JCTC Journal of Chemical Theory and Computation

ARTICLE

pubs.acs.org/JCTC

A Coarse-Grained Model for Molecular Dynamics Simulations of Native Cellulose

Jakob Wohlert* and Lars A. Berglund

Wallenberg Wood Science Center, Royal Institute of Technology, SE-10044 Stockholm, Sweden

Supporting Information

ABSTRACT: We have constructed a coarse-grained model of crystalline cellulose to be used in molecular dynamics simulations. Using cellobiose from the recently published MARTINI coarse-grained force field for carbohydrates [Lopez, C. A. et al. *J. Chem. Theory Comput.* **2009**, *5*, 3195−3210] as a starting point, we have reparameterized the nonbonded interactions to reproduce the partitioning free energies between water and cyclohexane for a series of cellooligomers, cellobiose through cellopentaose. By extrapolating the model to longer cellooligomers, and by assigning special cellulose−cellulose nonbonded interactions, we obtain a model which gives a stable, ordered structure in water that closely resembles the crystal structure of cellulose Iβ. Furthermore, the resulting model is compatible with an existing coarse-grained force field for proteins. This is demonstrated by a simulation of the motion of the carbohydrate-binding domain of the fungal cellulase Cel7A from *Trichoderma reesei* on a crystalline cellulose surface. The diffusion coefficient at room temperature is calculated at $D_l = 3.1 \times 10^{-11}$ cm^2 s^{-1}, which is in good agreement with experimental numbers.

INTRODUCTION

Cellulose, the 1−4 linked polymer of β-D-glucopyranose (see Figure 1), is synthesized in nature in slender fibrils, whose lateral dimensions typically range from 3 nm up to 5 nm, depending on the synthesizing species. These fibrils are usually assembled into larger structures, called fibril aggregates. Within the fibrils, the glucan chains are packed in a specific crystal form, named cellulose I, or native cellulose, which is known from X-ray crystallography and neutron diffraction.[1,2] Furthermore, there are two distinct allomorphs of native cellulose,[3] namely, cellulose Iα and cellulose Iβ, with the former being dominant in bacterial and algal cellulose and the latter being dominant in cellulose produced in higher plants.

Cellulose, being the most abundant, naturally occurring polymer on earth, has in recent years been the target of great, renewed interest. On one hand, cellulose-based biocomposites offer the potential to replace petroleum-based plastics and composites.[4,5] The use of cellulose nanofibrils extracted from plant cell walls is of particular interest since the potential for a significantly extended property range is vast for nanostructured cellulose materials.[6] On the other hand, cellulose offers potential as an important feedstock for large-scale production of bioethanol by enzymatic and chemical degradation. Great efforts are directed toward a better understanding of the enzymatic pathways, and the engineering of both enzymes and cellulosic material, with the goal of achieving better turnover rates.[7,8]

Molecular dynamics computer simulation has been shown to be an excellent tool to contribute to a molecular-level understanding of crystalline cellulose, its structure and its dynamics, as well as its interactions with solvents and other biomolecules.[9−14] At present, there exist several force fields that are being used in simulations of cellulose, such as GROMOS 45a4,[15] CHARMM36,[16] PCFF,[17] and GLYCAM06,[18] to mention a few. These are all atomistic force fields, meaning that all atoms, except for nonpolar hydrogen atoms in the case of GROMOS, are represented explicitly by one interaction site, making the simulations limited to fairly small systems and short simulation times.

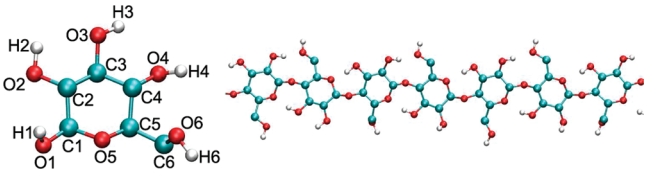

Figure 1. β-D-Glucose with the atom names used in this paper (left) and a cellulose chain (right).

Coarse-grained (CG) modeling is an alternative to all-atom (AA) modeling, wherein some of the atomic detail is averaged out, so that much longer time and length scales come within reach. During recent years, coarse-grained approaches have been highly successful for a wide range of biomolecular systems such as proteins, lipid bilayers, and carbohydrates.[19]

There are a few different approaches to construct CG models. Typically, a number of atoms are grouped together to form a CG superatom, or bead. The beads interact through effective pairwise potentials, which are constructed to reproduce relevant properties from the atomistic representation. This can be achieved in different ways, e.g., using Monte Carlo schemes to optimize potential parameters,[20] or through force-matching procedures.[21] These procedures generally lead to models that are highly specialized, meaning that they are usually very good at describing the system for which they are originally developed, but not easily transferable to other systems.

The MARTINI force field[22] was parametrized using a different approach. Here, a number of predefined CG beads interact via a

Received: August 27, 2010
Published: January 26, 2011

fixed set of pairwise potentials. They are used to construct small molecular building blocks, which are calibrated against thermodynamic data, in particular partitioning coefficients between polar and apolar phases. This leads to a model that can be extended to include new molecular species, in a fairly straightforward way, while at the same time retaining its internal consistency and compatibility.

There exist a couple of CG models for carbohydrates in the literature. The M3B model by Molinero and Goddard[20] was developed to simulate malto-oligosaccharides in solution, using Boltzmann inversion to construct potentials for the bonded interaction and a Monte Carlo scheme to optimize the nonbonded interactions which are described with Morse potentials. The model reproduces thermodynamic data for glucose very well and gives a helical structure for amylose which is stable for tens of nanoseconds. Liu et al.[21] have constructed a CG model for glucose in solution using a force-fitting procedure. Their model shows good structural properties, such as radial distribution functions, as well as some thermodynamic properties, like isothermal compressibility. Recently, the MARTINI force field was extended to include carbohydrates as well.[23] The extended model provides parameters for the monosaccharides glucose and fructose, a number of disaccharides, including cellobiose, and the oligosaccharides maltoheptose (a short strand of amylose) and laminaraheptabiose (Curdlan).

Bu et al.[24] used the M3B model as a starting point to construct a model which, to the best of our knowledge, is the only CG model for crystalline cellulose in the literature. They refitted the bonded interaction terms and scaled the nonbonded terms in order to get a crystal structure that closely matches that of cellulose Iβ. The model was used to study the interactions between a crystalline cellulose surface and the carbohydrate-binding domain (CBD) of a fungal cellulase, which was represented in atomic detail, together with an implicit water model. They show that it is indeed possible to get a stable cellulose crystal using only three interaction sites per glucose residue. However, since different force fields were mixed in their study, the results they obtain for cellulose−protein interactions are somewhat questionable. For instance, they report a diffusion coefficient for the CBD which is around 3 orders of magnitude too large compared to experimental values.[25]

To really benefit from a CG approach, it is desirable to have a model that is internally consistent. Furthermore, treating the system as whole at a CG level will also benefit from that simulation time, and length scales can be significantly extended. To that end, we have developed a model for crystalline cellulose based on the MARTINI parameter set.[22] By using the same basic methodology in the parametrization process, i.e., ensuring that the partitioning of the cellulose chains between polar and apolar phases are correct, we obtain a model that is inherently compatible with the existing models for proteins, lipids, carbohydrates, and solvents. This is demonstrated by simulating the diffusion of the CBD from cellobiohydrolase I (Cel7A) from *Trichoderma reesei* on a crystalline cellulose surface.

METHODS

Model. We took as our starting point the recently published MARTINI force field for carbohydrates.[23] The MARTINI parameter set[22] includes 18 predefined standard particle types, with different levels of polarity, which interact through pairwise Lennard-Jones potentials. Each particle, or bead, is used to represent three or four heavy atoms. Cellobiose is represented by six beads (see Figure 2), three beads per glucose unit, where each bead is positioned at the center of mass of its constituting AA atoms. To decide which bead types to use, we chose to parametrize our model using the partitioning between pure water and cyclohexane for a series of oligomers, cellobiose through cellopentaose, with the published model for cellobiose[23] as the starting point. Since there are, to the best of our knowledge, no experimental data for the partitioning of longer cellooligomers, we chose to use simulation data from an atomistic model as our reference data set. Bonded interactions were adapted directly from ref 23, with the exception of the dihedral angles, as described in the Results section.

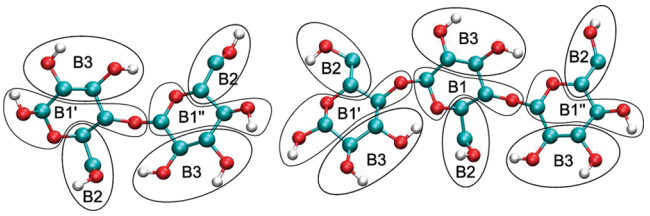

Figure 2. Mapping of atoms to CG beads for cellobiose and cellotriose, following López et al.[23]

Simulation Details. The systems considered in this present work are the following:

1. Cellopentaose, cellotetraose, cellotriose, and cellobiose in pure water and in cyclohexane, using both atomistic and coarse-grained modeling. These simulations were used for the parametrization of the nonbonded interactions, i.e., choice of bead types for the CG model. The size of the simulation box was roughly 3.6 × 3.6 × 3.6 nm in all cases.
2. A cellulose crystal in water, in a CG representation, which was used to optimize cellulose−cellulose interactions. The crystal consisted of 36 chains, each chain 40 glucose units in length, and was solvated in a box with dimensions 6.5 × 6.5 × 25.0 nm.
3. A CG model of a crystalline cellulose surface with the carbohydrate-binding domain of Cel7A attached to it, in solution.

All simulations were performed using GROMACS 4.0,[26] in a NPT ensemble. The temperature was maintained at 300 K, unless stated otherwise, using stochastic velocity rescaling,[27] and the pressure was kept at 1 atm using a Parrinello−Rahman barostat.[28] For the case of the crystal simulations, pressure scaling was applied in the lateral (x/y) directions only. Molecular graphics were produced using VMD.[29]

Coarse-Grained Simulations. For the coarse-grained simulations, we used the standard settings for the MARTINI parameter set. Nonbonded interactions were made to go smoothly to zero between 0.8 and 1.2 nm using a shift function. Integration was performed using a leapfrog algorithm with a 20 fs time step. Coordinates for the CBD were taken from Kraulis et al.[30] (PDB id: 1CBH) and were converted to a CG description using the MARTINI conversion tools. To preserve the structure of the CBD in solution, an elastic network approach was used. All pairs of backbone beads that were separated by a distance between 0.5 and 0.9 nm were restrained around that value using a harmonic potential with force constant 500 kJ mol^{-1} nm^{-2}. It has been shown that the elastic network potentials can be optimized to obtain a model that is better at preserving the internal motions of the protein,[31] but for our purposes this simple approach is

sufficient. Finally, parameters for the solvents, water and cyclohexane, were taken from the MARTINI force field.[22]

Atomistic Simulations. The atomistic simulations used the GROMOS united atoms force fields G45a3[32] for cyclohexane and G45a4[15] for carbohydrates, together with SPC water.[33] Despite GROMOS being a united-atoms force field, which means that aliphatic hydrogens are not modeled explicitly, the atomistic simulations will be referred to as AA simulations throughout this manuscript. The nonbonded interactions were handled with a twin-range cutoff approach. Inside 0.8 nm, the nonbonded interactions were updated every step, and between 0.8 nm and an outer cutoff of 1.4 nm they were updated once every 10 steps, which is the same frequency as the neighbor list updates. To account for long-range electrostatic interactions, a reaction field correction was applied with a relative permittivity of 66 in water. The cyclohexane model has no explicit charges, and for that reason no long-range correction is needed. The basic time step used was 2 fs. All bonds were kept at their equilibrium values using P-LINCS.[34]

Partitioning Free Energies. The partitioning free energy of a solute between water and cyclohexane, $\Delta\Delta G_{wc}$, can be calculated from the difference between the solvation free energies of the solute in the respective solvents, ΔG_w and ΔG_c. The partitioning free energy is related to the partitioning coefficient P_{wc} through $\Delta\Delta G_{wc} = -k_B T \log P_{wc}$, where k_B is Boltzmann's constant and T is the temperature. To this end, ΔG_w and ΔG_c were calculated using thermodynamic integration (TI), in which the solute−solvent interactions are controlled by a coupling parameter, λ. The coupling parameter λ was varied between 0 (all interactions intact) and 1 (fully decoupled state) in 25 discrete steps. Note however that, for AA simulations in water, the decoupling was actually performed in twice that number of steps, first the electrostatic and then the Lennard-Jones interactions. To avoid singularities when the Van der Waal's radii approach zero, the Lennard-Jones interactions were gradually switched to a soft-core potential with $\alpha = 0.5$, $p = 1$, and $\sigma = 0.3$. Intermolecular interactions of the solute were not coupled, meaning that the fully decoupled state corresponds to the gas phase of the solute, and the resulting energies are proper solvation free energies. For each value of λ, a 5 ns simulation was run. The output from these simulations is the derivative of G with respect to λ, $dG/d\lambda$, as a function of λ, which enables ΔG to be calculated using numerical integration over all λ's. For a more detailed description of TI, and the meaning of the soft-core parameters, we refer the reader to the GROMACS manual.

RESULTS AND DISCUSSION

The corresponding bead types are listed in Table 1. Note the difference between terminal and nonterminal residues, and also between the two different terminal residues, where B1′ is used at the reducing end and B1″ is used at the nonreducing end. Longer cellooligomers are constructed by repeated insertion of nonterminal residues.

Partitioning. As already mentioned, our starting point for the parametrization was the model for cellobiose from López et al.[23] However, this model turned out to be overall too polar, giving a partition free energy between water and cyclohexane, $\Delta\Delta G_{wc}$, of 70 kJ mol^{-1}, which is off by 50% compared to the reference value of 44 kJ mol^{-1}, calculated at an AA level. To overcome this, we settled on the model shown in Figure 2 and Table 1, which has the same basic features of the cellobiose from López et al.,[23] in that each glucose unit consists of three polar beads, one of which is more polar than the others, representing the two hydroxyl groups on C2 and C3. This model gives a $\Delta\Delta G_{wc}$ of 43 kJ mol^{-1}, in good agreement with the AA result (see Table 2). When one more glucose unit is attached to a cellobiose to make a cellotriose, the middle residue has one hydroxyl group less than the terminal residues. Consequently, the bead representing the glycosidic linkage in our model for cellotriose has a less polar bead type for residue number two than corresponding beads in residues one and three. The remaining beads in that residue are the same as in the terminal residues. The model that gives the best correspondence to the reference data is shown in Table 1. Longer cellooligomers are modeled according to the same scheme, using slightly different representations of terminal and nonterminal residues. Table 2 shows that our model matches the reference data for the whole series of cellooligomers investigated in this work, cellobiose through cellopentaose, very closely.

Table 1. Beads and Optimized Bead Types for the Model in Figure 2[a]

bead	atoms	mass (amu)	optimized bead type	original bead type[b]
B1	C1, C4, O4, O5	58.0368	Na (neutral, h-bond acceptor)	
B1′	C1, O1, H1, C4, O4, O5	75.0442	P1 (polar)	P2
B1″	C1, C4, O4, H4, O5	59.0448	P1 (polar)	P2
B2	C5, C6, O6, H6	44.0534	P1 (polar)	P1
B3	C2, O2, H2, C3, O3, H3	60.0528	P2 (more polar)	P4

[a] See the original MARTINI ref 22 for their definitions and interactions. [b] From ref 23.

Table 2. Partitioning Free Energies of Cellooligomers between Water and Cyclohexane in kJ mol^{-1}[a]

	ΔG_w^{CG}	ΔG_c^{CG}	$\Delta\Delta G_{wc}^{CG}$	ΔG_w^{AA}	ΔG_c^{AA}	$\Delta\Delta G_{wc}^{AA}$
cellobiose	−104	−61	43	−103	−59	44
cellotriose	−146	−88	58	−147	−86	61
cellotetraose	−189	−115	74	−183	−113	70
cellopentaose	−229	−139	90	−223	−140	83

[a] Statistical errors are approximately 3 kJ mol^{-1} for the CG simulations, and 5 kJ mol^{-1} for AA.

It is somewhat surprising that our model for cellobiose is different from the one in López et al.[23] After all, they should represent the same molecule, using the same force field. While the original model was parametrized against the partitioning between water and water-saturated octanol, we used water and cyclohexane instead, which may explain a large part of this difference. One reason for our choice is that we could not reproduce the results from López et al.[23] for octanol in the AA representation; in fact, we did not even get the correct sign for $\Delta\Delta G$. The reason for this might be that the water/octanol mixture, with a solute, is a fairly complicated system, which is hard to bring to convergence. Cyclohexane, being a much simpler system, does not suffer from this to the same extent. We also note a small difference in ΔG for cellobiose in water, in the AA representation, between ref 23 and the present result. This is perhaps surprising, since the same force fields are used in both

studies. It is however possible that this difference can be explained by small differences in temperature and simulation protocols.

Cellulose Crystal Structure. The next step was to investigate our model's ability to represent crystalline cellulose. To this end, a model crystal was constructed by replicating the experimental unit cell for cellulose Iβ, which is known from X-ray crystallography and neutron fiber diffraction,[1] in its three principal directions, creating a structure consisting of 36 chains (6×6), 40 glucose units long. The crystal was constructed with the two staggered faces, corresponding to the (110) and (1−10) crystallographic planes, facing out. This structure was then converted to a CG representation using the mapping in Figure 2 and placed in an orthorhombic simulation box filled with CG water. After energy minimization, the system was run for approximately 100 ns. The system quickly drifted away from the initial crystal structure and ended up in a structure in which the cellulose chains were stacked directly on top of each other, rather than the staggered arrangement of the native crystal. This structure was then stable for the rest of the simulation. This behavior is not very surprising considering that all interactions between the chains in this model are fairly attractive. In a stacked arrangement, beads of the same type can interact with each other in an optimal way. In an all-atom representation of cellulose, there exists forces that make a slightly offset arrangement more optimal, such as the ability to form hydrogen bonds between sheets, but these depend on details that the CG model evidently is too coarse to capture. To overcome the problem with artificial stacking, we introduced a repulsive component to the cellulose−cellulose interaction. In practice, we made all interactions between beads of type B1 strongly repulsive, while at the same time interactions between the side-chain beads B2 and B3 were made more attractive. Interactions between cellulose bead types and all other bead types were left untouched. The new cellulose−cellulose interaction matrix is shown in Table 4. That the optimized potential for the oligomers does not give a reasonable crystal structure is perhaps a bit disappointing, but not at all unexpected. As a matter of fact, this is analogous to that when modeling proteins using MARTINI; restraints are often needed to maintain the structural integrity of the protein, which, in a way, means making the structure a model parameter. Since a realistic representation of the structure, of both cellulose and proteins, is crucial for modeling their interactions with their respective surroundings, this is a necessary compromise. However, the fact that the internal energetics of the cellulose crystal are not correct is important to keep in mind when choosing suitable problems for our model.

Even though the interactions between the cellooligomers and the solvents (water and cyclohexane) are the same after the reparameterization as before, it is of course possible that the new intramolecular interactions have an impact on the solvation free energies and the partitioning. For that reason, a new series of simulations was performed for the CG model, identical to the ones described in the previous section with the exception that they were using the new interaction matrix. The resulting partitioning free energies are the same as those in Table 2, within the given error range.

Another feature of the resulting structure was that it developed a quite pronounced twist. This twist stemmed from the equilibrium dihedral angle between successive glucose units not being 180° in the original parameter set. Twists have been observed previously, using both CG[24] and AA[9] modeling, and also in experiments using both X-ray[1] and microscopic[35] methods. There seem to be no consensus about the nature of these twists, or how large they are. Since the predictive power of our model concerning cellulose structure is very limited, and moreover, since the twist is very easy to relate to a single model parameter, we chose to redefine the dihedral angles for the side chains around the glycosidic bonds (see Table 3), to force the cellulose chains into a flat conformation, as in ref 1.

Table 3. Energy Minima of the Dihedral Angles between the Side-Chain Beads of Two Neighboring Cellulose Monomers, along the B1−B1 Bond, for the Present Model and the Original Reference[23]

	present model	original
B2−B2	180°	−150°
B3−B3	180°	−150°
B2−B3	0°	30°

Table 4. Interaction Matrix for Cellulose−Cellulose Interactions[a]

	B1	B2	B3
B1	super repulsive	almost attractive	almost attractive
B2		attractive	attractive
B3			attractive

[a] See ref 22 for definitions of the interactions.

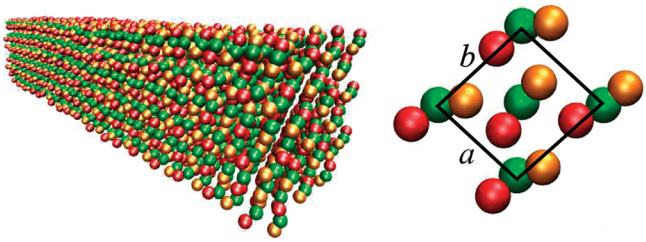

Figure 3. Snapshot of a stable coarse-grained cellulose crystal structure (left) and a representation of the cellulose Iβ crystal unit cell (right). The crystallographic axes a and b are also indicated in the figure. The third crystal axis (c) is directed parallel to the fibril long axis, which is pointing out from the figure to the right. CG beads of type B1 are green, beads of type B2 are orange, and beads of type B3 are red. Water beads are omitted for clarity.

As can be seen in Figure 3, the resulting structure is an ordered structure which retains the nearly quadratic cross-section of the initial conformation. It is possible to compare this structure to that of cellulose Iβ. A close examination of the structure shows that the unit cell parameters a and b (see Figure 3), which are predominantly a result of the nonbonded parameters, are on the high side (see Table 5). The third crystal axis, c, which is directed parallel to the chain direction and thus is determined by the B1−B1 bond distance, is in very good agreement with the experimental value. Even the unit cell angles are fairly close to the X-ray results. The time evolutions of the unit cell parameters a, b, and c are shown in Figure 4. They are stable over several nanoseconds, and their fluctuations are fairly small.

An attempt was made to fine-tune the interactions to see if the unit cell parameters could be improved. However, any other combination tried than the one in Table 4 only disrupted the desired staggered conformation. Consequently, we decided to

Table 5. Unit Cell Parameters of the CG Model, Compared to X-Ray and Neutron Diffraction Data for Cellulose Iβ[1]

	cellulose Iβ	CG model
a (nm)	0.78	0.90
b (nm)	0.82	0.97
c (nm)	1.04	1.02
α (deg)	90	90
β (deg)	90	90
γ (deg)	96.5	92

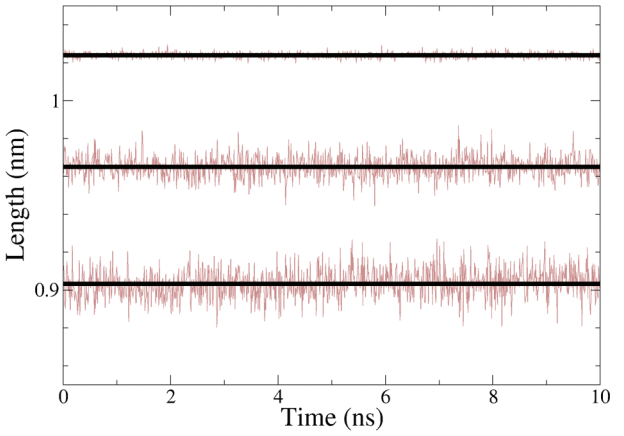

Figure 4. Time evolution of the unit cell parameters, from bottom to top, a, b, and c and their respective mean values.

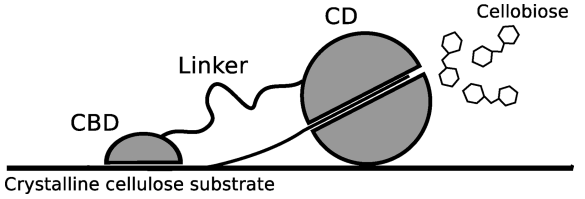

Figure 5. An illustration of the fungal cellulase Cel7A acting on a crystalline cellulose surface. The cellulase is composed of three distinct domains: a carbohydrate-binding domain (CBD) and a catalytic domain (CD) interconnected by a flexible linker peptide. During procession of the cellulase along the substrate, a cellulose chain is fed into the catalytic tunnel where the glycosidic bonds are hydrolyzed, with cellobiose as the end product.

keep the interactions from Table 4 and accept deviations from the experimental unit cell parameters.

Diffusion of a CBD on a Cellulose Surface. The carbohydrate-binding domain (CBD) of the fungal cellulase Cel7A from *Trichoderma reesei* is responsible for anchoring the cellulase to the cellulose substrate. It is attached to the catalytic domain (CD) by a flexible linker peptide (see Figure 5), and it has been shown to be essential for effective binding of the cellulase to cellulose.[36] There has also been speculation about whether the CBD also can assist the CD more directly by facilitating the lifting of the cellulose chains from the surface,[37] but recent experimental data confirm that its main purpose is to increase the local concentration of CDs on the cellulose surface.[38] However, computer simulations suggest that it might be involved directly in the recognition of loose chain ends.[39]

A common motif for all fungal CBDs is a cellulose-binding surface featuring three solvent-exposed aromatic residues, tyrosines in the case of Cel7A, which has been shown to be critical for the binding of the CBD to cellulose.[36] The spacing of these residues coincides with the spacing of the glucose units in cellulose, and it has therefore been proposed that the CBD binds to the cellulose surface essentially by favorable van der Waals interactions between the aromatic residues and the pyranose rings exposed on the flat surfaces of crystalline cellulose.[36,40] This has been partly confirmed by computer simulations that indicate the existence of localized binding grooves on the cellulose surface,[41] which makes it probable that the motion of the CBD on the surface takes place in discrete steps of about 0.5 nm in length.

Jervis et al.[25] give a value for the diffusion constant of a bacterial family 2CBD of 2×10^{-11} to 1.2×10^{-10} cm^2 s^{-1} at room temperature, which translates to an approximate rate constant of 4.8×10^4 s^{-1} to 2.9×10^5 s^{-1} assuming diffusion on a hexagonal lattice with a lattice constant of 0.5 nm. This rate is very fast compared to the intrinsic rate of the CD, which has been measured to be approximately 3.5 nm s^{-1} for the case of Cel7A,[38] suggesting that surface diffusion of the CBD does not limit cellulase activity.[25,42] Still, this rate is extremely slow compared to the time scales normally accessible in computer simulations. Even though the diffusion constant is for a different family, it is the only one that is currently available in the literature. We will assume that it is a reasonable estimate also for the diffusion constant of the CBD of Cel7A.

As mentioned above, it has been suggested that the flat face of the cellulose crystal (corresponding to the (110) crystallographic plane in the case of cellulose Iα and (100) in the case of cellulose Iβ) is the most likely to which the CBD will attach.[36] Starting from the X-ray structure of crystalline cellulose Iβ[1], we built a surface consisting of two layers of cellulose chains. Each layer was composed of 10 chains, and each chain consisted of 16 glucose units. The surface was made periodic in all directions, with each chain covalently bonded to its own periodic image in the chain direction, effectively mimicking a surface infinite in size. Using only two layers of cellulose saves computational time, but at the same time it makes the structure unstable. To prevent the crystal structure from being disrupted, one of the two layers, the "bottom" one, was made subject to harmonic restraints in all directions with a force constant of 10^3 kJ mol^{-1} nm^2. This was enough to stabilize the layer on top of it as well, even in the high temperature simulations described below. Next, the CBD was placed on top of the surface, with the three tyrosines mentioned above facing down, toward the cellulose surface. Finally, the system was solvated using a water layer approximately 5-nm-thick. Since the cutoff for interactions is only 1.2 nm, this is sufficient to prevent any interactions between surfaces and their periodic images and interactions between the CBD and surfaces to which it is not directly attached. This system was then simulated for 1 μs, during which the CBD after some small initial displacement sat completely still in one spot. Figure 6 shows the system after equilibration. As can be seen, the alignment of the three tyrosines is not perfectly parallel to the direction of the chains, which has also been noted in atomistic models.[43]

A step rate of 10^5 s^{-1} means that we would expect the CBD to take one step every 10 μs. Indeed, the equilibrated structure is stable for several microseconds. A simple way of speeding up the dynamics of the system is to raise the temperature. For this particular setup, 380 K seemed to be a lower limit to be able to see sufficient motion of the CBD during the time scales accessible. For

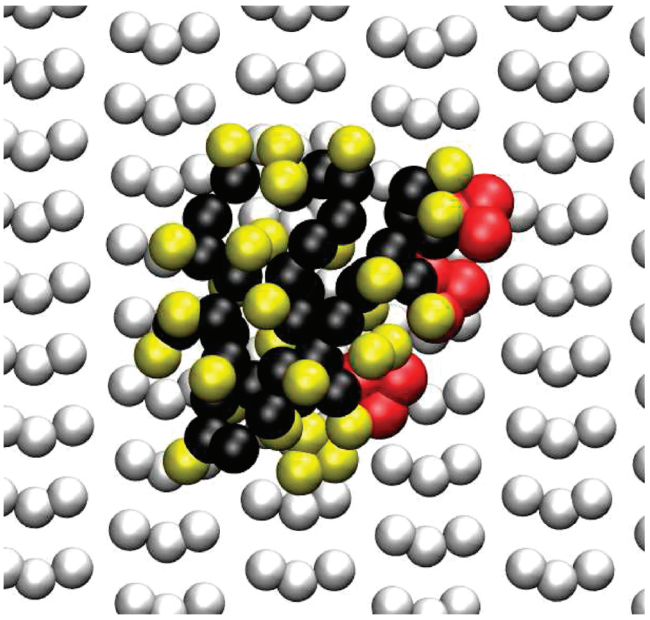

Figure 6. Top view of the CG model of a CBD on a cellulose surface. Backbone beads are black, side chains yellow, and the important tyrosine residues are colored red. The cellulose surface is light gray. Water is omitted for clarity. The chain direction is running vertically in the figure.

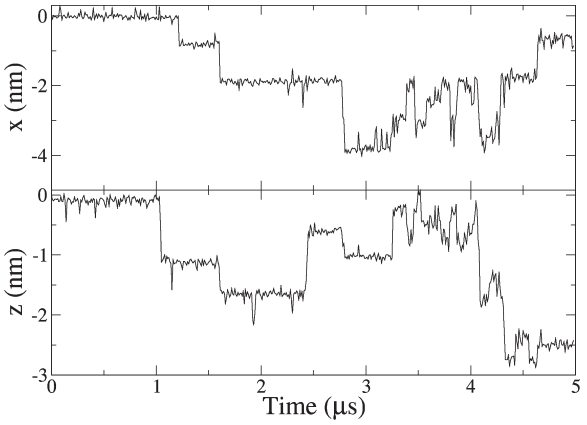

Figure 7. Position of the center of mass of the CBD as a function of simulation time at 410 K. The motion does, to a large degree, occur in discrete steps.

this reason, we ran several 5 μs simulations at elevated temperatures, between 380 and 450 K in 10 K increments. Of course, for a real system, such high temperatures would be devastating for both the crystalline cellulose and the CBD, but this is fortunately not necessarily the case in simulations. As explained above, the position restraints imposed on the cellulose were sufficient to preserve its structure even at the highest temperatures used. At the same time, the elastic network type restraints do the same thing for the CBD. For this reason, we assume that the main effect of raising the temperature is to speed up the dynamics.

Figure 7 shows the positions of the center of mass of the CBD during the simulation at 410 K. From the figure, it is clear that the motion to a high degree takes place in steps which are more or less, but certainly not always, multiples of 0.5 nm. Why the step lengths differ between steps becomes evident from looking at the trajectory in a molecular graphics representation. The motion of

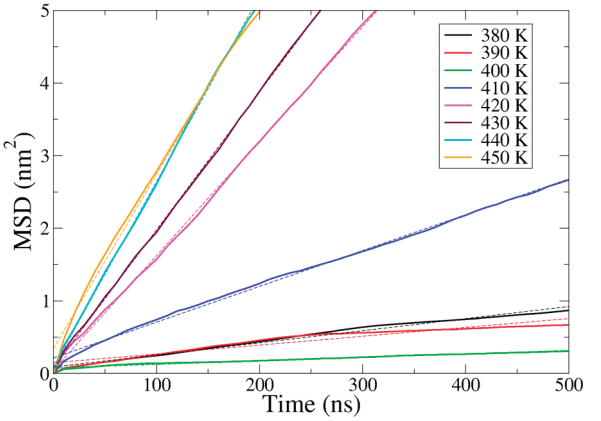

Figure 8. Calculated lateral mean square displacements of the CBD (solid) along with linear fits which were used to calculate D_l.

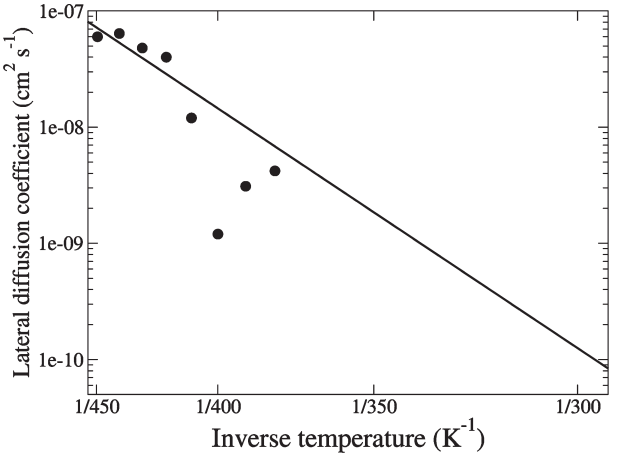

Figure 9. Linear regression, using least-squares fitting, of $\log(D_l)$ plotted against $1/T$, where D_l is calculated from the lateral mean square displacement at different temperatures using eq 1.

the CBD does not only involve pure translations but also rotations of the CBD around the surface normal, and the combination of translations and rotations between stable states leads to a variable step length.

The lateral diffusion constant D_l was calculated from the mean square displacements (MSD) of the CBD, using the Einstein relation in two dimensions:

$$D_l = \lim_{x \to \infty} \frac{1}{4} \frac{d}{dt} \langle [\mathbf{r}(t + t_0) - \mathbf{r}(t_0)]^2 \rangle_{t_0} \quad (1)$$

Figure 8 shows the calculated lateral MSD along with a least-squares fit of eq 1 to the data. Only the first 500 ns of the MSD were used for fitting due to the statistics quickly becoming too poor at longer times, especially at the lower temperatures.

Assuming stepwise diffusion where the diffusion rate follows the simple Arrhenius' law, we can write

$$D_l \propto \exp\{-\Delta G/k_B T\} \quad (2)$$

where ΔG is the free energy barrier between stable states. This means that the diffusion constant at room temperature can be obtained from a linear regression of $\log[D_l(T)]$ plotted against $1/T$. From the data presented in Figure 9, we calculate the diffusion coefficient at room temperature to be $D_l = 1.2 \times 10^{-10}$ cm^2 s^{-1}.

dx.doi.org/10.1021/ct100489z |J. Chem. Theory Comput. 2011, 7, 753–760

This is in very good agreement with the experimental value[25] stated above. Furthermore, the slope of the fit in Figure 9 gives the average barrier height between states, $\Delta G = 48$ kJ mol^{-1}, which of course poses a significant obstacle to the motion.

Until now, we have not mentioned the fact that time scales in CG simulations generally differ from time scales in atomistic simulations. This is due to the fact that coarse-graining tends to smoothen the energy landscape, which leads to faster dynamics. By calculating the self-diffusion of water, time in the MARTINI model has been found to run about 4 times faster than in atomistic models.[22] This means that the diffusion coefficients calculated above should be scaled by a factor 1/4, which leads to $D_l = 3.1 \times 10^{-11}$ cm^2 s^{-1} at room temperature, which is still within the experimental range.

CONCLUSIONS

We have constructed a coarse-grained molecular model of cellulose, consisting of three interaction sites per glucose residue, by optimizing the nonbonded parameters against the partitioning of a set of cellooligomers between water and cyclohexane, using a recently published model for cellobiose[23] as our starting point. The reference values for the partitioning free energies, which were obtained from atomistic simulations, were reproduced for a series of cellooligomers, cellobiose through cellopentaose. This indicates that the resulting model has the right balance between polar and apolar parts. The internal (cellulose−cellulose) nonbonded potentials were modified in order to obtain a crystal structure similar to the one for cellulose Iβ obtained by X-ray crystallography.[1] This means that the model probably does not correctly represent the internal energetics of the cellulose crystal. However, this was deemed a necessary compromise since a good representation of the crystal structure is important when modeling interactions between crystalline cellulose and its surroundings, e.g., solvents and other biomolecules such as proteins.

By building our model within the framework provided by the MARTINI force field,[22] we get a model that is inherently compatible with the existing model for proteins.[44] This was demonstrated by simulating the diffusive motion of the carbohydrate binding domain from *Trichoderma reesei* on a crystalline cellulose surface in explicit water. The calculated lateral diffusion coefficient is within the experimental range,[25] which is a clear indicator that the interactions between the protein model and the present model for cellulose are realistic.

Many applications directly involving cellulose are dependent on the interplay between crystalline cellulose and a large array of other molecules, e.g., solvents and proteins, as well as both synthetic and naturally occurring polymers. The present model can, in that context, provide useful information on a molecular level, which may be used to, for instance, optimize the controlled degradation of cellulose or help in tailoring new materials based on renewable resources.

ASSOCIATED CONTENT

Supporting Information. A complete GROMACS topology and coordinates for the CG cellulose crystal described in this work are provided. This information is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author
*E-mail: jacke@kth.se.

ACKNOWLEDGMENT

Computational resources were kindly provided by PDC Center for High Performance Computing under project title *Computer Modeling of Crystalline Cellulose*. The authors thank Malin Bergenstråhle for helpful discussions.

REFERENCES

(1) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.
(2) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300–14306.
(3) Atalla, R. H.; VanderHart, D. L. *Science* **1984**, *223*, 283–285.
(4) Samir, M. A. S. A.; Alloin, F.; Dufresne, A. *Biomacromolecules* **2005**, *6*, 612–626.
(5) Berglund, L. A.; Peijs, T. *MRS Bull.* **2010**, *35*, 201–207.
(6) Eichhorn, S. J.; Dufresne, A.; Aranguren, M. *J. Mater. Sci.* **2010**, *45*, 1–33.
(7) Ragauskas, A. J.; Williams, C. K.; Davison, B. H.; Britovsek, G.; Cairney, J.; Eckert, C. A.; Frederick, W. J., Jr.; Hallett, J. P.; Leak, D. J.; Liotta, C. L.; Mielenz, J. R.; Murphy, R.; Templer, R.; Tschaplinski, T. *Science* **2006**, *311*, 484–489.
(8) 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.
(9) Matthews, J. F.; Scopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.; Sugiyama, J.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2006**, *341*, 138–152.
(10) Bergenstråle, M.; Berglund, L. A.; Mazeau, K. *J. Phys. Chem. B* **2007**, *111*, 9138–9145.
(11) Bergenstråhle, M.; Wohlert, J.; Larsson, P. T.; Mazeau, K.; Berglund, L. A. *J. Phys. Chem. B* **2008**, *112*, 2590–2595.
(12) Mazeau, K.; Rivet, A. *Biomacromolecules* **2008**, *9*, 1352–1354.
(13) Zhong, L.; Matthews, J. F.; Crowley, M. F.; Rignall, T.; Talón, C.; Cleary, J. M.; Walker, R. C.; Chukkapalli, G.; McCabe, C.; Nimlos, M. R.; C., L. B., III; Himmel, M. E.; Brady, J. W. *Cellulose* **2008**, *15*, 261–273.
(14) Nishiyama, Y.; Johnson, G. P.; French, A. D.; Forsyth, V. T.; Langan, P. *Biomacromolecules* **2008**, *9*, 3133–3140.
(15) Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400–1412.
(16) 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*, 2543–2564.
(17) Sun, H.; Mumby, S. J.; Maple, J. R.; Hagler, A. T. *J. Am. Chem. Soc.* **1994**, *116*, 2978–2987.
(18) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, B. L.; Woods, R. J. *J. Comput. Chem.* **2007**, *29*, 622–655.
(19) Voth, G. A. *Coarse-graining of condensed phase and biomolecular systems*; CRC Press: Boca Raton, FL, 2009.
(20) Molinero, V.; Goddard, W. A., III. *J. Phys. Chem. B* **2004**, *108*, 1414–1427.
(21) Liu, P.; Izvekov, S.; Voth, G. A. *J. Phys. Chem. B* **2007**, *111*, 11566–11575.
(22) Marrink, S. J.; Risselada, H. J.; Yemov, S.; Tieleman, D. P.; de Vries, A. H. *J. Phys. Chem. B* **2007**, *111*, 7812–7824.
(23) López, C. A.; Rzepiela, A. J.; de Vries, A. H.; Dijkhizen, L.; Hünenberger, P. H.; Marrink, S. J. *J. Chem. Theory Comput.* **2009**, *5*, 3195–3210.
(24) 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*, 10994–11002.
(25) Jervis, E. J.; Haynes, C. A.; Kilburn, D. G. *J. Biol. Chem.* **1997**, *272*, 24016–24023.
(26) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput* **2008**, *4*, 435–447.
(27) G. Bussi, D. D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.

(28) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
(29) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
(30) Kraulis, P. J.; Clore, G. M.; Nilges, M.; Jones, T. A.; Pettersson, G.; Knowles, J.; Gronenborn, A. M. *Biochemistry* **1989**, *28*, 7241–7257.
(31) Periole, X.; Cavalli, M.; Marrink, S.; Ceruso, M. A. *J. Chem. Theory Comput.* **2009**, *5*, 2531–2543.
(32) Schuler, L. D.; Daura, X.; van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205–1218.
(33) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B. E., Ed.; Riedel: Dordrecht, The Netherlands, 1981; p 331.
(34) Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
(35) Hanley, S. J.; Revol, J.-F.; Godbout, L.; Gray, D. G. *Cellulose* **1997**, *4*, 209–220.
(36) Lehtiö, J.; Sugiyama, J.; Gustafsson, M.; Fransson, L.; Linder, M.; Teeri, T. T. *Proc. Natl. Acad. Sci.* **2003**, *100*, 484–489.
(37) Ruovinen, J.; Bergfors, T.; Teeri, T.; Knowles, J. K. C.; Jones, T. A. *Science* **1990**, *249*, 380–386.
(38) Igarashi, K.; Koivula, A.; Wada, M.; Kimura, S.; Penttilä, M.; Samejima, M. *J. Biol. Chem.* **2009**, *284*, 36186–36190.
(39) Nimlos, M. R.; Matthews, J. F.; Crowley, M. F.; Walker, R. C.; Chukkapalli, G.; Brady, J. W.; Adney, W. S.; Cleary, J. M.; Zhong, L.; Himmel, M. E. *Protein Eng. Des. Sel.* **2007**, *20*, 179–187.
(40) Reinikainen, T.; Ruohonen, L.; Nevanen, T.; Laaksonen, L.; Kraulis, P.; Jones, T. A.; Knowles, J. K. C.; Teeri, T. T. *Proteins* **1992**, *14*, 475–482.
(41) 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*, 1447–1453.
(42) Ting, C. L.; Makarov, D. E.; Wang, Z. *J. Phys. Chem. B* **2009**, *113*, 4970–4977.
(43) 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.* **2010**, *344*, 1984–1992.
(44) Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. *J. Chem. Theory Comput.* **2008**, *4*, 819–834.