AIP | The Journal of Chemical Physics

Transfer matrix approach to the hydrogen-bonding in cellulose Iα fibrils describes the recalcitrance to thermal deconstruction

Heinrich C. R. Klein, Xiaolin Cheng, Jeremy C. Smith, and Tongye Shen

Citation: The Journal of Chemical Physics **135**, 085106 (2011); doi: 10.1063/1.3626274
View online: http://dx.doi.org/10.1063/1.3626274
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/135/8?ver=pdfcov
Published by the AIP Publishing

Transfer matrix approach to the hydrogen-bonding in cellulose I_α fibrils describes the recalcitrance to thermal deconstruction

Heinrich C. R. Klein,[1,a)] Xiaolin Cheng,[2,3,b)] Jeremy C. Smith,[2,3,c)] and Tongye Shen[2,3,d)]
[1]*Computational Molecular Biophysics Group, Interdisciplinary Center for Scientific Computing, University of Heidelberg, 69120 Heidelberg, Germany*
[2]*UT/ORNL Center for Molecular Biophysics, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37830, USA*
[3]*Department of Biochemistry, Cellular and Molecular Biology, University of Tennessee, Knoxville, Tennessee 37996, USA*

(Received 17 May 2011; accepted 29 July 2011; published online 30 August 2011)

Cellulosic biomass has the potential to serve as a major renewable energy source. However, its strong recalcitrance to degradation hampers its large-scale use in biofuel production. To overcome this problem, a detailed understanding of the origins of the recalcitrance is required. One main biophysical phenomenon leading to the recalcitrance is the high structural ordering of natural cellulose fibrils, that arises largely from an extensive hydrogen-bond network between and within cellulose polymers. Here, we present a lattice-based model of cellulose I_α, one of the two major natural forms, at the resolution of explicit hydrogen bonds. The partition function and thermodynamic properties are evaluated using the transfer matrix method. Two competing hydrogen-bond patterns are found. This plasticity of the hydrogen-bond network leads to an entropic contribution stabilizing the crystalline fibril at intermediate temperatures. At these temperatures, an enhanced probability of bonding between the individual cellulose chains gives rise to increased resistance of the entire cellulose fibril to degradation, before the final disassembly temperature is reached. The results are consistent with the available crystallographic and IR spectroscopic experiments on the thermostability of cellulose I_α.
 [doi:10.1063/1.3626274]

I. INTRODUCTION

Cellulosic biomass, the most abundant biomaterial on earth,[1–3] has a vast yet mostly untapped potential for biofuel production.[4–9] However, in its natural setting as a major component of plant cell walls, cellulose is recalcitrant to deconstruction.[8,9] This recalcitrance strongly limits the efficiency of cellulase enzymes in breaking cellulose down into glucose, as is required for subsequent conversion into biofuels, thus necessitating costly and energy-intensive pretreatment of biomass before conversion.[2]

Generally, hydrogen bonds (H-bonds) have been recognized as one of the most important physical forces in the assembly of macro(bio)molecular complexes. They also play a very important role in the stability of cellulose. Native cellulose occurs in two different crystalline forms, I_α and I_β,[3] which are biosynthesized as polymer fibrils. This crystallinity, owing largely to an extensive and robust network of H-bonds between and within cellulose polymer chains, is one of the main physical sources of biomass recalcitrance.[8] For this reason, a fundamental understanding of the thermodynamic properties of the H-bond network in cellulose fibrils is essential for understanding the origins of recalcitrance and identifying suitable strategies for biomass pretreatment.

Just like ice having a plethora of phases, cellulose exists in multiple phases/polymorphs due to multiple possibilities of hydrogen bonding patterns. I_α and I_β are two distinct phases naturally produced. Whereas cellulose produced by higher plants consists predominately of the I_β phase, cellulose produced by algae and bacteria is rich in I_α.

High-resolution crystal structures of cellulose I_α and I_β obtained by neutron scattering at room temperature reveal the two distinct and complex H-bond networks between and within cellulose strands in each crystal phase.[10,11] These extensive H-bond networks are the major contributors to the stability of the fibrils. Intra-chain H-bonds stiffen each polymer chain, while inter-chain H-bonds form two-dimensional sheet structures. In contrast to the extensive H-bonding within individual sheets, neighboring sheets pack together mainly through van der Waals interactions, no H-bonds of the OH···O type are involved.[10,11] Hence, the native H-bond networks in cellulose I_α and I_β can be treated as two-dimensional problems. Unlike polypeptides and polynucleotides, in which the majority of the H-bonds can often be unambiguously specified in any given native structure, multiple H-bonding configurations are potentially possible in cellulose. Since some of these H-bonds cannot coexist, frustration appears in the low-energy state structures.

Various computational studies of crystalline cellulose have been reported. Molecular dynamics (MD) simulations have been widely used to probe structure and dynamics,[12–14] but significant differences have been observed between the structures obtained in simulations and in diffraction

a)Present address: Institute for Theoretical Physics, University of Heidelberg, 69120 Heidelberg, Germany. Electronic mail: heinrich.klein@bioquant.uni-heidelberg.de.
b)Electronic mail: chengx@ornl.gov.
c)Electronic mail: smithjc@ornl.gov.
d)Electronic mail: tshen@utk.edu.

experiments.[13,14] Alternative analytical models based on statistical physics have the potential to provide required insight.

In the current study, we focus on the case of cellulose I_α. A statistical, lattice-based model has been employed recently to examine the temperature-dependent behavior of the H-bond network in cellulose I_β.[15] This model is based on the evaluation of the free energy by exploiting the quasi-one-dimensional structure of cellulose I_β. However, several features of the H-bond network of cellulose I_α have not been addressed in the framework of the I_β model, including an alternative scheme of bifurcated and single H-bonds and an I_α specific H-bond with both inter- and intra-chain contributions. Here, we examine the temperature-dependent properties of the H-bond network of cellulose I_α, taking into account these special features, and compare the results with experimental findings.[16]

The theoretical approach used here is based on the evaluation of a partition function with the transfer matrix method. The method was introduced in the context of the Ising model[17,18] and has been previously employed in the studies of many one-dimensional biological systems and polymer chains,[19,20] such as the peptide helix-coil transition[21,22] and DNA-melting.[23,24] As will be demonstrated, the physics of H-bonding in cellulose I_α can be described by a one-dimensional system. The full cellulose sheet is decomposed into a sequence of subnetworks of H-bonds formed within and between adjacent glucose units. Each particular state of the H-bond network is assigned a statistical weight based on its free energy: the H-bond formation is associated with an energy gain, whereas the absence of H-bonds leads to an entropy gain. Possible H-bonds are limited to the native H-bonds found in I_α,[10] and thus no transition to other crystal structures can be investigated in the model. The strength of this lattice-based model lies in the simplicity of the physical concept, with only a small number of phenomenological parameters, and, as will be shown, in its ability to capture the thermodynamic changes of the structures characterized by H-bond networks, from a low temperature stable phase to the onset of the disassembly. The model suggests the presence of two competing H-bond patterns in cellulose I_α appearing before its final disassembly, thus revealing a mechanism which might help to stabilize the cellulose fibril in the intermediate temperature range.

II. MODEL AND METHODS

We use the IUPAC-IUB convention on polysaccharide chains[25] for naming the cellulose atoms and specifying the polymer direction. Cellulose I_α has a triclinic unit cell containing one chain consisting of two glucose rings. Although there is no requirement for space group symmetry, contiguous residues along a cellulose chain adopt a similar conformation after a twofold screw axis rotation. All hydroxymethyl groups are reported in the *tg* conformation with no evidence for inter-sheet H-bonds.[3,10] Thus, the individual cellulose molecules, referred to as chains, form two-dimensional planar assemblies connected by strong H-bonds, which hereafter we will denote as sheets.

TABLE I. Potential H-bonds.

Bond name	Location	Atoms (D-H···A)
A(a)/A(e)	Intra/Inter	$O_2H_2\cdots O_3/O_6$
B_1	Intra	$O_6H_6\cdots O_1$
B_2	Intra	$O_6H_6\cdots O_1/O_2$
C_1	Intra	$O_2H_2\cdots O_6$
C_2	Intra	$O_2H_2\cdots O_6/O_1$
D_1	Intra	$O_3H_3\cdots O_5$
D_2	Intra	$O_3H_3\cdots O_5/O_1$
G	Inter	$O_6H_6\cdots O_3/O_2$

A fibril consists of stacked sheets of cellulose chains, and the H-bond network in cellulose I_α can be considered as essentially two dimensional. Within one sheet, we can define a "row" as a collection of glucose monomers perpendicular to the chain direction (Fig. 1(a)). Each chain of cellulose I_α is made up of two types of conformationally distinct monomers. In comparison, cellulose I_β possesses two conformationally different chains, each chain consisting of one type of glucose conformation. Taking a closer look at Figs. 1(b) and 1(c), various H-bonds can be identified. These are listed in Table I and can be classified as inter-chain and intra-chain H-bonds with G being the only pure inter-chain H-bond. The bifurcated H-bond A, specific to cellulose I_α, contains both an inter-chain contribution, A(e) and an intra-chain contribution, A(a). Moreover, A(a) pointing along the chain direction is distinct from the other intra-chain H-bonds as it does not connect two glucose monomers but can rather be considered as an intra-monomer H-bond. This character of H-bond A requires a special treatment, which will be discussed below.

In the crystal structure of cellulose I_α, two competing patterns of H-bonds have been identified (Fig. 1(c)). Pattern I has a probability of 55% and pattern II a probability of 45% at room temperature.[10] Most of the hydrogen atoms involved in the formation of H-bonds can be found in more than one position, and this leads to various possibilities of frustration in the H-bond network.

Furthermore, there is an interesting scheme of alternation of single (B_1,C_1,D_1) and bifurcated (B_2,C_2,D_2) H-bonds in neighboring monomers along a chain reported in the crystal structure, which is a characteristic feature of cellulose I_α. The exact nature of bifurcated H-bonds exhibiting one donor and two potential acceptors is debated.[29,30] However, this is not essential to this work. The energy difference between the two types of H-bonds appears to be small. Thus, it seems likely that, with increasing temperature and thermal fluctuations, the alternating scheme disappears and a more random formation of single and bifurcated H-bonds occurs, a phenomenon similar to an antiferromagetic-to-paramagnetic phase transition. Hence, to constrain one row to exhibit only the bifurcated or the single type of H-bond may be a too restrictive treatment.

In order to allow treatment of a sufficiently large system size to be computationally feasible, we condense the alternating H-bond scheme into one row (depicted in Fig. 1(d)). In this condensed model, either the single or the bifurcated type of those alternating H-bonds can be formed within the same row. As both types are competing and hence are mutu-

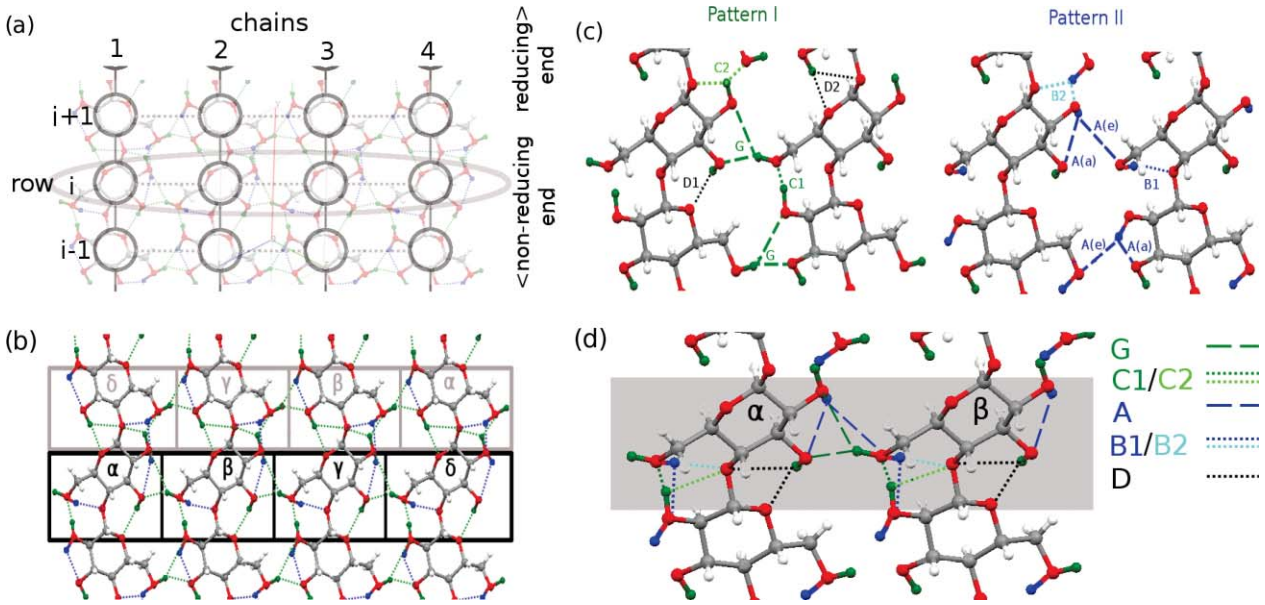

FIG. 1. An illustration of cellulose I_α structure and the corresponding H-bonding modeling. (a) Schematic view of a cellulose sheet. Individual glucose monomers are represented by circles. They are connected by vertical lines representing the chain axes of the cellulose strands. Potential inter-chain H-bonds are depicted by horizontal, dashed lines. (b) H-bond network of a four-chain system with switching subunits labeled by Greek letters. (c) Close up view of the two competing H-bond patterns. (d) A zoomed-in view of the complex H-bond network.

ally exclusive, the total probability of finding such an H-bond regardless of its single or bifurcated nature, is given by the sum of probabilities of individual types.

As depicted in Fig. 1(c), the hydrogens involved in the formation of H-bonds of type D (D_1 or D_2) do not exhibit different positions in pattern I and pattern II. They do not directly interact with the rest of the H-bond network. Therefore, we define an effective H-bond D, neglecting its single or bifurcated nature. This implicit treatment of H-bond D reduces the number of configurations without losing cooperative effects.

With the above simplifications, we define a set of base states representing possible combinations of H-bonds that can occur between two neighboring chains (shaded box in Fig. 1(d)). These possible states are enumerated by the following:

$$\text{Base states} = \{\phi, \mathrm{B}_1, \mathrm{B}_2, \mathrm{C}_1, \mathrm{C}_2, \mathrm{G}, \mathrm{G}+\mathrm{C}_1, \mathrm{G}+\mathrm{C}_2, \mathrm{A}, \mathrm{A}+\mathrm{B}_1, \mathrm{A}+\mathrm{B}_2, \mathrm{A}+\mathrm{C}_1, \mathrm{A}+\mathrm{C}_2\}. \quad (1)$$

Here the symbol $+$ in $X + Y$ represents the simultaneous formation of both H-bonds X and Y, and ϕ denotes the absence of any explicitly specified H-bonds. In a simplified notation, we can enumerate the configurations and define a set of 13 canonical base vectors $|0\rangle, \cdots, |12\rangle$. The number of base states for cellulose I_α is greater than that of the I_β form,[15] due to the larger number of H-bond permutations of I_α.

It is possible to incorporate the alternating scheme for the H-bonds of type B and C in the model using a phenomenological energy coupling parameter, J, which mimics the conformational preference of adjacent rows. Parameter J can be considered as analogous to the coupling energy in antiferromagnetic models. Thus, J represents an additional energy contribution ensuring the persistency of the switching scheme, i.e., relating the formation of a H-bond in row i to the conformation of row $i-1$. For example, the total formation energy U'_{C_1} of H-bond C_1 in row i consists of the energy of formation U_{C_1} and the additional coupling energy related to the presence of H-bonds C_1 and C_2 in row $i-1$,

$$U'_{C_1} = U_{C_1} + J(\delta_{C_2^{i-1}} - \delta_{C_1^{i-1}}). \quad (2)$$

Here, $\delta_{C_j^{i-1}}$ is 1 if H-bond C_j $(j = 1, 2)$ is formed in row $i-1$, and 0 otherwise. A positive J promotes the alternating scheme.

Hydrogen bonds with an intra-chain character can also be formed on the boundary of the sheet in contact with the surrounding environment. To lower the number of configurations, these H-bonds are implicitly treated and their interactions with adjacent rows are neglected. Given the approximate treatment of these boundary H-bonds, it is reasonable to focus on the H-bonds located inside individual sheets that are important for the high crystallinity of the fibril. Finally, we note that the approximate twofold screw symmetry of cellulose I_α leads to a quasisymmetric switching of the positions of the individual subunits. The subunits are labeled by Greek letters and their switching is schematically shown in Fig. 1(b).

As shown in Fig. 1(a), the topology of the sheet with M chains and L glucose units per chain is essentially described by a square lattice of size $M \times L$. Typical plant micro-fibrils have a value of $M < 10$.[26] In contrast, individual cellulose chains have lengths $L = 2\,000$–$15\,000$.[13,27,28] Thus, we can exploit the fact that the fibril diameter is much smaller than its length ($M \ll L$) to reduce the dimensionality further. Here in this quasi-one-dimensional model, we explicitly describe all states of the H-bond network in one row by a state vector. For an M-chain system, the state vector is spanned by the direct product of the $M-1$ base states ($n_{\text{base}} = 13$) identified for a two-chain system (Eq. (1)). Thus, the number of states in an M-chain system is given by $n = 13^{M-1}$ with each of

the states $i = 1, 2, \ldots, n$ describing a specific and complete configuration of a row. In this representation, we can define a transfer matrix $\mathbf{T}$ of size $n \times n$. To set up the transfer matrix, the free energy of each state is evaluated, taking into account the effects of bonding energy as well as the conformational entropy changes associated with the H-bond formation.

The bonding energy of an individual H-bond depends not only on the nature of donor and acceptor atoms involved and their relative positions, but also on the local environment. Thus, it is impractical to draw a reliable correlation between the crystallographic data and the energy of formation of individual H-bonds, especially for the bifurcated ones. Moreover, no calorimetric data on the individual H-bond energies are available. Therefore, we assign a uniform bonding energy U of 5 kcal/mol to the formation of each H-bond, regardless of the single or bifurcated character. Assigning this typical energy for intermediately strong H-bonds[30,31] is in agreement with the previous work on I_β.[15]

Evaluating the entropic contribution proves to be more complicated. H-bonds constrain the configuration space a molecule can explore. Thus, breaking up H-bonds leads to a gain in entropy because the glucose ring and its hydroxyl groups can rotate and move more freely. Therefore, we propose an entropic gain, S associated with the disruption of all intra-chain H-bonds connecting one glucose unit to its predecessor along the chain. This entropic gain also strongly depends on the coordinating presence of related inter-chain H-bonds,

$$S = S_a + (4 - z)/4 \times S_x; \quad z \in [0, 4]. \tag{3}$$

Similar to the previous work on I_β,[15] S_a denotes the entropic gain arising from the disruption of all intra-chain H-bonds connecting two glucose monomers, regardless of the presence of any inter-chain H-bonds. S_x is a collective entropy parameter taking into account the inter-chain H-bonds. The counting variable z is equal to the sum of inter-chain H-bonds connecting a glucose unit and its predecessor to the neighboring chains. For cellulose I_β, the total entropy difference per glucose between the crystalline state and the onset of disorder was estimated using the Lindeman melting criterion to be $\sim 20k_B$,[15,32] where k_B is the Boltzmann constant. As the unit cell volume per formula unit for cellulose I_α and cellulose I_β is almost identical,[10,11] we assume the same total entropic gain for cellulose I_α and I_β, which is evenly distributed between the two entropic parameters S_a and S_x ($S_a = S_x = 10k_B$).

For H-bond A in cellulose I_α, we have to treat its inter-chain and intra-chain (intra-monomer) entropic contributions differently. The intra-chain contribution A(a) does not constrain the motion of the whole glucose unit, and hence it has a much smaller impact on the entropy than a regular intra-chain H-bond. Thus, in contrast to the formation of regular intra-chain H-bonds, the formation of A(a) only reduces the entropic contribution S'_a by a reduction factor of 0.5 ($S'_a = 0.5 \times S_a$). Varying the reduction factor by 10% is found not to significantly affect the results. Moreover, we have to account for the inter-chain part A(e). The coordinating effect of A(e) is assumed to be weaker than the effect of the pure inter-chain H-bond G. This is incorporated into the model via the above introduced variable z, which counts the number of inter-chain H-bonds. Whereas the formation of the pure inter-chain H-bond G increases z by 1, the formation of A(e) increases z by 0.5.

To aid in understanding, individual transfer matrix elements for a two-chain system are exemplarily computed in the supplementary material.[38] For one sheet consisting of more than two chains, the setup is slightly more complicated due to the rapidly increased number of configurations, and the approximate twofold screw symmetry of the fibril, leading to a switch in the positions of the subunits in neighboring rows (Fig. 1(b)). However, the basic concept remains very similar. As the partition function and hence the transfer matrix contains all information about possible configurations and their statistical weights, it is possible to calculate the equilibrium joint probabilities for each pair of configurations, referred to as the pair state probability[19] $P_{\mu\nu} = (S^{-1}_{1\mu} T_{\mu\nu} S_{\nu 1}) \times \lambda_1^{-1}$. Here μ refers to one particular state of the upstream row of the sheet, ν refers to another particular state of the downstream row, and λ_1 is the largest eigenvalue of the transfer matrix $\mathbf{T}$. $\mathbf{S^{-1}}$ and $\mathbf{S}$ are the matrices used to diagonalize $\mathbf{T}$. The rows of $\mathbf{S^{-1}}$ are the left, transposed eigenvectors of $\mathbf{T}$, and the columns of $\mathbf{S}$ are the right eigenvectors.

The pair state probabilities also satisfy two sum rules $\sum_{\mu\nu} P_{\mu\nu} \equiv 1$ and $\sum_\mu P_{\mu\nu} = \sum_\mu P_{\nu\mu}$; the former can be understood as the normalization property.[19] This formalism provides an easy way to calculate the probability of formation of any given explicitly modeled H-bond,[15] say Y, by summing up all matrix elements of the pair state probability matrix $\mathbf{P}$ that correspond to configurations in which Y is formed: $P[Y] = \sum_{\mu\nu} P_{\mu\nu} \delta_{\mu,Y}$. Here, $\delta_{\mu,Y}$ is 1 if H-bond Y is part of the configuration μ and 0 otherwise. Pair-state probabilities also allow calculations of the probability of formation of any implicitly modeled H-bonds, say Y'. Using an additional weighting factor $q_{\mu\nu,Y'} \in [0, 1]$, which specifies the relative amount the implicitly modeled H-bond Y' contributes to $T_{\mu\nu}$, the probability for the implicitly modeled H-bond Y' is given by: $P[Y'] = \sum_{\mu\nu} P_{\mu\nu} q_{\mu\nu,Y'}$. The LAPACK routine DGEEVX was used to diagonalize the transfer matrix.[33]

III. RESULTS AND DISCUSSION

A. The two-chain system ($M = 2$)

1. The probabilities of formation of individual H-bonds

We initially investigate the H-bond network in a sheet consisting of two chains only. Since solvent effects are not explicitly incorporated, we will focus on the H-bonds located between the two chains that are assumed to be mainly responsible for the high crystallinity of the fibril. The probabilities of formation of these H-bonds as a function of temperature are shown in Fig. 2, in which we adopt the temperature unit kcal/mol/k_B ($\approx$503 K), which will be used throughout this work.

Figure 2 shows different H-bonds competing at low temperatures, whereas at higher temperatures the entropy dominates, leading to the disruption of all the H-bonds. The nonfrustrated H-bond D is formed with nearly 100% probability at lower temperatures and exhibits a simple decay

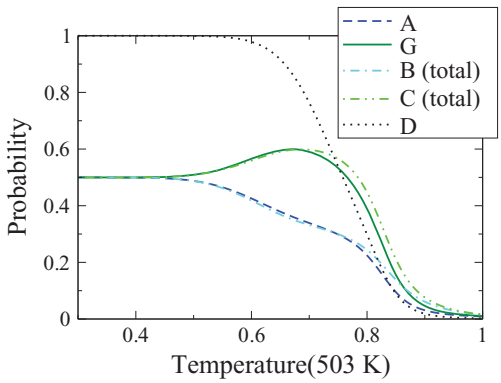

FIG. 2. Probabilities of H-bond formation in a two-chain system. Only H bonds formed between the two chains are enumerated. For H-bonds B and C, the total probability of formation regardless of their bifurcated or single nature, is shown.

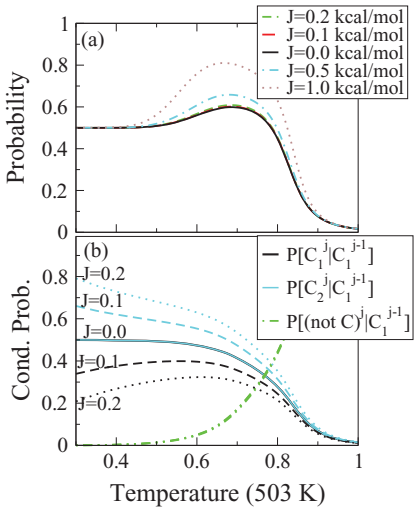

FIG. 3. Effect of parameter J on (a) the total probability of formation of H-bond C and (b) on the alternating scheme of H-bond C using conditional probabilities.

characteristic with rising T. In contrast, H-bonds A and G show a more complex behavior. At low temperatures, both H-bonds are formed with equal probability, which is expected as both are competing and their energy of formation is assumed to be equal. However, bonds A and G behave differently with increasing temperature. In the intermediate temperature range, enhanced formation of the pure inter-chain H-bond G is observed, whereas the probability of finding the partly inter-chain H-bond A constantly decreases. The concerted behavior of H-bonds A and B (pattern I) and H-bonds G and C (pattern II) is observed, in agreement with the two competing patterns identified experimentally.[10] Here, B and C refer to the total probability of formation of H-bonds of types B and C. This notation will be used from here on unless we specifically address the bifurcated and single forms, which will be indicated by a subscript, e.g., B_1 and B_2.

2. Alternating scheme of H-bonds in consecutive rows

In this subsection, we investigate the possibility of an alternating scheme of H-bonds along the chain. Thus far, the energy parameter J (Eq. (2)), which can be used to favor the alternating scheme, has been set to 0. In the crystal structure, however, the occurrence of an alternating scheme has been observed due to small conformational differences in consecutive rows.[10] As the energy coupling parameter J is of phenomenological nature, it is important to verify that the overall probabilities of formation for the H-bond network remain broadly unaffected when a finite value is assigned to J. From Fig. 3(a), it is evident that small values of J (0.1–0.2 kcal/mol) do not affect the total probability of formation significantly. However, larger J values lead to the divergence in the total probability profiles. Similar behavior was found for the other H-bonds (not shown). Therefore, in order not to bias the total probability occurrence with this phenomenological parameter, we only consider sufficiently small values of J.

Restricting ourselves to small energy values for J (compared to the energy of formation of an individual H-bond, U), we next examine how J affects the probability of finding the single H-bond C_1, the bifurcated H-bond C_2 or none of the two in one row if the neighboring row is constrained to H-bond C_1. This question can be expressed using the conditional probabilities: P_{C_1,C_1}, P_{C_2,C_1}, and $P_{\text{not C},C_1}$. Here, generally $P_{X,C_1} = P[X^i|C_1^{i-1}]$ with X^i being the state at position i and C_1^{i-1} denoting C_1 being formed in row $i-1$. Figure 3(b) shows a clear impact of J on these conditional probabilities. The solid lines depict the probabilities in the degenerated case of $J = 0$ kcal/mol, showing an equal formation of the single and bifurcated H-bond C, which corresponds to a random distribution over adjacent rows. By increasing J slightly, however, we can see a clear preference for the alternating scheme at low temperatures. This preference is weakened with increasing temperature, but stays preserved up to the final disassembly. Similar behavior was found for H-bond B, which is not shown here. These results show that it is possible to model the alternating scheme of single and bifurcated H-bonds by introducing a phenomenological energy parameter, which can be chosen sufficiently small so that the overall probabilities of formation are not affected significantly.

B. Multichain cases

Due to an exponential increase of the matrix size with the number of chains in a sheet, the maximum system size in the model is limited to one sheet consisting of four chains. Nevertheless, a clear impact of the system size on the probabilities of formation can be observed.

1. Effect of the system size on the probabilities of formation

In Fig. 4, the effect of the system size (number of chains in one sheet) on the averaged probabilities of formation of the individual H-bonds is depicted. The average was performed over all H-bond positions in one row (Fig. 1), excluding surface H-bonds to avoid potential artifacts. The system

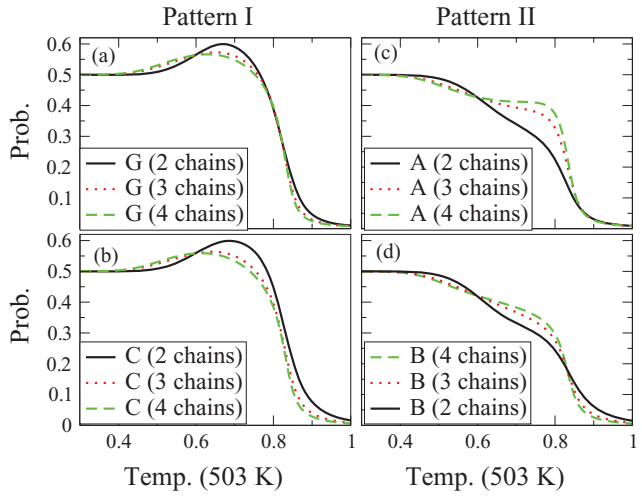

FIG. 4. Spatially averaged probabilities of formation. Pattern I: (a) H-bond G and (b) H-bond C. Pattern II: (c) H-bond A and (d) H-bond B (surface H-bonds excluded).

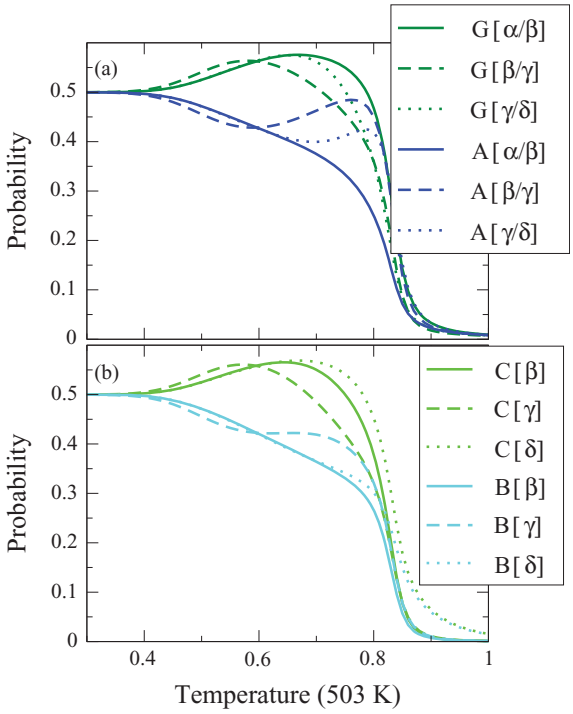

FIG. 5. The probabilities of formation of the H bonds at the resolution of individual subunits for a four-chain system. (Subunits labeled in Fig. 1(b)). (a) H-bonds A (dashed lines) and G (solid lines) with inter-chain character. (b) Intra-chain H-bonds B (dashed lines) and C (solid lines).

size shows the largest effect for H-bonds A and B belonging to pattern II (Figs. 4(c) and 4(d)). In the two-chain case, H-bond A exhibits a monotonic decay. However, with increasing system size a plateau is reached after an initial decay and before the final disassembly. In this plateau region, the averaged probability of formation of A remains nearly constant over a large temperature range. Furthermore, the initial decrease in the averaged probability of formation of A is shifted to slightly lower temperatures for an increased system size. Similar behavior is also observed for H-bond B in pattern II. However, the system size has weaker effects on this H-bond; and no plateau region occurs. In Figs. 4(a) and 4(b), the impact of the system size on the averaged probabilities of formation of H-bonds G and C in pattern I is depicted. The behavior for both H-bonds remains qualitatively similar for all system sizes. However, for larger system sizes, the increase in the probability of formation of H-bonds G and C is less pronounced and shifts to lower temperatures. The formation of H-bond D (data not shown) is insensitive to the system size at all temperatures, with only a very small shift of the decaying probability profiles to higher temperatures when the system size increases. For all H-bonds, the difference between the two- and three-chain systems is greater than that between the three- and four-chain systems, indicating fast convergence of the averaged probabilities of formation with increasing system size, as was observed in a similar model for cellulose I_β.[15]

We also studied a reduced model (data not shown), in which we neglected to distinguish between the single and bifurcated H-bonds, making it impossible to observe the alternating scheme but permitting a larger system (i.e., a five-chain system) to be computationally tractable. The averaged probabilities of formation calculated with this reduced model were found to well reproduce those calculated in the full model. Moreover, the results for the four- and five-chain systems in the reduced model are not qualitatively different, suggesting that the results from the four-chain model are sufficiently converged, and thus can be meaningfully compared with experimental data obtained from larger systems.

2. A closer look at the four-chain system

In the previous subsection, we have studied the impact of the system size on the averaged probabilities of formation. Here, we take a closer look at the H-bond formation at the individual H-bond positions in a four-chain system. The corresponding probabilities of formation are depicted in Fig. 5 (without surface H-bonds). The positions or subunits are labeled by Greek letters according to the black row in Fig. 1(b).

The probabilities for the H-bonds A and G connecting different subunits are shown in Fig. 5(a). The pure inter-chain H-bond G exhibits similar behavior at all positions. However, the increase of the probability of formation is more pronounced and occurs at higher temperatures for those H-bonds connecting the outer chains to the inner chains ($\alpha\cdots\beta$, $\gamma\cdots\delta$) than for those connecting the two inner chains ($\beta\cdots\gamma$). As the cohesive effect of the pure inter-chain H-bond G might be stronger than that of the part inter-chain H-bond A, the above observation suggests that at intermediate temperatures the outer chains could be tied more tightly to the core of the fibril. This mechanism might contribute to the recalcitrance of the cellulose I_α fibril in the intermediate temperature range because glycosyl hydrolases work on the surface of the cellulose fibril. Compared with H-bond G, the behavior of H-bond A is more complex. At the position connecting subunits α and β, it exhibits a monotonous decay similar to the two-chain case. However, at the other positions, after the initial decrease, an increase of probability is observed

at high temperatures, leading to a small temperature range, in which H-bond A exhibits a higher formation probability than H-bond G. The monotonic decay of the probability of H-bond A at position $\alpha \cdots \beta$ coincides with the rise in the formation H-bond G at the same position in relatively high temperatures before the final disassembly of the fibril. As both H-bonds are mutually exclusive, the increased stability of G at this position suppresses any increase in H-bond A. At the other positions, however, H-bond G is already decaying, making a higher chance of forming H-bond A. Although H-bond A has presumably a weaker ability to "glue" two chains together (due to its only partly inter-chain character and the relatively larger donor-acceptor distance), its increase at high temperatures might still help compensate some stability loss due to the decreased formation of G.

The probabilities of formation of the intra-chain H-bonds B and C are shown in Fig. 5(b). H-bond C behaves very similar to H-bond G, as both belong to the same pattern. The probability of H-bond B also resembles that observed in H-bond A. However, the increase toward high temperature regions at the γ position is not significant. Thus, H-bond B does not exhibit a higher probability of formation than that of H-bond C at any temperature.

C. Comparison with experiments

Here we compare the calculated thermodynamic quantities for the four-chain model with available experimental data. Although the experimental data were obtained for larger systems, the comparison is justified as the convergence of the averaged probabilities of formation with increasing number of chains was observed.

1. The disassembly temperature

An experimentally well-known property of cellulose I_α is its annealing to cellulose I_β or other high temperature phases under heat treatment.[34,35] Studies on cellulose I_α suggest that the intermolecular H-bonds break at 200 °C–220 °C.[16,36] Here, we define the disassembly temperature, T_D as the temperature at which the averaged probability of formation $P[X]$ for an H-bond X has decreased to half of its low temperature value ($P[X]|_{T_D} := 0.5 \times P[X]|_{T=0.3\ \mathrm{kcal/mol/k_B}}$). The convergence of the probabilities of formation at low temperatures makes this a plausible definition. Using this definition, we obtain a disassembly temperature of ∼145 °C for most H-bonds (A, G, C), while the calculated disassembly temperatures for H-bonds D and B are slightly lower. Thus, the calculated T_D is ∼60 °C lower than that experimentally measured. However, this deviation can be corrected since T_D depends critically on the ratio of the entropic to energetic contributions associated with the H-bond formation. Decreasing the overall entropy in a symmetric fashion (i.e., keeping the ratio S_a/S_x fixed (see Eq. (3))) leads to a trivial rescaling of the temperature axis and hence an increased T_D without affecting the shapes of the underlying curves. This is depicted in Fig. 6(a) for H-bonds A and G. With improved estimations of the H-bond energies, the reproduction of T_D would in principle result in a better estimation of the entropies connected with H-bond disruption. For further comparisons discussed below, we will employ the scaled temperature, T' to account for the differences between the measured and calculated disassembly temperatures. T' is derived from the experimental temperature by multiplying it by the ratio of the calculated and measured disassembly temperatures:

$$T' = T_{\mathrm{exp}} \cdot T_{D\,\mathrm{cal.}}/T_{D\,\mathrm{exp.}} = 0.8646 T_{\mathrm{exp}}. \quad (4)$$

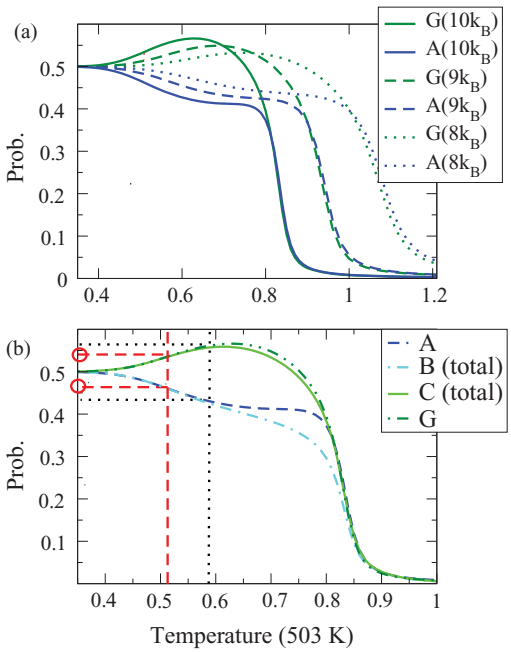

FIG. 6. Comparison with experiments. (a) Effect of a symmetric decrease of entropy on H-bonds A and G. (b) Average H-bond probabilities of formation (four-chain system). Here room temperature T_{room} and scaled room temperature T'_{room} are labeled with the dotted and dashed vertical lines, respectively.

2. Comparison to the crystal structure

Apart from the disassembly temperature, the occurrence probabilities of the H-bonds can also be compared to those in the crystal structure,[10] which was obtained at room temperature. As described above, two competing patterns have been identified in the crystal structure of cellulose I_α: pattern I consisting of H-bonds G and C, and pattern II consisting of H-bonds A and B, with probabilities of occurrence of ∼55% and ∼45%, respectively. Figure 6(b) displays the averaged probabilities of H-bond formation in a four-chain system. Here, pattern I exhibits a higher probability of formation than pattern II over a broad range of temperatures. At the scaled room temperature $T'_{\mathrm{room}} = 0.51$, depicted by a dashed line, pattern I occurs with a probability of ≈54% and pattern II of ≈46%, in good agreement with the experimentally measured values. Lacking crystallographic data of cellulose I_α at other temperatures, it is not possible to verify the theoretical increase of the probability of formation of pattern I to 60% before the final fibril disruption.

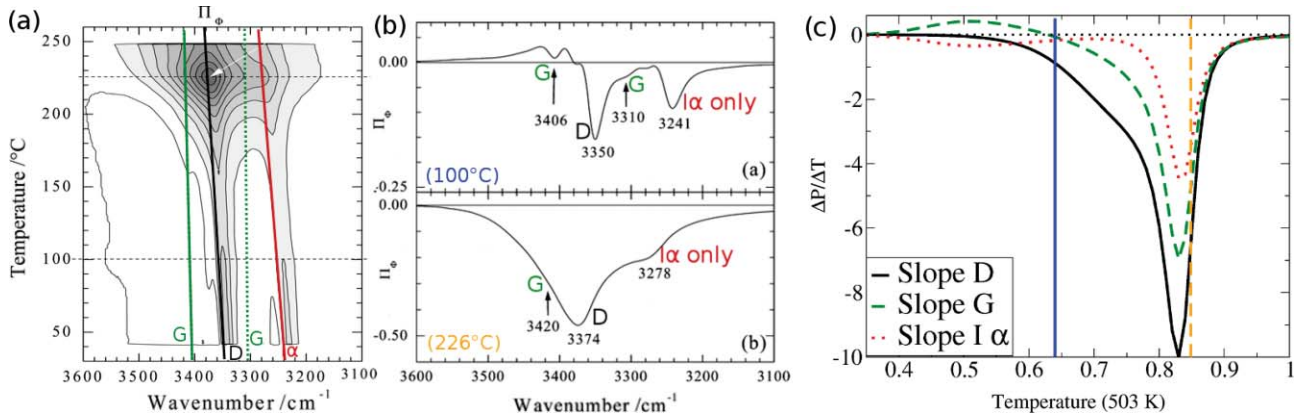

FIG. 7. Reproduced IR spectra plots from Watanabe *et al.*[16] are shown in (a) and (b). (a) Perturbation-correlation moving window two-dimensional (PCMW2D) spectrum with a temperature range of 30 − 260 °C. Positive and negative correlations are represented by white and gray areas, respectively. Peak positions related to H-bond D, G, and those formed in I_α solely are depicted by black, green, and red lines, respectively. (b) Two temperature slices of the PCMW2D at 100 ° (upper panel) and 226 ° (lower panel). (c) For a comparison to our data, the approximate slopes of average probabilities of formation of H-bonds D (solid line), G (dashed line), and H-bonds formed in cellulose I_α solely (dotted line) are shown. Vertical lines correspond to the rescaled temperatures of the sliced spectra shown in (b).

3. Comparison to IR studies

The temperature dependence of the H-bond network in cellulose I_α has been studied by IR spectroscopy combined with perturbation-correlation moving-window two-dimensional (PCMW2D) spectroscopy.[16] The PCMW2D provides a 2D correlation spectrum plotted on a plane between the wave number axis and the perturbation temperature axis. Figures 7(a) and 7(b) are reproduced from Figures 3 and 4 of Watanabe *et al.*[16] for a direct comparison. It has been demonstrated that the correlation spectra obtained with PCMW2D are proportional to the derivative of the IR absorbance or the spectral intensity y with respect to T, $\Pi_\Phi(\nu, T) \sim \partial y(\nu, T)/\partial T$.[37] Thus, the PCMW2D spectra can be related to the slopes of the averaged probabilities of H-bond formation with respect to temperature.

In the study of Watanabe *et al.*, those O–H stretching vibration peaks corresponding to H-bonds $O_3H_3\cdots O_5$, $O_6H_6\cdots O_3/O_2$ and H-bonds formed in cellulose I_α exclusively (not found in I_β) were identified. We can relate the first peak at 3348 cm^{-1} for H-bond $O_3H_3\cdots O_5$ to H-bond D. Although both the single and the bifurcated types of this H-bond have been reported in the crystal structure,[10] no bifurcated form was found in the IR spectroscopic study,[16] consistent with our approximate treatment of H-bond D.

The H-bond $O_6H_6\cdots O_3/O_2$ can be unambiguously assigned to the pure inter-chain H-bond G. The comparison of the crystal structures of cellulose I_α[10] and cellulose I_β[11] shows that H-bonds A and B_1 exist exclusively in cellulose I_α. Thus, we assume that the peak identified experimentally for the H-bonds solely formed in cellulose I_α can be related to a simple combination of H-bonds A and B_1, i.e., $P[I\alpha_{\text{solely}}] = 0.5(P[A] + P[B_1])$. In Fig. 7(a), the peaks related to H-bond D, G and those solely formed in I_α are depicted with black, green, and red lines, respectively.

The slopes of the averaged probabilities of formation of H-bonds G, D_1, and $I\alpha_{\text{solely}}$ as a function of T are depicted in Fig. 7(c) and exhibit several features that can be compared with the PCMW2D spectrum displayed in Fig. 7(a).[16] First, among all the H-bonds shown, only H-bond G exhibits a region of positive slope. Moreover, qualitatively, the calculated slopes correspond well to those observed in the PCMW2D spectra over the entire temperature range. Watanabe *et al.* presented two representative slices of the PCMW2D spectrum at 100 °C and 226 °C, as shown in Fig. 7(b). The corresponding scaled temperatures are marked in Fig. 7(c) by vertical lines. The sliced experimental spectrum at 100 °C shown in the upper part of Fig. 7(b) exhibits a slightly positive correlation for one and a slightly negative correlation for the other of the two peaks corresponding to H-bond G. This feature corresponds well with Fig. 7(c), where the lower scaled temperature (100 °C) marks the change from a positive to a negative slope of H-bond G. However, two vibrational peaks were identified spectroscopically for H-bond G, making the quantitative comparison to the single probability of formation difficult. Similarly, the single vibrational peak identified for the H-bonds solely formed in I_α has been related to a combination of the two H-bonds A and B_1, which leads to some ambiguity in the comparison. Except the aforementioned H-bonds A, B_1, D, and G, other H bonds, such as B_2 and C, are difficult to identify using the IR spectra. Thus, no comparison will be made for those bonds here.

IV. CONCLUSION

Understanding the origins of biomass recalcitrance to deconstruction, and in particular the molecular mechanisms for the mechanical stability of the cellulose fibril, is central to the physical science of biomass degradation.[8,15] The high structural ordering of natural cellulose fibrils largely owing to an extensive H-bond network between and within cellulose chains is a main biophysical origin of the recalcitrance.

Here, to investigate the thermostability of cellulose I_α, a model based on the H-bond network has been constructed. The calculations show that the two competing H-bond patterns experimentally observed at room temperature[10] exist throughout the whole temperature range before the onset of degradation. Whereas both patterns have equal probabilities

of occurrence at low temperatures, at intermediate temperatures, an increased probability of formation of pattern I and a corresponding decreased formation of pattern II are observed. This may lead to higher stability of the cellulose I_α fibril at intermediate temperatures because the pure inter-chain H-bond $O_6H_6 \cdots O_3/O_2$ of pattern I presumably contributes more to the stability of the fibril by "gluing" two chains together than does the partial inter-chain H-bond $O_2H_2 \cdots O_3/O_6$ of pattern II. The observed increase in the formation of pattern I is strongest at the position connecting the surface chains to the inner chains. This is of particular relevance as the outer chains are the most vulnerable to degradation. For a comparison, two competing patterns were also observed in the model of cellulose I_β.[15] However, in the case of I_β, one pattern (represented by inter-chain H-bond g), which is favored by energy, approaches 100% occupancy at the low temperature limit; while the other pattern (represented by inter-chain H-bond e+f) is more entropy driven and dominates over a high temperature range, until the onset of the disassembly.

To conclude, a transfer matrix method has been used to study the temperature-dependent changes of the H-bond network in cellulose I_α, which has not been attempted previously. Just like ice, cellulose occurs in multiple crystalline phases with distinct H-bond networks. The multitude of different phases and the occurrence of multiple H-bond patterns within each phase make the cellulose fundamentally interesting to chemical physicists, as well as of industrial importance. In contrast to I_β, in I_α, an alternating scheme of two distinct glucose conformations is found along a single chain. Moreover, I_α can form the partial inter-chain, intra-chain H-bond $O_2H_2 \cdots O_3/O_6$ which was not observed in I_β. These remarkable differences between cellulose I_α and I_β require a completely different setup compared to a previous work.[15]

The comparison of our results with available experimental data shows excellent agreement.[10,16] Particularly, the comparison of our theoretical model to continuous temperature IR spectra is important for model validation, and no such comparison has been reported before. The experimental measurement of changes in the absorption is an indirect probe of H-bonding and not all the H-bonds were individually resolved. Moreover, the observed H-bonds were not spatially resolved, so that all spectra have to be understood as spatially averaged. The present model calculates the absolute probability of formation of all individual H-bonds. The spatial resolution allows us to detect significant difference of bond formation at different positions. Furthermore, the model identified and allowed systematic investigation of intriguing, qualitative changes of the alternating pattern over a wide range of temperatures. Thus, our results go beyond previous findings and provide a microscopic explanation for the thermostability of cellulose I_α over a wide range of temperatures. They represent a significant step in the understanding of H-bonding in cellulose and will lead to new experimental and theoretical endeavors to further explore the H-bonding of these systems.

Some further aspects of the cellulose assembly could be examined within the framework of this model. First, as natural cellulose fibrils consist predominately of a mixture of the crystalline phases I_α and I_β,[3,13] and probably some amorphous regions, it would be interesting to expand the current model to a more realistic model of a natural fibril by incorporating the different phases. Further, only native H-bonds have been considered in the current model, and it would be interesting to include additional H-bond schemes from other crystalline forms of cellulose, e.g., cellulose II and III.[3] Furthermore, solvent effects were not incorporated into the current model. However, MD simulations of cellulose I_β have shown that water can compete with (interrupt) the H-bonds located on the surface of a fibril.[13,14] Therefore, explicit incorporation of solvent effects may be necessary for a more detailed characterization of natural cellulose systems.

ACKNOWLEDGMENTS

This work was funded by the U. S. Department of Energy, Scientific Discovery through Advanced Computing (SciDAC) program, and Office of Biological and Environmental Research (OBER) under FWP ERKJE84.

[1] A. E. Farrell, R. J. Plevin, B. T. Turner, A. D. Jones, M. O. Hare, and D. M. Kammen, Science **311**, 506 (2006).
[2] D. Klemm, B. Heublein, H.-P. Fink, and A. Bohn, Angew. Chem., Int. Ed. **44**, 3358 (2005).
[3] S. Dumitriu, *Polysaccharides: Structural Diversity and Functional Versatility*, 2nd ed. (Marcel Dekker, New York, 2005).
[4] U. S. DOE, International Energy Outlook 2009, DOE/EIA-0484(2009), U. S. Department of Energy: Energy Information Administration Office of Integrated Analysis and Forecasting, 2009.
[5] A. J. Ragauskas, C. K. Williams, B. H. Davison, G. Britovsek, J. Cairney, C. A. Eckert, W. J. Frederick, Jr., J. P. Hallett, D. J. Leak, C. L. Liotta, J. R. Mielenz, R. Murphy, R. Templer, and T. Tschaplinski, Science **311**, 484 (2006).
[6] C. E. Wyman, Trends Biotechnol. **25**, 153 (2007).
[7] U. S. DOE, Biomass as Feedstock for a Bioenergy and Bioproducts Industry: The Feasibility of a Billion-Ton Annual Supply, DOE/GO-102995-2135, ORNL/TM-2005/66, U. S. Departments of Energy and U. S. Department of Agriculture, 2005.
[8] U. S. DOE, Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda, DOE/SC-0095, U. S. Department of Energy Office of Science and Office of Energy Efficiency and Renewable Energy, 2006.
[9] M. E. Himmel, S.-Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady, and T. D. Foust, Science **315**, 804 (2007).
[10] Y. Nishiyama, J. Sugiyama, H. Chanzy, and P. Langan, J. Am. Chem. Soc. **125**, 14300 (2003).
[11] Y. Nishiyama, J. Sugiyama, H. Chanzy, and P. Langan, J. Am. Chem. Soc. **124**, 9074 (2002).
[12] B. J. Hardy and A. Sarko, Polymer **37**, 1833 (1996).
[13] J. F. Matthews, C. E. Skopec, P. E. Mason, P. Zuccato, R. W. Torget, J. Sugiyama, M. E. Himmel, and J. W. Brady, Carbohydr. Res. **341**, 138 (2006).
[14] T. Yui, S. Nishimura, S. Akiba, and S. Hayashi, Carbohydr. Res. **341**, 2521 (2006).
[15] T. Shen and S. Gnanakaran, Biophys. J. **96**, 3032 (2009).
[16] A. Watanabe, S. Morita, and Y. Ozaki, Biomacromolecules **8**, 2969 (2007).
[17] H. A. Kramers and G. H. Wannier, Phys. Rev. **60**, 252 (1941).
[18] E. W. Montroll, J. Chem. Phys. **9**, 706 (1941).
[19] P. J. Flory, *The Statistical Mechanics of Chain Molecules* (Hanser, Munich, 1988).
[20] P. J. Flory and R. L. Jernigan, J. Chem. Phys. **42**, 3509 (1965).
[21] B. H. Zimm and J. K. Bragg, J. Chem. Phys. **28**, 1246 (1958).
[22] B. H. Zimm, J. Chem. Phys. **33**, 1349 (1960).
[23] M. Ya. Azbel, Phys. Rev. A **20**, 1671 (1979).
[24] C. Kittel, Am. J. Phys. **37**, 917 (1969).
[25] IUPAC and IUB Committee, Pure Appl. Chem. **55**, 1269 (1983).
[26] R. H. Newman, Solid State Nucl. Magn. Reson. **15**, 21 (1999).
[27] E. Sjöström, *Wood Chemistry. Fundamentals and Applications*, 2nd ed. (Academic, San Diego, 1993).

[28]S. Kuga and R. M. Brown, Jr., in *Biosynthesis and Biodegradation of Cellulose*, edited by C. H. Haigler and P. J. Weiner (Marcel Dekker, New York, 1991), pp 125–142.
[29]Y. Marechal, *The Hydrogen Bond and the Water Molecule: The Physics and Chemistry of Water, Aqueous and Bio-Media*, 1st ed. (Elsevier, Amsterdam, 2007).
[30]G. A. Jeffrey, *An Introduction to Hydrogen Bonding* (Oxford University Press, New York, 1997).
[31]G. R. Desiraju and T. Steiner, *The Weak Hydrogen Bond In Structural Chemistry and Biology* (Oxford University Press, New York, 1999).
[32]J. N. Shapiro, Phys. Rev B **1**, 3982 (1970).
[33]E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen, *LAPACK Users' Guide*, 3rd ed. (Society for Industrial and Applied Mathematics, Philadelphia, 1999).
[34]E. M. Debzi, H. Chanzy, J. Sugiyama, P. Tekely, and G. Excoffier, Macromolecules **24**, 6818 (1991).
[35]M. Wada, J. Polym. Sci., Part B: Polym. Phys. **40**, 1095 (2002).
[36]M. Wada, T. Kondo, and T. Okano, Polym. J. **35**, 155 (2003).
[37]S. Morita, H. Shinzawa, I. Noda, and Y. Ozaki, Appl. Spectrosc. **60**, 398 (2006).
[38]See supplementary material at http://dx.doi.org/10.1063/1.3626274 for the construction of the elements of the transfer matrix in the two-chain case.