JCTC Journal of Chemical Theory and Computation

Article

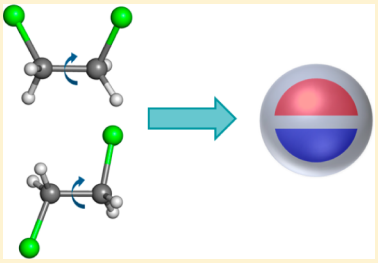
pubs.acs.org/JCTC

The Theory of Ultra-Coarse-Graining. 2. Numerical Implementation

Aram Davtyan, James F. Dama, Anton V. Sinitskiy, and Gregory A. Voth*

Department of Chemistry, The James Franck Institute, Institute for Biophysical Dynamics, and Computation Institute, The University of Chicago, Chicago, Illinois 60637, United States

ABSTRACT: The increasing interest in the modeling of complex macromolecular systems in recent years has spurred the development of numerous coarse-graining (CG) techniques. However, many of the CG models are constructed assuming that all details beneath the resolution of CG degrees of freedom are fast and average out, which sets limits on the resolution of feasible coarse-grainings and on the range of applications of the CG models. Ultra-coarse-graining (UCG) makes it possible to construct models at any desired resolution while accounting for discrete conformational or chemical changes within the CG sites that can modulate the interactions between them. Here, we discuss the UCG methodology and its numerical implementation. We pay particular attention to the numerical mechanism for including state transitions between different conformations within CG sites because this has not been discussed previously. Using a simple example of 1,2-dichloroethane, we demonstrate the ability of the UCG model to reproduce the multiconfigurational behavior of this molecular liquid, even when each molecule is modeled with only one CG site. The methodology can also be applied to other molecular liquids and macromolecular systems with time scale separation between conformational transitions and other intramolecular motions and rotations.



I. INTRODUCTION

Many vital biological processes involve large-scale conformational changes occurring over long periods of time. Common examples of such processes are protein folding and multicomplex assembly, which may occur on the time scales of seconds or even minutes. Despite continuous improvement in computational techniques,[1−10] it is still difficult, if not impossible, to study processes such as these using ordinary atomistic simulations. Thus, in recent years, there has been extensive effort to develop coarse-grained (CG) models that can address the complex macromolecular processes at the fraction of the computational cost by significantly reducing the number of the degrees of freedom in the system.[5,11−31] However, in practice, most of the existing coarse-graining techniques based on reproducing mechanical and structural correlations are limited in their ability to effectively decrease model resolution. This is due to the possibility of conformational and configurational changes that cannot be represented in terms of the positions of CG sites at lower resolution. As the CG sites represent effectively larger and larger subsystems of the overall system, simple CG models become increasingly unable to capture changes that occur within those subsystems or on the interfaces between them. Furthermore, these CG variables are typically unsuited to capturing the more and more nonlinear structural behavior typical of proteins when viewed at lower and lower resolution (for example, in the phenomenon of allostery).

Ultra-coarse-grained (UCG) modeling[32] addresses the problem described above by bringing together elements of two heretofore separate classes of systematic coarse-graining approaches. In the first class, the system is represented as a collection of kinematically interacting CG particle sites and the interaction potentials between these CG sites is derived based on finer-grained (such as all-atom) models using statistical properties of some type.[33−47] In the second class, the dynamics of the fine-grained (FG) system is modeled as a series of transitions between discrete states, with state-to-state transition rates fit from the projected dynamics of FG simulations.[48−58] UCG unifies both approaches[32] by introducing internal states associated with the CG particle configurations and by allowing the CG systems to undergo stochastic state transitions coupled to the CG particle conformations during the course of UCG simulation. The resulting framework is designed to enable study of a wide range of biological and soft material problems with more permissive limits on the resolution of feasible CG models. More permissive limits can enable the mesoscale simulation of very large macromolecular complexes, such as cytoskeletal actin networks,[59] that include chemical reactions and switch-like conformational changes as well as large-scale spatial and structural evolutions.

The idea of using discrete state change together with continuous motion has seen increased application in recent years. In a biophysical context, notable recent examples include work by Murtola et al.[42] to model lipid membrane using Monte Carlo simulations, by Enciso et al.[60] to capture the effect of changing pH on peptides. Also notable is the continuing work of Best, Hummer, and co-workers[61,62] that approximates multistate protein dynamics using a smooth mixing rule to capture major conformational changes in proteins. In polymer physics, recent work by Sing and Alexander-Katz[63−65] on self-associating polymers and by Müller, Morse et al.[66] on slip-link models for polymer entanglement both demonstrate the potential of such modeling in materials science as well. The

Received: September 13, 2014
Published: November 18, 2014

defining, differentiating feature of the UCG approach, as will be discussed below, is that it allows for the systematic derivation of state-dependent CG models from arbitrary FG representations without assuming specific functional forms for the interaction potentials. Thus, it promises to provide (i) new diagnostics, in that it can be used to check if models such as those above actually correspond directly to the fine-grained models that they are supposed to represent or are related only more indirectly through universality-like concepts, and (ii) a new means for rapid development and comparison of models of this type.

The underlying theory of UCG was introduced in a previous paper[32] (hereafter referred to as Paper I). Paper I defines a systematic UCG procedure and obtains necessary and sufficient conditions for consistency between UCG and FG representations. The formulation of this exact consistency condition then makes it possible to establish frameworks for deriving approximately consistent UCG models. Paper I[32] introduced two variational approaches by analogy to force-matching[34,35,39,40] and relative entropy,[41,45] respectively, either of which can be used to derive optimally consistent many-body UCG interaction potentials and state energies given a choice of how to assign FG structures to UCG states and configurations. The UCG modeling approach is agnostic as to whether the FG model is atomistic or itself coarse-grained at a higher resolution, making it possible to construct a hierarchy of CG models in which each model is derived from a higher-resolution model, starting from all-atom simulations and proceeding through a series of successively lower-resolution CG models.

While providing the theoretical framework, Paper I,[32] however, does not provide a numerical algorithm for simulating a UCG model for a specific system. The current paper will fill this gap by providing the algorithmic details of a UCG implementation based on the force-matching approach. We will pay particular attention to the methodology for internal CG state sampling and to the systematic procedures used for the derivation of the UCG interaction potentials and state transition rates.

To demonstrate the features of the UCG approach, we have chosen a simple system that includes both spatial and statelike motion: liquid 1,2-dichloroethane (DCE). DCE is a small molecule with well-defined long-lived dihedral conformational states and a large time-scale separation between transition rates between those conformational states and the rates of other intramolecular motions and rotations. The characteristic time scale for a conformational transition between the gauche and anti states is 0.1 ns, while the other motions are on the scale of femtoseconds to a few picoseconds. As we will show, the resulting UCG model accurately reproduces the local microstructure of the centers-of-mass of the DCE molecules as well as the key characteristics of the spontaneous conformational state dynamics that exist "within" the CG sites. The one-bead UCG model of DCE liquid, complemented with one internal variable per molecule, accurately captures equilibrium properties of the system that cannot be captured with a single CG site without internal states.

The remainder of this article is organized as follows: Section II starts with a short overview of the key UCG ideas from Paper I,[32] a discussion of state dynamics fitting and simulation methodology, and a detailed description of a UCG force matching implementation. This section also summarizes the results of preliminary all-atom simulation which motivate a particular definition for the DCE liquid UCG model. Section III presents and discusses the main results of UCG modeling of DCE liquid and compares the consistency of the results to the results of all-atom simulation. Section IV concludes the paper.

II. METHODS

A. Methodology of Ultra-Coarse-Graining. The theory of Ultra-Coarse-Graining (UCG) is discussed in detail in Paper I.[32] Here, we will provide a brief overview of the key ideas relevant to this work. The state of a UCG model is defined by collective spatial coordinates $\mathbf{R}^{N_\nu} \stackrel{\text{def}}{=} \{\mathbf{R}_I\}$ and discrete variables $\nu \in \Sigma$. These discrete variables can characterize certain implicit (hidden) properties of the CG system that cannot otherwise be deduced from the CG site representation directly. For example, ν can describe a conformational state, a ligand binding, or a chemical state. However, to avoid a possibly arbitrary definition, the value of ν is defined in a probabilistic manner. Given the FG coordinates $\mathbf{r}^n$ and a set of order parameters defined with them, we can state that there is a certain probability for the system to be in each of the possible states. Those probabilities are called membership functions, and they are denoted by $p_\Sigma(\nu;\mathbf{r}^n)$. In the most general case, the CG representation (including the number of CG sites) can vary with ν; however, in this paper, we will examine a case when the number of CG sites is the same across all states of the system. Nonetheless, we will continue to use N_ν to denote the number of CG sites, following the notation of a previous paper, for the sake of consistency.

Paper I[32] defined necessary and sufficient criteria for CG and FG models to be statistically consistent with each other. Equation 1 below states this criterion in terms of probability distributions for each state of the system, where $P_{R,\nu}(\mathbf{R}^{N_\nu})$ and $p_r(\mathbf{r}^n)$ are the CG and FG configurational probabilities, respectively, and $\mathrm{M}_{R,\nu}^{N_\nu}$ are mapping operators for coordinates in each state:

$$P_{R,\nu}(\mathbf{R}^{N_\nu}) = \int \mathrm{d}\mathbf{r}^n\, \delta(M_{R,\nu}^{N_\nu}(\mathbf{r}^n) - \mathbf{R}^{N_\nu}) p_\Sigma(\nu;\, \mathbf{r}^n) p_r(\mathbf{r}^n) \quad (1)$$

This consistency condition can also be written in terms of forces acting in CG and FG representations. In eq 2, $U(\nu,\mathbf{R}^{N_\nu})$ is the CG many-body potential of mean force (PMF), $u(\mathbf{r}^n)$ is the FG interaction potential, and the $\mathrm{M}_{R,\nu}^{N_\nu+}$ are mapping operators for forces for each state. The left-hand side represents the CG forces for a particular state of the system. Those forces should be equal to the expectations of effective forces, which consist of the FG forces minus terms accounting for the entropy of the UCG state distribution, i.e.,

$$\frac{\partial}{\partial \mathbf{R}^{N_\nu}} U(\nu,\, \mathbf{R}^{N_\nu}) = \left\langle \mathrm{M}_{R,\nu}^{N_\nu+}\left(\frac{\partial}{\partial \mathbf{r}^n}(u(\mathbf{r}^n) - \beta^{-1} \ln p_\Sigma(\nu;\, \mathbf{r}^n))\right)\right\rangle_{\mathbf{R}^{N_\nu},\nu} \quad (2)$$

Besides having fundamental significance, this consistency condition provides a practical modeling scheme in the form of a variational method that can be used to derive CG interactions from FG simulation data that approximate the many-body PMF $U(\nu,\mathbf{R}^{N_\nu})$ well. The basic idea behind this method is to minimize the mean-squared error between forces acting in CG and FG representations. In other words, we must minimize the following expression among trial CG potentials (W):

$$\chi_\rho^2[W] = \sum_{\nu\in\Sigma}\int \mathrm{d}\mathbf{R}^{N_\nu}\,\rho(\nu,\mathbf{R}^{N_\nu}) \times \|\mathbf{G}^{N_\nu}(\nu,\mathbf{R}^{N_\nu}) - \langle \mathrm{M}_{R,\nu}^{N_\nu+}(\mathbf{f}^n(\nu,\mathbf{r}^n))\rangle_{\mathbf{R}^{N_\nu},\nu}\|^2 \tag{3}$$

where

$$\mathbf{G}^{N_\nu}(\nu,\mathbf{R}^{N_\nu}) \overset{\text{def}}{=} \frac{\partial}{\partial \mathbf{R}^{N_\nu}} W(\nu,\mathbf{R}^{N_\nu}) \tag{4}$$

$$\mathbf{f}^n(\nu,\mathbf{r}^n) \overset{\text{def}}{=} \frac{\partial}{\partial \mathbf{r}^n}(u(\mathbf{r}^n) - \beta^{-1}\ln p_\Sigma(\nu;\mathbf{r}^n)) \tag{5}$$

Here, $\rho(\nu,\mathbf{R}^{N_\nu})$ is an arbitrary, everywhere-positive, normalized probability distribution. For the canonical ensemble and for the choice of $\rho(\nu,\mathbf{R}^{N_\nu}) = P_{R,\nu}(\nu,\mathbf{R}^{N_\nu})$ we can rewrite the above equation in terms of the time expectation of an unbiased simulation in the following way:

$$\chi_\rho^2[W] = \lim_{\tau\to\infty}\sum_{t=0}^{\tau}\sum_{\nu\in\Sigma} p_\Sigma(\nu;\mathbf{r}_t^n) \times \|\mathbf{G}^{N_\nu}(\nu,\mathbf{R}_t^{N_\nu}) - \mathrm{M}_{R,\nu}^{N_\nu+}\mathbf{f}^n(\nu,\mathbf{r}_t^n)\|^2 \tag{6}$$

Equation 6 defines a least-squares problem that can be solved using various numerical methods given a form for the trial potential W. Similar to the previous work on multiscale coarse-graining (MS-CG),[34,35,39,40] one very important case is when W can be represented as a linear combination of basis functions for each given state of the system. Equation 7 below shows such a representation of potential W_ζ of type ζ, where the type of the potential corresponds to a particular type of interatomic interaction (i.e., bonded, angle, dihedral, or nonbonded). The outer sum is over a set of sites participating in the type of interaction ζ. The inner sum goes over the basis functions; each basis function is weighted by state-dependent coefficients λ. The w_d denote the basis functions, which are defined as functions of scalar variables $x_{\gamma,\zeta}$ corresponding to the particular interaction γ and type of interaction ζ. These can correspond to bond lengths, angle values, or interbead distances for nonbonded interactions, such that

$$W_\zeta(\nu,\mathbf{R}^{N_\nu}) = \sum_\gamma\sum_d \lambda_d(\nu) w_d(x_{\gamma,\zeta}(\mathbf{R}^{N_\nu})) \tag{7}$$

Typical choices for the basis functions include step functions, B-splines, Gaussians, and polynomials of second order and higher. In that case, we can use standard linear least-squares numerical techniques to determine the λ coefficients and, thus, the interaction potential.

Another important aspect of UCG implementation involves the formulation of the state sampling procedure (i.e., state dynamics). This was not discussed in Paper I,[32] which dealt exclusively with the definition of UCG potentials. However, one can formulate the main criteria that the state dynamics must obey. First, the UCG state dynamics should reproduce FG equilibrium distributions. This means they must satisfy the detailed balance criterion, and, therefore, the state transitions must be dependent on the current CG configuration. Second, the overall rate of transition in both the UCG and FG models should be the same. Other more-sophisticated requirements can also be added. For instance, one could require that not only the overall rate of transitions but also specific conditional rates should match between the models. However, here, we will consider only those first two, in the interest of clarity.

Taking into account what has been said above, one possible way to represent the instantaneous transition rates is described by the following expressions:

$$K_{\nu_1\to\nu_2} = k_{\nu_1\to\nu_2}\min[\exp[-\beta(U(\nu_2,\mathbf{R}^{N_\nu}) - U(\nu_1,\mathbf{R}^{N_\nu})) - h], 1] \tag{8}$$

$$K_{\nu_2\to\nu_1} = k_{\nu_2\to\nu_1}\min[\exp[-\beta(U(\nu_1,\mathbf{R}^{N_\nu}) - U(\nu_2,\mathbf{R}^{N_\nu})) + h], 1] \tag{9}$$

In eqs 8 and 9, $k_{\nu_1\to\nu_2}$ and $k_{\nu_2\to\nu_1}$ are constant prefactors that must be determined in some way. The second parts of those expressions are Metropolis–Hastings-like terms that are dependent on the difference between the interaction energies in the current and candidate states. The term h is a third free constant that must also be determined. It corresponds to a one-body state-dependent energy in the UCG PMF; fitting it to match the average population in each state amounts to a systematic relative entropy minimization fit for the state distribution using that one-body term as a basis set for the fit.[67,68] Here, the ν_1 and ν_2 represent the overall states of the system that are one transition apart. For example, if there is one discrete variable for each of the N particles in a UCG system and the molecules make transitions between these states individually, then ν_1 and ν_2 must differ by the state of only one of the particles.

By construction eqs 8 and 9 can preserve the equilibrium distributions for each state, because they are dependent on the energy difference between the states. The choice of the constant prefactors and the free constant h will allow us to satisfy the conditions fully. Equation 10 below represents the detailed balance criterion required to ensure consistency, given accurate state-specific models:

$$\frac{K_{\nu_1\to\nu_2}}{K_{\nu_2\to\nu_1}} = \frac{P(\nu_2,\mathbf{R}^{N_\nu})}{P(\nu_1,\mathbf{R}^{N_\nu})} \tag{10}$$

where the probability distributions $P(\nu,\mathbf{R}^{N_\nu})$ are defined as follows:

$$P(\nu,\mathbf{R}^{N_\nu}) = \frac{\exp[-\beta U(\nu,\mathbf{R}^{N_\nu})]}{\sum_{\nu\in\mathrm{Z}}\int \mathrm{d}\mathbf{R}^{N_\nu}\exp[-\beta U(\nu,\mathbf{R}^{N_\nu})]} \tag{11}$$

Plugging eqs 8 and 9 into eq 10, we then arrive at the following expression:

$$\frac{k_{\nu_1\to\nu_2}}{k_{\nu_2\to\nu_1}}\exp[-\beta(U(\nu_2,\mathbf{R}^{N_\nu}) - U(\nu_1,\mathbf{R}^{N_\nu})) - h] = \exp[-\beta(U(\nu_2,\mathbf{R}^{N_\nu}) - U(\nu_1,\mathbf{R}^{N_\nu}))] \tag{12}$$

Using eq 12, we can find $h = \ln[k_{\nu_1\to\nu_2}/k_{\nu_2\to\nu_1}]$, which leads to the following expressions for the transition rates:

$$K_{\nu_1\to\nu_2} = k_{\nu_1\to\nu_2} \times \min\left[\frac{k_{\nu_2\to\nu_1}}{k_{\nu_1\to\nu_2}}\exp[-\beta(U(\nu_2,\mathbf{R}^{N_\nu}) - U(\nu_1,\mathbf{R}^{N_\nu}))], 1\right] \tag{13}$$

dx.doi.org/10.1021/ct500834t | J. Chem. Theory Comput. 2014, 10, 5265−5275

$$K_{\nu_2\to\nu_1} = k_{\nu_2\to\nu_1} \times \min\left[\frac{k_{\nu_1\to\nu_2}}{k_{\nu_2\to\nu_1}}\exp[-\beta(U(\nu_1,\mathbf{R}^{N_\nu}) - U(\nu_2,\mathbf{R}^{N_\nu}))], 1\right] \quad (14)$$

The constant prefactors k can be estimated from all-atom simulations by matching the average (over the all-atom trajectory) instantaneous rates K_{AA} with the effective transition rates $R^{AA}_{\nu_1\to\nu_2}$ and $R^{AA}_{\nu_2\to\nu_1}$ calculated independently, based on the same all-atom simulations only. In other words, neglecting recrossing effects, we need to solve the following equations, where the averaging is over the realizations of the corresponding states in the all-atom simulation:

$$R^{AA}_{\nu_1\to\nu_2} = \langle K_{\nu_1\to\nu_2}\rangle_{AA,\nu_1} \quad (15)$$

$$R^{AA}_{\nu_2\to\nu_1} = \langle K_{\nu_2\to\nu_1}\rangle_{AA,\nu_2} \quad (16)$$

where averages $\langle\cdots\rangle_{AA,\nu}$ can be defined in the most general case using the following expression:

$$\langle A\rangle_{AA,\nu} = \frac{\int d\mathbf{r}^n A(\mathbf{r}^n) p_r(\mathbf{r}^n) p_\Sigma(\nu;\mathbf{r}^n)}{\int d\mathbf{r}^n p_r(\mathbf{r}^n) p_\Sigma(\nu;\mathbf{r}^n)} \quad (17)$$

At first glance, this approach may seem impractical for large composite systems (e.g., liquids), because if the state is defined by the composition of a large number of individual discrete variables, properly sampling all of the possible states may become very hard. However, when those variables describe corresponding properties of different molecules or subunits, then the average can also be taken over the discrete variables themselves. Doing so requires an assumption that the states have underlying symmetry, but no assumptions of statistical independence.

Equation 6 provides a rigorous method to determine the state-specific interaction potentials up to a constant per state-specific potential. This makes it possible to find the forces acting in the system at any given moment in a given state, and to propagate it in space over time as long as the state remains fixed. However, for the state dynamics governed by eqs 13 and 14, we must know the relative energies before and after the transition. Thus, in order to be able to formulate the transition dynamics in terms of the potential energies $U_r(\nu,\mathbf{R}^{N_\nu})$ derived through the force-matching technique, we will introduce an additional correction factor in the Metropolis−Hastings-like terms of eqs 13 and 14. We will denote this factor by ε, such that

$$K_{\nu_1\to\nu_2} = k_{\nu_1\to\nu_2} \times \min\left[\frac{k_{\nu_2\to\nu_1}}{k_{\nu_1\to\nu_2}}\exp[-\beta(U_r(\nu_2,\mathbf{R}^{N_\nu}) - U_r(\nu_1,\mathbf{R}^{N_\nu}) - \varepsilon)], 1\right] \quad (18)$$

$$K_{\nu_2\to\nu_1} = k_{\nu_2\to\nu_1} \times \min\left[\frac{k_{\nu_1\to\nu_2}}{k_{\nu_2\to\nu_1}}\exp[-\beta(U_r(\nu_1,\mathbf{R}^{N_\nu}) - U_r(\nu_2,\mathbf{R}^{N_\nu}) + \varepsilon)], 1\right] \quad (19)$$

One possible way to define ε is through the following criterion, which, besides accounting for the ambiguity at the zero-energy level, ensures the maximal surface hopping acceptance rate and the least effort required to calculate the Metropolis−Hastings-like terms.

$$\left\langle\frac{K_{\nu_1\to\nu_2}}{k_{\nu_1\to\nu_2}}\right\rangle_{AA,\nu_1} = \left\langle\frac{K_{\nu_2\to\nu_1}}{k_{\nu_2\to\nu_1}}\right\rangle_{AA,\nu_2} \quad (20)$$

The most straightforward approach to determine the prefactors (k) and the relative energies (ε) is to carry out an extensive all-atom simulation to get a system of three equations with three unknowns from eqs 15−20. However, in practice, the recrossing effects in the FG state transitions, which are neglected in these equations, result in non-negligible differences between the FG and CG models; consequently, the solutions of the prefactors k and ε will not immediately reproduce the equilibrium distribution of the FG system. Thus, as we will discuss later, one typically must perform a self-consistent optimization by running UCG simulations and comparing the effective transition rates calculated from those simulations with those obtained from the FG all-atom model, adjusting the parameters each time. This type of systematic procedure ensures that the equilibrium distribution of the UCG system matches that of the all-atom one.

B. All-Atom Simulations of DCE. The all-atom simulations of DCE in liquid and gas phases were carried out in GROMACS,[69] using the OPLS-AA[70,71] (Optimized Potentials for Liquid Simulations, All-Atom) force field with long-range electrostatics calculated through Ewald summation. Liquid simulations contained 1000 DCE molecules, placed in a cubic box with periodic boundary conditions. The system was energy-minimized and then equilibrated for 2 ns, with equilibration under constant temperature (NVT) and under constant temperature and pressure (NPT) for 1 ns each. During the constant NPT run the box size and the density changed, from 5.11 nm to 5.04 nm and from 1.23 g/cm^3 to 1.28 g/cm^3, respectively. A constant NVT production simulation then was run for 100 ns. Gas-phase simulations contained one molecule that was placed in a nonperiodic box. The system was minimized and equilibrated under constant temperature for 1 ns, followed by a constant temperature production simulation that was run for 1 μs.

For temperature coupling, we used the V-Rescale thermostat, which is similar to a Berendsen thermostat but has an additional stochastic term which ensures that a proper canonical ensemble will be sampled.[72] For the NPT simulations, a Parrinello−Rahman pressure coupling was also used. All simulations were carried out under a temperature of 298 K and a constant pressure of 1 atm, when applicable. A time step of 1 fs was used. Using the all-atom simulations described above, we obtained the distribution of dihedral angle ϕ (formed by carbons and chlorines) for liquid and gas phases. Those results are summarized in Figure 1. As was expected, for both liquid and gas, we found three peaks corresponding to the two gauche states and one anti state (see Figure 2a). The two important features are noted here: (i) the peaks are well-separated and (ii) the heights of those peaks are very different for liquid and gas phases. The former means that, in this particular case, there is no ambiguity in how to assign the state of the molecule when given the value of the dihedral angle ϕ. Thus, the definition of the UCG membership functions is not utterly critical, so long as they distinguish between peaks corresponding to each of the states. The latter means that the state distribution is quite different for liquid and gas phases. In the gas phase, the anti state is dominant and occurs 81% of the

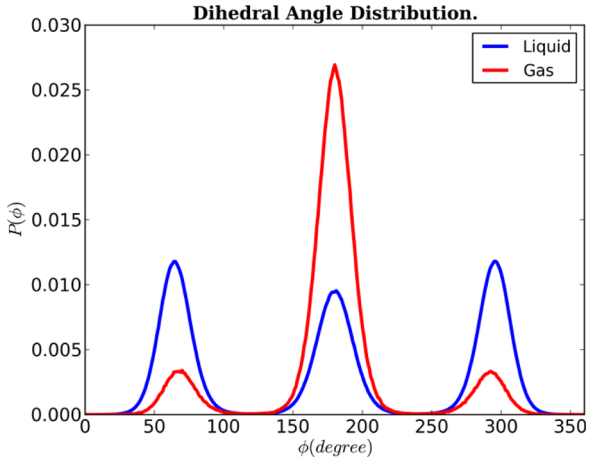


Figure 1. Normalized distribution of the carbon and chlorine dihedral angle of 1,2-dichloroethane for liquid and gas phases. The middle peaks correspond to the anti state, while the two peaks on the side correspond to gauche states.

time, while in the liquid phase, it occurs in only 32% of the time. This difference illustrates the importance of many-body correlations in state dynamics.

We also calculated the transition rates between gauche and anti states. We found that in the liquid phase the transition rate from gauche to anti was 9.34 ns^{-1} and the rate from anti to gauche was 20.12 ns^{-1}. In the gas phase, those rates were determined to be 14.69 and 3.53 ns^{-1}, respectively. The obtained rates show a clear separation of time scales between the state transitions and the other molecular motions.

C. UCG Model of DCE. 1,2-Dichloroethane is a small molecule with the chemical formula $C_2H_4Cl_2$. It can be found in three well-separated states defined by the value of the dihedral angle (ϕ) formed by carbons and chlorines. The anti state corresponds to a ϕ value of ~180° and two gauche states correspond to ϕ values of approximately ±65° (see Figure 2a). For simplicity, we will consider only two states: the anti state ($\phi \approx 180°$) and a joint gauche state with $|\phi| \approx 65°$. This type of simplification is based on two assumptions:

(1) the transitions between the two gauche states and the anti state occur at the same rate, and
(2) interactions between gauche states can be described by one potential (instead of three).

The numerical calculations (based on all-atom simulations) show that the first assumption holds almost exactly. The second one is more ambiguous; however, as the results will later show, it does not greatly affect the accuracy of the derived UCG model, shown schematically in Figure 2b.

Given the considerations above and the results of the all-atom simulations, we have chosen to represent each DCE molecule with one bead placed at its center of mass and one internal variable corresponding to either gauche or anti states. The state of each molecule was defined in a deterministic manner using step membership functions given by

$$p_{\Sigma,\mathrm{mol},I}(g;\,\mathbf{r}^n) = \begin{cases} 0 & \text{if } |\phi_I(\mathbf{r}^n)| > \phi_0 \\ 1 & \text{if } |\phi_I(\mathbf{r}^n)| \le \phi_0 \end{cases} \tag{21}$$

$$p_{\Sigma,\mathrm{mol},I}(a;\,\mathbf{r}^n) = \begin{cases} 1 & \text{if } |\phi_I(\mathbf{r}^n)| > \phi_0 \\ 0 & \text{if } |\phi_I(\mathbf{r}^n)| \le \phi_0 \end{cases} \tag{22}$$

Equation 21 describes the probability of the particular molecule to be in gauche state, and eq 22 describes the probability of the particular molecule to be in anti state. Here, ϕ_I is the dihedral angle corresponding to molecule I, and ϕ_0 is the dihedral angle value at the transition point. Based on all-atom simulations, we defined ϕ_0 to be 120°. Using the per-molecule membership functions defined in eqs 21 and 22, we now define the membership functions for the entire system.

The full DCE system is composed of N_ν identical molecules, and each can be in either gauche or anti state. Therefore, as system state labels, we can choose to use sets of N_ν single-molecule labels; $\nu_{\mathrm{sys}} = \{\nu_{\mathrm{mol},I}\}$, where $\nu_{\mathrm{mol},I}$ is either g or a. Thus, we can define the system membership function as a product of per-molecule membership functions

$$p_\Sigma(\{\nu_{\mathrm{mol},I}\};\,\mathbf{r}^n) = \prod_{I=1}^{N_\nu} p_{\Sigma,\mathrm{mol},I}(\nu_{\mathrm{mol},I};\,\mathbf{r}^n) \tag{23}$$

According to this definition, there will be 2^{N_ν} membership functions, with only one of them being equal to 1 and the rest being equal to zero, for any given configuration of the system. So, for example, in a system of three DCE molecules where the first one is gauche and the two others are in the anti state, only $p_\Sigma(\{g,a,a\};\mathbf{r}^n)$ will have a nonzero value of 1. In a system of 1000 identical molecules, using this sort of divide-and-conquer definitional scheme is essential to taming the complexity of the 2^{1000}-member state space.

The potential of mean force can be approximated as a sum of pairwise interactions that are dependent on the state of the interacting CG beads and the distances between them plus a sum of one-body terms dependent only on each molecule's state:

$$U(\nu_{\mathrm{sys}},\,\mathbf{R}^{N_\nu}) = \sum_{I<J} U_{IJ}(\nu_{\mathrm{pair},IJ},\,R_{IJ}) + \sum_I U_I(\nu_I) \tag{24}$$

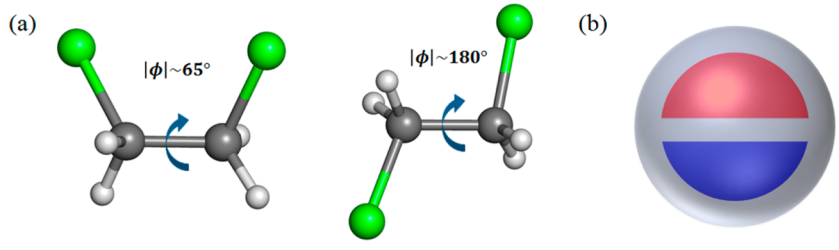


Figure 2. (a) Illustrations of the 1,2-dichloroethane molecule in gauche state (left) and anti state (right). (b) Schematic of the 1-site UCG DCE model; the two shades of color denote the hidden states in the single UCG site.

dx.doi.org/10.1021/ct500834t | J. Chem. Theory Comput. 2014, 10, 5265−5275

where $\nu_{\text{pair},IJ}$ denote the pair of per-molecule state labels $\{\nu_{\text{mol},I}\nu_{\text{mol},J}\}$. Again, using state symmetries is essential to taming the 2^{1000}-dimensional state space; deriving independent force fields for each would be unfeasible.

We chose to represent the pairwise potential as a linear combination of sixth-order B-splines,[73] such that

$$U(\nu_{\text{sys}}, \mathbf{R}^{N_\nu}) = \sum_{I<J} \sum_{d} \lambda_d(\nu_{\text{pair},IJ}) u_d(R_{IJ}) \tag{25}$$

The $\lambda_d(\nu_{\text{pair},IJ})$ coefficients are obtained through force-matching for every possible combination of the states of the interacting sites.

The state transitions are governed by eqs 18 and 19, where the prefactors $k_{\nu_1\to\nu_2}$ and $k_{\nu_2\to\nu_1}$ set the overall time scales with which the transitions between gauche and anti states were attempted, while the Metropolis−Hastings-like terms define the acceptance rates of the corresponding transitions. The scalar ε corresponds to the one-body term. Each accepted transition was immediately applied to the system and thus affected any subsequent transitions.

D. Force-Matching Procedure. Here, we will discuss the force-matching procedure that was used to derive the interaction potential for DCE liquid. In this regard, we will focus only on nonelectrostatic, pairwise CG interactions. However, everything we will discuss can easily be extended for interaction potentials of the most-general linear form.

Equation 6 defines the linear least-squares problem in terms of the trial potential W. The negative gradients of this potential with regard to the positions of the UCG sites give us the forces $\mathbf{G}^{N_\nu}$ acting on those sites. Following eq 25 we can write potential W in terms of a linear combinations of sixth order B-splines basis functions[73]

$$W(\nu_{\text{sys}}, \mathbf{R}^{N_\nu}) = \sum_{I<J} \sum_{d=1}^{N_d} \lambda_d(\nu_{\text{pair},IJ}) w_d(R_{IJ}) \tag{26}$$

The outer sum in eq 26 goes over all pairs of sites, while the inner sum goes over the N_d basis functions. The coefficients $\lambda_d(\nu_{\text{pair},IJ})$ are dependent on the states of the interacting molecules. Thus, with two possible states per molecule, we will, in reality, have three $\lambda_d(\nu_{\text{pair},IJ})$ parameters for each value of d. To make this more explicit, and to simplify the equations, we can replace the sum over d by a sum over all parameters λ. Therefore, we can write the interaction potential as

$$W(\nu_{\text{sys}}, \mathbf{R}^{N_\nu}) = \sum_{D=1}^{N_D} \sum_{\gamma} \lambda_D w_D(R_\gamma) \tag{27}$$

where the inner sum goes over only the pairs in corresponding states. Also, one should note that the set of new basis functions $\{w_D\}$ will simply contain the three copies of the old basis functions w_d, specialized by state, similar to how pair interactions are usually specialized by type. In this regard, it can be useful to consider the DCE liquid as a system of molecules of two different types, with the types of the molecules dynamically changing over time.

The forces acting on UCG site I will be

$$\mathbf{G}_I(\nu, \mathbf{R}^{N_\nu}) = \sum_{D=1}^{N_D} \sum_{\gamma} \lambda_D g_D(R_\gamma) \frac{\partial R_\gamma}{\partial \mathbf{R}_I} \tag{28}$$

where we define

$$g_D(R_\gamma) = -\frac{\mathrm{d}w_D(R_\gamma)}{\mathrm{d}R_\gamma} \tag{29}$$

This can further be transformed to

$$\mathbf{G}_I(\nu, \mathbf{R}^{N_\nu}) = \sum_{D=1}^{N_D} \lambda_D \mathcal{G}_{I;D}(R^{N_\nu}) \tag{30}$$

where

$$\mathcal{G}_{I;D}(\mathbf{R}^{N_\nu}) = \sum_{\gamma} g_D(R_\gamma) \frac{\partial R_\gamma}{\partial \mathbf{R}_I} \tag{31}$$

Using eq 30, we can rewrite eq 6 for a finite number of sampled configurations n_t. The sum over ν reduces to a single component because only one membership function $p_\Sigma(\nu_{\text{sys}};\mathbf{r}_t^n)$ has a nonzero value of 1 for any given configuration, such that

$$\chi_\rho^2[W] = \sum_{t=0}^{n_t} \sum_{I=1}^{N_\nu} \left\| M_{R,\nu}^{N_\nu+} \mathbf{f}_I^n(\nu_{\text{sys};t}, \mathbf{r}_t^n) - \sum_{D=1}^{N_D} \lambda_D \mathcal{G}_{I;D}(\mathbf{R}_t^{N_\nu}) \right\|^2 \tag{32}$$

The above linear least-squares problem (see eq 32) can be solved with a variety of numerical techniques. Analogous to earlier work on MS-CG,[40] we solved this least-squares problem using a singular value decomposition of the corresponding normal equations. The primary difference from an MS-CG calculation was that the types of UCG sites changed dynamically from one system configuration to another.

In order to estimate the error in the UCG force field, we divided the all-atom trajectory into five equal parts and derived interaction potentials for each of those subtrajectories by solving the least-squares problems defined in eq 32 for each. We found that the error in the pairwise potentials did not exceed 0.1 kJ/mol for distance separations larger than 0.35 nm. For distances smaller than 0.35 nm, we found the error to be much larger, presumably because of a lack of sampling.[74] Therefore, we fitted that portion of the pairwise potentials with functions of the form A/r^L. The final force field and potentials were constructed from the averages over the obtained five potentials for distances larger than 0.35 nm, and from fitted functions for distances smaller than 0.35 nm.

We described this force-matching procedure for a system with only nonelectrostatic pairwise CG interactions. However, everything discussed above can easily be generalized for more-complex interaction potential basis sets. This procedure was implemented in the Multi-Scale Coarse-Graining Force-Matching (MSCGFM) package[73] and can be used to derive general molecular interaction potentials including bonded potentials (distance, angle, and dihedral) and three-body nonbonded potentials.

E. Transition Rate Calculations and Optimization. The transition rates for the all-atom simulations were calculated using state correlation functions (see Figure 3) that are defined as the average fractions of molecules found in a particular state α at time $t + t'$ given that they were in that state at time t. In eq 33, N_α is the number of molecules in state α at time t and θ_α is an indicator function that is unity if molecule I is in state α and zero otherwise. The sum goes over all the molecules in the system, i.e.,

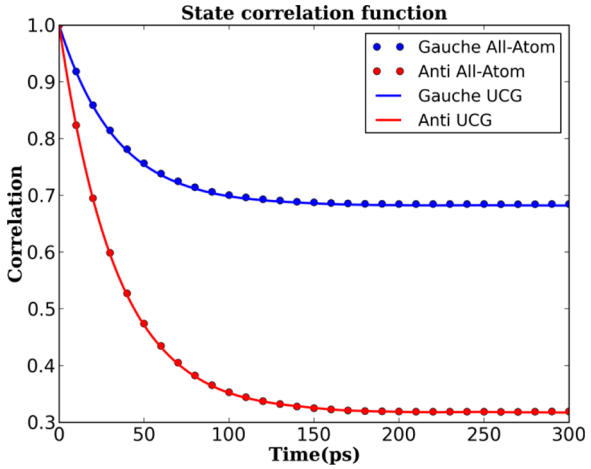


Figure 3. Plot showing the state correlation functions obtained from the all-atom (circles) and UCG (solid lines) simulations.

$$Y_\alpha(t') = \left\langle \frac{1}{N_\alpha} \sum_{I=1}^{N_\nu} \theta_\alpha(\mathbf{R}_{I,t}) \theta_\alpha(\mathbf{R}_{I,t+t'}) \right\rangle_t \tag{33}$$

Figure 3 shows the state correlation functions obtained for DCE liquid. Starting from a value of unity, these functions exponentially decay to stationary values within a few 100 ps. Knowing those stationary values and the rates of decay, one can then estimate the transition rates between the corresponding states. For this system, the dynamics are simple and each of the correlation functions can be fit with a function of a form

$$y = a + (1 - a)\exp(-bx) \tag{34}$$

The parameters a correspond to ratios of the rates, while the parameters b correspond to their sum.

For the UCG model, the rates can be calculated simply by taking the ratio of the number of transitions to the number of realizations of the initial state, such that

$$R^{\mathrm{UCG}}_{\nu_1 \to \nu_2} = \frac{N_{\nu_1 \to \nu_2}}{N_{\nu_1}} \tag{35}$$

$$R^{\mathrm{UCG}}_{\nu_2 \to \nu_1} = \frac{N_{\nu_2 \to \nu_1}}{N_{\nu_2}} \tag{36}$$

Because of the impossibility of recrossing in the UCG simulations, the rates calculated using eqs 35 and 36 will be equal to the rates calculated using the state correlation functions as described above. This seemingly unimportant detail actually makes it straightforward to establish a systematic rate optimization procedure.

Using eqs 15−20, initial estimates of $k_{\nu_1 \to \nu_2}$, $k_{\nu_2 \to \nu_1}$, and ε can be found. However, as we mentioned previously, those values will not exactly reproduce the equilibrium distribution of the all-atom system in the UCG simulations, so we adapted a self-consistent Robbins−Monro[75] optimization scheme, which consists of the following:

(1) run a 10-ns-long UCG simulation, using the parameters obtained from the all-atom trajectory;
(2) use the last 9 ns of the simulation to calculate the effective transition rates using eqs 35 and 36;
(3) compare the calculated transition rates with those estimated from the all-atom simulations, and then calculate the corrections to the prefactors k, as described in eq 38;
(4) obtain new prefactors by mixing the corrections with the old values using a decaying mixing parameter σ (see eqs 37 and 39 below);
(5) use eq 20 to obtain a new ε;
(6) run new UCG simulations, using the newly adjusted parameters; and
(7) repeat steps 2−6 until the optimization converges.

Relevant equations are given by

$$k^{(i+1)} = k^{(i)} + \sigma_i \Delta k^{(i)} \tag{37}$$

$$\Delta k^{(i)} = k^{(i)} \left(\frac{R^{\mathrm{AA}}}{R_i^{\mathrm{UCG}}} - 1 \right) \tag{38}$$

$$\sigma_i = \frac{1}{1 + q(i - 1)} \tag{39}$$

Equation 39 describes the mixing parameter σ that starts from a value of one at the initial iteration $i = 1$ and gradually decays according to a chosen positive rate q. Upon the first iteration, we completely accept the new values without mixing them with the previous values; however, starting with the second iteration, the new parameters are composed of the old parameters together with fractions of the new estimates. The decay rate q must be small enough to allow the parameters to reach their equilibrium values before saturating, but also must be large enough for the optimization procedure to converge in a reasonable time. In the current work, we chose $q = 0.5$. Figure 4 shows the effective transition rate from the anti to gauche

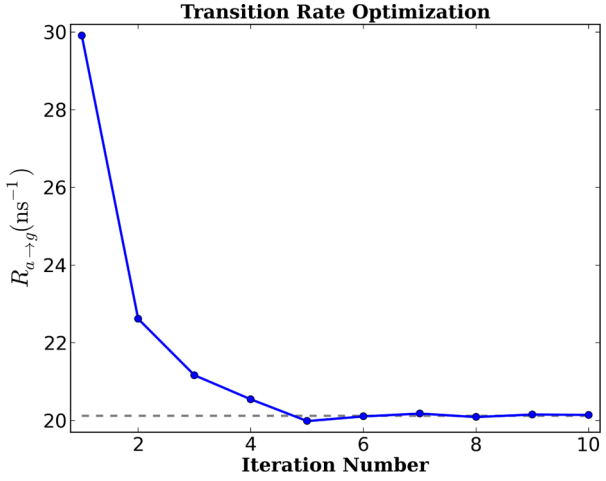


Figure 4. Illustration of the rate optimization procedure for the UCG anti to gauche transition rate over the course of 10 iterations. The dashed gray line indicates the rate value obtained from all-atom simulations toward which the UCG rates are optimized.

state calculated according to the above procedure over 10 iterations. In the first iteration, the parameters are estimated solely based on all-atom simulations. As illustrated, the optimization converges within 5 or 6 steps, reaching the desired value measured from the all-atom simulations, marked on the plot by the horizontal dashed line.

F. UCG Simulations of DCE. The UCG simulations of liquid 1,2-dichloroethane were carried out in LAMMPS[76] (http://lammps.sandia.gov) using a newly implemented

dx.doi.org/10.1021/ct500834t | J. Chem. Theory Comput. 2014, 10, 5265−5275

module for UCG state dynamics (UCG-SDM). This module works by changing the LAMMPS atom types of stochastically selected CG beads at each time step; changing the type changes the interactions. The CG sites interact through the tabular pair potentials obtained by the force-matching procedure described above. We used a Nose−Hoover thermostat to run constant NVT simulations of liquid DCE in a periodic box with the same dimensions (and number of molecules) as in the all-atom production runs. UCG runs were up to 100 ns long, with the first 1 ns used for equilibration. A time step of 1 fs was used, although care should be taken in relating CG time scales to FG time scales—because the continuous dynamics lack dissipation, the diffusive dynamics of the CG system are qualitatively different from those of the FG system.[77]

Because of the reduction in the number of degrees of freedom, the UCG simulations were significantly more efficient than the FG ones. Namely, using 16 processors, FG simulations were running with the rate of 1.93 ns/h, while, for UCG simulations, the rate of 16.93 ns/h (also using 16 processors) was achieved. The cost of the UCG-SDM module was found to be ∼10% from the total computational time.

One additional feature to note is that the instantaneous state transitions move the system from one potential energy surface to another, which, in principle, can lead to deviation of the energy and the creation of local "hot spots", with some of the UCG beads gaining an unphysical large kinetic energy. The use of constant NVT simulations ensures that the former problem will not occur, and, for the example discussed here, we have not observed the occurrence of any "hot spots." However, we recognize that caution should be taken for each particular system.

III. RESULTS AND DISCUSSION

To illustrate the UCG approach described above, we applied it to the molecular liquid consisting of 1000 1,2-dichloroethane

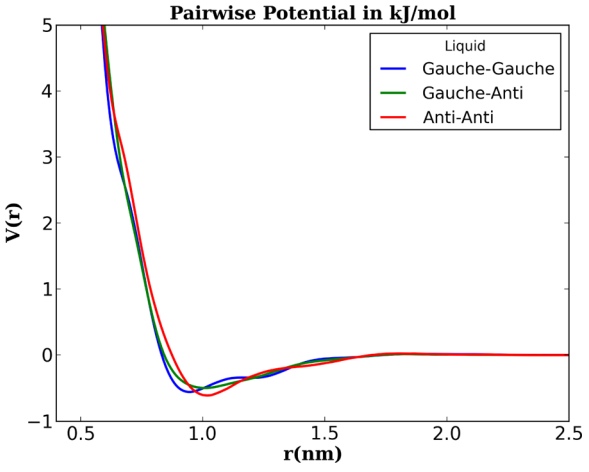


Figure 5. Pairwise interaction potentials between DCE beads, given the conformational state of the pair.

Table 1. Parameter Values for State Transition Equations[a]

$k_{g\to a}$ (ns^{-1})	$k_{a\to g}$ (ns^{-1})	self-energy factor, ε (kJ/mol)
83.73	20.36	0.0787

[a]The terms $k_{g\to a}$ and $k_{a\to g}$ are the rate prefactors for gauche-to-anti and anti-to-gauche transitions, respectively.

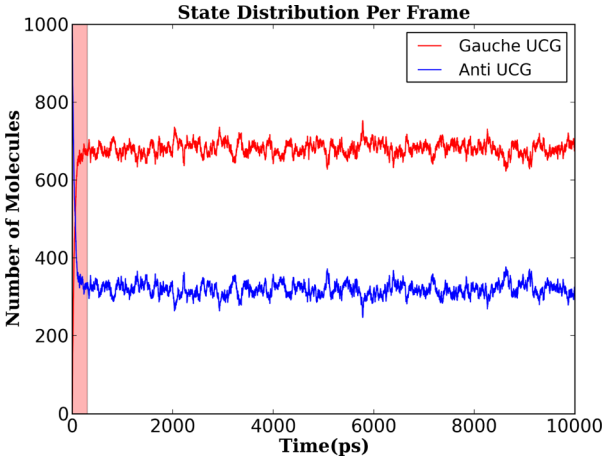


Figure 6. State dynamics of liquid DCE obtained from UCG simulations. All molecules were initialized in the anti state.

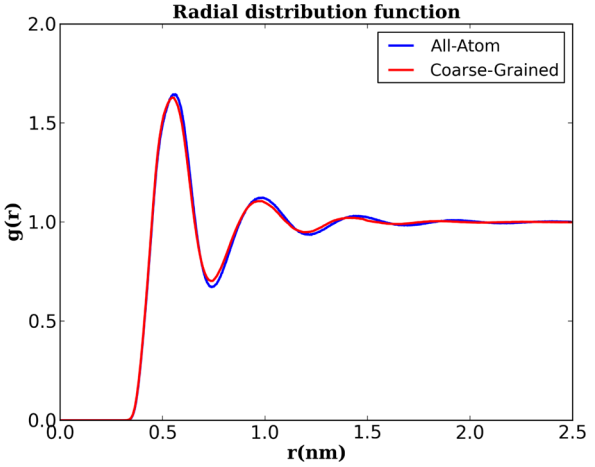


Figure 7. Comparison of overall center-of-mass (COM) radial distribution functions (RDFs) obtained from all-atom and UCG simulations.

(DCE) molecules described earlier. Using one CG bead per molecule to describe the spatial dynamics and one internal variable per molecule to describe the conformational dynamics, we derived a UCG model that reproduces key equilibrium properties of the DCE liquid.

Using the force-matching procedure described above, we obtained pairwise interaction potentials between UCG DCE molecules for the three possible combinations of states of the CG sites involved (see Figure 5). Knowing the interaction potential allowed us to then determine the prefactor k and the energy correction ε (see the Methods section and Table 1). Therefore, having everything needed, we ran UCG simulations of DCE liquid using UCG-SDM in LAMMPS.

Figure 6 shows the state dynamics of liquid DCE starting from an all-anti configuration. It can be seen that the system reaches equilibrium within only a few 100 ps (shown by the vertical red bar). After that, the numbers of CG molecules in gauche and anti states fluctuate around their average values. To evaluate the quality of the UCG model, we also calculated the radial distribution functions (RDFs) between centers-of-mass (COM) of DCE molecules. Figure 7 compares the overall RDFs between UCG and all-atom simulations, while Figure 8 shows the state differentiated RDFs side by side. As one can

dx.doi.org/10.1021/ct500834t | J. Chem. Theory Comput. 2014, 10, 5265−5275

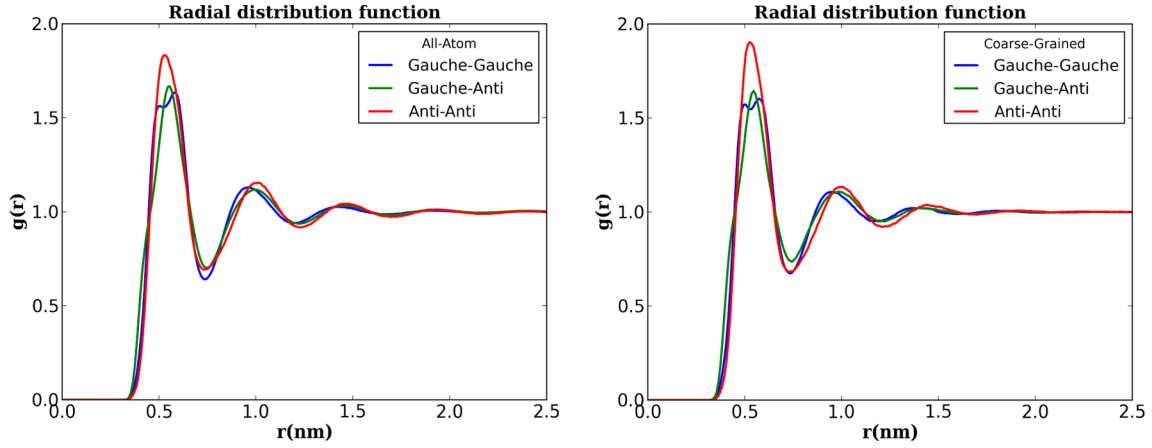


Figure 8. State-differentiated COM RDFs of liquid DCE obtained from all-atom (left) and UCG (right) simulations.

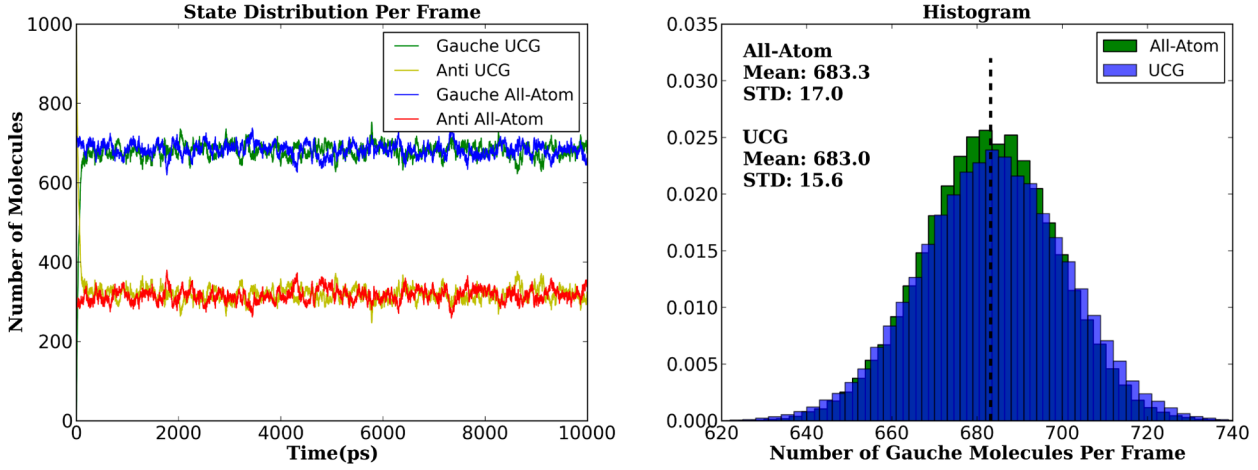


Figure 9. Comparison of UCG equilibrium dynamics to all-atom dynamics. The figure on the left shows the fluctuations of number of molecules in gauche and anti states over time. The figure on the right shows the histogram of the number of the molecules in the gauche state from the equilibrated part of the trajectory. In the histogram, the green and blue colors correspond to the UCG and the all-atom distributions, and the dark blue corresponds to the overlap in the two distributions.

see, the RDFs obtained from the UCG simulations capture all the major details of the all-atom RDFs, including the splitting of the first peak in the gauche−gauche correlation. However, small differences exist in the locations and the magnitudes of the minimums and maximums. One possible reason for these discrepancies is the neglect of small many-body effects in the UCG potential. The inclusion of three-body interactions may help to improve the agreement between the RDFs.[78] Another possibility would be to treat the DCE UCG molecules as a three-state system (see the Methods section). Either of those two additions, however, will complicate the model while only marginally improving it.

Figure 9 shows the quality of agreement in the equilibrium state dynamics. As one can judge from the plots and the histograms, the agreement is indeed good and the UCG simulations reproduce the average and the standard deviation of the state distribution. We also return here to the state correlation functions (see Figure 3) that were used to calculate the rates of transitions. Using them, we determined the gauche-to-anti and the anti-to-gauche transition rates in UCG simulations to be equal to 9.36 ns^{-1} and 20.12 ns^{-1}, respectively, which agree well with the rates found from the all-atom simulations (9.34 ns^{-1} and 20.12 ns^{-1}, respectively).

The results discussed above show that a one-bead UCG model that includes only pairwise interactions can adequately reproduce the main equilibrium properties of DCE liquid, despite its long-lived conformational multistability and the dependence of local liquid structure on the conformation of each molecule. Nevertheless, the dihedral angle distributions obtained from all-atom simulations for the gas and liquid phases (see Figure 1) may imply that capturing many-body effects will be important for developing transferable models that could describe the system well over a range of physical conditions (such as density). Such a model may need to include many-body interaction terms or many-body corrections in the state dynamics equations. However, further discussion of these effects is outside the scope of this paper and will be considered in future publications.

IV. CONCLUSIONS

In the current work, we have developed numerical methods to implement the methodology of ultra-coarse-graining (UCG), as originally discussed in the first paper in this series.[32] We used a combination of force matching for fitting state-specific configuration distributions and relative entropy minimization for fitting the distribution of states. The addition of state dynamics made it possible to simulate a 1-site UCG model for

1,2-dichloroethane liquid. The resulting model reproduces the key equilibrium properties of the all-atom system. Namely, the COM RDFs reproduced all the major features of the corresponding all-atom functions while only slightly differing from them in the positions and magnitudes of minima and maxima. The equilibrium state distribution also agreed well with the all-atom results, reproducing the averages of the state occupancies, their standard deviations, and the time correlation functions of the state. This is particularly important given that only the macroscopic transition rates were optimized toward those measured from all-atom simulations, i.e., the standard deviations were not targeted, and suggests that this modeling strategy is indeed a robust one.

In future work, we plan to apply the UCG methodology to charged systems (e.g., ionic liquids and simple polymers), as well as to elaborate more on many-body effects. Inclusion of many-body correlations may also allow us to formulate more-transferable models that could be applied over a range of physical conditions. Also, as large-scale conformational changes are common in biomolecular systems, it will be natural to consider modeling them using the UCG approach as well.

AUTHOR INFORMATION

Corresponding Author

*E-mail: gavoth@uchicago.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This material is based on work supported by the National Science Foundation through the Center for Multiscale Theory and Simulation (Grant No. CHE-1136709). We gratefully acknowledge the contribution of Andrew White, whose force-matching code ForcePy was used in the early stages of this work. We would also acknowledge Hans C. Andersen for helpful discussion on transition-rate calculations and Jacob W. Wagner, who helped implement support for UCG in the MSCGFM software package. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation (Grant No. ACI-1053575), as well as the resources of the University of Chicago Research Computing Center.

REFERENCES

(1) Allen, F.; Almasi, G.; Andreoni, W.; Beece, D.; Berne, B. J.; Bright, A.; Brunheroto, J.; Cascaval, C.; Castanos, J.; Coteus, P.; et al. *IBM Syst. J.* **2001**, *40*, 310−327.
(2) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(3) Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. *J. Comput. Chem.* **2007**, *28*, 2618−2640.
(4) Anderson, J. A.; Lorenz, C. D.; Travesset, A. *J. Comput. Phys.* **2008**, *227*, 5342−5359.
(5) Voth, G. A. *Coarse-Graining of Condensed Phase and Biomolecular Systems*; CRC Press: Boca Raton, FL, 2009.
(6) Brooks, B. R.; Brooks, C. L.; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; et al. *J. Comput. Chem.* **2009**, *30*, 1545−1614.
(7) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; et al. *Science* **2010**, *330*, 341−346.
(8) Zhmurov, A.; Dima, R. I.; Kholodov, Y.; Barsegov, V. *Proteins* **2010**, *78*, 2984−2999.
(9) Zuckerman, D. M. *Annu. Rev. Biophys.* **2011**, *40*, 41−62.
(10) Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. *Annu. Rev. Biophys.* **2012**, *41*, 429−452.
(11) Clementi, C.; Maritan, A.; Banavar, J. R. *Phys. Rev. Lett.* **1998**, *81*, 3287−3290.
(12) Tozzini, V. *Curr. Opin. Struct. Biol.* **2005**, *15*, 144−150.
(13) Wilson, M. R. *Int. Rev. Phys. Chem.* **2005**, *24*, 421−455.
(14) Ayton, G. S.; Noid, W. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2007**, *17*, 192−198.
(15) Scheraga, H. A.; Khalili, M.; Liwo, A. *Annu. Rev. Phys. Chem.* **2007**, *58*, 57−83.
(16) Stein, M.; Gabdoulline, R. R.; Wade, R. C. *Curr. Opin. Struct. Biol.* **2007**, *17*, 166−172.
(17) Liwo, A.; Czaplewski, C.; Ołdziej, S.; Scheraga, H. A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 134−139.
(18) Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869−1892.
(19) Balaz, S. *Chem. Rev.* **2009**, *109*, 1793−1899.
(20) Tozzini, V. *Q. Rev. Biophys.* **2010**, *43*, 333−371.
(21) Bahar, I.; Lezon, T. R.; Yang, L.-W.; Eyal, E. *Annu. Rev. Biophys.* **2010**, *39*, 23−42.
(22) de Pablo, J. J. *Annu. Rev. Phys. Chem.* **2011**, *62*, 555−574.
(23) Merchant, B. A.; Madura, J. D. *Annu. Rep. Comput. Chem.* **2011**, *7*, 67−87.
(24) Riniker, S.; Allison, J. R.; van Gunsteren, W. F. *Phys. Chem. Chem. Phys.* **2012**, *14*, 12423−12430.
(25) Takada, S. *Curr. Opin. Struct. Biol.* **2012**, *22*, 130−137.
(26) Saunders, M. G.; Voth, G. A. *Curr. Opin. Struct. Biol.* **2012**, *22*, 144−150.
(27) Shen, H.; Xia, Z.; Li, G.; Ren, P. *Annu. Rep. Comput. Chem.* **2012**, *8*, 129−148.
(28) Saunders, M. G.; Voth, G. A. *Annu. Rev. Biophys.* **2013**, *42*, 73−93.
(29) Noid, W. G. *J. Chem. Phys.* **2013**, *139*, 090901.
(30) Baaden, M.; Marrink, S. J. *Curr. Opin. Struct. Biol.* **2013**, *23*, 878−886.
(31) Chodera, J. D.; Noé, F. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135−144.
(32) Dama, J. F.; Sinitskiy, A. V.; McCullagh, M.; Weare, J.; Roux, B.; Dinner, A. R.; Voth, G. A. *J. Chem. Theory Comput.* **2013**, *9*, 2466−2480.
(33) Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(34) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469−2473.
(35) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.
(36) Arkhipov, A.; Freddolino, P. L.; Schulten, K. *Structure* **2006**, *14*, 1767−1777.
(37) Murtola, T.; Falck, E.; Karttunen, M.; Vattulainen, I. *J. Chem. Phys.* **2007**, *126*, 075101.
(38) Lyman, E.; Pfaendtner, J.; Voth, G. A. *Biophys. J.* **2008**, *95*, 4183−4192.
(39) Noid, W. G.; Chu, J.-W.; Ayton, G. S.; Krishna, V.; Izvekov, S.; Voth, G. A.; Das, A.; Andersen, H. C. *J. Chem. Phys.* **2008**, *128*, 244114.
(40) Noid, W. G.; Liu, P.; Wang, Y.; Chu, J.-W.; Ayton, G. S.; Izvekov, S.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2008**, *128*, 244115.
(41) Shell, M. S. *J. Chem. Phys.* **2008**, *129*, 144108.
(42) Murtola, T.; Karttunen, M.; Vattulainen, I. *J. Chem. Phys.* **2009**, *131*, 055101.
(43) Savelyev, A.; Papoian, G. A. *Biophys. J.* **2009**, *96*, 4044−4052.
(44) Savelyev, A.; Papoian, G. A. *J. Phys. Chem. B* **2009**, *113*, 7785−7793.
(45) Shell, M. S. *J. Chem. Phys.* **2012**, *137*, 084503.
(46) Carmichael, S. P.; Shell, M. S. *J. Phys. Chem. B* **2012**, *116*, 8383−8393.
(47) Davtyan, A.; Schafer, N. P.; Zheng, W.; Clementi, C.; Wolynes, P. G.; Papoian, G. A. *J. Phys. Chem. B* **2012**, *116*, 8494−8503.
(48) Jayachandran, G.; Vishal, V.; Pande, V. S. *J. Chem. Phys.* **2006**, *124*, 164902.

(49) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. *J. Chem. Phys.* **2007**, *126*, 155101.
(50) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. *J. Chem. Phys.* **2007**, *126*, 155102.
(51) Pan, A. C.; Roux, B. *J. Chem. Phys.* **2008**, *129*, 064107.
(52) Rains, E. K.; Andersen, H. C. *J. Chem. Phys.* **2010**, *133*, 144113.
(53) Lane, T. J.; Bowman, G. R.; Beauchamp, K. A.; Voelz, V. A.; Pande, V. S. *J. Am. Chem. Soc.* **2011**, *133*, 18413−18419.
(54) Lau, A. Y.; Roux, B. *Nat. Chem. Biol.* **2011**, *7*, 130−131.
(55) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. *J. Chem. Phys.* **2011**, *134*, 174105.
(56) Kellogg, E. H.; Lange, O. F.; Baker, D. *J. Phys. Chem. B* **2012**, *116*, 11405−11413.
(57) Vitalis, A.; Caflisch, A. *J. Chem. Theory Comput.* **2012**, *8*, 1108−1120.
(58) Keller, B. G.; Kobitski, A.; Jäschke, A.; Nienhaus, G. U.; Noé, F. *J. Am. Chem. Soc.* **2014**, *136*, 4534−4543.
(59) Schmidt, A.; Hall, M. N. *Annu. Rev. Cell Dev. Biol.* **1998**, *14*, 305−338.
(60) Enciso, M.; Schutte, C.; Delle Site, L. *Soft Matter* **2013**, *9*, 6118−6127.
(61) Best, R. B.; Chen, Y.-G.; Hummer, G. *Structure* **2005**, *13*, 1755−1763.
(62) Knott, M.; Best, R. B. *J. Chem. Phys.* **2014**, *140*, 175102.
(63) Sing, C. E.; Alexander-Katz, A. *Phys. Rev. Lett.* **2011**, *107*, 198302.
(64) Sing, C. E.; Alexander-Katz, A. *Macromolecules* **2011**, *44*, 6962−6971.
(65) Chen, H.; Fallah, M. A.; Huck, V.; Angerer, J. I.; Reininger, A. J.; Schneider, S. W.; Schneider, M. F.; Alexander-Katz, A. *Nat. Commun.* **2013**, *4*, 1333.
(66) Chappa, V. C.; Morse, D. C.; Zippelius, A.; Müller, M. *Phys. Rev. Lett.* **2012**, *109*, 148302.
(67) Chaimovich, A.; Shell, M. S. *J. Chem. Phys.* **2011**, *134*, 094112.
(68) Bilionis, I.; Zabaras, N. *J. Chem. Phys.* **2013**, *138*, 044313.
(69) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(70) Jorgensen, W. L.; Maxwell, D. S.; Tirado-Rives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225−11236.
(71) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474−6487.
(72) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
(73) Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 954−965.
(74) Das, A.; Lu, L.; Andersen, H. C.; Voth, G. A. *J. Chem. Phys.* **2012**, *136*, 194115.
(75) Robbins, H.; Monro, S. *Ann. Math. Stat.* **1951**, *22*, 400−407.
(76) Plimpton, S. *J. Comput. Phys.* **1995**, *117*, 1−19.
(77) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2006**, *125*, 151101.
(78) Larini, L.; Lu, L.; Voth, G. A. *J. Chem. Phys.* **2010**, *132*, 164107.