Prediction of Vibrational Spectra of Polysaccharides—Simulated IR Spectrum of Cellulose Based on Density Functional Theory (DFT)

Søren Barsberg*

Faculty of Life Sciences, University of Copenhagen, Rolighedsvej 23, DK-1958 Frederiksberg C, Denmark

Received: May 9, 2010; Revised Manuscript Received: July 28, 2010

The continuing developments of electronic structure methods may provide insight into the vibrational spectroscopy of polysachharides, which was not accessible to older works on this subject. The present work shows for the first time how main features of cellulose infrared spectra can be predicted and assigned using simple single chain models of cellulose combined with density functional theory prediction of their vibrational properties. The results provide a more informed basis for assigning cellulose IR bands and may resolve some of the challenges associated with the molecular origin of "marker" bands, which are commonly used to measure properties such as crystallinity or crystalline forms. The theoretical approach can be seen as a first-order approximation, which can be further improved.

Introduction

Vibrational spectroscopy of polysaccharides provides basic knowledge on their structure, properties, and interactions with the environment. Among the polysaccharides, cellulose accounts for most of the biomass in the plant kingdom. The ultrastructure of its crystalline forms is a highly complex subject. Here, we point out its main features with emphasis on natural cellulose I. For more details, the reader can consult several reviews on the subject and the references contained herein.[1−3]

The polyglucan chains, which make up cellulose, are synthesized in parallel at the cell surface and associate noncovalently to form microfibrils. Their transverse dimension varies depending on the source from 2−3 nm up to 20 nm in algal and tunicate samples, whereas the longitudinal dimension is much larger. Cellulose I is thus a semicrystalline arrangement of parallel polyglucan chains and is (from a crystallographic viewpoint) practically of infinite dimension only in the chain direction. In addition, various treatments are capable of modifying this arrangement and lead to the polymorphs cellulose II, III, and IV, which differ in characteristics such as chain directions (e.g., antiparallel for cellulose II) and/or arrangement within the unit cell.

Cellulose I occurs in two different crystal forms (allomorphs) I_α and I_β both assembled from parallel chains in flat-ribbon conformation with alternating glycosyl units locked in opposite orientation by two intramolecular hydrogen bonds. One of these is enabled by the rotational *tg* position of the hydroxymethyl group (O6C6 relative to O5C5 and C4C5), which brings the O6 in close proximity to O2 of the neighboring residue. The chains are held together edge to edge in flat sheets by intermolecular hydrogen bonds. The I_α and I_β forms differ in the way chains are organized and here mainly in the stacking and relatively weak bonding of chain sheets to form the microfibrils. This leads to a triclinic single chain unit cell for I_α, whereas I_β forms a monoclinic unit cell of two crystallographic independent center and origin chains. The I_α form can be converted to the thermodynamically more stable I_β form by thermal activation or (in principle) by shear forces. For each of these three types of chain sheets (I_α, I_β center, and I_β origin), the hydroxyl rotational positions can be described in terms of two different hydrogen bonding networks, which both organize the sheet. Thus, the sheets which stack to form cellulose I can be organized in either of these six (three by two) "sub-forms" based on qualitatively the *same* single-chain heavy atom organization.

Thus, for cellulose, IR spectroscopy can provide information on the ratio of I_α/I_β forms,[4−6] degree of crystallinity,[7] molecular orientation in plant cells,[8,9] and interactions with water.[10,11] For native cellulose, present day interpretation of IR spectroscopy results are largely based on older works on the assignments of its vibrational bands.[12−16] These works could not take advantage of more recent developments of basic electronic structure methods and their implementation by ever growing computational resources. A more recent work is notable in its thorough attempt to assign cellulose IR bands from the perturbations induced by EM field polarization, H/D exchange, and controlled heating.[17] Still, the capabilities of electronic structure methods to provide valuable complementary information have to the best of the author's knowledge not been explored within this field.

The present work demonstrates for cellulose I how the vibrations of polysaccharide crystals can be predicted and assigned from density functional theory (DFT) calculations. Cellulose IR bands are predicted from single chain models of finite degree of polymerization (DP) of the pyranose unit. These simple models do in principle not distinguish between the I_α and I_β forms, which would require several chains in crystal models, but do reflect the distinction in terms of hydrogen bond networks. Thus, the single chain cellulose I models adopt two possible orientational states of the hydroxyl groups, which are truncated representations (as intermolecular bonds are not modeled) of the corresponding hydrogen bond networks. It will be shown that such single chain models capture the essential features of the cellulose IR spectrum and provide valuable information on the nature of corresponding vibrations (normal modes). A refined description, which is not explicitly considered in the present work, would obtain the I_α or I_β spectra by considering at an appropriate lower level of theory how the single chain spectrum is perturbed by intermolecular couplings, i.e., mainly *via* hydrogen bonding, to neighboring chains within

* Phone: +45 35 33 15 00. Fax: +45 35 33 15 08. E-mail: sbar@life.ku.dk.

10.1021/jp104213z
Published on Web 08/20/2010

a chain sheet or full crystal model. In order to model single chain based cellulose IR bands, the well documented B3LYP functional and a relatively small basis set was chosen (details are found below). A critical point is how the vibrations of cellulose itself, consisting of orderly arranged molecular chains of DP $\sim 10^3-10^4$, can be modeled by DFT which in present day implementations usually handle at most $\sim 10^2$ atoms (or DP $\sim 5-10$ depending mainly on basis set and hardware). Thus, a method to approximate the vibrations of a DP $> 10^3$ system from DP $< 10^1$ systems is needed. Such a method must at least address the covalent intrachain interactions which are relatively strong and essential for predicting the vibrations as compared to the weaker noncovalent interchain interactions. In the present work, interchain interactions and chain arrangements in the crystallographic unit cell are, as explained above, not explicitly considered. Since this work focuses on *proof of principle* rather than an exhaustive assignment work, we limit the considerations to the 1500−600 cm^{-1} "fingerprint" interval. It will be shown that such single chain modeling captures most of the features of the cellulose IR spectrum in this interval.

Rather than studying simulated spectra of limited DP = 2, 3, etc., model systems, which all contain significant contributions from vibrations located *at or near the ends* of the models, the method focuses on the differences or increments of one simulated spectrum for a model with DP = n to the next obtained for a model with DP = $n + 1$. Thus, for $n = 4$, the incremental spectrum, which is a simple subtraction of the spectrum of the $n = 4$ model from that of the $n = 5$ model, effectively displays the vibrations pertaining to the middle residue in the DP = 5 model structure as the contributions from the penultimate and terminal residues are minimized (subtracted). It is shown that for several simulated bands displayed in the incremental spectra a practical convergence of position and IR band strength is obtained for the limited size (DP) models used. Such converged or extrapolated simulated spectral features are representative of DP $\sim \infty$ cellulose within the premises and limitations of the *overall* model.

The results show that this procedure is a reasonable, first-order approximation. Its usefulness is exemplified by showing how cellulose I_α, I_β and crystallinity "marker" bands can be qualitatively related to the nature of corresponding predicted vibrations. It is also found that not all vibrational modes lend themselves to simple interpretation.

Methods

Microcrystalline cellulose (from Merck), which is derived from wood sources and mainly consists of the I_β form, was used as cellulose reference material. A cellulose spectrum was obtained as the average of three spectra (100 scans, 4 cm^{-1}) obtained using a dry air purged Nicolet 6700 FT-IR spectrometer equipped with an ATR (Golden Gate) unit maintained at $T =$ 30 °C. Dispersion effects were corrected for by the spectrometer software using cellulose refractive index $n = 1.46$ obtained by (exponential decay) extrapolation of $n(\lambda)$ to the mid-IR range.[18] This corrects both for the wavelength dependence of penetration depth and for dispersion induced shifts of band positions.

The Gaussian 03W (Rev. D.01) and Gaussview 3.0 software were used for the calculation of vibrational properties and displaying the vibrational nature of normal modes, respectively.[19] All DFT calculations employed the default (75, 302) grid, and tight geometry convergence criteria ($<10^{-6}$ au rms force, $<4 \times 10^{-6}$ au rms displacement). All calculations used tight SCF convergence ($<10^{-8}$ au rms density matrix).

The model chain dihedral angles, i.e., $\beta(1\rightarrow4)$ linkage, exocyclic MeOH, and pyranose ring OH groups, all reflect the crystal structure values. The coordinates were adopted from the I_β crystal structure,[20] and two chains were constructed to reflect the two H bonding networks,[21] i.e., network A (NWA) where the hydroxymethyl oxygen is an acceptor in an *intra*molecular H bond and network B (NWB) where it is acceptor in an *inter*molecular H bond; see Figure 1. The models were built of $n = 2$, 3, 4, or 5 $\beta(1\rightarrow4)$ linked pyranose units, which at the terminal positions were substituted with an OMe group to mimic a glycosidic linkage.

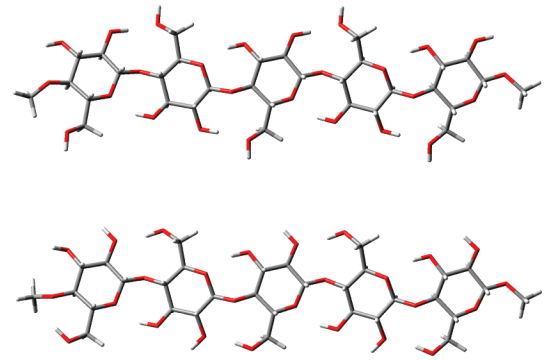

Figure 1. Cellulose chain models ($n = 5$) adopting a H bonding network A (top) and B (bottom).

OMe group conformations were chosen on the basis of conformers defined by the C_R-O dihedral angle φ, where C_R is the pyranose ring (R) C1 or C4 carbon. They were estimated by scanning the B3LYP/pc-1′ energy of a corresponding single ring fragment, as a function of φ, where the glycosidic linkage was replaced with a hydrogen, i.e., C1−H or C4−H. The basis set pc-1′ is the DFT optimized pc-1 basis set with all hydrogen p functions omitted.[22] It is of similar size and construction (3s2p1d on C and O, 2s on H) as the commonly used 6-31G(d) basis set but has been shown to be more adequate for DFT calculations.[23] Structures close to the local energy minima were then fully optimized at the same level, and frequency calculations confirmed these as true minima. These final conformers were then subjected to MP2(fc)/aug-cc-pVDZ single point energy calculations. The MP2 minimum energy conformer, i.e., of C1 or C4 substitution for a cellulose ring fragment, then provided the φ values used for the full model chains. The MP2 method was chosen in place of B3LYP for its expected better ability to account for weak dispersion interactions and conformational energies.[24] This choice is, however, of little practical consequence, as differences in simulated spectra due to different OMe group conformations both will be small and will effectively cancel out in the incremental spectra.

All model chains were fully B3LYP/pc-1′ geometry optimized, and harmonic frequency calculations provided vibrational frequencies and IR intensities. The frequency results confirmed all structures as true local energy minima. The data was converted to simulated IR spectra by the use of a Lorenzian line shape function (fwhm of 4 cm^{-1}). The single chain $n \rightarrow \infty$ asymptotic IR spectrum is approximated by extrapolating the *incremental* changes $n \rightarrow n + 1$ of the simulated IR spectra. The incremental IR spectrum is defined as $\delta I_{n+1} = I_{n+1} - I_n$, where I_{n+1} and I_n are the simulated IR spectra of models with chain length $n + 1$ and n. This representation allows for the identification of consistent spectral features and aims at reducing end group effects.

Results

In Figure 2, a cellulose spectrum is depicted together with its second derivative (2der) spectrum, which has been sign-

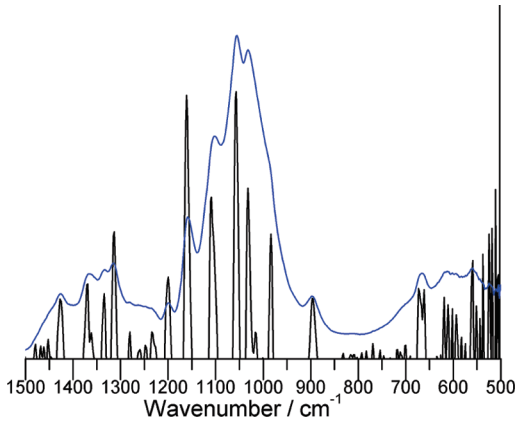

Figure 2. Experimental FT-IR spectrum of microcrystalline cellulose and its sign-inverted 2der spectrum.

inverted and scaled by a multiplicative constant. The effect of the dispersion correction on IR band positions as determined from the 2der spectra was small. Thus, the *largest* deviation is a 3 cm^{-1} downshift of the very strong (uncorrected) 1060 cm^{-1} band. The three individual spectra were practically indentical with band position differences <0.5 cm^{-1}. Hence, all results refer to the average spectrum.

Above ~600 cm^{-1}, where the 2der spectrum is not dominated by noise, the spectrum is resolved into many well-defined components, which can be compared with predictions.[25]

The predicted spectra are based on the following structural results.

The OMe end groups have minimum MP2 energy conformations (results not shown), which correspond qualitatively with the linkage dihedral angles. The effect of chain length n on structural parameters is small. For each chain, an average and standard deviation (STD) of each type of structural parameter was calculated and extrapolated average values for $n \rightarrow \infty$ considered, where error values of this extrapolation are estimated from the four $n = 2, \ldots, 5$ values (including STD). Thus, for linkage dihedral angles $\{\varphi, \psi\} = \{\varphi(\mathrm{O5C1O_LC4'}), \psi(\mathrm{C1O_LC4'C3'})\}$ convergence is achieved within ±0.5° with $\{\varphi, \psi\} \sim \{-91°, 95°\}_{\mathrm{NWA}}$ and $\{-91°, 93°\}_{\mathrm{NWB}}$, and for the linkage angle $\theta(\mathrm{C1O_LC4'})$, the values $\theta_{\mathrm{NWA}} = 119.4 \pm 0.1°$ and $\theta_{\mathrm{NWB}} = 119.0 \pm 0.1°$ are obtained. The C1C4 and the $\mathrm{C1(-O_L-)C4'}$ separation distances across a single pyranose ring and the linkage, abbreviated as d_{R} and d_{L}, respectively, are measures which mainly reflect bond lengths and crystallographic (c axis) repeat distances along the chain. These converge to the values $d_{\mathrm{R}}(\mathrm{NWA}) = 2.861 \pm 0.006$ Å, $d_{\mathrm{R}}(\mathrm{NWB}) = 2.886 \pm 0.006$ Å and $d_{\mathrm{L}}(\mathrm{NWA}) = 2.433 \pm 0.001$ Å, $d_{\mathrm{L}}(\mathrm{NWB}) = 2.426 \pm 0.001$ Å.

The predicted structural parameters correspond well with the experimental values, e.g., the center chain values are $\{\varphi, \psi\} = \{-89°, 95°\}$, $\theta = 116°$, $d_{\mathrm{R}} = 2.87$ Å, and $d_{\mathrm{L}} = 2.43$ Å, remembering that glycosidic linkage and exocyclic dihedral angles are sensitive to interchain interactions and differ up to ~10° between the origin and center chains.[20]

The behavior of the simulated IR spectra for the cellulose model chain structures is exemplified for the NWA ($n = 2$ up to $n = 5$) model chains in Figure 3. For each pyranose unit by which the model is extended, the majority of bands become increasingly well-defined and increase steadily in intensity. For some bands, however, there is practically no change in intensity, e.g., the group of bands at ~390, 430, and 460 cm^{-1}. These are all caused by vibrations with significant contribution of the τOH (torsional mode) of the C2OH or C3OH adjacent to the

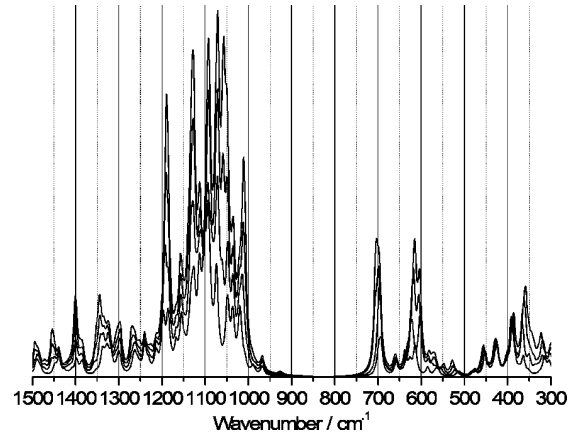

Figure 3. Predicted IR spectra for the four NWA model chains $n = \{2, 3, 4, 5\}$. The X axis is *not* scaled, and the intensity scale (Y axis) is arbitrary but identical for all four model chains.

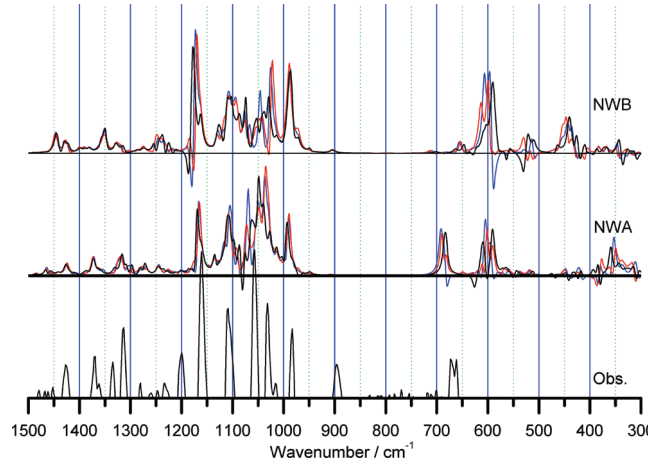

Figure 4. Incremental predicted IR spectra (δI_3, black line; δI_4, blue line; δI_5, red line) for the model chains (the X axis is scaled by $f = 0.98$). The sign-inverted 2der ATR-IR spectrum is depicted for comparison. The intensity scale is arbitrary but identical for NWA and NWB.

MeO end groups. Thus, by depicting instead the incremental spectra, such spectral contributions caused by vibrations localized near the ends of the model chains effectively cancel out. The convergence limit $n \rightarrow \infty$ has effectively been reached if the incremental spectrum is invariant toward adding another pyranose unit to the model chain. In that case, a simulated spectrum more representative of cellulose has been obtained.

In Figure 4, the observed (>650 cm^{-1}) and predicted incremental spectra are depicted. Predicted intensities are qualitative, as the pc-1′ basis set in this respect cannot be expected to perform well, since it contains no diffuse functions. Table 1 lists observed and predicted positions and qualitative assignments. The table also includes the observed band positions of I_β Valonia cellulose as reported by Maréchal and Chanzy.[17] The six characteristic bands observed in the 850−1200 cm^{-1} range at 896, 983, 1032, 1057, 1110, and 1161 cm^{-1} (2der peak positions) provide an optimum wavenumber scaling factor f by assigning these to predicted bands. The practically identical $f = 0.978$ and $f = 0.982$ for NWA and NWB, respectively, minimized the averaged rms error of their predicted positions to ~10 cm^{-1} in both cases. All predicted positions are henceforth obtained as $f = 0.98$ scaled values.

The scale of wavenumbers 300−1500 cm^{-1} facilitates a full comparison of band positions and relative strengths between dominant predicted and observed bands. Figure 5 provides a

TABLE 1: Predicted IR Band Frequencies for NWA (A) and NWB (B) Assigned to Observed Bands of This Work and of ref 17[a]

A, cm^{-1}	assignment	B, cm^{-1}	assignment	obs, cm^{-1}	ref 17
1465	C2, C3: υCC, CH and OH wag	1450	R: CH and OH wag		1450
1425	all: CH and OH wag	1425	MeOH: CH and OH wag	1427	1430
1375	C3, C4: υCC, CH wag			1370	
		1350	R: CH and OH wag	1335	1335
1320	all: CH and OH wag	1325	all: CH and OH wag	1314	1315
1170	L: υCO asym	1170	L: υCO asym	1161	1160
1110	complex	1105	complex	1110	1115
1050	MeOH: υCO	1080	MeOH: υCO	1057	1060
1040	R: υC5O	1020	R: υC5O	1032	1035
990	R: υC1O	990	R: υC1O	983	(1000)
910	MeOH: CH rock	905	MeOH: CH rock	896	
720	C1: umbrella	720	C1: umbrella	710	705
690	C2: τOH			670	665
650	L: δCOC	655	L: δCOC	610	
600	C3: τOH	600	C3,C6: τOH	550	
~350	C6: τOH	~450	C2: τOH		

[a] Note that the figures of reference[17] were rounded to 5 cm^{-1} in that work.

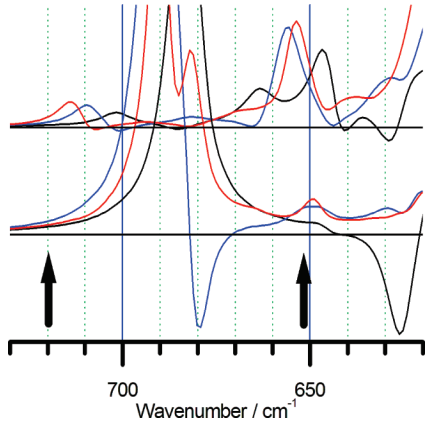

Figure 5. Close-up on the incremental spectra of Figure 4. The color coding is the same as that for Figure 4. The *Y* axis has been scaled by a factor of 10. The arrows indicate extrapolated positions of the relatively weak linkage δCOC (650 cm^{-1}) and C1 umbrella (720 cm^{-1}) vibrations.

close-up of a region of the predicted spectra with relatively weak but important bands. The intensity scale is here enhanced by a factor of 10 (relative to Figure 4), and two weak bands at ~650 and 720 cm^{-1} can be more easily observed; see Table 1. Both bands are significantly weaker for NWA compared to NWB, and the NWA 720 cm^{-1} band is effectively buried in the tail of a strong τOH band.

Discussion

Whereas the structural parameters behave well in respect to convergence with chain length, the vibrational properties, i.e., band positions and intensities, show slower convergence depending on the type of vibration. The necessity of correcting for chain end effects on vibrational properties is born out clearly by comparing Figures 3 and 4, and the incremental spectrum representation depicted in Figure 4 is moreover effectively normalized to the single pyranose unit.

The positions of the most characteristic bands are well predicted, and for many bands, the δI_3 and δI_4 spectra are nearly identical, indicating that the δI_4 spectrum is not far from convergence. In the 850−1200 cm^{-1} interval, the NWA and NWB spectra are similar except for a few bands in the 1050−1100 cm^{-1} interval which appear less well converged. The observed 1057 cm^{-1} band is tentatively assigned to the predicted NWA 1050 or NWB 1045 cm^{-1} band. In the interval 900−1500 cm^{-1}, the 2der observed spectrum corresponds well with the NWA spectra, which is believed to be dominant for the I_β form,[21] and superposition with the NWB spectra may improve this. Thus, three well-defined bands observed at 1315, 1335, and 1370 cm^{-1} may correspond to the 1325 cm^{-1} (NWA, NWB), the 1350 cm^{-1} (NWB), and the 1375 cm^{-1} (NWA) bands, respectively.

Such detailed assignments remain conjectural at the present level of theory. Several factors are limiting, most importantly the relatively small basis set, which gives at most qualitative information on IR intensities, and the fact that the model does not account for interchain interactions and how they perturb band positions and intensities. The fact that the predicted NWA and NWB spectra differ significantly in the 1200−1500 cm^{-1} interval suggests that vibrational modes in this interval are not only sensitive to the single chain conformation but also to interchain interactions. Considering these limitations, it is not known whether the same type of vibrational mode, e.g., the CH and/or OH wagging dominated modes present in the 1300−1500 cm^{-1} interval, have similar or very different *intrinsic* associated intensities when these are compared between NWA and NWB at a better level of theory. The intrinsic intensity is proportional to the change of dipole moment coupled with the vibration of interest, and is by the incremental approach effectively intensity *per* glucopyranose unit. The reliable quantitative prediction of intensities would be an immense task, and is outside the scope of the present work.

The most dominant observed IR bands are tentatively assigned in Table 1. A main problem concerning assignments in the 1000−1500 cm^{-1} range is, however, the fact that many cellulose vibrations are composed of bending and stretching contributions from most bond types. This can make it *in principle* impossible to assign specific νCO vibrations, e.g., νCO of secondary (C2−O2H or C3−O3H) or primary ($C6H_2$−OH) alcohols, to observed IR bands, although they may of course contribute partially. Inspection of the vibrational nature of the normal modes shows this to be true to varying degrees. The interval 1000−1100 cm^{-1} is especially problematic also in relation to convergence of the simulated incremental spectra.

In a relatively recent experimental work, most of the bands in Table 1 were assigned.[17] The present work is in good accordance with these assignments except for the 1000−1100 cm^{-1} interval. It is possible to find predicted νCO vibrations of

C2−O2H and C3−O3H in this interval, but they contribute only partially to vibrational modes responsible for simulated bands at 1095 and 1010 cm^{-1}, respectively, for both NWA and NWB. The vibrational mode with nearly full νCO contribution of $C6H_2-OH$ is predicted at 1050 or 1080 cm^{-1}; see Table 1. The order of these three band types is in the present work predicted as 1010 cm^{-1} (νC3O3) < 1050 cm^{-1} −1080 cm^{-1} (νC6O6) < 1095 cm^{-1} (νC2O2), whereas in the recent experimental work they are suggested as 1000−1035 cm^{-1} (νC6O6) < 1060 cm^{-1} (νC3O3) < 1115 cm^{-1} (νC2O2), i.e., νC6O6 and νC3O3 interchanged. The νC2O2 and νC3O3 modes show practically no variation of their predicted positions between NWA and NWB, whereas the νC6O6 position is predicted to vary by ∼30 cm^{-1}. This variation may be taken as an upper limit estimate of the degree of band position variation of νCO vibrations due to *additional* interchain H bonding interactions, since the single chain model does not account for the fact that a COH may also be a donor or acceptor in an interchain H bond. This error source, due to the expected interchain coupling effects, must be added to the rms fitting error of 10 cm^{-1} determined from the most marked predicted and observed IR bands (see the Results section). It can therefore not be excluded that the relatively low rms error of 10 cm^{-1} results from error cancellation effects. Thus, it appears that only the νC3O3 mode position is not in reasonable correspondence to its presumed experimental position. This can, however, be related to the fact that *no predicted mode exists* with a predominant νC3O3 contribution. The assumption of the existence of a vibrational mode with dominant νC3O3 contribution is a premise for discussing its assignment.

When the lower wavenumber interval 300−900 cm^{-1} is studied, the convergence and correspondence with observed bands are found more problematic, and δI spectra more sensitive to the NWA or NWB configuration. This last observation is related to the fact that the τOH (torsion) modes give rise to all high intensity bands in the 300−700 cm^{-1} interval. For both networks, the C3OH adopts a similar conformation and intramolecular H bonding to the ring oxygen. Its τOH mode, which for NWB is coupled to the C6 τOH, is thus found at the same ∼600 cm^{-1} position, whereas the C2 and C6 τOH modes differ significantly between NWA and NWB according to the bonding state of the OH groups in the *isolated* chain models; see Table 1. However, these positions may suffer corrections in a more extended model, which takes interchain interactions into account. Especially the NWA C6 τOH would be affected, as this group interacts *via* H bonding with another adjacent chain. The τOH modes of OH groups which in the present model are already engaged in strong H bonding would suffer relatively smaller perturbations.

The two relatively weak bands predicted at 650 and 720 cm^{-1}, see Figure 5 and Table 1, show less positional but a very large intensity dependence on the NWA or NWB conformation. These two vibrational types, i.e., linkage COC bending and C1 umbrella vibration, have to the best of the author's knowledge not been considered before. In addition to the τOH vibrations, the predictions bear out their existence in this region of the spectrum. Their positions and intensities may suffer corrections in a more extensive model of cellulose. The 720 cm^{-1} vibration will be considered in more detail below.

Whereas bands contained within the interval 1100−1000 cm^{-1} show a high degree of overlap and are difficult to characterize in terms of simple group or bond vibrations, the bands within the 1500−1100 and <1000 cm^{-1} intervals are better defined and separated. The typical cellulose "marker bands", which are used for "property measures" of cellulose, are placed within these intervals.

In the following, two cases of cellulose "marker" bands, often used and discussed in the literature, will be examined.

I_α and I_β bands at 750 and 710 cm^{-1}, respectively, are often mentioned but with different proposals as to their origin. The predictions make clear which vibrations can and cannot cause these bands. The proposal that a CH_2 rocking mode is responsible[16] can be dismissed, as this mode is predicted at ∼905−910 cm^{-1} for both NWA and NWB. The *next lowest* predicted mode is a vibration at 720 cm^{-1} (both) followed by the 690 cm^{-1} τOH mode (NWA) and the ∼650 cm^{-1} linkage bend mode (both). In accordance with the literature, the (doublet) band observed at 670 cm^{-1} is assigned a τOH mode, possibly the NWA C2 τOH mode.[14] The linkage bend mode appears too separated in position to explain any of the I_α and I_β bands. Therefore, it is proposed that these are caused by the 720 cm^{-1} vibration which is split into two components due to the different interchain couplings experienced in the two crystal forms. This mode is an out of phase "umbrella" type bending of the C−C1, O−C1, O−C1 and C−C4, C−C4, O−C4 bonds, respectively, which effectively contracts and expands the chain in the direction of the ring CH bonds. It can thus be sensitive to interactions with neighboring chains in the crystal, especially the sheet organization, which is the main distinguishing factor between the I_α and I_β forms.

A cellulose crystallinity index, defined, e.g., as the 1426 cm^{-1}/896 cm^{-1} absorbance ratio, measures intensity perturbations of one mode relative to the other. The predicted bands at ∼910 and ∼1425 cm^{-1} correspond well with these bands. The intensity of the 910 cm^{-1} mode is a *factor of 6* larger for the NWB as compared to the NWA, whereas the 1426 cm^{-1} band is predicted to be of similar intensity. The absence of crystalline environment, i.e., disordered chains or surface chains, is believed to force the H bonding toward NWB.[21] According to the predictions, this would imply a decrease of the ratio. This measure of crystallinity thus appears based on the sensitivity of the total 896 cm^{-1} band intensity, which is the superposed NWA and NWB intensities, to the relative occurrence of the NWA and NWB hydrogen bonding networks.

The nature of the vibrations of native cellulose is thus born out by fairly simple single chain models, and qualified suggestions as to the molecular origin of cellulose marker bands can be given. It appears that the crystallinity index reflects the relative occurrence of NWA and NWB, which modulates band intensities, and the I_α versus I_β marker bands reflect the different interactions with neighboring chain sheets, which presumably modulate both band position and intensities. As these two structural characteristics, hydrogen bonding networks versus crystal I_α or I_β phases, are only properly defined for multichain crystal models, the cellulose property measures should more appropriately be considered using such extended structure models.

Conclusions

The use of IR spectroscopy for analysis of cellulose containing samples can thus be supported and developed by the present DFT based approach. The strength of the use of isolated model chains together with the incremental simulated spectrum approach is demonstrated by the fact that it reproduces most of the features of the cellulose IR spectrum in the 1500−700 cm^{-1} interval, and identifies the vibrational nature of commonly used cellulose "marker bands". Apart from the DFT functional and limited size basis set, the method appears so far mainly limited

by the neglect of interchain interactions, and here most importantly the H bonding interactions. The procedure can be improved in at least two directions, i.e., (1) by higher quality DFT functionals and basis sets and/or (2) by extending the physical scale of the model to include more pyranose units ($n \gg 5$) arranged, e.g., in a crystal lattice model.

The last direction would quickly require lower levels of theory. Thus, a molecular mechanics force field (MMFF) could be applied alone for a multichain crystal model to predict absolute frequencies, etc., at this level of theory. However, absolute frequencies and intensities could be calculated by high level theory for single chain models, as in the present work, and low level corrections to these then calculated by comparing MMFF calculation results for the same single chain models with those for the more complex multichain crystal models. This would be a composite approach unifying two different levels of theory, one for obtaining more accurate absolute frequencies and intensities and a lower level for predicting the shifts or perturbations of these quantities due to the replacement of a less realistic single chain model with a more realistic multichain model, which takes interchain interactions explicitly into account.

References and Notes

(1) Klemm, D.; Heublein, B.; Fink, H.-P.; Bohn, A. Cellulose: Fascinating Biopolymer and Sustainable Raw Material. *Angew. Chem., Int. Ed.* **2005**, *44*, 3358–3393.

(2) Habibi, Y.; Lucia, L. A.; Rojas, O. J. Cellulose Nanocrystals: Chemistry, Self-Assembly, and Applications. *Chem. Rev.* **2010**, *110*, 3479–3500.

(3) Zugenmaier, P. Conformation and Packing of Various Crystalline Cellulose Fibers. *Prog. Polym. Sci.* **2001**, *26*, 1341–1417.

(4) Sugiyama, J.; Persson, J.; Chanzy, H. Combined Infrared and Electron-Diffraction Study of the Polymorphism of Native Celluloses. *Macromolecules* **1991**, *24*, 2461–2466.

(5) Horikawa, Y.; Sugiyama, J. Localization of Crystalline Allomorphs in Cellulose Microfibril. *Biomacromolecules* **2009**, *10*, 2235–2239.

(6) Akerholm, M.; Hinterstoisser, B.; Salmen, L. Characterization of the Crystalline Structure of Cellulose using Static and Dynamic FT-IR Spectroscopy. *Carbohydr. Res.* **2004**, *339*, 569–578.

(7) Nelson, M. L.; O'connor, R. T. Relation of Certain Infrared Bands to Cellulose Crystallinity and Crystal Lattice Type. Part II A New Infrared Ratio for Estimation of Crystallinity in Celluloses I and II. *J. Appl. Polym. Sci.* **1964**, *8*, 1325–1341.

(8) Kataoka, Y.; Kondo, T. FT-IR Microscopic Analysis of Changing Cellulose Crystalline Structure During Wood Cell Wall Formation. *Macromolecules* **1998**, *31*, 760–764.

(9) Stevanic, J. S.; Salmen, L. Orientation of the Wood Polymers in the Cell Wall of Spruce Wood Fibres. *Holzforschung* **2009**, *63*, 497–503.

(10) Watanabe, A.; Morita, S.; Kokot, S.; Matsubara, M.; Fukai, K.; Ozaki, Y. Drying Process of Microcrystalline Cellulose Studied by Attenuated Total Reflection IR Spectroscopy with Two-Dimensional Correlation Spectroscopy and Principal Component Analysis. *J. Mol. Struct.* **2006**, *799*, 102–110.

(11) Watanabe, A.; Morita, S.; Ozaki, Y. Study on Temperature-Dependent Changes in Hydrogen Bonds in Cellulose I Beta by Infrared Spectroscopy with Perturbation-Correlation Moving-Window Two-Dimensional Correlation Spectroscopy. *Biomacromolecules* **2006**, *7*, 3164–3170.

(12) Tsuboi, M. Infrared Spectrum and Crystal Structure of Cellulose. *J. Polym. Sci.* **1957**, *25*, 159–171.

(13) Mann, J.; Marrinan, H. J. Crystalline Modifications of Cellulose. Part II. A Study with Plane-Polarized Infrared Radiation. *J. Polym. Sci.* **1958**, *32*, 357–370.

(14) Liang, C. Y.; Marchessault, R. H. Infrared Spectra of Crystalline Polysaccharides. II. Native Celluloses in the Region from 640 to 1700 cm^{-1}. *J. Polym. Sci.* **1959**, *39*, 269–278.

(15) Higgins, H. G.; Steward, C. M.; Harrington, K. J. Infrared Spectra of Cellulose and Related Polysaccharides. *J. Polym. Sci.* **1961**, *51*, 59–84.

(16) Blackwell, J.; Vasko, P. D.; Koenig, J. L. Infrared and Raman Spectra of the Cellulose from the Cell Wall of Valonia Ventricosa. *J. Appl. Phys.* **1970**, *41*, 4375–4379.

(17) Maréchal, Y.; Chanzy, H. The Hydrogen Bond Network in Iβ cellulose as observed by infrared spectrometry. *J. Mol. Struct.* **2000**, *523*, 183–196.

(18) Kasarova, S. N.; Sultanova, N. G.; Ivanov, C. D.; Nikolov, I. D. Analysis of the Dispersion of Optical Plastic Materials. *Opt. Mater.* **2007**, *29*, 1481–1490.

(19) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G.E.; Robb, M. A.; Cheeseman, J. R.; Montgomery, J. A., Jr.; Vreven, T.; Kudin, K. N.; Burant, J. C.; et al. *Gaussian 03*, revision D.01; Gaussian, Inc.: Wallingford, CT, 2004.

(20) Nishiyama, Y.; Langan, P.; Chanzy, H. Crystal Structure and Hydrogen-Bonding System in Cellulose Iβ from Synchrotron X-ray and Neutron Fiber Diffraction. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.

(21) Nishiyama, Y.; Johnson, G. P.; French, A. D.; Trevor Forsyth, V.; Langan, P. Neutron Crystallography, Molecular Dynamics, and Quantum Mechanics Studies of the Nature of Hydrogen Bonding in Cellulose I_β. *Biomacromolecules* **2008**, *9*, 3133–3140.

(22) Jensen, F. Polarization consistent basis sets. II. Estimating the Kohn-Sham Basis Set Limit. *J. Chem. Phys.* **2002**, *116*, 7372–7379.

(23) Jensen, F. Polarization consistent basis sets. IV. The Basis Set Convergence of Equilibrium Geometries, Harmonic Vibrational Frequencies, and Intensities. *J. Chem. Phys.* **2003**, *118*, 2459–2463.

(24) Riley, K. E.; Holt, B. T. O.; Merz, K. M. Critical Assessment of the Performance of Density Functional methods for Several Atomic and Molecular Properties. *J. Chem. Theory Comput.* **2007**, *3*, 407–433.

(25) A reviewer pointed out that the production process of microcrystalline cellulose may introduce surface sulfation or carboxymethylation of the resulting product. However, a low level of absorption in the 1750−1600 cm^{-1} interval, where strong intrinsic absorption due to both sulfate and carboxymethyl groups is expected, ensures that this is not the case to any significant extent. The ratio of the maximum (ATR corrected) absorption in this interval to the strongest 1057 cm^{-1} and close lying 1430 cm^{-1} band is ~1/160 and ~1/30, respectively.

JP104213Z