energy as a function of C in the case of ethanol.

To begin the discussion of the results, we divide all studied liquids into two groups. The first group includes aprotic liquids, i.e., acetone, benzene, and toluene. Note that for those liquids, $U/k_bT \approx 10-13$ (where U denotes the average potential energy per bead and k_bT is the thermal energy). Moreover, as is

TABLE 2: Comparison of the Thermodynamic Properties of Aprotic Molecules Computed from the CG and Atomistic Simulations[a]

molecule	property	atomistic simulation	coarse-grained simulation
acetone	E_{pot} (kcal/mol)	-6.15 ± 0.008	-6.15 ± 0.001
	ΔH_{vap} (kcal/mol)	6.74 ± 0.008	6.74 ± 0.001
	C_p (cal/(mol K))	20.09 ± 0.73	10.54 ± 0.07
	d (g/cm^3)	0.7 ± 0.001	0.77 ± 0.0003
benzene	E_{pot} (kcal/mol)	-6.79 ± 0.007	-6.81 ± 0.0009
	ΔH_{vap} (kcal/mol)	7.38 ± 0.007	7.4 ± 0.0009
	C_p (cal/(mol K))	16.34 ± 0.7	11.59 ± 0.17
	d (g/cm^3)	0.85 ± 0.0006	0.87 ± 0.00001
toluene	E_{pot} (kcal/mol)	-7.72 ± 0.004	-7.73 ± 0.0008
	ΔH_{vap} (kcal/mol)	8.31 ± 0.004	8.32 ± 0.0008
	C_p (cal/(mol K))	18.86 ± 1.1	10.55 ± 0.1
	d (g/cm^3)	0.84 ± 0.0005	0.93 ± 0.0003

[a] Abbreviations: E_{pot} denotes the average potential energy per bead, ΔH_{vap} is the heat of vaporization, C_p is the isobaric heat capacity, and d denotes density.

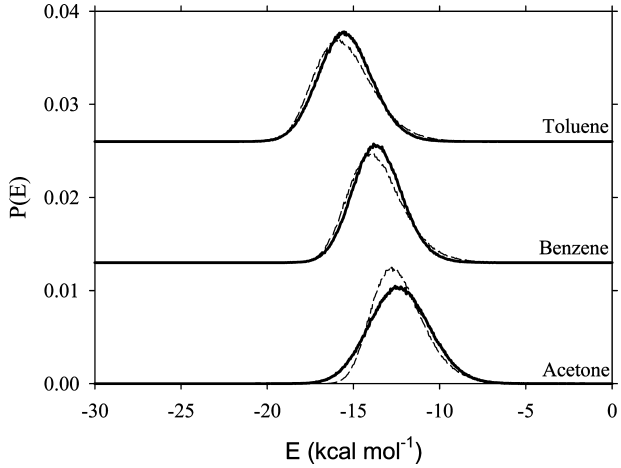

Figure 3. Distributions of total bonding energies for monomers in aprotic liquids computed from CG (dashed lines) and atomistic (solid lines) simulation at 298 K and 1 bar. The interaction potential in the CG simulation is such that the average energies in the CG and atomistic simulations are the same. Note that the most probable energies are slightly different.

commonly known, those molecules are not able to form H-bonds by themselves. The second group of studied liquids includes protic liquids, i.e., methanol, acetamide, ethanol, and water. For those liquids, the molecular interactions are stronger, i.e., $U/k_bT \approx 16.5-17$. Moreover, at the atomistic level, the interactions between protic liquids are highly directional (i.e., they are dominated by short-range H-bonds). As will be intuitively obvious, the integration of molecular interactions over all rotational degrees of freedom is more realistic for aprotic solvents. Indeed, our CG method reproduces structural and thermodynamic properties of liquid acetone, benzene, and toluene reasonably well, as is presented in Table 2 and Figures 3 and 4. The effective CG potential is qualitatively similar to the (12,6) Lennard-Jones one, as is displayed in Figures 1 and 2. However, it is characterized by a smoother repulsive core. For toluene, the shape of the effective CG potential near the minimum is more complex. Nevertheless, for all studied aprotic molecules, the attractive well depth of the trial effective potential computed directly from atomistic simulations is too shallow (see Figures 1 and 2). Thus, the adjustment of the effective CG potential is necessary to reproduce the average potential energy per bead computed in atomistic simulation (see Table 2 and

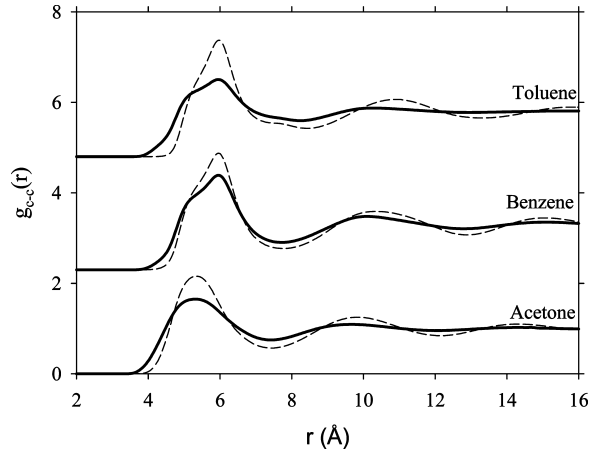

Figure 4. Radial distribution functions (mass-center-to-mass-center) for aprotic liquids computed form CG (dashed lines) and atomistic (solid lines) simulation at 298 K and 1 bar. The interaction potential in the CG simulation is such that the average energies in the CG and atomistic simulations are the same. Note that for the optimized average potential energy the CG liquid is more structured and denser (see Table 2) than the atomistic liquid.

Figures 1 and 2). As expected, the isobaric heat capacities computed in the CG simulations are lower than their atomistic counterparts, as is presented in Table 2. Since the isobaric heat capacity is proportional to the second moment of the probability of states in the N,P,T ensemble, this means that the spherically symmetric effective CG potential washed out some fluctuations, as discussed in the Introduction. Although the studied aprotic molecules are far from the quasi-spherical shape, the total bonding energy profiles for monomers are almost quantitatively reproduced at the CG level. Moreover, all radial distribution functions computed from atomistic simulations are reproduced by the CG simulations with reasonable accuracy, as is presented in Figure 4 (see also the comparison between atomistic and CG snapshots of liquid benzene in the Supporting Information). Taking into account the agreement between the total bonding energy profiles for aprotic monomers displayed in Figure 3, one can conclude that the reduction of the isobaric heat capacity at the CG level results from the nonspherical shape of an atomistic aprotic molecule.

On the basis of the results summarized in Table 2 we conclude that the lack of directional interactions as well as their strength in comparison to thermal energy is the source of the observed reasonable agreement between structure, density, and heat of vaporization. We stress once more that our results rely on the particular CG procedure described in section II. Since there may exist better constructions of the CG potential, still better agreement with atomistic simulations could be obtained. This makes the prospects of modeling the aprotic molecules by the spherically symmetric effective potentials even better.

Now we turn to the studied protic molecules. The effective CG potentials for methanol and acetamide are compared in Figure 5, and in Figure 6, the effective CG potential for ethanol is shown. The effective CG potential for protic molecules is generally more complex in comparison to aprotic molecules. As previously, we found that the trial effective potential obtained from simulations underestimates the average potential energy per bead measured directly in atomistic simulation. Thus, the adjustment of the attractive well depth is necessary if one wishes to preserve the average potential energy of interaction per bead in atomistic/CG simulation, as is displayed in Figures 5 and 6. It is obvious that, for the protic molecules, the complex shape of the effective CG potential results from directional short-range

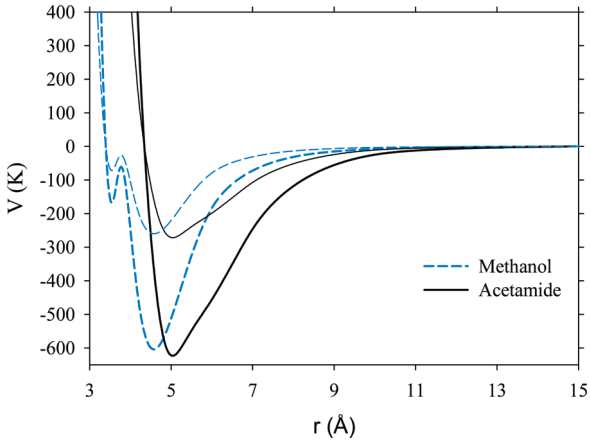

Figure 5. Effective CG (bold lines) and trial effective potential (thin lines) computed for methanol and acetamide at 298 K and 1 bar. The effective sizes of the methanol and the acetamide beads are 3.39 and 4.34 Å, respectively (i.e., we define the effective size of the spherical bead as a value for which the effective potential is zero).

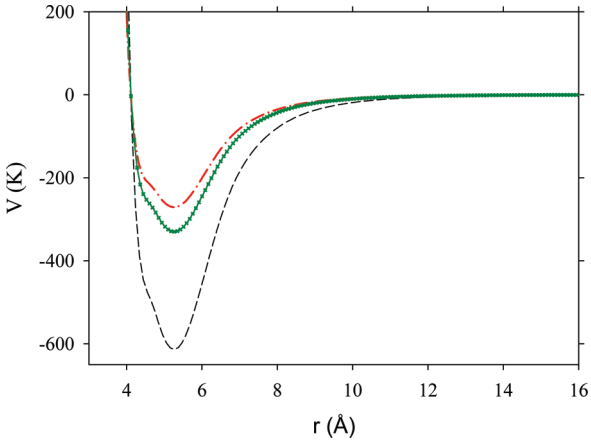

Figure 6. Effective CG potentials computed for ethanol liquid. The red line corresponds to rescaling constant $C = 1.0$, the green line corresponds to $C = 1.218\,54$, and the black line corresponds to $C = 2.265\,495$. The effective size of the ethanol bead is 4.105 Å (i.e., we define the effective size of the spherical bead as a value for which the effective potential is zero).

interactions. Let us first consider the total interaction energy profiles for protic monomers displayed in Figure 7. Although our CG model predicts the position of the maxima of the protic liquids with reasonable accuracy, their dispersion is much lower in comparison to atomistic simulation. This observation can be understood in terms of differences in fluctuations of the two molecular models. Simply, in every case, the spherically symmetric effective CG potential is not able to reproduce a broad spectrum of fluctuations that are present in highly anisotropic atomistic systems of protic liquids. The reader perhaps noticed that spherical beads tend to pack better then atomistic molecules. This results in higher density as well as structure of the CG liquids in comparison to atomistic ones, as is presented in Table 3. Note the disagreement between the radial distribution functions computed from the atomistic and CG simulations (see Figure 8). The rotational motion of protic molecules in atomistic simulation broadens the correlation functions. Thus, it is not surprising that the imposed spherical symmetry of beads and equal average potential energy per molecule result in denser and more structured CG liquids, as is displayed in Figures 8 and 9. Obviously, one can improve a description of all radial distribution functions of protic liquids by reduction of the effective CG potential. This however results

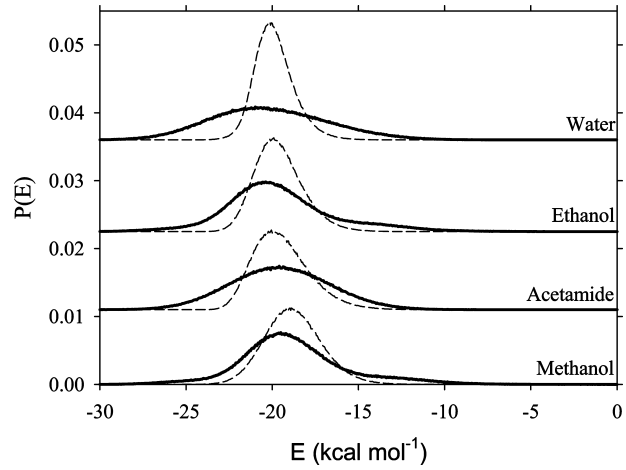

Figure 7. Distributions of total bonding energies for monomers in protic liquids computed from CG (dashed lines) and atomistic (solid lines) simulation at 298 K and 1 bar. The interaction potential in the CG simulation is such that the average energies in the CG and atomistic simulations are the same. Note that the most probable energies are higher (except from acetamide), and the distribution is much narrower in the case of the CG simulation compared to the atomistic simulation.

TABLE 3: Comparison of the Thermodynamic Properties of Protic Molecules Computed from the CG and Atomistic Simulations[a]

molecule	property	atomistic simulation	coarse-grained simulation
methanol	E_{pot} (kcal/mol)	-9.39 ± 0.02	-9.38 ± 0.003
	ΔH_{vap} (kcal/mol)	9.98 ± 0.02	$9,97 \pm 0.003$
	C_p (cal/(mol K))	17.11 ± 1.8	13.04 ± 0.52
	d (g/cm^3)	0.67 ± 0.01	1.27 ± 0.0007
acetamide	E_{pot} (kcal/mol)	-9.76 ± 0.02	-9.72 ± 0.003
	ΔH_{vap} (kcal/mol)	10.36 ± 0.02	10.31 ± 0.003
	C_p (cal/(mol K))	23.2 ± 3.9	11.98 ± 0.2
	d (g/cm^3)	0.93 ± 0.003	1.13 ± 0.0003
ethanol	E_{pot} (kcal/mol)	-9.79 ± 0.03	-9.8 ± 0.002
	ΔH_{vap} (kcal/mol)	10.38 ± 0.03	10.39 ± 0.002
	C_p (cal/(mol K))	21.31 ± 6.2	10.63 ± 0.15
	d (g/cm^3)	0.91 ± 0.01	1.35 ± 0.0005
water	E_{pot} (kcal/mol)	-10.06 ± 0.02	-9.93 ± 0.04
	ΔH_{vap} (kcal/mol)	10.65 ± 0.02	10.52 ± 0.04
	C_p (cal/(mol K))	20.46 ± 1.5	-
	d (g/cm^3)	1.0 ± 0.004	1.59 ± 0.004

[a] The abbreviations are as in Table 2.

in incorrect energetic/cohesive properties measured in CG simulations. Comparison of the radial distribution function and the distribution of the total bonding energy for monomer computed from the effective CG potentials for the values of the rescaling constant: $C = 1.0$, $1.218\,54$, and $2.265\,495$ with the results of atomistic simulations is given in Figures 9 and 10 for ethanol. As expected (see Appendix A), the radial distribution function is quite well reproduced for $C = 1.0$ (for separations $r > 1.5\sigma$, the agreement is very good); the energy, however, deviates very strongly from the results of atomistic simulation. In contrast, for $C = 2.265\,495$, the bonding energy for ethanol monomer is reproduced with a reasonable accuracy, but the structure is overestimated.

Correct description of the energetic properties could be obtained by the CG potential $\Phi(r)$ equal to the interaction potential between the two molecules averaged over their orientations and over orientations and positions of the remaining molecules (see B5 in Appendix B). On the other hand, the structure is quite well reproduced by the CG potential associated

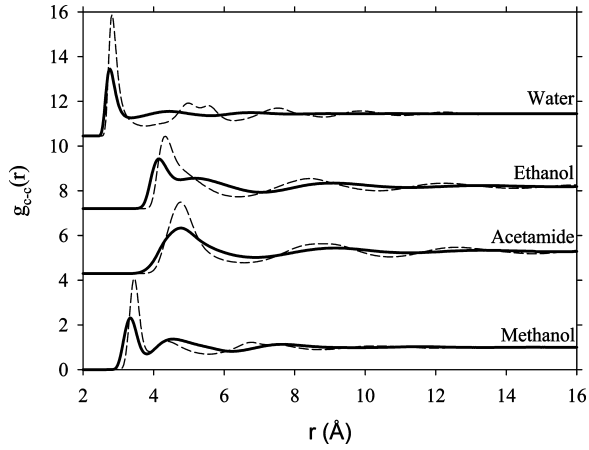

Figure 8. Radial distribution functions (mass-center-to-mass -center) for protic liquids computed from CG (dashed lines) and atomistic (solid lines) simulation at 298 K and 1 bar. The interaction potential in the CG simulation is such that the average energies in the CG and atomistic simulations are the same. Note that for optimized average potential energy the CG liquid is much more structured and denser (see Table 3) than the atomistic liquid.

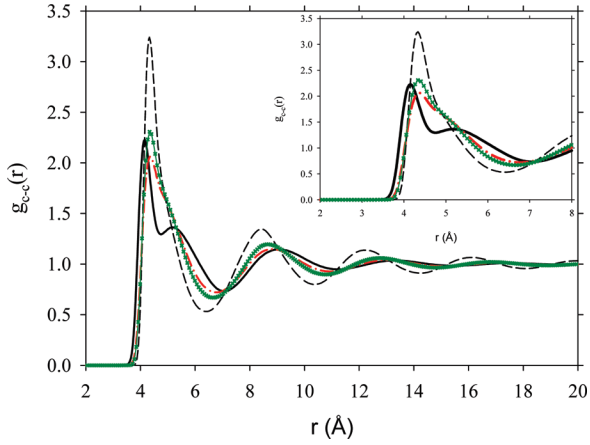

Figure 9. Comparison of the radial distribution functions computed for atomistic liquid ethanol (bold solid lines) and three effective CG potentials (line styles as in Figure 6). Note that effective CG potential with $C = 1.0$ and 1.218 54 correctly reproduces the structure of the atomistic liquid ethanol in CG simulations. In contrast, effective CG potential with $C = 2.265\ 495$ (i.e., optimized to reproduce the average potential energy per bead) overestimates the structure in CG simulations (see Appendix A).

with the force between the two molecules averaged over their orientations and over orientations and positions of the remaining molecules (see section 2.1 and Appendix A). Poor representation of energetic properties by the potential $\Psi(r)$ shows that for ethanol (and the remaining protic liquids) the potentials $\Phi(r)$ and $\Psi(r)$ are significantly different. The difference between the average force $-\nabla_1\Psi$ and the force $-\nabla_1\Phi$ associated with the average potential is equal to the correlation function between the interaction potential of a pair of molecules and total force acting on one of the two molecules in the pair (see B8). From the result (B8) of Appendix B, it follows that a simultaneous precise description of the structural and energetic properties by spherically symmetric effective potentials is not possible if there are strong correlations between the interaction potential of a pair of molecules and total force acting on one of the two molecules in the pair. Such correlations are expected for highly directional interactions, present in protic liquids.

We stress again that our results rely on the assumed form of the CG potential, and performance of the optimized CG could be better.

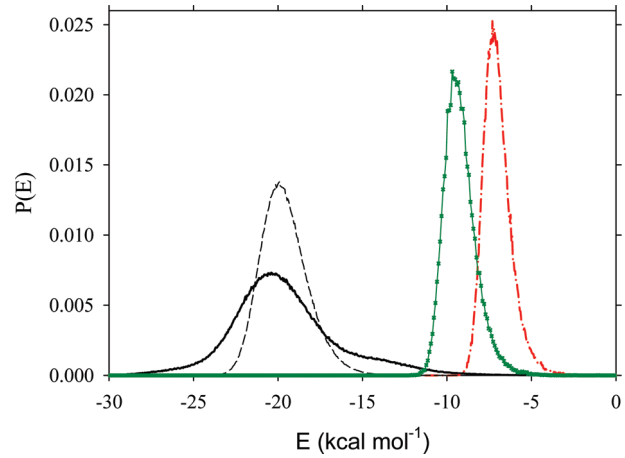

Figure 10. Comparison of total bonding energies for monomers computed for atomistic liquid ethanol (bold solid line) and three effective CG potentials (line styles as in Figure 6). Note that effective CG potential with $C = 1.0$ and 1.218 54 underestimates the total bonding energies for monomers as well as washes out some fluctuations (see Appendix B).

In any case, however, it seems that simultaneous precise description of the structural and energetic properties of the protic liquids by spherically symmetric effective potentials is not possible. Thus we want to emphasize that for protic liquids (or more generally clusters of atoms whose interactions in atomistic simulation are dominated by short-range highly directional ones) some microscopic information has to be lost. Which information? This question is related to the purpose of the CG. For example, if one wishes to preserve a correlations/structure on the CG level (i.e., assume $C = 1.0$), the interaction energy per bead in the CG simulation will be reduced in comparison to atomistic simulation. One can also argue that in the best CG method the deviations of all quantities of interests computed in the CG simulation from the corresponding quantities measured in atomistic systems should be minimized. Within this method, this can be achieved by choosing some intermediate value of the parameter C, between $C = 1.0$ and the value give in Table 1 (see the results displayed in Table 4 and in Figures 9 and 10). One thing is certain from our study, namely that even for a highly anisotropic atomistic system the researcher can fully control and design the CG method. On the other hand, for aprotic liquids (or more generally clusters of atoms whose interactions in atomistic simulation are not specific in comparison to thermal energy) the proposed CG method reproduces satisfactorily fundamental structural and thermodynamic properties, excluding isobaric heat capacity. Taking into account our simulation results, one can expect that many synthetic polymers (such as for instance polyethylene), *n*-alkanes, noble gases, and simple molecules which interact via weak electrostatic forces can be easily CG using the proposed method. On the other hand, polysaccharides and other macromolecules that are characterized by short-range directional interactions between clusters of atoms are more difficult to CG and some compromise between the accuracy and computational cost has to be made. Further extension of our CG method for more complex systems will be a subject of future works.

IV. Conclusions

We study the role of short-range directional interactions in CG of protic (i.e., acetamide, methanol, ethanol, and water) and aprotic (i.e., acetone, benzene, and toluene) liquids at normal conditions. With the new CG method implemented in the *N*,*P*,*T* ensemble, we obtain and compare with atomistic simulations several structural/thermodynamic properties including: heat of vaporization, density,

TABLE 4: Comparison of the Thermodynamic Properties of Ethanol Computed from Atomistic Simulations and Selected CG Effective Potentials[a]

property	atomistic simulation	CG simulation for $C = 2.265495$	CG simulation for $C = 1.0$	CG simulation for $C = 1.21854$
E_{pot} (kcal/mol)	-9.79 ± 0.03	-9.8 ± 0.002	-3.5 ± 0.001	-4.6 ± 0.001
ΔH_{vap} (kcal/mol)	10.38 ± 0.03	10.39 ± 0.002	4.09 ± 0.001	5.2 ± 0.001
C_p (cal/(mol K))	21.31 ± 6.2	10.63 ± 0.15	11.24 ± 0.11	11.27 ± 0.26
d (g/cm^3)	0.91 ± 0.01	1.35 ± 0.0005	1.07 ± 0.0002	1.16 ± 0.0003

[a] The abbreviations are as in Table 2. Note that simple rescaling of the trial effective potential (see eqs 4 and 5) allows full control of thermodynamic properties computed in CG simulations.

radial distribution function, and total bonding energy profile for monomer. For aprotic liquids, the spherically symmetric CG effective potentials correctly reproduce the properties mentioned above, except for the isobaric heat capacity. For protic liquids, the spherically symmetric effective CG potential produces more structure, enhanced packing of beads and underestimated isobaric heat capacity of the CG liquids. We explain this fundamental difference between protic and aprotic liquids by taking into account the presence of the short-range directional interactions in atomistic protic liquids. Although some information from atomistic protic liquids has to be lost on the CG level, the understating of the action of short-range directional interactions between protic molecules seems to be the key for their effective CG. Balance between the average potential energy per bead (i.e., strength of attractive potential) and correlations in liquid (i.e., radial distribution function) allows controllable CG of protic liquids. Aprotic beads characteristic for many polymers or *n*-alkanes seem to be easily CG by the proposed method. We should mention that all our results are based on the simple method of optimization of the interaction potential in the CG simulations which is not unique and other relations between the function $U_c(r)$ and the trial effective potential could yield better results for the structure and density. For example, one could assume $U_c(r) = \Psi(r) - C$ for $r < r_0$ and $U_c(r) = 0$ for C_p (cal/(mol K)), where $\Psi(r) = 0$ for $r > r_0$. C is a constant to be determined in the way described in section II.1. Different methods of determining the trial potential are also possible.

Acknowledgment. The authors acknowledge the use of the computer cluster VPAC (Victorian partnership for advanced computing). P.K. acknowledges the Royal Melbourne Institute of Technology for a postdoctoral fellowship (Academic Level B, 2008−2010). P.A.G. acknowledges the use of the computer cluster at Poznan Supercomputing and Networking Centre as well as the Information and Communication Technology Centre of the Nicolaus Copernicus University (Torun, Poland). A.C. acknowledges partial support by the Polish Ministry of Science and Higher Education, Grant No. NN 202 006034. The authors also acknowledge an anonymous reviewer for helpful comments and fruitful discussions. The project was partially supported by grant NN 204 288634 (2008−2011).

Appendix A: Construction of the Effective Potential

We proceed as in ref 55. Let us consider the probability density $p_N(\mathbf{r}_1, ..., \mathbf{r}_N)$ of finding the centers of mass of N molecules in the positions $(\mathbf{r}_1, ..., \mathbf{r}_N)$ regardless of their orientations

$$p_N(\mathbf{r}_1, ..., \mathbf{r}_N) \equiv \frac{e^{-\beta[U_e(\mathbf{r}_1,...,\mathbf{r}_N)]}}{Z} = \int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N \frac{e^{-\beta[U(\mathbf{r}_1,\boldsymbol{\omega}_1;...;\mathbf{r}_N,\boldsymbol{\omega}_N)]}}{Z} \quad \text{(A1)}$$

where

$$U(\mathbf{r}_1, \boldsymbol{\omega}_1;...;\mathbf{r}_N, \boldsymbol{\omega}_N) = \sum_{i<j} U_a(||\mathbf{r}_i - \mathbf{r}_j||, \boldsymbol{\omega}_i, \boldsymbol{\omega}_j) \quad \text{(A2)}$$

and the notation is introduced in section II.1. The exact pair distribution function is given by

$$g(r_{1,2}) = \frac{V^2 \int d\mathbf{r}_3 ... \int d\mathbf{r}_N e^{-\beta U_e}}{Z} \quad \text{(A3)}$$

In (A1), we introduce the effective energy $U_e(\mathbf{r}_1, ..., \mathbf{r}_N)$, which in general is not pairwise additive and difficult to calculate. In order to find an approximate pairwise-additive effective potential that reproduces the structure as given by the pair distribution function, we make the assumption that $U_e(\mathbf{r}_1, ..., \mathbf{r}_N)$ can be written in the form

$$U_e(\mathbf{r}_1, ..., \mathbf{r}_N) = u_e(\mathbf{r}_1, ..., \mathbf{r}_N) \sum_{i<j} \Psi(r_{i,j}) \quad \text{(A4)}$$

where $r_{i,j} = ||\mathbf{r}_i - \mathbf{r}_j||$ and we make the Ansatz that the deviation from the pair additivity is small, i.e.

$$u_e(\mathbf{r}_1, ..., \mathbf{r}_N) \approx 1 \quad \text{and} \quad \Psi(r_{1,j})\nabla_1 u_e(r_1, ..., r_N) \ll \nabla_1 \Psi(r_{1,j}) \quad \text{(A5)}$$

Assumptions A5 are not generally valid and difficult to justify in particular cases; they can be verified by comparison of the results of the approximate and the exact theories.

The above equations do not define the functions $u_e(\mathbf{r}_1, ..., \mathbf{r}_N)$ and $\Psi(r_{i,j})$ in a unique way. The optimal choice of $\Psi(r_{i,j})$ should yield the best approximation for the distribution functions.

For a particular choice of the functions $u_e(\mathbf{r}_1, ..., \mathbf{r}_N)$ and $\Psi(r_{i,j})$ that satisfy (A5), we substitute (A4) for $U_e(\mathbf{r}_1, ..., \mathbf{r}_N)$ and (A2) for $U(\mathbf{r}_1,\boldsymbol{\omega}_1; ...; \mathbf{r}_N,\boldsymbol{\omega}_N)$ in (A1), take the logarithm of both sides of the resulting equation, and then differentiate with respect to $\mathbf{r}_1$. The result is

$$\sum_{j>1} u_e \nabla_1 \Psi(r_{1,j}) + \text{corr} = \sum_{j>1} \frac{\int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N e^{-\beta U} \nabla_1 U_a(r_{1,j}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_j)}{\int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N e^{-\beta U}} \quad \text{(A6)}$$

where the derivative of $u_e(\mathbf{r}_1, ..., \mathbf{r}_N)$ is included in the correction term. We neglect this term according to assumption A5.

Each term in the sum on the left-hand side (LHS) of the above equation has the same form as a function of its argument, and each term in the sum on the right-hand side (RHS) of the above equation has the same structure. We thus assume equality between each term in the sum on the LHS of (A6) and the corresponding term in the sum on the RHS. For $j = 2$, we obtain

$$u_e \nabla_1 \Psi(r_{1,2}) \approx \frac{\int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N e^{-\beta U} \nabla_1 U_a(r_{1,2}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2)}{\int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N e^{-\beta U}} \tag{A7}$$

we multiply both sides of the resulting equation by the denominator of the RHS and, next, integrate both sides with respect to the positions of the centers of mass of the remaining molecules

$$\nabla_1 \Psi(r_{1,2}) \int d\mathbf{r}_3 ... \int d\mathbf{r}_N u_e \int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N e^{-\beta U} \approx \int d\mathbf{r}_3 ... \int d\mathbf{r}_N \int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N e^{-\beta U} \nabla_1 U_a(r_{1,2}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \tag{A8}$$

Using again assumption A5 in (A8), we obtain the final approximation

$$\nabla_1 \Psi(r_{1,2}) \approx \langle \nabla_1 U_a(r_{1,2}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \rangle_{r_{1,2}} \tag{A9}$$

where we introduced the notation

$$\langle X \rangle_{r_{1,2}} = \frac{\int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N \int d\mathbf{r}_3 ... \int d\mathbf{r}_N X e^{-\beta U}}{\int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N \int d\mathbf{r}_3 ... \int d\mathbf{r}_N e^{-\beta U}} \tag{A10}$$

Note that up to the minus sign the RHS in eq A9 is the force between molecules 1 and 2, averaged over their orientations, and over the positions and orientations of the remaining molecules. The potential Ψ of this orientation-averaged force should yield the optimised pairwise-additive approximation for the exact many-body effective potential (see A4). Since the latter yields an exact pair-distribution function, Ψ is the effective pair potential optimized in order to predict the correct structure. When the effective pairwise interaction is assumed, the corrections associated with nonadditivity of the effective orientation-independent interactions (i.e. with $u_e(\mathbf{r}_1, ..., \mathbf{r}_N) \neq 1$) are disregarded. The deviations between the predictions of the effective and the exact theories will occur in the case of strong many-body correlations, i.e. when the probability of finding the centres of mass of three molecules at $(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$, $p_3(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3)$, is significantly different from $p_2(\mathbf{r}_1, \mathbf{r}_2)p_2(\mathbf{r}_1, \mathbf{r}_3)p_2(\mathbf{r}_3, \mathbf{r}_2)$ where $p_2(\mathbf{r}_1, \mathbf{r}_2)$ is the probability of finding the centers of mass of two molecules at $(\mathbf{r}_1, \mathbf{r}_2)$. Such correlations are expected for short distances, and in such cases, the contributions to $U_e(\mathbf{r}_1, ..., \mathbf{r}_N)$ due to $u_e(\mathbf{r}_1, ..., \mathbf{r}_N) \neq 1$ cannot be neglected. The spherically symmetric CG potentials cannot reproduce the structure resulting from strong correlations between positional and rotational degrees of freedom.

Appendix B: Comparison between Average Energy per Particle in the Exact and CG Cases

The average energy per particle is given by[56]

$$\bar{u} = \frac{\bar{U}}{N} = \frac{\int dr_2 ... dr_N \int d\omega_1 \int d\omega_N U_a(r_{1,2}, \omega_1, \omega_2) e^{-\beta U} \rho V^2}{Z} \tag{B1}$$

After multiplying both the numerator and denominator of (B1) by

$$\int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_N \int d\mathbf{r}_3 ... \int d\mathbf{r}_N e^{-\beta U} \tag{B2}$$

we obtain

$$\bar{u} = \int d\mathbf{r}_2 \Phi(r_{1,2}) \rho \bar{g}(r_{1,2}) \tag{B3}$$

where

$$\bar{g}(r_{1,2}) = \frac{V^2 \int d\boldsymbol{\omega}_1 ... \int d\boldsymbol{\omega}_2 \int d\mathbf{r}_3 ... \int d\mathbf{r}_N e^{-\beta U}}{Z} \tag{B4}$$

is the radial distribution function, and

$$\Phi(r_{1,2}) = \langle U_a(r_{1,2}, \boldsymbol{\omega}_1, \boldsymbol{\omega}_2) \rangle_{r_{1,2}} \tag{B5}$$

is the pair potential between molecules 1 and 2 separated by the distance $r_{1,2}$, averaged over their orientations and over orientations and positions of the remaining molecules (see (A10)).

In the CG case, the average energy per molecule is given by

$$\bar{u}_{CG} = \int d\mathbf{r}_2 \Psi(r_{1,2}) \rho g_{CG}(r_{1,2}) \tag{B6}$$

where Ψ is the potential of the orientation-averaged force (see (A9)), and,

$$g_{CG}(r_{1,2}) = \frac{V^2 \int dr_3 ... dr_N e^{-\beta \sum_{i<j} \Psi(r_{ij})}}{Z} \approx \bar{g}(r_{1,2}) \tag{B7}$$

where the last approximate equality in (B7) holds when (A5) is valid. The difference between $\bar{u}$ and $\bar{u}_{CG}$ results from the difference between Φ and Ψ. It is easier to compare gradients of the above two functions, where the gradient of the latter is given in (A9), and from (B5) and (A10), we obtain

$$\nabla_1 \Phi = \nabla_1 \Psi - \beta \sum_{j>1} \left[\langle U_a(r_{1,2}, \omega_1, \omega_2) \nabla_1 U_a(r_{1,2}, \omega_1, \omega_j) \rangle_{r_{1,2}} - ... \langle U_a(r_{1,2}, \omega_1, \omega_2) \rangle_{r_{1,2}} \langle \nabla_1 U_a(r_{1,2}, \omega_1, \omega_j) \rangle_{r_{1,2}} \right] \tag{B8}$$

Equation B8 shows that if there are correlations between the potential of a pair of molecules, $U_a(r_{1,2}, \omega_1, \omega_2)$ and the total force acting on one of the molecules in the pair, the average force $-\nabla_1\Psi$ is different from the force $-\nabla_1\Phi$ associated with the average potential. From the fact that $\Phi \neq \Psi$ and from eqs B3 and B6, it follows that the exact and CG average energies per molecule are different. Another source of disagreement between energies per molecule in the atomistic and CG systems results from the difference between $\bar{g}$ and g_{CG}, but this difference should be of comparable order as the difference between the structural properties.

Supporting Information Available: Figures 1S−7S and movie file discussed in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Voth, G. A. In *Coarse-graining of condensed phase and biomolecular systems*; Voth, G. A., Ed.; CRC Press: Boca Raton, 2008.
(2) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812.
(3) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469.
(4) Izvekov, S.; Parrinello, M.; Burnham, C. J.; Voth, G. A. *J. Chem. Phys.* **2004**, *120*, 10896.
(5) Hooper, J. B.; Bedrov, D.; Smith, G. D. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2034.
(6) Reith, D.; Meyer, H.; Müller-Plathe, F. *Comput. Phys. Commun.* **2002**, *148*, 299.
(7) Reith, D.; Pütz, M.; Müller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.
(8) Lyubartsev, A. P *Eur. Biophys. J.* **2005**, *35*, 53.
(9) Lyubartsev, A. P.; Karttunen, M.; Vattulainen, I.; Laaksonen, A. *Soft Mater.* **2003**, *1*, 121.
(10) Wang, Y.; Noid, W. G.; Liu, P.; Voth, G. A. *Phys. Chem. Chem. Phys.* **2009**, *11*, 2002.
(11) Tòth, G. *J. Phys.: Condens. Matter* **2007**, *19*, 335220.
(12) Nielsen, S. O.; Ensing, B.; Ortiz, V.; Moore, P. B.; Klein, M. L *Biophys. J.* **2005**, *88*, 3822.
(13) Saiz, L.; Klein, M. L. *Acc. Chem. Res.* **2002**, *35*, 482.
(14) de Vries, A. H.; Mark, A. E.; Marrink, S. J. *J. Am. Chem. Soc.* **2004**, *126*, 4488.
(15) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750.
(16) Srinivas, G.; Shelley, J. C.; Nielsen, S. O.; Discher, D. E.; Klein, M. L. *J. Phys. Chem. B* **2004**, *108*, 24.
(17) Bandyopadhyay, S.; Tarek, M.; Klein, M. L. *J. Phys. Chem. B* **1999**, *103*, 10075.
(18) Nielsen, S. O.; Lopez, C. F.; Moore, P. B.; Shelley, J. C.; Klein, M. L. *J. Phys. Chem. B* **2003**, *107*, 13911.
(19) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464.
(20) Müller-Plathe, F. *Soft Mater.* **2003**, *1*, 1.
(21) Müller-Plathe, F. *Chem. Phys. Chem.* **2002**, *3*, 754.
(22) Faller, R.; Schmitz, H.; Biermann, O.; Müller-Plathe, F. *J. Comput. Chem.* **1999**, *20*, 1009.
(23) Santangelo, G.; Di Matteo, A.; Müller-Plathe, F.; Milano, G. *J. Phys. Chem. B* **2007**, *111*, 2765.
(24) Milano, G.; Müller-Plathe, F. *J. Phys. Chem. B* **2005**, *109*, 18609.
(25) Spyriouni, T.; Tzoumanekas, Ch.; Theodorou, D.; Müller-Plathe, F.; Milano, G. *Macromolecules* **2007**, *40*, 3876.
(26) Chaimovich, A.; Shell, M. S. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1901.
(27) Kamio, K.; Moorthi, K.; Theodorou, D. N. *Macromolecules* **2007**, *40*, 710.
(28) Baron, R.; Trzesniak, D.; de Vries, A. H.; Elsener, A.; Marrink, S. J.; van Gunsteren, W. F. *Chem. Phys. Chem.* **2007**, *8*, 452.
(29) Ungerer, P.; Beauvais, C.; Delhommelle, J.; Boutin, A.; Rousseau, B.; Fuchs, A. H. *J. Chem. Phys.* **2000**, *112*, 5499.
(30) Bourasseau, E.; Ungerer, P.; Boutin, A.; Fuchs, A. H. *Mol. Simul.* **2002**, *28*, 317.
(31) Mognetti, B. M.; Oettel, M.; Yelash, L.; Virnau, P.; Paul, W.; Binder, K. *Phys. Rev. E* **2008**, *77*, 041506.
(32) Mognetti, B. M.; Virnau, P.; Yelash, L.; Paul, W.; Binder, K.; Müller, M.; MacDowell, L. G. *J. Chem. Phys.* **2009**, *130*, 044101.
(33) Mognetti, B. M.; Yelash, L.; Virnau, P.; Paul, W.; Binder, K.; Müller, M.; MacDowell, L. G. *J. Chem. Phys.* **2008**, *128*, 104501.
(34) Stell, G.; Rasaiah, J. C.; Narang, H. *Mol. Phys.* **1974**, *27*, 1393.
(35) Johnson, M. E.; Head-Gordon, T.; Louis, A. A *J. Chem. Phys.* **2007**, *126*, 144509.
(36) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.
(37) Allen, E. C.; Rutledge, G. C. *J. Chem. Phys.* **2008**, *128*, 154115.
(38) Chen, L. J.; Qian, H. J.; Lu, Z. Y.; Li, Z. S.; Sun, Ch. Ch. *J. Phys. Chem. B* **2006**, *110*, 24093.
(39) Izvekov, S.; Voth, G. A. *J. Chem. Theory Comput.* **2006**, *2*, 637.
(40) Onsager, L. *Ann. N.Y. Acad. Sci.* **1949**, *51*, 627.
(41) Chandler, D.; Weeks, J. D.; Andersen, H. C. *Science* **1983**, *220*, 787.
(42) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237.
(43) Chandler, D. *Annu. Rev. Phys. Chem.* **1978**, *29*, 441.
(44) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon: Oxford, 1987.
(45) Frenkel, D.; Smit, B. *Understanding Molecular Simulation From Algorithms To Applications*; Academic Press: London, 1996.
(46) Ercolessi, F.; Adams, J. B. *Europhys. Lett.* **1994**, *26*, 583.
(47) Ercolessi, F.; Parrinello, M.; Tosatti, E. *Philos. Mag. A* **1988**, *58*, 213.
(48) Michalewicz, Z. *Genetic Algorithms + Data Structures = Evolution Programs*; Springer-Verlag: Berlin, 1995.
(49) Stone, A. J. *Science* **2008**, *321*, 787.
(50) Ungerer, Ph.; Tavitian, B.; Boutin, A. *Applications of molecular simulation in the oil and gas industry. Monte Carlo methods*; IFP publications: Paris, 2005.
(51) Wick, C. D.; Siepmann, J. I.; Klotz, W. L.; Schureb, M. R. *J. Chromatogr. A* **2002**, *954*, 181.
(52) Jorgensen, W. L.; Jenson, C. *J. Comput. Chem.* **1998**, *19*, 1179.
(53) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
(54) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(55) Chu, J. W.; Ayton, G. S.; Izvekov, S.; Voth, G. A. *Mol. Phys.* **2007**, *105*, 167.
(56) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover Publications, Inc.: New York, 1960.

JP9038368