Discovery and characterization of a new family of lytic polysaccharide mono-oxygenases

Glyn R. Hemsworth[1], **Bernard Henrissat**[2], **Gideon J. Davies**[*,1], and **Paul H. Walton**[*,1]

[1]Department of Chemistry, University of York, Heslington, York, UK, YO10 5DD

[2]Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille Université, 13288 Marseille, France

Abstract

Lytic polysaccharide mono-oxygenases (LPMOs) are a recently discovered class of enzymes capable of oxidizing recalcitrant polysaccharides. They currently attract much attention due to their potential use in biomass conversion, notably in the production of biofuels. Past work has identified two discrete sequence-based families of these enzymes termed AA9 (formerly GH61) and AA10 (formerly CBM33). Here we report the discovery of a third family of LPMOs. Using a chitin-degrading exemplar from *Aspergillus oryzae*, we show that the 3-D structure of the enzyme shares some features of the previous two classes of LPMOs, including a copper active centre featuring the *histidine brace* active site, but is distinct in terms of its active site details and its EPR spectroscopy. The new AA11 family expands the LPMO clan with the potential to broaden both the range of potential substrates and the types of reactive copper-oxygen species formed at the active site of LPMOs.

Biopolymers such as cellulose, chitin and diverse marine polysaccharides are widespread and abundant. They offer the potential to become *the* primary feedstocks in the production of biofuels and/or sustainable commodities.[1] Notwithstanding this potential however, the chemical and mechanical resistance of these biopolymers constrains their commercial viability to a significant extent. In this regard the recent discovery of a class of enzymes now known as *lytic polysaccharide mono-oxygenases* is a breakthrough. Formerly classified as GH61 and CBM33 in the CAZy database,[2] LPMOs are part of the consortium of enzymes secreted by organisms that obtain their energy from dead biomass. What marks out LPMOs is their unique oxidative mode of action, differing from that of canonical glycoside hydrolases. As seminally demonstrated on a chitin-active LPMO from the bacterium *Serratia marcescens*, this mode of action is an oxidation of the polysaccharide chain which cleaves the chains at the surface of the crystalline polymer, thus disrupting the structure to

such an extent that it becomes tractable to further enzymatic action and eventual degradation.[3] In a discovery with important consequences for the future of biofuels, this oxidative mode of action was subsequently also shown on cellulose for a GH61 enzyme[4,5] and a CBM33 enzyme,[6] leading to a reclassification of LPMOs in the CAZy database of carbohydrate active enzymes (www.cazy.org) from families GH61 and CBM33 to "Auxiliary Activity" families AA9 and AA10 respectively.[7]

LPMOs studied thus far have an active site that is located near the center of an extended flat face, where the face is presumed to interact with the crystalline surface of the substrate (Fig. 1a). Following initial AA9[8] and AA10[3] 3-D structure solutions, X-ray crystallographic structural details of the Cu-active site were first fully shown in a cellulose-active fungal LPMO from *Thermoascus aurantiacus* in which a single copper ion is chelated by two nitrogen atoms of the N-terminal histidine (through the NH_2 terminus and an N-atom of the side chain) and a further nitrogen atom of another histidine side chain in an overall T-shaped N_3 configuration coined as the *histidine brace* (Fig. 1b).[5] It was further demonstrated that additional differences exist between the fungal AA9 and bacterial AA10 versions, where the fungal AA9 enzymes have a tyrosine residue adjacent to the copper coordination sphere and, when expressed in filamentous fungal systems, carry an unusual methylation on the τ-*N* atom of the N-terminal histidine.[9] The methylation is not essential for activity, as non-methylated versions expressed in *Pichia pastoris* are known to be active,[10,11] but its exact role is yet to be defined. In a further refinement, a classification of LPMOs based upon whether the principal site of oxidation is C1 or C4 has been proposed,[9] while oxidation at C6 has been proposed for some AA9 LPMOs. [5,10]

LPMOs have major industrial significance, highlighted by the two-fold reduction in total enzyme load and cost for cellulose degradation when AA9 is incorporated.[12] Indeed, the initial work on chitin degradation showed a greater than six-fold increase in the release of chitobiose when an AA10 was co-incubated with a β-chitin nanowhisker substrate in the presence of ascorbic acid (reducing agent) and an *endochitinase*. [3] Given their demonstrated commercial potential in the enhancement of biomass degradation there is great interest in the discovery and analysis of LPMOs. This interest extends to revealing further details of their oxidative modes of action, which promises to uncover more of the diversity of biological copper-oxygen chemistry (recently reviewed in ref 13.[13]) In this context we now report a new LPMO family, to be termed AA11 in the CAZy classification (www.cazy.org), that is phylogenetically distinct from the previous AA9 and AA10 families. The AA11 example described here, from *Aspergillus oryzae* [hereafter *Ao*(AA11)], has activity on chitin and shares some structural and spectroscopic characteristics with AA9s and with AA10. This new family thereby extends the structural range of LPMOs and adds to the complexity of active site variations which LPMOs exhibit.

RESULTS

Discovery of the AA11 Family

Family AA11 was identified by "module walking", in which potential activity as a carbohydrate-active enzyme is probed *via* analysis of the component domains, "walking" from one example to another. Many carbohydrate active enzymes are multi-modular with, in

addition to catalytic modules, one or more additional domains which are often substrate-targeting *carbohydrate binding modules* (CBMs).[14,15] CBM modularity of LPMOs has also been explicitly analysed by Horn *et al*.[16] Within this context, an analysis of family AA9 LPMO sequences reveals that several of them carry a short conserved domain of unknown function (termed herein X278) characterized by the conservation of 4 cysteines and 1 aromatic residue (X278 sequence alignment in supplementary information, Supplementary Results, Supplementary Fig. 1). A search for other proteins that contain the X278 domain then retrieved a large number of secreted modular proteins in addition to those corresponding to the starting AA9s, hinting that some of these proteins may be carbohydrate-active enzymes. Furthermore, a subset of the X278-containing proteins was found to possess a conserved N-terminal region commencing with an N-terminal histidine (as indicated by signal peptide analyses, such as with SignalP[17]) but exhibited no significant sequence similarity to LPMOs of families AA9 and AA10.

Analysis of the sequences in this candidate new LPMO family shows that these proteins exist either appended to X278 or as a single module, Fig. 2. The AA11 domain of BAE61530.1 was used as the query sequence for a BLAST search (23 September 2013) on the non-redundant protein sequence database of the NCBI, and this retrieved approximately 450 sequences of the putative AA11 family with *e*-values better than 4×10^{-4} (data not shown). Only three of the several hundred known AA9 sequences were found in the BLAST report, but with *e*-values worse than 10^{-3}, the commonly accepted threshold of significance. Similarly a Pfam search on the server of the Sanger Institute (pfam.sanger.ac.uk) with the sequence of *A. oryzae* AA11 (GenBank BAE61530) failed to identify a significant hit with the profile "glyco_hydro_61" corresponding to family AA9, whereas a search conducted with the *bona fide A. oryzae* AA9 protein (GenBank BAE64395) gave a significant match (e-value of 3.9×10^{-80}) with this profile. Collectively this indicates that AA11 defines a sequence family separate from AA9 and one which already contains several hundred fungal proteins (Supplementary Data Set). An alignment of thirty AA11 sequences derived from finished genomes is shown in Supplementary Fig. 2. Furthermore, we identified three metazoan AA11 genes in the recently published genome of the bdelloid rotifer *Adineta vaga* (PMID=23873043).[18] The asexual reproductive biology of this organism has been shown to be accompanied by large amounts of horizontal gene transfers (HGT) from bacteria, plants and fungi.[18] Notably all polysaccharide *lyase* and 40% of all *glycoside hydrolase* genes in *A. vaga* are from HGT. It is very likely therefore that the three AA11 genes in *A. vaga* have a fungal origin.

Expression and characterization of Ao(AA11)

Sequence alignments of diverse AA11 and X278 domains indicated that the AA11 catalytic domain could be delineated, encompassing approximately 220 residues from His20 to Cys235 (in the case of *Ao*(AA11), below, data not shown). Initially, a panel of representative AA11 domains was cloned and expressed using *E. coli* as the initial expression host. Of these, the *Ao*(AA11) domain from sequence BAE61530.1 was successfully cloned and over-expressed, via periplasmic secretion, as a soluble protein to high levels in *E.coli*.

Previous studies on AA9[4,5] and AA10[19,20] have shown that these LPMOs are copper-dependent enzymes which bind copper with tight binding constants. Accordingly, isothermal titration calorimetry on *apo-Ao*(AA11) exhibited the same tight binding of copper with binding too tight to be determined accurately and thus estimated to have a $K_D < 1$ nM (data not shown). Accurate determination of the copper binding constant was achieved by displacement calorimetry[21] in which a weaker binding and kinetically labile metal, Zn^{2+} (K_D 8 μM at pH 5 with *Ao*(AA11), data not shown), was first bound to *apo-Ao*(AA11) and then the Cu^{2+} binding titrated competitively, allowing accurate determination of a dissociation constant, K_D, for Cu-*Ao*(AA11) of 790 ±150 pM at pH 5, Fig. 3a.

The substrate specificity of *Ao*(AA11) was determined through incubation of the Cu-loaded enzyme (the copper was kept slightly sub-stoichiometric to avoid any free-copper catalysed side reactions) with various polysaccharide substrates in the presence of ascorbate as the electron donor. MALDI-TOF analysis of the products shows that the *E. coli*-expressed Cu-*Ao*(AA11) enzyme is chitinolytic, Fig. 3b and Supplementary Fig. S3, but has no activity on other substrates including diverse mannans, cellulose and starch (data not shown). As with AA10 the product mass is consistent with a primary chain cleavage to yield predominantly aldonic acid oligosaccharides with even-numbered degrees of polymerization (DP = 4,6,8,10…).[3] In contrast to AA10, however, there is also significant presence of unmodified oligosaccharides (DP = 5,6,7,8,9…) and species with a mass difference of −2 Da, which could represent unopened lactones, or perhaps C4 oxidation products.[22]

Ao(AA11) structure and its comparison with AA9 and AA10

Screening for crystallization conditions for *apo-Ao*(AA11) identified a condition containing 10 mM ZnCl$_2$ (see methods) which gave diffraction quality crystals. The resulting structure was solved by single wavelength anomalous dispersion at λ = 1.282 Å, optimising the *f''* component of the zinc anomalous scattering. The Zn-*Ao*(AA11) structure was refined to a final resolution of 1.55 Å (Supplementary Table 1) with a single Zn^{2+} ion in the active site. The necessity for Zn^{2+} in the crystallisation condition was revealed by the direct coordination to the zinc ion by the side chain of Glu74 from a symmetry-related molecule, thereby augmenting the packing interaction between adjacent AA11 molecules. The structure of Cu-*Ao*(AA11), Fig. 4a, was obtained via soaking of the Zn-crystal in cryo-protectant containing 2 mM Cu^{2+}. Consistent with the ITC described above and the expected relative binding constants of zinc and copper from the Irving-Williams series, this allowed access to the Cu-bound form of *Ao*AA11. Structural analysis of Cu-*Ao*(AA11) at 1.4 Å (Supplementary Table 1) shows that it has a similar tertiary structure to the AA9 and AA10 classes of LPMOs. The core of the protein is formed by a largely antiparallel β-sandwich fold and is stabilised by three disulfide bonds. In both Zn-*Ao*(AA11) and Cu-*Ao*(AA11) structures, residues 99–109 and 151–169 are highly disordered and have not been modelled in the final structure. The intact nature of the enzyme was confirmed however by electrospray mass spectrometry of apo-*Ao*(AA11). Both of these mobile regions are adjacent to the C-terminal region of the structure where the AA11 domain was truncated to remove the X278 module. It is possible that these regions will be ordered in the intact multi-domain protein.

The Cu-*Ao*(AA11) structure is similar to the previous Cu-LPMO structures from families AA9 and AA10. As with AA9 and AA10 classes the N-terminal copper active site in *Ao*(AA11) sits near the center of an extended flat face, albeit one that is slightly convex in AA11, Fig. 4a. Unlike AA9s the face is free of aromatic residues resembling more the binding face of AA10s with residues capable of forming hydrogen-bonds to a potential polysaccharide substrate. The active site N-terminus lies at the end of a small strand which itself interacts in a face-to-face manner with a central four-stranded beta sheet that is the structural core of all LPMOs. Structural similarity searches using PDBeFold[23] (with the default PDBeFOLD Q-score as the criteria) show that the closest structural match is with the AA9 enzyme from *Thielavia terrestris* (PDB: 3eii). This structure yields a Q-score of 0.31 and shows 17% sequence identity and over 145 aligned residues with a Cα r.m.s deviation of 2.6 Å (Fig. 4b). On the whole, AA10 structures match fewer residues in the structural overlaps with AA11, with the closest AA10 structure (Q score 0.28) the *Enterococcus faecaelis* enzyme (PDB: 4alt) which has 118 residues overlapping with a Cα r.m.s deviation of 2.3 Å (Fig. 4c).

In light of the fact that LPMOs are oxidoreductases, there has been focus in recent publications about the role of electron transport chains within the structures.[9] In AA9s conserved tyrosines are important in this context, whereas in AA10s a conserved patch of tryptophan residues has been implicated in this role.[20] In AA11s, conserved tryptophans along with a methionine connect the active site to the distal face of the enzyme and could fulfil a similar function (Supplementary Fig. S4).

The copper active site of *Ao*(AA11)

The copper active site of Cu-*Ao*(AA11), Figure 4d, indicates that the copper ion has been photoreduced by the incident X-ray beam to the copper(I) oxidation state. This is in common with other LPMO structures, as shown clearly for an AA10 from *Bacillus amyloliquefaciens*[20] and observed in a range of PDB files: 4alc, 4ale, 4alq, 4alr, 4als, 4alt by Gudmundsson *et al*, reviewed in reference 13[13]. The copper ion's immediate coordination sphere from endogenous ligands is T-shaped with the three coordinating nitrogen atoms of the *histidine brace*. The oxygen atom of a tyrosine is near one of the copper ion's potential axial coordination sites, although a Cu…O distance of 3.1 Å precludes a formal Cu–O bond. An oxygen atom of a glutamate residue from an adjacent AA11 in the crystal lattice lies 2.5 Å from the copper ion bringing the oxygen atom close to the copper's primary coordination sphere. This contact does not exist in solution since SEC-MALLS shows that the enzyme is monomeric (Supplementary Fig. S6). Accordingly, this interaction is perhaps indicative of how open the copper ion is to coordination by various potential ligands. This is in contrast to the active sites of AA10 enzymes which are more sterically encumbered.[13] Detailed comparison of the active sites of AA9, AA10 and AA11 shows that the AA11 active site has some features of both AA9 and AA10, namely the tyrosine seen in AA9 (Fig. 4e) and the conserved secondary coordination sphere alanine of AA10 (Fig. 4f). In previous studies with AA9 enzymes (for example[10,11,24]) it has been shown that the *N*-methylation is not required for activity, and is absent in enzymes expressed in hosts other than filamentous fungi.[10,11] Consistent with such observations, the *E. coli* expressed protein described herein shows no τ-*N*-methylation of the N-terminal histidine in its active site structure.

EPR spectroscopic analysis of the copper site in *Ao*(AA11)

The X-band EPR spectra of Cu-Ao(AA11) at 150 K in the absence and presence of azide are shown in Fig. 5. The azide-free structure exhibits a near-axial spectral envelope revealing a singly-occupied molecular orbital at copper with significant d(x^2-y^2) character. Accurate simulation of the Cu-*Ao*(AA11) spectrum could be achieved in the parallel region with g_z = 2.27 and $|A_z|$ = 157 G (0.0166 cm^{-1}, Supplementary Table 2), placing Cu-*Ao*(AA11) squarely within the Peisach-Blumberg classification of a type 2 copper centre.[25] Given the lack of resolution in the perpendicular region, accurate simulation of the $A_{x,y}$ and of $g_{x,y}$ values was not possible, however an overall fit to the spectral envelope could only be achieved by introducing a degree of rhombicity, with g_x = 2.03 and g_y = 2.10. It is notable that there is some resolution of superhyperfine coupling to ligating nitrogen atoms observable in the perpendicular region (~15 G, 0.0014 cm^{-1}, 43 MHz) with coupling constants typical for coordination to copper(II) by sp^2 hybridised nitrogen atoms.[26] Addition of excess sodium azide causes a shift in the EPR spectrum of Cu-*Ao*(AA11) to $g_x = g_y$ = 2.06, g_z = 2.24 and $|A_z|$ = 175 G (0.018 cm^{-1}, Fig. 5, Supplementary Table 2) demonstrating that azide coordinates directly to the copper ion, accompanied by a shift in coordination geometry symmetry towards axial. A comparison of the known EPR parameters for the different LPMO classes and substrates, using the standard axial type 2 copper parameters in AA9 as a reference (Supplementary Table 3) shows that AA10 enzymes which are active on chitin demonstrate a reduced $|A_z|$ value and some rhombicity in perpendicular g values. AA11 enzymes have type 2 copper $|A_z|$ values and rhombicity in g values, thereby lying somewhat in-between AA9 and AA10 in terms of their EPR spectroscopic features, commensurate with the structural variations in the active sites of the different classes.

Discussion

We have reported herein the discovery of a new LPMO family using a module-walking approach from a common domain of unknown function, termed X278. This approach demonstrates that it is possible to find LPMOs with different activities/sequences from those already known, thereby opening-up further the discovery potential for this important class of enzymes. In the example described herein, the observation and occurrence of an X278 module on a *chitinase* (Fig. 2), along with the observed activity of *Ao*(AA11) on chitin suggest that X278 could be a chitin-binding domain but it is worth noting that such speculation demands further experimental evidence, given that CBM specificity does not always correlate with that of the catalytic module.[14,15] Using an exemplar from *Aspergillus oryzae* we showed that the new family has structural, activity and spectroscopic characteristics that are an intriguing mixture of those seen in enzymes of the previous AA9 and AA10 classifications, sharing active centre features that are common to one or the other family. A significant implication of the AA11 family is that distinct LPMO families now form a "clan", consisting of CAZy families AA9-AA11. With this expansion of the LPMO clan it is possible to envisage that other LPMO members will be found, with the potential to broaden both the range of potential substrates and the types of reactive copper-oxygen species formed at the active site of LPMOs.

On line methods

Discovery of AA11

Blast searches conducted with AA11 sequences did not retrieve AA9s nor AA10s with significant scores and the same vice-versa. BLAST searches starting with either AA9 or AA10 sequences do not pick AA11s with significant scores. PSI-BLAST analyses suggest distant relatedness to AA9, but insufficient to build a reliable alignment. The structural relationship between the current three families (AA9, AA10 and AA11) of LPMOs suggests that AA9, AA10 and AA11 form a clan of families. The sequences were aligned using MUSCLE[27] and the output was made using ESPRIPT[28].

Expression and Purification of Ao*(AA11)*

Nucleotides 457 to 1104 from NCBI reference sequence: XM_001822611.2 which code for mature *Ao*(AA11) (residues 20 - 235 of BAE61530.1) were codon optimised for expression in *E. coli*, synthesised with the pelB leader sequence and cloned into the pET-26a vector between NdeI and XhoI restriction sites by GenScript. Protein was expressed in BL21* (DE3) *E. coli* in 6×500 ml LB cultures at 37 °C shaking at 180 rpm. At an A_{600} of 0.4 the temperature was lowered to 16 °C before the addition of IPTG to a final concentration of 1 mM when $A_{600} = 0.6\text{-}0.8$. The following day cells were harvested by centrifugation at 11,000 g for 20 min at 4 °C.

The cell paste was re-suspended in three volumes of ice-cold 50 mM Tris p*H* 8, 20% *w/v* sucrose. 40 μl of 10 mg ml^{-1} lysozyme was added for every gram of cell paste and left on ice for 1 h with occasional agitation. 60 μl of 1 M $MgSO_4$ per gram of cell paste was then added and left on ice for a further 30 min. Following centrifugation at 10,000 g for 20 min, 4 °C and removal of the supernatant to a fresh tube, the pellet underwent osmotic shock by re-suspension in 3 volumes of ice cold MilliQ water and was left on ice for a further hour. The cell debris was removed by centrifugation at 10,000 g for 20 minutes and the supernatant combined with that from the previous step. This was sonicated on ice in an MSE Soniprep 150 sonicator to reduce the viscosity before the p*H* was lowered by adding a small amount of 1 M Na-acetate p*H* 5.0. The protein was passed through a 5 ml HiTrap SP HP (GE Healthcare) column equilibrated in 50 mM Na-acetate p*H* 5. Solid $(NH_4)_2SO_4$ was added directly to the flow through which contained *Ao*AA11 to give a final concentration of 1 M. The protein was then passed through a 5 ml Phenyl Sepharose HP column (GE Healthcare), in 50 mM Na-acetate p*H* 5.0, 1M $(NH_4)_2SO_4$. The protein was then precipitated by the addition of solid $(NH_4)_2SO_4$ to 85% saturation at 4°C and isolated by centrifugation at 38,000 g for 20 min, 4 °C. The pellet was dissolved in 10 volumes of 20 mM sodium acetate p*H* 5.0, 250 mM NaCl, 5 mM EDTA and then concentrated to less than 10 ml for size exclusion chromatography on a HiLoad 26/60 Superdex 75 column (GE Healthcare) in 20 mM Na-acetate p*H* 5.0, 250 mM NaCl. Peak fractions were pooled and concentrated by centrifugation on a Sartorius 10 kDa molecular weight cut off concentrator. Protein concentrations for all subsequent experiments were determined by measuring the A_{280} with an extinction coefficient of 25,815 dm^3 mol^{-1} cm^{-1} and a molecular weight of 23,055.8 Da. The protein was judged to be >95% pure on SDS-PAGE gel and gave a single species at 23,049 Da by electrospray mass spectrometry.

Isothermal titration calorimetry

In order to obtain a Cu^{2+} binding constant, competition isothermal titration calorimetry was performed using an ITC-200 calorimeter (GE Healthcare) at 298K. First Zn^{2+} binding was measured with the protein in the cell at 150 μM and 10-fold that concentration of $ZnCl_2$. Fitting these data using Origin 7 (MicroCal) gave N = 0.753 ± 0.008, K = $1.25 \times 10^5 \pm 9.8 \times 10^3 M^{-1}$, H = 4034 ± 57cal/mol, and S = 36.8 cal/mol/K. Copper binding was then measured with protein (and $ZnCl_2$ present in the cell at 50 μM and 500 μM respectively) and $CuCl_2$ at 500 μM. 1μl injections were used with 2 min between each injection. 20 mM sodium acetate p*H* 5.0, 250 mM NaCl, was the buffer used throughout with the metal solution prepared in exactly the same buffer, ensuring that no competing hydrolysis of metals was occurring. Fitting with the competitive binding model using the above parameters for zinc gave N = 0.760 ± 0.0025, K = $1.27 \times 10^9 \pm 2.1 \times 10^8 M^{-1}$, H = −7401 ± 77.7 cal/mol for copper binding. As noted in previous work, the stoichiometry of the interaction with metals was typically less than 1:1. This is likely due to the incomplete removal of copper from the buffers leaving a sub-population of protein copper bound given the high affinity of the interaction. Assuming a binding model of a single copper ion displacing a single zinc ion from *Ao*AA11 the thermogram data were fitted using a nonlinear regression procedure within Origin 7 for data from a single experiment.

Reaction Product Analysis

Squid pen chitin was chosen as a substrate as it is purer than crab and shrimp chitins, in terms of the presence of contaminating chito-oligosaccharides, and because it is widely regarded as the standard substrate for mass spectrometry in the field.[3] 1 ml reactions were set up with 0.2% w/v solid substrate (kind gift from Dominique Gillet of Mahtani Chitosan Pvt Ltd) in 10 mM ammonium acetate p*H* 5.0, 1 mM ascorbic acid, 1 μM $CuCl_2$ and 1 μM Cu-*Ao*(AA11) and were incubated at 30 °C rotating overnight. Remaining substrate was removed by centrifugation at 16,000 g, 4 °C for 5 min and the supernatant used for the analysis. 1 μl of sample was mixed with an equal volume of 10 mg/ml 2,5-dihydroxybenzoic acid (DHB) in 50% acetonitrile, 0.1% trifluoroacetic acid (TFA) on a SCOUT-MTP 384 target plate (Bruker). The spotted samples were then dried in a vacuum desiccator before being analysed by mass spectrometry on an Ultraflex III matrix-assisted laser desorption ionization–time of flight/time of flight (MALDI-TOF/TOF) instrument (Bruker), as described in Vaaje-Kolstad *et al.*[3]

Crystallisation of Ao(AA11)

Pure *Ao*(AA11) was buffer exchanged on a Sartorius 10 kDa molecular weight cut off concentrator into 20 mM sodium acetate p*H* 5.0 to remove salt and was finally concentrated to 25 mg/ml for crystallization. Crystal screens were performed at this concentration using a Mosquito robot (TTP Labtech). Crystals were obtained in several conditions all containing $ZnCl_2$, the best of which were in PACT (Qiagen) condition B12: 10 mM $ZnCl_2$, 100 mM MES p*H* 6.0, 20% w/v PEG-6000. These were readily reproduced in larger hanging drops and were used for subsequent soaks and data collection.

Diffraction Data Collection, Processing and Structure Determination

Crystals were cryo-cooled for data collection by first soaking for 30 s in mother liquor containing 15% *v/v* ethylene glycol before plunging in liquid nitrogen. Diffraction data were collected from Zn^{2+}-containing crystals at Diamond Light Source, beamline I02 at a wavelength of 1.282 Å. Data were indexed and integrated using XDS[29] with subsequent processing steps performed in the CCP4 software suite[30]. The structure was solved using SAD phasing from the anomalous Zn^{2+} signal with SHELX[31]. The initial model was re-built using ARP/wARP[32] before subsequent rounds of rebuilding and refinement using COOT[33] and REFMAC5 [33] respectively.

To obtain the copper complex, crystals were soaked for 2 h in cryo-protectant in which the $ZnCl_2$ had been replaced with 2 mM $CuCl_2$ and again were cryo-cooled by plunging in liquid nitrogen. Diffraction data were collected at Diamond Light Source beamline I04 at a wavelength of 0.980 Å. These data were indexed and integrated using XDS [29] and further downstream processing performed in the CCP4 software suite.[30] The original model was refined against these new data having removed flexible loops and the metal ion to avoid model bias. Rebuilding and refinement were again performed in COOT[33] and REFMAC5[34] respectively. The quality of the models was monitored throughout using MOLPROBITY,[35] with both models showing no Ramachandran outliers and 98.3% in the favoured region for both the Cu and Zn complexes. Data processing and structure refinement statistics can be found in Supplementary table 1.

Structures have been deposited in the Protein Data Bank with codes 4MAH and 4MAI.

Size Exclusion Chromatography with Multi-Angle Laser Light Scattering (SEC-MALLS)

50 μl samples of *Ao*(AA11) were applied onto a BioSep-SEC-S 3000 column (Phenomenex) at a concentration of 2 mg/ml. For the copper-loaded sample the buffer used was 20 mM sodium acetate p*H* 5.0, 250 mM NaCl, and for copper free 1 mM EDTA was also included in the buffer. MALLS data were collected on a Dawn Heleos II 18-angle light scattering detector with an in-line OptilabrEX refractive index monitor (Wyatt Technology). ASTRA software was used to analyze the data fitting with the Zimm model and an estimated dn/dc value of 0.190 mL/g. The errors were estimated within the ASTRA software and are based on a single measurement.

Electron paramagnetic resonance spectroscopy

Continuous wave X-band frozen solution EPR spectra of single sample of 0.2 to 0.5 mM solutions of Cu(II)-*Ao*(A11) and with 1000-fold excess of sodium azide (10% *v/v* glycerol) at p*H* 5.0 (acetate buffer) and 150 K were acquired on a Bruker EMX spectrometer operating at ~9.30 GHz, with a modulation amplitude of 4 G and microwave power of 5.02 mW. Spectral simulation was carried out using Easyspin 4.0.0 on a desktop PC. Simulation parameters are given in Supplementary table 2. g_z and $|A_z|$ values were determined accurately from the three absorptions at low field. It was assumed that *g* and *A* tensors were axially coincident. Accurate determination of the g_x, g_y, $|A_x|$ and $|A_y|$ was not possible due to the second order nature of the perpendicular region, although it was noted that satisfactory simulation could only be achieved with one particular set of *g* values.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

REFERENCES

1. Gelfand I, et al. Sustainable bioenergy production from marginal lands in the US Midwest. Nature. 2013; 493:514–517. [PubMed: 23334409]

2. Cantarel BL, et al. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic acids research. 2009; 37:D233–8. [PubMed: 18838391]

3. Vaaje-Kolstad G, et al. An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides. Science. 2010; 330:219–222. [PubMed: 20929773]

4. Phillips CM, Beeson WT, Cate JH, Marletta MA. Cellobiose Dehydrogenase and a Copper-Dependent Polysaccharide Monooxygenase Potentiate Cellulose Degradation by Neurospora crassa. ACS Chemical Biology. 2011; 6:1399–1406. [PubMed: 22004347]

5. Quinlan RJ, et al. Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. Proceedings of the National Academy of Sciences. 2011; 108:15079–15084.

6. Forsberg Z, et al. Cleavage of cellulose by a CBM33 protein. Protein Science. 2011; 20:1479–1483. [PubMed: 21748815]

7. Levasseur A, Drula E, Lombard V, Coutinho PM, Henrissat B. Expansion of the enzymatic repertoire of the CAZy database to integrate auxiliary redox enzymes. Biotechnology for biofuels. 2013; 6:41. [PubMed: 23514094]

8. Karkehabadi S, et al. The first structure of a glycoside hydrolase family 61 member, Cel61B from Hypocrea jecorina, at 1.6 Å resolution. Journal of molecular biology. 2008; 383:144–154. [PubMed: 18723026]

9. Li X, Beeson, William T, Phillips, Christopher M, Marletta, Michael A, Cate, Jamie HD. Structural Basis for Substrate Targeting and Catalysis by Fungal Polysaccharide Monooxygenases. Structure (London, England : 1993). 2012; 20:1051–1061.

10. Bey M, et al. Cello-oligosaccharide oxidation reveals differences between two lytic polysaccharide monooxygenases (family GH61) from Podospora anserina. Applied and environmental microbiology. 2013; 79:488–96. [PubMed: 23124232]

11. Wu M, et al. Crystal structure and computational characterization of the lytic polysaccharide monooxygenase GH61D from the basidiomycota fungus Phanerochaete chrysosporium. Journal of Biological Chemistry. 2013; 288:12828–12839. [PubMed: 23525113]

12. Harris PV, et al. Stimulation of Lignocellulosic Biomass Hydrolysis by Proteins of Glycoside Hydrolase Family 61: Structure and Function of a Large, Enigmatic Family. Biochemistry. 2010; 49:3305–3316. [PubMed: 20230050]

13. Hemsworth GR, Davies GJ, Walton PH. Recent insights into copper-containing lytic polysaccharide mono-oxygenases. Current opinion in structural biology. 2013; 23:660–668. [PubMed: 23769965]

14. Gilbert HJ, Knox JP, Boraston AB. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. Current Opinion in Structural Biology. 2013; 23:669–677. [PubMed: 23769966]

15. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochemical Journal. 2004; 382:769. [PubMed: 15214846]

16. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VGH. Novel enzymes for the degradation of cellulose. Biotechnology for Biofuels. 2012; 5:45. [PubMed: 22747961]

17. Petersen TN, Brunak S, von Heijne G, Nielsen H. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Meth. 2011; 8:785–786.

18. Flot J-F, et al. Genomic evidence for ameiotic evolution in the bdelloid rotifer Adineta vaga. Nature. 2013; 500:453–457. [PubMed: 23873043]

19. Aachmann FL, Sørlie M, Skjåk-Bræk G, Eijsink VGH, Vaaje-Kolstad G. NMR structure of a lytic polysaccharide monooxygenase provides insight into copper binding, protein dynamics, and substrate interactions. Proceedings of the National Academy of Sciences. 2012; 106:18779–18784.

20. Hemsworth GR, et al. The copper active site of CBM33 polysaccharide oxygenases. Journal of the American Chemical Society. 2013; 135:6069–77. [PubMed: 23540833]

21. Wilcox DE. Isothermal titration calorimetry of metal ions binding to proteins: An overview of recent studies. Inorganica Chimica Acta. 2008; 361:857–867.

22. Beeson WT, Phillips CM, Cate JHD, Marletta MA. Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases. Journal of the American Chemical Society. 2012; 134:890–892. [PubMed: 22188218]

23. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallographica Section D. 2004; 60:2256–2268.

24. Kittl R, Kracher D, Burgstaller D, Haltrich D, Ludwig R. Production of four Neurospora crassa lytic polysaccharide monooxygenases in Pichia pastoris monitored by a fluorimetric assay. Biotechnology for biofuels. 2012; 5:79. [PubMed: 23102010]

25. Peisach J, Blumberg WE. Structural implications derived from the analysis of electron paramagnetic resonance spectra of natural and artificial copper proteins. Archives of Biochemistry and Biophysics. 1974; 165:691–708. [PubMed: 4374138]

26. Iwaizumi M, Kudo T, Kita S. Correlation between the hyperfine coupling constants of donor nitrogens and the structures of the first coordination sphere in copper complexes as studied by nitrogen-14 ENDOR spectroscopy. Inorganic Chemistry. 1986; 25:1546–1550.

27. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 2004; 32:1792–7. [PubMed: 15034147]

28. Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics. 1999; 15:305–8. [PubMed: 10320398]

29. Kabsch W. Xds. Acta crystallographica. Section D, Biological crystallography. 2010; 66:125–32.

30. Winn MD, et al. Overview of the CCP4 suite and current developments. Acta crystallographica. Section D, Biological crystallography. 2011; 67:235–42.

31. Sheldrick G. A short history of SHELX. Acta Crystallographica Section A. 2008; 64:112–122.

32. Langer G, Cohen SX, Lamzin VS, Perrakis A. Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nature protocols. 2008; 3:1171–9.

33. Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. Acta Crystallographica Section D-Biological Crystallography. 2004; 60:2126–2132.

34. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica D. 1997; 53:240–255.

35. Davis IW, et al. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res. 2007; 35:W375–83. [PubMed: 17452350]
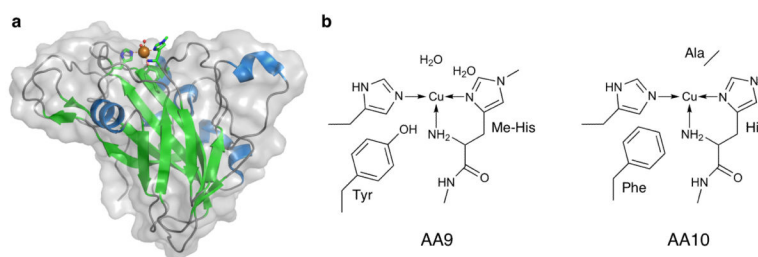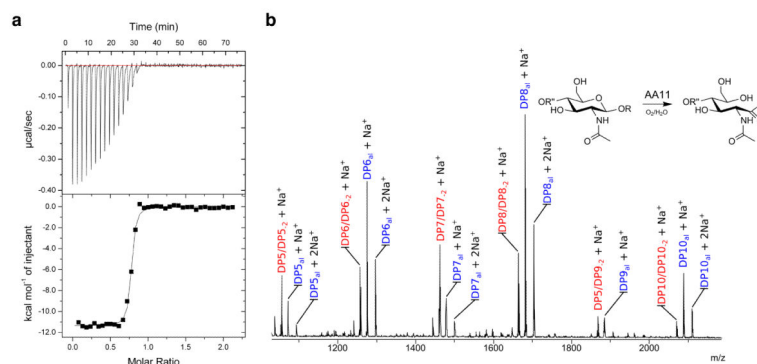
Figure 1.
Structure of typical AA9 and active sites of AA9 and AA10. (a) Overall structure of AA9 from *Thermoascus aurantiacus*[5] with the active site copper shown as a sphere and active site residues shown as sticks, (b) schematic representations of the Cu active sites observed in AA9 and AA10 structures.
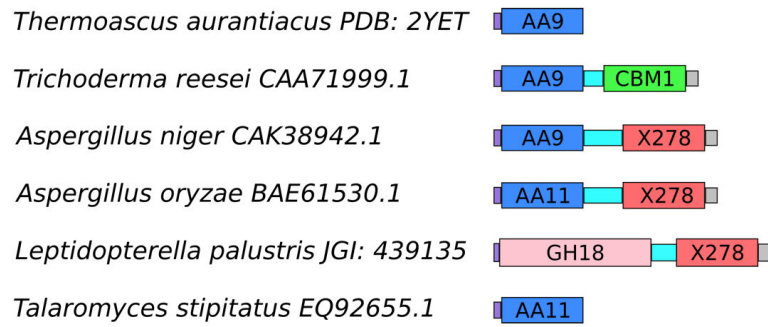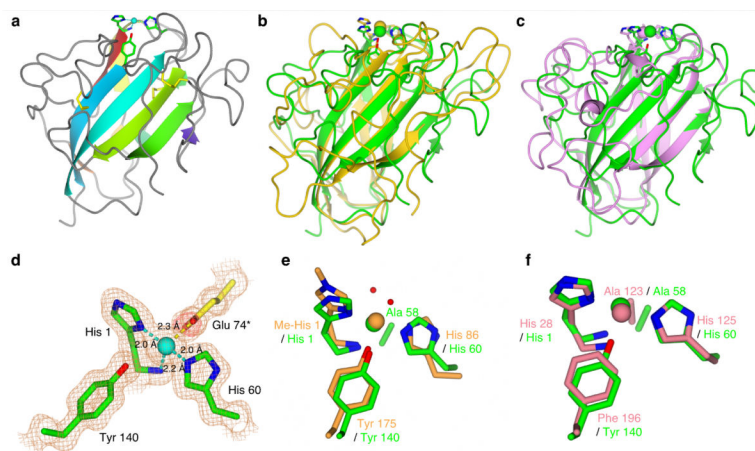
Figure 2.
"Module walking" to discover new LPMOs.

Figure 3.
Copper binding affinity and oxidative activity of *Ao*(AA11). (a) ITC thermogram of Zn^{2+} displacement by Cu^{2+} from the active site of *Ao*(AA11). (b) MALDI-TOF analysis of *Ao*(AA11) action on squid-pen chitin. DPn_{al} = aldonic acid, DPn_{-2} = oxidation from R-OH to R=O, (measured MW). DP5/DP5$_{-2}$ +Na$^+$ (1056.4, 1054.4), DP5$_{al}$ +Na$^+$ (1072.4), DP5$_{al}^-$ +2Na$^+$ (1094.4), DP6/DP6$_{-2}$ +Na$^+$ (1259.5, 1257.5), DP6$_{al}$ +Na$^+$ (1275.5), DP6$_{al}^-$ +2Na$^+$ (1297.5), DP7/DP7$_{-2}$ +Na$^+$ (1462.6, 1460.6), DP7$_{al}$ +Na$^+$ (1478.6), DP7$_{al}^-$ +2Na$^+$ (1500.6), DP8/DP8$_{-2}$ +Na$^+$ (1665.6, 1663.6), DP8$_{al}$ +Na$^+$ (1681.6), DP8$_{al}^-$ +2Na$^+$ (1703.6), DP9/DP9$_{-2}$ +Na$^+$ (1868.7, 1866.7), DP9$_{al}$ +Na$^+$ (1884.7), DP10/DP10$_{-2}$ +Na$^+$ (2071.8, 2069.8), DP10$_{al}$ +Na$^+$ (2087.8), DP10$_{al}^-$ +2Na$^+$ (2109.8). Magnified region of DP6 is shown in Supplementary Fig. 3.

Figure 4.

Structural comparisons of *Ao*(AA11) with known AA9 and AA10 enzymes. (a) 3D structure of Cu-*Ao*(AA11), ribbon depiction. The conserved active site residues are shown as sticks with green carbons and disulfide bonds (from conserved cysteines) as yellow sticks. (b) Overall superposition of Cu-*Ao*(AA11) (green) with Zn-(AA9) from *T. terrestris* (yellow) with rmsd = 2.6 Å over 145 Cα's (c) Superposition of Cu-*Ao*(AA11) (green) with Cu-(AA10) from *E. faecaelis* (pink) with r.m.s.d = 2.3 Å over 118 residues overlapping with a Cα's. (d) The electron density maps contoured at 1σ in the active site of Cu-*Ao*(AA11), Cu-N(His 1) = 1.97 Å, Cu-NH$_2$(His1) = 2.19 Å, Cu-N(His60) = 1.98 Å, N(His1)-Cu-NH$_2$ = 90.5°, N(His60)-Cu-NH$_2$ = 103.0°, N(His1)-Cu-N(His1) = 164.8°. Glu74, marked with asterisk is from a symmetry related molecule and is shown with yellow carbon atoms. (e) Active site overlay of *Ao*(AA11) (green carbons/copper) with Cu-AA9 from *T. aurantiacus* (orange carbons/copper), note side chain of conserved alanine 58, depicted as green rod in AA11. (f) The active site overlap of Cu-*Ao*(AA11) (green carbons/copper) with Cu-(AA10) from *B. amyloliquefaciens* (pink carbons/copper). See Supplementary figure 5 for stereo views of d-f.
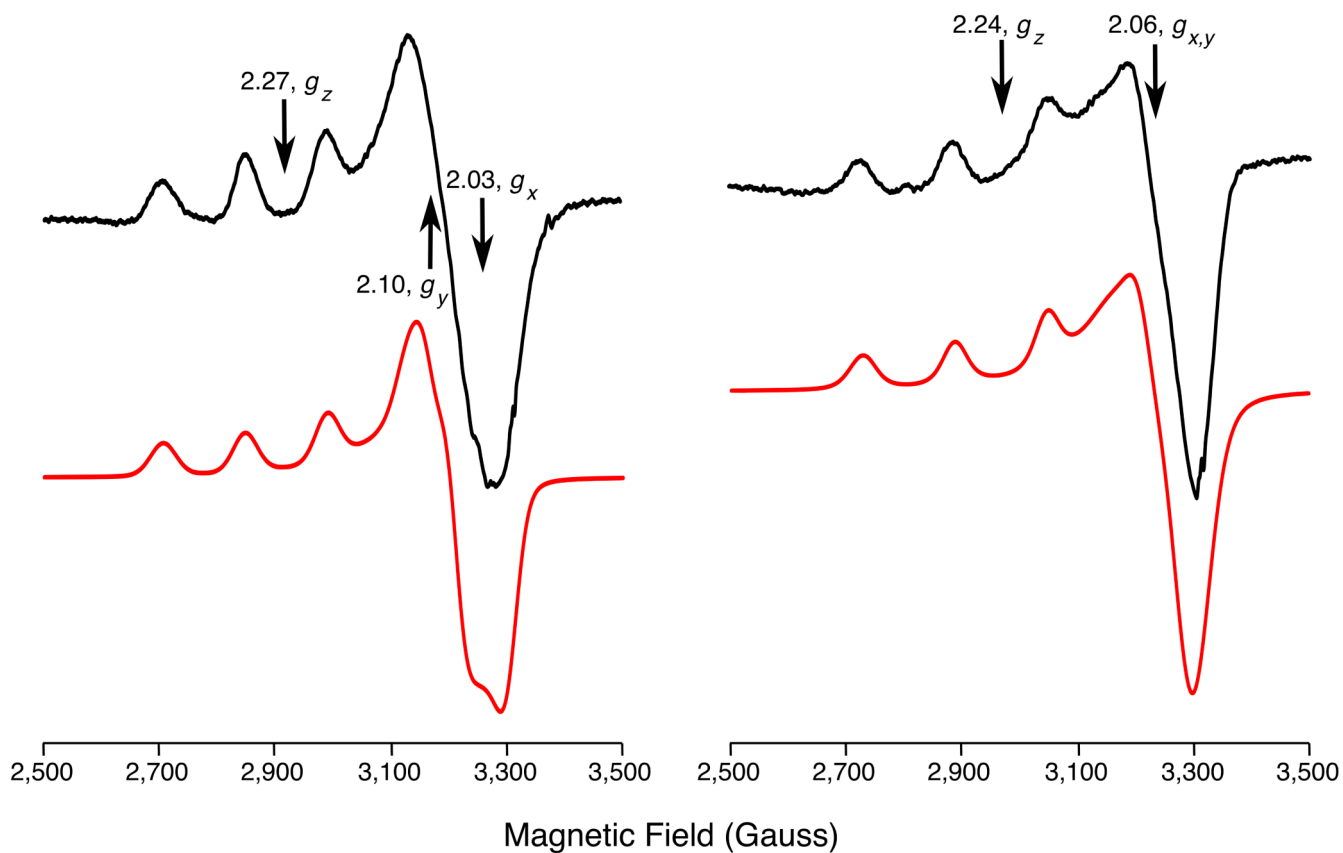
Figure 5.
The X-band EPR spectra of Cu enzymes (2500-3500 G, 9.3 GHz, 150 K) with simulations (red) of a) Cu-*Ao*(AA11), p*H* 5, 10% *v/v* glycerol and b) Cu-*Ao*(A11), p*H* 5, 10% *v/v* glycerol with 1000 equivalents of azide.