

Quantum mechanical modeling of the structures, energetics and spectral properties of I α and I β cellulose

James D. Kubicki ·
Mohamed Naseer-Ali Mohamed ·
Heath D. Watts

Received: 18 July 2012 / Accepted: 30 November 2012 / Published online: 12 December 2012
© The Author(s) 2012. This article is published with open access at Springerlink.com

Abstract Periodic planewave and molecular cluster density functional theory (DFT) calculations were performed on I α and I β cellulose in four different conformations each. The results are consistent with the previous interpretation of experimental X-ray and neutron diffraction data that both I α and I β cellulose are dominantly found in the *tg* conformation of the hydroxymethyl group with a H-bonding conformation termed “Network A”. Structural and energetic results of the periodic DFT calculations with dispersion corrections (DFT-D2) are consistent with observation suggesting that this methodology is accurate to within a few percent for modeling cellulose. The structural and energetic results were confirmed by comparison of calculated vibrational frequencies against observed infrared and Raman frequencies of I α and I β cellulose. Structures extracted from the periodic DFT-D2 energy minimizations were used to calculate the ^{13}C nuclear magnetic resonance chemical shifts ($\delta^{13}\text{C}$), and the *tg*/Network A conformations of both I α and I β cellulose

produced excellent correlations with observed $\delta^{13}\text{C}$ values.

Keywords Cellulose · Infrared · Raman · NMR · DFT

Introduction

The structure and chemistry of cellulose has been studied for decades (O’Sullivan 1997; Payen 1838), and interest in this biologically-produced polymer has increased recently due to its potential use as a source of renewable fuels (Himmel et al. 2007; Himmel et al. 1999; Lynd et al. 1991) and materials (Moon et al. 2011). Utilization of cellulose from plants is inhibited by the fact that it is difficult to extract and biodegrade because of its own structure and its interactions with other plant cell wall components. Our interest is mainly in the key role cellulose plays in plant cell wall architecture. Before one can create a realistic model of the plant cell wall, the structure of cellulose and its surfaces must be known.

Although the X-ray and neutron diffraction patterns of I α and I β cellulose were collected and interpreted a decade ago (Nishiyama et al. 2002; Nishiyama et al. 2003), the nature of cellulose is such that these methods do not fully define the details of its structure (Atalla 1999). The small cellulose fiber diameter and partially disordered regions lead to uncertainty regarding several questions. For example, does cellulose twist (Fernandes et al. 2011; French and Johnson

J. D. Kubicki (✉) · H. D. Watts
Department of Geosciences and the Earth and
Environmental Systems Institute, The Pennsylvania State
University, University Park, PA 16802, USA
e-mail: jdk7@psu.edu

M. N.-A. Mohamed
School of Advanced Sciences, VIT University, Vellore
632 014, India

2009)? If this twist exists, does it contribute to periodic disruption of order (Moon et al. 2011)? Are the H-bonding distances accurate and how does H-bonding relate to the observed vibrational spectra (Nishiyama et al. 2008)? Lastly, how do the hydroxymethyl groups on each glucose residue unit rotate with respect to the ring atoms and what are the energy barriers to rotation (Gonzalez-Outeiriño et al. 2006)?

In addition to X-ray and neutron diffraction, vibrational (i.e., infrared and Raman) and nuclear magnetic resonance (NMR) spectroscopies have added complementary information that shed further light on details of cellulose structure (Atalla 1999; Atalla and VanderHart 1984; Blackwell 1977; Earl and VanderHart 1981, 1984; Heiner et al. 1995; Hesse-Ertelt et al. 2008; Koch et al. 2000; Sternberg et al. 2003; Witter et al. 2006). Consequently, the goal of this paper is to use quantum mechanical calculations to determine model structures that match the XRD structures while producing vibrational frequencies consistent with observed vibrational spectra and ^{13}C chemical shifts ($\delta^{13}\text{C}$) that match experimental NMR values. Once this agreement between observation and computation is achieved, the computational methods can be used to address problems in cellulose surface chemistry, formation, and amorphization.

Although DFT has had numerous successes throughout computational chemistry, modeling van der Waals forces has been problematic because standard DFT does not account for these intermolecular electron correlation energies (Grimme 2006). A number of research groups (Grimme 2006; Kim et al. 2012; Zhao et al. 2006) have addressed this issue by adding empirical corrections to account for van der Waals forces based on DFT-calculated electron densities. For this work, the DFT-D2 method of Grimme (2006) has been employed because it has been implemented in the Vienna Ab-initio Simulation Package, VASP 5.4 (Kresse and Furthmüller 1996), and been shown to accurately reproduce cellulose I α and I β structures (Bućko et al. 2011; Li et al. 2011). In this study, we take the additional steps of calculating vibrational frequencies for comparison against observed vibrational spectra and $\delta^{13}\text{C}$ NMR chemical shifts to compare against observed NMR spectra.

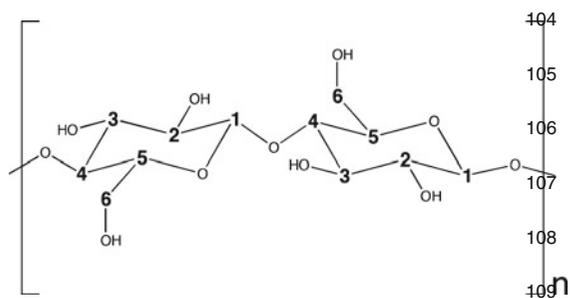
In addition to testing and verifying the cellulose structures published by Nishiyama and co-workers based on X-ray and neutron diffraction, our goal was to demonstrate that DFT-D2 methods can reliably

reproduce the observed vibrational frequencies of cellulose because this ability will allow us to trust this methodology when simulating less well understood materials such as “amorphous” or disordered cellulose. In combination with hybrid molecular orbital/density functional theory calculations (MO/DFT) to predict NMR chemical shifts, DFT-D2 can be used to interpret the molecular-level structures of cellulose when order is insufficient for XRD. For example, knowledge of cellulose surface structures is critical when attempting to understand cellulose interaction with water (Li et al. 2011; Matthews et al. 2006; Newman and Davidson 2004), hemicelluloses (Hanus and Mazeau 2006; Mazeau and Charlier 2012) or cellulose binding domains (Fernandes et al. 2011; Harris et al. 2012; Tavagnacco et al. 2011; Tormo et al. 1996). With detailed atomistic models benchmarked against a variety of analytical data, the DFT-D2 results can then be used as benchmarks for classical molecular mechanics simulations of cellulose and its interactions with other plant cell wall components.

Methods

Carbohydrate main-chain torsion angle conformations (Φ , Ψ) (i.e., $\Phi = \text{O5-C1-O-C4}$; $\Psi = \text{C1-O-C4-C5}$) are defined based on the notation shown in Scheme 1. In carbohydrates, the exocyclic hydroxymethyl torsions can collectively be represented by the orientation of the O5–C5 bond with respect to the C6–O6 bond ($\chi_1 = \text{O5-C5-C6-O6}$) and the C4–C5 bond with respect to the C6–O6 bond ($\chi_2 = \text{C4-C5-C6-O6}$). In cellulose, the values of hydroxymethyl torsions (χ) vary depending on residue 1 (χ_1) and residue 2 (χ_2) in the I α polymorph/center chain (χ_1 , χ_2) and the origin chain (χ_1 , χ_2) and center chain (χ'_1 , χ'_2) in the I β polymorph (Nishiyama et al. 2002, 2003). The χ_1 and χ'_1 are referred to as trans-gauche (*tg*; $\chi_1 = 180^\circ$), gauche-trans (*gt*; $\chi_1 = 60^\circ$) and gauche-gauche (*gg*; $\chi_1 = 300^\circ$) depending on the positions of the O5–O6 and H5–H6 atoms (Gonzalez-Outeiriño et al. 2006).

The initial structures of I α and I β cellulose were created using Materials Studio 5.5 (Accelrys Inc., San Diego, CA, USA) based on CIF files published by Nishiyama et al. (2002) and (2003), respectively. Cartesian coordinate files generated with Materials Studio 5.5 were converted into VASP 5.2 input files via a Perl script written by A.V. Bandura



Scheme 1 The Arabic numerals represent the C-atoms in a two-glucose residue unit of a cellulose chain showing the nomenclature used in this paper for the torsion angles in cellulose. *Note* that the “n” does not imply that cellobiose is the repeating unit of cellulose, but that the structure of the polymer continues in both directions through $\beta(1 \rightarrow 4)$ linkages

(St. Petersburg State University). Manipulations of χ to three different conformations (*tg*, *gg* and *gt*) and H-bonding networks (NetA and NetB; ‘Net’ here refers to ‘Network’) were performed manually in Materials Studio 5.5 based on two possible H-atom positions centered about O6 suggested by Nishiyama and co-workers (Nishiyama et al. 2002). The $I\alpha$ and $I\beta$ simulation cell stoichiometries were $C_{24}O_{20}H_{40}$ and $C_{48}O_{40}H_{80}$, respectively. The $I\alpha/tg/NetA$ conformation was also doubled in size along the *b*-axis (Fig. 1) to test the effect of system size on crystal structure; this model is designated $1 \times 2 \times 1$.

Finite clusters were extracted from the energy-minimized periodic structures and methyl groups were added to terminate broken C–O bonds to the next glucose residue unit. Atomic coordinates were fixed in the positions determined by the DFT-D2 energy minimizations. The models were 7 and 12 chains of glucose tetramers for $I\alpha$ and $I\beta$, respectively ($C_{182}O_{147}H_{322}$ and $C_{312}O_{252}H_{552}$). Only the glucose residues in the center of the finite clusters were used to predict ^{13}C NMR chemical shifts ($\delta^{13}C$), because those C-atoms should best reflect those found in the interior of cellulose.

Periodic calculations were performed with the Vienna Ab-initio Simulation Package (VASP) (Kresse and Furthmüller 1996; Kresse et al. 1994; Kresse and Hafner 1993, 1994). Projector-augmented planewave pseudopotentials were used with the PBE gradient-corrected exchange correlation functional for the 3-D periodic DFT calculations. The choice of electron density and atomic structure optimization parameters

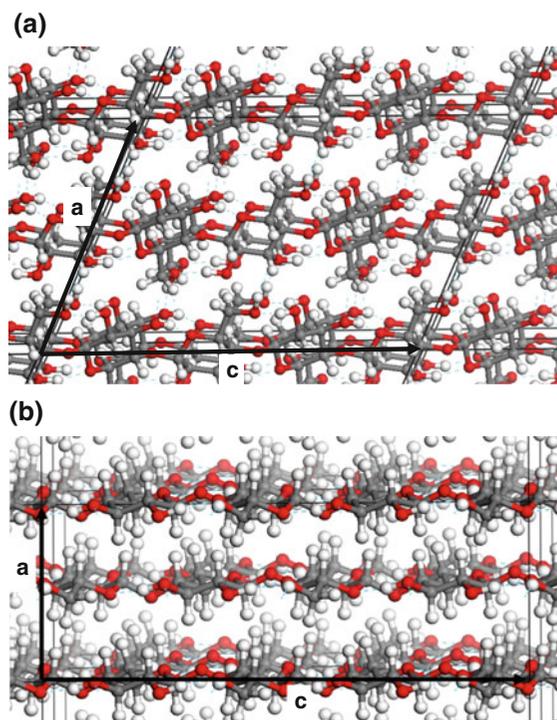


Fig. 1 Calculated minimum energy structures of **a** $I\alpha$ and **b** $I\beta$ cellulose. *Blue dashed lines* are H-bonds, and the *black boxes* outline the simulation cells. *C gray; O red; H white*. (Color figure online)

were based on Li et al. (2011) and Bućko et al. (2011). An energy cut-off of $77,190 \text{ kJ mol}^{-1}$ was used with an electronic energy convergence criterion of $9.65 \times 10^{-6} \text{ kJ mol}^{-1}$. Atomic structures were relaxed until the energy gradient was $<1.93 \text{ kJ}/(\text{mol} \cdot \text{\AA})$. $2 \times 2 \times 2$ k-point samplings were used. Atoms were first allowed to relax with the lattice parameters constrained to the experimental values, and then the atoms and lattice parameters were allowed to relax to obtain the structures, energies and spectroscopic properties reported herein. The D2 dispersion-correction parameters were 40 \AA for the cutoff distance (Bućko et al. 2011) and 0.75 for the scaling factor (s_6) and 20 for the exponential coefficient (d) in the damping function (Grimme 2006).

Frequency analyses were performed on the energy minimized structures as predicted using VASP. Second derivatives of the potential energy matrix with respect to atomic displacements were calculated using two finite-difference steps (NFREE = 2) and atomic movements of 0.015 \AA (POTIM = 0.015). Calculated frequencies were scaled by a factor of 0.97 for

comparison with experiment based on comparison of vibrational modes observed in sum frequency generation (SFG) spectra (Lee et al. 2012) and the calculated vibrational modes. Vibrational modes were analyzed using the free version of the program wxDragon 1.8.0 (Eck 2012).

Finite cluster calculations and shielding tensor calculations on the finite clusters were carried out using Gaussian 09 (Frisch et al. 2009). Gauge-independent atomic orbitals (GIAO) (Wolinski et al. 1990) were employed with the modified Perdue-Wang exchange-correlation functional mPW1PW91 and the 6-31G(d) basis set (Rassolov et al. 2001). Chemical shifts were calculated relative to methanol because this secondary standard produces $\delta^{13}\text{C}$ in better agreement with experiment (Sarotti and Pellegrinet 2009; Watts et al. 2011) than does a direct comparison of the tensors with the tetramethylsilane standard (Cheeseman et al. 1996). This multi-standard reference method also uses an empirical correction (50.4 ppm) for the difference between the $\delta^{13}\text{C}$ in methanol and (TMS) commonly used as an experimental ^{13}C NMR standard (Sarotti and Pellegrinet 2009).

Benchmark testing of the computational methodology on reproducing observed H-bond energies and O–H stretching frequencies was conducted on the H_2O – H_2O and CH_3OH – CH_3OH dimers. These are excellent test systems for H-bonding accuracy because the interactions between the molecules are dominated by a single H-bond and because there are experimental data on the energy of dimerization and vibrational frequency shifts from monomer to dimer. Gaussian 09 (Frisch et al. 2009) calculations were performed with the $\omega\text{B97X-D}$ (Chai and Head-Gordon 2008) functional and the 6-311G(d,p), 6-311+G(d,p), and 6-311G++(d,p) basis sets. The $\omega\text{B97X-D}$ exchange-correlation functional has proven reliable for reproducing H-bonding (Cirtog et al. 2011), and the basis sets selected are reasonably robust for these compounds. The diffuse functions (“+” in the basis set designations) were added to evaluate the effect they have on H-bond energies and O–H stretching frequencies because these types of functions are missing from the DFT-D2 calculations. The above three dimers were fully energy minimized and subjected to frequency analyses in Gaussian 09, then these structures were re-minimized and subjected to frequency analyses in VASP 5.2 using the same method described above for the cellulose models.

Results

H_2O – H_2O and CH_3OH – CH_3OH benchmark tests

As a test of the accuracy of the DFT-D2 computational methodology employed in the VASP calculations, comparisons of $\omega\text{B97X-D}$ and DFT-D2 results were made for the H_2O – H_2O and CH_3OH – CH_3OH dimers. The water dimerization energies (ΔE_{dimer}) are equal to -32 , -26 and -26 kJ mol^{-1} for the 6-311G(d,p), 6-311+G(d,p) and 6-311++G(d,p) basis sets, respectively. The DFT-D2 ΔE_{dimer} result was -25 kJ mol^{-1} . The calculated ΔE_{dimer} is approximately within the range of reported experimental data of -20.4 to -12.3 kJ mol^{-1} with reported uncertainties of approximately ± 3.5 kJ mol^{-1} (Fiadzomor et al. 2008; Rocher-Casterline et al. 2011). Addition of diffuse functions did improve the precision of the calculated results with respect to the water dimerization energy by -6 kJ mol^{-1} .

More relevant to the results on cellulose considered in this study are the H-bond distances and O–H stretching frequencies of the H-bonded OH group. The $\omega\text{B97X-D}/6-311\text{G(d,p)}$ and $6-311++\text{G(d,p)}$ basis sets resulted in O–O bond lengths of 2.811 and 2.835 Å and unscaled harmonic O–H symmetric stretching frequencies of 3,785 and 3,779 cm^{-1} , respectively. Thus, for H-bond distances and symmetric O–H stretching frequencies, the lack of diffuse functions in the basis set does not make a significant difference on the calculated values. In comparison, the DFT-D2 calculated values for ΔE_{dimer} , O–O H-bond distance and O–H stretching frequency associated with the H-bond in the H_2O – H_2O dimer are -25 kJ mol^{-1} , 2.856 Å, and 3,814 cm^{-1} , respectively. The experimental O–H frequency is 3,735 cm^{-1} (Huisken et al. 1996), and O–O H-bond distance is 2.976 Å (Odutola and Dyke 1980). Therefore, DFT-D2 results agree well with the observed interaction energy, O–O distance and O–H stretching frequency. This is especially true if the DFT-D2 frequency is scaled by the 0.97 factor ($3,814 \times 0.97 = 3,670$ cm^{-1}) derived by correlating DFT-D2 frequencies with SFG frequencies (see “Methods” section).

The CH_3OH – CH_3OH dimer is a test system that is even more analogous to the H-bonds found in cellulose. Although experimental data on the ΔE_{dimer} is not available for the methanol dimer, we suggest that the $\omega\text{B97X-D}/6-311++\text{G(d,p)}$ calculations on the methanol dimer are as accurate as they are for the

Table 1 Comparison of $I\alpha$ and $I\beta$ calculated lattice parameters with observed structure via X-ray and neutron diffraction by Nishiyama et al. (2002, 2003)¹ and low-temperature structure of Nishiyama et al. (2008)²; Nishiyama calculated values fromNishiyama et al. (2008); Li values from Li et al. (2011); and Bućko values from Bućko et al. (2012). CHARMM values based on structure of $6 \times 6 \times 40$ glucose residue microfibril MD simulation in water in the *tg*/NetA conformation

$I\alpha$	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	α (°)	β (°)	γ (°)	$\Phi_1, \Psi_1, \Phi_2, \Psi_2$ (°)
Expt ¹	13.434	5.962	10.400	118.08	114.80	80.37	−98, −138, −99, −140
Li	12.760	5.970	10.470	116.90	114.30	80.70	−95, −143, −95, −144
Bućko/NetA*	13.391	5.800	10.391	117.90	114.00	81.50	−94, −143, −94, −145
Bućko/NetB*	13.565	5.035	10.300	nr	nr	nr	nr
<i>tg</i> /NetA	13.132	5.664	10.386	116.20	112.30	83.00	−95, −141, −93, −143
<i>tg</i> /NetA $1 \times 2 \times 1$	13.117	5.652	10.386	116.00	112.00	83.10	−95, −145, −93, −143
<i>tg</i> /NetB	14.679	5.004	10.335	118.70	99.00	81.60	−97, −157, −83, −140
<i>gt</i>	15.341	5.038	10.404	118.20	115.70	79.10	−85, −156, −89, −156
<i>gg</i>	15.046	4.804	10.377	111.10	107.10	83.30	−93, −160, −81, −140
$I\beta$	<i>a</i>	<i>b</i>	<i>c</i>	α	β	γ	$\Phi_1, \Psi_1, \Phi_2, \Psi_2$ (°)
Expt ¹	7.784	8.201	10.380	96.50	90.00	90.00	−99, −142, −89, −147
CHARMM	7.680	8.330	10.460	90.00	90.00	97.50	−91, −150, −92, −149
Nishiyama ²	nr	8.230	10.460	90.00	90.00	89.90	−95, −144, −95, −144
Li	7.850	8.180	10.470	96.60	90.00	90.00	−95, −144, −95, −143
Bućko/NetA	7.570	8.140	10.390	96.50	90.00	90.00	−93, −143, −94, −145
Bućko/NetB	7.510	8.550	10.300	98.20	90.10	90.00	nr
<i>tg</i> /NetA	7.547	8.137	10.396	96.40	90.00	90.00	−93, −143, −94, −145
<i>tg</i> /NetB	7.538	8.407	10.359	100.00	90.00	90.00	−85, −158, −98, −135
<i>gt</i>	8.092	7.924	10.400	95.80	90.00	90.00	−93, −145, −92, −151
<i>gg</i>	7.949	8.070	10.411	94.50	90.00	90.00	−87, −141, −85, −156

$I\alpha$ phase was transformed in Bućko et al. (2011) in the setting of $I\beta$, by defining a new set of lattice vectors related to the original lattice vectors via following transformations: $a_0 = a + b$, $b_0 = a - b$, and $c_0 = c$

* The glycosidic torsion angles, Φ and Ψ which describe the relative orientation of adjacent glycosyl residues in the same chain are defined by (O5–C1–O1–C4) and (C1–O1–C4–C5)

water dimer. Hence, comparison of DFT-D2 and ω B97X-D/6-311++G(d,p) results can be used to evaluate the accuracy of the DFT-D2 energy results. The DFT-D2 ΔE_{dimer} is -29 kJ mol^{-1} which compares favorably with -28 kJ mol^{-1} calculated using ω B97X-D/6-311++G(d,p). Similarly, the O–H–O H-bond distance using DFT-D2 is 1.826 \AA whereas the ω B97X-D/6-311++G(d,p) value is 1.875 \AA . The experimental value for this bond length is 2.034 \AA (Lovas et al. 1995), so the calculated results were within 0.2 \AA of the observed H-bond length. The unscaled, calculated O–H stretching frequencies associated with this H-bond are $3,504$ and $3,767 \text{ cm}^{-1}$ for DFT-D2 and ω B97X-D/6-311++G(d,p), respectively. In this case, the ω B97X-D/6-311++G(d,p) result overestimated the observed value of $3,527 \text{ cm}^{-1}$

(Han et al. 2011) by 126 cm^{-1} (3 %, after scaling by 0.97); furthermore, when scaling is considered, and the DFT-D2 result underestimates the frequency by 128 cm^{-1} (3 %) after scaling. We conclude that the DFT-D2 methodology used on cellulose in this study is accurate to approximately 10 and 3 % for the O–H distances and O–H stretching frequencies based on these comparisons with experimental and the ω B97X-D/6-311++G(d,p) results.

Structures

Table 1 contains the lattice parameters and glycosidic torsion angles from experiment (Nishiyama et al. 2002, 2003), 2-D periodic DFT calculations (Nishiyama et al. 2008), previous 3-D periodic DFT-D2

Table 2 Calculated total electronic energies (in kJ mol^{-1} per glucose unit) including the D2 dispersion correction for van der Waals forces and relative energies (ΔE) compared to the $I\beta/tg/NetA$ model

	$I\alpha$ (kJ mol^{-1})	ΔE (kJ mol^{-1})	$I\beta$ (kJ mol^{-1})	ΔE (kJ mol^{-1})
<i>tg/NetA</i>	-12,124	+4	-12,128	0.0
$1 \times 2 \times 1$	-12,125	-	-	-
<i>tg/NetB</i>	-12,105	+23	-12,104	+24
<i>gt</i>	-12,113	+15	-12,109	+19
<i>gg</i>	-12,100	+28	-12,102	+26

calculations (Bućko et al. 2011; Li et al. 2011) and the present study for $I\alpha$ and $I\beta$ cellulose. Although variations on the order of a few percent are present throughout when comparing calculated results versus experimental data and our calculated results against previous calculations, the model values are reasonably precise. The *tg/NetA* results agree best with experimental lattice parameters compared to *tg/NetB*, *gt* and *gg*, for both $I\alpha$ and $I\beta$ (Table 1). In general, the calculated lattice parameters for the *tg/NetA* model are less than observed and some of this discrepancy may be accounted for the temperature difference between experiment and theory (i.e., observations at 298 K and calculations at 0 K) causing thermal expansion in the former as compared with the latter. In addition, the *tg/NetA* models are predicted to have the lowest total electronic energies for both $I\alpha$ and $I\beta$ (Table 2), consistent with the interpretation of experimental data (Nishiyama et al. 2003). For $I\alpha$ and $I\beta$, the energy differences between *NetA* and *NetB* conformations are -20 and -24 $\text{kJ}/\text{glucose}$ residue, respectively.

One of the larger variations from observation is for the *b*-lattice parameter of $I\alpha$ where our closest calculated value is 5 % shorter (Table 1) than found in Nishiyama et al. (2003). These discrepancies are not surprising because the *b*-lattice parameter is strongly influenced by van der Waals forces between sheets which are likely to be the least accurate component of the DFT-D2 methodology.

The structure of $I\beta$ cellulose is predicted more accurately than $I\alpha$ (Table 1). For the *tg/NetA* conformation, the largest discrepancy is the *a*-lattice vector with a 3 % error. The better agreement for $I\beta$ versus $I\alpha$ is likely due to the monoclinic symmetry of the $I\beta$ unit cell serving to limit the relaxation of the simulation cell. This symmetry effect seems to be significant in spite of the fact that the simulation cell was doubled in

size along the *c*-direction compared to the $I\alpha$ simulation cell.

Glycosidic torsions (Scheme 1— Φ_1 , Φ_2 , Ψ_1 , Ψ_2) are critical parameters in determining whether or not cellulose twists (French and Johnson 2009), and these calculated structural parameters are also compared with experiment and previous calculations in Table 1. The discrepancies of our calculated values with those determined by Nishiyama et al. (2002, 2003) are relatively small. For both $I\alpha$ and $I\beta$, the calculated torsion angles differ from experiment by at most 6° . The reported experimental standard deviation for the glycosidic torsion angles are approximately 3° for $I\alpha$ and up to 20° for $I\beta$ cellulose (Nishiyama et al. 2003). Hence, these model deviations from experiment of $\leq 6^\circ$ are reasonably precise. Previous calculations resulted in similar differences from experiment of approximately 5 % (Bućko et al. 2011; Li et al. 2011; Nishiyama et al. 2008). Unfortunately, due to the limited size of the simulation cell, the possible twist along the cellulose chains cannot be directly investigated. In addition, the experimental uncertainty and computational error will not allow for evaluation of potential subtle, long-range twisting. Nonetheless, this model of cellulose that strictly excludes twisting will be shown to reproduce experimentally observed spectra accurately. Models including any twist must reproduce experimental observables more accurately in order for twisting to be considered a necessary component of the cellulose structure.

Nishiyama et al. (2002, 2003) determined that the torsions of the hydroxymethyl groups (χ) are in the *tg* conformation for both $I\alpha$ and $I\beta$ cellulose based on X-ray diffraction patterns and structural refinements. Consequently, previous studies using DFT-D2 methods to model these structures focused solely on the *tg* conformations of the hydroxymethyl groups (Bućko et al. 2011; Li et al. 2011). However, based on classical

Table 3 Comparison of experimental (Nishiyama et al. 2002, 2003) and calculated hydroxymethyl torsions. CHARMM values based on structure of $6 \times 6 \times 40$ glucose residue microfibril MD simulation in water in the *tg*/NetA conformation

	χ_1, χ_2	χ_1', χ_2'
Iα		
Expt	167, -75	166, -74
Li	162, -80	166, -76
Bučko	164, -79	166, -76
<i>tg</i> /NetA	165, -76	165, -76
$1 \times 2 \times 1$	165, -77	165, -76
<i>tg</i> /NetB	154, -87	156, -86
<i>gt</i>	58, 177	77, 164
<i>gg</i>	-68, 52	-62, 58
Iβ		
Expt	170, -70	158, -83
CHARMM	166, -73	171, -69
Li	164, -78	165, -77
Bučko	168, -74	165, -77
<i>tg</i> /NetA	168, -74	165, -76
<i>tg</i> /NetB	159, -83	148, -92
<i>gt</i>	58, 178	52, 170
<i>gg</i>	-64, 57	-67, 54

$\chi_1 = \text{O5-C5-C6-O6}$

$\chi_2 = \text{C4-C5-C6-O6}$

molecular dynamics simulations of cellulose, Matthews et al. (2012) have suggested that I α undergoes a conformational change from *tg* to *gg*, and I β switches from *tg* to *gt* at temperatures of 500 K. Furthermore, χ of native cellulose surfaces may not be constrained to the values observed in the interior due to the interactions with water and other plant biopolymers (Fernandes et al. 2011; Harris et al. 2012; Newman and Davidson 2004; Viëtor et al. 2002; Štrucova et al. 2004). Hence, we have also investigated the *gt* and *gg* conformations in addition to *tg* of I α and I β cellulose in order to predict energy and structural differences accompanying hydroxymethyl group rotations. Although the calculated lattice parameters, glycosidic torsions and energies of these other conformations do not suggest that they exist in large percentages in the particular samples analyzed (Tables 1, 2), consideration of these other forms is worthwhile.

Table 3 lists the experimental and calculated χ_1 and χ_2 torsion angles for the hydroxymethyl groups in I α

and I β . Our calculated values are similar to the Nishiyama et al. (2003) values for the *tg*/NetA conformations, especially considering experimental uncertainty. All values deviate from the ideal values of 180° and -60° . The *gt* and *gg* torsion angles also are predicted to deviate by only a few degrees from the ideal values of $+60^\circ/+60^\circ$ and $-60^\circ/-60^\circ$, respectively. These changes in the hydroxymethyl group torsion angles induce changes in the *a* and *b* lattice parameters of I α cellulose that are outside of the range of computational accuracy, but the changes in lattice parameters for I β cellulose are smaller than for I α (Table 1). This result suggests that mixtures of *tg*/*gt*/*gg* conformers may be more difficult to detect via X-ray and neutron diffraction in I β compared to I α cellulose. Relative to the *tg* conformers, the calculated energy changes are +11 and +24 kJ mol $^{-1}$ for the *gt* and *gg* conformers in I α , respectively, and +19 and +26 kJ/glucose residue for the *gt* and *gg* conformers in I β , respectively (Table 2). Hence, higher temperatures could induce transitions to other conformations when entropic effects are considered.

Hydrogen bonding

Selected H-bond and O–O distances are listed in Table 4. In general, the DFT-D2 values are up to 0.2 Å less than the experimental values, whereas the CHARMM-based values from a $6 \times 6 \times 40$ I β microfibril are typically 0.1 Å greater than the experimental values. The calculated distances are comparable to those obtained with DFT-D2 and wB97X-D/6-311++G(d,p) for the water and methanol dimers discussed above. These discrepancies can be considered small for computational methods, especially considering the uncertainties in the experimental values. However, the DFT-D2 model overestimation of the H-bond strengths could have an effect on the calculated O–H stretching frequencies discussed below because O–H vibrations are highly sensitive to H-bond strengths. This will be particularly problematic for the I α O3–H3–O5 and O6–H2–O2 O–H stretches and the I β O6–H6–O3 stretch where the frequencies may be 200 cm $^{-1}$ lower than observation because of the stronger calculated H-bonding (see “Vibrational frequencies” section below). In addition, the DFT-D2 calculations do not predict an O6–H6–O1 H-bond in I α cellulose observed by Nishiyama et al. (2003) (Table 4).

Table 4 $I\alpha$ and $I\beta$ H-bonding parameters compared to experimental values of Nishiyama et al. (2002, 2003)

$I\alpha$	Expt. O–H–O (Å)	tg/NetA O–H–O (Å)	
O3d–H3d–O5u	1.954	1.785	
O2d–H2d–O6u	1.689	1.753	
O6d–H6d–O3d	2.176	1.719	
	MUE (Å)	0.215	
	RMSE (Å)	0.282	
	Max error (Å)	0.456	
$I\alpha$	Expt. O–O (Å)	tg/NetA O–O (Å)	
O3d–O5u	2.918	2.744	
O2d–O6u	2.465	2.676	
O6d–O3d	2.821	2.700	
	MUE (Å)	0.169	
	RMSE (Å)	0.173	
	Max error (Å)	0.211	
$I\beta$	Expt. O–H–O (Å)	CHARMM O–H–O (Å)	tg/NetA O–H–O (Å)
O3o–H3o–O5o	1.966	1.768	1.780
O2o–H2o–O6o	1.832	1.959	1.756
O6o–H6o–O3o	2.040	1.969	1.691
O3c–H3c–O5c	1.752	1.723	1.766
O2c–H2c–O6c	1.904	1.950	1.683
O6c–H6c–O3c	1.779	2.062	1.757
	MUE (Å)	0.126	0.145
	RMSE (Å)	0.155	0.188
	Max error (Å)	0.283	0.349
$I\beta$	Expt. O–O (Å)	CHARMM O–O (Å)	tg/NetA O–O (Å)
O3o–O5o	2.764	2.667	2.744
O2o–O6o	2.765	2.876	2.749
O6o–O3o	2.892	2.948	2.673
O3c–O5c	2.705	2.667	2.736
O2c–O6c	2.865	2.887	2.676
O6c–O3c	2.711	2.837	2.734
	MUE (Å)	0.075	0.083
	RMSE (Å)	0.084	0.120
	Max error (Å)	0.126	0.219
$I\alpha$	Expt. O–H–O (Å)	tg/NetA O–H–O (Å)	
O3u–H3u–O5d	2.072	1.779	
O2u–H2u–O6d	1.763	1.675	

Table 4 continued

$I\alpha$	Expt.O–H–O (Å)	tg/NetA O–H–O (Å)
O6u–H6u–O3u	1.853	1.720
	MUE (Å)	0.171
	RMSE (Å)	0.193
	Max error (Å)	0.293
$I\alpha$	Expt. O–O (Å)	tg/NetA O–O (Å)
O3u–O5d	2.868	2.744
O2u–O6d	2.480	2.676
O6u–O3u	2.770	2.700
	MUE (Å)	0.130
	RMSE (Å)	0.140
	Max error (Å)	0.196

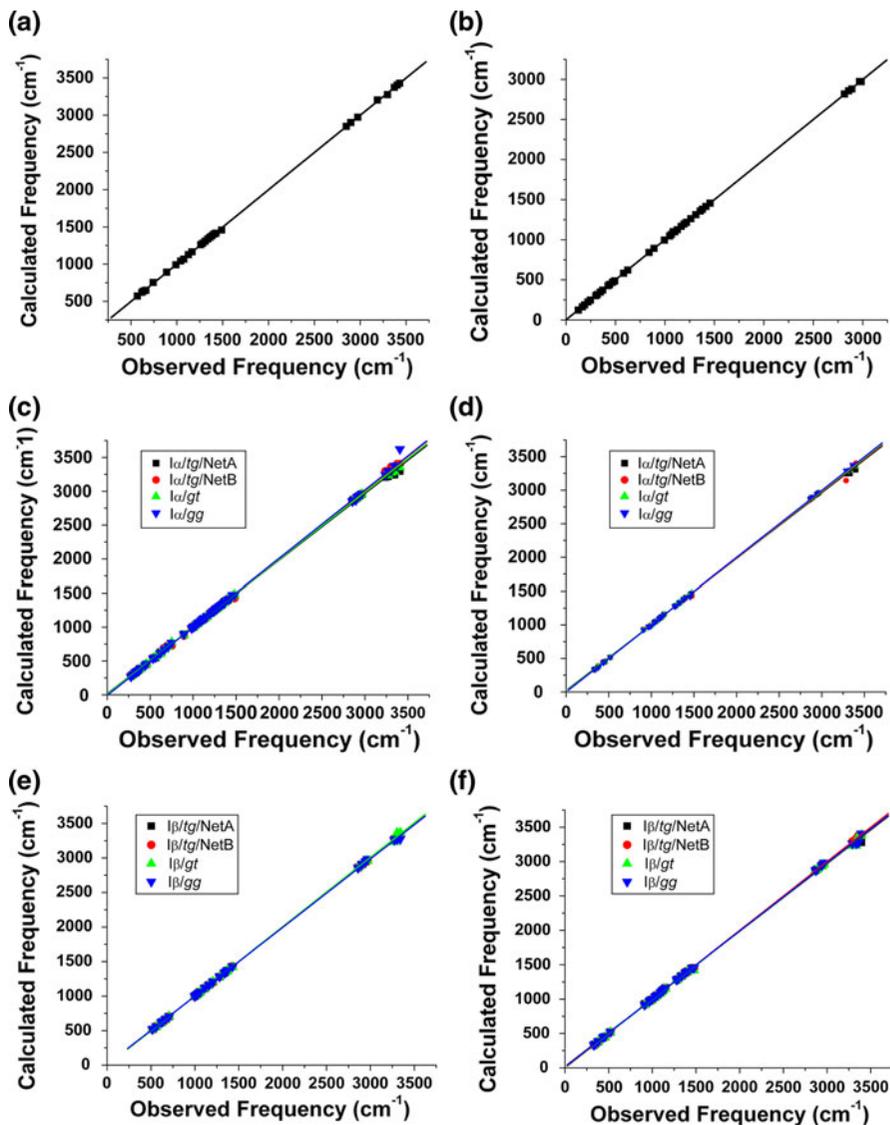
CHARMM values based on structure of $6 \times 6 \times 40$ glucose residue microfibril MD simulation in water in the *tg/NetA* conformation (Zhao et al. 2012)

Vibrational frequencies

To test the accuracy of our frequency calculations on crystalline solids, we compared calculated and observed vibrational frequencies of crystalline cellobiose. There is no controversy about the cellobiose crystal structure, so this model provides a firm link between structure and vibrational frequencies. Figure 2a, b illustrates that the correlation between the modeled and observed frequencies is excellent (IR: slope 1.00, intercept -2 cm^{-1} , R^2 0.998; Raman: slope 1.00, intercept 2 cm^{-1} , R^2 0.998). A perfect 1:1 correlation would result in a slope of 1.0, intercept of 0.0 and R^2 value of 1.0. Hence, we conclude that our computational methodology is accurate for carbohydrate vibrational frequencies.

The correlations of calculated vibrational frequencies with observed IR and Raman spectra are excellent for both $I\alpha$ and $I\beta$ cellulose (Fig. 2c–f). The slopes and R^2 values of the *tg/NetA* conformations deviate from their ideal values by $<2\%$. The maximum error for the intercept is 20 cm^{-1} for the $I\alpha$ Raman spectrum (Fig. 2d; Table 5). Paradoxically, the maximum error is found for the *tg/NetA* conformation that best matches the crystal structure (Table 1) and ^{13}C NMR chemical shifts (see below). This result makes it difficult to distinguish the correct model of cellulose structure based on IR and Raman spectra because all the models

Fig. 2 Correlations of calculated with observed **a** IR and **b** Raman frequencies for cellobiose, **c**, **d** I α cellulose and **e**, **f** I β cellulose. Cellobiose crystal IR and Raman from Xie et al. (2011), Valonia I α IR frequencies from Blackwell (1977), Valonia I α and Ramie I β Raman frequencies from Wiley and Atalla (1987), and Valonia I β (converted from I α) IR frequencies from Maréchal and Chanzy (2000)



are reasonably consistent with observed frequencies. We think the reason for the discrepancy between observation and the model *tg/NetA* is that the higher frequency O–H stretches that appear in the calculated *tg/NetB* conformation and observed spectra. Natural cellulose samples are probably a mixture of different conformations (Nishiyama et al. 2008); hence, the observed spectra may pick up frequencies that are not due to the most stable and predominant conformation.

Vibrational modes visualized are generally consistent with previous spectral assignments of IR and Raman bands, but detailed analysis of vibrational modes is beyond the scope of this paper. Based on these statistics it could be impossible to distinguish

among the four conformations (*tg/NetA*, *tg/NetB*, *gt* and *gg*). Many vibrational modes such as C–C stretches and CH₂ angle bends should not be sensitive to the long-range order because short-range covalent forces dominate them. However, the O–H stretching region between 3,000 and 3,500 cm⁻¹ should be affected by longer-range structure because H-bonding distances would be affected by changes in lattice parameters. H-bond distances are known to have a readily observable correlation with O–H stretching frequencies.

Consequently, we performed correlations between calculated and observed O–H stretching modes separately. The correlation statistics for all models were

Table 5 Correlation statistics for IR and Raman frequencies of I α and I β cellulose

	<i>tg</i> /NetA	<i>tg</i> /NetB	<i>gt</i>	<i>gg</i>
Iα-IR				
Slope	0.984	1.006	0.992	1.010
Intercept	15.5	-7.9	9.0	-10.1
R ²	0.998	0.998	0.998	0.998
RMSE	20.5	15.6	14.6	30.1
Max error	109.4	58.6	62.2	214.4
Iα-Raman				
Slope	0.981	0.988	0.993	0.999
Intercept	20.2	11.0	8.9	-0.3
R ²	0.998	0.998	0.998	0.998
RMSE	29.6	28.9	12.9	11.2
Max error	103.4	142.7	46.6	37.2
Iβ-IR				
Slope	0.99	1.00	1.004	0.992
Intercept	5.53	0.20	-4.0	9.0
R ²	0.998	0.998	0.998	0.998
RMSE	14.0	6.3	16.4	8.2
Max error	60.5	19.0	61.4	24.5
Iβ-Raman				
Slope	0.984	1.000	0.990	0.990
Intercept	17.8	-0.7	6.6	11.9
R ²	0.998	0.998	0.998	0.998
RMSE	28.3	8.2	24.6	21.9
Max error	114.5	34.0	100.3	86.2

poorer compared to the overall correlations as expected, but the *gt* and *gg* models resulted in better fits compared to the I α IR and Raman frequencies, respectively. The *tg*/NetB model better matched observed I β IR and Raman O–H stretching frequencies.

There could be a number of reasons for these discrepancies. For example, we note that the IR spectrum for I α cellulose of Blackwell (1977) exhibits a peak near 3,408 cm⁻¹. The model that best produces this higher frequency (i.e., less H-bonded) O–H stretch is the I α /*tg*/NetB conformation. We have concluded above based on comparisons of structures to experiment and calculated energetics, that the I α /*tg*/NetA conformation should be dominant, but mixtures of networks A and B are possible in cellulose (Nishiyama et al. 2008). The fact that our calculations reproduce other observed frequencies so accurately, while this one vibrational mode can deviate from observation by

over 100 cm⁻¹, leads us to conclude that it is not inaccuracy of the model calculations but rather structural heterogeneity that leads to the presence of this 3,408 cm⁻¹ peak in the observed spectrum. A similar discrepancy exists between the calculated and observed IR spectrum of I β cellulose; but in this case, the network A conformation is approximately 60 cm⁻¹ under-estimated compared to experiment (Table 4). Another potential reason why the preferred I α /*tg*/NetA conformation does not reproduce some of the observed higher frequency O–H stretches such as that at 3,408 cm⁻¹ could be that the observed spectra are detecting O–H groups at the surface of the cellulose. The H-bonding of surface OH groups could be less strong than internal OH groups and lead to higher frequency O–H stretches. Lastly, the scaling factor for the O–H stretches could be different from that obtained based on the CH₂ stretches.

A crystallinity index for cellulose I based on the intensities of the 380 and 1,096 cm⁻¹ Raman peaks is available (Agarwal et al. 2010). Since these two vibrational modes may be related to the order/disorder of cellulose I, we examined the model vibrational modes of frequencies in these regions. Calculated frequencies were scaled by 0.97 based on matching observed and calculated sum frequency generation (SFG) modes (Lee et al. 2012), and we found that the observed 380 cm⁻¹ vibration is due to C1–C2–C3, C1–O1–C4 and C1–O5–C5 bending modes in our calculations. The 1,096 cm⁻¹ Raman peak consists of C1–O1, C1–C2 and C4–C5 stretches, C1–O1–C4 angle bends, and C5–C6H₂ twists. Note that Raman intensities were not calculated, so we cannot definitively assign these modes to Raman bands nor examine the intensity ratios, but these motions do provide a hypothesis for amorphization of cellulose I. The fact that the internal ring modes of the 380 cm⁻¹ peak lose Raman intensity suggests that the glucose residues may be distorting from the chair conformation. The predicted motions of the 1,096 cm⁻¹ peak involve both the glycosidic and hydroxymethyl torsions, which will affect the order both along and between the glucose residue chains.

¹³C NMR chemical shifts

As discussed above for the IR and Raman spectra, the *tg*/NetA conformation results in the best overall

Table 6 Correlation values for $\delta^{13}\text{C}$ calculations versus observation (Witter et al. 2006; Sternberg et al. 2003)

$I\alpha$	<i>tg</i> /NetA	<i>tg</i> /NetB	<i>gt</i>	<i>gg</i>	
Slope	0.99/1.01	0.88/0.84	0.99/0.83	0.92/0.87	
Intercept	0.08/−2.24	7.6/12.6	−1.1/11.1	3.3/7.6	
R^2	0.997/0.996	0.893/0.970	0.933/0.961	0.921/0.875	
RMSE	1.3/1.7	4.7/3.0	4.0/4.0	4.8/5.7	
Max error	2.6/2.9	10.6/5.4	7.0/6.4	10.5/10.9	
$I\beta$	<i>tg</i> /NetA	CHARMM	<i>tg</i> /NetB	<i>gt</i>	<i>gg</i>
Slope	1.01/1.00	0.99/1.03	0.98/1.02	1.02/1.06	0.93/0.95
Intercept	−1.4/−0.9	−8.3/−11.2	1.3/−1.1	−2.3/−5.2	3.5/2.4
R^2	0.996/0.998	0.943/0.946	0.984/0.978	0.986/0.984	0.939/0.931
RMSE	1.1/0.8	10.0/9.3	1.7/2.1	1.8/2.0	4.0/3.7
Max error	1.9/1.7	12.9/12.6	3.0/3.7	2.7/2.9	8.0/7.3

Two sets of experimental values are for different glucose residue units. CHARMM values based on structure of $6 \times 6 \times 40$ glucose residue microfibril MD simulation in water in the *tg*/NetA conformation. A center section was extracted and subjected to NMR calculations identical to those based on the DFT-calculated structures

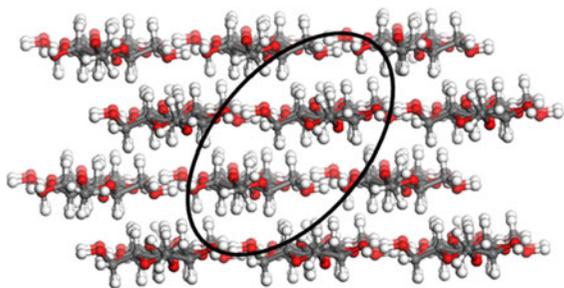


Fig. 3 Example extracted cluster used to calculate ^{13}C NMR chemical shifts based on central atoms. Glucose residues in the center of the cluster (*circled*) were used for $\delta^{13}\text{C}$ correlations versus experiment. C gray; O red; H white. (Color figure online)

correlation of calculated and observed $\delta^{13}\text{C}$ values for both $I\alpha$ and $I\beta$ cellulose (Table 6; Figs. 3, 4). In this case, the slopes deviate from 1.0 by <1 % and the intercepts deviate from 0.0 by <1 ppm. The maximum error between computation and observation for individual peaks is no greater than 3 ppm for both $I\alpha$ and $I\beta$ *tg*/NetA (Table 6). The correlations of the other conformations are reasonable, but all have greater root-mean-squared error (RMSE) and maximum errors than the *tg*/NetA conformations. These results represent a significant improvement over previous semi-empirical calculations (Koch et al. 2000) and are of similar quality to the results in Witter et al. (2006) based on empirical correlations of structures versus observed $\delta^{13}\text{C}$ values.

C1 and C4 ^{13}C chemical shifts and glycosidic torsions

The C1 and C4 atoms involved in the glycosidic torsions are in agreement with observed values well within the accuracy of our methodology. This could have three explanations. One, the $\delta^{13}\text{C}$ signals from these atoms could be relatively insensitive to the 10 % error in Φ and Ψ mentioned above in our model. Two, there could be compensating errors in computing the $\delta^{13}\text{C}$ values, allowing an inaccurate structure to provide fortuitous agreement with experiment. Three, model Φ and Ψ values could be more accurate than Table 1 would suggest.

The first explanation was tested by varying the Φ and Ψ values of a cellulose dimer model through $(-86^\circ, 151^\circ)$, $(-87^\circ, 156^\circ)$ and $(-88^\circ, 161^\circ)$ with other atomic coordinates held constant. The C1 and C4 $\delta^{13}\text{C}$ values varied by 2.6 and 1.1 ppm over this range, respectively. In contrast, ranges of approximately 15 and 12 ppm for the C1 and C4 $\delta^{13}\text{C}$ values have been reported (Suzuki et al. 2009), but the range of Φ and Ψ values was from 0 to 360° in that study. The Suzuki et al. (2009) results within the window of Φ and Ψ values we investigated show small $\delta^{13}\text{C}$ dependence, so the current results are actually consistent with this previous study. Hence, one can conclude that the C1 and C4 chemical shifts are relatively insensitive to changes in the glycosidic torsion angles within our computational accuracy. The second argument can be

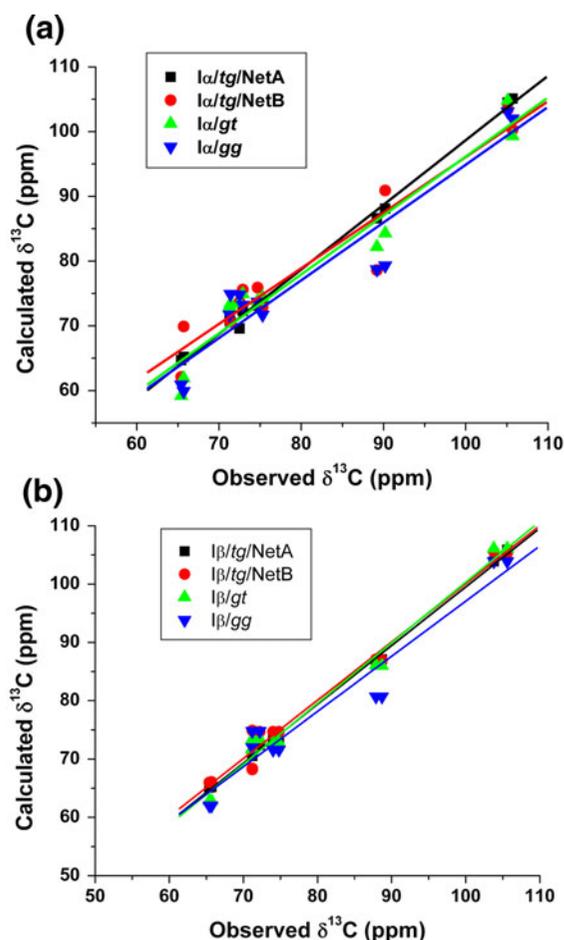


Fig. 4 Correlations of calculated $\delta^{13}\text{C}$ values with observed chemical shifts for **a** $\text{I}\alpha$ cellulose from Witter et al. (2006) and **b** $\text{I}\beta$ cellulose from Erata et al. (1997) as cited in Sternberg et al. (2003)

discounted because similar methodologies applied to simple molecules with known structures (e.g., cellobiose) predict ^{13}C chemical shifts to within approximately 2 ppm. The third explanation cannot be tested with our methods because of the insensitivity of the C1 and C4 $\delta^{13}\text{C}$ values mentioned above. We conclude that uncertainty in measuring Φ and Ψ angles and the actual variation in these cellulose structural parameters forces one to accept this level of uncertainty.

C4, C5 and C6 ^{13}C chemical shifts and hydroxymethyl torsions

In a similar vein, the C4, C5 and C6 $\delta^{13}\text{C}$ values should be sensitive to the χ_1 and χ_2 torsion angles (Suzuki et al.

2009). The fact that the calculated $\delta^{13}\text{C}$ values for C4, C5 and C6 match experiment closely (Fig. 4) is another indication that the *tg* conformation predominates and that the model structures predicted are consistent with experiment. For example, the $\text{I}\alpha/\text{tg}/\text{NetA}$ C5 and C6 chemical shifts differ from experiment by <1 ppm, whereas as those for the $\text{I}\alpha/\text{tg}/\text{NetB}$, *gt*, and *gg* conformations are in error by up to 4.2, 6.2 and 4.8 ppm, respectively. The $\delta^{13}\text{C}_4$ value changes from 87.1 ppm in the *tg* conformation to 80.6 ppm in the *gg* conformation compared to the 87.9 and 88.7 ppm values of Sternberg et al. (2003). Consequently, our results strongly support the *tg* conformation.

The sensitivity of the C5 and C6 chemical shifts to hydroxymethyl group rotations was also tested by varying χ_1 from 160 to 170°. NMR calculations were performed without any further energy relaxation. The 10° difference between the 160 and 170° conformations resulted in changes of 1.2 and 0.4 ppm for the C5 and C6 atoms, respectively. These results are consistent with Suzuki et al. (2009) who predicted that the C5 atom should be most sensitive to changes in χ angle in this range. This magnitude of change should be observable in the ^{13}C NMR spectra but is within the uncertainty of our computational methodology, so we cannot use the calculated $\delta^{13}\text{C}$ values to further refine the expected χ_1 and χ_2 torsion angles.

Discussion

The main purpose of this study was to test the ability of the Nishiyama et al. (2002, 2003) cellulose structures to reproduce the IR, Raman (Fig. 2), and ^{13}C NMR (Fig. 4) spectra of $\text{I}\alpha$ and $\text{I}\beta$ cellulose. It is not possible to distinguish the correct structure based on our correlations of calculated and observed vibrational frequencies because many frequencies are similar in each model; furthermore, no model accurately reproduces all O–H stretching frequencies. However, based on the excellent correlations of calculated and observed $\delta^{13}\text{C}$ values, the Nishiyama structures are confirmed as our results match observation within computational accuracy. The conformations of the hydroxymethyl groups and the H-bonding networks suggested in Nishiyama et al. (2002, 2003) resulted in better fits than the alternative structures in most cases.

Another goal was to evaluate the DFT-D2 methodology for modeling cellulose structure. Although

improvements are possible, the structures produce good agreement with experimental observables such as lattice parameters, vibrational spectra and ^{13}C NMR chemical shifts. This benchmarking step is significant as it allows one to trust DFT-D2 results when modeling cellulose behavior for cases that are not as well constrained experimentally (e.g., disordered cellulose and surface interactions). DFT-D2 can then be used to help interpret experimental data and to benchmark classical force fields.

Significant limitations in the present study should be addressed, however. First, due to computational constraints, the size of the simulation cell does not adequately test the possibility of twisting along the length of the cellulose chains. Larger scale, classical MD simulations have shown that twisting is an effect of model dimensions, so larger model structures should be examined in the future. Furthermore, the fully ordered 3-D periodic simulation cells constrain the possible structures obtained in energy minimizations. Models of finite cellulose fibers surrounded by vacuum, water, or other plant cell wall components would be useful. Finite cellulose fibers, such as those used in classical simulations (Matthews et al. 2006, 2012; Zhong et al. 2008) allow one to study the surface relaxation that likely occurs. This surface relaxation may be the cellulose component observed as “amorphous” or disordered cellulose (Harris et al. 2012), so including this possibility is critical. In addition, surface relaxation may induce changes in the prediction of bulk cellulose structure, so comparison with experimental observables could be more realistic in model systems of finite cellulose fibers. Lastly, results in this study were based on energy minimizations, so DFT-D2 MD simulations would be an important additional methodology. The inclusion of a finite temperature in the simulation would allow for a more direct comparison to experimental thermodynamics, structures and spectra.

Acknowledgments This work was supported by the U.S. Department of Energy grant for the Energy Frontier Research Center in Lignocellulose Structure and Formation (CLSF) from the Office of Science, Office of Basic Energy Sciences under Award Number DE-SC0001090. The authors also thank Ilana Bar Ben Gurion University of the Negev for providing IR and Raman frequencies of crystalline cellobiose and Yoshiharu Nishiyama for suggesting DFT-D2 calculations as a methodology for modeling cellulose. Computational support was provided by the Research Computation and Cyberinfrastructure group at The Pennsylvania State University.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Agarwal UP, Reiner RS, Ralph SA (2010) Cellulose I crystallinity determination using FT-Raman spectroscopy: univariate and multivariate methods. *Cellulose* 17(4):721–733
- Atalla R (1999) Carbohydrates and their derivatives including tannins, cellulose and related lignins. In: Barton D, Nakanishi K (eds) *Comprehensive natural products chemistry*, 3rd edn. Elsevier, Amsterdam, pp 529–598
- Atalla R, VanderHart D (1984) Native cellulose: a composite of two distinct crystalline forms. *Science* 223:283–285
- Blackwell J (1977) Infrared and Raman spectroscopy of cellulose. In: Aurther J (ed) *Cellulose chemistry and technology*, ACS symposium series. American Chemical Society, Washington, DC, pp 206–218
- Bučko T, Tunega D, Ángyán JG, Hafner J (2011) Ab initio study of structure and interconversion of native cellulose phases. *J Phys Chem A* 115:10097–10105
- Chai J-D, Head-Gordon M (2008) Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections. *Phys Chem Chem Phys* 10(44):6615–6620
- Cheeseman JR, Trucks GW, Keith TA, Frisch MJ (1996) A comparison of models for calculating nuclear magnetic resonance shielding tensors. *J Chem Phys* 104(14):5497–5509
- Cirtog M, Alikhani ME, Madebène B, Soulard P, Asselin P, Tremblay B (2011) Bonding nature and vibrational signatures of oxirane: (water) $n = 1–3$. Assessment of the performance of the dispersion-corrected DFT methods compared to the ab initio results and Fourier transform infrared experimental data. *J Phys Chem A* 115(24):6688–6701
- Earl WL, VanderHart DL (1981) Observations by high-resolution carbon-13 nuclear magnetic resonance of cellulose I related to morphology and crystal structure. *Macromolecules* 14:570–574
- Earl WL, VanderHart DL (1984) Studies of microstructure in native celluloses using solid-state carbon-13 NMR. *Macromolecules* 17(8):1465–1472
- Eck B (2012) wxDragon. Retrieved from www.wxdragon.de
- Erata T, Shikano T, Yunoki S, Takai M (1997) The complete assignment of the ^{13}C CP/MAS NMR spectrum of native cellulose by using ^{13}C labeled glucose. *Cellul Commun* 4:128–131
- Fernandes AN, Thomas LH, Altaner CM, Callow P, Forsyth VT, Apperly DC, Kennedy CJ et al (2011) Nanostructure of cellulose microfibrils in spruce wood. *Proc Natl Acad Sci USA* 108(47):E1195–E1203
- Fiadzomor PAY, Keen AM, Grant RB, Orr-Ewing AJ (2008) Interaction energy of water dimers from pressure broadening of near-IR absorption lines. *Chem Phys Lett* 462(4–6):188–191
- French AD, Johnson GP (2009) Cellulose and the twofold screw axis: modeling and experimental arguments. *Cellulose* 16(6):959–973

- Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, Montgomery JA Jr et al (2009) Gaussian 09 Revision B.01. Gaussian, Inc, Wallingford, CT
- Gonzalez-Outeiriño J, Kirschner KN, Thobhani S, Woods RJ (2006) Reconciling solvent effects on rotamer populations in carbohydrates—a joint MD and NMR analysis. *Can J Chem* 84(4):569–579
- Grimme S (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J Comput Chem* 27(15):1787–1799
- Han H, Camacho C, Witek HA, Lee Y (2011) Infrared absorption of methanol clusters (CH₃OH)_n with n = 2 – 6 recorded with a time-of-flight mass spectrometer using infrared depletion and vacuum-ultraviolet ionization. *J Chem Phys* 134(14):144309–144319
- Hanus J, Mazeau K (2006) The xyloglucan—cellulose assembly at the atomic scale. *Biopolymers* 82:59–73
- Harris DM, Corbin K, Wang T, Gutierrez R, Bertolo AL, Carloalberto P, Smilgies D-M et al (2012) Cellulose microfibril crystallinity is reduced by mutating C-terminal transmembrane region residues CESA1A903 V and CESA3T942I of cellulose synthase. *Proc Natl Acad Sci USA* 109(11):4098–4103
- Heiner AP, Sugiyama J, Teleman O (1995) Crystalline cellulose I α and I β studied by molecular-dynamics simulation. *Carbohydr Res* 273(2):207–223
- Hesse-Ertelt S, Witter R, Ulrich AS, Kondo T, Heinze T (2008) Spectral assignments and anisotropy data of cellulose I α : ¹³C-NMR chemical shift data of cellulose I α determined by INADEQUATE and RAI techniques applied to uniformly ¹³C-labeled bacterial celluloses of different Gluconacetobacter xylinus strains. *Magn Reson Chem* 46: 1030–1036
- Himmel ME, Ruth MF, Wyman CE (1999) Cellulase for commodity products from cellulosic biomass. *Curr Opin Biotechnol* 10:358–364
- Himmel ME, Di S-Y, Johnson DK, Adney WS, Nimlos MR, Brady JW, Foust TD (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels. *Science* 315:804–807
- Huisken F, Kaloudis M, Kulcke A (1996) Infrared spectroscopy of small size-selected water clusters. *J Chem Phys* 104(17): 17–25
- Kim H, Choi J, Goddard WA (2012) Universal correction of density functional theory to include London dispersion (up to Lr, element 103). *J Phys Chem Lett* 3(3):360–363
- Koch F-T, Priess W, Witter R, Sternberg U (2000) Calculation of solid-state ¹³C NMR spectra of cellulose I α , I β and II using a semi-empirical approach and molecular dynamics. *Macromol Chem Phys* 201(15):1930–1939
- Kresse G, Furthmüller J (1996) Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys Rev B* 54:11169–11186
- Kresse G, Hafner J (1993) Ab initio molecular dynamics for open-shell transition metals. *Phys Rev B* 48(17):13115–13118
- Kresse G, Hafner J (1994) Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Phys Rev B* 49(20):14251–14269
- Kresse G, Furthmüller J, Hafner J (1994) Theory of the crystal structures of selenium and tellurium: the effect of generalized-gradient corrections to the local-density approximation. *Phys Rev B* 50(18):13181–13185
- Lee CM, Park YB, Mohamed MNA, Kubicki JD, Roberts E, Cosgrove et al (2012) Structural understanding of cellulose from sum-frequency-generation (SFG) spectroscopy analyses. In: 243rd American chemical society national meeting and exposition. American Chemical Society, San Diego
- Li Y, Lin M, Davenport JW (2011) Ab Initio studies of cellulose I: crystal structure, intermolecular forces, and interactions with water. *J Phys Chem C* 115(23):11533–11539
- Lovas F, Belov S, Tretyakov M, Stahl W, Suenram RD (1995) The a-Type K = 0 microwave spectrum of the methanol dimer. *J Mol Spectrosc* 170:478–492. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0022285285710867>
- Lynd LR, Cushman JH, Nichols RJ, Wyman CE (1991) Fuel ethanol from cellulosic biomass. *Science* 251:1318–1323
- Maréchal Y, Chanzy H (2000) The hydrogen bond network in I β cellulose as observed by infrared spectrometry. *J Mol Struct* 523(1–3):183–196
- Matthews JF, Skopec CE, Mason PE, Zuccato P, Torget RW, Sugiyama J, Himmel ME et al (2006) Computer simulation studies of microcrystalline cellulose I β . *Carbohydr Res* 341:138–152
- Matthews JF, Beckham GT, Bergensträhle-Wohlert M, Brady JW, Himmel ME, Crowley MF (2012) Comparison of cellulose I β simulations with three carbohydrate force fields. *J Chem Theory Comput* 8(2):735–748
- Mazeau K, Charlier L (2012) The molecular basis of the adsorption of xylans on cellulose surface. *Cellulose* 19(2):337–349
- Moon RJ, Martini M, Nairn J, Simonsen J, Youngblood J (2011) Cellulose nanomaterials review: structure, properties and nanocomposites. *Chem Soc Rev* 40(7):3941–3994
- Newman RH, Davidson TC (2004) Molecular conformations at the cellulose–water interface. *Cellulose* 11:23–32
- Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose I β from synchrotron X-ray and neutron fiber diffraction. *J Am Chem Soc* 124(31):9074–9082
- Nishiyama Y, Sugiyama J, Chanzy H, Langan P (2003) Crystal structure and hydrogen bonding system in cellulose I α from synchrotron X-ray and neutron fiber diffraction. *J Am Chem Soc* 125(47):14300–14306
- Nishiyama Y, Johnson GP, French AD, Forsyth V, Langan P (2008) Neutron crystallography, molecular dynamics, and quantum mechanics studies of the nature of hydrogen bonding in cellulose I β . *Biomacromolecules* 9(11):3133–3140
- O’Sullivan AC (1997) Cellulose: the structure slowly unravels. *Cellulose* 4(3):173–207
- Odotola JA, Dyke TR (1980) Partially deuterated water dimers: microwave spectra and structure. *J Chem Phys* 72(9): 5062–5070
- Payen A (1838) Memoir on the composition of the tissue of plants and of woody material. *CR Biol* 7:1052–1056
- Rassolov VA, Ratner MA, Pople JA, Redfern PC, Curtiss LA (2001) 6–31G* basis set for third-row atoms. *J Comput Chem* 22:976–984
- Rocher-Casterline BE, Ch’ng LC, Mollner AK, Reisler H (2011) Communication: determination of the bond

- dissociation energy (D_0) of the water dimer, $(\text{H}_2\text{O})_2$, by velocity map imaging. *J Chem Phys* 134(21):211101–211104
- Sarotti AM, Pellegrinet SC (2009) A multi-standard approach for GIAO (^{13}C) NMR calculations. *J Org Chem* 74(19):7254–7260
- Sternberg U, Koch F-T, Priess W, Witter R (2003) Crystal structure refinements of cellulose polymorphs using solid-state ^{13}C chemical shifts. *Cellulose* 10:189–199
- Štrucova A, His I, Apperley DC, Sugiyama J, Jarvis MC (2004) Structural details of crystalline cellulose from higher plants. *Biomacromolecules* 5:1333–1339
- Suzuki S, Horiia F, Kurosu H (2009) Theoretical investigations of ^{13}C chemical shifts in glucose, cellobiose, and native cellulose by quantum chemistry calculations. *J Mol Struct* 921(1–3):219–226
- Tavagnacco L, Masona PE, Schnupfa U, Piticia F, Zhong L, Himmel ME, Crowley M et al (2011) Sugar-binding sites on the surface of the carbohydrate-binding module of CBH I from *Trichoderma reesei*. *Carbohydr Res* 346(6):839–846
- Tormo J, Lamed R, Chirino AJ, Morag E, Bayer EA, Shoham Y, Steitz TA (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *J Eur Mol Biol Organ (EMBO)* 5:5739–5751
- Viëtor RJ, Newman RH, Ha M-A, Apperely DC, Jarvis MC (2002) Conformational features of crystal-surface cellulose from higher plants. *Plant J* 30:721–731
- Watts HD, Mohamed MNA, Kubicki JD (2011) Comparison of multistandard and TMS-standard calculated NMR shifts for coniferyl alcohol and application of the multistandard method to lignin dimers. *J Phys Chem B* 115(9):1958–1970
- Wiley JH, Atalla RA (1987) Band assignments in the Raman spectra of celluloses. *Carbohydr Res* 160:113–129
- Witter R, Sternberg U, Hesse S, Kondo T, Koch F-T, Ulrich AS (2006) ^{13}C chemical shift constrained crystal structure refinement of cellulose I α and its verification by NMR anisotropy experiments. *Macromolecules* 38(19):6125–6132
- Wolinski K, Hinton JF, Pulay P (1990) Efficient implementation of the gauge-independent atomic orbital method for NMR chemical shift calculations. *J Am Chem Soc* 112:8251–8260
- Xie H, Pincu M, Brauer B, Gerber RB (2011) Raman and infrared spectra of cellobiose in the solid state: what can be learned from single-molecule calculations? *Chem Phys Lett* 514:284–290
- Zhao Y, Schultz NE, Truhlar DG (2006) Design of density functionals by combining the method of constraint satisfaction with parametrization for thermochemistry, thermochemical kinetics, and noncovalent interactions. *Design* 2:364–382
- Zhong L, Matthews JF, Crowley MF, Rignall T, Talón C, Cleary JM, Walker RC et al (2008) Interactions of the complete cellobiohydrolase I from *Trichoderma reesei* with microcrystalline cellulose I β . *Cellulose* 15(2):261–273