Article:
Frandsen, KEH, Tovborg, M, Jørgensen, CI et al. (2019) Insights into an unusual Auxiliary Activity 9 family member lacking the histidine brace motif of lytic polysaccharide monooxygenases. The Journal of Biological Chemistry, 294 (45). pp. 17117-17130. ISSN: 0021-9258

https://doi.org/10.1074/jbc.ra119.009223

Insights into an unusual Auxiliary Activity 9 family member lacking the histidine brace motif of lytic polysaccharide monooxygenases

Kristian E. H. Frandsen[1,2], Morten Tovborg[3], Christian I. Jørgensen[3], Nikolaj Spodsberg[3], Marie-Noëlle Rosso[2], Glyn R. Hemsworth[4§,5], Elspeth F. Garman[6], Geoffrey W. Grime[7], Jens-Christian N. Poulsen[1], Tanveer S. Batth[8], Shingo Miyauchi[2], Anna Lipzen[9], Chris Daum[9], Igor V. Grigoriev[9,10], Katja S. Johansen[11], Bernard Henrissat[12,13,14] Jean-Guy Berrin[2], Leila Lo Leggio[1*].

1. Department of Chemistry, University of Copenhagen, Copenhagen, Denmark. **2**. INRA, Aix-Marseille Université, UMR1163 BBF (Biodiversité et Biotechnologie Fongiques) Marseille, France. **3.** Novozymes A/S, Bagsvaerd, Denmark. **4.** School of Molecular and Cellular Biology and Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK **5.** Department of Chemistry, University of York, York, YO10 5DD, UK. **6.** Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, UK. **7.** The Ion Beam Centre, Advanced Technology Institute, University of Surrey, Guildford, GU2 7XH, U.K. **8.** The Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Copenhagen, Denmark. **9**. US Department of Energy Joint Genome Institute, Walnut Creek, CA, USA. **10.** Department of Plant and Microbial Biology, University of California Berkeley, Berkeley, CA, USA. **11.** Department of Geosciences and Natural Resource Management University of Copenhagen, Copenhagen, Denmark **12.** Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, Aix-Marseille Université, Marseille, France. **13.** INRA, USC 1408 AFMB, Marseille, France. **14.** Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia

Running title: An unusual AA9 naturally lacking the His-brace motif

*leila@chem.ku.dk, Universitetsparken 5, 2100 Copenhagen, DK (Phone: +45 35 32 02 95; Fax: +45 35 32 03 22).
[§]Current Address.

Keywords:
His-brace, N-terminal Arg AA9, Biomass degradation, Lytic polysaccharide monooxygenase, Phosphorylation, Xylooligosaccharide.

Abstract

Lytic polysaccharide monooxygenases (LPMOs) are redox-enzymes involved in biomass degradation. All characterized LPMOs possess an active site of two highly conserved histidine residues coordinating a copper ion (the histidine brace), which are essential for LPMO activity. However, some protein sequences that belong to the AA9 LPMO family, display a natural N-terminal His to Arg substitution (Arg-AA9). These are found almost entirely in the phylogenetic fungal class *Agaricomycetes*, associated with wood-decay, but no function has been demonstrated for any Arg-AA9.

Through bioinformatics, transcriptomic and proteomic analyses we present data, which suggest that Arg-AA9 proteins could have a hitherto unidentified role in fungal degradation of lignocellulosic biomass in conjunction with other secreted fungal enzymes. We present the first structure of an Arg-AA9, *Ls*AA9B, a naturally occurring protein from *Lentinus similis*. The *Ls*AA9B structure reveals gross changes in the region equivalent to the canonical LPMO copper binding site, whilst features implicated in carbohydrate binding in AA9 LPMOs have been maintained. We obtained a structure of *Ls*AA9B with xylotetraose bound on the surface of the protein although with considerably different binding mode compared to other AA9 complex structures. In addition, we have found indications of protein phosphorylation near the N-terminal Arg and the carbohydrate binding site, for which the potential function is currently unknown.

Our results are strong evidence that Arg-AA9s function markedly different from canonical AA9 LPMO, but nonetheless may play a role in fungal conversion of lignocellulosic biomass.

Fungal degradation of plant biomass involves a collection of carbohydrate active enzymes (CAZymes) which include glycoside hydrolases, carbohydrate esterases, polysaccharide lyases and oxidoreductases. Lytic polysaccharide monooxygenases (LPMOs) are copper-dependent oxidoreductases shown to be pivotal in efficient plant biomass degradation (1). These enzymes activate molecular oxygen or hydrogen peroxide, to introduce oxidative chain breaks into polysaccharide chains (2,3) (Fig. 1). Some LPMOs are single domain enzymes while others are linked to CBMs (Carbohydrate Binding Modules) or domains of unknown function (4,5). LPMOs are classified into seven Auxiliary Activity families in the CAZy database, AA9-AA11 and AA13-AA16 (6), and have been found in bacteria, viruses, fungi and recently in arthropod species and in plants (7,8). In the active site a copper ion is held by a motif denoted the histidine brace (His brace) formed by the highly conserved N-terminal histidine and a second histidine later in the sequence (9). The His brace motif is strictly conserved in all LPMOs studied to date (10-12) and is considered essential for LPMO activity. This is supported by mutational analysis of both AA9 and AA10 LPMOs (first performed on *Tt*AA9E and *Sm*AA10A, followed by later studies on *Mt*AA9D and *Sc*AA10C) in which replacement of these critical residues within the His brace or the secondary coordination sphere of the copper ion resulted in reduced LPMO activity and hence reduced glycoside hydrolase boosting (13-16).

LPMOs display different substrate specificities, with activity demonstrated on cellulose, hemicelluloses, chitin and starch (2,9,14,17-19) and interact with their polysaccharide substrates through a surface on which the His brace is found (20-24). This surface can have varying contour and polarity, presumably dictating substrate specificity (19,23,25,26). Additionally, for some AA9 LPMOs activity has been demonstrated on soluble β-1,4-linked polysaccharides and also oligosaccharides (17,22,23,27,28), a feature which was instrumental in the determination of the first enzyme-substrate complex structures of an AA9 from *Lentinus similis* (*Ls*AA9A)(22,29). The structures revealed polar residues in a loop denoted L3 (after notation in (25)) interacting with cellooligosaccharides near the active site and demonstrated for the first time that a conserved Tyr (Y203 in *Ls*AA9A) located on the same surface was involved in substrate binding (22,23). In addition these structures revealed a new and highly unusual lone pair … π aromatic interaction between the pyranose ring O5 and the imidazole of the N-terminal His of the His brace, further highlighting its importance for LPMO function (29).

Both in *Basidiomycetes* and *Ascomycetes* fungal species multiple AA9-encoding genes are found (more than twenty and thirty in *Schizophyllum commune* and *Podospora anserina*, respectively), the vast majority of which encode proteins with N-terminal His (either demonstrated or predicted). For several of these genes the expression and subsequent secretion of the proteins is readily induced by growth on plant biomass (30). Interestingly, some AA9 members have an intriguing substitution of the N-terminal His to an Arg (from here on referred to as Arg-AA9). This can be observed in genomes of species belonging to the *Lentinus* genus, where for example *Lentinus tigrinis* has sixteen AA9s one of which is an Arg-AA9 (https://genome.jgi.doe.gov/mycocosm/proteins-browser/browse;3gQkyP?p=Lenti7_1).
Similarly, for *L. similis* at least seven proteins classified as AA9s are described in patent literature (31,32)), and one of these proteins (GH61-5 in (32) and denoted *Ls*AA9B from this point) has an N-terminal Arg.

This type of His to Arg substitution was first noted by Yakovlev and co-workers in the *Russulales* fungus *Heterobasidion irregulare* in an AA9 denoted *Hi*GH61G (*Hi*AA9G from this point) (33). Yakovlev et al. found that genes encoding six AA9s, including *Hi*AA9G, were upregulated during growth on lignocellulosic biomass (33). We have previously reported comprehensive integrative omics studies of the saprotrophic white-rot fungi *Pycnoporus coccineus* BRFM310 (34) and *Polyporus brumalis* BRFM985 (35) of the Polyporales order. There we reported co-regulation of glycoside hydrolases (GH) commonly involved plant biomass breakdown (e.g. GH5_5) and Arg-

AA9 members (protein ID #1430659 and #1403153 on the JGI Mycocosm public database) which in each species was highly upregulated during fungal growth on wheat straw (for the latter during solid-state fermentation). In addition, in a recent bioinformatics study (5) it was shown that Arg-AA9s are found in several fungal species almost entirely restricted to one phylogenetic class of largely wood-decaying members namely the *Agaricomycetes*. These Arg-AA9 sequences clustered together (cluster 63 in (5)), and appear to have been maintained for a novel function rather than simply being a decaying AA9 gene product. This indicates a putative biological function for the Arg-AA9s related to lignocellulosic biomass degradation, but to date no function has been ascribed to this novel AA9 sub-group.

Using bioinformatics, transcriptomic and secretomic approaches, we have obtained data supporting a potential role for Arg-AA9s from *Agaricomycetes* species in plant biomass conversion. We present a high-resolution X-ray crystal structure of a naturally occurring protein from the *Polyporale* fungus *L. similis, Ls*AA9B, the first for any AA9 with an N-terminal Arg. The *Ls*AA9B structure has an overall fold highly similar to AA9 LPMOs, but in addition to the N-terminal His to Arg substitution, reveals extensive changes in residues equivalent to those forming the copper binding site and the secondary coordination sphere of the copper (essential for fully functioning AA9 LPMOs). Thus, perhaps predictably, no LPMO activity could be detected for *Ls*AA9B. Unexpectedly however, the structure appears with a post-translational modification in the form of a phosphoserine (pS25), in a position adjacent to the N-terminal Arg sidechain. In addition, we have obtained a complex structure of *Ls*AA9B with xylotetraose (Xyl4) bound at the protein surface in a conserved cleft near the speculated phosphorylation site. Sequence alignment indicates that also other Arg-AA9 proteins contain potentially similar putative phosphorylation sites and carbohydrate binding sites.

Our results strongly suggest a role related to biomass degradation for these Arg-AA9 proteins, which seem to be strikingly different compared to canonical AA9 LPMOs. The functional significance of both the N-terminal His to Arg substitution, the phosphoserine and the Xyl4 ligand require further investigation, but the characterization and structure determination of *Ls*AA9B presented here will serve as an important starting point for the elucidation of the function and exact biological role of Arg-AA9s.

Results

Phylogenetic distribution and regulation of AA9s with N-terminal Arg

In order to assess the phylogenetic distribution of fungal AA9 family sequences with N-terminal Arg (Arg-AA9s), we analysed AA9 sequences belonging to cluster 63 (5) in the CAZy database (36), which we found were restricted to species of the *Polyporales*, *Agaricales* and *Russulales* orders (all of the *Agaricomycetes* class in *Basidiomycetes* fungi). Notably, the sequence alignment showed that residues critical for function in canonical AA9 LPMOs (H1, H68, H142, Q151 and Y153 in *Tt*AA9E (14)) were all changed in the Arg-AA9 sequences (Fig. S1). Interestingly, among the *Agaricales* and *Russulales* sequences a small number displayed N-terminal Lys instead (and in a few occasions Asn). The Arg and Lys sequences clustered together phylogenetically; meaning that for species of these orders Arg and Lys are perhaps interchangeable at the N-terminal position. We found that many Arg-AA9 members have signal peptides that indicate they would be secreted. As an example, we found that for the Arg-AA9 of *Pycnoporus coccineus BRFM310* (JGI#1430659) the SignalP server (37) reports with high certainty a putative signal peptide with predicted cleavage before the Arg, meaning that the protein is most likely targeted for the secretory pathway.

To augment previous reports on transcriptional regulation of Arg-AA9-encoding genes during fungal growth on plant biomass (33-35), we sought additional evidence in other fungal species. Using RNAseq analysis (as previously described in (38)), we identified a number of genes in several *Polyporales* species (with around 70% sequence identity to *Ls*AA9B, Fig. 2), that were upregulated when the fungi

were cultivated on cellulose or complex biomass compared to control growth on maltose as an easily uptakable carbon source (Table 1). We found Arg-AA9s in the *Polyporales* fungi of *Trametes ljubarskyi* BRFM 1659 (JGI#971189), *Leiotrametes sp.* BRFM 1775 (JGI#236437), *Pycnoporus sanguineus* BFRM 1264 (JGI#1740610), *Pycnoporus coccineus* BRFM 1662 (JGI#248104) and *Trametes elegans* BRFM 1663 (JGI#360271), which were highly upregulated at day 3 on Avicel, pine, aspen and wheat straw. In addition, these were co-regulated with several other secreted CAZymes targeting plant biomass polysaccharides (Table 1 and Fig. 2).

These results strongly suggest that these Arg-AA9s may have a role in the fungal adaptive response to (ligno)cellulosic substrates in coordination with the plant cell well targeting CAZymes that they are coexpresssed alongside (i.e. GH5_5, GH5_7 and GH12 endoglucanases and endomannanases, GH7 cellobiohydrolases, CE16 deacetylases, GH3 glucosidases, GH28 polygalacturonases and canonical AA9, Table 1).

_Ls_AA9B, an AA9 with an N-terminal Arg _from Lentinus similis,_ is secreted upon growth on hardwood pulp.

Since the bioinformatics analysis and transcriptomic data suggested that Arg-AA9s might be functionally expressed during fungal growth on biomass, we sought to structurally and functionally characterise one of these proteins. The *Polyporales* fungus *L. similis* contains at least seven genes encoding proteins that are classified as AA9 of which *Ls*AA9A has been well characterized (22,23,29). Also among these proteins is a single domain Arg-AA9 protein of 221 amino acids (aa) residues which we denote *Ls*AA9B (GH61-5 in (32), GenBank accession MN265867). The *Ls*AA9B sequence includes a signal peptide (MKTWAVLSSLALLASSVSA) which suggest that this protein is secreted by *L.similis*. Indeed, from secreted fractions obtained after induction experiments of *L.similis* with hardwood pulp, we identified peptides corresponding to the mature *Ls*AA9B protein by electrospray ionization tandem mass spectrometry (ESI MS/MS) (Fig. 3 and Table S1), conclusively showing that the *Ls*AA9B protein is produced extracellularly when the fungus grows on biomass. This is to our knowledge only the second report of identification of upregulation of an Arg-AA9 at the protein level. The first one was of *Polyporus brumalis* BRFM985 (JGI#1403153), which as reported in the supplementary table 3 of (35) was found in the fungal secretome when *P. brumalis* was grown on wheat straw.

Sequence comparison with well characterized canonical AA9 LPMOs indicated that the residues R1, N84, L158, Q167 and F169 in *Ls*AA9B were equivalent to the His brace, conserved Gln and His residues in the secondary coordination sphere, and a Tyr that occupies the axial position of the active site copper (which in *Tt*AA9E are H1, H68, H142, Q151 and Y153 (14)). Thus, it appeared that, except for Q167, the residues important for AA9 LPMO activity (13,15) were changed in *Ls*AA9B. *Ls*AA9B was heterologously expressed in *Aspergillus oryzae* and purified for further biochemical studies (see experimental procedures or (32)). As expected, with the purified *Ls*AA9B protein no enzymatic activity could be detected on cellulosic and hemicellulosic substrates (PASC, AZCL-cellulose or AZO-Xylan), under the experimental conditions tested as compared to the previously characterized *Ls*AA9A that showed clear activity on PASC and AZCL-cellulose.

The _Ls_AA9B X-ray crystal structure reveals a fold very similar to active AA9 LPMOs.

Crystals of *Ls*AA9B were obtained in a range of conditions with combinations of polyethylene glycols (PEGs) (see Table 2 for further details). Structures could be determined by molecular replacement (MR) in space group $P2_12_12_1$ to better than 1.60 Å resolution. We obtained structures of both a naturally glycosylated and a partly deglycosylated form of *Ls*AA9B (*Ls*AA9B and *Ls*AA9B_deglyc). In, addition, using the glycosylated *Ls*AA9B batch, we obtained structures after soaking experiments with either a solution containing transition metals (*Ls*AA9B_metalsoak) or a solution containing xylotetraose (*Ls*AA9B-Xyl4). For all structures, all 221 residues corresponding to the mature protein could be modelled in the electron density and the final structures showed good

refinement statistics (Table 3). The structure of *Ls*AA9B revealed a typical LPMO topology (Fig. 4a and Fig. S2) consisting of a central immunoglobulin-like β-sandwich (β1:3-9; β2:52-55; β3:59-65; β4:88-94; β5:108-114; β6:134-138; β7:146-156; β8:167-178). The AA9 structurally most closely related to *Ls*AA9B (based on RMSD of the C_α trace calculated with Superpose(39), see Table S2) is *Nc*AA9C(25) (r.m.s.d. of 1.04 Å) (PDB entry 4D7U) followed by *Nc*AA9D(40) (PDB entry 4EIR) and *Ls*AA9A(22) (PDB entry 5ACH) all harbouring an extended L3 loop(25) implicated in cellulose binding. *Nc*AA9F (PDB entry 4QI8) and in fact *Tt*AA9E (PDB entry 3EJA, used as MR search model) are the more distantly related structures, while the *Ta*AA9A(9), *Nc*AA9M(40) and *Tr*AA9B(41) structures (PDB entries 3ZUD, 4EIS and 2VTC, respectively) which all possess an extended L2 loop(40) (which is absent in *Ls*AA9B) show intermediate RMSDs. However, all are structurally quite similar to *Ls*AA9B with RMSDs of less than 2 Å (Fig. S2 and Table S2).

The *Ls*AA9B structures show striking differences compared to canonical LPMO active sites, but share structural features with AA9 LPMOs active on oligosaccharides.

In all the *Ls*AA9B structures the N-terminal Arg (R1) and the residues, N84, F169, L158 and Q167 (equivalent to H1, H68, H142, Q151 and Y153 in *Tt*AA9E important for AA9 LPMO activity) were clearly defined in the electron density map (Fig. 4b). These residues in *Ls*AA9B (Fig. 4c and Fig. S2) are arranged in a similar configuration compared to *Nc*AA9C (the closest structural match, PDB 4D7U), as defined by the backbone positions, but are not associated with any metal or positioned compatibly with a likely metal binding function (guanidinium group of R1 points away from N84)(Fig 4). This confirms that residues important for AA9 LPMO activity are changed in *Ls*AA9B, as was indicated by sequence comparison (note that in other LPMO families with demonstrated activity Phe is found in positions equivalent to F169). Furthermore, the structure does not indicate that nearby residues could functionally compensate for the loss of the conserved residues to provide a copper binding site/LPMO activity.

Near the N-terminal Arg, *Ls*AA9B Y206 is found (Fig. 2) in a position equivalent to Y204 in *Nc*AA9C and Y203 in *Ls*AA9A, shown to be involved in AA9 LPMO oligosaccharide interactions (21-23). In those structures the Tyr residues are well defined and make hydrogen bonds to the backbone amide of the conserved His of the secondary copper coordination sphere (H147 in *Ls*AA9A and H155 in *Nc*AA9C). However, in *Ls*AA9B the rotamer of the apolar L158 sidechain interferes with the hydrogen bonding of the backbone amide to Y206, which in many of the structures was not well defined (Fig. S3ab). In addition, *Ls*AA9B has a number of residues, such as N26 and H66 which are in positions equivalent to those involved in oligosaccharide binding in *Ls*AA9A (*Ls*AA9A N26, H66, N67, S77, E148, D150, Y203 are equivalent to *Ls*AA9B N26, H66, T67, T83, N159, E161, Y206) (Fig. S2). Almost all of the Arg-AA9 sequences have Asn, His/Arg and Tyr/Phe in the position of N26, H66 (Arg is found in this position in the oligosaccharide active *Cv*AA9A LPMO (23)) and Y206 (Fig. S1). However, for N159 and E161 (equivalent to residues in *Ls*AA9A interacting with the substrate in the minus subsites) there seems to be virtually no conservation (Fig. 5 and Fig. S1).

Other structural features of *Ls*AA9B

O-linked and N-linked glycosylation at T59 and N134, respectively were also observed in all of the *Ls*AA9B structures. From T59 one α-linked mannosyl unit could be modelled in all cases. In the *Ls*AA9B structure, two N-acetylglucosamine (GlcNAc) units as well as one poorly defined mannosyl unit could be modelled from N134. In the *Ls*AA9B_deglyc structure, one well-defined and one less defined GlcNAc unit were modelled, indicating that the protein batch was not fully deglycosylated (as confirmed by mass spectrometry (MS)).

Near N134 a small pocket was found (mainly formed by Y71 and a GlcNAc unit from the N-glycosylation site) (Fig. 5) which was often occupied by electron density that could be attributed to crystallization condition components (e.g. glycine, sulphate or a MES molecule). In the *Ls*AA9B_metalsoak structure a MES molecule could quite confidently be modelled in this position (Fig. 5b) interacting

with H65 and Y71, and in part with the GlcNAc unit (Fig. 5c). A putative functional role is supported by full conservation of H65 and Y71 (in loop L3) among AA9-Arg, while the glycosylated N134 can be substituted with Phe/Tyr which could assume a similar structural role in forming a pocket. Additional full conservation of three prolines (P76, P79 and P82) suggests that the extended L3 loop has conserved rigid structure in this AA9 subgroup (Fig. 5d and Fig. S1). In some canonical AA9 LPMOs charged or polar residues in the extended L3 loop are putative determinants of specificity towards soluble substrates.

For many of the crystal structures determined during this work, Fourier difference map density was visible proximal to S25. Since additional evidence suggested a putative phosphorylation site in *Ls*AA9B (see section *Indications of a potential phosphorylation site in LsAA9B*), the final structure includes a phosphoserine (pS25), which fits well the electron density in two of the four structures (*Ls*AA9B_deglyc and *Ls*AA9B_metalsoak). In these structures, the pS25 and S25 are modelled in alternative conformations with 70% and 30% occupancy, respectively (Fig. 4b).

*Ls*AA9B is not a copper binding protein

When inspecting the *Ls*AA9B structures no bound metals were found. We speculated whether R1 would be able to adopt a different conformation in response to addition of transition metals allowing for *Ls*AA9B to coordinate a metal cofactor at this surface. Co-crystallization and soaking experiments with transition metals were attempted with solution mixtures containing Fe^{2+}, Cu^{2+}, Mn^{2+} and Co^{2+}. X-ray diffraction data were collected at an appropriate wavelength of 1.35Å~9.184 keV to ensure an anomalous signal would be obtained to allow location of any bound metal ions in the crystal structure. Following map generation, no peaks that could be interpreted as bound metals could be found when inspecting an anomalous Fourier difference map. The highest peaks were found near the sulphur atoms in the Cys residues of the protein, confirming that these metals did not bind in any of these experiments. None of the structures inspected during this work showed any sign of bound metals.

Metal binding was also investigated by thermal shift assays with differential scanning fluorimetry (DSF) and differential scanning calorimetry (DSC). For three completely independent measurements with DSF, the T_i *for* *Ls*AA9B was 65.5 ± 0.75 °C. The presence of copper ions at slightly over stoichiometric amounts (0.048 mM copper acetate) resulted in a decrease of T_i to 62.4 °C. A similar approximate 4 °C decrease of T_m was found with the addition of copper measured by DSC. No significant change of T_m was found for a range of other metals by DSC.

As a final step, microPIXE (Proton Induced X-ray Emission) analysis (42) was used to detect any additional elements present in the sample with no added metals, in order to ascertain whether *Ls*AA9B copurified with any metal ions which might hint at function. The analysis did not reveal significant amounts of copper, iron or any other metal typically associated with redox enzymes present in the sample above trace levels (Fig. S4 and Table S3) (calculations suggest a lower detection limit for copper of 4×10^{-3} atoms per protein molecule).

Indications of a potential phosphorylation site in *Ls*AA9B

Although it did not reveal any natively bound metals, the PIXE analysis detected the presence of a single atom of phosphorus in two point spectra (Fig. S4 and Table S3) for every protein molecule in the sample. In many of the structures we obtained during this study, density was observed in close proximity to the sidechain of S25, upon inspection of the Fourier difference map. In light of these data, this was therefore interpreted as being a phosphoserine. Modelling a phosphorylated Ser (pS25), with the phosphate group hydrogen bonding to the sidechain of the N-terminal Arg (2.9 Å), led to an improved fit to the electron density map and model quality (improved R-factor and clash score). The position of the phosphate group does not appear to interfere with possible copper binding (Fig. S3c)

Following this observation, we tried to identify phosphorylation using MS. Making several measurements of the EndoH deglycosylated *Ls*AA9B batches with Matrix Associated Laser Desorption Ionization Time Of

Flight (MALDI-TOF) we identified the largest peaks with m/z values that could correspond to *Ls*AA9B with glycosylation (N-GlcNAc and O-Man) and a phosphorylation (23.255 Da + 203 Da + 162 Da + 80 Da = 23.670 Da). Using MALDI-TOF to measure the glycosylated *Ls*AA9B we found that the m/z values could correspond with a single phosphorylation, two GlcNAc and varying degrees of mannosylation (m/z increasing by 162 Da for every mannosyl unit). MS techniques with a trypsin-digested preparation of *Ls*AA9B only identified the non-modified peptide (total sequence coverage of 55%), thus conclusive evidence for phosphorylation is lacking.

However, it is interesting to observe that in positions corresponding to S24 and S25 in *Ls*AA9B most of the analysed Arg-AA9 sequences have Ser or Thr (Fig 4b, Fig 5 and Fig S1) which are potential phosphorylation sites, as predicted by the NetPhos 3.1 Server (43). Thus, at least theoretically, it seems that a number of the Arg-AA9 sequences could be phosphorylated.

Investigation of oligosaccharide binding

As already stated *Ls*AA9B contains residues (e.g. N26, H66 and Y206) equivalent to some of those involved in *Ls*AA9A oligosaccharide binding (Fig. S2). When superimposing either the *Ls*AA9A-Cell5 (PDB 5NLS) or *Ls*AA9A-Xyl5 (PDB 5NLO) complex onto the *Ls*AA9B structure, we found that some of these residues in *Ls*AA9B came within hydrogen bonding distance of the Cell5 and Xyl5 ligands of the *Ls*AA9A complex structures. In addition, we observed that the sugar moiety of the +2 subsite in *Ls*AA9A (the C6-hydroxyl and the C2-hydroxyl for Cell5 and Xyl5, respectively) could interact with the putative phosphorylation in *Ls*AA9B. Thus, *Ls*AA9B carbohydrate binding was investigated by cocrystallization with monosaccharides, and crystal soaking experiments and thermal shift assays with oligosaccharides.

Addition of oligosaccharides at 50 mM concentration in solution caused very little change in T_i (66.1 °C with maltopentaose; 65.9 °C with Cell5; 64.7 °C for Xyl5; 65.0 °C for chitopentaose). Addition of stoichiometric amounts of copper (with or without 200 mM NaCl as in (22)) decreases the T_i to 62.4 °C (T_i of 61.9 °C in the presence of NaCl), still with no upward thermal shift on addition of oligosaccharides (T_is of 59.9-61.4°C with copper alone and T_is of 61.3-62.2 °C with copper and NaCl). No crystals appropriate for data collection were obtained from cocrystallization experiments. Soaking experiments with cellotetraose (Cell4), cellopentaose (Cell5) or xylotetraose (Xyl4) oligosaccharides were done in conditions with $MgCl_2$ and $CaCl_2$ (30-100 mM) since the presence of chloride ions has been shown to enhance oligosaccharide binding to canonical AA9 (22). No electron density that could be interpreted as cellooligosaccharides was identified at the protein surface.

However, we did identify a Xyl4 ligand bound at the surface of *Ls*AA9B, but the binding mode is distinct compared to *Ls*AA9A-Xyl5 (Fig. 6a). One explanation for this could be that glycosylation of a symmetry related molecule occludes this binding site (Fig. 6b). Instead the Xyl4 ligand bound in a small cleft formed by the residues 41-49 (TGFIQPVSK) and 99-10 (TQS). Three xylosyl units of the Xyl4 ligand (numbered 1-3 from the reducing end) are well defined and could easily be modelled in the electron density, while the fourth xylosyl unit at the non-reducing end is only partially defined (Fig. 6c). Many of the glycosidic torsion angles of the Xyl4 ligand deviate from ideal values e.g. between xylosyl unit 1 and 2 (in particular the Ψ- torsion angle) (Table S4).

Discussion

*Ls*AA9B can be heterologously expressed and is a well folded and stable protein, with melting temperatures comparable to enzymatically active members of the family (44-47), but showed no detectable LPMO activity under the condition tested. The X-ray crystal structure of *Ls*AA9B in fact confirms a complete disruption of the copper binding site that is the hallmark of LPMOs (9) and absence of other metal binding sites, even though the overall fold remains very similar to that expected. We clearly demonstrated that residues R1 and N84 in *Ls*AA9B are positioned equivalently to the His

brace of AA9 LPMOs, but that neither of these (nor any other) residues coordinate copper. In all structures (with or without the putative phosphorylation site modelled and/or metals added), R1 in fact assumes a rotamer conformation which is not optimal for forming a His-brace like metal binding site. PIXE spectroscopy (42), crystal soaking experiments with metal ions and thermal shift analyses in solution failed to indicate binding of any metal. This is consistent with a survey of metal binding sites in protein structures reporting that Asn and Arg interact with elements such as K, Na, Mg and Ca more frequently than with transition metals (of which only Mn is reported), while in contrast His is the most common protein residue to bind Cu (48). Thus, our work confirms the expectation that *Ls*AA9B and other Arg-AA9s are very unlikely to be Cu-dependent LPMOs.

At the same time, evidence presented here and elsewhere shows that a number of genes (in wood-degrading fungal *Polyporales* species) encoding Arg-AA9 proteins are upregulated during fungal growth on plant biomass alongside lignocellulose degrading CAZymes (most commonly members of GH and AA families). The majority of the Arg-AA9 sequences possess a signal peptide (with predicted processing before Arg) expectedly targeting the mature N-terminal Arg proteins for the secretory pathway. Indeed both *Ls*AA9B and the Arg-AA9 of *Polyporus brumalis* BRFM 985 (JGI#1403153) have been detected in fungal secretomic data. In addition, the mature *Ls*AA9B protein with N-linked glycosylation is secreted when recombinantly produced in *Aspergillus* from a construct with the native *L.similis* signal peptide. Thus, there is circumstantial evidence that Arg-AA9 proteins, like conventional LPMOs, could be involved in plant biomass degradation. The open question remains how, and while this work does not provide a conclusive answer, the structural analysis suggest a number of new avenues for further investigation.

For example, we found that the relatively conserved L3 loop is likely structurally rigid in the majority of Arg-AA9 proteins and together with the N134 glycosylation forms a small pocket. In the *Ls*AA9B_metalsoak structure, a MES molecule is interacting with the completely conserved H65, Y71 in this pocket. It is possible that the MES molecule mimics biologically relevant interactions and notably, MES bear some resemblance to phenolic compounds (e.g. p-coumaryl- coniferyl, sinapyl alcohols) that make up lignin, and thus the possible connection to lignin degradation/detoxification pathways should be investigated in the future.

Structural comparison suggested that *Ls*AA9B could bind oligosaccharides in a manner similar to *Ls*AA9A, but this could not be confirmed by our structural studies, perhaps because of glycosylation from a symmetry-related molecule occluding this binding site. In contrast, we obtained an *Ls*AA9B structure with a Xyl4 ligand bound at a distinct binding site, in a small cleft made up mostly by relatively conserved residues (no similar cleft is present in *Ls*AA9A where this space is occupied by the peptide stretch VDNRVV formed by residues 43 to 48). It is difficult to establish whether the binding site is biologically relevant: the expected structural conservation of this region in Arg-AA9 and the fact that there is only one hydrogen bond of Xyl4 with a symmetry-related molecule suggest this is not a crystal artefact. On the other hand the bound conformation of Xyl4 deviates considerably from what is usually observed in xylooligosaccharides and xylan (49) (Table S4), and Xyl5 in solution caused no thermal shift. The path connecting this binding site and the binding site in *Ls*AA9A is definitely occluded by crystal contacts, and thus binding of longer oligosaccharides (connecting the two binding sites) is not possible in this crystal form. However the discovery of a Xyl4 binding site warrants additional investigations of polysaccharide binding by this and other Arg-AA9.

Finally, we found indications that S25 could be phosphorylated in *Ls*AA9B as supported by some high-resolution structures and by PIXE data, and consistent with MALDI-TOF analysis, though a phosphorylated peptide could not be conclusively identified. Intriguingly, the phosphoserine modelled in the structure interacts with the N-terminal Arg (2.9 Å hydrogen bond) and could additionally potentially interact with bound polysaccharides near the binding sites observed for *Ls*AA9A and *Ls*AA9B.

Potential similar phosphorylation sites (equivalent to S24 and S25 of *Ls*AA9B, Fig. 4b and Fig. 5d) are also predicted for Arg-AA9 protein sequences (presented here in Fig. 2 and Table 1 and previously (34,35,38)), for which the transcription of the corresponding genes were upregulated during fungal growth on plant biomass. In addition, in the Fungi Phosphorylation Database (FPD) (50) documented Ser phosphorylation sites in fungal species (including *Aspergillus*) can be found which are consistent with the *Ls*AA9B sequence surrounding the S25 phosphorylation, supporting the notion that phosphorylation could play a role in fungal plant biomass degradation. It is known that proteins can be phosphorylated in the secretory pathway (51,52), in mammalian cells by the Fam20C kinase (53,54), and phosphorylation does appear to play a role extracellularly during some biological events (e.g. during microbial host infection (55-57)). Thus, it is possible that also phosphorylated Arg-AA9s could exist extracellularly, since the evidence points to an extracellular location for these proteins.

In summary, our results indicate importance of Arg-AA9s in fungal adaptation to biomass degradation. We have shown that the transcription of Arg-AA9 genes are upregulated in fungal species belonging to the phylogenetic class of *Agaricomycetes* (of which many are wood or litter decayers) when cultivated on plant biomass. We have determined this first structure of an Arg-AA9 protein, *Ls*AA9B, (belonging to the wood-decaying *Lentinus* genus (58)) and have identified interesting structural features, including a potential carbohydrate-binding cleft, a small conserved pocket and a potential phosphorylation site. Sequences conservation suggests that these features are found in the majority of the Arg-AA9s as well. These findings provide a solid framework for future investigations of this subgroup of AA9 and for the pursuit of their potential role in fungal biology. The original structures of AA9 (14) and AA10 (24) family members were of key importance towards the subsequent identification of the nature of LPMO action on biomass. We hope that the structure of *Ls*AA9B will likewise act as the starting point for further characterisation of these unusual AA9 family members.

Experimental procedures

Fungal transcriptomic and secretomic data.

Lentinus similis was cultivated at 30 °C on a basic fungal media including 0.5% (w/v) hardwood BCTMP (bleached chemi-thermo mechanical pulp) for induction experiments. Samples were taken after 0, 3, 5, 7 and 10 days in order to analyse secreted proteins and identify *Ls*AA9B with ESI-MS/MS. Tryptic digests were prepared by a Filter-aided sample preparation (FASP) method. Following digestion the extracted peptides were analyzed on a nano LC-MS/MS system: UltiMate 3000 RSLCnano / LTQ Orbitrap Velos Pro (Thermo/Dionex). For protein identification the data were searched against the complete *L. similis* proteome (internal database) using the Mascot search engine (Matrix science) in the Genedata Expressionist software (1% False Discovery Rate cutoff). Relative protein concentrations were calculated by label free quantification from peptide volumes using a Hi3 standard method in Genedata Expressionist.

Pycnoporus coccineus BRFM 1662, *Trametes ljubarskyi* BRFM 1659, *Leiotrametes* sp. BRFM 1775 and *Trametes elegans* BRFM 1663 strains were obtained from the CIRM collection at the National Institute of Agricultural Research[1]. Transcriptome and secretome data were collected from triplicated independent three day-cultures in the presence of either 20 $g \times l^{-1}$ maltose, 15 $g \times l^{-1}$ Avicel, 15 $g \times l^{-1}$ ground wheat straw, 15 $g \times l^{-1}$ ground pine wood, or 15 $g \times l^{-1}$ ground aspen wood as the sole carbon source. RNA libraries were prepared and sequenced on Illumina HighSeq-2500 as described in (59). Secreted proteins were collected from the same cultures, dia-filtered and identified by ESI-MS/MS (as in (59)). Transcript reads were analysed as described in (38). For each strain, the genes with similar transcription profiles on the five carbon sources were grouped into nodes using the Self-organizing maps Harboring Informative Nodes with Gene Ontology Pipeline (SHIN+GO; (34)). All

sequence data are available on the Mycocosm public database at Joint Genome Institute[2] (60).

_Ls_AA9B protein production, purification and activity measurements

An AA9 protein from *L. similis* (*Ls*AA9B) was recombinantly produced in *Aspergillus oryzae* MT3568 and purified as described in the patent literature (32) (similar to the previously characterized *Ls*AA9A (22,31). The mycelium was removed by filtration and the broth collected for protein purification by chromatography with a 50 mm diameter 167 ml Butyl-ToyoPearl 650 column (ToSoh BioSciences, Stuttgart, Germany) with a gradient of 0-100% of buffer A (25 mM Tris-HCl, 1.0 M ammonium sulphate, pH 7.5) and buffer B (25 mM Tris-HCl pH 7.5). To ensure purity suitable for structural studies, two further purification steps were applied. After buffer change by ultrafiltration (Vivaspin, Sartorius) the sample was applied to a 100 ml Q-Sepharose column (Sigma-Aldrich) in 20 mM Tris-HCl pH 8.0. The protein was eluted with a gradient of 0 to 500 mM NaCl. Fractions containing protein were further purified on a 26 mm Superdex75 size exclusion (Sigma-Aldrich) with isocratic 20 mM MES, 125 mM NaCl, pH 6.0. All protein batches were buffered in 20 mM MES pH 6.0. Activity was measured as previously described by depolymerisation of AZCL-cellulose (61) and PASC (9) in parallel on *Ls*AA9A and *Ls*AA9B in the presence of a variety of electron donors (pyrogallol, 4-OH-5-CH3-3-furanone, ascorbate, cysteine) and under similar conditions. In addition, *Ls*AA9B activity was measured (according to manufacturer instructions) on AZO-xylan (purchased from Megazyme; product code S-AXBL) without and with equimolar concentration (1 µM) of Cu or Mn in the presence of ascorbate (1 mM).

Mass spectrometry analysis of purified _Ls_AA9B

MS was performed using MALDI-TOF. The matrix was prepared by saturating a TA solution (0.1% (v/v) triflouricacteic acid (TFA) and acetonitrile (ACN) in a 2:1 (v/v) ratio) with sinapic acid. Matrix and protein sample were sequentially applied to the target plate (in a final 2:1 ratio) and left for solvent to evaporate. The TOF experiments were performed in linear mode calibrated for 10-50 kDa molecules.

Products following a trypsin digest of the heterologously expressed *Ls*AA9B batches were analysed using a Thermo Fisher Scientific Q-Exactive HF-X Orbitrap mass spectrometer which was operated in positive mode with a MS1 resolution of 120,000. A top 10 method was utilized and a resolution of 45,000 for MS2 fragment scans was utilized. Peptides were separated on a 15 cm column (75 µM inner diameter) packed in-house with 1.9 µM C18 particles. EASY-nLC 1200 nano-liquid chromatography system coupled to the mass spectrometer was utilized to separate injected peptides over an increasing gradient of buffer B (80% acetonitirile, 0.1% formic acid). The collected raw data were analyzed using MaxQuant (version 1.6.1.11).

Crystallization and data collection

Deglycosylation of protein for crystallization was performed in 20 mM MES pH 6.0, 125 mM NaCl, by incubating about 10 µl endoglycosidase H (from Roche Diagnostics, 11643053001) per mg *Ls*AA9B protein overnight at room temperature. Crystallization was carried out by sitting drop vapour diffusion in 96-well MRC-2 plates using an Oryx8 crystallization robot (Douglas Instruments) with drop sizes of 0.3-0.4 µl (with protein:reservoir:mili-Q water ratios of either 3:1:1, 3:1:0, 1:1:1 or 1:1:0) and with 100 µl reservoirs in 24-well VDX plates with drop sizes of 2-4 µl and reservoirs of 1 ml. Morpheus screens (62) were set up using both a naturally glycosylated batch (11 mg/ml) and a deglycosylated (with endoglycosidase H from Roche Diagnostics, 11643053001) batch (4-5 mg/ml) of *Ls*AA9B. For all batches, crystals could be obtained in quite similar conditions (Table 2).

With an initial protein batch a crystal grew in Morpheus condition #28 in well C4 (Table 2), from which a complete 1.6 Å resolution X-ray diffraction dataset of 155 frames (155°) was collected. With a glycosylated *Ls*AA9B, intergrown crystal plates were obtained in Morpheus condition #85 (Table 2). A single plate was mounted separately from a cluster and a dataset of 200 images (200°) was

collected. Crystals were also obtained with a deglycosylated *Ls*AA9B (Morpheus condition #6 in well A6) and were reproducible in MRC-2 plates with protein concentration from 2-5 mg/ml and 24-48 %(w/v) PPT4 (a 1:1:1 mixture of MPD(racemic), PEG100 and PEG3350), and a dataset of 324 images (162°) was collected (Table 2). For all datasets the data collection was carried out with the crystals mounted in nylon loops at 100K without additional cryoprotectant, since the Morpheus screen composition is already cryo-protecting.

Cocrystallization experiments with transition metals were performed in VDX plates with 33–42%(w/v) of PPT1 (a 2:1 mixture of PEGMME500 and PEG20.000) or PPT4 at pH 6.5, 7.5 or 8.5 and a metal mixture additive of 5–50mM $CuSO_4$, $MnCl_2$ and $CoCl_2$, but failed to provide crystals of sufficient quality. For soaking experiments, crystals were grown in PPT4 at pH 7.5 (with either carboxylic acids or amino acids or monosaccharides additives). Transition metal mixture solutions were added to an approximate final concentration of 6 – 7 mM of each of Fe^{2+}, Cu^{2+}, Mn^{2+} and Co^{2+}. Crystals were mounted at different time points from 1 min to 5h20min and datasets (generally to better than 1.7 Å resolution) were collected as before but at a wavelength of 1.35Å (9.184 keV) at which the anomalous signal should be significant to detect all of the metals. No peaks that could account for bound metals were found when inspecting an anomalous Fourier difference map. In fact, the highest peaks were found near the sulphur atoms in the cysteine residues.

Optimized cocrystallization with monosaccharides additives (20-100 mM) were carried out in VDX plates with 33-42 %(w/v) of either PPT1 or PPT4 at pH 6.5, 7.5 or 8.5 (since intergrown protein crystals in conditions with monosaccharide additives were observed in the screens). However, no crystals appropriate for data collection were obtained. Soaking experiments with oligosaccharides were carried out in MRC-2 plates using crystals grown with 6–11 mg/ml protein in 27-42 %(w/v) PPT1 at either pH 6.5 or pH 8.5 with additives of either 30 mM or 100 mM of $MgCl_2$ and $CaCl_2$. Crystals were transferred to reservoir solutions containing 400-800 mM of either cellotetraose (Cell4), cellopentaose (Cell5) or xylotetraose (Xyl4) and soaked for 40 – 80 minutes.

All data were collected at the MX beamline I911-3 at the MAX-IV laboratory, Lund (SE) or by remote access at beamline ID23-1 at the ESRF, Grenoble, (FR). All crystals were isomorphous and processed and scaled with XDS/XSCALE (63) in $P2_12_12_1$ with similar cell dimensions (Table 3).

Structure determination and refinement

From sequence alignment the closest AA9s with structures available at the time were *Tt*AA9E (pdb entry 3EJA (14)) and *Nc*AA9D (pdb entry 4EIR (40)) with 42.9% and 45.1% sequence identity, respectively. Molecular Replacement (MR) with MOLREP (64) using only the protein coordinates (or Sculptor (65) modified models) of either of PDB entry 3EJA or 4EIR as a search models gave solutions which following one round (10 cycles) of restrained refinement (with Refmac5(66), CCP4 suite(67)) resulted in R-factors of ~ 40% after which parts of the map could be relatively easily interpreted. Modelling in COOT (68,69) followed by several rounds of restrained refinement resulted in an initial preliminary structure, which was used to refine additional structures obtained later. Structures of *Ls*AA9B were refined in Refmac5 (CCP4 suite) (66,67) using R_{free} flags imported from the previous structure factor file. All *Ls*AA9B structures obtained after this point were refined with anisotropic B factors for all protein atoms (including glycosylation) and isotropic B-factors for all other atoms.

Construction of multiple sequence alignments and structural comparison

Superpose (CCP4 suite) was used for calculation of the RMSD C_α trace using Secondary-structure matching (SSM) (39,67). Multiple sequence alignment was constructed using a standalone version of STRAP (STRuctural Alignments of Proteins) and programs within (70,71). For multiple sequence alignment, MAFFT was used (72). For the structure based sequence alignment TMalign was used to align the C_α trace of the AA9 3D structures (73). Mapping of sequence on the *Ls*AA9B structure was done using CAMPO (74)

available in the PyMod 2.0 plugin for PyMOL (75).

Thermal shift assays

Thermal shift assays with DSF were carried out using intrinsic fluorescence using the NanoTemper Tycho NT.6 (Nano Temper technologies) according to the manufacturer's instructions. Tycho NT.6 follows the unfolding process by recording sample fluorescence at 330 nm and 350 nm during thermal unfolding. A constant heating rate of 30 °C / min is applied to the sample, heating from 35 °C to 95 °C. A shift in T_i to higher temperatures is usually indicative of ligand binding. 10 µl of each sample were prepared, incubated for 5 min and then loaded in capillaries for measurements. *Ls*AA9B was at a concentration of 1 mg/ml in 20 mM MES, pH 6.0. All measurements were carried out at least in triplicates. Thermal shift assays with capillary DSC (Northampton, MA) were carried out with heating rate of 1.5 °C / min with *Ls*AA9B at pH 5.0 in the presence of 1 mM divalent metal ions (Cu^{2+}, Ni^{2+}, Fe^{2+}, Zn^{2+}, Mn^{2+}, Ca^{2+}) or 1 mM DTPA (as chelating agent).

Particle Induced X-Ray Emission (MicroPIXE) Analysis

*Ls*AA9B, purified as described above, was prepared for microPIXE analysis by passing down a 16/600 Superdex 75 (GE Healthcare) column in order to buffer exchange into 20 mM ammonium acetate pH 5.5, 200 mM KBr. The final sample was then concentrated to 7.7 mg/ml using a 10-kDa cut off VivaSpin Concentrator (Sartorius).

0.1 µl of sample was placed onto a 4 µm polypropylene film (Prolene supplied by Fluxana GmbH & Co., Germany. www.fluxana.com) which was then allowed to dry naturally in a closed environment to prevent dust contamination. The sample was mounted in the path of a 2.5 MeV proton beam with a beam diameter of approximately 2.5 µm at the Ion Beam Centre, University of Surrey[3]. Prior to the experiment a glass standard was analysed and validated to ensure accurate elemental quantitation. Data were collected by first scanning the proton beam across the sample to generate a set of elemental maps. Point spectra were then measured from two distinct regions of the sample based on the distribution of sulphur in the map, signifying where protein was located. The collected spectra were processed using the Q-factor method(76) as implemented in the OMDAQ-3 software[4] and converted into the number of metal atoms per protein molecule as described in (42).

Acknowledgements: This work was partly performed during the Critical Enzymes for Sustainable Biofuels from Cellulose (CESBIC) project supported by a grant from the European Research Agency—Industrial Biotechnology Initiative (http://www.era-ib.net/cesbic) and the Danish Council for Strategic Research (grant numbers 12-134923 and 12-134922). G.R.H was supported by the Biotechnology and Biological Sciences (BBSRC) through grant BB/L000423/1 awarded to Paul H. Walton and Gideon Davies. The work was also partly performed with the support of the Novo Nordisk Foundation (NNF17SA0027704). K.E.H.F. thanks the Carlsberg Foundation for financial support through an Internationalisation Postdoc Fellowship (grant n°CF16-0673 and n°CF17-0533), and has received the support of the EU in the framework of the Marie-Curie FP7 COFUND People Programme, through the award of an AgreenSkills+ fellowship (under grant agreement n°609398). 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 N283570) funded travel to synchrotrons. MAX-IV laboratory, Lund, Sweden and the ESRF, Grenoble, France, are thanked for beamtime and staff assistance. The work conducted by the U.S. Department of Energy Joint Genome Institute (JGI), a DOE Office of Science User Facility, was supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We thank Rebecca A. Batters for assistance with the MicroPIXE data analysis and Jesper V. Olsen at the Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, for his involvement with advanced mass spectrometry and phosphoproteomic data. K.E.H.F., J.-C.N.P. and L.L.L. are members of ISBUC Integrative Structural Biology at the University of Copenhagen (www.isbuc.ku.dk).

Conflicts of interest: M.T., C.I.J. and N.S. are employed by Novozymes A/S, which is a major enzyme producing company. The remaining authors declare that they have no conflicts of interest with the contents of this article.

References

1. Johansen, K. S. (2016) Lytic Polysaccharide Monooxygenases: The Microbial Power Tool for Lignocellulose Degradation. *Trends in plant science* **21**, 926-936
2. Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sorlie, M., and Eijsink, V. G. (2010) An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. *Science* **330**, 219-222
3. Bissaro, B., Rohr, A. K., Muller, G., Chylenski, P., Skaugen, M., Forsberg, Z., Horn, S. J., Vaaje-Kolstad, G., and Eijsink, V. G. H. (2017) Oxidative cleavage of polysaccharides by monocopper enzymes depends on H2O2. *Nature chemical biology* **13**, 1123-1128
4. Hemsworth, G. R., Henrissat, B., Davies, G. J., and Walton, P. H. (2014) Discovery and characterization of a new family of lytic polysaccharide monooxygenases. *Nature chemical biology* **10**, 122-126
5. Lenfant, N., Hainaut, M., Terrapon, N., Drula, E., Lombard, V., and Henrissat, B. (2017) A bioinformatics analysis of 3400 lytic polysaccharide oxidases from family AA9. *Carbohydrate research*
6. Levasseur, A., Drula, E., Lombard, V., Coutinho, P. M., and Henrissat, B. (2013) Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. *Biotechnol Biofuels* **6**
7. Sabbadin, F., Hemsworth, G. R., Ciano, L., Henrissat, B., Dupree, P., Tryfona, T., Marques, R. D. S., Sweeney, S. T., Besser, K., Elias, L., Pesante, G., Li, Y., Dowle, A. A., Bates, R., Gomez, L. D., Simister, R., Davies, G. J., Walton, P. H., Bruce, N. C., and McQueen-Mason, S. J. (2018) An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion. *Nature communications* **9**, 756
8. Yadav, S. K., Archana, Singh, R., Singh, P. K., and Vasudev, P. G. (2019) Insecticidal fern protein Tma12 is possibly a lytic polysaccharide monooxygenase. *Planta*
9. Quinlan, R. J., Sweeney, M. D., Lo Leggio, L., Otten, H., Poulsen, J. C. N., Johansen, K. S., Krogh, K. B. R. M., Jorgensen, C. I., Tovborg, M., Anthonsen, A., Tryfona, T., Walter, C. P., Dupree, P., Xu, F., Davies, G. J., and Walton, P. H. (2011) Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 15079-15084
10. Frandsen, K. E. H., and Lo Leggio, L. (2016) Lytic polysaccharide monooxygenases: a crystallographer's view on a new class of biomass-degrading enzymes. *IUCrJ* **3**, 448-467
11. Vaaje-Kolstad, G., Forsberg, Z., Loose, J. S., Bissaro, B., and Eijsink, V. G. (2017) Structural diversity of lytic polysaccharide monooxygenases. *Current opinion in structural biology* **44**, 67-76
12. Tandrup, T., Frandsen, K. E. H., Johansen, K. S., Berrin, J. G., and Lo Leggio, L. (2018) Recent insights into lytic polysaccharide monooxygenases (LPMOs). *Biochemical Society transactions* **46**, 1431-1447
13. Lo Leggio, L., Welner, D., and De Maria, L. (2012) A structural overview of GH61 proteins - fungal cellulose degrading polysaccharide monooxygenases. *Computational and structural biotechnology journal* **2**, e201209019
14. Harris, P. V., Welner, D., McFarland, K. C., Re, E., Navarro Poulsen, J. C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S., and Lo Leggio, L. (2010) Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family. *Biochemistry* **49**, 3305-3316
15. Span, E. A., Suess, D. L. M., Deller, M. C., Britt, R. D., and Marletta, M. A. (2017) The Role of the Secondary Coordination Sphere in a Fungal Polysaccharide Monooxygenase. *ACS chemical biology* **12**, 1095-1103

16. Forsberg, Z., Rohr, A. K., Mekasha, S., Andersson, K. K., Eijsink, V. G., Vaaje-Kolstad, G., and Sorlie, M. (2014) Comparative study of two chitin-active and two cellulose-active AA10-type lytic polysaccharide monooxygenases. *Biochemistry* **53**, 1647-1656
17. Agger, J. W., Isaksen, T., Varnai, A., Vidal-Melgosa, S., Willats, W. G., Ludwig, R., Horn, S. J., Eijsink, V. G., and Westereng, B. (2014) Discovery of LPMO activity on hemicelluloses shows the importance of oxidative processes in plant cell wall degradation. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 6287-6292
18. Vu, V. V., Beeson, W. T., Span, E. A., Farquhar, E. R., and Marletta, M. A. (2014) A family of starch-active polysaccharide monooxygenases. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 13822-13827
19. Couturier, M., Ladeveze, S., Sulzenbacher, G., Ciano, L., Fanuel, M., Moreau, C., Villares, A., Cathala, B., Chaspoul, F., Frandsen, K. E., Labourel, A., Herpoel-Gimbert, I., Grisel, S., Haon, M., Lenfant, N., Rogniaux, H., Ropartz, D., Davies, G. J., Rosso, M. N., Walton, P. H., Henrissat, B., and Berrin, J. G. (2018) Lytic xylan oxidases from wood-decay fungi unlock biomass degradation. *Nature chemical biology* **14**, 306-310
20. Aachmann, F. L., Sorlie, M., Skjak-Braek, G., Eijsink, V. G., and Vaaje-Kolstad, G. (2012) NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 18779-18784
21. Courtade, G., Wimmer, R., Rohr, A. K., Preims, M., Felice, A. K., Dimarogona, M., Vaaje-Kolstad, G., Sorlie, M., Sandgren, M., Ludwig, R., Eijsink, V. G., and Aachmann, F. L. (2016) Interactions of a fungal lytic polysaccharide monooxygenase with beta-glucan substrates and cellobiose dehydrogenase. *Proceedings of the National Academy of Sciences of the United States of America* **113**, 5922-5927
22. Frandsen, K. E., Simmons, T. J., Dupree, P., Poulsen, J. C., Hemsworth, G. R., Ciano, L., Johnston, E. M., Tovborg, M., Johansen, K. S., von Freiesleben, P., Marmuse, L., Fort, S., Cottaz, S., Driguez, H., Henrissat, B., Lenfant, N., Tuna, F., Baldansuren, A., Davies, G. J., Lo Leggio, L., and Walton, P. H. (2016) The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nature chemical biology* **12**, 298-303
23. Simmons, T. J., Frandsen, K. E. H., Ciano, L., Tryfona, T., Lenfant, N., Poulsen, J. C., Wilson, L. F. L., Tandrup, T., Tovborg, M., Schnorr, K., Johansen, K. S., Henrissat, B., Walton, P. H., Lo Leggio, L., and Dupree, P. (2017) Structural and electronic determinants of lytic polysaccharide monooxygenase reactivity on polysaccharide substrates. *Nature communications* **8**, 1064
24. Vaaje-Kolstad, G., Houston, D. R., Riemen, A. H., Eijsink, V. G., and van Aalten, D. M. (2005) Crystal structure and binding properties of the Serratia marcescens chitin-binding protein CBP21. *The Journal of biological chemistry* **280**, 11313-11319
25. Borisova, A. S., Isaksen, T., Dimarogona, M., Kognole, A. A., Mathiesen, G., Varnai, A., Rohr, A. K., Payne, C. M., Sorlie, M., Sandgren, M., and Eijsink, V. G. (2015) Structural and Functional Characterization of a Lytic Polysaccharide Monooxygenase with Broad Substrate Specificity. *The Journal of biological chemistry* **290**, 22955-22969
26. Lo Leggio, L., Simmons, T. J., Poulsen, J.-C. N., Frandsen, K. E. H., Hemsworth, G. R., Stringer, M. A., von Freiesleben, P., Tovborg, M., Johansen, K. S., De Maria, L., Harris, P. V., Soong, C.-L., Dupree, P., Tryfona, T., Lenfant, N., Henrissat, B., Davies, G. J., and Walton, P. H. (2015) Structure and boosting activity of a starch-degrading lytic polysaccharide monooxygenase. *Nature communications* **6**
27. Isaksen, T., Westereng, B., Aachmann, F. L., Agger, J. W., Kracher, D., Kittl, R., Ludwig, R., Haltrich, D., Eijsink, V. G., and Horn, S. J. (2014) A C4-oxidizing lytic polysaccharide

monooxygenase cleaving both cellulose and cello-oligosaccharides. *The Journal of biological chemistry* **289**, 2632-2642
28. Bennati-Granier, C., Garajova, S., Champion, C., Grisel, S., Haon, M., Zhou, S., Fanuel, M., Ropartz, D., Rogniaux, H., Gimbert, I., Record, E., and Berrin, J. G. (2015) Substrate specificity and regioselectivity of fungal AA9 lytic polysaccharide monooxygenases secreted by Podospora anserina. *Biotechnol Biofuels* **8**
29. Frandsen, K. E. H., Poulsen, J.-C. N., Tandrup, T., and Lo Leggio, L. Unliganded and substrate bound structures of the cellooligosaccharide active lytic polysaccharide monooxygenase LsAA9A at low pH. *Carbohydrate research*
30. Berrin, J. G., Rosso, M. N., and Abou Hachem, M. (2017) Fungal secretomics to probe the biological functions of lytic polysaccharide monooxygenases. *Carbohydrate research* **448**, 155-160
31. Spodsberg, N., Shaghasi, T., SWEENEY, M., Xu, F., and Schnorr, K. (2014) Polypeptides having cellulolytic enhancing activity and polynucleotides encoding same. Google Patents, WO2014066141 A3
32. Spodsberg, N. (2015) Polypeptides having cellulolytic enhancing activity and polynucleotides encoding same. Google Patents, US2015/0247137 A1
33. Yakovlev, I., Vaaje-Kolstad, G., Hietala, A. M., Stefanczyk, E., Solheim, H., and Fossdal, C. G. (2012) Substrate-specific transcription of the enigmatic GH61 family of the pathogenic white-rot fungus Heterobasidion irregulare during growth on lignocellulose. *Applied microbiology and biotechnology* **95**, 979-990
34. Miyauchi, S., Navarro, D., Grisel, S., Chevret, D., Berrin, J. G., and Rosso, M. N. (2017) The integrative omics of white-rot fungus Pycnoporus coccineus reveals co-regulated CAZymes for orchestrated lignocellulose breakdown. *PloS one* **12**, e0175528
35. Miyauchi, S., Rancon, A., Drula, E., Hage, H., Chaduli, D., Favel, A., Grisel, S., Henrissat, B., Herpoel-Gimbert, I., Ruiz-Duenas, F. J., Chevret, D., Hainaut, M., Lin, J., Wang, M., Pangilinan, J., Lipzen, A., Lesage-Meessen, L., Navarro, D., Riley, R., Grigoriev, I. V., Zhou, S., Raouche, S., and Rosso, M. N. (2018) Integrative visual omics of the white-rot fungus Polyporus brumalis exposes the biotechnological potential of its oxidative enzymes for delignifying raw plant biomass. *Biotechnol Biofuels* **11**, 201
36. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic acids research* **42**, D490-495
37. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nature methods* **8**, 785-786
38. Miyauchi, S., Navarro, D., Grigoriev, I. V., Lipzen, A., Riley, R., Chevret, D., Grisel, S., Berrin, J. G., Henrissat, B., and Rosso, M. N. (2016) Visual Comparative Omics of Fungi for Plant Biomass Deconstruction. *Front Microbiol* **7**, 1335
39. Krissinel, E., and Henrick, K. (2004) Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta crystallographica. Section D, Biological crystallography* **60**, 2256-2268
40. Li, X., Beeson, W. T. t., Phillips, C. M., Marletta, M. A., and Cate, J. H. (2012) Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure* **20**, 1051-1061
41. Karkehabadi, S., Hansson, H., Kim, S., Piens, K., Mitchinson, C., and Sandgren, M. (2008) The first structure of a glycoside hydrolase family 61 member, Cel61B from Hypocrea jecorina, at 1.6 A resolution. *Journal of molecular biology* **383**, 144-154
42. Garman, E. F., and Grime, G. W. (2005) Elemental analysis of proteins by microPIXE. *Progress in biophysics and molecular biology* **89**, 173-205

43. Blom, N., Gammeltoft, S., and Brunak, S. (1999) Sequence and structure-based prediction of eukaryotic protein phosphorylation sites. *Journal of molecular biology* **294**, 1351-1362
44. Kadowaki, M. A. S., Varnai, A., Jameson, J. K., AE, T. L., Costa-Filho, A. J., Kumagai, P. S., Prade, R. A., Polikarpov, I., and Eijsink, V. G. H. (2018) Functional characterization of a lytic polysaccharide monooxygenase from the thermophilic fungus Myceliophthora thermophila. *PloS one* **13**, e0202148
45. Kittl, R., Kracher, D., Burgstaller, D., Haltrich, D., and Ludwig, R. (2012) Production of four Neurospora crassa lytic polysaccharide monooxygenases in Pichia pastoris monitored by a fluorimetric assay. *Biotechnol Biofuels* **5**, 79
46. Lo Leggio, L., Weihe, C. D., Poulsen, J. N., Sweeney, M., Rasmussen, F., Lin, J., De Maria, L., and Wogulis, M. (2018) Structure of a lytic polysaccharide monooxygenase from Aspergillus fumigatus and an engineered thermostable variant. *Carbohydrate research* **469**, 55-59
47. Kracher, D., Andlar, M., Furtmuller, P. G., and Ludwig, R. (2018) Active-site copper reduction promotes substrate binding of fungal lytic polysaccharide monooxygenase and reduces stability. *The Journal of biological chemistry* **293**, 1676-1687
48. Zheng, H., Chruszcz, M., Lasota, P., Lebioda, L., and Minor, W. (2008) Data mining of metal ion environments present in protein structures. *Journal of inorganic biochemistry* **102**, 1765-1776
49. Martinez-Abad, A., Berglund, J., Toriz, G., Gatenholm, P., Henriksson, G., Lindstrom, M., Wohlert, J., and Vilaplana, F. (2017) Regular Motifs in Xylan Modulate Molecular Flexibility and Interactions with Cellulose Surfaces. *Plant physiology* **175**, 1579-1592
50. Bai, Y., Chen, B., Li, M., Zhou, Y., Ren, S., Xu, Q., Chen, M., and Wang, S. (2017) FPD: A comprehensive phosphorylation database in fungi. *Fungal biology* **121**, 869-875
51. Klement, E., and Medzihradszky, K. F. (2017) Extracellular Protein Phosphorylation, the Neglected Side of the Modification. *Molecular & cellular proteomics : MCP* **16**, 1-7
52. Yalak, G., and Vogel, V. (2012) Extracellular phosphorylation and phosphorylated proteins: not just curiosities but physiologically important. *Science signaling* **5**, re7
53. Tagliabracci, V. S., Engel, J. L., Wen, J., Wiley, S. E., Worby, C. A., Kinch, L. N., Xiao, J., Grishin, N. V., and Dixon, J. E. (2012) Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science* **336**, 1150-1153
54. Tagliabracci, V. S., Wiley, S. E., Guo, X., Kinch, L. N., Durrant, E., Wen, J., Xiao, J., Cui, J., Nguyen, K. B., Engel, J. L., Coon, J. J., Grishin, N., Pinna, L. A., Pagliarini, D. J., and Dixon, J. E. (2015) A Single Kinase Generates the Majority of the Secreted Phosphoproteome. *Cell* **161**, 1619-1632
55. de Sa Pinheiro, A. A., Amazonas, J. N., de Souza Barros, F., De Menezes, L. F., Batista, E. J., Silva, E. F., De Souza, W., and Meyer-Fernandes, J. R. (2007) Entamoeba histolytica: an ecto-phosphatase activity regulated by oxidation-reduction reactions. *Experimental parasitology* **115**, 352-358
56. Yalak, G., Ehrlich, Y. H., and Olsen, B. R. (2014) Ecto-protein kinases and phosphatases: an emerging field for translational medicine. *Journal of translational medicine* **12**, 165
57. Crane, J. K., and Vezina, C. M. (2005) Externalization of host cell protein kinase C during enteropathogenic Escherichia coli infection. *Cell death and differentiation* **12**, 115-127
58. Karunarathna, S. C., Yang, Z. L., Zhao, R.-L., Vellinga, E. C., Bahkali, A. H., Chukeatirote, E., and Hyde, K. D. (2011) Three new species of Lentinus from northern Thailand. *Mycological Progress* **10**, 389-398
59. Couturier, M., Navarro, D., Chevret, D., Henrissat, B., Piumi, F., Ruiz-Duenas, F. J., Martinez, A. T., Grigoriev, I. V., Riley, R., Lipzen, A., Berrin, J. G., Master, E. R., and Rosso, M. N. (2015) Enhanced degradation of softwood versus hardwood by the white-rot fungus Pycnoporus coccineus. *Biotechnol Biofuels* **8**, 216

60. Grigoriev, I. V., Nikitin, R., Haridas, S., Kuo, A., Ohm, R., Otillar, R., Riley, R., Salamov, A., Zhao, X., Korzeniewski, F., Smirnova, T., Nordberg, H., Dubchak, I., and Shabalov, I. (2014) MycoCosm portal: gearing up for 1000 fungal genomes. *Nucleic acids research* **42**, D699-704
61. Badino, S. F., Christensen, S. J., Kari, J., Windahl, M. S., Hvidt, S., Borch, K., and Westh, P. (2017) Exo-exo synergy between Cel6A and Cel7A from Hypocrea jecorina: Role of carbohydrate binding module and the endo-lytic character of the enzymes. *Biotechnology and bioengineering* **114**, 1639-1647
62. Gorrec, F. (2009) The MORPHEUS protein crystallization screen. *Journal of applied crystallography* **42**, 1035-1042
63. Kabsch, W. (2010) Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125-132
64. Vagin, A., and Teplyakov, A. (2010) Molecular replacement with MOLREP. *Acta crystallographica. Section D, Biological crystallography* **66**, 22-25
65. Bunkoczi, G., and Read, R. J. (2011) Improvement of molecular-replacement models with Sculptor. *Acta crystallographica. Section D, Biological crystallography* **67**, 303-312
66. Murshudov, G. N., Skubak, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F., and Vagin, A. A. (2011) REFMAC5 for the refinement of macromolecular crystal structures. *Acta crystallographica. Section D, Biological crystallography* **67**, 355-367
67. Winn, M. D., Ballard, C. C., Cowtan, K. D., Dodson, E. J., Emsley, P., Evans, P. R., Keegan, R. M., Krissinel, E. B., Leslie, A. G., McCoy, A., McNicholas, S. J., Murshudov, G. N., Pannu, N. S., Potterton, E. A., Powell, H. R., Read, R. J., Vagin, A., and Wilson, K. S. (2011) Overview of the CCP4 suite and current developments. *Acta crystallographica. Section D, Biological crystallography* **67**, 235-242
68. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta crystallographica. Section D, Biological crystallography* **60**, 2126-2132
69. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501
70. Gille, C., and Frommel, C. (2001) STRAP: editor for STRuctural Alignments of Proteins. *Bioinformatics* **17**, 377-378
71. Gille, C., Fahling, M., Weyand, B., Wieland, T., and Gille, A. (2014) Alignment-Annotator web server: rendering and annotating sequence alignments. *Nucleic acids research* **42**, W3-6
72. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772-780
73. Zhang, Y., and Skolnick, J. (2005) TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic acids research* **33**, 2302-2309
74. Paiardini, A., Bossa, F., and Pascarella, S. (2005) CAMPO, SCR_FIND and CHC_FIND: a suite of web tools for computational structural biology. *Nucleic acids research* **33**, W50-55
75. Janson, G., Zhang, C., Prado, M. G., and Paiardini, A. (2017) PyMod 2.0: improvements in protein sequence-structure analysis and homology modeling within PyMOL. *Bioinformatics* **33**, 444-446
76. Grime, G. W. (1996) The " Q factor" method: quantitative microPIXE analysis using RBS normalisation. *Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms* **109-110**, 170-174

FOOTNOTES:

[1] https://www6.inra.fr/cirm_eng/Filamentous-Fungi; [2] https://genome.jgi.doe.gov/programs/fungi; [3] https://www.surrey.ac.uk/ion-beam-centre; [4] http://www.microbeams.co.uk/. The atomic coordinates for the X-ray crystal structures of *Ls*AA9B have been deposited in the Protein Data Bank (https://www.wwpdb.org/) and given the PDB entry codes; 6RS6 (*Ls*AA9B), 6RS7 (*Ls*AA9B_deglyc), 6RS8 (*Ls*AA9B_metalsoak), 6RS9 (*Ls*AA9B_Xyl4). The *Ls*AA9B nucleotide sequence has been submitted to the NCBI GenBank and given accession number MN265867. Sequences of proteins encoded by regulated Arg-AA9 genes (Figure 2) have been deposited to the MycoCosm Database at the Joint Genome Institute (JGI) and given the following JGI protein IDs; 1430659 (*Pycnoporus coccineus BRFM 310*), 971189 (*Trametes ljubarskyi BRFM 1659*), 248104 (*Pycnoporus coccineus BRFM 1662*), 236437 (*Leiotrametes sp. BRFM 1775*), 1740610 (*Pycnoporus sanguineus BRFM 1264*), 360271 (*Trametes elegans BRFM 1663*) and 1403153 (*Polyporus brumalis BRFM 1820*). The sequence of *Hi*AA9G is also available at the JGI (JGI protein ID 173373).

The abbreviations used are: LPMO, Lytic polysaccharide monooxygenase; AA, Auxiliary Activity; GH, Glycoside Hydrolases; CE, Carbohydrate Esterases; CAZymes, Carbohydrate Active enZymes; RNA-seq, RNA sequencing; MES, 2-(N-morpholino)ethanesulfonic acid; HEPES, 2-[4-(2-hydroxyethyl)piperazin-1-yl]ethanesulfonic acid; MOPS, 3-(N-Morpholino)propanesulfonic acid; MPD, 2-Methyl-2,4-pentanediol; PEG, Polyethylene Glycol; PIXE, Proton Induced X-ray Emission; MR, Molecular Replacement; ITC, Isothermal Titration Calorimetry; Cell4, Cellotetraose; Cell5, Cellopentaose; Cell6, Cellohexaose; Xyl5, Xylopentaose; Xyl4, Xylotetraose; DTPA, 2-[Bis[2-[bis(carboxymethyl)amino]ethyl]amino]acetic acid; PASC, phosphoric acid swollen cellulose; MS, Mass Spectrometry; MALDI-TOF, Matrix Associated Laser Desorption Ionization Time Of Flight; ESI-MS/MS, Electrospray ionization tandem mass spectrometry; BCTMP, bleached chemi-thermo mechanical pulp. *Ti*, Inflection point temperature; ASU, Asymmetric Unit.

Table 1. Transcriptional regulation of genes coding for Arg-AA9 in six *Polyporales* species

Organism	**JGI Prot ID**	**x-fold upregulation on**				**Co-regulated genes**	**Reference**
		Av	As	Pi	WS		
Pycnoporus coccineus BRFM 310	1430659	110	10	52	110	GH5_5, GH7, AA9, CE16, GH28, AA2	(34)
Trametes ljubarskyi BRFM 1659	971189	675	388	64	256	GH5_5, GH7, GH5_7, GH12, AA9, CE16	this manuscript
Pycnoporus coccineus BRFM 1662	248104	14	9	5	6	-	this manuscript
Leiotrametes sp. BRFM 1775	236437	104	10	3	10	GH1, GH3, GH131	this manuscript
Pycnoporus sanguineus BRFM 1264	1740610	78	7	nd	3	GH7, AA9	(38)
Trametes elegans BRFM 1663	360271	16	48	11	34	GH10	this manuscript

Normalized transcript read counts after 3 day-growth on Avicel (Av), aspen, (As), pine (Pi) or wheat straw (WS) where compared to normalized read counts after 3 day-growth on maltose. Coregulated genes were identified that shared similar transcript levels and transcription profiles on the five carbon sources.

Table 2. Crystallization conditions for crystals from which datasets were collected

Protein batch	Protein concentration	Primary Precipitant	Additive	Buffer (pH)	Notes	Structure
*Ls*AA9B (initial)	4 mg/ml	PPT4 37.5%(w/v)	NPS	Imidazole/MES 0.1 M (pH6.5)	-	-
*Ls*AA9B (glycosylated)	5 mg/ml	PPT1 30%(w/v)	AA	Imidazole/MES 0.1 M (pH6.5)	-	*Ls*AA9B
*Ls*AA9B (deglycosylated)	5 mg/ml	PPT4 48%(w/v)	NPS	Imidazole/MES 0.1 M (pH6.5)	-	*Ls*AA9B_deglyc
*Ls*AA9B (glycosylated)	11 mg/ml	PPT4 37.5%(w/v)	CA	HEPES/MOPS 0.1M (pH 7.5)	30 min metal soak	*Ls*AA9B_metalsoak
*Ls*AA9B (glycosylated)	11 mg/ml	PPT1 36%(w/v)	Divalents	HEPES/MOPS 0.1M (pH 7.5)	1 hour 20 min. xylotetraose soak	*Ls*AA9B_Xyl4

Drop sizes were 0.3 µl and with protein:reservoir ratios of 3:1
Polyethylene glycol monomethyl ether (PEGMME), Polyethylene glycol (PEG)
PPT1: 2:1 ratio of PEGMME 500 and PEG 20.000
PPT4: 1:1:1 ratio of MPD (racemic); PEG 1000; PEG 3350
Amino Acids (AA): 20 mM of Glutamate, Alanine, Glycine, Lysine, and Serine (some racemic).
NPS: 30 mM of each of $NaNO_3$; Na_2HPO_4; $(NH_4)_2SO_4$
Carboxylic Acids (CA): 0.02M of each of Na-formate, NH_4-acetate, Na-citrate, Na/K-tartrate, Na-oxamate
Divalents: 30 mM each of $CaCl_2$ and $MgCl_2$
Buffer: 2-(N-morpholino)ethanesulfonic acid (MES), 3-(N-morpholino)propanesulfonic acid (MOPS), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)

Table 3. Crystallographic data and refinement statistics

Data collection	***Ls*AA9B**	***Ls*AA9B_deglyc**	***Ls*AA9B_metalsoak**	***Ls*AA9B_Xyl4**
Beamline Synchrotron	I911-3 MAX-IV	I911-3 MAX-IV	I911-3 MAX-IV	ID32-1 ESRF
Wavelength (Å)	0.9900	1.0000	1.3499	0.9760
temperature of data collections (K)	100	100	100	100
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell Dimensions a, b, c (Å) α=β=γ (°)	 35.19 72.48 78.56 90	 35.08 72.99 79.06 90	 35.16 72.58 78.68 90	 35.41 72.79 78.59 90
Molecules/ASU	1	1	1	1
Resolution (Å)	20.00-1.60 (1.70-1.60)	20.00-1.60 (1.70-1.60)	40.00-1.58 (1.62-1.58)	40.00-1.40 (1.44-1.40)
Rrim (%)	10.8 (67.8)	8.4 (82.0)	5.5 (39.1)	10.0 (156.0)
I/σ(I)	14.07 (2.74)	14.39 (1.94)	28.46 (3.0)	13.57 (1.36)
CC½ (%)	99.8 (76.4)	99.8 (76.6)	99.9 (82.6)	99.9 (45.4)
Completeness (%)	99.6 (98.5)	98.5 (91.7)	92.4 (72.4)	100 (100)
Multiplicity	7.83 (6.01)	5.81 (3.59)	8.60 (2.84)	7.17 (7.27)
No. total reflections	212379 (26267)	157803 (14758)	239233 (4271)	292531 (40817)
No. unique reflections	27139 (4369)	27139 (4106)	27805 (1501)	40817 (2961)
Rfactor (%)	13.67 (22.0)	15.48 (29.9)	11.27 (22.2)	13.31 (29.9)
Rfree (%)	18.22 (27.5)	21.14 (37.5)	17.31 (24.4)	17.51 (34.9)
No. atoms / B-factors ($Å^2$) Protein[§] Solvent[#]	 1703 / 14.84 241 / 30.10	 1737 / 17.41 291 / 31.47	 1740 / 12.39 270 / 26.94	 1741 / 19.21 231 / 31.12
RMSD Bond lengths (Å) Bond angles (°)	 0.018 1.90	 0.009 1.48	 0.015 1.83	 0.013 1.81
Ramachandran[†] (%) favored / allowed/ outliers	90 / 10 / 0	88 / 12 / 0	88 / 11 / 0	89 / 11 / 0
PDB accession codes	6RS6	6RS7	6RS8	6RS9

Highest resolution shell is in parenthesis

§ Glycosylation (N-acetylglucosamine units, mannosyl units) are included in "Protein"
# PEG, MES, Acetate, Cl^- and water molecules are included under "Solvent"
† Calculated using Procheck within the CCP4 suite

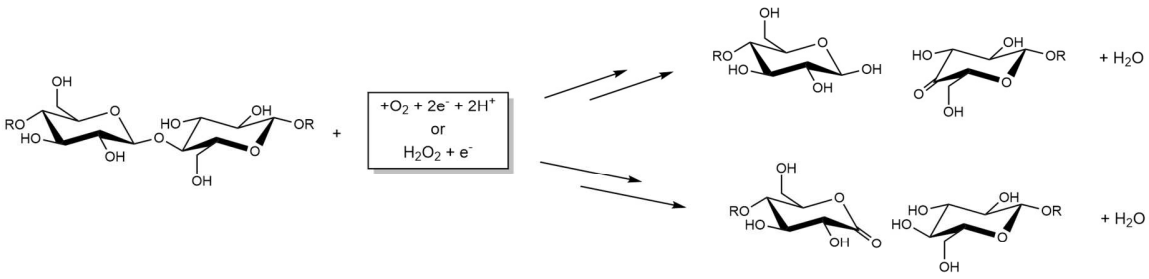

Figure 1 – Schematic representation of oxidative cleavage of cellooligosaccharides/cellulose by LPMOs.

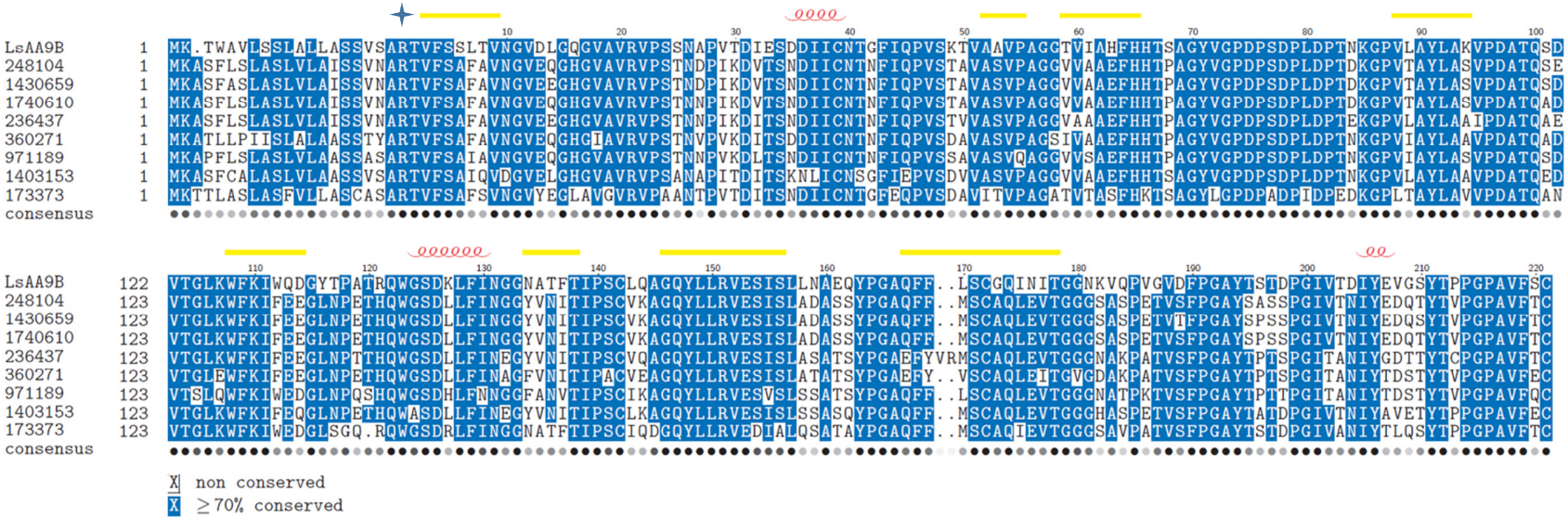

Figure 2. AA9s with N-terminal Arg upregulated on lignocellulosic biomass
Multiple sequence alignment of Arg-AA9 sequences found to be upregulated during growth on aspen, pine and wheat straw. The signal peptides are included for each sequence. The N-terminal Arg of the mature proteins is indicated with star above the sequence. Numbering above the sequence alignment is according to the mature *Ls*AA9B protein with Arg in the N-terminus. Secondary structure elements of *Ls*AA9B are shown above the alignment in yellow and red for β-strands and α-helices, respectively. The JGI protein ID is given next to each sequence. Residues with more than 70 % sequence identity are indicated with blue (i.e. residues shared by at least seven of the nine protein sequences).

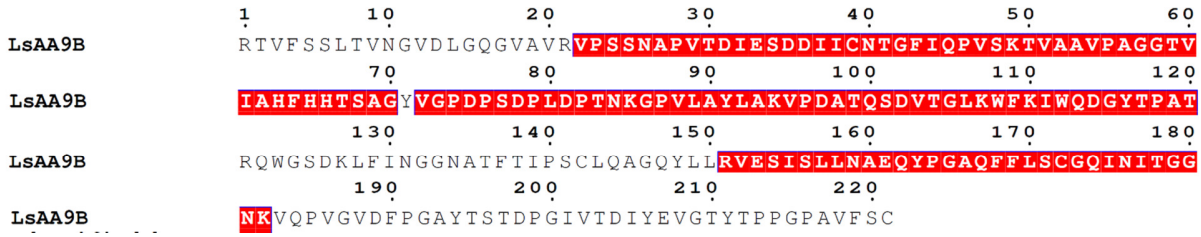

Figure 3 – Sequence coverage of *Ls*AA9B peptides found in the *L. similis* secretome following growth on hardwood pulp

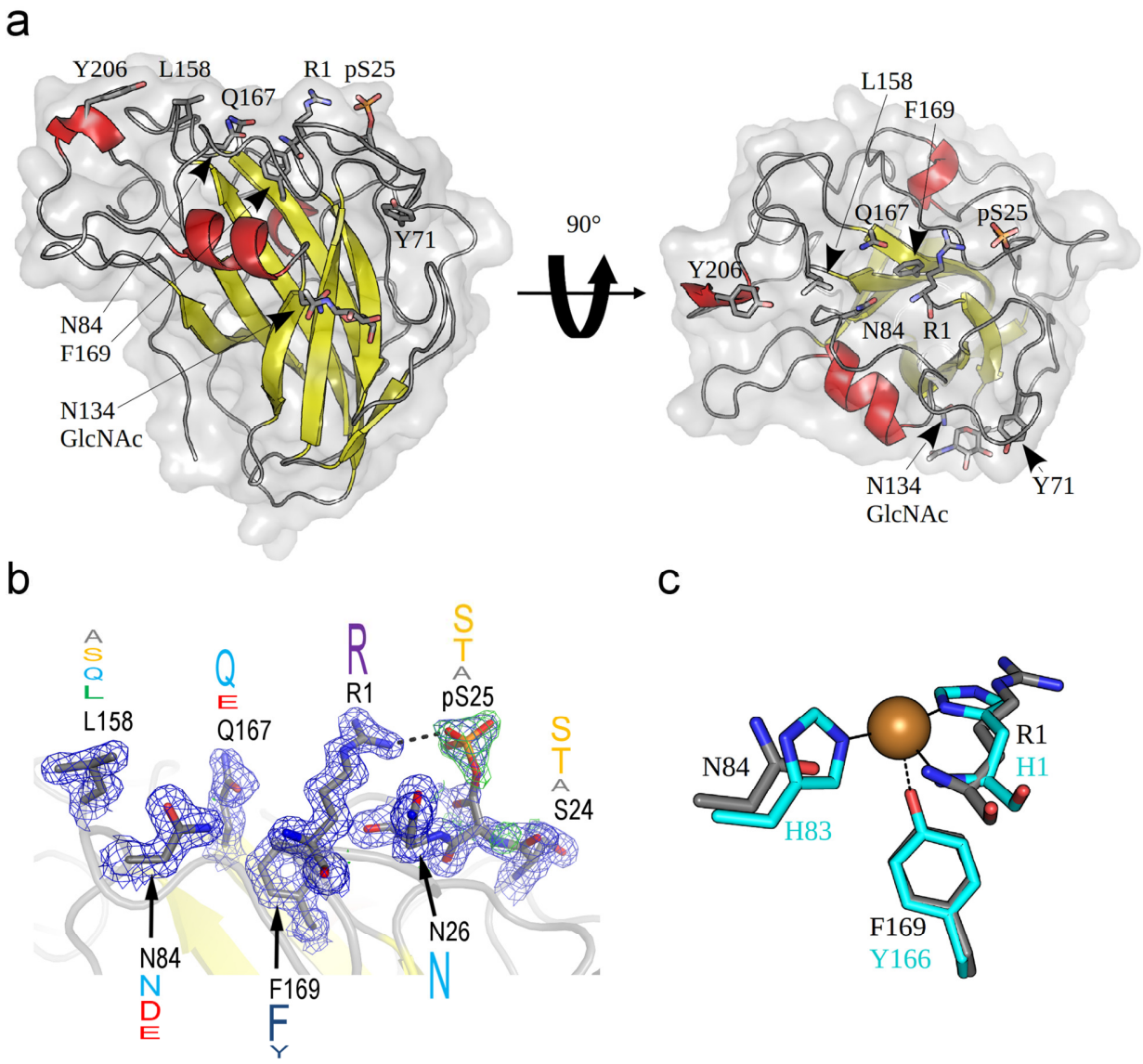

Figure 4. Structure of *Ls*AA9B and positions equivalent to critical residues in AA9 LPMOs
a) Side (left) and top (right) view of *Ls*AA9B showing the common immunoglobulin-like β-sandwich fold. **b)** Detailed view with electron density near the N-terminal Arg (R1), including phosphorylated Ser25 (pS25). Ser25 is modelled in two conformations (70% and 30% occupancies) with the highest occupied conformation being phosphorylated. The sidechain of R1 makes a 2.9 Å interaction with the phosphorylation on Ser25. *Ls*AA9B R1, N84, L158, Q167 and F169 are equivalent to H1, H68, H142, Q151 and Y153 in *Tt*AA9E (residues critical for AA9 LPMO function). The electron density maps (2Fo-Fc in blue and Fo-Fc in green, contoured at 1.0σ and 3.0σ, respectively) are calculated from a structure before modelling the phosphorylation. The logo-conservation depict most common residues at a given position and their frequency (letter size) in Arg-AA9 sequences. **c)** Active site residues of the structurally closest homolog *Nc*AA9C (cyan, PDB 4D7U) superimposed onto the equivalent residues of *Ls*AA9B.

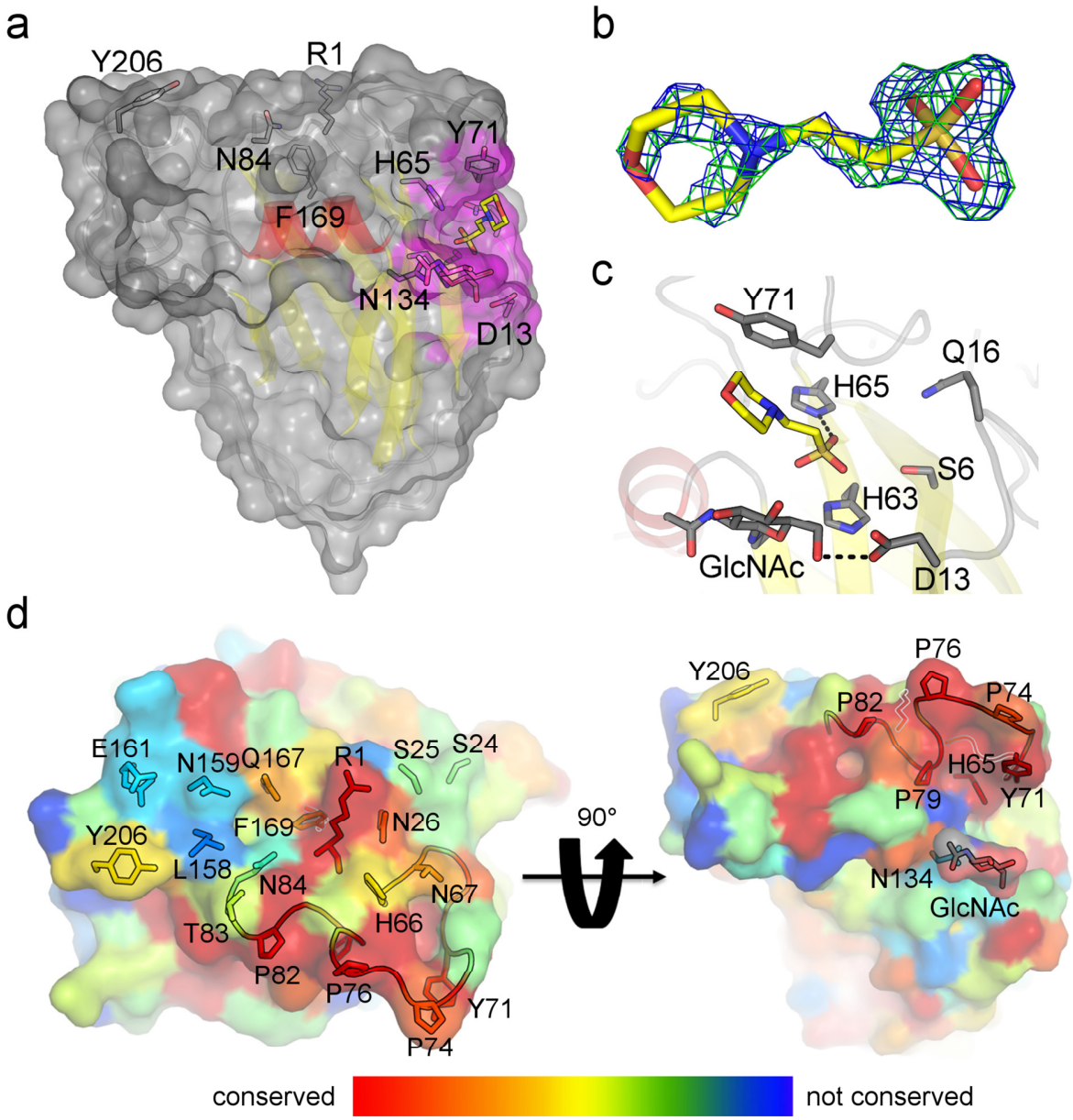

Figure 5. Surface features and conservation of *Ls*AA9B
a) Structure of *Ls*AA9B interacting with a MES buffer molecule in a small pocket (magenta). The MES molecule makes an H-bond to H65 while interacting with Y71 and a GlcNAc molecule (N-linked glycosylation at N134) **b)** Electron density (2Fo-Fc (1.0σ) in blue and Fo-Fc (3.0σ) in green) for MES before modelling the buffer molecule. **c)** View of *Ls*AA9B S6, D13, Q16, H63, H65, Y71 and N134-GlcNAc that form the pocket. **d)** Residues of *Ls*AA9B colored in a rainbow spectrum according to their conservation within Arg-AA9 sequences (red: completely conserved; yellow: intermediately conserved; blue: not conserved). Positions equivalent to residues involved in Cu coordination and secondary coordination sphere in canonical AA9 LPMOs (R1, N84, F169, L158, Q167) and in substrate interactions in *Ls*AA9A (N26, H66, N67, T83, N159, E161, Y206) are shown in sticks. The speculated phosphorylation sites at S25 and S24 and the completely conserved residues in the L3 loop (H65, Y71, P76, P79 and P82) are also shown in sticks representation.

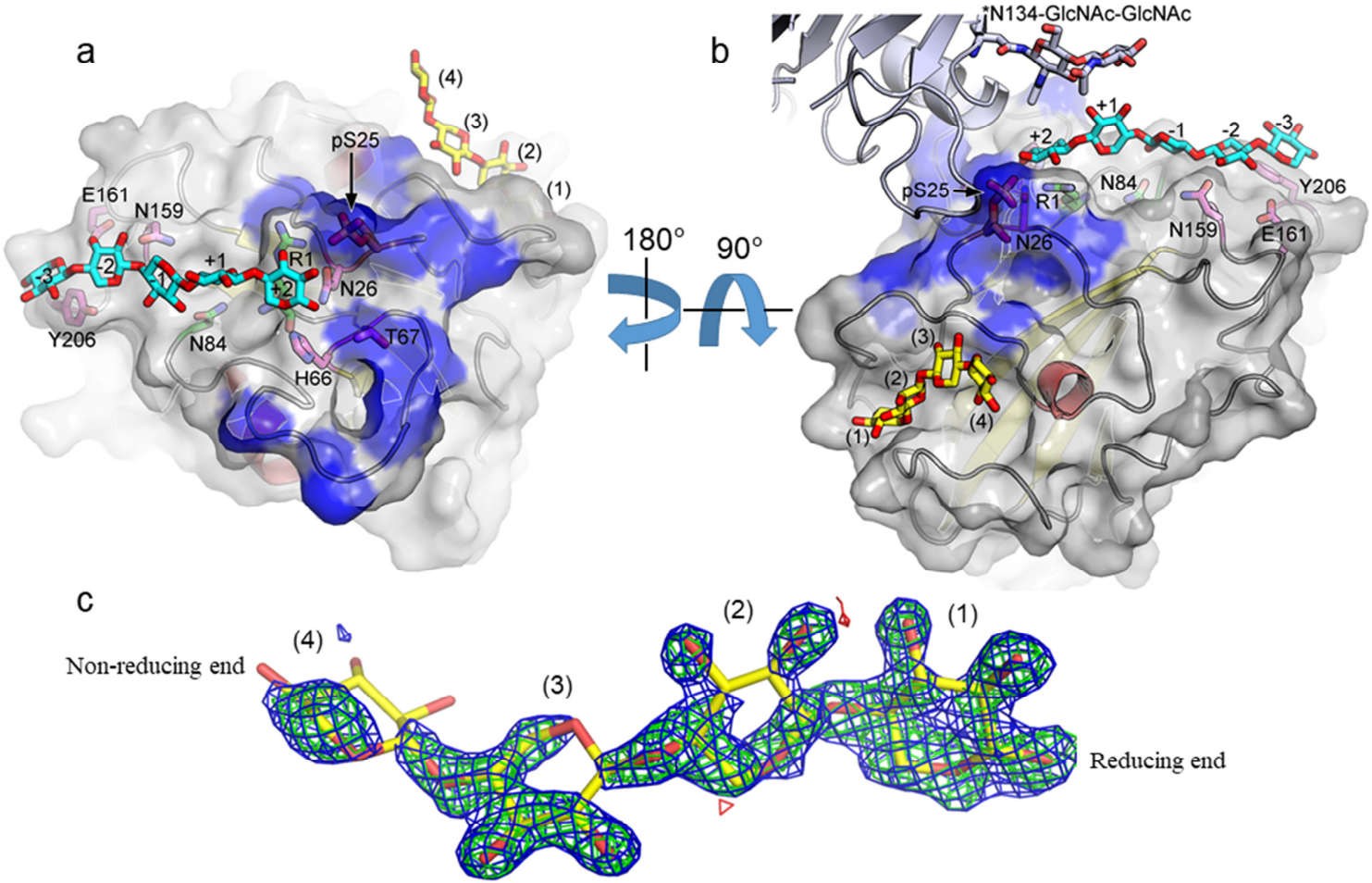

Figure 6. Xylotetraose bound on the surface of *Ls*AA9B
a-b) Structure of *Ls*AA9B-Xyl4 showing interaction with the xylotetraose ligand (Xyl4; yellow). The N-terminal Arg (R1) and Asn84 (equivalent to the His brace of canonical AA9s) are shown in green. Positions involved in AA9-substrate interactions are colored magenta. The xylopentaose ligand (cyan) from the *Ls*AA9A-Xyl5 complex structure (PDB 5NLO) is shown following superimposition onto *Ls*AA9B. Blue surface indicates *Ls*AA9B crystal contacts. The pS25 involved in crystal contacts are highlighted with an arrow. N-linked glycosylation from a symmetry molecule (indicated with an asterisk) is found close to the Xyl5 ligand **c)** Electron density (2Fo-Fc (1.0σ) in blue and Fo-Fc (3.0σ) in green) calculated from a structure before modelling the Xyl4 ligand.