

A Solvent-Free Coarse Grain Model for Crystalline and Amorphous Cellulose Fibrils

Goundla Srinivas, Xiaolin Cheng, and Jeremy C. Smith*

UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, 1 Bethel Valley Road, Oak Ridge, Tennessee 37831, United States

 Supporting Information

ABSTRACT: Understanding biomass structure and dynamics on a range of time and length scales is important for the development of cellulosic biofuels. Here, to enable length and time scale extension, we develop a coarse grain (CG) model for molecular dynamics (MD) simulations of cellulose. For this purpose, we use distribution functions from fully atomistic MD simulations as target observables. A single bead per monomer level coarse graining is found to be sufficient to successfully reproduce structural features of crystalline cellulose. Without the use of constraints the CG crystalline fibril is found to remain stable over the maximum simulation length explored in this study ($>1\ \mu\text{s}$). We also extend the CG representation to model fully amorphous cellulose fibrils. This is done by using an atomistic MD simulation of fully solvated individual cellulose chains as a target for developing the corresponding fully amorphous CG force field. Fibril structures with different degrees of crystallinity are obtained using force fields derived using a parameter coupling the crystalline and amorphous potentials. The method provides an accurate and constraint-free approach to derive CG models for cellulose with a wide range of crystallinity, suitable for incorporation into large-scale models of lignocellulosic biomass.

I. INTRODUCTION

Polysaccharides are the most abundant form of biomaterial on Earth. A polysaccharide of particular importance is cellulose, which is widely present in plants as cell wall or extra-cellular material.¹ Cellulose plays a crucial role in plant growth and cell wall function.^{2,3}

Cellulose has recently gained increasing attention due to its potential application in generating biomass-based renewable energy.^{4,5} Various methods have been proposed for converting cellulose-based biomass to ethanol and other transportation fuels,⁶ typically involving three steps: (i) biomass pretreatment, (ii) conversion of extracted cellulose into sugars, typically through enzymatic hydrolysis, and (iii) fermentation of sugars to obtain ethanol. However, the natural recalcitrance of biomass to hydrolysis has led to the need to improve pretreatment and hydrolysis methods.

A characterization of cellulose structure and assembly at the molecular level is important for understanding the recalcitrance of biomass to hydrolysis. There have been substantial experimental efforts to understand cellulose structure. For example, a combination of X-ray diffraction analysis and computational modeling has revealed that native cellulose (cellulose I) occurs in two forms: I α and I β .⁷ The I α phase contains a single cellulose chain in a triclinic cell, while a two-chain monoclinic cell is found for I β . The relative ratio of the two forms depends on the origins of the cellulose. Recently, extensive synchrotron X-ray and neutron diffraction measurements on I α and I β structures have provided detailed crystal and molecular structures together with the hydrogen-bonding network within each phase.^{8,9} Cellulose also occurs in amorphous phases.⁸

Although extensive experimental studies have been undertaken, simulation studies of cellulose microfibrils have been relatively rare. This is not surprising, as the complexity and the

size of cellulose fibril structures have rendered them difficult to study with traditional molecular simulation approaches. The smallest cellulose microfibrils consist of approximately 36 individual chains with a large degree of polymerization. These large system sizes, together with the associated long-time dynamics often needed to examine cell wall processes, put simulation studies of cell wall decomposition beyond the scope of a complete atomistic representation.^{10–12} Hence, recent simulation studies have focused on using alternative methodologies based on lower resolution models, such as coarse grain (CG) molecular dynamics (MD),^{13–15} that have been proven to be useful in exploring long time and length scale processes involving proteins,^{15,16} lipids,^{17–20} polymers,²¹ and other materials.²²

In one recent study, a three-site CG model for monosaccharides was developed based on chemical information on the underlying monomers,²³ and the model was employed to study water dynamics in the glassy state formed by sugar–water complexes. Subsequently, a similar but improved three-site CG model was proposed that successfully reproduces the atomistic behavior of a single chain cellulose molecule of 14 monomer units.²⁴ In later work the model of ref 24 was applied to study interactions between cellulose (40 monomer long) and carbohydrate-binding modules.²⁵ However, in ref 25, in order to preserve the fibril structure of the cellulose assembly, cylindrical reflecting boundary conditions needed to be applied. Most recently, another three-site CG model for native cellulose was introduced based on the principles of Martini force field.²⁶ However, this force field was unable to reproduce stable crystalline cellulose fibril structures (additional interaction terms and backbone restraints were needed

Received: March 15, 2011

Published: July 01, 2011

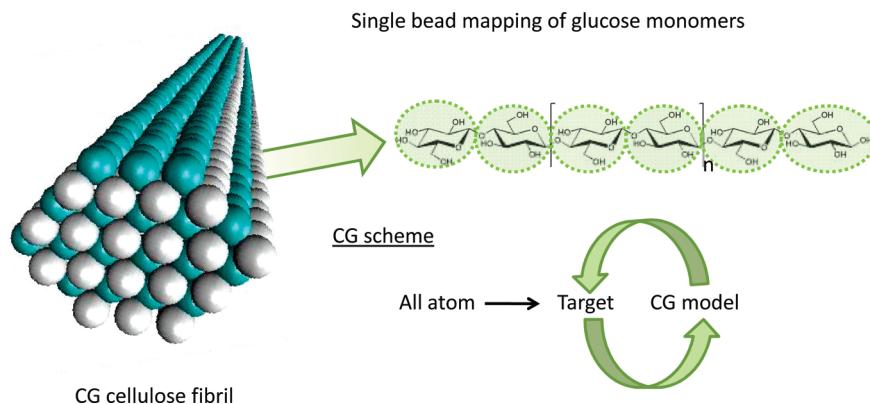


Figure 1. Single bead CG model of a cellulose chain is shown along with the underlying chemical structure. A CG cellulose fibril is also shown. Within the cellulose fibril, chains belong to origin (cyan) and center (white) sheets. A schematic representation of the CG methodology used in this work is also depicted (please see text for details).

to stabilize the crystalline fibril). Hence, to the best of our knowledge, there has hitherto been no CG force field able to model the crystalline cellulose fibril without constraints.

Here, we propose an unconstrained CG model of the cellulose fibril. The polysaccharide is modeled as a flexible chain consisting of single-site monomer units. In other words, each monosaccharide is replaced with a single CG unit, as compared to three sites in the previous studies.^{23–26} The CG model is developed using information obtained from a 20 ns long atomistic trajectory of a cellulose fibril in water.²⁷ In lignocellulosic biomass cellulose exists in varying degrees of crystallinity,^{5,6} and the crystallinity can change after pretreatment.²⁸ Hence, it is important to be able to generate models of cellulose in amorphous and fully crystalline forms as well as various states of intermediate crystallinity. The present CG model also addresses this need. We use a system of individual cellulose chains desolved in water as a model for the fully amorphous/noncrystalline form. Further, we obtain several noncrystalline forms of cellulose fibril by introducing a coupling parameter between CG force fields of fully crystalline and fully amorphous cellulose systems.

The rest of the paper is organized as follows: In the next section we describe the CG methodology and its application to crystalline cellulose. The simulation details are presented in Section III. The results and discussion for the crystalline models are presented in Section IV along with parameterization and CG results for noncrystalline fibrils. We conclude the paper with closing remarks in Section V.

II. CG METHODOLOGY

A schematic representation of the present CG method is shown in Figure 1. For development of the CG model we make use of the chemical information from all-atom simulations of a representative cellulose system, the details of which are described elsewhere.²⁷ We summarize important aspects here. A cellulose fibril was constructed with 36 chains based on the I β crystal structure⁸ with a degree of polymerization of 80 (containing 80 glucose monomers or 40 cellobiose units). The cellulose fibril was represented with the CHARMM force field²⁹ and solvated with more than 200 000 TIP3P water molecules³⁰ in a rectangular simulation box. Atomistic simulations of this system were performed for 20 ns using the NAMD simulation package.³¹ Five such independent 20 ns simulations using a 2 fs time step were performed for the analysis. The convergence of individual simulations

was monitored by calculating the total dipole moment of the cellulose fibril, which converged in less than 5 ns in all the simulations.²⁷ Since our parametrization is completely based on atomistic crystalline fibril data, the present CG potential is expected to be less transferrable but accurately represent crystalline fibril structure.

A. CG Model for Crystalline Cellulose. In the I β fibril structure the cellulose chains are categorized as belonging to origin or center planes based on the location of the chain in the crystalline unit cell. Macroscopically, these two different chains constitute alternate layers or sheets of the fibril, as represented in Figure 1. In the CG model, the origin and center chains are distinguished and assigned separate parameter sets: the monomers in the origin chains are denoted as “OR”, while those in the center chains are denoted as “CE”. The need for separate parameter sets for crystalline and noncrystalline structures is examined in detail in the sections relevant to amorphous fibrils.

Bonded Parameters. A CG cellulose chain is constructed using the center of mass of each monomer derived from the corresponding atomistic representation. Hence, each glucose monomer is represented by a single CG bead, as shown in Figure 1. Clearly then, a single cellulose chain with a degree of polymerization 80 would be represented as an 80-bead chain.

The complete CG interaction potential is given by

$$V_{\text{cg}} = V_{\text{bond}} + V_{\text{angle}} + V_{\text{torsion}} + V_{\text{nonbond}} \quad (1)$$

In eq 1, any two consecutive monomers in a chain are connected by harmonic pseudobond potentials V_{bond} and similarly, harmonic pseudobond angle potentials are employed for three consecutively connected beads. A torsional potential, V_{torsion} , is applied to any four consecutive CG beads. V_{nonbond} represents the nonbonded interactions. Since we represent each glucose monomer as a single CG site, there is no net charge associated with any of the CG units. Hence, there are no explicit electrostatic interaction terms.

In the following we describe each of the terms in eq 1. The harmonic bond interactions are given by

$$V_{\text{bond}} = \sum_{\text{bonds}} k_b (r - r_0)^2 \quad (2)$$

where k_b and r_0 represent the force constant and equilibrium bond distance for the pseudobond connecting monomers; k_b and r_0 were adjusted until the CG-bond distance distributions reproduced the corresponding distributions in the atomistic MD to the

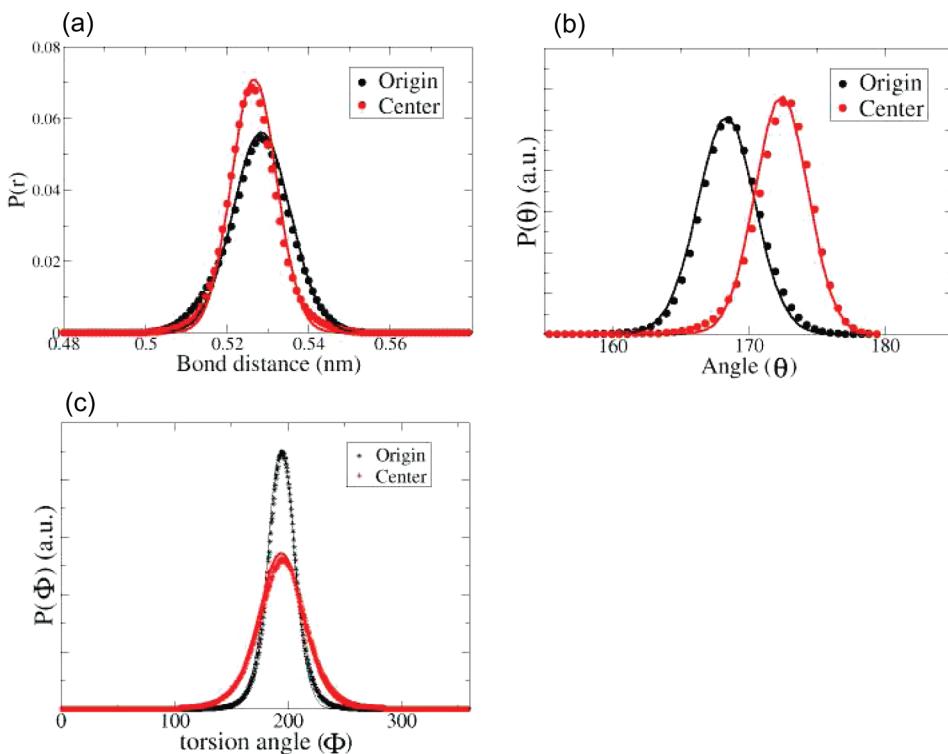


Figure 2. Comparison of distributions obtained from all-atom simulations (symbols) and CG simulations (lines): (a) Bond distance (OR–OR and CE–CE); (b) bond angle (OR–OR–OR and CE–CE–CE); and (c) torsion angle (OR–OR–OR–OR and CE–CE–CE–CE) distribution comparisons. In all the figures, comparisons for origin and center sheets are shown in black and red, respectively.

desired accuracy. The bond distance distributions obtained from the CG simulations using the final parameters for origin and center chains are shown along with the corresponding atomistic distributions in Figure 2a.

The bond angle potential is given by

$$V_\Theta = \sum_{\text{angles}} k_\Theta (\Theta - \Theta_0)^2 \quad (3)$$

where the equilibrium virtual bond angle Θ_0 and the corresponding force constant k_Θ were determined in a similar fashion, as described above. The bond angle distribution comparisons for origin and center chains are shown in Figure 2b. As can be seen from the figures, the chosen set of virtual bond and angle CG parameters reproduces underlying atomistic distributions quite well.

To represent the long-range order associated with the cellulose crystalline structure, it was found necessary to include a torsional potential. A harmonic potential for the torsion angles was found to reproduce the atomistic torsional distributions reasonably well, as shown in Figure 2c for both origin and center chains. The torsional potential is given by

$$V_\Phi = \sum_{\text{dihedrals}} k_\Phi (\Phi - \Phi_0)^2 \quad (4)$$

where Φ_0 and k_Φ represent the equilibrium torsion angle and the corresponding force constant for four consecutively connected monomers. CG parameters for the bond, angle, and torsional potentials are listed in Table 1.

Nonbonded Parameters. While the development of bonded CG parameters is relatively straightforward, treating nonbonded interactions is more complex. There exist different approaches in developing the nonbonded CG parameters for any given system,^{14,17,19} and the specific need of the intended application of the developed CG force field also influences the approach

Table 1. CG-Bonded Parameters for Crystalline and Amorphous Cellulose Fibrils

bonds	k_b (kJ mol ⁻¹ nm ⁻²)	r_0 (nm)
OR–OR	4.588×10^4	0.5283
CE–CE	7.700×10^4	0.5250
AM–AM	2.420×10^4	0.5228
angles	k_Θ (kJ mol ⁻¹ rad ⁻²)	Θ_0 (°)
OR–OR–OR	840	168.7
CE–CE–CE	1080	173.2
AM–AM–AM	266.8	163.5
torsions	k_Φ (kJ mol ⁻¹ rad ⁻²)	Φ_0 (°)
OR–OR–OR–OR	23	191.6
CE–CE–CE–CE	8	187.2
AM–AM–AM–AM	5.6	224.0

chosen. If the model needs to be developed for a particular application, one may choose to parametrize the system based solely on the information obtained from corresponding, specific

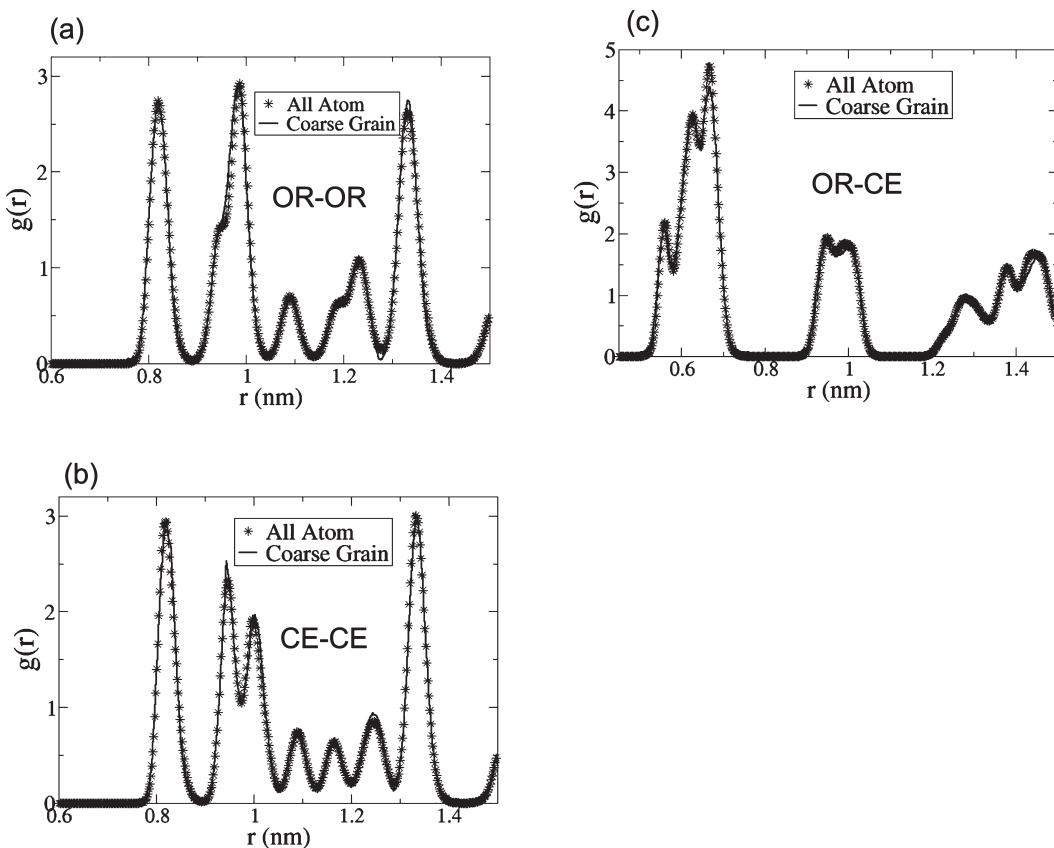


Figure 3. Comparison of all-atom radial distribution functions with those of the CG simulations: (a) $g(r)$ between the CG monomers in origin sheets; (b) $g(r)$ for the CG monomers in center sheets; and (c) $g(r)$ obtained for the cross-interactions between origin and center sheets. In all the figures, CG distributions (lines) agree with the underlying atomistic distributions (symbols) remarkably well.

higher-resolution simulations. In one such example, a CG force field was developed for biological lipid molecules based on Boltzmann inversion criteria by matching the distributions at the potential level that captures specific biological properties including self-assembly into various morphologies.¹⁷ In another approach, systematic CG models for a variety of biological systems have been developed by introducing multiscale force-matching CG methods.¹⁴ A further method captures correlated motion by matching covariance Hessian matrices.^{15,16} These approaches widen the use of CG methods. Alternatively, CG parameters can be developed based on previously available generalized parameters for the individual components of the system. In an example of this, the Martini CG force field contains a database of CG parameters for the naturally occurring amino acids based solely on their active coefficients in water and oil.³² These, more general CG parameters are more transferable for simulating systems not considered during the force field development process but may fall short of accuracy if the system-specific details have not been incorporated into the parametrization. For example, a recent CG model of cellulose was unable to reproduce a crystalline fibril structure based on MARTINI parameters.²⁶ The present one-site model is based on Boltzmann inversion criteria, an approach that Klein et al. have used extensively for developing CG force fields for copolymers,^{21,22} and was also adopted for a previous three-site model of cellulose.¹³

Distance distributions between the nonbonded units of interest from corresponding atomistic simulations were used as target observables. The goal is to reproduce the atomistic distributions

by carrying out CG simulations, adjusting the underlying potentials. To this end, the radial distribution function $g(r)$ is the target observable from the atomistic simulations. Since there are two different CG units, OR and CE, three separate nonbonded potentials must be determined (OR–OR, OR–CE, and CE–CE); $g(r)$ for these interactions from the atomistic simulations are shown in Figure 3.

Here, we obtain nonbonded CG potentials iteratively using the target observable (the atomistic $g(r)$) as follows

$$V_{\text{new}} = V_{\text{old}} + k_B T \ln(g_{\text{cg}}(r)/g_{\text{aa}}(r)) \quad (5)$$

where the updated potential V_{new} between corresponding CG units in each step is obtained by modifying the potential from the previous simulation, V_{old} , k_B is the Boltzmann constant, T the absolute temperature, and the distributions from the atomistic and CG simulations are $g_{\text{aa}}(r)$ and $g_{\text{cg}}(r)$, respectively. We began with a simple Lennard-Jones 12–6 potential as the starting CG potential and then modified it according to eq 5. In each iteration the CG simulation is carried out using the updated potential, the distribution functions obtained from the updated trajectory are inserted in eq 5, and the potentials updated. Iteration continues until a satisfactory comparison with the target observables is obtained.

The $g(r)$ for crystalline cellulose is complex with multiple discrete distinct peaks and thus contrasts with the simple, relatively smoother functions of typical bulk/liquid systems. Consequently, it is impractical to employ simple Lennard-Jones-type

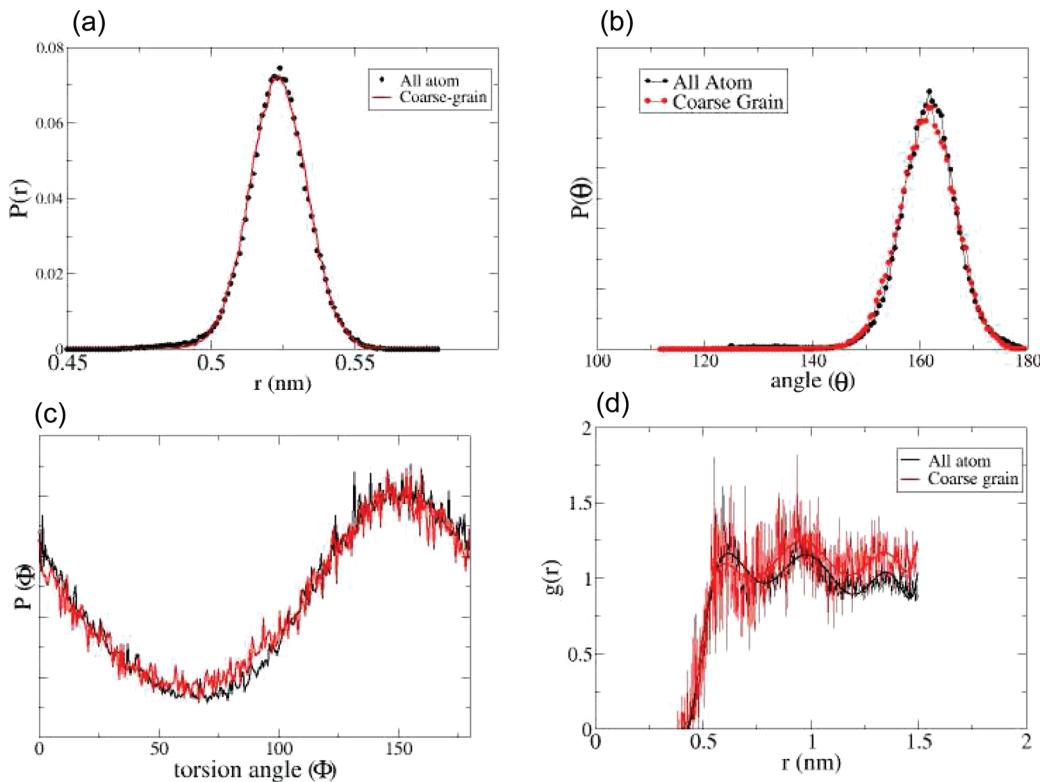


Figure 4. Comparison between atomistic (black) and CG (red) distribution functions of a single cellulose in water: (a) Bond distance, (b) bond angle, and (c) torsional angle, and (d) radial distribution function.

potentials in this case. Using such simplified potentials may also cause instabilities in cellulose crystalline structure, as was observed in previous CG simulation studies of the cellulose fibril.²⁵

Comparison (after five or six iterations) of the target and CG observables for the origin chains indeed showed that this method is unable to capture the fine structure of $g_{aa}(r)$ arising from the crystalline nature of cellulose. Such problems are common when dealing with crystalline materials, as, for example, in the recent study of crystalline fatty acids.³³ As noted in ref 33, eq 5 needs to be modified by introducing a damping factor (δ_{ij}) in order to suppress abrupt changes in the potentials:

$$V_{new} = V_{old} + \delta k_B T \ln[g_{cg}(r)/g_{aa}(r)] \quad (6)$$

where δ is assigned a positive value between 0 and 1. When δ is 1, we recover eq 5 corresponding to a noncrystalline system.

The $g(r)$ obtained including the damping factor was compared with the atomistic $g(r)$ functions, and the iterative procedure repeated with eq 6 until the target and CG distributions were in satisfactory agreement. Final comparison is shown in Figure 3. The CG distribution functions reproduce their atomistic counterparts remarkably well. Importantly, the positions of all the peaks corresponding to the crystal structure are reproduced faithfully by the present CG simulations. Excellent agreement between the atomistic and CG structures demonstrates the ability of the single-bead model to capture the structural features at the monomer level. Hence, this comparison validates the present CG force field for the I β crystalline cellulose fibril.

The potentials obtained using eq 6 do not correspond to any simple analytical form, and hence all three nonbonded potentials were tabulated. The Supporting Information gives more details together with the tabulated potentials.

B. CG Model for Noncrystalline Cellulose. The CG force field for the crystalline cellulose fibril was used as a basis from which to develop CG parameters for noncrystalline cellulose. The development procedure used is similar in spirit to that of the crystalline fibril. However, while there is a unique experimental structure corresponding to any given crystalline phase of cellulose, for amorphous cellulose, this is no longer the case. Hence, in order to develop CG parameters for amorphous cellulose, the approach adopted was modified. To this end, two simulation systems representing two extreme scenarios were considered: one for fully crystalline cellulose and the other for fully amorphous. For the fully amorphous systems an atomistic model was constructed consisting of 9 fully hydrated single cellulose chains in TIP3P water, each chain consisting of 10 glucose monomers. An MD simulation of this system was carried out for 20 ns, as was the case for the crystalline system.

Bonded Parameters. A CG parameter set for the amorphous system was mapped from the atomistic system in a similar manner to the crystalline case. Again, the bond distance and the bond and torsion angle potentials were obtained by iterative CG simulations. The obtained parameter set for the single chains differs significantly from that for crystalline cellulose. One major difference is clearly that there is no distinction of origin or center chains, all chains being indistinguishable. Hence, there is a single distribution for each of the bond distances and angles and the torsion angles.

The bond distance and angle distributions are represented with harmonic potentials as before. However, the torsion angle is more flexible than in the crystalline phase and is thus represented using a periodic potential of the following form, corresponding to a periodic potential with single multiplicity.

$$V_{am}(\Phi) = \sum_{dihedrals} k_{am}(1 + \cos(\Phi - \Phi_0)) \quad (7)$$

The comparisons of the bond distance and angle and the torsional angle distributions obtained from the atomistic and CG simulations are shown in Figure 4a–c, respectively. The final bonded parameter values obtained are listed in Table 1.

Nonbonded Parameters. The complexity of developing nonbonded interaction parameters is greatly reduced in the amorphous case compared to the crystalline fibril due to the presence of a single type of CG bead (named AM). Hence, nonbonded interaction potentials needed to be developed only for AM–AM interactions. Accordingly, the radial distribution function for AM obtained from the atomistic simulations is shown in Figure 4d. As can be seen, the $g_{am}(r)$ in this case does not show discontinuities, in contrast to the crystalline system. The CG distributions were obtained starting with the following Lennard-Jones potential:

$$V_{am}(r_{ij}) = 27/4(\varepsilon_{ij})((\sigma/r_{ij})^9 - (\sigma/r_{ij})^6) \quad (8)$$

where ε_{ij} and σ_{ij} represent the potential well depth and the contact distance between i and j . The $g_{ij}(r)$ so obtained was used with eq 6 to refine the potential, which in turn was used to carry out new CG simulations. By following the iterative procedure described in Section II, final ε and σ values were determined. The final parameters for the fully amorphous system are listed in Table 1. Note that the scaling factor δ (eq 6) was not needed in this case.

III. SIMULATION DETAILS

The CG simulations were performed using stochastic dynamics in order to incorporate solvent effects in an implicit fashion. The dynamics of the individual CG particles was governed by a Langevin equation:

$$m_i(dv_i/dt) = F_i(r_i) - m_i\xi_i v_i + \mathbf{R}_i \quad (9)$$

where the CG particle's mass, velocity, and position are represented by m_i , v_i , and r_i , respectively. F_i is the systematic force on particle i calculated using the interaction potentials described in the previous sections. ξ_i is the friction coefficient, and \mathbf{R}_i is the random force. While longer time step (10–20 fs) was possible for noncrystalline CG fibrils, we were unable to use a time step larger than 1 fs for fully crystalline fibrils because of the discrete nature of the fitted potential. Nevertheless, for the sake of consistency, we have decided to use 1 fs time step for both the crystalline and the noncrystalline fibril simulations. As a result we have chosen a GROMACS default value of 0.2 ps⁻¹ for the damping coefficient in all the simulations.

IV. RESULTS AND DISCUSSION

A. Crystalline Cellulose Fibril. Previous CG simulation studies of cellulose fibrils have encountered problems with the stability of the fibril. Hence, as a first test of the present CG model, we examine this aspect. To this end, two different CG simulations were carried out, with different initial conditions. Both of these simulations started with a CG cellulose fibril in the crystalline I β form. In general, CG simulations would allow larger time steps

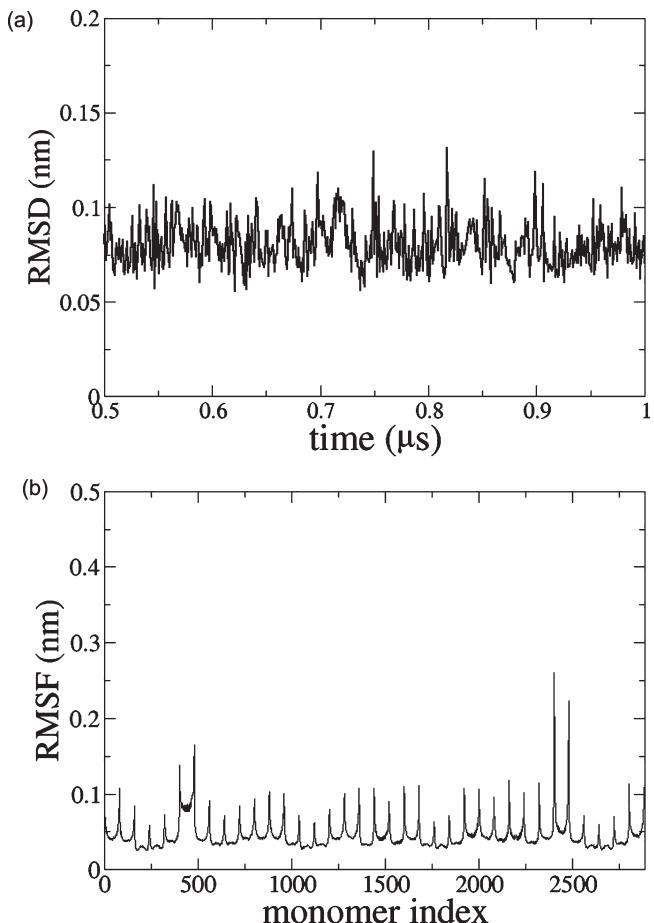


Figure 5. Stability of the CG crystalline cellulose fibril. (a) RMSD is plotted as a function of simulation time. (b) RMSF for each CG monomer within crystalline cellulose fibril.

compared to atomistic simulations. However, in the present study, crystalline nature of the fibril required smaller time steps, as the potentials are relatively rough compared to traditional empirical potentials. After several test runs with different time steps, a time step of 1 fs was chosen, and the production runs were carried out for at least 1 μ s after the equilibration. The results presented in the following were obtained by averaging over the two separate trajectories.

The stability of the crystalline structure can be assessed by calculating the fluctuations within the fibril (Figure 5a). The root-mean-square displacement (RMSD) is <0.11 nm over the entire simulation time (>1 μ s). This result is in good agreement with a recent atomistic simulation study of crystalline cellulose fibril.¹² The RMSD for the individual (origin and center) sheets shows similar behavior. Considering the fact that each CG site represents a glucose monomer of diameter \sim 0.52 nm, these fluctuations are relatively small. The stability of the cellulose crystalline structure is confirmed by the root-mean-square fluctuations (RMSF) of individual monomers, shown in Figure 5b. The monomers at the chain ends exhibit relatively larger fluctuations compared to monomers in the middle part of the chain. In other words, the chain ends are relatively flexible in the fibril structure and might disrupt the crystal were the cellulose chains to be too short. Using longer cellulose chains avoids such unwanted effects. Further, the largest fluctuations were observed for the monomers 401–480 and 2401–2480, which belong to corner chains in the

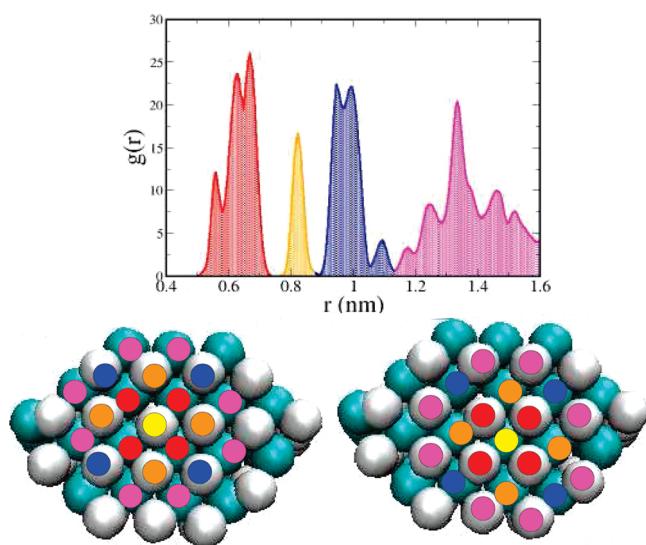


Figure 6. Radial distribution function for a crystalline cellulose fibril. Peaks are color coded. The cellulose fibril is shown in two schematic representations. The fibril structure on the right reveals the nearest neighbors around a selected origin chain monomer, while that on the left shows the same for a center chain monomer.

fibril. This suggests that the corner chains are relatively flexible compared to other chains in the fibril.

In order to examine the structural details we have calculated the pair distribution functions. In Figure 6 the radial distribution function [$g(r)$] for crystalline fibril is plotted. This system shows rich structural features even beyond 2 nm. Individual peaks were analyzed in detail and are schematically shown in the figure, color coded so as to indicate the responsible CG sites within the fibril. The broader set of peaks around 0.64 nm arises from first neighbor interactions between opposite-type monomers (origin and center or center and origin). The first and second nearest neighbors of the same kind (origin or center monomers) produce two peaks, at 0.81 and 0.98 nm, respectively. Second nearest neighbors of different kinds give rise to the broader peak at 1.38 nm. These structural details are important when comparing crystalline and amorphous cellulose structures, as discussed in detail in the later sections pertaining to amorphous structures.

B. Noncrystalline Cellulose Fibril. As a first check, we applied the fully amorphous CG parameter set to a crystalline fibril. The fibril lost its structure and rapidly collapsed to form a near spherical aggregate (<100 ns). The RMSD for simulations using crystalline and fully amorphous potentials sets is shown in Figure 7. The fully amorphous structure shows very high RMSD values (12 nm) compared to that of crystalline fibril (0.11 nm).

C. Generating Noncrystalline Cellulose Fibril Structures. In this section we describe the study of noncrystalline fibrils, generated by using a combination of fully crystalline and fully amorphous potentials with a coupling parameter as follows

$$V_{\text{cg}}(r_{ij}) = \lambda V_{\text{cr}}(r_{ij}) + (1 - \lambda)V_{\text{am}}(r_{ij}) \quad (10)$$

When $\lambda = 0$ the above potential represents a fully amorphous state, whereas for $\lambda = 1$ we recover crystalline behavior. The above combination rule is applied for both bonded and nonbonded interactions separately.

Fifteen different simulation systems were constructed with a nonlinear spacing of λ values between 0 and 1. Final snapshots

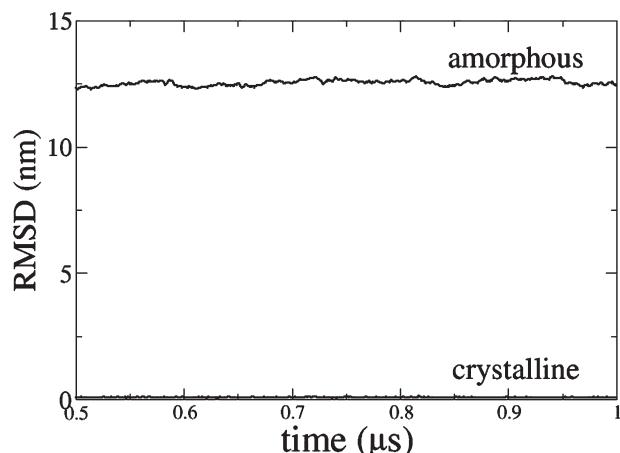


Figure 7. RMSD for the fully amorphous system for the 1 μ s CG simulation. For comparison, RMSD for the fully crystalline (CG) cellulose is also shown.

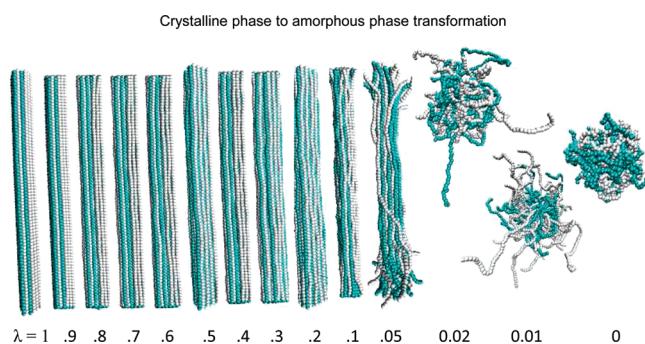


Figure 8. Transition of cellulose fibril from crystalline structure to amorphous (noncrystalline) structure with decreasing coupling parameter (λ) values. Final snapshots obtained from simulations with different λ values are shown; $\lambda = 1$ and 0 correspond to crystalline and fully amorphous phases, respectively.

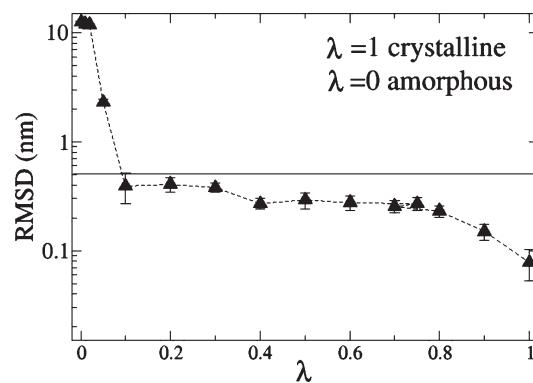


Figure 9. RMSD values for systems with different λ values.

obtained for the selected λ values after 1 μ s simulation are shown in Figure 8. The cellulose deviates from the fully crystalline structure for all values of $\lambda < 1$. However, the transition to a fully amorphous structure accelerates when decreasing λ below 0.1. This is confirmed in Figure 9 in which the average RMSD values for the systems with different λ values are plotted. The RMSD in the fully crystalline phase is near zero. Decreasing λ values increase the RMSD, indicating the emergence of noncrystalline

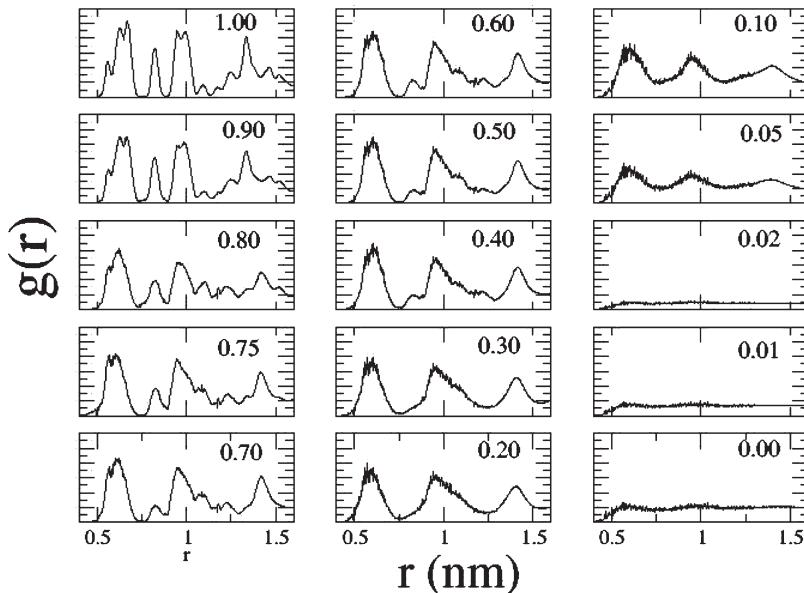


Figure 10. Gradual progression of crystalline to amorphous transition as evident by radial distribution functions. Each figure corresponds to a simulation carried out with the specified λ value.

behavior. However, the RMSD shows a significant crossover between $\lambda = 0.1$ and 0.02 , suggesting a major disruption in the crystalline structure. A similar trend was also observed in RMSF (Supporting Information). For $\lambda < 0.1$, single chain behavior becomes dominant, and the RMSD fluctuations are larger than the size of the CG monomers (>0.52 nm). In fully amorphous structures ($\lambda = 0$) the RMSD is found to be as large as 12 nm. Together, these results provide guidelines for the interaction range over which a possible transition from crystalline to amorphous structure occurs. In the present formalism, fibrils with $\lambda < 0.1$ represent fully noncrystalline structures.

Analyzing fibril structures as a function of λ reveals interesting features. In Figure 10 the pair distribution function $g(r)$ for all the systems with different λ values is plotted. The crystalline structure ($\lambda = 1.0$) gradually disappears with decreasing λ , as evidenced by a gradual disappearance of corresponding peaks in $g(r)$. With decreasing λ , the first peak to show changes is the peak at 0.81 nm, which starts to diminish when λ is decreased by as little as 10%. The peak height then gradually further decreases and ultimately merges with the neighboring peak at 0.98 nm, thereby forming a single broader peak. As explained above, the peak at 0.81 nm corresponds to first nearest neighbors of the same kind, and the peak at 0.98 nm corresponds to second nearest neighbors of different kinds. Combined with the peak position analysis, the merging of these two peaks illustrates how the distinction between the second neighbors of the same or different kind of beads becomes negligible. The structure defining origin and central chains is gradually lost at distances greater than 0.8 nm. Similarly, the peaks from distinct third neighbors merge to a single broader peak around 1.33 nm. Thus the simulations trace the disappearance of long-range structure with decreasing λ . Nevertheless, the presence of peak around 0.63 nm indicates that a short-range structure within the fibril is still present. Further decreasing λ results in the complete loss of crystalline order, as evident from the $g(r)$ at $\lambda < 0.1$. For the λ values 0.01–0.0, the $g(r)$ resembles that of typical dense liquid systems, with the peaks corresponding to neighboring shells of the monomers.

V. CONCLUSIONS

In this paper we have presented efficient and reliable CG models for crystalline and amorphous cellulose fibrils. Each monomer in the cellulose chain is mapped on to a single CG bead. In order to model the corresponding experimental cellulose I β structure, the cellulose fibril is constructed with distinct chains of origin and center sheets. The target observables were obtained from extensive atomistic simulations, and the optimized CG model obtained is in good agreement with the corresponding atomistic distributions. The robustness of the present CG force field for crystalline cellulose is demonstrated by the stability of cellulose fibrils for several hundreds of nanoseconds without using constraints. To our knowledge, this is the first such unconstrained CG model that can be used to study cellulose fibrils.

Over the course of the CG simulation ($>1 \mu\text{s}$), mean bead fluctuations are ~ 0.1 –0.3 nm, indicating that the fluctuations in the crystalline structure are much smaller than the size of the CG glucose monomer itself (0.52 nm). In natural samples, amorphous and crystalline cellulose coexist.³⁵ Hence, in order to enable a more versatile model of cellulose structure, we extended the crystalline CG model to amorphous cellulose. An individual cellulose chain completely solvated (in water) was considered as an “extreme” case of noncrystalline fully amorphous cellulose, and a corresponding CG parameter set was developed. The use of a coupling parameter λ allows a combinations of potentials for fully crystalline ($\lambda = 1$) and fully amorphous ($\lambda = 0$) systems, thereby producing a series of structures intermediate between crystalline and amorphous states. This λ coupling approach allows the CG study of various partially crystalline and noncrystalline cellulose structures. Decreasing λ by as little as 10% introduces significant noncrystallinity into the cellulose structures, and the simulations exhibit a clear crossover from partially crystalline to fully amorphous behavior around $\lambda < 0.1$.

Previous experimental and simulation studies have shown distinct behavior between the center and the origin planes due to their slightly different hydrogen-bonding patterns.^{12,36} The present results provide noncrystalline models in which the specific

hydrogen bonding is disrupted, leading to the breakdown of short-range order. Analysis of the radial distribution functions shows how as a function of λ the distinction between the center and the origin planes evolves. For $\lambda < 0.3$, the distinction between the two types of planes exists up to the first nearest neighbors, while no distinction was found beyond 0.8 nm, a distance corresponding to second and further neighbors.

In conclusion, the present CG model allows the exploration of both crystalline and amorphous cellulose fibril structures for length- and time-scales beyond the reach of atomistic simulations. A systematic method is presented for generating and representing both crystalline and amorphous cellulose states. Natural cellulose fibrils consist predominately of a mixture of the crystalline phases I α and I β . After biomass treatment, other crystalline forms, such as cellulose II, III, can be formed. Although the present study has focused on the I β cellulose, similar methodology can be easily extended to other crystalline forms. Further, the introduction of a coupling parameter λ provides a unified way of generating cellulose fibril structures with different degrees of crystallinity, thereby enabling the modeling of extensive cellulose fibrils with both crystalline and amorphous characteristics. The mixed CG models could also allow the study of the crystalline-to-amorphous transition of celluloses using an adaptive resolution scheme.³⁷ Future studies will include the importance of explicit solvent incorporation along with the interactions between crystalline and amorphous cellulose structures. Furthermore, increasing the complexity of the CG models to incorporate other biomolecules, such as lignin and hemicellulose, will be important in understanding biomass recalcitrance, a central theme in biomass-based renewable energy. Work in these directions is presently underway.

■ ASSOCIATED CONTENT

S Supporting Information. Details on tabulated potentials used in this work along with supporting results that were not presented in manuscript are presented in the Supporting Information. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: smithjc@ornl.gov.

■ ACKNOWLEDGMENT

The authors would like to thank L. Petridis for help and valuable scientific discussions. G.S. thanks R. Schulz for the help with GROMACS software and B. Lindner for helping with structural characterizations and for providing the trajectories of the atomistic crystalline cellulose fibril. This research is sponsored by DOE's Scientific Discovery through Advanced Computing (SciDAC) program through DOE's Office of Advanced Scientific Computing Research (ASCR) and Biological and Environmental Research (BER) under FWP ERKJE84 and performed at Oak Ridge National Laboratory (ORNL). ORNL is managed by UT-Battelle, LLC, for the U.S. Department of Energy under contract DE-AC05-00OR22725.

■ REFERENCES

- (1) Habibi, Y.; Lucia, L. A.; Rojas, O. J. *Chem. Rev.* **2010**, *110*, 3479–3500.
- (2) Metthew, 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*, 138–152.
- (3) Ding, S. Y.; Himmel, M. E. *Agri. Food. Chem.* **2006**, *54*, 597–606.
- (4) Demain, A. L.; Newcomb, M.; Wu, J. H. D. *Microbiol. Mol. Biol. Rev.* **2005**, *69*, 124.
- (5) Lynd, L. R.; Cushman, J. H.; Nicholos, R. J.; Wyman, C. E. *Science* **1991**, *251*, 1318.
- (6) Himmel, M. E.; Ding, S.-Y.; Johnson, D. K.; Adney, W. S.; Nimlos, M. R.; Brady, J. W.; Foust, T. D. *Science* **2007**, *315*, 804–807.
- (7) Vasella, A.; Davics, G. J.; Bohm, M. *Curr. Opin. Chem. Bio.* **2008**, *12*, 539–629.
- (8) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.
- (9) Nishiyaman, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300–14306.
- (10) 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., III; Himmel, M. E.; Brady, J. W. *Cellulose* **2008**, *15*, 261–273.
- (11) Zhong, L.; Matthews, J. F.; Hansen, P. I.; Crowley, M. F.; Cleary, J. M.; Walker, R. C.; Nimlos, M. R.; Brooks, C. L., III; Adney, W. S.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2009**, *344*, 1984–1992.
- (12) Gorss, A. S.; Chu, J.-W. *J. Phys. Chem. B* **2010**, *114*, 13333–13341.
- (13) Nielsen, S.; Lopez, C.; Srinivas, G.; Klein, M. L. *J. Phys.: Condens. Matter* **2004**, *16*, R481.
- (14) 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.
- (15) Moritsugu, K.; Smith, J. C. *Biophys. J.* **2008**, *95*, 1639–1648. *Biophys. J.* **2007**, *93*, 3460–3469.
- (16) Voltz, K.; Trylska, J.; Tozzini, V.; Kurkal-Siebert, V.; Langowski, J.; Smith, J. C. *J. Comput. Chem.* **2008**, *29*, 1429–1439.
- (17) Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 4464. Shelley, J. C.; Shelley, M. Y.; Reeder, R. C.; Bandyopadhyay, S.; Moore, P. B.; Klein, M. L. *J. Phys. Chem. B* **2001**, *105*, 9785.
- (18) 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.
- (19) Marrink, S. J.; de Vries, A. H.; Mark, A. E. *J. Phys. Chem. B* **2004**, *108*, 750–760.
- (20) Queyroy, S.; Neyertz, S.; Brown, D.; Muller-Plathe, F. *Macromolecules* **2004**, *37*, 7338–7350.
- (21) Srinivas, G.; Discher, D. E.; Klein, M. L. *Nat. Mat.* **2004**, *3*, 638–644.
- (22) Srinivas, G.; Pitera, J. W. *Nanolett* **2008**, *8*, 611–618.
- (23) Molinero, V.; Goddard, W. A., III *J. Phys. Chem. B* **2004**, *108*, 1414–1427.
- (24) Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566–11575.
- (25) Bu, L.; Beckham, G. T.; Crowley, C. 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*, 10994–11002.
- (26) Wohlert, J.; Berglund, L. A. *J. Chem. Theory Comput.* **2011**, *7*, 753–760.
- (27) Schultz, R.; Lindner, B.; Petridis, L.; Smith, J. C. *J. Chem. Theory Comput.* **2009**, *5*, 2798–2808.
- (28) Pingali, S. V.; Urban, V. S.; Heller, W. T.; McGaughey, J.; O'Neill, H.; Foston, M.; Myles, D. A.; Ragauskas, A.; Evans, B. R. *Biomacromolecules* **2010**, *11*, 2329–2335.
- (29) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **2004**, *4*, 187–217.
- (30) Jorgensen, W. L.; Chandrashekar, J.; Madhura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (31) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (32) Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P.; deVries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.

- (33) Hadley, K. R.; McCabe, C. *J. Chem. Phys.* **2010**, *132*, 134505.
- (34) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (35) Dumitriu, S. *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed.; Marcel Dekker: New York, 2005.
- (36) Yui, T.; Nishimura, S.; Akiba, S.; Hayashi, S. *Carbohydr. Res.* **2006**, *341*, 2521–2530.
- (37) Praprotnik, M; Site, L. D.; Kremer, K. *Annu. Rev. Phys. Chem.* **2008**, *59*, 545–571.