Chemistry
A European
Journal

Chemistry Europe
European Chemical Societies Publishing

Accepted Article

Title: Chemoenzymatic synthesis of fluorinated cellodextrins identifies a new allomorph for cellulose-like materials

Authors: Peterson de Andrade, Juan C. Muñoz-García, Giulia Pergolizzi, Valeria Gabrielli, Sergey A. Nepogodiev, Dinu Iuga, László Fábián, Rinat Nigmatullin, Marcus A. Johns, Robert Harniman, Stephen J. Eichhorn, Jesús Angulo, Yaroslav Z Khimyak, and Robert A. Field

To be cited as: *Chem. Eur. J.* 10.1002/chem.202003604

Link to VoR: https://doi.org/10.1002/chem.202003604

WILEY-VCH

10.1002/chem.202003604

Chemoenzymatic synthesis of fluorinated cellodextrins identifies a new allomorph for cellulose-like materials

Peterson de Andrade[1,7‡], Juan C. Muñoz-García[2‡], Giulia Pergolizzi[1,3‡], Valeria Gabrielli[2], Sergey A. Nepogodiev[1], Dinu Iuga[4], László Fábián,[2] Rinat Nigmatullin[5], Marcus A. Johns[5], Robert Harniman[6], Stephen J. Eichhorn[5], Jesús Angulo[2], Yaroslav Z. Khimyak[*2], Robert A. Field[*1,3,7]

[1]*Department of Biological Chemistry, John Innes Centre, Norwich NR4 7UH, UK*
[2]*School of Pharmacy, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK*
[3]*Iceni Diagnostics Ltd, Norwich Research Park Innovation Centre, Colney Lane, Norwich, Norfolk NR4 7GJ, United Kingdom*
[4]*Department of Physics, University of Warwick, Coventry CV4 7AL, UK*
[5]*Bristol Composites Institute, CAME School of Engineering, University of Bristol, Bristol BS8 1TR, UK*
[6]*School of Chemistry, University of Bristol, Bristol BS8 1TS, UK*
[7]*Present address: Department of Chemistry and Manchester Institute of Biotechnology, University of Manchester, Manchester M1 7DN, UK*

‡ Equal contribution

[*]Corresponding author e-mail address: y.khimyak@uea.ac.uk; robert.field@manchester.ac.uk

Abstract

Understanding the fine details of self-assembly of building blocks into complex hierarchical structures represents a major challenge *en route* to the design and preparation of soft matter materials with specific properties. Enzymatically synthesised cellodextrins are known to have limited water solubility beyond DP9, a point at which they self-assemble into particles resembling the antiparallel cellulose II crystalline packing. We have prepared and characterised a series of site selectively fluorinated cellodextrins of different degrees of fluorination and substitution patterns by chemoenzymatic synthesis. Bearing in mind the potential disruption of the hydrogen bond network of cellulose II, we have prepared and characterised a multiply 6-fluorinated cellodextrin. In addition, a series of single site selectively fluorinated cellodextrins were synthesised to assess the structural impact upon addition of one fluorine atom per chain. The structural characterisation of these materials at different length scales, combining advanced NMR and microscopy methods, showed that a 6-fluorinated donor substrate yielded multiply 6-fluorinated cellodextrin chains that assembled into particles presenting morphological and crystallinity features, and intermolecular interactions, that are unprecedented for cellulose-like materials.

Introduction

Cellulose is an abundant natural biopolymer used extensively in industry as a raw material for the production of paper, textile, food thickeners, dietary fibre, etc.[1,2] The current use of cellulose increasingly involves nanosized cellulose particles (nanocellulose), which is a promising class of renewable material due to its intrinsic characteristics and potential for a broad range of industrial applications.[3–5] The development of nanocellulose-based materials relies on assembly-driven processes, the manipulation of which can impact on mechanical properties or bring additional functionality to the material.[6–10]

The production of cellulose nanocrystals and nanofibrillated cellulose, the main classes of nanocellulose, both rely on top-down bioprocessing methodologies, based on the isolation of nanocellulose from cellulosic biomass, which requires high energy consumption.[5] In addition, the functionalisation of nanocellulose to meet requirements for specific applications often requires harsh chemical conditions (*i.e.* strong acids and bases). As an alternative, enzymatic synthesis presents an attractive approach,[11,12] enabling the bottom-up preparation of site-specifically modified oligo- and poly-saccharides in a regio- and stereo-controlled manner.[13–16] Specifically in relation to glucose-based materials, glycoside phosphorylases (GPs)[17–21] have shown substantial potential for the synthesis of amylose- and cellulose-like materials. In particular, cellodextrin phosphorylase (CDP, EC 2.4.1.49)[22–24] has emerged as a powerful tool for the synthesis of differently functionalised cellulose oligomers, giving rise to a variety of nanostructures (sheets,[25,26] rods,[27] or ribbons[28]) depending on the nature of the substrate.

The ability to systematically modify the structure of oligo- and poly-saccharides presents new opportunities to gain insight into the hierarchical self-assembly of carbohydrate-based materials.[29] In connection with the present study, we had a need for the site-specific introduction of probes into cellulose to report on local structure and solvation, and potentially to modulate material properties. Fluorine is well-known for its unique physicochemical properties, such as small size, high electronegativity, great polarity and stability of the C–F bond.[30] In addition, the absence of fluorine in biological systems and in the majority of materials makes the introduction of ^{19}F nuclei a powerful reporter of local structure and environment. For instance, ^{19}F NMR has been used to monitor crystallisation in nanoporous materials[31] and fibrillation of intrinsically disordered proteins,[32] to characterise polymeric biomaterials,[33] and to map the interactions of fluorinated oligosaccharides with protein targets.[34] Nonetheless, the use of fluorine remains under-explored with respect to carbohydrate-based materials.

The top-down derivatisation of cellulose[8,35] is complicated by solubility challenges, resulting in incomplete control of the sites and extent of fluorination.[36] The bottom-up chemical synthesis of structurally defined cellodextrins, including 3-fluorinated compounds,[37] has been achieved recently by automated chemical synthesis approaches.[38] In the present study, we wanted to investigate the impact of incorporation of fluorine in place of the primary alcohols in cellulose, which are more accessible and have a fundamental role in the hydrogen-bonding network that gives rise to the native cellodextrin structure (cellulose type II).[39] Recognising the detrimental electronic impact of fluorination on sugar reactivity, we reasoned that enzymatic polymerisation *per se* may be inefficient: Kobayashi *et al.* note the reduced reactivity of 6-fluorinated GlcNAc oxazoline towards chitinase-mediated polymerisation, for instance.[40] We therefore opted to exploit well-studied cellodextrin phosphorylase (CDP) as it is known to produce DP 9 cellodextrin, which due to its limited water solubility results in anti-parallel glucan chain association (cellulose II-like material).[22] Our expectation therefore was that the

need for only a limited number of glycosylation events with CDP might still be achievable even allowing for the impact of fluorination on donor and/or acceptor substrate reactivity. Likewise, we expected to achieve a higher structural impact on the fluorinated cellodextrin produced from the modified donor, as it allows the introduction of multiple fluorine atoms along the chains. In addition, single site fluorination was also investigated to test the hypothesis of the effect of one fluorine atom per chain on the native cellodextrin structure. Furthermore, single site fluorination can be used as a potential probe for future multicomponent aggregation studies.

Herein, we enzymatically produced cellodextrins (EpCs) with different fluorination patterns. Monofluorinated EpCs (2F-, 3F- and 6F-EpC) were obtained by CDP-mediated oligomerisation of α-D-glucose 1-phosphate (Glc-1P) as donor and deoxy-fluoro-cellobioses as acceptor substrates. Multiply 6-fluorinated EpC (multi-6F-EpC) was prepared from 6-deoxy-6-fluoro-α-D-glucose 1-phosphate (6F-Glc-1P) and cellobiose as donor and acceptor substrates, respectively. We demonstrate that the presence of a single fluorine atom per cellodextrin chain did not exert a substantial impact on the morphology and crystalline structure of the material, while the presence of multiple 6-deoxy-6-fluoroglucose units yielded an unprecedented crystalline allomorph never reported before for a cellulose-like material.

Results and discussion

Enzymatic synthesis of fluorinated cellodextrins

*Synthesis of 2-, 3- and 6-monofluorinated cellodextrins (2F-EpC, **4**; 3F-EpC, **5**; and 6F-EpC, **6**)*

CDP uses glucose as an acceptor substrate only poorly, compared to cellobiose and longer cello-oligosaccharides,[22,24] to produce cellodextrins containing deoxy-fluoro-glucose at the reducing terminus (compounds **4**-**6**). We therefore initially used cellobiose phosphorylase (CBP) (PRO-GH94-004) to synthesise monofluorinated cellobiose analogues (**1**-**3**) (Figure 1A) for use as acceptors for CDP (Figure 1B). CBP was incubated at 37 °C with Glc-1P (100 mM) and deoxy-fluoro-glucose (2F-, 3F- or 6F-Glc) (100 mM) for 16 h, at which point TLC showed *ca.* 80% conversion into the disaccharides **1** and **3**, and *ca.* 60% into **2**. The different conversion efficiencies may be rationalised based on a study of *Cellulomoinas uda* cellobiose phosphorylase,[41] in which k_{cat}/K_m values for 2F-Glc (2.4%), 3F-Glc (0.013%) and 6F-Glc (31%) acceptors are substantially lower than that of the parent Glc substrate, but all three compounds are indeed productive substrates. CBP was removed from the reaction mixture by affinity chromatography (His6-tag nickel column purification) and the desired products were purified by gel filtration chromatography. The purification successfully removed residual deoxy-fluoro-glucose acceptors, but small amounts of cellobiose required removal by HPLC to obtain compounds **1**-**3** (4-9 mg) in high purity for characterisation purpose (†ESI Figures S1-S3). It is important to highlight that the monofluorinated cellodextrins **4**-**6** could be obtained in one-pot reactions from the respective sugar-1P and fluorinated glucoses without HPLC purification.

Once Glc-1P consumption was almost complete in the CBP reactions, more Glc-1P (4 eq.) was added together with CDP and the reactions were incubated at 37 °C shaking for 16 h. A white precipitate was formed and isolated by centrifugation, followed by resuspension and washing with MQ water to remove enzyme, salts and any soluble sugars. Further 4 eq. of Glc-1P were added to the supernatant, and CDP reaction was further incubated to produce more fluorinated EpCs. In this manner, monofluorinated cellodextrins were obtained (*ca*. 40 mg) with

10.1002/chem.202003604

reasonable overall yield based on consumed fluoro-glucose [47% (2F-EpC, **4**), 30% (3F-EpC, **5**), 32% (6F-EpC, **6**)]. MALDI-TOF mass spectrometry analysis showed these materials to have an average DP *ca*. 9, while the unsubstituted cellodextrin (EpC, **8**) produced under the same reaction conditions averaged *ca.* DP 8 (†ESI Figure S7). Traces of longer fluorinated cellodextrins were evident in the mass spectrometry data, which may reflect greater water solubility of the monofluorinated materials, thus resulting in further enzymatic extension. Solution-state ^{19}F NMR analysis in 1 M NaOD (†ESI Figure S4) showed two singlets for each material, reflecting reducing terminal anomers, with peaks at -195.21 and -195.26 ppm (2F-EpC, **4**), -190.86 and -197.19 ppm (3F-EpC, **5**) and -232.55 and -234.05 ppm (6F-EpC, **6**).

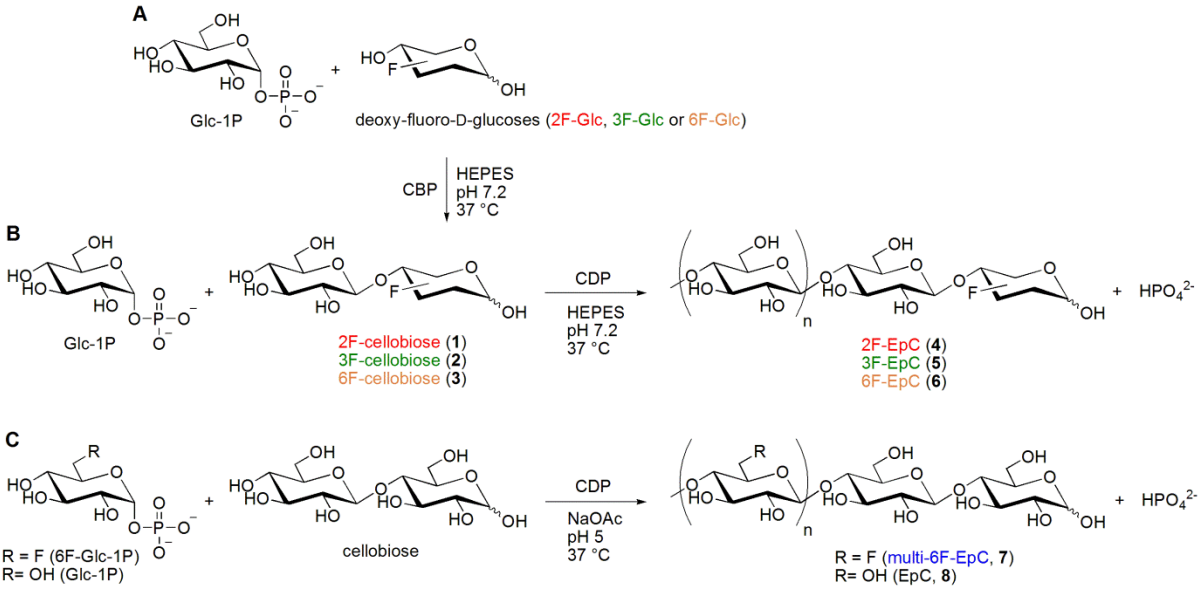


Figure 1. **Enzymatic synthesis of fluorinated cellodextrins**. (**A**) Cellobiose phosphorylase (CBP) catalysed reaction of α-D-glucose 1-phosphate (Glc-1P) and deoxy-fluoro-D-glucose (2F-Glc, 3F-Glc or 6F-Glc), followed by (**B**) cellodextrin phosphorylase (CDP) catalysed oligomerisation with Glc-1P and monofluorinated cellobioses, to afford enzymatically-produced fluorinated cellodextrins (2F-EpC, **4**; 3F-EpC, **5**; and 6F-EpC, **6**). (**C**) CDP-catalysed reaction of 6-deoxy-6-fluoro-α-D-glucose 1-phosphate (6F-Glc-1P) or Glc-1P and cellobiose as acceptor, to produce multiply 6-fluorinated cellodextrin (multi-6F-EpC, **7**) or the parent Enzymatically produced Cellodextrin (EpC, **8**), respectively.

Synthesis of multiply 6-fluorinated cellodextrin (multi-6F-EpC, 7)

We also investigated CDP-mediated oligomerisation using the chemically modified glucosyl donor 6F-Glc-1P (Figures S5-S6)[42] and cellobiose as acceptor (Figure 1C) to achieve higher structural impact by placing multiple fluorine atoms along the cellodextrin (multiply 6-fluorinated cellodextrin, multi-6F-EpC, **7**). The initial tests using 6F-Glc as an acceptor to obtain a fully 6F-substitued cellodextrin proved very slow and inefficient. Alternatively, CBP was tested with 6F-Glc as an acceptor to generate a difluorinated cellobiose, which could be a better substrate for CDP. However, only trace amounts of the product were detected, prompting us to choose the natural acceptor cellobiose. CDP was incubated at 37 °C with 6F-Glc-1P (200 mM) and cellobiose (30 mM) for 72 h. The resulting white precipitate was isolated by centrifugation, followed by re-suspension and washing with MilliQ water to give **7** with 64% yield. ^{19}F solution state NMR analysis of **7** dissolved in 1 M NaOD (†ESI Figure S4) showed one major singlet at -233.25 ppm, which may correspond to fluorine from the 6F-Glc internal repeating units, and three smaller singlets at -233.29, -233.31 and -233.35 ppm from 6F-Glc

close to the reducing terminal and the non-reducing terminal unit. Analysis by MALDI-TOF mass spectrometry revealed that multi-6F-EpC **7** (17 mg) had a higher average DP (*ca.* 10) than the parent EpC (*ca.* DP 8) and that longer chains, up to DP 15, could also be observed in the multiple 6-fluorinated material (†ESI Figure S7). These data are comparable to the monofluorinated compounds and, more importantly, the presence of multiple fluorine atoms clearly had a higher impact on the DP of the cellodextrin products. The quantities of multiply 6-fluorinated material obtained in these proof of concept studies enabled us to carry on to detailed structural characterisation at different length scales; scale up of enzymatic syntheses to provide materials for bulk physical properties assessment will be reported in due course.

Morphological characterisation

Electron Microscopy (EM) and Atomic Force Microscopy (AFM)

Transmission electron microscopy (TEM) was initially used to observe the morphological differences between EpC and fluorinated EpCs, which were prepared for analysis only by dilution of concentrated suspensions obtained after purification of precipitates formed during enzymatic synthesis. As expected, the TEM images of the monofluorinated 2F-EpC (**4**), 3F-EpC (**5**) and 6F-EpC (**6**) (†ESI Figure S8) show a very similar morphology to EpC (**8**, Figure 2A, a). This crystalline sheet-like morphology is well-known for enzymatically synthesised cello-oligosaccharides,[22,43] including derivatised cellulose, such as acrylated cellulose[26] and cellulose conjugated with oligo(ethylene glycol).[28] On the other hand, multi-6F-EpC (**7**) particles formed predominantly into significantly shorter platelets (< 100 nm length) (Figure 2B, a). These differences were further confirmed by AFM imaging using samples prepared by depositing diluted sample suspensions on freshly cleaved mica (Figure 2B, b and c). Although a few long platelets are present in multi-6F-EpC (**7**), their fraction is smaller than in EpC (**8**). As reported in the literature,[22,43] the thickness of EpC (**8**) platelets was found to be ~5 nm. Similar thicknesses were observed for long platelets of multi-6F-EpC (**7**, Figure 2B, c).

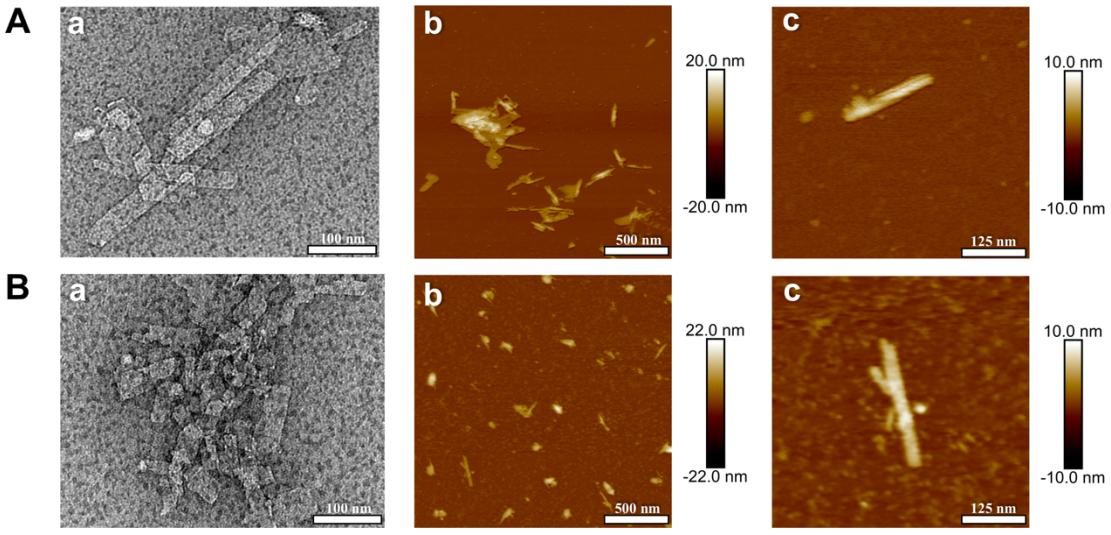


Figure 2. TEM (**a**) and AFM (**b** and **c**) images of EpC (**8**, row **A**) and multi-6F-EpC (**7**, row **B**). Scale bars are shown at the bottom of each image. The gradient bars next to **b** and **c** correspond to height measurements.

Long-range structural characterisation by powder X-ray diffraction (PXRD)

The PXRD patterns of the monofluorinated 2F-EpC (**4**), 3F-EpC (**5**) and 6F-EpC (**6**) are virtually indistinguishable from the diffraction pattern of EpC (**8**) (Figure 3). This result indicates that the monofunctionalised cellodextrin-like molecules arrange as a cellulose type II allomorph, with three intense and sharp peaks located at 2θ = 12°, 20° and 23° (d-spacings of 0.74, 0.44 and 0.39 nm, respectively) representing $(1\bar{1}0)$, (110) and (020) planes.[39,44] On the other hand, the experimental PXRD pattern reported for multi-6F-EpC (**7**) does not correspond to any allomorph previously described for cellulose[44–49] (Figure 3). The pattern shows two well defined peaks at 2θ = 15° and 23° (d-spacings of 0.59 and 0.39 nm, respectively), as well as four different broad components at 2θ = 21°, 25°, 30° and 36° (d-spacings of 0.42, 0.36, 0.30 and 0.25 nm, respectively).

In order to verify possible similarities with previously reported cellulose structural organisations, we predicted and compared the PXRD spectra of multi-6F-EpC (**7**) to each known allomorph (†ESI Figure S9 and Table S1). Remarkably, the observed peak positions of multi-6F-EpC (**7**) are unique when compared to the diffraction patterns of the known allomorphs (Figure 3, †ESI Figure S9 and Table S1), hence demonstrating the formation of a new crystalline structure for this new cellulose-like material.

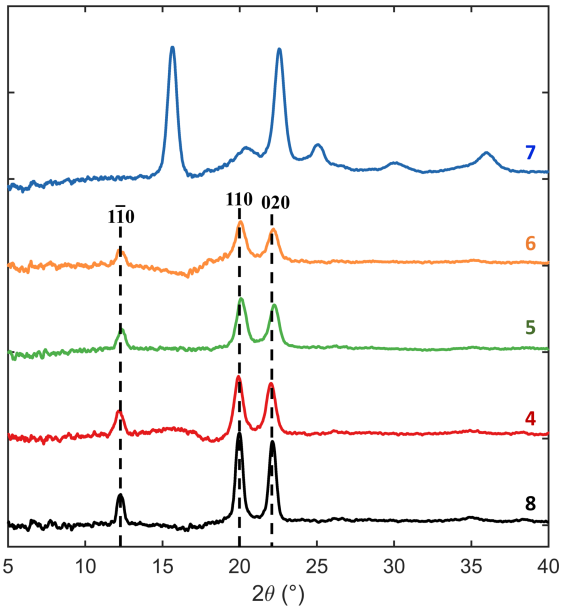


Figure 3. Powder Wide Angle X-ray diffraction patterns of EpC (**8**, black), 2F-EpC (**4**, red), 3F-EpC (**5**, green), 6F-EpC (**6**, orange) and multi-6F-EpC (**7**, blue).

Molecular level characterisation

Raman spectroscopy

Figure 4 shows typical Raman spectra of 2F-EpC (**4**), 3F-EpC (**5**), 6F-EpC (**6**), multi-6F-EpC (**7**) and EpC (**8**). The bands located at *ca.* 1462 (HOC and HCH stretching), 1265 (HCC and HCO stretching) and 576 cm^{-1} (heavy atom stretching) and the dominance of the band located at *ca.* 354 cm^{-1} over the band located at *ca.* 379 cm^{-1} (both heavy atom stretching) confirmed that EpC (**8**) arranges into a cellulose type II structure.[50] **4**, **5** and **6** are very similar to **8**, as expected for a single fluorine atom (at the reducing end) per oligosaccharide chain. The weak band located at 487 cm^{-1} for the monofluorinated EpCs is probably an amalgamation of the 480 and 496 cm^{-1} bands as a result of the single fluorine present in each chain.

In contrast, the multi-6F-EpC (**7**) spectrum is significantly different owing to the presence of multiple fluorine atoms. Multi-6-fluorination results in new Raman bands located at 480, 496 and 924 cm^{-1}. The first two bands correspond to the presence of CH_xF, whilst the third relates to CH_2F, confirming modification at the C6 position.[51] It is likely that the presence of the additional bands is simply due to the more extensive presence of these groups within the multi-6F-EpC sample, since these are relatively more polar in nature, say compared to the C-O-C vibrations which being symmetric are present in all samples. Multi-6-fluorination also results in the shift of multiple bands, including 1462 to 1451 cm^{-1} and 1265 to 1268 cm^{-1}, as well as the loss of others, such as the band located at 576 cm^{-1} (Table S2). Most significantly, the band associated with the glycosidic linkage located at 1097 cm^{-1} is shifted to 1088 cm^{-1}. This provides some evidence that the crystal structure of the multi-6F-EpC material is neither cellulose type I nor type II. In any case, it should be noted that these changes most likely reflect differences in the packing of the oligosaccharide chains only, and do not involve conformational changes at the level of the glucose units. In this regard, quantum mechanics calculations carried out for fluorinated cellobiose showed that neither the puckering of the glucose units ($^4C_1$) nor the conformation of the glycosidic bonds are affected by the substitution of all hydroxyl groups by fluorine atoms.[52]

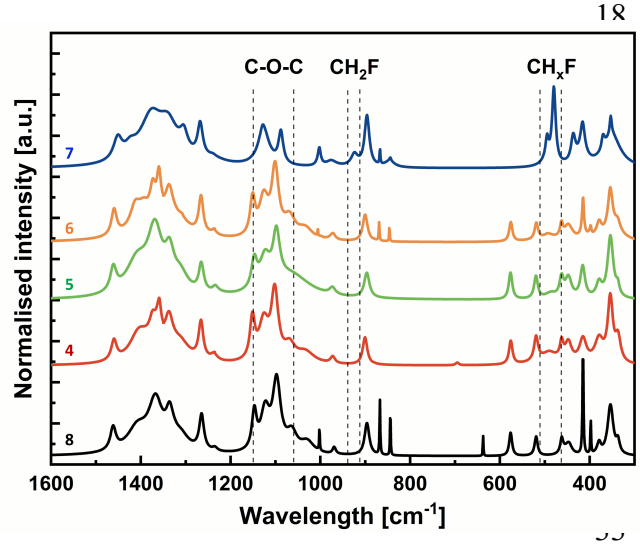


Figure 4. Deconvoluted and normalised Raman spectra for EpC (**8**, black), 2F-EpC (**4**, red), 3F-EpC (**5**, green), 6F-EpC (**6**, orange) and multi-6F-EpC (**7**, blue). Dashed lines correspond to boundaries of bands associated with C-O-C stretching (C-O-C) and presence of fluorinated carbon groups (CH_2F and CH_xF). Each Raman spectrum represents the average of three Lorentzian-deconvoluted spectra upon noise removal.

Solid-state Nuclear Magnetic Resonance (SSNMR)

Direct polarisation ^{19}F NMR experiments (without ^1H decoupling) were carried out at 60 kHz MAS rate for the 2F-EpC (**4**), 3F-EpC (**5**), 6F-EpC (**6**) and multi-6F-EpC (**7**) (Figure 5a). A single very broad and asymmetric peak was observed for the monofluorinated materials, centred at -190 (2F-EpC, **4**), -197 (3F-EpC, **5**) and -232 ppm (6F-EpC, **6**), respectively, in good agreement with the solution ^{19}F NMR data (†ESI Figure S4). 3F-EpC (**5**) and 6F-EpC (**6**) showed broad peaks, with line widths at half height of 11.9 and 9.4 kHz, respectively (Figure 5c), while multi-6F-EpC (**7**) showed a sharper (3.8 kHz width at half height) Lorentzian-shaped peak (Figure 5c). ^1H-decoupled ^{19}F ($^{19}F\{^1H\}$) NMR spectra of 3F-EpC recorded at slower MAS rate showed an even broader ^{19}F peak (†ESI Figure S10), indicating that (i) fast MAS is more efficient at decoupling that radiofrequency decoupling (fast MAS decouples both ^{19}F-^{19}F homonuclear dipolar coupling as well as heteronuclear ^1H-^{19}F coupling), and (ii) the large ^{19}F line widths of 2F-EpC (**4**), 3F-EpC (**5**), 6F-EpC (**6**) and multi-6F-EpC (**7**) in the fast MAS

10.1002/chem.202003604

spectra (Figure 5a) are mostly due to the large heterogeneity of ^{19}F chemical environments. This can be easily understood considering that these materials assemble into particles with a specific crystalline packing (cellulose type II[44] for 2F-EpC (**4**), 3F-EpC (**5**) and 6F-EpC (**6**), and a new organisation for muti-6F-EpC (**7**); Figures 5d and 6a), and the ^{19}F nucleus is extremely sensitive to chemical environment. Upon assembly of nanocellulose the ^{19}F atoms of each cellulose chain can occupy any position within the nanofibril (surface, core, far from or nearby other fluorinated residues, etc.), hence presenting non-equivalent environments within the packing of EpC (Figure 5c). Assuming that ^{19}F-^{1}H dipolar interactions are reduced considerably at fast MAS, the peak broadening reflects a multitude of orientations sampled by the C-F bonds.

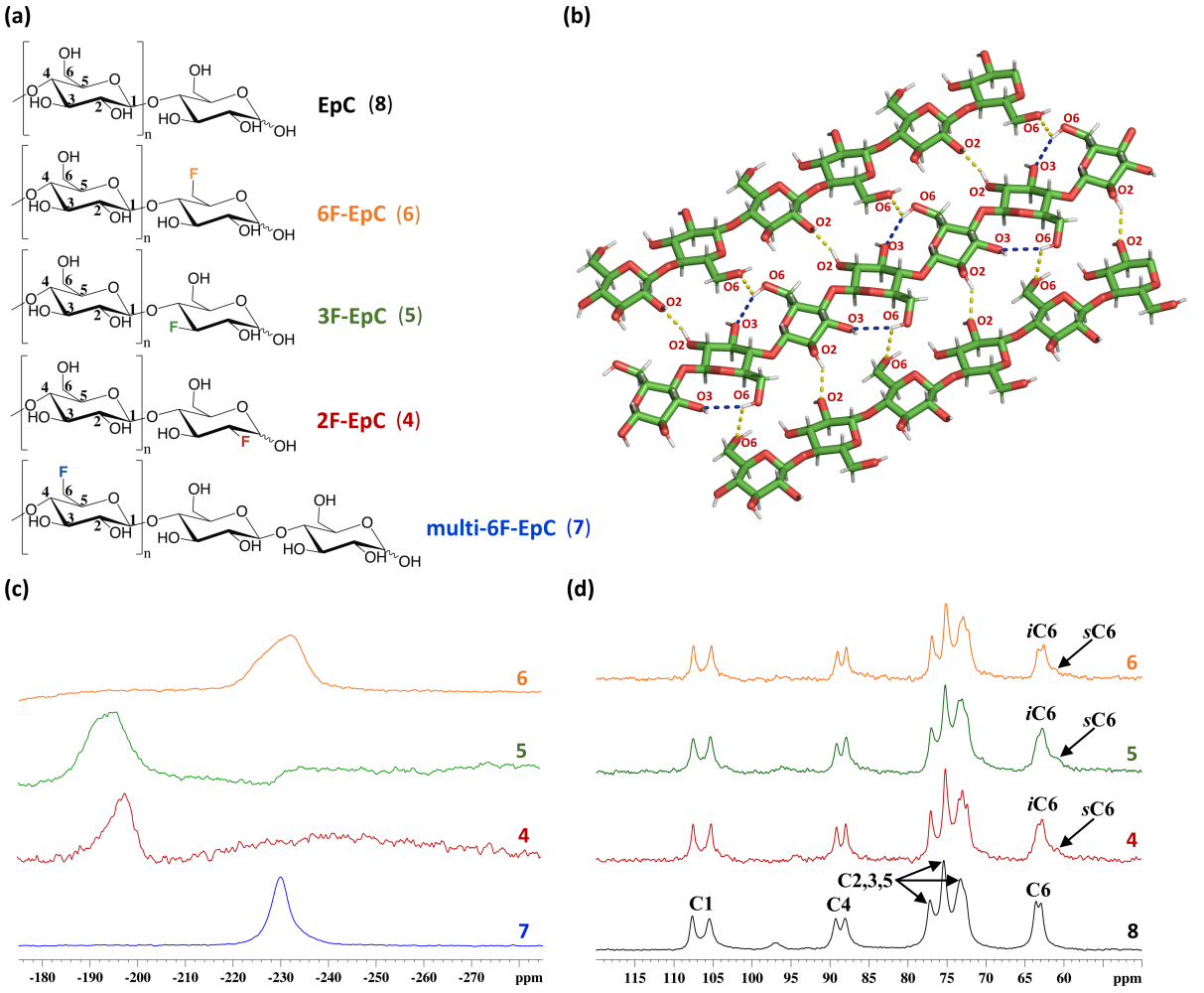


Figure 5. (a) Chemical structures of non-modified EpC (**8**), monofluorinated 2F- (**4**), 3F- (**5**) and 6F-EpC (**6**), and multi-6F-EpC (**7**). **(b)** 3D model of the crystalline packing of cellulose II allomorph based on the origin-center-origin (o-c-o) chains.[39] The O3-O6 intra-chain (blue dashes) and O2-O2 and O6-O6 inter-sheet (yellow dashes) hydrogen bonds are shown. It should be noted that the substitution of all -OH groups at C6 with fluorine atoms precludes the formation of O6-O6 inter-sheet hydrogen bonds during self-assembly. Note: the intra-chain hydrogen bonds are only shown for the center chain for simplicity.[39] **(c)** Direct detection ^{19}F MAS NMR spectra of multi-6F-EpC (**7**, black) and 2F- (**4**, red), 3F- (**5**, green) and 6F-EpC (**6**, orange) powders, acquired at 60 kHz MAS rate and 800 MHz ^{19}F frequency (20 T magnetic field). **(d)** ^{1}H-^{13}C CP/MAS NMR spectra of EpC powder (**8**, black) acquired at 10 kHz MAS rate, and 2F- (**4**, red), 3F- (**5**, green) and 6F-EpC (**6**, orange) 10 wt% dispersions acquired at 6 kHz MAS, and 100 MHz ^{13}C frequency.

To characterise the structural organisation of 2F-EpC (**4**), 3F-EpC (**5**), 6F-EpC (**6**) and multi-6F-EpC (**7**) materials at the molecular level, ^{1}H-^{13}C CP/MAS experiments were carried out. Each type of cellulose allomorph presents a characteristic ^{13}C NMR fingerprint.[53,54] Monofluorinated EpCs (**4**, **5** and **6**) 10 wt% dispersions showed the characteristic cellulose II ^{1}H-^{13}C CP fingerprint, typical of non-modified EpC (**8**) (Figure 5d). The only noticeable difference was the presence of a broad peak at *ca.* 61 ppm, which is characteristic of a surface/disordered population of C6 (*s*C6, Figure 5d).[54,55] Hence, the peak at 63 ppm represents the interior/ordered domains of C6 (*i*C6, Figure 5d). The *s*C6 broad peak is typically observed in bacterial cellulose (BC), which consists of cellulose particles containing both I_α and I_β crystalline domains and disordered regions.[55] Indeed, surface/disordered and interior/ordered domains are typically found in nanocrystalline cellulose,[54] bacterial cellulose[55] and plant cell walls.[56] We note that the presence of a fluorine atom substituting the 3-hydroxyl group of glucose might affect the formation of the characteristic O3-O6 intra-chain hydrogen bond between adjacent glucose residues in cellulose II allomorph (Figure 5b). On the other hand, the formation of the O2-O2 and O6-O6 inter-sheet hydrogen bonds of cellulose II would be affected in 2F- and 6F-EpC, respectively (Figure 5b). Spectral deconvolution of the *s*C6 and *i*C6 peaks of ^{1}H-^{13}C CP/MAS NMR spectra enabled us to estimate the relative surface area (RSA) of the EpC particles constituting each material (†ESI Figure S11, Table S3). RSA values of about 16-23 % were obtained for the three monofluorinated EpCs, with 3F-EpC showing the highest value (23 %; †ESI Table S3). Although the ^{1}H-^{13}C CP/MAS NMR experiments were not fully quantitative, similar RSA values have been determined before for BC.[55]

The ^{1}H-^{13}C CP/MAS NMR spectrum of multi-6F-EpC (**7**) showed a pattern of peaks that does not correspond to either cellulose I, II or III allomorphs (†ESI Table S4, S5). This is evidenced by the appearance of C1 (105.9 ppm) as a singlet peak in **7**, which is a singlet in cellulose type I_α and III_I, and a doublet in cellulose type I_β, II and and III_{II} (†ESI Table S4).[47,48] The small shoulder observed at *ca.* 107 ppm cannot be assigned unambiguously to a specific structural feature. The different multiplicity of the peak corresponding to C1 indicates the presence of only one non-equivalent anomeric carbon per unit cell in multi-6F-EpC (Figure 6a, †ESI Figure S12). This is different from non-modified EpC, which shows a doublet for C1 (Figure 5d) due to the presence of two non-equivalent anomeric carbons per unit cell. Also, the pattern of ^{13}C chemical environments in multi-6F-EpC does not fully match any of the cellulose structures reported so far (†ESI Table S4). Hence, our solid-state NMR data demonstrate that the inter-chain interactions in multi-6F-EpC are different from non-modified EpC or any other cellulose-like structure.

Importantly, the PXRD pattern of multi-6F-EpC (**7**, Figure 3) does not correspond to any cellulose allomorphs reported so far (†ESI Table S1). Hence, **7** assembles into a crystalline organisation which is unprecedented for a cellulose-type material. The formation of this novel structural motif is also supported by the new features observed in the Raman spectra, which do not correspond to either cellulose I or II (Figure 3, †ESI Table S4).

The combination of fast MAS ^{1}H-^{13}C CP, low MAS ^{1}H,^{19}F-decoupled ^{19}F-^{13}C CP, water-polarisation transfer (WPT) solid-state NMR and ^{13}C, COSY and HSQC solution NMR experiments enabled the assignment of the ^{13}C spectrum of multi-6F-EpC (**7**, Figure 6) to be made. ^{1}H,^{19}F-decoupled ^{19}F-^{13}C CP experiments enabled the assignment of C6, C4 and C5 peaks of the fluorinated residues (Figure 6a). The highest intensity peak was assigned to C6 (83.8 ppm), as it is the carbon atom closest to 6F (1.3 Å). The peak at 73.1 ppm corresponds to C4 and C5 sites, based on their proximity to fluorine (Figure 6a, †ESI Figure S12), while C2 and C3 are too far away to cross-polarise from fluorine effectively. ^{13}C DEPT135, COSY and

HSQC solution NMR experiments confirmed this assignment (†ESI Figure S13), with the methylene carbons of the fluorinated (C6) and non-fluorinated (C6*) glucose units appearing in antiphase with respect to the CH carbons (Figure 6a). Importantly, the ^{13}C peaks at 81.9 and 73.1 ppm observed in the CP spectrum did not appear on the ^{13}C DEPT135 nor ^{1}H-^{13}C HSQC solution NMR experiments carried out for a diluted dispersion of multi-6F-EpC (**7**). Hence, these peaks most likely correspond to the immobile interior carbons (*i*C6 and *i*C2,3,4,5, respectively) that are too broad to be detectable by solution NMR. The solution-NMR-observed C6 and C2,3,4,5 peaks were therefore assigned to surface/disordered domains (*s*C6 and *s*C2,3,4,5, respectively). The assignment of *s*C6 and *i*C6 was further validated by water polarisation transfer CP (WPT-CP) NMR experiments (Figure 6b).[57] The peak intensity in WPT-CP experiments depends on the distance and relative mobility of bound water at the particle surface and the number of interacting water molecules at a particular site. Hence, peaks corresponding to surface domains will show faster WPT growth at short mixing times than interior domains, as we have recently observed for BC.[55] At sufficiently long mixing times, WPT become homogeneous for both surface and interior domains due to the efficient spin diffusion. Figure 6b shows the WPT factors for a 25 wt% dispersion of multi-6F-EpC (**7**) at 16 ms mixing time (under our experimental conditions, homogenisation of surface-interior water polarisation transfer is achieved around 200 ms). A much higher WPT factor was observed for the *s*C6 (83.8 ppm) compared to the *i*C6 peak (81.8 ppm), confirming the assignment of *s*C6 and *i*C6 peaks to surface and interior domains, respectively (Figure 6b). Also, *s*C2,3,4,5 showed higher WPT compared to *i*C2,3,4,5 (Figure 6b), in agreement with solution NMR data where the *s*C2,3,4,5 and *i*C2,3,4,5 peaks are visible and invisible, respectively (Figure 6a). Spectral deconvolution of *s*C6 and *i*C6 peaks of the ^{1}H-^{13}C CP spectrum acquired at 60 kHz, indicated that multi-6F-EpC (**7**) presents an RSA of ~54 % (†ESI Figure S11, Table S3). Although this experiment was not fully quantitative, the RSA determined for **7** is similar to what we have reported before for nanocrystalline cellulose.[58]

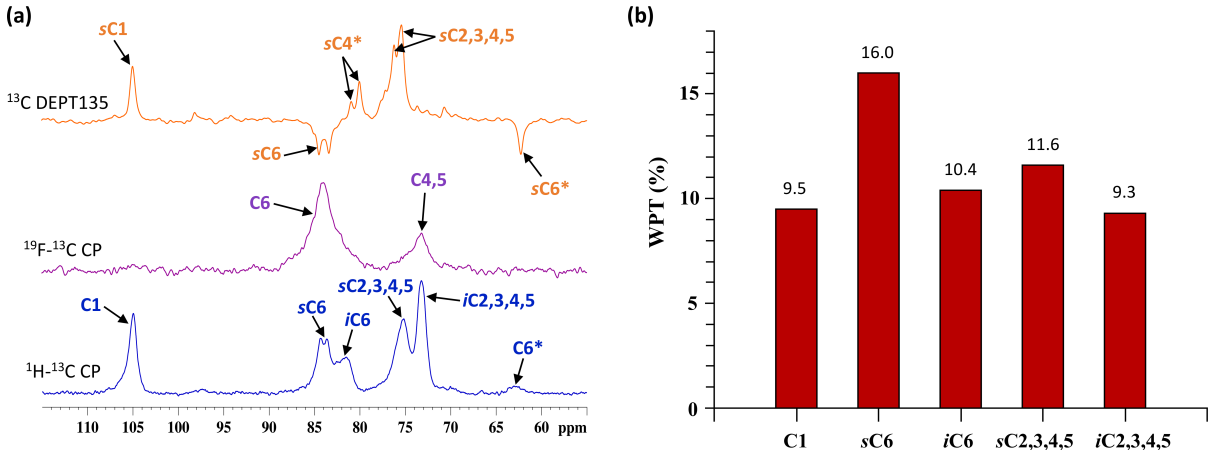


Figure 6. (a) ^{1}H-^{13}C CP (blue) and ^{1}H,^{19}F-decoupled ^{19}F-^{13}C CP (purple) NMR spectra of multi-6F-EpC (**7**) powder acquired at 60 and 15 kHz MAS spinning, respectively, and 212.5 MHz ^{13}C frequency. The ^{13}C DEPT135 spectrum of a 1 wt% dispersion of multi-6F-EpC (**7**) in D_2O is shown for comparison (orange). * Low intensity peaks corresponding to the non-fluorinated glucose units of **7** at the reducing terminal of each cellodextrin chain. **(b)** Bar graph showing the water polarisation transfer (WPT) factors, in percentage, determined for each carbon peak of multi-6F-EpC 25 wt% hydrogel using a mixing time of 16 ms. The higher WPT factor observed for *s*C6 and *s*C2,3,4,5 compared to *i*C6 and *i*C2,3,4,5 demonstrates the increased solvation of the former, hence being assigned to surface domains. The individual WPT values appear on top of each bar.

To summarise, we have demonstrated that the presence of multiple 6-deoxy-6-fluoro glucose residues precludes the formation of the cellulose type II crystallinity and hydrogen bond patterns that defines EpC.[39] To understand this at the molecular level, we should note that two different types of chains (centre, c, and origin, o) leading to three different types of hydrogen bond patterns (o-o-o, c-c-c and o-c-o) define the inter-chain interactions of the EpC cellulose II structure. In particular, the O2-H···O6, O6-H···O2, O6-H···O6 and O2-H···O2 intermolecular hydrogen bonds characterise the cellulose II packing (Figure 5b, †ESI Figure S15). Thus, the presence of multiple 6-deoxy-6-fluoro-glucose residues per oligosaccharide chain will most likely prevent the formation of the O6-H···O6 and O6-H···O2 hydrogen bonds. In addition, the weaker character of the C-F···H-O hydrogen bonds and the partially hydrophobic nature of fluorine will probably affect the formation of the O2-H···O6 hydrogen bonds, leading to the formation of different hydrogen bond interactions and/or partial structural disorder.

The unprecedented PXRD pattern of multi-6F-EpC clearly shows that this material presents a crystalline organisation different from any known cellulose-like material (Figure 3). In addition, the Raman spectra confirm that multi-6F-EpC long-range ordering does not correspond to either cellulose I or II (Figure 4). This, together with the solid-state NMR data showing different inter-chain interactions, indicates that multi-6F-EpC assembles into a novel crystalline packing defined by a distinct pattern of hydrogen bond interactions never reported before for a cellulose-like structure. In addition, both the PXRD and NMR data show that the multi-6F-EpC particle network present both ordered and disordered domains.

Conclusion

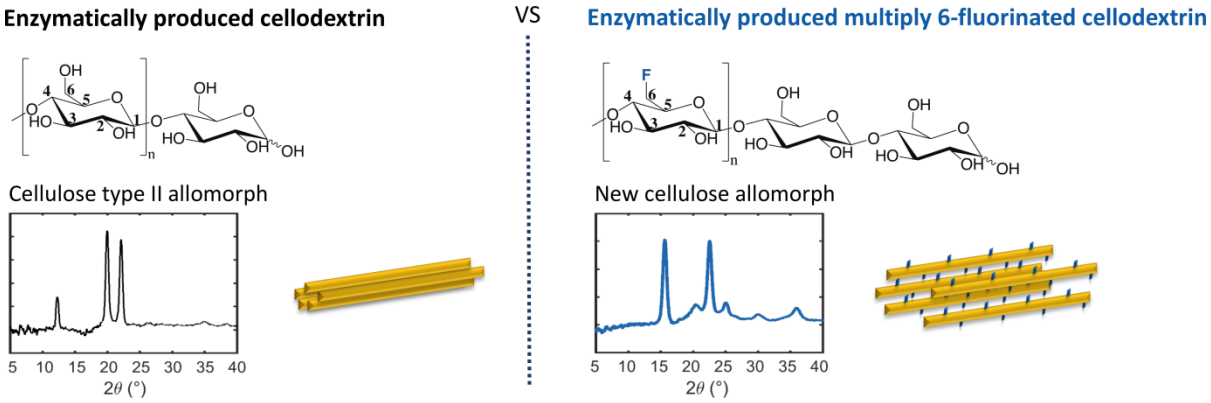
In the current study, we have demonstrated the enzymatic incorporation of single and multiple fluorinated glucose residues into cellodextrin chains. The OH to F substitution is tolerated by the cellodextrin phosphorylase, albeit at low efficiency. Nonetheless, we were able to produce selectively fluorinated cellodextrins averaging ca DP 9 in size, which self-assemble into crystalline materials. Singly fluorinated cellodextrins display structural features reminiscent of cellulose II, as judged by solid-state NMR, powder X-ray diffraction and Raman spectroscopy. In contract, multiply 6-fluorinated cellodextrin gave rise to a new allomorph, not previously reported for either native celluloses or cellulose-like materials. Advanced solid-state NMR methods have enabled the detailed characterisation of these novel materials, deciphering the water exposed and interior chemical environments for different carbon sites. Our findings highlight the considerable potential of chemoenzymatic synthesis for generating novel glycomaterials of controlled molecular structure and morphology.

Acknowledgements

We thank the GelEnz consortium, which is funded by EPSRC (Grant Research Number: IUK 59000 442149). Work at the John Innes Centre is supported by the UK BBSRC Institute Strategic Program on Molecules from Nature - Products and Pathways [BBS/E/J/000PR9790] and the John Innes Foundation; the BBSRC. BBSRC, EPSRC, and Innovate UK: IBCatalyst (Grant BB/M02903411). The UK 850 MHz solid-state NMR Facility used in this research was funded by EPSRC and BBSRC (contract reference PR140003), as well as the University of Warwick including via part funding through Birmingham Science City Advanced Materials Projects 1 and 2 supported by Advantage West Midlands (AWM) and the European Regional Development Fund (ERDF). The Engineering and Physical Sciences Research Council (EPSRC) is acknowledged for provision of financial support (EP/N033337/1) for J.C.M.-G.,

J.A. and Y.Z.K. We are also grateful for UEA Faculty of Science NMR facility. V.G. would like to acknowledge the support of BBSRC Norwich Research Park Bioscience Doctoral Training Grant (BB/M011216/1). Additional research data supporting this publication are available as electronic supplementary files at the following link: XXX.

Graphical abstract



References

1 Y. Habibi, *Chem. Soc. Rev.*, 2014, **43**, 1519–1542.
2 A. D. French and H. J. Kim, in *Cotton Fiber: Physics, Chemistry and Biology*, ed. D. D. Fang, Springer International Publishing, Cham, 2018, pp. 13–39.
3 S. J. Eichhorn, A. Dufresne, M. Aranguren, N. E. Marcovich, J. R. Capadona, S. J. Rowan, C. Weder, W. Thielemans, M. Roman, S. Renneckar, W. Gindl, S. Veigel, J. Keckes, H. Yano, K. Abe, M. Nogi, A. N. Nakagaito, A. Mangalam, J. Simonsen, A. S. Benight, A. Bismarck, L. A. Berglund and T. Peijs, *J. Mater. Sci.*, 2010, **45**, 1–33.
4 T. Abitbol, A. Rivkin, Y. Cao, Y. Nevo, E. Abraham, T. Ben-Shalom, S. Lapidot and O. Shoseyov, *Curr. Opin. Biotechnol.*, 2016, **39**, 76–88.
5 A. Dufresne, *Curr. For. Reports*, 2019, **5**, 76–89.
6 D. Klemm, E. D. Cranston, D. Fischer, M. Gama, S. A. Kedzior, D. Kralisch, F. Kramer, T. Kondo, T. Lindström, S. Nietzsche, K. Petzold-Welcke and F. Rauchfuß, *Mater. Today*, 2018, **21**, 720–748.
7 S. J. Eichhorn, *Soft Matter*, 2011, **7**, 303–315.
8 T. Heinze, O. A. El Seoud and A. Koschella, in *Cellulose Derivatives*, 2018, pp. 259–292.
9 A. C. Genix and J. Oberdisse, *Soft Matter*, 2018, **14**, 5161–5179.
10 C. Dannert, B. T. Stokke and R. S. Dias, *Polym.*, 2019, **11**, 275.
11 B. Shiro, *Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci.*, 2007, **83**, 215–247.
12 X. Pérez, M. Faijes and A. Planas, *Biomacromolecules*, 2011, **12**, 494–501.
13 E. Okamoto, T. Kiyosada, S. I. Shoda and S. Kobayashi, *Cellulose*, 1997, **4**, 161–172.
14 A. Makino, M. Ohmae and S. Kobayashi, *Macromol. Biosci.*, 2006, **6**, 862–872.
15 J. I. Kadokawa, *Chem. Rev.*, 2011, **111**, 4308–4345.
16 S. Kobayashi and A. Makino, *Chem. Rev.*, 2009, **109**, 5288–5353.
17 E. C. O'Neill and R. A. Field, *Carbohydr. Res.*, 2015, **403**, 23–37.
18 G. Pergolizzi, S. Kuhaudomlarp, E. Kalita and R. A. Field, *Protein Pept. Lett.*, 2017, **24**, 696–709.
19 J. I. Kadokawa, *Catalysts*, 2018, **8**, 473.
20 J. I. Kadokawa, *Pure Appl. Chem.*, 2018, **90**, 1045–1054.
21 K. Loos and J. Kadokawa, in *Enzymatic Polymerization towards Green Polymer Chemistry*, eds. S. Kobayashi, H. Uyama and J. Kadokawa, Springer Singapore, Singapore, 2019, pp. 47–87.

22 M. Hiraishi, K. Igarashi, S. Kimura, M. Wada, M. Kitaoka and M. Samejima, *Carbohydr. Res.*, 2009, **344**, 2468–2473.
23 E. Samain, C. Lancelon-Pin, F. Férigo, V. Moreau, H. Chanzy, A. Heyraud and H. Driguez, *Carbohydr. Res.*, 1995, **271**, 217–226.
24 E. C. O'Neill, G. Pergolizzi, C. E. M. Stevenson, D. M. Lawson, S. A. Nepogodiev and R. A. Field, *Carbohydr. Res.*, 2017, **451**, 118–132.
25 Y. Yataka, T. Sawada and T. Serizawa, *Chem. Commun.*, 2015, **51**, 12525–12528.
26 J. Wang, J. Niu, T. Sawada, Z. Shao and T. Serizawa, *Biomacromolecules*, 2017, **18**, 4196–4205.
27 Y. Yataka, T. Sawada and T. Serizawa, *Langmuir*, 2016, **32**, 10120–10125.
28 T. Nohara, T. Sawada, H. Tanaka and T. Serizawa, *Langmuir*, 2016, **32**, 12520–12526.
29 T. Tyrikos-Ergas, G. Fittolani, P. H. Seeberger and M. Delbianco, *Biomacromolecules*, 2020, **21**, 18–29.
30 D. O'Hagan, *Chem. Soc. Rev.*, 2008, **37**, 308–319.
31 K. P. Nartowski, D. Malhotra, L. E. Hawarden, J. Sibik, D. Iuga, J. A. Zeitler, L. Fábián and Y. Z. Khimyak, *Angew. Chemie*, 2016, **128**, 9050–9054.
32 C. Li, E. A. Lutz, K. M. Slade, R. A. S. Ruf, G. F. Wang and G. J. Pielak, *Biochemistry*, 2009, **48**, 8578–8584.
33 J. R. Pinney, G. Melkus, A. Cerchiari, J. Hawkins and T. A. Desai, *ACS Appl. Mater. Interfaces*, 2014, **6**, 14477–14485.
34 T. Diercks, A. S. Infantino, L. Unione, J. Jiménez-Barbero, S. Oscarson and H. J. Gabius, *Chem. - A Eur. J.*, 2018, **24**, 15761–15765.
35 V. López Durán, P. A. Larsson and L. Wågberg, *Carbohydr. Polym.*, 2018, **182**, 1–7.
36 N. Kasuya, K. Iiyama, G. Meshitsuka and T. Okano, *Carbohydr. Polym.*, 1997, **34**, 229–234.
37 Y. Yu, T. Tyrikos-Ergas, Y. Zhu, G. Fittolani, V. Bordoni, A. Singhal, R. J. Fair, A. Grafmüller, P. H. Seeberger and M. Delbianco, *Angew. Chemie - Int. Ed.*, 2019, **58**, 13127–13132.
38 Y. Yu, S. Gim, D. Kim, Z. A. Arnon, E. Gazit, P. H. Seeberger and M. Delbianco, *J. Am. Chem. Soc.*, 2019, **141**, 4833–4838.
39 P. Langan, Y. Nishiyama and H. Chanzy, *J. Am. Chem. Soc.*, 1999, **121**, 9940–9946.
40 A. Makino, J. Sakamoto, M. Ohmae and S. Kobayashi, *Chem. Lett.*, 2005, **35**, 160–161.
41 B. Nidetzky, C. Eis and M. Albert, *Biochem. J.*, 2000, **351**, 649–659.
42 J. S. Zhu, N. E. McCormick, S. C. Timmons and D. L. Jakeman, *J. Org. Chem.*, 2016, **81**, 8816–8825.
43 T. Serizawa, M. Kato, H. Okura, T. Sawada and M. Wada, *Polym. J.*, 2016, **48**, 539–544.
44 P. Langan, Y. Nishiyama and H. Chanzy, *Biomacromolecules*, 2001, **2**, 410–416.
45 Y. Nishiyama, J. Sugiyama, H. Chanzy and P. Langan, *J. Am. Chem. Soc.*, 2003, **125**, 14300–14306.
46 Y. Nishiyama, P. Langan and H. Chanzy, *J. Am. Chem. Soc.*, 2002, **124**, 9074–9082.
47 M. Wada, L. Heux, A. Isogai, Y. Nishiyama, H. Chanzy and J. Sugiyama, *Macromolecules*, 2001, **34**, 1237–1243.
48 M. Wada, L. Heux, Y. Nishiyama and P. Langan, *Biomacromolecules*, 2009, **10**, 302–309.
49 M. Wada, H. Chanzy, Y. Nishiyama and P. Langan, *Macromolecules*, 2004, **37**, 8548–8555.
50 K. Schenzel, H. Almlöf and U. Germgård, *Cellulose*, 2009, **16**, 407–415.
51 M. Tammer, *Colloid Polym. Sci.*, 2004, **283**, 235–235.
52 A. D. French, G. P. Johnson, A. M. Kelterer and G. I. Csonka, *Tetrahedron Asymmetry*, 2005, **16**, 577–586.
53 A. Mittal, R. Katahira, M. E. Himmel and D. K. Johnson, *Biotechnol. Biofuels*, 2011, **4**, 1–16.
54 P. T. Larsson, K. Wickholm and T. Iversen, *Carbohydr. Res.*, 1997, **302**, 19–25.
55 J. C. Muñoz-García, K. R. Corbin, H. Hussain, V. Gabrielli, T. Koev, D. Iuga, A. N. Round, D. Mikkelsen, P. A. Gunning, F. J. Warren and Y. Z. Khimyak, *Biomacromolecules*, 2019, **20**, 4180–4190.
56 T. Wang and M. Hong, *J. Exp. Bot.*, 2016, **67**, 503–514.
57 P. B. White, T. Wang, Y. B. Park, D. J. Cosgrove and M. Hong, *J. Am. Chem. Soc.*, 2014, **136**, 10399–10409.

58 R. Nigmatullin, R. Harniman, V. Gabrielli, J. C. Muñoz-García, Y. Z. Khimyak, J. Angulo and S. J. Eichhorn, *ACS Appl. Mater. Interfaces*, 2018, **10**, 19318–19322.