JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

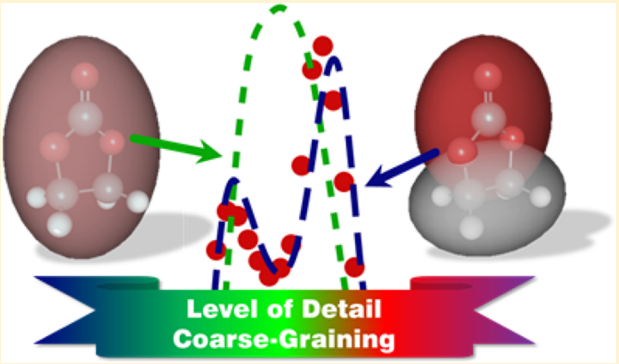
# Systematic Generation of Anisotropic Coarse-Grained Lennard-Jones Potentials and Their Application to Ordered Soft Matter

Andreas F. Tillack, Lewis E. Johnson, Bruce E. Eichinger, and Bruce H. Robinson*

Department of Chemistry, University of Washington, Seattle, Washington 98195-1700, United States

Supporting Information

**ABSTRACT:** We have developed an approach to coarse-grained (CG) modeling of the van der Waals (vdW) type of interactions among molecules by representing groups of atoms within those molecules in terms of ellipsoids (rather than spheres). Our approach systematically translates an arbitrary underlying all-atom (AA) representation of a molecular system to a multisite ellipsoidal potential within the family of Gay−Berne type potentials. As the method enables arbitrary levels of coarse-graining, or even multiple levels of coarse-graining within a single simulation, we describe the method as a Level of Detail (LoD) model. The LoD model, as integrated into our group's Metropolis Monte Carlo computational package, is also capable of reducing the complexity of the molecular electrostatics by means of a multipole expansion of charges obtained from an AA force field (or directly from electronic structure calculations) of the charges within each ellipsoid. Electronic polarizability may additionally be included. The present CG representation does not include transformation of bonded interactions; ellipsoids are connected at the fully atomistic bond sites by freely rotating links that are constrained to maintain a constant distance. The accuracy of the method is demonstrated for three distinct types of self-assembling or self-organizing molecular systems: (1) the interaction between benzene and perfluorobenzene (dispersion interactions), (2) linear hydrocarbon chains (a system with large conformational flexibility), and (3) the self-organization of ethylene carbonate (a highly polar liquid). Lastly, the method is applied to the interaction of large (∼100 atom) molecules, which are typical of organic nonlinear optical chromophores, to demonstrate the effect of different CG models on molecular assembly.




## INTRODUCTION

The search for accurate and systematic coarse-grained (CG) methods is an ongoing challenge in classical chemical simulation. An ideal potential should represent a complex molecule and capture the interactions that determine complex self-assembly processes, but minimize computational expense to enable calculations on mesoscale systems. The most fundamental method for representing complex molecules using classical mechanics is via all atom (AA) force fields.[1−4] Extensive efforts have been made to reduce the complexity of AA force fields by combining atoms into rigid units (e.g., united atom approaches).[5−7] Herein, we will focus on a systematic method for fitting coarse-grained nonbonded interactions, in particularly van der Waals (vdW) type interactions. Typical AA force fields use spherically symmetric potentials as the fundamental representation of atoms within molecules; each of these spheres has an associated vdW radius and one or more parameters representing the depth of the potential well.[8−12] Most coarse-grained potentials, such as the MARTINI force field[13−16] and even SPC/E,[17] which is an all-charge representation of water, also use spherical steric potentials to represent arrangements of multiple atoms. Spheres have also been used by others to represent complex systems such as lipids in a membrane.[18−26]

Ellipsoids have the potential to provide a wide range of representations for anisotropic or planar systems, but require additional computational cost for generating a reduced potential and for solving two-body interaction potentials. We have used ellipsoids to represent molecules or subcomponents thereof in much of our prior work,[27−30] wherein we focused on using ellipsoids to describe organic chromophores used in nonlinear optical applications (ONLO chromophores). Such molecules have rigid conjugated cores, which typically contain aromatic moieties and have large net dipole moments; these chromophores also often contain flexible pendant groups. They are assembled as either neat materials or in host polymers, and need to have a bulk acentric order for electro-optic applications, the order being usually generated using an external poling field. A proper description of the intermolecular interactions requires the use of effective CG methods. Here we demonstrate that ellipsoids are computationally efficient (see Supporting Information, section S.E), can be systematically generated

Received: February 27, 2016
Published: July 19, 2016






DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

from an underlying AA force field of choice, and can be added to the pallet of tools available for molecular modeling. Under identical simulation conditions, our CG parametrization closely reproduces the intermolecular radial and angular distribution functions as well as other statistical properties such as the system density and the dielectric constant when compared to simulations using an underlying AA force field.

One of the most robust, and popular, methods for simulating the van der Waals (vdW) type interactions among coarse-grained representations of multiatom moieties is the Gay–Berne (GB) potential, which has a general form to represent the structure of a vdW type interaction between two objects of arbitrary shape, though generally applied to spheroids.[31−35] The parameters used in a GB-type potential are the depth of the potential energy minimum between two GB type objects, $\varepsilon_{AB}$, the width of the potential well, $\sigma^0$, the center to center distance between the two objects, $R_{AB} = |\vec{R}_B - \vec{R}_A|$, and a shape-dependent term, $\sigma_{AB}$, describing the distance at which the surfaces of the two GB type objects touch. The two objects are touching when the LJ potential is zero. The two objects are further defined by the positions of their centers, $\vec{R}_B$ and $\vec{R}_A$, as well as how they are rotated in space from the internal coordinates, as defined by the rotation matrices, $\mathcal{R}_B$ and $\mathcal{R}_A$. In general, both the energy minimum, $\varepsilon_{AB}$, and the contact parameter, $\sigma_{AB}$, are functions of these properties, $\varepsilon_{AB} = \varepsilon_{AB}(\vec{R}_A, \vec{R}_B, \mathcal{R}_A, \mathcal{R}_B)$ and $\sigma_{AB} = \sigma_{AB}(\vec{R}_A, \vec{R}_B, \mathcal{R}_A, \mathcal{R}_B)$, while the potential well width, $\sigma^0$, is typically fixed. The resulting potential of the GB form between two objects A and B is given as

$$V_{AB} = 4\varepsilon_{AB}\left\{\left(\frac{\sigma^0}{R_{AB} - \sigma_{AB} + \sigma^0}\right)^{12} - \left(\frac{\sigma^0}{R_{AB} - \sigma_{AB} + \sigma^0}\right)^{6}\right\} \quad (1)$$

This GB-type potential is a generalized form of the Lennard-Jones potential extended to represent objects of arbitrary shape. (The LJ potential can be obtained when one sets $\sigma_{AB} = \sigma^0 =$ const.). This is a general CG form with, as yet, no obvious connection between the structure of the object that it is meant to embody and the AA representation of LJ interactions for the underlying complex molecules. The difficulty comes in finding proper values for all of the parameters for the CG potential. Several different examples of fitting these parameters are given in the literature, with varying levels of complexity in the fitting procedure.[36−39] While the form given in eq 1 is general, fitting the parameters can be challenging, and the resulting parameters may have less intuitive physical meaning than the standard Lennard-Jones parameters. For these reasons, we have chosen to develop parameters based on the fundamental assumption that the general shape of each of the two interacting structures can be described by an ellipsoid. These parameters will be determined by comparing the GB form of the LJ potential, embodied in eq 1, with the potential generated by the underlying AA potential (Supporting Information, section S.A).

One approach is to optimize the parameters of all CG moieties based on running small-scale simulations and fitting the radial and orientational distribution functions, or the related potentials of mean force, between all CG unit pairs. In contrast, our approach, follows the notion of a transferrable force field by separately fitting each CG unit. This is achieved by optimizing the agreement at the pair potential level for each individual CG moiety using varying sizes of LJ test spheres as interaction partners. The parameters are then assembled into interaction potentials between CG moiety pairs by using a set of combination rules. This approach limits the overall number of parameters fitted to one set of parameters per CG unit rather than one set of parameters per CG unit pair in the potential of mean force approach. *Prima facie*, reducing the complexity of subunits interacting with a generic potential would be expected to be less accurate than fitting CG moieties in terms of explicit pair potentials. Surprisingly, our approach is able to efficiently recover a high level of accuracy with respect to all-atom pair potentials. Should further optimization be desired, one can most certainly use our parameters as initial guesses for more elaborate fitting schemes.

The parametrization of the vdW potential is one example of the larger problem of representing complex molecules with discrete anisotropic components such as ellipsoids. At the highest level of detail (LoD), each atom is represented by a single ellipsoid. This is the basic solution, as the parameters are identical to the AA force field, representing the molecule as a collection of spheres. At the other extreme, the entire molecule may be represented as a single ellipsoid. This is the most coarse-grained, or lowest level of detail. There is a sequence of possible LoDs for representing a molecule in between these two extremes. Nevertheless, once a LoD has been chosen, the resulting set of parameters describing the potential energy of the system should be systematically based on the chosen AA force field.

Our goal is to demonstrate that one can construct a CG representation of an AA force field, that accurately mimics the properties and short-range structure, by matching the radial correlation function, $g(r)$, for the interaction of two AA systems with the associated CG potential. In a given LoD, a set of atoms within a molecule is described by a single ellipsoid, with the total energy defined as a sum over pairs of ellipsoids both within a molecule and with other molecules. One way in which two ellipsoids spatially relate can be described by the contact function, $F_{AB}$, between them (see section S.D of the Supporting Information). The contact function (CF), whose analytic form was given by Perram and Wertheim, was initially developed for hard ellipsoids[40,41] and has previously been adapted to a Lennard-Jones-like form,[29,38]

$$V_{LJ}^{CF} = 4\varepsilon_{AB}\{F_{AB}^{-6} - F_{AB}^{-3}\} \quad (2)$$

where the contact function depends on the effective contact distance between the sites, and the contact function is

$$F_{AB} = \frac{\vec{R}_{AB}^{\dagger}\cdot\vec{R}_{AB}}{\vec{R}_{AB}^{0\dagger}\cdot\vec{R}_{AB}^{0}} = \left(\frac{|\vec{R}_{AB}|}{|\vec{R}_{AB}^{0}|}\right)^2 = \left(\frac{R_{AB}}{\sigma_{AB}}\right)^2 \quad (3)$$

Here, similar to the definitions above, $\vec{R}_{AB}$ is the vector between the centers of two ellipsoids. $\vec{R}_{AB}^0$ is the vector between the centers if the two ellipsoids were moved along $\vec{R}_{AB}$ to the point where the ellipsoid surfaces just contact/touch, its magnitude is $\sigma_{AB}$. When the contact function is 1, the two surfaces are touching at a single point. In the Lennard-Jones-like form of eq 2, an additional assumption had been made that $\varepsilon_{AB}$ can be represented as a single, orientation-independent parameter, and defined as the geometric mean of scalar energy parameters for each ellipsoid as per the common Berthelot combination rules.[29,38,42]



DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

Our goal at this juncture is to show how the contact function can be merged into the GB form of the potential. Specific examples have been given in the literature.[29,30,36−39,43] We will show that there are various strategies and approximations that can be made within this approach. We divide the approach into three different methods: The first method is the simplest one expressed by eq 2 where the LJ energy parameter, $\varepsilon_{AB}$, is a constant; the second method adds a dependence of the shape of the two ellipsoids into the LJ energy parameter, $\varepsilon_{AB} = \varepsilon_{AB}(\vec{R}_A, \vec{R}_B, \mathcal{R}_A, \mathcal{R}_B)$; and the third method expressed by eq 1 further includes the effective width parameter, $\sigma^0$. While the focus of this work is on potential energy calculations as used in our Metropolis Monte Carlo simulations, forces and torques needed in molecular dynamics calculations with ellipsoids were derived in analytic form by Paramonov and Yaliraki,[38] and could be implemented in MD methods in future work. Such additional calculations for ellipsoids in MD simulations are likely to be costlier than the force calculations using typical spherical potentials.

We outline below a systematic, deterministic, and automated approach to defining the various terms in the Gay-Berne potential, eq 1, in terms of the physical structure of the ellipsoids, using the contact function in such a way that all terms best represent the underlying AA force field. We use the pairwise potential energy functions to compare the approximate GB potential with the AA potential. The radial distribution function $g(r)$ is the primary metric used to determine the reproducibility of the AA force field in NPT ensemble simulations owing to its direct relationship with the potential of mean force.

We illustrate the LoD parametrization method using benzene as an example of a molecule that can be heavily reduced in complexity through coarse-graining. We then benchmark the model for three different systems: (1) liquid benzene interacting with perfluorobenzene, (2) a linear hydrocarbon chain, and (3) liquid ethylene carbonate. In the first case, we compare the CG representation with the AA force fields for interactions of two benzene molecules as well as the radial distribution function for the 1:1 stoichiometric mixture of benzene and perfluorobenzene. This system was chosen because benzene and perfluorobenzene are nonpolar and charge-neutral and yet the nature of the interactions between them is distinct from the pure molecules alone. In the second comparison, we evaluate the ability of LoD models of different flexibility to represent the conformational flexibility of a hydrocarbon chain as measured by its end-to-end distance. Such a system has long been used as a model for polymer chain distributions and several intuitive models have emerged to relate polymer flexibility, and the surface area of the individual units of a polymer. Finally, ethylene carbonate is studied as an example of an organic liquid in which electrostatic interactions dominate, leading to an extremely large dipolar density and dielectric constant.[44−46] Strongly interacting liquids are challenging to simulate even with AA potentials.[9−11]

## METHODS

All simulations in this report were run using our custom implementation of the Metropolis Monte Carlo technique. In addition to our previously published version[29,47] the current implementation features (a) the ability to definite molecules from common structure file formats, (b) coarse-graining of molecular subunits to an arbitrary, user-definable level, (c) the usage of charges (in addition to dipoles) in our reaction field treatment (see Supporting Information, section S.A, for more details), and (d) the isothermal−isobaric (NPT) ensemble. We use parts (see Supporting Information, section S.E, for the parameters used) of the OPLS-AA[48] force field to provide the underlying potential, but note that any fully atomistic force field or united atom force field or even a quantum mechanically based calculation can be used. Specifically, in the simulation results shown, the OPLS-AA force field provided only the Lennard-Jones energy parameters and radii, but the molecular geometry and charges were obtained from DFT calculations using Gaussian '09 at the B3LYP/6-31G(d) level of theory with ESP charges obtained from the CHELPG fitting method.[49,50]

**Development of CG Parameters.** Using the CG method, each and every molecule in the system is divided up into various subunits of arbitrary size. We begin by suggesting that we cluster together a set of N atoms within a molecule (which is often a small subset of all the atoms in the molecule) to be represented by a single ellipsoid. Figure 1 illustrates our general approach, using benzene as an example, in which all ($N = 12$) atoms are replaced by a single ellipsoid; closely approximating the shape of the benzene ring by an oblate ellipsoid.

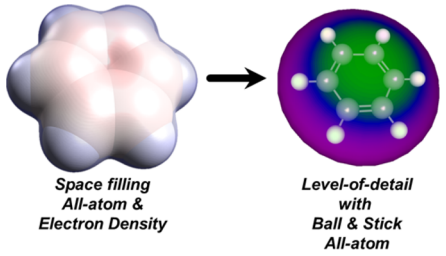



**Figure 1.** Progression from space filling all-atom (AA) model (shown here with overlaid electron density isosurface at 0.0025 level) to a single ellipsoid CG model overlaid on the underlying AA model (ball and stick representation). Ellipsoid coloring represents LJ well depth as a function of the molecular position of an interacting test sphere using eq 18, going from smallest (magenta) to largest (green) in magnitude.

The ellipsoid is constructed by first finding a center of expansion, $\vec{r}_{LoD}$ from the vectors describing the coordinates of each atom, $\vec{r}_i$. This center is taken to be the weighted centroid of the set of N atoms: $\vec{r}_{LOD} = \sum_{i=1}^{N} w_i \vec{r}_i$ where the weights must sum to unity. We have considered uniform weighting and weights proportional to the atomistic LJ energies $w_i \propto \sqrt{\epsilon_i}$, and find the latter slightly better. The method to generate the axes of the ellipsoid begins with the gyration tensor,

$$S = \frac{3}{N}\sum_{i=1}^{N}(\vec{r}_i - \vec{r}_{LOD})(\vec{r}_i - \vec{r}_{LOD})^{\dagger} \tag{4}$$

For example, if there are six points located at $\pm a$ on $x$, $\pm b$ on $y$, and $\pm c$ on $z$, and $\vec{r}_{LOD} = \vec{0}$ then

$$S = \begin{pmatrix} a^2 & 0 & 0 \\ 0 & b^2 & 0 \\ 0 & 0 & c^2 \end{pmatrix}$$

This tensor defines an ellipsoid (see Supporting Information, sections S.B and S.D) where the points in space, $\vec{r}_{El}$, that define the surface of the ellipsoid are given by $\vec{r}_{El}^{\dagger}\cdot S^{-1}\cdot\vec{r}_{El} = 1$ and $a$, $b$, and $c$ are its semiaxes. A rotation of the original six points


DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

around the center will still return the same diagonalized form of the gyration tensor shown (the eigenvalues of a square matrix are invariant under rotation). Moreover, one can use the eigenvectors to reconstruct the proper rotation matrix used (more information can be found in the Supporting Information, section S.B, eqs SB.2−6). Therefore, each atom is represented by six points along the major axes at the surface of its LJ sphere. This guarantees that the gyration tensor will never be singular and will provide a shape and a proper rotation matrix upon diagonalization.

Construction of the gyration tensor, as just described, provides a good starting point to get ellipsoid semiaxes ratios and their orientation in space relative to the underlying AA structure. However, we found that using the semiaxes obtained from the eigenvalues of the gyration tensor typically leads to ellipsoid sizes that are too small as evidenced by inaccurate densities in simulations using the model. The two systematic approaches we use to correct this behavior are scaling of semiaxes by a constant factor to match the excluded volume of the underlying AA system (as done in our prior work[30]) and individually rescaling each of the semiaxis based on the boundaries of the underlying AA system (details in the Supporting Information, section S.B, eqs SB.7−11). We have found that the orientation of the ellipsoid generated by the gyration tensor is satisfactory and does not require further adjustment. As shown later for the benzene/perfluorobenzene system, the AA-based rescaling approach is better able to match the results from the underlying AA system than the excluded volume approach.

*Fitting Parameters for a GB-Derived Potential.* In terms of the Gay−Berne potential, the contact function between two ellipsoids, given in eq 3, is not only a function of the semiaxes of the two ellipsoids, but also depends on the relative orientation of the two ellipsoids and the distance between centers, $R_{AB} = |\vec{R}_B - \vec{R}_A|$. If the two ellipsoids are both simple spheres with radii $a$ and $b$, then the contact function gives $\sigma_{AB} = a + b$. Thus, the contact function is consistent with the LJ potential for spheres. We now use the contact function in the GB formulation of the LJ potential. The relative, or fractional, contact distance from the contact function then is $f_{AB} = \frac{\sigma_{AB}}{R_{AB}} = \frac{1}{\sqrt{F_{AB}}}$ Thus, the GB type potential of eq 1, with a fixed, adjusted width (AW) of the potential, now containing the contact function for the interaction between two ellipsoids, A and B, may be written as

$$V_{AB}^{AW} = 4\varepsilon_{AB}\left\{\left(\frac{\sigma^0}{R_{AB}(1 - f_{AB}) + \sigma^0}\right)^{12} - \left(\frac{\sigma^0}{R_{AB}(1 - f_{AB}) + \sigma^0}\right)^{6}\right\} \quad (5)$$

The general strategy then is to compare the potential energy for the LoD of interest, as defined by eq 5, with that given by the AA force field for the underlying system of atoms, (see Supporting Information, section S.A). We reduce the problem by considering how a single ellipsoid interacts with a predetermined test sphere, p, with known radius $w_p$ and energy parameter, $\varepsilon_p$. The parameters $\{\sigma^0$ and $\varepsilon_{AB}\}$ of the GB-type potential (eq 5) will then be optimized so that the potential agrees with the potential generated by the underlying AA force field interacting with the same test sphere, eq 8, below. To adapt the LJ potential, of eq 5, for the case of an ellipsoid interacting with a test sphere, we assume the combination rules used for the pairwise interaction energy, $\varepsilon_{A,p}$, and the potential well width, $\sigma^0 = w_A + w_p$. The parameter, $w_A$, like $\sigma^0$, is considered to be a constant, and is the adjusted width (AW) of the ellipsoid. With these two additional assumptions, eq 5 is simplified to

$$V_{A,p}^{AW} = 4\sqrt{\varepsilon_p\varepsilon_A}\left\{\left(\frac{w_A + w_p}{R_{Ap} - \delta}\right)^{12} - \left(\frac{w_A + w_p}{R_{Ap} - \delta}\right)^{6}\right\} \quad (6)$$

Here, ellipsoid A is interacting with a test sphere, p. The ellipsoid shape and orientation dependence is contained in the parameter $\delta$. Unlike that for Paramonov−Yaliraki[38] we consider $\delta$ to be identical in both terms of eq 6 and more akin to the original formulation of Gay and Berne[31,32] as used in eqs 1 and 5:

$$\delta = R_{Ap}f_{AB} - \sigma^0 = R_{Ap}f_{Ap} - (w_A + w_p) \quad (7)$$

The point at which $V_{A,p}^{AW} = 0$, when the two objects are touching, occurs when $f_{Ap} = 1$, and therefore the value of $w_A$(or $\sigma^0$) does not affect this point. The all atom (AA) LJ potential, from the AA representation, in which the atoms are spheres with radii $w_i$ and LJ energy parameters $\varepsilon_i$, interacting with the test sphere is given by the potential, $V_{AA,p}$:

$$V_{AA,p} = \sum_{i=1}^{N} 4\sqrt{\varepsilon_p\varepsilon_i}\left\{\left(\frac{w_i + w_p}{R_{ip}}\right)^{12} - \left(\frac{w_i + w_p}{R_{ip}}\right)^{6}\right\} \quad (8)$$

The two potential energies of eqs 6 and 8, are functions of the distance between the center of the ellipsoid (which is coincident with the reference point for the set of N atoms), the center of the test sphere, and the orientation of the ellipsoid (and associated N atoms) with respect to the sphere. We begin by comparing the two potential energies, eqs 6 and 8, as a function of the orientation of the ellipsoid, $\Omega$, with respect to the sphere. At fixed orientation, two locations on the potential energy surface are particularly useful. The first is the zero-crossing point, where $V_{AA,p} = 0$, and the second is the point at which potential energy is minimized. We identify those two unique places as $R_{A,p} = R_1(\Omega)$ and $R_{A,p} = R_2(\Omega)$, respectively. We require that $V_{A,p}$ be equal to the AA potential at these two points. Therefore

$$\left(\frac{w_A(\Omega) + w_p}{R_1(\Omega) - \delta}\right)^6 = 1 \quad \text{and} \quad \left(\frac{w_A(\Omega) + w_p}{R_2(\Omega) - \delta}\right)^6 = \frac{1}{2} \quad (9)$$

$\delta = \delta(\Omega)$ and does not depend on the distance between the centers and thus is the same at the two positions we are considering. Thus,

$$w_A(\Omega) = \frac{(R_2(\Omega) - R_1(\Omega))}{(2^{1/6} - 1)} - w_p$$

. This provides a width term for the potential well that depends on the orientation of the ellipsoid. In general, and perhaps surprisingly, the width term is not strongly dependent on the orientation, which is one reason why the GB form is a particularly robust functional form. Therefore, a single parameter $w_A$ may be obtained as the average over all ellipsoidal orientations:



DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

$$w_A = \frac{\int w_A(\Omega)\, d\Omega}{\int d\Omega} \tag{10}$$

Of the systems we have examined, the relative standard error of the width term, computed over all orientations, is about 10%. The estimate to the width adjustment constant, $w_A$, is determined only from the all atom potential; no other feature of the GB functional form impacts this choice. The width adjustment term, $w_A$, is also dependent on the size of the test sphere. This dependence on the size of the test sphere is small, as shown in Figure S.4 of the Supporting Information, and hence typically ignored. For practical purposes, $w_A$ is determined using a test sphere with a LJ radius corresponding to a carbon atom.

Using a constant $w_A$, determined by eq 10, in eq 6, gives a numerator, $\sigma^0$, that is a constant, which is common among GB potentials. This now leaves the energy parameter, $\sqrt{\varepsilon_A}$, the final remaining parameter to be determined. With a constant $w_A$ eq 6 becomes

$$V_{A,p}^{LoD}(R, \Omega) = \sqrt{\varepsilon_p \varepsilon_A(R, \Omega)}\, \nu_{A,p}^{LoD} \tag{11A}$$

with either

$$\nu_{A,p}^{LoD} = \nu_{A,p}^{CF} = 4\{F_{A,p}{}^{-6} - F_{A,p}{}^{-3}\} \tag{11B}$$

or

$$\nu_{A,p}^{LoD} = \nu_{A,p}^{AW} = 4\left\{\left(\frac{w_A + w_p}{R_{Ap} - \delta}\right)^{12} - \left(\frac{w_A + w_p}{R_{Ap} - \delta}\right)^{6}\right\} \tag{11C}$$

Equation 11B uses the contact function of eq 2, and eq 11C uses the adjusted width terms of eq 6. The optimal energy parameter, $\varepsilon_A(R,\Omega)$, which is the only remaining term that is a function of R and Ω, is found by setting the two potentials from eqs 8 and 11A equal:

$$\sqrt{\varepsilon_p \varepsilon_A(R, \Omega)} = \frac{V_{AA,p}}{\nu_{A,p}^{LoD}} \tag{12}$$

A constant $w_A$ in eqs 11 may appear to place undue burden on the energy parameter to track, in a sensible fashion, the functional dependence of the underlying AA force field. Equation 12 shows that the optimal energy, $\varepsilon_A(R,\Omega)$, depends on the AA force field. We will demonstrate in what follows, that this dependence can be well approximated by the product of a best average energy and a simple orientation dependent term. Of the many strategies we considered and tested, to determine a functional form of $\varepsilon_A(R,\Omega)$, which suitably followed the AA potential, the best choice started with a Boltzmann-weighted average energy parameter. The average energy, eq 13, is obtained by averaging over the radial distance over the range of $R \geq R_m = \min\{R_1(\Omega)\}$ for the test sphere placed radially from the center of the ellipsoid. The resulting energy parameter (after averaging over the radial distance) is then reduced to a single value by averaging over all possible orientations.

$$\sqrt{\varepsilon_p \varepsilon_A^0} = \frac{\int_\Omega \int_{R=R_m}^{R_{max}} \sqrt{\varepsilon_p \varepsilon_A(R, \Omega)}\, \omega(R, \Omega) R^2\, dR\, d\Omega}{\int_\Omega \int_{R=R_m}^{R_{max}} \omega(R, \Omega) R^2\, dR\, d\Omega} \tag{13}$$

Thus, a single optimal effective energy, $\varepsilon_A^0$, is determined (analogously to the adjusted width). The weighting function, $\omega(R,\Omega)$, in eq 13 may subsume spherical coordinates, allows for a quadrature or uneven spacing of points on $R$ and improves convergence at large $R_{max}$. It is crucial for the weighting function to represent the CG potential as a self-consistent, average representation of the underlying AA potential. After trying many different weighting functions,

$$\omega(R, \Omega) = e^{-\beta \sqrt{\varepsilon_p \varepsilon_A^0}\, \nu_{A,p}^{LoD}(R,\Omega)}$$

was found to be the best choice. The weights depend on the averaged CG potential energies, and therefore the integral must be done iteratively (usually a few times) until the average energy converges to a consistent quantity. Further note that in the weighting function $\nu_{A,p}^{LoD}$ is either $\nu_{A,p}^{CF}$ from eq 11B, if one uses the simple contact function of eq 2, or $\nu_{A,p}^{GB}$ from eq 11C if the GB type potential of eq 6 is used. The upper range on the radial integration is set to a sufficiently large value that it does not substantially affect the result: $R_{max} \approx 14 \cdot R_m$ was found to be adequate.

While the size of the test sphere does not significantly affect the value of $w_A$, see Figure S.4 in the Supporting Information, the size of the test sphere dramatically changes the energy, $\varepsilon_A^0$, see Figure 2. The effective energy from the LoD model

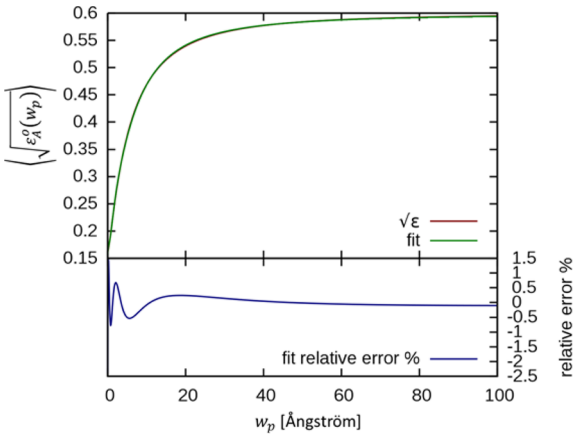



**Figure 2.** Average square root of the effective Lennard-Jones potential well depth, eq 13, is shown as a function of the interacting Lennard-Jones sphere of radius $w_p$, along with the fit to the curve using eq 14, for benzene.

(computed from eq 13) needed to match the underlying potential as a function of the test sphere radius is shown in Figure 2; the effective energy increases with increasing sphere radius, and eventually limits asymptotically to the sum of the individual LJ energies in the AA force field ($\lim_{r_p \to \infty} \sqrt{\varepsilon_A^0} = \sum_i \sqrt{\varepsilon_i}$, see Supporting Information, section S.C, and Figure S.4). Therefore, $\varepsilon_A^0$ is a function of the size of the test sphere: $\varepsilon_A^0 = \varepsilon_A^0(w_p)$. While an exact, analytic form of $(\varepsilon_A^0(w_p))^{1/2}$ is elusive, a good fit, by least-squares adjusting the six fit parameters, $\alpha_l$, can be obtained using the following function:

$$\sqrt{\varepsilon_A^0(w_p)} \approx \frac{[\sum_{l=1}^{6} \alpha_l w_p^{l-1}] + w_p{}^6 \sum_i \sqrt{\varepsilon_i}}{(w_p + w_A)^6} \tag{14}$$

The optimum well depth, as given by the energy parameter, $\varepsilon_A(R,\Omega)$, is typically a strong function of the orientation of the


DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

ellipsoid with respect to the test sphere. The Lennard-Jones energy between two objects depends on the amount of surface contact.[44,46] The greater the surface, the greater the attraction energy becomes. The surface area then varies with the orientation of the ellipsoid. Therefore, we suggest that the LJ energy parameter is dependent on an interaction area factor, $\eta_{IA}^{A}(p)$:

$$\eta_{IA}^{A}(B) = \frac{s_A(B)}{\langle s_A \rangle_\Omega} \tag{15}$$

Here, $s_A(B)$ describes the effective surface area of A seen by object B and $\langle s_A \rangle_\Omega$ is the surface area averaged over all orientations. Initially object B is the test sphere. As a computationally efficient approximation to $s_A(B)$ we use the surface area of the ellipse obtained by slicing through the center of ellipsoid A normal to the vector between the centers of A and B:[51]

$$s_A(B) = \frac{\pi abc}{\sqrt{a^2 n_x^{\,2} + b^2 n_y^{\,2} + c^2 n_z^{\,2}}} \tag{16}$$

Here, $a$, $b$, and $c$ are the semiaxes of ellipsoid A and $n_{x,y,z}$ are the components of the unit-vector, $\hat{n}$, between the centers of A and B, and

$$\hat{n} = \mathcal{R}_A^T \frac{(\vec{r}_B - \vec{r}_A)}{\| \vec{r}_B - \vec{r}_A \|} \tag{17}$$

where $\mathcal{R}_A^T$ is the transposed rotation matrix of A. The orientation dependence of the energy parameters expressed with the interaction area factor, $\eta_{IA}^{A}(p)$, for A interacting with the test sphere is

$$\sqrt{\varepsilon_A^{IA}(p)} = \eta_{IA}^{A}(p)\sqrt{\varepsilon_A^0(w_p)} \tag{18}$$

If ellipsoid A is itself a sphere, then the interaction area factor, $\eta_{IA}^{A}(p) = 1$. Equation 18 provides a replacement formula for $(\varepsilon_A(R,\Omega))^{1/2}$ in eq 11A. With the substitution of eq 18, into 11A, the calculation of $V_{A,p}^{LoD}(R,\Omega)$ no longer explicitly requires the AA force field.

Figure 3 shows Lennard-Jones potential energy $V_{LJ}$ for a single-ellipsoid CG model of benzene interacting with a simple test sphere, compared with potential energy of the underlying all-atom model interacting with the same test sphere. The parameters of the LoD were fitted to the all-atom potential as described above. Specifically, eq 11A was used, with the values for the LJ energy term from eq 18, and the adjusted width from eq 10. The interaction was calculated for three different positions of the test sphere around the ellipsoid, corresponding to the three Cartesian axes, where the $z$ axis is normal to the plane of the molecule; $y$ is nearly coincident with a CH bond; and $x$ and $y$ are perpendicular and in the plane of the benzene. The $d_{6h}$ symmetry of benzene is lost in the gyration tensor because two eigenvalues are degenerate, resulting in identical potentials along the $x$ and $y$ axes. For the fully atomistic model, the $x$-axis is not coincident with a CH bond, leading to differing projections along the two axes. Similar results are obtained for perfluorobenzene.

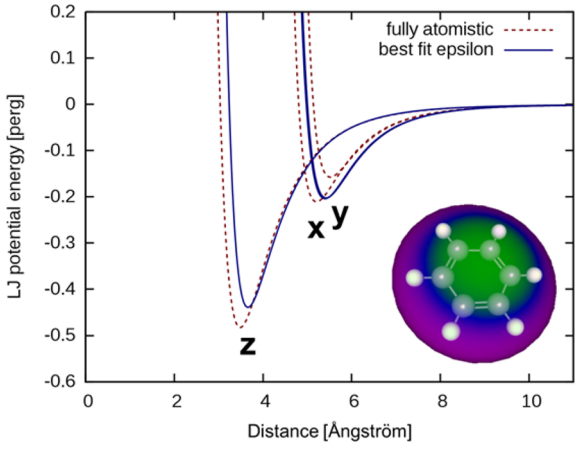



**Figure 3.** Interaction potential of fully atomistic model (red, dashed) and LoD model (blue solid) for benzene with a characteristic test sphere ($w_p$ = 1.58 Å) for three different orientations of benzene using eq 11C and 18

We now adapt the equations, 10, 11, and 18, previously developed for the interaction of an ellipsoid with a test sphere, to the interaction between a pair of ellipsoids, A and B. The interaction strength is calculated through a generalization of the Lorentz−Berthelot combination rules for Lennard-Jones type potentials;[9,11,42,52,53] in which the well-depth parameter for the interaction between a pair of ellipsoids of different types is equal to the geometric mean of the well depths for the two interacting ellipsoids, $\varepsilon_{AB} = \sqrt{\varepsilon_A \varepsilon_B}$. Therefore, we generalize eq 18 for the interaction of two ellipsoids $\varepsilon_{AB}$ as

$$\varepsilon_{AB} = \sqrt{\varepsilon_A^{IA}(B)\varepsilon_B^{IA}(A)} = \eta_{IA}^{A}(B)\eta_{IA}^{B}(A)\sqrt{\varepsilon_A^0(w_B)\varepsilon_B^0(w_A)} \tag{19}$$

In this way, each ellipsoid has an energy determined by the interaction with an effective sphere, and then the effective width of the other ellipsoid replaces the sphere width. The product of the two interaction area factors, $\eta_{IA}^{A}(B)\eta_{IA}^{B}(A)$, called the joint interaction area factor, depends only on the unit vector between the two ellipsoids and their semiaxes; by depending on the relative orientation of the two ellipsoids, eq 19 provides a way to mimic the AA pairwise LJ interaction. The average energy, $\varepsilon_A^0(w_B)$, for ellipsoid A depends on the average size of ellipsoid B but not the relative orientation. The relative orientation is accounted for in the joint interaction area term: The best average energy should be represented by an averaged value. The average energy is then corrected by the joint interaction area depending on the orientation of the respective partner ellipsoid. If the partners are both spheres all correction factors are unity and the typical combination rules apply.

In summary, there are three methods in which the AA force field can be reproduced, with varying levels of fidelity: (Method 1) The simplest way is with the contact function substituting for the ratio of the contact term to the distance between centers, (i.e., eq 2 replaces eq 6) using a single (orientation independent) LJ energy parameter, obtained from eq 13. Such an LJ energy parameter, $\varepsilon_A^0(p)$, depends on the size of the test sphere. The average radius of the partner sphere then is found from $w_B$ of eq 10 and the energy then is $\varepsilon_A^0(w_B)$. Thus, the pairwise energy parameter, $\varepsilon_{AB}$, is a constant derived from the underlying AA force field. (Method 2) The next level of complexity is to include the joint interaction area factor, which includes both ellipsoids, eq 19. This gives the simplest orientation dependence to $\varepsilon_{AB}$ in eq 2. (Method 3) The most detailed method (which takes full advantage of the form of eq 1) uses eq 5 with the adjusted widths (found using eq 10) and the energy parameter (determined by eq 19). In the


DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

Results and Discussion section, we will demonstrate the successive accuracy of these three methods by comparing the correlation with the exact AA simulations and the three different methods for using the GB-type force field at various LoD levels for three different systems.

## RESULTS AND DISCUSSION

We now demonstrate the CG modeling approach developed in the previous section as applied to three different test systems. The fidelity of the various CG models is determined by comparison of the radial distribution functions, $g(r)$ for an NPT ensemble of molecules at the AA level with the various LoD levels and the various methods. The objective of the comparisons is to reproduce the results of the AA force field, and not necessarily to obtain accurate experimental results for the systems considered here. Details of the simulations, parameters used, and some additional results, including $g_\mu(r)$, are provided in the Supporting Information, section S.F.

**Benzene and Perfluorobenzene.** The first system to consider is the interaction of benzene with perfluorobenzene. Figure 4 shows how one benzene interacts with a second benzene as a function of mutual orientation. The side−side interaction energy along the $x$ axis is slightly different from that on $y$ axis because the two hydrogens are aligned with the $x$ axis, whereas the interaction along the $y$ axis involves four hydrogens. The fit is the coarse grained, single ellipsoid LoD using eq 5 with the joint interaction area (eq 19) and the offset parameters (eq 10). Figure 4 is analogous to Figure 3, but now two benzenes interact rather than a single benzene with a calibration test sphere.

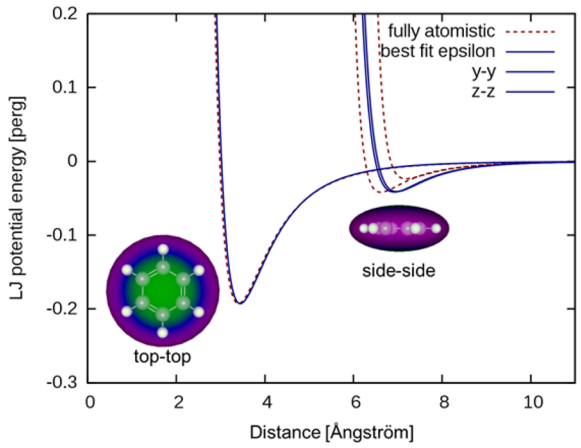



**Figure 4.** Top−top and side−side LJ interaction potential of fully atomistic model (red, dashed) and the single ellipsoid LOD model (blue solid) for the interaction of two benzene molecules using eqs 5 and 19.

We now consider the interaction of a 1:1 mixture of benzene and perfluorobenzene: The system consists of 108 molecules of each type; simulations were run in the NPT ensemble at 298 K, and 1 bar. The $g(r)$ of pure benzene is not shown because it is not sensitive to the parameters that constitute the LoD; and many different types of LoDs that were tried gave results very similar to those of the AA simulations. The pair correlation between the centers of benzene and perfluorobenzene in the 1:1 material proved to be much more sensitive to the values of the parameters used in the LJ potential than that of the benzene-to-benzene pair correlation function. Figure 5 shows the radial distribution function between benzene and perfluorobenzene for the different methods described above to simulate the pairwise LJ potential. Identical electrostatic

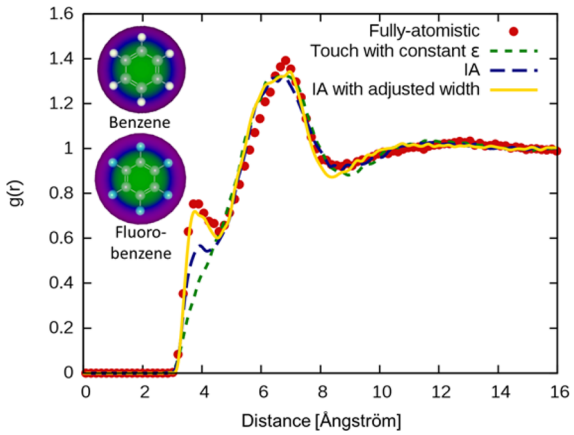



**Figure 5.** Radial distribution function $g(r)$ between benzene and perfluorobenzene centers in simulations run with single ellipsoid LoD using (1) the contact function with constant LJ $\varepsilon$ (green dashed line, method 1), (2) the joint interaction area, IA, correction to the LJ $\varepsilon$ (blue dashed line, method 2), and (3) IA and adjusted width (solid gold line, method 3). The AA simulations are also shown (red dots) for reference.

**Table 1. AA Force Field Lennard-Jones Parameters Used in This Work and Infinite Frequency Dielectric $n^2$ and the Temperature Used in the NPT Ensemble MC Simulations[a]**

| molecule system (DFT method) | atom | LJ radius [Å] | LJ energy [kcal/mol] | $n^2$ | temp [K] |
|---|---|---|---|---|---|
| benzene/hexafluorobenzene (B3LYP/6-31G(d)) | C | 1.65 | 0.066 | 2.25 | 293 |
| | H | 1.25 | 0.015 | | |
| | F | 1.42 | 0.061 | | |
| ethylene carbonate (B3LYP/6-311G(d,p)) | C | 1.75 | 0.066 | 2.0164 | 313 |
| | H | 1.21 | 0.015 | | |
| | O | 1.5 | 0.17 | | |
| | C (C═O) | 1.875 | 0.105 | | |
| | O (C═O) | 1.48 | 0.21 | | |
| hydrocarbon chain (B3LYP/6-31G(d)) | C | 1.65 | 0.066 | 2.89 | 298 |
| | H | 1.25 | 0.015 | | |

[a]Molecular geometries and partial charges were obtained from DFT calculations in vacuum at the B3LYP/6-31G(d) level of theory; see Supporting Information for details.



DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

representations are used for all models, with point charges centered on each atom. The charges on the fluorine atoms are nearly the same as those on the hydrogens, but with opposite sign. This charge reversal gives rise to the first peak in $g(r)$ and arises from the "top to top" interactions (in the "$z$" direction of each molecule, as defined in Figure S.2) between benzene and perfluorobenzene (see the Supporting Information, section S.F. for a plot of the orientational distribution functions $g_\mu(r)$) and $g_\mu^2(r)$.

**Flexible Hydrocarbon Chain.** The treatment of internal dynamics and conformational flexibility by CG methods is an important application and extension of the LoD approach. We use a saturated alkyl chain with 32 carbon atoms, shown in Figure 6, as a straightforward test of the internal LJ interactions in a flexible system.

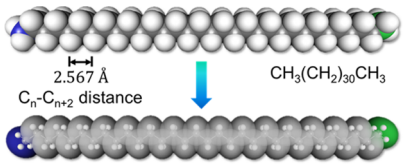



**Figure 6.** (Top) Space filling AA model of a 32mer hydrocarbon chain in the transoid configuration. (Bottom) Converted to the LoD model in which a single $CH_2$ unit is contained within an ellipsoid. The two end $CH_3$ units are treated separately, and are similarly contained in a single ellipsoid. The structure and charges were determined by DFT using B3LYP/6-31G(d).

Alkyl chains are widely used as flexible linkers in large molecular systems, and are a good model for entropically driven molecular components. The reference system is the all trans configuration. Only CC bond rotations (altering the internal dihedral angles) are allowed in our simulations. Bond distances and $\overline{CCC}$ bond angles are held constant. The AA system uses OPLS-AA Lennard-Jones energy parameters, and charges (centered at the atoms) from DFT calculations (see Supporting Information). Free bond rotations around the carbon−carbon bonds are allowed. We compared this reference system with three different CG models. Figure 7 shows $g(r)$ for the three different models for the case where the ellipsoids are used to replace a single $CH_2$ unit. The two end $CH_3$ units are treated with a differently parametrized ellipsoid. The LJ interactions were not used for second or third nearest neighbors; a method often used in similar systems at the AA LoD. Calculations were performed on individual molecules; only the intramolecular moves are of importance.

Figure 7 shows the histogram of end-to-end distances for the hydrocarbon chain. The bond distance between adjacent carbons and the $\overline{CCC}$ bond angles, obtained from DFT geometry, are maintained at $l$ = 1.534 Å and $\theta$ = 113.6°; only rotations around bonds are allowed. The theoretical expectation of the mean square end-to-end distance, $R$, for a freely jointed chain[45] is $\langle R^2 \rangle = Nl^2$, which predicts a RMS distance of 8.7 Å. The standard width of the distribution is predicted to be about 3.5 Å. A more relevant model for our simulations is the freely rotating chain model, in which the angles are fixed but the bonds are free to rotate.[54] The freely rotating chain model predicts a RMS distance of 13.2 Å. In the present simulations, in addition to free bond rotations, electrostatic interactions and steric restrictions imposed by LJ interactions are included. Therefore, the present simulations are considered to be a hindered rotating chain. Consequently, the end-to-end distance

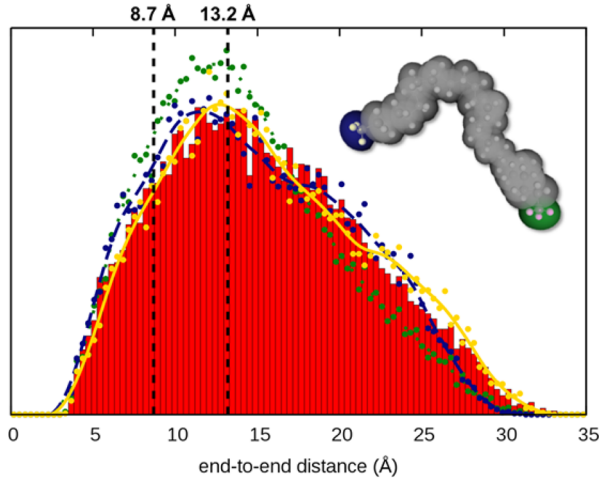



**Figure 7.** End-to-end distance histograms between terminal carbon centers for a fully atomistic reference (red bars) and the LoD simulations, in which a single $CH_2$ unit is replaced with a single ellipsoid, with the three different methods (see Methods section). The "simple contact function" LJ potential (eq 2), with a constant LJ energy parameter (eq 13, green dots and dotted line); with the joint interaction area (eq 19, blue dots and dashed line) included in the LJ energy parameter; and with both IA and adjusted width included (eq 5, gold dots and solid line). The lines are present only to aid the eye. The RMS expectation values for the freely jointed and freely rotating chain models are indicated with vertical (black dashed) lines.

distribution is expected to be skewed toward longer end-to-end distances relative to the freely jointed chain.[54]

The various LoD models are constructed from (a) single ellipsoids around the two $CH_3$ end-groups groups and (b) from the 30 $CH_2$ repeat units grouped from 1 to 4 $CH_2$ units into a single ellipsoid. The partial charges in each ellipsoid are placed on the atom centers, as they were in the underlying AA model. LoD Lennard-Jones and electrostatic interactions are calculated from the second-nearest repeat units and up. For the AA model fourth nearest atom interactions and up were included in the calculation, similar to what has been done elsewhere,[3,4] matching the comparable interaction distances in the LoD models.

In the above simulation, only a single $CH_2$ unit was grouped into an ellipsoid. In principle, any number of such units can be grouped into a single ellipsoid, for a different LoD. Therefore, we also demonstrate the distributions that arise when one through four $CH_2$ units are grouped into a single ellipsoid. Figure 8 shows the distribution for the full LoD model compared to the AA model for a different number of $CH_2$ units grouped together. For the case where three $CH_2$ units were grouped into a single ellipsoid (making 10 ellipsoids, plus 1 on either end) nothing further was done for the resulting fit.

For the even numbered grouping of $CH_2$ repeat units (2 and 4) bonds connecting the ellipsoids are transoid, and therefore, rotation around such bonds will not significantly shorten the chain. Therefore, additional measures had to be taken to achieve a modicum of similarity with the underlying AA model. The bonds connecting the ellipsoids then were allowed to bend freely. The bending is hinged at the atoms from the underlying model on both ends of the linkage within each ellipsoid. To disallow neighboring units from folding on themselves and to maintain the minimal stiffness required values of the calculated LJ energy parameters for nearest neighbor interactions were determined from the LJ energy parameters but scaled by 0.1%


DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

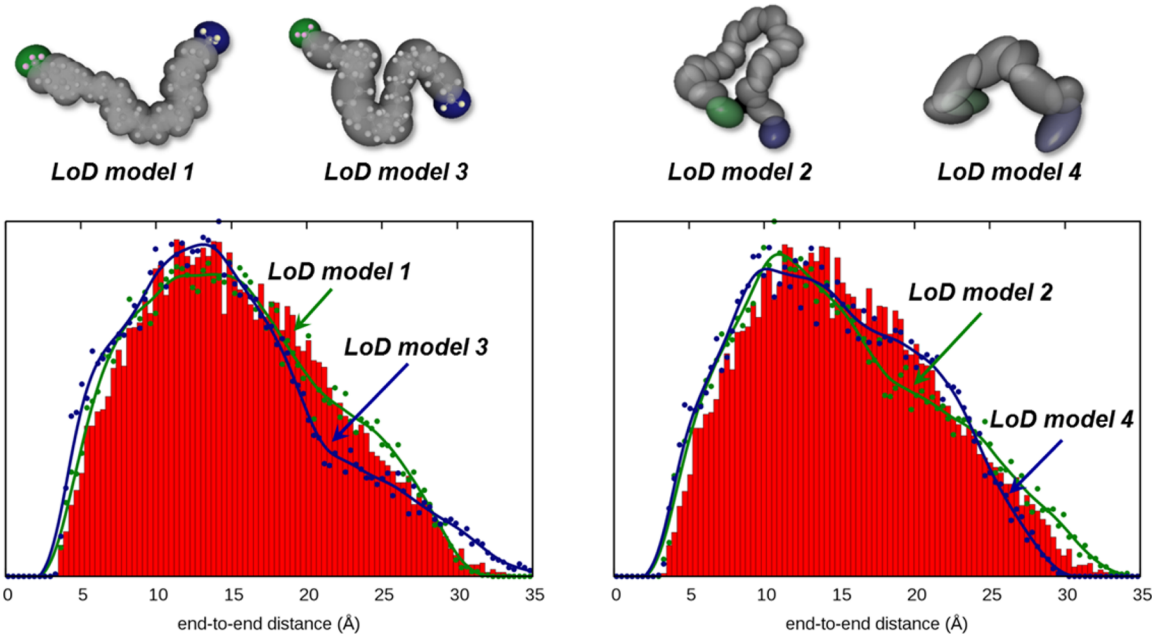



**Figure 8.** End-to-end distance histograms between the two terminal carbon centers (as the blue and green ellipsoids on the representative chains, top of figure) for the AA (red bars). (Left) The LoD simulations grouping 1 (green, same as Figure 7) and 3 (blue) $CH_2$ repeat units into a single ellipsoid. Right: The LoD simulations grouping 2 (green) and 4 (blue) repeat units into a single ellipsoid. All LoD simulations use both the IA and the adjusted widths.

and 5% for the 2 and 4 groupings of $CH_2$ units, respectively. The bending moves augmented the same rules as allowed for the other models.

**Ethylene Carbonate.** Ethylene carbonate was chosen for testing the LoD model due to its high dipole density yielding unusually strong short-range interactions. Ethylene carbonate has the largest dielectric constant (90 at 40 °C) of common organic liquids. Finding the proper interactions and local structure in such a material represents a severe challenge to any CG method, requiring proper treatment of both electrostatics and dispersion interactions.

One LoD model for ethylene carbonate was created by decomposing the molecule into two subunits: one for the carbonate group and one for the ethyl backbone; parameters were generated using the rules specified above. The two ellipsoids within each molecule are heavily overlapped and rigidly locked together. Therefore, their interactions are not computed and have no effect on the simulation. The charges and structure of ethylene carbonate are determined by DFT and the residual charges are kept fixed at the atoms. We previously evaluated a model in which the charge distribution was reduced to a point charge and a dipole at the center of the ellipsoids; these results are similar to the current model using AA charges.[30] LJ energy parameters were obtained from the OPLSA-AA force field (given in section S.F of the Supporting Information). The structure and charges were determined by B3LYP/6-311+G(d,p) DFT, with the PCM method using the dielectric of ethylene carbonate. The charges remained fixed on the atoms in all simulations to demonstrate the differences among the different LoD models for the LJ potentials.

Table 2 lists the simulation results for the individual ethylene carbonate simulations shown in Figure 9 (see the Supporting Information for a plot of the orientational distribution function $g_\mu(r)$). The properties determined from the fully atomistic simulation are very close to the experimental values of 1.321 g/cc for the density, 134 J/(mol K)[55] for the heat capacity,[56] and 90 for the dielectric constant.[55]

**Table 2. Ethylene Carbonate Simulation Results**[a]

| model | density [g/cc] | $C_p$ [J/(mol K)] | static dielectric constant |
|---|---|---|---|
| AA model | 1.265 ± 0.001 | 123 ± 7 | 91 ± 35 |
| single ellipsoid LoD | 1.274 ± 0.001 | 115 ± 3 | 103 ± 32 |
| two-ellipsoid LoD | 1.212 ± 0.001 | 119 ± 5 | 89 ± 33 |

[a]Results are averages of eight individual simulations with associated standard deviations using the NPT ensemble under 1 atm at 40 °C with the models as described in the text (see Supporting Information for model details). Note that $(12/2)k_B$ was added to heat capacities in order to account for internal degrees of freedom.

The two-ellipsoid and three-ellipsoid (shown in section S.F of the Supporting Information) models fit the AA model extremely well and give excellent estimates to the density and the static dielectric constant, computed from fluctuations in the system's net dipole at zero applied field.[29,47,57] The two-ellipsoid LoD has concave regions in its vdW surface and correctly represents end-to-end and side-to-side stacking as separate peaks at 4 and 6 Å, consistent with the AA model. The orientation dependent $g_\mu(r)$ (section S.F, Figure S.7) also illustrates this reorientation. It is perhaps surprising that the single ellipsoid (a fully convex surface) provides an averaged interaction and an even larger dielectric constant. If and when the high resolution in detail is not needed, (such as second nearest neighbor interactions and beyond) then the simpler LoD models may be used.

**Large Prolate Chromophores.** As a final demonstration to underscore the importance of the interaction area, we show different morphologies that follow from considering the interaction area as a correction to the LJ energy parameters, and the adjusted width correction to the Lennard-Jones form, for molecules with highly elongated shapes. ONLO chromo-


DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

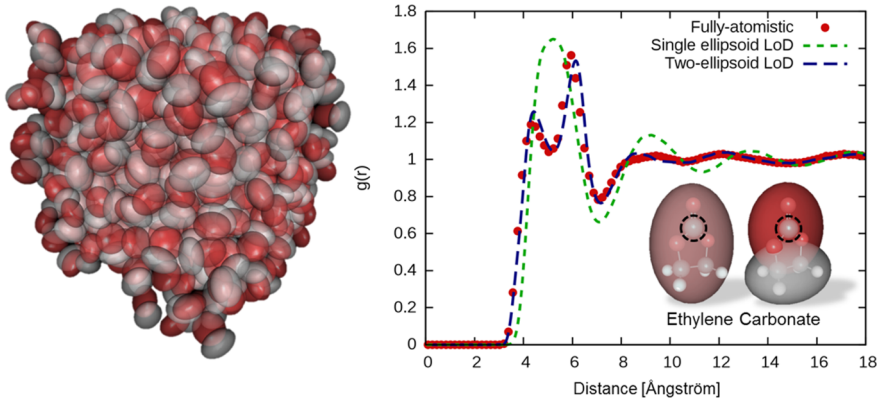



**Figure 9.** Simulation snapshot of 432 ethylene carbonates with radial distribution function resulting from an NPT ensemble MC simulation. Comparison among the AA (red dots), single ellipsoid LoD (dashed green line), and two-ellipsoid LoD (dashed blue line) simulations using the contact function with IA and adjusted widths, model 3.

phores, such the CLD and FTC families of dyes used for electro-optic switching, typically possess a highly prolate geometry as well as a very large dipole moment.[58−60] Much of our prior work[29,47] has involved simulating such chromophores using single-ellipsoid models. We have demonstrated the types of packing and order that result from such ellipsoids and provided reasonable estimates to the acentric order developed by such ellipsoids under a poling field. Now we compare the morphology of the packing of representative ellipsoids when using different methods of describing the LJ interaction for ellipsoids with a single dipole in the center (Figure 10). When an external poling field is present the

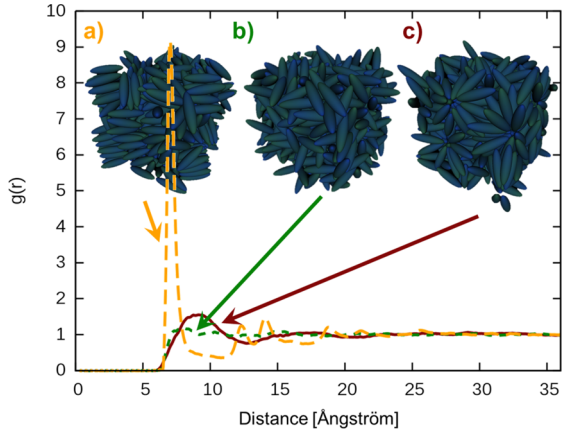



**Figure 10.** Simulation comparison using system snapshot and radial distribution function g(r) using (a) both the IA and adjusted width corrections to the LJ potential, as Method 3, (b) only the adjusted width to the LJ potential but a constant, best fit LJ energy term, $\varepsilon$, and (c) simple touch LJ potential with constant, best fit LJ $\varepsilon$, as Method 1.

dipoles of the chromophores induce partial order. The radial distribution functions and snapshots under equilibrium demonstrate the markedly different morphologies that result from including the interaction area correction and adjusted widths. We have demonstrated for small molecules that the joint interaction area corrects the simple LJ energy parameters and makes the CG representation model the AA interaction for the simpler systems above. However, the IA correction becomes increasingly important with increased molecular aspect ratio.

Specifically, the IA correction of eq 19 makes the side−side interaction more important than the end−end one. This removes the micelle-like, florid arrangements of the ellipsoids seen when the IA correction is not used. As we have demonstrated in smaller molecules by comparing to the AA structures that the IA is necessary, we conclude that the structures simulated with the IA are more realistic representations of larger molecules with higher eccentricities.

We have also included the adjustable width in the full LJ interaction potential. However, for systems of nearly comparable sizes the adjustable width causes little change in the LJ interaction. The use of just the IA with no adjusted width (Method 2, not shown) gives a result nearly identical to that of the simulation including the adjusted width. The adjustable width is most necessary for interactions between molecules of greatly differing sizes, such as a large ONLO chromophore interacting with an explicit solvent.

## SUMMARY AND CONCLUSIONS

The goal of any coarse-graining approach is to reproduce the simulation results of the original AA description it replaces. We have shown that in addition to typical simulation properties such as densities, dielectric constants, and system energies, radial and orientational correlation functions can be reproduced very well at various levels of coarse-graining by a systematic approach that builds up ellipsoidal CG representation using spheres as probes to transform an AA representation of a molecule (or molecular subunit) into a CG representation. The interactions between ellipsoids are then evaluated using combination rules similar to those of established, transferrable force fields. This "pair wise, bottom up" approach avoids the need to optimize the parameters by simulating the potential of mean force for complex systems. We note that the bottom-up approach is very much like the experimental approach to determining molecular structure, in which a molecule is probed by a test molecule, such as the oxygen atom in a single CO molecule.[61,62] Simulation results demonstrate that CG matching at the pairwise level can lead to excellent results at the many body level.

The use of ellipsoids provides a powerful and flexible route to coarse-grained models of complex molecules. The computational cost has been considered by others to be a primary reason for avoiding such models. We show (see Supporting Information, section S.E, Figure S.5) that our implementation of an algorithm (see section S.D, using eqs SD.15) to compute


DOI: 10.1021/acs.jctc.6b00219
J. Chem. Theory Comput. 2016, 12, 4362−4374

the contact function at most costs a factor of 5 in computational time compared to models using spheres. The performance penalty per subunit is more than made up for by bundling 2 to 3 atoms into a single ellipsoid as the calculations depend on the square of the number of entities.[30] In the case of benzene, a 12-atom molecule can be replaced with a single ellipsoid without significant loss in accuracy, (as shown in Figure 5) resulting in a net speed increase of over 37 fold compared to the Lennard-Jones summation of the underlying AA model. Even including AA charges inside the single ellipsoid LoD representation, the overall calculation time (as a relative time using our custom MC codes) is reduced by an order of magnitude. (Details of the relative timing of the LJ part of the simulations are given in section S.E)

The steric envelope of benzene is well mimicked by a single ellipsoid and an examination of the electron density of benzene argues strongly that the ellipsoid may be fundamentally a better model for the overall shape than the 3-sphere structure from the MARTINI force field. Of other simple shapes, perhaps, only a superellipsoid would better simulate the curvature of the molecule on its outer edges. One drawback to using spheres is that the spheres need to be of similar sizes otherwise void spaces in between connected spheres could allow penetration by adjacent molecules. For example, the LoD of the hydrocarbon chains using ellipsoids regardless of the number of units bundled into one ellipsoid have inherently less void space than an all sphere basis and hence can encompass a larger variety and volume of molecular subunits. Moreover, because the ellipsoids offer a wide array of shapes to correspond to the underlying structures, ellipsoid-based CG potentials have a built-in anisotropy that follows the underlying structure. Ellipsoids can represent the shape of the molecule without retaining atomistic details.

The LoD model allows for closely representing the Lennard-Jones interactions for a molecule with a reduced number of nonbonded interactions. The LoD model developed here produced results consistent with the underlying AA force field for a variety of systems and properties spanning the basic units needed for complex organic molecules such as ONLO chromophores. The model has been validated for aromatic systems (benzene/perfluorobenzene), aliphatic systems (hydrocarbon chain), and finally, strongly polar organic liquids (ethylene carbonate). The LoD model performs well for all of these types of systems, and provides a systematic, bottom-up approach to CG models. The use of ellipsoidal shapes rather than spheres allows for greater flexibility in model design. The models are automatically generated from first-principles and generate results that are highly consistent with the underlying AA force field. Moreover, the present work sets the stage to implement the primary feature of LoD, which is to use different levels of detail for the same underlying structure in a single simulation depending the interaction distances between pairs of molecules.

## ■ ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.6b00219.

Full simulation Hamiltonian (section S.A), mathematical framework to calculate the ellipsoid shape (section S.B), the dependence of LoD parameters on the size of the test sphere (section S.C), the contact function algorithm (section S.D), relative computational cost of ellipsoids vs spheres (section S.E), and the set of model parameters used (section S.F) (PDF)

## ■ AUTHOR INFORMATION


### Corresponding Author

*E-mail: robinson@chem.washington.edu. Tel.: +1 (206) 543-1773.



### Funding

We gratefully acknowledge the financial support of the National Science Foundation (Grant No. DMR-1303080); and the Air Force Office of Scientific Research (FA9550-10-1-0558 and FA9550-15-1-0319).


### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS


Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank Drs. Lutz Maibaum and Larry Dalton for their advice and enthusiasm for this project.


## ■ REFERENCES


(1) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 490−519.

(2) Jorgensen, W. L.; Tirado-Rives, J. The OPLS [optimized Potentials for Liquid Simulations] Potential Functions for Proteins, Energy Minimizations for Crystals of Cyclic Peptides and Crambin. *J. Am. Chem. Soc.* **1988**, *110* (6), 1657−1666.

(3) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.

(4) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114* (8), 2549−2564.

(5) Kaminski, G.; Jorgensen, W. L. Performance of the AMBER94, MMFF94, and OPLS-AA Force Fields for Modeling Organic Liquids. *J. Phys. Chem.* **1996**, *100* (46), 18010−18013.

(6) Peter, C.; Kremer, K. Multiscale Simulation of Soft Matter Systems − from the Atomistic to the Coarse-Grained Level and Back. *Soft Matter* **2009**, *5* (22), 4357.

(7) Noid, W. G. Perspective: Coarse-Grained Models for Biomolecular Systems. *J. Chem. Phys.* **2013**, *139* (9), 090901.

(8) Bondi, A. Van Der Waals Volumes and Radii. *J. Phys. Chem.* **1964**, *68* (3), 441−451.

(9) Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids; Clarendon Press*; Oxford University Press: Oxford, England; New York, 1989.

(10) Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: Orlando, FL, 2001.

(11) Leach, A. R. *Molecular Modelling: Principles and Applications*, 2nd ed.; Prentice Hall: Dorchester, England, 2001.

(12) Dunn, N. J. H.; Noid, W. G. Bottom-up Coarse-Grained Models That Accurately Describe the Structure, Pressure, and Compressibility of Molecular Liquids. *J. Chem. Phys.* **2015**, *143* (24), 243148.

(13) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **2007**, *111* (27), 7812−7824.




DOI: 10.1021/acs.jctc.6b00219
*J. Chem. Theory Comput.* 2016, 12, 4362−4374

(14) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S.-J. The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J. Chem. Theory Comput.* **2008**, *4* (5), 819−834.
(15) de Jong, D. H.; Singh, G.; Bennett, W. F. D.; Arnarez, C.; Wassenaar, T. A.; Schäfer, L. V.; Periole, X.; Tieleman, D. P.; Marrink, S. J. Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J. Chem. Theory Comput.* **2013**, *9* (1), 687−697.
(16) Vögele, M.; Holm, C.; Smiatek, J. Coarse-Grained Simulations of Polyelectrolyte Complexes: MARTINI Models for Poly(styrene Sulfonate) and Poly(diallyldimethylammonium). *J. Chem. Phys.* **2015**, *143* (24), 243151.
(17) Berendsen, H. J. C.; Grigera, J. R.; Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **1987**, *91* (24), 6269−6271.
(18) Smit, B.; Hilbers, P. a. J.; Esselink, K.; Rupert, L. a. M.; van Os, N. M.; Schlijper, A. G. Computer Simulations of a Water/oil Interface in the Presence of Micelles. *Nature* **1990**, *348* (6302), 624−625.
(19) Smit, B.; Esselink, K.; Hilbers, P. A. J.; Van Os, N. M.; Rupert, L. A. M.; Szleifer, I. Computer Simulations of Surfactant Self-Assembly. *Langmuir* **1993**, *9* (1), 9−11.
(20) Shelley, J. C.; Shelley, M. Y. Computer Simulation of Surfactant Solutions. *Curr. Opin. Colloid Interface Sci.* **2000**, *5* (1−2), 101−110.
(21) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. A Coarse Grain Model for Phospholipid Simulations. *J. Phys. Chem. B* **2001**, *105* (19), 4464−4470.
(22) Marrink, S. J.; de Vries, A. H.; Mark, A. E. Coarse Grained Model for Semiquantitative Lipid Simulations. *J. Phys. Chem. B* **2004**, *108* (2), 750−760.
(23) Murtola, T.; Falck, E.; Patra, M.; Karttunen, M.; Vattulainen, I. Coarse-Grained Model for Phospholipid/cholesterol Bilayer. *J. Chem. Phys.* **2004**, *121* (18), 9156−9165.
(24) Venturoli, M.; Maddalena Sperotto, M.; Kranenburg, M.; Smit, B. Mesoscopic Models of Biological Membranes. *Phys. Rep.* **2006**, *437* (1−2), 1−54.
(25) Pasqua, A.; Maibaum, L.; Oster, G.; Fletcher, D. A.; Geissler, P. L. Large-Scale Simulations of Fluctuating Biological Membranes. *J. Chem. Phys.* **2010**, *132* (15), 154107.
(26) MacDermaid, C. M.; Kashyap, H. K.; DeVane, R. H.; Shinoda, W.; Klauda, J. B.; Klein, M. L.; Fiorin, G. Molecular Dynamics Simulations of Cholesterol-Rich Membranes Using a Coarse-Grained Force Field for Cyclic Alkanes. *J. Chem. Phys.* **2015**, *143* (24), 243144.
(27) Robinson, B. H.; Dalton, L. R. Monte Carlo Statistical Mechanical Simulations of the Competition of Intermolecular Electrostatic and Poling-Field Interactions in Defining Macroscopic Electro-Optic Activity for Organic Chromophore/Polymer Materials. *J. Phys. Chem. A* **2000**, *104* (20), 4785−4795.
(28) Rommel, H. L.; Robinson, B. H. Orientation of Electro-Optic Chromophores under Poling Conditions: A Spheroidal Model. *J. Phys. Chem. C* **2007**, *111* (50), 18765−18777.
(29) Johnson, L. E.; Barnes, R.; Draxler, T. W.; Eichinger, B. E.; Robinson, B. H. Dielectric Constants of Simple Liquids: Stockmayer and Ellipsoidal Fluids. *J. Phys. Chem. B* **2010**, *114* (25), 8431−8440.
(30) Tillack, A. F.; Johnson, L. E.; Rawal, M.; Dalton, L. R.; Robinson, B. H. Modeling Chromophore Order: A Guide For Improving EO Performance. In *MRS Online Proc. Libr.*; **2014**; *1698*, 10.1557/opl.2014.795
(31) Berne, B. J.; Pechukas, P. Gaussian Model Potentials for Molecular Interactions. *J. Chem. Phys.* **1972**, *56* (8), 4213−4216. Gay, J. G.; Berne, B. J. Modification of the Overlap Potential to Mimic a Linear Site−Site Potential. *J. Chem. Phys.* **1981**, *74* (6), 3316−3319.
(32) Cao, J.; Berne, B. J. Theory of Polarizable Liquid Crystals: Optical Birefringence. *J. Chem. Phys.* **1993**, *99*, 2213.
(33) Berardi, R.; Orlandi, S.; Zannoni, C. Monte Carlo Simulation of Discotic Gay−Berne Mesogens with Axial Dipole. *J. Chem. Soc., Faraday Trans.* **1997**, *93* (8), 1493−1496.
(34) Houssa, M.; Rull, L. F.; McGrother, S. C. Effect of Dipolar Interactions on the Phase Behavior of the Gay−Berne Liquid Crystal Model. *J. Chem. Phys.* **1998**, *109* (21), 9529.
(35) Varga, S.; Szalai, I.; Liszi, J.; Jackson, G. A Study of Orientational Ordering in a Fluid of Dipolar Gay−Berne Molecules Using Density-Functional Theory. *J. Chem. Phys.* **2002**, *116* (20), 9107. Kushick, J.; Berne, B. J. Computer Simulation of Anisotropic Molecular Fluids. *J. Chem. Phys.* **1976**, *64* (4), 1362−1367.
(36) Berardi, R.; Fava, C.; Zannoni, C. A Generalized Gay-Berne Intermolecular Potential for Biaxial Particles. *Chem. Phys. Lett.* **1995**, *236* (4−5), 462−468. Wallqvist, A. Computer Simulation of Hydrophobic Hydration Forces on Stacked Plates at Short Range. *J. Phys. Chem.* **1995**, *99* (9), 2893−2899.
(37) Fukunaga, H.; Takimoto, J.; Aoyagi, T.; Shoji, T.; Sawa, F.; Doi, M. Parameterization of the Gay-Berne Potential for nCB. *Prog. Theor. Phys. Suppl.* **2000**, *138*, 396−397.
(38) Paramonov, L.; Yaliraki, S. N. The Directional Contact Distance of Two Ellipsoids: Coarse-Grained Potentials for Anisotropic Interactions. *J. Chem. Phys.* **2005**, *123* (19), 194111.
(39) Lee, C. K.; Hua, C. C.; Chen, S. A. Parametrization of the Gay−Berne Potential for Conjugated Oligomer with a High Aspect Ratio. *J. Chem. Phys.* **2010**, *133* (6), 064902.
(40) Perram, J. W.; Wertheim, M. S. Statistical Mechanics of Hard Ellipsoids. I. Overlap Algorithm and the Contact Function. *J. Comput. Phys.* **1985**, *58* (3), 409−416.
(41) Perram, J. W.; Rasmussen, J.; Præstgaard, E.; Lebowitz, J. L. Ellipsoid Contact Potential: Theory and Relation to Overlap Potentials. *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.* **1996**, *54* (6), 6565−6572.
(42) Berthelot, D. Sur Le Mélange Des Gaz. *CR Hebd. Séances Acad. Sci.* **1898**, *126*, 1703−1855.
(43) Tillack, A. F. Electro-Optic Material Design Criteria Derived from Condensed Matter Simulations Using the Level-of-Detail Coarse-Graining Approach. Ph.D. Thesis, University of Washington: Seattle, 2015.
(44) Flory, P. J.; Orwoll, R. A.; Vrij, A. Statistical Thermodynamics of Chain Molecule Liquids. I. An Equation of State for Normal Paraffin Hydrocarbons. *J. Am. Chem. Soc.* **1964**, *86* (17), 3507−3514.
(45) Flory, P. J. *Statistical Mechanics of Chain Molecules*; Wiley Interscience: New York, 1969.
(46) Dill, K. A.; Bromberg, S. *Molecular Driving Forces: Statistical Thermodynamics in Chemistry and Biology*; Garland Science: New York, 2002.
(47) Johnson, L. E. Multi-Scale Modeling of Organic Electro-Optic Materials. Ph.D. Thesis, University of Washington: Seattle, 2012.
(48) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225−11236.
(49) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, N. J.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. *Gaussian 09*; Gaussian, Inc.: Wallingford, CT, USA, 2009.
(50) Breneman, C. M.; Wiberg, K. B. Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. The Need for High Sampling Density in Formamide Conformational Analysis. *J. Comput. Chem.* **1990**, *11* (3), 361−373.
(51) Klein, P. P. On the Ellipsoid and Plane Intersection Equation. *Appl. Math.* **2012**, *03* (11), 1634−1640.


DOI: 10.1021/acs.jctc.6b00219
*J. Chem. Theory Comput.* 2016, 12, 4362−4374

(52) Lorentz, H. A. Ueber Die Anwendung Des Satzes Vom Virial in Der Kinetischen Theorie Der Gase. *Ann. Phys.* **1881**, *248* (1), 127−136.
(53) Rowlinson, J. S.; Swinton, F. L. *Liquids and Liquid Mixtures*; Butterworth Scientific: 1982.
(54) *Physical Properties of Polymers Handbook*; Mark, J. E., Ed.; Springer Science + Business Media: New York, 2007.
(55) *CRC Handbook of Chemistry and Physics*, 96th ed.; Haynes, W. M., Lide, D. R., Bruno, T. J., Eds.; CRC Press: Boca Raton, FL, 2015.
(56) Peppel, W. J. Preparation and Properties of the Alkylene Carbonates. *Ind. Eng. Chem.* **1958**, *50* (5), 767−770.
(57) Kirkwood, J. G. The Dielectric Polarization of Polar Liquids. *J. Chem. Phys.* **1939**, *7* (10), 911−919.
(58) Dalton, L. R. Rational Design of Organic Electro-Optic Materials. *J. Phys.: Condens. Matter* **2003**, *15* (20), R897.
(59) Dalton, L. R.; Sullivan, P. A.; Bale, D. H. Electric Field Poled Organic Electro-Optic Materials: State of the Art and Future Prospects. *Chem. Rev.* **2010**, *110* (1), 25−55.
(60) Kuzyk, M. G.; Pérez-Moreno, J.; Shafei, S. Sum Rules and Scaling in Nonlinear Optics. *Phys. Rep.* **2013**, *529* (4), 297−398.
(61) Gross, L.; Mohn, F.; Moll, N.; Liljeroth, P.; Meyer, G. The Chemical Structure of a Molecule Resolved by Atomic Force Microscopy. *Science* **2009**, *325* (5944), 1110−1114.
(62) Gross, L.; Mohn, F.; Moll, N.; Schuler, B.; Criado, A.; Guitián, E.; Peña, D.; Gourdon, A.; Meyer, G. Bond-Order Discrimination by Atomic Force Microscopy. *Science* **2012**, *337* (6100), 1326−1329.


DOI: 10.1021/acs.jctc.6b00219
*J. Chem. Theory Comput.* 2016, 12, 4362−4374