950 *Macromolecules* **2011**, *44*, 950–957
DOI: 10.1021/ma102240r

Macromolecules
ARTICLE

X-ray Structure of Anhydrous β-Chitin at 1 Å Resolution

Yoshiharu Nishiyama,*,† Yasutomo Noishiki,‡,§ and Masahisa Wada‡,∥

†*Centre de Recherches sur les Macromolécules Végétales (CERMAV-CNRS), BP 53, F-38041Grenoble Cedex 9, France,* ‡*Department of Biomaterials Science, The University of Tokyo, Tokyo 113-8657, Japan, and* ∥*College of Life Sciences, Kyung Hee University, Gyeonggi-do 446-701, Korea.* §*Current address: Oji Paper Co., Tokyo 135-8558, Japan.*

Received September 30, 2010; Revised Manuscript Received January 7, 2011

ABSTRACT: The crystal structure of β-chitin, a homopolymer of *N*-acetylglucosamine, was determined using highly resolved X-ray fiber diffraction data obtained from uniaxially oriented specimen of diatom spines. The labile *N*-acetyl groups and hydroxymethyl groups were directly located in the Fourier omit maps constructed using the glycosyl backbone as phasing model. Using the 216 independent intensities extending to a resolution of 1 Å, we could refine the structure without any restraints on bond angles to obtain a reasonable model with high accuracy. The carbonyl oxygen of acetamide group formed intermolecular hydrogen bonding between the primary hydroxyl and amine groups forming a two-dimensional hydrogen bonding network in a plane perpendicular to the pyranosyl plane. The hydrogen bonding pattern explains why the conformation of the acetamide groups is slightly rotated from standard conformation.

Introduction

Polysaccharides having simple molecular structure such as cellulose, chitin, and amylose represent the major part of biomass and have great potential as renewable resources. They have different reactivity such as enzymatic digestibility or solvent resistance, which depends on their solid state structure. For example, cellulose III has been shown to be hydrolyzed much faster than cellulose I by a cellobiohydrolase;[1] β-chitin is considered to be more reactive than α-chitin;[2] a wide range of polar molecules can intercalate β-chitin crystals[3−10] but not those of α-chitin. Thus, the structure and intermolecular interaction in solid is a key to understand and modify their properties in view of their optimal utilization. Although the explicit interaction at molecular level is difficult to characterize experimentally for poorly crystalline samples, molecular details can be, in principle, experimentally determined for crystals. In the case of cellulose and chitin, certain biological organisms are known to produce highly crystalline samples that diffract beyond 1 Å resolution. Yet, it is still a challenge to derive structural information from diffraction data from polycrystalline samples, since the nanometric size of each crystal do not allow a single crystal study in a conventional way.

Many crystal structures of polymers have been refined since the 1960's based on limited number of fiber diffraction intensity data combined with stereochemical constraints or restraints which gave structures that reproduce measured diffraction intensities with reasonable molecular geometry.[11] Yet the uniqueness of these structural determination is not guaranteed, since most often only a limited number of models are screened on a trial and error basis. A typical example demonstrating this pitfall, is the occurrence of two distinct models of cellulose I proposed in the 1970's based on similar refinement technique, where the molecules were running in opposite directions between the two models.[12−14]

Recently we reinvestigated a series of cellulose allomorphs using almost uniaxially oriented highly crystalline cellulose samples with fiber diffraction techniques. The two-dimensional intensity data with high spatial resolution and high dynamic range allowed us to improve the accuracy and number of measured diffraction intensities. With this improved data set, we could directly locate labile atoms or solvent molecules in the Fourier omit maps and thus unambiguously determine the basic structure.[15−21]

We are on the way to further improve the method to obtain maximum structural information from fiber diffraction data, limiting the bias introduced by model construction. This is done by careful background evaluation, texture analysis, and improved peak fitting procedures to obtain accurate diffraction intensities, combined with phasing using rigid fragment of the molecule.

In this study, we reinvestigated the structural details of β-chitin, one of the two known crystalline form of pure chitin found so far. The β-chitin allomorph has never been recrystallized *in vitro*, but can only be found as a direct result of biosynthesis,[22,23] in the same way as cellulose I is only formed by biological machinery. In contrast to cellulose biosynthesis in which only parallel chain cellulose I is produced in nature, most chitinous biomass resources, namely arthropod cuticles such as crab shells are composed of α-chitin[24] which is also the crystalline form when produced *in vitro*,[25,26] regenerated from solution[27] or from swollen state.[28,29]

Although much smaller in quantity as readily available biomass, pure chitin with virtually no contaminant can be obtained as extracellular β-chitin fibrils of certain species of diatoms,[30] namely *Thalassiosira weissflogii*.This kind of highly crystalline pure chitin can have important applications when purity is of importance such as for medical applications. The unique property of easily accommodating polar molecules reversibly in the crystal lattice to form various crystallosolvates can also be an advantage of this allomorph.

The structure of β-chitin has been studied by combination of X-ray fiber diffraction and polarized infrared spectroscopic data, leading to the currently accepted structural model by Gardner and Blackwell (GB model).[31] The unit cell of chitin anhydrous form is monoclinic with, $a = 4.85$ Å, $b = 9.26$ Å, c (fiber repeat) = 10.38 Å, and $\gamma = 97.5°$, with the space group $P2_1$. This unit cell contains only one chain and thus the chains of β-chitin are packed

*To whom correspondence should be addressed E-mail: yoshi@cermav.cnrs.fr.

 Published on Web 01/26/2011

in parallel arrangement, composed of only one anhydroglucose as asymmetric unit. Gardner and Blackwell refined three structural models with fixed bond length and bond angles against 61 reflection data and compared the final agreement using Hamiltonian test to discriminate among the plausible models. Their final model was obtained by imposing certain intermolecular hydrogen bonds, deduced from polarized infrared spectra, by applying a restraint on oxygen−oxygen distance. When this restraint was lifted, the model converged to a structure where the hydroxymethyl oxygen was not forming hydrogen bonding. It showed slightly better agreement with X-ray data, but the difference was not significant enough to decide between the two models based on X-ray data only. The only structural model of β-chitin in the Cambridge Structural Database (CSD), coded LILVUX published by Dweltz et al.,[32] was rejected on the basis of crystallographic agreement factor over the GB model.

The GB model (the published coordinate apparently contains typographical error, and does not reproduce their figure with C2−N bond distance and C2−N−C7 bond angle beyond the acceptable range) does not have contradiction with any experimental data available today, including infrared-spectroscopy, solid state NMR spectroscopy,[33] and can explain the unique property to incorporate polar molecules. However, as the structure determination is based on model building and selection among limited number of models, there is no guarantee that the GB structural model is the only acceptable one. The structural model was built to explain different experimental data and thus there is no redundancy in data. Yui et al.[34] carried out an exhaustive search to refine the structure from many possible starting conformation, and give credit to the gross structure of GB model but do not provide reliable coordinates. We thus reinvestigated the crystal structure of anhydrous β-chitin using high resolution data to determine the structure with minimum a priori suppositions.

Experimental Section

β-Chitin Sample. A centric diatom, *T. weissflogii* strain CCMP1051, was cultured in f/2 medium under artificial illumination and CO_2 enriched aeration for 1 month. The culture suspension was lightly homogenized by a waring blender to dislodge the chitin spines from diatom cells, followed by centrifugation at 2000*g* to sediment the cells, while leaving chitin spines in the supernatant, which was further centrifuged at 19000*g* to collect cell-free chitin spines as pellet. The recovered chitin was successively treated with 5% KOH (25 °C, overnight), methanol (65 °C, 2 h), 0.3% NaClO2 (pH 4.9, 70 °C, 6 h), 0.1 N HCl (boiling, 1 h), and finally 1% HF (25 °C, overnight), with centrifugal rinsing with water after each step. Purified chitin was freeze-dried and kept in a desiccator until use.

Preparation of Oriented Fibers. Approximately 5 mg of dry chitin was dispersed in 7 mL of water and mixed with 3 mL of 1% fibrinogen solution in 3% sodium chloride. This mixture was supplemented by several drops of concentrated thrombin aqueous solution, immediately spread in a glass Petri dish to form approximately a 3 mm thick layer, and allowed to stand at room temperature. When the layer formed a soft gel, it was cut into a 50 mm ×10 mm strip and slowly stretched by hands to about 3 times with continuous removal of water by contacting with a filter paper. This string like specimen was softened by immersing in 1% KOH and further stretched to 1.5−2 times. The string was cut into 20 mm long pieces, and a proper number of pieces were bundled and inserted into a glass capillary of 1 mm inner diameter. Fibrinogen in the specimen was removed by alternately injecting 5% KOH solution and water into the capillary several times. The specimen was finally dried in a vacuum at 120 °C.

Data Collection. Synchrotron X-ray fiber diffraction data was collected at the BL40B2 beamline at SPring-8 (Hyogo, Japan).

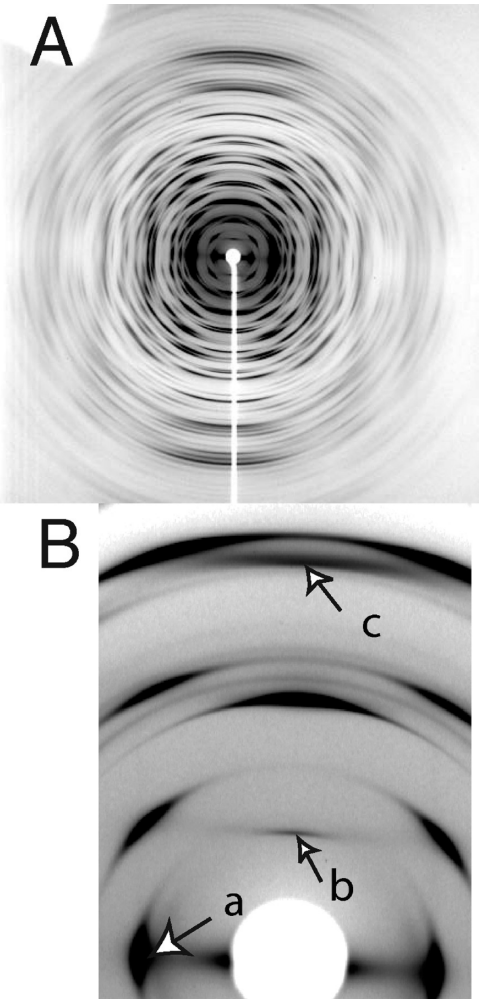

Figure 1. (A) Fiber diffraction diagram of anhydrous β-chitin and (B) its close up of central zone. The fiber axis is approximately vertical. The arrows indicate spots that cannot be explained by anhydrous β-chitin unit cell with $P2_1$ symmetry.Arrow a corresponds to equatorial reflection of hydrated form, and arrows b and c correspond to 0 0 1 and 0 0 3, respectively.

The oriented fibers thus prepared were mounted on a goniometer head and synchrotron radiated X-rays (λ = 0.7293 Å) were irradiated orthogonally to the fiber axis. The diffraction patterns were recorded on flat imaging plates (R-AXIS IV++, Rigaku) with a sample to detector distance of 150 mm for 60 s. The data was read out as a matrix of 3000 by 3000 pixels with 100 μm resolution. For callibration purpose, Si powder pattern was also recorded at the same condition in a capillary of 1 mm diameter for 10 s.

Figure 1 shows the raw fiber diffraction data after polarization correction and flat-plate Lorentz correction with intensity in logarithmic scale to account for large dynamic range. In the equator, there was a small peak shoulder at lower angle of 0 1 0 and noticeable intensity in the meridian 0 0 1, 0 0 3, which is prohibited for $P2_1$ symmetry. However, the absolute intensity of these peaks in linear scale were very small (~1% of 0 0 2 intensity) and these spots showed a streaking perpendicular to the fiber axis, absent on other reflections. This streaking is an indication that these peaks arise from thin domains parallel to the chain direction. The domains can be defects due to imperfect drying or rehydration of the sample or the lateral surfaces of the crystals. Note that the side groups exposed on the surface alters every two residues. If these groups take conformations different from those in the bulk crystal or correlated vapor adsorption takes place on the surface, significant electron density

fluctuation can arise corresponding to this odd order. For a chitin crystal with a thickness of 30 nm, about 10% of the chains are on the surface, which is probably sufficient to explain the faint intensities observed on the odd order meridian. On the basis of these considerations, the prohibited streak was neglected in the further analysis.

Unit Cell Determination. Peak positions of the 25 strong diffraction of the fiber pattern were measured using R-AXIS display software (Rigaku). After indexing the *d*-spacings according to one chain monoclinic unit cell, the unit cell parameters were determined by a least-squares refinement method. Unit cell parameters thus determined were almost the same as those reported by Blackwell.[3] They were $a = 4.820(3)$ Å, $b = 9.247(4)$ Å, $c = 10.390(7)$ Å, and $\gamma = 97.2\ 1(7)°$, with unit cell volume of 459.3 Å^3.

Data Processing. The beam center coordinate, detector tilt, and detector distance was determined from Si powder pattern using Fit2D software (http://www.esrf.eu/computing/scientific/FIT2D).

Polarization Correction. The synchrotron X-ray source has linear polarization whose polarization factor, *P*, depends on the azimuthal angle ρ, taken with respect to the polarization axis and the scattering angle θ with respect to the incident beam as follows[35]

$$P = \frac{1+\cos^2 2\theta}{2} - A\frac{\cos 2\rho \sin^2 2\theta}{2} \quad (1)$$

The linear polarization coefficient, *A*, was 0.83 under our experimental condition. When the pixel intensity was corrected for this polarization correction, the oval-shaped background scattering became circularly symmetric.

Background Subtraction. The diffraction pattern was converted with the above polarization correction to polar coordinate intensity matrix of 360 by 2000 elements, corresponding to azimuthal angle every 1° and radial distance with pixel unit (100 μm) interval, respectively. The intensity value and its standard deviation at each polar coordinate was evaluated from all detected intensities corresponding to the same bins. This polar coordinate data format has the advantage that the two-dimensional peak intenisty profiles can be expressed as a product of two orthogonal function as.

$$I(r,\theta) = I(r)I(\theta) \quad (2)$$

r and θ being the radial distance and azimuthal angles, respectively. Only one-dimensional profiles need to be stored as opposed to the standard fiber diffraction representation where a two-dimensional profile has to be calculated for each peak.

The background was evaluated for each radial trace using bicubic spline function with grid points every 50 pixels. A Bayesian approach was used to evaluate the background in order to avoid systematic overestimation of weak peak intensities as has been the case in our previous studies.[16,17] A bicubic spline was used to express the background intensity distribution. The first guess was given by Sonnevelt algorithm, and linear least-squares fit cycles with weighting scheme to scale down the intensity from diffraction peaks were applied to maximize the likelihood function.

Peak Fitting. The orientation distribution of the fiber axis was approximated by a single Gaussian function. The width of the Gaussian function was estimated by simulating the azimuthal intensity distribution of equatorial reflection. The azimuthal intensity distribution of any diffraction spots can be calculated by numerical integration of the population contributing to the intensity at each position in reciprocal space.[36] The radial width of each peak was given depending on the peak position according to the following formula

$$w = w_o + w_c(d^*,\sigma)/\cos^2 2\theta \cos\theta \quad (3)$$

where w_0 corresponds to the broadening due to the beam size, w_c corresponds to the broadening due to the crystal size and disorder.

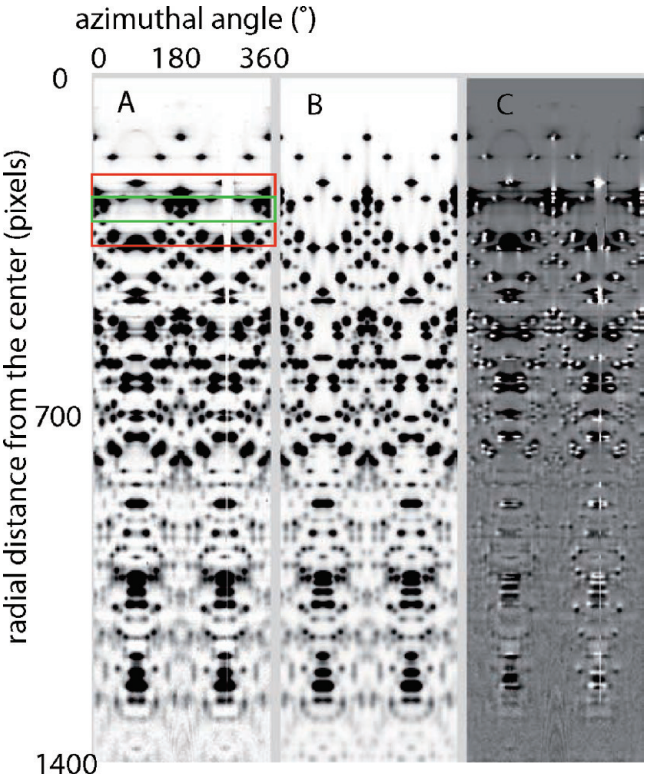

Figure 2. Visual representation of diffraction diagrams transformed into polar coordinate for intensity fitting purpose. Key: (A) observed; (B) fitted; (C) A − B. In this polar coordinate system, the origin is at the center of incident beam and the azimuthal angles are taken with respect to the horizontal left of the fiber pattern in Figure 1 in a clockwise manner. Thus, the equator in Figure 1 corresponds to azimuthal angles of 0 and 180°, whereas the meridian corresponds to 90 and 270°. The size of data region used for each fit cycle is represented by a red rectangle while only intensities of the peaks found in the green rectangle are retained.

Since the crystal size along the fiber direction is expected to be very large, we took the following form to describe the broadening w_c depending on the angle σ between the fiber axis and the lattice position in the reciprocal coordinate.

$$w_c = w_m + w_{eq} \sin\sigma + cd^{*2} \quad (4)$$

where w_m is the width of the peak along the meridian and w_{eq} is the width along the equator, and c is the paracrystallinity parameter. Given the peak profile of each diffraction, the diffraction data should be a linear combination of the peak profile with positive peak intensities. Since the peak overlapping are limited to the neighboring area, these intensities can be determined in a block-wise manner by least-squares fitting. The block size was chosen to contain at least several entire peaks but small enough for reasonable calculation time and memory allocation. In this case a block of 360 × 150 pixels was chosen. The diffraction close to the border of a block can be influenced by the neighboring peaks that are out of the block. To avoid this artifact, the calculation block was moved so that it overlaps with the preceding block. Only the peak intensities in the inner part of each block were taken.

A LAPACK driver function DGELSD that computes minimum norm solution for under-determined linear least-squares problem was used. The fitting was iterated on the residuals, with negative intensities set to zero.

Figure 2 shows the observed (A) and calculated (B) diffraction data in polar coordinate as well as their difference (C).

The standard error estimation of intensities obtained in this way is problematic since the rank deficient normal matrix cannot be inverted. Thus, the error estimation was done by calculating the

variation of sum of squared residuals χ^2 when each intensity was modified by unity from the final solution, and arbitrarily given by the following.

$$\sigma I = \text{const.} + 0.01/\delta(\chi^2) \quad (5)$$

In this fitting procedure, intensities of very close reflections in the cylindrically averaged reciprocal space are given evenly divided values which does not reflect the physical reality. Furthermore, a slight modification in peak position due to detector distortion or an error in unit cell parameters would easily result in false intensity evaluation. On the other hand, intensity data for each Miller index is needed to construct Fourier synthesis, so this list was kept as a "crude intensity list". This is almost equivalent to equi-partitioning overlapping peak intensities, employed in the direct method analysis.[37] For refinement purpose, the reflections that belong to the same layerline and that are too close to each other were combined into one intensity data to make a "regrouped list". Since the broadening is mainly in the direction perpendicular to the fiber axis, the distance from the fiber axis in the reciprocal space, r, was compared and if the distance between the two spots, r_i-r_j, was smaller than 0.01(Å^{-1}), the intensities were summed into a composite peak intensity. The highest estimated standard error among the contributing reflections were chosen as the standard error of the composite data.

Refinement. *Construction of Rigid Backbone*. Chitin is a homopolymer of *N*-acetyl-D-glucosamine (GlcNAc) residues, with a molecular formula of ($C_8H_{13}NO_5$). The unit cell volume of 459 Å^3 can only contain two residues. With the $P2_1$ symmetry selection, the asymmetric unit is one GlcNAc residue. Since the pyranose ring is known to take a rigid $^4C_1$ conformation, the pyranose ring excluding the labile hydroxymethyl and acetamide groups was used as phasing model. The backbone has some degree of freedom around the glycosidic linkage, but the $P2_1$ symmetry and fiber periodicity almost uniquely determine the glycosidic torsion angles.

To construct the starting model, the pyranose ring was energy minimized using Tripos 5.2 force field.[38] The pyranose ring was then rotated around a virtual bond linking two subsequent glycosidic oxygen atoms keeping the $P2_1$ symmetry. The rotation giving the smallest energy was chosen as the starting structure. Then the motif, involving six carbon atoms, three oxygen atoms, and one nitrogen atom, was placed in the unit cell with the $P2_1$ symmetry axis coinciding with the z axis. All hydrogen atoms and labile side groups including two oxygen and two carbon atoms were not included in the model. The chain was placed either with the z coordinate of O5 higher than that of C5 of the same residue or in the opposite direction, corresponding to "up" or "down" structure, respectively. These motifs were rotated around the chain axis with steps of 1° and the agreement with experimental intensities was calculated.

To facilitate the evaluation of agreement between the experimental value and intensity calculated from model structure, the following indicators for exhaustive search were defined[39]as

$$E_I = 1 - \left(\frac{\mathbf{I}_{\text{obs}}\mathbf{I}_{\text{cal}}}{|\mathbf{I}_{\text{obs}}||\mathbf{I}_{\text{cal}}|}\right)^2; \quad E_F = 1 - \left(\frac{\mathbf{F}_{\text{obs}}\mathbf{F}_{\text{cal}}}{|\mathbf{F}_{\text{obs}}||\mathbf{F}_{\text{cal}}|}\right)^2$$

which are equivalent to the square of wR^2 used in SHELX refinement with a weight of unity given to all data and global scale factor optimized. Here the $\mathbf{I}_{obs}$and $\mathbf{I}_{cal}$, or $\mathbf{F}_{obs}$and $\mathbf{F}_{cal}$ are the vectorial representation of observed and calculated intensity or structure factor amplitude list, respectively. The calculation was also done using both "crude" (c) and "regrouped" (r) list resulting in four indicators, E_{Ic}, E_{Ir}, E_{Fc}, E_{Fr} as a function of the chain rotation.

Figure 3 shows the variation of these E values using different resolution shells as a function of rotation angle around the z axis. In both the "up" and "down" cases there is only one region of global minimum independent of resolution limit or grouping of intensities. However, with the rigid model, it was impossible

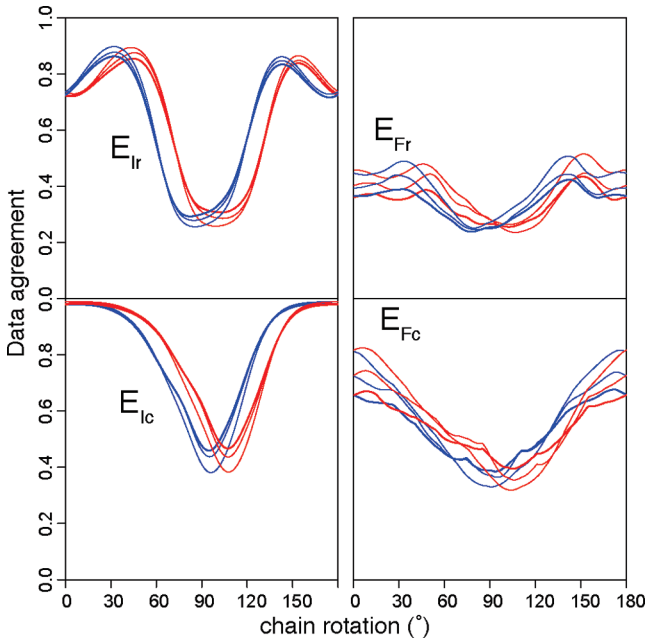

Figure 3. E values as a funtion of rotation angle for up (blue lines) and down (red lines) chain with a resolution extending to 2.5 Å (thick lines), 2 Å, and 1.5 Å (thin lines). Intensity (left) or structure factor amplitude (right) with "regrouped" (top) or "crude" (bottom) data lists were used.

to distinguish between "up" and "down" arrangement based on the intensity agreement only. Thus, two "up" models with rotation of 75 and 95°, and a down model with 110° rotation were considered as starting model for further refinement using SHELX.

First, the bond angles and bond distances were restrained to the equilibrium values of the Tripos5.2 force field. When refined against the "crude" intensity list up to 1.3 Å resolution, both "up" models converged to the same structure with the classic crystallographic R-factor $R_1 = 0.38$ whereas the down structure converged to a structure with $R_1 = 0.55$.

Fourier Omit Map. The Fourier omit maps (Figure, 4) were visualized using *Coot*[40]from the phase and normalized observed intensities and calculated intensities obtained in the above refinement. Difference map using $F_o - F_c$ as amplitude and sigma weighted Fourier map (SIGMAA) with amplitude of $2mF_o - F_c$ with m being the figure of merit were used. Here F_o is the square root of observed intensity and F_c the calculated counterpart. With the up model, the SIGMAA map clearly showed electron densities corresponding to the O6 and the oxygen and carbon atoms of acetyl groups, while the difference map located the position of O6, that of carbonyl oxygen O7 and the carbon C8 of the methyl group (Figure 4). With the down model, only the O6 close to the gg position could be recognized on the map. Thus, the down structure was discarded and only the up model was considered in the further refinement.

The missing atoms were progressively introduced using the crude data up to 1.3 Å at the position of residual electron density peaks in the SHELX output file, followed by refinement with bond angle and bond distance restrained. At each addition of atoms the Fourier difference map was calculated (Figure 4b−d). The addition of O6 resulted in $R_1 = 0.32$ and further addition of C7 in $R_1 = 0.3$. The last C8 and O7 are interchangeable. In the first place the O7 was incorporated to form the most frequent Z-conformation where the carbonyl oxygen is *trans* to the amide hydrogen. Finally with all carbon and oxygen atoms incorporated, $R_1 = 0.17$ with 104 intensities $F_o > 4\sigma(F_o)$ within 1.3 Å was obtained.

Structure Refinement. Further structure refinement proceeded using the "regrouped" list. With bond distance, bond angle restrained and isotropic temperature factor restrained with SIMU

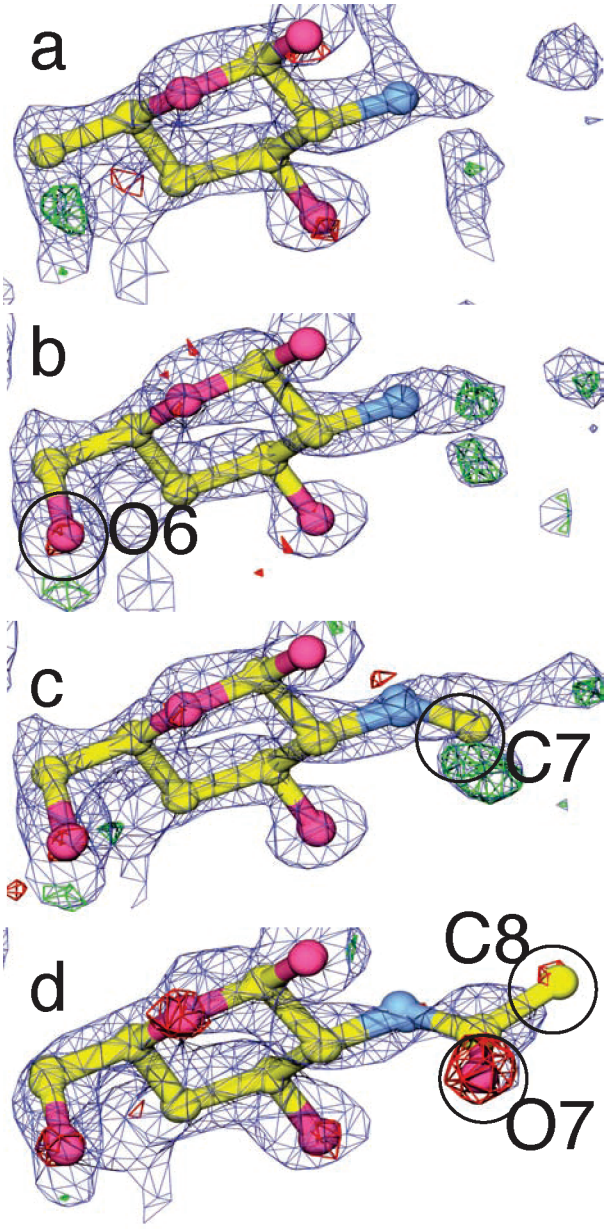

Figure 4. Fourier omit maps of "up" model. The side group atoms were successivly added to the initial phase model (a) in order of O6 (b), C7 (c), and O7 and C8 (d). Blue wires are SIGMAA map, green for positive intensity and red for negative intensity in $F_o - F_c$ maps.

command, the classical agreement factor was $R_1 = 0.11$ for $73F_o > 4\sigma(F_o)$. Then all geometric restraints were lifted and the atom positions were refined against intensity data up to 1 Å resolution. The intensity data included 216 intensities out of which 105 intensities significantly higher than the noise level ($F_o > 4\sigma(F_o)$). At this stage $R_1 = 0.12$ for 105 intensities. Incorporation of hydrogen using HFIX commands on carbon and nitrogen atoms resulted in R1 = 0.109.

Then the C8 and O7 were exchanged and refined to check for the alternative conformation which gave $R_1 = 0.116$: fixing the parameters and exchanging C8 and O7 again resulted in $R_1 = 0.110$. The alternative conformation (*E* conformation) can be rejected at 99.5% confidence level according to the Hamilton's test.[41]

Finally, when the structure was refined with anisotropic temperature factor for non-hydrogen atoms with DELU and SIMU restraints, with hydrogen atoms on oxygen added with HFIX command, $R_1 = 0.064$ was obtained. However, the bond distances

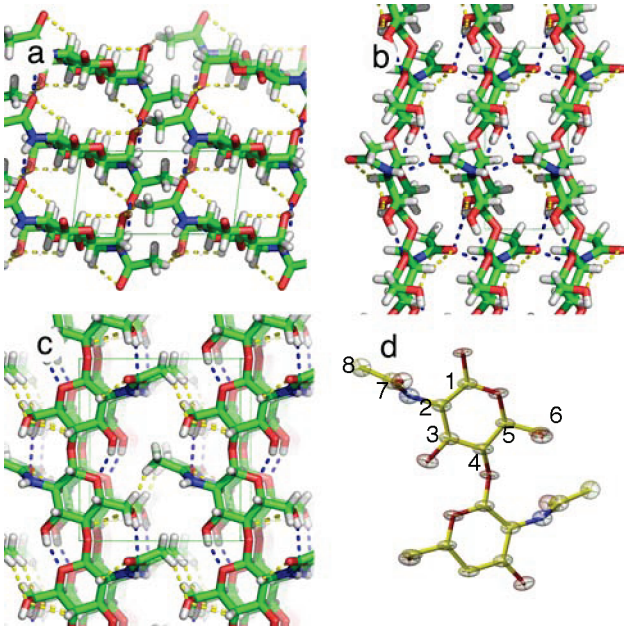

Figure 5. Final structural model. Key (a) *ab* projection; (b) view along *b* axis; (c) *ac* projection; (d) thermal ellipsoid representation of chitobiosyl residue. The hydrogen bonds are indicated by broken lines with blue for OH or NH donor and yellow for CH donor. The carbon numbers are shown in part d.

C2−C3, C3−C4, and C7−C8 were refined to a value nearly 10% smaller than the standard value. With the current data quality, the standard uncertainty of the atomic coordinate was about 0.05 Å perpendicular to the chain direction and 0.03 Å along the chain direction. The refined bond distance can be far off from reasonable values due to this uncertainty. Thus, we reintroduced the restraint on the bond distance using standard values, resulting in $R_1 = 0.074$. When bond distances were restrained, the atom positions were defined with a standard uncertainty of 0.03 Å perpendicular to chain direction and 0.02 Å along the chain direction. This model, considered as the final model is detailed in the CIF file in the Supporting Information. When this CIF file generated by SHELX was checked with the PLATON software,[42] no geometric anomaly was reported. The final structure seen from different directions and the molecular geometry with thermal displacement ellipsoids are shown in Figure 5a−d, respectively. The atomic coordinates and geometrical parameters are given in the CIF file in the Supporting Information.

Discussion

Molecular Directionality. From Figure 3, it is clear that using rigid backbone without side groups, it is not possible to distinguish between up and down structure with data set of any resolution. However, once the model is given the rotational freedom around covalent bonds, the difference in agreement with experimental intensities became significant.

In fact when up and down molecules are superimposed without the side groups, the overall morphology overlaps to a large extent and the differences in atom positions are within a fraction of an ångström. When the models were relaxed, only the model placed in correct direction can fill the gap to fit the diffraction data, and the improvement in agreement with observed intensities were much more important than for the motif placed in the wrong direction. We could thus construct the phasing model with minimum bias from the starting structure.

The fact that two starting model rotated by 20° converged to the same structure further shows the robustness of the

Table 1. Conformational Parameters of *N*-Acetyl Glucosamine-Based Structures

	parameters (deg)[g]									
	τ	ω	ω'	ϕ	ψ	χ	χ'	ξ	ξ'	Θ
this structure[a]	117(2)	−65(3)	58(3)	−89(3)	−152(2)	95(3)	−136(3)	16(4)	−170(2)	10(3)
LIVUX[b] ref[32]	116	105	−133	−92	−150	−114.3	133	1.	180	0.30
GB[c] ref[31]	111.6	−66	55	−103	−142	106	−135	0	180	0.30
Yui[34]	114	−56		−88	−152	112		16		
chitobiose[47] (*R*)[d]	116	−75	46	−80	−107	139	−98	−2	180	0.54
chitobiose[47] (NR)[e]		−66	57			100	−137	3	−174	2.87
ACEGLUA11[f] ref[48]		−60	59			140	−97	−10	170	3.09

[a] Standard errors between parentheses. [b] LIVUX coordinates correspond to *N*-acetyl-L-glucosamine instead of *N*-acetyl-D-glucosamine. [c] After correction of a presumed typographical error. [d] Reducing sugar. [e] Nonreducing sugar. [f] *N*-acetyl-α-D-glucosamine. [g] τ: glycosidic bond angle C4OC1. Torsion angles: ω, O5C5C6O6; ω', C4C5C6O6; φ, O5C1O1C4; ψ, C1O1C4C; χ, C1C2N2C7; χ', C3C2N2C7; ξ, C2N2C7O7; ξ', C2N2C7C8. Ring puckering parameter: Θ.

refinement procedure of the backbone. In this way, we reconfirmed the present model of parallel-up structure.

For β-chitin, attempts to experimentally determine the molecular directionality by combining single crystal electron diffraction with staining of the reducing ends or enzymatic digestion of nonreducing ends have been reported.[43−45] In the last description,[45] using biotinylation of the aldehyde group of the reducing end followed by streptavidin gold labeling and a series of tilt diffraction experiment unambiguously confirmed the parallel-up structure of β-chitin. This was also supported by an exhaustive structure search based on diffraction intensity reported by Gardner and Blackwell and packing energy consideration.[34] Our result is in agreement with the basic structure of the current model supported by both direct identification of reducing end and exhaustive search, but is based on very little a priori knowledge, demonstrating the advantage of using the omit refinement procedure.

Molecular Conformation. Selected conformational parameters are summerized in Table 1. The torsion angles ϕ and ψ around the glycosidic linkage were close to the values refined by Yui and differed significantly from those of the Gardner and Blackwell model. The azimuthal sugar puckering parameter[46] Θ indicating the deviation from the $^4C_1$ chair conformation is larger than the one in the single crystal studies of *N*-acetylglucosamine and chitobiose. The Θ value is similar to that of cellulose III_I and of the center chain of cellulose II. It is interesting to note that the structure of β-chitin and that of cellulose III_I are quite similar having one chain monoclinic unit cell and identical chain arrangement. In this case, the puckering parameter is probably conditioned by the neighboring pyranose rings.

We compared the conformational parameters of structures containing C8H3−C7O7−NH−C2HCC motifs in the Cambridge Structural Database (CSD)[49,50] using ConQuest software[51] and *N*-acetylglucosamine residue in the protein data bank (PDB) records. NDG or GNC residue in the PDB, solved with a resolution above 1.5 Å, corresponded to 51 files and 127 residues. Figure 6 shows the histogram of torsion angles of analogues found in the CSD and PDB. The torsion angles O7−C7−N−C2 (ξ) and C8−C7−N−C2 (ξ') deviating from 0° or 180° indicates the nonplanarity of *N*-acetyl groups. The combination (ξ, ξ') of (0, 180) corresponds to the *E*, or *cis*, conformation and (180, 0) the *Z*, or *trans*, conformation.[52,53] The combination of torsion angles C1−C2−N−C7 (χ) and C3−C2−N−C7 (χ') of +120 and −120° indicates the antiperiplanar and syn-periplanar orientations of C2H−NH, respectively.

Among the 487 occurrence of the above motif in 403 structural records, all but one structure in the CSD had the C−CO−NH−C segment in *Z* conformation with a standard deviation of 7°. PDB structures had very similar tendency

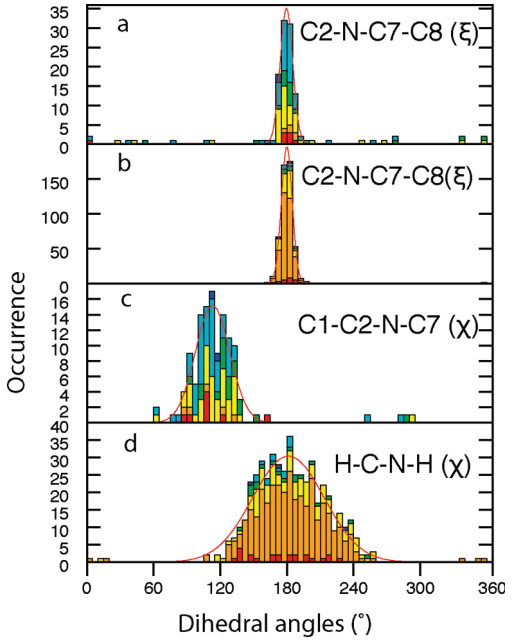

Figure 6. Distribution of dihedral angles observed in GlcNac residues in the PDB as a function of resolution (a, c) or analogue sequence in the CSD as a function of R factor (b, d). The resolution is coded as follows: red,1 Å; orange, 1.1 Å; yellow, 1.2 Å; green, 1.3 Å; magenta, 1.4 Å; blue, 1.5 Å. The *R* factors thresholds are as follows: red, 3%; orange, 6%; yellow, 9%; green, 12%; cyan, 15%; blue, 18%.

except for a handful of outliers. The *Z* and *E* conformations are difficult to distinguish from X-ray data at low resolution because the methyl group and the oxygen have 9 and 8 electrons respectively and thus the electron density is quite similar. With our data, the *Z* conformation could only be validated by reducing the number of variables fixing all other atom positions and exchanging positions of methyl group and oxygen. The dominance of Z-conformation in solution structure of *N*-acetylglucosamine was recently confirmed by NMR spectroscopy with only 1.8% of *E*-conformers.[54] When the PDB files of any resolution is chosen, the conformation distribution is more widespread.[54] This is probably due to the experimental uncertainties that comes from the limited resolution rather than a physical reality. In chitobiose single crystal study,[47] *N*-acetylglucosamine has the angles ξ and ξ' close to 180° and 0°, respectively. The deviations from ideal *Z* conformation of 16 and 10° in our structure are higher than the deviations found in the single crystal studies of chitobiose, 3 and 6° for the nonreducing ring, but coincide with the refinement by Yui et al.,[34] and correspond to the limit of reasonable structure with respect to similar motifs.

Table 2. Hydrogen Bonding Geometry

donor	acceptor	H···A distance (A)	D···A distance	D−H···A angle	acceptor residue
N2H	O7	2.02	2.71(4)	136	$x+1, y, z$
O3H	O5	1.92	2.71(2)	162	$-x, -y, z-{}^1/_2$
O6H	O7	2.22	3.01(3)	162	$-x-1, -y, z-{}^1/_2$
C2H	O7	2.33	2.81(4)	109	x, y, z
C4H	O6	2.46	2.83(4)	102	x, y, z
C8H	O6	2.54	3.35(4)	142	$x, y-1, z$

The torsion around C2−N2 is more flexible. In most reported structurs H−C−N−H is antiperiplanar, with a standard deviation of 46° for the CSD, where only 6 occurrences of synperiplanar conformation have been reported. In *N*-acetylglucosamine residues, the standard deviation of torsion angle C1−C2−N2−C7 was 20° centered at 113°. The overall morphology of the molecule drastically changes with torsion angle around the C2−N bond. The value of $\chi = 91°$ in our structure is smaller than the 100.5° and 138.7° observed in the chitobiose nonreducing sugar and reducing sugar respectively but is within the standard deviation of structures found in analogous structure segments.

Hydrogen Bonds. The hydrogen bonding geometry in our structure is listed in Table 2. The structure is stabilized by intramolecular hydrogen bonding of O3H···O5 rigidifying the glycosidic linkage and two intermolecular hydrogen bonding along the *a*-axis that stabilize the otherwise hydrophobically bound sheet structure. There is no hydrogen bonding along *b*-axis as has been mentioned by numerous authors, except for the weak polar interaction from methyl hydrogen C8H to the pendant primary alcohol O6. Any polar molecule would have better affinity, which explains the ease of polar molecules to intercalate between the sheets to expand the structure in *b*-axis direction.

The O6H···O7 hydrogen bond explains in part the small torsion angle χ of the structure as the amide oxygen is pulled to the O6 of the neighbor chain. The refinement by Yui et al.[34] involved a distance criteria for nonbonded atoms, but not for any polar interaction, while the diffraction data was not rich enough to locate the side groups with high precision. The GB model on the other hand, explicitly biased the structure for the hydrogen bonding by constraining the O6 to O7 distance. They obtained a value closer to our refinement, although the X-ray data was not rich enough to determine the structure on its own.

Advantage of High Resolution Fiber Diffraction for Crystal Structure Determination. During the first step of refinement, the phasing model contained 67 electrons out of 108 electron in the asymmetric unit. This means that nearly 40% of the structure was missing. The omit map clearly showed all positions of these missing atoms without relying on model building. This was facilitated by several factors in our case. The unit cell was relatively small so that the average distance between diffraction peaks was relatively large. The crystal size was exceptionally large for a polymer sample so the diffraction peaks were narrow. Thus, the overlaps of diffraction peaks were limited and peaks could be reasonably separated without sophisticated peak shape analysis. Although these conditions might not be fulfilled for many polymer samples, there are several advantages in determining structures at atomic resolution using samples with fiber texture, including small molecules that does not easily grow into macroscopic crystals.

Several alternative methods to determine structures when a crystal cannot be obtained at a macroscopic size are being developed: single crystal analysis using micro- or nano-focused X-ray[55] or electron[56] beams and high resolution powder diffraction.[57,58] Apart form technical difficulty associated with manipulation of tiny objects, collecting diffraction from small sample volume inevitably encounters the problem of beam damage. In the powder diffraction method, the three-dimensional diffraction pattern collapses into one-dimension, leading to peak overlaps. In fiber diffraction, the diffraction pattern collapses into two-dimension, and thus the peak overlap problem lies between single crystal and powder diffraction. Since almost the whole reciprocal space is covered in one shot, the data acquisition is the most efficient. The instrumental peak broadening can be smaller in a powder diffractometer, but if the peak broadening due to crystal size is more important than instrumental broadening, diffraction on uniaxially oriented samples (fiber diffraction) has a great advantage over powder diffraction. As the sample preparation and the acquisition is quite simple, fiber diffraction can be used for high throughput structure determination for small crystals with a morphology of high aspect ratio.

Acknowledgment. This study was supported by a grant from Agence National de la Recherche. We thank Dr. H. Chanzy for valuable discussions. The synchrotron radiation experiments were performed at BL40B2 in SPring-8 with the approval of the Japan Synchrotron Research Institute (JASRI).

Supporting Information Available: CIF file of the refined structure of β-chitin. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Igarashi, K.; Wada, M.; Samejima, M. *FEBS J.* **2007**, *274*, 1785–1792.
(2) Lamarque, G.; Viton, C.; Domard, A. *Biomacromolecules* **2004**, *5*, 1899–1907.
(3) Blackwell, J. *Biopolymers* **1969**, *7*, 281–298.
(4) Saito, Y.; Okano, T.; Putaux, J.-L.; Gaill, F.; Chanzy, H. Crystallosolvates of beta chitin and alcohols. In *Advances in Chitin Science II*; Domard, E. A., Roberts, G., Vårum, K., Eds.; Jacque André Publisher: Lyon, France, 1998; pp 507−512.
(5) Saito, Y.; Okano, T.; Gaill, F.; Chanzy, H.; Putaux, J.-L. *Int. J. Biol. Macromol.* **2000**, *28*, 81–88.
(6) Saito, Y.; Kumagai, H.; Wada, M.; Kuga, S. *Biomacromolecules* **2002**, *3*, 407–410.
(7) Rössle, M.; Flot, D.; Engel, J.; Burghammer, M.; Riekel, C.; Chenzy, H. *Biomacromolecules* **2003**, *4*, 981–986.
(8) Noishiki, Y.; Nishiyama, Y.; Wada, M.; Okada, S.; Kuga, S. *Biomacromolecules* **2003**, *4*, 944–949.
(9) Noishiki, Y.; Kuga, S.; Wada, M.; Hori, I.; Nishiyama, Y. *Macromolecules* **2004**, *37*, 6839–6842.
(10) Yoshifuji, A.; Noishiki, Y.; Wada, M.; Heux, L.; Kuga, S. *Biomacromolecules* **2006**, *7*, 2878–2881.
(11) Arnott, S. *ACS Symp. Ser.* **1980**, *141*, 1–30.
(12) Gardner, K. H.; Blackwell, J. *Biopolymers* **1974**, *13*, 1975–2001.
(13) Woodcock, C.; Sarko, A. *Macromolecules* **1980**, *13*, 1183–1187.
(14) French, A. D.; Roughead, W. A.; Miller, D. P. *ACS Symp. Ser.* **1987**, *340*, 15–37.
(15) Langan, P.; Nishiyama, Y.; Chanzy, H. *Biomacromolecules* **2001**, *2*, 410–416.
(16) Nishiyama, Y.; Langan, P.; Chanzy, H. *J. Am. Chem. Soc.* **2002**, *124*, 9074–9082.
(17) Nishiyama, Y.; Sugiyama, J.; Chanzy, H.; Langan, P. *J. Am. Chem. Soc.* **2003**, *125*, 14300–14306.
(18) Nishiyama, Y.; Johnson, G. P.; French, A. D.; Forsyth, V. T.; Langan, P. *Biomacromolecules* **2008**, *9*, 3133–3140.
(19) Nishiyama, Y. *J. Wood Sci.* **2009**, *55*, 241–249.
(20) Wada, M.; Heux, L.; Nishiyama, Y.; Langan, P. *Cellulose* **2009**, *16*, 943–957.
(21) Nishiyama, Y.; Langan, P.; Wada, M.; Forsyth, V. T. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 1172–1177.
(22) Herth, W.; Barthlott, W. *J. Ultrastruct. Res.* **1979**, *68*, 6–15.
(23) Herth, W. *J. Ultrastruct. Res.* **1979**, *68*, 16–27.
(24) Blackwell, J. The macromolecular organization of cellulose and chitin. In *Cellulose and other Natural Polymer Systems*; Brown, R. M. , Jr., Ed.; Plenum Press: New York, 1982; pp 403−428.

(25) Gay, L.; Chanzy, H.; Bulone, V.; Girard, V.; Févre, N. *J. Gen. Microbiol.* **1993**, *139*, 2117–2122.
(26) Sakamoto, J.; Sugiyama, J.; Kimura, S.; Imai, T.; Itoh, T.; Watanabe, T.; Kobayashi, S. *Macromolecules* **2000**, *33*, 4155–4160.
(27) Persson, J. E.; Domard, A.; Chanzy, H. *Int. J. Biol. Macromol.* **1992**, *23*, 221–224.
(28) Saito, Y.; Putaux, J.-L.; Okano, T.; Gaill, F.; Chanzy, H. *Macromolecules* **1997**, *30*, 3867–3873.
(29) Noishiki, Y.; Takami, H.; Nishiyama, Y.; Wada, M.; Okada, S.; Kuga, S. *Biomacromolecules* **2003**, *4*, 896–899.
(30) McLachlan, J.; McInnes, A. G.; Falk, M. *Can. J. Bot.* **1965**, *43*, 707–713.
(31) Gardner, K. H.; Blackwell, J. *Biopolymers* **1975**, *14*, 1581–1595.
(32) Dweltz, N. E.; Colvin, J. R.; McInnes, A. G. *Can. J. Chem.* **1968**, *46*, 1513–1521.
(33) Tanner, S. F.; Chanzy, H.; Vincendon, M.; Roux, J. C.; Gaill, F. *Macromolecules* **1990**, *23*, 3576–3583.
(34) Yui, T.; Taki, N.; Sugiyama, J.; Hayashi, S. *Int. J. Biol. Macromol.* **2007**, *40*, 336–344.
(35) Kahn, R.; Fourme, R.; Gadet, A.; Janin, J.; Dumas, C. *J. Appl. Crystallogr.* **1982**, *15*, 330–337.
(36) Nishiyama, Y.; Langan, P. *Fibre Diffr. Rev.* **2000**, *9*, 18–23.
(37) Dorset, D. L. *Rep. Prog. Phys.* **2003**, *66*, 305–338.
(38) Clark, M.; Cramer-III, R. D.; Van Obdenbosch, N. *J. Comput. Chem.* **1989**, *10*, 982–1012.
(39) Nishiyama, Y.; Mazeau, K.; Morin, M.; Cardoso, M. B.; Chanzy, H.; Putaux, J.-L. *Macromolecules* **2010**, *43*, 8628–8636.
(40) Emsley, P.; Cowtan, K. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2004**, *60*, 2126–2132.
(41) Hamilton, W. C. *Acta Crystallogr.* **1965**, *18*, 502–510.
(42) Spek, A. L. *Acta Crystallogr., Sect. D: Biol. Crystallogr* **2009**, *65*, 148–155.
(43) Sugiyama, J.; Boisset, C.; Hashimoto, M.; Watanabe, T. *J. Mol. Biol.* **1999**, *286*, 247–255.
(44) Imai, T.; Watanabe, T.; Yui, T.; Sugiyama, J. *FEBS Lett.* **2002**, *510*, 201–205.
(45) Imai, T.; Watanabe, T.; Yui, T.; Sugiyama, J. *Biochem. J.* **2003**, *374*, 755–760.
(46) Cremer, D.; Pople, J. A. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
(47) Mo, F.; Jensen, L. H. *Acta Crystallogr., Sect. B: Struct. Sci.* **1978**, *34*, 1562–1569.
(48) Mo, F.; Jensen, L. H. *Acta Crystallogr., Sect. B: Struct. Sci.* **1975**, *31*, 2867–2873.
(49) Fletcher, D. A.; McMeeking, R. F.; Parkin, D. *J. Chem. Inf. Comp. Sci.* **1996**, *36*, 746–749.
(50) Allen, F. H. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380–388.
(51) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Kessler, M.; Macrae, C. F.; McCabe, P.; Pearson, J.; Taylor, R. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 389–397.
(52) Mazeau, K.; Pérez, S.; Rinaudo, M. *J. Carbohydr. Chem.* **2000**, *19*, 1269–1284.
(53) Fowler, P.; Bernet, B.; Vasella, A. *Helv. Chim. Acta* **1996**, *79*, 269–287.
(54) Hu, X.; Carmichael, I.; Serianni, A. S. *J. Org. Chem.* **2010**, *75*, 4899–4910.
(55) Popov, D.; Buléon, A.; Burghammer, M.; Chanzy, H.; Montesanti, N.; Putaux, J.-L.; Potocki-Véronèse, G.; Riekel, C. *Macromolecules* **2009**, *42*, 1167–1174.
(56) Dorset, D. L. *Structural Electron Crystallography*; Plenum Press: New York, 1995.
(57) Margiolaki, I.; Wright, J. P.; Wilmanns, M.; Fitch, A. N.; Pinotsis, N. *J. Am. Chem. Soc.* **2007**, *129*, 11865–11871.
(58) Margiolaki, I.; Wright, J. P.; Fitch, A. N.; Fox, G. C.; Von Dreele, R. B. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2005**, *61*, 423–432.