

Coarse-Grain Model for Glucose, Cellobiose, and Cellotetraose in Water

Antti-Pekka Hynninen,[†] James F. Matthews,[‡] Gregg T. Beckham,^{†,§} Michael F. Crowley,[†] and Mark R. Nimlos^{*†}

[†]National Bioenergy Center

[‡]Biosciences Center

National Renewable Energy Laboratory, Golden, Colorado 80401, United States

[§]Department of Chemical Engineering, Colorado School of Mines, Golden, Colorado 80401, United States

S Supporting Information

ABSTRACT: We present a coarse-grain (CG) simulation model for aqueous solutions of β -D-glucose, cellobiose, and cellotetraose, based on atomistic simulation data for each system. In the model, three spherical beads are used to represent glucose, and a single bead is used to represent water. For glucose, the force field is calculated using force matching by minimizing the sum of the square differences between forces calculated from atomistic and CG simulations. For cellobiose and cellotetraose, we use a hybrid method where the nonbonded interactions are obtained using force matching and the bonded interactions are obtained using Boltzmann inversion. We demonstrate excellent agreement in the structural properties between the atomistic simulations and the CG simulations. This model represents the first step in developing a CG force field for cellulose, as it is of significant interest to study cellulose behavior at much longer time and length scales relative to atomistic simulations.

INTRODUCTION

Carbohydrate polymers are the most abundant organic materials on Earth because of their prevalence in plants (cellulose, hemicellulose, pectin) and structural materials in fungi and insects (chitin). Cellulose is the linear polymer of β -D-glucose, and a significant component of the global carbon cycle and potential source of sugars for the production of transportation fuels. Most enzymatic strategies for deconstructing cellulose involve the hydrolysis of accessible glycosidic bonds between sugars.¹ However, accessing the individual linkages between sugars is a significant challenge because cellulose packs into dense crystalline fibrils. Understanding the structural nature of cellulose is thus important for determining how enzymes deconstruct plants in the biosphere and in designing enhanced properties to convert cellulose to glucose more efficiently for bioenergy processes.

To date, molecular simulation has been used extensively to understand how glucose, short polymers of cellulose, and cellulose crystals behave in solution. Most studies to date have treated cellulose and celldextrin chains with fully atomistic models.^{2–18} A few studies to date have developed coarse-grained (CG) models of varying resolutions for studying monosaccharides and cellulose.^{19–23} However, there still remains a significant need to develop improved CG models for cellulose (and other carbohydrate polymers) because of the large number of atoms and the long time scales required to study these polymers. As pointed out recently for simulations of cellulose crystals, it is unlikely that any MD simulations to date have been conducted long enough (hundreds of nanoseconds) to equilibrate.^{10,24} It is quite likely that CG models will be necessary to study interesting problems such as (1) phase transitions in cellulose,²⁵ (2) microfibril dissolution and behavior in nonaqueous solvents,^{26–28} (3) material properties

of long cellulose microfibrils and cellulose chains in nonaqueous solvents,²⁹ (4) the nucleation of cellulose crystals during biosynthesis,^{30,31} (5) microfibril aggregation, (6) interactions of cellulase and cellulosomal systems with the plant cell wall during biological deconstruction,^{32–34} and (7) the interactions of cellulose with hemicellulose, pectin, and lignin as a first step in developing models of the whole plant cell wall.

There are multiple methods for developing CG models in the literature.³⁵ One class includes methods based on the inverse Boltzmann rule where the probability distribution (or the radial distribution function) obtained from the atomistic simulations is matched using the CG model. This class includes the inverse Monte Carlo³⁶ and the iterative inverse Boltzmann methods.³⁷ In another class of methods, called force-matching methods, the forces calculated from the atomistic model are matched by the CG model.^{38,39} The force-matching method was originally developed by Ercolessi and Adams for producing atomistic interaction potentials from *ab initio* simulations.⁴⁰ Typically, the matching is done by minimizing the sum of squared differences in forces generated by atomistic and CG simulations. Izvekov and Voth termed this method multiscale coarse-graining (MS-CG) because the CG model is constructed using information from the atomistic scale.

Force matching has emerged as a powerful method to develop the force fields necessary to conduct CG MD simulations. However, many of these studies⁴¹ focus on systems that only require two-body interactions. Going toward more complex systems with angles and dihedrals, Noid et al. used force matching to model ethylmethylimidazolium (EMIM^+)/ NO_3^-

Received: February 8, 2011

Published: June 07, 2011

ionic liquid.⁴² In their work, the EMIM⁺ molecule was modeled using four CG beads, and the NO₃⁻ molecule was modeled using a single CG bead. In addition, Izvekov and Voth have recently applied the force-matching method on lipid bilayers.^{43,44} In particular, the work in ref 43 shows that the force-matching method can be used to coarse-grain out the solvent to obtain a solvent-free model. Application of the force-matching method to ion channels shows that the method can be used to calculate the interactions for a mixed system with atomistic and CG parts.⁴⁵ In a recent article, Lu et al.⁴⁶ discuss the various options for the implementation of the force-matching method, which has been extended for transfer between temperatures,⁴⁷ the isothermal-isobaric ensemble,⁴⁸ and three-body potentials.⁴⁹

Using their own method that is based on reproduction of atomistic structural and thermodynamic properties, Molinero and Goddard developed a three-bead model to describe α -D-glucose and mapped the forces from atomistic simulations on the CG model.^{20,50} They used this approach to model glucose in aqueous solution and in glasses. Their CG model yields good agreement with the atomistic simulations for density, structural properties, and cohesive energy. Molinero and Goddard later used their CG model to study water diffusion in glucose glasses.⁵⁰ Liu et al.²¹ used a model similar to that used by Molinero and Goddard to coarse-grain α -D-glucose. In their method, the bonded interactions were fitted to harmonic potential functions, whereas the nonbonded interactions were tabulated and calculated using force matching. Liu et al. obtained good structural agreement between the CG and the atomistic simulations. Furthermore, they showed that their CG model could be applied to other temperatures and pressures than where it was originally derived.

Additionally, our group recently developed a preliminary CG force field to study the hydrophobic face of crystalline cellulose in the presence of an atomistic carbohydrate-binding module.^{23,51} The original model for cellulose was based on the glucose CG model from Molinero and Goddard,²⁰ and Boltzmann inversion was used to fit the bond, angles, and dihedral parameters. For the nonbonded terms, rescaling factors were applied to a standard 6–12 Lennard-Jones potential to represent the directionality in a cellulose I β crystal from atomistic simulations. The solvent was represented by the generalized Born model with switching (GBSW).^{52,53} The CG model was found to agree well with the atomistic simulations, except in the a lattice parameter (or layer spacing) for cellulose I β . In our previous studies, we used a multiscale model to study enzyme behavior on the hydrophobic face of cellulose I β , and thus this model was adequate for the intended purposes. We used this model to examine the behavior of a Family 1 carbohydrate-binding module (CBM) on cellulose and showed that the CBM has regions of thermodynamic stability along a cellulose chain on the surface every 1 nm, which was later confirmed with an atomistic study.⁵⁴

Recently, Wohlert and Berglund developed a coarse-grained force field for native cellulose within the MARTINI force field suite.⁵⁵ They parametrized their model with partition free energy calculations for cellobextrin oligomers between water and cyclohexane and then extended the parameters to model cellulose I β . They found that to maintain the staggered conformation of origin and center chains in cellulose I β ,⁵⁶ an additional repulsive term was needed in the force field between certain beads in cellulose. From this, they conducted MD simulations of a cellulose microfibril and compared the lattice parameters to the experimentally determined structure.⁵⁶ They found good agreement down the length of the cellulose chains but observed

substantial deviation from the a and b lattice parameters relative to the experimental structure. The authors used their model to study the surface diffusion of the Family 1 Cel7A CBM and found good agreement with the experimentally measured diffusion coefficient for a Family 2 CBM. However, the authors pointed out that, similar to the model of Bu et al.,²³ their model is not useful for studying cellulose microfibril properties beyond surface properties.

In this article, we describe the development of coarse grain models of β -D-glucose, cellobiose, and cellotetraose as a first step to developing a CG model for cellulose. Glucose is the smallest repeat unit of a cellulose chain, and a CG model of this molecule serves as a logical starting point toward developing a model of cellulose. Cellobiose (1,4-O- β -D-glucopyranosyl-D-glucose) is the conformational monomer in crystalline cellulose, and cellotetraose is the smallest cello-oligomer that has an internal cellobiose unit. Thus, we anticipate that the force field for cellotetraose will be a reasonable first iteration for a CG force field of cellulose. The method of Noid et al.^{42,57} was originally applied directly for each of these molecules. However, we found that this approach is only useful for glucose, and a hybrid of force matching and Boltzmann inversion was needed for cellobiose and cellotetraose.

METHODS

Atomistic Simulations. All of the molecular dynamics (MD) simulations were performed using CHARMM.⁵⁸ For all three atomistic systems (glucose, cellobiose, and cellotetraose), the molecules were initially placed on a cubic grid and solvated with TIP3P water.⁵⁹ The system was then equilibrated using constant pressure simulations. The production runs were conducted in the canonical ensemble where the size of the simulation box was set to be equal to the average density from the constant pressure simulations. The time step in the atomistic MD simulations was 2 fs, and the nonbonded interaction cutoff distance was 12 Å. The electrostatic interactions were calculated using the particle mesh Ewald method with sixth-order spline, $\kappa = 0.32$, and a $32 \times 32 \times 32$ grid.⁶⁰ The SHAKE algorithm was used to fix the bond lengths to hydrogen atoms.⁶¹ The carbohydrates were modeled using the C35 force field.^{62,63} The atomistic MD simulations of glucose consisted of 64 glucose molecules and 551 TIP3 water molecules (total 3189 atoms) in a cubic box with side length of 30.79 Å. The length of the production run was 120 ns. The atomistic MD simulations of cellobiose included 32 cellobiose molecules and 1169 water molecules (total 4947 atoms). The production run length was 240 ns. The atomistic MD simulations of cellotetraose consisted of 32 cellotetraose molecules and 4954 water molecules (total 17 646 atoms). The production run length was 60 ns.

Coarse-Grain Model. In our coarse-grain model, each glycan is represented with three beads (AT, BB, C), and each water molecule is replaced with one bead (W) as originally used by Molinero and Goddard²⁰ and later by Liu et al.²¹ for modeling glucose and Bu et al.²³ for modeling cellulose. Figure 1a shows the arrangement of the coarse grain beads in glucose and the corresponding atoms.

Generally, the CG force field consists of nonbonded pair interactions, bonded pair interactions, angular three-body interactions, and dihedral four-body interactions and is represented numerically using lookup tables. This allows the interaction potentials to be of any shape, and no functional form is assumed. As described in the Results section, this feature is essential in capturing the nonharmonic nature of the bonded interactions and the shape of the nonbonded pair interactions. In all three

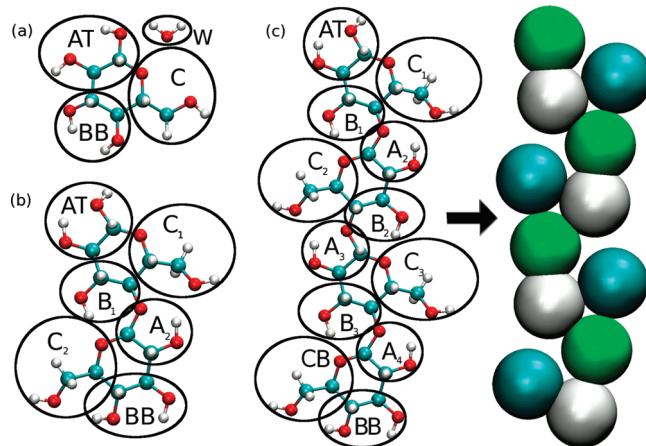


Figure 1. Atomistic structure of (a) glucose and water, (b) cellobiose, and (c) cellotetraose with definition of the CG beads. For cellotetraose, the mapping to CG beads is indicated by the arrow.

models, each CG bead is positioned at the center-of-mass of the atoms that correspond to the beads, and the mass is set to be the sum of the masses of these atoms. We did not include explicit point charges in the CG model. Instead, the Columbic and van der Waals interactions are joined together in the CG force field. Throughout the description of the beads in the model, we use subscripts to differentiate specific beads of the same type (e.g., C_1 and C_2 are the same bead type but have different locations in cellobiose and cellotetraose).

In the CG model for glucose, there are bonded interactions between all three beads and nonbonded interactions between all beads in *separate* glucose molecules. There are no nonbonded interactions between beads in the same CG glucose molecule because hydrogen bonding interactions between the hydroxyl groups on a glucose residue are incorporated into the bonded interactions. The nonbonded interactions between each glucose bead and water were also calculated.

Figure 1b shows the atomistic structure of cellobiose and the definition of the CG beads. To differentiate between the reducing and nonreducing ends of cellobiose, we use bead names AT, or “A top”, and BB, or “B bottom”. We note that the B_1 bead in cellobiose has one oxygen atom compared to glucose, where it has two. In the CG cellobiose model, there are bonded interactions between all neighboring beads, i.e., $AT-C_1$, $AT-B_1$, B_1-C_1 , B_1-A_2 , C_2-A_2 , C_2-BB , and A_2-BB . In addition, the model contains a bonded interaction between C_1 and C_2 beads to mimic the nonbonded interaction. The model also has an angular potential between beads $C_1-B_1-A_2$ and $C_2-A_2-B_1$ and dihedral potentials $B_1-C_1-C_2-A_2$, $AT-B_1-C_1-A_2$, and $BB-A_2-C_2-B_1$. In the cellobiose CG model, water is also represented using a single bead, and the nonbonded interactions are only included between beads in different cellobiose units.

The CG model for cellotetraose is shown in Figure 1c, and the definition of the beads is similar to the cellobiose model. The only difference is that we introduce an additional bead type, CB or “C bottom”. That is, for cellotetraose, we distinguish CB from C_1 , C_2 , and C_3 because we noticed that this model gave slightly better agreement with the atomistic results than where all four C beads were of the same type. Note that cellotetraose has two glycans (2 and 3 in Figure 1c) in the core of the molecule that will not have reducing end or nonreducing end interactions. Thus,

the force field for these middle glycans may be more similar to what is found in the core of cellulose chains.

In setting up the bonded interactions for the CG models, one must be careful to choose the set of bonds, angles, and dihedrals that retain the correct atomistic molecular configuration; i.e., if one or more bonds are missing, the CG model will have degrees of freedom that are not constrained by the force field. Such “underdetermined” systems have internal degrees of freedom that are not constrained in the same way as they are in the atomistic system. For example, if we were to take away the $B-C-C'-A$ dihedral interaction from the cellobiose model, the system would freely rotate around its dihedral. Force matching would still give us potentials that look reasonable, but the behavior of the system would not correspond to the atomistic model for cellobiose. One must also be careful not to over-determine the interactions. Interactions are overdetermined if one of the bonds, angles, and dihedrals can be left out and the structure can still be uniquely identified. Overdetermined sets of bonded interactions typically give rise to interaction potentials that do not exhibit minima even though the total potential energy still has a minimum. In the case of glucose, setting up the bonded interactions is simple. Clearly, the three bonds uniquely define the molecular configuration, and none of them can be left out. One way to overdetermine the CG glucose model would be to add an angle between A, B, and C beads. In the case of cellobiose, setting up the bonded interactions is more complicated, and the set of bonded interactions we use is not unique. We developed this set by trying seven different combinations of bonded interactions. In this trial-and-error process, we calculated the bonded interaction potentials using the force-matching method. Some of the sets were discarded because they were overdetermined and yielded interaction potentials that do not have a minimum. Some others were discarded because they were underdetermined.

Force-Matching Method. The CG force field is represented by a lookup table where the forces are defined on a linear grid with a fixed bin width. We use N_D to denote the total number of bins in all lookup tables. That is, N_D includes all lookup tables for both bonded and nonbonded potentials.

In the force-matching method, the forces between CG beads are matched to the force data from atomistic MD simulations using a least-squares fit. As shown by Noid et al.,⁴² the least-squares problem reduces to a linear equation of the form

$$\mathbf{g}\mathbf{x} = \mathbf{f} \quad (1)$$

where \mathbf{g} is a matrix of size $3Nn_t \times N_D$, \mathbf{x} is a vector of length N_D , and \mathbf{f} is a vector of length $3Nn_t$. Here, N is the number of CG beads, and n_t is the number of frames in the atomistic trajectory. Vector \mathbf{f} consists of forces acting on each CG bead in each frame, which is obtained from the atomistic trajectory. Vector \mathbf{x} is the unknown, consisting of the table lookup values. That is, once eq 1 is solved, \mathbf{x} contains the CG force field. To explain the structure of matrix \mathbf{g} , we consider a simple case of two particles A and B connected by a bond. In this case, the elements of \mathbf{g} are given

$$g_{A,d}^t = \frac{c_A - c_B}{r_{AB}} \theta(r_{AB} - r_d) \quad (2)$$

where c is one of the Cartesian coordinates $c = x, y, z$; t is the frame $t = 1, \dots, n_t$; r_{AB} is the distance between particles A and B; $r_d = r_{\min} + \delta x(d-1)$ is the bin location, $d = 1, \dots, N_D$; and θ is a δ function with width δx . To reduce the size of matrix \mathbf{g} , one can normalize eq 1 by multiplying it on both sides with \mathbf{g}^T .

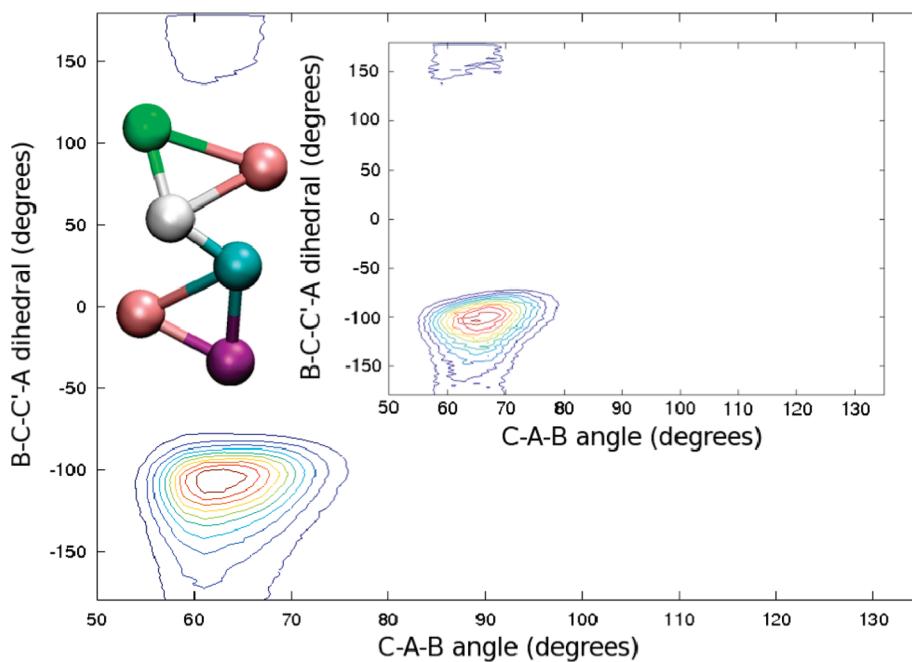


Figure 2. Contour plots of the probability distribution of the $B_1-C_1-C_2-A_2$ dihedral angle versus the $C_2-A_2-B_1$ angle from MD simulations of cellobiose. The main figure shows the result for the CG model where bonded interactions are obtained using Boltzmann inversion. The inset shows the result for the atomistic model. The snapshot shows the configuration of the cellobiose corresponding to the basin of attraction at $(60^\circ, -100^\circ)$.

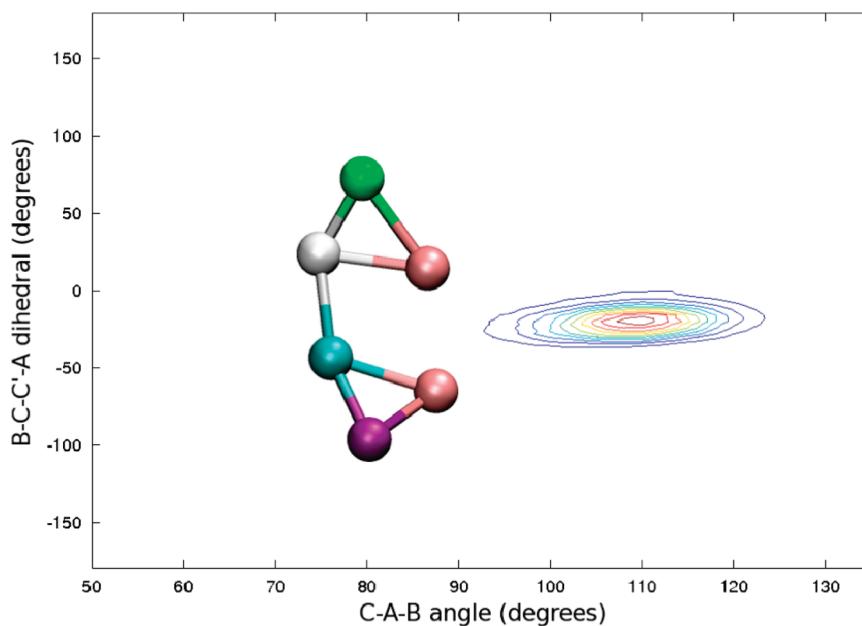


Figure 3. Contour plots of the probability distribution of the $B_1-C_1-C_2-A_2$ dihedral angle versus the $C_2-A_2-B_1$ angle from MD simulations of cellobiose showing the result for the CG model where the bonded interactions are obtained from force matching. The snapshots correspond to the basin of attraction at $(110^\circ, -25^\circ)$.

Using the notation $\mathbf{G} = \mathbf{g}^T \mathbf{g}$ and $\mathbf{b} = \mathbf{g}^T \mathbf{f}$, we obtain

$$\mathbf{Gx} = \mathbf{b} \quad (3)$$

where now \mathbf{G} is an $N_D \times N_D$ matrix and \mathbf{b} is a vector of length N_D . Noid et al.⁴² showed how to construct \mathbf{G} and \mathbf{b} without constructing the (usually very large) matrix \mathbf{g} . The advantage of using eq 3 versus eq 1 is that the size of matrix \mathbf{G} is usually much

smaller than that of matrix \mathbf{g} . This is because the size of matrix \mathbf{g} increases linearly with the size of the atomistic trajectory.

Finally, once \mathbf{x} is solved from eq 3, the force between particles A and B at distance r is given by

$$\mathbf{f}_{AB}(\mathbf{r}) = \sum_d x_d \theta(r - r_d) \frac{\mathbf{r}}{r} \quad (4)$$

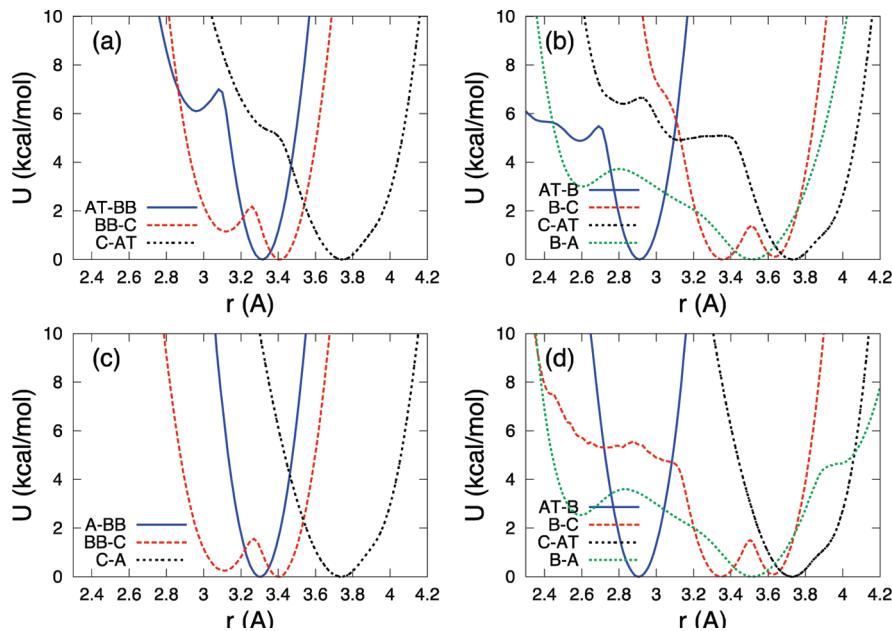


Figure 4. Bonded pair potentials for (a) glucose, (b and c) cellobiose, and (d) cellotetraose CG models. In b, “B–A” refers to the “ B_1 – A_2 ” bond between glucose residues, as shown in Figure 1 (b). Similarly, in d, “B–A” refers to the “ B_1 – A_2 ”, “ B_2 – A_3 ”, and “ B_3 – A_4 ” bonds between glucose residues, as shown in Figure 1c.

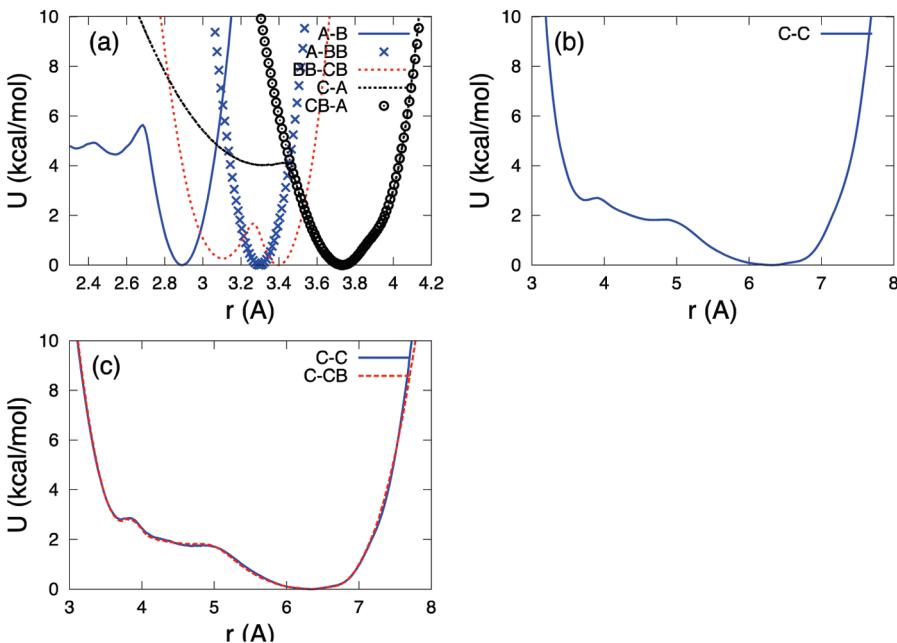


Figure 5. Bonded pair potentials for (a) cellobiose, (b) cellobiose, and (c) cellotetraose CG models. In a, “A–B” refers to bonds “ A_2 – B_2 ” and “ A_3 – B_3 ” within the glucose residue, as shown in Figure 1c.

where x_d is the d th element of vector \mathbf{x} , and the role of the δ function θ is to select the contribution at the correct distance range $r = r_d - \delta x/2, \dots, r_d + \delta x/2$. A more detailed description on how to construct \mathbf{G} and \mathbf{b} is given in the original article by Noid et al.⁴²

We wrote a C-language program to perform the force matching. The program takes as input the atomistic PSF (CHARMM protein structure file) file, atomistic trajectory file, CG PSF file, and a file defining which atom belongs to which CG bead. The program then goes through the trajectory and builds the \mathbf{G} matrix and \mathbf{b} vector.

This part of the program is by far the most time-consuming. Therefore, we parallelized the program such that it splits up the trajectory evenly for all CPUs, and each CPU builds \mathbf{G} and \mathbf{b} for their part of the trajectory. At the end, the program combines the \mathbf{G} and \mathbf{b} matrices and vectors from all CPUs into the final result. Finally, the linear equation in eq 1 is solved using singular value decomposition (SVD), which was implemented using the CLAPACK library.⁶⁴

Boltzmann Inversion. We found that the direct application of force matching for cellobiose and cellotetraose produces bonded

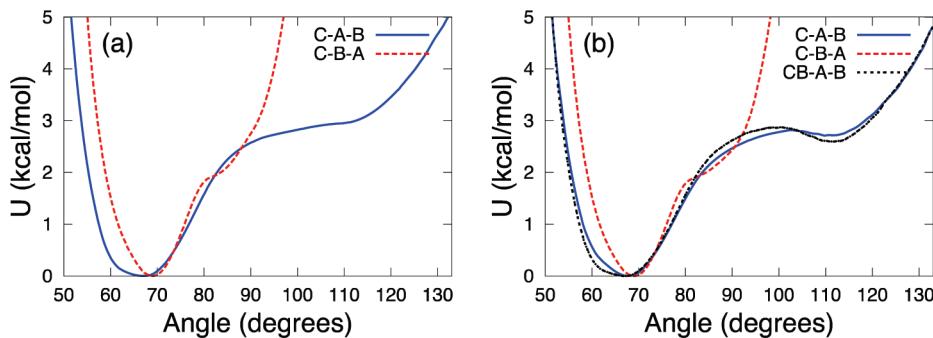


Figure 6. Angular potential for (a) cellobiose and (b) cellotetraose CG models.

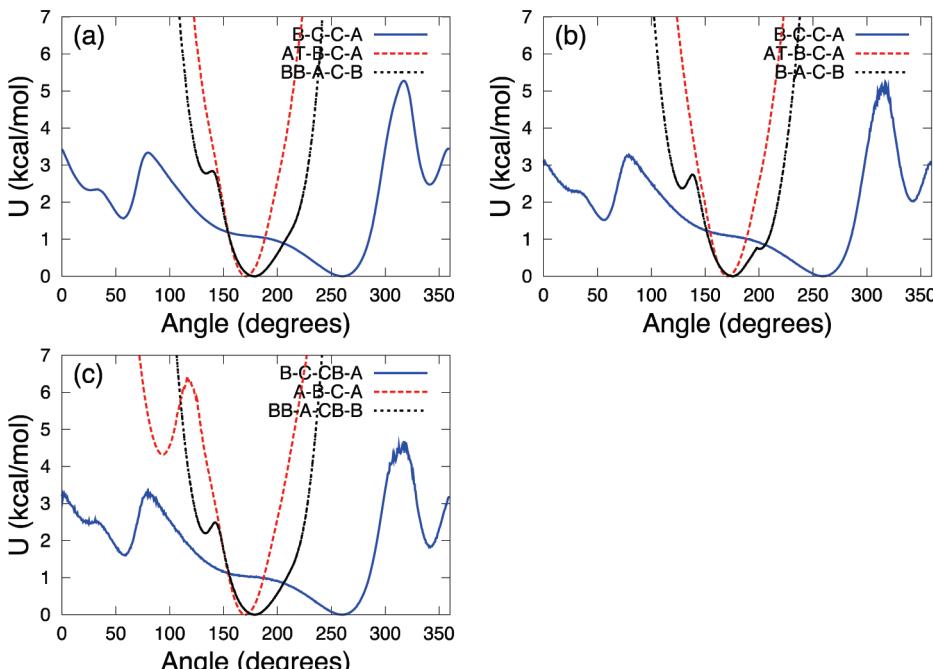


Figure 7. Dihedral potentials for (a) cellobiose and (b and c) cellotetraose CG models.

interactions that give rise to nonphysical conformations. The reason for this is that the force-matching method retains the correlations between bonded interactions, but these correlations are not taken into account in the subsequent MD simulations of the CG model. This problem is highlighted in Figure 2, where we plot contours of the probability distribution of the $B_1-C_1-C_2-A_2$ dihedral angle as a function of the $C_2-A_2-B_1$ angle from MD simulations of cellobiose. The inset in Figure 2 shows the contour plot from atomistic MD simulations. Comparing the inset in Figure 2 with Figure 3, which shows the contour plot from CG MD simulations where the bonded interactions are obtained using the force-matching method, we clearly see that CG simulations do not agree with the atomistic simulations. Instead, the CG simulations favor the “flipped” configuration shown in the snapshot in Figure 3.

To work around this issue, we used a hybrid method where the bonded interactions are obtained from Boltzmann inversion and the nonbonded interactions are obtained from force matching. This approach has been used previously by Liu et al.²¹ in the coarse-graining of glucose. In Boltzmann inversion, we first calculate the histograms $h(r)$ for each bond length, angle, and

dihedral angle and then apply the following equation:

$$U(r) = -k_B T \ln[h(r)] \quad (5)$$

to obtain the interaction potential $U(r)$, where k_B is the Boltzmann constant and T is the temperature. Finally, we use numerical derivatives to calculate the force from the potential in eq 5. In the hybrid method, we first calculate both the bonded and the nonbonded interactions using force matching. We then replace the force-matched bonded interactions with the Boltzmann inverted bonded interactions. Note that there is a minor inconsistency here: the nonbonded interactions are no longer completely consistent since the force-matched bonded interactions have been exchanged for the Boltzmann inverted bonded interactions. To check that our approach is accurate, we again conducted force matching with a new force data set, where the histogram-based bonded interactions were subtracted. The result is a new set of nonbonded interactions that was then used with the histogram-based bonded interactions. We then ran CG MD simulations using this new force field. For both cellobiose and cellotetraose, the results were indiscernible from the results with the original, inconsistent hybrid force

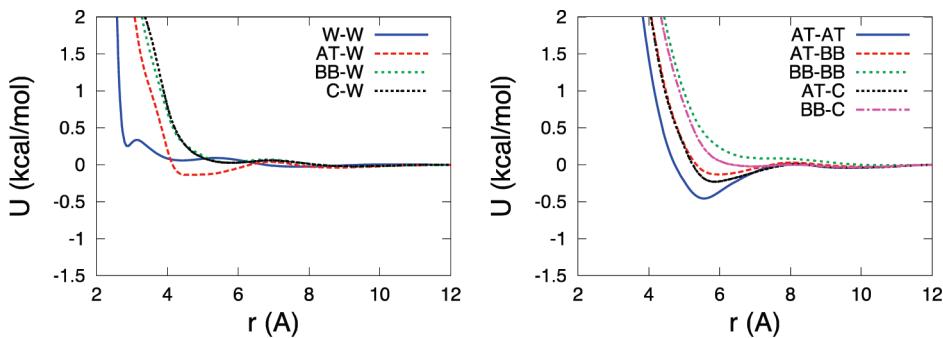


Figure 8. Nonbonded pair-potential between CG water (W) and beads AT, BB, and C, in the CG glucose model.

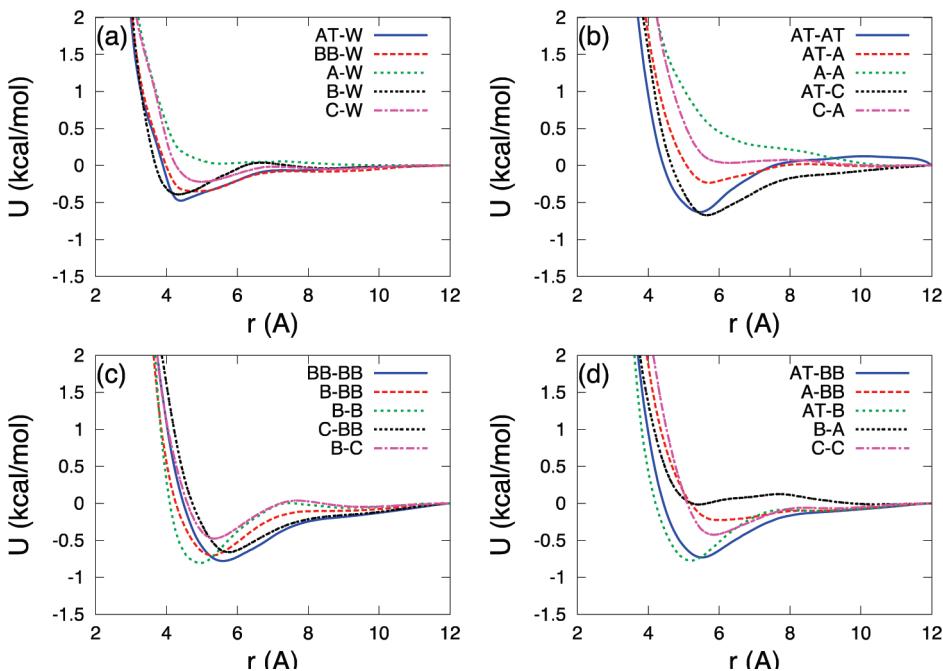


Figure 9. Nonbonded interactions in CG cellobiose model.

field without the subtraction of the histogram-based bonded interactions. Therefore, we decided to use the original hybrid force field.

Figure 2 shows the contour plot for the hybrid method. Comparing the main figure with the atomistic result in the inset, we see that the hybrid method results in cellobiose configurations that closely agree with the atomistic data. In summary, the coarse-graining process for cellobiose and cellotetraose was as follows:

- Step 1 Perform atomistic simulations
- Step 2 Calculate nonbonded CG force field using force matching
- Step 3 Calculate bonded CG force field using Boltzmann inversion
- Step 4 Perform the CG simulations using the force fields from steps 2 and 3

Implementation in CHARMM. To perform the CG MD simulations using CHARMM, we modified CHARMM to use lookup tables for bond length, angular, and dihedral interactions. To prevent CHARMM from assigning nonbonded interactions between beads in the same cellobiose or cellotetraose unit, we added “zero-force” bonded interactions to the model. These are simply bonded force lookup tables that are set to zero at all distances. For cellobiose, we

added AT–A₂, AT–BB, and B₁–BB bonds and AT–BB–C, B₁–BB–C, and C–B₁–BB angles. For cellotetraose, we added AT–A₃, AT–A₄, B₁–B₂, B₁–B₃, B₂–BB, C₁–B₂, C₁–B₃, C₁–B₄, C₃–B₁, A₃–C₁, A₄–C₁, C₁–BB, C₂–BB, C₃–BB, A₁–A₂, A₁–A₃, A₁–CB, B₁–CB, B₂–CB, B₃–CB, C₁–BB, C₂–BB, and C₃–BB bonds and AT–A₂–B₂, AT–A₂–B₃, AT–A₂–BB, AT–A₂–C₂, AT–A₂–C₃, AT–A₂–CB, B₁–B₂–A₃, B₁–B₂–A₄, C₁–AT–C₃, C₁–AT–CB, A₂–AT–B₃, A₂–AT–BB, B₂–AT–A₄, and A₃–AT–CB angles.

In our force-matching program, the nonbonded force field was represented using lookup tables with bin width $\Delta x = 0.05 \text{ \AA}$ and low and high cutoffs of 2.3 \AA and 12.0 \AA , respectively. The bonded, angular, and dihedral force lookup tables had bin widths of $\Delta x = 0.02 \text{ \AA}$, $\Delta\theta = 1.0^\circ$, and $\Delta\varphi = 1.0^\circ$, respectively. The program automatically detects and sets the upper and lower bounds of bonded, angular, and dihedral force tables.

RESULTS AND DISCUSSION

Coarse-Grained Force Fields. Figures 4 and 5 show the bonded pair potentials for glucose, cellobiose, and cellotetraose.

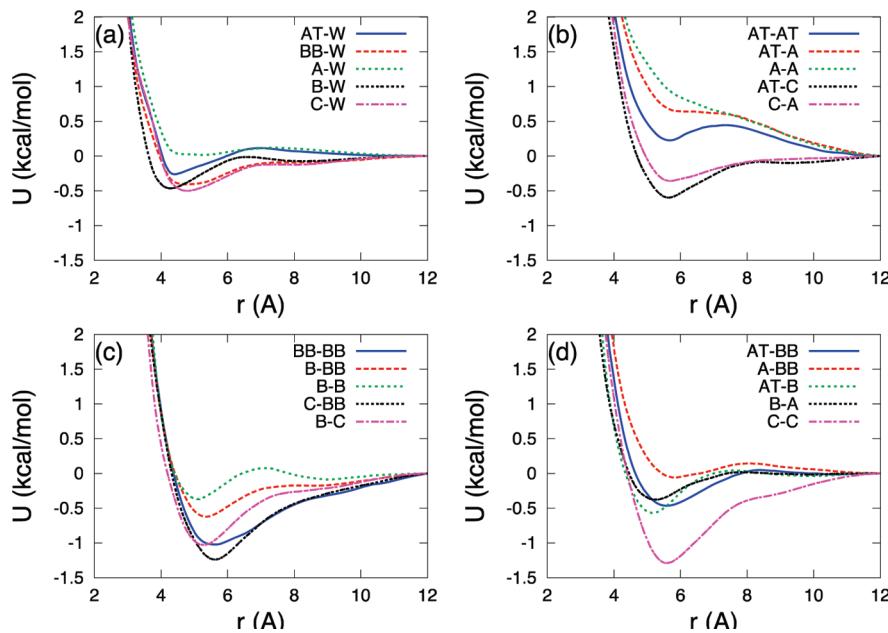


Figure 10. Nonbonded interactions in the CG cellotetraose model.

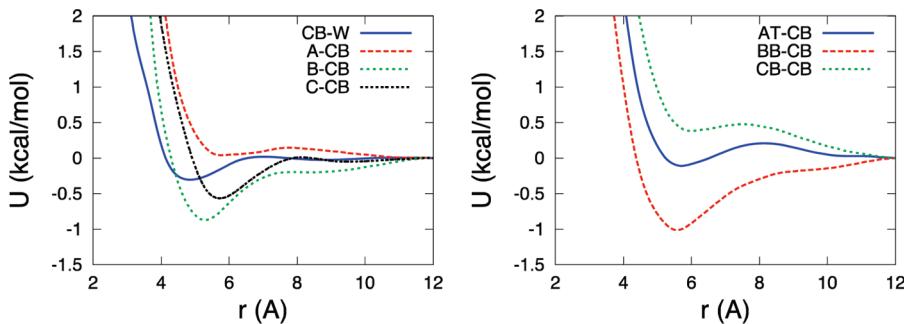


Figure 11. Nonbonded interactions in the CG cellotetraose model (continued from Figure 10).

We are showing the postprocessed potentials where we have added repulsive barriers on the edges of the potentials. In practice, this was done via a linear extrapolation of the force data. This gives rise to quadratic repulsion in the potential energy.

To facilitate a comparison between the pair potentials, the graphs in Figures 4 and 5 are plotted on the same axis scale. The location of the global minimum of the glucose AT–BB potential matches with the cellobiose and cellotetraose A–BB potentials, but does not match with the cellobiose and cellotetraose AT–B or A–B potentials. This can be explained by the difference in the center of mass between the AT and A and the BB and B beads. The AT bead only has an additional hydrogen atom compared to A; hence the center of mass of AT is very similar to that of A. Compared to B, the BB bead has an additional oxygen and hydrogen atoms; hence the center of mass of BB is different from that of B.

The B–A bonded potential is shown Figure 4b and d for cellobiose and cellotetraose, respectively. We see that the B–A potential agrees well between cellobiose and cellotetraose, both having a global minimum at around 3.5 Å, and a local minimum at around 2.6 Å. The C–C bonded pair potential is shown in Figure 5b and c for cellobiose and cellotetraose, respectively. As

can be seen, the C–C potential agrees well between cellobiose and cellotetraose. The minimum in the C–C potential is at 6.3 Å, which compares well with the minimum in the crystalline cellulose where the equilibrium distance is 6.7 Å.

The double well behavior shown in Figure 4a for the BB–C interaction is due to the change in conformation of the primary alcohol group in the atomistic simulations. This behavior is seen in multiple cases for glucose, cellobiose, and cellotetraose in Figure 4b–d and in Figure 5a. The local minimum in Figure 4a for the AT–BB interaction is due to the carbon ring puckering. The same effect of ring puckering is seen in Figure 4b and Figure 5a for cellobiose and cellotetraose. Ring puckering is a relatively rare event occurring approximately once for one of the 64 glucose molecules during a 20 ns simulation. Therefore, the lack of puckering in Figure 4c and d is most likely due to inadequate sampling in these rarely visited regions.

Figure 6a and b show the angular potentials for cellobiose and cellotetraose, respectively. Comparing the figures, we see that the angular potentials for the two models are in good agreement.

Figure 7a shows the dihedral potentials for cellobiose, and Figure 7b,c show the dihedral potentials for cellotetraose. Comparing Figure 7a–c, we see that these dihedral potentials

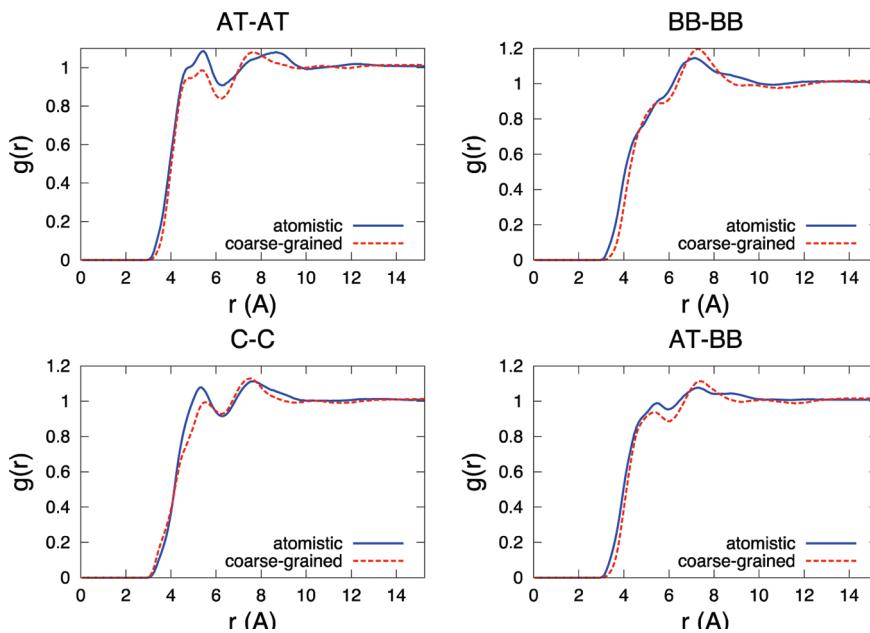


Figure 12. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of glucose for AT–AT, BB–BB, C–C, and AT–BB bead pairs.

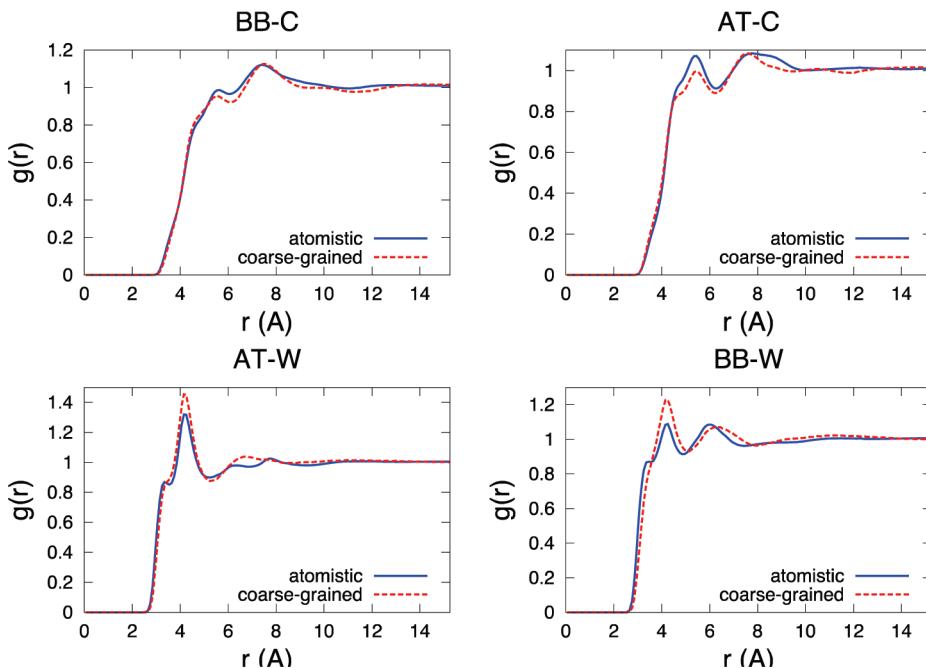


Figure 13. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of glucose for BB–C, AT–C, AT–W, and BB–W bead pairs.

agree quite well between cellobiose and cellotetraose. The only major difference is in the A–B–C–A potential shown in Figure 7c, which has a secondary local minimum at around 90°.

It is interesting to note that, as can be seen from Figures 4–7, almost all bonded potentials are nonharmonic and therefore cannot be fitted with an analytical harmonic function such as has been done in some previous work.^{20,21}

Figure 8 shows the nonbonded interaction potentials for the CG glucose model. As shown in Figure 8, the nonbonded

interaction potentials cannot be easily described by a functional potential, such as the Lennard-Jones potential.

Figure 9 shows the nonbonded interactions in the CG cellobiose model. Figures 10 and 11 show the nonbonded interactions for the CG cellotetraose model. For the cellobiose force field to be suitable for use as a force field for CG cellulose, it should have nonbonded interactions that are similar to the nonbonded interactions in the cellotetraose model. It is therefore interesting to compare the nonbonded potentials in the cellobiose

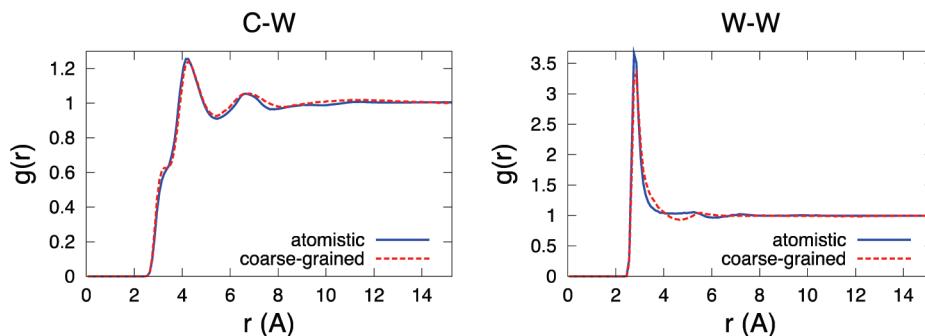


Figure 14. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of glucose for C–W and W–W bead pairs.

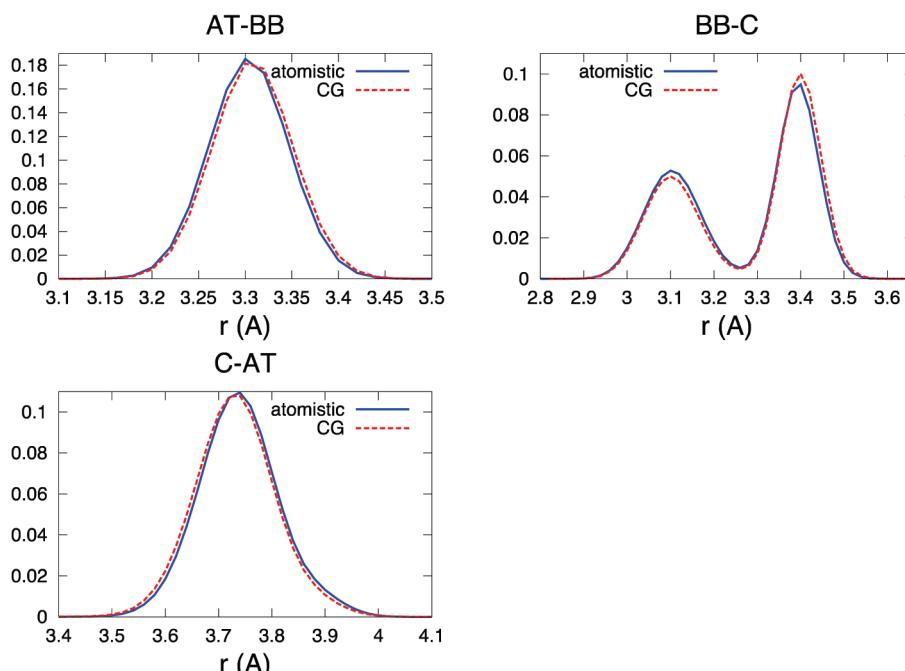


Figure 15. Histograms of bonds from the atomistic (solid line) and CG (dashed line) simulations of glucose.

and cellooligosaccharide models. We make the comparison for the B–A, B–C, and C–A nonbonded potentials. These potentials are shown in Figure 9b–d and Figure 10b–d for cellobiose and cellooligosaccharides, respectively. For example, we compare the B–C interaction potential. As shown, the B–C potential has an attractive well around 5.5 Å for both cellobiose and cellooligosaccharides. However, the depth of the well for cellooligosaccharides is about twice as much as it is for cellobiose. A similar comparison for B–A and C–A nonbonded potentials shows the same trend where the cellooligosaccharide potential is always more attractive than the cellobiose potential.

Comparison of Coarse-Grain and Atomistic Simulations. To validate the coarse-grain models, we compare the radial distribution functions (RDF) from the atomistic and coarse-grain simulations. The atomistic RDF was calculated by mapping the CG beads onto the atomistic trajectory. To avoid the RDFs being dominated by the bonded interactions, we exclude the contributions from beads in the same glucose, cellobiose, or cellooligosaccharide units.

Figures 12–14 show the comparison for glucose. As we can see, the CG and atomistic results are in good agreement for glucose.

Figure 15 shows the histograms of the bonds AT–BB, BB–C, and C–AT from the atomistic and CG simulations of glucose. As we can see, the agreement between the atomistic and CG simulations is excellent.

Figures 16 and 17 show selected RDFs for cellobiose and cellooligosaccharides, respectively. As we can see, the agreement with the atomistic results is not as good for cellobiose and cellooligosaccharides than for glucose. From Figures 16 and 17, we see that the agreement is very good for the RDF between the C bead and CG water (W). This is true in general for both cellobiose and cellooligosaccharides; i.e., the agreement is very good for all RDFs that are between a bead and CG water (including the water–water RDF). The rest of the RDFs for both cellobiose and cellooligosaccharides are shown in the Supporting Information.

Our eventual goal is to create a CG model for cellulose. To model the cellulose fibril crystal structure correctly, one has to correctly model the intrachain A–B, A–C, and B–C interactions. Figure 17 shows the comparison between the atomistic and CG RDFs for A–B, A–C, and B–C, respectively. As can be seen from the figures, the location of the first peak matches well between the two results. This suggests that our CG force field for

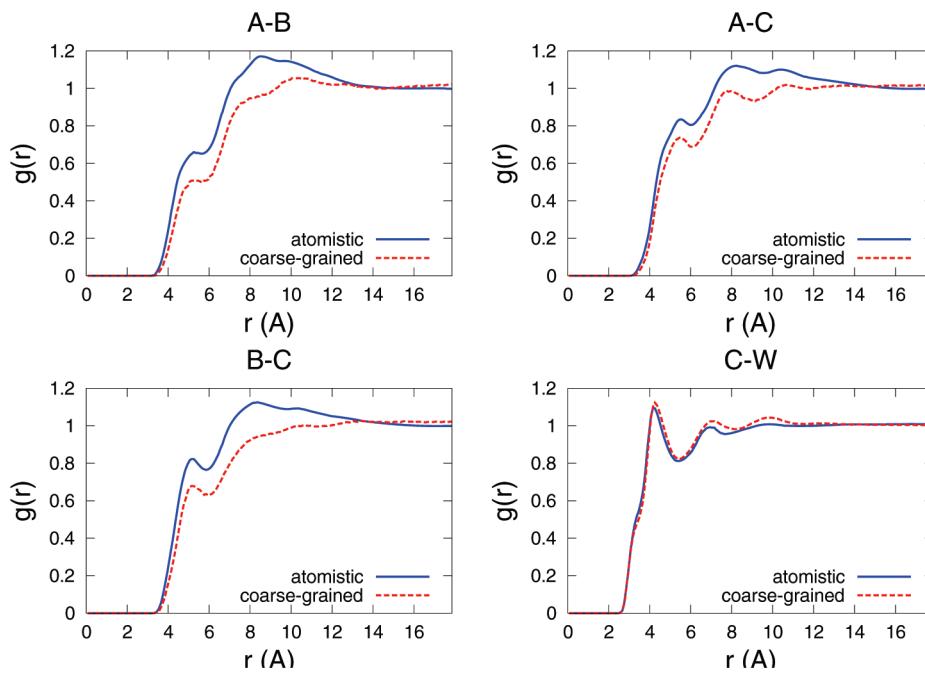


Figure 16. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of cellobiose.

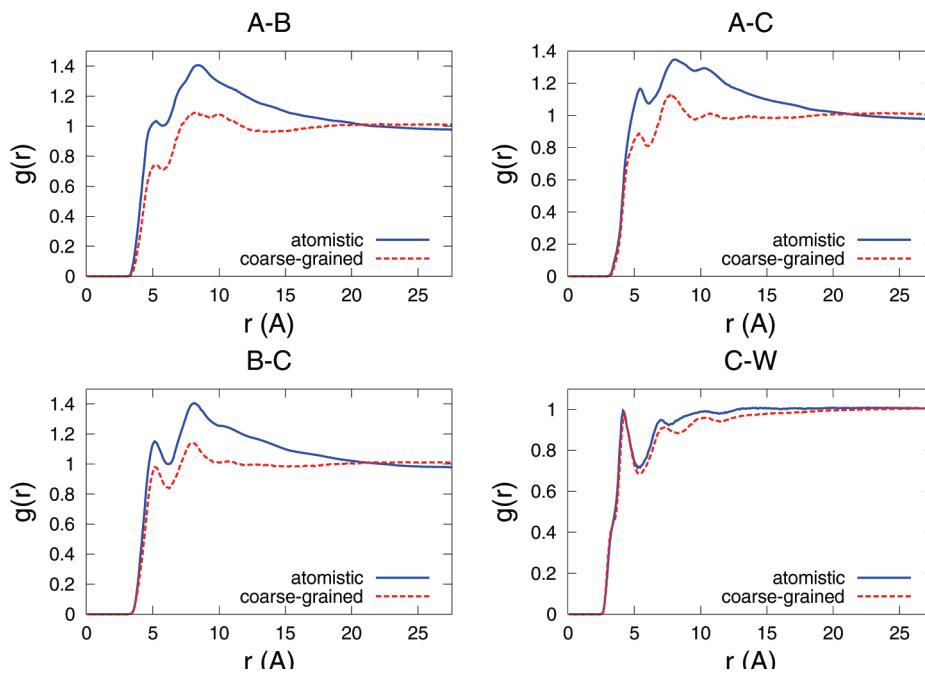


Figure 17. Radial distribution functions from atomistic (solid line) and CG (dashed line) simulations of cellotetraose.

cellotetraose has the potential for correctly reproducing the cellulose fibril crystal structure, which we will examine in a future study.

Figure 18 shows the B–C–C–A dihedral angle probability distribution for cellotetraose from atomistic and CG simulations. As can be seen, the shapes of the distributions match fairly well at the most stable configuration, around -100° . We also see that the finer details of the distribution, such as the metastable configuration at -25° , is not reproduced by the CG force field.

Finally, we make a comparison for the end-to-end distance for a single cellotetraose chain in water. We ran both atomistic and

CG simulations of a single cellotetraose chain and 1470 water molecules in a box with a size of $29.22 \times 38.75 \times 39.10 \text{ \AA}^3$. Figure 19 shows the probability distribution $p(L/n;n) = n \times P(L/n;n)$ of the end-to-end distance L . (Note that this is the same order parameter that was used in the work by Shen et al.⁴) The solid (blue) and dashed (red) lines plot the probability distribution between AT and BB beads from the atomistic and CG simulations, respectively. As can be seen, the location of the maximum of the probability distribution agrees fairly well, although the shape of the probability distribution is broader in the CG results.

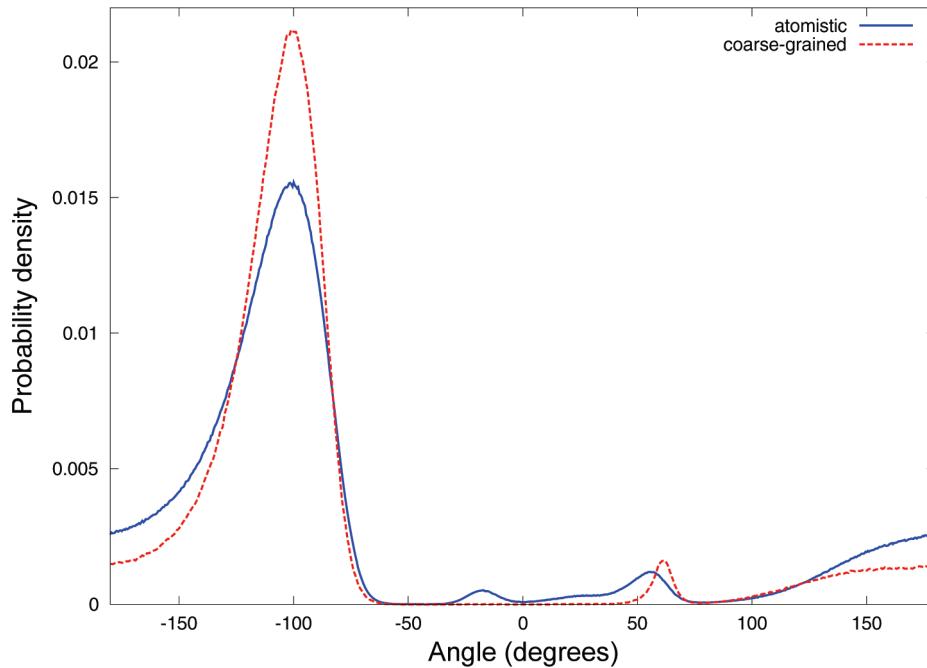


Figure 18. B–C–C–A dihedral angle distribution for cellobetaose from atomistic (solid line) and CG (dashed line) MD simulations.

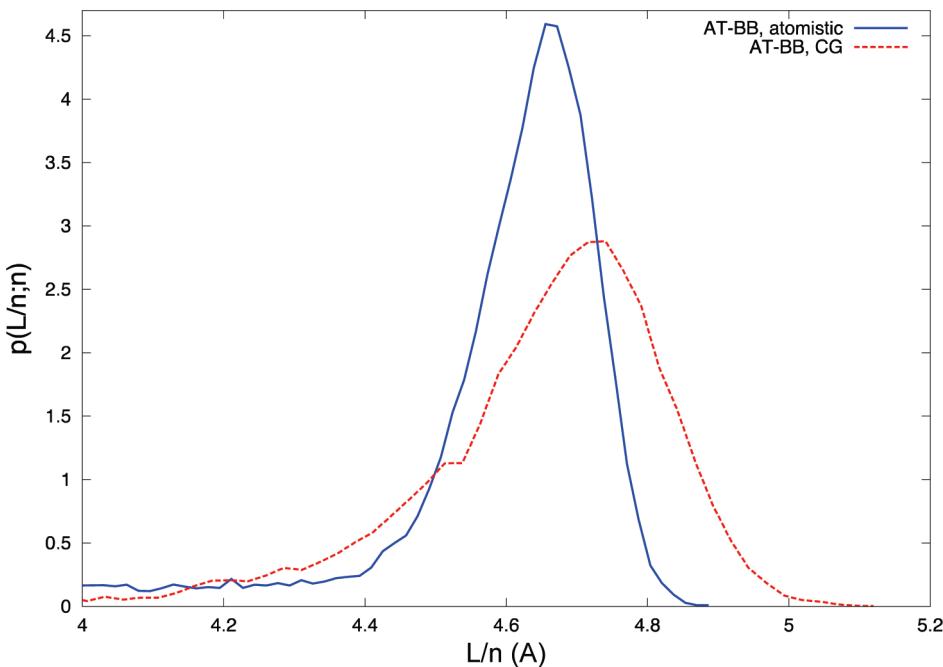


Figure 19. Probability distribution of the end-to-end chain distance for cellobetaose. The solid (blue) and dashed (red) lines show the result for the distance between AT and BB beads.

CONCLUSIONS

We have developed a CG force field for glucose, cellobiose, and cellobetaose using a combination of force matching and Boltzmann inversion. In our CG model, every glucose unit consists of three spherical beads, and water is modeled as a single bead. We note that for all three models, both the bonded and the nonbonded interaction potentials had highly nonanalytical features; i.e., fitting of the CG force field with Lennard-Jones type

and harmonic potentials would likely not be accurate. This model represents the first step to develop an accurate CG model for cellulose crystals in solution, which will make tractable multiple studies on cellulose biosynthesis, interconversion between different cellulose structures, and material properties of cellulose structure. Future research involves extending the CG model for longer cellulose fibrils by extending the cellobetaose model to cellohexaose, then applying the force field to long cellulose fibrils.

■ ASSOCIATED CONTENT

S Supporting Information. Additional radial distribution functions are given, and the bonded and nonbonded interactions force field is provided in an electronic form. This information is available free of charge via the Internet at <http://pubs.acs.org/>

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: Mark.Nimlos@nrel.gov.

■ ACKNOWLEDGMENT

This work was supported by the National Renewable Energy Laboratory Directed Research & Development program. Computational resources for this research were supported in part by the Golden Energy Computing Organization and the Colorado School of Mines using resources acquired with financial assistance from the National Science Foundation and the National Renewable Energy Laboratory. Computer time was provided in part by the NREL Computational Sciences Center supported by the DOE Office of EERE under contract number DE-AC36-08GO28308. We thank Professor Jhih-Wei Chu of UC Berkeley for helpful discussions.

■ REFERENCES

- (1) Himmel, M. E.; Ding, S. Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804.
- (2) Matthews, J. F.; Skopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.; Sugiyama, J.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2006**, *341*, 138.
- (3) Bergenstrahle, M.; Wohlert, J.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2010**, *345*, 2060.
- (4) Shen, T.; Langan, P.; French, A. D.; Johnson, G. P.; Gnanakaran, S. *J. Am. Chem. Soc.* **2009**, *131*, 14786.
- (5) Yui, T.; Hayashi, S. *Biomacromolecules* **2007**, *8*, 817.
- (6) Yui, T.; Hayashi, S. *Cellulose* **2009**, *16*, 151.
- (7) Yui, T.; Nishimura, S.; Akiba, S.; Hayashi, S. *Carbohydr. Res.* **2006**, *341*, 2521.
- (8) Yui, T.; Okayama, N.; Hayashi, S. *Cellulose* **2010**, *17*, 679.
- (9) Gross, A. S.; Chu, J.-W. *J. Phys. Chem. B* **2010**, *114*, 13333.
- (10) Bellesia, G.; Asztalos, A.; Shen, T.; Langan, P.; Redondo, A.; Gnanakaran, S. *Acta Crystallogr., Sect. D* **2010**, *66*, 1184.
- (11) Ford, Z. M.; Stevens, E. D.; Johnson, G. P.; French, A. D. *Carbohydr. Res.* **2005**, *340*, 827.
- (12) French, A. D.; Johnson, G. P. *Cellulose* **2009**, *16*, 959.
- (13) Yoneda, Y.; Mereiter, K.; Jaeger, C.; Brecker, L.; Kosma, P.; Rosenau, T.; French, A. D. *J. Am. Chem. Soc.* **2008**, *130*, 16678.
- (14) Zhong, L.; Matthews, J. F.; Crowley, M. F.; Rignall, T.; Talon, C.; Cleary, J. M.; Walker, R. C.; Chukkapalli, G.; McCabe, C.; Nimlos, M. R.; Brooks, C. L.; Himmel, M. E.; Brady, J. W. *Cellulose* **2008**, *15*, 261.
- (15) Zhong, L. H.; Matthews, J. F.; Hansen, P. I.; Crowley, M. F.; Cleary, J. M.; Walker, R. C.; Nimlos, M. R.; Brooks, C. L.; Adney, W. S.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2009**, *344*, 1984.
- (16) Bergensträhle, M.; Berglund, L. A.; Mazeau, K. *J. Phys. Chem. B* **2007**, *111*, 9138.
- (17) Beckham, G. T.; Matthews, J. F.; Peters, B.; Bomble, Y. J.; Himmel, M. E.; Crowley, M. F. *J. Phys. Chem. B* **2011**, *115*, 4118.
- (18) Payne, C. M.; Himmel, M. E.; Crowley, M. F.; Beckham, G. T. *J. Phys. Chem. Lett.* **2011**, *2*, 1546.
- (19) Lopez, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhuizen, L.; Huenenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195.
- (20) Molinero, V.; Goddard, W. *J. Phys. Chem. B* **2004**, *108*, 1414.
- (21) Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566.
- (22) Shen, T.; Gnanakaran, S. *Biophys. J.* **2009**, *96*, 3032.
- (23) Bu, L.; Beckham, G. T.; Crowley, M. F.; Chang, C. H.; Matthews, J. F.; Bomble, Y. J.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R. *J. Phys. Chem. B* **2009**, *113*, 10994.
- (24) Beckham, G. T.; Bomble, Y. J.; Bayer, E. A.; Himmel, M. E.; Crowley, M. F. *Curr. Opin. Biotech.* **2010**, *22*, 231.
- (25) Kroon-Batenburg, L. M. J.; Bouma, B.; Kroon, J. *Macromolecules* **1996**, *29*, 5695.
- (26) Liu, H. B.; Sale, K. L.; Holmes, B. M.; Simmons, B. A.; Singh, S. *J. Phys. Chem. B* **2010**, *114*, 4293.
- (27) Ramakrishnan, S.; Collier, J.; Oyetunji, R.; Stutts, B.; Burnett, R. *Bioresour. Technol.* **2010**, *101*, 4965.
- (28) Swatloski, R. P.; Spear, S. K.; Holbrey, J. D.; Rogers, R. D. *J. Am. Chem. Soc.* **2002**, *124*, 4974.
- (29) Singh, S.; Simmons, B. A.; Vogel, K. P. *Biotechnol. Bioeng.* **2009**, *104*, 68.
- (30) Hu, S.-Q.; Gao, Y.-G.; Tajima, K.; Sunagawa, N.; Zhou, Y.; Kawano, S.; Fujiwara, T.; Yoda, T.; Shimura, D.; Satoh, Y.; Munekata, M.; Tanaka, I.; Yao, M. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 17957.
- (31) Paredez, A. R.; Somerville, C. R.; Ehrhardt, D. W. *Science* **2006**, *312*, 1491.
- (32) Fontes, C. M. G. A.; Gilbert, H. J. *Annu. Rev. Biochem.* **2010**, *79*, 655.
- (33) Bayer, E. A.; Belaich, J.-P.; Shoham, Y.; Lamed, R. *Annu. Rev. Microbiol.* **2004**, *58*, 521.
- (34) Bomble, Y. J.; Beckham, G. T.; Matthews, J. F.; Nimlos, M. R.; Himmel, M. E.; Crowley, M. F. *J. Biol. Chem.* **2011**, *286*, 5614.
- (35) Murtola, T.; Bunker, A.; Vattulainen, I.; Deserno, M.; Karttunen, M. *Phys. Chem. Chem. Phys.* **2009**, *11*, 1869.
- (36) Lyubartsev, A. P.; Laaksonen, A. *Phys. Rev. E* **1995**, *52*, 3730.
- (37) Reith, D.; Putz, M.; Muller-Plathe, F. *J. Comput. Chem.* **2003**, *24*, 1624.
- (38) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2005**, *109*, 2469.
- (39) Izvekov, S.; Voth, G. A. *J. Chem. Phys.* **2005**, *123*, 134105.
- (40) Ercoleossi, F.; Adams, J. B. *Europhys. Lett.* **1994**, *26*, 583.
- (41) Ruhle, V.; Junghans, C.; Lukyanov, A.; Kremer, K.; Andrienko, D. *J. Chem. Theory Comput.* **2009**, *5*, 3211.
- (42) 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.
- (43) Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2009**, *113*, 4443.
- (44) Izvekov, S.; Voth, G. A. *J. Chem. Theory Comput.* **2006**, *2*, 637.
- (45) Shi, Q.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2006**, *110*, 15045.
- (46) Lu, L.; Izvekov, S.; Das, A.; Andersen, H. C.; Voth, G. A. *J. Chem. Theory Comput.* **2010**, *6*, 954.
- (47) Krishna, V.; Noid, W. G.; Voth, G. A. *J. Chem. Phys.* **2009**, *131*.
- (48) Das, A.; Andersen, H. C. *J. Chem. Phys.* **2010**, *132*, 164106.
- (49) Larini, L.; Lu, L.; Voth, G. A. *J. Chem. Phys.* **2010**, *132*, 164107.
- (50) Molinero, V.; Goddard, W. A. *Phys. Rev. Lett.* **2005**, *95*, 045701.
- (51) Bu, L.; Himmel, M. E.; Nimlos, M. R. In *Computational Modeling in Lignocellulosic Biofuel Production*; Nimlos, M. R., Crowley, M. F., Eds.; American Chemical Society: Washington, DC, 2010; pp 99–117.
- (52) Im, W.; Feig, M.; Brooks, C. L. I. *Biophys. J.* **2003**, *85*, 2900.
- (53) Im, W.; Lee, M. S.; Brooks, C. L. I. *J. Comput. Chem.* **2003**, *24*, 1691.
- (54) Beckham, G. T.; Matthews, J. F.; Bomble, Y. J.; Bu, L.; Adney, W. S.; Himmel, M. E.; Nimlos, M. R.; Crowley, M. F. *J. Phys. Chem. B* **2010**, *114*, 1447.
- (55) Wohlert, J.; Berglund, L. A. *J. Chem. Theory Comput.* **2011**, *7*, 753.
- (56) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074.
- (57) 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.
- (58) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.

- (59) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D. *J. Chem. Phys.* **1983**, *79*, 926.
- (60) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (61) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (62) Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2008**, *29*, 2543.
- (63) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. *J. Chem. Theory Comput.* **2009**, *5*, 2353.
- (64) Anderson, E.; Bai, Z.; Bischof, C.; Blackford, S.; Demmel, J.; Dongarra, J.; Croz, J. D.; Greenbaum, A.; Hammarling, S.; McKenney, A.; Sorensen, D. *LAPACK Users' Guide*, 3rd ed.; Society for Industrial and Applied Mathematics: Philadelphia, PA, 1999.