Available online at www.sciencedirect.com
SCIENCE DIRECT®
Carbohydrate Research 340 (2005) 1165–1174
Carbohydrate RESEARCH

Determination of the three-dimensional structure of oligosaccharides in the solid state from experimental ^{13}C NMR data and ab initio chemical shift surfaces

Ivan Sergeyev and Guillermo Moyna*

Department of Chemistry and Biochemistry, University of the Sciences in Philadelphia, 600 South 43rd Street, Philadelphia, PA 19104-4495, USA

Received 23 November 2004; accepted 16 February 2005

Abstract—A novel method for the determination of the three-dimensional (3D) structure of oligosaccharides in the solid state using experimental ^{13}C NMR data is presented. The approach employs this information, combined with ^{13}C chemical shift surfaces (CSSs) for the glycosidic bond carbons in the generation of NMR pseudopotential energy functions suitable for use as constraints in molecular modeling simulations. Application of the method to trehalose, cellobiose, and cellotetraose produces 3D models that agree remarkably well with the reported X-ray structures, with ϕ and ψ dihedral angles that are within 10° from the ones observed in the crystals. The usefulness of the approach is further demonstrated in the determination of the 3D structure of the cellohexaose, an hexasaccharide for which no X-ray data has been reported, as well as in the generation of accurate structural models for cellulose II and amylose V_6.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: ^{13}C NMR spectroscopy; Ab initio calculations; Chemical shift surfaces; GIAO; Molecular modeling; NMR constraints; Oligosaccharide conformation

1. Introduction

Carbohydrates are perhaps the most versatile class of compounds in Nature, with functions ranging from energy storage in living systems and the control of cellular shape, to participation in cell and protein signaling processes.[1] Understanding the conformational preferences of these molecules is crucial to explain their function at the atomic level, and this information is extremely valuable for researchers in an equally diverse number of disciplines. For example, elucidating the solution dynamics of oligosaccharide epitopes is a requirement for the rational development of carbohydrate-based vaccines,[1–3] and knowledge of the structural characteristics of polysaccharides such as cellulose and its derivatives is desirable to those in the fiber, membrane, paint, and paper industries.[4] Therefore, the development of new methods capable of providing three-dimensional (3D) structural information on oligosaccharides and related molecules continues to be an active area of carbohydrate research.

Several well-established approaches that combine NMR spectroscopy and molecular modeling simulations are available to study oligosaccharide conformation and dynamics in solution, and a large number of carbohydrate and carbohydrate-containing molecules have been investigated using these methods.[5–8] On the other hand, the determination of accurate solid-state 3D structures for these molecules is somewhat limited to crystallographic techniques. As a result, and due in part to the difficulties associated with the preparation of crystalline samples suitable for X-ray studies, only a few hundred 3D structures of oligosaccharides at atomic resolution have been deposited in the Cambridge Structural Database (CSD).[9] In many instances, solid-state NMR provides an alternative route for the investigation of carbohydrate samples not amenable to X-ray crystallography. It has

* Corresponding author. Tel.: +1 215 596 8526; fax: +1 215 596 8543; e-mail: g.moyna@usip.edu

0008-6215/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.carres.2005.02.022

long been known that the chemical shifts of the glycosidic bond carbons hold a periodic dependency with the ϕ and ψ glycosidic linkage angles.[10–13] Since these dihedral angles are the principal determinants of the global fold of oligosaccharide chains, the relationship has been utilized in combination with ^{13}C NMR data from cross-polarization magic angle spinning (CP/MAS) experiments to study, at least qualitatively, the 3D structure of repetitive carbohydrates in the solid state.[12–15] A vast number of morphological studies of cellulose and amylose polymorphs by CP/MAS ^{13}C NMR spectroscopy attest to the predictive power of this approach.[14–19] Recently, we have shown that accurate equations relating the variation of the chemical shifts of the glycosidic carbons with the glycosidic linkage conformation can be derived through the use of ab initio calculations.[20] These empirical functions, of mathematical form $^{13}C\delta = f(\phi,\psi)$ and which we refer to as chemical shift surfaces (CSSs), can be used to accurately back-calculate the glycosidic carbon chemical shifts of oligosaccharides in the solid state.[20,21] As it has been the case for other empirical equations relating NMR parameters with molecular geometry, we have postulated that CSSs could be employed to derive structural information from experimental ^{13}C NMR chemical shifts. In this report we describe an approach that combines CP/MAS ^{13}C NMR measurements, pseudopotential energy functions based on CSSs, and molecular modeling simulations in the generation of accurate 3D structural models of oligosaccharides in the solid state, and present its validation on a number of systems (Fig. 1).

The results outlined below indicate that the method yields 3D models that agree remarkably well with reported X-ray structures in cases where these are available. In addition, sound models were obtained for oligosaccharides lacking crystallographic data and for the polysaccharides cellulose and amylose. Most importantly, our findings serve to further demonstrate the potential of ^{13}C chemical shifts as accurate sources of structural information in the study of the conformational preferences of oligosaccharides and related molecules.

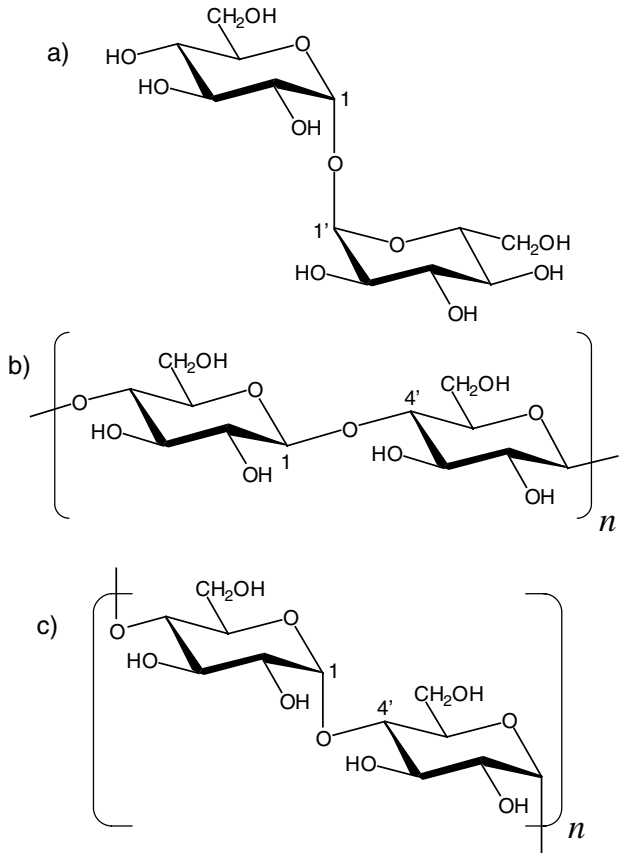

Figure 1. Chemical structures of trehalose (a), cellobiose (b, $n = 1$), cellotetraose (b, $n = 2$), cellohexaose (b, $n = 3$), cellulose (b, $n > 100$), and amylose (c, $n > 100$).

2. Computational methods

2.1. Chemical shift surface derivation

The methodology employed in the derivation of the CSSs for the glycosidic carbons of the α-(1→1), α-(1→4), and β-(1→4)-D-Glc*p* linkages has been described in detail elsewhere,[20] and is briefly presented here. For each model D-Glc*p*-D-Glc*p* disaccharide, a matrix in the [180°↔180°] range for ϕ and ψ at 20° intervals was built, resulting in a grid with 324 conformers. While maintaining the monosaccharides in the $^4C_1$ conformation and the $\langle\phi,\psi\rangle$ angles frozen with dihedral constraints on heavy atoms, the geometry of each structure in the grid was optimized using the AM1 semiempirical Hamiltonian (Spartan 5.0.1, Wavefunction, Inc.). The ab initio ^{13}C chemical shieldings for the optimized structures were computed with the Gauge-Including Atomic Orbital (GIAO) method as implemented in Gaussian 98,[22] utilizing Hartree–Fock (HF) theory and a 321G basis set. The results obtained for the glycosidic carbons were converted into chemical shifts by subtracting the chemical shielding computed at the same level of theory for the methyl carbons of tetramethylsilane (TMS), and subsequently scaled to results from a series of reference 6-311G** level calculations. The raw ab initio chemical shift data were then fitted to trigonometric series expansions of the form (Eq. 1):

$$\begin{aligned}{}^{13}C\delta(\phi,\psi) = A_0 &+ \sum_i \left[A_i\cdot\sin(i\cdot\phi) + B_i\cdot\cos(i\cdot\phi)\right.\\ &\left. + C_i\cdot\sin(i\cdot\psi) + D_i\cdot\sin(i\cdot\psi)\right]\\ &+ \sum_{i,j,\alpha,\beta}\left[A_{i,j,\alpha,\beta}\cdot\sin(i\cdot\alpha)\cdot\cos(j\cdot\beta)\right.\\ &+ B_{i,j,\alpha,\beta}\cdot\sin(i\cdot\alpha)\cdot\sin(j\cdot\beta)\\ &\left. + C_{i,j,\alpha,\beta}\cdot\cos(i\cdot\alpha)\cdot\cos(j\cdot\beta)\right] \qquad (1)\end{aligned}$$

where A_0, A_i, B_i, C_i, D_i, $A_{i,j,\alpha,\beta}$, $B_{i,j,\alpha,\beta}$, and $C_{i,j,\alpha,\beta}$ are the fit parameters and α and β can be either ϕ or ψ. The i and j indexes were varied between 1 and 3 to give the 91-term series used in the fits, which were carried out with Mathematica 4.1 (Wolfram Research, Inc.). Graphical representations of the resulting functions are presented in Figure 2.

2.2. Molecular mechanics simulations

Models for all the oligosaccharides considered in the study were built and analyzed with Sybyl 6.8 (Tripos Associates, Inc.). Relaxed potential energy surfaces (PESs) in $\langle\phi,\psi\rangle$-space for trehalose (D-Glc*p*-α-(1→1)-D-Glc*p*) and cellobiose (D-Glc*p*-β-(1→4)-D-Glc*p*) were computed in vacuo with the MMFF94 force field.[23] As done in the generation of the input structures for the GIAO chemical shielding calculations, 20° steps were used for the ϕ and ψ angles, and the conformations of the monosaccharide units were held in the $^4C_1$ conformation with dihedral constraints on heavy atoms throughout the derivation of the PESs.

Conformer ensembles consisting of 2000 structures for cellotetraose and 4000 structures for cellohexaose

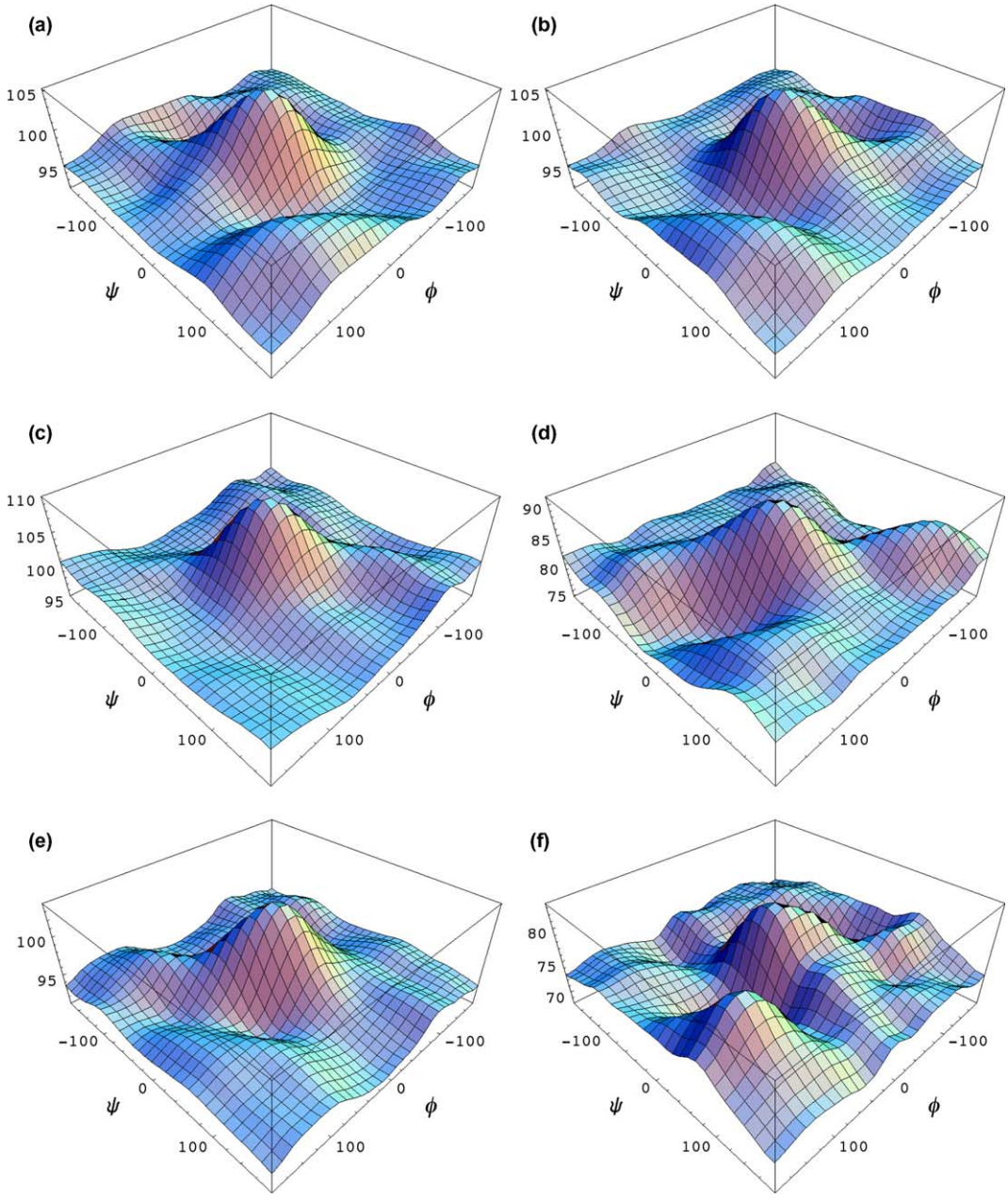

Figure 2. ^{13}C CSSs for the carbons of the α-(1→1) (a and b), β-(1→4) (c and d), and α-(1→4) (e and f) D-Glc*p* glycosidic linkages as a function of the ϕ and ψ dihedral angles, defined as ⟨H-1–C-1–O-1–C-*n*′⟩ and ⟨C-1–O-1–C-*n*′–H-*n*′⟩, respectively.

and maltohexaose were obtained through simulated annealing (SA). In all cases, each SA cycle involved a 1 ps interval of equilibration to 1000 K, followed by a 1 ps period of exponential cooling to 200 K. The resulting structures were then energy minimized to a gradient below 0.05 kcal/mol. The same force field and conditions used in the derivation of the PESs described above, including constraining the monosaccharides in the $^4C_1$ conformation, were employed for these simulations.

2.3. Generation of 3D structural models from CP/MAS ^{13}C NMR data

In order to take into account the chemical shifts of the glycosidic linkage carbons in the molecular modeling simulations, the potential energy function of the force field was augmented with an NMR pseudopotential energy term (Eq. 2):

$$E_{\text{Total}} = E_{\text{FF}} + \sum_i E_{^{13}\text{C}\delta i} \tag{2}$$

E_{Total} represents the augmented potential energy function of the system, E_{FF} is the standard force field potential (MMFF94 in our case), and $E_{^{13}\text{C}\delta i}$ is the NMR pseudopotential energy, or constraint, for the ith glycosidic carbon being considered. These constraints are defined as simple quadratic well penalty functions of the form (Eq. 3):

$$E_{^{13}\text{C}\delta} = \begin{cases} k \cdot [^{13}\text{C}\delta_{\text{exp}} - {}^{13}\text{C}\delta(\phi,\psi)]^2 \\ \quad \text{if } |^{13}\text{C}\delta_{\text{exp}} - {}^{13}\text{C}\delta(\phi,\psi)| > \Delta\delta \\ 0 \quad \text{if } |^{13}\text{C}\delta_{\text{exp}} - {}^{13}\text{C}\delta(\phi,\psi)| \leqslant \Delta\delta \end{cases} \tag{3}$$

where $^{13}\text{C}\delta_{\text{exp}}$ is the experimental chemical shift of the glycosidic carbon (i.e., the target value of the constraint), $^{13}\text{C}\delta(\phi,\psi)$ is the empirical function representing the CSS (Eq. 1), k is a force constant that determines the weight of the constraint, and $\Delta\delta$ defines the limits of the quadratic well. This type of function, commonly used to incorporate constraints based on NMR parameters into molecular modeling simulations, has no effect when the calculated parameter is within the specified range, but imposes an energy penalty that steepens progressively as the estimation deviates further from the target value. Thus, its effect is to guide the simulation toward conformers which best reproduce the experimental data. The force constant k is usually chosen so as to give the constraints and the remaining terms of the force field potential comparable weights. In all our studies k was set to 10.0 kcal/mol, and chemical shifts were constrained to a 1.0 ppm range centered around their experimental values ($\Delta\delta$ = 0.5 ppm).

For trehalose and cellobiose, E_{Total} was minimized in $\langle\phi,\psi\rangle$-space with the 'FindMinimum' routine available in Mathematica 4.1, using the appropriate MMFF94 PESs to represent the E_{FF} force field potential (vide supra). For the larger oligosaccharides, $E_{^{13}\text{C}\delta i}$ was computed using the ϕ and ψ angles from each conformer in the ensembles obtained by SA, and applied as a correction to the total energy of the structures. For each ensemble, the 10 conformers with the lowest E_{Total} obtained from this process were averaged. With the exception of the glycosidic linkage ϕ and ψ angles, the remaining bond lengths, angles, and torsions on the average structures were allowed to relax, and the resulting models employed for the comparisons discussed in the following sections. In all cases, the $^{13}\text{C}\delta_{\text{exp}}$ target values used in the calculations came from CP/MAS ^{13}C NMR measurements found in the literature.[12,14,24]

3. Results and discussion

3.1. Method description and evaluation on disaccharides

As evidenced by Figure 2 and Eq. 1, the functions defining the CSSs are not invertible. In other words, use of widely different $\langle\phi,\psi\rangle$ angle pairs as input for these equations could lead to identical ^{13}C chemical shift estimations. This is clearly seen in Figure 3a, which shows contours on the CSS for the anomeric carbon of trehalose (Fig. 1a) plotted at ±0.5 ppm from the experimentally measured chemical shift of 93.0 ppm.[24] Any of the $\langle\phi,\psi\rangle$ angle pairs in the region delimited by the two contours will yield ^{13}C chemical shift estimates, which are within ±0.5 ppm from experiment. Therefore, a single chemical shift cannot be employed to derive accurate information regarding the conformation of the glycosidic linkage. However, the range of possible $\langle\phi,\psi\rangle$ angles can be substantially reduced if the CSSs of both glycosidic linkage carbons are taken into account simultaneously. For trehalose, the C-1 and C-1′ carbons are identical due to internal symmetry, and thus the CSS for one is the transpose of the other and vice versa (i.e., $^{13}\text{C}\delta_{\text{C-1}}(\phi,\psi) = {}^{13}\text{C}\delta_{\text{C-1}'}(\phi,\psi)^{\text{T}}$, Figs. 2a,b and 3a,b). As expected, if the two CSS plots depicting contours at ±0.5 ppm from the experimental C-1 and C-1′ chemical shifts are intersected, the resulting $\langle\phi,\psi\rangle$ region which satisfies both chemical shift estimations, is considerably smaller (Fig. 3c). This intersection is in practice the graphical representation of the constraint function $E_{^{13}\text{C}\delta i}$ presented in Eq. 3.

It is worth noting that the ϕ and ψ angles obtained from trehalose X-ray structures are both $-51.9 \pm 10.3°$ on average[25–27] and belong to the intersection highlighted in Figure 3c. Finally, superposition of this region on the MMFF94 PES plotted at 1.5 kcal/mol from the global energy minimum leads to a further reduction of the region of allowed ϕ and ψ angles (Fig. 4a). This is indeed a simplified depiction of E_{Total} (Eq. 2), which after full minimization as described earlier, leads to $\phi = \psi = -49.0°$. These values agree well with the

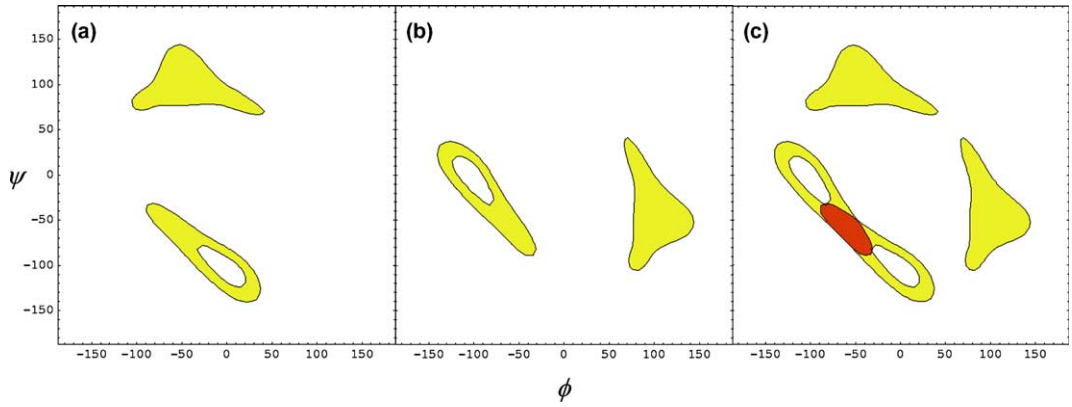

Figure 3. CSS contour plots for the D-Glc*p*-α-(1→1)-D-Glc*p* C-1 and C-1′ carbons (a and b, respectively). Areas holding $\langle\phi,\psi\rangle$ angle pairs giving chemical shift estimations within ±0.5 ppm from experiment are shown in yellow. The region of $\langle\phi,\psi\rangle$-space obtained from the intersection of these two surfaces is shaded orange (c).

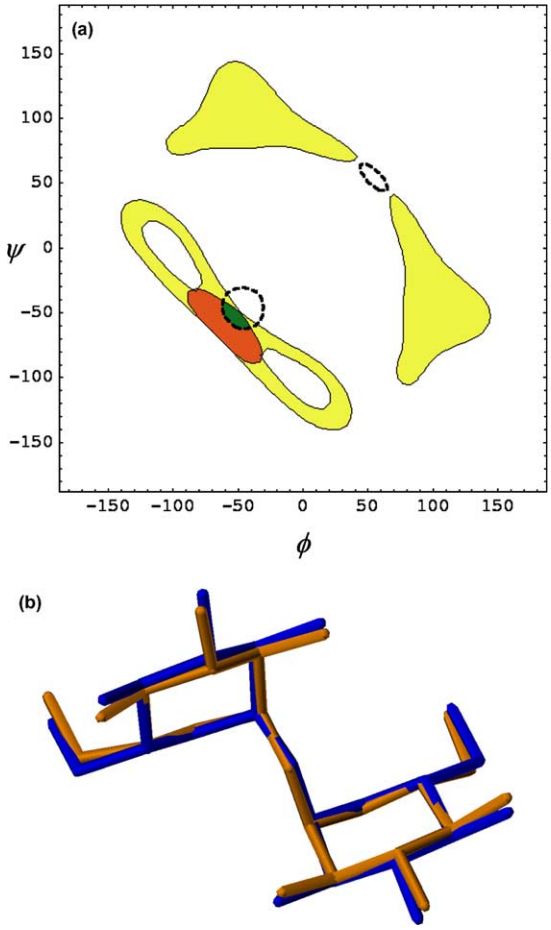

Figure 4. (a) Superposition of the region of $\langle\phi,\psi\rangle$-space presented in Figure 3c on the 1.5 kcal/mol contour of the MMFF94 PES for trehalose (dashed line). (b) 3D structure of trehalose obtained using ^{13}C chemical shift constraints (blue), superimposed to its average crystal structure (orange). The heavy-atom rmsd between the two is 0.14 Å.

glycosidic linkage dihedral angles measured on crystal structures (vide supra). Similarly, the root-mean-square deviation (rmsd) between heavy atoms of the trehalose 3D model derived using ^{13}C chemical shift constraints and the average X-ray structure is less than 0.14 Å (Fig. 4b).

While trehalose serves to describe the methodology, internal symmetry makes it a rather simple and well-behaved system. On the other hand, cellobiose (Fig. 1b) possesses no such symmetry and requires consideration of two independent CSSs for the C-1 and C-4′ carbons (Fig. 2c,d). As seen in Figure 5a,b, contours plotted at ±0.5 ppm from the chemical shifts measured by CP/MAS yield very similar regions of $\langle\phi,\psi\rangle$-space.[14] In addition, their intersection gives two distinct areas with glycosidic linkage angles that can satisfy experimental chemical shifts for C-1 and C-4′ simultaneously (Fig. 5c).

Despite this, superposition of this intersection with the 1.5 kcal/mol energy level contour of the MMFF94 PES for cellobiose results in a single and smaller area of $\langle\phi,\psi\rangle$-space (Fig. 6a). Minimization of the associated E_{Total} potential leads to a unique conformer with ϕ = 40.6° and ψ = −16.4°, both of which compare well with the average angles computed from X-ray structures (ϕ = 41.2 ± 3.9° and ψ = −12.9 ± 4.6°).[28–30] As was the case for trehalose, the rmsd for heavy atoms between the cellobiose structural model obtained from ^{13}C chemical shift constraints and the average X-ray structure is only 0.12 Å (Fig. 6b).

3.2. Application to linear β-(1→4)-linked D-Glc*p* oligomers

The approach presented above can only be employed for disaccharides with well-defined PESs. In larger oligosaccharides, the internal energy of the system is not only the combination of the individual PES for the each glycosidic linkage present in the molecule, but also includes non-bonded interactions between atoms in non-contiguous residues.[6] In order to optimize the conformation of these molecules, energy minimizations with suitable

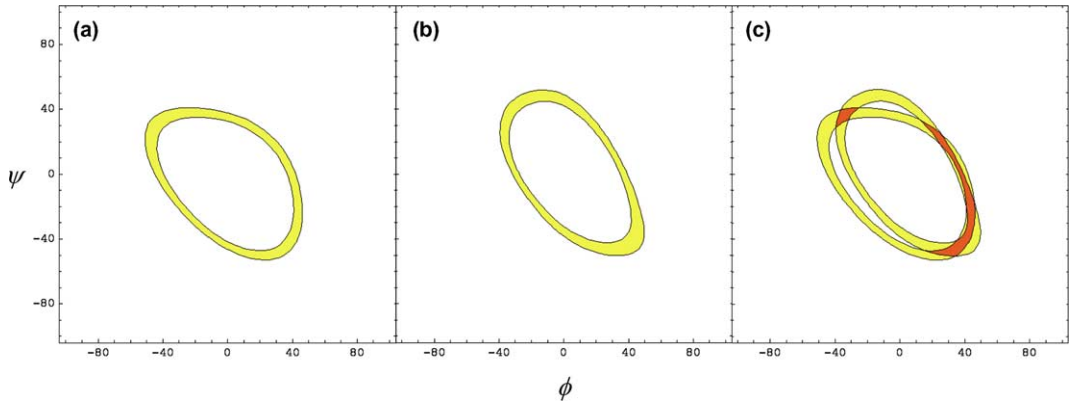

Figure 5. CSS contour plots for the D-Glc*p*-β-(1→4)-D-Glc*p* C-1 and C-4′ carbons (a and b, respectively). $\langle\phi,\psi\rangle$ angle pairs yielding chemical shift estimations within ±0.5 ppm from experiment are found in the areas shaded in yellow. The regions of $\langle\phi,\psi\rangle$-space obtained from the intersection of these two surfaces is shown in orange (c).

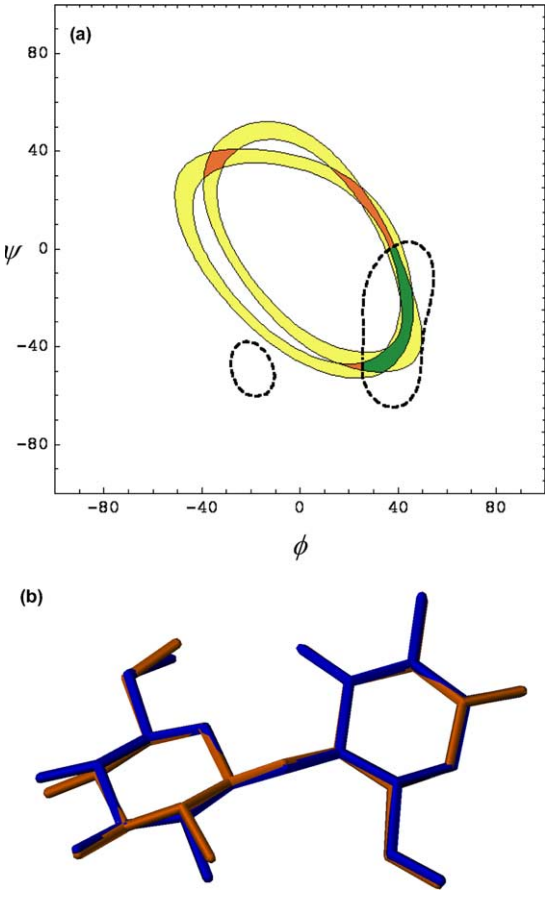

Figure 6. (a) Overlay of the $\langle\phi,\psi\rangle$-space region presented in Figure 5c on the 1.5 kcal/mol contour of the MMFF94 PES for cellobiose (dashed line). (b) 3D structure of cellobiose obtained using ^{13}C chemical shift constraints (blue), superimposed to its average crystal structure (orange). The heavy-atom rmsd between the two is 0.12 Å.

force fields are required. Ideally, the chemical shift constraints derived from the CSSs described above can be incorporated into the force field potentials to guide the geometry optimization process toward conformers satisfying the experimental ^{13}C NMR data.[31,32] However, the complexity of the equations defining the CSSs makes this task a considerable software development undertaking. Furthermore, lack of source code renders the integration of these pseudopotentials with commercial molecular modeling packages unfeasible. We thus followed an alternative route to take ^{13}C chemical shifts into account in the generation of oligosaccharide 3D models. First, the conformational space available to these molecules was sampled extensively through SA. Subsequently, the ϕ and ψ dihedral angles derived from the resulting conformers were used as input for the CSSs to calculate theoretical ^{13}C chemical shifts for the glycosidic linkage carbons. These estimations were employed together with the corresponding experimental CP/MAS ^{13}C NMR data to compute energy penalties according to Eq. 3, which were then added to the force field energy of each conformer in the ensembles. The conformers with the lowest total energy were selected and averaged as described in the previous section to give the final oligosaccharide 3D models.

Figure 7a shows the structural model obtained as outlined above for cellotetraose (Fig. 1b), superimposed to the average of the two structures reported for the crystal unit cell of the tetrasaccharide.[33] The mean ϕ and ψ angles for the three glycosidic linkages for the conformer derived from ^{13}C chemical shifts are 25.8 ± 6.8° and −28.9 ± 1.9°, respectively, which compare well with the corresponding average angles computed from the crystal structure (ϕ = 27.3 ± 2.4° and ψ = −25.5 ± 4.8°), and the rmsd for heavy atoms between the two structures is only 0.25 Å. Most importantly, use of these average dihedral angles as input for the D-Glc*p*-β-(1→4)-D-Glc*p* CSSs gives back-calculated chemical shifts for C-1 and C-4′ of 107.2 and 88.7 ppm, respectively. These two estimates are in excellent agreement with the averages obtained from experimental CP/MAS measurements (107.4 and 89.7 ppm).[14]

For comparison purposes, the lowest energy conformer obtained by SA without the consideration of

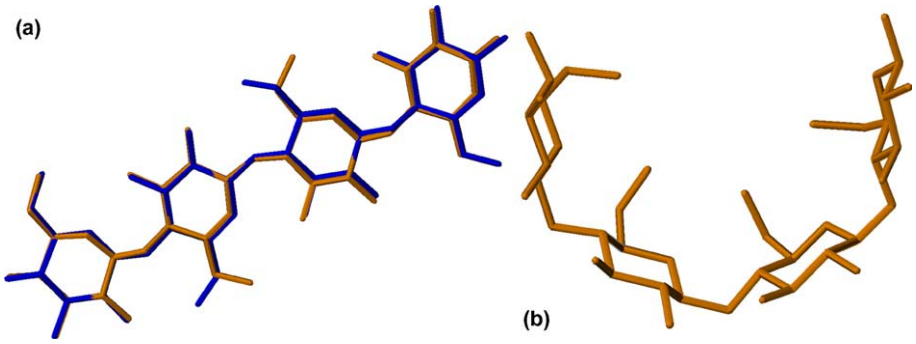

Figure 7. (a) Superposition of the cellotetraose structural model derived using ^{13}C chemical shifts constraints (blue) and its average X-ray structure (orange). (b) Lowest energy conformer for cellotetraose obtained from SA without the consideration of chemical shift constraints.

^{13}C chemical shifts constraints is shown in Figure 7b. The ϕ and ψ angles for all glycosidic linkages in this structure are virtually identical, and centered around 177.3° and 3.3°, respectively. While these values correspond to a local minima in the MMFF94 PES of the β-(1→4)-D-Glc*p* linkage, they deviate considerably from the angles found by X-ray crystallography. This is not surprising, as in vacuo molecular modeling simulations cannot properly reproduce crystal packing forces, and usually yield results which are highly dependent on the force field potentials employed.[6] Furthermore, if these angles are used with the appropriate CSSs for the back-calculation of the C-1 and C-4′ chemical shifts, estimates of 98.7 and 76.9 ppm are obtained, respectively, both of which depart significantly from experiment.[14]

We then focused our attention on cellohexaose (Fig. 1b), an hexasaccharide lacking crystallographic data but for which high-resolution CP/MAS ^{13}C NMR measurements have been reported.[14] An analogous approach to the one followed for cellotetraose was used in the derivation of a structural model in this case, the only exception being that a larger number of SA cycles were employed for the generation of the initial conformational ensemble. The resulting 3D structure for the hexasaccharide is shown in Figure 8a, and has $\phi = 31.3 \pm 5.2°$ and $\psi = -35.0 \pm 7.1°$. Although this model cannot be compared directly to a high-resolution X-ray structure, use of the computed ϕ and ψ dihedral angles and the corresponding CSSs in the back-calculation of the C-1 and C-4′ chemical shifts results in predictions of 106.2 and 87.9 ppm, respectively, which compare well with data from CP/MAS ^{13}C NMR experiments (107.1 and 89.3 ppm). In addition, cellotetraose and cellohexaose can be considered as minimal model systems for cellulose.[14,33] As discussed in the following section, the structure presented here for the β-(1→4)-linked D-Glc*p* hexasaccharide closely resembles 3D models proposed for the cellulose II polymorph.[34,35]

As it was found for the β-(1→4)-linked D-Glc*p* tetrasaccharide, the lowest energy structure obtained for cellohexaose when ^{13}C chemical shift constraints are omitted from the simulations has $\phi = 176.8 \pm 5.4°$ and $\psi = -1.9 \pm 1.0°$ (Fig. 8b). Use of these dihedral angles in the back-calculation of the C-1 and C-4′ chemical shifts also leads to estimations that are, respectively,

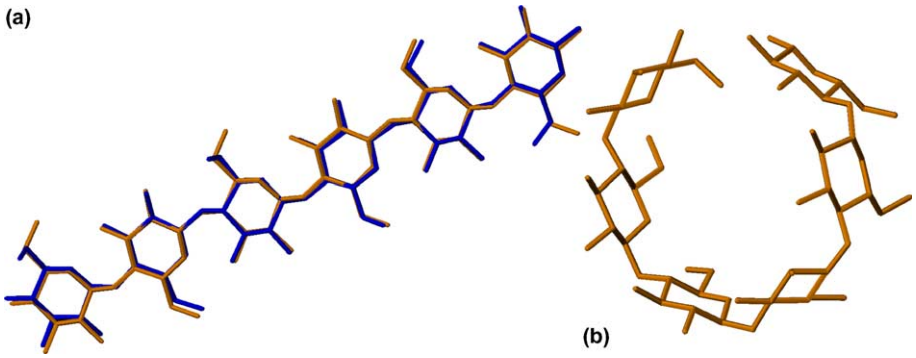

Figure 8. (a) Superposition of cellohexaose and cellulose II 3D models derived using ^{13}C chemical shifts constraints (blue) and a cellulose II polymorph structural model built from data reported in Ref. 35 (orange). (b) Lowest energy conformer for cellohexaose obtained through SA when chemical shift constraints are omitted.

Table 1. Comparison of calculated NMR and structural parameters to their experimental counterparts for the different carbohydrates considered in the study

Molecule	Calculated[a,b]				Experimental[d]			
	ϕ	ψ	$^{13}C\delta_{C\text{-}1}$	$^{13}C\delta_{C\text{-}n'}$	ϕ	ψ	$^{13}C\delta_{C\text{-}1}$	$^{13}C\delta_{C\text{-}n'}$
Trehalose	−49.0	−49.0	93.4	93.4	−51.9 ± 10.3	−51.9 ± 10.3	93.0	93.0
Cellobiose	40.6	−16.4	104.8	85.3	41.2 ± 3.9	−12.9 ± 4.6	104.6	85.0
Cellotetraose[c]	25.8 ± 6.8	−28.9 ± 1.9	107.2	88.7	27.3 ± 2.4	−25.5 ± 4.8	107.4 ± 1.3	89.7
Cellohexaose[c]	31.3 ± 5.2	−35.0 ± 7.1	106.2	87.9	—	—	107.1 ± 1.8	89.3
Cellulose II[c]	31.3 ± 5.2	−35.0 ± 7.1	106.2	87.9	21.9 ± 2.1	−29.0 ± 5.7	107.1 ± 1.2	88.9
Amylose V_6[e]	−7.4 ± 2.7	−22.2 ± 3.2	102.3	82.0	−9.6 ± 7.8	−9.0 ± 2.8	103.8 ± 0.5	82.4 ± 1.6

[a] ^{13}C chemical shifts are in parts per million and dihedral angles in degrees.
[b] C-n' refers to the C-1′ carbon in the case of trehalose, and the C-4′ carbon for molecules with β-(1→4) and α-(1→4) linkages.
[c] For cellulose and the D-Glc*p*-β-(1→4)-linked oligomers, experimental ϕ and ψ angles are the averages for the two structures reported in the unit cell, and the mean of the two resonances for the anomeric carbons observed by CP/MAS ^{13}C NMR is shown.
[d] Sources of experimental CP/MAS ^{13}C NMR and crystallographic data can be found in the text.
[e] For amylose V_6, the ϕ and ψ angles shown are the averages calculated from structures reported in Refs. 36 and 40, and the glycosidic carbon chemical shifts correspond to the range reported for the V_6 polymorph in Ref. 12.

8.4 and 12.6 ppm upfield from literature values (vide supra). Once again, these results clearly indicate that the incorporation of experimental data is essential for the derivation of meaningful structural models through approaches based on molecular mechanics force fields.

3.3. Generation of cellulose II and amylose V_6 structural models

To further assess the predictive power of the method, it was applied to the derivation of structural models of β- and α-(1→4)-linked-D-Glc*p* polymers, namely cellulose and amylose (Fig. 1b,c). While single-crystal X-ray structures for the various polymorphs of these polysaccharides have not been reported, fiber diffraction studies have allowed high-resolution 3D models to be developed,[34–39] many of which are supported by the crystal structures of smaller systems. For example, the X-ray structure of cellotetraose is considered to represent the cellulose II polymorph,[33] and the helical regions observed in the crystal for cyclomaltohexaisocaose, a cyclic oligosaccharide composed of 26 α-(1→4)-linked-D-Glc*p* units, have been postulated as a model for amylose V_6.[40] In addition, extensive ^{13}C NMR data in the solid state are available for cellulose and amylose polymorphs, making them ideal for our purposes.[12,14–19]

As mentioned earlier, cellohexaose can be considered as a minimal model system for cellulose. In fact, if CP/MAS ^{13}C NMR data for the cellulose II polymorph are used in combination with the conformational ensemble employed earlier to derive the 3D model of cellohexaose,[14] a structure identical to the one described above for the hexasaccharide is obtained. This is not surprising, as the chemical shifts for the C-1 and C-4′ carbons in cellulose II are virtually the same as those observed for the corresponding glycosidic carbons in cellohexaose (Table 1). The ϕ and ψ angles found for the cellulose II model derived from ^{13}C chemical shifts, which were presented in the previous section, compare favorably with recent estimates reported by Langan et al. ($\phi = 21.9°$ and $\psi = -29.0°$).[35] As evidenced in Figure 8a, our model is remarkably similar to a theoretical structure of the polymorph built from the reported ϕ and ψ angles, the two having a heavy-atom rmsd of less than 0.30 Å.

Working by analogy to the cellulose case, the α-(1→4)-D-Glc*p* hexasaccharide maltohexaose was employed as a model system for amylose. A conformational ensemble was obtained as described for the other oligosaccharides, and a structural model for amylose V_6 was derived using CP/MAS ^{13}C NMR data reported for this polymorph and CSSs for the D-Glc*p*-α-(1→4)-D-Glc*p* linkage (Fig. 2e,f).[12] The ϕ and ψ angles for the resulting structure are $-7.4 \pm 2.7°$ and $-22.2 \pm 3.2°$, respectively, and use of these dihedral angles in the back-calculation of chemical shifts yields estimations of 102.3 ppm for the C-1 carbons and 82.0 ppm for the C-4′ carbons. While the ^{13}C chemical shift predictions compare well with experiment, the ϕ and ψ angle estimates show somewhat larger deviations from reported values than those found for other systems considered in our study (Table 1). However, these are consistent with the variations observed among the geometries of different amylose V_6 polymorph structural models reported in the literature.[36,37,40] Indeed, the 3D structure derived from ^{13}C chemical shifts appears to be flanked by two of these models, one derived from X-ray diffraction studies and the other based on the helical regions observed in the crystal structure of cyclomaltohexaisocaose (Fig. 9).[36,40]

4. Conclusions

The results obtained for all the systems considered in this study are summarized in Table 1. In the vast majority of

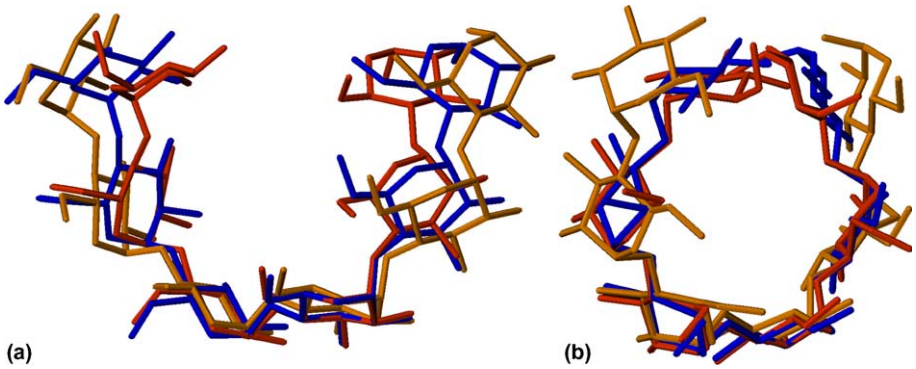

Figure 9. Side (a) and top (b) views of the 3D structural model for amylose V_6 derived from chemical shift constraints (blue). The heavy-atom rmsd between this structure and models built based on crystallographic data from Refs. 36 (orange) and 40 (red) is 0.67 and 0.78 Å, respectively.

cases, the glycosidic linkage ϕ and ψ angles in models derived from ^{13}C chemical shift data are well within 10° from those measured on crystal structures and other existing 3D models. When larger deviations were observed, such as for the amylose V_6 model described above, these were similar to those seen among structural models found in the literature. In addition, structures derived without the consideration of chemical shift constraints departed considerably from reported models, stressing the importance of including experimental data related to the molecular geometry in the study of carbohydrate conformation. Overall, our findings indicate that the methodology outlined here represents a robust and accurate tool for the study of oligosaccharide structure in the solid state, and these findings serve to further demonstrate the utility of ^{13}C chemical shifts as important sources of structural information for these molecules.

Finally, some limitations of our approach deserve to be pointed out. First, the ω angles, and thus the relative geometries of the hydroxymethyl groups, are currently not considered. While these dihedral angles do not determine the fold of the carbohydrate chain and are known to exist in a limited number of conformations (i.e., *gt*, *gg*, and *tg*), they could be easily incorporated into the simulations. Since the chemical shift of the C-6 carbons has been shown to vary systematically with the ω angle,[41] empirical functions relating the two can be readily derived.[42] These would allow experimental chemical shifts for the hydroxymethyl carbons to be taken into account during the generation of the 3D structural models. Similarly, the CSSs described here are independent from the glycosidic bond lengths. A recent report by Sternberg et al. indicates that these can have a significant effect on the chemical shifts of the C-4′ carbons in cellulose,[32] and comparable effects on the C-*n*′ carbons of other oligosaccharides are not unlikely. Inclusion of these geometric parameters into the ^{13}C shielding models could result in small but non-negligible improvements in the accuracy of the chemical shift calculations, as well as in the quality of the 3D structures determined based on these estimations. Perhaps the most important drawback of our method involves the approach employed for the incorporation of chemical shift data into the calculations. As mentioned earlier, simultaneous minimization of the chemical shift penalty functions and force field potentials would rapidly lead to structures consistent with experimental NMR data,[31,32] avoiding the need to carry out extensive sampling of the conformational space available to the oligosaccharides. Due to this limitation, and owing to the similarities in the chemical shifts of their glycosidic carbons, the present implementation of the method yields identical structural models for cellohexaose and cellulose II. Furthermore, attempts to determine the two structures found in the unit cells of the β-(1→4)-D-Glc*p* oligomers and polymers were not successful, and only average structures were computed. Although the resolution of the 3D models presented here is very satisfactory when compared to that of other NMR structure determination strategies, improved versions to be announced in due course will allow for the simultaneous optimization of force field potentials and chemical shift target functions.

Acknowledgements

The authors wish to thank the H. O. West Foundation for providing the funds necessary for the construction of the Beowulf supercomputer clusters used to carry out the simulations presented in this report. Chet W. Swalina provided invaluable comments and suggestions regarding the manuscript for which we are thankful. Financial support from the Pfizer Summer Undergraduate Research Fellowship Program (I.S.) and the Camille and Henry Dreyfus Foundation (G.M.) is also acknowledged.

References

1. Dwek, R. A. *Chem. Rev.* **1996**, *96*, 683–720.
2. Kuberan, B.; Linhardt, R. J. *Curr. Org. Chem.* **2000**, *4*, 653–677.

3. Nitz, M.; Bundle, R. D. The Unique Solution Structure and Immunochemistry of the *Candida albicans* β1,2-Mannopyranan Cell Wall Antigen. In *NMR Spectroscopy of Glycoconjugates*; Jiménez-Barbero, J., Peters, T., Eds.; Wiley-VCH: Weinheim, 2003; pp 111–144.
4. French, A. D.; Bertoniere, N. R.; Brown, M. R.; Chanzy, H.; Gray, D.; Hattori, K.; Glasser, W. Cellulose In *Kirk–Othmer Encyclopedia of Chemical Technology, 4th ed.*; John Wiley & Sons: New York, 1993; Vol. 5, pp 476–496.
5. Bush, C. A.; Matrin-Pastor, M.; Imberty, A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 269–293.
6. Imberty, A.; Pérez, S. *Chem. Rev.* **2000**, *100*, 4567–4588.
7. Wormald, M. R.; Petrescu, A. J.; Pao, Y.-L.; Glithero, A.; Elliott, T.; Dwek, R. A. *Chem. Rev.* **2002**, *102*, 371–386.
8. Weimar, T.; Woods, R. J. Combining NMR and Simulation Methods in Oligosaccharide Conformational Analysis. In *NMR Spectroscopy of Glycoconjugates*; Jiménez-Barbero, J., Peters, T., Eds.; Wiley-VCH: Weinheim, 2003; pp 111–144.
9. Allen, F. H. *Acta Crystallogr.* **2002**, *B58*, 380–388.
10. Bock, K.; Brignole, A.; Sigurskjold, B. W. *J. Chem. Soc., Perkin Trans. 2* **1986**, 1711–1713.
11. Saitô, H. *Magn. Reson. Chem.* **1986**, *24*, 835–852.
12. Gidley, M. J.; Bociek, S. M. *J. Am. Chem. Soc.* **1988**, *110*, 3820–3829.
13. Jarvis, M. C. *Carbohydr. Res.* **1994**, *259*, 311–318.
14. Dudley, R. L.; Fyfe, C. A.; Stephenson, P. J.; Deslandes, Y.; Hamer, G. K.; Marchessault, R. H. *J. Am. Chem. Soc.* **1983**, *105*, 2469–2472.
15. Gidley, M. J.; Bociek, S. M. *J. Am. Chem. Soc.* **1985**, *107*, 7040–7044.
16. Atalla, R. H.; Gast, J. C.; Sindorf, D. W.; Bartuska, V. J.; Maciel, G. E. *J. Am. Chem. Soc.* **1980**, *102*, 3249–3251.
17. Earl, W. L.; VanderHart, D. L. *J. Am. Chem. Soc.* **1980**, *102*, 3251–3252.
18. Kono, H.; Yunoki, S.; Shikano, T.; Fujiwara, M.; Erata, T.; Takai, M. *J. Am. Chem. Soc.* **2002**, *124*, 7506–7511.
19. Kono, H.; Erata, T.; Takai, M. *Macromolecules* **2003**, *36*, 3589–3592.
20. Swalina, C. W.; Zauhar, R. J.; DeGrazia, M. J.; Moyna, G. *J. Biomol. NMR* **2001**, *21*, 49–61.
21. O'Brien, E. P.; Moyna, G. *Carbohydr. Res.* **2004**, *339*, 87–96.
22. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T. A.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Gonzalez, C.; Challacombe, M.; Gill, P. M. W.; Johnson, B. G.; Chen, W.; Wong, M. W.; Andres, J. L.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian 98 (Revision A.7), Gaussian, Inc., Pittsburgh PA, 1998.
23. Halgren, T. A. *J. Comput. Chem.* **1996**, *17*, 490–519.
24. Zhang, P.; Klymachyov, A. N.; Brown, S.; Ellington, J. G.; Grandinetti, P. J. *Solid State Nucl. Magn. Reson.* **1998**, *12*, 221–225.
25. Brown, G. M.; Rohrer, D. C.; Berking, B.; Beevers, C. A.; Gould, R. O.; Simpson, R. *Acta Crystallogr., Sect. B* **1972**, *28*, 3145–3158.
26. Taga, T.; Senma, M.; Osaki, K. *Acta Crystallogr., Sect. B* **1972**, *28*, 3258–3263.
27. Jeffrey, G. A.; Nanni, R. *Carbohydr. Res.* **1985**, *137*, 21–30.
28. Jacobson, R. A.; Wunderlich, J. A.; Lipscomb, W. N. *Acta Crystallogr.* **1961**, *14*, 598–607.
29. Brown, C. J. *J. Chem. Soc. A* **1966**, 927–932.
30. Chu, S. S. C.; Jeffrey, G. A. *Acta Crystallogr., Sect. B* **1968**, *24*, 830–838.
31. Moyna, G.; Zauhar, R. J.; Williams, H. J.; Nachman, R. J.; Scott, A. I. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 702–709.
32. Sternberg, U.; Koch, F.-T.; Prieß, W.; Witter, R. *Cellulose* **2003**, *10*, 189–199.
33. Gessler, K.; Krauss, N.; Steiner, T.; Betzel, C.; Sarko, A.; Saenger, W. *J. Am. Chem. Soc.* **1995**, *117*, 11397–11406.
34. Raymond, S.; Kvick, A.; Chanzy, H. *Macromolecules* **1995**, *28*, 8422–8425.
35. Langan, P.; Nishiyama, Y.; Chanzy, H. *Biomacromolecules* **2001**, *2*, 410–416.
36. Winter, W. T.; Sarko, A. *Biopolymers* **1974**, *13*, 1461–1482.
37. Brisson, J.; Chanzy, H.; Winter, W. T. *Int. J. Biol. Macromol.* **1991**, *13*, 31–39.
38. Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.
39. Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300–14306.
40. Gessler, K.; Usón, I.; Takaha, T.; Krauss, N.; Smith, S. M.; Okada, S.; Sheldrick, G. M.; Saenger, W. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 4246–4251.
41. Horii, F.; Hirai, A.; Kitamaru, R. *Polym. Bull. (Berlin)* **1983**, *10*, 357–361.
42. Mazeau, K.; Taravel, F. R.; Tvaroska, I. *Chem. Pap.* **1996**, *50*, 77–83.