

# Structure of a C1/C4-oxidizing AA9 lytic polysaccharide monooxygenase from the thermophilic fungus *Malbranchea cinnamomea*

Scott Mazurkewich,<sup>a‡</sup> Andrea Seveso,<sup>a‡</sup> Silvia Hüttner,<sup>a</sup> Gisela Brändén<sup>b</sup> and Johan Larsbrink<sup>a\*</sup>

Received 22 May 2021

Accepted 25 June 2021

Edited by M. Czjzek, Station Biologique de Roscoff, France

‡ These authors contributed equally.

**Keywords:** lytic polysaccharide monooxygenases; LPMOs; auxiliary activity family 9; AA9; *Malbranchea cinnamomea*; enzyme structure.

**PDB reference:** McAA9F, 7ntl

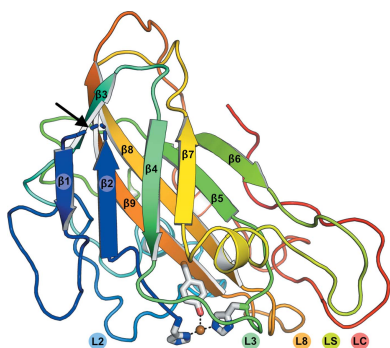
**Supporting information:** this article has supporting information at journals.iucr.org/d

<sup>a</sup>Wallenberg Wood Science Center, Division of Industrial Biotechnology, Department of Biology and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden, and <sup>b</sup>Department of Chemistry and Molecular Biology, University of Gothenburg, SE-405 30 Gothenburg, Sweden. \*Correspondence e-mail: johan.larsbrink@chalmers.se

The thermophilic fungus *Malbranchea cinnamomea* contains a host of enzymes that enable its ability as an efficient degrader of plant biomass and that could be mined for industrial applications. This thermophilic fungus has been studied and found to encode eight lytic polysaccharide monooxygenases (LPMOs) from auxiliary activity family 9 (AA9), which collectively possess different substrate specificities for a range of plant cell-wall-related polysaccharides and oligosaccharides. To gain greater insight into the molecular determinants defining the different specificities, structural studies were pursued and the structure of McAA9F was determined. The enzyme contains the immunoglobulin-like fold typical of previously solved AA9 LPMO structures, but contains prominent differences in the loop regions found on the surface of the substrate-binding site. Most significantly, McAA9F has a broad substrate specificity, with activity on both crystalline and soluble polysaccharides. Moreover, it contains a small loop in a region where a large loop has been proposed to govern specificity towards oligosaccharides. The presence of the small loop leads to a considerably flatter and more open surface that is likely to enable the broad specificity of the enzyme. The enzyme contains a succinimide residue substitution, arising from intramolecular cyclization of Asp10, at a position where several homologous members contain an equivalent residue but cyclization has not previously been observed. This first structure of an AA9 LPMO from *M. cinnamomea* aids both the understanding of this family of enzymes and the exploration of the repertoire of industrially relevant lignocellulolytic enzymes from this fungus.

## 1. Introduction

Biorefineries based on the conversion of lignocellulose will play an important role in the move towards a biobased, circular economy, in which renewable plant biomass is converted into valuable products such as fuels, materials and chemical precursors (Ubando *et al.*, 2020). One significant challenge in the field is the recalcitrance of plant biomass to deconstruction, remodeling or extraction of plant fibers (Sánchez & Cardona, 2008). Recalcitrance is a multi-factor and multi-scale property emerging from the complex molecular and structural features of the plant cell wall (McCann & Carpita, 2015), and polysaccharide crystallinity has been identified as an important factor impeding the decoupling of this intricate architecture (Himmel *et al.*, 2007). In nature, microorganisms harness a vast repertoire of enzymatic activities to access, deconstruct and utilize plant fibers as carbon sources and thus represent a reservoir of enzymes that could be mined for industrial biomass-deconstruction purposes. Fungi in particular are regarded as efficient plant cell-wall

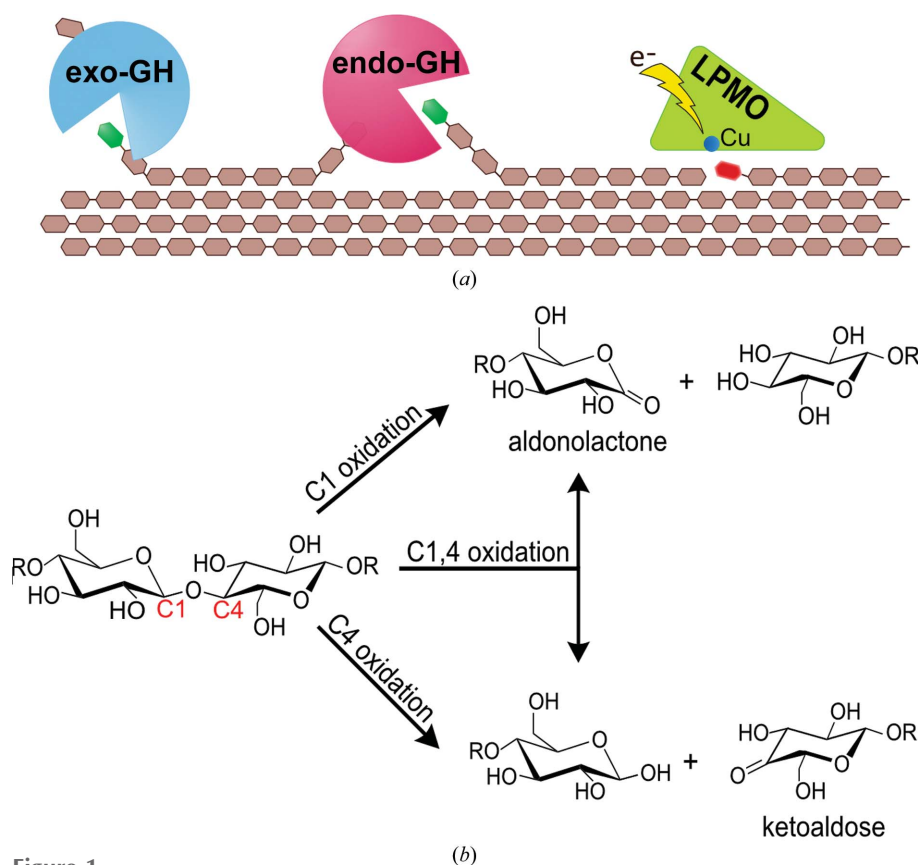


metabolizers and employ a wide range of enzymatic activities in the process. Lytic polysaccharide monooxygenases (LPMOs) are a relatively newly discovered class of enzymes and represent one such important activity as they can target highly recalcitrant motifs in the biomass. These include the flat surfaces of crystalline substrates such as cellulose, where an LPMO can generate new polysaccharide ends that aid in accessibility for further depolymerization by glycoside hydrolases (GHs; Fig. 1a).

Discovered a little over a decade ago, LPMOs are metallo-enzymes that utilize a copper ion to catalyze the oxidative cleavage of oligosaccharides and polysaccharides (Vaaje-Kolstad *et al.*, 2010; Morgenstern *et al.*, 2014). Interest in LPMOs has grown steadily since their discovery, especially given their importance in increasing biomass valorization yields when supplementing commercial enzyme cocktails (Harris *et al.*, 2014). LPMOs belong to the auxiliary activities (AAs) class in the Carbohydrate Active enZymes database (CAZy; <http://www.cazy.org>; Lombard *et al.*, 2014) and are divided into seven families on the basis of amino-acid sequence similarity [AA family 9 (AA9), AA10, AA11, AA13, AA14, AA15 and AA16]. While their exact catalytic mechanism remains under debate, LPMOs cleave glycosidic bonds by incorporating an O atom in the substrate (Vaaje-Kolstad *et al.*, 2010; Bissaro *et al.*, 2017). Proposed models

suggest that these enzymes use activated oxygen species to perform a redox reaction with the transfer of two electrons via the Cu atom coordinated by the so-called histidine brace in the active site. Catalysis is thought to be based on the utilization of molecular oxygen, with electrons delivered by an external donor (Vaaje-Kolstad *et al.*, 2010; Beeson *et al.*, 2012), although an alternative mechanism has recently been presented which is instead based on the utilization of hydrogen peroxide (Bissaro *et al.*, 2017; Hangasky *et al.*, 2018).

Within the AA class, AA9 has received considerable attention as it contains members with diverse substrate specificities ranging from crystalline cellulose to complex glucuronoxylans and also enzymes that are active on oligosaccharides (Lombard *et al.*, 2014). AA9 enzymes have exclusively been found in fungi and are categorized into subgroups based on the position of oxidation in their reaction products. These subgroups are C1-oxidizing enzymes that generate aldonic acids as products, C4-oxidizing enzymes that generate ketoaldoses as products, and C1,4-oxidizing enzymes, which are more promiscuous and can perform both C1-type and C4-type oxidations, leading to varying mixtures of the two product types (Fig. 1b). A phylogenetic analysis of AA9 showed that enzymes with different activities fell into distinct clades, and thus a type 1, 2 and 3 nomenclature for C1-, C4- and C1,4-oxidizing AA9 LPMOs, respectively, was proposed (Vu *et al.*, 2014). However, the preferred site of oxidation may not follow a strict phylogenetic relationship as several exceptions have been found (Vu *et al.*, 2014; Hüttner *et al.*, 2019). Filamentous fungi are regarded as major environmental biomass degraders and often encode a surprisingly large number of predicted AA9 enzymes. For example, *Podospira comata* contains at least 33 unique AA9 genes (Lombard *et al.*, 2014; Boucher *et al.*, 2017). While significant steps forward in our understanding of AA9 enzymes have been taken in recent years, the reasons behind their abundance in some genomes, and the molecular basis for their specificities and mechanism of action, require additional in-depth studies.



**Figure 1**  
Mode of action of AA9 LPMOs. (a) Model for cellulose degradation by LPMOs, which target crystalline regions, and endo- and exo-acting glycoside hydrolases (GHs), which target the ends of glycan chains or amorphous regions, respectively. (b) Different types of oxidation of the glycosidic bond performed by AA9 enzymes.

In a recent study, we sequenced the genome of the thermophilic fungus *Malbranchea cinnamomea*, which was isolated from a municipal waste-treatment factory in Vietnam (Hüttner *et al.*, 2017). This fungus is able to grow at temperatures above 50°C and has been found to be a great source of carbohydrate-active enzymes with potential use in industrial biomass-conversion applications (Mahajan *et al.*, 2016). *M. cinnamomea* encodes eight AA9

enzymes, several of which were upregulated during growth on wheat bran, xylan and cellulose when compared with glucose, thus indicating important roles for these enzymes in biomass turnover (Hüttner *et al.*, 2017). In a follow-up study, four of the AA9 LPMOs that were possible to heterologously produce in *Pichia pastoris* (*McAA9A*, *McAA9B*, *McAA9F* and *McAA9H*) were biochemically characterized. The LPMOs had different substrate specificities and were collectively active on a range of crystalline and soluble plant cell wall-related polysaccharides, as well as oligosaccharides (Hüttner *et al.*, 2019). *McAA9* enzymes are diverse in primary structure (sequence identities of between 29% and 61%), and while all were predicted to share the expected immunoglobulin G-like  $\beta$ -sandwich fold, considerable differences were predicted among their active-site surfaces. These differences were also reflected in the oxidation patterns, where *McAA9A*, *McAA9B* and *McAA9F* were predominantly C4-oxidizers, but also produced C1- and C1/C4-oxidized cellooligosaccharides from phosphoric acid-swollen cellulose (PASC), while *McAA9H* primarily produced C1-oxidized cellooligosaccharides. *McAA9A*, *McAA9B* and *McAA9F* (but not *McAA9H*) were shown to act on soluble tamarind xyloglucan (TXG;  $\beta$ -1,4-linked glucose backbone decorated with  $\alpha$ -1,6-linked xylose moieties which can be further appended with  $\beta$ -1,2-linked galactose moieties), with concomitant generation of C4-oxidized oligosaccharides, indicating that these enzymes can cleave between any two glucosyl residues in the xyloglucan backbone independently of the side chains present. Furthermore, only *McAA9A* and *McAA9F* showed clear activity on soluble cellooligosaccharides (cellohexaose), generating C4-oxidation products. The observation that each *McAA9* enzyme appears to have its own activity profile on soluble and crystalline substrates indicates that they each have a different biological role. The differences in activity are likely to stem from differences in the substrate-binding surfaces, and structural studies can thereby facilitate the identification of the molecular motifs responsible for enzyme–substrate interactions.

In this work, we present the structure of *McAA9F* and make comparisons with previously characterized AA9 enzymes, highlighting key differences and exploring the possible determinants defining substrate specificity in *McAA9F*.

## 2. Methods

### 2.1. Protein crystallization

*McAA9F* was produced heterologously in *Pichia pastoris* SMD1168H and purified by anion-exchange chromatography as described previously (Hüttner *et al.*, 2019). The protein was dialyzed into 50 mM tris(hydroxymethyl)aminomethane buffer pH 8.0 containing 50 mM NaCl. Crystallization conditions were screened with a Mosquito robot (SPT Labtech) using the JCSG+ screening kit (Molecular Dimensions) in MRC sitting-drop plates. Screens were set up with a reservoir volume of 40  $\mu$ l and protein mixed with reservoir solution in a 1:1 ratio in 0.6  $\mu$ l drops. After six months, needle-like crystals

**Table 1**

Table of crystallographic statistics for *McAA9F*.

Data collection	
Date	11 June 2020
Source	BioMAX at MAX IV
Wavelength (Å)	0.97625
Space group	<i>P12<sub>1</sub></i>
<i>a</i> , <i>b</i> , <i>c</i> (Å)	38.96, 43.13, 52.34
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00, 101.81, 90.00
No. of measured reflections	220814 (13034)
No. of independent reflections	34403 (2911)
Resolution (Å)	38.13–1.38 (1.43–1.38)
<i>R</i> <sub>merge</sub> (%)	6.38 (87.3)
CC <sub>1/2</sub> (%)	99.9 (71.5)
Mean <i>I</i> / $\sigma$ ( <i>I</i> )	13.7 (1.13)
Completeness (%)	92.3 (65.4)
Multiplicity	6.4 (4.5)
Refinement	
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.161/0.193
No. of atoms	
Protein	1661
Ligand/ions	22
Water	243
<i>B</i> factors (Å <sup>2</sup> )	
Protein	20.3
Ligand/ions	36.9
Water	31.2
R.m.s.d.	
Bond lengths (Å)	0.006
Bond angles (°)	0.91
PDB code	7ntl

formed in a condition (C1) which was further optimized to yield clusters of needles with increased dimensions that grew after four months. The final crystallization condition utilized consisted of 100 mM phosphate–citrate buffer pH 4.2 with 100 mM NaCl, 23.5% PEG 8000. A crystal was separated from the cluster, mounted and flash-cooled in liquid nitrogen in the absence of additional cryoprotectant.

### 2.2. Data collection and structure determination

A data set was collected to beyond 1.38 Å resolution from a crystal on the BioMAX beamline at MAX IV Laboratory on 11 June 2020. The data set was processed in *XDS* (Kabsch, 2010) and the structure was determined by molecular replacement with *Phaser* (McCoy *et al.*, 2007) in *Phenix* (Adams *et al.*, 2010) using the model of an LPMO from *Aspergillus fumigatus* (*AfAA9B*; PDB entry 6ha5; Lo Leggio *et al.*, 2018) as the search template. The initial phases were sufficient to enable direct manual rebuilding in *Coot* (Emsley *et al.*, 2010). *Coot* and *phenix.refine* (Afonine *et al.*, 2012) were used in iterative cycles of real-space and reciprocal-space refinement. The electron density is well defined, with only the last two residues (Ser221–Gly222) being unable to be resolved in the final model. The data-collection, processing and refinement statistics can be found in Table 1.

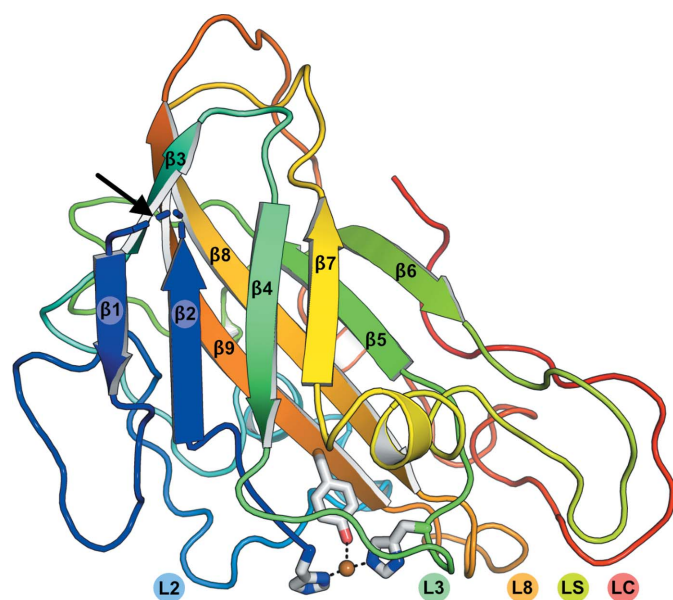
## 3. Results

### 3.1. *McAA9F* structure

Currently, more than 700 sequences encoding putative AA9 LPMOs can be found in CAZY, although protein structures of

only 16 members have been determined (Lombard *et al.*, 2014). To reveal the key determinants of the different substrate specificities that we observed for the AA9 LPMOs from *M. cinnamomea* (Hüttner *et al.*, 2019), we pursued structural studies of the enzymes and were able to generate crystals of *McAA9F*. The protein crystals grew as needle clusters; a data set was collected to 1.38 Å resolution and the structure was solved by molecular replacement.

The overall structure of *McAA9F* is comprised of an immunoglobulin-like topology with a central  $\beta$ -sandwich and some small peripheral helical sections, as is typical of previously solved AA9 structures (Fig. 2). The two  $\beta$ -sheets are composed of four and five  $\beta$ -strands, respectively, which are interconnected by several loops. The loops referred to as L2, L3, LS, L8 and LC (Wu *et al.*, 2013; Borisova *et al.*, 2015) participate in forming the flat surface typical of LPMOs acting on crystalline substrates, which also hosts the copper-binding site. Within L2 there are two  $3_{10}$ -helices ( $\eta$ -helices) separated by a turn. Another  $\eta$ -helix can be found in the long C-terminal loop LC, while a short  $\alpha$ -helix is present in the loop denominated LS. A disulfide bridge between Cys51 and Cys171, which is conserved amongst all structurally determined AA9 LPMO structures, connects loop L2 to strand  $\beta_9$ , while another disulfide bridge between Cys92 and Cys96 anchors a short unstructured region located after  $\beta_6$ . A fluorescence scan revealed the presence of a copper ion in the crystallized protein, which as expected was found to be coordinated by the hydroxyl of Tyr169 and the histidine brace: N $^{\delta}$  of His1 and N $^{\epsilon 2}$  of His81 (Supplementary Fig. S1). Fungal LPMOs are reported to carry an unusual methylation of the N $^{\epsilon 2}$  atom of the N-terminal histidine residue of the histidine brace, which



**Figure 2**  
Overall structure of *McAA9F*. The structure is colored blue to red from the N-terminus to the C-terminus. The copper ion and the residues coordinating it are shown. The position of the succinimide found in place of Asp10 in the loop between strands  $\beta_1$  and  $\beta_2$  is identified by a black arrow and the loops forming the flat surface comprising the substrate-binding face are labeled (L2, L3, L8, LS and LC).

has been suggested to convey protection against oxidative damage (Petrović *et al.*, 2018; Quinlan *et al.*, 2011). This post-translational modification is not seen in *McAA9F*, and is also not expected from heterologous protein production in *P. pastoris* (Petrović *et al.*, 2018). A succinimide is observed in place of Asp10, which is found in a turn connecting  $\beta_1$  and  $\beta_2$  (Supplementary Fig. S2). Succinimides can form as a result of a spontaneous cyclizing dehydration resulting from nucleophilic attack by the main-chain N atom on the  $\gamma$ -carbon of asparagine and aspartate side chains (Haley *et al.*, 1966; Geiger & Clarke, 1987). It is extremely rare to find such a motif in protein structures; it has currently been found in only 44 protein entries deposited in the PDB, and to date it has not been encountered in other LPMO structures.

### 3.2. Comparison with other structurally determined AA9 LPMOs

A structural alignment was made using the *DALI* server to aid in the comparison of *McAA9F* with other AA9 structures available in the PDB (Holm, 2020; Supplementary Table S1). The structure-based sequence identities ranged from 31% to 56% and the three enzymes with closest similarity were all predicted C1/C4-oxidizing AA9 LPMOs: *AfAA9B* from *A. fumigatus* (PDB entry 5x6a; 56% identity; Q. Shen, unpublished work), *TaAA9A* from *Thermoascus aurantiacus* (PDB entry 2yet; 54% identity; Quinlan *et al.*, 2011) and *TrAA9A* from *Trichoderma reesei* (PDB entry 5o2w; 52% identity; Hansson *et al.*, 2017). The expected conserved overall  $\beta$ -sandwich architecture of AA9 LPMOs was clearly evident, especially the internal  $\beta$ -sheets of the molecule. The active-site geometry is also highly conserved, with a close-to-identical arrangement of the histidine residues forming the brace to coordinate the Cu atom together with the tyrosine residue (Tyr169 in *McAA9F*), which is found in an axial position pointing towards the flat surface (Fig. 3a). Similar to all previously determined AA9 LPMO structures, a glutamine and a histidine residue (Gln167 and His158 in *McAA9F*), which have been suggested to be involved in O<sub>2</sub> activation, are positioned in proximity to the copper-coordinating residues (O'Dell *et al.*, 2017; Span *et al.*, 2017). While containing considerable similarities to other AA9 LPMOs, particularly in the catalytic core, significant differences are observed in *McAA9F*, especially in the various loops involved in forming the substrate-binding surface from the N-terminal end of the molecule.

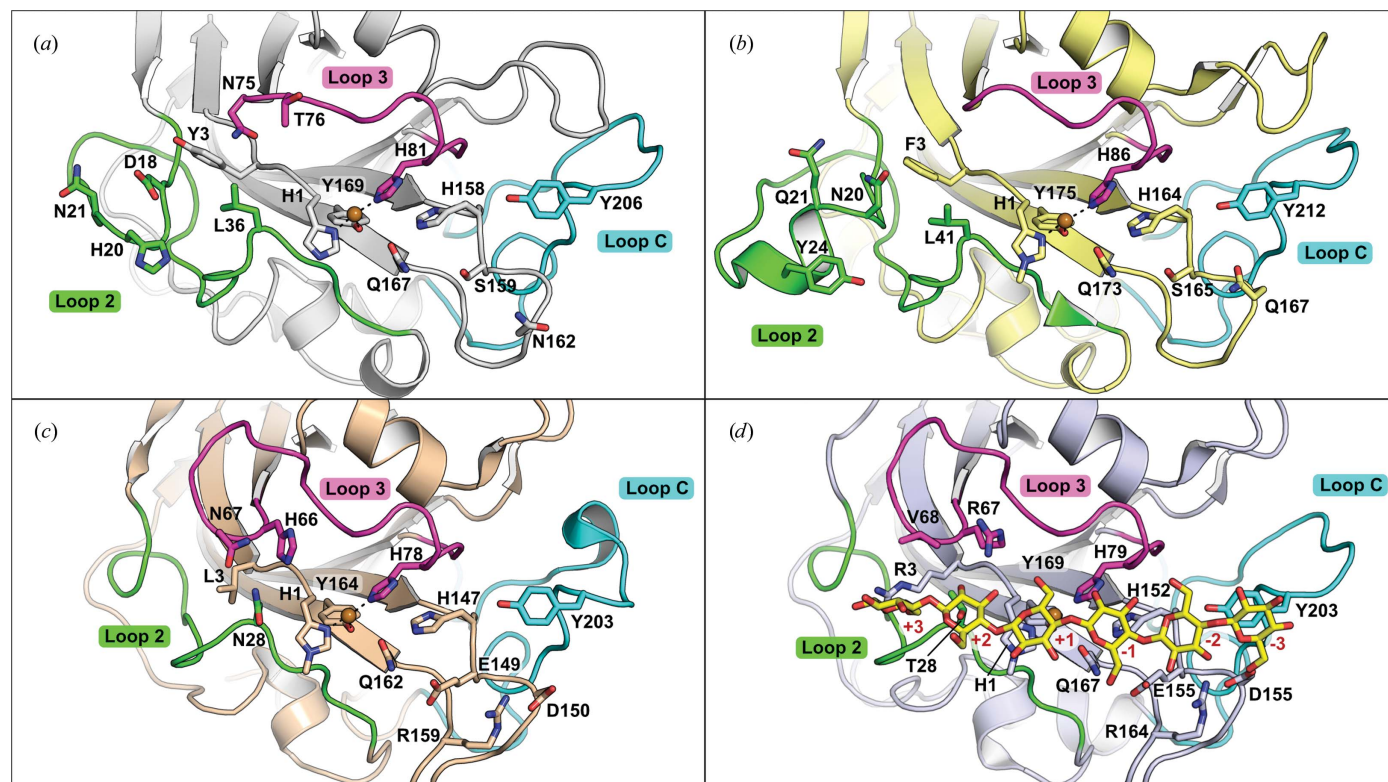
A prominent feature found amongst some C1/C4-oxidizing AA9 enzymes (for example *TaAA9A*) is the presence of an extended L2 to include a short (six-residue)  $\alpha$ -helix that packs parallel to the flat surface and contains a tyrosine residue (Tyr24 in *TaAA9A*) which has been suggested to be involved in stacking interactions with the substrate (Li *et al.*, 2012; Fig. 3b). The extended L2 responsible for this motif was shown to be important for regioselectivity and suggested to govern C4-oxidation in certain AA9 enzymes. A variant of the C1/C4-oxidizing NCU07760 from *Neurospora crassa* (UniProt entry Q7S111) lacking this insert predominantly produced

C1-oxidative products from PASC (Vu *et al.*, 2014). In comparison to *TaAA9A*, the L2 region in *McAA9F* is shorter by five residues and lacks the helical motif. *McAA9F* also lacks the conserved tyrosine residue of the motif, but instead possesses a histidine residue (His20), the imidazole side chain of which occupies an equivalent position and is likely to fulfill an equivalent proposed substrate-stacking role. However, the molecular determinants that define C4-oxidative regioselectivity in the AA9 family may in fact not depend upon factors in L2 since other members such as *CvAA9A* from *Collariella virescens* (PDB entry 5nlt) and *LsAA9A* from *Lentinus similis* (PDB entry 5acf), two AA9 enzymes with significantly shorter L2 regions, are capable of performing both C1- and C4-oxidation (Simmons *et al.*, 2017; Frandsen *et al.*, 2016). A more recent proposal suggests that since similar distances are observed between the C1 and C4 axial protons relative to the active oxygen species in the AA9–substrate complexes, small perturbations in substrate binding may lead to a preferred site of attack, rather than the regioselectivity being more strictly defined by phylogenetic relationships (Simmons *et al.*, 2017; Frandsen *et al.*, 2016).

Another characteristic of *McAA9F* is that it has a very short L3 loop compared with those found in C4-oxidizing AA9 structures (*NcAA9A*, PDB entry 5foh; *NcAA9C*, PDB entry 4d7u; *NcAA9D*, PDB entry 4eir; Petrović *et al.*, 2019; Borisova *et al.*, 2015; Li *et al.*, 2012) and in the aforementioned *LsAA9A* and *CvAA9A*, which are C1/C4-oxidizing AA9 enzymes that

are active on oligosaccharides (Fig. 3c). The presence of an extended L3 loop, which can even harbor a helix formation, has been associated with a capability to act on soluble substrates such as celooligosaccharides and xyloglucan (Simmons *et al.*, 2017; Frandsen *et al.*, 2021). It has also been observed that a truncation of L3 in *NcAA9C* abolished the activity on xyloglucan (Laurent *et al.*, 2019). Furthermore, Frandsen and coworkers showed that in *LsAA9A* the extended L3 loop contributes to the formation of a ridge protruding from the binding surface, which is involved in the interaction with soluble substrates (Frandsen *et al.*, 2016). In the same study it was also speculated that the presence of this formation caused a lack of enzymatic activity on crystalline substrates. This would be due to the ridge affecting the flatness of the typical LPMO surface, which is suggested to be one of the key factors for the activity on flat, crystalline fibers (Frandsen *et al.*, 2016). Interestingly, *McAA9F* showed activity on both crystalline cellulose and on soluble substrates while containing an extremely short L3 loop aligned to the surface level, indicating that the absence of an extended L3 is not sufficient to define a lack of activity towards soluble substrates and oligosaccharides.

*McAA9F* is the first reported AA9 to contain a succinimide modification. The aspartate of this position (Asp10) is conserved in many other AA9 enzymes or is replaced by an asparagine, which is equally prone to cyclization (Clarke, 1987; Clarke *et al.*, 1992; Supplementary Table S1). A proposed



**Figure 3**

Comparison of the substrate-binding sites of *McAA9A* and selected AA9 LPMOs. Key residues lining the flat surface of (a) *McAA9F*, (b) *TaAA9A* (PDB entry 2yet), (c) *LsAA9A* (PDB entry 5acf) and (d) *CvAA9A* in complex with cellohexaose (PDB entry 6yde) are shown. Loop regions corresponding to L2, L3 and LC for each protein are colored green, magenta and cyan, respectively.

model for succinimide formation (Bornstein & Balian, 1977) suggests the absence of steric hindrance close to the main-chain N atom, which performs the attack on the side-chain carbonyl, as an important requirement for cyclization. In *McAA9F*, Asp10 is preceded by an alanine and is followed by a glycine, which fulfills these requirements. However, identical or very similar compositions, with valine or isoleucine substituting for the alanine, are found in the majority of AA9 structures, suggesting that this XDG (or XNG) pattern in the sequence alone might not be sufficient. The occurrence and formation of succinimides can indeed vary depending on the environmental conditions, such as buffer composition (Geiger & Clarke, 1987; Hooi *et al.*, 2013), and it is possible that the observed succinimide in the structure of *McAA9F* is a result of the crystallization conditions or of the protein-production process.

### 3.3. Comparisons to LPMOs with bound ligands

In the crystal form of *McAA9F* that we obtained, the putative substrate-binding surface is heavily hindered by crystal contacts and thus it would not be possible to pursue ligand complexes using this crystal form (Supplementary Fig. S3). To gain greater insights into the potential key substrate-interacting residues in *McAA9F*, we compared the structure with those of other AA9 enzymes in complex with ligands. Both *LsAA9A* (32% structure-based sequence identity to *McAA9F*) and *CvAA9A* (35% structure-based sequence identity to *McAA9F*) are C1/C4-oxidizing LPMOs with an activity profile similar to that of *McAA9F*, and structures of both have been solved in complex with cellobiose (Frandsen *et al.*, 2016; Tandrup *et al.*, 2020), towards which *McAA9F* also has activity (Hüttner *et al.*, 2019). Although they share <40% sequence identity, *CvAA9A*–cellobiose (PDB entry 6yde) and *LsAA9A*–cellobiose (PDB entry 5aci) are closely related in structure to *McAA9F*, with root-mean-square deviations of 1.7 and 1.9 Å, respectively, making them good candidates for comparison.

*LsAA9A* binds the oligosaccharide from subsites –4 to +2, while in *CvAA9A* the cellobiose molecule is bound from subsites –3 to +3; however, the orientations and positions of the glucose monomers are conserved. Many of the key residues that make interactions with the substrate from the +1 to –3 sites are either conserved or functionally similar in *McAA9F*, including a highly conserved tyrosine residue in LC (Tyr206 in *McAA9F*), which has been suggested in multiple studies to be involved in stacking interactions with the flat pyranose ring of the substrates (Li *et al.*, 2012; Harris *et al.*, 2010; Asensio *et al.*, 2013; Nishio, 2011) and is likely to support substrate binding in a similar fashion in *McAA9F* (Fig. 3*d*). Greater differences in *McAA9F* are present in the +2 sugar-binding site and further towards the substrate reducing-end sites, particularly originating from the differences in L2 and L3 described above. In the extended L3 region of *LsAA9A*, His66 and Asn67 (PDB entry 5aci) interact with the oligosaccharide by hydrogen bonding to the hydroxyl moieties of C2 and C3 of the sugar at the +2 site. In *CvAA9A*, leucine and

arginine residues (Leu66 and Arg67 in PDB entry 6yde) are found at equivalent positions, with the arginine making a hydrogen bond to the C3 atom of the sugar at the +2 site. While L3 in *McAA9F* contains hydrophilic residues capable of interacting with substrates, its considerably smaller size results in the loop being set back >5.5 Å from the +2 sites observed in *CvAA9A* and *LsAA9A*. Without considerable rearrangement, it is unlikely that the residues of the short L3 in *McAA9F* will be in direct interaction with substrates at the +2 site positions as observed in *CvAA9A*, *LsAA9A* and other AA9 LPMOs. Further, in both *CvAA9A* and *LsAA9A* a hydrophilic residue in L2 (Asn28 in PDB entry 5aci and Thr28 in PDB entry 6yde) forms an additional hydrogen bond to the C2 hydroxyl of the sugar at the +2 site. The L2 region in *McAA9F* is modeled differently and has a leucine residue (Leu36) in the corresponding position with its side chain rotated towards the core of the protein and contributing to creating a flat surface. As previously mentioned, packing of the protein in this lattice leads to considerable contacts across the putative substrate-binding site and may cause some distortions in loop and residue positioning. However, the L2 and L3 regions of *McAA9F* are distinct compared with *CvAA9A* and *LsAA9A*, which may lead to differences in the positioning of oligosaccharides from the +2 site to the reducing end compared with the other complexes.

## 4. Discussion

The number of AA9 LPMOs in the genomes of filamentous fungi is often very high, especially in species growing on complex biomass (Lombard *et al.*, 2014). The multiplicity of AA9 LPMOs in fungal genomes has been highlighted previously (Jagadeeswaran *et al.*, 2016; Hüttner *et al.*, 2019; Petrović *et al.*, 2019), but the reason behind this multiplicity remains unclear. To date, characterization of the full AA9 LPMO repertoire from a single species has not been achieved, which precludes a deeper understanding of their respective biological roles. In many species, however, there are differences in substrate and product specificities among the studied LPMOs, with enzymes acting on both crystalline and insoluble cellulose and soluble heteroglycans, and with C1- and C4- as well as C1/C4-oxidation of the products. *N. crassa*, which is one of the most extensively studied species with respect to its LPMO repertoire, encodes 14 AA9 members in its genome, nine of which have been characterized and shown to be active on cellulose with C1-, C4- and C1/C4-oxidizing activity (Zhou & Zhu, 2020). Similarly, the thermophilic fungus *M. cinnamomea* encodes eight putative AA9 enzymes, four of which have been biochemically characterized and shown to oxidize a range of both insoluble and soluble substrates with different oxidation patterns (Hüttner *et al.*, 2019). Curiously, none of the *M. cinnamomea* AA9 enzymes contain appended carbohydrate-binding modules, which have been shown to affect regioselectivity in other LPMOs (Chalak *et al.*, 2019).

The structure of *McAA9F* represents the first structurally determined AA9 LPMO from *M. cinnamomea*. The overall fold of the enzyme is similar to those of structurally

determined AA9 members, with the typical  $\beta$ -sandwich fold and the expected histidine brace holding the catalytic Cu atom. A peculiar and unexpected feature of the *McAA9F* structure is the presence of a succinimide within the protein sequence, which is the first to be reported in a LPMO structure and which originates from the spontaneous cyclization of the aspartate at position 10. No apparent implications for the enzyme activity were observed (Hüttner *et al.*, 2019), but further structural studies on this and other AA9 LPMOs might help to elucidate the reason behind the formation of such a motif.

One of the most intriguing observations in *McAA9F* is the lack of the extended L3 that has been reported to play a major role in interaction with soluble substrates, despite *McAA9F* being active on oligosaccharides. A superposition with other structures in complex with celohexaose showed that certain amino-acid residues that represent contact points with the oligosaccharide substrates do not have an equivalent residue in *McAA9F*, suggesting that there might be other regions in *McAA9F* that are involved in binding that have not yet been identified and described. Alternatively, celohexaose could be bound on the *McAA9F* surface in a slightly different position, particularly in the +2 site and towards the reducing end, due to the lack of the constraints imposed by the ridge that is observed in members with an extended L3, thus possibly exploiting new points of interaction not described to date. *McAA9F* also has activity on the hemicellulosic heteropolysaccharide xyloglucan (Hüttner *et al.*, 2019). No AA9 LPMO structure in complex with a xyloglucan oligosaccharide has yet been obtained and reported, making it difficult to speculate how it would be bound to the surface of the enzyme considering the steric volume represented by the xyloside substituents, although these would be expected to lie in the same flat plane as the glucose-based backbone. The xyloglucan-active *LsAA9A* and *CvAA9A* require that xylosyl substitutions occur only on the backbone pyranose monomers at the -3, -2, -1, +2 and +3 subsites, with the +1 subsite occupied by an unsubstituted glucose moiety (Simmons *et al.*, 2017), while *McAA9A*, *McAA9B* and *McAA9F* can cleave the substrate regardless of the position of the substitutions. These three LPMOs from *M. cinnamomea* all have a predicted significantly shorter L3 loop compared with those exhibited by *LsAA9A* and *CvAA9A* (Hüttner *et al.*, 2019; Supplementary Fig. S4), which suggests that the ridge protruding from the flat surface of the latter enzymes might pose the steric constraints limiting the binding of C6-substituted sugars at the +1 site.

Comparing the sequences of all characterized AA9 enzymes from *M. cinnamomea*, the other type 3 AA9 LPMOs (*McAA9A* and *McAA9B*) both have a long L2 region similar to that observed in *LsAA9A* and *CvAA9A*, with a conserved tyrosine residue likely to aid in binding polysaccharide chains beyond the +2 site (Supplementary Fig. S4). The L2 region in *McAA9F* is shorter and is modeled differently but contains a histidine residue that can likely fulfill the functionality of the conserved tyrosine. *McAA9A*, *McAA9B* and *McAA9F* are all active towards PASC, celooligosaccharides and xyloglucan. Each contains a short L3 region which, as seen in *McAA9F*,

creates more space around the +2 subsite than is observed in other LPMOs and is likely to contribute to the broad specificity amongst the enzymes. *McAA9H*, which lacks activity on both PASC and celooligosaccharides but has activity towards xylan, lacks the majority of the L2 region, particularly the portion containing the typically conserved L2 tyrosine residue, and has a short L3 region similar to the other characterized AA9 enzymes from *M. cinnamomea*. The determinants resulting in the shift from PASC to xylan activity in *McAA9H* remain elusive and further research is needed to explore the molecular features defining this interesting and thus far rare substrate specificity.

## 5. Related literature

The following reference is cited in the supporting information for this article: Liebschner *et al.* (2017).

## Acknowledgements

X-ray diffraction data were collected on the BioMAX beamline at MAX IV Laboratory (proposal No. 20190298).

## Funding information

This work was supported by the Knut and Alice Wallenberg Foundation via the Wallenberg Wood Science Center.

## References

- Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L.-W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). *Acta Cryst.* **D66**, 213–221.
- Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H. & Adams, P. D. (2012). *Acta Cryst.* **D68**, 352–367.
- Asensio, J. L., Ardá, A., Cañada, F. J. & Jiménez-Barbero, J. (2013). *Acc. Chem. Res.* **46**, 946–954.
- Beeson, W. T., Phillips, C. M., Cate, J. H. D. & Marletta, M. A. (2012). *J. Am. Chem. Soc.* **134**, 890–892.
- Bissaro, B., Røhr, K., Müller, G., Chylenski, P., Skaugen, M., Forsberg, Z., Horn, S. J., Vaaje-Kolstad, G. & Eijsink, V. G. H. (2017). *Nat. Chem. Biol.* **13**, 1123–1128.
- Borisova, A. S., Isaksen, T., Dimarogona, M., Kognole, A. A., Mathiesen, G., Várnai, A., Røhr, K., Payne, C. M., Sørli, M., Sandgren, M. & Eijsink, V. G. H. (2015). *J. Biol. Chem.* **290**, 22955–22969.
- Bornstein, P. & Balian, G. (1977). *Methods Enzymol.* **47**, 132–145.
- Boucher, C., Nguyen, T.-S. & Silar, P. (2017). *Cryptogam. Mycol.* **38**, 485–506.
- Chalak, A., Villares, A., Moreau, C., Haon, M., Grisel, S., d'Orlando, A., Herpoël-Gimbert, I., Labourel, A., Cathala, B. & Berrin, J. (2019). *Biotechnol. Biofuels*, **12**, 206.
- Clarke, S. (1987). *Int. J. Pept. Protein Res.* **30**, 808–821.
- Clarke, S., Stephenson, R., Lowenson, J., Ahern, T. & Manning, M. (1992). *Pharm. Biotechnol.* **2**, 1–29.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Frandsen, K. E. H., Haon, M., Grisel, S., Henrissat, B., Lo Leggio, L. & Berrin, J.-G. (2021). *J. Biol. Chem.* **296**, 100086.
- Frandsen, K. E. H., Simmons, T. J., Dupree, P., Poulsen, J. N., 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. & Walton, P. H. (2016). *Nat. Chem. Biol.* **12**, 298–303.
- Geiger, T. & Clarke, S. (1987). *J. Biol. Chem.* **262**, 785–794.
- Haley, E. E., Corcoran, B. J., Dorer, F. E. & Buchanan, D. L. (1966). *Biochemistry*, **5**, 3229–3235.
- Hangasky, J. A., Iavarone, A. T. & Marletta, M. A. (2018). *Proc. Natl Acad. Sci. USA*, **115**, 4915–4920.
- Hansson, H., Karkehabadi, S., Mikkelsen, N., Douglas, N. R., Kim, S., Lam, A., Kaper, T., Kelemen, B., Meier, K. K., Jones, S. M., Solomon, E. I. & Sandgren, M. (2017). *J. Biol. Chem.* **292**, 19099–19109.
- Harris, P. V., Welner, D., McFarland, K., Re, E., Navarro Poulsen, J.-C., Brown, K., Salbo, R., Ding, H., Vlasenko, E., Merino, S., Xu, F., Cherry, J., Larsen, S. & Lo Leggio, L. (2010). *Biochemistry*, **49**, 3305–3316.
- Harris, P. V., Xu, F., Kreel, N. E., Kang, C. & Fukuyama, S. (2014). *Curr. Opin. Chem. Biol.* **19**, 162–170.
- Himmel, M. E., Ding, S.-Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W. & Foust, T. D. (2007). *Science*, **315**, 804–807.
- Holm, L. (2020). *Protein Sci.* **29**, 128–140.
- Hooi, M. Y., Raftery, M. J. & Truscott, R. J. (2013). *Mech. Ageing Dev.* **134**, 103–109.
- Hüttner, S., Nguyen, T. T., Granchi, Z., Chin-A-Woeng, T., Ahrén, D., Larsbrink, J., Thanh, V. N. & Olsson, L. (2017). *Biotechnol. Biofuels*, **10**, 265.
- Hüttner, S., Várnai, A., Petrović, D. M., Bach, C. X., Kim Anh, D. T., Thanh, V. N., Eijssink, V. G. H., Larsbrink, J. & Olsson, L. (2019). *Appl. Environ. Microbiol.* **85**, e01408-19.
- Jagadeeswaran, G., Gainey, L., Prade, R. & Mort, A. J. (2016). *Appl. Microbiol. Biotechnol.* **100**, 4535–4547.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Laurent, C. V. F. P., Sun, P., Scheiblbrandner, S., Csarman, F., Cannazza, P., Frommhagen, M., van Berkel, W. J. H., Oostenbrink, C., Kabel, M. A. & Ludwig, R. (2019). *Int. J. Mol. Sci.* **20**, 6219.
- Li, X., Beeson, W. T., Phillips, C. M., Marletta, M. A. & Cate, J. H. (2012). *Structure*, **20**, 1051–1061.
- Liebschner, D., Afonine, P. V., Moriarty, N. W., Poon, B. K., Sobolev, O. V., Terwilliger, T. C. & Adams, P. D. (2017). *Acta Cryst.* **D73**, 148–157.
- Lo Leggio, L., Weihe, C. D., Poulsen, J. N., Sweeney, M., Rasmussen, F., Lin, J., De Maria, L. & Wogulis, M. (2018). *Carbohydr. Res.* **469**, 55–59.
- Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. (2014). *Nucleic Acids Res.* **42**, D490–D495.
- Mahajan, C., Basotra, N., Singh, S., Di Falco, M., Tsang, A. & Chadha, B. S. (2016). *Bioresour. Technol.* **200**, 55–63.
- McCann, M. C. & Carpita, N. C. (2015). *J. Exp. Bot.* **66**, 4109–4118.
- McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.
- Morgenstern, I., Powlowski, J. & Tsang, A. (2014). *Brief. Funct. Genomics*, **13**, 471–481.
- Nishio, M. (2011). *Phys. Chem. Chem. Phys.* **13**, 13873–13900.
- O'Dell, W. B., Agarwal, P. K. & Meilleur, F. (2017). *Angew. Chem. Int. Ed.* **56**, 767–770.
- Petrović, D. M., Bissaro, B., Chylenski, P., Skaugen, M., Sørliie, M., Jensen, M. S., Aachmann, F. L., Courtade, G., Várnai, A. & Eijssink, V. G. H. (2018). *Protein Sci.* **27**, 1636–1650.
- Petrović, D. M., Várnai, A., Dimarogona, M., Mathiesen, G., Sandgren, M., Westereng, B. & Eijssink, V. G. (2019). *J. Biol. Chem.* **294**, 15068–15081.
- Quinlan, R. J., Sweeney, M. D., Lo Leggio, L., Otten, H., Poulsen, J. N., Johansen, K. S., Krogh, K. B., Jørgensen, C. I., Tovborg, M., Anthonsen, A., Tryfona, T., Walter, C. P., Dupree, P., Xu, F., Davies, G. J. & Walton, P. H. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 15079–15084.
- Sánchez, Ó. J. & Cardona, C. A. (2008). *Bioresour. Technol.* **99**, 5270–5295.
- 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. & Dupree, P. (2017). *Nat. Commun.* **8**, 1064.
- Span, E. A., Suess, D. L. M., Deller, M. C., Britt, R. D. & Marletta, M. A. (2017). *ACS Chem. Biol.* **12**, 1095–1103.
- Tandrup, T., Tryfona, T., Frandsen, K. E. H., Johansen, K. S., Dupree, P. & Lo Leggio, L. (2020). *Biochemistry*, **59**, 3347–3358.
- Ubando, A. T., Felix, C. B. & Chen, W. H. (2020). *Bioresour. Technol.* **299**, 122585.
- Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørliie, M. & Eijssink, V. G. H. (2010). *Science*, **330**, 219–222.
- Vu, V. V., Beeson, W. T., Phillips, C. M., Cate, J. H. D. & Marletta, M. A. (2014). *J. Am. Chem. Soc.* **136**, 562–565.
- Wu, M., Beckham, G. T., Larsson, A. M., Ishida, T., Kim, S., Payne, C. M., Himmel, M. E., Crowley, M. F., Horn, S. J., Westereng, B., Igarashi, K., Samejima, M., Ståhlberg, J., Eijssink, V. G. H. & Sandgren, M. (2013). *J. Biol. Chem.* **288**, 12828–12839.
- Zhou, X. & Zhu, H. (2020). *Bioresour. Bioprocess.* **7**, 11.