The Stability of Cellulose: A Statistical Perspective from a Coarse-Grained Model of Hydrogen-Bond Networks

Tongye Shen[†‡] and S. Gnanakaran[†*]
[†]Theoretical Biology & Biophysics Group, and [‡]Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico

ABSTRACT A critical roadblock to the production of biofuels from lignocellulosic biomass is the efficient degradation of crystalline microfibrils of cellulose to glucose. A microscopic understanding of how different physical conditions affect the overall stability of the crystalline structure of microfibrils could facilitate the design of more effective protocols for their degradation. One of the essential physical interactions that stabilizes microfibrils is a network of hydrogen (H) bonds: both intrachain H-bonds between neighboring monomers of a single cellulose polymer chain and interchain H-bonds between adjacent chains. We construct a statistical mechanical model of cellulose assembly at the resolution of explicit hydrogen-bond networks. Using the transfer matrix method, the partition function and the subsequent statistical properties are evaluated. With the help of this lattice-based model, we capture the plasticity of the H-bond network in cellulose due to frustration and redundancy in the placement of H-bonds. This plasticity is responsible for the stability of cellulose over a wide range of temperatures. Stable intrachain and interchain H-bonds are identified as a function of temperature that could possibly be manipulated toward rational destruction of crystalline cellulose.

INTRODUCTION

Biofuels derived from lignocellulosic biomass offer a cleaner and sustainable alternative to fossil fuels, with several advantages over ethanol derived from corn and sugar cane. Challenges encountered during the conversion of biomass to fuels are critically linked to uncertainties in the physical properties of the feedstock (1). In particular, a better understanding of cellulose and its biodegradation will allow the factors preventing an efficient fuel economy based on cellulosic biomass conversion to be addressed. A central problem is how cellulose, a polymer of glucose, can be effectively isolated and disassembled (2), to its basic building block, glucose.

For plants, cellulose is one of the major components of the resistance to external chemical, mechanical, or biological perturbations. The resistance of cellulose to depolymerization is due to the fact that it occurs as highly crystalline polymer fibers. In nature, the synergetic action of endoglucanase, exoglucanase, and β-glucosidase enzymes, collectively called cellulases, is able to slowly degrade the surfaces of cellulose fibers (3). Cellulose occurs naturally in plants in two crystalline forms, I-α and I-β (4,5). Their crystal structures suggest that hydrogen (H) bonding plays a key role in determining the properties of cellulose.

Here we investigate the role of H-bonding in I-β, the main form found in higher plants. Known crystal structures of cellulose show strong intrachain and interchain H-bonds that provide thermostability for the complex (4). Intrachain H-bonds raise the stiffness of each polymer, whereas an interchain H-bond network links chains together to form a two-dimensional sheet. In contrast, sheets are mainly packed together by weak van de Waals interactions. There are no H-bonds of type O-H…O between sheets for I-β (4). Thus, the interaction between adjacent sheets is much weaker compared to the strong hydrogen-bond interaction within one sheet, as demonstrated experimentally by the strong anisotropic thermal expansion property of this structure (6). Unlike polypeptides or polynucleotides, which have a set of uniquely specified H-bonds and other contact interactions between monomers for a biologically functioning structure, in cellulose I-β several distinct H-bonds can be formed. However, some of these bonds cannot coexist, which creates frustration for the favorable low energy state of this structure.

In this study, the stability of a sheet of chains in I-β is considered using a statistical description of H-bonding networks on a square lattice, as shown in Fig. 1. We will use this lattice model to answer questions such as what are the favorable H-bonding patterns and how they vary with rising temperature until disassembly. This theoretical model describes the state of the system as a collection of discrete representation of the state of each potential H-bonding position, i.e., each type of H-bond is either formed or not formed at a position. In such a binary description, when a bond is formed, we associate it with a gain in bonding energy. Similarly, a gain in entropy is assigned when H-bonds are absent for a particular bonding position. The specific gain in entropy also depends on its neighboring states. Finally, we can associate each particular state of the hydrogen-bond network with a statistical weight based on its free energy.

A major simplification of this work enables the evaluation of the partition function of this lattice model to obtain an exact solution. The interaction is limited to native H-bonds, i.e., only those potentially present in I-β. This simplification effectively limits the phase space of the system to either the assembled ordered I-β structure or a disassembled one. A

Submitted August 12, 2008, and accepted for publication December 31, 2008.

*Correspondence: gnana@lanl.gov

Editor: Kathleen B. Hall.

0006-3495/09/04/3032/9 $2.00

doi: 10.1016/j.bpj.2008.12.3953

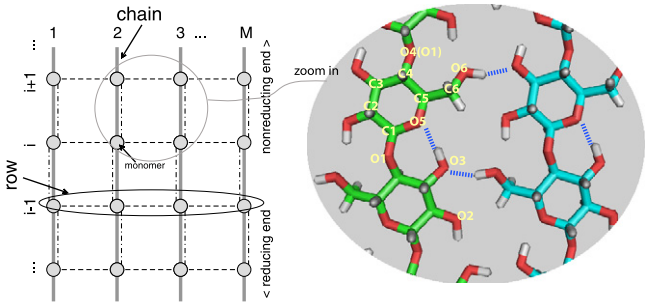

FIGURE 1 An illustration of the sheet structure of cellulose I-β. Cellulose chains (*solid vertical rods*) are linear collection of monomers (*solid circles*), which can be potentially linked using H-bonds residing intrachain positions (*vertical dot-dashed lines*) and interchain positions (*horizontal dashed lines*). The assembly (or disassembly) process is expressed by the formation (or disruption) of H-bonds at these positions. A small part (four monomers) of the sheet structure is viewed with an atomistic representation.

similar approximation of native-interaction-only, which ignores major misfolded structures, is often used to evaluate the partition function of large-scale motions of a protein's internal configurations, i.e., folded/disordered ones (7). For the cellulose assembly, this assumption is valid in the situation of small thermal agitation motion inside the native basin of the free energy landscape, up to the level that the system is about to hop out of the native basin, a stage leading to the onset of the disassembly of polymer chains. The strength of this model is that it can capture the transition from a low temperature phase of the H-bond network to a high temperature phase of H-bond disruption. Any possibility of forming H-bonds that are not in the I-β crystal structure basin or distortions of the network (due to defects in crystals) is completely ignored. Thus, it is out of the scope of this model when one wants to explore the transition from I-β to another ordered form such as antiparallel cellulose II.

Results obtained from the lattice model reported here suggest, instead of only one stable pattern, different H-bonding patterns dominate at different temperatures until disassembly at very high temperatures. Furthermore, intrachain H-bonding is more stable than the interchain H-bonding throughout the temperature range considered. This statistical model is also able to capture important equilibrium properties at the resolution of individual bonds. Recently, the structural microheterogeneities of crystalline I-β and associated complex hydrogen-bonding networks were investigated using an elegant atomistic model (8). The advantage of such atomistic simulations is that both native and nonnative hydrogen-bonding networks can be treated on equal footing. Some uncertainties in these structures are: whether the small size scale of the system properly represents the thermal behavior of the I-β bulk phase due to limited size with artificial boundary conditions; ongoing force-field issues of polysaccharides; and short simulation timescales. As shown in this study, using a partition-function-based statistical method to study the coarse-grained description of cellulose at a wide range of temperatures is a valuable complimentary approach to atomistic simulations.

MODEL AND METHOD

Similar to proteins and nucleic acids, polysaccharides are directional polymers. We follow the conventions recommended by the IUPAC-IUB Joint Commission on Biochemical Nomenclature (9) for naming atoms and defining the direction of cellulose. By this convention, each polymer starts with a reducing end (the end close to O_5, the ring oxygen) and terminates with a nonreducing end. For example, monomer $i + 1$ is located toward the nonreducing end relative to monomer i. For simplicity, each cellulose molecule is termed a chain. Further, a sheet is defined as the strongly H-bonded assembly of chains. The direction of the intermolecular H-bond is also termed the interchain direction, or the rung direction—a generalized phrase from the original ladder- or zipper-type of two-chain systems. A row is defined as a collection of monomers linked by interchain H-bonding positions. Finally, sheets stack together to form a three-dimensional cellulose microfibril. The crystal structure of cellulose I-β, has the following geometry features: ~5 Å between sequential monomers in a chain; ~8 Å between neighboring chains in a sheet; and ~4 Å between neighboring sheets. Typically, the width of microfibrils in woody plants is of 25–60 Å, which corresponds to 3–7 chains per sheet. However, there are larger microfibrils such as those obtained from ramie and tunicate (10,11).

As listed in Table 1, seven possible H-bonds (basic units of the H-bond network) can be deduced from neutron diffraction experiments on the I-β crystal structure (4). Classified by connectivity, four (a, b, c, d) are intrachain H-bonds and the other three (e, f, g) are interchain H-bonds. Four (c, d, e, f) are simple H-bonds (12) and the rest (a, b, g) are bifurcated H-bonds (one donor group contacts two acceptors simultaneously). It is important to point out that the microscopic nature of the bifurcated H-bond is still being debated. Although crystallographers provide us with such structures of one donor interacting with two acceptors, it has been questioned whether it represents a fast (picosecond timescale) switching between two single H-bonds (13). Our treatment is quite general and will not suffer from such alternative interpretations.

As mentioned previously, we have devised a set of forbidden rules since not all combinations of these H-bonds can be simultaneously formed. The statistical weight of the H-bonding network will be reduced to zero if any part of the network violates these rules:

1. For intrachain H-bonds, the simultaneous formation of the pair of H-bonds (a, b), (a, c), or (b, c) is forbidden.
2. For interchain H-bonds, the simultaneous formation of the pair (e, g) or (f, g) is not possible.
3. For the interrelation between intra- and inter-H-bonds, the forbidden pairs are listed in Table 2.

Note that only one intrachain H-bond of the two adjacent chains can potentially interfere with the interchain H-bonding that connects these two chains.

TABLE 1 All possible H-bonds (of the type OH$\cdots$O) in crystalline cellulose I-β

i	Position	Atoms (A$\cdots$H-D)	Nature/strength
a	intra	$O_6\cdots H_2$-O_2, $O_1\cdots H_2$-O_2	u_2
b	intra	$O_2\cdots H_6$-O_6, $O_1\cdots H_6$-O_6	u_2
c	intra	$O_6\cdots H_2$-O_2	u_1
d	intra	$O_5\cdots H_3$-O_3	u_1
e	inter	$O_3\cdots H_6$-O_6	u_1
f	inter	$O_6\cdots H_2$-O_2	u_1
g	inter	$O_3\cdots H_6$-O_6, $O_2\cdots H_6$-O_6	u_2

See article and Table 2 for the forbidden combinations of these elements to form a hydrogen-bond network. Energy of single and bifurcated bonds is set to be u_1 and u_2, respectively.

TABLE 2 H-bonding forbidden rules that capture the incompatible states of intrachain H-bonding given the specified states of interchain H-bonding

i\($i+1$)	ø	e	f	g	ef
ø	ø	ø	a,c	ø	a,c
e	b	b	a,b,c	b	a,b,c
f	a,c	a,c	a,c	a,c	a,c
g	b	b	a,b,c	b	a,b,c
ef	a,b,c	a,b,c	a,b,c	a,b,c	a,b,c

The row state on the table is the i^{th} (upstream) row of the sheet, while the column state is the $(i+1)^{th}$ (downstream) row of the sheet. The content is the struck intramolecular H-bond(s). Notice that only the intrachain H-bonding of one of the two chains will be affected. Here ø is the state without any H-bond.

All these rules naturally arise from the geometry of crystalline cellulose I-β and the nature of H-bonding. For example, between two hydroxyl groups, -O_xH_x and -O_yH_y, we can form either $O_xH_x\cdots O_y$ or $O_yH_y\cdots O_x$; while forming, one H-bond excludes the formation of the other.

As shown in Fig. 1, the topology of the system is essentially described by a square lattice. A lattice of size $M \times L$ consists of M chains of L glucose units each. If each monomer is labeled using index (m, l) with $m = 1, \ldots, M$ and $l = 1, \ldots, L$, except the ones at boundary, it can form interchain H-bonds with monomer $(l \pm 1, m)$, as well as form intrachain H-bonds with monomer $(l, m \pm 1)$. Since the number of chains per sheet, M, is generally small (<10), we explicitly describe all of the states of H-bonding in a row by a state vector and reduce the problem to a quasi-one-dimensional problem.

Each discrete state of the lattice will be assigned a free-energy value based on the status of the H-bonding positions described by that state. The free energy takes into account both the effects of bonding energy and conformational entropy changes associated with bonding. The energy parameters are set as the following: uniform bonding energies u_1 and u_2 are assigned for single H-bonds (c, d, e, f) and bifurcated H-bonds (a, b, g), respectively. We use the following scheme to assign entropic differences between different states of the network, since its value depends on the states of both inter- and intrachain H-bonds connecting to that monomer. Three entropy parameters are introduced to accommodate the entropy changes due to H-bonding: $s_|$ and s_- are linked to the basal value of entropy loss due to intra- and interchain H-bond formation, respectively; and $s_\times$ describes the intrinsic interplay between entropic changes of intra- and interchain H-bonding. The term associated with $s_\times$ alleviates the effect of double counting of the entropy gain from breaking bonds. For example, the entropy change associated with the formation of an intrachain H-bond should be smaller when neighboring interchain H-bonds are already formed. We associate $s_\times$ with the intrachain H-bond term without losing generality. Thus, we give the total intrachain H-bond entropy an extra term $\Delta s = [(4 - z)/4] \cdot s_\times$. Here Δs is simply assumed to be linear in z (= 0, 1, 2, 3, or 4), the number of interchain H-bonds formed at the nearest neighboring positions. We note that in addition to the term Δs, the rules for forbidden bonding patterns further contribute to the cooperativity among different H-bonds, making the statistical mechanics of this system more than a simple collection of independent individual H-bonds.

The transfer matrix method (14) is adopted to solve the current quasi-one-dimensional system with short-range interactions. It was initially developed for solving simple one-dimensional magnetic systems (15) and extended later to solve (bio)polymer configurations by the pioneering studies of many researchers (16–19). In addition to applications of transfer matrix to linear biomolecular systems, there are also applications of this method in the dynamic assembly aspect of biofiber problems (20).

For the cellulose system, the linear size, n, of the transfer matrix is that of the state vector describing the complete state of one row along the (quasi)-one-dimensional lattice. A two-chain system will have five possible interchain H-bonding states: ø, e, f, g, or e+f for each row, hence, $n = 5$. Here, state ø means none of the H-bonds is present, state e means only H-bond e is formed, and state e+f means both e and f are simultaneously present. The order of states ({ø, e, f, g, e+f}) is adopted throughout the following descriptions of the transfer matrix. For an M-chain system, the state vector is spanned by a direct product of $M - 1$ interchain H-bonding positions, i.e., $n = 5^{M-1}$. Each of the states $i = 1, 2, \ldots, n$ describes a specific and complete configuration of a row.

The transfer matrix $\mathbf{T}$ of size $n \times n$ describes all of the statistical weights of how one connects the state of one row to the next. It is composed of two parts, $\mathbf{T} = \mathbf{UV}$. Here $\mathbf{U}$ is an on-site, diagonal matrix that describes the statistical weight owing to interchain interactions of one row of the sheet, and $\mathbf{V}$ is the matrix that describes the interactions along the chain direction. The on-site matrix for a two-chain system ($M = 2$) is

$$\mathbf{U}^{(2)} = \mathrm{Diag}\left[e^{s_-}, e^{-\beta\varepsilon_e}, e^{-\beta\varepsilon_f}, e^{-\beta\varepsilon_g}, e^{-\beta(\varepsilon_e+\varepsilon_f)}\right]. \quad (1)$$

Again, we can easily generalize $\mathbf{U}$ to the case of multichain sheet by direct product, e.g., $\mathbf{U}^{(3)} = \mathbf{U}^{(2)} \otimes \mathbf{U}^{(2)}$.

Here $\mathbf{V}$ assigns the statistical weights of the transition from one row to the next, containing nonzero off-diagonal elements. Due to its complicated expression, matrix $\mathbf{V}$ is explicitly shown only for the case $M = 2$, where a computer was used to enumerate the states and generate $\mathbf{V}$ for multichain cases. For simplicity, several shorthand notations are introduced before the explicit expression of $\mathbf{V}^{(2)}$. A useful expression of the partition function, $G(s)$, describing one intrachain H-bonding position without having any of potential intrachain H-bonding states eliminated, is defined as

$$\begin{aligned} \mathrm{G(s)} = {} & \mathrm{e^s} + \mathrm{F_+(\varepsilon_a)} + \mathrm{F_+(\varepsilon_b)} + \mathrm{F_+(\varepsilon_c)} \\ & + \mathrm{F_+(\varepsilon_d)} + F_+(\varepsilon_a + \varepsilon_d) + F_+(\varepsilon_b + \varepsilon_d) \\ & + F_+(\varepsilon_c + \varepsilon_d), \end{aligned} \quad (2)$$

where $F_+(x) = \exp(-\beta x)$, and the entropy s associated with the state depends on its neighboring interchain H-bonding state. Further, we define a function

$$\begin{aligned} Y^z_{\mathrm{i_a,i_b,i_c}} = {} & \left[e^{s_| + (4-z)/4\cdot s_\times} + F_+(\varepsilon_d) + F_{\mathrm{i_a}}(\varepsilon_a)\right. \\ & + F_{\mathrm{i_a}}(\varepsilon_a + \varepsilon_d) + F_{\mathrm{i_b}}(\varepsilon_b) + F_{\mathrm{i_b}}(\varepsilon_b + \varepsilon_d) \\ & \left. + F_{\mathrm{i_c}}(\varepsilon_c) + F_{\mathrm{i_c}}(\varepsilon_c + \varepsilon_d)\right] \cdot G\big(s_| \\ & + (4 - z)/4 \cdot s_\times\big). \end{aligned} \quad (3)$$

Here i_a, i_b, and i_c are either $-$ or $+$, and the subscript indicates that the H-bond element is forbidden and reduces the partition function to $F_- = 0$. Matrix $\mathbf{V}$ can finally be expressed as

$$\mathbf{V}^{(2)} = \begin{pmatrix} Y^0_{+++} & Y^1_{+++} & Y^1_{-+-} & Y^1_{+++} & Y^1_{-+-} \\ Y^1_{+-+} & Y^2_{+-+} & Y^2_{---} & Y^2_{+-+} & Y^2_{---} \\ Y^1_{-+-} & Y^2_{-+-} & Y^2_{-+-} & Y^2_{-+-} & Y^2_{-+-} \\ Y^1_{+-+} & Y^2_{+-+} & Y^2_{---} & Y^2_{+-+} & Y^2_{---} \\ Y^1_{---} & Y^2_{---} & Y^2_{---} & Y^2_{---} & Y^2_{---} \end{pmatrix}. \quad (4)$$

Matrix $\mathbf{V}$ has an intrinsic description of the upstream and downstream directions of the microfibril, and is generally asymmetric. Once the transfer matrix is specified, one can compute and express all equilibrium properties of the system with the elements of $\mathbf{T}$. At the limit of chain length $L \to \infty$, the bulk properties of cellulosic microfibrils can be simply described by the information contained in the several largest eigenvectors of asymmetric matrix $\mathbf{T}$. The right eigenvectors of $\mathbf{T}$ form a matrix $\mathbf{A}$, and the left eigenrow forms $\mathbf{A}^{-1}$. Together, $\mathbf{T}$ can be diagonalized with $\mathbf{A}^{-1}\mathbf{TA} = \mathbf{\Lambda}$.

After some computation, one can show (18) that the pair state probability is

$$\mathrm{p_{\mu,v}} = \mathrm{T_{\mu,v}A_{v1}}\left[\mathrm{A^{-1}}\right]_{1\mu}/\lambda_1, \quad (5)$$

where μ refers to one particular state of the total n possible states for the upstream row of the sheet and ν refers to another particular state for the downstream row. Here, λ_1 is the largest eigenvalue of $\mathbf{T}$. The value $p_{\mu\nu}$ is the equilibrium joint probability distribution between two neighboring rows, not a conditional distribution. It satisfies the following sum rules: $\sum_{\mu,\nu} p_{\mu\nu} \equiv 1$ and $\sum_{\mu} p_{\mu\nu} = \sum_{\mu} p_{\nu\mu}$.

The details of the fractions of intrachain and interchain H-bonds formed can be further computed based on $p_{\mu\nu}$ and individual bond energies. Since the state of each interchain H-bond is explicitly labeled by the subscript in $p_{\mu\nu}$, one can obtain the J^{th} individual interchain H-bond in the state i (i = ø, e, f, g, or e+f) with the probability

$$P^{\mathrm{E}}_{\mathrm{Ji}} = \sum_{\mu\nu} p_{\mu\nu} \cdot \delta_{\mu,\mathrm{Ji}}. \tag{6}$$

Here $\delta_{\mu,\ \mathrm{Ji}} = 1$ if bonding position J is in state i for the row state μ, and 0 otherwise. Equivalently, one can define $P^{\mathrm{E}}_{\mathrm{Ji}} = \sum_{\mu\nu} p_{\mu\nu} \cdot \delta_{\nu,\mathrm{Ji}}$. The calculation of the intrachain H-bonds is slightly more complicated, since the states of individual intrachain H-bonds are not directly shown in $p_{\mu\nu}$. However, for a specified pair state of μ and ν, i.e., the states of all the interchain H-bonds in two adjacent rows, any intrachain H-bond is formed with a probability that can be enumerated according to its weight, i.e., $q_{\mathrm{Ji},\ \mu\nu} \propto \exp(-\beta\varepsilon_{\mathrm{Ji},\ \mu\nu})$. Here i = ø, a, b, c, d, a+d, b+d, or c+d. Together we have for the probability of the J^{th} intrachain H-bond in status i,

$$P^{\mathrm{A}}_{\mathrm{Ji}} = \sum_{\mu\nu} p_{\mu\nu} \cdot q_{\mathrm{Ji},\mu\nu}. \tag{7}$$

One particularly relevant property to compute is the communal probability of a specific bonding position, which is defined as the chance of forming any H-bond (regardless of type) at that position. It is also the complement of the probability that none of the H-bonds is formed, i.e.,

$$Q^{\mathrm{A/E}}_{\mathrm{J}} = \sum_{\mathrm{i} \neq \phi} P^{\mathrm{A/E}}_{\mathrm{Ji}} = 1 - P^{\mathrm{A/E}}_{\mathrm{J}\phi}. \tag{8}$$

Throughout the article, superscripted E and A refer to inter- and intra-, respectively. Beyond the properties purely obtained from the largest eigenvector of $\mathbf{T}$, we have also calculated the two-point correlation length, ξ, of the H-bond network, along the chain direction. It is defined by the ratio of the first two largest eigenvalues of $\mathbf{T}$, i.e., $\xi = -1/\ln(|\lambda_2|/\lambda_1)$.

RESULTS AND DISCUSSIONS

In the following subsections, we will first present the results for a two-chain cellulose system. At this level, the composition of individual H-bonding elements is discussed. Later, we present the results at the level of a multichain sheet, a complex assembly with fewer well-studied models. At the multichain level, we focus on 1), how the basic features introduced for two-chain systems vary with the total number of chains; and 2), how these properties behave as functions of individual chain indices. It is also important to realize that all the results using parameter $s_\times = 0$ for $M = 2$ will not change for $M > 2$. Thus, these calculations can be treated as the mean-field results that are not sensitive to the number of chains.

In this study, the energy unit is kcal/mol. The temperature unit is kcal/mol/k_{B}, which equals 503 K (230°C), and the corresponding entropy unit is the Boltzmann constant k_{B}. Based on various studies (13), we assume that H-bond formation lowers energy $u_1 = u_2 = 5$ kcal/mol throughout this article, unless explicitly specified otherwise. This value corresponds to the lower boundary of the intermediate strength of hydrogen bonding.

The estimation of entropy parameters presents more of a challenge. A simple estimation of the entropy loss associated with the localization of the (center-of-mass) position of each monomer is given as $\Delta s_{\mathrm{t}} \simeq \ln(d \times A/\delta v)$. Here d is the spacing between the sheets, and A is the area of the unit-cell encompassing a glucose, while volume δv is a value directly linked to the scale of the vibrational motion. For the Lindemann melting criterion with critical fraction = 0.1 (21), we have $\Delta s_{\mathrm{t}} \simeq 6.9$. Similarly, the localization of the rotational motion of each torsional angle can be given $\Delta s_{\mathrm{r}} \simeq \ln(2\pi/\Delta\theta)$. Taken together, the total entropy difference between the crystalline state and the onset of the disordered form is ~20 k_{B} per glucose unit. Since the entropy parameters are, intrinsically, functions of the spacing between sheets, a part of the three-dimensional nature of the assembly enters indirectly through our implementation of entropy as well.

Two-chain system ($M = 2$)

The model of a two-chain polysaccharide assembly has some similarities to the zipper model of dsDNA melting (16) and to the helix-coil transition of peptides (19). The major difference here is that cellulose can exhibit many different H-bonds at each edge of the lattice topology. Additionally, the dynamic interplay between interchain H-bonding and intrachain H-bonding is lacking in the nucleic acid systems.

We first consider the communal probabilities as functions of temperature. For $M = 2$, each row consists of two intrachain H-bonding positions and one interchain H-bonding position. The communal probabilities for a two-chain system are $Q^{\mathrm{E}} = 1 - P^{\mathrm{E}}_{\ \phi}$ and $Q^{\mathrm{A}} = \sum_{\mu\nu} p_{\mu\nu}[1 - (q_{1\phi,\mu\nu} + q_{2\phi,\mu\nu})/2]$. Due to the twofold helical (planar zigzag) structure within each chain, these two intrachain H-bonding positions switch every row and we only report the mean value for Q^{A}. The decreases in communal probabilities of intrachain H-bonding (*shaded*) and interchain H-bonding (*solid*) with rising T are shown in Fig. 2. In the low T region, there is a high probability of forming the native H-bond network, whereas virtually none of the H-bonds is formed at very high T, and the assembly no longer exists. As shown in Fig. 2 *a*, two shaded temperature areas (from low to high) indicate the critical temperature for breaking interchain H-bonding and intrachain H-bonding, respectively. There are also two minor events of synchronization, shown by a small drop in the probability of the complimentary H-bond accompanying the major event of breaking. Hence, these two types of H-bonding are mutually supportive of each other to a lesser degree. Overall, the rupture of H-bonds occurs at a very high temperature, ~500 K. This melting temperature is a measure of the strong stability of this assembly and an indication of the rare probability of H-bond breaking at room temperature.

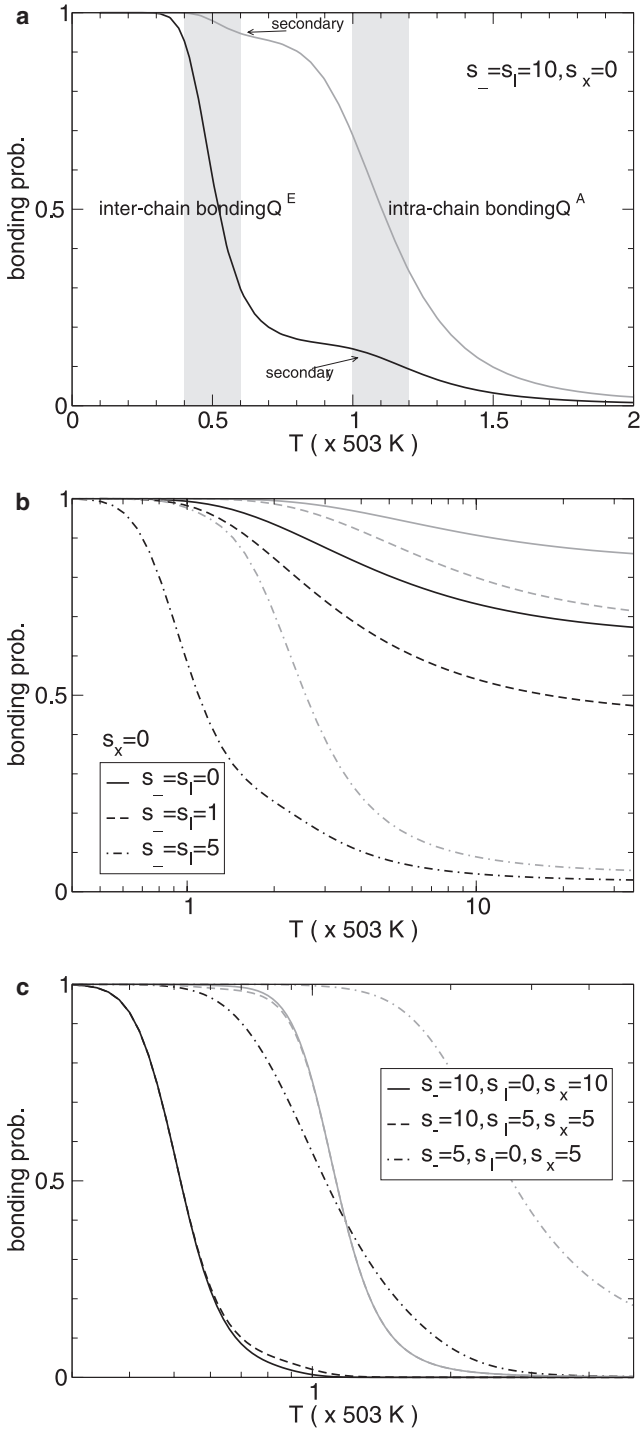

FIGURE 2 (*a*) For a two-chain system ($M = 2$), the intrachain (*shaded*) and interchain (*solid*) H-bonding probabilities are shown as functions of T. (*b* and *c*) Effects of entropy parameters.

We further test the robustness of the above results to changes in the parameters responsible for the loss in entropy due to the formation of interchain and intrachain H-bonds. As shown in Fig. 2 *b*, when the entropy parameters are decreased or eliminated, a finite population of H-bonds are formed even at the high T limit. An increase in sharpness of the disassembly transition is seen with the inclusion of these entropic terms. Additionally, Fig. 2 *c* shows a more complicated effect due to assigning the entropy to intrachain H-bonding using $s_\times$ versus $s_|$. For a given s_- and a fixed amount of $s_\times + s_|$ as a normalization, when the parameter $s_\times$ is increased, intrachain H-bonding becomes stronger. At the same time, interchain H-bonding becomes weaker. This conclusion can be drawn from a careful analysis of the Boltzmann factors for different scenarios. Indeed, as shown in Fig. 2 *c*, the results from comparison of two cases that both preserve $s_| + s_\times = 10$, solid line, ($s_| = 10$ and $s_\times = 0$); and dashed line, ($s_| = 0$ and $s_\times = 10$), confirm the above statement. In all of the cases studied here, intrachain H-bonding is always more stable and has a higher chance of being formed at any given temperature. On the other hand, the corresponding interchain H-bond will be the first to break with increasing temperature.

Next, we examine the probabilities of specific microscopic states of H-bonding for the two-chain system. First we consider the individual intrachain H-bonds for the entropy parameters $s_- = s_| = 10$ and $s_\times = 0$. There are eight intrachain states (a, b, c, d, a+d, b+d, c+d, ø). These states are mutually exclusive at a particular intrachain H-bonding position. For example, state d means only H-bond d is formed while no other intrachain H-bond is present. The probabilities of these states, P_i^A, are shown by the solid lines as functions of T in Fig. 3, *a* and *b*, in normal and logarithmic scales, respectively. Note that the sum of these probabilities is unity.

An interesting result is that different H-bonding states tend to dominate at different temperatures, instead of just one specific populated state of H-bonding. The states a+d and c+d are the most populated states at low T. On the contrary, the population of the state b+d or d increases with temperature until the onset of disassembly. There is a complete coincidence between curves p_{a+d} and p_{c+d}, indicating equal population between them, regardless of temperature. Similarly, there is a degeneracy between p_a and p_c, though the latter pair has values very close to zero, and thus not shown clearly in Fig. 3 *a*. These degeneracies are due to the symmetry of forbidden rules and the energy parameters used. When the H-bonding energy of bifurcated H-bonds, u_2, is reduced from 5.0 to 4.9 kcal/mol, state c+d becomes the sole dominating state at low temperature, as shown by dotted lines in Fig. 3 *a*.

The plasticity of the hydrogen-bond network is also evident from these results. It appears that the multiplicity of H-bonds acts as a buffer to maintain the consistent stability of the assembly over a wide range of temperatures. The microscopic picture of the transition from dominating states a+d and/or c+d to the state of b+d can be considered as the participating atoms switching their roles as donor and acceptor. Alternatively, it can be visualized as the direction of H-bond for the two (intrachain) hydroxyl groups O_2H_2 and O_6H_6 changing. Also shown as dashed lines in Fig. 3 *b* is a case with another set of entropy parameters demonstrating that the above

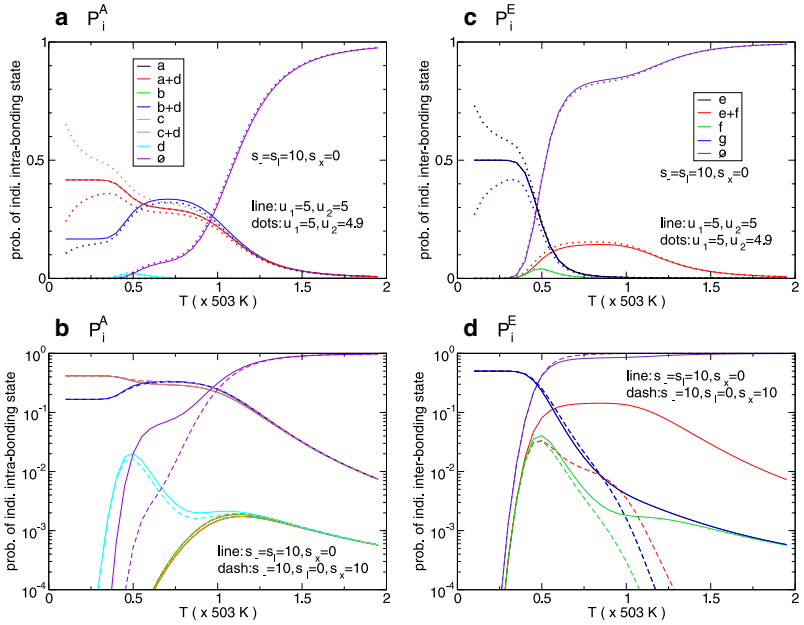

FIGURE 3 (*a* and *b*) For a two-chain system ($M = 2$), the probabilities of individual states (*different colors*) of intrachain H-bonding as a function of T, shown for three sets of parameters. (*Lines*) Results for $u_1 = u_2 = 5$, $s_- = s_| = 10$, $s_\times = 0$; (*dots*) $u_1 = 5$, $u_2 = 4.9$, $s_- = s_| = 10$, $s_\times = 0$; and (*dashes*) $u_1 = u_2 = 5$, $s_- = s_\times = 10$, $s_| = 0$. (*c* and *d*) The corresponding results for states of interchain H-bonds. Panels *b* and *d* are displayed in logarithmic scale for a better view of the region of low probability.

observation is relatively robust and not a peculiar effect of the entropy parameter chosen in Fig. 3 *a*. The H-bonding element d, $O_5\cdots H\text{-}O_3$, is the most independent since it does not participate in interactions with any other H-bonds, and only weakly contributes to the cooperative H-bonding network through the entropy term, $s_\times$.

Interchain H-bonding states also exhibit similar features to those observed in the case of intrachain H-bonds. Completely parallel to Fig. 3, *a* and *b*, for the intrachain H-bonding states, we show the corresponding results of interchain H-bonding for the same sets of parameters. In Fig. 3, *c* and *d*, the probability of interchain H-bonding for states (e, f, g, e+f, ø) as a function of temperature is shown (*solid lines*). Here curves p_e and p_g are degenerate. Again, when the degeneracy is broken by the assignment of the parameters $u_1 = 5 > u_2 = 4.9$ (*dots*), p_e approaches 1 at low T. With rising temperature, there is a crossover involving the most populated state of H-bonds, from state e or g to the state e+f. One may wonder why e+f is not the most dominating at low T. After all, simultaneous formation of e and f gains the most interchain H-bonding energy, 10 kcal/mol. As it turns out, the presence of H-bonds e and f will eliminate the possibility of forming intrachain H-bond a, b, or c in one of the two chains. Hence, it is not favored overall at low temperatures. However, at transient temperatures, H-bonds a, b, and c are less likely to form, and consequently we see a rise in the population of e+f.

Again, the compensation or buffering mechanism to maintain a consistent interchain H-bonding network is observed until disassembly at high temperature. Specifically, H-bonding element e, ($O_3\cdots H_6\text{-}O_6$), is present in both low (in state e) and high (in state e+f) temperatures. Therefore, we conclude that element e is the most stable interchain H-bond. The major shift of the H-bonding pattern with increasing T is associated with the switch from $O_2\cdots H_6\text{-}O_6$, one of two H-bonds in g to $O_6\cdots H_2\text{-}O_2$ in f. Thus, the overall picture is $O_3\cdots H_6\text{-}O_6$, which is persistently stable until disassembly. The H-bond between two (interchain) hydroxyl groups, $O_6\text{-}H_6$ and $O_2\text{-}H_2$, changes direction of the donor and the acceptor with rising temperature until disassembly. This movement echoes the corresponding switching motions of the intrachain H-bonds.

Multiple-chain sheet ($M > 2$)

For cellulose sheets made up of more than two chains, the chains at the edges will naturally have higher entropy values as they have only a single neighboring chain. This effect will propagate and give each chain a different value of stability depending on its position in the sheet. As mentioned in Model and Method, it is important to set $s_\times \neq 0$ to ensure that the statistics of H-bonding networks remains a cooperative one, rather than that of a collection of independent chains. Otherwise, the results of a multichain sheet will not depend on the chain index. A combination of nonzero $s_\times$ and the natural boundary condition of H-bonding ensures that each chain sense its environment.

We first show the mean of intra- and interchain communal probabilities as functions of T for systems with different numbers of chains, $2 \leq M \leq 6$, in Fig. 4. The entropy parameters are $s_- = 5$, $s_| = 0$, and $s_\times = 10$. For the multichain case, there are $M - 1$ interchain positions as well as M intrachain positions specified in the transfer matrix. The results are qualitatively consistent among sheets with different M. With increasing M, the stability gradually increases for a given T. However, the increment gradually decreases with increasing M, i.e., the difference between $M = 2$ and $M = 3$ is much larger than that of $M = 5$ and $M = 6$. This trend quickly converges for

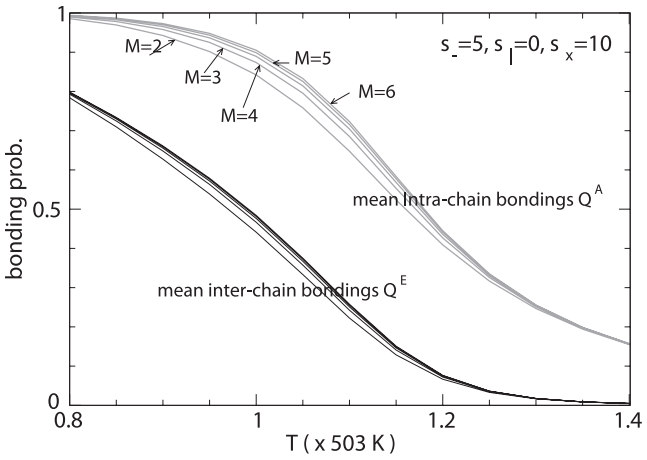

FIGURE 4 The average probability of H-bonding over chains, Q^{A} (*shaded*) and Q^{E} (*solid*) are shown as functions of temperatures for a set of sheets with total chain number, $M = 2 - 6$.

large M and/or high T and validates the use of a model with just a few chains.

Next, we consider the composition of H-bonding among different chains at a given temperature. The results for a six-chain system (under the same condition as those used in Fig. 4) are shown in Fig. 5 at two discrete temperatures, $T = 1.0$ and $T = 1.2$. The communal probabilities for individual positions: six intrachain H-bonding, Q^{A}_{J} ($J = 1,\ldots,6$), and five interchain H-bonding, Q^{E}_{J} ($J = 1,\ldots,5$), are displayed. As expected, the H-bonding of chains in the middle of the sheet is more stable than those at the edges. The H-bonding probabilities are sharply decreased at the edges, whereas the central chains show almost uniformly high values.

We also subjected the system to a sensitivity study on the chosen energy parameters. Two sets of perturbations were used. One family of parameters, controlled by R, scales the relative strength of single versus bifurcated H-bonds. The strength of single H-bonds c, d, e, and f is amplified by R, $u_1(R) = u_1 \times R$, whereas that of the bifurcated H-bonds a, b, and g is decreased by the same ratio, $u_2(R) = u_2/R$. The second family of parameters, controlled by R', tunes the relative strength of intra- versus interchain H-bonding. Thus, energies of a, b, c, and d are amplified by R' while e, f, and g are decreased by R'. The results for a six-chain system at $T = 1$ are shown in Fig. 6. The entropy parameters are the same as those in Fig. 5. With increasing R', Q^{A} is monotonously increasing whereas Q^{E} is monotonously decreasing. Since the intrachain H-bonding is more stable than that of the interchain H-bonding ($Q^{A} > Q^{E}$) during the range of R' considered, such a conclusion is quite robust. The nonmonotonous behavior of $Q^{E/A}(R)$ is another evidence of the complexity of the network. Still, the conclusion that $Q^{A} > Q^{E}$ is robust within the range of R considered here.

Finally, the dependence of the correlation length, a property contained beyond the largest eigenvector, on the total number of chains and temperature is considered. The correlation length, ξ, as a function of T, is shown in Fig. 7. Here, ξ measures the size scale of the fluctuation, i.e., the size of a local melted region or a local orderly assembled region. We see that, for a system with the energy degeneracy $u_1 = u_2$, a convergent and finite correlation length can be obtained at the low T limit. As expected, with the breaking of the degeneracy (here again, we set $u_1 = 5 > u_2 = 4.9$), the correlation length diverges to infinity with decreasing temperature. Keeping this asymmetrical energy setup and with increasing chain number M from two through four, ξ is slightly increased for the high T region. It appears that the correlation length is not greatly affected by varying M for the low T phase. Throughout the physically relevant temperature region, we observe that the correlation length is short,

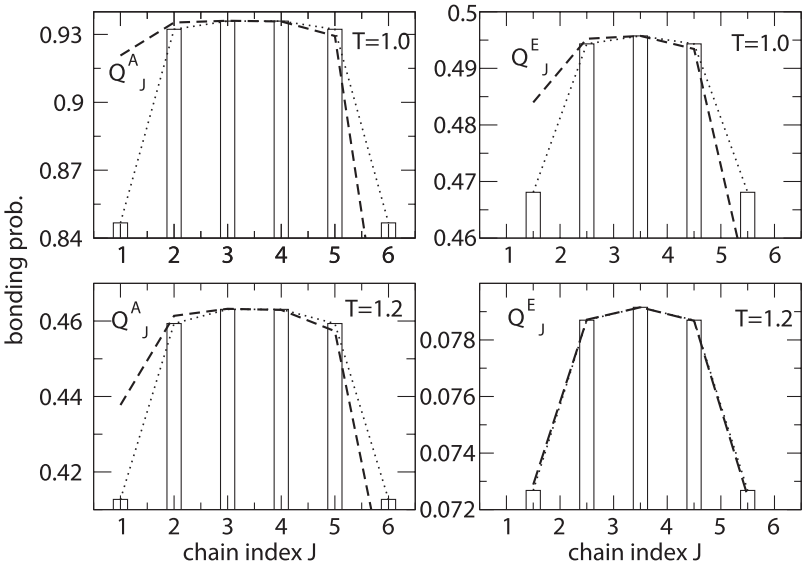

FIGURE 5 The probabilities of intrachain (*left panels*) and interchain (*right panels*) H-bonding for a six-chain sheet with individual chain resolution at $T = 1.0$ (*upper panels*) and $T = 1.2$ (*lower panels*). The entropy and energy parameters are set the same as the previous figure. The symmetrized values are shown with bars linked by dots, whereas the original sets are shown in dashes.

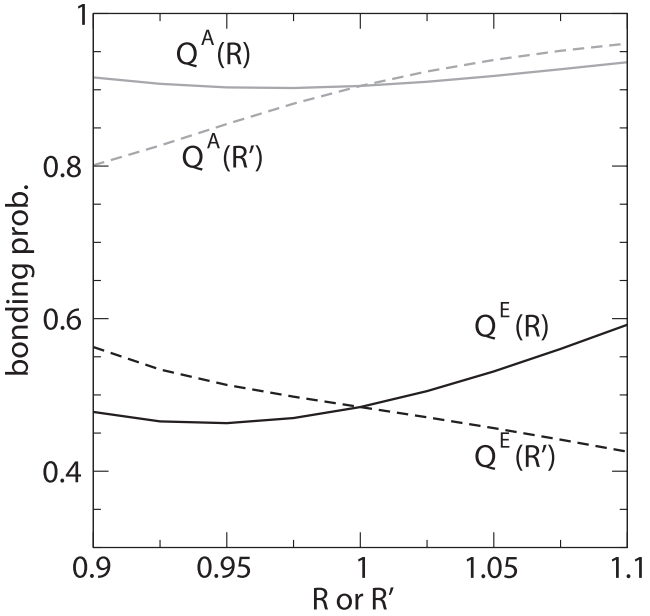

FIGURE 6 The mean probabilities of H-bonding of intrachain Q^{A} (*shaded*) and interchain Q^{E} (*solid*) for a six-chain sheet at $T = 1.0$ as functions of varying ratio parameter R (*solid lines*), which controls the relative strength of single versus bifurcated H-bonding, and R' (*dashes*), which controls the relative strength of intrachain H-bonding versus interchain H-bonding.

an indication that the fluctuations of H-bonding and disruption of H-bonds are isolated to specific rows (sites) rather than regions of many rows along the polymer.

CONCLUSIONS

Understanding the mechanical stabilities of polymer assemblies maintained by hydrogen-bond networks is important to various problems, from the efficiency of biomass degradation (22) to the design of biomimetic nanocomposites (23). In this study, cellulose, one of the most stable biomechanical structures constructed by nature, has been studied using a lattice model of the H-bonding networks. The underlying stability comes from the dense H-bond network constructed among the crystalline-ordered polysaccharide chains. An interesting and unique feature of this assembly is the multiple possibilities of H-bonding patterns within a single stable structure of cellulose. The mutual exclusion rules of some of these H-bonds listed in Model and Method prohibit them from forming simultaneously.

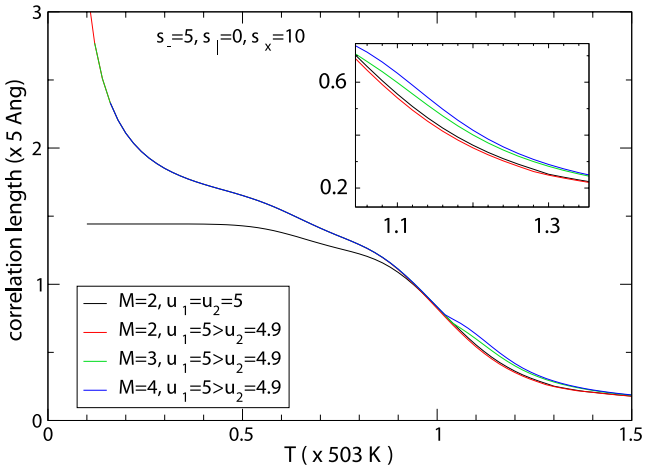

FIGURE 7 The correlation length (along the polymer direction) as a function of T for several sets of energy parameters and system sizes. (*Inset*) Zoom-in view of a part of the plot.

Our calculations suggest that such plasticity of H-bonding networks (owing to frustration and seeming redundancy of the H-bonding elements) keeps the overall H-bonding stability persistent over a large range of temperatures. Specifically, we found that intrachain H-bond $O_3H_3\cdots O_5$ is always stable until disassembly. Intrachain H-bond $O_2H_2\cdots O_6$, and interchain H-bonds $O_6H_6\cdots O_3$ and $O_6H_6\cdots O_2$, are stable at low temperature. However, with increasing temperature, the H-bonding network undergoes a switch in the connectivity of H-bonds. A role exchange between donor and receptor for the two hydroxyl groups O_6H_6 and O_2H_2 is observed. As a result, H-bond $O_2H_2\cdots O_6$ switches off as an intrachain H-bond and shows up as a interchain H-bond (H-bond $c \rightarrow f$).

Our findings on the thermal response of H-bonding networks are consistent with those obtained using simulations and experimental methods. Previous atomistic MD simulations in which crystalline cellulose was considered at two discrete temperatures (24) also found the appearance of H-bond f at high temperature, consistent with our calculation. Our predictions are also qualitatively consistent with the conclusions drawn from temperature-dependent infrared spectroscopy measurements (25,26).

There are several further interesting aspects of cellulose systems that could be examined in the future. Throughout this article, we have focused on a model of hydrogen-bonding network that can only be formed between the cellulose polymers in the crystal structure of I-β. However, cellulose can exist in other crystalline forms (I-α, II, III), and different types of intrinsic disorders may also appear (27,28). The method reported here could be easily applied to other crystalline phases and may even be extended, by an analogy to the random Ising model, to disorder cellulose.

It could be extended beyond native-only H-bond interaction by incorporating additional H-bonding schemes from all-atom models (8). At this time, the effect of the aqueous environment is also not explicitly treated. Based on previous all-atom simulation studies (29,30), the water media interacts with the outside layer of the cellulose microfibril and makes the H-bond network of the cellulose at the boundary much weaker. One possibility is to explore the solvent influence by altering the boundary condition of the current model.

Finally, we want to point out that similar lattice-partition-function-based theoretical methodologies can be used to address a wide range of problems in supramolecular

assembly in general (and especially for other biological microfibrils), where the intra- and intermolecular H-bonds are extremely important for stability (31). For example, collagen, the most abundant eukaryotic protein, also forms stable microfibril structures (11). By tuning the level of hydroxylation of its proline residues (adding extra hydroxyl group) and thus changing the total number of intermolecular H-bonds, collagen is able to alter its thermal stability (32). The current coarse-grained methods can be easily extended to study these situations as well.

In summary, we have provided a microscopic explanation on how cellulose is able to exhibit such a stability and plasticity over destabilizing environments. With the help of a new lattice-based model, we capture the plasticity of the H-bonding network in cellulose due to frustration and redundancy in the replacement of H-bonds that give rise to stability over a wide range of temperatures. These findings are facilitated by construction of a statistical mechanical model at the resolution of explicit H-bonds that take into account both intrachain and interchain H-bonds. This theoretical work is also of great relevance to pretreatment and enzyme degradation processes that represent the major cost hurdles to developing biofuels from lignocellulosic biomass. In that regard, we are able to identify stable interchain and intrachain H-bonds that can be manipulated toward rational destruction crystalline cellulose.

We thank Drs. P. Langan, G. P. Johnson, J. Gubernatis, C. J. Unkefer, and A. Redondo for helpful discussions.

This work was supported in parts by an LANL-LDRD grant from the United States Department of Energy and Center for Nonlinear Studies.

REFERENCES

1. Wyman, C. E. 2007. What is (and is not) vital to advancing cellulosic ethanol. *Trends Biotechnol.* 25:153–157.
2. Klemm, D., B. Heublein, H. -P. Fink, and A. Bohn. 2005. Cellulose: fascinating biopolymer and sustainable raw material. *Angew. Chem. Int. Ed.* 44:3358–3393.
3. Bayer, E. A., H. Chanzy, R. Lamed, and Y. Shoham. 1998. Cellulose, cellulases and cellulosomes. *Curr. Opin. Struct. Biol.* 8:548–557.
4. Nishiyama, Y., P. Langan, and H. Chanzy. 2002. Crystal structure and hydrogen-bonding system in cellulose I-β from synchrotron x-ray and neutron fiber diffraction. *J. Am. Chem. Soc.* 124:9074–9082.
5. Nishiyama, Y., J. Sugiyama, H. Chanzy, and P. Langan. 2003. Crystal structure and hydrogen bonding system in cellulose I-α from synchrotron x-ray and neutron fiber diffraction. *J. Am. Chem. Soc.* 125:14300–14306.
6. Wada, M. 2002. Lateral thermal expansion of cellulose I-β and III–I polymorphs. *J. Polym. Sci. [B].* 40:1095–1102.
7. Bryngelson, J. D., and P. G. Wolynes. 1987. Spin glasses and the statistical mechanics of protein folding. *Proc. Natl. Acad. Sci. USA.* 84:7524–7528.
8. Mazeau, K. 2005. Structural micro-heterogeneities of crystalline Iβ-cellulose. *Cellulose.* 12:339–349.
9. IUPAC and IUB 1983. Symbols for specifying the conformation of polysaccharide chains. *Pure Appl. Chem.* 55:1269–1272.
10. Jakob, H. F., D. Fengel, S. E. Tschegg, and P. Fratzl. 1995. The elementary cellulose fibril in *Picea abies*: comparison of transmission, electron microscopy, small-angle x-ray scattering, and wide-angle x-ray scattering results. *Macromolecules.* 28:8782–8787.
11. Fratzl, P. 2003. Cellulose and collagen: from fibers to tissues. *Curr. Opin. Coll. Int. Sci.* 8:32–39.
12. Pauling, L. 1960. The Nature of the Chemical Bond, 3rd Ed. Cornell University Press, Ithaca, NY.
13. Marechal, Y. 2007. The Hydrogen Bond and the Water Molecule: The Physics and Chemistry of Water, Aqueous and Bio-Media Elsevier, Amsterdam, The Netherlands.
14. Baxter, R. J. 1982. Exactly Solved Models in Statistical Mechanics Academic Press, San Diego, CA.
15. Kramers, H. A., and G. H. Wannier. 1941. Statistics of the two-dimensional ferromagnet. Part I. *Phys. Rev.* 60:252–262.
16. Zimm, B. H., and J. K. Bragg. 1958. Theory of the one-dimensional phase transition in polypeptide chains. *J. Chem. Phys.* 28:1246–1247.
17. Flory, P. J., and R. L. Jernigan. 1965. Second and fourth moments of chain molecules. *J. Chem. Phys.* 42:3509–3519.
18. Flory, P. J. 1989. Statistical Mechanics of Chain Molecules Hanser-Gardner, Munich, Germany.
19. Poland, D., and H. A. Scheraga. 1970. Theory of Helix-Coil Transitions in Biopolymers Academic, New York.
20. Zong, C., T. Lu, T. Shen, and P. G. Wolynes. 2006. Nonequilibrium self-assembly of linear fibers: microscopic treatment of growth, decay, catastrophe and rescue. *Phys. Biol.* 3:83–92.
21. Shapiro, J. N. 1970. Lindemann law and lattice dynamics. *Phys. Rev. B.* 1:3982–3989.
22. DOE. 2006. Breaking the Biological Barriers to Cellulosic Ethanol: A Joint Research Agenda. U.S. Department of Energy, Rockville, MD. DOE/SC-0095.
23. Capadona, J. R., K. Shanmuganathan, D. J. Tyler, S. J. Rowan, and C. Weder. 2008. Stimuli-responsive polymer nanocomposites inspired by the sea cucumber dermis. *Science.* 319:1370–1374.
24. Bergenstrahle, M., L. Berglund, and K. Mazeau. 2007. Thermal response in crystalline I-β; cellulose: a molecular dynamics study. *J. Phys. Chem. B.* 111:9138–9145.
25. Watanabe, A., S. Morita, and Y. Ozaki. 2006. Study on temperature-dependent changes in hydrogen bonds in cellulose I-β by infrared spectroscopy with perturbation-correlation moving-window two-dimensional correlation spectroscopy. *Biomacromolecules.* 7:3164–3170.
26. Watanabe, A., S. Morita, and Y. Ozaki. 2007. Temperature-dependent changes in hydrogen bonds in cellulose I-α studied by infrared spectroscopy in combination with perturbation-correlation moving-window two-dimensional correlation spectroscopy: comparison with cellulose I-β. *Biomacromolecules.* 8:2969–2975.
27. Wada, M., H. Chanzy, Y. Nishiyama, and P. Langan. 2004. Cellulose III–I crystal structure and hydrogen bonding by synchrotron x-ray and neutron fiber diffraction. *Macromolecules.* 37:8548–8555.
28. Wada, M., Y. Nishiyama, and P. Langan. 2006. X-ray structure of ammonia-cellulose I: new insights into the conversion of cellulose I to cellulose III–I. *Macromolecules.* 39:2947–2952.
29. Yui, T., S. Nishimura, S. Akiba, and S. Hayashi. 2006. Swelling behavior of the cellulose I[β] crystal models by molecular dynamics. *Carbohydr. Res.* 341:2521–2530.
30. Matthews, J. F., C. E. Skopec, P. E. Mason, P. Zuccato, R. W. Torget, et al. 2006. Computer simulation studies of microcrystalline cellulose I-β. *Carbohydr. Res.* 341:138–152.
31. Mingos D. M. P., editor. (2004). Supramolecular Assembly via Hydrogen Bonds. I. Springer-Verlag, Berlin, Germany.
32. Walsh, C. 2006. Posttranslational Modification of Proteins: Expanding Nature's Inventory Roberts and Company, Greenwood Village, CO.