Structural and electronic determinants of lytic polysaccharide monooxygenase reactivity on polysaccharide substrates

T.J. Simmons, K.E.H. Frandsen, L. Ciano, T. Tryfona, N. Lenfant, J.C. Poulsen, L.F.L. Wilson, T. Tandrup, M. Tovborg, K. Schnorr, K.S. Johansen, B. Henriassat, P.H. Walton, L Lo Leggio & P. Dupree

Lytic polysaccharide monooxygenases (LPMOs) are industrially important copper-dependent enzymes that oxidatively cleave polysaccharides. Here we present a functional and structural characterization of two closely related AA9-family LPMOs from Lentinus similis (LsAA9A) and Collariella virescens (CvAA9A). LsAA9A and CvAA9A cleave a range of polysaccharides, including cellulose, xyloglucan, mixed-linkage glucan and glucomannan. LsAA9A additionally cleaves isolated xylan substrates. The structures of CvAA9A and of LsAA9A bound to cellulolic and non-cellulosic oligosaccharides provide insight into the molecular determinants of their specificity. Spectroscopic measurements reveal differences in copper co-ordination upon the binding of xylan and glucans. LsAA9A activity is less sensitive to the reducing agent potential when cleaving xylan, suggesting that distinct catalytic mechanisms exist for xylan and glucan cleavage. Overall, these data show that AA9 LPMOs can display different apparent substrate specificities dependent upon both productive protein–carbohydrate interactions across a binding surface and also electronic considerations at the copper active site.
The need for sustainable sources of energy and materials has spurred significant research efforts towards a greater understanding of the biological catabolism of lignocellulose, the world’s most abundant source of renewable material and bioenergy. The inherent recyclability of lignocellulose, however, is one of the major barriers to the utilization of biomass. This recyclability is a consequence of both the heterogeneous composition and the often semi-crystalline association of the polymers. In addressing the problem of recyclability, multiple potential means have been proposed and assessed, including chemical, mechanical and enzymatic methods. Advances in enzyme cocktail formulations that accelerate the saccharification step of cell wall breakdown, in particular the inclusion of the lytic polysaccharide mono-oxygenases (LPMOs), are helping cellulosic-ethanol biorefineries move towards both commercial and environmental viability.

LPMOs are reducing agent- and O2-dependent copper metalloenzymes now classified as auxiliary activity families AA9-AA11 and AA13. Extensive spectroscopic and structural studies on LPMOs have shown that the enzyme’s active site contains a single copper ion, which is coordinated by the amino terminus nitrogen atom, by a side chain nitrogen atom of the N-terminal histidine, and by the side chain nitrogen atom of an additional histidine, in a structural motif known as the histidine brace. What is distinctive about LPMOs is that they oxidatively rather than hydrolytically cleave polysaccharides producing saccharides with oxidized ends. LPMOs augment the action of other polysaccharide-degrading enzymes, and accordingly much research attention is devoted to a greater understanding of the enzymatic mechanism and the range of LPMO saccharide substrates.

It was first shown that LPMOs could boost the action of cellulases on cellulose and chitin by increasing the rate of hydrolysis. But LPMOs are now known to act on several crystalline substrates such as chitin, cellulose and retrograded starch. Later, enzymes with activity against non-crystalline and oligomeric structures were identified. Furthermore, fungal AA9 LPMOs have been shown to be active on soluble substrates such as xyloglucan, mixed-linkage glucan and glucomannan, and on cellulose-bound xylan. A LPMO active on isolated xylan has not been reported; however, this range of reported substrates will likely grow. The large number and sequence diversity of LPMOs that individual fungi maintain, and their disparate expression profiles when the fungi are grown on different polysaccharide substrates, signal that AA9 LPMOs do have distinct, and functionally significant, polysaccharide substrate specificities, although some evolutionary diversity of LPMOs likely arises through their use of different reducing systems.

The root causes of LPMO substrate specificity remain poorly understood. This is because LPMO chemistry is a subtle and complex combination of structural and electronic factors, both of which must be taken into account when developing an understanding of the mechanism of action. The structure–function relationship of substrate specificity and regioselectivity has been recently reviewed. Insight into LPMO substrate binding can be gained from the structures of LPMOs and combined structural and spectroscopic studies of LPMOs in contact with substrate. Recent ITC, NMR and docking studies of an AA9 LPMO from Neurospora crassa in contact with oligosaccharides revealed that more extended substrates had significantly higher binding affinities. This is in accord with a multi-point interaction of the substrate with the LPMO surface where the surface loops in some LPMOs remote from the active site enhance binding affinity.

The study also showed that a single cellohexaose (Cell6) chain likely spans the copper active site from the −3 to +3 or −2 to +4 subsites (subdivisions of the binding cleft numbered relative to the site of cleavage), in which the L3 loop (important for interactions with the +3/+4 subsites) and the LC loop (important for binding to approximately −4 subsite) lie at somewhat extended distances from the copper active site. Detailed insight into an AA9 LPMO–substrate interaction came from the first crystal structures of LPMO:oligosaccharide complexes: Lentinus similis AA9A (LsAA9A) bound to Cell5 and Cell3, Cell6 was shown to bind at subsites −4 to +2 via interactions with aromatic residues, the N-terminal His and a conserved Tyr as well as a number of hydrogen-bonds with other residues in a contoured binding surface on the LPMO. The +2 glucosyl residue exhibits a set of well-defined hydrogen-bonding interactions with amino-acid side chains (Asn28, His66 and Asn67) essentially locking this residue into a fixed position with respect to the active site.

Electronic factors around the active site also play a key role in determining the reactive mechanism. Changes in the electronic structure of the copper ion, an important factor in the ability of the copper ion to activate O2, occur upon substrate binding to LsAA9A. Furthermore, in an illustration of the complexity of substrate–LPMO interaction and the subtle interplay of electronic and structural factors, Cell5 is bound synergistically with an exogenous ligand on the copper ion. It is likely that the oxidative mechanism adopted by LPMOs can proceed via one or more of several different routes, the determinants of which depend on the varying extents of the substrate, the reducing agent reducing the potential and the positioning of the substrate on the LPMO surface. For instance, the means by which electrons are donated to the LPMO active site modulate the apparent range of reactivity.
The detailed molecular and electronic insights of the LPMO–substrate interaction afforded by combined biochemical, X-ray diffraction and electron paramagnetic resonance (EPR) spectroscopic studies can significantly enhance our understanding of LPMO reactivity. We report herein a study into the principal structural and electronic factors of the reactivity of two AA9 LPMOs with a range of substrates. Through X-ray crystal structures studies of LsAA9A with bound substrates, we illustrate how binding cleft interactions dictate the site of polysaccharide attack. Through comparison with the CvAA9A structure, also determined here, we suggest some structural determinants of specificity for the two enzymes. LsAA9A is active on isolated xylan, but this activity is associated with a distinct low sensitivity to reducing agent potential and a different copper co-ordination at the active site, which together reveal an alternative mechanistic pathway for LPMO action on this substrate. These data show how AA9 LPMO substrate cleavage is dependent upon both productive protein:carbohydrate interactions across a binding surface and also electronic considerations at the active site.

Results

CvAA9A is an AA9 LPMO active on cello-oligosaccharides. To help understand the basis of AA9 substrate specificity, we searched for enzymes related to LsAA9A that might also cleave soluble oligosaccharides. Because LPMOs exhibit high variability in their C-termini, we performed a large-scale alignment of LPMO protein sequences using their N-terminal portion. We selected 98 AA9 sequences that were highly similar in their N-terminal half to LsAA9A and 326 sequences that were highly similar to TaAA9A in the same region. TaAA9A was used for comparison purposes, as it does not show the ability to cleave soluble oligosaccharides. After adding the sequences of 20 AA9 enzymes studied in the literature, a distance tree was built with the resulting 444 AA9 sequences (Fig. 1a, Supplementary Table 1). The tree clearly places LsAA9A and TaAA9A in distinct clades. From the LsAA9A clade, we identified an LPMO from Collariella virescens (CvAA9A; 46% sequence similarity to LsAA9A) that lacks some residues observed by Frandsen et al. as being involved in enzyme–substrate interactions (Supplementary Fig. 1). All three of the substrate +2 substrate-binding residues in LsAA9A (Asn28, His66 and Asn67) are different in CvAA9A (Thr28, Arg67 and Val68) (Supplementary Fig. 1). To study the activity of CvAA9A, the enzyme was expressed in Aspergillus oryzae and successfully purified from the fermentation broth. (Note that the expression in this fungal host preserves the natural side-chain methylation at the N-terminal histidine, in contrast to fungal LPMO expression in Pichia pastoris and bacterial systems.) On phosphoric acid-swollen cellulose (PASC), CvAA9A produced a range of cello-oligosaccharides (Fig. 1b). The cello-oligosaccharide product profile of CvAA9A was similar to that of LsAA9A and notably shorter than those produced by TaAA9A. Indeed, CvAA9A readily degraded Cellul-(2-aminobenzenamide) (Cellc−2AB) using a C4-oxidising mechanism to yield Cellul and oxidized Cellul−2AB (Supplementary Fig. 2), like LsAA9A but unlike TaAA9A. Therefore, the distance relationships between the three enzymes, as measured using the N-terminal comparison method above, mirror the similarities in activities of the enzymes.

Position-specific cleavage of a range of hemicelluloses. We next determined whether the LsAA9A and CvAA9A enzymes are active on a range of β-(1 → 4)-d-glucan-related polysaccharides (Fig. 2, Table 1). Mixed-linkage glucan (MLG) is a β-d-glucan in which three to four (1 → 4)-linked residues (Cellul, Cellul) are separated by single (1 → 3) bonds, glucomannan has a backbone randomly composed of β-(1 → 4)-d-glucosyl and β-(1 → 4)-d-mannosyl residues, xylolglucan is a β-(1 → 4)-d-glucan with α-(1 → 6)-d-xylosyl branches, and xylan is a polymer of β-(1 → 4)-d-xylosyl residues that is similar to β-(1 → 4)-d-glucan but lacks C6 groups (Fig. 2b). Both LsAA9A and CvAA9A showed activity against MLG, glucomannan and xylolglucan, producing a range of oligosaccharide products (Fig. 2a). LsAA9A also showed some activity on xylan, whereas CvAA9A showed no measurable activity on this substrate. No LPMO activity was observed on starch (α-(1 → 4)-d-glucan), laminarin (β-(1 → 3)-d-glucan) or chitin (poly β-(1 → 4)-d-GlcNAC) (Supplementary Fig. 3). Altogether, these activities indicate that both LsAA9A and CvAA9A enzymes only cleave near β-(1 → 4) bonds, and that some variation to the cellulotic β-(1 → 4)-d-glucan, including substitution, linkage and backbone residue, can be accommodated at or near the site of cleavage by both of the enzymes.

To identify precise substrate cleavage sites, we studied the products of both LsAA9A and CvAA9A cleavage of MLG, glucomannan, xylolglucan and xylan (in the case of LsAA9A) polysaccharides by MALDI-ToF MS. Minor double oxidation products were observed, indicating cleavage of these hemicelluloses and PASC using both C1- and C4-oxidising mechanisms (Fig. 3; Supplementary Fig. 4, Table 1). We further investigated the site of attack on these different hemicelluloses using differing
regions (Fig. 3). The inability of each enzyme favours cleaving within Cell4 regions over Cell3 regions (Fig. 3). The inability of LsAA9A and CvAA9A to cleave β-Glc-(1 → 4)-β-Glc-(1 → 3)-β-Glc-(1 → 4)-Glc (G4G3G4G), despite their ability to cleave Cell3 (G4G4G4G) (Supplementary Fig. 5), supports the hypothesis that neither enzyme can cleave at β-(1 → 3) bonds and require substantial β-(1 → 4)-linked regions for cleavage. On glucomannan, we employed High-Performance Anion-Exchange Chromatography (HPAEC) analysis of trifluoroacetic acid (TFA) hydrolysates of digestion products to assess the site of cleavage. Notably, the data indicated that cleavage can occur not only between glucosyl residues, but also with mannose at the +1 or −1 site (Supplementary Fig. 6). In order to deduce site of attack on xylglucan we employed xylglucan DP14-18 oligosaccharides (Supplementary Fig. 7). Inspection of the position of substituted glucose (Glc) in the products indicated that xyloloxyl substitution of Glc at O-6 was accommodated at the −3, −2, −1, +2 and +3 subsites but unsubstituted Glc was always required at subsite +1. In contrast to the LsAA9A and CvAA9A products on polysaccharides, LsAA9A degraded Xyl6-2AB to yield two trimers using solely a C4-oxidising mechanism (Supplementary Fig. 8), analogous to cleavage of Cell6-2AB by both LsAA9A and CvAA9A (Supplementary Fig. 2).

To allow a semi-quantitative determination of the influence of sugar structures on enzyme activity, we probed LsAA9A and CvAA9A cleavage of the soluble Cell6, xylohexaose (Xyl6) and mannohexaoxide (Man6) oligosaccharides (Supplementary Table 4). LsAA9A activity against Cell6 was substantially (~100-fold) better than its activity on Xyl6. Consistent with the absence of activity on glucuronoxylan, CvAA9A activity on Xyl6 was almost undetectable (~1000-fold less than Cell6 activity). Although both enzymes showed activity on glucomannan and can cleave adjacent to mannose, activity was scarcely detectable on Man6 (~10,000-fold less than Cell6 activity), indicating that the enzymes require some Glc residues within a mannan backbone for activity.

Recent results show dependence of the LPMO action on reductant strength.34, 35. We found that cleavages of MLG, glucuronoxylan and xyloloxyl by LsAA9A were sensitive to reducing agent potential, with ascorbate as reductant yielding much higher amount of product (Fig. 4a). In contrast, cleavage of xylan was not sensitive. We corroborated this finding with oligosaccharides, observing that Xyl6 was poorly sensitive to reducing agent potential, unlike cleavage of Cell6, where LsAA9A showed much greater activity with ascorbate than pyrogallol (Fig. 4b).

Table 1 Summary of activity assays on different substrates

| Polysaccharide | LsAA9A Activity | Notes | CvAA9A Activity | Notes |
|---------------|----------------|-------|----------------|-------|
| Cellulose and cello-oligosaccharides | ++ | Activity on both cellulose oligosaccharides and insoluble cellulose material (PASC) | ++ | Activity on both cellulose oligosaccharides and insoluble cellulose material (PASC) |
| MLG | ++ | Pattern suggests β-(1 → 3) bonds accommodated at specific places within active site, but not between −1 and +1. Cleavage can occur with Glc or mannose at −1 or +1. Inactivity on Man6 indicates some O2-C2 hydroxyl orientation needs to be present between −3 and +3. Cleavage occurs with unsubstituted Glc at subsite +1. Xylooloxyl substitution at −3, −2, −1, +2 and +3. Galactosyl-xylooloxyl substitutions can occur at either −2, −1 and/or +3. | ++ | Pattern suggests β-(1 → 3) bonds accommodated at specific places within active site, but not between −1 and +1. Cleavage can occur with Glc or mannose at −1 or +1. Inactivity on Man6 indicates some O2-C2 hydroxyl orientation needs to be present between −3 and +3. Cleavage occurs with unsubstituted Glc at subsite +1. Xylooloxyl substitution at −3, −2, −1, +2 and +3. Galactosyl-xylooloxyl substitutions can occur at either −2, −1 and/or +3. |
| Glucomannan and Man6 | ++ | Activity on both glucuronoxylan, Laminarin and G4G3G4G (MLG oligosaccharide) | ++ | Activity on both glucuronoxylan, Laminarin and G4G3G4G (MLG oligosaccharide) |
| Xyloglucan | ++ | Activity on both Weak Xyl activity compared to Cell6; suggests that, while Glc C6 is not required for activity, it is very important at certain sites, such as +1. | ++ | Activity on both Weak Xyl activity compared to Cell6; suggests that, while Glc C6 is not required for activity, it is very important at certain sites, such as +1. |
| Xylan and Xyl6 | + | Activity on both | +/− | Activity suggests that CVAA9A necessarily cleaves β-(1 → 4) bonds |
| Starch | − | Absence of activity indicates that LLsAA9A necessarily cleaves β-(1 → 4) bonds | − | Absence of activity indicates that CVAA9A necessarily cleaves β-(1 → 4) bonds |
| Laminarin | − | Absence of activity indicates that LLsAA9A either requires O2 interactions or cannot accommodate N-acetyl on amino C2 | − | Absence of activity indicates that CVAA9A either requires O2 interactions or cannot accommodate N-acetyl on amino C2 |

To help understand the structural basis of LPMO attack on different substrates, we employed crystallographic analyses. We report here an LsAA9A:Cell6 complex (Fig. 5a, b; Supplementary Table 4; Table 2), which, owing to a lack of significant substrate contacts protocols. On MLG, we observed in the MALDI data a predominance of DP 4, 7 and 10 oligosaccharides indicating that each enzyme favours cleaving within Cell4 regions over Cell3 regions (Fig. 3). The inability of LsAA9A and CvAA9A to cleave β-Glc-(1 → 4)-β-Glc-(1 → 3)-β-Glc-(1 → 4)-Glc (G4G3G4G), despite their ability to cleave Cell6 (G4G4G4G) (Supplementary Fig. 5), supports the hypothesis that neither enzyme can cleave at β-(1 → 3) bonds and require substantial β-(1 → 4)-linked regions for cleavage. On glucomannan, we employed High-Performance Anion-Exchange Chromatography (HPAEC) analysis of trifluoroacetic acid (TFA) hydrolysates of digestion products to assess the site of cleavage. Notably, the data indicated that cleavage can occur not only between glucosyl residues, but also with mannose at the +1 or −1 site (Supplementary Fig. 6). In order to deduce site of attack on xylglucan we employed xylglucan DP14–18 oligosaccharides (Supplementary Fig. 7). Inspection of the position of substituted glucose (Glc) in the products indicated that xyloloxyl substitution of Glc at O-6 was accommodated at the −3, −2, −1, +2 and +3 subsites but unsubstituted Glc was always required at subsite +1. In contrast to the LsAA9A and CvAA9A products on polysaccharides, LsAA9A degraded Xyl6-2AB to yield two trimers using solely a C4-oxidising mechanism (Supplementary Fig. 8), analogous to cleavage of Cell6-2AB by both LsAA9A and CvAA9A (Supplementary Fig. 2).

To allow a semi-quantitative determination of the influence of sugar structures on enzyme activity, we probed LsAA9A and CvAA9A cleavage of the soluble Cell6, xylohexaose (Xyl6) and mannohexaoxide (Man6) oligosaccharides (Supplementary Table 4). LsAA9A activity against Cell6 was substantially (~100-fold) better than its activity on Xyl6. Consistent with the absence of activity on glucuronoxylan, CvAA9A activity on Xyl6 was almost undetectable (~1000-fold less than Cell6 activity). Although both enzymes showed activity on glucomannan and can cleave adjacent to mannose, activity was scarcely detectable on Man6 (~10,000-fold less than Cell6), indicating that the enzymes require some Glc residues within a mannan backbone for activity.

Recent results show dependence of the LPMO action on reductant strength.34, 35. We found that cleavages of MLG, glucuronoxylan and xyloloxyl by LsAA9A were sensitive to reducing agent potential, with ascorbate as reductant yielding much higher amount of product (Fig. 4a). In contrast, cleavage of xylan was not sensitive. We corroborated this finding with oligosaccharides, observing that Xyl6 was poorly sensitive to reducing agent potential, unlike cleavage of Cell6, where LsAA9A showed much greater activity with ascorbate than pyrogallol (Fig. 4b).
to symmetry-related molecules, is a more faithful depiction of the binding conformation of a single oligosaccharide to \textit{LsAA9A} as compared to the original \textit{LsAA9A}:Cell\textsubscript{6} structure described by Frandsen et al.\textsuperscript{35}. Tyr203 stacking is still a major interaction in \textit{LsAA9A}:Cell\textsubscript{6} but a new hydrogen bond is seen between O6 and Asp150 at subsite -3, and glycosidic torsion angles are closer to ideal values (Supplementary Table 2). Other interacting residues at the negative subsites are Glu148, Arg159 and Ser77 (Fig. 5; Supplementary Table 3). Like the \textit{LsAA9A}:Cell\textsubscript{6} structure\textsuperscript{35}, the main interactions to Cell\textsubscript{6} are a network of hydrogen bonds by Asn28, His66 and Asn67 interacting with O2 and O3 at subsite +2, and the interaction with MeHis1 at subsite +1\textsuperscript{38}.

To understand better how protein structure might influence the similarities and differences in \textit{CvAA9A} and \textit{LsAA9A} substrate cleavage patterns, the X-ray crystal structure of \textit{CvAA9A} was solved (Supplementary Figs. 10 and 11; Supplementary Tables 3 and 4; Table 2). The Cu-coordinating amino-acid residues are MeHis1 and His79 (with equatorial distances to the Cu ranging from 2.0 to 2.1 Å), while a non-coordinating Tyr169 occupies the axial position (2.6–2.8 Å). No exogenous ligands are evident within 3.0 Å of the Cu-ion indicating that the active site is mostly in a photoreduced Cu(I) state. A ‘pocket-water’ is bound in an H-bond network with the amide-nitrogen and oxygen of Asp76 and MeHis1, respectively. The active-site geometry of \textit{CvAA9A} thus closely resembles that of \textit{LsAA9A} (Supplementary Fig. 10c).

However, there are some amino-acid differences in \textit{CvAA9A} compared to \textit{LsAA9A} at subsites +2 and –1. Crystals of \textit{CvAA9A} were soaked with Cell\textsubscript{6} and Cell\textsubscript{6} oligosaccharides but this did not result in any catalytically relevant complex.

\textbf{\textit{LsAA9A}:hemicellulose oligosaccharide structures.} To study the structural determinants of the \textit{LsAA9A} positional specificity of cleavage, a number of \textit{LsAA9A} crystal structures in complex with MLG, glucomannan and xylo-oligosaccharides were solved (see Table 2, Supplementary Fig. 12 and Supplementary Tables 2 and 4 for experimental and crystallographic data and refinement information, hydrogen-bonding interactions between enzyme and ligand, and ligand conformations). Soaking experiments with commercially available xyloglucan fragments failed to produce crystallographic complexes, possibly because the substrate oligosaccharides are large and binding likely to be impeded by crystal contacts.

\textbf{Complexes with MLG tetrasaccharide.} \textit{LsAA9A} crystals were soaked with two different MLG tetrasaccharides, each with a single β-(1→3) linkage: G4G4G3G and G4G3G4G. Interestingly, the \textit{LsAA9A}:G4G4G3G complex did not reveal any β-(1→3) linkages. An apparent Cell\textsubscript{6} substrate appears to be bound from subsite –2 to +2 (Supplementary Fig. 13) giving essentially identical interactions as the –2 to +2 glucosyl residues in the \textit{LsAA9A}:Cell\textsubscript{6} complex. We interpret this result as the β-(1→4)-glucan (Cell\textsubscript{6}) part of the substrate being bound in two overlapping conformations in different asymmetric units from subsites –2 to +1 and –1 to +2, while the β-(1→3)-glucosidic residues are completely disordered in both cases. A structure of \textit{LsAA9A} crystals soaked with G4G3G4G could not be convincingly modelled, further indicating that the enzyme needs at least two consecutive β-(1→4) linkages (a Cell\textsubscript{6} unit) for recognition and efficient binding.

\textbf{Complexes with glucomannan oligosaccharides.} \textit{LsAA9A} crystals were soaked with a mixture of glucomannan oligosaccharides. The resulting difference density was well defined clearly showing glycosyl units occupying subsites –3 to +2, additional density at –4 and some residual density occupying subsite +3 (Fig. 5c, e). Consistent with the activity data, the structure unequivocally showed a mannosyl unit at the +1 subsite, while glucosyl units were clearly observable at –2, –1 and +2 subsites. Moreover the C2 hydroxyl of the mannosyl unit at subsite +1 points towards the face of the imidazole side chain of MeHis1, and the axial water molecule is displaced (Fig. 5d, f). The identity of the glycosyl unit at subsite –3 is ambiguous though best modelled as mannose. The density of the glycosyl unit at subsite –4 is weak and occupies a very similar position as the corresponding unit in the \textit{LsAA9A}:Cell\textsubscript{6} complex, as does the glycosyl unit at subsite –3, due to similar crystal constraints.

\textbf{Complexes with xylo-oligosaccharide.} Although in crystals soaked with Xyl\textsubscript{3} and Xyl\textsubscript{4} the oligosaccharides did not fully span the active site, \textit{LsAA9A}:Xyl\textsubscript{3} crystals revealed very well-defined density from subsites –3 to +2 (Fig. 6a). The oligosaccharide position at subsites –3 to –1 are similar to \textit{LsAA9A}:Cell\textsubscript{6}, but with a translation of about half a pyranose unit in the non-reducing end direction. In contrast, the plane of the xyloside unit at subsite +1 is rotated ~90° compared to the corresponding glucosidic unit, while the xyloside residue binding +2 is rotated ~180° (Fig. 6d; Supplementary Table 2). As a result...
xylose at subsite +1 does not stack with MeHis1, and in fact appears to have no interactions with the enzyme, while the same residues that bind the subsite +2 glucosyl residue in the LsAA9A:Cell5 structure, Asn28, His66 and Asn67, interact here with O1, O5 and O1 of the +2 xyloside residue, respectively (Fig. 6c; Supplementary Table 3). A structure of LsAA9A:Xyl5, Cu (II) determined from a low X-ray dose data collection showed the substrate bound similarly, and revealed a mix of water/Cl\(^-\) in the axial position and a fully occupied equatorial water on the active site copper (Fig. 6b). Thus, in contrast to binding of cellulose or glucomannan oligosaccharides, the axial water was not displaced by binding of Xyl5.

**EPR data suggest alternative LsAA9A substrate binding modes.** We studied substrate binding on both LsAA9A and CvAA9A using EPR spectroscopy (Table 3) to investigate the electronic state of the active site copper upon binding. As has been shown by Frandsen et al.\textsuperscript{32} and Courtade et al.\textsuperscript{34}, the binding affinity of oligosaccharide substrates is significantly affected by the presence of the exogenous ligand on the copper ion. Accordingly, EPR experiments were carried out in both the absence and presence of 200 mM chloride (1.0 M chloride for Xyl6 studies). Furthermore, experiments were carried out at high substrate concentration to maximize substrate binding. For LsAA9A a wide range of substrates was tested. In all cases, the parallel region of the spectra could be modelled with reliable g\(_z\) and |A\(_z\)| values, giving some insight into the electronic nature of the copper ion. Perpendicular values were less reliable due to the second-order nature of the spectra in this region, and are therefore not used in the analysis, although the appearance of superhyperfine coupling to ligands in this region was apparent in some cases (Table 3, Supplementary Fig. 14) and used as an indication of increased metal-ligand covalency in the singly-occupied molecular orbital (SOMO), as previously discussed by Frandsen et al.\textsuperscript{33}. In all cases apart from xylan, the addition of substrate gave perturbations of the Cu spin Hamiltonian parameters similar to that already reported by Frandsen et al.\textsuperscript{33}. In particular, shifts in g\(_z\) values to ca 2.23 (along with the appearance of strong superhyperfine coupling) were seen upon addition of Avicel, glucomannan and xyloglucoman, indicative of chloride coordination to the copper ion in the equatorial position of the copper coordination sphere. These shifts are analogous to those of LsAA9A interacting with Cell5, and PASC33.

In contrast, addition of Xyl6 did not give significant shifts in g\(_z\), but did give perturbations in the |A\(_z\)| value, with the appearance of superhyperfine coupling indicative of a second species different from that formed with Cell6. The EPR spectra of LsAA9A binding to Xyl6 and xylan are indicative of substrate binding to the enzyme (although binding of Xyl6 could be achieved only at high chloride concentrations), but without the chloride occupying the equatorial coordination position on the copper ion, revealing that these substrates drive an electronic state at the copper ion that is different to that of the other substrates. EPR perturbation was seen upon addition of Cell6 to CvAA9A but not with Xyl6, consistent with the observed activity on Cell6 and not Xyl6 (Fig. 7; Supplementary Fig. 15).

**Discussion**

Our understanding of the molecular basis for substrate binding and cleavage has been aided by the recent report of a crystal structure of LsAA9A in complex with Cell6 and Cell5, as well as biochemical and EPR data for LsAA9A on cellulosic substrates\textsuperscript{33}. Here we have extended this biochemical, EPR and structural analysis by using a range of substrates as well as an additional related enzyme, CvAA9A, to provide a better insight into substrate specificity.

Extensive probing of LsAA9A and CvAA9A substrate specificity showed that both cleave a range of cellulosic and non-cellulosic substrates, some of which have been shown for other AA9s\textsuperscript{19, 20, 22, 23, 27, 29, 35, 39–41}. We made a number of important observations. Notably, whereas LPMO activity on cellulosic-associated xylan has been previously observed for MILPOMO9\textsuperscript{23}, we report activity of LsAA9A on xylo-oligosaccharides and isolated xylan; this may have important implications for the use of LPMOs in biotechnological contexts. We also observe that both LsAA9A and CvAA9A are able to cleave glycosidic bonds adjacent to mannosyl residues (Supplementary Fig. 6), which occur interspersed randomly with glycosyl residues in glucomannan, a biochemical observation supported by the LsAA9A:glucomannan oligosaccharide structure, which unambiguously shows a mannosyl residue at subsite +1. We also noticed substrate-specific...
oxidation profiles, namely that LsAA9A and CvAA9A cleaved small oligosaccharides using a C4-oxidising mechanism whereas they cleaved polysaccharides with both C1- and C4-oxidising mechanisms in varying proportions. Assuming a copper-based oxidative species, the similar distances between both C1 and C4 axial protons and the active oxygen species, as noted in Frandsen et al.\(^3\), may allow slight differences in substrate binding to switch the C–H bond that is closest to attack. Substrate binding differences may also subtly alter the electronics at the copper site, which potentially could also favour a specific oxidation site. Oxidation regioselectivity is therefore less likely to be a strong functional constraint. This is in agreement with the presence of C1 and C4 regiospecificity in several clades in the N-terminal sequence similarity tree (Fig. 1). On the other hand, the tree enabled us successfully to predict that LsAA9A and CvAA9A might have similarities in having activity on a range of soluble substrates.

Our investigation highlighted many examples of the way in which substrate specificity and the site of attack on a polysaccharide is dictated by binding cleft interactions with the substrate. For example, LsAA9A’s preference to cleave Cell\(_2\) into Cell\(_3\) (as shown for NcLPMO9C\(^9\)), compared to the product profile of CvAA9A of Cell\(_3\), Cell\(_2\) and Glc (Supplementary Fig. 5), could be attributed to binding cleft interactions at the +2 subsite. This is because, though CvAA9A shares the same fold, active site co-ordination and overall structure with LsAA9A, it lacks all three of the substrate +2 substrate-binding residues in LsAA9A (Asn28, His66 and Asn67) (Supplementary Figs. 1 and 10). This suggests that, while LsAA9A binds Cell\(_2\) from subsite −2 to +2, CvAA9A binds between substrates −3 and +1.

The structural data provide a molecular rationale for how LsAA9A is able to catalyse the unexpected cleavage of mannos-containing bonds. The LsAA9A:glucomannan oligosaccharide crystal structure shows the presence of a mannose (the C2 epimer of Glc), and essentially no Glc, at subsite +1. Talose (C4 epimer of mannose) and galactose (C4 epimer of Glc) arose from the reduction of C4-oxidised cleavage products (Supplementary Fig. 6), also suggesting both mannosyl and glucosyl occupation of

---

**Table 2 Scaled crystallographic data statistics and refinement statistics**

| Data collection | LsAA9A: Cell\(_5\) | LsAA9A: G4G4G3G | LsAA9A: Xyl\(_5\) | LsAA9A: Xyl\(_4\) | LsAA9A: Xyl\(_5\) Cu(II) | LsAA9A: GM | CvAA9A |
|----------------|-------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Synchrotron    | ESF               | MAX-lab         | ESF             | ESF             | ESF             | ESF             | ESF             |
| Beamline       | ID23-1            | 1911-2          | ID23-1          | ID23-1          | ID23-2          | ID23-2          | ID30-B          |
| Wavelength (Å) | 0.97625           | 1.03841         | 0.97625         | 0.97625         | 0.87260         | 0.87260         | 0.96862         |
| Space group    | P\(_4\)2          | P\(_4\)2         | P\(_4\)2         | P\(_4\)2         | P\(_4\)2         | P\(_4\)2         | P\(_4\)2         |
| Cell dimensions a, b, c (Å) | 126.30           | 124.91          | 125.18          | 125.01          | 125.35          | 125.30          | 125.29          |
| α, β, γ (°)   | 90                | 90              | 90              | 90              | 90              | 90              | 90              |
| Resolution (Å) | 50.00-1.75        | (1.80-1.75)\(^a\) | 50.00-2.00      | (2.05-2.00)     | 50.00-1.50      | (1.54-1.50)     | 44.20-1.59      |
| R\(_{meas}\)   | 1.18              | (1.81)          | 0.238           | (1.64)          | 0.114           | (1.67)          | 0.097           |
| I/a            | 11.53             | (1.02)          | 12.31           | (1.85)          | 16.84           | (1.81)          | 17.53           |
| Completeness (%) | 99.7 (99.4)       | 99.9 (99.9)     | 99.9 (99.9)     | 99.7 (99.6)     | 99.8 (100)      | 100 (100)       | 100 (100)       |
| R\(_{cryst}\)  | 0.0135            | 0.0183          | 0.0186          | 0.0174          | 0.0245          | 0.0167          | 0.026           |
| B-factors (Å)  | 21.45             | 23.91           | 25.0           | 25.01           | 21.66           | 21.88           | 8.67            |
| Water          | 32.1              | 35.9            | 36.0            | 52.9            | 45.26           | (8102)          | 21.66           |
| R.m.s. deviations (Å) | 10.36            | 12.00           | 10.41           | 9.64            | 10.51           | (10.82)         | 8.67            |
| Bond angles (°) | 1.5018            | 1.9101          | 1.8982          | 1.7407          | 2.0278          | 1.7393          | 2.3372          | 1.7245          |

\(^a\)Highest-resolution shell is shown in parentheses

\(^b\)Glycosylation (a single N-acetylglucosamine unit) and the active site copper are included in ‘Protein’

---
subsection +1. We have previously described that the glucosyl unit at the +1 subsite in LsAA9A-CelL and CelL complexes interacts with MeHis1 through its β-face, and while glucose can make carbohydrate-aromatic stacking interactions through both faces of the pyranose ring, β-mannose is believed to have absolute preference for interactions through its α-face due to its axial C2-hydroxyl. Nonetheless, determination of the crystal structure of glucomannan fragments with LsAA9A confirmed that this type of interaction takes place, and the mannose residue at the +1 subsites interacts with MeHis through its β-face with an O5-imidazole ring centre distance of 3.6 Å (Fig. 5f) (compared to 3.4–3.5 Å for the cello-oligosaccharide complexes). No similar interactions could be found through a search in the PDB. We did not observe a mannose residue at the −1 subsite in the structure, but it would cause no steric clash and so could be readily accommodated (though it would cause the loss of a hydrogen bond interaction with Ser77).

Although we were unable to obtain a structure with xyloglucan oligosaccharides bound, our observation that both LsAA9A and CvAA9A cleaved xyloglucan DP14–18 oligomers with the sole unsubstituted backbone glucosyl residues at subsite +1 (XXX/XXXXGol; as found for NcLPMO9C19) is consistent with the binding of xyloglucan’s celluloseyl backbone being similar to the binding of cello-oligosaccharides. This would suggest that the LsAA9A could tolerate glucosyl residues with C6 xylosyl substitutions at subsites −1 or +2, but not at +1 where the C6 hydroxymethyl group occludes the copper axial binding site, and displaces the axial water.

LsAA9A under the selected conditions degraded Xyl6 with about 1/100 the efficiency as CelL, while CvAA9A left Xyl6 essentially untouched at all conditions tested (Supplementary Fig. 9). The differences in key amino acids involved in defining the LsAA9A and CvAA9A subsites, particularly the +2 subsite (vide supra) (Supplementary Fig. 1 and Supplementary Fig. 10), are likely an important factor in LsAA9A’s superior xylan-degrading activity.

Our observation in LsAA9A complexes that MLG oligosaccharides were unable to bind with β-(1 → 3)-glucan bonds near the active site are consistent with our observation that both LsAA9A and CvAA9A favour the cleavage of celluloseyl regions in MLG.

Not all aspects of substrate specificity could be explained through binding cleft interactions. Rather, aspects of the specificity differences appear to be mechanistic in origin and relate to the reactivity of different substrates. The high activity, spectroscopy data and structures of LsAA9A and CvAA9A with β-(1 → 4)-glucan substrates leads us to suggest the effective oxidative mechanism deployed in these situations may be regarded as a ‘canonical pathway’. It is clear, however, that LPMOs may also have other ‘non-canonical pathway’ mechanisms, as exemplified by the differences between binding, spectroscopy and structures of LsAA9A with Xyl6. The crystallographic and EPR data show that a chloride ion—an oxygen species mimic—is not recruited into the copper’s equatorial binding site upon xyloglucosaccharide substrate binding, as happens in the canonical mechanism described by Frandsen et al.33. The aldopentose nature of xylose categorically excludes the synergistic binding of saccharide ligand and molecular oxygen, which is brought about by a bridging ‘pocket’ water molecule between the C6-hydroxymethyl group of Glc and the amino terminus of the enzyme. This suggests a different oxidative mechanism may well be in operation for the cleavage of xylose-based substrates by LsAA9A. Indeed, as has already been proposed by Kjaergaard et al.13, activation of O2 by an AA9 from Thermoascus aurantiacus probably gives formation of a copper-bound superoxide or hydroperoxo (HO2) through associative displacement of a superoxide anion by a water molecule through the axial coordination site on the copper ion. In particular, a superoxide ion bound to the copper in the axial position would be in position to cleave a saccharidic chain by direct attack. Such a mechanism is expected when the axial water molecule on the copper ion is not displaced by the binding of substrate, as is the case with the binding of Xyl1 to LsAA9A. From the low-dose LsAA9A:Xyl3 structure described herein, the axial ligand is clearly present on the copper ion, though it is best modelled as a mixture of chloride and water, and the Tyr-O distance (2.86 Å) is not shortened compared to the un-complexed low dose structure (2.72 Å—PDB 5ACG). This is in contrast to the low dose LsAA9A:CelL structure where the Tyr-O distance is 2.47 Å (PDB 5ACF). Furthermore, the equatorial position in the low dose LsAA9A:Xyl3 is occupied by a water molecule, not a chloride ion, as corroborated by the EPR spectroscopy. Thus, a mechanism by which a copper-bound superoxide is generated next to the substrate is possible within the LsAA9A-Xyl3 complex. Such a mechanism may be expected to be rate-independent on the redox potential of the reducing agent, since the rate-limiting step is likely to be hydrogen atom abstraction by the superoxide from the substrate rather than reductive cleavage of the O–O bond. Therefore, the fact that the rate of cleavage of xylan and Xyl6 by LsAA9A is less dependent on reducing agent while the cleavage of the other substrates is strongly dependent (Fig. 4) illustrates that a different oxidative mechanism is in operation. Thus the extent of activity on certain substrates is a function of the oxidative species which can be formed at the copper ion which is—in turn—dependent on the substrate. This means that for some substrates the use of reducing agents with different potentials can profoundly affect apparent substrate specificity. But, more importantly, LPMOs appear to have more than one oxidative mechanism available for substrate cleavage, governed to some extent by the nature of the substrate—LPMO interaction. Indeed, the existence of multiple oxidative mechanisms for a
single LPMO is an intriguing contribution to the ongoing debate about LPMO mode of action.

Results obtained in this study broaden the known substrate specificity of AA9 LPMOs to include isolated xylan and xylo-oligosaccharides, and mannosyl-containing bonds within glucomannan. We further show that oxidation type (C1/C4) is influenced by substrate type, and in this work differed between oligo- and polysaccharides. This investigation into the molecular causes of AA9 LPMO substrate specificity demonstrated the existence of multiple influences. As with carbohydrate-acting hydrolases, for example, LPMO substrate specificity is dictated by binding cleft protein:carbohydrate interactions. But in addition, the fact that activity on some substrates is differentially responsive to reducing agent potential suggests that these carbohydrates do not properly activate the active site copper, and are cleaved through an alternative oxidative pathway. Combinations of canonical and non-canonical mechanisms greatly extend the range of potential substrates for LPMOs and offer new insight into their biochemical mode of action.

Table 3 Spin-Hamiltonian parameters (parallel region) for LsAA9A and CvAA9A in contact with substrates

| Enzyme-substrate combination | \( g_z \) | \( A_z \) (MHz) | Comments |
|------------------------------|--------|----------------|----------|
| No NaCl                      |        |                |          |
| LsAA9A-H\(_2\)O              | 2.279  | 458            | Weak superhyperfine (SHF) coupling |
| LsAA9A + Cell\(_6\)          | 2.273  | 515            | Intense SHF coupling |
| LsAA9A + xylan               | 2.278  | 470            | Weak SHF coupling |
| LsAA9A + solubilised xylan   | 2.272  | 470            | Spectrum complicated by organic-based radicals in perpendicular region |
| LsAA9A + glucomannan         | 2.232  | 518            | Very likely NaCl contamination in the substrate. Intense SHF coupling |
| LsAA9A + xyloglucan          | 2.270  | 515            | Very intense SHF coupling. |
| 200 mM NaCl                  |        |                |          |
| LsAA9A-Cl                    | 2.258  | 455            | Likely mixture of \( H_2O \) and Cl species. |
| LsAA9A + Cell\(_6\)          | 2.234  | 517            | Intense SHF coupling |
| LsAA9A + avicel              | 2.232  | 522            | Slight change in perpendicular region, some appearance of SHF coupling |
| LsAA9A + xylan               | 2.270  | 470            | Spectrum complicated by organic-based radicals in perpendicular region |
| LsAA9A + solubilised xylan   | 2.272  | 470            | Radical impurities present in the perpendicular region |
| LsAA9A + glucomannan         | 2.231  | 520            | Intense SHF coupling |
| LsAA9A + xyloglucan          | 2.228  | 530            | Intense SHF coupling |
| LsAA9A + Xyl\(_6\)           | 2.268  | 400            | Very rhombic, different from both Cell\(_6\)-bound and unbound protein, intense SHF coupling. Could only be achieved with very high Xyl\(_6\) concentrations. |
| CvAA9A                       |        |                |          |
| CvAA9A + Cell\(_6\)          | 2.273  | 476            | Likely no Cl species present |
| CvAA9A + Xyl\(_6\)           | 2.228  | 522            | Mixture of C\(_6\)-bound and unbound CvAA9A. Some SHF coupling visible. Full binding could not be achieved even with large excess of Cell\(_6\) |
| CvAA9A-1 M NaCl              | 2.273  | 468            | Likely no Cl species present even in the presence of 1 M NaCl |
| CvAA9A + Xyl\(_6\)-1 M NaCl   | 2.273  | 468            | Spectrum identical to the unbound form, even at very high Xyl\(_6\) concentrations. |

The experiments were performed with or without 0.2 M chloride. For xylose hexaose, 1.0 M chloride was used. Spectra are shown in Supplementary Figs. 14 and 15. Due to the high amount of protein required by the technique, the data presented are from single EPR experiments, although the spectra with Cell6, Xyl6 and avicel were performed in at least duplicate.

Methods

Phylogenetic tree. AA9 is a family with more than 6000 sequences listed in NCBI nr and JGI databases in 2016. Because of high variability in the N-terminal portion of LPMO amino-acid sequences, no significant global alignment of LPMOs can be obtained—thereby limiting global downstream phylogenetic analyses. We chose therefore to extract the highly variable N-terminal half of these sequences (which includes two histidine residues involved in the coordination of the copper atom) for phylogenetic analysis as well as to limit the analysis to sequences that are closely related to each other and to those that have been biochemically characterized in the literature. We reduced the set of AA9 sequences to those that gave BLAST bit-scores greater than or equal to a value of 200, using LsAA9A and TaA9A as queries. A Jaccard distance matrix was compiled from BLAST bit scores and represented as a tree, built according to the principle of neighbor-joining method displaying the resulting 444 sequences (Fig. 1).

Protein production. The gene encoding LsAA9A was amplified from genomic DNA of *Lentinus similis* and the gene encoding CvAA9A was amplified from genomic DNA of *Collariella virescens* (formerly known as *Chaetomium virescens*). Both were cloned in *E. coli* using primers containing insertion sites for the vector pDau109 used for cloning. The fragments were then cloned into BamHI- and XhoI-digested pDau109 using an IN-FUSION Cloning Kit. Cloning of the genes into BamHI- and XhoI-digested pDau109 resulted in transcription of the recombinant.
enzymes encoding gene under the control of a NA2-tpi double promoter. The treated plasmids and inserts were transformed into One Shot TOP10F® Chemically Competent E. coli (Stratagene) according to the manufacturer’s protocol, spread onto LB plates supplemented with 0.1 mg/ml ampicillin and incubated at 37 °C overnight. Colonies of each transformation were cultivated in LB medium supplemented with 0.1 mg/ml ampicillin and plasmids were isolated using a QIAQuick Miniprep Kit (Qiagen).

LA9AA and CA9AA was expressed in Aspergillus oryzae MT3568. Transformants producing the recombinant enzymes were inoculated in 21 of Dap- 4 medium and incubated at 30 °C for 4 days. Mycelia were removed by filtration, and the medium was collected for purification. Ammonium sulfate was added to the supernatant to a conductivity of 1.2 ml/cm and the pH was adjusted to 7.5. The broth was applied to a 50×15 Butyl Toyopearl column (TosoBio Biosiences) equilibrated with 25 mM Tris, 1.5 M ammonium sulfate, pH 7.5. The column was washed in the same buffer and eluted with a gradient to 25 mM Tris, pH 7.5.

Fractions containing recombinant enzymes were combined and washed with milliQ water by ultrafiltration (10 kDa MWCO, PES filter, Sartorius) to a conductivity of 1.2 ml/cm and the pH adjusted to 8.0 and applied to a 50/40 Q Sepharose FF column (GE Healthcare) equilibrated with 20 mM Tris, pH 8.0. The column was washed in the same buffer and the enzyme eluted with a gradient from 0 to 0.5 M sodium chloride. Fractions containing LA9AA were combined and concentrated by ultrafiltration using VIVASPIN 20 (10 kDa MWCO) spin concentrators.

**Enzyme assays.** Apo-LA9AA and apo-CA9AA were pre-incubated for 0.5–1 h at 5°C in 0.9 stoichiometric Cu(II)(NO3)2 immediately before enzyme reactions. AA9 enzyme reactions on oligosaccharides were in 10 μl containing 5 nmol oligosaccharide, 100 mM ammonium formate pH 6.0, ±10 μmol H5G XEG and were incubated at 20°C for 4 h. Oligosaccharides were purchased from Megazyme (see also following section). In general, enzyme reactions on polysaccharides were in 100 μl containing 0.5% (w/v) polysaccharide, 100 mM ammonium formate pH 6.0, ±4 μmol ascorbate, pyrogallol or cysteine, ±5 pmol LA9AA or TA9AA and were incubated at 20°C for 4 h. Xyloglucan endoglucanase (XEG) reactions were in 10 μl containing 5 nmol oligosaccharide, 100 mM ammonium formate pH 6.0, ±10 pmol H5G XEG and were incubated at 20°C for 4 h. Oligosaccharides were purchased from Megazyme (see also following section). In general, enzyme reactions on polysaccharides were in 100 μl containing 0.5% (w/v) polysaccharide, 100 mM ammonium formate pH 6.0, ±4 μmol ascorbate, pyrogallol or cysteine, ±63 pmol PMP0, and were incubated at 20°C for 16 h. Avicel cellulose was purchased from Sigma-Aldrich, UK; barley beta-glucan medium viscosity (mixed-linked glucan), konjac glucomannan, tamarind xyloglucan, birchwood xylan, corn starch and laminarin were purchased from Megazyme, Ireland; squid pen–beta-chitin was a kind gift from Dominique Gillet of Mahatni Chitosan. Phosphoric acid-swollen cellulose (PASC) was prepared by making a slurry of 1 g Avicel cellulose (Sigma-Aldrich) with 3 ml H2O before adding 30 ml ice-cold phosphoric acid and incubating at 0°C for 1 h. The cellulose was then washed numerous times with water until the pH reached 3.0 in solution and transferred to a milder pH before use in reactions. Mixed-linkage glucan, glucomannan, xyloglucan, xylan, starch and laminarin were boiled for 5 min to make solubilized 1% (w/v) stock solutions before reactions. To aid solubilisation where necessary, water was added to a methanol: polysaccharide slurry before boiling, which improved dispersion throughout the water. Reactions were routinely stopped by addition of three- to fourfold volumes of 96% (v/v) ethanol before precipitation of the undigested substrates, and separation of the reaction products for further analysis. For polysaccharide analysis by carbohydrate electrophoresis (PACE), reaction products and oligosaccharide standards (Megazyme) were reductively aminated with 8-amino-saccharide, 100 mM ammonium formate pH 6.0, 0.1 M NaCl and incubated at 55°C for 90 min. All experiments were carried out at least in triplicate. See Supplementary Table 1 for details of carbohydrate electrophoresis conditions.

**Molecular Replacement using MOLREP with modified coordinates of the highest resolution structure of LsAA9A: oligosaccharide complex** is shown in Supplementary Fig. 12. LsAA9A was modelled in the low-resolution electron density and refined isotropically for all atoms. For all crystals with an Rmerge > 30% a number of water molecules were modelled with 80% occupancy. For different data set (collected on a crystal grown in presence of 0.1 M NaCl, Supplementary Table 4) a structure solved using the low-resolution electron density showed a Cell A molecule, which however was not bound at the active site. Solves with Cell B, damaged the crystals.

**Data extraction and refinement.** All data was processed using XDS (the resolution cutoff was chosen on the basis of a CC1/2 around 50%) and subsequently scaled using SCALEX. Refmac5 was used for restrained refinement of the structures in which LsAA9A: XylS was refined anisotropically, while LsAA9A: XylS was refined anisotropically for protein atoms and isotropically for all other atoms. All other structures were refined isotropically for all atoms. For LA9AA/G4G4G/G4G4G the structure was modelled by the G4G4G portion of the structure bound mainly in subsite −1 to +2 of the catalytic cleft (20% occupancy). Near subsite, −2 a number of water molecules were modelled with 80% occupancy. Ligands and structures were modelled in COOT and validated using MolProbity (within COOT) and Procheck (CPK4 suite), which reported Ramachandran plots with 99% of residues in allowed regions for all structures. Calculations of stereochemical statistics (also in XDS) methods are refined in the presence of 100 mM NaCl for all structures. To identify potential stacking interactions of the β-face of β-mannose with His, the PDB database was searched with Glyvicinity. All first protein/β-mannose interactions.
interactions within a distance cut-off of 4.0 Å were identified. Among these, only two structures were found where the interactions involved His residues and the pyranose O5. The interactions between the imidazole and the pyranose rings were side by side or almost perpendicular, and thus not comparable with the +1 subsite interactions of the LsAA9A complexes.

**EPR spectroscopy.** Continuous wave (cw) X-band frozen solution EPR spectra of 0.2–0.3 mM solution of LsAA9A or CmAA9A (in 10% v/v glycerol) at pH 6.0 (50 mM sodium phosphate buffer with or without addition of 200 mM NaCl or 20 mM MES buffer, 200 mM NaCl) and 165 K were acquired on a Bruker EMX spectrometer operating at ~9.30 GHz, with modulation amplitude of 4 G, modulation frequency 100 kHz and microwave power of 10.02 mW (three scans). Aviced celulose, konjac glucomannan, tamarind xyloglucan and birchwood xylan were added to the EPR tube containing the protein as solids. Alternatively, glucomannan and xylan were heated until dissolution (ca. 2 min) to make solubilized 1% (w/v) stock solutions in water, which were then used for addition of excess polysaccharides to LsAA9A. Cellohexose and xylohexose were added to the protein solution either from stock solutions in water or as a solid up to 60- or 150-fold excess, respectively. For the experiments in the presence of xylohexose, additional NaCl was added to the protein alone or the protein:Xyl6 mixture from a 5 M stock solution. Due to the high amount of protein required by the technique, the data presented are from single EPR experiments, although the spectra with CmAA9 and avicel were performed in at least duplicate. Spectral simulations were carried out using EasySpin 5.0.3.19 integrated into MATLAB 2016a software on a desktop PC. Simulation parameters are given in Table 3. g and |A| values were determined accurately from the absorptions at low field. It was assumed that g and A tensors were axial coincident.

**Data availability.** Protein Data Bank: Atomic coordinates and structure factors for the reported crystal structures were deposited under accession codes 5C1T (CmAA9A, SNLQ (LsAA9A-Cellulose), SNLQ (LsAA9A-Xyl5)), 5D98 (LsAA9A:CuII), 5SNL (LsAA9A-Xyl5), 5SNL (LsAA9A-Xyl5-CuII), GenBank: Sequence data for AA9A-Xyl5-CuII), GenBank: Sequence data for AA9A-Xyl5, GenBank:Sequence data for LsAA9A were deposited under accession codes 5NLT (AA9A), 761 (2015).

**References**

1. Carroll, A. & Somerville, C. Cellulosic biofuels. Annu. Rev. Plant. Biol. 60, 165–182 (2009).
2. Perlack, R. D. & Stokes, B. J., U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry. Technical Report ORNL/TM-2011/224, 227 (US Department of Energy, Oak Ridge National Laboratory: Oak Ridge, TN, 2011).
3. Marrriott, P. E., Gomez, L. D. & McQueen-Mason, S. J. Unlocking the potential of lignocellulosic biomass through plant science. New Phytol. 209, 1566–1581 (2016).
4. Simmons, T. J. et al. Folding of xylose onto cellulose fibrils in plant cell walls revealed by solid-state NMR. Nat. Commun. 7, 13902 (2016).
5. Correa, T. R. L., dos Santos, L. V. & Pereira, G. A. G. A9A and AA10: from enzymatic to essential enzymes. Appl. Microbiol. Biotechnol. 100, 9–16 (2016).
6. Muller, G., Varnai, A., Johansen, K. S., Eijsink, V. G. H. & Horn, S. J. Harnessing the potential of LPMO-containing cellulosic cocktails poses new demands on processing conditions. Biotechnol. Biofuels. 8, 817 (2015).
7. Hensworth, G. R., Henriksen, B., Davies, G. J. & Walton, P. H. Discovery and characterization of a new family of lytic polysaccharide monooxygenases. Nat. Chem. Biol. 10, 122–126 (2014).
8. Cantarel, B. L. et al. The carbohydrate-active EnZymes database (CAZy): an expert resource for glycoecogenics. Nucleic Acids Res. 37, D233–D238 (2009).
9. Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. The molecular basis of lytic polysaccharide monooxygenases. J. Bacteriol. 191, 5453–5467 (2009).
10. Frommberger, M. et al. The combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. Biotechnol. Biofuels. 8, 101 (2015).
11. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M. & Henriksen, B. Expansion of the enzymatic repertoire of the CAZy database to include auxiliary activities. Curr. Opin. Biotechnol. 25, 415–421 (2013).
12. Yakovlev, I. et al. Substrate-specific transcription of the enigmatic GH61 family of the pathogenic white-rot fungus Heterobasidion irregulare during growth on lignocellulose. Appl. Microbiol. Biotechnol. 95, 979–990 (2012).
13. Podevin, L. et al. Comparative analyses of Podospora anserina secretomes reveal a large array of lignocellulose-active enzymes. Appl. Microbiol. Biotechnol. 98, 195–204 (2016).
14. Frommberger, M. et al. Lytic polysaccharide monooxygenases from Myceliophthora thermophila C1 differ in substrate preference and reducing agent specificity. Biotechnol. Biofuels. 9, 186 (2016).
15. Kracher, D. et al. Extracellular electron transfer systems fuel cellulose oxidative degradation. Science 352, 1098–1101 (2016).
16. Borisova, A. S. et al. Structural and functional characterization of a lytic polysaccharide monooxygenase with broad substrate specificity. J. Bacteriol. Chem. 290, 22955–22969 (2015).
17. Frandsen, K. E. H. & Lo Leggio, L. Lytic polysaccharide monooxygenases: a crystallographer’s view on a new class of biomass-degrading enzymes. JUCr 3, 4154–4167 (2016).
18. Vaaje-Kolstad, G., Forsberg, Z., Loose, J. S., Bissaro, B. & Eijsink, V. G. Structural diversity of lytic polysaccharide monooxygenase. Curr. Opin. Struct. Biol. 44, 67–76 (2017).
19. Davies, G. J., Wilson, K. S. & Henriisat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. Curr. Opin. Struct. Biol. 32, 537–539 (2015).
20. Frandsen, K. E. H. et al. The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. Nat. Chem. Biol. 12, 298–302 (2016).
21. Walton, P. H. & Davies, G. J. On the catalytic mechanisms of lytic polysaccharide monooxygenases. Curr. Opin. Chem. Biol. 31, 195–207 (2016).
22. Cannella, D. et al. Light-driven oxidation of polysaccharides by photosynthetic pigments and a metalloenzym. Nat. Commun. 7, 11134 (2016).
23. Lenfant, N. et al. A bioinformatics analysis of 3400 lytic polysaccharide monooxygenases from family AA9. Carbohydr. Res. 448, 166–174 (2017).
24. Loose, J. S. M. et al. Activation of bacterial lytic polysaccharide monooxygenases with cellodiose dehydrogenase. Protein. Sci. 25, 2175–2186 (2016).
25. Frandsen, K. E. H., Poulsen, J. N., Tandrup, T. & Lo Leggio, L. Unliganded and substrate bound structures of the cellodioxygenase active lytic polysaccharide monooxygenase LsAA9A at low pH. Carbohydr. Res. 448, 187–190 (2017).
26. Koijima, Y. et al. A lytic polysaccharide monooxygenase with broad xyloglucan specificity from the brown-rot fungus Gloeophyllum trabeum and its action on cellulose-xyloglucan complexes. Appl. Environ. Microbiol. 82, 6557–6572 (2016).
27. Nekuinaite, L., Arntzen, M. O., Svensson, B., Vaaje-Kolstad, G. & Abuachem, M. Lytic polysaccharide monooxygenases and other oxidative enzymes are abundantly secreted by Aspergillus nidulans grown on different carbon sources. Biotechnol. Biofuels. 8, 138 (2015).
28. Fanuel, M. et al. The Podospora anserina lytic polysaccharide monooxygenase PaLPM0H catalyzes oxidative cleavage of diverse plant cell wall matrix glyans. Biotechnol. Biofuels. 10, 63 (2017).

**Received:** 23 April 2017 **Accepted:** 29 August 2017 **Published online:** 20 October 2017
42. Asensio, J. L., Arda, A., Canada, F. J. & Jimenez-Barbero, J. Carbohydrate-aromatic interactions. Acc. Chem. Res. 46, 946–954 (2013).
43. Kjaergaard, C. H. et al. Spectroscopic and computational insight into the activation of O-2 by the mononuclear Cu center in Polysaccharide monoxygenases. Proc. Natl Acad. Sci. USA 111, 8797–8802 (2014).
44. Lefort, V., Desper, R. & Gascuel, O. FastME 2.0: a comprehensive, accurate, and fast distance-based phylogeny inference program. Mol. Biol. Evol. 32, 2798–2800 (2015).
45. Kabash, W. XDS. Acta Crystallogr. D66, 125–132 (2010).
46. Vagin, A. A. et al. REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use. Acta Crystallogr. D60, 2184–2195 (2004).
47. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. Acta Crystallogr. D60, 2126–2132 (2004).
48. Rojas-Macias, M. A. & Lutteke, T. Statistical analysis of amino acids in the vicinity of carbohydrate residues performed by GlyVicinity. Methods Mol. Biol. 1273, 215–226 (2015).
49. Stoll, S. & Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. J. Magn. Reson. 178, 42–55 (2006).
50. MATLAB. MATLAB and Statistics Toolbox Release (The MathWorks, Inc., Natick, MA, USA, 2014).

Acknowledgements
We thank MAX-lab, Sweden and the European Synchrotron Radiation Facility (ESRF), France, for synchrotron beamtime and assistance. Travel to synchrotrons was supported by the Danish Ministry of Higher Education and Science through the Instrument Center DANS.CATT and the European Community’s Seventh Framework Programme (FP7/2007-2013) under BioStruct-X (grant agreement 283570). This work was supported by the UK Biotechnology and Biological Sciences Research Council (grant numbers BB/L00423/1 to P.D. and P.H.W., and BB/L021633/1 to P.H.W.) and the Danish Council for Strategic Research (grant numbers 12-134923 to L.L.L. and 12-134922 to K.S.J.).

Author contributions
T.J.S. carried out most of the activity assays, assisted by T.Tr. and L.F.L.W. K.E.H.F. carried out most of the structural studies with J.C.P., T.Ta. and L.L.L. L.C. carried out EPR spectroscopy. L.N. and B.H. carried out phylogenetic studies. M.T. and K.S. carried out target protein identification and production. P.D., L.L.L. and P.H.W. supervised the experimental work. T.J.S., K.E.H.F., K.S.J., P.H.W., L.L.L. and P.D. analysed the data and wrote the paper.

Additional information
Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-017-01247-3.

Competing interests: M.T. and K.S. are employees of Novozymes, a producer of enzymes for industrial use. The remaining authors declare no conflict of interest.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.