J|A|C|S ARTICLES

Published on Web 09/29/2009

Conformational Flexibility of Soluble Cellulose Oligomers: Chain Length and Temperature Dependence

Tongye Shen,[†,‡] Paul Langan,[§] Alfred D. French,[⊥] Glenn P. Johnson,[⊥] and S. Gnanakaran*[,†]

Theoretical Biology & Biophysics Group, Center for Nonlinear Studies, and Biosciences, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, and Cotton Structure and Quality Research Unit, USDA, ARS, SRRC, New Orleans, Louisiana 70124

Received April 27, 2009; E-mail: gnana@lanl.gov

Abstract: Structures, dynamics, and stabilities of different sized cellulosic oligomers need to be considered when designing enzymatic cocktails for the conversion of biomass to biofuels since they can be both productive substrates and inhibitors of the overall process. In the present work, the conformational variability, hydrogen bonding, and mechanical properties of short, soluble cellulose chains are investigated as a function of chain length. Cellulose oligomers consisting 2, 4, and 6 β-D-glucose units are examined in explicit solvent using replica exchange molecular dynamics (REMD) which provides a rigorous evaluation of the relative stabilities of different conformations and their temperature dependencies. This application of REMD to oligosaccharides in solution also allows evaluation of the quality of the force-field and its suitability for sampling carbohydrates efficiently. Simulation results are analyzed in synergy with polymer theory and compared to known measurements of oligomers and crystals. As the chain length is increased, the conformations of the oligomers become more rigid and likely to form intrachain hydrogen bonds, like those found in crystals. Several other conformations and hydrogen bonding patterns distinguish these short cellulose chains from those in cellulose crystals. These studies have also addressed the key role played by solvent on shifting the conformational preferences of the oligosaccharides with respect to vacuum and crystals. Correlation between pyranose ring flipping and the conformation of the 1,4-glycosidic bond was observed.

Introduction

Cellulose occurs as microfibrils in the cell walls of higher plants and is also produced by some bacteria, algae, fungi, and sea animals. Two distinct crystal phases, namely, I_α and I_β, are found in nature, collectively called cellulose I, in proportions that depend on the origin of the cellulose.[1] For many applications, raw cellulosic material is pretreated or processed so that the cell wall architecture is disrupted and so that the cellulose crystal structure is transformed from cellulose I to other crystal phases that have improved properties for their specific application. The processes of regeneration and mercerization with alkali yield cellulose II. Processing cellulose I and cellulose II with amines yields cellulose III_I and III_{II}, collectively referred to as cellulose III. By using a combination of X-ray and neutron crystallographic techniques with various spectroscopic measurements, the detailed atomic structures of the various forms of cellulose, and the complexes of those forms with various amines, have recently been reported.[2−7] These different crystal phases have very different properties, such as different susceptibilities to digestion by cellulases. That is of importance in the conversion of biomass into sugars that can be used in the production of biofuels.[8−10]

As part of our ongoing research in biomass conversion, we have been studying the molecular aspects, both structural and dynamical, of the formation of the different cellulose polymorphs and how these molecular aspects are also important for the interaction of the different polymorphs with cellulases.[11] The present work concerns individual molecules of cellulose oligomers in aqueous solution. Several questions remain unanswered regarding the conformational dynamics of these short chains. For example, do the glycosidic linkages have greater flexibility when freed from the restrictions imposed by the crystal environment? How do the local conformational propensi-

† Theoretical Biology & Biophysics Group, Los Alamos National Lab.
‡ Center for Nonlinear Studies, Los Alamos National Lab.
§ Biosciences, Los Alamos National Lab.
⊥ USDA.

(1) Atalla, R. H.; Vanderhart, D. L. *Science* **1984**, *223*, 283–285.
(2) Langan, P.; Nishiyama, Y.; Chanzy, H. *J. Am. Chem. Soc.* **1999**, *121*, 9940–9946.
(3) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.
(4) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300–14306.
(5) Wada, M.; Chanzy, H.; Nishiyama, Y.; Langan, P. *Macromolecules* **2004**, *37*, 8548–8555.
(6) Wada, M.; Nishiyama, Y.; Langan, P. *Macromolecules* **2006**, *39*, 2947–2952.
(7) Wada, M.; Heux, L.; Nishiyama, Y.; Langan, P. *Cellulose*, published online July 11, 2009, http://dx.doi.org/10.1007/s10570-009-9338-5.
(8) Amano, Y.; Nozaki1, K.; Araki, T.; Shibasaki, H.; Kuga, S.; Kanda, T. *Cellulose* **2001**, *8*, 267–274.
(9) Igarashi, K.; Wada, M.; Samejima, M. *FEBS J.* **2007**, *274*, 1785–1792.
(10) Chen, Y.; Stipanovic, A. J.; Winter, W. T.; Wilson, D. B.; Kim, Y.-J. *Cellulose* **2007**, *14*, 283–293.
(11) Flugge, L. A.; Blank, J. T.; Petillo, P. A. *J. Am. Chem. Soc.* **1999**, *121*, 7228–7238.

 10.1021/ja9034158 CCC: $40.75

ties change with increasing degree of polymerization? Which internal motions are the most sensitive to thermal perturbation? It is useful to address these questions for different lengths of cellulose chains because during enzymatic depolymerization, varying lengths of cellulose chain fragments are generated. Exoglucanases predominantly produce cellobiose and single glucose units due to their processive action on cellulose microfibrils. On the other hand, catalysis by endoglucanases produces a large fraction of cellulose chain fragments due to random and nonprocessive actions. Short glucose oligomers produced by these enzymes can hinder the overall efficiency of biomass conversion to sugar. These effects include product inhibition, competition with substrates of other enzymes, and interference with the synergetic action of cellulases by adsorption back onto the cellulose microfibril. Therefore, structural and dynamical characterization of different oligomer sizes and their relative stabilities needs to be taken into consideration when designing enzymatic cocktails.

For the current study, chain lengths (based on the monosaccharide unit, β-D-glucose) $n = 2$ (dimer), $n = 4$ (tetramer), and $n = 6$ (hexamer) have been considered. We report our investigation of the conformational variability and hydrogen bonding of soluble cellulose chains as a function of chain length. Their relative stabilities and temperature dependence are studied as well. Unlike peptides, short carbohydrate chains exhibit complex local modes involving pyranose ring structures. Thus, it has been difficult to structurally characterize oligoglucoses or short cellulose chain fragments beyond the dominating conformation at room temperature.[12−14]

In the present work, all-atom replica exchange molecular dynamics (REMD) simulations have been used to probe the molecular aspects of the soluble oligoglucoses. The all-atom details of the simulations ensure that intrachain and solvent hydrogen bonding, hydrophobicity and solvent ordering, and dynamics are properly taken into account. REMD has been successfully applied to study the folding of peptides and proteins.[15−17] Recently, REMD was utilized to probe the shift in secondary structural propensities of peptides as a function of peptide length and to evaluate the ability of force fields to accurately reproduce trends observed by experiments.[18,19]

This work is one of only a few theoretical studies of the conformational dynamics of varying lengths of carbohydrate chains.[12,20] Therefore, our effort will also serve as a measure of the quality of the force field and the capability of the chosen theoretical method (REMD) to provide efficient sampling of carbohydrates. To our knowledge, this work stands out as the first REMD study on polysaccharides in aqueous environment. Previously only disaccharides in vacuum[21] have been studied with REMD. The closest system studied in aqueous solution was glycerol.[22] The use of REMD and long simulation times should provide more thorough sampling than in previous studies, and the multiple temperatures considered add considerable new information. Comparisons to known measurements of oligomers and crystals are made whenever possible.

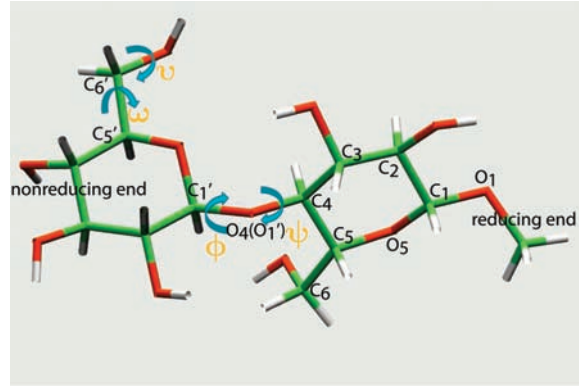

Figure 1. The structure of methyl β-D-cellobioside.

Methods

A. Cellulose Oligomers. Molecular models of dimer (methyl β-D-cellobiose), tetramer (methyl β-D-cellotetraose), and hexamer (methyl β-D-cellohexaose) are considered in this study. The online carbohydrate builder at www.glycam.com of the Woods Group was used to create the initial models.[23] After a careful visual inspection using VMD,[24] we used the tleap module of the AMBER software to build the topology and parameter files.[25] Throughout the article, we follow the conventions recommended by IUPAC-IUB Joint Commission on Biochemical Nomenclature[26] for naming atoms and defining the direction of cellulose; each polymer starts with the reducing end (the end close to O5) and terminates with the nonreducing end. For example, monomer $i + 1$ is located toward the nonreducing end relative to monomer i as illustrated with the dimer in Figure 1.

B. Force Field and Initial Simulation Setup. The AMBER/GLYCAM force field was used with the parameter file, glycam04l.dat and the prep file, GLYCAM04.prep, downloaded from the GLYCAM web portal, www.glycam.com. These files were used with the local tleap module from AMBER to create prmtop files.[25] The TIP3P model of water was used.[27] We rely on this well-tested atomistic water model to properly describe any potential interactions between solvent and oligosaccharide. It is important to have a good description of the interaction of water and oligosaccharide, from hydrogen bonding to desolvation effects, as previously demonstrated with quantum mechanical calculations[28,29] and molecular simulations.[13,30,31] Unlike the case for proteins, where AMBER force

(12) Hardy, B. J.; Sarko, A. *J. Comput. Chem.* **1993**, *14*, 831–847.
(13) Hardy, B. J.; Sarko, A. *J. Comput. Chem.* **1993**, *14*, 848–857.
(14) DeMarco, M. L.; Woods, R. J. *Glycobiology* **2008**, *18*, 426–440.
(15) Gnanakaran, S.; Nymeyer, H.; Portman, J.; Sanbonmatsu, K. Y.; Garcia, A. E. *Curr. Opin. Struct. Biol.* **2003**, *13*, 168–174.
(16) Gnanakaran, S.; Hochstrasser, R. M.; Garcia, A. E. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 9229–9234.
(17) Garcia, A. E.; Onuchic, J. N. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13898–13903.
(18) Gnanakaran, S.; Garcia, A. E. *J. Phys. Chem. B* **2003**, *107*, 12555–12557.
(19) Gnanakaran, S.; Garcia, A. E. *Proteins* **2005**, *59*, 773–782.
(20) Queyroy, S.; Muller-Plathe, F.; Brown, D. *Macromol. Theory Simul.* **2004**, *13*, 427–440.
(21) Campen, R. K.; Verde, A. V.; Kubicki, J. D. *J. Phys. Chem. B* **2007**, *111*, 13775–13785.
(22) Yongye, A. B.; Foley, B. L.; Woods, R. J. *J. Phys. Chem. A* **2008**, *112*, 2634–2639.
(23) Woods, R. J. *GLYCAM Web*; Complex Carbohydrate Research Center, University of Georgia: Athens, GA, 2008.
(24) Humphrey, W.; Dalke, A.; Schulten, K. *J. Mol. Graphics* **1996**, *14*, 33–38.
(25) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
(26) IUPAC; IUB. *Pure Appl. Chem.* **1983**, *55*, 1269−1272.
(27) Jorgensen, W. L.; Chandreshkar, J.; Madura, J. D.; Imprey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
(28) Cramer, C. J.; Truhlar, D. G. *J. Am. Chem. Soc.* **1993**, *115*, 5745–5753.
(29) Bosma, W. B.; Appell, M.; Willett, J. L.; Momany, F. A. *J. Mol. Struct. THEOCHEM* **2006**, *776*, 1–19.
(30) Leeflang, B. R.; Vliegenthart, J. F. G.; Kroon-Batenburg, L. M. J.; van Eijck, B. P.; Kroon, J. *Carbohydr. Res.* **1992**, *230*, 41–61.
(31) Rohfritsch, P. F.; Frank, M.; Sandstrom, C.; Kenne, L.; Vliegenthart, J. F. G.; Kamerling, J. P. *Carbohydr. Res.* **2007**, *342*, 597–609.

fields are implemented with 1−4 nonbonded scaling, the GLYCAM carbohydrate force field is normally used without 1−4 nonbonded scaling.[32] Long molecular dynamics simulations of all our systems with and without the 1−4 scaling gave fairly minor differences between the two cases, but these effects were not deemed significant. With 1−4 scaling, the sampling is slightly broader in terms of dihedral angles, similar to the effects of slightly raising the temperature. To be consistent with the previous simulations using GLYCAM, we carried out our simulations with no 1−4 scaling.

The correct box size for each solvated system was determined by carrying out simulations at a constant pressure (1.0 atm) starting with the initial tleap built solvent box size. The size of the initial tleap built solvent box was chosen to ensure that a layer of water of at least 8 Å thick surrounded the solute. These simulations were carried out for 500 ps for the dimer and tetramer systems and 1 ns for the hexamer. Simulations were then carried out at constant volume for 1 ns for the dimer system and approximately 3 ns for the tetramer and the hexamer systems. The final configuration from these simulations was used as the initial configuration for replica exchange molecular dynamics simulations at constant volume.

C. Replica Exchange Molecular Dynamics. As described above, cellulose oligomers were simulated in explicit aqueous solution under periodic boundary conditions by REMD methodology. The suite of programs in AMBER 6 was modified to incorporate REMD.[33−35] REMD was implemented for all systems at constant volume and a fixed number of atoms. The number of replicas and the temperature range depend primarily on the number of water molecules. The dimer was solvated with 714 water molecules and the linear dimension of the cubic box was set 28.0 Å, consistent with the density of water at room temperature and 1 atm. For the dimer, REMD was implemented with 24 replicas that covered a temperature range of 276−469 K in intervals of 6−12 K. The tetramer was solvated with 1833 water molecules and the dimension of the cubic box was 38.35 Å. For this system, REMD was implemented with 42 replicas that covered a temperature range of 275−551 K in intervals of 4−10 K. The hexamer was solvated with 3741 water molecules and the dimension of the solvent box was 48.5 Å. For this system, REMD was implemented with 48 replicas that covered a temperature range of 297−557 K in intervals of 3−10 K.

A time step of 2 fs was used in all replicas. The long-range electrostatic interactions were treated by the particle-mesh Ewald method.[36] All bond distances involving hydrogen atoms were constrained by using SHAKE.[37] The temperature was regulated by the Nose−Hoover method.[38,39] Nonbonded interactions were updated every 10 integration steps. For all three systems, the exchange between replicas was attempted every 0.25 ps and trajectories were collected every 0.5 ps. After preparing the system according to the procedure described above, we performed REMD simulations long enough to obtain reasonable sampling. We verified that the properties of interest such as dihedral angles and end-to-end distance all converged. The initial run times of another 2−4 ns are not considered in the analysis in each of the oligomer systems. The subsequent times of the simulations used for this study are 8 ns per replica for the dimer, 12 ns per replica for the tetramer, and 14 ns per replica for the hexamer.

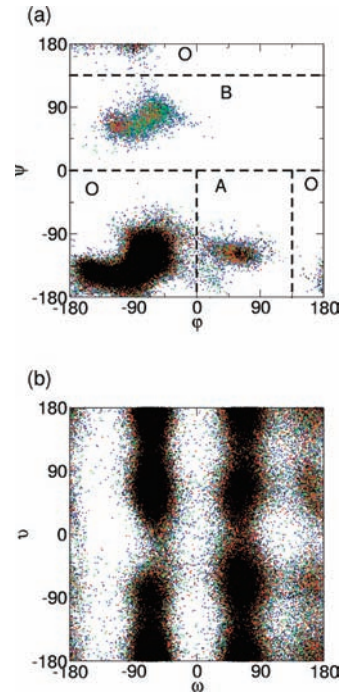

Figure 2. (a) The distribution of important torsion angles for the dimer is shown for the pair $\phi-\psi$ and (b) the pair $\omega-\nu$. The division of backbone torsion conformational space into three discrete regions O, A, and B is also displayed in panel a.

Results and Discussion

A. Backbone and Side-Chain Torsion Angles. First, we examine the torsion angles that characterize the backbone conformation of the carbohydrate chains. Figure 2 shows how the populations of four important torsion angles are distributed in the dimer. Two backbone torsion angles that specify the glycosidic linkage between two monomers are ϕ (O5′−C1′−O4−C4) and ψ (C1′−O4−C4−C5). Throughout this paper, the prime symbol is used to label atoms of monomer $i + 1$, which is downstream of monomer i. Figure 2a gives an overview of the $\phi-\psi$ distribution of the dimer collected from all 24 replicas covering the temperature range of 276−469 K. The color of the data points indicates the four temperature groups containing six temperature replicas each (low: black, midlow: red, midhigh: green, high: blue). As demonstrated in Figure 2a for the dimer, the ϕ and ψ angles are quite localized even at high temperatures, though a larger torsion angle space is sampled, illustrated by the fact that the blue dots are more widespread than the black dots. There are three major basins based on the sampling of $\phi-\psi$. The major basin, basin O, shaped like a boot, is located in the left-lower corner of the plot. The other two basins, A and B, are centered around (60°, −120°) and (−80°, 70°), respectively. Each conformation can then be associated with a coarse-grained discrete representation that is classified according to these basins. For practical purposes, we define a conformation as being in state A, when (ϕ,ψ) satisfies $0° < \phi < 135°$ and $\psi < 0°$, and state B, if ψ satisfies $0° < \psi < 135°$. The remaining region belongs to state O. The global minimum of the free energy profile is located around (−75°, −120°), which is consistent with the result (−77.5°, −128.6°) of a previous

(32) Kirschner, K. N.; Woods, R. J. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10541–10545.
(33) Nymeyer, H.; Gnanakaran, S.; Garcia, A. E. *Methods Enzymol.* **2004**, *383*, 119–149.
(34) Garcia, A. E.; Sanbonmatsu, K. Y. *Proteins* **2001**, *42*, 345–354.
(35) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
(36) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
(37) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comp. Phys.* **1977**, *23*, 327–341.
(38) Nose, S. *Mol. Phys.* **1984**, *52*, 255–268.
(39) Hoover, W. *Phys. Rev. A* **1985**, *31*, 1695–1697.

simulation of cellobiose in solution with a different force-field.[40] None of the results concerning the populations of these states presented below are sensitive to the precise definition of the boundaries, since the areas near these boundaries are almost void of data points.

The temperature dependence of the backbone conformational population can be examined in terms of these three discrete states. It is difficult to plot directly the changes of two-dimensional free energy profile with varying temperature. Instead, conformations are binned into one of the three states as a function of temperature. For the purpose of comparison, only the middle linkage of the oligomer is considered since the dimer only has a single glycosidic linkage while the tetramer and hexamer have three and five, respectively. Specifically, these are the linkages between the second and the third rings in the tetramer and between the third and the fourth rings in the hexamer. As shown in Figure 3, the state O dominates the population at low temperature. States A and B are increasingly populated with rising temperature. Comparing across oligomers, the dimer is the most flexible molecule of the three oligomers indicated by a higher percentage of time spent sampling away from the ground state basin O. Around 300 K for the dimer, we can estimate from $\Delta G = -k_B T \times \ln$ (%A/%O) that the free energy of basin A is about 2.9 kcal/mol above the ground basin, O. Similarly, the basin B is about 3.7 kcal/mol above at the same time.

Interestingly, the free energy profiles of backbone torsions of solvated cellulose oligomers reported here capture the same three basins seen in the quantum mechanical potential energy surface calculations of cellobiose in vacuum.[41] Quantum mechanics (QM) calculations identify basin A as the global minimum. However, basin O is the global minimum in our studies of solvated oligosaccharides. The bootlike shape and the landscape within this basin O is similar to that obtained through QM calculations of the cellobiose in vacuum.[41] Crystalline cellulose I and II belong to this basin and it is proximal to the 2-fold screw axis.

The ground state (O, the basin to which crystal cellulose I[3] and II[2] belong) of (ϕ,ψ) obtained from this study is consistent with various experimental data. Several crystal structures of cellulose oligomers have backbone torsion angles in this basin, such as cellobiose[42] (−76.3°, −132.3°), methyl β-cellotrioside,[43] averaged over eight pairs (−94.4°, −146.3°), cellotetraose,[44] averaged over six pairs (−94.4°, −146.7°), and dicyclohexyl cellobioside,[45] averaged over three pairs of the central linkage (−94.2°, −152.4°). Several oligomers bound to enzyme are also located in basin O.[46] There are various NMR data supporting our finding as well. NMR chemical shift data of cellotriose to cellohexaose in aqueous solution,[47] and cellobiose and cellotetraose in ionic liquid solutions,[48] point to the backbone similarities between these oligomers and longer cellulose polymers that adopt the conformation of cellulose II. NMR *J*-coupling data of cellobiose[49] for $^3J_{H,C}$, 4.08 and 4.75 for ϕ_H and ψ_H, respectively, are consistent with the values of (ϕ,ψ) reported here for this basin.

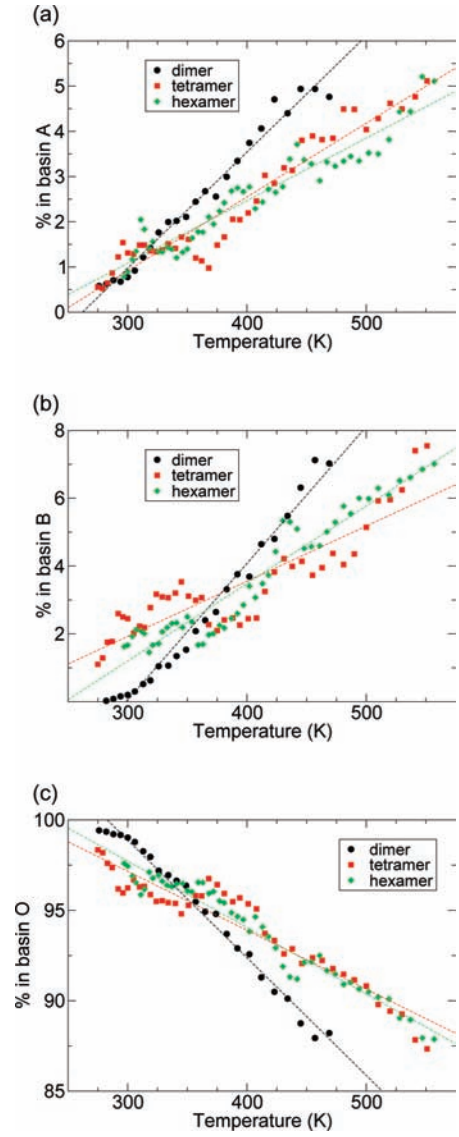

Figure 3. The population of the torsion angle of the middle linkage of the oligomers for the three regions (a) A, (b) B, and (c) O is shown as a function of temperature. Fitted dashed lines are included as a guide to capture the overall temperature profiles.

Besides these two backbone torsion angles, side chain torsion angles ω and ω* have been analyzed. These torsion angles ω (O5−C5−C6−O6) and ω* (C4−C5−C6−O6), describe the conformations of the hydroxymethyl side chains. Conventionally, a two-letter code is used to describe these three conformations; *gg* indicates that ω is gauche, while ω* is gauche+; *gt* indicates that ω is gauche+ and ω* is trans; *tg* indicates ω is trans and ω* is gauche−. By the nature of the tetrahedral carbon, $\omega \simeq \omega^* + 120$ is always satisfied. Thus, we focus on the differences among the one-dimensional profiles of ω with temperature. As shown in Figure 4, the *gg* conformation is the most populated among the three conformations. The dominance of *gg* (and followed by *gt*) is consistent with previous NMR

(40) Pereira, C. S.; Kony, D.; Baron, R.; Muller, M.; van Gunsteren, W. F.; Hunenberger, P. H. *Biophys. J.* **2006**, *90*, 4337–4344.
(41) French, A. D.; Johnson, G. P. *Can. J. Chem.* **2006**, *84*, 603–612.
(42) Chu, S. S. C.; Jeffrey, G. A. *Acta Crystallogr.* **1968**, *B24*, 830–838.
(43) Raymond, S.; Henrissat, B.; Qui, D. T.; Kvick, A.; Chanzy, H. *Carbohydr. Res.* **1995**, *277*, 209–229.
(44) Gessler, K.; Krauss, N.; Steiner, T.; Betzel, C.; Sarko, A.; Saenger, W. *J. Am. Chem. Soc.* **1995**, *117*, 11397–11406.
(45) Yoneda, Y.; Mereiter, K.; Jaeger, C.; Brecker, L.; Kosma, P.; Rosenau, T.; French, A. *J. Am. Chem. Soc.* **2008**, *130*, 16678–16690.
(46) French, A. D.; Johnson, G. P.; Kelterer, A.-M.; Dowd, M. K.; Cramer, C. J. *Int. J. Quantum Chem.* **2001**, *84*, 416–425.
(47) Dudley, R. L.; Fyfe, C. A.; Stephenson, P. J.; Deslandes, Y.; Hamer, G. K.; Marchessault, R. H. *J. Am. Chem. Soc.* **1983**, *105*, 2469–2472.
(48) Moulthrop, J. S.; Swatloski, R. P.; Moyna, G.; Rogers, R. D. *Chem. Commun.* **2005**, *2005*, 1557–1559.
(49) Cheetham, N. W. H.; Dasgupta, P.; Ball, G. E. *Carbohydr. Res.* **2003**, *338*, 955–962.

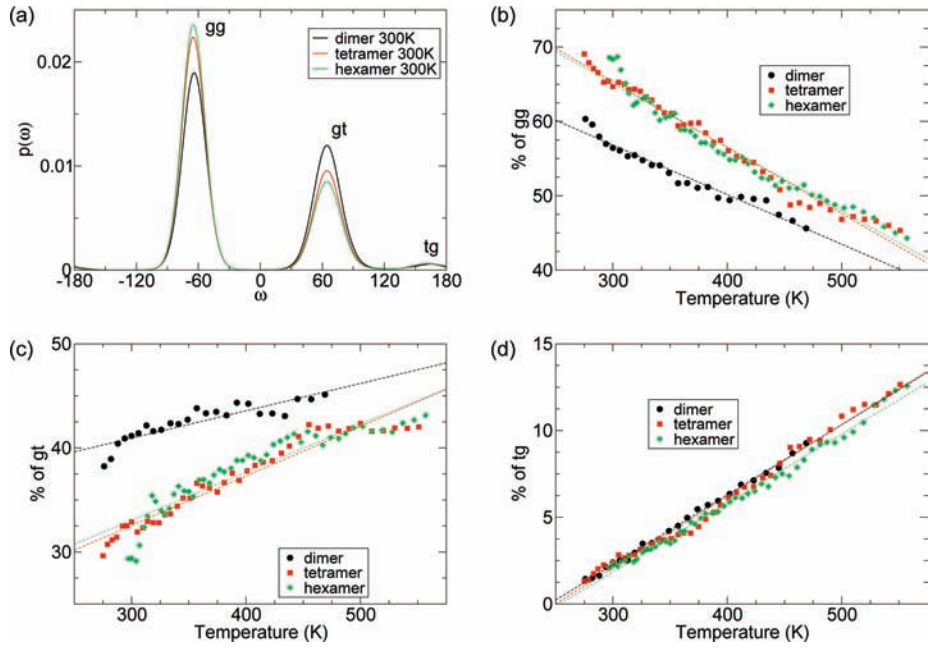

Figure 4. (a) The distribution of the torsion angle of the hydroxymethyl side chain ω, $p(\omega; 300\ \mathrm{K})$, of the middle cellobiose unit of the chain. The discrete population of (b) *gg*, (c) *gt*, and (d) *tg* is shown as a function of temperature.

data on oligosaccharides in solution[48,47,45] and computer simulation of cellobiose in solution.[40] The population of *gg* decreases with increasing temperature, whereas the populations of *gt* and *tg* grow, as expected. Comparing across the three oligomers at the same temperature, the hexamer would appear to have the most rigid hydroxymethyl side chain, based on the high dominance of population *gg*, while the dimer has the most flexible side chain, since it has a broader distribution, and correspondingly a larger entropy ($-k_B \int d\omega\, p \ln p$). The population of *tg* was not sensitive to the oligomer length. Another torsion angle, ν (C5−C6−O6−H6), further specifies the orientation of the hydroxyl group of the tip of the side chain. The relationship between ω and ν are shown in Figure 2b for the dimer. Although the torsion angle ν is less critical for deciding the conformations of the oligomers, its value can be important for hydrogen bond formation (subsection C).

B. Sugar-Ring Puckering. Another important conformational property associated with oligosaccharides is the individual glucose ring conformation. The most stable conformation of glucopyranose is the $^4C_1$ (normal chair). However, the ring can potentially flip to other less energetically favored states at higher temperatures. Although there are more complicated ways of classifying the six-membered ring conformations[50] using a three-parameter set, one convenient way to determine binarily whether the ring is in the usual chair conformation is to examine three alternate torsion angles of the total six backbone torsion angles of the ring. Specifically, values of the O5−C1−C2−C3, C2−C3−C4−C5, and C4−C5−O5−C1 torsion angles are calculated. For a chairlike conformation of the ring, these torsion angles have stable values of about +55°. We consider the ring to be in a normal, chairlike configuration if all three angles are positive. If any one of these three torsion angles becomes less than zero, the ring is considered to be in a strongly distorted non-$^4C_1$ conformation. Figure 5a shows the mean percentage of rings that did not meet this generous criterion as a function of temperature for dimer, tetramer, and hexamer. At low

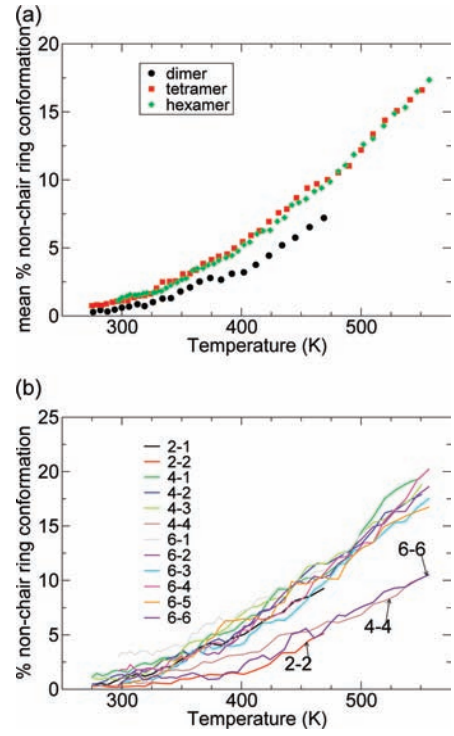

Figure 5. (a) The mean percentage of glucopyranose rings in strongly distorted non-$^4C_1$ conformations is shown as a function of temperature; (b) values for individual rings.

temperature, the occurrence of ring flipping (departure from the normal chair) is very rare. As expected, the occurrence of ring flipping increases with temperature, which is consistent with a previous finding.[12] On average, ring flipping seems to occur more frequently with an increasing level of polymerization as shown in Figure 5a. The tendency to flip can be divided to fall into two different classes: the ring at the nonreducing end and all other rings. Indeed, Figure 5b shows that the glucose ring at the nonreducing end undergoes fewer puckering flips than the

(50) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

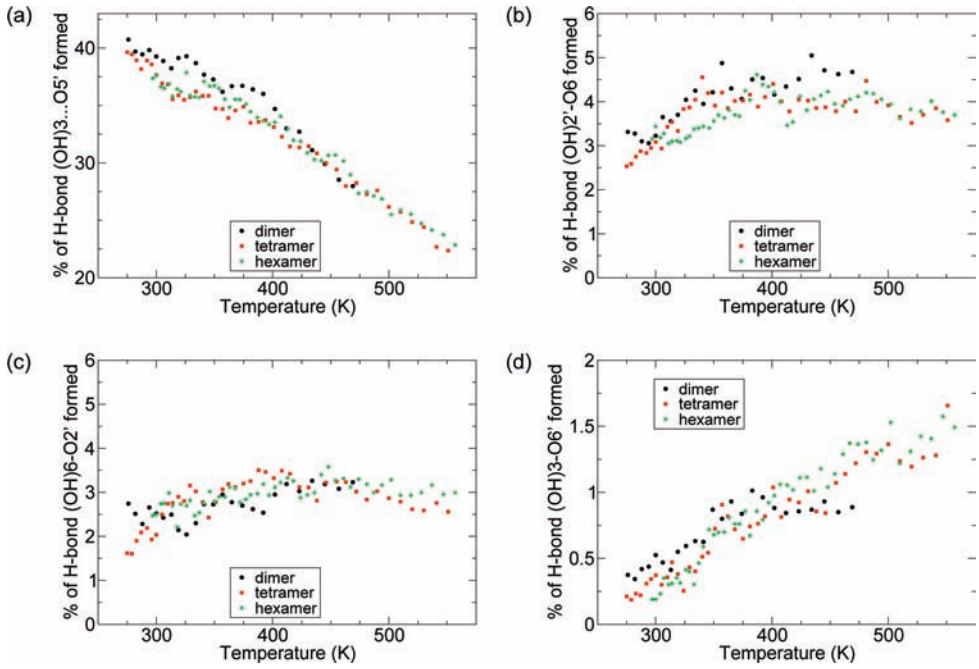

Figure 6. The intrachain hydrogen bonding percentage (between the middle glucoses i and $i + 1$, $i = 1, 2, 3$ for dimer, tetramer, and hexamer respectively) is shown as a function of temperature, for selected cases.

other rings. All other rings have a similar ratio of flipping among themselves. All three oligomers display a similar dependence on temperature. The apparent stability of rings in the dimer shown in the mean percentage plot is due to the fact that the nonreducing end ring is weighted more in the dimer compared to the tetramer and hexamer.

We further examined the cross-correlation between the state of ring-flipping and the state of backbone torsion angles. Using the case of the dimer as an example, we calculated what percentage of the dimer has at least one of the two rings in the nonchair like conformation for a given torsion angle state of the glycosidic linkage. It turns out that state A has the highest probability (11.2%), which is followed by state O (4.7%). If the molecule is in state B, then it has very low chance of any ring anomaly (1.4%). Earlier work,[51] based on the study of an analogue of the dimer that lacked all hydroxyl groups, led to similar conclusions: (1) the reducing ring is more likely to assume a shape other than the usual chair and (2) the alternative ring shapes are most likely for the A conformation.

C. Hydrogen Bonding. Hydrogen bonds (H-bonds) are one of the most dominant types of interactions that can stabilize conformations in sugar molecules. For cellulose chains, there are potential intrachain H-bonds between sequential monomers i and $i + 1$ and competing hydrogen bonding to solvent water. Again, since the properties of the H-bonds may depend on the position of the linkage for the tetramer and the hexamer, we focus on the H-bonds formed between the middle monomer units. Thus, H-bonds formed between the second and the third monomer units for the tetramer and the third and fourth monomer units for the hexamer are considered. The existence of hydrogen bonding is defined in terms of the following geometric criteria: (1) the distance between the donor and acceptor oxygen atoms must be less than 3.5 Å and (2) the angle of donor−hydrogen−acceptor must be more than 135°. This criterion has been typically used to capture H-bond formation from snapshots of classical MD simulation, and it may be a bit strict from the viewpoint of experimental measurements or quantum mechanical calculations.[52]

As expected, the H-bonding to solvent competes with intrachain H-bonds in oligosaccharides. Previous studies have also focused on the importance of competing hydrogen bonding interactions with water.[13,32,53,30,31] Figure 6 shows the percentages of intrachain H-bonds formed as a function of temperature for some representative H-bonds listed in Table 1. The H-bond O3−H···O5′, widely observed in crystals of cellulose and related molecules, has the highest percentage of formation (about 40% at 280 K) in the current study. The percentage of intrachain H-bond formation is much higher in a crystalline environment for the same temperature.[54] The observation that solvent lowers the percentage of intrachain H-bond formation is consistent with previous NMR and MD studies.[30,53] Particularly, by computing radial distribution functions, Leeflang et al. noted the ordering of first shell water around hydroxy group O3−H and its exposure to the solvent.[30] With increasing temperature, the percentage of O3−H···O5′ decreases, but it still has a moderate probability of formation at very high temperature. With increasing temperature, both H-bonds O2′−H′···O6 and O6−H···O2′ have small and consistent percentages throughout the temperature range considered. The H-bond, O3−H···O6′, has a very low probability of formation, but its population slightly increases with increasing temperature. Although a hydrogen bond between O2′ and O6 is usually observed with a *tg* orientation of O6, the dominant orientation in these models is *gg*, even when the O6−O2′ hydrogen bonds are formed. Recently, an instance of O6−O2′ hydrogen bonding in the crystalline state with O6 *gg* was observed.[45]

(51) Stortz, C. A.; French, A. D. *Mol. Simul.* **2008**, *34*, 373–389.
(52) Kirschner, K. N.; Woods, R. J. *J. Phys. Chem. A* **2001**, *105*, 4150–4155.
(53) Kroon-Batenburg, L. M. J.; Kroon, J.; Leeflang, B. R.; Vliegenthart, J. F. G. *Carbohydr. Res.* **1993**, *245*, 21–42.
(54) Shen, T.; Gnanakaran, S. *Biophys. J.* **2009**, *96*, 3032–3040.

Table 1. Intrachain H-Bonds Investigated in this Study and the Ratio of Each Rotamer State, *gt*, *gg*, and *tg* under the Condition when the Specific H-Bond is Formed[a]

		300 K				450 K			
atoms (A···H−D)	in I_β, I_α ?	population %	*gt*%	*gg*%	*tg*%	population %	*gt*%	*gg*%	*tg*%
O6···H2′−O2′	Y	3.0	0	83	17	4.1	0.1	61.9	38
O1′···H2′−O2′	Y	<0.01				<0.01			
O2′···H6−O6	Y	2.5	0	85	15	3.2	0.3	68.6	31.1
O1′···H6−O6	Y	0.05	0	0	100	0.23	0	2	98
O5′···H3−O3	Y	37.4				30.3			
O6′···H3−O3	N	0.2	86′	14′	0′	1.2	85′	15′	0′
O3···H6′−O6′	N	0.3	83′	17′	0′	1.3	79′	21′	0′
O6···H6′−O6′	N	0.2				0.4			
O6′···H6−O6	N	0.1				0.2			
O3···H2′−O2′	N	0.5				0.9			
O2′···H3−O3	N	0.1				1.0			

[a] The percentages of H-bonding at 300 and 450 K for the middle position (between monomer 3 and 4) of hexamer are shown. Prime labels properties for downstream monomer. Rotamer state is not listed for the cases with poor statistics and when it does not involve the H-bonds directly, i.e., neither O6 nor O6′ is involved.

For the most dominant H-bond, O3−H···O5′, H-bond formation does not directly correlate with the rotamer state since neither O6 nor O6′ participates. As a result, almost two-thirds of the population of the rotamer are seen in the *gg* conformation, regardless of whether this hydrogen bond is formed or not. Other hydrogen bond formations do have a strong correlation with the status of the rotamer, and we therefore examine the correlation between H-bonding and the conformations of the side chain. Table 1 provides the status of side chain torsion population distribution in terms of *gg*, *gt*, and *tg* for all formed intrachain H-bonds at low (300 K) and high (450 K) temperatures when O6 is directly involved in the H-bonds. The results are similar for dimer and tetramer (data not shown). Interestingly, it is still possible to form H-bonds, for O2′−H′···O6 and O6−H···O2′, in the *gg* conformation even though they are geometrically more feasible with the *tg* conformation. Other intrachain H-bonds are not populated to a sufficient degree for a meaningful quantitative analysis. Many of them are not observed because they are only feasible with the least populated side chain conformation, *tg*. For example, the table shows that intrachain H-bond, O6−H···O1′, is formed predominantly with O6 in the *tg* orientation.

D. Chain End-to-End Distance. The end-to-end distance, L, of these oligomers is measured by the distance between the glycosdic and aglycon oxygen atoms at the nonreducing and reducing ends of the molecule. For each system at a given temperature, one can obtain the probability distribution of length, $P(L;n)$ with $n = 2$ for dimer, $n = 4$ for tetramer and $n = 6$ for hexamer. The distribution for the normalized end-to-end distance $l = L/n$, $p(l;n)$, is shown in Figure 7a. Here $p(l;n) = n \times P(L/n;n)$ with n being the number of monomer glucose units. Furthermore, as shown in Figure 7b, the mean value of end-to-end distance, $\langle l(T;n)\rangle$, can also be obtained as a function of temperature. The observed changes in end-to-end distances may come from two effects: (1) changes in the individual monomer glucose unit lengths and (2) changes in the flexibility of the glycosidic linkage between glucose units. For the case of cellulose, the size variance of the glucose unit is relatively small. As the majority of the fluctuation in end-to-end distance comes from the flexibility of the glycosidic linkage between monomers, we focus on the second effect.

E. Rigidity. The elasticity of these cellulose oligomers can be extracted from a combination of simulation results and polymer theory. We will first characterize the spontaneous fluctuation of the oligomers and link these properties to elasticity with known equations from the fluctuation-response theorem.

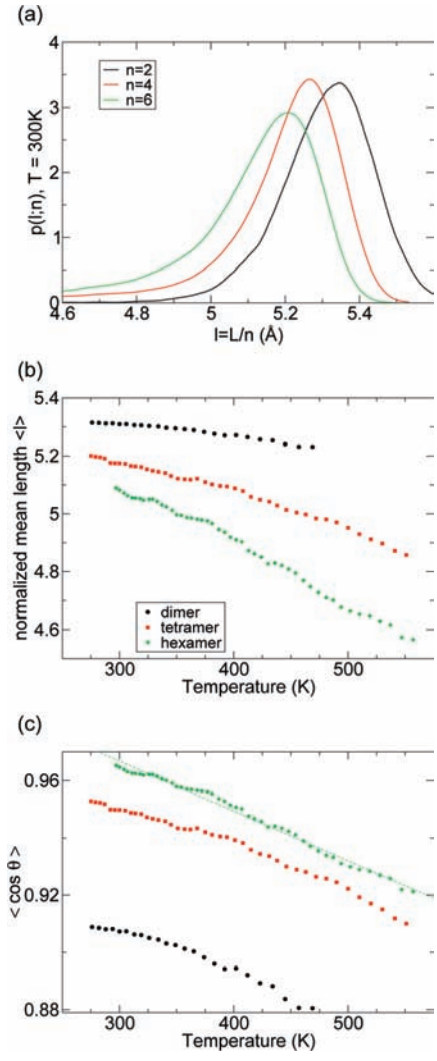

Figure 7. The normalized end-to-end distance distribution at 300 K, $p(l)$, is shown in panel a and the normalized mean distance as a function of temperature is shown in panel b. The effective parameter $\langle \cos\theta \rangle$ of a freely rotating chain model is shown in panel c, which reveals the reversed order compared to that in panel b.

More specifically, in order to compare the backbone flexibility among three types of polymers, a simple model is used to extract the effective parameter $\langle \cos(\theta) \rangle$ of a polymer model of freely

rotating chains.[55] Here θ is the angle change between the neighboring units, which is held constant throughout the chain in this simple model. Assuming the vector of glucose unit i is $\vec{r}_i = \vec{r}_{O1(i+1)} - \vec{r}_{O1(i)}$, the angle between $\vec{r}_i$ and $\vec{r}_{i+1}$ is $\theta = \arccos(\vec{r}_i \cdot \vec{r}_{i+1}/|\vec{r}_i||\vec{r}_{i+1}|)$. This model parameter $\langle\cos(\theta)\rangle$ can be obtained by fitting the mean value of the end-to-end distance, $\langle l\rangle$. For this model, the mean length changes as a function of monomer number n and $\langle\cos(\theta)\rangle$ as $\langle l(n)\rangle = l_o(1 - \langle\cos(\theta)\rangle^{n+1})/(n - n\langle\cos(\theta)\rangle)$. The unit length of the glucose unit, l_o, has been set to 5.55 Å, which is approximately the distance from O1 to O4 for a glucose residue in AMBER/GLYCAM.[56] The mean corresponding distance in small molecule crystal structures is slightly less than this value. Given n and $\langle l\rangle$, we can calculate the corresponding values of $\langle\cos\theta(T;n)\rangle$ that fit the freely rotating chain model as a function of temperature as depicted in Figure 7c. Note that although a simple freely rotating chain only has one parameter $\langle\cos(\theta)\rangle$, here its extension to a function of temperature originates from the temperature dependence of $\langle l\rangle$. Even though the value for $\langle\cos\theta(T;n)\rangle$ obtained by fitting is a simplification to a complicated problem, it provides a direct connection to the mechanical properties via polymer theory. The hexamer is the most rigid of the three oligomers since the case with $n = 6$ yields the largest $\langle\cos\theta\rangle$ at any given temperature. Vice versa, the dimer is the most flexible one.

F. Young's Modulus. Besides directly probing the responses of a polymer under external stretching forces,[57,58] the Young's modulus can also be obtained from the effective spontaneous bending spring constant of the polymer,[59] based on how $\langle\cos(\theta)\rangle$ varies with temperature. Using a simple model and assuming the distribution of the solid angle $p(\Omega)d\Omega \propto \exp(-\beta c\theta^2/2)\sin\theta\, d\theta\, d\phi$, which should be valid for the small θ of semirigid polymers, one obtains $\langle\cos\theta\rangle(T) \approx 1 - 1/(c\beta) = 1 - k_B \times T/c$. Here c is the bending spring constant and $\beta = 1/(k_B T)$.

In Figure 7c, the hexamer has the most linear behavior over the range of temperature studied. The fit results give, for $n = 6$, $\langle\cos(\theta)\rangle(T) = 1.019 - 1.74 \times 10^{-4} \times T$. Thus the spring constant $c = 5.75 \times 10^3 \times k_B$ K $= 7.94 \times 10^{-20}$ J $= 47.8$ kJ/mol. Further, the bending spring constant c is normalized with unit length, l_o, to obtain the flexural rigidity $\kappa = cl_o = 265.3$ kJ/mol · Å. The persistence length ξ can be obtained further as $\xi = \beta\kappa$. At room temperature, $\xi \approx 19.0 \times l_o = 10.6$ nm. Our value of ξ is comparable with experimental data on cellulose chains in aqueous metal complexes,[60] 15.8 nm for Cd-tren, 10.2 nm for Ni-tren, 13.1 nm for cuoxam, and 6−8 nm for xyloglucan chain.[61]

With one final step, we can estimate the bulk material property represented by the Young's modulus, Y. Briefly, the energy cost to bend an elastic rod of length l_o at very small bending angle $\Delta\theta$ can be defined as $\Delta E = c/2 \cdot (\Delta\theta)^2$. On the other hand, assuming the direction of bending is on x-axis, we have $\Delta E = \int dx\, dy \cdot (Y/l_o) \cdot [z(x,y) - z^o(x,y)]^2/2$. This integration is performed on the cross section labeled by the xy plane, and $\Delta z(x,y)$ is the height change due to the stretching or shrinking of that volume element at the microscopic level due to the macroscopic bending. Since in this case $\Delta z(x,y) = x \cdot \Delta\theta$, we have $\Delta E l_o = (\Delta\theta)^2 Y(dx\, dy\, x^2)/2 = (\Delta\theta)^2 \cdot Y \cdot J/2$. Here, $J = \int x^2\, dx\, dy$ is the moment of inertia of the cross section (second moment of area), for example, a cylindrical rod of radius R has $J = \pi R^4/4$. A rough estimation of the radius of the cross-section of the polymer with $R \approx 3$ Å yields $J \approx 60$ Å^4. This radius gives a cross-section area $\simeq 30$ Å^2 to that of bulk cellulose I_β. Finally, by canceling $(\Delta\theta)^2$ from both sides of the equation of ΔE, we have $Y = \kappa/J \simeq 7$ GPa $(= 10^9$ J/m$^3)$, under the assumption of an isotropic elastic media,[59,62] which is about 1 order of magnitude above most filaments, ~1 GPa.[62] Because of the nature of the model we used, both the flexural rigidity and the Young's Modules obtained here are temperature independent.

The dry cellulose Young's modulus is quite high compared to normal biofibers, a value almost half that of steel. With the approximations mentioned above, we obtain a value for the Young's modulus on the order of 10 GPa, which is an order of magnitude smaller than the measured value of dry cellulose (80 GPa).[62] Our lower value for Young's modulus may indicate that in addition to the fundamental properties of a single chain of cellulose, interactions between chains strengthen the cellulose as a bulk material. Such an effect can give rise to a higher Young's modulus of cellulose. Meanwhile, a direct mechanical perturbation of a cellulose chain in vacuum by molecular mechanics with classical force-field calculations have resulted in values for Young's modulus that are closer to the measured values.[57] Many of these discrepancies in Young's modulus may arise from assumptions made, in order to extend the flexibility of individual chains of cellulose to a bulk material property. Another explanation could be that the experiments were performed under dry conditions, which are a better match with the vacuum conditions used in the classical mechanics calculation than the current solvated environment.

Conclusions

In a general noninteracting polymer,[55] an increase in the number of linkages should increase the total entropy, and decrease the overall ordering while maintaining the same conformational flexibility about each (glycosidic) linkage. However, in the case of cellulose oligomers, we find that conformations become more rigid as the chain length is increased. This trend arises due to various factors that contribute to the overall conformational rigidity. One of the main contributions comes from the intrachain H-bond network. The rigidity of cellulose oligomers has been characterized in terms of a simple polymer model of freely rotating chains. The calculated persistence length of 10.6 nm is comparable to the experimentally determined values on cellulose chains. In cellulose crystals, the flexibility of the β-(1,4)-glycosidic bond is restrained by hydrogen bonding between adjacent units. This leads to a significant planar-strand structure. Many of the intrachain H-bonds that stabilize cellulose crystals are not found in individual cellulose chains, due to greater flexibility involving local conformational modes and the absence of the cooperativity between intrachain and interchain H-bonds. However, as the length of these cellulose fragments increases, the tendency to form "crystalline like" intrachain H-bonds increases.

(55) Doi, M.; Edwards, S. *The Theory of Polymer Dynamics*, 2nd ed.; Oxford U. Press: Oxford, 1986.
(56) Nishiyama, Y.; Johnson, G. P.; French, A. D.; Forsyth, V. T.; Langan, P. *Biomacromolecules* **2008**, *9*, 3133–3140.
(57) Eichhorn, S. J.; Young, R. J.; Davies, G. R. *Biomacromolecules* **2005**, *6*, 507–513.
(58) Shen, T.; Hamelberg, D.; McCammon, J. A. *Phys. Rev. E.* **2006**, *73*, 041908(1−6).
(59) Gittes, F.; Mickey, B.; Nettleton, J.; Howard, J. *J. Cell Biol.* **1993**, *120*, 923–934.
(60) Saalwachter, K.; Burchard, W.; Klufers, P.; Kettenbach, G.; Mayer, P.; Klemm, D.; Dugarmaa, S. *Macromolecules* **2000**, *33*, 4094–4107.
(61) Picout, D. R.; Ross-Murphy, S. B.; Errington, N.; Harding, S. E. *Biomacromolecules* **2003**, *4*, 799–807.
(62) Boal, D. *Mechanics of the Cell*; Cambridge University Press, New York, 2002.

The conformational variability of cellulose oligomers is greatly influenced by solvent, which competes for H-bonding and causes the disruption of intrachain H-bonding. With respect to internal conformational preferences, the most preferred β-(1,4)-glycosidic linkage population basin for solvated oligomers (basin O) is the same as that found in the cellulose crystals. This finding is consistent with a series of structural data on cellulose oligomers, which concludes that oligomers have a set of backbone torsion angles that is similar to those in crystalline cellulose. This basin is proximal to the 2-fold screw axis. QM potential energy surface calculations of cellobiose in vacuum capture the same locations of the three basins as seen in our free energy profile of solvated cellulose oligomers. At lower temperature, monomer ring conformations stay in chairlike conformation as expected. At higher temperatures, deviations are seen. One of the important findings is the correlation between ring flipping and the 1,4-glycosidic bond. Most of the ring flips tend to occur in the 1,4-glycosidic basin A.

Several other interesting properties are also identified that distinguish cellulose oligomers from the commonly known conformation and hydrogen bonding patterns observed in cellulose crystalline forms. The side chain torsion angle ω influences the interplay between inter- and intrachain hydrogen bonding in solvated cellulose oligomers since the *tg*, *gg*, and *gt* conformations allow alternative hydrogen bonding possibilities. We find that the preference for ω torsion angles in soluble cellulose chains is very different from that in cellulose crystals. Our studies show that the preference is toward *gg* compared to *tg* in cellulose I_β and I_α crystals or *gt* in cellulose II and III. Our analysis of the correlation between side chain torsion conformations and intrachain H-bonding indicates that many of the H-bonds seen in the cellulose I crystal are feasible only with *tg*, which is the least populated conformation for oligosaccharides in water. In cellulose I_β and I_α crystals, the side chain conformation is *tg* and thereby allows formation of many of the intrachain H-bonds listed in Table 1. In solvent, the preference shifts mainly to the *gg* conformation and, consequently, reduces the frequency of certain intrachain H-bonds.

This study can also be considered as a systematic and rigorous evaluation of the chosen carbohydrate force field. The reliability of molecular dynamics simulations depends on the accuracy of the force field including proper treatment of solvent and efficient sampling of conformational space. The sampling concerns have been accommodated by the REMD approach. That leaves some of the interesting observations from this study to the nature of the force field chosen. For example, we have been able to capture interesting position-dependent ring flipping. While puckering flips do occur more frequently at higher temperature, the tendency of ring flipping is much lower at the nonreducing-end-ring compared to all other positions in the oligomer chain. Since both ends are not identical it is possible that they exhibit different ring-flipping behavior. No experimental measurements are currently available to verify such an effect. Thus, it is not yet clear it if this due to a deficiency in the force field or a real effect revealed by an accurate force field and an efficient sampling. Furthermore, most carbohydrate force fields are parametrized at room temperature. For many studies that should be adequate to provide reasonable results. However, it is imperative to examine whether or not the force fields are adequately representing the higher temperature case as done here.

This work represents, to our knowledge, the first study using REMD on oligosaccharides in explicit solution. Currently, the application of REMD to peptides and small proteins has become routine. Several extensions of REMD are also being applied to other biomolecular systems. However, one of the relatively untested territories for REMD is carbohydrates. In REMD, high temperatures are often used as a way to get over barriers and sample extensive conformational space. Peptides are somewhat robust and temperature can be used as a measure to enhance sampling. However, carbohydrates contain relatively soft ring modes that can easily be perturbed by temperature. At lower temperature the most stable ring conformation is chairlike. At higher temperatures rings can flip into one of the many boatlike conformations. With an accurate force field and the proper implementation of REMD, equilibrated ensembles at different temperatures should exhibit the proper populations. We have shown that this is indeed the case here. While ring conformations can deviate from the chair conformation at high temperature, with the exchanges between replicas, these excited ring conformations flip back into the chair conformation in a self-consistent manner as the temperature is decreased.

Finally, we have shown that cellulose oligomers have different properties from chains in the crystalline forms of cellulose. The results presented here may be relevant for understanding the properties of cellulose fragments during degradation by cellulases. During this enzymatic process, cellulose chains are frayed and extracted from the crystalline environments. They can potentially form alternate conformations with cellulase or when transferred from the crystalline phase to the aqueous phase.

Acknowledgment. We thank Dr. R. J. Woods for very helpful discussions on the GLYCAM force-field. We would also like to thank Dr. M. Mustyakimov, Dr. D. Fox, and other members of the cellulosic biofuels team at Los Alamos National Lab for their support. This work was supported in part by an LANL-LDRD grant from the United States Department of Energy and Center for Nonlinear Studies. Funding for A.D.F. and G.P.J. was from the U.S. Department of Agriculture, Agricultural Research Service project 6435-44000-070-00D.

JA9034158