



Recrystallization and size distribution of dislocated segments in cellulose microfibrils—a molecular dynamics perspective

Ali Khodayari · Ulrich Hirn · Stefan Spirk · Aart W. Van Vuure · David Seveno

Received: 10 March 2021 / Accepted: 3 May 2021 / Published online: 27 May 2021
© The Author(s) 2021

Abstract The arrangement of cellulose molecules in natural environment on the nanoscale is still not fully resolved, with longitudinal disorder in cellulose microfibrils (CMF) being one relevant question. Particularly the length of the dislocated cellulose segments in CMFs is still under debate. Using molecular dynamics simulations, we are first investigating the phenomenon of pseudo-recrystallization of dislocated cellulose regions after cleavage of CMFs. Based on our simulations we propose that 3–4 glucose residues bordering to each side of a cellulose nanocrystal are actually reorganizing to a quasi-

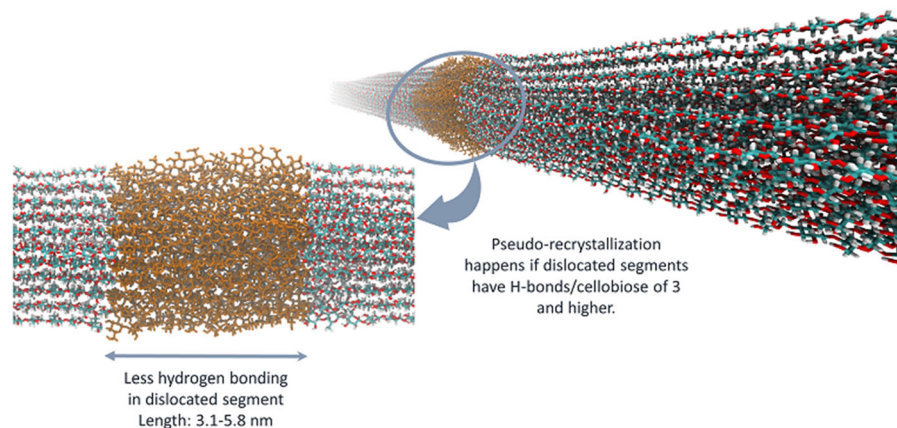
crystalline state, which are corroborating recent analytical investigations reporting an increase in crystallinity after acid vapor hydrolysis of CMFs. Combining our molecular dynamics simulation results with these analytical data we can estimate the length of the dislocated cellulose segments in CMFs. We propose that, for the investigated sources of biomass (cotton and ramie), the dislocation lengths are between 3.1–5.8 nm equaling to 6–11 glucose residues in the cellulose crystallites.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10570-021-03906-7>.

A. Khodayari (✉) · A. W. Van Vuure · D. Seveno
Department of Materials Engineering, KU Leuven,
Kasteelpark Arenberg 44, bus 2450, Leuven 3000,
Belgium
e-mail: ali.khodayari@kuleuven.be

U. Hirn · S. Spirk
Institute of Bioproducts and Paper Technology, Graz
University of Technology, Inffeldgasse 23, A-8010 Graz,
Austria

Graphic abstract



Keywords Recrystallization · Dislocated cellulose · Hydrogen bonding · Cellulose nanocrystal · Cellulose microfibril · Molecular dynamics

Introduction

Cellulose nanomaterials, often termed nanocellulose, are emerging materials and are integral parts of more sustainable and biodegradable lightweight composites with exceptional mechanical properties. Their preparation mainly relies on disassembly of macroscopic lignocellulosic biomass such as pulp and natural fibres and involves mechanical and/or chemical treatment to liberate the nanocellulose. In lignocellulosic biomass, the nanocellulose materializes in microfibrils (CMF), also occasionally termed as cellulose nanofibrils, which form the smallest structural unit in plant cell walls. They provide structural integrity and are capable to assemble into larger structures often called macrofibrils (Nishiyama 2009; Gibson 2012). In CMFs, the cellulose chains are aligned in a parallel fashion with respect to their reducing ends, resulting in an extremely high aspect ratio. The width of the CMFs is usually monodisperse for a specific species. For highly developed plants such as trees, CMF width is typically low and ranges from 2-5 nm, depending on the chosen analysis technique and sample preparation. While there is widespread agreement on the crystal structures of native cellulose ($I\beta/I\alpha$) (Nishiyama et al. 2002, 2003b), many details on CMF composition, shape and geometry are still under discussion. For

instance, it is still under debate how many chains build up a single CMF. The most prominent model (6×6 chains) (Endler and Persson 2011) has recently been challenged by new computational and experimental evidence and points at 24- and 18-chain models (Jarvis 2013; Oehme et al. 2015). Even longitudinal disorder in CMFs has been a major source for confusion, largely due to the adaption of the fringed-fibrillar models to cellulose. In these models, CMF consists of alternating highly ordered crystalline and disordered 'amorphous' domains, which feature different reactivity and accessibility (Mark 1940; Hearle 1958). Many properties of cellulose materials and particularly water uptake and chemical reactivity were explained and rationalized by this model. One of the first papers that challenged the fringed-fibrillar model was a SANS study on ramie fibres (Nishiyama et al. 2003a; Müller et al. 2000). The major result was that the 'amorphous' regions are very short (ca. 2 nm, 4–5 glucose residues) compared to the crystalline fraction with dimensions of tens to hundreds of nanometers. However, this means that these 'amorphous' domains are defects or dislocations in the cellulose crystals, rather than extended regions as in synthetic polymers where these can be even observed by optical microscopy. The fringed-fibrillar model was also used to explain acidic hydrolysis of CMFs as it predicts different reactivities for the 'amorphous' and crystalline domains, resulting in different kinetics with the LODP (levelling off degree of polymerization) being the end point. The LODP is also an important parameter in cellulose nano-crystal (CNC) preparation. However, the major difference between LODP

determination and CNC preparation is the employed acid strength. For sulfuric acid, there is only a narrow process window upon which CNCs are formed (i.e. 64–65% at 45°C (Borrega et al. 2018)). Yields in aqueous hydrolysis to produce CNCs are typically low which has been justified by the fringed-fibrillar model which implies losses due to removal and subsequent hydrolysis of 'amorphous' domains from the CMFs. However, these low yields are a direct consequence of the harsh reaction conditions and work-up conditions (centrifugation, dialysis) in aqueous medium, whilst the use of vaporous HCl in hydrolysis revealed nearly quantitative yields (Lorenz et al. 2017). Washing the resulting CNC solid with water (Pääkkönen et al. 2018) did not give significant amounts of glucose (<1%), which is in clear contrast to the fringed fibrillar model (Kontturi et al. 2016).

The formation of such dislocated regions can be rationalized by the accumulation of stress on the chains due to the twist of the fibrils. This can deteriorate the hydrogen bonds within the crystal regions, leading to dislocated segments. As a result, the term 'dislocation' better reflects the nature of such regions than 'amorphous cellulose'. In particular, the amount of two intra-chain hydrogen bonds (H-bonds), namely O3-HO3...O5 and O2-HO2...O6, and the more dominant inter-chain H-bond in cellulose I β , O6-HO6...O3, drop significantly in these dislocated regions (Khodayari et al. 2020b). The inherent twist of the fibrils is expected to be removed due to this dependency (Bu et al. 2015; Zhao et al. 2013; Matthews et al. 2006). Nonetheless, degree of twist in microfibrils is still a matter of debate due to lack of enough experimental observations. For instance, we have measured a constant degree of twist of $\sim 1.8^\circ/\text{nm}$ for the outer surface of crystalline cellulose by molecular dynamics simulations, whereas the values are shown to be different should an average over the cross-section be taken, the fibrils have different number of chains or different lengths (Hadden et al. 2013; Kannam et al. 2017). This value is also reported to alter under various circumstances, such as changes in humidity levels, throughout the length of the fibrils (Ogawa 2019). Despite all said, theoretical powder diffraction patterns have shown that twisting does not excessively disorder the unit cell of the cellulose I β (Hadden et al. 2014).

Several atomistic models for, or dislocated cellulose have already been published (Mazeau and Heux

2003; Mohammad et al. 2018; Zhang et al. 2013; Kulasinski et al. 2014, 2015a, b, 2017). In previous work, we modelled the dislocated regions within CMFs using molecular dynamics (MD) simulations, and characterized the models based on the conformational preferences and presence of intra- and inter-chain H-bonds (Khodayari et al. 2020b). We further showed that the inverse rule of mixtures can be employed to compute the effective elastic modulus of fibrils composed of crystalline and dislocated segments (Khodayari et al. 2020a).

In our work we first address re-organization of dislocated regions in CMFs after cleavage, as suggested by Kontturi et al. (2016) and Spiliopoulos et al. (2021), reflected by the increase in the crystallinity index (Kargazadeh et al. 2012). Using MD models we explore how H-bonds in short-length dislocated cellulose regions can hold the chains together through intra- and inter-chain bonding, keeping the properties of the segments comparable to those of the crystalline regions. It is then predicted that there is a direct relationship between the elastic modulus of dislocated segments and the hydrogen bonding in these regions. Our numerical results suggest that 3-4 GRs bordering to crystalline regions are actually returning to a pseudo-crystalline state after cleaving the CMF. Based on literature data of mass loss in acid vapor cleavage we then estimate the length of the dislocated regions in CMFs ranges from 3.1 to 5.8 nm, considering pseudo-recrystallization of such regions with shorter sizes. These results are providing a molecular dynamics based explanation for the experimentally observed increase of the crystallinity of cellulose after degradation by vaporous acid hydrolysis (Kontturi et al. 2016). Based on our result of 3-4 GRs recrystallizing upon cleavage, the size of the dislocated segments in CMFs from different source materials can be estimated as well, if the mass loss after hydrolysis to LODP is accurately determined.

Methods and models

Molecular dynamics simulations

Molecular dynamics simulations are performed using GROMACS 2018.7 version (van der Spoel et al. 2005). GLYCAM06 force field parameters are used (Kirschner et al. 2008), and the conversion from

Amber (Case et al. 2018) topology to GROMACS is performed through the Acypye python script (Sousa Da Silva and Vranken 2012). Materials studio is used to generate the initial conformations (BIOVIA 2017). Minimization is performed using the steepest decent algorithm. Integration of Newton's equation of motion is carried out by the leap-frog algorithm in all steps of the simulations. A step size of 2 fs is used for majority of the simulations. Bonded hydrogens are constrained through the LINear Constraint Solver (LINCS) algorithm. Cut-off of neighbour searching for short-range electrostatics and van der Waals interactions are set to 1.2 nm. Particle mesh Ewald (PME) is used to treat the electrostatics. Velocity rescaling thermostat is used for the temperature coupling. All models are positioned in the middle of a box of $23 \times 30 \times 100 \text{ nm}^3$.

Structural configuration of the $I\beta$ crystalline cellulose was taken from the crystallographic data reported by Nishiyama et al. (2002).

A geometric criterion is considered to distinguish the hydrogen bonds, being the donor-acceptor distance lower than 3.5 Å and hydrogen-donor-acceptor angle of less or equal to 30° (Jeffrey 1997).

The molecular dynamics simulations to measure the elastic modulus of the fibrils were performed according to the methodology described in previous work (Khodayari et al. 2020b). In a first approach, models are simulated in vacuum. It is already shown that presence or absence of water has minor effects on the elasticity of CNCs, while not generally on that of the CMFs. Crystalline cellulose is impenetrable to water, which hinders any effects on the intra- and inter-chain H-bond patterns controlling the chiral properties such as the twist of fibrils of 6×6 arrangements (Bu et al. 2015). Water should interact more strongly with dislocated segments but as we are most interested in the transition region between crystalline and dislocated cellulose. As discussed by French (2014) GLYCAM parameters might lead to an overestimation of values in vacuum as they have been adjusted for water models. This could possibly then lead to an overestimation of the recrystallization effect. In a previous work (Khodayari et al. 2020b), we have demonstrated that the hydrogen bonding pattern, as well as the conformational preferences of the models remain intact, when modelled in water, compared to current work in vacuum. This simplification therefore seems justified.

Models

Crystalline/disordered models

A $6 \times 6 I\beta$ cellulose crystal with exposed surfaces formed from the $1\bar{1}0$ and 110 planes was modelled, having a DP of 119, showing an initial length of 61.25 nm before equilibration, and 63.52 nm of axial length when energy minimized and equilibrated using the GLYCAM06 parameter sets. This minor change in the length of the model is expected as minimization and equilibration is shown to have a shrinking effect on a - and an expanding effect on c -unit cell dimensions (Nishiyama et al. 2012). Following an established methodology to disrupt the perfect arrangement of the chains (Khodayari et al. 2020a, b), the generation of the dislocated segments in the middle of the fibril was performed by step-wise tempering the segment to temperatures above its melting temperature, being 750, 900, and a maximum temperature of between 1000 and 1050 K. The last step was followed by quenching down the temperature to 300 K, which concluded with 4 ns of simulation (1 ns for each tempering step). Note that, the annealing and quenching steps are used as tricks to disrupt the hydrogen bonding pattern in specific segments of the fibril, and no measurements are performed at high temperatures, to avoid unstable and highly kinetic structures. Otherwise stated, the tempering cycles are to generate enough thermal vibration to the systems and to artificially disrupt the crystallinity of the segments (Morthomas et al. 2019). Structures analysed at room temperature are hence equilibrated and stable. Moreover, no disruption of the pyranose conformations either in the crystalline or the dislocated segments was observed. 11 models including different lengths of dislocated segments of 2.6 - 18.5 nm were modelled using the above methodology, having lengths of 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, and 35 glucose residues (GRs) in the middle of the fully crystalline model. The models were then named by adding the letter r before the number of GRs being dislocated. Figure 1 shows an example of the model including a 10-glucose residue dislocated segment in the middle of the fibril (r10).

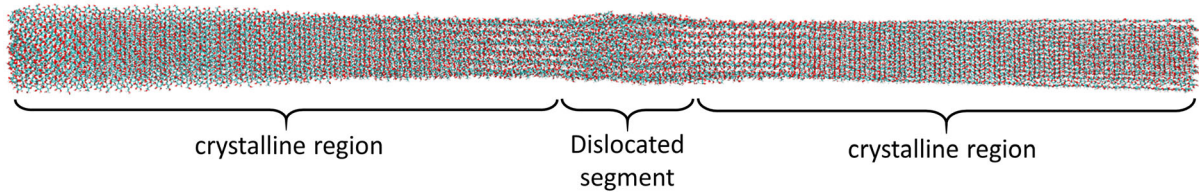


Fig. 1 The model including 10 glucose residue of dislocated segment in the middle (r10)

Cleaved models

To generate the cleaved models from complete CMFs, the right hand side crystalline region in Fig. 1 is removed manually and O4 oxygen on the ending residues are capped by hydrogen. Reorganization of the models is then studied by equilibrating the structures for 90 ns at room temperature. A step size of 3 fs is used for this set of simulations.

Models to study temperature dependency

To study temperature dependency of hydrogen bonding, nine models (from r5 to r25) were selected to investigate deterioration of the hydrogen bonding pattern in the dislocated segment by temperature. Models were all tempered to 750 K, and subsequently to a maximum temperature of 900, 1000, 1020, 1030, 1040, and 1050 K (1 ns equilibration time for each tempering step). Models were quenched down after reaching each maximum temperature, leading to generation of 54 models.

Results and discussion

H-bond and the elastic modulus

Using 1020 K as the maximum tempering temperature, dislocated regions were generated for all r models. The extent of disorder can be rationalized by the number of H-bonds per cellobiose. Figure 2a shows the number of H-bonds per cellobiose in the dislocated region of all the models. Accordingly, the value drops sensibly when the r10 model is considered, and it remains almost constant by increasing the length of the dislocated region in the r15, r20, r25, r30, and r35 models. This indicates a tipping point, at which H-bonds are irreversibly lost. The elastic moduli of the fibrils including dislocated segments

shorter and equal to 10 GRs are then measured and represented in Fig. 2b. The first four models show very similar values to those of the fully crystalline model (152 GPa for 3.7% of strain). The modulus slightly drops in the r9 model, and reaches its lowest value in the r10 model. As the tensile modulus of the fibrils depends on the amount of hydrogen bonds (Santiago Cintrón et al. 2011), the moduli of the dislocated regions larger than 10 GRs would subsequently be almost equal to 76 GPa, as measured for the r10 model (Khodayari et al. 2020b). Comparing the two plots, it can be concluded that the elastic modulus follows a similar trend with the change in the H-bonds. Hence, H-bond analysis can be used as a strong indication for distinguishing the crystalline parts from the dislocated regimes.

The elastic modulus of the dislocated segments can be computed using the inverse rule of mixtures (IROM), and the volume fractions of the segments (Khodayari et al. 2020a). The IROM to calculate the modulus of the dislocated segments (shorter than r10) can be expressed as $\frac{1}{E_{eff}} = \frac{v_c}{E_c} + \frac{v_d}{E_d}$, where E_{eff} is the effective elastic modulus of the fibril, v_c and v_d are the volume fractions of the crystalline and dislocated segments, and E_c and E_d are the elastic moduli of the crystalline and dislocated segments, respectively. It is worth noting that the elastic modulus of the r5, r6, r7, r8, and r9 dislocated regions are slightly lower than those of the fully crystalline model as evidenced in Fig.S1 from the supplementary information. It drops by 10–20% in the first five dislocated models, while it reaches a lowest value of one half the fully crystalline model in the r10 model, being 76 GPa. This trend also completely follows the pattern of the normalized hydrogen bonding number.

Transition region

The number of hydrogen bonds per cellobiose in the 54 tempered-quenched regions (explained in “Models

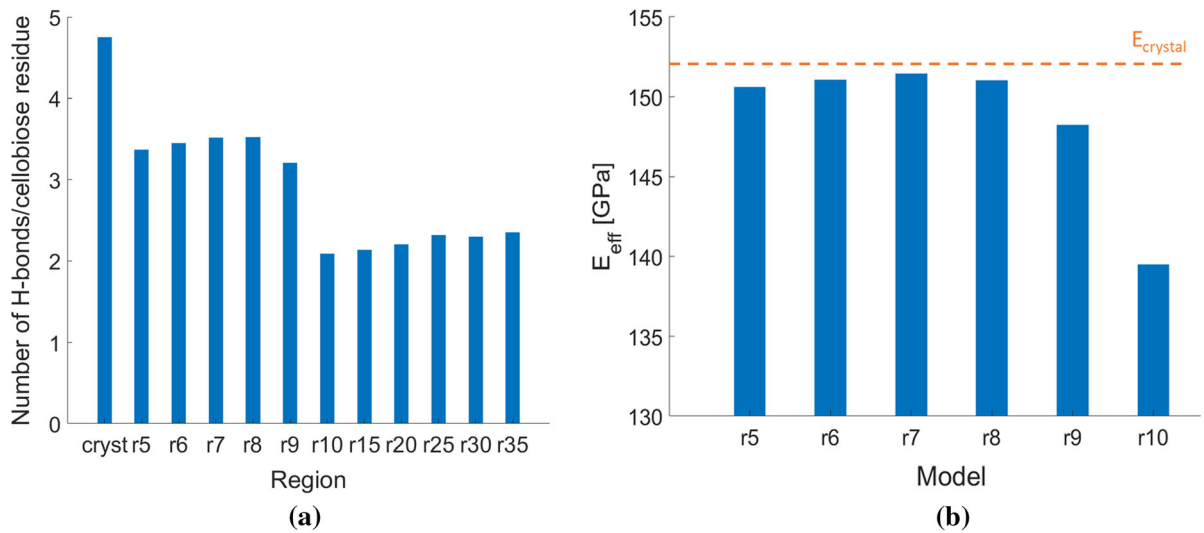


Fig. 2 **a** Number of hydrogen bonds per cellobiose in the dislocated region of all the r models and in the fully crystalline model, and **b** the elastic modulus of the models including dislocated segments up to 10 GRs

to study temperature dependency”) is calculated and depicted in Fig. 3. As can be seen, trends from different models seem to start diverging when temperature rises. Increasing the temperature to values over 1020 K does not necessarily decrease the hydrogen bonds in the r5 and r6 regions, as well as for models longer than r10. In other words, the most disturbed structures for all the models equal to and longer than 10 GRs can be generated by reaching a temperature of 1020 K. Increasing the maximum temperature to values over 1050 either generates very high kinetic energies, leading to detachment of the

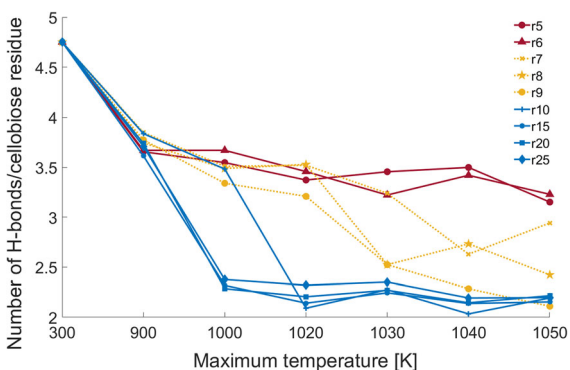


Fig. 3 Temperature dependence of the hydrogen bonding number per cellobiose. Most models reach a minimum number of hydrogen bonds per cellobiose at 1020 K and higher. r7, r8, and r9 models behave as transition regions. Error bars are not shown for the sake of clarity. A standard deviation of ± 0.25 can be considered

chains in models with dislocated segments longer than 7 GRs, or only marginally decreases the H-bonds in r5 and r6 models.

According to Fig. 3, three regions can be defined. The stable region, at which H-bonds intensely resist breaking, including the r5 and r6 models. On the other hand, heat-treated segments larger than 10 GRs hold the minimum number of H-bonds, at temperatures 1000 K and above, having the least resistance against keeping H-bonds. A transition section can also be observed (dashed lines), where the persistence in keeping the H-bonds decreases as the imposed kinetic energy increases. At lower kinetic energies, r7, r8, and r9 models keep a high number of H-bonds. As the temperature rises, the models tend to lose their H-bonds gradually. This can be possibly the region, where dislocated segments can reform, or recrystallize back to crystalline cellulose, depending on the number of hydrogen bonds. Stated differently, the transition regime is a section where chains form a crystal at lower temperatures, while they gradually transform into dislocated segments when temperature rises. Taking the example at 1020 K, it is shown in Fig. 2b that the mechanical properties can be very close to that of a crystalline cellulose for the dashed models in Fig. 3.

Dependency of chain reorganization on the H-bonds

To check if reorganization occurs or not, 14 structures were selected from the ones in Fig. 3. The choice was made based on different lengths and different H-bonds/cellobiose ratios. Models were cut to mimic cleavage after acid hydrolysis in experiments. The models are cleaved perfectly, i.e. into a single crystalline segment on the left and a corresponding dislocated section on the right (removing the right-hand side crystalline segment) as shown in Fig. 4. Each model is then equilibrated at 300 K for 90 ns. Particularly, models were chosen to cover a variety of initial H-bonds/cellobiose ratios (blue bars in Fig. 5). Figure 5 shows the change in the H-bonds/cellobiose ratio in the dislocated region before and after cleavage. As can be seen, the gain in the H-bonds/cellobiose ratio is more pronounced in the first four models than for the rest. Note that, models having lower number of H-bonds/cellobiose ratios before cleavage such as the first three models in the red background zone, also show a growth in the H-bonds, but they did not reach values in the range 3.5–4. This implies that the mechanical properties of these models, despite gaining some H-bond, cannot be as high as for the crystalline segments. However, models which already have H-bonds/cellobiose higher than 3.2 ± 0.2 , i.e. those in the green background zone, experience higher gain in H-bonds. As also shown in Fig. 2, a model having 4 H-bonds/cellobiose can show very close values of elasticity to that of a crystal. The results show that reorganization of the simulated models is rather an H-bond dependent phenomenon than being length dependent. The argument can be better explained in Fig. 6. This figure includes four models which are trimmed out of the previously cleaved dislocated segments in Fig. 5. For example, 6_8 denotes a dislocated segment with 6 GRs length, cut out of a

dislocated model with a length of 8 GRs. This is mimicking a situation when hydrolysis has caused cleavage of a 8 GRs length dislocated segment into two parts of 6 and 2 GRs. Interestingly, in spite of all models having the same length, the ones which originally have had H-bonds/cellobiose ratio of 2.5 or lower did not gain any H-bonds. This confirms the H-bond dependency of the phenomenon.

It must be noted that, despite H-bond dependency of the reorganization, the assumption of having dislocated segments with lengths in the transition region (dashed lines in Fig. 3) is still a valid argument. This is because models longer than 10 GRs are more prone to lose hydrogen bonds, in contrast to models shorter than 6 GRs. It can hence be concluded that dislocated segments are intrinsically expected to have lengths shorter than 10 GRs, but not less than 6 GRs. This is further investigated when H-bond distribution in the dislocated regions is analysed.

To check whether reorganization of the models occurs at all chains or at the interior chains of the fibril, the number of H-bonds/cellobiose in the 16 interior chains are also measured and compared with the values before reorganization. In particular, two sets of data were prepared: the gain in H-bonds/cellobiose for the exterior chains of the dislocated segments including 20 chains, which is the difference of the initial H-bonds/cellobiose and its value after cleavage, and the gain of H-bonds/cellobiose of the 16 core chains. To compare the two sets, values are normalized by the considered number of chains, i.e. the first set of data is divided by 20, while the second set is divided by 16. The reorganization intensity (RI) can be defined as:

$$RI = \frac{(Hbonds/cellobiose)_0 - (Hbonds/cell)_t}{chain\ number} \quad (1)$$

where 0 refers to the state before cleavage, t to the state after reorganization, and chain number is the number of chains in the considered section of analysis.

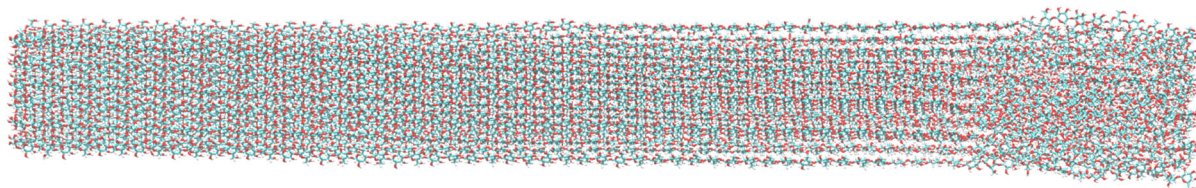


Fig. 4 Cleaved model to study chain reorganization is simulated by removing the right-hand side crystalline segment, resulting in a crystalline segment on the left and a dislocated segment connected to it on the right

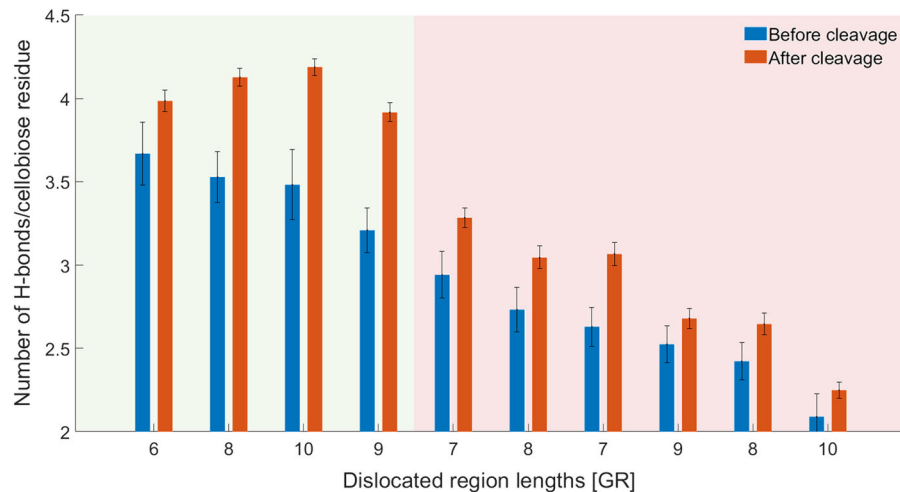


Fig. 5 The change in the H-bonds/cellobiose in the dislocated segments before and after cleavage by removing the right hand side crystalline region in Fig. 1. Horizontal axis values show the lengths of the dislocated segments in original H-

bonds/cellobiose decreasing order. Models having H-bonds/cellobiose higher than 3.2 ± 0.2 show a higher gain in H-bond after equilibration. Repeated numbers refer to different points on Fig. 3. Error bars show the standard deviation of the mean

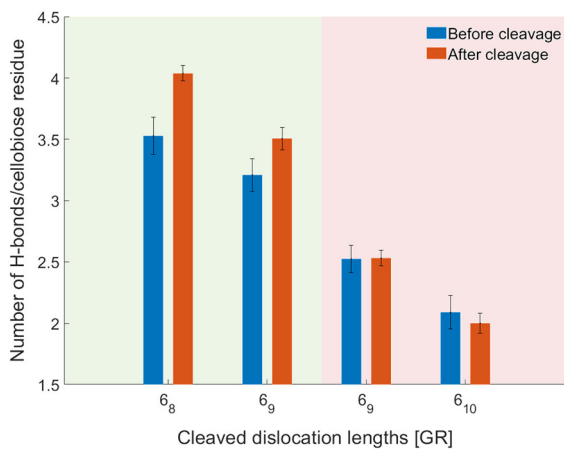


Fig. 6 The change in the H-bonds/cellobiose for models where cleavage of the dislocated region has occurred inside the region rather than at the end of it. Reorganization, regardless of the length of the models, is a H-bond dependent phenomenon. Error bars show the standard deviation of the mean

Figure 7 illustrates the reorganization intensity for the interior and exterior chains of the reorganized models in Figs. 5 and 6. Results show that for the crystallized models, reorganization mostly happens at the core, rather than the exterior of the fibrils. It however must be noted that, the results brought here concern individually isolated fibrils, whereas CNCs and in general CMFs in the plant cell wall aggregate to build super clusters connected via other amorphous

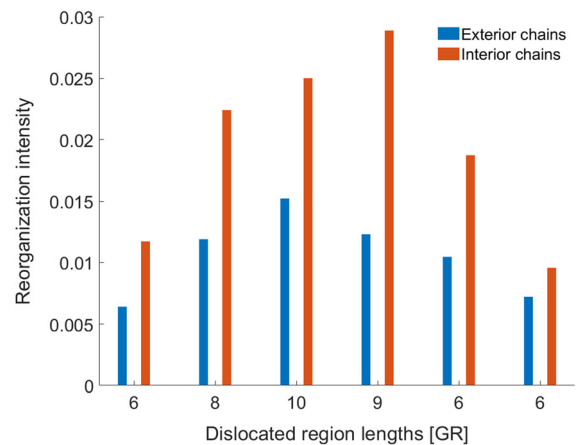


Fig. 7 Intensity of reorganization for different lengths of cleaved dislocated segments. Majority of the models show reorganization at the core, twice that of the exterior chains

media, e.g. hemicellulose and lignin, which alter the surface characteristics of the individual fibrils. These results are hence not relevant when elementary or technical fibres are considered.

Reorganization or recrystallization?

The idea of recrystallization of the dislocated segments of CNFs has been proposed as a result of occasional observations at the macroscale. Studies have already been performed to evaluate the structural

reorganization of cellulose crystals (Miyamoto et al. 2009). Our simulations tried to elucidate whether a full recrystallization of the disordered segments can happen if enough freedom is given to the structure to reform, e.g. due to a cleavage. Nonetheless, the analysis on the change of the H-bonds reveals that full retrieval of the lost H-bonds does not occur, at least within the 90 ns time frame given to the simulations. On the other hand, H-bonds reform to the extent that the so called recrystallized models show mechanical properties close to those of a fully crystalline segment. To further elaborate on this, another set of analyses was done to check conformational changes of the main dihedral angles about the glycosidic linkage. Crystallinity of CNCs is shown to be mostly controlled via three dihedral angles $\phi = O5 - C1 - O4 - C4$, $\psi = C1 - O4 - C4 - C5$, $\omega = O5 - C5 - C6 - O6$, and the glycosidic angle $\theta = C1 - O4 - C4$, discussed thoroughly in the literature (Heiner et al. 1995; Nishiyama et al. 2002, 2003b). These parameters have been demonstrated to show persistent ranges of values for different cellulose crystals (French et al. 2021; Mazeau and Heux 2003; Chen et al. 2012). The reported ranges for these torsional angles are $\phi : 262^\circ - 272^\circ$, $\psi : 213^\circ - 221^\circ$, and $\omega : 158^\circ - 170^\circ$.

Here, we maintain our focus on ϕ and ψ values (see Fig. 8a), as they mainly affect the intra-chain H-bonds controlling specific behaviour of CNCs, such as the leverage effect in stretched cellulose (Djahedi et al. 2016). Note that, synergistic changes of these two torsional angles have shown to control the elasticity of cellulose fibrils (Khodayari et al. 2020b). The crystalline models in our study show canonical values of $268^\circ (-91.9^\circ \pm 0.3)$ and $218^\circ (-142.7^\circ \pm 0.3)$ for ϕ and ψ , respectively (see Fig. 8b).

Distribution of ϕ and ψ for the first four models in Fig. 5 are illustrated before and after cleavage (Fig. 8c). According to the results, there are slight deterioration in the mean values and standard deviation (SD) of these two torsional angles before cleavage, compared to those of the crystalline segments. In particular, reorientation of the chains after cleavage leads to minor shifts in the canonical values of the dihedral angle preferences and/or increase of the population of angles around the canonical values, reflected as a decrease in the standard deviation of the mean values. Here, it must be noted that, even small

changes in SD imposes noticeable variations in the distribution of the torsional angles, as for instance the deviation in the dihedrals of the crystalline segments is only about 0.3. To elaborate further, the mean values and respective SDs of ϕ and ψ before and after cleavage are brought in Table.1. As can be seen, the mean values fluctuate more or less around the canonical values reported for the crystalline models. Generally, relaxing the models after cleavage might lead to one of the following two conditions. The mean of the distributions changes to values closer to those of the CNCs, or the standard deviation of the mean decreases, while the mean does not change noticeably. As an example, standard deviation of r6, r10, and r9 all decrease for both ϕ and ψ , meaning that more torsional angles are distributed about the canonical values, due to reorganization. On the other hand, r8 model does not show a decrease of SD, however, the mean values of ϕ and ψ get closer to those of the crystals. It must also be added that, the longer the model is, the easier it would be to retrieve the lost H-bonds. In other words, not only is it difficult to generate dislocated segments shorter than 6 GRs, but it is also not possible for the chains to reorient and form back to a crystalline formation. Interestingly, The last 6 models in Fig. 5 show no or marginal change in the ϕ and ψ percentages, and are still far from those of a crystalline model (refer to section S.2 of the supplementary information for more). In a study by French et al. (2021), authors presented the distribution of ϕ and ψ values for different celooligosaccharide conformations and derived the energy surfaces obtained by DFT quantum mechanics method for the experimental observation of these values at the β -(1 \rightarrow 4) linkages. Interestingly, the reorganized models here exhibit an agreement of dihedral preferential values with those reported by French et al. (2021) (Fig.S.3).

To increase the reliability of the conclusions, further analyses are performed and theoretical X-ray powder diffraction (XRD) patterns are calculated by Debyer Wojdyr (2012) for the dislocated and reorganized structures (6, 8, 10, and 9 GR) and compared to those of the crystalline segments (Fig. 9). Crystalline segments (gray patterns in Fig. 9) show three peaks at 14.55° , 23.35° , and 33.75° , corresponding to the $1\bar{1}0$, 200, and 004 planes. The patterns are in good agreement with those reported for the twisted models by Hadden et al. (2014), even though models of 81

Fig. 8 **a** Illustration of the dihedral angles ϕ and ψ in a section from a cellulose chain. **b** Distribution of the dihedral angles ϕ and ψ for CNC. **c** The blue scatter plots show the distribution of the dihedral angles within the first four dislocated models in Fig. 5 (6, 8, 10, and 9 GR), whereas the right orange scatter plots depict the same distribution for the reorganized models. The canonical values for the crystalline segments is measured to be $268^\circ (-91.9^\circ \pm 0.3)$ and $218^\circ (-142.7^\circ \pm 0.3)$ for ϕ and ψ , respectively

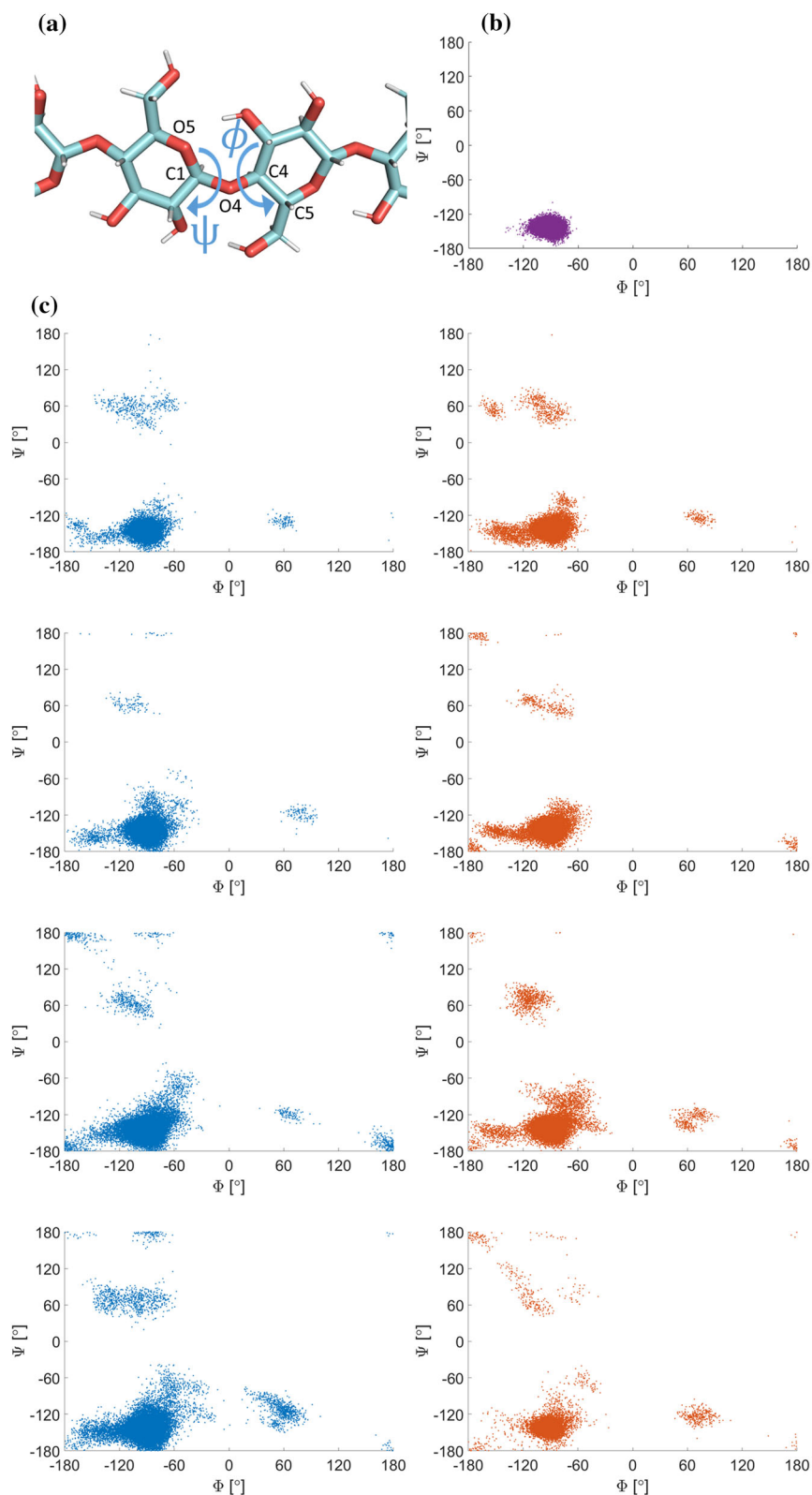


Table 1 The change in dihedral angle values and respected standard deviation (SD) due to reorganization of the models shown in Fig. 8c

Model	Before cleavage		After cleavage (reorganized)	
	Mean [°]	SD	Mean [°]	SD
r6(Ψ)	-139.8	0.8	-138	0.5
r6(Φ)	-92	0.7	-92.1	0.5
r8(Ψ)	-143.9	0.8	-140.9	0.9
r8(Φ)	-90.9	0.5	-92.1	1.1
r10(Ψ)	-135.3	1.7	-134	1.1
r10(Φ)	-88.7	0.8	-89.2	0.7
r9(Ψ)	-139.9	1.4	-139	0.6
r9(Φ)	-89.6	1	-90.8	0.6

total chains were used in that study. The same agreement can be seen about the minor low-angle scattering in the 10° - 13° range of 2θ . This low-angle scattering is seen for the twisted models, as well as linearly constrained models by Hadden et al. (2014). Nishiyama et al. (2012) argued that the low-angle observations diminish when larger number of chains are considered alongside water solvation shells around the models. However, for the sake of comparison between the three types of structures here, their existence is not of major concern. Our diffraction patterns for the crystalline cellulose also fit well with those observed for the 6×6 models in the study of Nishiyama et al. (2012). As can be seen in Fig. 9 dislocated models exhibit larger low-angle scattering,

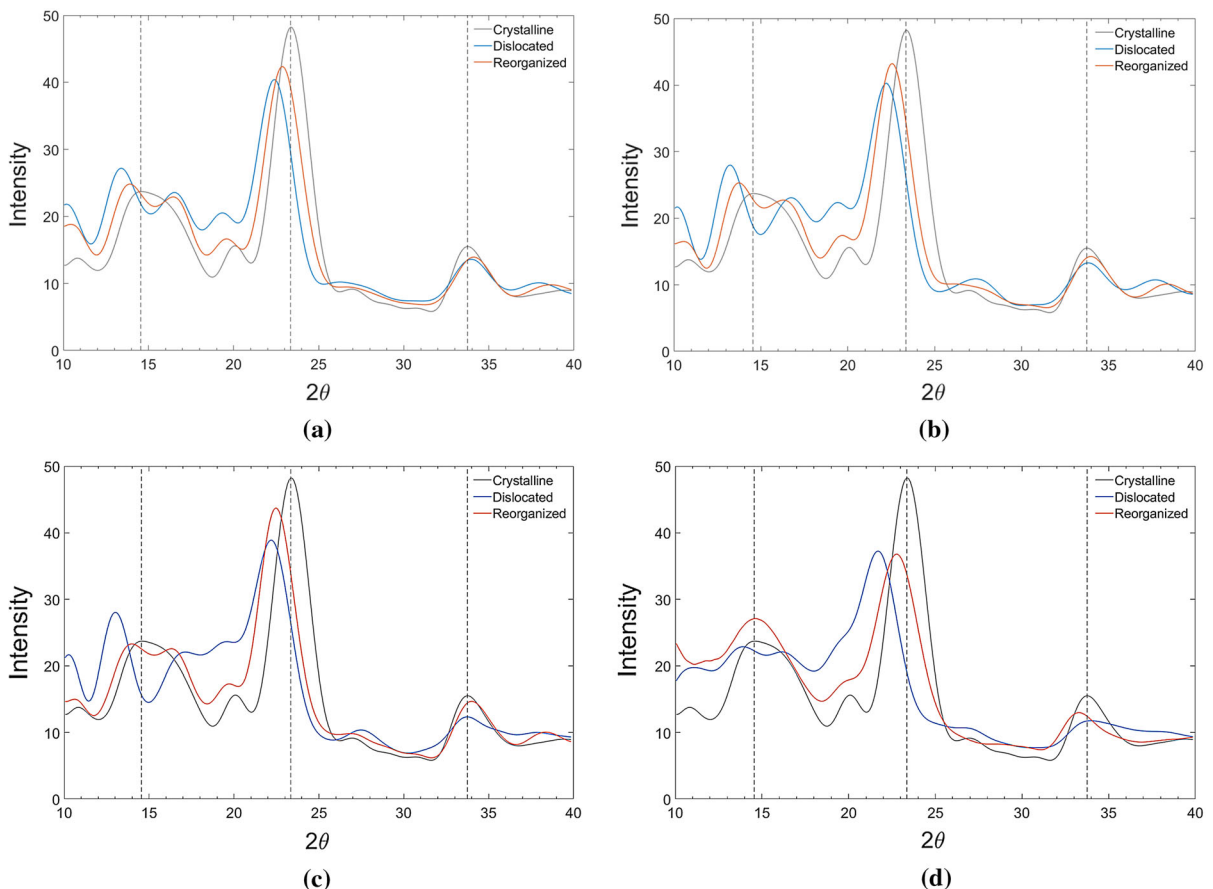


Fig. 9 Theoretical diffraction patterns of the crystalline cellulose (shown in gray) and those of the dislocated and reorganized models of **a** 6 GR, **b** 8 GR, **c** 10 GR, and **d** 9 GR in length. The peaks of the corresponding patterns to the $1\bar{1}0$, 200,

and 004 planes are shown with dashed lines. The patterns are calculated by Debyer (Wojdyr 2012). A Wavelength of $\text{CuK}\alpha$, 1.5418 Å, is used

influencing the peak position and intensities of the $1\bar{1}0$ plane. The 6 GR and 9 GR models still show minor low-angle scattering after reorganization, which obviously deteriorates the further peaks. Interestingly, in all the models, $1\bar{1}0$ and 200 peak positions move closer to those of the crystalline cellulose. All reorganized models show a shift in 200 peak position, towards $2\theta = 23.35^\circ$. This increase in 2θ values can be attributed to the decrease in the intermolecular spacing. In particular, reorganization of the chains leads to a more compact arrangement and hence higher values of 2θ at the 200 plane. Peak positions at 004 do not change significantly, especially for the first three models, despite an increase in the intensity. This suggests that the chain length remains nearly constant in the disordered or the reorganized models. Based on the XRD patterns observed here, we can conclude that there is certainly some degree of reorder in the dislocated segments after cleavage.

Above observations generate another set of conclusions that provide more evidence that a so-called “pseudo-recrystallization” phenomenon can happen only if the degree of disorder is below a certain level, i.e. H-bonds/cellobiose higher than our reported value. That is to say, despite the fact that the reordered chains never exhibit full similarities to a crystalline cellulose, as the H-bond analysis, the dihedral analysis, and the diffraction patterns show, the structures tend to reorient to arrangements close to a crystal. Hence, for the sake of simplicity, the term “recrystallization” is used instead of “pseudo-recrystallization”, hereafter.

H-bond distribution in the dislocated regions

Distribution of the H-bonds can be a decisive parameter in defining the cleavage and recrystallization trend in the dislocated segments, as well as in defining the length of these segments. In dislocated regions featuring less than 10 GRs (see Fig. 10a), there is homogeneous distribution of H-bonds to a major extent. However, if longer dislocated segments do exist, the distribution varies throughout their length. Meaning that, the number of H-bonds will be at its highest near the adjacent crystalline segments, and reach a minimum in the middle of the dislocations. Figure 10b shows the distribution of H-bonds if dislocated models would have existed as long as 25

GRs. The GR length of each bar in Fig. 10 is considered to be 3, for two reasons: shorter GRs cannot include enough inter- and intra-chain hydrogen bonds from a statistical aspect, and longer GRs could not provide clear insight when dislocation length is short, i.e. 8 GRs or less (refer to S.3 of the supplementary information for the rest of the distributions).

According to the previous results, recrystallization of the dislocated regions can happen if the H-bonds/cellobiose in that region is above an approximate value of 3.2 ± 0.2 . Hence, considering the model in Fig. 10b, there would be recrystallization under two conditions: cleavage would happen at the weakest point, which is the middle of the dislocated segment; only the two ends of the dislocated segment, each containing ~ 3 GRs, recrystallize, and the remaining 19 GRs would have to be considered as mass loss after the hydrolysis. This is different for models shorter than 9 GRs, as distribution of H-bonds is almost constant throughout the length, and in case of recrystallization, all the dislocated region can be recrystallized, and little to no mass loss is expected. As a conclusion, results show how difficult it would be to break the H-bonds within the ca. 3–4 GRs attached to the crystalline regions in such a way that H-bonds/cellobiose average in those regions falls below the observed margin of 3.2 ± 0.2 .

The lengths of dislocated cellulose segments in cellulose microfibrils

Combining the information about the number of dislocated GRs that recrystallize after cleavage with the mass loss upon vapour phase acid hydrolysis of CMFs, the length of the segments with dislocated cellulose in CMFs can be estimated.

When performing acid hydrolysis with vaporous HCl, only the dislocated regions of the CMFs are cleaved off. The resulting glucose mass loss of the CMFs can be measured, for example, for cotton and ramie CMFs, mass loss was found to be 0.1% (Kontturi et al. 2016) and 1.7% (Nishiyama et al. 2003a), respectively. The investigated CMFs had a LODP of 170 for cotton and 300 for ramie. Assuming there is *no recrystallization* of dislocated GRs, it is straightforward to calculate the number of dislocated GRs in these CMFs by multiplying the glucose mass loss m [%] by the degree of polymerization p [-], here

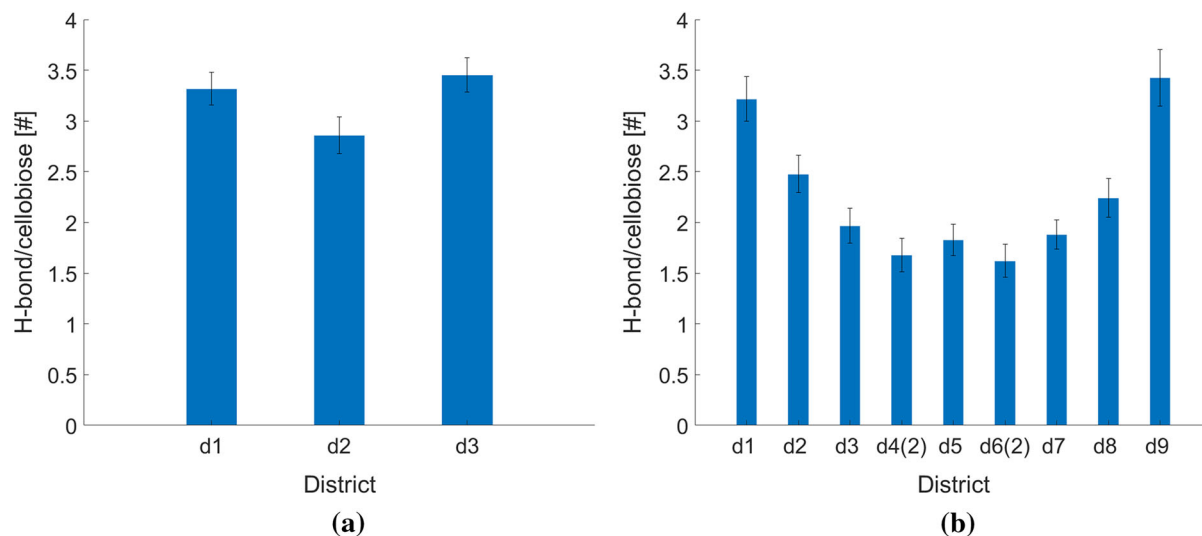


Fig. 10 Distribution of H-bonds in the dislocated regions. $dn(i)$ denotes the n^{th} district having 3 number of GRs for **a** r9, and **b** r25 models. The exception districts with 2 GRs in length are mentioned. The error bars represent the standard deviation of the mean

measured as LODP. For the cotton CMFs, this is $0.1\% \times 170 = 0.17$, meaning that the number of dislocated GRs would be essentially zero, for the ramie CMFs it is $1.7\% \times 300 = 5.1$, that is 5 GRs.

When recrystallization is considered, the number of dislocated GRs is larger for the same glucose mass loss. According to our results discussed above, on each side of a CNC there are 3 GRs having the potential to recrystallize after acidic cleavage. In reality, this number may be lower, as the dislocated cellulose may be cleaved off directly next to the crystalline region, which makes recrystallization impossible. If we assume that r GRs can recrystallize, the most optimistic scenario is that all of them are actually recrystallizing. The number of dislocated GRs g between CNCs in this case is given by $g = mp + 2r$. This scenario is an upper bound for the number of dislocated GRs, as it assumes the highest possible amount of recrystallization. The most pessimistic scenario for recrystallization is that there is equal probability for each cellobiose in the dislocated GRs to be cleaved. It is the most pessimistic because the distribution of H-bonds tends to be higher next to the crystalline region as shown in Fig. 10. Thus, in reality, there might not be equal probability for splitting, instead, the GRs next to the crystalline region are less prone to cleavage. Assuming equal probability for r GRs to be cleaved, the average amount of cleaved

GRs is $r/2$ ¹ implying that on average $r/2$ GRs are recrystallizing on each end of a CNC. Thus the lower bound for the number of dislocated GRs g between two CNCs is given by $g = mp + r$.

Utilizing the equations for upper bound $g = mp + 2r$ and lower bound $g = mp + r$ one can calculate the number of GRs in the dislocated regions of CMFs. Figure 11 provides the results for this type of analysis based on the vapour phase acid hydrolysis data for cotton CMFs (Kontturi et al. 2016) and ramie CMFs (Nishiyama et al. 2003a), as already discussed previously. The first two variables, m and p in the equations are data on the glucose mass loss from vapour phase acid hydrolysis. The variable r is the number of recrystallized GRs from dislocated segments in the CMF. Please note that the slope mp of the resulting lines is defined by the results from hydrolysis, while the offset r is defined by the number of recrystallizing GRs.

¹ The average amount of GRs cleaved off is the sum of the probability p_i of each scenario times the number of GRs cleaved off g_i in this scenario, i.e. $\sum_{i=0} p_i g_i$. Assuming a maximum number of r GRs that can be cleaved off (minimum is zero), and equal probability for splitting, p_i is constant as $p_i = 1/(r+1)$, we obtain $\sum_{i=0}^r i/(r+1) = 1/(r+1) \sum_{i=0}^r i = r(r+1)/2(r+1) = r/2$. Thus the resulting average amount of GRs cleaved off is $r/2$. This result was also confirmed by simple numerical modelling

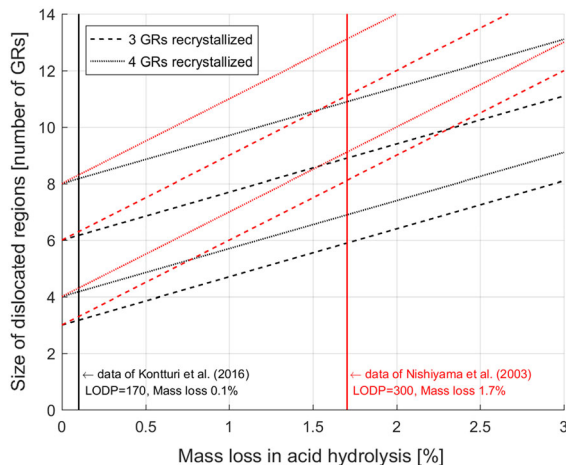


Fig. 11 Size of the dislocated regions in CMFs as a function of the mass loss in acid hydrolysis. The plot shows the results for recrystallization of 3 GRs (dashed lines) or 4 GRs (dotted lines), indicating the upper and lower bound as discussed in the text. Black lines represent results on cotton fibres based on data published by Kontturi et al. (2016), red lines are results on ramie fibres based on data from Nishiyama et al. (2003a). The intersection with the vertical solid lines gives the resulting number of dislocated GRs based on these data

According to our analysis from previous sections, in our models, it is most likely that about 3 GRs are recrystallizing on each end of a CNC, however for dislocated segments shorter than 9 GRs it is envisaged that this value increases to 4 GRs. Thus results are calculated for $r = 3, 4$, where the dashed lines are the results for 3 recrystallized GRs, the dotted lines for 4 recrystallized GRs. The intersection of the glucose mass loss (vertical solid lines) with the lines for upper and lower bound gives the resulting region estimating the number of GRs in the dislocated CMF segments. For the cotton CMFs (black lines) the length of dislocated cellulose regions is calculated to be between 3 and 8 GRs. Considering the dislocated segments with less than 7 GRs are really stable in terms of hydrogen bonds (Figure 3), which impedes cleavage, we tend to think that the length of the dislocated regions in the analyzed cotton fibres is around 6 to 8 GRs. For the ramie fibre CMFs (red lines) the size of the dislocated segment is estimated to be 7 to 11 GRs, with 3 GRs recrystallizing on each end of a CNC.

Conclusions

In this work, we have been addressing the recrystallization of amorphous regions in CMFs after cleavage, a process which has been proposed after observing an increase in crystallinity after acid vapor hydrolysis of CMFs (Kontturi et al. 2016; Spiliopoulos et al. 2021). Using molecular dynamics we modelled a 63.5 nm CMF in a 6×6 model. The model was exploited to generate dislocated segments with lengths varying between 2.6–18.5 nm. We showed that the models losing hydrogen bonds inherently result in fibrils with lower elastic modulus. We found that a hydrogen bond threshold exists below which pseudo-recrystallization cannot take place, and based on our models, it can happen when H-bonds/cellobiose ratio is higher than 3.2 ± 0.2 . The change in the distribution of the dihedral angles, as well as the results of the theoretically calculated powder diffraction patterns suggest that there is some degree of reorder for these models. The term pseudo-recrystallization is hence used as the respective GRs are in a state which is similar to crystallinity in terms of (hydrogen-)bonding, conformation, and mechanical rigidity, yet there is some remaining disorder. In summary we are suggesting that 3–4 GRs are re-crystallizing after cleavage.

Based on analysis of mass loss after acid vapor hydrolysis in the literature we calculate the length of dislocated segments in the CMFs. Depending on the source plant, we found a length of the dislocated segments of 6–8 GRs for cotton and 7–11 GRs for ramie, equivalent to 3.1–4.2 nm and 3.6–5.8 nm, respectively. Providing the considered recrystallization threshold of 3–4 GRs, our approach can be further used to calculate the dislocated segments content in other CMF samples, knowing their mass loss in acid hydrolysis.

It is worth emphasizing that, our conclusions rely on 6×6 chain models of the CMFs, which is arguably one of the most accepted arrangements. We are however aware that other arrangements of cellulose (18- and 24-chain models) are viable options. We cannot hence precisely predict how other arrangements could affect the length and size of dislocations, which will be part of future work. We think that this molecular dynamics work provides a new perspective for a quantitative understanding on the structure of dislocated cellulose in CMFs and the composition of

the plant cell wall. Further experimental and modeling work will be necessary to confirm our findings.

Acknowledgments Authors are grateful to the reviewers for their helpful comments on earlier drafts of the manuscript. The computational resources and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by the Research Foundation - Flanders (FWO) and the Flemish Government — department EWI. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 764713, project FibreNet.

Declarations

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

Data Availability All data that support this study are kept available at KU Leuven and can be accessed upon reasonable request from the corresponding author.

Ethical approval Not applicable

Consent to participate Not applicable

Consent for publication Not applicable

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Biovia DS (2017) Materials studio 2017 r2. Dassault Systèmes, San Diego
- Borrega M, Ahvenainen P, Kontturi E (2018) Impact of hydrothermal and alkaline treatments of birch kraft pulp on the levelling-off degree of polymerization (LODP) of cellulose microfibrils. *Cellulose* 25(11):6811–6818. <https://doi.org/10.1007/s10570-018-2017-7>
- Bu L, Himmel ME, Crowley MF (2015) The molecular origins of twist in cellulose I-beta. *Carbohydrate Polymers* 125:146–152. <https://doi.org/10.1308/rcsann.2016.0321>
- Case DA et al (2018) AMBER 2018. University of California, San Francisco
- Chen P, Nishiyama Y, Mazeau K (2012) Torsional entropy at the origin of the reversible temperature-induced phase transition of cellulose. *Macromolecules* 45(1):362–368. <https://doi.org/10.1021/ma201954s>
- Djahedi C, Bergenstråhle-Wohlert M, Berglund LA, Wohlert J (2016) Role of hydrogen bonding in cellulose deformation: the leverage effect analyzed by molecular modeling. *Cellulose* 23(4):2315–2323. <https://doi.org/10.1007/s10570-016-0968-0>
- Endler A, Persson S (2011) Cellulose synthases and synthesis in arabidopsis. *Mol Plant* 4(2):199–211. <https://doi.org/10.1093/mp/ssq079>
- French AD (2014) Computerized models of carbohydrates. *Polysaccharides*. Springer, Cham, pp 1–38. https://doi.org/10.1007/978-3-319-03751-6_33-1
- French AD, Montgomery DW, Prevost NT, Edwards JV, Woods RJ (2021) Comparison of cellooligosaccharide conformations in complexes with proteins with energy maps for cellobiose. *Carbohydrate Polymers* 264:118004. <https://doi.org/10.1016/j.carbpol.2021.118004>
- Gibson LJ (2012) The hierarchical structure and mechanics of plant materials. *J The Royal Soc Interface* 9(76):2749–2766. <https://doi.org/10.1098/rsif.2012.0341>
- Hadden JA, French AD, Woods RJ (2013) Unraveling cellulose microfibrils: A twisted tale. *Biopolymers* 99(10):746–756. <https://doi.org/10.1002/bip.22279>
- Hadden JA, French AD, Woods RJ (2014) Effect of microfibril twisting on theoretical powder diffraction patterns of cellulose I β . *Cellulose* 21(2):879–884. <https://doi.org/10.1007/s10570-013-0051-z>
- Hearle J (1958) A fringed fibril theory of structure in crystalline polymers. *J Polym Sci* 28(117):432–435. <https://doi.org/10.1002/pol.1958.1202811722>
- Heiner AP, Sugiyama J, Teleman O (1995) Crystalline cellulose I α and I β studied by molecular dynamics simulation. *Carbohydr Res* 273(2):207–223. [https://doi.org/10.1016/0008-6215\(95\)00103-Z](https://doi.org/10.1016/0008-6215(95)00103-Z)
- Jarvis MC (2013) Cellulose biosynthesis: counting the chains. *Plant Physiol* 163(4):1485–1486. <https://doi.org/10.1104/pp.113.231092>
- Jeffrey GA (1997) An introduction to hydrogen bonding. Oxford University Press, Oxford
- Kannam SK, Oehme DP, Doblin MS, Gidley MJ, Bacic A, Downton MT (2017) Hydrogen bonds and twist in cellulose microfibrils. *Carbohydr Polym* 175:433–439. <https://doi.org/10.1016/j.carbpol.2017.07.083>
- Kargarzadeh H, Ahmad I, Abdullah I, Dufresne A, Zainudin SY, Sheltami RM (2012) Effects of hydrolysis conditions on the morphology, crystallinity, and thermal stability of cellulose nanocrystals extracted from kenaf bast fibers. *Cellulose* 19(3):855–866. <https://doi.org/10.1007/s10570-012-9684-6>
- Khodayari A, Hirn U, Van Vuure A, Seveno D (2020a) Inverse rule of mixtures at the nanoscale: prediction of elastic properties of cellulose nanofibrils. *Compos Part A: Appl Sci Manuf* 138(July):106046. <https://doi.org/10.1016/j.compositesa.2020.106046>
- Khodayari A, Van Vuure AW, Hirn U, Seveno D (2020b) Tensile behaviour of dislocated/crystalline cellulose fibrils at the nano scale. *Carbohydrate Polymers* 235:115946. <https://doi.org/10.1016/j.carbpol.2020.115946>

- Kirschner KN, Yongye AB, Tschampel SM, González-Outeirio J, Daniels CR, Foley BL, Woods RJ (2008) GLY-CAM06: a generalizable biomolecular force field carbohydrates. *J Comput Chem* 29:622–655
- Kontturi E, Meriluoto A, Penttilä PA, Baccile N, Malho JM, Potthast A, Rosenau T, Ruokolainen J, Serimaa R, Laine J, Sixta H (2016) Degradation and crystallization of cellulose in hydrogen chloride vapor for high-yield isolation of cellulose nanocrystals. *Angewandte Chemie - International Edition* 55(46):14455–14458. <https://doi.org/10.1002/anie.201606626>
- Kulasinski K, Keten S, Churakov SV, Derome D, Carmeliet J (2014) A comparative molecular dynamics study of crystalline, paracrystalline and amorphous states of cellulose. *Cellulose* 21(3):1103–1116
- Kulasinski K, Guyer R, Derome D, Carmeliet J (2015a) Water adsorption in wood microfibril-hemicellulose system: role of the crystalline-amorphous interface. *Biomacromolecules* 16(9):2972–2978
- Kulasinski K, Guyer R, Derome D, Carmeliet J (2015b) Water diffusion in amorphous hydrophilic systems: a stop and go process. *Langmuir* 31(39):10843–10849
- Kulasinski K, Derome D, Carmeliet J (2017) Impact of hydration on the micromechanical properties of the polymer composite structure of wood investigated with atomistic simulations. *J Mech Phys Sol* 103:221–235
- Lorenz M, Sattler S, Reza M, Bismarck A, Kontturi E (2017) Cellulose nanocrystals by acid vapour: towards more effortless isolation of cellulose nanocrystals. *Faraday Discuss* 202:315–330. <https://doi.org/10.1039/c7fd00053g>
- Mark H (1940) Intermicellar hole and tube system in fiber structure. *J Phys Chem* 44(6):764–788. <https://doi.org/10.1021/j150402a009>
- Matthews JF, Skopec CE, Mason PE, Zuccato P, Torget RW, Sugiyama J, Himmel ME, Brady JW (2006) Computer simulation studies of microcrystalline cellulose I β . *Carbohydr Res* 341(1):138–152. <https://doi.org/10.1016/j.carres.2005.09.028>
- Mazeau K, Heux L (2003) Molecular dynamics simulations of bulk native crystalline and amorphous structures of cellulose. *J Phys Chem B* 107(10):2394–2403. <https://doi.org/10.1021/jp0219395>
- Miyamoto H, Umemura M, Aoyagi T, Yamane C, Ueda K, Takahashi K (2009) Structural reorganization of molecular sheets derived from cellulose II by molecular dynamics simulations. *Carbohydr Res* 344(9):1085–1094. <https://doi.org/10.1016/j.carres.2009.03.014>
- Mohammad AS, Biernacki JJ, Northrup S, Adenson M (2018) Diffusion of CO₂ and fractional free volume in crystalline and amorphous cellulose. *J Anal Appl Pyrolysis* 134:43–51
- Morthomas J, Gonçalves W, Perez M, Foray G, Martin CL, Chantrenne P (2019) A novel method to predict the thermal conductivity of nanoporous materials from atomistic simulations. *J Non-Cryst Sol* 516(April):89–98. <https://doi.org/10.1016/j.jnoncrsol.2019.04.017>
- Müller M, Czihak C, Schober H, Nishiyama Y, Vogl G (2000) All disordered regions of native cellulose show common low-frequency dynamics. *Macromolecules* 33(5):1834–1840. <https://doi.org/10.1021/ma9912271>
- Nishiyama Y (2009) Structure and properties of the cellulose microfibril. *J Wood Sci* 55(4):241–249. <https://doi.org/10.1007/s10086-009-1029-1>
- Nishiyama Y, Langan P, Chanzy H (2002) Crystal structure and hydrogen-bonding system in cellulose I β from synchrotron X-ray and neutron fiber diffraction. *J Am Chem Soc* 124(31):9074–9082. <https://doi.org/10.1021/ja0257319>
- Nishiyama Y, Kim UJ, Kim DY, Katsumata KS, May RP, Langan P (2003a) Periodic disorder along ramie cellulose microfibrils. *Biomacromolecules* 4(4):1013–1017. <https://doi.org/10.1021/bm025772x>
- Nishiyama Y, Sugiyama J, Chanzy H, Langan P (2003b) Crystal Structure and Hydrogen Bonding System in Cellulose I α from Synchrotron X-ray and Neutron Fiber Diffraction. *J Am Chem Soc* 125(47):14300–14306. <https://doi.org/10.1021/ja037055w>
- Nishiyama Y, Johnson GP, French AD (2012) Diffraction from nonperiodic models of cellulose crystals. *Cellulose* 19(2):319–336. <https://doi.org/10.1007/s10570-012-9652-1>
- Oehme DP, Downton MT, Doblin MS, Wagner J, Gidley MJ, Bacic A (2015) Unique aspects of the structure and dynamics of elementary I β cellulose microfibrils revealed by computational simulations. *Plant Physiol* 168(1):3–17. <https://doi.org/10.1104/pp.114.254664>
- Ogawa Y (2019) Electron microdiffraction reveals the nanoscale twist geometry of cellulose nanocrystals. *Nanoscale* 11(45):21767–21774. <https://doi.org/10.1039/c9nr06044h>
- Pääkkönen T, Spiliopoulos P, Knuts A, Nieminen K, Johansson LS, Enqvist E, Kontturi E (2018) From vapour to gas: optimising cellulose degradation with gaseous HCl. *React Chem Eng* 3(3):312–318. <https://doi.org/10.1039/c7re00215g>
- Santiago Cintrón M, Johnson GP, French AD (2011) Young's modulus calculations for cellulose I β by MM3 and quantum mechanics. *Cellulose* 18(3):505–516. <https://doi.org/10.1007/s10570-011-9507-1>
- Sousa Da Silva AW, Vranken WF (2012) ACPYPE - AnteChamber PYthon Parser interface. *BMC Research Notes* 5(367)
- Spiliopoulos P, Spirk S, Pa T, Awais M, Kontturi E (2021) Visualizing degradation of cellulose nano fibers by acid hydrolysis. *Biomacromolecules*. <https://doi.org/10.1021/acs.biomac.0c01625>
- van der Spoel Y, Lindahl E, Hess B, Groenho G, Mark AE, Berendsen HJC (2005) GROMACS: fast, flexible and free. *J Comput Chem* 19(6):1701–1718
- Wojdyr M (2012) Debye. <https://debyer.readthedocs.io/en/latest/https://github.com/wojdyr/debyer>
- Zhang X, Tschopp MA, Horstemeyer MF, Shi SQ, Cao J (2013) Mechanical properties of amorphous cellulose using molecular dynamics simulations with a reactive force field. *Int J Model, Identif Control* 18(3):211–217
- Zhao Z, Shklyav OE, Nili A, Mohamed MNA, Kubicki JD, Crespi VH, Zhong L (2013) Cellulose microfibril twist, mechanics, and implication for cellulose biosynthesis. *J Phys Chem A* 117(12):2580–2589. <https://doi.org/10.1021/jp3089929>

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