ORIGINAL RESEARCH

Impact of hemicelluloses and crystal size on X-ray scattering from atomistic models of cellulose microfibrils

Aleksi Zitting · Antti Paajanen · Paavo A. Penttilä

Abstract X-ray scattering methods allow efficient characterization of cellulosic materials, but interpreting their results is challenging. By creating molecular dynamics models of cellulose microfibrils and calculating the scattering from them, we investigated how different properties of the structures affect their scattering intensities. We studied the effects of hemicelluloses and crystal size on small-angle and wide-angle X-ray scattering (SAXS, WAXS). Microfibril models with and without surface-bound hemicelluloses were built based on the chemical composition of spruce secondary cell walls. The effect of fibril size was investigated by comparing the scattering from fibrils with 14 to 40 cellulose chains. The hemicelluloses appeared in the SAXS region as an increase in the fibril radius and as a clear contribution of a shell around the fibril. The hemicelluloses also increased the crystal size as determined from the broadening of the 200 diffraction peak of cellulose I_β. The SAXS and WAXS analysis provided consistent estimates

for the size of the microfibrils, and their special features and challenges were discussed. In particular, the results of 18-chain microfibrils were consistent with prior experimental results. Carrying out the simulations in wet and dry environments had the most pronounced effect on fibrils with a hemicellulose coating. Twisting of the fibril had very little impact on most properties, except for a minor effect on the WAXS peaks. The results allow for more correct interpretation of experimental scattering results, leading to more accurate descriptions of microfibril structures in natural and processed cellulosic materials.

Keywords Cellulose microfibrils · Cellulose crystals · Crystal size · X-ray scattering · Molecular modelling · Hemicellulose

A. Zitting · P. A. Penttilä (✉)
Department of Bioproducts and Biosystems, Aalto University, Espoo, Finland
e-mail: paavo.penttila@aalto.fi

A. Paajanen
VTT Technical Research Centre of Finland Ltd, Espoo, Finland

Introduction

Wood is a complex and hierarchical material used in a variety of applications. Ultimately, most properties of wood arise from its cell wall nanostructure, which consists of several biomolecules, primarily cellulose, hemicelluloses and lignin. While there have been many strides in understanding the fundamental structure and composition of the cell wall (Fernando et al. 2023), many aspects are still uncertain due to the difficulty of characterizing the structures in their natural state. As an example, the number of cellulose chains in a microfibril, and how much the number varies

within or between plants, is still uncertain (Cosgrove 2022). The structural characterization of nanocelluloses, such as cellulose nanofibrils (CNF) and cellulose nanocrystals (Balea Martin et al. 2021; Foster et al. 2018), is similarly challenging, particularly when water is present. Both fields would benefit from dedicated method development.

X-ray and neutron scattering are non-destructive means to characterize the nanoscale fibrillar structures in the cell wall and in processed cellulosic materials (Martínez-Sanz et al. 2015). However, interpreting the scattering results requires knowledge of the underlying structure. By utilizing molecular models and the scattering calculated from them, one can estimate how certain structures scatter, which helps to interpret experimental scattering results (Hadden et al. 2014; Inouye et al. 2014; Kubicki et al. 2018; Langan et al. 2014; Lindner et al. 2015; Newman 2008; Newman et al. 2013; Nishiyama et al. 2012; Paajanen et al. 2022; Penttilä et al. 2021; Rosén et al. 2020; Yang and Kubicki 2020; Zhang et al. 2015). For instance, this approach can be used to determine the scattering contribution from hemicelluloses, which surround the crystalline cellulose microfibrils in secondary cell walls of wood (Terrett et al. 2019). Scattering intensities calculated from well-defined model structures may give new insights into specific factors contributing to experimental scattering data from wood and other cellulosic materials, such as inherent variation of the size and shape of cellulosic fibrils in real samples and the presence of hemicelluloses around the microfibrils.

X-ray diffraction (XRD) and wide-angle X-ray scattering (WAXS) are sensitive to atomic-level crystal structures in the sample. Their data is interpreted most simply in terms of the locations of the diffraction peaks on the scattering angle or scattering vector axis, which correspond to the spacings between crystallographic planes. At the same time, the radial broadening of the diffraction peaks is affected by several factors, some of which carry information on the sample structure. These factors include instrumental resolution (e.g., due to a finite distribution of X-ray wavelength or beam collimation), mechanical strains and other crystal imperfections, and eventually the number of the stacked lattice planes in a crystal (Cullity and Stock 2001; Mittemeijer and Welzel 2008). The effect of the crystal size on the peak broadening can be quantified with the aid of the Scherrer equation

(Langford and Wilson 1978; Scardi et al. 2004), which is commonly used in studies of cellulose and other polymers (Daicho et al. 2018; Driemeier and Bragatto 2013; Elazzouzi-Hafraoui et al. 2008; Fang and Catchmark 2014; Fernandes et al. 2011; Fink et al. 1995; Jakob et al. 1995; Jarvis 2018; Leppänen et al. 2009; Newman et al. 2013; Thomas et al. 2012). In most cases, the other sample-related sources of peak broadening are neglected, and all peak broadening that remains after correcting for instrumental broadening is attributed to the coherence length of the crystal in the direction perpendicular to the respective lattice plane (*hkl*) (crystal size L_{hkl}). In the case of cellulose, there are specific factors that could contribute to the peak broadening. These include the presence of bound hemicelluloses on the fibril surfaces, conformational disorder in the surface chains, and the shape of the fibril cross-section. Moreover, the absolute value given by the Scherrer equation is slightly affected by the choice of the Scherrer constant K (Langford and Wilson 1978), the shape of the diffraction peak and the definition of its width, and whether the fitting was done to the intensity as a function of the scattering angle 2θ or the scattering vector q. Quantifying the effects of these factors would be important to assess the reliability of crystal size values determined for fibrillar cellulose crystals based on diffraction peak broadening.

The exact contribution of cellulose microfibrils and the surrounding hemicelluloses to lower scattering angles, i.e., to small-angle X-ray scattering (SAXS) has also remained elusive. This is partially due to difficulties in determining the geometry of a representative fibril cross-section, which could be used in the analysis of the data by model fitting. Cross-sections varying from circular to rectangular and more customized shapes have been utilized to approximate the form factor representing the scattering from cellulosic microfibrils in wood-based samples (Elazzouzi-Hafraoui et al. 2008; Chen et al. 2021; Mao et al. 2017; Rosén et al. 2020; Schmitt et al. 2018; Su et al. 2014; Viell et al. 2020). Typically, a circular cross-section has been inferred a sufficient approximation for microfibrils in wood (Jakob et al. 1995; Leppänen et al. 2009), and any possible deviation from this may be taken into account by including statistical variation (polydispersity) to the value of the fibril diameter (Penttilä et al. 2019). A slightly more sophisticated form factor model for the cellulose microfibrils would

include a core-shell structure (Pedersen 1997). In such a structure, the core could represent the denser core of the microfibril and the shell the combined contribution of the deviation from circular geometry and the chains at the fibril surface, either hemicelluloses or weakly ordered cellulose chains. In order to develop such more accurate approximations of the cellulose microfibril for small-angle scattering analysis, it is necessary to investigate the specific effects of the cross-sectional shape and hemicellulose coating on the scattering intensities.

In this work, we have created molecular models of cellulose microfibrils that are surrounded by hemicellulose chains, as well as ones with different lateral size. They thus represent ideal single microfibrils, separated from the plant cell wall or other source. We investigate how the presence of hemicelluloses impacts scattering both in the WAXS and SAXS regimes and in wet and dry conditions, especially in terms of the apparent fibril size. In addition, we look at different fibril sizes to see how accurately scattering analysis reproduces the structural features of cellulose microfibrils. Utilizing these results, the impact of fibril size and hemicelluloses on scattering becomes more apparent, making the interpretation of experimental scattering data easier and more accurate.

Methods

Molecular models

Microfibrils Molecular models of cellulose microfibrils were created starting from crystallographic unit cell data for the cellulose I_{β} polymorph (Nishiyama et al. 2002). Two types of models were created: (1) straight fibrils formed by periodic chains, and (2) twisting fibrils formed by non-periodic chains. Periodicity refers here to covalent bonding between the first and last repeat unit of a chain across the periodic boundary of the simulation domain. The results shown in the main article are for the straight fibrils. Corresponding data for the twisting fibrils are given in the Supplementary Information (SI).

Six different chain arrangements were used to build microfibrils of different diameter but a similar hexagonal cross-section (Fig. 1a). Here the 18-chain microfibril with the 2-3-4-4-3-2 chain arrangement

and larger ($1\bar{1}0$) than (110) surface was used as the basis for building the models, because it was considered the most representative of microfibrils in wood (Cosgrove 2022; Nixon et al. 2016; Yang and Kubicki 2020). It exposes both hydrophobic and hydrophilic crystallite surfaces and is nearly circular (Kubicki et al. 2018). Using a similar hexagonal shape for all fibrils reduces the effect of the cross-sectional shape and allows to compare only the influence of the crystal size on the scattering. The number of chains in consecutive hydrogen-bonded layers is given by 2-$\cdots$-($N-1$)-N-($N-1$)-$\cdots$-2 (odd number of layers) and 2-$\cdots$-($N-1$)-N-N-($N-1$)-$\cdots$-2 (even number of layers) with $N = \{4, 5, 6\}$, which results in fibrils with a chain count between 14 and 40. When the number of layers is even, the relative position of the two middle layers can be chosen in two different ways that affects the proportions of the ($1\bar{1}0$) and (110) surfaces (Daicho et al. 2018; Rosén et al. 2020). The one leading to a more circular cross-section was chosen here. In each model, the chains were 34 glucose units long. The initial atomic coordinates of the straight fibrils were cut from a model of a cellulose single crystal that was first equilibrated at normal conditions. Equilibration refers to the simulation of molecular dynamics until the system reaches an equilibrium state with respect to chosen observables, in this case the various components of potential energy and the unit cell vectors. The initial coordinates of the twisting fibrils were created directly from the unit cell data. The helical twist would develop spontaneously during the subsequent production simulation in water.

After the initial coordinates and molecular topologies were created, the systems were subjected to the following simulation workflow: (1) energy minimization, (2) solvation with water and energy minimization, (3) solvent equilibration, (4) production simulation in water, (5) removal of water, (6) energy minimization, and (7) production simulation in vacuum. The production simulation in water was carried out for 600 ns in the isothermal-isobaric ensemble at 300 K temperature and 1 atm pressure. The production simulation in vacuum was carried out for 400 ns in the canonical ensemble at 300 K temperature. These are later referred to as the wet and dry environment, respectively. Further details of the simulation set-up are given at the end of this section.
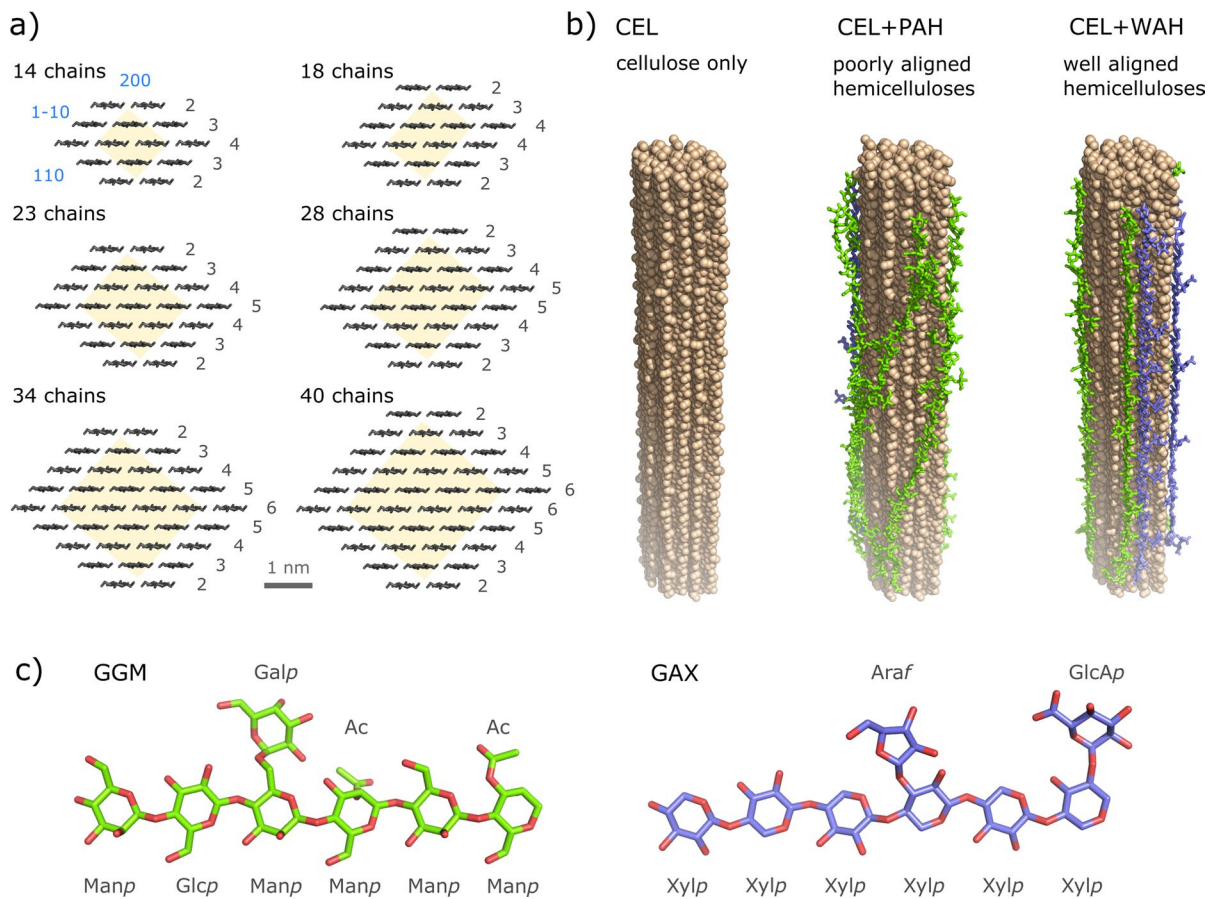
Fig. 1 Models of neat and hemicellulose-coated cellulose microfibrils. **a** Cross-sections of fibrils of different size. The total number of chains and the number of chains in each hydrogen-bonded layer are indicated for each cross-section. The (200), (1̄10) and (110) crystallographic planes are indicated for the top left cross-section. Core-chains are highlighted in yellow. **b** Straight fibrils with no surface-bound hemicelluloses (CEL), poorly aligned hemicelluloses (PAH) and well aligned hemicelluloses (WAH) in the dry state. The fibrils have 18 cellulose chains in the 2-3-4-4-3-2 arrangement (top right cross-section in **a**). The cellulose chains consist of 34 repeat units and the GGM and GAX chains of 30 repeat units. The PAH and WAH models were built with 16 different hemicellulose configurations, only one of which is shown in the figure. **c** Chemical structures of the hemicelluloses, exemplified by chain segments of six repeat units

Microfibrils with surface-bound hemicelluloses
The effect of surface-bound hemicelluloses on the scattering intensity was studied using two series of models: (1) fibrils with poorly aligned hemicelluloses (PAH), and (2) fibrils with well aligned hemicelluloses (WAH) (see Fig. 1b). Both were based on the 18-chain microfibril that was used as a basis for the fibrils with different crystal sizes. In the PAH series, hemicelluloses were adsorbed on the fibril through simulation in water environment (at the beginning of step (4) of the simulation workflow). Four galactoglucomannan (GGM) and two glucuronoarabinoxylan (GAX) chains were uniformly distributed around the

fibril at roughly 1 nm distance from its surface. The order of the chains and their rotation about their longitudinal axis was random. The chains were initially in the extended twofold helical screw conformation and parallel to the fibril axis. During the adsorption simulation, they would attach to the fibril surface while remaining roughly aligned with the fibril axis. In the WAH series, the hemicelluloses were generated directly on the fibril surface so that they conform to the cellulose crystal lattice (shared backbone conformation, orientation and lattice spacing).

Both series of hemicellulose-coated fibrils consisted of 16 variants with different hemicellulose

configurations. The mass fraction and chemical structure of the hemicelluloses were chosen to mimic those found in Norway spruce (*Picea abies*) (Zitting et al. 2021). The distribution of repeat units and side groups was random, with certain exceptions set by known substitution patterns. Both the GGM and the GAX chains consisted of 30 repeat units. After the initial placement of the hemicelluloses, the simulation protocol was similar to that used for the neat cellulose fibrils.

Hemicelluloses The chemical structures used for GGM and GAX are as follows (see Fig. 1c). The GGM chains have a random backbone of β-(1→4)-linked D-glucopyranosyl (Glc$p$) and D-mannopyranosyl (Man$p$) units (Glc$p$:Man$p$ = 1:4). The Man$p$ C2 and C3 hydroxyls are randomly acetylated (Ac) (Ac:Man$p$ = 1:4). The Man$p$ units have a random pattern of α-(1→6)-linked D-galactopyranosyl (Gal$p$) side groups (Gal$p$:Glc$p$ = 1:1). The GAX chains have a backbone of β-(1→4)-linked D-xylopyranosyl (Xyl$p$) units. The Xyl$p$ units have a regular pattern of α-(1→2)-linked glucuronic acid (GlcAp) side groups, with a GlcAp attached to every 6th Xyl$p$ unit (GlcAp:Xyl$p$ = 1:6). Lastly, Xyl$p$ units have a random pattern of α-(1→3)-linked L-arabinofuranosyl (Ara$f$) side groups, which only occur two Xyl$p$ units removed from the GlcAp groups towards the non-reducing end of the chain (Ara$f$:Xyl$p$ = 1:12).

Details of the simulation set-up All simulations were performed using GROMACS (Abraham et al. 2015) and the GLYCAM06H force field (Kirschner et al. 2008). Covalent bonds involving hydrogen were constrained using the LINCS algorithm (Hess et al. 1997). Water was described using the TIP3P model (Jorgensen et al. 1983). Temperature control was implemented using a stochastic variant of the Berendsen thermostat (Bussi et al. 2007) with a 200 fs time constant. Pressure control was implemented with the Berendsen barostat (Berendsen et al. 1984) with a 2 ps time constant. The total linear momentum of the system was set to zero at 2 ps intervals. The equations of motion were integrated using the velocity-Verlet algorithm with a 2 fs time step. GROMACS utilities and MDTraj (McGibbon et al. 2015) were used for trajectory analysis. The doGlycans tool (Danne et al. 2017) was used for creating the models of hemicelluloses.

Direct analysis of molecular models

The molecular trajectories were post-processed to determine (1) the lattice spacings and (2) the fibril width along the different crystallographic directions, as well as their local variation, and (3) the radial electron density distribution with respect to the fibril axis. For calculating the lattice spacings, the trajectories were coarse-grained from atomistic to sugar unit resolution. This was done by replacing each glucose unit with a single point particle at the center of mass of its pyranose ring. The vectors adjoining the glucose units and their nearest neighbors were used to calculate local estimates for the unit cell parameters, which were then converted to interplanar distances. These were averaged over the glucose units of each chain and over time, to obtain lattice spacing distributions within the fibril cross-section. The first and last pair of glucose units along each chain were omitted from the analysis to exclude possible distortion at the fibril ends. The fibril widths were calculated in a similar fashion by determining the distances between outermost glucose unit pairs in each crystallographic direction. It should be noted that when determined in this way, the fibril widths are slightly underestimated in comparison to the full outer dimensions of the molecular models.

For calculating the radial electron density distributions, the electrons were counted at the locations of atomic nuclei and the principal axis of the fibril was used as the axis of reference. The used spatial resolution was 1 pm and time averaging was applied over the trajectory. A two-step function was fitted against each radial distribution to get a core-shell cylinder approximation of the fibril's electron density. The core-shell approximation is defined by the radius of the core cylinder (R), the thickness of the cylindrical shell (t) and the ratio of the shell and core densities (ρ_{rel}) (see Fig. 2). The uncertainties of the core radius (ΔR) and shell thickness (Δt), caused by the blurring of the density profile steps, were quantified based on a Gaussian approximation of the distribution near the steps. The trajectory analyses described above were applied to the last 200 ns of each production simulation. Further details of the analyses are given in the SI.
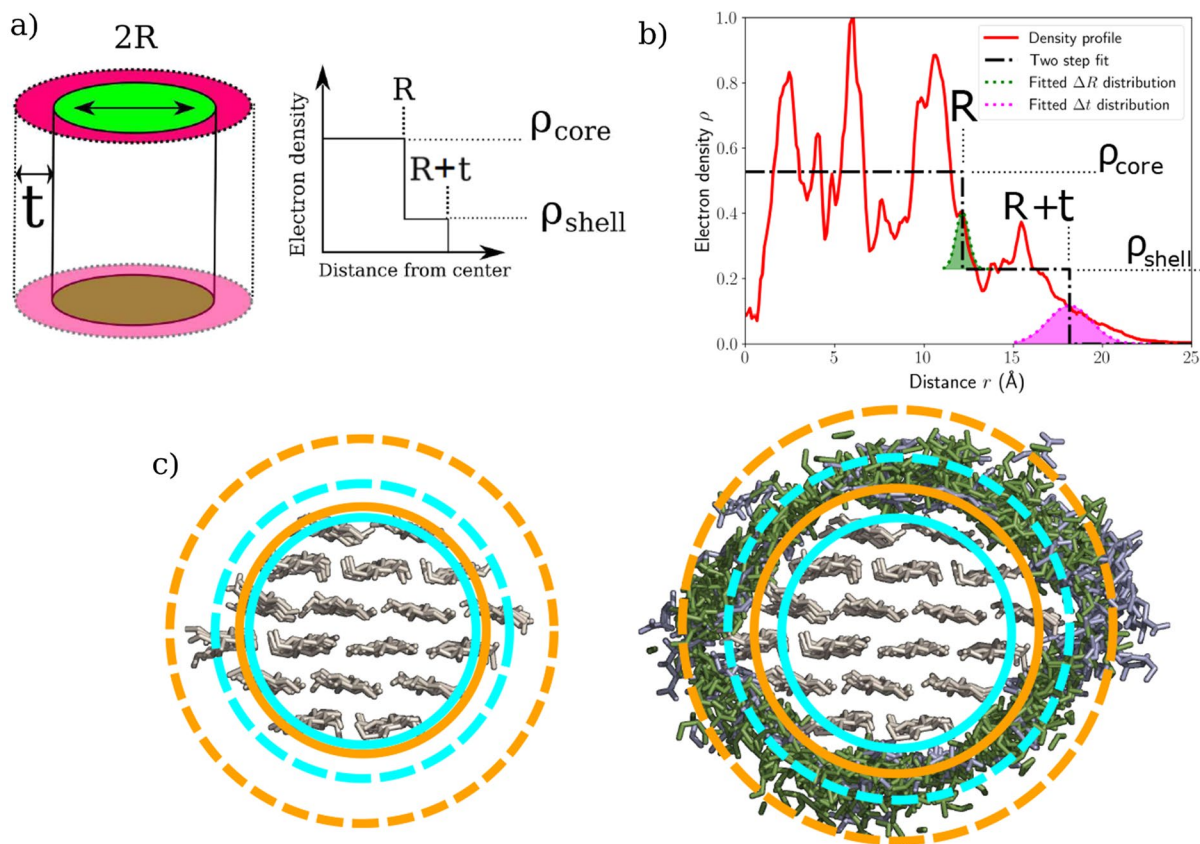
Fig. 2 Core-shell cylinder models and their parameters. **a** Visual clarification of the core-shell cylinder geometry and the parameters describing it. **b** Example of a radial electron density profile as obtained from the molecular models, and the step-function fit used to determine the parameters R, t and ρ_{rel}. **c** Results of a core-shell cylinder fit in the SAXS intensities (orange) and the radial electron density profiles (cyan) of a wet fibril with (right) and without (left) poorly aligned hemicelluloses. The cylinder core (solid line) and shell (dashed line) are visualized on top of a model snapshot (hemicelluloses visible from all configurations, cellulose crystal from only one)

Scattering computation

Equatorial and meridional scattering intensities were calculated using the cylindrically-averaged form of the Debye scattering equation (Zhang et al. 2016):

$$I(q_R, q_Z) = \sum_i^n \sum_j^n f_i(q_R, q_Z) f_j(q_R, q_Z) J_0(2\pi r_{ij} q_R) \cos(2\pi z_{ij} q_Z)$$
(1)

where q_R and q_Z are the magnitudes of the equatorial and meridional scattering vectors, respectively, n is the number of atoms, f_i and f_j are the atomic form factors of atoms i and j, respectively, r_{ij} and z_{ij} are the radial and axial components of interatomic distance between atoms i and j, respectively, and J_0 is a zeroth

order Bessel function of the first kind. For calculating the scattering intensities, the molecular trajectories were sampled at 10 ns intervals over the last 200 ns to obtain 20 snapshots.

For equatorial scattering, the fibrils were divided into axial segments of roughly one nanometer. Scattering intensities were calculated separately for each snapshot of each segment, and then averaged over the snapshots and the segments. For the two series of models with hemicelluloses, averaging was also applied over the different hemicellulose configurations. The first and last pair of glucose units along each cellulose chain were omitted from the scattering calculation to exclude possible distortion at the fibril ends and the part of the fibril not covered by hemicelluloses. The principal axis of the fibril was chosen as the symmetry axis.

For meridional scattering, the procedure was otherwise similar, but the fibrils were not divided into axial segments. The calculations were performed both with and without the surface-bound hemicelluloses. Water was not included in the analysis due to two kinds of difficulties arising from it. Including all water molecules would create a strong, unwanted contribution of water to the scattering, whereas only including a layer of them around the fibril would introduce a strong scattering from the water-vacuum interface.

Analysis of computed scattering

The computed scattering intensities were divided into two regions, the small-angle region ($q < 0.75 \text{ Å}^{-1}$) and the wide-angle region ($q > 0.75 \text{ Å}^{-1}$). The small-angle region corresponds to structures larger than 8 Å, mostly scattering from the microfibril, while the wide-angle region corresponds to structures smaller than 8 Å, mostly from the crystal structure inside the fibril. The boundary value of $q = 0.75 \text{ Å}^{-1}$ was chosen because it corresponds to a local minimum of the intensity and is immediately followed by the $1\bar{1}0$ peak of cellulose I_β in the WAXS region.

The SAXS region analysis was based on scattering from core-shell cylinders, where the main fitting parameters are the mean cylinder radius R, standard deviation of the radius ΔR, shell thickness t and the relative electron density of the shell ρ_{rel}. The equatorial scattering intensity is proportional to the average of the square of the equatorial form factor F with respect to the size distribution (here Gaussian):

$$I(q) = < |F(q)|^2 > . \tag{2}$$

The equatorial form factor for an infinitely long cylinder with radius R in vacuum is proportional to (Pedersen 1997, Hashimoto et al. 1994)

$$F(q) \propto \pi R^2 \frac{J_1(qR)}{qR} \tag{3}$$

where J_1 is the first order Bessel function of the first kind. The form factor of two concentric cylinders, i.e. a core-shell cylinder, is a sum of the form factors of the core and the shell. For a core-shell cylinder in vacuum, where the electron densities of the cylinder core and shell are ρ_{core} and ρ_{shell}, respectively, and the shell has a thickness t, the equatorial form factor becomes

$$F(q) \propto (\rho_{core} - \rho_{shell})\pi R^2 \frac{J_1(qR)}{qR}$$
$$+ \rho_{shell}\pi(R + t)^2 \frac{J_1(q(R + t))}{q(R + t)}. \tag{4}$$

When fitting the equatorial SAXS intensities of cellulosic fibrils, we expressed the electron density in Eq. 4 relative to the density of the cylinder core ($\rho_{rel} = \rho_{shell}/\rho_{core}$). In the fit to the equatorial SAXS intensity of the hemicellulose shell only, we set $\rho_{core} = 0$. The shape of the fibril in the molecular models deviates from a cylinder. In the scattering analysis, these deviations were mostly accounted for by including polydispersity ΔR in the model. Including a shell around the cylinder core improved the fit even in the case of neat cellulose, but it did not have any large effect on the cylinder radius. The fit also included a constant background term.

The WAXS intensities were analyzed by fitting Gaussian functions to the four strongest diffraction peaks indexed based on the crystal structure of cellulose I_β: $1\bar{1}0$, 110 and 200 in equatorial scattering, 004 in meridional scattering. The WAXS fit included a background component consisting of a sum of exponentially decreasing intensity $I(q) \propto q^{-4}$ and a constant background, scaled according to the level of the intensity around $q = 0.8$ and 2.0 Å^{-1}. The d-spacings corresponding to the peak locations were obtained using Bragg's law:

$$d_{hkl} = \frac{2\pi}{q_{hkl}}, \tag{5}$$

where d_{hkl} is the lattice spacing for planes represented by the Miller indices hkl and q_{hkl} is the location of the peak. The peak widths of the Gaussians were used to calculate the crystal size perpendicular to the planes (hkl) using the Scherrer equation (Cullity and Stock 2001; Langford and Wilson 1978; Scardi et al. 2004) as a function of scattering angle 2θ:

$$L_{hkl,2\theta} = \frac{K\lambda}{\beta cos\theta}, \tag{6}$$

where L_{hkl} is the crystal size, K is the Scherrer constant or shape factor ($K = 1$ in this work), β is the line broadening at half the maximum intensity (in radians) and θ is the Bragg angle of the peak (one-half of 2θ). The equation can be converted to a simpler form in the q-representation:
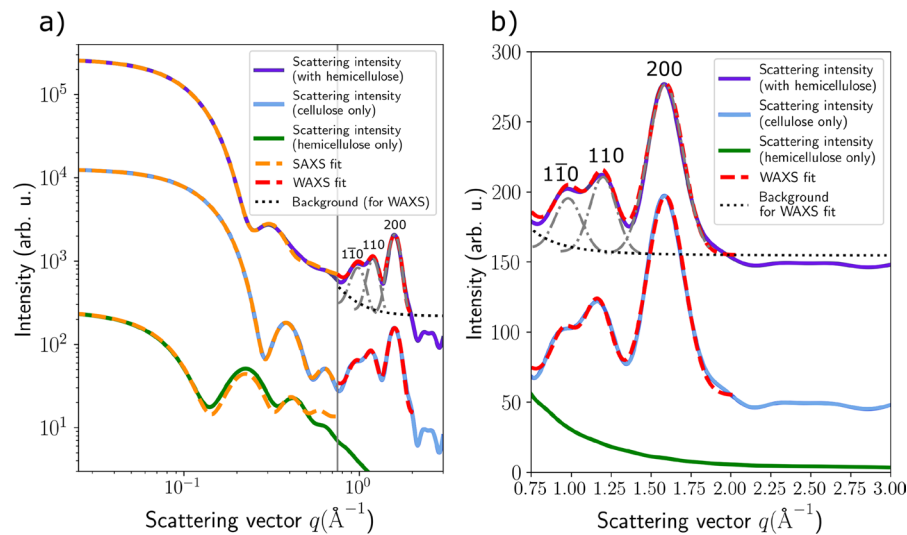
Fig. 3 Equatorial scattering intensities from wet, straight fibrils with 18 cellulose chains and poorly-aligned hemicellulose coating, plotted on fully logarithmic (**a**) and linear (**b**) axes. The scattering intensities calculated from the whole fibril (CEL+PAH), from the cellulose chains only (CEL+PAH, scattering from cellulose only) and from the hemicellulose chains only (CEL+PAH, scattering from hemicelluloses only) are presented, shifted vertically and scaled arbitrarily for clarity. Fits of the core-shell model (Eq. 4) in the SAXS regime and Gaussian peaks in the WAXS regime are also shown

$$L_{hkl,q} = \frac{2\pi K}{\Delta q_{hkl}}, \tag{7}$$

where we describe the peak broadening Δq by the integral breadth of the peak at q_{hkl}. We did not observe differences larger than 0.5% between the two ways of determining the crystal size from peak broadening (Eqs. 6 and 7) and thus present the results only based on Eq. 7.

Results and discussion

Overall effects of hemicelluloses on scattering

To determine the contribution of hemicelluloses to X-ray scattering from cellulose microfibrils, scattering intensities were computed from microfibril models both with and without hemicelluloses. For the former, two different levels of hemicellulose alignments were considered. The effect of hemicelluloses could be seen clearly in the equatorial SAXS intensities, with an example shown in Fig. 3a. The effect is visible in the q-range of 0.2–0.8 Å^{-1}, where the hemicelluloses dampen the fluctuations of the form-factor scattering in that region. Such clear intensity

fluctuations can be observed in scattering patterns from well-defined and monodisperse particles with sharp interfaces, whereas the loss of such features can be an indication of a less-clearly defined interface (Li et al. 2016). The scattering from the hemicelluloses alone is also presented in Fig. 3, together with a SAXS fit of a cylindrical shell with thickness 0.58 nm. The scattering from the hemicelluloses shows similar form-factor fluctuations as the cellulose fibrils but with the intensity maxima partially overlapping with the minima of the neat cellulose fibril intensities and vice-versa. This, together with the loss of a sharp interface between the fibril and the surrounding medium, explains why the hemicellulose coating leads to a smearing out of the sharp features of the scattering from the cellulose fibrils. The same applies even when the hemicelluloses are well-aligned on the surface of the microfibril. These observations indicate that the presence of hemicelluloses on microfibril surfaces can lead to detectable differences in SAXS patterns measured with a sufficiently high q-resolution and signal-to-noise ratio.

In the WAXS region, the effect of hemicelluloses is much less noticeable (Fig. 3b). The high-q tail of the SAXS contribution from hemicelluloses partially overlaps with the 1$\bar{1}$0 and 110 diffraction peaks, which

also partially overlap with each other. This makes investigating and fitting the low-q side of the WAXS region difficult. We also note, based on Fig. 3b, that the hemicellulose coatings tested in this work do not lead to any broad amorphous halo in the WAXS pattern, as would be expected for a polymer melt or other truly amorphous molecular configuration. Instead, the scattering from the hemicelluloses seems to blend in with the diffraction peaks of cellulose. This makes it impossible to distinguish their contribution from that of cellulose in the WAXS range.

Effects of hemicelluloses in the SAXS range

In order to quantify the effect of the hemicellulose coating on the SAXS intensities, a core-shell cylinder model (Fig. 2a, Eq. 4) was fitted to the equatorial SAXS intensities (example shown in Fig. 3a). The fitting results gave a cylinder core diameter

($2R$) of 2.6 nm for the neat 18-chain cellulose microfibril (Fig. 4a), which is similar to the typical SAXS-based values of around 2.5 nm reported for softwood samples (Jakob et al. 1995; Penttilä et al. 2019). The cylinder radius from the scattering fits can be compared with a similar parameter determined directly from the molecular models, obtained from fitting a step-function to the radial electron density profile (Fig. 2b). It is noteworthy that the optimization of the fit to the scattering intensity is done in reciprocal space, whereas fitting the electron density distribution takes place in real space. This can lead to systematic differences in the fitting results, as exemplified by the nearly 10% smaller (0.2 nm) cylinder core diameter $2R$ for the neat cellulose fibrils by the radial distribution fit (Fig. 4a). The fibril is only approximately cylindrical, and this deviation from a mathematically perfect circular cross-section is accounted for by the polydispersity
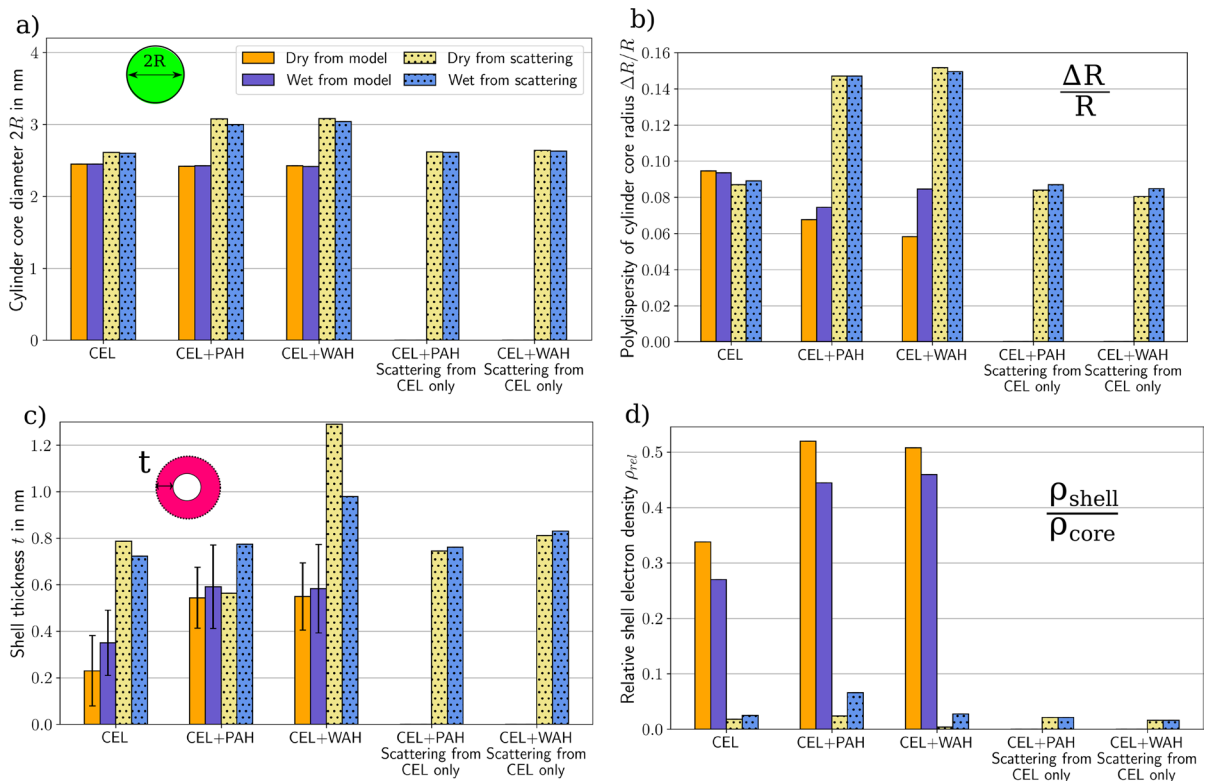
Fig. 4 SAXS-related parameters determined for fibrils of neat cellulose (CEL), fibrils with poorly aligned hemicelluloses (CEL+PAH) and fibrils with well aligned hemicelluloses (CEL+WAH): **a** Cylinder core diameter $2R$, **b** Polydispersity of cylinder core radius $\Delta R/R$, **c** Shell thickness t (uncertainty

Δt for the model shown as error bars), **d** Relative shell electron density ρ_{rel}. On the right-hand side of each plot: parameters determined from scattering intensities computed only from the cellulose chains of the CEL+PAH and CEL+WAH models

parameter $\Delta R/R$ (Fig. 4b). In the neat cellulose fibrils, the SAXS-based polydispersity $\Delta R/R$ was around 0.1, which is similar to the result of Jakob et al. (1995) for spruce wood by multiple methods but only about one-half of the SAXS fits of Penttilä et al. (2019).

Based on the SAXS fitting results, the hemicellulose coating increased the cylinder core diameter by roughly 20% (0.5 nm) (Fig. 4a). The hemicelluloses also increase the SAXS-based polydispersity of the radius ($\Delta R/R$) by roughly 70% (0.06) (Fig. 4b). This suggests that roughly half of the polydispersity of the fibril radius in hemicellulose-coated fibrils with the chosen geometry can be explained by the purely cellulosic microfibril itself deviating from a perfect cylinder, while half of it is due to the contribution of hemicelluloses. This total contribution brings the polydispersity parameter $\Delta R/R$ closer to the values around 0.22 reported by Penttilä et al. (2019). Unlike the scattering fits, the fits to the electron-density distribution do not exhibit any change in $2R$ due to the hemicellulose coating, and also the polydispersity $\Delta R/R$ remains similar. This difference between the two fitting-based approaches could be related to the fitting of the density distribution taking place in real space and the scattering fits in reciprocal space. It also leads to a difference in the way the two approaches detect the shell, which will be discussed in the following.

The hemicelluloses introduce a clear contribution of a shell to the scattering intensity, as is evident from the related fitting parameters (Fig. 4c, d). In the case of the neat cellulose fibril, the SAXS-based shell thickness in the wet state is roughly equal to that of fibrils with poorly aligned hemicelluloses (Fig. 4c), but the shell density is lower (Fig. 4d). The fibrils with well aligned hemicelluloses resulted in a thicker but sparser shell. There were clear differences between the shell parameters determined from the electron density profile and from the scattering fit. The density-profile approach produced shells with a much larger relative electron density (Fig. 4d) and a generally lower shell thickness (Fig. 4c), demonstrating again that the fitting methods are not fully equivalent. Nevertheless, judging from the examples shown in Fig. 2c, the results of both approaches can be considered reasonable, and none of them is clearly more accurate than the other.

Effects of hemicelluloses in the WAXS range

In the WAXS range of the scattering intensities, the lattice spacings d_{hkl} and crystal size L_{hkl} were obtained based on the location and width of the diffraction peaks, respectively. Similar parameters were determined directly from the molecular models, in which the hemicelluloses were not considered to contribute to L_{hkl}. The two approaches produced rather similar values for the d_{200} and d_{004} lattice spacings (Figs. 5a, S4), whereas the results for $d_{1\bar{1}0}$ and d_{110} deviated more (Figure S5). This might be due to the scattering from the hemicellulose shell occupying a region that partially overlaps with $1\bar{1}0$ and 110 diffraction peaks, making fitting to these peaks difficult.

The presence of hemicelluloses had minimal impact on the lattice spacing in most directions. Only a small but consistent increase in d_{200} can be seen in Fig. 5a, both based on scattering fits and a direct analysis of the molecular models. A more significant change can be seen in the width of the 200 diffraction peak. The peak becomes narrower with the inclusion of poorly aligned hemicelluloses, leading to an increase in the WAXS-based crystal size by 0.1 nm (Fig. 5b). This effect is even stronger when the hemicelluloses are well aligned on the crystal surface, which results in an increase of 0.2 nm in comparison to the neat cellulose microfibril. This shows that even though the hemicellulose coating does not produce any distinguishable intensity contribution in the WAXS regime, it can make a detectable contribution to the crystal size determined based on the broadening of the 200 diffraction peak of cellulose I_β. The hemicelluloses also seem to have an impact on the fitted locations of the $1\bar{1}0$ and 110 diffraction peaks, reducing the d-spacing in both cases (Figure S5). However, as the same parameters determined from the molecular models do not show any clear changes, and knowing the difficulties in fitting these WAXS peaks, the scattering-based result could be interpreted as an artifact of the fitting. This implies also that no conclusions can be made on the effects of hemicelluloses on the broadening of the $1\bar{1}0$ and 110 peaks. The WAXS fits show a minor decrease in d_{004} with the inclusion of hemicelluloses (Figure S4), which can be interpreted as a specific contribution of hemicelluloses in the scattering.

Figure 5c shows the distributions of the d_{200}-spacing, as determined directly from the molecular
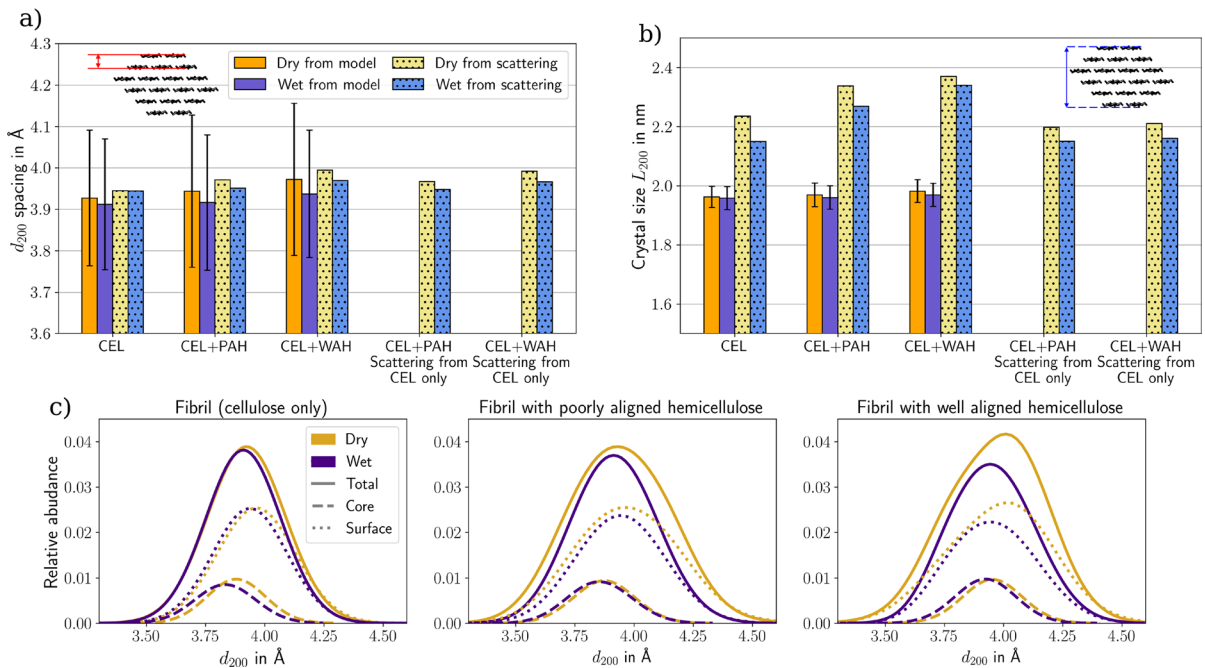
Fig. 5 Crystalline parameters determined for fibrils of neat cellulose (CEL), fibrils with poorly aligned hemicelluloses (CEL+PAH) and fibrils with well aligned hemicelluloses (CEL+WAH): **a** Lattice spacing d_{200} (standard deviation for the model shown in error bars, standard error negligibly small), **b** Crystal size L_{200}, **c** Distributions of the lattice spacing d_{200} as determined from the molecular models and separated into core and surface chains. On the right hand side of **a**, **b**: parameters determined from scattering intensities computed only from the cellulose chains in the CEL+PAH and CEL+WAH models

models and separated for cellulose chains on the microfibril surface and core (as indicated in Fig. 1a). The d_{200}-spacing of the surface chains is larger than that of the more tightly packed core chains, and the surface chains are especially affected by the hemicelluloses. In practice, the hemicellulose coating shifted the d_{200} distributions to larger values and broadened especially the distributions of the surface chains. The increase in d_{200} could be explained by the hemicellulose chains disturbing the organization of the cellulose surface chains, pulling them away from the core.

The degree of hemicellulose alignment on the fibril surface had a small effect on the results in the WAXS range. When the hemicelluloses were well aligned, d_{200} and the WAXS-based L_{200} were slightly larger than when they were poorly aligned. Therefore, the more tightly attached hemicellulose chains on the crystallite surface disturb the order of the surface chains more than a looser, less aligned

coating. Similarly, the larger L_{200} obtained for the fibril with well-aligned hemicelluloses results from the overall higher degree of order in this type of hemicellulose-coated fibril.

Effects of crystal size on scattering

The impact of crystal size on the scattering intensities was investigated by comparing scattering from neat cellulose fibrils of different cross-sectional size. The studied fibrils consisted of 14, 18, 23, 28, 34 and 40 cellulose chains in a hexagonal arrangement, as shown in Fig. 1a. A fibril with 18 chains was chosen as it was the basis of our model that included hemicelluloses. The minimum number of 14 was chosen because it is the smallest number of chains that could form a sensible hexagonal fibril that is not unreasonably small. The structures larger than 18 chains were chosen by stacking more layers in the direction

Fig. 6 Equatorial scattering intensities from wet, straight fibrils with a varying number of cellulose chains, plotted on fully logarithmic (**a**) and linear (**b**) axes and shifted vertically for clarity. Fits of the core-shell model (Eq. 4) in the SAXS regime (orange) and Gaussian peaks in the WAXS regime (red) are also shown
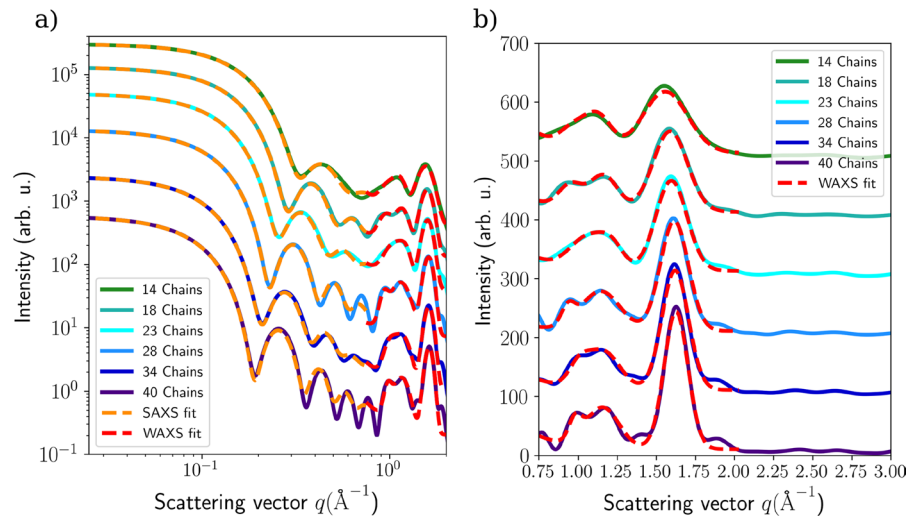
perpendicular to the (100) and (200) surfaces while maintaining a hexagonal shape.

The computed scattering intensities (Fig. 6) show the expected qualitative changes with crystal size. As the crystal cross-section becomes wider, the features in the equatorial SAXS intensities shift towards lower values of q (Fig. 6a) and the diffraction peaks in the equatorial WAXS intensities become sharper (Fig. 6b). The core-shell model reproduced the increase in the fibril diameter both based on the SAXS fits and the step function fit against the electron density distribution (Fig. 7a), although the values from the electron density distribution were slightly smaller like in the case of the 18-chain models before. Also the shell thickness in the SAXS fits increased with the crystal size (Fig. 7c), but this change was accompanied by a decreasing density of the shell (Fig. 7d). In connection with the sharpening of the diffraction peaks, the crystal size L_{200} obtained from the WAXS fits (Fig. 8b) increased by approximately 0.4 nm with the addition of each molecular layer into the fibril. At small crystal sizes (<30 chains), the crystal size determined directly from the molecular model was slightly smaller than the one obtained from the WAXS fits. At larger sizes (> 30 chains), however, this trend was reversed. This difference is probably related to the shape of the microfibril cross-section, as in the chosen model geometries (Fig. 1a) the top and bottom surfaces become relatively smaller with an increasing number of chains in the fibril. The crystal size perpendicular to the (1$\bar{1}$0) and (110)

lattice planes increased in the models as expected, whereas the WAXS fits yielded inconsistent results (Figure S6). Among the cross-section models studied in this work, the 18-chain model is in best agreement with a microfibril diameter of 2.5 nm observed for softwoods by SAXS (Jakob et al. 1995; Penttilä et al. 2019) and similar to atomic-force microscopy results from wood-based CNF (Daicho et al. 2018).

The lattice spacings in the cellulose crystals were affected by the crystal size, which is a well known result from diffraction experiments (Newman 2008; Huang et al. 2018). Similar size-dependent lattice expansion is known to take place in other types of nanoparticles (Diehm et al. 2012). The overall decreasing trend of d_{200} with crystal size (Fig. 8a) exhibits the effect of increasing cohesive interactions between an increasing number of chains. The d_{200} distributions in Fig. 8c clearly show that the chains in the fibril core are more tightly packed than at the surfaces. The ratio of core chains and surface chains also increases with the number of chains, which shifts the total distribution. At the same time, the d_{200} values determined from the WAXS peak locations and directly from the molecular model show a discrepancy that decreases with the crystal size (Fig. 8a). We interpret this to be caused by the violation of the common assumptions of diffraction theory by extremely small crystallites (Newman 2008). As a peculiar detail, the 18-chain case, which is thought to best represent the average microfibril in wood (Cosgrove 2022; Nixon et al. 2016; Yang and Kubicki 2020), has a noticeably lower d_{200}-spacing compared to the 23
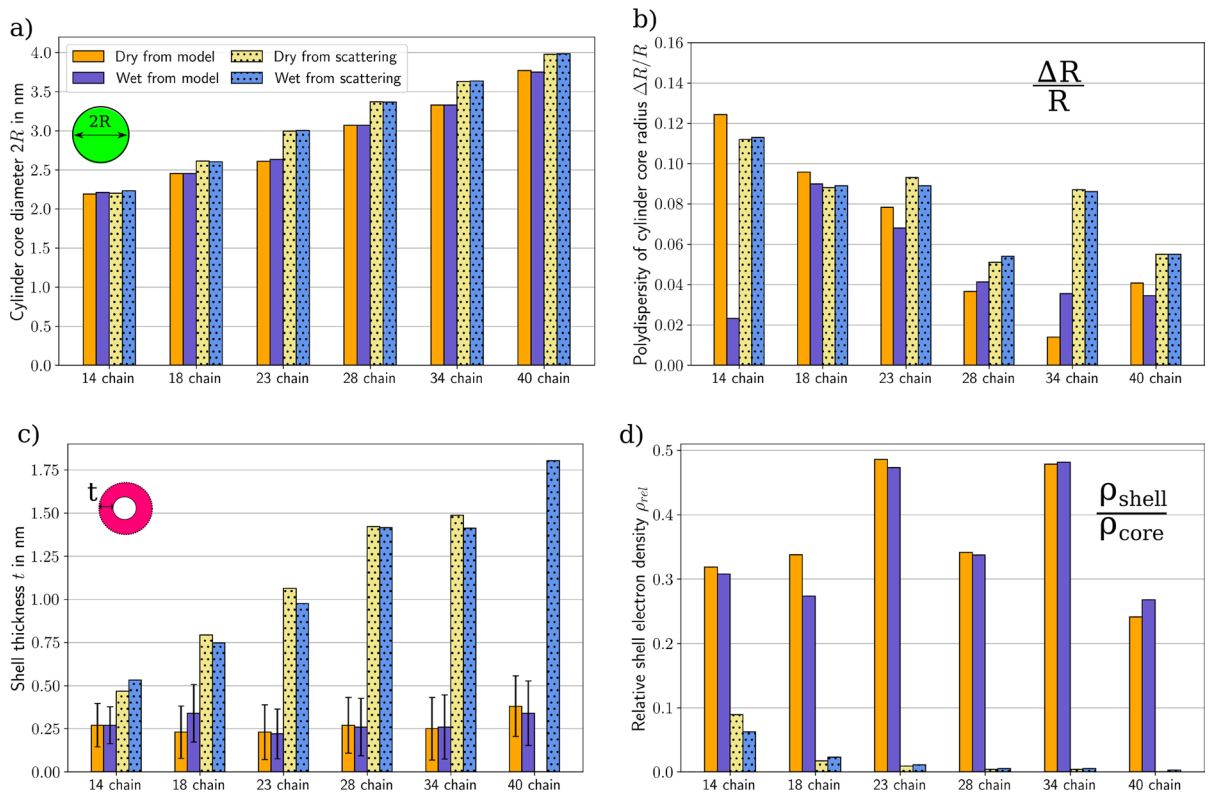
Fig. 7 SAXS-related parameters determined for fibrils with different number of cellulose chains: **a** Cylinder core diameter $2R$, **b** Polydispersity of cylinder core radius $\Delta R/R$, **c** shell thickness t (uncertainty Δt for the model shown as error bars), **d** Relative shell electron density ρ_{rel}

or 28 chain fibrils. However, this effect was observed only for the straight, non-twisting fibril, which indicates that it could be an artifact of the model or statistical variation between individual simulations. The d_{004}-spacing remained constant for a straight fibril (Figure S7) but increased slightly with the crystal size for a twisting fibril (as will be discussed later in "Effects of twisting" section), which is consistent with previous modelling results for a twisting fibril (Nishiyama et al. 2012).

Impact on determining the size of a microfibril

Accurate determination of the size of a cellulose microfibril cross-section is important for understanding the fundamental structure and biosynthesis of plant cell walls as well as for technological applications of cellulose in various forms. However, even its definition at the molecular level is not fully straightforward. In this work, the size of the microfibril was

determined in different ways from scattering intensities and molecular models. In the SAXS range, the equatorial scattering intensities and radial electron density profiles were fitted using a core-shell cylinder model. The thickness of the shell was always clearly smaller than the core radius, or its density was very small compared to that of the core. Therefore, we consider here the diameter of the cylinder core $2R$ representative of the fibril diameter based on SAXS and radial density profiles. In the WAXS region, the size of the microfibril cross-section can be defined as the mean of the crystal sizes determined in different directions, that is, the mean of $L_{1\bar{1}0}$, L_{110} and L_{200}. We determined these crystal sizes based on fits to the WAXS intensities by using the Scherrer equation (Eq. 7 with $K = 1$) and directly from the molecular models, as a distance between the centers of the outermost cellulose chains in each crystallographic direction. The fibril sizes determined for all model

a)

b)

c)
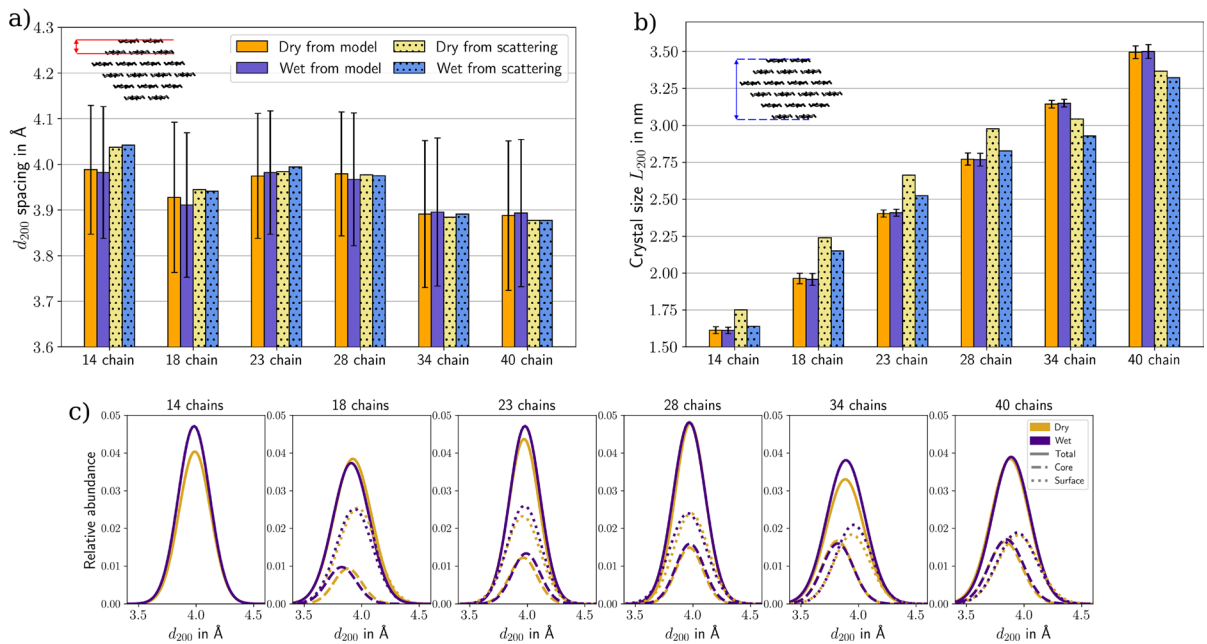
Fig. 8 Crystalline parameters determined for fibrils of varying size: **a** Lattice spacing d_{200} (standard deviation for the model shown in error bars, standard error negligibly small), **b** Crystal size L_{200}, **c** Distributions of the lattice spacing d_{200} as determined from the molecular models and separated into core and surface chains

structures in the above-listed ways are summarized Fig. 9.

All the methods that were used to determine the microfibril size show similar behavior with respect to the number of cellulose chains in the fibril (Fig. 9). The mean crystal size determined directly from the molecular model provides the smallest estimate for the fibril size, followed by determining the fibril diameter from the electron density distribution. The scattering-based mean crystal size and cylinder diameter provide roughly equal values, both of which are larger than with the methods based on analyzing the molecular model directly (Fig. 9). The fibril size based on the atomic coordinates (mean crystal size from model) can be considered the absolute minimum fibril size, while the value determined from the SAXS fit (cylinder diameter from scattering) gives the loose upper bound for the whole fibril cross-section. We note that using values of 0.7–0.8 for the constant K in the Scherrer equation (Eq. 7) would remove the difference in Fig. 9 between the mean crystal sizes determined from the WAXS peak fitting and directly from the model. However, as discussed before, fitting the $1\bar{1}0$ and 110 diffraction peaks was difficult, and in

practice it might often be possible to determine only the crystal size L_{200} reliably. Also the underestimating effect of determining the crystal size from the molecular backbone rather than from the outer dimensions of the molecules in our models (mentioned in "Direct
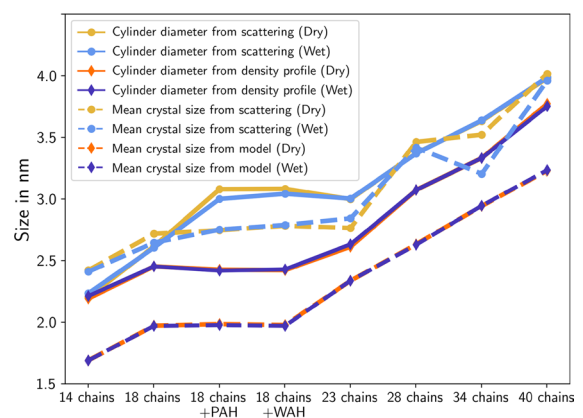
Fig. 9 Comparison of different methods of determining the size of a microfibril cross-section from the molecular model directly or from scattering intensities computed from it. The hemicellulose-containing models have been included between the 18 and 23 chain models

analysis of molecular models" section) is less significant in the [200] direction than in the [1$\bar{1}$0] and [110] directions. To remove the size difference for L_{200} (visible in Fig. 8b), values of $K = 0.9$–1.1 could be used. In particular, value $K = 0.9$ would give the best agreement for the 18-chain microfibril with the chosen cross-section geometry (Fig. 1a). The value of K would then roughly correct for both the shape of the fibril cross-section and the variation of the lattice spacings within a fibril (Figs. 5c, 8c), which also broadens the diffraction peaks.

Values of around 3 nm have been often determined for L_{200} based on fits to WAXS intensities from native woods (Fernandes et al. 2011; Jarvis 2018; Leppänen et al. 2009; Martínez-Sanz et al. 2015). This size is larger than expected for an 18-chain model of the cellulose microfibril (Jarvis 2018). Studies combining WAXS with SAXS have also pointed out that these values clearly exceed the fibril diameter determined by SAXS (Leppänen et al. 2009; Penttilä et al. 2019). However, these discrepancies can be removed by conducting the WAXS peak fitting in slightly different ways. For instance, subtracting the isotropic scattering contribution and fitting the 200 peak with two Gaussian functions in WAXS data from spruce wood yielded values consistent with a bimodal distribution of single and stacked fibrils of about 18 chains (Paajanen et al. 2022). Also, Daicho et al. (2018) observed for various cellulosic materials from pulp to CNF, that omitting any broad amorphous intensity contribution from fits to XRD intensities resulted in systematically smaller but still reasonable values for L_{200}. It is therefore possible that the details of the peak fitting, and in particular the presence and exact shape of the amorphous contribution, can explain why WAXS and XRD have typically yielded crystal sizes that imply larger crystals than those made of 18 cellulose chains. As a consequence, determining the exact number of chains in a representative cellulose microfibril should not depend too strictly on previously published crystal sizes obtained from diffraction peak broadening, especially if they are in contradiction with results from other methods.

Figure 9 also reveals something about the effect of hemicelluloses on the microfibril size. When hemicelluloses are included to coat the 18-chain fibril, a clear change in the fibril size is observed in the SAXS fits, so that the cylinder diameter of the 18-chain fibril becomes almost identical to that of the 23-chain fibril (Fig. 9). From the other parameters in Fig. 9, only the crystal size from WAXS shows a small increase with the introduction of hemicelluloses. Here the change of 0.1–0.2 nm in the crystal size perpendicular to the (200) planes (Fig. 5b) was partially compensated by a decrease in the other crystallographic directions, which were more difficult to fit. Thus, even though the hemicellulose coating broadens the diffraction peaks and increases especially the crystal size L_{200}, their effect is smaller than increasing the number of cellulose chains from 18 to 23 in the chosen model geometries.

Effects of twisting

In order to study the effects of twisting and how it would appear in the scattering data, we conducted the molecular simulations and all analyses for both periodic (straight) and non-periodic (twisting) fibrils. The twisting of the microfibril seems to have minimal to no impact on the parameters of the core-shell cylinder fits (Figures S8, S12) and the crystal size in any direction (Figures S9, S10, S11, S13, S14, S15). This was true for both the cases containing hemicelluloses and the cases with differently sized fibrils. However, there is a small yet consistent impact on the d-spacing of the (200) planes, where both the model and the WAXS fit show a slightly larger d_{200} for the straight fibrils (Figs. 5a, 8a) as compared to the twisting ones regardless of the presence of hemicelluloses (Figure S9a) or the fibril size (Figure S13a). A similar shift of the 200 peak due to twisting can be seen in the computed scattering intensities of Nishiyama et al. (2012) and Hadden et al. (2014). However, overall this change is minor and there is a lack of consistent twisting-related change in the crystal size L in any direction and the d-spacing of the (1$\bar{1}$0) and (110) lattice planes (Figures S5, S10). Considering this, distinguishing accurately between twisting and non-twisting fibrils based on cylindrically averaged experimental WAXS intensities is expected to be challenging. As mentioned previously ("Effects of crystal size on scattering" section) twisting fibrils were also the only ones to exhibit a small increase in d_{004} when the fibril size increased (Figure S15). The lack of significant differences observed in our

models of neat cellulose and those containing hemicellulose between the straight and twisting fibrils agrees with visual inspection of the models as well as recent electron diffraction results, showing that the crystal structure of cellulose is locally preserved within each cross-section of a twisting fibril (Ogawa 2019; Willhammar et al. 2021).

Effects of moisture

To address the effects of moisture changes on single microfibrils, we conducted the molecular simulations and all analyses for fibrils surrounded by water (wet) and in a vacuum environment (dry). Overall, the most significant effects of moisture can be seen in the results of the SAXS range describing the hemicellulose shell (Fig. 4). This is reasonable, because the presence of water swells the hemicelluloses especially in the poorly-aligned case, and allows them to be only loosely attached to the cellulose fibril surfaces (as shown on the right of Fig. 2c). In the vacuum, the hemicelluloses tightly adsorb on the cellulose fibril surfaces, regardless of their level of alignment. The swelling effect of the hemicelluloses is seen in the SAXS range in slightly different ways, depending on the method used to observe it. The SAXS core-shell fit shows a large increase in the electron density of the shell when going from the dry state to the wet (Fig. 4d), whereas the shell thickness either increases (CEL+PAH) or decreases (CEL+WAH) (Fig. 4c). The decrease of shell thickness in CEL+WAH is likely related to the shell being extremely sparse in the dry state. The fit to the radial electron density distribution shows an increase in the polydispersity of the cylinder core radius (Fig. 4b) and a decrease of the relative shell density (Fig. 4d) with swelling by water.

The differences in the crystalline structure between wet and dry fibrils are rather small in fibril models without hemicelluloses, since water is not able to penetrate into the crystallites. In systems with hemicelluloses, the hemicelluloses swell in response to water, which also affects the organization of the cellulose chains. Although the neat cellulose fibril shows a small increase in d_{200} from wet to dry state, a similar effect can be seen much stronger in the models including hemicelluloses (Fig. 5a). In more detail, the d_{200} distributions (Fig. 5c) show how in the neat fibril, both core and surface chains are affected by the

moisture change. Once hemicelluloses are included in the model, the core chains remain unaffected and the d_{200}-spacing of mainly the surface chains increases with drying. This effect is even stronger when the hemicelluloses are well aligned on the surface. This suggests that while water has minimal interaction with purely crystalline cellulose, it does interact with hemicelluloses which in turn interact with the surface cellulose chains. In the series of neat cellulose fibrils with different sizes, a small yet consistent increase of d_{200} with drying is seen only in the twisting case (Figure S13). The neat twisting fibrils also show a trend of an increasing d_{004} with drying (Figures S11, S15), whereas a minor decrease could be seen especially in the hemicellulose-coated straight fibrils (Figure S4). A decrease of d_{004} with drying is in agreement with experimental findings (Paajanen et al. 2022) and could be understood to be caused by a shrinking hemicellulose coating. Nevertheless, the overall moisture-related changes in the lattice parameters of the microfibrils seen in the current work were smaller (0.7% for d_{200}, 0.1% for d_{004}) than in similar systems with aggregated microfibrils (3% for d_{200}, 0.2% for d_{004}) (Paajanen et al. 2022). This supports the essential role of fibrillar aggregation to explain drying-induced shifts of diffraction peaks in cellulosic materials (Paajanen et al. 2022).

The presence of water around the fibril affects the degree of order of its constituent molecules. Among the parameters determined from the scattering intensities, this is most clearly illustrated by the changes in the crystal size L_{200} (Figs. 5b, 8b). Being determined from the diffraction peak broadening, this parameter is affected by changes in the lattice spacing distributions and the number of molecular layers in the crystal. Unlike in most experiments and our previous models of aggregated fibrils (Paajanen et al. 2022), the diffraction peaks from the individual fibrils in this work became narrower with drying, leading to an increase in L_{200}. The same effect can be seen regardless of the presence of hemicelluloses and with crystals of all sizes, although it is diminished in larger fibrils (Fig. 8b). It might therefore be related to the fraction of water-accessible surface chains, which decreases with the number of chains in the fibril. This increase in L_{200} by drying was clearly smaller in the current models of individual fibrils than the decrease observed in models of aggregated systems (Paajanen et al. 2022).

Conclusions

We analyzed scattering computed from microfibril models and examined the effects of hemicelluloses and crystal size on the fibril properties and the scattering. Analysis of SAXS and WAXS intensities provided structural parameters mostly consistent with those determined directly from the molecular models. The core-shell model used as the basis for SAXS analysis identified the hemicellulose contribution and changes in the number of chains in the crystal, although comparing the actual values with those determined directly from the model structure was not straightforward. The hemicelluloses contributed to the scattering intensities mostly in the q-region at the border of small-angle and wide-angle regions, which complicates the analysis of the $1\bar{1}0$ and 110 peaks of cellulose I_β in WAXS intensities. In the studied configurations, the hemicellulose coating narrows down the diffraction peaks from cellulose. The hemicellulose coating can therefore make a small contribution to the crystal size determined based on the broadening of the 200 diffraction peak of cellulose. Both the hemicellulose-coating and the size of the fibril cross-section influenced the crystal structure of cellulose, in which the surface chains were typically most affected. The effect of the fibril twist on scattering is minor and telling the difference based on cylindrically-averaged experimental scattering data would be exceedingly difficult. The effects of moisture on the crystal structure of the isolated fibrils were modified by the hemicellulose coating, but they were overall minor in comparison with those in previously studied aggregated systems. A comparison between different ways of determining the size of a microfibril cross-section from scattering data revealed systematic differences between the approaches, which should be taken into account when using the methods. Most of the results support the 18-chain model for a representative microfibril in wood.

The results of this work make the interpretation of experimental scattering data more accurate, enhancing the capability of scattering methods in structural characterization of cellulosic materials. This supports the use of these methods in providing better understanding of the nanostructure of plant cell walls and the development of various technological applications based on wood and lignocellulosic biomass.

Author Contributions A.P. and P.A.P. planned the study with contributions from A.Z. A.P. created the molecular models, conducted the simulations, computed the scattering data, created model visualizations, and wrote the corresponding experimental section. A.Z. conducted all the scattering data analysis, collected the results, and wrote the first full version of the manuscript. All authors participated in the interpretation of the results and writing the manuscript. A.P. and P.A.P. acquired funding and P.A.P. supervised the work of A.Z.

Availability of data The data is available from the authors upon reasonable request.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Code availability The code is available upon reasonable request.

Ethics approval No results of studies involving humans or animals are reported.

Consent to participate No results of studies involving humans or animals are reported.

Consent for publication No results of studies involving humans or animals are reported.

References

Abraham MJ, Murtola T, Schulz R, Páll S, Smith JC, Hess B, Lindahl E (2015) Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2:19–25. https://doi.org/10.1016/j.softx.2015.06.001

Balea Martin A, Blanco A, Delgado-Aguilar M, Monte M, Tarrés Q, Fuente E, Mutjé P, Negro C (2021) Nanocellulose characterization challenges. BioResources 16:4382–4410. https://doi.org/10.15376/biores.16.2.Balea

Berendsen HJC, Postma JPM, van Gunsteren WF, DiNola A, Haak JR (1984) Molecular dynamics with coupling to an external bath. J Chem Phys 81(8):3684–3690. https://doi.org/10.1063/1.448118

Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126(014101):1. https://doi.org/10.1063/1.2408420

Chen P, Li Y, Nishiyama Y, Pingali SV, O'Neill HM, Zhang Q, Berglund LA (2021) Small angle neutron scattering shows nanoscale pmma distribution in transparent wood biocomposites. Nano Lett 21(7):2883–2890. https://doi.org/10.1021/acs.nanolett.0c05038

Cosgrove DJ (2022) Building an extensible cell wall. Plant Physiol 189(3):1246–1277. https://doi.org/10.1093/plphys/kiac184

Cullity B, Stock S (2001) Elements of X-ray diffraction, 3rd edn. Prentice-Hall, Hoboken

Daicho K, Saito T, Fujisawa S, Isogai A (2018) The crystallinity of nanocellulose: dispersion-induced disordering of the grain boundary in biologically structured cellulose. ACS Appl Nano Mater 1(10):5774–5785. https://doi.org/10.1021/acsanm.8b01438

Danne R, Poojari C, Martinez-Seara H, Rissanen S, Lolicato F, Róg T, Vattulainen I (2017) doglycans-tools for preparing carbohydrate structures for atomistic simulations of glycoproteins, glycolipids, and carbohydrate polymers for gromacs. J Chem Inf Model 57(10):2401–2406. https://doi.org/10.1021/acs.jcim.7b00237

Diehm PM, Ágoston P, Albe K (2012) Size-dependent lattice expansion in nanoparticles: Reality or anomaly? Chem Phys Chem 13(10):2443–2454. https://doi.org/10.1002/cphc.201200257

Driemeier C, Bragatto J (2013) Crystallite width determines monolayer hydration across a wide spectrum of celluloses isolated from plants. J Phys Chem B 117(1):415–421. https://doi.org/10.1021/jp309948h

Elazzouzi-Hafraoui S, Nishiyama Y, Putaux JL, Heux L, Dubreuil F, Rochas C (2008) The shape and size distribution of crystalline nanoparticles prepared by acid hydrolysis of native cellulose. Biomacromology 9(1):57–65. https://doi.org/10.1021/bm700769p

Fang L, Catchmark J (2014) Structure characterization of native cellulose during dehydration and rehydration. Cellulose. https://doi.org/10.1007/s10570-014-0435-8

Fernandes AN, Thomas LH, Altaner CM, Callow P, Forsyth VT, Apperley DC, Kennedy CJ, Jarvis MC (2011) Nanostructure of cellulose microfibrils in spruce wood. Proc Natl Acad Sci USA 108(47):E1195–E1203. https://doi.org/10.1073/pnas.1108942108

Fernando D, Kowalczyk M, Guindos P, Auer M, Daniel G (2023) Electron tomography unravels new insights into fiber cell wall nanostructure; exploring 3D macromolecular biopolymeric nano-architecture of spruce fiber secondary walls. Sci Rep. https://doi.org/10.1038/s41598-023-29113-x

Fink HP, Hofmann D, Philipp B (1995) Some aspects of lateral chain order in cellulosics from x-ray scattering. Cellulose 2:51–70

Foster EJ, Moon RJ, Agarwal UP, Bortner MJ, Bras J, Camarero-Espinosa S, Chan KJ, Clift MJD, Cranston ED, Eichhorn SJ, Fox DM, Hamad WY, Heux L, Jean B, Korey M, Nieh W, Ong KJ, Reid MS, Renneckar S, Roberts R, Shatkin JA, Simonsen J, Stinson-Bagby K, Wanasekara N, Youngblood J (2018) Current characterization methods for cellulose nanomaterials. Chem Soc Rev 47(8):2609–2679. https://doi.org/10.1039/C6CS00895J

Hadden J, French A, Woods R (2014) Effect of microfibril twisting on theoretical powder diffraction patterns of cellulose Iβ. Cellulose 21:879–884. https://doi.org/10.1007/s10570-013-0051-z

Hashimoto T, Kawamura T, Harada M, Tanaka H (1994) Small-angle scattering from hexagonally packed cylindrical particles with paracrystalline distortion. Macromolecules 27(11):3063–3072. https://doi.org/10.1021/ma00089a025

Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: a linear constraint solver for molecular simulations. J Comput Chem 18(12):1463–1472

Huang S, Makarem M, Kiemle SN, Zheng Y, He X, Ye D, Gomez EW, Gomez ED, Cosgrove DJ, Kim SH (2018) Dehydration-induced physical strains of cellulose microfibrils in plant cell walls. Carbohydr Polym 197:337–348. https://doi.org/10.1016/j.carbpol.2018.05.091

Inouye H, Zhang Y, Yang L, Venugopalan N, Fischetti R, Gleber S, Vogt S, Fowle W, Makowski B, Tucker M, Ciesielski P, Donohoe B, Matthews J, Himmel M, Makowski L (2014) Multiscale deconstruction of molecular architecture in corn stover. Sci Rep 4:3756. https://doi.org/10.1038/srep03756

Jakob HF, Fengel D, Tschegg SE, Fratzl P (1995) The elementary cellulose fibril in picea abies: Comparison of transmission electron microscopy, small-angle x-ray scattering, and wide-angle x-ray scattering results. Macromolecules 28(26):8782–8787. https://doi.org/10.1021/ma00130a010

Jarvis MC (2018) Structure of native cellulose microfibrils, the starting point for nanocellulose manufacture. Philos Trans R Soc Lond Ser A 376(2112):20170045. https://doi.org/10.1098/rsta.2017.0045

Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935. https://doi.org/10.1063/1.445869

Kirschner KN, Yongye AB, Tschampel SM, González-Outeiriño J, Daniels CR, Foley BL, Woods RJ (2008) GLYCAM06: a generalizable biomolecular force field. Carbohydrates. J Comput Chem 29(4):622–655. https://doi.org/10.1002/jcc.20820

Kubicki J, Yang H, Sawada D, O'Neill H, Oehme D, Cosgrove D (2018) The shape of native plant cellulose microfibrils. Sci Rep. https://doi.org/10.1038/s41598-018-32211-w

Langan P, Petridis L, O'Neill HM, Pingali SV, Foston M, Nishiyama Y, Schulz R, Lindner B, Hanson BL, Harton S, Heller WT, Urban V, Evans BR, Gnanakaran S, Ragauskas AJ, Smith JC, Davison BH (2014) Common processes drive the thermochemical pretreatment of lignocellulosic biomass. Green Chem 16:63–68. https://doi.org/10.1039/C3GC41962B

Langford JI, Wilson AJC (1978) Scherrer after sixty years: a survey and some new results in the determination of crystallite size. J Appl Crystallogr 11(2):102–113. https://doi.org/10.1107/S0021889878012844

Leppänen K, Andersson S, Torkkeli M, Knaapila M, Kotelnikova N, Serimaa R (2009) Structure of cellulose and microcrystalline cellulose from various wood species, cotton and flax studied by x-ray scattering. Cellulose. https://doi.org/10.1007/s10570-009-9298-9

Li T, Senesi AJ, Lee B (2016) Small angle x-ray scattering for nanoparticle research. Chem Rev 116(18):11128–11180. https://doi.org/10.1021/acs.chemrev.5b00690

Lindner B, Petridis L, Langan P, Smith JC (2015) Determination of cellulose crystallinity from powder diffraction diagrams. Biopolymers 103(2):67–73. https://doi.org/10.1002/bip.22555

Mao Y, Liu K, Zhan C, Geng L, Chu B, Hsiao BS (2017) Characterization of nanocellulose using small-angle neutron, x-ray, and dynamic light scattering techniques. J Phys Chem B 121(6):1340–1351. https://doi.org/10.1021/acs.jpcb.6b11425

Martínez-Sanz M, Gidley MJ, Gilbert EP (2015) Application of x-ray and neutron small angle scattering techniques to study the hierarchical structure of plant cell walls: a review. Carbohydr Polym 125:120–134. https://doi.org/10.1016/j.carbpol.2015.02.010

McGibbon R, Beauchamp K, Harrigan M, Klein C, Swails J, Hernández C, Schwantes C, Wang LP, Lane T, Pande V (2015) Mdtraj: a modern open library for the analysis of molecular dynamics trajectories. Biophys J 109:1528–1532. https://doi.org/10.1016/j.bpj.2015.08.015

Mittemeijer EJ, Welzel U (2008) The "state of the art'' of the diffraction analysis of crystallite size and lattice strain. Z Kristallogr 223(9):552–560. https://doi.org/10.1524/zkri.2008.1213

Newman RH (2008) Simulation of x-ray diffractograms relevant to the purported polymorphs cellulose IV$_I$ and IV$_{II}$. Cellulose 15:769–778. https://doi.org/10.1007/s10570-008-9225-5

Newman RH, Hill SJ, Harris PJ (2013) Wide-angle X-ray scattering and solid-state nuclear magnetic resonance data combined to test models for cellulose microfibrils in mung bean cell walls. Plant Physiol 163(4):1558–1567. https://doi.org/10.1104/pp.113.228262

Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose I-beta from synchrotron X-ray and neutron fiber diffraction. J Am Chem Soc 124(31):9074–9082. https://doi.org/10.1021/ja0257319

Nishiyama Y, Johnson GP, French AD (2012) Diffraction from nonperiodic models of cellulose crystals. Cellulose 19:319–336. https://doi.org/10.1007/s10570-012-9652-1

Nixon B, Mansouri K, Singh A, Du J, Davis J, Lee JG, Slabaugh E, Vandavasi V, O'Neill H, Roberts E, Roberts A, Yingling Y, Haigler C (2016) Comparative structural and computational analysis supports eighteen cellulose synthases in the plant cellulose synthesis complex. Sci Rep 6:28696. https://doi.org/10.1038/srep28696

Ogawa Y (2019) Electron microdiffraction reveals the nanoscale twist geometry of cellulose nanocrystals. Nanoscale 11:21767–21774. https://doi.org/10.1039/C9NR06044H

Paajanen A, Zitting A, Rautkari L, Ketoja JA, Penttilä PA (2022) Nanoscale mechanism of moisture-induced swelling in wood microfibril bundles. Nano Lett 22(13):5143–5150. https://doi.org/10.1021/acs.nanolett.2c00822

Pedersen JS (1997) Analysis of small-angle scattering data from colloids and polymer solutions: modeling and least-squares fitting. Adv Colloid Interface Sci 70:171–210. https://doi.org/10.1016/S0001-8686(97)00312-6

Penttilä PA, Rautkari L, Österberg M, Schweins R (2019) Small-angle scattering model for efficient characterization of wood nanostructure and moisture behaviour. J Appl Crystsallogr 52(2):369–377. https://doi.org/10.1107/S1600576719002012

Penttilä PA, Paajanen A, Ketoja JA (2021) Combining scattering analysis and atomistic simulation of wood–water interactions. Carbohydr Polym 251:117064. https://doi.org/10.1016/j.carbpol.2020.117064

Rosén T, He H, Wang R, Zhan C, Chodankar S, Fall A, Aulin C, Larsson PT, Lindström T, Hsiao BS (2020) Cross-sections of nanocellulose from wood analyzed by quantized polydispersity of elementary microfibrils. ACS Nano 14(12):16743–16754. https://doi.org/10.1021/acsnano.0c04570

Scardi P, Leoni M, Delhez R (2004) Line broadening analysis using integral breadth methods: a critical review. J Appl Crystallogr 37(3):381–390. https://doi.org/10.1107/S0021889804004583

Schmitt J, Calabrese V, da Silva MA, Lindhoud S, Alfredsson V, Scott JL, Edler KJ (2018) Tempo-oxidised cellulose nanofibrils; probing the mechanisms of gelation via small angle x-ray scattering. Phys Chem Chem Phys 20:16012–16020. https://doi.org/10.1039/C8CP00355F

Su Y, Burger C, Hsiao B, Chu B (2014) Characterization of tempo-oxidized cellulose nanofibers in aqueous suspension by small-angle x-ray scattering. J Appl Crystallogr. https://doi.org/10.1107/S1600576714005020

Terrett O, Lyczakowski J, Yu L, Iuga D, Franks W, Brown S, Dupree R, Dupree P (2019) Molecular architecture of softwood revealed by solid-state nmr. Nat Commun. https://doi.org/10.1038/s41467-019-12979-9

Thomas LH, Forsyth VT, Šturcová A, Kennedy CJ, May RP, Altaner CM, Apperley DC, Wess TJ, Jarvis MC (2012) Structure of cellulose microfibrils in primary cell walls from collenchyma. Plant Physiol 161(1):465–476. https://doi.org/10.1104/pp.112.206359

Viell J, Székely NK, Mangiapia G, Hövelmann C, Marks C, Frielinghaus H (2020) In operando monitoring of wood

transformation during pretreatment with ionic liquids. Cellulose. https://doi.org/10.1007/s10570-020-03119-4

Willhammar T, Daicho K, Johnstone DN, Kobayashi K, Liu Y, Midgley PA, Bergström L, Saito T (2021) Local crystallinity in twisted cellulose nanofibers. ACS Nano 15(2):2730–2737. https://doi.org/10.1021/acsnano.0c08295

Yang H, Kubicki JD (2020) A density functional theory study on the shape of the primary cellulose microfibril in plants: effects of c6 exocyclic group conformation and h-bonding. Cellulose. https://doi.org/10.1007/s10570-020-02970-9

Zhang Y, Inouye H, Yang L, Himmel M, Tucker M, Makowski L (2015) Breakdown of hierarchical architecture in cellulose during dilute acid pretreatments. Cellulose. https://doi.org/10.1007/s10570-015-0592-4

Zhang Y, Inouye H, Crowley M, Yu L, Kaeli D, Makowski L (2016) Diffraction pattern simulation of cellulose fibrils using distributed and quantized pair distances. J Appl Crystallogr 49(6):2244–2248. https://doi.org/10.1107/S1600576716013297

Zitting A, Paajanen A, Penttilä PA, Rautkari L (2021) Deswelling of microfibril bundles in drying wood studied by small-angle neutron scattering and molecular dynamics. Cellulose 28(17):10765–10776. https://doi.org/10.1007/s10570-021-04204-y

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.