Carbohydrate Research 345 (2010) 2060–2066

Contents lists available at ScienceDirect

Carbohydrate Research

journal homepage: www.elsevier.com/locate/carres

Simulation studies of the insolubility of cellulose

Malin Bergenstråhle [a], Jakob Wohlert [a,†], Michael E. Himmel [b], John W. Brady [a,*]

[a] Department of Food Science, Cornell University, Ithaca, NY 14853, United States
[b] National Renewable Energy Laboratory, 1617 Cole Boulevard, Golden, CO 80401-3393, United States

ARTICLE INFO

Article history:
Received 4 February 2010
Received in revised form 5 June 2010
Accepted 25 June 2010
Available online 6 July 2010

Keywords:
Cellulase
Cellobiohydrolase I
Cellulose
Computer modeling
Molecular dynamics

ABSTRACT

Molecular dynamics simulations have been used to calculate the potentials of mean force for separating short cellooligomers in aqueous solution as a means of estimating the contributions of hydrophobic stacking and hydrogen bonding to the insolubility of crystalline cellulose. A series of four potential of mean force (pmf) calculations for glucose, cellobiose, cellotriose, and cellotetraose in aqueous solution were performed for situations in which the molecules were initially placed with their hydrophobic faces stacked against one another, and another for the cases where the molecules were initially placed adjacent to one another in a co-planar, hydrogen-bonded arrangement, as they would be in cellulose Iβ. From these calculations, it was found that hydrophobic association does indeed favor a crystal-like structure over solution, as might be expected. Somewhat more surprisingly, hydrogen bonding also favored the crystal packing, possibly in part because of the high entropic cost for hydrating glucose hydroxyl groups, which significantly restricts the configurational freedom of the hydrogen-bonded waters. The crystal was also favored by the observation that there was no increase in chain configurational entropy upon dissolution, because the free chain adopts only one conformation, as previously observed, but against intuitive expectations, apparently due to the persistence of the intramolecular O3–O5 hydrogen bond.

1. Introduction

Cellulose, the β-(1→4)-linked polymer of D-glucose that is the primary structural component of plant cell walls, has been one of the most studied of all biopolymer molecules. In the plant cell wall, cellulose acts as the load-bearing component, and co-exists with other cell wall polymers, such as hemicelluloses and lignin, as well as with water.[1] As the principal cell wall component, it is the single most abundant biological molecule in the biosphere, and thus represents the most important feedstock for the industrial production of liquid fuels using biomass conversion technologies, a topic of considerable current interest.[2] Its biosynthesis takes place on the surfaces of the plasma membranes of plant cell walls, where newly formed cellulose chains are organized (coalesced) into elementary fibrils with lateral dimensions of 3–5 nm.[3] The nature of the aggregated structure of these microfibrils is complex and still an open question. Cellulose may aggregate in at least six different crystal polymorphs, and in addition exists in less-ordered structures, sometimes referred to as amorphous or paracrystalline cellulose.[1,4] Native crystalline cellulose is found in a metastable form called cellulose I, with two polymorphs, cellulose Iβ and Iα, differing slightly in their unit cells.[5] These crystalline polymorphs are found to co-exist in nature with source-dependent ratios.[6,7] When native cellulose is processed through regeneration or mercerization, it will irreversibly restructure to form so-called cellulose II. The unit cell of cellulose II differs from that of cellulose I in that neighboring chains are probably oriented in an antiparallel fashion, instead of parallel as they are in the native crystals.[1,8,9]

The most significant limitation for the use of cellulosic material for fuel production is its insolubility in water, which hinders the action of cellulase enzymes. Cellulose remains completely insoluble below temperatures of about 300 °C, where it rapidly decomposes.[10] This insolubility is a strong function of the chain length; as the degree of polymerization (DP) of cellooligomers increases, their solubility rapidly drops to zero beyond celloheptaose.[11] Cellulose insolubility makes evolutionary sense in terms of its biological role. As the structural framework of plant cell walls, it would be highly disadvantageous if cellulose dissolved on contact with water. However, while the biological advantages arising from insolubility are clear, the physical reasons for this insolubility of cellulose are not obvious. The fact of cellulose insolubility has been so well-known for so long that rarely are the features of this polymer that contribute to its insolubility examined.

When attempting to understand the solubility of a molecule, it is necessary to compare the many contributions to the free energy of both the solid and solution states. For example, cellulose in solution would be expected to have much larger translational and rotational entropy than when confined in the crystal, and the solution of course is also favored by the entropy of mixing. These

* Corresponding author. Tel.: +1 (607) 255 2897.
E-mail address: jwb7@cornell.edu (J.W. Brady).
† Present address: Wallenberg Wood Science Center, Teknikringen 56-58, Royal Institute of Technology, SE-10044 Stockholm, Sweden.

0008-6215/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2010.06.017

contributions tend to favor solvation over a highly ordered crystalline state. In addition, in principle, the solution should also be favored by an increase in the configurational entropy for the polymer, because the chain should have more conformational possibilities in solution than when packed into a regular crystal lattice with a single conformation. On the other hand, the 'tops' and 'bottoms' of the cellulose chains are hydrophobic, since all of the polar hydroxyl groups in this polymer are in equatorial positions, while all of the aliphatic protons on the ring are in axial positions, as is true for the monomer β-D-glucopyranose (see Fig. 1). Thus, when the chains are packed into the reported Iβ or Iα crystal structures,[5,12,13] their non-polar surfaces can be organized into hydrophobic sheets paired against one another, rather than structuring large amounts of water in solution. (Note that in this type of interaction, the driving force for association is not simple van der Waals interactions, but rather hydrophobic association driven by the liberation of structured water molecules.[14]) Furthermore, in the proposed crystal structures for cellulose, the chains make several hydrogen bonds, both within the chain and between adjacent chains in the same crystal layer (although not between layers). However, these hydrogen bonds amount to little more than one per hydroxyl group on average, whereas in aqueous solution each hydroxyl group could make approximately 2–2.5 hydrogen bonds of equal or greater energy to water molecules,[15] which would superficially suggest that hydrogen bonding should strongly favor solvation over the crystal state.

Taken all together, this notional bookkeeping (without as yet assigning actual numbers to the various contributions) would seem to favor solubility over crystallinity. Because the opposite case is actually observed, it would be useful to be able to get an order of magnitude estimate of some of the contributions to the solubility/insolubility of cellulose and cellooligomers. In particular, it would be helpful to determine if the association of cellooligomers is favored by hydrophobic forces, and to be able to assign a magnitude to this effect. It would also be useful to determine whether or not the hydrogen bonding pattern reported for cellulose crystals might be expected to contribute significantly to the stability of those crystals relative to the solvated state. Unfortunately, however, the quantitative assessment of the actual lattice energies for the various cellulose crystals would be quite sensitive to the details of the models, as well as to the reported crystal packings.

Here we have attempted to calculate highly approximate estimates of the association-free energies by computing the potential of mean force (pmf), or work function, for the separation of pairs of cellooligomers of different sizes in aqueous solution along two different reaction coordinates. The particular pathways chosen are illustrated in Figure 2, for the example of cellobiose. Such pmfs were computed for a series of four molecules: glucose, cellobiose, cellotriose, and cellotetraose, using molecular dynamics (MD) computer simulations. The goal is to obtain a qualitative picture of the interaction of cellulose chains in specific geometries, rather than to deconstruct the actual lattice energy of cellulose Iβ, or even to comprehensively model the interactions of cellooligomer chains in solution. While relatively modest, these goals are tractable with present technology, and the results provide a framework for further progress in understanding cellulose interactions in an aqueous environment.

2. Methods

Potentials of mean force for the separation along specific pathways of pairs of the molecules β-D-glucopyranose, β-cellobiose, β-cellotriose, and β-cellotetraose in aqueous solution were computed using constrained MD simulations. In one set of simulations, referred to here as A1–A4, the molecules were stacked with their hydrophobic faces directly on top of one another, and were then moved apart along a pathway perpendicular to the mean plane of the twofold screw ribbon (or the mean plane of the single ring in the case of glucose), as illustrated in Figure 2A. The pmf calculated along this pathway estimates the free energy from hydrophobic association. Note, however, that this stacked arrangement does not exactly reproduce the relative arrangements of two chains in different layers of cellulose I, where chains in different layers are offset from one another. For the second set of calculations, called B1–B4, the pairs of molecules were started in relative positions that mimicked the relationship that two adjacent chains would have in the same plane of crystalline cellulose Iβ,[12] and were then moved apart in that plane to estimate the contribution to the

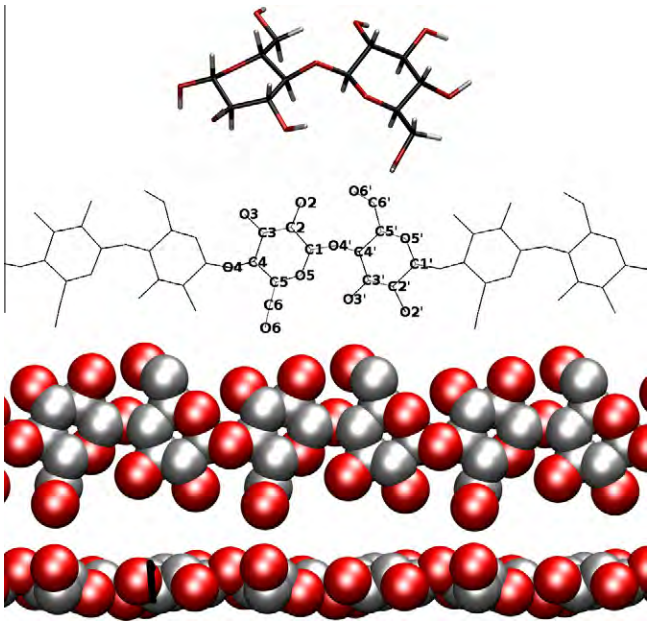

Figure 1. Top: an all-atom licorice representation of a cellobiose molecule. Note that the aliphatic protons are all in axial positions pointing perpendicular to the ring. Middle: a 'stick' figure indicating the atom names used in this paper. Bottom: a van der Waals surface representation of the cellulose chain, seen from 'above' and from the side, with oxygen atoms colored red and the non-polar carbon atoms shaded gray. Hydrogen atoms are not shown in this representation for clarity.

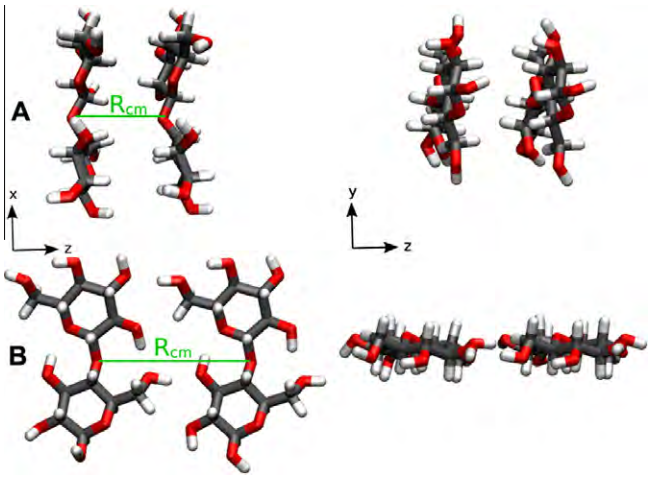

Figure 2. Orientation of sugars in simulations A1–A4 (top) and B1–B4 (bottom) as viewed from two different angles, indicated by coordinate axes. The reaction coordinates used for pmf calculations are also indicated. Several constraints apart from the umbrella potential were used to make the sugars retain their relative orientation (see text).

association energy arising from hydrogen bonding. One additional pmf calculation, called F1, was conducted for the separation of two glucose molecules without constraints on their relative orientations as an internal control. Finally, a 10 ns standard MD simulation of single free cellotetraose molecule in solution was carried out as a control to characterize the preferred conformations that this oligomer adopts in aqueous solution.

The pmf simulations were performed using the program GROMACS 4,[16] employing the CSFF force field for the carbohydrates[17,18] and the CHARMM version of the TIP3P model for water molecules,[19,20] with a time step of 2 fs. The temperature was maintained at 300 K using velocity rescaling,[21] and the pressure was controlled with a Parrinello–Rahman barostat[22] with a reference pressure 1 atm. All bonds involving hydrogen atoms were constrained to their equilibrium values using P-LINCS,[23] and water molecules were kept completely rigid using SETTLE.[24] Particle-mesh Ewald summation[25] was used for electrostatic interactions, and van der Waals interactions were made to go smoothly to zero between 10 and 12 Å, using a switching function.

Starting coordinates were generated using CHARMM,[26,27] but in different ways for the A1–A4 simulations and the B1–B4 simulations. For the A1–A4 simulations, two identical sets of coordinates for the sugar molecule pair were generated, and their long axes aligned with the x axis of the simulation box. Finally one of them was translated 5 Å along the y axis. For the B1–B4 simulations, experimental coordinates for the cellulose Iβ crystal structure[12] were used, and the crystallographic **b** axis was aligned with the z axis of the simulation box. Each structure was placed in a fully periodic box of dimensions $30 \times 30 \times 30$ Å ($35 \times 30 \times 30$ Å in the case of cellotetraose), and solvated using an equilibrated box of TIP3P water. This lead to total system sizes between 2436 and 3309 atoms.

The pmfs were calculated using umbrella sampling, with center-of-mass separation of the pairs of molecules as the reaction coordinate (z), using an harmonic umbrella potential, $U(z) = k_2(z - z_0)^2$. The force constant k_z was 5 kcal mol^{-1} Å^{-2}, with z_0 ranging from 5 to 14 Å for simulations A1–A4, and from 8 to 14 Å in the simulations B1–B4, in 0.5 Å increments. The simulation time for each umbrella window was 5 ns, and the simulations were unbiased using WHAM[28] to get the pmf as function of z. Since the (one-dimensional) reaction coordinate in F1 was applied in three-dimensional space, as opposed to A1–A4 and B1–B4, which were one-dimensional due to the constraints, a correction of $2k_BT\cdot\ln(z)$, originating from the Jakobian of the coordinate transformation,[29] was added to the pmf.

Obviously, the chosen reaction paths are highly artificial, in a sense that as the real molecules separate, they would likely rotate relative to one another about each of their three principal axes, displace 'sideways' in the case of the A series, and out of their common plane in the case of the B series, and possibly undergo conformational changes away from a planar ribbon shape for the longer oligomers. For this reason, several more restraining potentials had to be used. To maintain the relatively flat conformation of the crystal structure, the two dihedral angles ϕ and ψ over the glycosidic linkage (O5–C1–O4′–C4′ and C1–O4′–C4′–C5′, respectively) were constrained, where applicable, to −98.5° and −142.3°, respectively, using harmonic potentials with a force constant of 25 kcal mol^{-1} rad^{-1}. Further, all heavy ring atoms (C1–C5 and O5) were constrained in their x and y coordinates, using harmonic restraining potentials with force constant 2 kcal mol^{-1} Å^{-2}, thus allowing the molecules to move freely only in the direction parallel to the reaction coordinate. To account for the contribution of these additional constraints to the pmf, the total constraint potential energy was recorded separately. This energy was subsequently mapped to the reaction coordinate and finally added to the pmf.

3. Results and discussion

Figure 3 presents the results of the pmf calculations for the cases of hydrophobic stacking for the series of oligomers spanning DP 1–4 (simulations A1–A4). For all four molecules, the solvation free energy favors hydrophobic face-to-face association of the oligomers. In the case of glucose, the magnitude of this association energy is 1.0 kcal/mol, whereas for the disaccharide cellobiose, this energy is slightly greater, 3.3 kcal/mol. The association energies for cellotriose and cellotetraose are much larger, nearly 5.5 and 7.5 kcal/mol, respectively. All four curves display similar qualitative features, with a global minimum between 4.6 and 5.0 Å, a barrier around 6.5 Å, followed by a secondary minimum centered between 7 and 8 Å. The heights of the primary barriers relative to the energy of the first minimum are 1.4, 3.7, 5.5, and 7.5 kcal/mol, respectively, but only 0.4, 0.4, 0.1, and 0.0 kcal/mol relative to infinite separation. For the monomer and dimer, there is no significant structure beyond the secondary minimum, but for the larger oligomers there is a low second barrier around 9.5 Å and a very weak, broad third minimum centered around 11 Å.

The barriers between the primary and secondary minima in these curves may be in part artifacts of the artificial reaction coordinates chosen. In general, these barriers arise from the energy needed to pull a vacuum between the two chains as they are separated, or to squeeze out the last solvation layer between them as they are brought together, and is a general feature of pmfs for the approach of surfaces in a liquid solvent.[30] In a real solution of these oligomers, however, the chains would rotate relative to one another, and possibly change conformation as well, to reduce these effects. In this case, little or no such barrier would exist, as can be seen from the pmf for two glucose molecules unrestricted in their relative orientations, calculated as a reference and internal control (simulation F1, Fig. 3). Glucose does not exhibit any measurable tendency to associate in aqueous solution at dilute concentrations up to 5 molal, in either experiments or simulations,[31] and the unrestricted pmf for two glucose molecules indeed exhibits no tendency for aggregation. The extremely shallow well, if statistically significant, is significantly less than k_BT. Similar considerations may have affected the pmfs calculated for the larger oligomers, but unfortunately, completely unconstrained calculations are not computationally feasible for these larger oligomers with presently available resources.

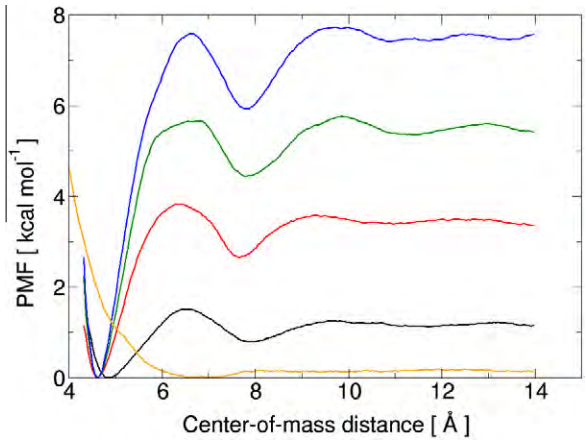

Figure 3. The calculated pmfs from simulations A1–A4, together with the calculated pmf for two glucose molecules whose relative orientations were unconstrained (simulation F1, shown in orange). The curve for glucose is shown in black; that for cellobiose is red; for cellotriose, green; and for cellotetraose, blue. Note that unlike the glucose in A1, there is no contact minimum, and no net attraction between the two glucoses in F1 at any distance.

Figure 4 displays the pmf functions calculated for the series of simulations for oligomers of DP 1–4 (simulations B1–B4) in which the molecules are initially placed alongside one another in the same plane, with the relative positions and intermolecular hydrogen bonds reported for the cellulose Iβ crystal.[12] As can be seen, these pmfs are qualitatively different from those found for hydrophobic stacking. For glucose in this orientation, there is very little net binding affinity, with the limiting value being only about 0.2 kcal/mol above the very shallow minimum value centered between 8.5 and 9.0 Å. This result is not surprising, because for a pair of glucose molecules in this orientation, there is only one cellulose-like hydrogen bond formed between the rings, and this is easily replaced by hydrogen bonds to water.

The other pmf curves exhibit more structure than the one for glucose. All exhibit a primary minimum centered between 8.5 and 9.0 Å. All three also exhibit a substantial broad barrier to final bonding as they approach, with the barrier magnitude becoming higher as the DP increases. There is also a significant secondary minimum around 11.0–11.5 Å. In all three cases, the net association energy along this separation axis in water is much smaller in magnitude than for the hydrophobic pairing. This result is not particularly surprising since the hydrogen bonds between the molecules can be easily replaced by hydrogen bonds to water, as in the case for glucose, as the molecules separate. Given this ease of replacement, it is in fact somewhat surprising that there is any net affinity between the two chains at all resulting from hydrogen bonding. For the cellotetraose chains, this net association energy is a little over 1.0 kcal/mol. This is interesting, considering that one might expect a value not much different from zero. This net affinity is only slightly less for the cellotriose and cellobiose pairs. Apparently the cooperative and ordered nature of the hydrogen-bonding pattern in the higher oligomers stabilizes this configuration over the more distorted and fluctuating hydrogen bonds to water.

The major process contributing to the structure of the curves in Figure 4 is the breaking and formation of the hydrogen bonds. At small separations, around 8–8.5 Å, no water is present between the two sugars, and they form the predominant hydrogen-bonding network present in the cellulose Iβ crystal structure[12,32] for cellobiose and the larger oligomers. It is characterized by three cooperative hydrogen bonds. There are intramolecular hydrogen bonds between HO2 and O6 and between HO3 and O5, and an intermolecular hydrogen bond between the two chains connecting HO6 and O3 (Fig. 5). This pattern implies a difference between the two oligomers in the number of intramolecular hydrogen bonds when the number of rings is even (i.e., in B2 and B4).

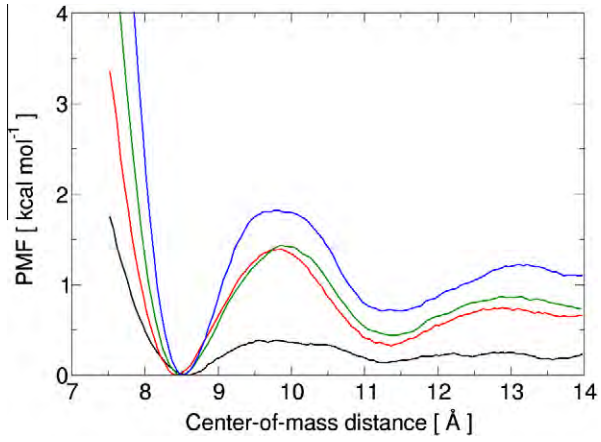

Figure 4. The calculated pmfs from simulations B1–B4. The color scheme is the same as in Figure 3.

The origin of the barriers exhibited in the curves for DP 2–4 in Figure 4 is the breaking of these intermolecular hydrogen bonds as the chains separate. Figure 6 plots the number of inter- and intramolecular hydrogen bonds made between the two chains for each of the four sets of simulations. Hydrogen bonds were defined using geometric criteria, with a hydrogen-acceptor cutoff distance of 2.4 Å. The cellulose Iβ hydrogen bond pattern remained stable throughout the simulations at separations smaller than 9 Å. However, as can be seen, when the separation distance exceeded a threshold of approximately 9–10 Å, all of the intermolecular hydrogen bonds broke. That the total cellulose crystal hydrogen bonding pattern is cooperative can be seen from the fact that the breaking of the intermolecular hydrogen bonds was accompanied by a significant decrease in the number of intramolecular hydrogen bonds, predominantly between HO2 and O6. This decrease occurred even though in principle such intramolecular hydrogen bonds were still as geometrically possible as before. There was also a small decrease of the HO3–O5 hydrogen bond population, which would be expected to be weaker than the others, because the partial charge assigned to the O5 ether oxygen atom of the pyranose rings is substantially lower than that for hydroxyl oxygen atoms. This bond is, however, remarkably stable and is almost always present even for free molecules in solution. It has been argued that in real cellulose this cooperative effect arises from a combination

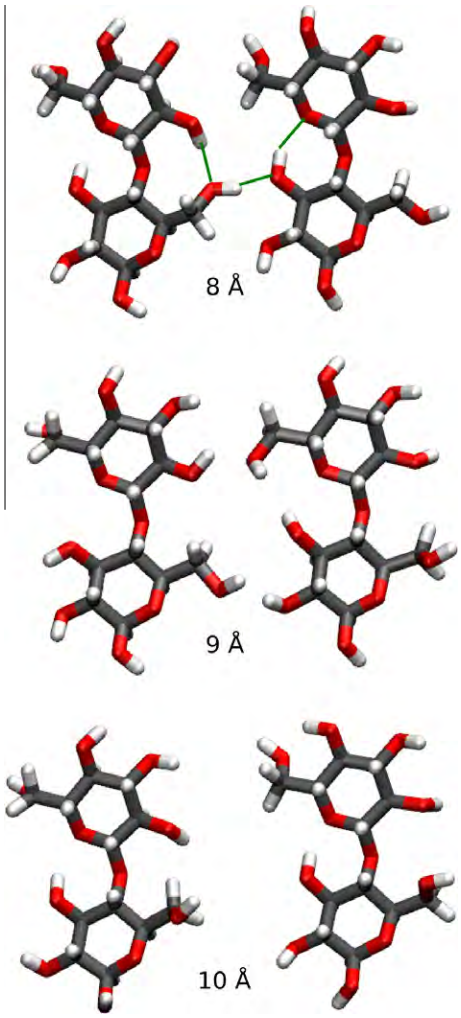

Figure 5. The rupture of the cooperative hydrogen-bonding network in cellulose Iβ, shown for the cellobiose example from the B2 simulations as the sugars separate in water.

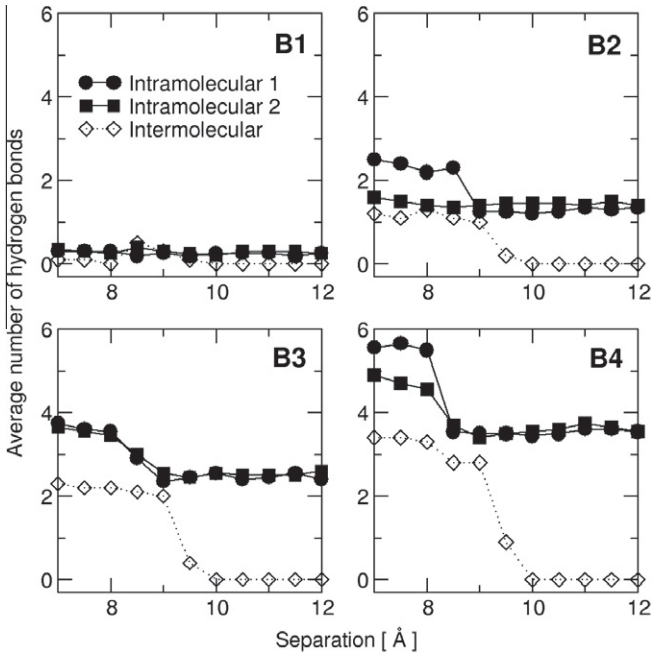

Figure 6. Data from simulations B1–B4, showing the average number of intra- and intermolecular hydrogen bonds formed within and between the sugars, glucose (B1), cellobiose (B2), cellotriose (B3), and cellotetraose (B4) as a function of separation. Intramolecular 1 and 2 refer to intramolecular hydrogen bonds within the two sugars, respectively.

of charge transfers and dipole–dipole interactions. Recent ab initio calculations have quantified the quantum mechanical contribution to cooperativity in this pattern,[33] but of course, in the present MD simulations, with fixed potential energy functions, any cooperativity arises from classical effects.

As the chains separated and their intermolecular hydrogen bonds were broken, they were simultaneously replaced by hydrogen bonds to water molecules as shown in Figure 7 for the cellotriose example (B3). As can be seen from this figure, the number of hydrogen bonds to water for the O2, O3, and O6 groups increased between 8.0 and 10.0 Å for the central glucose of the trimer chain which had these functional groups facing in, leveling off at the same value as for those facing out into the solution (which did not change upon separation). Even in the crystal-like configuration at very close distances, the O2 and O6 groups made a significant number of hydrogen bonds to surrounding water (an average of about 0.75 such bonds), while the O3 group made almost 0.5 such hydrogen bonds to water on average. This situation for these groups is very different from that which would exist in the actual crystal, where no waters are generally present, and where there are no alternate hydrogen bond partners available. The small difference between the number of hydrogen bonds to water made by the O2 and O3 hydroxyl groups can be understood in terms of their relative solvent accessibilities, since in the crystal-like geometry, the O3 group has the lowest solvent accessibility due to the topology of its intramolecular hydrogen bond.

The shallow secondary minimum in the pmf curves for cellotriose and cellotetraose results from a situation in which water molecules are interposed between the two chains, making hydrogen bonds to both. Such bridging hydrogen bonds have been observed in MD simulations of carbohydrates before,[31,34,35] and would be expected to produce broad and shallow minima of the type seen here.

As a final probe of the solvation behavior of these cellooligomers, a simple MD simulation of a single cellotetraose molecule in water was conducted. This molecule was also started from the planar twofold ribbon conformation of the cellulose crystal, but during the equilibration phase quickly adopted a slightly twisted conformation (Fig. 8) such as has been seen in other MD simulations of cellooligomers using different force fields.[36] This twist resulted when the ϕ angles went from $-98.5°$ to approximately $-75°$, and ψ went from $-142.3°$ to approximately $-124°$. After this initial conformational change, however, the chain was quite rigid, as was also observed in previous simulations of cellooligomers.[36,37] Figure 8 also displays the histories of the ϕ angles for all three glycosidic linkages over the stable 10.0 ns period of the trajectory. As can be seen, with the exception of brief fluctuations, this conformation remained stable throughout the simulation. This somewhat surprising result suggests that there is no significant increase in chain configurational entropy upon solvation, which would help contribute to the explanation of the insolubility of cellooligomers.

In the simulation of the isolated cellotetraose molecule, the cooperative hydrogen-bonding network is of course disrupted by the absence of the adjacent chains, but the intramolecular HO3–O5 hydrogen bond was nonetheless quite persistent. Using a

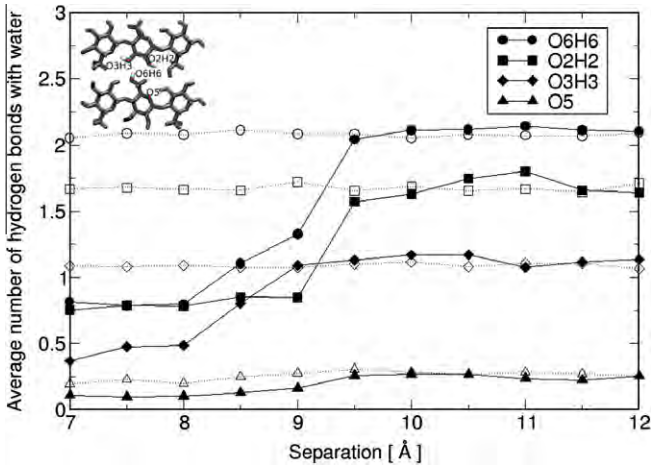

Figure 7. Data from the B3 simulation, showing the average number of hydrogen bonds between water and the three different hydroxyl groups and the ring oxygen atom, pointing inwards from the two middle glucose units in the cellotriose molecules as a function of the reaction coordinate (center-of-mass separation). Dotted lines show the respective reference values for the different groups when free in solution.

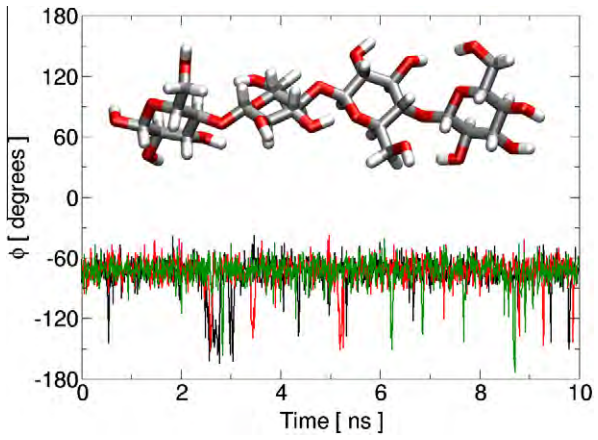

Figure 8. A snapshot view of the final conformation from an MD simulation of a single unconstrained cellotetraose molecule in aqueous solution. Also shown are the time series for each of the three glycosidic angles ϕ from the simulation of the single unconstrained cellotetraose molecule in aqueous solution.

distance cutoff criterion of 3.5 Å between heavy atoms, and an angular cutoff of 30° away from linear, the cellotetraose had an average of 1.4 intramolecular hydrogen bonds, and approximately 26 hydrogen bonds to the solvent water. The maximum possible number of intramolecular hydrogen bonds would be 3.0, which indicates the degree to which the HO3–O5 hydrogen bonds are stretched by the conformational fluctuations. These hydrogen bonds should be weaker than hydrogen bonds to solvent due to the much lower partial atomic charge on the ether oxygen atom, but nevertheless these interactions may play an important role in stabilizing the conformational fluctuations of the chain and keeping the chain in a single conformation.

In the reported crystal structures for cellulose, the exocyclic primary alcohol groups, C6–HO6, all have the so-called *tg* conformation about the C5–C6 bond,[38] which allows the formation of the cooperative hydrogen bond network of the crystal. NMR studies of the conformation for this group of glucose in aqueous solution have found an approximate 60:40 equilibrium of the *gg* and *gt* forms, with little or no measurable *tg* population.[39] Accordingly, the carbohydrate force field used in the present simulations was parameterized to reproduce this conformational distribution in glucose, with the *gg* form having the lowest free energy in solution.[17,18] The occurrence of the *tg* conformation in cellulose is presumably due to the packing requirements of the crystal, and in particular to the formation of the set of cooperative hydrogen bonds along and between adjacent chains. Most primary alcohol groups in the present simulations that were not involved in intermolecular hydrogen bonding rotated to either the *gg* or *gt* conformations, as would be expected for this potential energy function. In fact, the distribution of these conformers averaged over the simulations approximated the 60:40 distribution of glucose, since there was no crystal lattice to enforce the *tg* conformation. However, in recent simulations of crystalline cellulose with this force field,[40] fewer than half of the primary alcohol groups remained in the *tg* conformation, with alternating planes of the crystal lattice adopting primarily the *gg* conformation, making hydrogen bonds between layers. It is not clear whether or not this reorganization is an artifact of the force field employed, because simulations using the GROMOS force field found similar primary alcohol conformational changes.[41] Such considerations would not significantly affect the energies reported here, but would contribute to the overall stabilization energy of the crystal form.

4. Conclusions

Cellulose insolubility is so well-known that it is taken for granted, and the factors contributing to its insolubility are therefore not generally examined in detail. Naturally, understanding the solid/solution equilibrium requires a quantitative understanding of the lattice energies of the various proposed crystal forms, as well as of the amorphous solid phase, and a characterization of the cellulose chains in solution. From a consideration of the likely properties of the hydrated polysaccharide chain, it might naively be expected that the solution would be favored by both hydrogen bonding and chain configurational entropy, making it somewhat less clear that it should be so insoluble.

However, all of the energetic contributions examined in the present study either favor the crystal form of cellulose or at least do not favor the solution form. Both hydrophobic association and hydrogen bonding favor the crystal, and the chain configurational entropy apparently favors neither. It thus seems less surprising that cellulose is so insoluble. These favorable energetic contributions are partially a consequence of the planar topology resulting from the twofold screw conformation about the β-(1→4) linkage, and the hydrophobic character of the tops and bottoms of the flat ribbons.

The free energy simulations reported here find a significant hydrophobic pairing energy favoring the stacking association of cellooligomer chains in a manner similar to that found in the various proposed crystal structures for cellulose. Such an effect is not surprising, given that previous studies have shown that the hydrophobic 'tops' and 'bottoms' of the glucose molecule impose significant structuring on the adjacent water molecules when they are in solution.[31,42,43] The magnitude of this pairing energy for the longer oligomers from these pmf simulations can be estimated as approximately 2.0 kcal/mol/residue.

It may seem somewhat more surprising that hydrogen bonding in the aqueous environment also slightly favors the pairing of the cellooligomers in the manner found in the cellulose crystal structures. In part, this effect may be due to the cooperative nature of the hydrogen-bonding pattern in the cellulose crystal, since disruption of the intermolecular component of this hydrogen bond network also led to the partial disruption of the intramolecular hydrogen bonds as well. The entropy of the water molecules hydrogen bonding to the carbohydrate chains in solution would also tend to favor crystal pairing. Previous simulations have shown that the severe constraints placed upon the possible positions of water molecules that are hydrogen bonded to the adjacent hydroxyl groups of sugars localize and structure these solvent molecules even more than do hydrophobic surfaces of the tops and bottoms of the sugar rings.[35,42,43] Thus, when two sugars pair up by hydrogen bonding, the water molecules that are released as a consequence experience a large gain in entropy, just as in hydrophobic pairing, favoring the aggregation.

From the simulation of the unconstrained cellotetraose molecule in solution, it appears that chain configurational entropy also does not significantly favor the solution. The assumption that cellulose chains are essentially flexible coils in solution appears to be incorrect, since both this present MD simulation, as well as others,[36,37] have suggested that the chain actually exists primarily in a single extended, twisted conformation even in solution. This rigidity reduces the supposed configurational entropy that would favor solvation over crystal packing if the chains were writhing random coils, although the solvated chains would still have greater rotational and translational freedom than in the crystal.

The present studies do not permit quantitative estimates of the total free energy difference between the solvated and crystalline states. This, however, was not the objective of the present study, which was designed to attempt to deconvolute some of the components of this energy in a qualitative manner. Consequently, these calculations used artificial, highly restricted reaction coordinates selected to illustrate individual components to the total energy. Calculating actual solution/solid energy differences would also require computing the crystal lattice energies, which has proved unexpectedly difficult, in part because the reported crystal structures are not stable with any of the presently-available force fields (CSFF-CHARMM,[17,18] CHARMM36,[44] GROMOS45a4,[45] or GLYCAM06).[40,46]

The arguments and calculations presented here are all predicated upon the assumption that the alternative state to aqueous solvation is the reported native crystal structures. However, natural cellulose fibers can also be substantially amorphous in character. The actual nature of the less-ordered regions of microfibrils of cellulose has not been definitively determined beyond the observation that they are not sufficiently regular to diffract X-rays or neutrons. It is likely, however, that even in the amorphous regions of cellulose microfibrils, the hydrophobic surfaces of the chains are paired together in such a fashion as to exclude water, and that many intermolecular hydrogen bonds exist, so that many of the arguments presented here would also apply to cellulose in the native plant cell wall.

Of course, the present pmf results will depend in their quantitative details upon the force field and water model used. The results may also have been affected by the limited duration of the simulations, particularly for the tetraose case, which may have prevented the sampling of any possible alternate conformations. Much more importantly, they will be affected by the artificiality of the reaction paths studied. They do provide, however, a framework for discussing the various contributions, and these qualitative results are less likely to be significantly affected by the specific parameters used here. Thus, the present results suggest general explanations for why cellulose is so insoluble in water, an observation that, in spite of the conventional wisdom, is not entirely intuitive.

Acknowledgments

The authors thank P.E. Mason, D.B. Wilson, and P.I. Hansen for helpful discussions. This work was supported by the DOE Office of Science, Office of Biological and Environmental Research through the BioEnergy Science Center (BESC), a DOE Bioenergy Research Center. M.B. and J.W. also thank the Sweden–America Foundation for financial support.

References

1. O'Sullivan, A. C. *Cellulose* **1997**, *4*, 173–207.
2. 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.
3. Bergenstråhle, M. *Polymer Technology*; KTH Chemical Science and Engineering: Stockholm, Sweden, 2008.
4. Samir, M. A. S. A.; Alloin, F.; Dufresne, A. *Biomacromolecules* **2005**, *6*, 612–626.
5. Nishiyama, Y. *J. Wood Sci.* **2009**, *55*, 241–249.
6. Wada, M.; Sugiyama, J.; Okano, T. *J. Appl. Polym. Sci.* **1993**, *49*, 1491–1496.
7. Newman, R. H. *Holzforschung* **1999**, *53*, 335–340.
8. Langan, P.; Nishiyama, Y.; Chanzy, H. *J. Am. Chem. Soc.* **1999**, *121*, 9940–9946.
9. Langan, P.; Nishiyama, Y.; Chanzy, H. *Biomacromolecules* **2001**, *2*, 410–416.
10. Ciolacu, D.; Popa, V. I. *Cellul. Chem. Technol.* **2006**, *40*, 445–449.
11. Huebner, A.; Ladisch, M. R.; Tsao, G. T. *Biotechnol. Bioeng.* **1978**, *XX*, 1669–1677.
12. Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.
13. Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300–14306.
14. Cousins, S. K.; Brown, R. M. *Polymer* **1995**, *36*, 3885–3888.
15. Brady, J. W. *J. Am. Chem. Soc.* **1989**, *111*, 5155–5165.
16. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
17. Palma, R.; Zuccato, P.; Himmel, M. E.; Liang, G.; Brady, J. W. In *Glycosyl Hydrolases in Biomass Conversion*; Himmel, M. E., Ed.; American Chemical Society: Washington, 2001; pp 112–130.
18. Kuttel, M.; Brady, J. W.; Naidoo, K. J. *J. Comput. Chem.* **2002**, *23*, 1236–1243.
19. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926–935.
20. Durell, S. R.; Brooks, B. R.; Ben-Naim, A. *J. Phys. Chem.* **1994**, *98*, 2198–2202.
21. Donadio, D.; Bussi, G.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 14101.
22. Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182–7190.
23. Hess, B. *J. Chem. Theory Comput.* **2008**, *4*, 116–122.
24. Miyamoto, S.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 952–962.
25. Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577–8593.
26. Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187–217.
27. Brooks, B. R.; Brooks, C. L., III; MacKerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
28. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
29. Trzesniak, D.; Kunz, A.-P. E.; van Gunsteren, W. F. *ChemPhysChem* **2007**, *8*, 162–169.
30. Israelachvili, J. *Intermolecular and Surface Forces*, 2nd ed.; Academic Press: London, 1992.
31. Mason, P. E.; Neilson, G. W.; Enderby, J. E.; Saboungi, M.-L.; Brady, J. W. *J. Phys. Chem. B* **2005**, *109*, 13104–13111.
32. Nishiyama, Y.; Johnson, G. P.; French, A. D.; Forsyth, V. T.; Langan, P. *Biomacromolecules* **2008**, *9*, 3133–3140.
33. Qian, X. *Mol. Simul.* **2008**, *34*, 183–191.
34. Ueda, K.; Brady, J. W. *Biopolymers* **1996**, *38*, 461–469.
35. Liu, Q.; Schmidt, R. K.; Teo, B.; Karplus, P. A.; Brady, J. W. *J. Am. Chem. Soc.* **1997**, *119*, 7851–7862.
36. Umemura, M.; Yuguchi, Y.; Hirotsu, T. *J. Phys. Chem. A* **2004**, *108*, 7063–7070.
37. Shen, T.; Langan, P.; French, A. D.; Johnson, G. P.; Gnanakaran, S. *J. Am. Chem. Soc.* **2009**, *131*, 14786–14794.
38. Marchessault, R. H.; Peréz, S. *Biopolymers* **1979**, *18*, 2369–2374.
39. Nishida, Y.; Hori, H.; Ohrui, H.; Meguro, H. *J. Carbohydr. Chem.* **1988**, *7*, 239–250.
40. Matthews, J. F.; Skopec, C. E.; Mason, P. E.; Zuccato, P.; Torget, R. W.; Sugiyama, J.; Himmel, M. E.; Brady, J. W. *Carbohydr. Res.* **2006**, *341*, 138–152.
41. Bergenstråhle, M.; Berglund, L. A.; Mazeau, K. *J. Phys. Chem. B* **2007**, *111*, 9138–9145.
42. Ha, S.; Gao, J.; Tidor, B.; Brady, J. W.; Karplus, M. *J. Am. Chem. Soc.* **1991**, *113*, 1553–1557.
43. Liu, Q.; Brady, J. W. *J. Am. Chem. Soc.* **1996**, *118*, 12276–12286.
44. Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell, A. D. *J. Comput. Chem.* **2008**, *29*, 2543–2564.
45. Lins, R. D.; Hünenberger, P. H. *J. Comput. Chem.* **2005**, *26*, 1400–1412.
46. Yui, T.; Nishimura, S.; Akiba, S.; Hayashi, S. *Carbohydr. Res.* **2006**, *341*, 2521–2530.