Article:
Simmons, Thomas J, Frandsen, Kristian E H, Ciano, Luisa et al. (2017) Structural and electronic determinants of lytic polysaccharide monooxygenase reactivity on polysaccharide substrates. Nature Communications. 1064 (2017). ISSN: 2041-1723

https://doi.org/10.1038/s41467-017-01247-3

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

Simmons TJ[1]*, Frandsen KEH[2]*, Ciano L[3], Tryfona T[1], Lenfant N[4,5], Poulsen JC[2], Wilson LFL[1], Tandrup T[2], Tovborg M[6], Schnorr K[6], Johansen KS[7], Henrissat B[4,5,8], Walton PH[3], Lo Leggio L[2] & Dupree P[1].

1. Department of Biochemistry, University of Cambridge, Cambridge, UK.
2. Department of Chemistry, University of Copenhagen, Copenhagen, Denmark.
3. Department of Chemistry, University of York, York, UK.
4. Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, Aix-Marseille Université, Marseille, France.
5. Institut National de la Recherche Agronomique (INRA), AFMB, Marseille, France.
6. Novozymes A/S, Bagsvaerd, Denmark.
7. Department of Geoscience and Natural Resources Management, Copenhagen University, Frederiksberg, Denmark
8. Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

*These authors contributed equally to this work

Correspondence and requests for material should be sent to L.L.L. (email: leila@chem.ku.dk) or P.D. (email: pd101@cam.ac.uk)

Abstract

Lytic polysaccharide monooxygenases (LPMOs) are industrially important copper-dependent enzymes that oxidatively cleave polysaccharides. We studied two closely related AA9-family LPMOs from *Lentinus similis (Ls*AA9A) and *Collariella virescens* (*Cv*AA9A). *Ls*AA9A and *Cv*AA9A cleave a range of polysaccharides, including cellulose, xyloglucan, mixed-linkage glucan, and glucomannan. *Ls*AA9A additionally cleaves isolated xylan substrates, the first LPMO to show such activity. Insights into the determinants of specificity come from the structures of *Cv*AA9A and of *Ls*AA9A bound to cellulosic and non-cellulosic oligosaccharides. EPR spectra further reveal differences in copper co-ordination on binding of xylan compared to glucans. *Ls*AA9A activity is notably less sensitive to reducing agent potential on xylan when compared to other substrates, suggesting a different mechanistic pathway for the cleavage of xylan. 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 toward a greater understanding of the biological catabolism of lignocellulose, the world's most abundant source of renewable material and bioenergy[1,2]. The inherent recalcitrance of lignocellulose, however, is one of the major barriers to the utilization of biomass. This recalcitrance is a consequence of both the heterogeneous composition and the often semi-crystalline association of the polymers[3,4]. In addressing the problem of recalcitrance, 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[5], in particular the inclusion of the lytic polysaccharide mono-oxygenases (LPMOs)[6], are helping cellulosic-ethanol biorefineries move toward both commercial and environmental viability.

LPMOs are O_2 and reducing-agent dependent copper metalloenzymes now classified as Auxiliary Activity families AA9–AA11 and AA13[7-11]. 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[12]. What is distinctive about LPMOs is that they oxidatively rather than hydrolytically cleave polysaccharides producing saccharides with oxidized ends[13]. 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[14-16], but LPMOs are now known to act on several crystalline substrates such as chitin, cellulose and retrograded starch[10-12,17]. Later, enzymes with activity against non-crystalline and oligomeric structures were identified[18,19]. Furthermore, fungal AA9 LPMOs have been shown to be active on soluble substrates such as xyloglucan, mixed-linkage glucan and glucomannan[19-22] and on cellulose-bound xylan[23]. Conspicuously, an LPMO active on isolated xylan has not been reported. This range of reported substrates however will likely grow. The large number and sequence diversity of LPMOs that individual fungi maintain[24],

and their disparate expression profiles when the fungi are grown on different polysaccharide substrates[25,26], 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[27,28].

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[29]. The structure-function relationship of substrate specificity and regiospecificity has been recently reviewed[30,31]. Insight into LPMO:substrate binding can be gained from the structures of LPMOs[30] 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[21]. The study also showed that a single $Cell_6$ 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[32]), 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 (*Ls*AA9A) bound to cellohexaose ($Cell_6$) and $Cell_3$[33]. $Cell_6$ 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 reactive mechanism. Changes in the electronic structure of the copper ion, an important factor in the ability of the copper ion to activate O_2, occur upon substrate binding to *Ls*AA9A[33]. Furthermore, in an illustration of the complexity of substrate-LPMO interaction and the

subtle interplay of electronic and structural factors, $Cell_6$ 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[34], the determinants of which depend to varying extents on the substrate, the reducing agent reducing 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[28,35].

The detailed molecular and electronic insights of the LPMO-substrate interaction afforded by *combined* biochemical, X-ray diffraction and 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 *Ls*AA9A with bound substrates we illustrate how binding cleft interactions dictate the site of polysaccharide attack. Through comparison with the *Cv*AA9A structure, also determined here, we suggest some structural determinants of specificity for the two enzymes. *Ls*AA9A 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

*Cv*AA9A is an additional AA9 LPMO active on cello-oligosaccharides

To help understand the basis of AA9 substrate specificity, we searched for enzymes related to *Ls*AA9A which 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[36]. We selected 98 AA9 sequences that were highly similar in their N-terminal half to *Ls*AA9A and 326 sequences that were highly similar to *Ta*AA9A in the same region. *Ta*AA9A was used for comparison purposes, since 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 *Ls*AA9A and *Ta*AA9A in distinct clades. From the *Ls*AA9A clade, we identified an LPMO from *Collariella virescens* (*Cv*AA9A; 46% sequence similarity to *Ls*AA9A) that lacks some residues observed by Frandsen et al.[33] as being involved in enzyme-substrate interactions (Supplementary Fig. 1). All three of the subsite +2 substrate-binding residues in *Ls*AA9A (Asn28, His66 and Asn67) are different in *Cv*AA9A (Thr28, Arg67 and Val68) (Supplementary Fig. 1). To study the activity of *Cv*AA9A, the enzyme was expressed in *Aspergillus oryzae* and successfully purified from the fermentation broth. (Note that 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), *Cv*AA9A produced a range of cello-oligosaccharides (Fig. 1b). The cello-oligosaccharide product profile of *Cv*AA9A was similar to that of *Ls*AA9A and notably shorter than those produced by *Ta*AA9A. Indeed, *Cv*AA9A readily degraded $Cell_6$-2-aminobenzamide ($Cell_6$-2AB) using a C4-oxidising mechanism to yield $Cell_3$ and oxidized $Cell_3$-2AB (Supplementary Fig. 2), like *Ls*AA9A but unlike *Ta*AA9A[33]. 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 *Ls*AA9A and *Cv*AA9A 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 ($Cell_3$, $Cell_4$) 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, xyloglucan 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 *Ls*AA9A and *Cv*AA9A showed activity against MLG, glucomannan and xyloglucan, producing a range of oligosaccharide products (Fig. 2a). *Ls*AA9A also showed some activity on xylan whereas *Cv*AA9A 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 *Ls*AA9A and *Cv*AA9A enzymes only cleave near β-(1→4)-bonds, and that some variation to the cellulosic 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 *Ls*AA9A and *Cv*AA9A cleavage of MLG, glucomannan, xyloglucan and xylan (in the case of *Ls*AA9A) 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 protocols. On MLG, we observed in the MALDI data a predominance of DP 4, 7 and 10 oligosaccharides indicating that each enzyme favors cleaving within $Cell_4$ regions over $Cell_3$ regions (Fig. 3). The inability of *Ls*AA9A and *Cv*AA9A to cleave β-Glc-(1→4)-β-Glc-(1→3)-β-Glc-(1→4)-Glc (G4G3G4G), despite their ability to cleave $Cell_4$ (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 subsite (Supplementary Fig. 6). In order to deduce site of attack on xyloglucan we employed xyloglucan DP14–18 oligosaccharides (Supplementary Fig. 7). Inspection of the position of substituted glucose (Glc) in the products indicated that xylosyl 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 *Ls*AA9A and *Cv*AA9A products on polysaccharides, *Ls*AA9A degraded Xyl_6-2AB to yield two trimers using solely a C4-oxidising mechanism (Supplementary Fig. 8), analogous to cleavage of $Cell_6$-2AB by both *Ls*AA9A[33] and *Cv*AA9A (Supplementary Fig. 2).

To allow a semi-quantitative determination of the influence of sugar structures on enzyme activity, we probed *Ls*AA9A and *Cv*AA9A cleavage of the soluble $Cell_6$, xylohexaose (Xyl_6) and mannohexaose (Man_6) oligosaccharides (Supplementary Fig. 9). The *Ls*AA9A activity against $Cell_6$ was substantially (~100-fold) better than its activity on Xyl_6. Consistent with

the absence of activity on glucuronoxylan, *Cv*AA9A activity on Xyl_6 was almost undetectable (~1,000-fold less than $Cell_6$ activity). Although both enzymes showed activity on glucomannan and can cleave adjacent to mannose, activity was scarcely detectable on Man_6 (~10,000-fold less than $Cell_6$), 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[28,37]. We found that cleavages of MLG, glucomannan and xyloglucan by *Ls*AA9A 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 Xyl_6 was poorly sensitive to reducing agent potential, unlike cleavage of $Cell_6$ where *Ls*AA9A showed much greater activity with ascorbate than pyrogallol[33] (Fig. 4b).

*Ls*AA9A: and *Cv*AA9A:cello-oligosaccharide structures

To help understand the structural basis of LPMO attack on different substrates, we employed crystallographic analyses. We report here an *Ls*AA9A:$Cell_5$ complex (Fig. 5 and Supplementary Table 4 and 5) which, owing to a lack of significant substrate contacts to symmetry-related molecules, is a more faithful depiction of the binding conformation of a single oligosaccharide to *Ls*AA9A as compared to the original *Ls*AA9A:$Cell_6$ structure described by Frandsen et al[33]. Tyr203 stacking is still a major interaction in *Ls*AA9A:$Cell_5$ 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 *Ls*AA9A:$Cell_6$ structure[33], the main interactions to $Cell_5$ 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[38].

To understand better how protein structure might influence the similarities and differences in *Cv*AA9A and *Ls*AA9A substrate cleavage patterns, the X-ray crystal structure of *Cv*AA9A was solved (Supplementary Fig. 10 and Supplementary Tables 3-5). The Cu-coordinating amino acid residues are MeHis1 and His79 (with equatorial distances to the

Cu ranging from 2.0-2.1Å), while a non-coordinating Tyr169 occupies the axial position (2.6-2.8Å). 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 *Cv*AA9A thus closely resembles that of *Ls*AA9A (Supplementary Fig. 10c). However, there are some amino acid differences in *Cv*AA9A compared to *Ls*AA9A at subsites +2 and −1. Crystals of *Cv*AA9A were soaked with $Cell_3$ and $Cell_6$ oligosaccharides but this did not result in any catalytically relevant complex.

*Ls*AA9A:hemicellulose oligosaccharide structures

To study the structural determinants of the *Ls*AA9A positional specificity of cleavage, a number of *Ls*AA9A crystal structures in complex with MLG, glucomannan and xylo-oligosaccharides were solved (see Supplementary Tables 2–5 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.

Complexes with MLG tetrasaccharide

*Ls*AA9A crystals were soaked with two different MLG tetrasaccharides, each with a single β-(1→3)-linkage: G4G4G3G and G4G3G4G. Interestingly, the *Ls*AA9A:G4G4G3G complex did not reveal any β-(1→3)-linkages. An apparent $Cell_4$ substrate appears to be bound from subsite –2 to +2 (Supplementary Fig. 11) giving essentially identical interactions as the −2 to +2 glucosyl residues in the *Ls*AA9A:$Cell_5$ complex. We interpret this result as the β-(1→4)-glucan ($Cell_3$) 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 *Ls*AA9A crystals soaked with G4G3G4G (not shown) showed very little difference density, which could not be convincingly modelled, further indicating that the

243 enzyme needs at least two consecutive β-(1→4)-linkages (a $Cell_3$ unit) for recognition and
244 efficient binding.

245 *Complexes with glucomannan oligosaccharides*

246 *Ls*AA9A crystals were soaked with a mixture of glucomannan oligosaccharides. The
247 resulting difference density was well defined clearly showing glycosyl units occupying
248 subsites –3 to +2, additional density at –4 and some residual density occupying subsite +3
249 (Fig. 6a and 6c). Consistent with the activity data, the structure unequivocally showed a
250 mannosyl unit at the +1 subsite, while glucosyl units were clearly observable at –2, –1 and
251 +2 subsites. Moreover the C2 hydroxyl of the mannosyl unit at subsite +1 points towards
252 the face of the imidazole side chain of MeHis1, and the axial water molecule is displaced
253 (Fig. 6b and 6d). The identity of the glycosyl unit at subsite –3 is ambiguous though best
254 modelled as mannose. The density of the glycosyl unit at subsite –4 is weak and occupies
255 a very similar position as the corresponding unit in the *Ls*AA9A:$Cell_6$ complex, as does the
256 glycosyl unit at subsite –3, due to similar crystal constraints.

257 *Complexes with xylo-oligosaccharide*

258 Whereas in crystals soaked with Xyl_3 and Xyl_4 (Supplementary Tables 2–-5) the
259 oligosaccharides did not fully span the active site, *Ls*AA9A:Xyl_5 crystals revealed very well-
260 defined density from subsites –3 to +2 (Fig. 7a). The oligosaccharide position at subsites
261 –3 to –1 are similar to *Ls*AA9A:$Cell_5$, but with a translation of about half a pyranose unit in
262 the non-reducing end direction. In contrast, the plane of the xyloside unit at subsite +1 is
263 rotated approximately 90° compared to the corresponding glucosidic unit, while the
264 xyloside residue binding +2 is rotated approximately 180° (Fig. 7d, Supplementary Table
265 2). As a result xylose at subsite +1 does not stack with MeHis1, and in fact appears to
266 have no interactions with the enzyme, while the same residues that bind the subsite +2
267 glucosyl residue in the *Ls*AA9A:$Cell_5$ structure, Asn28, His66 and Asn67, interact here with
268 O1, O5 and O1 of the +2 xyloside residue, respectively (Fig. 7c Supplementary Table 3). A
269 structure of *Ls*AA9A:Xyl_5 determined from a low X-ray dose data collection showed the
270 substrate bound similarly, and revealed a mix of water/Cl^- in the axial position and a fully
271 occupied equatorial water on the active site copper (Fig. 7b). Thus, in contrast to binding

of cello- or glucomannan oligosaccharides, the axial water was not displaced by binding of Xyl_5.

EPR data suggest alternative *Ls*AA9A substrate binding modes

We studied substrate binding on both *Ls*AA9A and *Cv*AA9A using electron paramagnetic resonance (EPR) spectroscopy (Table 2) to investigate the electronic state of the active site copper upon binding. As has been shown by Frandsen et al[33] and Courtade et al[21], 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 Xyl_6 studies). Furthermore, experiments were carried out at high substrate concentration to maximize substrate binding. For *Ls*AA9A 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 2, Supplementary Fig. 12) and used as an indication of increased metal-ligand covalency in the singly-occupied molecular orbital (SOMO), as previously discussed by Frandsen et al.[33]. In all cases apart from xylan, the addition of substrate gave perturbation of the Cu spin Hamiltonian parameters similar to that already reported by Frandsen et al.[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 xyloglucan, indicative of chloride coordination to the copper ion in the equatorial position of the copper coordination sphere. These shifts are analogous to those of *Ls*AA9A interacting with $Cell_6$ and PASC[33]. In contrast, addition of Xyl_6 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 $Cell_6$. The EPR spectra of *Ls*AA9A binding to Xyl_6 and xylan are indicative of substrate binding to the enzyme (although binding of Xyl_6 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

303 $Cell_6$ to *Cv*AA9A but not with Xyl_6, consistent with the observed activity on $Cell_6$ and not
304 Xyl_6 (Fig.8 and Supplementary Fig.13).

305

306 ## Discussion

307 Our understanding of the molecular basis for substrate binding and cleavage has been
308 aided by the recent report of a crystal structure of *Ls*AA9A in complex with $Cell_3$ and $Cell_6$,
309 as well as biochemical and EPR data for *Ls*AA9A on cellulosic substrates[33]. Here, we have
310 extended this biochemical, EPR and structural analysis by using a range of substrates as
311 well as an additional related enzyme, *Cv*AA9A, to provide a better insight into substrate
312 specificity.

313 Extensive probing of *Ls*AA9A and *Cv*AA9A substrate specificity showed that both cleave a
314 range of cellulosic and non-cellulosic substrates, some of which have been shown for
315 other AA9s[19,20,22,23,27,29,35,39-41]. We made a number of novel observations. Notably,
316 *Ls*AA9A activity on xylan and xylo-oligosaccharides is the first report of LPMO cleavage of
317 isolated xylan; this may have important implications for the use of LPMOs in
318 biotechnological contexts. LPMO activity on xylan has been observed before for
319 *Mt*LPMO9A[23], but only on xylan associated with cellulose. We also observe that both
320 *Ls*AA9A and *Cv*AA9A are able to cleave glycosidic bonds adjacent to mannosyl residues
321 (Supplementary Fig. 6), which occur interspersed randomly with glucosyl residues in
322 glucomannan, a biochemical observation supported by the *Ls*AA9A:glucomannan
323 oligosaccharide structure, which unambiguously shows a mannosyl residue at subsite +1.
324 We also noticed substrate-specific oxidation profiles, namely that *Ls*AA9A and *Cv*AA9A
325 cleaved small oligosaccharides using a C4-oxidising mechanism whereas they cleaved
326 polysaccharides with both C1- and C4-oxidising mechanisms in varying proportions.
327 Assuming a copper-based oxidative species, the similar distances between both C1 and
328 C4 axial protons and the active oxygen species, as noted in Frandsen et al.[33], may allow
329 slight differences in substrate binding to switch the C-H bond that is closest to attack.
330 Substrate binding differences may also subtly alter the electronics at the copper site, which
331 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 *Ls*AA9A and *Cv*AA9A 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, *Ls*AA9A's preference to cleave $Cell_4$ into $Cell_2$ (as shown for *Nc*LPMO9C[18]), compared to the product profile of *Cv*AA9A 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 *Cv*AA9A shares the same fold, active site co-ordination and overall structure with *Ls*AA9A, it lacks all three of the subsite +2 substrate-binding residues in *Ls*AA9A (Asn28, His66 and Asn67) (Fig. 5; Supplementary Fig. 1). This suggests that, while *Ls*AA9A binds $Cell_4$ from subsite -2 to +2, *Cv*AA9A binds between subsites −3 and +1.

The structural data provide a molecular rationale for how *Ls*AA9A is able to catalyse the unexpected cleavage of mannose-containing bonds. The *Ls*AA9A: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 subsite +1. We have previously described that the glucosyl unit at the +1 subsite in *Ls*AA9A $Cell_3$ and $Cell_6$ complexes[33,38] interacts with MeHis1 through its β-face, and while glucose can make carbohydrate-aromatic stacking interactions[42] 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 *Ls*AA9A confirmed that this type of interaction takes place, and the mannosyl residue at the +1 subsites interacts with MeHis through its β-face with an O5-imidazole ring centre distance of 3.6 Å (Fig 6d) (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 mannosyl 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 *Ls*AA9A and *Cv*AA9A cleaved xyloglucan DP14–18 oligomers with the sole unsubstituted backbone glucosyl residues at subsite +1 (XXX/GXXXGol; as found for *Nc*LPMO9C[19]) is consistent with the binding of xyloglucan's cellulosyl backbone being similar to the binding of cello-oligosaccharides. This would suggest that the *Ls*AA9A 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.

*Ls*AA9A under the selected conditions degraded Xyl_6 with about 1/100 the efficiency as $Cell_6$, while *Cv*AA9A left Xyl_6 essentially untouched at all conditions tested (Supplementary Fig. 9). The differences in key amino acids involved in defining the *Ls*AA9A and *Cv*AA9A subsites, particularly the +2 subsite (vide supra) (Supplementary Fig. 1; Fig. 5), are likely an important factor in *Ls*AA9A's superior xylan-degrading activity.

Our observation in *Ls*AA9A 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 *Ls*AA9A and *Cv*AA9A favour the cleavage of cellulosyl 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 *Ls*AA9A and *Cv*AA9A 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 *Ls*AA9A with Xyl_6.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 xylooligosaccharide 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 *Ls*AA9A. Indeed, as has already been proposed by Kjaergaard et al[43], activation of O_2 by an AA9 from *Thermoascus aurantiacus* probably gives formation of a copper-bound superoxide or hydrosuperoxide (HO_2) 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 Xyl_5 to *Ls*AA9A. From the low dose *Ls*AA9A:Xyl_5 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 (2.72Å - PDB 5ACG). This is in contrast to the low dose *Ls*AA9A:$Cell_3$ structure where the Tyr-O distance is 2.47 Å (PDB 5ACF). Furthermore, the equatorial position in the low dose *Ls*AA9A:Xyl_5 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 *Ls*AA9A-Xyl_5 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 Xyl_6 by *Ls*AA9A is less dependent on reducing agent while the cleavage of the other substrates is strongly dependent (Figure 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 on-going 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.

Materials and 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 *Ls*AA9A and *Ta*AA9A 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[44] displaying the resulting 444 sequences (Fig. 1).

Protein production

Cloning, expression, and purification of *Ls*AA9A was done as described previously[33]. The gene encoding *Cv*AA9A was amplified from genomic DNA of *Collariella virescens* (formerly known as *Chaetomium virescens*) and expressed in *Aspergillus oryzae* MT3568. The secreted LPMO was purified using a Butyl Toyopearl resin followed by purification on a Q-Sepharose FF resin followed by ultrafiltration with 10 kDa cutoff filter.

Enzyme assays

Apo-*Ls*AA9A and apo-*Cv*AA9A were pre-incubated for 0.5–1 h at 5 °C in 0.9 stoichiometric $Cu(II)(NO_3)_2$ immediately before enzyme reactions. AA9 enzyme reactions on oligosaccharides were in 10 µL containing 5 nmol oligosaccharide, 100 mM ammonium formate pH 6, ±4 mM ascorbate, pyrogallol or cysteine, ±5 pmol *Ls*AA9A or *Ta*AA9A 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, ±10 µmol GH5 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, ±4 mM ascorbate, pyrogallol or cysteine, ±63 pmol LPMO, and were incubated at 20 °C for 16 h. Avicel cellulose was purchased from Sigma-Aldrich, UK; barley beta-glucan medium viscosity (mixed-linkage glucan), konjac glucomannan, tamarind xyloglucan, birchwood xylan, corn starch and laminarin were purchased from Megazyme, Ireland; squid-pen β-chitin was a kind gift from Dominique Gillet of Mahtani Chitosan. Phosphoric acid-swollen cellulose (PASC) was prepared as described previously[33]. Mixed-linkage glucan, glucomannan, xyloglucan, xylan, starch and laminarin were boiled for 5 mins 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 reaction volumes of 96% (v/v) ethanol before precipitation of the undigested substrates, and separation of the reaction products for further analysis. Polysaccharide Analysis by Carbohydrate Electrophoresis (PACE) using high concentration acrylamide gels (for resolution of small oligosaccharides) and lower concentration gels (for resolution of larger oligosaccharides) was carried out as described by Frandsen et al.[33] and Goubet et al.[45] respectively. MALDI-ToF MS was performed as described previously[46]. 2-Aminobenzamide (2-AB) labelling was performed as described previously[33]. All experiments were carried out at least three times.

Sodium borohydride reducing agent experiments were performed as described by Frandsen et al.[33]. HPAEC was performed on a CarboPac PA1 column (Dionex) with injections of 20 µL and elution at 0.4 mL min^{-1}. The elution profile was: 0–3 min, 10 mM NaOH (isocratic); 3–6 min, 10→1 mM NaOH (linear gradient); 6–19 min, 1 mM NaOH (isocratic); 19–37 min, 45 mM NaOH, 225 mM sodium acetate (isocratic). A pulsed amperometric detector (PAD) with a gold electrode was used. PAD response was calibrated using markers (500 pmol).

X-ray crystallography and PDB database searches

All crystallization trials were set up in MRC 2-well plates at room temperature using an Oryx-8 robot (Douglas Instrument). Crystals were obtained by sitting-drop vapor diffusion technique in drops of 0.3-0.5 µL with a reservoir volume of 100 µL. Pre-incubation with 1-2 mM Cu(II) acetate for 30-60 min was carried out for all crystallization trials. Crystallization and post-crystallization experimental details are shown in Supplementary Table 4. Crystals were cryocooled in liquid nitrogen and all datasets were collected at cryogenic temperatures (100 K) at either the MX beamlines I911-2/I911-3 at MAX-lab in Lund, Sweden, or at the MX beamlines ID23-1, ID23-2 or ID30-B at ESRF, Grenoble, France (Supplementary Table 5). *Ls*AA9A crystallization was performed as described in Frandsen et al.[33]. Oligosaccharide substrates used for soaking were purchased from Megazyme (MLG (G4G4G3G and G4G3G4G), xylotriose (Xyl_3), xylotetraose (Xyl_4), xylopentaose (Xyl_5), xyloglucan heptasaccharide (XXXG), cellotriose ($Cell_3$) or provided by Novozymes A/S (cellopentaose ($Cell_5$)). Data were initially collected on crystals soaked with G4G4G3G (*Ls*AA9A:G4G4G3G; PDB 5NLR), Xyl_3 (*Ls*AA9A:Xyl_3; PDB 5NLQ), Xyl_4 (*Ls*AA9A:Xyl_4; PDB 5NLP) and $Cell_5$ (*Ls*AA9A:$Cell_5$; 5NLS). On a crystal soaked with Xyl_5 a dataset with reduced X-ray dose ((*Ls*AA9A:Xyl_5Cu(II); PDB 5NLN; 40 frames of 5.7% transmission, 0.05s exposure/frame, 1° oscillation with a beamsize of 10x10 µm) was collected using helical collection to minimize photoreduction of the active site copper. Subsequently, on similar crystals another full dose dataset was collected to high resolution (*Ls*AA9A:Xyl_5; PDB 5NLO). Ladders of glucomannan (GM), from konjac, and of xyloglucan (XG), from tamarind, were prepared by partial acid hydrolysis (20-200 mM TFA for 20 min at 120°C) of polysaccharide substrates purchased from Megazyme. Hydrolyzed products were isolated using ethanol precipitation to remove the remaining polysaccharides. The oligosaccharides were dried thoroughly using a SpeedVac. Data were collected on crystals soaked in GM (*Ls*AA9A:GM; PDB 5NKW) or XG stock solutions (in 3.8 M NaCl, 0.1 M citric acid pH 5.5). Crystals were also soaked in the presence of 0.3 M XXXG, and up to 1.2 M of XG oligosaccharide purchased from Megazyme (consisting primarily of XXXGXXXG, see Courtade et al[21]). No complex structures were obtained from any of the crystals soaked with XG substrates, either because no binding was observed or only cellooligosaccharides were bound (presumably because acid hydrolysis caused debranching).

*Cv*AA9A was deglycosylated in 20 mM MES, pH 6.0, 125 mM NaCl by incubation with approximately 0.03 units mg^{-1} *Cv*AA9A of endoglycosidase H from (Roche Diagnostics, 11643053001), and then buffer exchanged to 20 mM Na-acetate pH 5.5. Intergrown crystals were initially obtained in an index screen in conditions of 1.5–2.0 M $(NH_4)_2SO_4$ (and in some cases 0.1 M NaCl) in pH 6.5-8.5 (0.1 M of either Bis-Tris, HEPES or Tris). The crystals diffracted to 2.0–3.5 Å resolution but were multiple. Crystal conditions were optimized in a range of 1.2 M – 2.6 M $(NH_4)_2SO_4$ (+/– 0.1 M NaCl) in pH 6.5–8.5, which produced crystal plates suitable for mounting. A dataset collected at I911-3 on a crystal grown in 0.1 M Bis-Tris pH 6.5, 2.0 M $(NH_4)_2SO_4$ could be processed in $P2_1$ to 2.5 Å (PDB 5NLT). A preliminary *Cv*AA9A structure with four molecules in the asymmetric unit was solved by Molecular Replacement using MOLREP with modified coordinates of the high resolution structure of *Ls*AA9A (PDB 5ACH) which is 41% identical, as a model and refined isotropically to an R_{free} of 32%. From another dataset (collected on a crystal grown in presence of 0.1 M NaCl; Supplementary Table 4) a structure solved (using the preliminary one) with six molecules in the asymmetric unit could be fully modelled and refined resulting in the complete *Cv*AA9A structure (Supplementary Table 5). The 6 molecules in the asymmetric unit are very similar (average RMSDs of 0.08Å). The density of MeHis1 is less clear in chains C and F, and in particular methylation is not as obvious in all chains. Soaks (with 1.2 M $Cell_3$) of *Cv*AA9A were also prepared. Data were collected to 2.1 Å and the electron density showed a $Cell_3$ molecule, which however was not bound at the active site (not shown). Soaks with $Cell_6$ damaged the crystals.

Each dataset was processed using XDS (the resolution cutoff was chosen on the basis of a CC½ around 50%) and subsequently scaled using XSCALE. Refmac5 was used for restrained refinement of the structures in which *Ls*AA9A:Xyl_5 was refined anistropically, while *Ls*AA9A:Xyl_3 was refined anisotropically for protein atoms and isotropically for all other atoms. All other structures were refined isotropically for all atoms. For *Ls*AA9A:G4G4G3G the structure was best modelled by the G4G4G portion of the substrate bound mainly in subsite –1 to +2 (80% occupancy) and with a minor conformation occupying subsite –2 to +1 (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 (CCP4 suite) which reported Ramachandran plots with 99% of residues in allowed regions for all

structures. Scaled data statistics and refinement statistics are summarized in Supplementary Table 5.

To identify potential stacking interactions of the β-face of β-mannose with His, the PDB database was searched with Glyvicinity[47]. First all protein/β-mannose interactions within a distance cut-off of 4.0 Å for structures determined at a resolution better than 3.0 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 *Ls*AA9A complexes.

Electron Paramagnetic Resonance (EPR) Spectroscopy

Continuous wave (cw) X-band frozen solution EPR spectra of 0.2 to 0.3 mM solution of *Ls*AA9A or *Cv*AA9A (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 (3 scans). Avicel cellulose, 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 polysaccharide to *Ls*AA9A. Cellohexaose and xylohexaose 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 xylohexaose, additional NaCl was added to the protein alone or the protein:Xyl_6 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 $Cell_6$, Xyl_6 and avicel were performed in at least duplicate.

Spectral simulations were carried out using EasySpin 5.0.3[48] integrated into MATLAB R2016a[49] software on a desktop PC. Simulation parameters are given in Table 2. g_z and $|A_z|$ values were determined accurately from the absorptions at low field. It was assumed that *g* and A tensors were axially coincident.

Data availability

Protein Data Bank: Atomic coordinates and structure factors for the reported crystal structures were deposited under accession codes 5NLT (*Cv*AA9A), 5NLS (*Ls*AA9A-$Cell_5$), 5NLR (*Ls*AA9A-G4G4G3G), 5NKW (*Ls*AA9A-GM), 5NLQ (*Ls*AA9A-Xyl_3), 5NLP (*Ls*AA9A-Xyl_4), 5NLO (*Ls*AA9A-Xyl_5) and 5NLN (*Ls*AA9A-Xyl_5-Cu_{II}), GenBank: Sequence data for *Cv*AA9A were deposited under accession code KY884985. Raw EPR data are available on request through the Research Data York (DOI: 10.15124/5810c962-148c-4328-ab92-895e2dae4d3c).

References

1 Carroll, A. & Somerville, C. Cellulosic Biofuels. *Annu. Rev. Plant Biol.* **60**, 165-182, (2009).

2 Perlack, R. D. & Stokes, B. J., Leads. *U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry*. Technical Report ORNL/TM-2011/224, 227. (U.S. Department of Energy, Oak Ridge National Laboratory, Oak Ridge, TN, 2011).

3 Marriott, P. E., Gomez, L. D. & McQueen-Mason, S. J. Unlocking the potential of lignocellulosic biomass through plant science. *New Phytol.* **209**, 1366-1381, (2016).

4 Simmons, T. J. *et al.* Folding of xylan onto cellulose fibrils in plant cell walls revealed by solid-state NMR. *Nat. Commun.* **7**, (2016).

5 Correa, T. L. R., dos Santos, L. V. & Pereira, G. A. G. AA9 and AA10: from enigmatic to essential enzymes. *Appl. Microbiol. Biot.* **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 cellulase cocktails poses new demands on processing conditions. *Biotechnol. Biofuels* **8**, (2015).

7 Hemsworth, G. R., Henrissat, B., Davies, G. J. & Walton, P. H. Discovery and characterization of a new family of lytic polysaccharide rnonooxygenases. *Nat. Chem. Biol.* **10**, 122-126, (2014).

8 Cantarel, B. L. *et al.* The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233-D238, (2009).

9 Lombard, V., Ramulu, H. G., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490-D495, (2014).

10 Vu, V. V., Beeson, W. T., Span, E. A., Farquhar, E. R. & Marletta, M. A. A family of starch-active polysaccharide monooxygenases. *P. Natl. Acad. Sci. USA* **111**, 13822-13827, (2014).

11 Lo Leggio, L. *et al.* Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nat. Commun.* **6**, (2015).

12 Quinlan, R. J. *et al.* Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *P. Natl. Acad. Sci. USA* **108**, 15079-15084, (2011).

13 Hemsworth, G. R., Johnston, E. M., Davies, G. J. & Walton, P. H. Lytic Polysaccharide Monooxygenases in Biomass Conversion. *Trends Biotechnol.* **33**, 747-761, (2015).

14 Vaaje-Kolstad, G., Horn, S. J., van Aalten, D. M. F., Synstad, B. & Eijsink, V. G. H. The non-catalytic chitin-binding protein CBP21 from Serratia marcescens is essential for chitin degradation. *J. Biol. Chem.* **280**, 28492-28497, (2005).

15 Merino, S. T. & Cherry, J. Progress and challenges in enzyme development for Biomass utilization. *Adv. Biochem. Eng. Biot.* **108**, 95-120, (2007).
16 Harris, P. V. *et al.* Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family. *Biochemistry-Us* **49**, 3305-3316, (2010).
17 Vaaje-Kolstad, G. *et al.* An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science* **330**, 219-222, (2010).
18 Isaksen, T. *et al.* A C4-oxidizing Lytic Polysaccharide Monooxygenase Cleaving Both Cellulose and Cello-oligosaccharides. *J. Biol. Chem.* **289**, 2632-2642, (2014).
19 Agger, J. W. *et al.* Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *P. Natl. Acad. Sci. USA* **111**, 6287-6292, (2014).
20 Bennati-Granier, C. *et al.* Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by Podospora anserina. *Biotechnol. Biofuels* **8**, (2015).
21 Courtade, G. *et al.* Backbone and side-chain H-1, C-13, and (15) N chemical shift assignments for the apo-form of the lytic polysaccharide monooxygenase NcLPMO9C. *Biomol. NMR Assign.* **10**, 277-280, (2016).
22 Jagadeeswaran, G., Gainey, L., Prade, R. & Mort, A. J. A family of AA9 lytic polysaccharide monooxygenases in Aspergillus nidulans is differentially regulated by multiple substrates and at least one is active on cellulose and xyloglucan. *Appl. Microbiol. Biot.* **100**, 4535-4547, (2016).
23 Frommhagen, M. *et al.* Discovery of the combined oxidative cleavage of plant xylan and cellulose by a new fungal polysaccharide monooxygenase. *Biotechnol. Biofuels* **8**, (2015).
24 Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M. & Henrissat, B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol. Biofuels* **6**, (2013).
25 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. Biot.* **95**, 979-990, (2012).
26 Poidevin, L. *et al.* Comparative analyses of Podospora anserina secretomes reveal a large array of lignocellulose-active enzymes. *Appl. Microbiol. Biot.* **98**, 7457-7469, (2014).
27 Frommhagen, M. *et al.* Lytic polysaccharide monooxygenases from Myceliophthora thermophila C1 differ in substrate preference and reducing agent specificity. *Biotechnol. Biofuels* **9**, (2016).
28 Kracher, D. *et al.* Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science* **352**, 1098-1101, (2016).
29 Borisova, A. S. *et al.* Structural and Functional Characterization of a Lytic Polysaccharide Monooxygenase with Broad Substrate Specificity. *J. Biol. Chem.* **290**, 22955-22969, (2015).
30 Frandsen, K. E. H. & Lo Leggio, L. Lytic polysaccharide monooxygenases: a crystallographer's view on a new class of biomass-degrading enzymes. *Iucrj* **3**, 448-467, (2016).
31 Vaaje-Kolstad, G., Forsberg, Z., Loose, J. S., Bissaro, B. & Eijsink, V. G. Structural diversity of lytic polysaccharide monooxygenases. *Curr Opin Struct Biol* **44**, 67-76, (2017).
32 Davies, G. J., Wilson, K. S. & Henrissat, B. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* **321 (Pt 2)**, 557-559, (1997).
33 Frandsen, K. E. H. *et al.* The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nat. Chem. Biol.* **12**, 298-+, (2016).
34 Walton, P. H. & Davies, G. J. On the catalytic mechanisms of lytic polysaccharide monooxygenases. *Curr. Opin. Chem. Biol.* **31**, 195-207, (2016).
35 Cannella, D. *et al.* Light-driven oxidation of polysaccharides by photosynthetic pigments and a metalloenzyme. *Nat. Commun.* **7**, (2016).
36 Lenfant, N. *et al.* A bioinformatics analysis of 3400 lytic polysaccharide oxidases from family AA9. *Carbohydr Res*, (2017).

37 Loose, J. S. M. *et al.* Activation of bacterial lytic polysaccharide monooxygenases with cellobiose dehydrogenase. *Protein Sci.* **25**, 2175-2186, (2016).
38 Frandsen, K. E. H., Poulsen, J. C. N., Tovborg, M., Johansen, K. S. & Lo Leggio, L. Learning from oligosaccharide soaks of crystals of an AA13 lytic polysaccharide monooxygenase: crystal packing, ligand binding and active-site disorder. *Acta Crystallogr. D.* **73**, 64-76, (2017).
39 Kojima, 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. Microb.* **82**, 6557-6572, (2016).
40 Nekiunaite, L., Arntzen, M. O., Svensson, B., Vaaje-Kolstad, G. & Abou Hachem, M. Lytic polysaccharide monooxygenases and other oxidative enzymes are abundantly secreted by Aspergillus nidulans grown on different starches. *Biotechnol. Biofuels* **9**, (2016).
41 Fanuel, M. *et al.* The Podospora anserina lytic polysaccharide monooxygenase PaLPMO9H catalyzes oxidative cleavage of diverse plant cell wall matrix glycans. *Biotechnol Biofuels* **10**, 63, (2017).
42 Asensio, J. L., Arda, A., Canada, F. J. & Jimenez-Barbero, J. Carbohydrate-Aromatic Interactions. *Accounts 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 monooxygenases. *Abstr. Pap. Am. Chem. S.* **248**, (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 Goubet, F., Jackson, P., Deery, M. J. & Dupree, P. Polysaccharide analysis using carbohydrate gel electrophoresis: A method to study plant cell wall polysaccharides and polysaccharide hydrolases. *Anal. Biochem.* **300**, 53-68, (2002).
46 Tryfona, T. *et al.* Carbohydrate structural analysis of wheat flour arabinogalactan protein. *Carbohyd. Res.* **345**, 2648-2656, (2010).
47 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).
48 Stoll, S. & Schweiger, A. EasySpin, a comprehensive software package for spectral simulation and analysis in EPR. *J Magn Reson* **178**, 42-55, (2006).
49 MATLAB and Statistics Toolbox Release 2014a (The MathWorks, I., Natick, Massachusetts, United States).

Acknowledgements

We wish to thank MAXLAB, 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 DANSCATT 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/L000423/1 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.T. and L.F.L.W. K.E.H.F. carried out most of the structural studies with J.C.P., T.T. 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.W. supervised the experimental work. T.J.S., K.E.H.F., K.S.J.O., P.W., L.L.L. and P.D. analysed the data and wrote the paper.

Competing Financial Interests

M.T. and K.S. are employees of Novozymes, a producer of enzymes for industrial use.

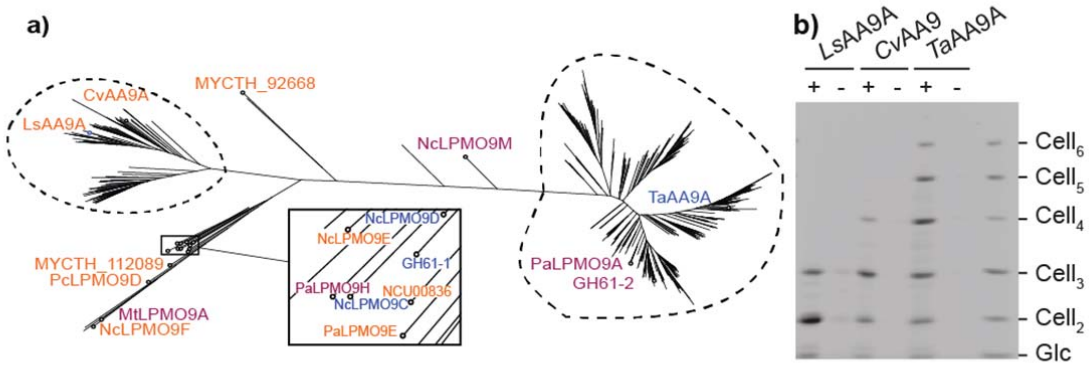
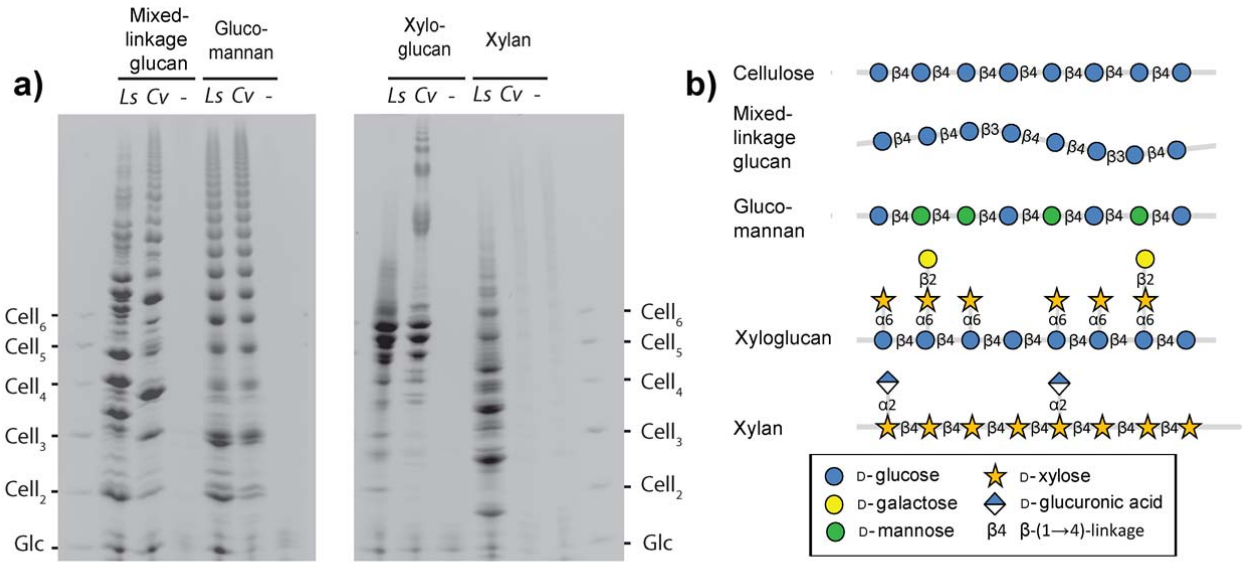


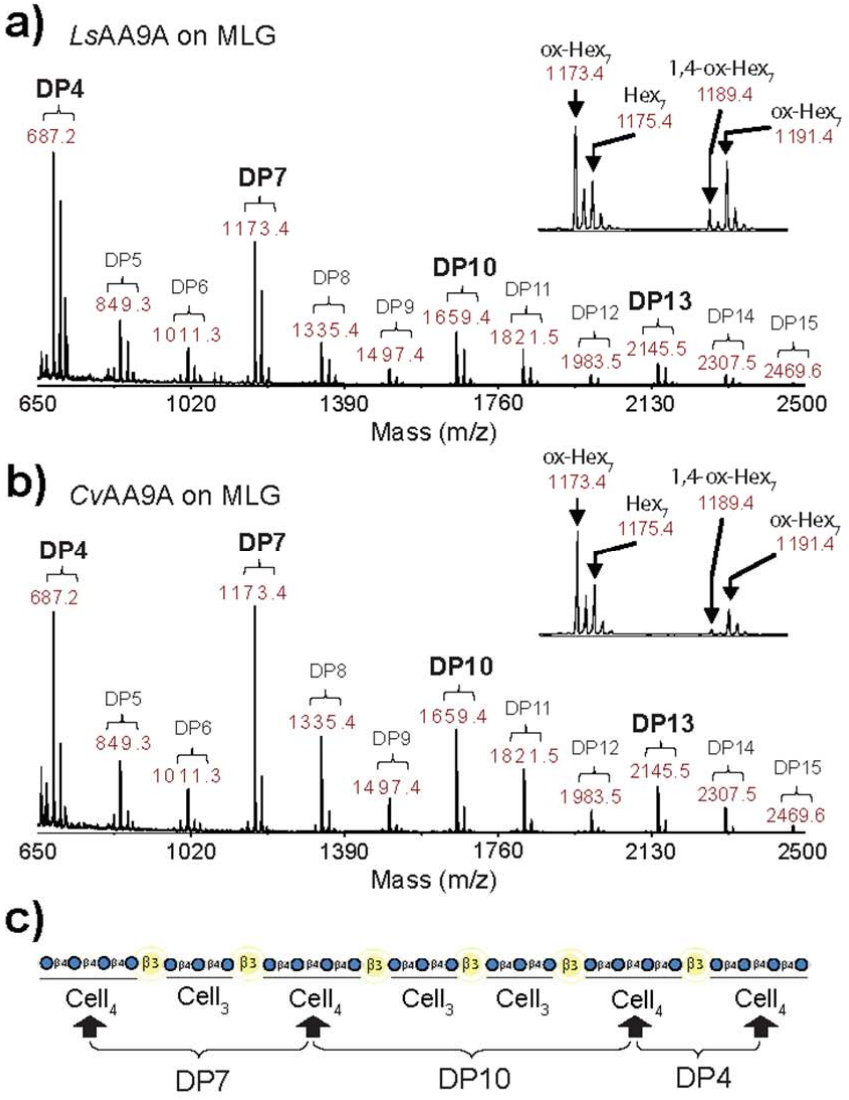
Figure 1: Sequence similarity between *Ls*AA9A, *Cv*AA9A and *Ta*AA9A and analysis of their reaction products. a, Distance tree of 444 selected AA9 sequences (see methods). Blue, purple and orange labels designate AA9 enzymes that oxidize the sugar ring at C1, C4 and C1+C4, respectively. See Supplementary Table 1 for protein accession numbers. Unlabelled branches represent AA9 enzymes for which the regioselectivity of oxidation is not available from the literature. **b,** PACE gel showing reaction products of the three enzymes on PASC; +, incubation with 4mM ascorbate; -, incubation without ascorbate (performed in triplicate).



742

743 **Figure 2: Comparison of *Ls*AA9A and *Cv*AA9A action on non-cellulosic substrates.**
744 **a**, PACE gel showing digestion products on lignocellulosic polysaccharides with 4mM
745 ascorbate reducing agent. *Ls*, *Ls*AA9A; *Cv*, *Cv*AA9A; -, no enzyme. **b**, Structures of
746 polysaccharides.

747



748

749 **Figure 3: *Ls*AA9A and *Cv*AA9A digestion products of MLG suggest preference for**
750 **Cell$_4$ region cleavage.** Products of *Ls*AA9A (**a**) and *Cv*AA9A (**b**) activity on barley MLG
751 with 4mM ascorbate were analysed by MALDI-ToF MS. Both enzymes can produce both
752 C1 and C4 oxidation on MLG (1,4-ox; oxidized C1 and C4. See insets). Further,
753 oligosaccharide profiles show a distinct pattern indicative of the mechanism of attack and
754 substrate specificity of each enzyme on MLG. **c**, proposed region of cleavage.

755

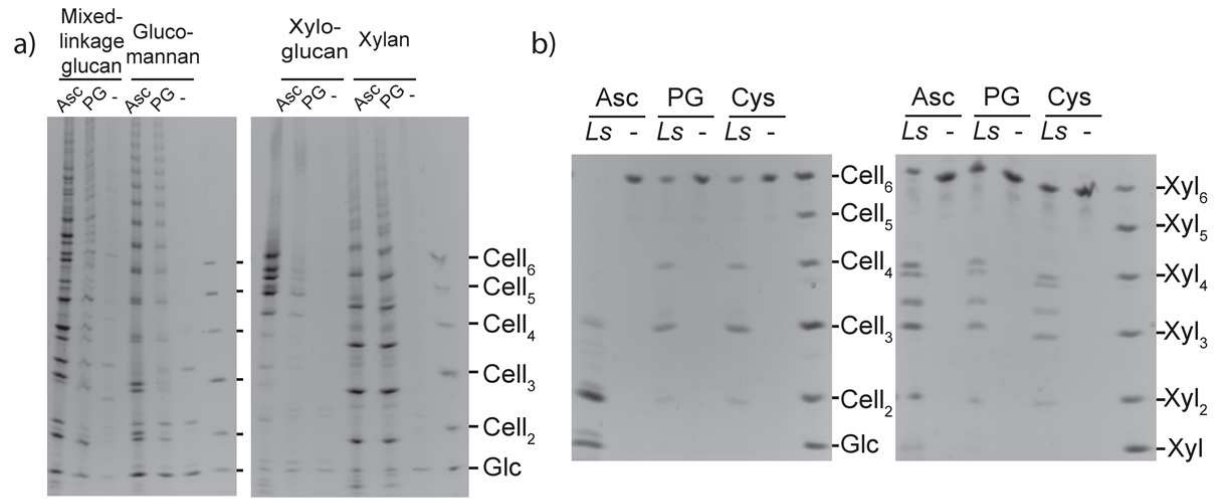
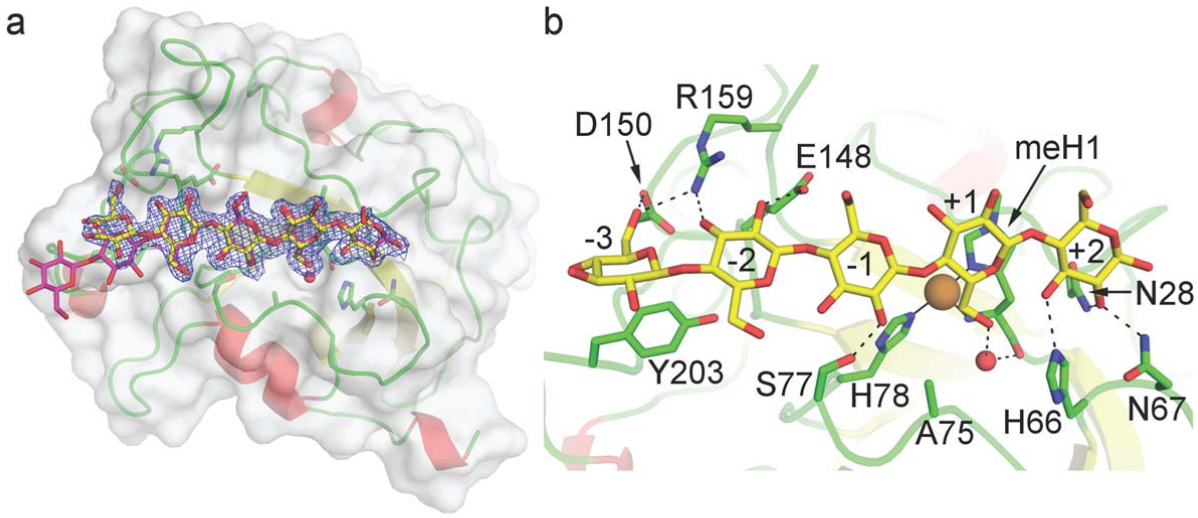


Figure 4: *Ls*AA9A activity on different poly- and oligosaccharide substrates show differing sensitivity to reducing agent potential. a, PACE gels showing products of *Ls*AA9A activity on MLG, glucomannan, xyloglucan and xylan polysaccharides using 4mM ascorbate or 4mM pyrogallol as reductants. The migration standards are cello oligosaccharides. b, PACE gels showing products of *Ls*AA9A activity on $Cell_6$ and Xyl_6 oligosaccharides using 4mM ascorbate, pyrogallol and cysteine as reductants. Asc, Ascorbate; PG, pyrogallol; Cys, cysteine.



765

766 **Figure 5: *Ls*AA9A:Cell$_5$ structure shows interactions at subsites -3 to +2. a,** Cell$_5$
767 (yellow) is well defined in subsites -3 to +2. A $2F_{obs}$-F_{calc} electron density map is shown at
768 1σ contour level. The structure shows no crystal contact induced distortion of the Cell$_5$
769 substrate when compared to Cell$_6$ (magenta). **b,** *Ls*AA9A-Cell$_5$ interactions are shown as
770 dashes. An additional interaction between subsite -3 (O)6 and Asp150 is gained in the
771 absence of symmetry related contacts to the substrate.

772

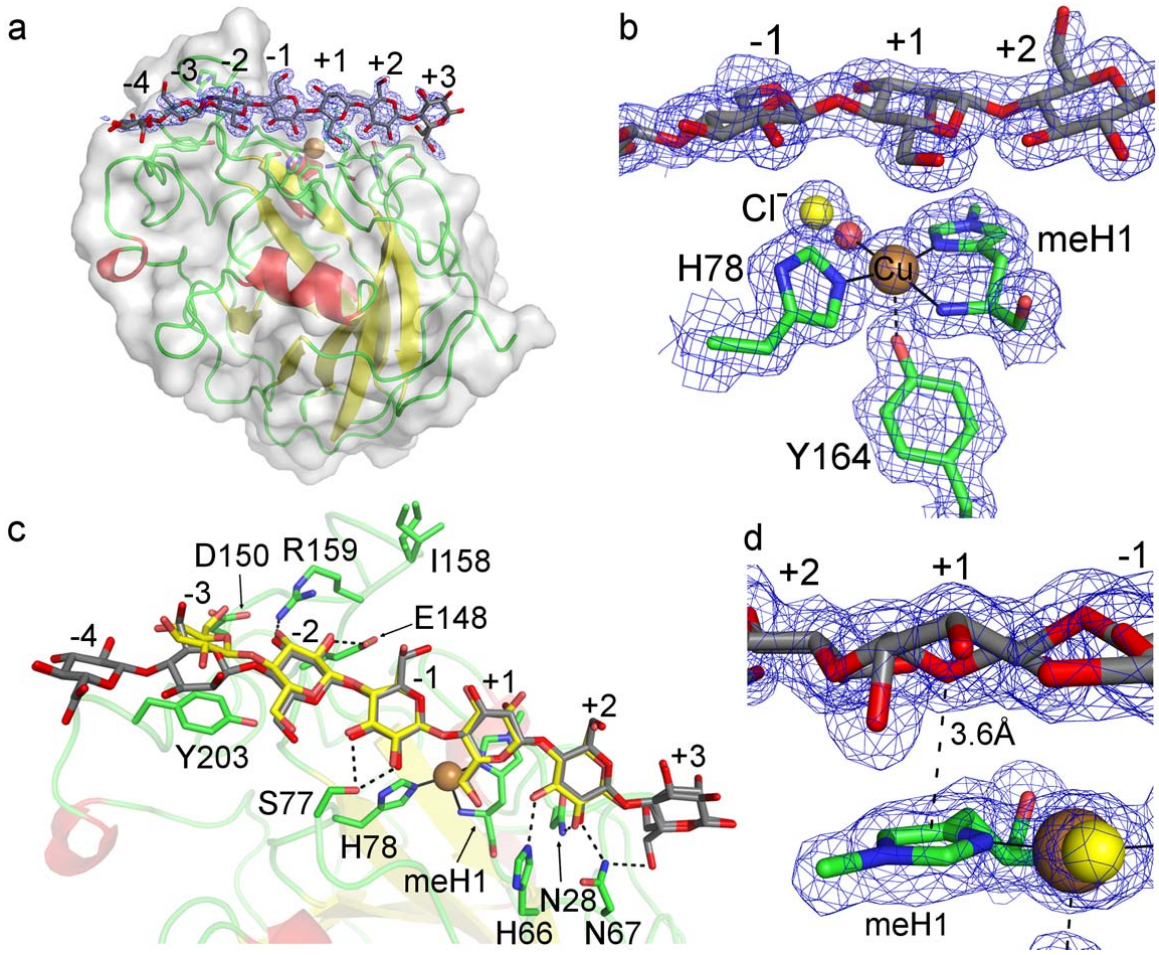
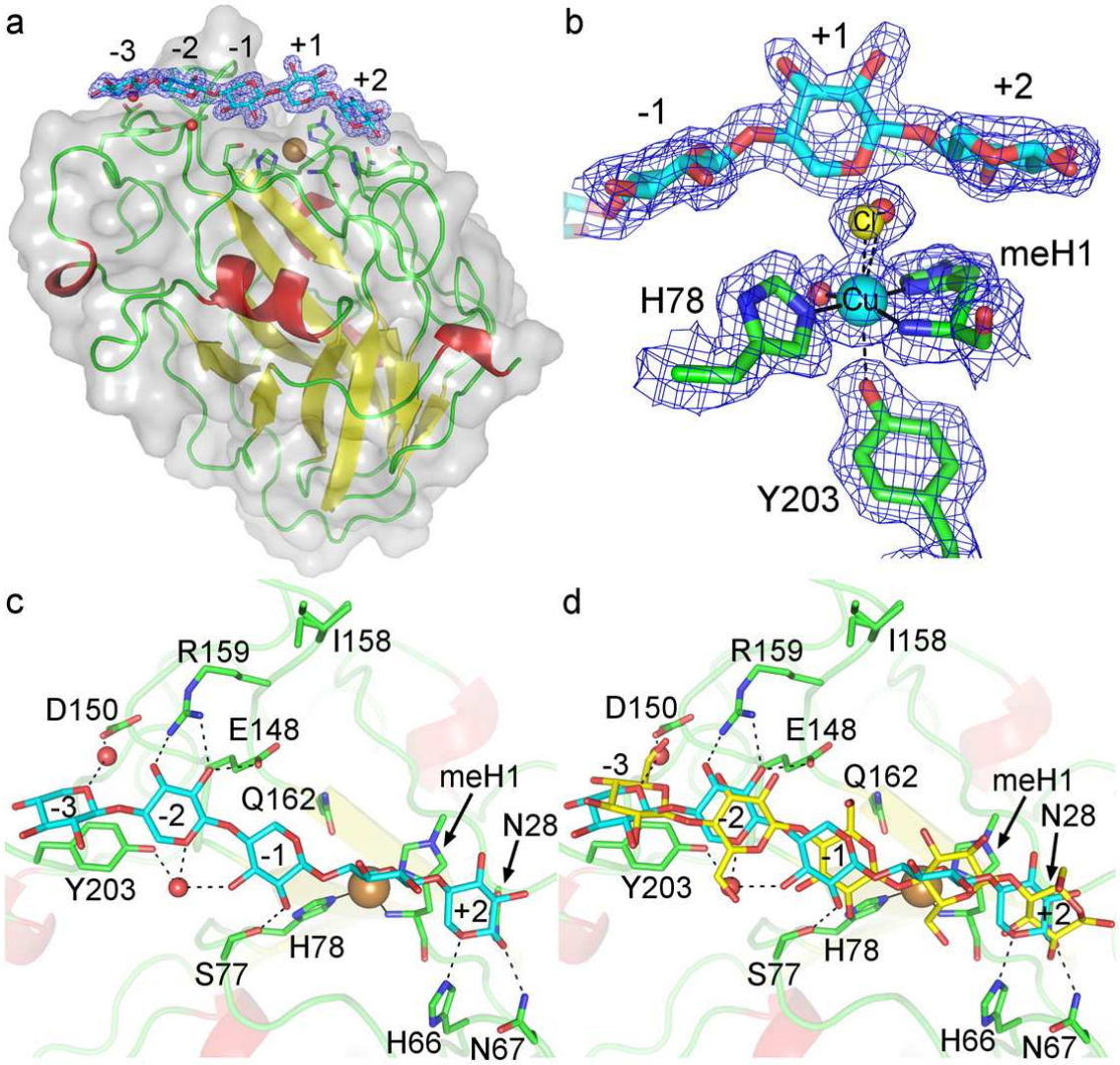


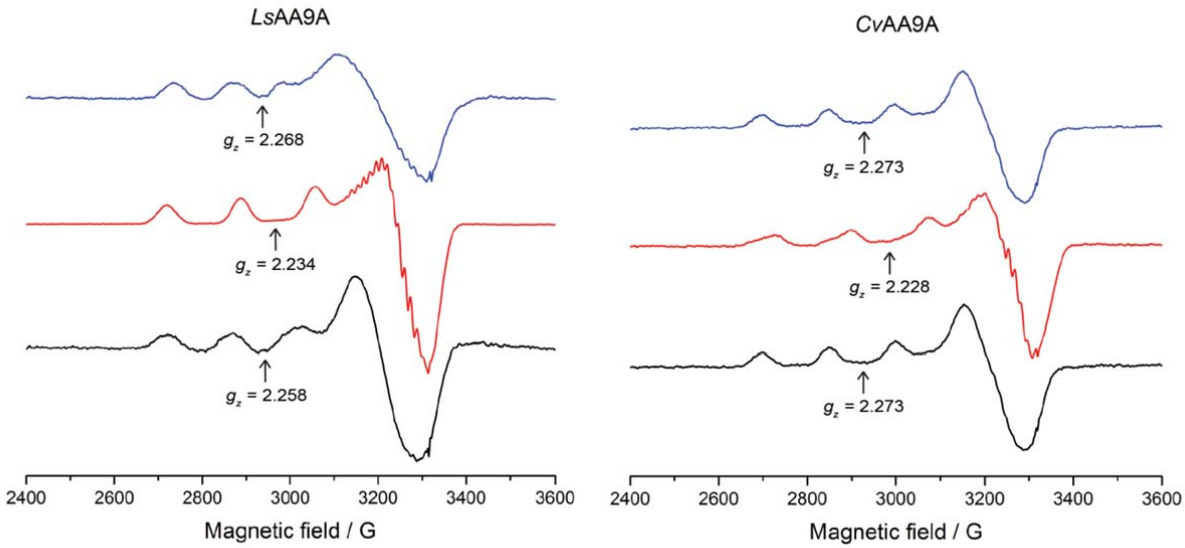
Figure 6: Structure of the *Ls*AA9A:glucomannan oligosaccharide complex. a, Overall structure of *Ls*AA9A with GM (blue) bound from subsite -4 to +3. **b,** Close up of the active site with GM fragment bound at subsite -1 to +2. Axial coordinations are in black dashes while equatorial coordinations are in full black lines. **c,** Top-down view of *Ls*AA9A:GM (GM in grey) and for comparison *Ls*AA9A:$Cell_5$ ($Cell_5$ in yellow). Dashed lines show interactions within hydrogen bond distance (2.8 Å). **d,** The C2-hydroxyl of mannose is clearly visible in the density at subsite +1. The pyranose O5-imidazole ring interaction (3.6 Å) is indicated with dashes. The interaction of MeHis and the mannosyl residue is very similar to the interaction with glucosyl residues in previous complexes[38]. A $2F_{obs}$-F_{calc} electron density map is shown at 1σ contour levelfor panels a, b and d**.**



784

785 **Figure 7: Structure of the *Ls*AA9A:xylo-oligosaccharide complex. a,** *Ls*AA9A:Xyl_5 with
786 bound substrate (in cyan) from subsite -3 to +2. **b,** Active site structure in the low dose
787 *Ls*AA9A:Xyl_5 structure, showing that the +1 xylosyl unit does not direct interact with the
788 enzyme or displace the axial ligand on the copper (modelled as chloride and water in
789 0.5:0.5 ratio). **c,** Top-down view *Ls*AA9A:Xyl_5 (in cyan). **d,** Top-down view *Ls*AA9A:Xyl_5 (in
790 cyan) and *Ls*AA9A:$Cell_5$ (in yellow) shown for comparison. A $2F_{obs}$-F_{calc} electron density
791 map is shown at 1σ contour level in panels a and b.

792



793

794 **Figure 8: X band cw EPR spectra of *Ls*AA9A (left) and *Cv*AA9A (right), 150 K.**
795 Spectra were collected in the presence of 1 M NaCl (black), $Cell_6$ and 200 mM NaCl (red),
796 or Xyl_6 and 1 M NaCl (blue).

797

Polysaccharide	*Ls*AA9A		*Cv*AA9	
	Activity	**Notes**	**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.	++	Pattern suggests β-(1→3)-bonds accommodated at specific places within active site, but not between -1 and +1.
Glucomannan and Man_6	++	Cleavage can occur with Glc or mannose at -1 or +1. Inactivity on Man_6 indicates some Glc C2 hydroxyl orientation needs to be present between -3 and +3.	++	Cleavage can occur with Glc or mannose at -1 or +1. Inactivity on Man_6 indicates some Glc C2 hydroxyl orientation needs to be present between -3 and +3.
Xyloglucan	++	Cleavage occurs with unsubstituted Glc at subsite +1. Xylosyl substitution at -3, -2, -1, +2 and +3. Galactosyl-xylosyl substitutions can occur at either -2, -1 and/or +3.	++	Cleavage occurs with unsubstituted Glc at subsite +1. Xylosyl substitution at -3, -2, -1, +2 and +3. Galactosyl-xylosyl substitutions can occur at either -2, -1 and/or +3.
Xylan and Xyl_6	+	Activity on both. Weak Xyl_6 activity compared to $Cell_6$ suggests that, while Glc C6 is not required for activity, it is very important at certain sites, such as +1.	+/-	Much poorer activity of *Cv*AA9A compared with *Ls*AA9A.
Starch	-	Absence of activity indicates that *Ls*AA9A necessarily cleaves β-(1→4)-bonds.	-	Absence of activity indicates that *Cv*AA9A necessarily cleaves β-(1→4)-bonds.
Laminarin	-		-	
G4G3G4G (MLG oligosaccharide)	-		-	
Chitin	-	Absence of activity indicates that *Ls*AA9A either requires O2 interactions or cannot accommodate N-Acetyl on amino C2.	-	Absence of activity indicates that *Cv*AA9A either requires O2 interactions or cannot accommodate N-Acetyl on amino C2.

Table 1: Summary of activity assays on different substrates. Semi-quantitative activity results summarising the activity of *Ls*AA9A and *Cv*AA9A on the range of different substrates used in this manuscript.

Enzyme-substrate combination	g_z	A_z (MHz)	Comments
No NaCl			
*Ls*AA9A-H_2O	2.279	458	Weak superhyperfine (SHF) coupling
*Ls*AA9A+$Cell_6$	2.273	515	Intense SHF coupling
*Ls*AA9A+avicel	2.278	470	Weak SHF coupling
*Ls*AA9A+xylan	2.272	480	Spectrum complicated by organic-based radicals in perpendicular region
*Ls*AA9A+glucomannan	2.232	518	Very likely NaCl contamination in the substrate. Intense SHF coupling
*Ls*AA9A+xyloglucan	2.270	515	Very intense SHF coupling.
200 mM NaCl			
*Ls*AA9A-Cl	2.258	455	Likely mixture of H_2O and Cl species.
*Ls*AA9A+$Cell_6$	2.234	517	Intense SHF coupling
*Ls*AA9A+avicel	2.232	522	Slight change in perpendicular region, some appearance of SHF coupling
*Ls*AA9A+xylan	2.270	470	Spectrum complicated by organic-based radicals in perpendicular region
*Ls*AA9A+solubilised xylan	2.272	470	radical impurities present in the perpendicular region
*Ls*AA9A+glucomannan	2.231	520	Intense SHF coupling
*Ls*AA9A+solubilised glucomannan	2.233	515	Intense SHF coupling
*Ls*AA9A+xyloglucan	2.228	530	Intense SHF coupling
*Ls*AA9A+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.
*Cv*AA9A	2.273	476	Likely no Cl species present
*Cv*AA9A+$Cell_6$	2.228	527	Mixture of C_6-bound and unbound *Cv*AA9A. Some SHF coupling visible. Full binding could not be achieved even with large excess of $Cell_6$
*Cv*AA9A-1 M NaCl	2.273	468	Likely no Cl species present even in the presence of 1 M NaCl
*Cv*AA9A+Xyl_6 -1 M NaCl	2.273	468	Spectrum identical to the unbound form, even at very high Xyl_6 concentrations.

Table 2: Spin-Hamiltonian parameters (parallel region) for *Ls*AA9A and *Cv*AA9A in contact with substrates. The experiments were performed with or without 0.2 M chloride. For xylohexaose 1.0 M chloride was used. Spectra are shown in Supplementary Fig. 12–13.