PCCP

Towards a Coarse-Grained Model of the Peptoid Backbone: The Case of N,N- dimethylacetamide

Journal:	*Physical Chemistry Chemical Physics*
Manuscript ID	CP-ART-05-2018-003283.R1
Article Type:	Paper
Date Submitted by the Author:	13-Aug-2018
Complete List of Authors:	Du, Pu; Louisiana State University, Chemistry Rick, Steven W; University of New Orleans, Department of Chemistry Kumar, Revati; Louisiana State University, Chemistry

Towards a Coarse-Grained Model of the Peptoid Backbone: The Case of N, N - dimethylacetamide

Pu Du[+], Steven W. Rick[‡] and Revati Kumar[+,*]

[+]Department of Chemistry, Louisiana State University,

Baton Rouge, Louisiana 70803, United States

[‡]Department of Chemistry, University of New Orleans,

New Orleans, Louisiana 70148, United States

Abstract

In this study, a coarse-grained (CG) model for N, N-dimethylacetamide (DMA), which represents the polypeptoid backbone, is developed as a step towards establishing a CG model of the complex polypeptoid system. Polypeptoids or poly N-substituted glycines are a type of peptidomimetic polymers that are highly tunable, and hence an ideal model system to study self-assembly as a function of chemical groups in aqueous soft matter systems. The DMA CG model is parameterized to reproduce the structural properties of DMA liquid as well as a dilute aqueous solution of DMA using a reference all atom model, namely the OPLS-AA force-field. The intermolecular forces are represented by the Stillinger-Weber potential, that consists of both two- and three-body terms that are very short-ranged. The model is validated on thermodynamic properties of liquid and aqueous DMA, as well as the vapor-liquid interface of liquid DMA and the structure of a concentrated aqueous solution of DMA in water as well as a simple peptoid in water. Without long-ranged interactions and the absence of interaction sites on hydrogen atoms, the CG DMA model is an order of magnitude faster than the higher resolution all-atom (AA) model.

Corresponding author: revatik@lsu.edu

I. Introduction

Self-assembly of block copolymers in solution is relevant to a number of applications including biomedical uses.[1, 2] Understanding the effect of secondary/non-covalent interactions on self–assembly can be challenging given the complexity in macromolecular systems. Polypeptoids, a class of highly tunable biomimetic analogues of peptides, are an ideal prototypical model system to study self-assembly such as micelle formation in aqueous environments.[3-5] In addition, polypeptoids themselves have many potential applications including in the field of drug delivery.[6, 7] A number of block copolypeptoids have been synthesized and their aggregation behavior in liquid solution have been studied by experiments and computational investigations in recent years.[8-12] Atomistic (AA) molecular dynamics simulations have also been used to study solvation and self-assembly of peptoid systems. For example, the atomic-resolution structure of bilayer peptoid nanosheets was determined by molecular simulation with using CHARMM based atomistic force field developed by Mirijanian et al.[13] Prakash et al. used molecular simulations to compare the backbone flexibility of a model peptoid with a corresponding peptide in water and at surfaces.[14] The OPLS[15, 16] based force field has previously been used to model peptoid systems, including the studies of structural and dynamical characteristics of peptoid oligomers and mechanism of aggregation of cyclic polypeptoids in methanol solution.[17] However, the computational expense limits simulation times and system size to a few hundred nanoseconds and nanometers, respectively. Although the use of enhanced sampling algorithms can aid the sampling of configuration space, the dependence of these algorithm to a few select variables such as temperature, geometric collective

variables, can often pose a challenge to sampling rare events from atomistic simulations.[18-20] Coarse-grained (CG) models, on the other hand, are computationally cheaper alternatives wherein atomistic level details are removed while retaining the relevant important physics, thus making it easier to study more complicated systems, including self-assembly in soft matter.[21-23] By reducing some degrees of freedom and only representing the most important ones, CG modeling can span scales efficiently.[24-26]

Coarse-graining methodologies,[27, 28] especially in the area of biological applications,[29] have become increasingly important in recent times starting from the seminal work of Warshel and Levitt.[30] As already mentioned this typically involves a reduction in the number of degrees of freedom resulting in a lower resolution model that is significantly more efficient but nonetheless incorporates the relevant physics of the problem under study. At the very least, light atoms such as hydrogen are folded into the nearest heavy atoms, the so-called united atom approach. Further coarse-graining combines groups of heavy atoms to form pseudo-atoms, as is the case in the MARTINI force-field.[31, 32] The parameterization of the force-field can follow a number of different paths and can use experimental data or reference higher-resolution simulation data or a combination of both. The method adopted here is a so-called "physics-based force-field" wherein one retains the functional forms used in typical all atom simulations such as bonded and non-bonded interactions.[24] The higher resolution all-atom simulations provide data in the training set that can be used to parameterize the model using iterative Boltzmann,[33] force-matching,[34-36] or Newton inversion methods,[37] to name just a few.

To develop a coarse-grained model of complex peptoid systems, a bottom-up approach is being used for parameterizing CG models to accurately reproduce structural properties and is validated on thermodynamic properties of AA models. The typical polypeptoids used in micelle

formation consist of a backbone with hydrophobic hydrocarbons, neutral ether based hydrophilic and charged carboxylate based hydrophilic side chains.[10] The underlying coarse-graining philosophy is to develop CG models of each chemical group. The CG models of the hydrocarbon and ether chains have already been developed.[38, 39]

In this paper, a CG model of a simple amide system, N, N-dimethylacetamide (DMA) is developed as it contains the repeating motif of the peptoid backbone. Hence it presents an important step towards establishing CG models for peptoid systems. A key focus of this work is to reproduce not just the structure of liquid DMA but also the solvation structure of DMA in water. This is particularly important in order to develop peptoid CG models that can accurately describe the solvation properties of the polymer in aqueous solutions and hence enable accurate investigations of self-assembly in an aqueous environment. This is in contrast to a complex anisotropic implicit solvent CG model of peptoids developed by Haxton et al. to study two dimensional peptoid based nanosheets.[40]

It has been recognized that water plays an essential role in determining the structure and properties of peptoid systems.[14, 41] A CG model of water developed by Molinero and co-workers, based only on short-ranged Stillinger-Weber (SW) potential,[42] is used in this work to describe the interactions between DMA and water. The OPLS-UA[15] force field is adopted to describe the intramolecular interactions between DMA molecules with the elimination of hydrogen sites. The focus of this paper is to develop intermolecular parameters for DMA-DMA and DMA-water interactions using a "reduced range" approach that allows the retention of molecular features but the comparatively short-ranged interactions, involving three-body forms in some cases, leads to computational efficiency over conventional all atom models.[38]

In this paper, we develop a coarse-grained model of DMA, laying the groundwork for developing CG models of polypeptoids. This CG model is parameterized on atomistic simulations, using a functional form that is a hybridized approach involving the OPLS-UA force field and the Stillinger-Weber[42] "reduced range" potential form. The parameterization scheme is based on the method used by Kumar and Skinner to develop an all-atom model of liquid water with three-body interactions.[43] It is optimized to reproduce structural and thermodynamic properties of DMA in liquid solution, specifically features of a number of relevant radial distribution functions of pure DMA and dilute DMA aqueous solution and is validated on liquid/vapor properties as well as concentrated aqueous solutions of DMA. The reference data is from simulations using the all atom OPLS-AA force-field.[16] In addition, the computational efficiency of the CG model is discussed and compared to the higher resolution AA models.

The paper is organized in the following manner. The computational methods of the CG model parametrization scheme and the simulation set up are described in Section II. The results are presented and discussed in Section III and the conclusions are outlined in Section IV.

II. Methods

The functional form of the non-bonded interactions used in this study is based on the Stillinger-Weber (SW)[42, 44] formalism that includes the short-ranged two- and three-body terms. The potential energy of the system is given by

$$\mathrm{E} = \sum_i \sum_{j>i} \phi_2(r_{ij}) + \sum_i \sum_{j\neq i} \sum_{k>j} \phi_3(r_{ij}, r_{ik}, \theta_{jik}) \qquad (1)$$

Here the two-body term $\phi_2(r_{ij})$ is given by

$$\phi_2(r_{ij}) = A\epsilon \left[\mathrm{B} \left(\frac{\sigma}{r_{ij}}\right)^p - \left(\frac{\sigma}{r_{ij}}\right)^q \right] \exp\left(\frac{\sigma}{r_{ij} - a\sigma}\right) \qquad (2)$$

and the three-body term $\phi_3(r_{ij}, r_{ik}, \theta_{ijk})$ *by*

$$\phi_3(r_{ij}, r_{ik}, \theta_{jik}) = \lambda\epsilon\left[cos\theta_{jik} - cos\theta_0\right]^2 \exp\left(\frac{\gamma\sigma}{r_{ij}-a\sigma}\right)\exp\left(\frac{\gamma\sigma}{r_{ik}-a\sigma}\right) \quad (3)$$

Both terms are taken to be zero at distances greater than aσ. In this work, as with the original formalism, p is set to 4 and q to 0. The constants A and B are taken to be the same as in the original formalism with *A=7.049556277, B=0.60222455*. The two parameters in the two-body term, namely the energy scaling variable ε and the particle size σ, are part of the parameterization set. The three-body term is characterized by a scaling factor λ and an angular term θ, both of which are included in the parameterization procedure. The angle θ_0 takes into account the preferred orientational ordering of the solvation shell around the central i^{th} atom and in the case of the neat water model, mW, developed by Molinero and co-workers is 109.47°.[44] The term λ scales the three-body angular interaction. Both the three-body and two-body terms are relatively short-ranged, with the range controlled by the parameter *a* (since the potential goes to zero at aσ) and to a lesser extent by γ, and the two are set to the values used in the mW water model, namely a=1.8 and γ=1.2.

The parameterization involves optimizing the set $(\varepsilon, \sigma, \lambda, \theta_o)$ for the different non-bonded interactions based on data from all atom simulations. The details of the all atom simulations and the parameterization scheme is presented below.

All-atom atomistic reference simulations

The reference all atom simulations were performed using the TIP3P[45] water model along with the OPLS-AA force field for the DMA molecules. Two sets of all atom (AA) reference simulations were performed, the first of a dilute solution of DMA in water (0.2 mol/L) and the second of pure liquid DMA. The dilute aqueous simulation of DMA was carried out in a cubic box, with an initial box length of 20 Å, consisting of a single DMA and 524 TIP3P water

molecules. The liquid DMA liquid simulations consisted of 512 DMAs in a cubic box of initial length 40 Å. A third set of simulations of a more concentrated aqueous solution of DMA (0.5 mol/L) were carried out with the AA model. The system consisted of 20 DMA molecules along with 1700 water molecules in a box of length 40 Å. In addition to the above set of simulations, the DMA liquid-vapor system was also simulated. The system consisted of the same number of DMA molecules as above but with a box size of 40 Å x 40 Å x 80 Å, with the liquid-vapor interface perpendicular to the z-axis. Finally, simulations of a simple polypeptoid solution were also performed. The system contained two simple ten-mer polypeptoid chains (see Figure 1 b) and 1700 water molecules in a 40 Å cubic box.

In each case, except as noted, the simulations were carried out within the isobaric isothermal (NPT) ensemble at 300 K and 1 atm with a 1 fs timestep, using the LAMMPS MD package[46]. Each system was equilibrated for 10 ns and followed by a 10 ns production run. The temperature was controlled with a Nose-Hoover thermostat[47] and barostat.[48] Long-ranged Coulomb interactions were calculated using the Ewald summation, specifically the Particle-Particle-Particle-Mesh (PPPM) method with the desired relative error in forces set to 0.0001.[49] The Lennard-Jones cutoff distance was set to 10 Å. For the calculations of the diffusion constant of pure liquid DMA, a 5 ns production run in the NVE ensemble was carried out.

Free energy of solvation of DMA in water was carried out using finite difference thermodynamic integration (FDTI),[50] using a soft core potential to avoid singularities.[51-53] The free energy was calculated using a 10 ns NPT run with 20 integration points, equally spaced between 0 and 1, with a time step of 1 fs using the GROMACS[54] software.

Coarse-grained (CG) model development

Figure 1 a shows atom types and the mapping of the DMA molecule from AA to the CG representation. The interactions between DMA and water are represented by SW potentials. Intramolecular interactions are taken from the force field, OPLS-UA.[15] Initial guesses for the parameters of each type of atom in DMA were selected from the monatomic model of water (mW) and methane, which were previously developed by Molinero and co-workers.[55]

The CG simulations of dilute DMA with mW waters, liquid DMA, concentrated aqueous DMA, polypeptoid solution and the DMA liquid-vapor cases were performed in LAMMPS with a 5 fs time step under the same simulation conditions as the reference AA simulations. The intramolecular terms were the same as the united atom OPLSUA model.[15] The suitability of the 5 fs time step was tested using a 5 ns NVE simulation that showed virtually no energy drift, despite large fluctuations in the energy. The free energy of solvation was carried out using the dilute DMA system, again using the LAMMPS suite of programs. The free energy calculations were done using finite difference thermodynamic integration (FDTI), with a soft core potential to avoid singularities. The free energy was calculated using 100 integration points, equally spaced between 0 and 1, and were run for 5 million steps, with a time step of 5 fs. The subroutines to carry out TI calculations in LAMMPS were taken from the work of Gyawali et al.[38, 56] For the CG peptoid solution simulations, the parameters developed for DMA were used, with the parameters for the CH2 group in the peptoid set to the values obtained for the CH3N methyl group in DMA.

Parameterization of the DMA CG model

The development of the CG model is primarily focused on reproducing the solvation structure of water around the DMA molecules. The energy scales ϵ and λ, particle size σ, and

three-body angle θ_o are the parameters that govern the solvation behavior. One can obtain estimates of θ_o from either the angular distribution in the first solvation shell from AA simulations or from experimental information such as the tetrahedral geometry of water, leaving just the particle size and energy scale for parameterization. The overall parameterization process is outlined in the flowchart shown in Figure 2. An iterative parameterization procedure, described below, is adopted wherein one starts with a zeroth order model that is progressively optimized to reproduce properties in the training set, in this case details of important radial distribution functions, from the all atom reference simulation data.

As mentioned in the introduction, the parameterization scheme is based on the one adopted by Kumar and Skinner[43] in the development of an improved all atom water model and is similar in spirit to the Newton inversion method. In order to illustrate the method, consider a system that consists of just one type of particle that interacts via the SW potential form. In that case, the total intermolecular interaction energy can be expressed as a sum of two-body energies E_{12} and three-body energies E_{123}. Thus, one can write:

$$E_{total} = \omega_1 E_{ij} + \omega_2 E_{ijk}, \qquad (4)$$

where ω_1 and ω_2 are scaling parameters and are initially set to 1, the so-called zeroth order model with a particular set of parameters $(\varepsilon, \sigma, \lambda, \theta_o)$. Based on the scheme by Kumar et. al., ω_i can be determined in a systematic manner.[43] Since, for this simple case, there are two parameters that need to be determined, two observables/properties *(P_1 and P_2)* are chosen to be part of the training set. Each observable/property will in general be a function of both ω_1 and ω_2. Hence the total derivative of each P_j is given by

$$dP_j = \sum_i \frac{\partial P_j}{\partial \omega_i} d\omega_i \qquad (5)$$

The derivatives $\partial P_j/\partial \omega_i$ are calculated numerically using a series of 5 simulations with parameters $(\omega_1, \omega_2), (\omega_1 + \delta\omega_1, \omega_2), (\omega_1 - \delta\omega_1, \omega_2), (\omega_1, \omega_2 + \delta\omega_2), (\omega_1, \omega_2 + \delta\omega_2)$. Based on these numerical derivatives and the difference, ΔP_j, between the value of each P_j from the simulation at (ω_1, ω_2) and the reference/experimental value, the next generation values of (ω_1, ω_2) can be determined by solving the following set of equations for $\Delta\omega_i$:

$$\Delta P_j = \sum_i \frac{\partial P_j}{\partial \omega_i} \Delta\omega_i \qquad (6)$$

The new values of (ω_1, ω_2) now form the first-generation model and the process is repeated until convergence is reached. In the above example, only one type of particle was considered and hence only two scaling parameters, one scaling the three-body and the other the two-body interactions, are needed. The method can be extended to systems with more than one type of particle (as is the case for the DMA systems under study) leading to many two-body and three-body scaling parameters, with the corresponding number of properties in the training set. The training set can be from higher level simulation data, from experiments or from a combination of both and can be extended to more complicated systems with more than one type of particle. Therefore, for a given set of σ values, the energy scale factors for the two- and three-body terms can be determined.

In the current development of the CG model, data from the reference all atom OPLS-AA simulations on liquid DMA and the dilute solution of DMA in water are used. Specifically, the position of the first maximum of relevant radial distribution functions (O atom of water with the different DMA atoms in the dilute aqueous solution as well as intermolecular DMA atoms in the liquid DMA solution) as well as the associated coordination number around each atom type are the properties in the training set.

The initial σ and ε and values for O and C atoms were taken from the work of Molinero et al. while the initial N atom parameters set was taken from the OPLS-UA force-field.[15, 44, 55] The Lorentz-Bertholet mixing rules were used to obtain the remaining σ and ε values. The three-body angle θ was set to 109.47^{o} for all the atoms types, based on previous work on water and methane.[44, 55] The three-body energy scaling parameter λ was initially set to the value for the mW water model. The AA simulations did not show any preferred orientation of the solvation environment for the N, and since the N atom is buried in DMA molecule one does not expect that the three-body term will be significant. Starting with these initial values and the data from the all atom simulations, the two and three-body terms were scaled and the scaling parameters were determined using the iterative algorithm outlined above. First, the DMA-DMA parameters were determined using the iterative algorithm on simulations with pure liquid DMA. Ten interactions were scaled, namely five two-body terms (C-C, O-O, N-N, CH3N-CH3N and CH3N-CH3N) and five three-body terms (O-O-O, CH3N-O-CH3N, O-CH3N-CH3N, O-CH3-CH3, and CH3-O-CH3). For the three-body terms only those involving O with distinct sites were chosen. For the DMA-water interactions, the dilute DMA systems were simulated. In this case five two-body terms (C-O_w, CH3N-O_w, C-O_w, CH3-O_w, and O-O_w) and five three-body terms (O_w-O-O_w, O-O_w-O_w, O_w-CH_3N-O_w, CH3N-O_w-O_w, and O_w-C-O_w) were considered in the parameterization scheme. O_w refers to the mW CG water. In each case a 10 X 10 matrix was determined, solved to obtain the new set of parameters, and the procedure carried out iteratively until convergence was reached. For each case, convergence was achieved in less than ten iterations. After the first generation model was developed, a series of new starting points were developed by changing the various σ values. The iterative procedure was applied to each set. The final model was one for which the root mean square error was the least (less than 1).

It should be noted that for the two-body interactions that were not varied, the Lorentz-Bertholet mixing rules were used with the scaling factor of the scaled relevant two-body interactions absorbed into the corresponding ε values. The mixing rules are thus implicitly included in the parametrization procedure. Care should be taken when using mixing rules since they can lead to unphysical results. However, the bottom-up approach with increasing complexity that is being adopted to develop the force-field for real peptoid systems, starting from the alkane chains (hydrophobic side chains), to the peptoid backbone (current work) to more complex side chains (future work) to some extent mitigates this risk.

III. **Results and Discussion**

At the end of the parameterization procedure, the only three-body terms that were found to be non-zero were the water interactions with the carbonyl oxygen involving the $O_wO_{DMA}O_w$, and the $O_wO_wO_{DMA}$ angle, where O_w refers to the CG water (mW) and O_{DMA} to the carbonyl O atom of the DMA. Since the N atom is buried inside the molecule, there is no special preferred orientation of the water molecules or other DMA molecules with the N atom and hence the three-body terms involving the N atoms turn out to be essentially zero as expected. The C atoms, similar to the case of methane and other aliphatic chains in previous studies, also did not result in non-zero three-body terms. The values of each of the parameters are listed in Table 1.

The O_{DMA}-O_w, C_{DMA}-O_w, CH_3-$_{DMA}$-O_w, N_{DMA}-O_w and the CH_3N-$_{DMA}$-O_w radial distribution functions are shown in Figure 3 and the associated coordination numbers in Table 2. Note that these atom types are based on the representations in Figure 1. The CG models does indeed give a fair representation of the solvation environment of the DMA in water and reproduces reasonably well both the positions of the first peak and the corresponding peak height.

However, the CG model radial distribution functions are slightly shifted to smaller distances compared to the AA model, suggesting that the CG DMA molecules, on average, are more hydrophilic than the AA model. The liquid DMA radial distribution functions (C-C, O-O, N-N, CH3N-CH3N and CH3-CH3) are shown in Figure 4 and the coordination numbers are tabulated in Table 3. Once again, the representation of the intermolecular DMA-DMA structure is reproduced relatively well by the CG model when compared to the AA case. Specifically, the first peaks of the O-O and CH3N-CH3N show really good agreement, while for the N atom that is buried inside, the peak position is reproduced whereas the peak height is not. The same holds true for the C-C radial distribution functions. This is not necessarily surprising since these groups are further apart from each other. The CH3-CH3 radial distribution function is not very strongly structured and hence is difficult to reproduce very accurately using this method. The CG model seems to suggest a stronger attraction of the CH3-CH3 as compared to the AA model, since the radial distribution function is shifted to a smaller distance for the CG case as compared to the AA. Nonetheless, the dominant intermolecular interactions are fairly well reproduced by the CG model both for pure DMA and dilute DMA in water.

The radial distribution functions were part of the training set of the model and hence for validation different properties are calculated and compared to both experiment and the AA model in the following section.

Liquid DMA properties

A number of properties of pure liquid DMA were calculated and tabulated in Table 4. The self-diffusion coefficient of DMA was calculated from a 5 ns NVE run using the Einstein equation:[57]

$$D = \lim_{t\to\infty} \frac{1}{6t} < |r(t) - r(0)|^2 > \qquad (9)$$

where r(t) is the position of the center of mass of DMA at time t. The CG diffusion constant is a little more than twice the value of the AA, which is not surprising since the dynamics of CG models are faster than the corresponding all atom systems.[38, 44]

The isothermal compressibility of liquid DMA at 300K around density of $\rho_0 = 937\ kg/m^3$ was found from:[58]

$$\mathrm{K}_T = -\frac{1}{V}\left(\frac{\partial V}{\partial p}\right)_T \approx \left(\frac{\ln\left(\frac{\rho_2}{\rho_1}\right)}{<p_2>-<p_1>}\right)_{\mathrm{T}} \qquad (10)$$

where ρ_2 is $1.01\rho_0$ and ρ_1 is $0.99\ \rho_0$. p_2 and p_1 were the pressures computed from a 10 ns NVT simulation. Surprisingly the CG model reproduces the experimental value more accurately than the AA model to which it was parameterized. The AA model shows a lower compressibility as compared to the CG model. This could be related to the CH3-CH3 radial distribution for CG being shifted to smaller distances as compared to the AA model.

The enthalpy of vaporization was computed from:[57]

$$\Delta H_{vap} = < E_{pot}(g) > - < E_{pot}(l) > + k_b T \qquad (11)$$

where $E_{pot}(g)$ is the gas-phase potential energy and $E_{pot}(l)$ is the liquid-phase potential energy. The CG model gives a higher value for the enthalpy of vaporization as compared to the all atom model and the experimental value, but is within around 10 % of the experimental value.

The next set of studies focuses on aqueous solutions of DMA as well as the simplest polypeptoid that can be formed with DMA as the repeating unit.

Dilute aqueous DMA solution

The solvation free energy of DMA from the CG simulations is very close to the value of the AA model, although both are around 1 kcal/mol higher than the experimental value. The fidelity of the solvation free energy of the CG model to the AA model along with the agreement between

the radial distribution function data between the CG and AA models does suggest that the CG model reproduces the solvation environment around the DMA molecule as seen in the all atom simulations.

Concentrated DMA aqueous solution and simple peptoid solution

The various radial distribution functions of the concentrated DMA-water system and the simple peptoid systems are shown in Figure 5 and Figure 6 respectively. Despite not being part of the training set, the CG model gives a good representation of the solvation environment of the DMA and the simple peptoid in the aqueous system, although once again the CG model is slightly more hydrophilic than the corresponding AA model.

It is clear from both the agreement with the training data as well as the concentrated DMA simulations that the CG model provides a reasonable description of the DMA in bulk conditions. The next set of studies determines its suitability for interfacial regions. The air-liquid interface is an example for an extreme case of an asymmetrical environment and thus is an excellent system to test the applicability of the CG force-field.

Surface tension of liquid DMA

The surface tension of the liquid DMA liquid-air interface was also calculated from the liquid-DMA-air slab simulations at 300K. In order to determine the surface tension γ the following relationship between γ and the pressure tensors, evaluated from the simulation, was used:[57]

$$\gamma = \frac{L_z}{2}[< p_N > - < p_T >] \qquad (8)$$

where L_z is the box length of the direction perpendicular to the interface and $< p_N >$ and $< p_T >$ are pressure tensors perpendicular and tangential to the liquid-air interface,

respectively.[59] The surface tension of the CG model is compared to the value from the AA model along with the experimental data[60] in Table 4.

Liquid and vapor density of liquid DMA at 300K

In order to determine the density of the two phases of pure DMA at 300K, NVT simulations of the liquid DMA-air interface were carried out. The density profile of DMA (using the center of mass of each DMA), $\rho(z)$, is plotted in Figure 7 as a function of distance, z, from the center of mass of the liquid slab in the direction perpendicular to the interface. While both the AA and the CG models show density oscillations near the interfacial region, the oscillations are much more pronounced for the CG model. The stronger CH3-CH3 interactions of the CG model as compared to the AA model could result in this increased structuring at the interface. This can also result in differences in the orientation of the DMA molecules in the interfacial region. The distribution of the angle made by the CH3-C vector of DMA molecules (with z-coordinate of the center of mass between 19 and 21 Å, the beginning of the interfacial region as seen in Figure 7) with the normal to the xy plane (air-liquid interface is along the z-direction, perpendicular to the xy plane) is shown in Figure 8. Although both the AA and CG models show a broad distribution, the AA has a broad peak in the 80-90^o region whereas the CG is shifted to around 120^o.

The density profiles in Figure 7 are fit to the following function,

$$\rho(z) = \frac{\rho_l+\rho_v}{2} - \frac{\rho_l-\rho_v}{2} tanh \frac{z-z_o}{d} \qquad (7)$$

where ρ_l and ρ_v are the densities of the vapor phase, respectively, d is the width of the vapor-liquid interface and d is the position of the Gibbs dividing surface of the interface.[59, 61] The densities of the two phases are thereby determined from this fit and tabulated in Table 5 for both the CG model and the AA model along with the experimental value for the liquid phase.[59] The

CG model liquid density is in good agreement with the AA model, despite the differences in interfacial orientation, and the liquid density compares reasonably well with the experimental value, although both CG and AA densities are higher than the experimental value by around 3 %.

The good agreement between the CG and AA models for the surface tension and the liquid density, despite these parameters not being in the training set, does indicate that the reduced range formalism based CG model can reproduce the results for the structure and thermodynamics of the bulk and the interfacial region obtained from the higher level all atom simulations. This could be in part due to a fortuitous cancelation of errors but is probably also due to the attention to an accurate description of the solvation structure that is the underlying philosophy in the development of this CG model. Previous works with SW CG models have also found good agreement for water/vapor, liquid alkane/water, liquid alkane/vapor, and liquid ether/vapor surface tensions.[38, 39, 44, 55]

Finally, benchmark simulations on liquid DMA were performed to test the performance of the CG model against the AA. Table 6 shows our CG DMA model is at least 27 times faster than the AA DMA model. The scaling behavior clearly improves with increase in the system size.

Future work will concentrate on developing compatible CG models for different side chains of the peptoid system, namely hydrophilic neutral groups such as ethers as well as charged groups such as carboxylate.

IV. Conclusions

A new coarse-grained model of N-N dimethyl acetamide, a representative of the peptoid backbone, has been developed. The underlying philosophy of the CG model development is to reproduce the solvation structure in the liquid of a more complex all atom model, OPLSAA in

this case, using reduced range functional forms that include three-body terms. Without electrostatics or any long-ranged interactions, the CG DMA model is at least 27 times faster than AA model for the system sizes studied. The DMA molecules were represented as 6 heavy atoms interacting through short-ranged potentials. The CG DMA model gives a good representation of the liquid DMA system, aqueous DMA solution as well as the liquid DMA-vapor interface. The pure bulk DMA solvation structure as well as the dilute DMA aqueous solution solvation structure are part of the model training set. The CG model was validated by a comparison to results from AA simulations of the surface tension, enthalpy of vaporization, isothermal compressibility and liquid density of pure DMA, the solvation free energy of DMA in water as well as the solvation structure of both a concentrated aqueous solution of DMA as well as a simple peptoid aqueous solution.

Acknowledgments

This material is based upon work supported by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under EPSCoR Grant No. DE-SC0012432 with additional support from the Louisiana Board of Regents. We gratefully acknowledge computer time allotted by the high-performance computing center at LSU and the Louisiana Optical Network Initiative.

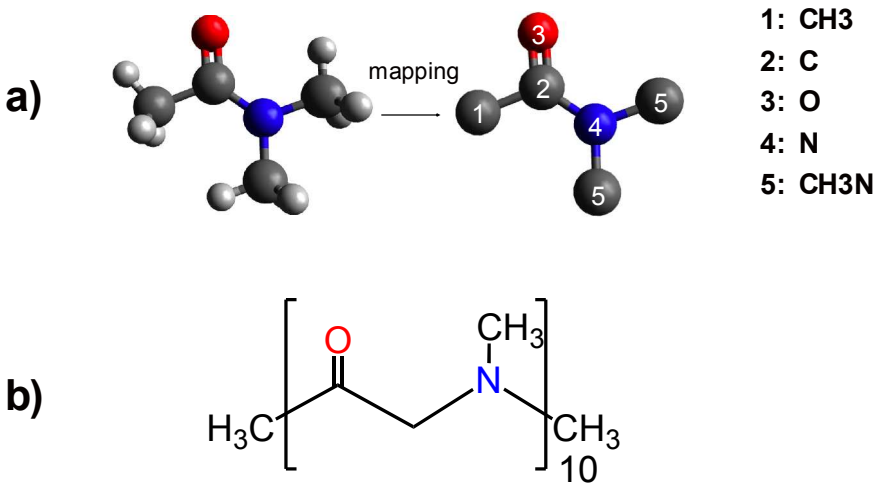


Figure 1. a) Mapping of DMA molecule from the AA to the CG model. The white balls represent H, the gray ones C, the blue ones N, and the red ones O. There are 5 total atom types in the CG DMA molecule, which are 1: CH3, 2: C, 3: O, 4: N, 5: CH3N. b) Schematic of the simple peptoid chain.

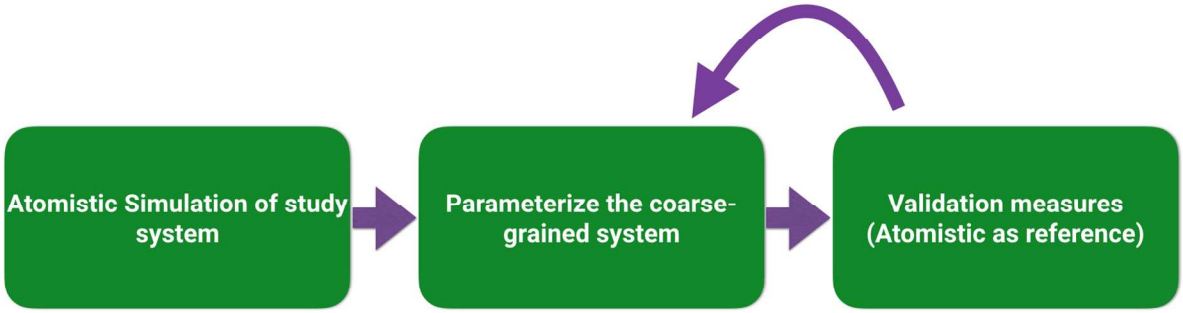


Figure 2. A flowchart of CG model parameterization scheme.

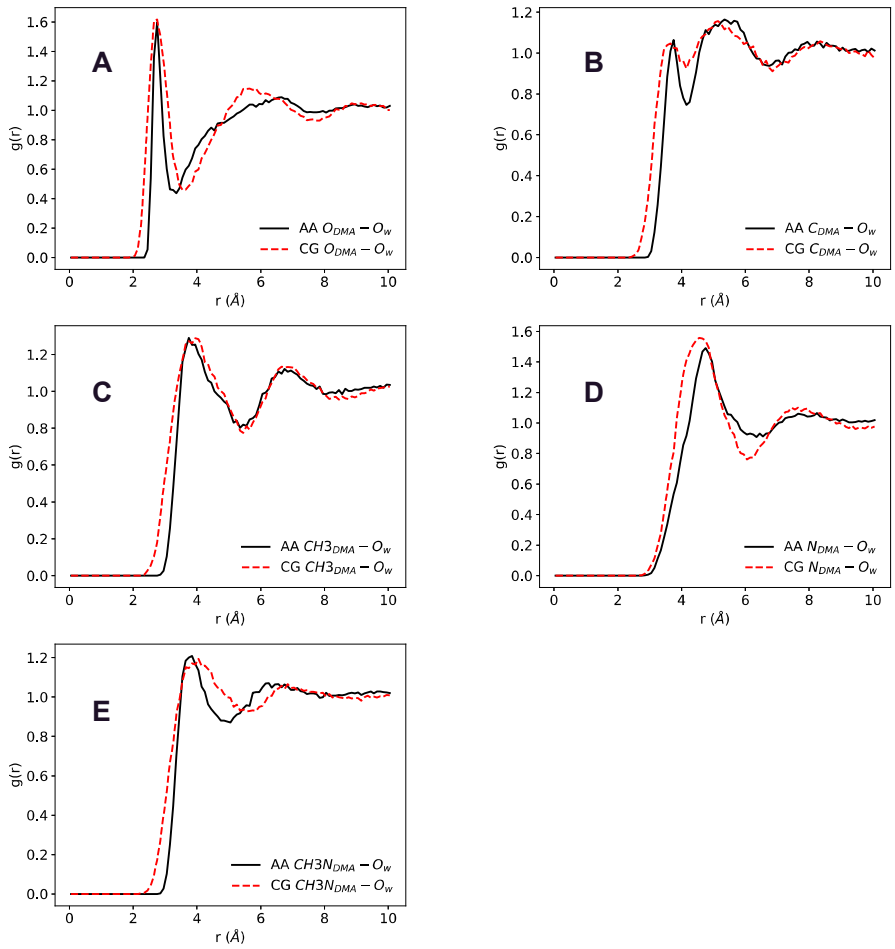


Figure 3. The water-DMA radial distribution functions for dilute DMA solution at 300 K for AA and CG models. Note for the CG case the atom types for DMA are based on Figure 1 and O_w refers to the mW CG water. For the AA, CH_3 refers to the C atom of the methyl group bonded to the carbonyl group, CH_3N refers to the carbon atom of methyl group bonded to the nitrogen atom,

C refers to the carbon atom of the carbonyl group and O to the oxygen atom of the carbonyl group, and O_w refers to the water oxygen atom.

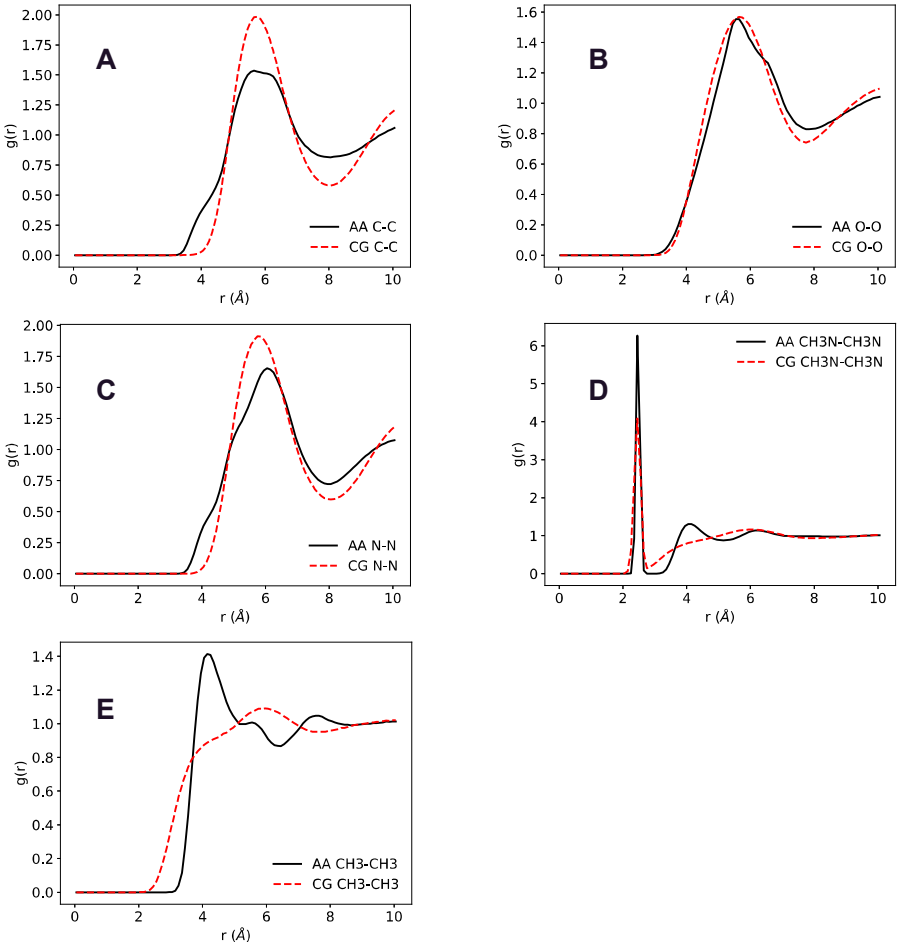


Figure 4. The liquid DMA radial distribution function at 300 K for AA and CG models. Note the atom types are based on Figure 1 for the CG case. For the AA case, CH_3 refers to the carbon atom of the methyl group bonded to the carbonyl group, CH_3N refers to the carbon atom of the methyl group bonded to the N atom, C refers to the carbon atom of the carbonyl group and O to the oxygen atom of the carbonyl group.

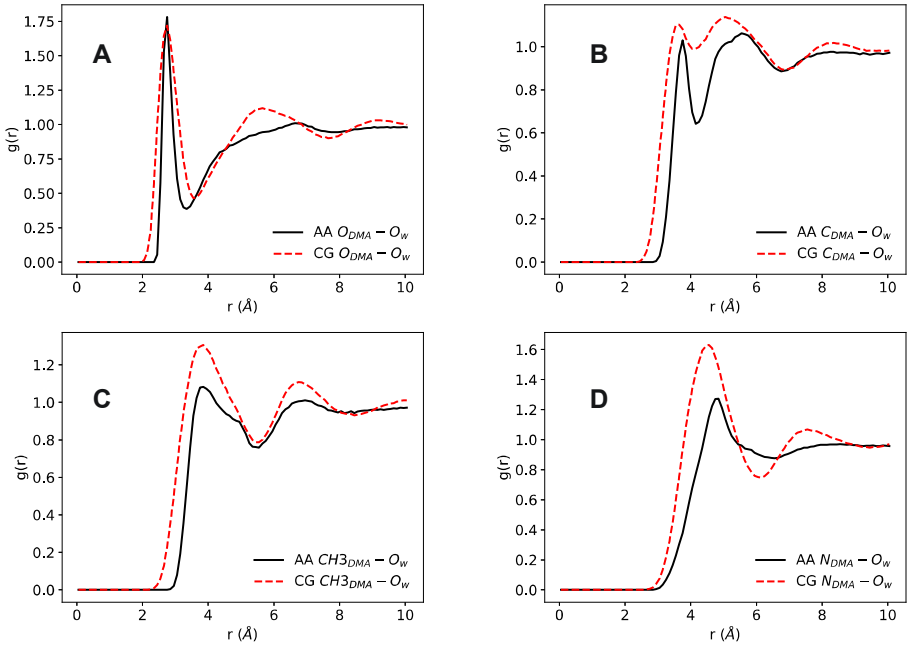


Figure 5. The radial distribution functions for the concentrated aqueous DMA solution at 300 K for AA and CG models. Note for the CG case the atom types for DMA are based on Figure 1 and O_w refers to the mW CG water. For the AA case, CH_3 refers to the carbon atom of the methyl group bonded to the carbonyl group, CH_3N refers to the carbon atom of the methyl group bonded to the nitrogen atom, C refers to the carbon of the carbonyl group and O to the oxygen atom of the carbonyl group, and O_w refers to the water oxygen atom.

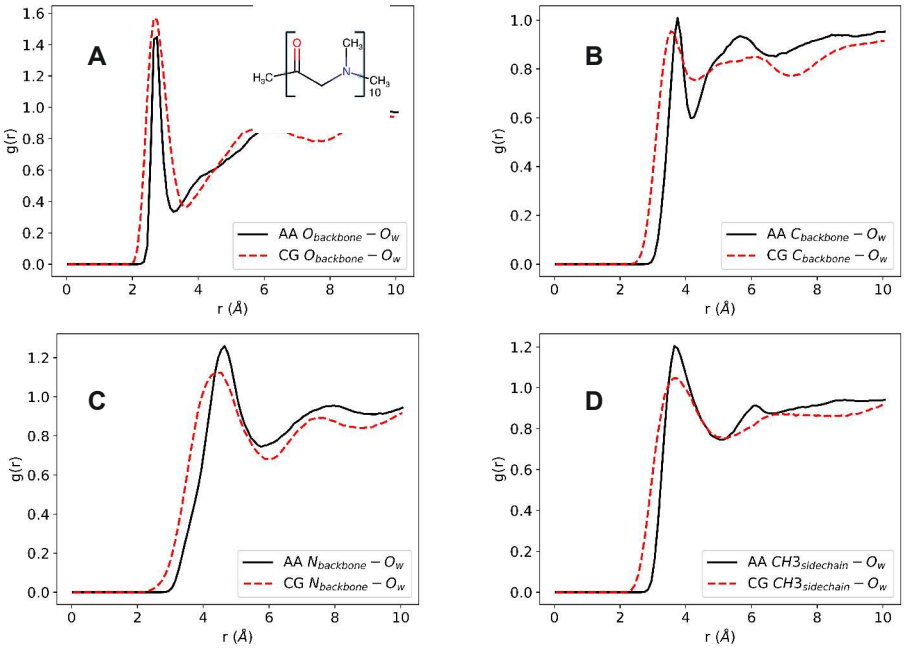


Figure 6. The radial distribution functions for the dilute polypeptoid solution at 300 K for AA and CG models. Note that the chemical structure of polypeptoid is shown in A. For both cases, $C_{backbone}$ and $O_{backbone}$ refer to the carbon and oxygen atom of the carbonyl group, $N_{backbone}$ refers to the nitrogen atom on the backbone, and $CH3_{sidechain}$ refers to the carbon atom of methyl group bonded to the nitrogen atom.

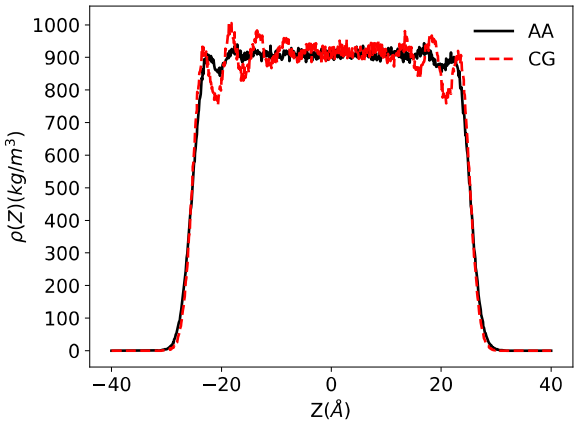


Figure 7. Density profile of DMA liquid-vapor interface at 300K for both the AA and CG simulations. Z represents the distance in the Z direction from the center of mass of the liquid and is perpendicular to the liquid-vapor interface.

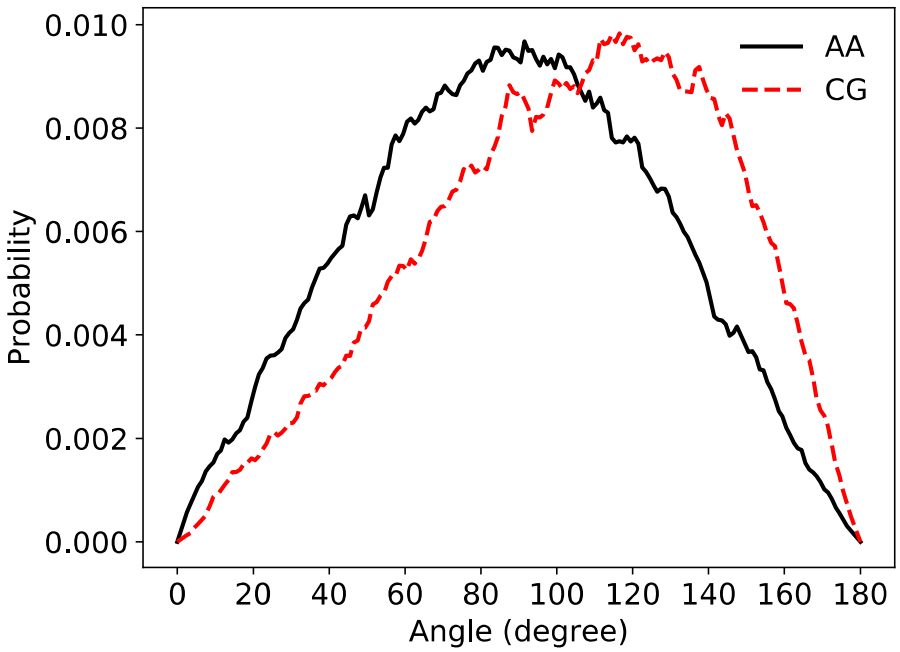


Figure 8: Distribution of the angle made by the CH3-C vector of DMA molecules (with the z-coordinate of the center of mass between 19 and 21 Å) with the normal to the xy plane. (Note the interface is along the z-axis)

(a)

Type i	Type j	ϵ (kcal / mol)	σ (Å)
mW	mW	6.189	2.3925
mW	CH3	0.40	3.7
mW	C	0.08	4.8
mW	O	3.20	2.2
mW	N	0.115	4.4
mW	CH3N	0.40	3.7
CH3	CH3	0.1	4.6
CH3	C	0.09	4.9
CH3	O	0.16	4.2
CH3	N	0.1	5.05
CH3	CH3N	0.1	4.6
C	C	0.09	5.2
C	O	0.1	4.5
C	N	0.09	5.35
C	CH3N	0.09	4.9
O	O	0.6	4.8
O	N	0.115	4.55
O	CH3N	0.16	4.2
N	N	0.1	5.5
N	CH3N	0.1	5.05
CH3N	CH3N	0.1	4.6

(b)

Type i	Type j	Type k	λ	$cos\theta$
mW	mW	mW	23.15	-0.333333333
O	mW	mW	10.00	-0.333333333
mW	O	mW	10.00	-0.333333333

Table 1. Parameters of a) two-body and (b) and three-body interactions for the CG DMA together with previously published parameters for water.[44, 55] The scaling terms obtained from the parameterization step are folded into the ϵ and λ parameters: $\lambda_i = \lambda_{i,initial}\omega_i$ and $\varepsilon_j = \varepsilon_{j,initial}\omega_j$.

	R (Å)		C	
	AA	CG	AA	CG
O_{DMA}-O_w	2.75	2.75	2.8	4.3
C_{DMA}-O_w	3.75	3.65	4.5	6.2
$CH_{3\text{-}DMA}$-O_w	3.75	3.95	19.2	19.9
N_{DMA}-O_w	4.75	4.55	31.5	30.5
$CH_3N_{\text{-}DMA}$-O_w	3.85	4.05	14.5	21.2

Table 2. The Position (R) of the First Maximum in the corresponding radial distribution function and the Coordination Number (C) for dilute DMA system.

	R (Å)		C	
	AA	CG	AA	CG
O-O	5.65	5.65	12.5	12.3
C-C	5.65	5.75	13.5	13.6
N-N	6.05	5.85	13.1	13.7
CH3N-CH3N	2.45	2.45	1.0	1.0
CH3-CH3	4.15	5.75	6.3	10.6

Table 3. The Position (R) of the First Maximum in the corresponding radial distribution function and the Coordination Number (C) for liquid DMA system.

Properties	**AA DMA (300K)**	**CG DMA (300K)**	**Experiment**
Surface tension γ (mN/m)	32±1	39±1	37.13 (293.15 K)[60]
Diffusion coefficient D (10^{-5} cm^2/s)	0.8	1.8	
Isothermal compressibility K_T (Gpa^{-1})	0.53±0.01	0.61±0.01	0.63 (298.15 K)[62]
Solvation free energy ΔG_{solv} (kcal/mol)	-7.5±0.2	-7.7±0.1	-8.5 (298.15 K)[63]
Enthalpy of vaporization ΔH_{vap} (kcal/mol)	12.40±0.01	13.4±0.2	12.01 (298.15 K)[62]

Table 4. Comparison of DMA models and Experiment.

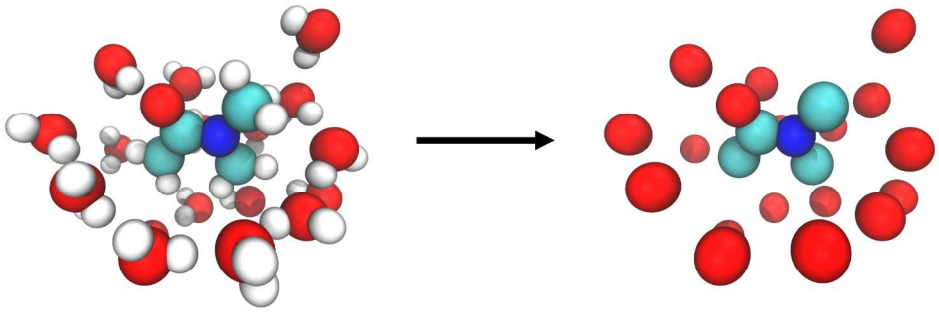
AA		**CG**		**Experiment[64]**
ρ_L (kg/m^3)	ρ_V (kg/m^3)	ρ_L (kg/m^3)	ρ_V (kg/m^3)	ρ_L (kg/m^3)
907.33	0.19	905.21	0.82	937

Table 5. The liquid density (ρL) and vapor density (ρV) for DMA at 300K together with experiment data for the liquid at 298.15K.

CPUS (cores)	2560 DMAs			12800 DMAs			64000 DMAs		
	AA	CG	CG/AA ratio	AA	CG	CG/AA ratio	AA	CG	CG/AA ratio
	ns/day			ns/day			ns/day		
20	4.44	118.38	27	0.89	25.80	29	0.17	5.72	34
40	8.14	218.61	27	1.71	47.39	28	0.33	10.88	33
60	10.80	304.60	28	2.44	69.26	28	0.50	14.89	30
80	11.94	361.60	30	3.16	89.54	28	0.67	19.58	29
100	15.05	437.72	29	4.04	109.30	27	0.82	24.20	30

Table 6. Computational efficiency for three different systems including 2560, 12800 and 64000 DMAs using 20, 40, 60, 80 and 100 CPU cores. All benchmark simulations were performed on the QB2 computational nodes of Louisiana Optical Network Infrastructure (LONI), each with 2.8 GHz E5-2680v2 Xeon processors and 64 GB memory using the LAMMPS software.

TOC image



References

1. S. P. Nunes, *Macromolecules*, 2016, **49**, 2905-2916.
2. H. Feng, X. Lu, W. Wang, N.-G. Kang and J. Mays, *Polymers*, 2017, **9**, 494.
3. J. Sun and R. N. Zuckermann, *ACS Nano*, 2013, **7**, 4715-4732.
4. C. Fetsch, A. Grossmann, L. Holz, J. F. Nawroth and R. Luxenhofer, *Macromolecules*, 2011, **44**, 6746-6758.
5. A. M. Rosales, R. A. Segalman and R. N. Zuckermann, *Soft Matter*, 2013, **9**, 8400.
6. D. Zhang, S. H. Lahasky, L. Guo, C.-U. Lee and M. Lavan, *Macromolecules*, 2012, **45**, 5833-5841.
7. N. Gangloff, J. Ulbricht, T. Lorson, H. Schlaad and R. Luxenhofer, *Chem. Rev.*, 2016, **116**, 1753-1802.
8. H. Otsuka, Y. Nagasaki and K. Kataoka, *Mater. Today*, 2001, **4**, 30-36.
9. C. Fetsch, J. Gaitzsch, L. Messager, G. Battaglia and R. Luxenhofer, *Scientific Reports*, 2016, **6**, 33491.
10. G. L. Sternhagen, S. Gupta, Y. Zhang, V. John, G. J. Schneider and D. Zhang, *J. Am. Chem. Soc.*, 2018, **140**, jacs.8b00461.
11. C. Fetsch and R. Luxenhofer, *Macromol. Rapid Commun.*, 2012, **33**, 1708-1713.
12. A. M. Rosales, H. K. Murnen, S. R. Kline, R. N. Zuckermann and R. A. Segalman, *Soft Matter*, 2012, **8**, 3673.
13. D. T. Mirijanian, R. V. Mannige, R. N. Zuckermann and S. Whitelam, *J. Comput. Chem.*, 2014, **35**, 360-370.
14. A. Prakash, M. D. Baer, C. J. Mundy and J. Pfaendtner, *Biomacromolecules*, 2018, **19**, 1006-1015.
15. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1988, **110**, 1657-1666.
16. W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225-11236.
17. P. Du, A. Li, X. Li, Y. Zhang, C. Do, L. He, S. W. Rick, V. T. John, R. Kumar, D. Zhang and P. D. Butler, *Phys. Chem. Chem. Phys.*, 2017, **131**, 2798-2799.
18. A. Barducci, J. Pfaendtner and M. Bonomi, in *Molecular Modeling of Proteins*, ed. A. Kukol, Springer New York, New York, NY, 2015, DOI: 10.1007/978-1-4939-1465-4_8, pp. 151-171.
19. A. Mitsutake, Y. Mori and Y. Okamoto, in *Biomolecular Simulations: Methods and Protocols*, eds. L. Monticelli and E. Salonen, Humana Press, Totowa, NJ, 2013, DOI: 10.1007/978-1-62703-017-5_7, pp. 153-195.
20. T. Maximova, R. Moffatt, B. Ma, R. Nussinov and A. Shehu, *PLOS Computational Biology*, 2016, **12**, e1004619.
21. M. G. Saunders and G. A. Voth, *Annual Review of Biophysics*, 2013, **42**, 73-93.
22. J. McCarty, I. Y. Lyubimov and M. G. Guenza, *J. Phys. Chem. B*, 2009, **113**, 11876-11886.

23. F. L. Carlos, O. N. Steve, B. M. Preston, C. S. John and L. K. Michael, *J. Phys.: Condens. Matter*, 2002, **14**, 9431.
24. S. Kmiecik, D. Gront, M. Kolinski, L. Wieteska, A. E. Dawid and A. Kolinski, *Chem. Rev.*, 2016, **116**, 7898-7936.
25. B. W. Boras, S. P. Hirakis, L. W. Votapka, R. D. Malmstrom, R. E. Amaro and A. D. McCulloch, *Frontiers in Physiology*, 2015, **6**, 250.
26. S. C. L. Kamerlin, S. Vicatos, A. Dryga and A. Warshel, *Annu. Rev. Phys. Chem.*, 2011, **62**, 41-64.
27. M. Guenza, *The European Physical Journal Special Topics*, 2015, **224**, 2177-2191.
28. S. Izvekov, J. M. J. Swanson and G. A. Voth, *J. Phys. Chem. B*, 2008, **112**, 4711-4724.
29. V. Tozzini, *Acc. Chem. Res.*, 2010, **43**, 220-230.
30. M. Levitt and A. Warshel, *Nature*, 1975, **253**, 694-698.
31. S. J. Marrink, H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. De Vries, *J. Phys. Chem. B*, 2007, **111**, 7812-7824.
32. S. J. Marrink and D. P. Tieleman, *Chem. Soc. Rev.*, 2013, **42**, 6801-6822.
33. W. Schommers, *Phys. Rev. A*, 1983, **28**, 3599-3605.
34. F. Ercolessi and J. B. Adams, *EPL (Europhysics Letters)*, 1994, **26**, 583.
35. S. Izvekov and G. A. Voth, *J. Phys. Chem. B*, 2005, **109**, 2469-2473.
36. S. Izvekov, M. Parrinello, C. J. Burnham and G. A. Voth, *J. Chem. Phys.*, 2004, **120**, 10896-10913.
37. A. Lyubartsev, A. Mirzoev, L. Chen and A. Laaksonen, *Faraday Discuss.*, 2010, **144**, 43-56.
38. G. Gyawali, S. Sternfield, R. Kumar and S. W. Rick, *J. Chem. Theory Comput.*, 2017, **13**, 3846-3853.
39. B. Raubenolt, G. Gyawali, W. Tang, K. S. Wong and S. W. Rick, *Polymers*, 2018, **10**, 475.
40. T. K. Haxton, R. N. Zuckermann and S. Whitelam, *J. Chem. Theory Comput.*, 2016, **12**, 345-352.
41. M. L. Hebert, D. S. Shah, P. Blake, J. P. Turner and S. L. Servoss, *Organic & Biomolecular Chemistry*, 2013, **11**, 4459-4464.
42. F. H. Stillinger and T. A. Weber, *Phys. Rev. B*, 1985, **31**, 5262-5271.
43. R. Kumar and J. L. Skinner, *J. Phys. Chem. B*, 2008, **112**, 8311-8318.
44. V. Molinero and E. B. Moore, *J. Phys. Chem. B*, 2009, **113**, 4008-4016.
45. W. L. Jorgensen, *J. Am. Chem. Soc.*, 1981, **103**, 335-340.
46. S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1-19.
47. D. J. Evans and B. L. Holian, *J. Chem. Phys.*, 1985, **83**, 4069-4074.
48. G. J. Martyna, D. J. Tobias and M. L. Klein, *J. Chem. Phys.*, 1994, **101**, 4177-4189.
49. B. A. Luty and W. F. Van Gunsteren, *J. Phys. Chem.*, 1996, **100**, 2581-2587.
50. M. Mezei, *J. Chem. Phys.*, 1987, **86**, 7084-7088.
51. P. V. Klimovich, M. R. Shirts and D. L. Mobley, *J. Comput. Aided Mol. Des.*, 2015, **29**, 397-411.
52. T. Steinbrecher, I. S. Joung and D. A. Case, *J. Comput. Chem.*, 2011, **32**, 3253-3263.
53. T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber and W. F. van Gunsteren, *Chem. Phys. Lett.*, 1994, **222**, 529-539.
54. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1-2**, 19-25.

55. L. C. Jacobson and V. Molinero, *J. Phys. Chem. B*, 2010, **114**, 7302-7311.
56. G. Gyawali and S. Rick, A git repository for the free energy routines is hosted at https://github.com/ggyawali/pair_sw_soft/tree/master).
57. C. Caleman, P. J. van Maaren, M. Hong, J. S. Hub, L. T. Costa and D. van der Spoel, *J. Chem. Theory Comput.*, 2012, **8**, 61-74.
58. K. A. Motakabbir and M. Berkowitz, *J. Phys. Chem.*, 1990, **94**, 8359-8362.
59. J. Alejandre, D. J. Tildesley and G. A. Chapela, *J. Chem. Phys.*, 1995, **102**, 4574-4583.
60. R. K. Shukla, A. Kumar, U. Srivastava, N. Awasthi and J. D. Pandey, *Can. J. Phys.*, 2013, **220**, 211-220.
61. R. Sakamaki, A. K. Sum, T. Narumi and K. Yasuoka, *J. Chem. Phys.*, 2011, **134**, 124708.
62. Y. Marcus, *The Properties of Solvents*, John Wiley & Sons, 1998.
63. D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley and W. Sherman, *J. Chem. Theory Comput.*, 2010, **6**, 1509-1519.
64. H. Iloukhani and K. Khanlarzadeh, *J. Chem. Eng. Data*, 2006, **51**, 1226-1231.