Oligosaccharide Binding and Thermostability of Two Related AA9 Lytic Polysaccharide Monooxygenases

Tobias Tandrup, Theodora T. Tryfona, Kristian Erik Høpfner Frandsen, Katja Salomon Johansen, Paul P. Dupree, Leila Lo Leggio

To cite this version:

Tobias Tandrup, Theodora T. Tryfona, Kristian Erik Høpfner Frandsen, Katja Salomon Johansen, Paul P. Dupree, et al.. Oligosaccharide Binding and Thermostability of Two Related AA9 Lytic Polysaccharide Monooxygenases. Biochemistry, 2020, 59 (36), pp.3347-3358. 10.1021/acs.biochem.0c00312 . hal-03167196

HAL Id: hal-03167196
https://hal.inrae.fr/hal-03167196

Submitted on 24 Mar 2021

Oligosaccharide binding and thermostability of two related AA9 Lytic Polysaccharide Monooxygenases

Tobias Tandrup[1], Theodora Tryfona[2], Kristian Erik Høpfner Frandsen[1,3], Katja Salomon Johansen[4], Paul Dupree[2], and Leila Lo Leggio[1]*

[1]Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100-DK, Copenhagen, Denmark. [2]Department of Biochemistry, University of Cambridge, Cambridge, CB2 1QW, UK. [3]INRAE, Aix-Marseille Université, Biodiversité et Biotechnologie Fongiques (BBF), Marseille, France. [4]Department for Geosciences and Natural resource management, University of Copenhagen, 1958-DK, Frederiksberg, Denmark.

KEYWORDS.

Crystal structure, substrate specificity, structural biology, oligosaccharide, copper monooxygenase, lytic polysaccharide monooxygenases (LPMOs).

ABSTRACT.

Lytic polysaccharide monooxygenases (LPMOs) are copper-dependent enzymes which cleave polysaccharide substrates oxidatively. First discovered because of their action on recalcitrant crystalline substrates (chitin and cellulose) a number of LPMOs are now reported to act on soluble substrates including oligosaccharides. However, crystallographic complexes with oligosaccharides have only been reported for a single LPMO so far, an enzyme from the basidiomycete fungus *Lentinus similis* (*Ls*AA9_A). Here we present a more detailed comparative study of *Ls*AA9_A and an LPMO from the ascomycete fungus *Collariella virescens* (*Cv*AA9_A) with which it shares 41.5% sequence identity. *Ls*AA9_A is considerably more thermostable than *Cv*AA9_A, and the structural basis for the difference has been investigated. We have compared the patterns of oligosaccharide cleavage and the patterns of binding in several new crystal structures explaining the basis for product preferences by the two enzymes. Obtaining structural information on complexes of LPMOs with carbohydrates has proven very difficult in general judging from the structures reported in the literature thus far and this can only partly be attributed to low affinity for small substrates. We have thus evaluated the use of differential scanning fluorimetry as a guide to obtaining complex structures. Furthermore, an analysis of crystal packing of LPMOs and glycoside hydrolases, corroborates the hypothesis that active site occlusion is a very significant problem for LPMO-substrate interaction analysis by crystallography, due to their relatively flat and extended substrate binding sites.

INTRODUCTION

Lytic polysaccharide monooxygenases (LPMOs) seem to play vital roles in the degradation of recalcitrant biomass. These copper containing enzymes cleave polysaccharides oxidatively rather than through the more wide-spread hydrolytic mechanism of glycoside hydrolases[1,2]. The discovery of cellulose-active LPMOs has led to their incorporation in commercial enzyme cocktails (such as Cellic® CTec) for a more efficient enzymatic degradation of biomass in the biorefinery industry[3,4].

Since 2013 LPMOs have been classified in Auxiliary Activity enzyme families 9-11 and 13-16 (AA9-AA11 and AA13-AA16) in the Carbohydrate-Active enzyme database (CAZy)[5–7]. The first LPMOs studied were of fungal or bacterial origin and demonstrated to be active on cellulose or chitin[1,8]. Later, enzymes in the AA13 family were found to act on starch-derived substrates[9]. The most recently discovered LPMO families are AA14 showing activity on cellulose-associated xylan[10], AA15 for which members are found in arthropod species (*e.g.* insects)[7,11] and the fungal, cellulose active AA16[7]. Recently an AA10 of fern origin has also been discovered[12] as has an AA9 LPMO with preferential activity on xylan[13].

Shared by all LPMOs is the Cu-coordinating histidine-brace motif (His-brace)[1] made up of an N-terminal His (often Nε-methylated in filamentous fungal species), and a second His later in the sequence. These occupy the ligand positions in the equatorial plane of the Cu and are often accompanied by a Tyr occupying the axial ligand position. In addition, exogenous ligands are often found in the axial- and/or equatorial coordination positions depending on the oxidation state of the Cu. LPMOs use a co-substrate (molecular oxygen or hydrogen peroxide) and an electron donor (such as small molecules, other oxidoreductases or lignin[14–16]) in order to produce an oxygen species able to oxidize its substrate[2,17].

A large number of high-resolution LPMO structures have been determined (more than 38 different proteins), which have provided useful information for understanding the LPMO function, as reviewed recently *e.g.* in[18–22], but despite this the crystallographic information on LPMO substrate binding is scarce compared to other CAZymes. A pivotal structural finding came with the first crystallographic information on carbohydrate ligand binding by LPMOs from studies on an AA9 family enzyme from the basidiomycete fungus *Lentinus similis* (*Ls*AA9_A) in 2016[23]. In that study enzyme kinetics, EPR spectroscopy and X-ray crystallography indicated that cello-oligosaccharide binding leads to displacement of the axial exogenous ligand of the copper, a shortening of the Tyr-Cu(II) distance, and that the substrate is connected to the LPMO N-terminus via a hydrogen bonding network, leading to activation of the copper. In several of the *Ls*AA9_A crystal structures, a Cl^- ion in the equatorial position (from crystallization conditions) has been proposed as an oxygen species mimic[23].

In a follow up biochemical and structural study, the same *Ls*AA9_A enzyme and a homologue from the ascomycete fungus *Collariella virescens* (*Cv*AA9_A) with which it shares 41.5% sequence identity, were further investigated. Both enzymes had activity on a range of soluble poly- and oligosaccharides[24], although with some notable differences in activity on xylan/xylo-oligosaccharides, with which *Cv*AA9_A exhibited a much weaker activity compared to *Ls*AA9_A. Furthermore, crystallographic complexes with a range of oligosaccharides could be obtained for *Ls*AA9_A. The crystal structure of *Cv*AA9_A in space group *P*1 was also determined.

However, a catalytically relevant complex structure could not be achieved at that time. This was likely due to a partial occlusion of the active site surface of *Cv*AA9_A in the *P*1 crystal form.

Here we have determined a new structure of *Cv*AA9_A in space group $P2_1$ with a less occluded active site surface. We present crystallographic complexes of both *Ls*AA9_A and *Cv*AA9_A with cello-oligosaccharides in different crystal forms, correlating the structures to the pattern of oligosaccharide binding and cleavage by the two enzymes. Generally, it has proven difficult to obtain complexes of LPMOs with their saccharide substrates. We suggest that this is because LPMOs tend to have flatter, generally more occluded substrate binding sites than glycoside hydrolases. Additionally, the difficulty in forming complexes is due to many LPMOs having low affinity for smaller substrates. Finally, we have evaluated the potential of native differential scanning fluorimetry (DSF) as a guide to crystallographic experiments of LPMOs with ligands. During this work it became apparent that *Ls*AA9_A is considerably more stable than *Cv*AA9_A and the structural basis of this additional stability was analyzed.

MATERIALS AND METHODS

Crystallization, X-ray data collection and refinement

Proteins for crystallization have been produced and deglycosylated as previously[23,24]. Crystals of *Cv*AA9_A and *Ls*AA9_A were grown using the sitting-drop vapor diffusion method. Crystallization drops were set up in MRC two-well plates (Hampton Research) using an Oryx-8 crystallization robot (Douglas Instruments). During both screening and optimization the reservoir volume was kept at 100 µL, while the drops were 0.3 µL during screening and 0.4-0.5 µL for optimization. Drops of *Cv*AA9_A were set up in a protein to reservoir ratio of 3:1 and 1:1 for both screening and optimization. 6.3 mg/ml *Cv*AA9_A in 20 mM sodium acetate was incubated with a slight molar excess of Cu(II)-acetate for 1 hour at 4 °C before being dispensed into the MRC two-well plates. Crystallization conditions for *Cv*AA9_A were established previously[24], but more open crystal forms (with an unoccluded active site) in different conditions (reservoir consisting of 0.1 M NaCl, 0.1 M Bis-Tris pH 6.5 and 1.5 M ammonium sulfate) were additionally identified in the INDEX screen (Hampton Research). These conditions were optimized by varying ammonium sulfate concentrations gradually from 1.0 M to 2.0 M. The final crystals resulting in the presented structures were grown from 1.4 M to 2.0 M ammonium sulfate (see Table 1). In these conditions *Cv*AA9_A produces crystals with a plate-like morphology. They were harvested and flash frozen in liquid N_2 without added cryoprotectant.

*Cv*AA9_A crystals used for soaking were adjusted to pH 5.5 by moving the crystal to similar conditions to the ones they were grown in, but with a different pH, since most of the previous complexes were obtained at pH 5.5. Oligosaccharides (Megazyme) were dissolved in mother-liquor-like conditions. The saturated oligosaccharide solution was gently spun down to remove any residual material before being pipetted into the drop for a final concentration of between 0.1 M and 1.0 M. Crystals were soaked in oligosaccharides for 5-60 minutes before being harvested and flash frozen without any added cryoprotectant.

Crystallization conditions of *Ls*AA9_A have also been established previously[23] as 3-4.4 M NaCl, 0.1 M citric acid pH 3.0-4.0. The crystal used for solving the *Ls*AA9_A-$Cell_4$ structure was grown in 3.3 M NaCl, 0.1 M citric acid pH 3.5. The drops were of 0.5 µl with a protein to reservoir to water ratio of 3:1:1. The crystal was pH adjusted and soaked using the same procedure as described for *Cv*AA9_A above. The crystal was soaked for 10 minutes before cryo-cooling without added cryoprotectant.

All datasets were processed and scaled using XDS and XSCALE[25], with resolution cutoff chosen based on $CC_{1/2}$ close to 50%. Phases for *Cv*AA9_A structures were obtained from molecular replacement, using Molrep[26] and the already published *Cv*AA9_A structure (without substrate, space group *P*1, PDB 5NLT)[24]. Phases for *Ls*AA9_A were obtained from the high resolution *Ls*AA9_A structure (PDB 5ACH)[23]. Restrained refinement was performed in Refmac5[27] of the CCP4 suite[28], with manual rebuilding and validation performed in COOT[29]. Ramachandran statistics and B-factors were calculated using Procheck (Wilson B-factor)[30] and Baverage (Average B-factors) of the CCP4 suite[28]. Figures were prepared in PyMOL[31].

Enzyme reactions and PACE analysis

Enzyme reaction conditions were as described by[24]. Briefly, apo-*Cv*AA9_A and apo-*Ls*AA9_A were pre-incubated for 1 h at 4 °C with stoichiometric amounts of $Cu(II)(NO_3)_2$ immediately before enzyme reactions. AA9 enzyme reactions on cello-oligosaccharides were performed at 20 °C for 4 h in 10 µl reaction volume containing: 5 nmol oligosaccharide, 100 mM ammonium formate pH 6.0, 4 mM ascorbate and 5 pmol *Cv*AA9_A or *Ls*AA9_A. Identical conditions without enzyme were used as the control. Reactions were stopped by addition of three reaction volumes of 96% (v/v) ethanol before precipitation of the undigested substrates at 4 °C, and separation of the reaction products for further analysis.

High concentration gel polysaccharide analysis by carbohydrate electrophoresis (PACE) was performed as described by[24] with some alterations. Briefly, reaction products and oligosaccharide standards (Megazyme) were reductively aminated with 8-aminonaphthalene-1,3,6-trisulfonic acid (ANTS, Invitrogen) and separated by acrylamide gel electrophoresis. To ensure uniform electrophoresis conditions during PACE, small impurities were removed from the gel by pre-running at 1000 V for 1 h immediately before electrophoresis. Electrophoresis was carried out at 10 °C at 1000 V for 220 min using a 192 mM glycine, 25 mM Tris, pH 8.5, running buffer. Gel composition was as follows: resolving gel, 37.5 ml 40% (w/v) acrylamide, 12.5 ml 375 mM Tris-HCl buffer, pH 8.8, 100 µl 10% ammonium persulfate, 50 µl tetramethylethylenediamine (TEMED); stacking gel, 2 ml 40% (w/v) acrylamide, 2.5 ml 375 mM Tris-HCl buffer, pH 8.8, 5.4 ml water, 100 µl 10% ammonium persulfate, 10 µl TEMED. Gel visualization was performed with a G-box (Syngene) equipped with a short pass detection filter (500–600 nm) and long-wave UV tubes (365 nm emission).

Mass spectrometry

Mass spectrometry was performed using MALDI-TOF. A 0.1% (v/v) TFA solution and an acetonitrile solution were mixed in a 2:1 (v/v) ratio and sinapic acid was added to the solution until saturated. The mixture was shortly vortexed and then spun down. The supernatant (synaptic acid matrix solution) and protein sample in solution were sequentially pipetted onto to the target plate (in a final 2:1 ratio) and left for the solvent to evaporate. The TOF experiments were performed in linear mode calibrated for 20-50 kDa molecules.

DSF measurements

The inflection point (T_i) of thermal unfolding was measured using differential scanning fluorimetry on a Tycho-NT6 (NanoTemper). The Tycho-NT6 measures the native fluorescence at 330 and 350 nm, the ratio of which changes as the enzyme unfolds. The measurement is performed from 35-95 °C over 3 minutes (these settings cannot be changed with this instrument). Changes in T_i were used to probe interactions between enzymes and substrates. *Cv*AA9_A and *Ls*AA9_A were incubated with a slight molar excess of Cu(II)-acetate for 1 hour at 4 °C before being mixed with oligosaccharides (Megazyme). The final enzyme concentration for the measurement was 0.2 mg/ml in 20 mM Na-acetate pH 5.5, with an oligosaccharide concentration of 10 mM, with the exception of $Cell_3$, $Cell_4$, Xyl_5 and Mal_5 which were also tested at 100 mM, and Xyl_5 at 1 M. The mixture was incubated for 5 min at RT before being transferred into a capillary and recorded on the instrument. T_i values are determined automatically by the instrument software based on the peak in the first derivative of the unfolding curve. All measurements were carried out at least in triplicates.

Sequence and structure analysis

CLUSTAL was used to determine the sequence identity between *Ls*AA9_A and *Cv*AA9_A[32]. AREAIMOL[33,34] from the CCP4 suite[28] was used to analyze buried surface on ligand complexation, respectively. For the subsite analysis, the analysis was carried out with an intact glucose residue for each subsite.

Crystal contacts were analysed using NCONT within the CCP4 suite[28]. For each LPMO (AA9-AA11 and AA13-AA15) and endo-acting GH5 enzyme one representative of isomorphous structures was chosen (the highest resolution structure without bound substrate). Structures inappropriate for analysis (such as mutants or with severely disordered active sites) were discarded resulting in almost 100 structures of proteins of both bacterial and eukaryotic origin. We searched for intermolecular contacts (*i.e.* between symmetry-related molecules and between different chains within the ASU) within 5 Å, 7.5 Å and 10 Å. The analysis were performed for active site residues (His-brace in LPMOs and catalytic dyad of Glu in GH5s) and aromatic residues near the minus subsites (conserved Tyr in AA9s equivalent of Tyr208 in *Cv*AA9_A) or plus subsites (Tyr or Trp in the L2 loop in AA10s; Tyr54 in *Sm*AA10A). In GH5s two Trp (Trp170 and Trp273 in *Ta*GH5_A) on each side of the active site (Glu133 and Glu240 in *Ta*GH5_A) are involved in substrate interactions near the plus and minus subsites, respectively.

For the analysis of possible thermostabilizing features, hydrogen bonds were identified using HBPLUS[35]. Secondary structure was identified using DSSP from CCP4[36]. Within PyMOL the solvent accessible surface area was calculated using the *get_area* command while *dot_solvent* was set to *on*, *dot_density* was set to 4 and water and ligands was ignored. Ion-pairs were estimated manually within PyMOL using side chain oxygen atoms from Asp/Glu residues as donor and any nitrogen atom from His/Arg/Lys as acceptor within a 4.0 Å distance from each other.

Molecular dynamics simulations were performed on the high resolution structure of *Ls*AA9_A (PDB-ID 5ACH), and on a modified model of *Cv*AA9_A (PDB-ID 5NLT). The *Cv*AA9_A structure (PDB-ID 5NLT) was submitted to Swiss-model[37] with *Ls*AA9_A (PDB-ID 5ACH) as a template, adding residues 224-227 to the C-terminal. Additionally, residues 228-236 were modeled manually in COOT[29] in a similar conformation as in the *Ls*AA9_A model. The modified *Cv*AA9_A model contributed to hydrogen bonding through the added residues, and conformed to favored torsion angles (only outlier being Glu227) which were validated with RAMPAGE[38]. The initial models were stripped of water and ions (including the active-site Cu). Using VMD[39], the proteins were placed at the center of a cubic water box, with 15 nm to the edges in each direction, and the charge of the proteins was balanced with Na^+ ions. The simulations were performed using NAMD-2.12[40] with the CHARMM36[41] force field and run for 10 ns at 400 K in 2 fs time steps. The simulated data was analyzed using VMDs *time line* tool[39]. Snapshots from the MD simulations were made using VMD[39].

RESULTS

A new crystal structure of *Cv*AA9_A with unoccluded active site

The previously published native structure of *Cv*AA9_A (in space group *P*1 with 6 mol/ASU) showed a partially occluded substrate binding surface and thus we looked to determine the *Cv*AA9_A structure in a different crystal form. In the course of optimization of crystallization conditions, several parameters were varied in the composition of the reservoir solution (originally 1.6 M ammonium sulfate, 0.1 M NaCl, 0.1 M HEPES, pH 7.5 as described in[24]). Analysis of X-ray diffraction data suggested that crystal forms were pH-dependent, with lower pHs favoring monoclinic ($P2_1$) or orthorhombic ($P2_12_12$) crystal forms with similar cell dimensions. This trend was however not completely consistent as in at least one occasion (see next section) crystals grown under identical conditions turned out both in space group *P*1 and $P2_12_12$.

With crystals grown in 0.1 M NaCl, 0.1 M Bis-Tris pH 6.5 and 1.5 M ammonium sulfate as reservoir, we managed to solve a *Cv*AA9_A structure in space group $P2_1$ with two molecules pr. ASU, to 2.1 Å resolution. Interestingly, this new *Cv*AA9_A structure exhibits a less occluded active site surface than the original crystals in *P*1. The His-brace motifs are equally free in both crystal forms, but in the *P*1 crystals a surface exposed Tyr208 equivalent to *Ls*AA9_A Tyr203 involved in oligosaccharide binding is occluded by crystal contacts. In the $P2_1$ structure crystal contacts leave Tyr208 completely free in one of the two molecules in the ASU. The $P2_12_12_1$ crystal form also has a relatively unoccluded active site surface: the N-terminal His has its closest crystal contact at more than 13 Å away, and Tyr208 at more than 9 Å away.

Novel oligosaccharide complex structures of *Ls*AA9_A and *Cv*AA9_A

Since the new crystal forms showed a more accessible substrate-binding surface, they were used as starting points for renewed efforts to obtain oligosaccharide complexes. Crystals were grown with 0.1 M NaCl, 0.1 M Bis-Tris pH 6.5 and 1.4-2.0 M ammonium sulfate as reservoir. The crystals were transferred to drops with cello-oligosaccharides ($Cell_4$, $Cell_5$ or $Cell_6$) and soaked for up to 29 minutes. From these we managed to solve *Cv*AA9_A-ligand complexes obtained with $Cell_4$,

$Cell_5$ (in $P2_1$ with four and two mol/ASU, respectively) and $Cell_6$ (in $P2_12_12$ with one mol/ASU) to 2.0-2.8 Å resolution.

In the *Cv*AA9_A-$Cell_4$ complex, $Cell_4$ is bound in subsites -3 to +1 in two of the four molecules of the ASU (Figure 1a). In the third molecule only three glucosyl units (subsites -3 to -1) could be modelled in the electron density. The fourth molecule of the ASU has the active site surface partially blocked by a symmetry related molecule around 4 Å from Tyr208 (likely involved with aromatic stacking interaction, similar to Tyr203 in *Ls*AA9_A[23,24]), and no electron density that could be interpreted as cello-oligosaccharide was present near the active site of this molecule. The *Cv*AA9_A-$Cell_5$ structure was more difficult to interpret with electron density only accommodating three glucose units in one molecule of the ASU (-1 to -3 subsites) and a fourth disordered glucose unit with higher B-factors (+1 subsite) (Figure 1b). In the other molecule of the ASU two glucose units can be clearly seen at the -1 and -2 subsites (Figure S1), with the active site surface being partially occluded directly above Tyr208 from a symmetry related molecule down to a distance of 4.7 Å. In the *Cv*AA9_A-$Cell_6$ complex structure, $Cell_6$ is bound in the -3 to +3 subsites. This structure exhibits the most accessible substrate surface of the *Cv*AA9_A structures determined (Figure 1c). Unlike several of the published complex structures with *Ls*AA9_A, no Cl^- ions could be identified by any anomalous signal, or could confidently be modeled into any difference density for the CvAA9_A structures. Soaking of *Cv*AA9_A crystals with $Cell_3$ or Xyl_5 did not yield any complexes. Addition of Xyl_5 in a 1.0 M concentration dissolved the crystals for soaks longer than 10 min, or abolished diffraction for shorter soaks.

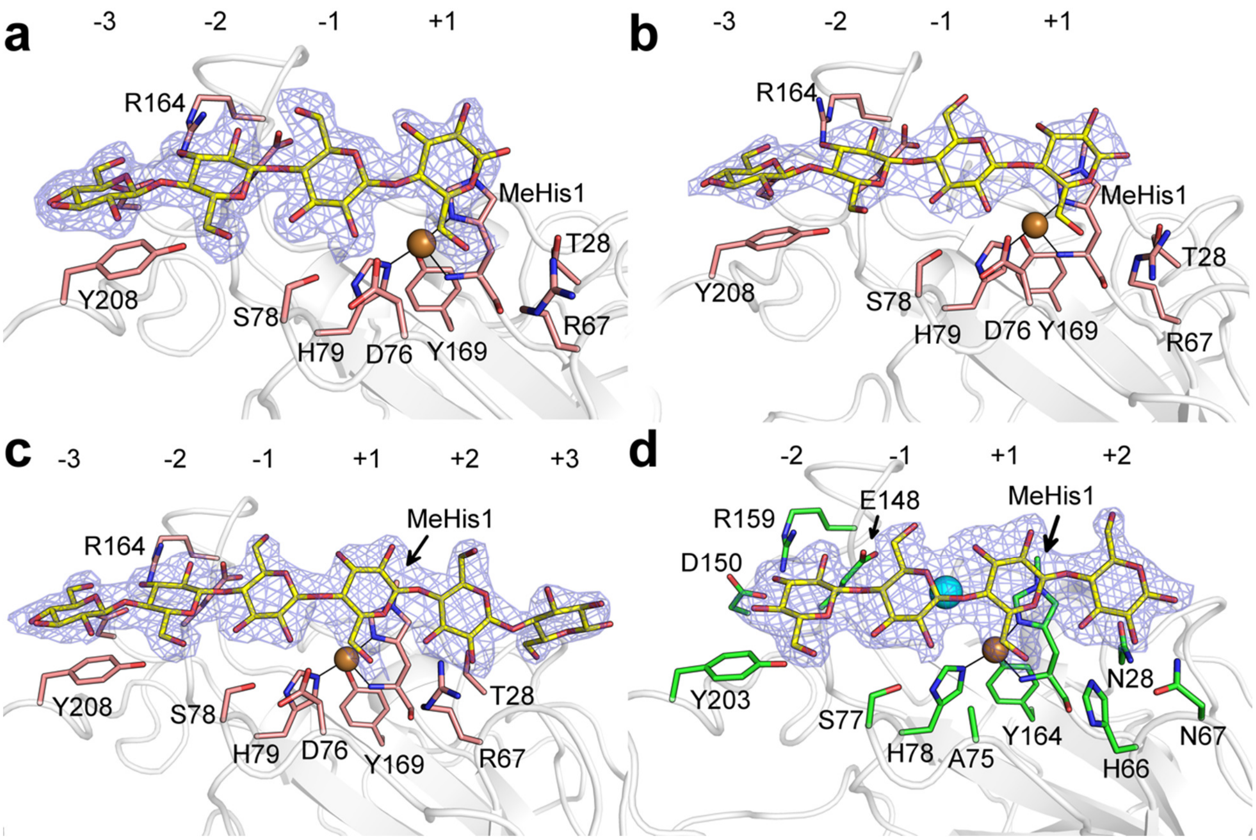


Figure 1. Complex structures of *Cv*AA9_A and *Ls*AA9_A. (a-c) *Cv*AA9_A with Cell$_4$, Cell$_5$ and Cell$_6$ respectively. Only molecule A of *Cv*AA9_A-Cell$_4$ and -Cell$_5$ complexes are shown. Cell$_5$ could only be built with four glucosyl units as no electron density was available to adequately accommodate the remaining glucosyl unit. (d) *Ls*AA9_A with Cell$_4$. 2F$_o$-F$_c$ electron density is shown as blue mesh at 1.0 σ contour level.

Previously, we have reported crystallographic complexes of *Ls*AA9_A with Cell$_3$, Cell$_5$ and Cell$_6$[23,24], which showed the substrates being bound from subsites -1 to +2, -3 to +2 and -4 to +2, respectively. For completeness, we here also determined the structure of *Ls*AA9_A in complex with Cell$_4$ (Figure 1d). In this Cell$_4$ complex the subsites -2 to +2 are occupied. The crystal for the *Ls*AA9_A-Cell$_4$ complex structure (determined in $P4_132$) was formed as described in[23]. For details on crystallization conditions, data collection- and refinement statistics see Table 1.

Table 1. Crystallographic data and refinement statistics.

	*Cv*AA9_A	*Cv*AA9_A :Cell$_4$	*Cv*AA9_A :Cell$_5$	*Cv*AA9_A :Cell$_6$	*Ls*AA9_A :Cell$_4$
Beamline	ID30A-3 (ESRF)	ID30A-3 (ESRF)	ID30B (ESRF)	ID30B (ESRF)	P11 (PETRA III)
Wavelength [Å]	0.96	0.96	0.90	0.90	1.03
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_12_12$	$P4_132$
Cell parameters					
(a, b, c)[Å]	51.12, 39.39, 119.93	39.79, 120.8, 136.36	52.14, 39.70, 120.65	39.55, 124.37, 51.67	125.67
(α,β,γ)[°]	90.0, 93.77, 90.0	90.0, 92.09, 90.0	90.0, 93.491, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution [Å]	50-2.12 (2.18-2.12)	50-2.0 (2.05-2.0)	50-2.80 (2.87-2.80)	50-2.20 (2.26-2.20)	50.0-1.9 (1.95-1.9)
Completeness [%]	94.5 (91.5)	98.4 (97.2)	98.2 (99.1)	96.9 (99.8)	100.0 (100.0)
R_{meas} [%]	12.4 (88.0)	23.2 (115.4)	26.0 (115.4)	27.9 (136.0)	7.7 (183.0)
I/ σ(I)	8.70 (1.58)	5.54 (1.44)	4.23 (1.08)	5.03 (1.28)	20.25 (1.28)
$CC_{1/2}$ [%]	99.5 (50.8)	98.9 (54.4)	96.9 (54.7)	98.2 (62.0)	100.0 (54.7)
Unique reflections	26050 (1853)	85419 (6184)	23359 (1783)	13174 (977)	50564 (3763)
Observed reflections	104767 (7161)	389031 (25746)	68143 (5290)	68277 (5185)	468655 (34891)
Redundancy	4.02 (3.86)	4.55 (4.16)	2.91 (2.96)	5.41 (5.61)	9.26 (9.27)
No. mol./ASU	2	4	2	1	1
R_{Work} [%]	19.73	21.46	19.78	19.40	21.90
R_{Free} [%]	24.94	26.21	26.09	26.36	26.11
Ramachandran[%]§					
Preferred	94.6	95.2	92.1	93.7	92.2
Allowed	5.4	4.8	7.2	6.3	7.3
Outlier	0.0	0.0	0.7	0.0	0.4
Avr. B-factors[Å^2]					
Protein	15.59	11.01	13.01	17.72	25.90
Ligand / ion	78.47	29.74	34.17	50.17	46,45
Water	39.11	31.43	28.42	31.39	45,26
Wilson B-factor[Å^2]	41.18	29.46	53.62	40.33	44.79
No. of atoms					
Protein	3503	7136	3506	1765	1816
Ligand / ion	15	144	68	67	46
Water	241	1461	82	178	197
RMSD					
Bond lengths [Å]	0.014	0.014	0.013	0.014	0.016
Bond Angles [°]	1.56	1.48	1.57	1.69	1.70
Crystallization conditions	0.1 M NaCl 0.1 M HEPES pH 6.5 1.5 M $(NH_4)_2SO_4$	0.1 M NaCl 0.1M Bis-Tris pH 6.5 1.8 M $(NH_4)_2SO_4$	0.1 M NaCl 0.1 M Bis-Tris pH 6.5 2.0 M $(NH_4)_2SO_4$	0.1 M NaCl 0.1 M Bis-Tris pH 6.5 1.4 M $(NH_4)_2SO_4$	3.3 M NaCl 0.1 M citric acid pH 3.5
Drop volume	0.4 µl	0.4 µl	0.4 µl	0.4 µl	0.5
Ratio§§	1:1:0	3:1:0	1:1:0	3:1:0	3:1:1
Soak concentration	-	1.0 M	0.1-0.5 M	0.1-0.5 M	1.0 M
Soak duration [min]	-	12	29	28	10
PDB-ID	6YDF	6YDC	6YDD	6YDE	6YDG

Highest resolution shell shown in parenthesis.

§ Ramachandran preferred/allowed/outlier regions calculated by Rampage[38]

§§ Protein to reservoir to water volume ratio

The modes for cello-oligosaccharides binding to *Ls*AA9_A and *Cv*AA9_A in the crystal structures from this and previous studies[23,24] are summarized in Figure 2. Table S1 shows the values of torsion angles found in the oligosaccharides in the complexes described in this paper, and indicate a conformation close to cellulose.

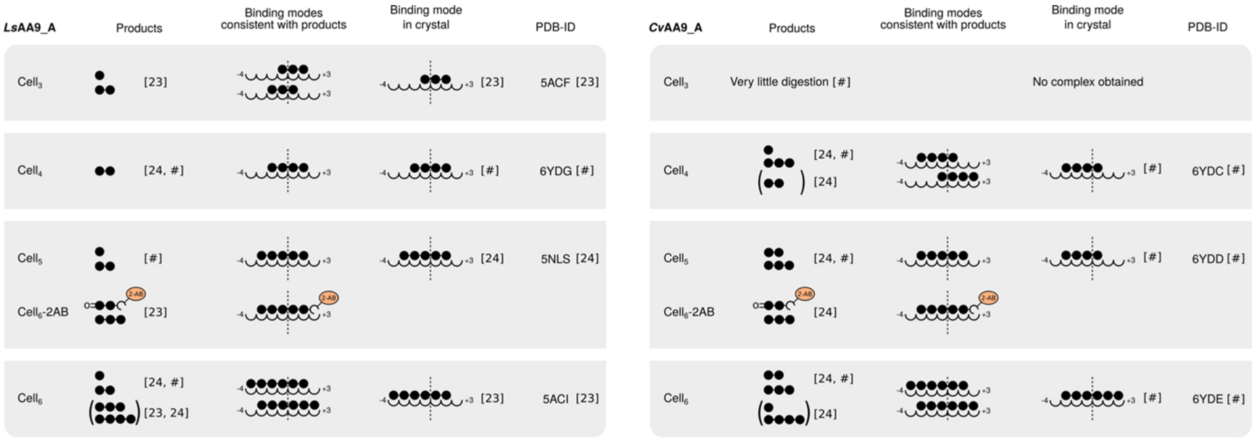


Figure 2. Schematic of cello-oligosaccharide binding in *Ls*AA9_A and *Cv*AA9_A. The active site surface is schematically shown as subsites -4 to +3 (semicircles). Filled circles indicate glucosyl units. For each substrate the predominant products identified in the PACE analysis (here, indicated by # and in previous work as indicated in parentheses and with reference number), and the binding mode found in the corresponding crystal structures (here, indicated by # and/or in previous work as indicated by the reference number) have been illustrated. Binding modes consistent with the product patterns are shown. For Cell$_3$ and Cell$_4$ the binding modes in the crystal can fully account for the solution products. Reaction products on Cell$_6$-2AB (Cell$_5$ analogue) determined by mass spectrometry[23,24] are also shown and suggest a main binding mode of Cell$_5$ for both enzymes from -3 to +2. For *Ls*AA9_A this is consistent with the crystal binding mode. The crystal structure of *Cv*AA9_A with Cell$_5$ has the two molecules pr. ASU shown with only four glucosyl units in molecule A, and two units in molecule B. The products of *Cv*AA9_A with Cell$_5$ digestion agrees with the crystal structure assuming the missing (presumably disordered) glucosyl unit of molecule A is found in subsite +2.

The protein-carbohydrate interactions at individual subsites were analyzed in terms of hydrogen bonds and buried accessible area on the protein using the *Ls*AA9_A and *Cv*AA9_A complexes with Cell$_6$ in Table S2. Total interactions in complexes with different oligosaccharides are summarized in Table S3. Figures S2 and S3 compare complex structures of *Ls*AA9_A and *Cv*AA9_A with Cell$_4$ and Cell$_6$. These figures highlight that residues interacting with the substrates are mostly the same with some notable differences at subsite -1 (Ala75 in *Ls*AA9_A and Asp76 in *Cv*AA9_A) and subsite +2 (Asn28/His66/Asn67 in *Ls*AA9_A and Thr28/Arg67/Val68 in *Cv*AA9_A).

*Ls*AA9_A and *Cv*AA9_A activity on cello-oligosaccharides

We have previously reported some of the patterns of cello-oligosaccharide degradation for *Ls*AA9_A and *Cv*AA9_A[24]. Here, we have further investigated digestion of cello-oligosaccharides (Cell$_3$, Cell$_4$, Cell$_5$ and Cell$_6$) by the two LPMOs, and analyzed the reaction products by Polysaccharide Analysis using Carbohydrate Electrophoresis (PACE). Since these two enzymes are known to oxidize cello-oligosaccharides at the C4 position, all products (oxidized and non-oxidized) are visible by PACE analysis (Figure 3), and by the relative migration of the products it is possible to investigate the site of cleavage. Under the reaction conditions tested, *Cv*AA9_A hardly degraded $Cell_3$, while *Ls*AA9_A readily degraded $Cell_3$ yielding considerable amounts of products migrating as both cellobiose (likely the C4-oxidized 4-ketoaldose form, based on $Cell_3$ binding in *Ls*AA9_A-$Cell_3$ structure, PDB-ID: 5ACF) and glucose.

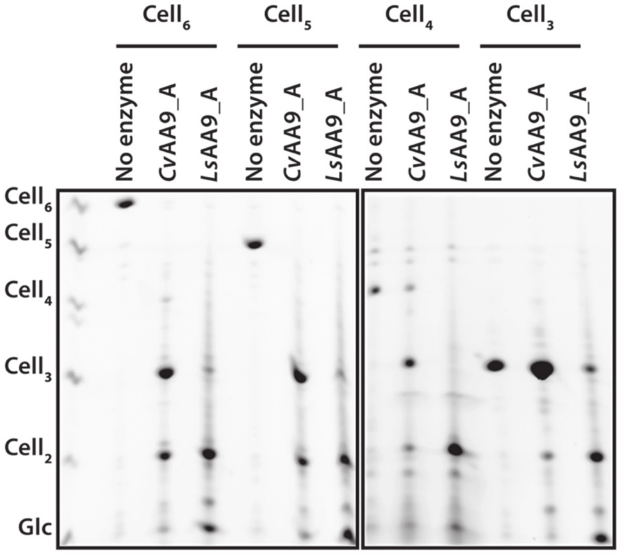


Figure 3. PACE gel showing the digestion products of *Cv*AA9_A and *Ls*AA9_A on $Cell_6$, $Cell_5$, $Cell_4$ and $Cell_3$ oligosaccharides. Enzymatic reactions with the indicated enzyme, 1.0 stochiometric amounts of Cu and excess reducing agent were run for 4 hours. Reaction products were aminated and separated using acrylamide gel electrophoresis. Oxidized and non-oxidized reaction products with the same degree of polymerization are expected to migrate similarly.

The pattern of degradation for $Cell_4$ is very different, with *Ls*AA9_A producing predominantly cellobiose, while *Cv*AA9_A produces mostly $Cell_3$, indicating different binding modes. With $Cell_5$ the products are consistent with what has been identified earlier using mass spectrometry[23,24], where Cell6-2AB was used as a $Cell_5$ analogue and produce $Cell_3$ and $Cell_2$ for both *Ls*AA9_A and *Cv*AA9_A. *Ls*AA9_A is then expected to degrade $Cell_3$ further to $Cell_2$ and glucose, which does not appear to be the case for *Cv*AA9_A. With $Cell_6$ as substrate *Ls*AA9_A produces mostly cellobiose and glucose, while *Cv*AA9_A produced mostly $Cell_3$ and cellobiose. In this case it is difficult to relate this pattern to a particular binding mode, since the initial products are likely to be cleaved further[24], and the binding modes in crystal may be affected by crystal contacts. Product profiles from this and previous work are summarized in Figure 2 together with the crystal binding modes.

Probing oligosaccharide binding in solution by differential scanning calorimetry (DSF)

Given the difficulties in obtaining crystallographic complexes with LPMOs, we were curious to see how DSF experiments in solutions correlate with the crystallographic results obtained for *Ls*AA9_A and *Cv*AA9_A. Thermal curves shifts on addition of ligands (even relatively poor ligands) can indicate binding in solution, as already shown for LPMOs[42], and guide crystallographic complex formation studies. In our case, we used native DSF, based on the differences in intrinsic fluorescence of the enzymes before and after thermal unfolding, in particular following the intrinsic fluorescence ratio (350 nm / 330 nm) and determining the inflection point in the thermal unfolding transition (T_i). For Cu-loaded *Ls*AA9_A the thermal unfolding is easily followed by this method, as the difference in the fluorescence ratio is pronounced between the folded and unfolded state. T_i was thus easily determined as 71.8 ± 1.7 °C (T_i values listed in Table 2 and visualized in Figure S4).

Table 2. Points of thermal inflection (T_i) for *Ls*AA9_A in the presence of oligosaccharides. Inflection temperatures extracted from DSF unfolding curves of *Ls*AA9_A. T_i measurements are listed in °C. Measurements were performed using 0.2 mg/ml LPMO in 20 mM Na-acetate pH 5.5. Values highlighted in bold font are calculated to be significantly different from the LPMO (without substrate) measurement, with a 95% confidence in Student's t-test. In the case of cello-oligosaccharides producing a biphasic unfolding curve, only the first T_i is listed.

		*Ls*AA9_A
Apo-LPMO		69.1 ± 0.1
LPMO		71.8 ± 1.7
+10 mM	Cell$_3$	72.4 ± 1.7
	Cell$_4$	72.3 ± 1.6
	Cell$_5$	**75.2 ± 0.9**
	Cell$_6$	**76.0 ± 0.2**
	Mal$_5$	71.9 ± 1.0
	Man$_6$	73.0 ± 0.3
	G4G3G4G	72.0 ± 0.3
	Xyl$_5$	72.8 ± 0.5
	XXXG	**78.7 ± 0.1**
	Chi$_5$	72.0 ± 0.1
+100 mM	Cell$_3$	72.2 ± 0.37
	Cell$_4$	73.5 ± 0.1
	Xyl$_5$	73.5 ± 0.1
	Mal$_5$	72.4 ± 0.36
+ 1 M	Xyl$_5$	**76.0 ± 0.57**

The addition of 10 mM Cell$_5$, Cell$_6$ or the xyloglucan oligosaccharide XXXG gives a clear thermal shift towards higher T_i, while the other oligosaccharides tested at this concentration do not (Figure 4). However, Cell$_4$ gives a biphasic unfolding curve when added at a concentration of 10 mM, indicating that Cell$_4$ likely binds. None of the other oligosaccharides significantly affected the T_i or shape of the curve (see Figure S5 and S6).

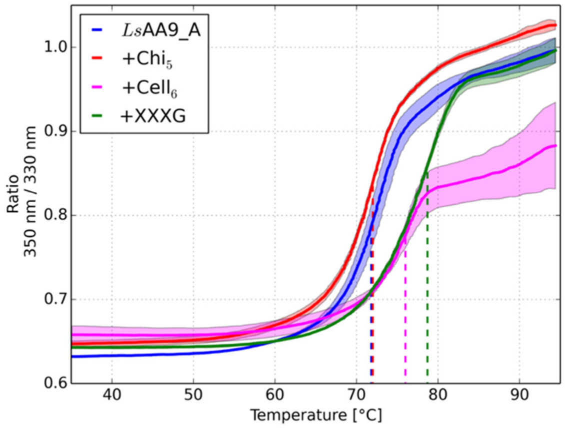


Figure 4. Unfolding curves of *Ls*AA9_A with selected oligosaccharides. Unfolding curves are shown for Cu-loaded *Ls*AA9_A with 10 mM of chitopentaose (Chi$_5$, red, negative control), cellohexaose (Cell$_6$, magenta), and Xyloglucan oligosaccharide (XXXG, green). Transparent color area indicate standard deviation. Dashed lines indicate the inflection point (T$_i$). *Ls*AA9_A alone has similar T$_i$ to *Ls*AA9_A with Chi$_5$, while with Cell$_6$ or XXXG the T$_i$ increases.

Binding of Cell$_3$ and Xyl$_5$ (xylopentaose) has previously been observed in crystal structures with *Ls*AA9_A, where the oligosaccharide concentrations used have been higher than 150 mM, and as high as 0.9 M for Xyl$_5$[23,24]. For Cell$_3$, Cell$_4$ and Xyl$_5$ we chose to repeat the *Ls*AA9_A thermal unfolding experiments with oligosaccharide concentrations closer to the ones used for complex formation (see Figure S7 and S8). At 100 mM concentrations Cell$_4$ still exhibit a biphasic curve with the largest inflection indicating a T$_i$ of 87.7 °C. Using Cell$_3$ at 100 mM concentration the curve is biphasic as well indicating that a population of the enzyme is shifted to higher T$_i$, however the shift is much less pronounced than when using Cell$_4$. Using Xyl$_5$ or Mal$_5$ (maltopentaose, an α-1,4-linked oligosaccharide of glucose, which is not expected to be cleaved by AA9 LPMOs[24]) at 100 mM concentrations the curves are unchanged in appearance and inflection points are consistent with the ones observed without added oligosaccharide. Since Xyl$_5$ is bound in a crystal structure with *Ls*AA9_A where the oligosaccharide concentration was 0.9 M[24] we wanted to approach this concentration for the DSF experiment as well. With 1.0 M of Xyl$_5$ the T$_i$ increased to 76.0 ± 0.57 °C, and the fluorescence ratio increased throughout the measurement (see Figure S7C and S8C). Similar observations were however made using 1-4 M of xylose (Figure S9), indicating that at these high concentrations the thermal shifts may be due to unspecific effects.

A similar study was done with *Cv*AA9_A, but the melting curve is shallower than for *Ls*AA9_A, and this gives challenges in precise determination of T$_i$ which is however clearly much lower than for *Ls*AA9_A, 57.7 ± 0.9 °C in the presence of copper. Small changes in folding curves are seen for some cello-oligosaccharides, but generally the results are less clear (Table S4 and Figure S10 to S13).

Analysis of AA9 family crystal packing and comparison to a glycosyl hydrolase family

We have previously noted a flat and extended active site surface in LPMOs compared to the deeper substrate binding clefts in glycoside hydrolases such as GH5 (Figure S14). Thus, we hypothesized that the aromatic platforms involved in LPMO-substrate interaction would be more frequently involved in crystal contacts compared to substrate-interacting aromatic residues in GH5s, providing a rationale as to why so few LPMO-complex structures have been obtained.

Therefore, we investigated crystal contacts in AA9 and GH5 structures involving key functional residues: the His-brace and aromatic platforms of LPMOs, and the catalytic dyad and substrate interacting Trp residues in GH5s. Indeed, we found that within a 5 Å distance a higher number of intermolecular contacts were observed for LPMOs compared to GH5s. In LPMO structures, an average of 6, 13, 33 and 37 contacts pr. chain were found for the N-terminal His, second His, the conserved AA9 Tyr and AA10 L2 Tyr/Trp, respectively. In contrast, in GH5 structures 8 contacts pr. chain were found on average for the Trp near the negative subsites (Trp273 in *Ta*GH5_A, PDB 1H1N), while the value was less than one for both the Trp near the positive subsite and for the active site Glu catalytic dyad (Acid/Base and Nucleophile) (Figure 5). A similar trend was observed when inspecting the number of contacts found within 7.5 Å and 10 Å, showing that LPMO structures have significantly more crystal contacts involving functional residues compared to GH5 structures (Figure S14d).

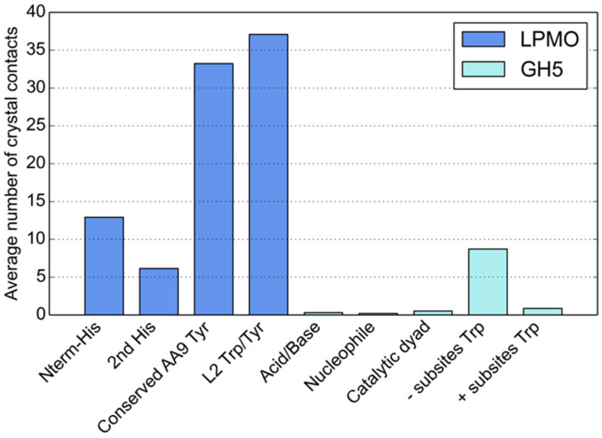


Figure 5. Crystal contacts of substrate interacting residues in LPMO and GH5 structures. Average number of crystal contacts pr. chain observed within 5 Å of the active site and substrate interacting aromatic residues in both LPMO and GH5 structures.

Structural basis of thermostability

In order to identify possible reasons for the significantly higher thermostability of *Ls*AA9_A compared to *Cv*AA9_A a quantification of structural features often correlated with thermostability was carried out, including analysis of Pro/Gly ratio, potential hydrogen bonds and ion pairs, secondary structure and accessible surface composition (Table S5). Although there is a clear difference in terms of *Cv*AA9_A having a more charged surface the other differences seem small and unable to explain the large difference in stability.

One clear difference is the construct length used for structure determination (252 amino acids for *Cv*AA9_A), however the number visible in the structure is only 224. In order to clarify if this is due to disorder in the structure or degradation during production, storage or crystallization we performed MALDI-TOF MS of a *Cv*AA9_A sample in solution (data not shown). The determined weight of 25974.15 Da corresponds to 236 residues suggesting indeed that degradation has occurred and that the preparations of *Cv*AA9_A and *Ls*AA9_A comprise polypeptide chains of similar length (235 residues for *Ls*AA9_A). Despite the similar length of the two enzymes, more residues are visible in the *Ls*AA9_A structure (Figure S15). The C-terminus of *Ls*AA9_A is able to form an additional β-strand which elongates one of the β-sheets of the *Ls*AA9_A β-sandwich. At the end of this additional β-strand residue Ile226 is buried in a small hydrophobic pocket which act as an anchor point for the extended C-terminus (Figure S15A). The equivalent position of Ile226 in *Ls*AA9_A is Cys230 in *Cv*AA9_A (Figure S15C), which is not able to make similar interactions. Furthermore, the hydrophobic pocket is already filled by Trp129 in *Cv*AA9_A. Thus, the much lower thermostability of the *Cv*AA9_A construct studied could be due to a more disordered C-terminus.

To investigate this hypothesis further, the missing residues of *Cv*AA9_A (225-236) were modeled through a combination of Swiss-model[37] (residues 224-227) and manual modeling (residue 228-236) using *Ls*AA9_A (PDB 5ACH) as a template (Figure S15C). 10 ns molecular dynamics simulations were performed for both *Ls*AA9_A and *Cv*AA9_A at 400 K (Figure S16 and S17). The root mean square fluctuations (RMSF) over the whole simulation indicates that the extended C-terminal of *Cv*AA9_A is more disordered than the equivalent segment of *Ls*AA9_A. Over the course of the simulation the β-strand modelled in *Cv*AA9_A (residues 226-230) becomes untethered, unlike the equivalent amino acids in *Ls*AA9_A (residues 222-226). Additionally, the surface L7 and LC loops (*Ls*AA9_A naming) are destabilized in *Cv*AA9_A as seen in the larger RMSF for residues 150-166/196-216, compared to the equivalent regions 145-161/195-212 in *Ls*AA9_A (see Figure S16 and S17).

DISCUSSION

As indicated now by a number of other studies, the clear difference in both binding pattern and cleavage pattern of *e.g.* Cell$_4$ by *Cv*AA9_A and *Ls*AA9_A highlight that LPMOs primarily cleave through binding to substrate and bringing of substrate and reactive oxygen species precisely together, and not merely by generation of reactive species. Binding of substrate is a prerequisite for efficient use of LPMOs – such precise and appropriate binding of substrate is in fact necessary for preventing untimely inactivation of LPMOs[43–45]. In this respect we have previously reported that binding of *Ls*AA9_A to xylo-oligosaccharides may be less ‘appropriate’ than for cello-oligosaccharides[24], in the sense that the binding mode is different, preventing a similar geometry of interaction with the oxygen species, and probably resulting in diminished protection from oxidative damage.

From this study we can extract some general features of oligosaccharide binding to LPMOs, with on average 97 Å^2 accessible area buried for an intact glucose unit and 1.5 H-bonds per subsite. Generally cello-oligosaccharides bind in a cellulose-like conformation. The analysis of interactions reflects well substrate preference in cleavage studies, for example we see both lower buried accessible area and fewer hydrogen bonds with Xyl$_5$ than Cell$_5$ for *Ls*AA9_A. The symmetrical cleavage of Cell$_4$ by *Ls*AA9_A compared to the asymmetrical cleavage by *Cv*AA9_A correlates well with extra hydrogen bonds interactions of *Ls*AA9_A at +2 and *Cv*AA9_A's extra interactions at -3.

Generally speaking, there is for *Ls*AA9_A an overall good correlation between the degraded substrates, oligosaccharides that could bind in the crystals, and oligosaccharides that gave a thermal shift or otherwise clear changes in the DSF analysis. For very soluble substrates in the DSF analysis it is possible to use very high concentrations, as in crystal soaks, so that even weakly binding fragments, such as xylo-oligosaccharides, can be identified. However, at molar concentrations non-specific interactions will also be reflected in the thermal shifts, as was found in measurements with >1 M xylose (Figure S9). DSF (in that case using a fluorescence dye) was first utilized in LPMO research for an in depth study of *Nc*AA9_C[42] where the effect of several additives on stability was investigated, including soluble polymeric substrates, but not oligosaccharides. In two recent publications, we have attempted to use native DSF with an LPMO and an LPMO-like protein to probe oligosaccharide binding, but with somewhat inconclusive results[46,47]. Here we establish that DSF, native and presumably also using fluorescent dyes, can be useful to pre-screen for useful binding fragments for crystallographic studies. However, as shown for *Cv*AA9_A, this is mostly useful where a clear inflection point can be shown in the melting curve.

One exception to the correlation between crystallographic results and DSF is with xyloglucan oligosaccharides. As previously reported[24], no crystallographic complexes could be obtained, though xyloglucan is degraded by *Ls*AA9_A and other AA9 LPMOs acting on soluble substrates, where direct interactions with the substrate have been clearly shown by NMR[48]. Thermal shift analysis confirm that XXXG is bound by *Ls*AA9_A, thus as discussed before, this oligosaccharide is probably too big to fit between crystal contacts in the crystal form used in this and previous studies.

The synergistic binding of chloride ions and oligosaccharides has been previously proposed[23], and chloride ions are often found in complex structures with *Ls*AA9_A. In the presented complex structures of *Cv*AA9_A no chloride ions could be identified. This is likely due to the much lower Cl^- concentration in conditions favoring *Cv*AA9_A crystal growth (100 mM compared to 3-4 M with *Ls*AA9_A).

In general, complex formation with LPMOs in crystals, just like for other enzymes, requires an accessible active site surface. Because LPMOs substrate binding sites are 'flatter' than for most other enzymes, we suggest that this is particularly problematic for crystallographic studies of LPMOs. Indeed, for the enzyme sets investigated here, LPMOs active sites in crystals are more often occluded by symmetry-related molecules compared to GH5 hydrolases.

In the course of this investigation we observed that *Cv*AA9_A is much less thermostable than *Ls*AA9_A. Thermostability is an important feature for industrial enzymes[49] including AA9 family LPMOs, and has previously been investigated for *Af*AA9_B[50], which had unfolding temperature in similar range as *Ls*AA9_A. Protein engineering of *Af*AA9_B generated variants with a 6-7 °C increase in melting temperature and improved industrial performance[50], probably due to a combination of improved electrostatic and packing interactions. These variants were inspired by the structure of *Ta*AA9_A, which is unique in terms of its exceptional thermostable nature, and a T_i of up to 89.5 °C has recently been demonstrated using DSF[51]. The engineered region of *Af*AA9_B is already similar to the equivalent regions in both the structures of *Cv*AA9_A and *Ls*AA9_A, so for these two enzymes further improvement of thermostability cannot be expected by applying the same *Ta*AA9_A inspired mutations.

In contrast, most other AA9 LPMOs where this has been investigated had lower melting temperatures[42,52]. In this case, comparison of the structures of *Ls*AA9_A and *Cv*AA9_A pointed to improved interactions of the C-terminal portion of *Ls*AA9_A with the rest of the structure, compared to *Cv*AA9_A, may be the main sources of additional stability. Molecular dynamics simulations at 400 K, while not necessarily biologically relevant per se, illustrate clearly the more flexible nature of the C-terminus in *Cv*AA9_A compared to *Ls*AA9_A already indicated by proteolysis during crystallization and lack of electron density for several C-terminal residues. In the hyperstable *Ta*AA9_A, the very C-terminal residue Gly228 corresponds to *Ls*AA9_A Gly220, preceding the flexible C-terminus. Our findings are thus is in line with previous work[52,53], highlighting the importance of C-terminal and linker regions in AA9 for their stability.

In summary, our further work on *Cv*AA9_A and *Ls*AA9_A, two AA9 LPMOs sharing 41.5% sequence identity, provides new details on similarities and differences. Despite their sequence and structural similarity, *Ls*AA9_A is considerably more thermostable. We proposed that the lower thermostability of *Cv*AA9_A is caused by a more flexible C-terminus, which is also prone to proteolysis. We further investigated interactions with oligosaccharide substrates and found native DSF to be a promising technique to guide crystallographic complex structure determination of LPMOs in the future, which however is likely to continue to be hampered by the involvement of the shallow active site of LPMOs in crystal contacts. We show that *Cv*AA9_A, only the second LPMO for which crystallographic information on substrate binding has been obtained, binds oligosaccharides similarly to *Ls*AA9_A overall, but with subtle differences mediated particularly by structural differences at the +2 subsites. These differences are reflected in an asymmetric pattern of $Cell_4$ degradation by *Cv*AA9_A, while *Ls*AA9_A produces only DP2 products.

ACCESSION CODES

Atomic coordinates and structure factors for the reported crystal structures were deposited in the Protein Data Bank under accession codes: 6YDC (*Cv*AA9_A-Cell$_4$), 6YDD (*Cv*AA9_A-Cell$_5$), 6YDE (*Cv*AA9_A-Cell$_6$), 6YDF (*Cv*AA9_A (without substrate), 6YDG (*Ls*AA9_A-Cell$_4$). UniprotKB Protein accession codes are *Cv*AA9_A: A0A223GEC9; *Ls*AA9_A: A0A0S2GKZ1.

AUTHOR INFORMATION

Corresponding Author

*Leila Lo Leggio (leila@chem.ku.dk)

Funding Sources

L.L.L., T.Tandrup., K.E.H.F. and K.S.J's research on LPMOs are funded by the Novo Nordisk Foundation HOPE project (NNF17SA0027704). K.E.H.F. is additionally funded by a Carlsberg foundation Postdoc Fellowship (CF-16-0673 and CF17-0533) and a Marie-Curie FP7 COFUND People Program, AgreenSkills+ fellowship (under grant agreement n°609398). L.L.L., T.Tandrup. and K.E.H.F. are members of ISBUC, Integrative Structural Biology at the University of Copenhagen (www.isbuc.ku.dk). Travel to synchrotrons was supported by the Danish Ministry of Higher Education and Science through the Instrument Center DANSCATT.

Conflict of interest

The authors declare that they have no conflicts of interest with the contents of this article.

ACKNOWLEDGEMENT

We acknowledge the previous CESBIC consortium, which included the University of Cambridge, the University of York, the CNRS Marseille, Novozymes A/S and the University of Copenhagen for providing the initial basis for this study.

We thank Novozymes A/S for providing protein samples.

We acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the provision of experimental facilities. Parts of this research were carried out at PETRA III and we would like to thank Saravanan Panneerselvam for assistance in using P11. Parts of this research were carried out at the European Synchrotron Radiation Facility (ESRF), Grenoble, France. We are grateful for the assistance in using beamline ID30A-3 and ID30B respectively.

Molecular dynamics simulations were performed at the Danish National Supercomputer for Life Sciences *Computerome*, installed at the National Life Science Center at Technical University of Denmark.

ABBREVIATIONS

AA, Auxiliary Activity; *Af*AA9_B, isoform B of Auxiliary Activity family 9 LPMO from *Aspergillus fumigatus*; ASU, Asymmetric Unit; CAZy, Carbohydrate Active enzyme; $Cell_3$, cellotriose oligosaccharide; $Cell_4$, cellotetraose oligosaccharide; $Cell_5$, cellopentaose oligosaccharide; $Cell_6$, cellohexaose oligosaccharide; Chi_5, chitopentaose oligosaccharide; *Cv*AA9_A, isoform A of Auxiliary Activity family 9 LPMO from *Collariella virecens*; DSF, native Differential Scanning Fluorimetry; G4G3G4G, Glu-1,4-Glu-1,3-Glu-1,4-Glu; GH5, Glycoside Hydrolase family 5; LPMO, Lytic Polysaccharide Monooxygenase; *Ls*AA9_A, isoform A of Auxiliary Activity family 9 LPMO from *Lentinus similis*; Mal_5, maltopentaose oligosaccharide; Man_6, mannohexaose oligosaccharide; *Nc*AA9_A, isoform A of Auxiliary Activity family 9 LPMO from *Neurospora crassa*; NCS, Non-Crystallographic Symmetry; PACE, polysaccharide analysis by carbohydrate electrophoresis; RMSF, root mean square fluctuations; *Ta*AA9_A, isoform A of Auxiliary Activity family 9 LPMO from *Thermoascus aurantiacus;* TEMED, tetramethylethylenediamine; T_i, point of thermal inflection; XXXG, xyloglucan oligosaccharide, cellotetraose backbone with xylose decorations on three consecutive glucose units from the non-reducing end, see nomenclature[54]; Xyl_5, xylopentaose oligosaccharide.

SUPPORTING INFORMATION

Tables S1 – S5

Figures S1 – S17

REFERENCES

(1) 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. *Proc. Natl. Acad. Sci. 108*, 15079–15084.

(2) Vaaje-Kolstad, G., Westereng, B., Horn, S. J., Liu, Z., Zhai, H., Sørlie, M., and Eijsink, V. G. H. (2010) An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science 330*, 219–222.

(3) Johansen, K. S. (2016) Lytic Polysaccharide Monooxygenases: The Microbial Power Tool for Lignocellulose Degradation. *Trends Plant Sci. 21*, 926–936.

(4) Cannella, D., Hsieh, C. W. C., Felby, C., and Jørgensen, H. (2012) Production and effect of aldonic acids during enzymatic hydrolysis of lignocellulose at high dry matter content. *Biotechnol. Biofuels 5*, 1–10.

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

(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*, No. 41, 1-14.

(7) Filiatrault-Chastel, C., Navarro, D., Haon, M., Grisel, S., Herpoël-Gimbert, I., Chevret, D., Fanuel, M., Henrissat, B., Heiss-Blanquet, S., Margeot, A., and Berrin, J.-G. (2019) AA16, a new lytic polysaccharide monooxygenase family identified in fungal secretomes. *Biotechnol. Biofuels 12*, 1–15.

(8) Vaaje-Kolstad, G., Horn, S. J., Sørlie, M., and Eijsink, V. G. H. (2013) The chitinolytic machinery of Serratia marcescens--a model system for enzymatic degradation of recalcitrant polysaccharides. *FEBS J. 280*, 3028–3049.

(9) Vu, V. V., Beeson, W. T., Span, E. A., Farquhar, E. R., and Marletta, M. A. (2014) A family of starch-active polysaccharide monooxygenases. *Proc. Natl. Acad. Sci. 111*, 13822–13827.

(10) Couturier, M., Ladevèze, S., Sulzenbacher, G., Ciano, L., Fanuel, M., Moreau, C., Villares, A., Cathala, B., Chaspoul, F., Frandsen, K. E., Labourel, A., Herpoël-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. *Nat. Chem. Biol. 14*, 306–310.

(11) 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. *Nat. Commun. 9*, No. 756, 1-12.

(12) 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 249*, 1987–1996.

(13) Hüttner, S., Várnai, A., Petrović, D. M., Bach, C. X., Kim Anh, D. T., Thanh, V. N., Eijsink, V. G. H., Larsbrink, J., and Olsson, L. (2019) Specific Xylan Activity Revealed for AA9 Lytic Polysaccharide Monooxygenases of the Thermophilic Fungus Malbranchea cinnamomea by Functional Characterization. *Appl. Environ. Microbiol.* (Master, E. R., Ed.) *85*, 1–13.

(14) Westereng, B., Cannella, D., Wittrup Agger, J., Jørgensen, H., Larsen Andersen, M., Eijsink, V. G. H., and Felby, C. (2015) Enzymatic cellulose oxidation is linked to lignin by long-range electron transfer. *Sci. Rep. 5*, No. 18561, 1-9.

(15) Li, X., Beeson IV, W. T., Phillips, C. M., Marletta, M. A., and Cate, J. H. D. (2012) Structural basis for substrate targeting and catalysis by fungal polysaccharide monooxygenases. *Structure 20*, 1051–1061.

(16) Kracher, D., Scheiblbrandner, S., Felice, A. K. G., Breslmayr, E., Preims, M., Ludwicka, K., Haltrich, D., Eijsink, V. G. H., and Ludwig, R. (2016) Extracellular electron transfer systems fuel cellulose oxidative degradation. *Science 352*, 1098–1101.

(17) Bissaro, B., Røhr, Å. K., Müller, 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. *Nat. Chem. Biol. 13*, 1123–1128.

(18) Chylenski, P., Bissaro, B., Sørlie, M., Røhr, Å. K., Várnai, A., Horn, S. J., and Eijsink, V. G. H. (2019) Lytic Polysaccharide Monooxygenases in Enzymatic Processing of Lignocellulosic Biomass. *ACS Catal. 9*, 4970–4991.

(19) Hangasky, J. A., Detomasi, T. C., and Marletta, M. A. (2019) Glycosidic Bond Hydroxylation by Polysaccharide Monooxygenases. *Trends Chem. 1*, 198–209.

(20) Forsberg, Z., Sørlie, M., Petrović, D., Courtade, G., Aachmann, F. L., Vaaje-Kolstad, G., Bissaro, B., Røhr, Å. K., and Eijsink, V. G. (2019) Polysaccharide degradation by lytic polysaccharide monooxygenases. *Curr. Opin. Struct. Biol. 59*, 54–64.

(21) Tandrup, T., Frandsen, K. E. H., Johansen, K. S., Berrin, J., and Lo Leggio, L. (2018) Recent insights into lytic polysaccharide monooxygenases (LPMOs). *Biochem. Soc. Trans. 46*, 1431–1447.

(22) Vaaje-Kolstad, G., Forsberg, Z., Loose, J. S., Bissaro, B., and Eijsink, V. G. (2017) Structural diversity of lytic polysaccharide monooxygenases. *Curr. Opin. Struct. Biol. 44*, 67–76.

(23) 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., and Walton, P. H. (2016) The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nat. Chem. Biol. 12*, 298–303.

(24) 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. *Nat. Commun. 8*, No. 1064, 1-12.

(25) Kabsch, W. (2010) XDS. *Acta Crystallogr. Sect. D Biol. Crystallogr. 66*, 125–132.

(26) Vagin, A., and Teplyakov, A. (1997) MOLREP : an Automated Program for Molecular Replacement. *J. Appl. Crystallogr. 30*, 1022–1025.

(27) Murshudov, G., A.Vagin, and E.Dodson. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood method of Maximum Likelihood Refinement. *Acta Cryst. D53*, 240–255.

(28) 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. W., 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 CCP 4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr. 67*, 235–242.

(29) Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. Sect. D - Biol. Crystallogr. 60*, 2126–2132.

(30) Vaguine, A. A., Richelle, J., and Wodak, S. J. (1999) SFCHECK : a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallogr. Sect. D Biol. Crystallogr. 55*, 191–205.

(31) Schrodinger, L. L. C. (2018) The PyMOL Molecular Graphics System, Version 2.0.1. Schrodinger, LLC.

(32) Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., Thompson, J. D., and Higgins, D. G. (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol. 7*, 539–539.

(33) Saff, E. B., and Kuijlaars, A. B. J. (1997) Distributing many points on a sphere. *Math. Intell. 19*, 5–11.

(34) Lee, B., and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility. *J. Mol. Biol. 55*, 379–400.

(35) McDonald, I. K., and Thornton, J. M. (1994) Satisfying Hydrogen Bonding Potential in Proteins. *J. Mol. Biol. 238*, 777–793.

(36) Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–2637.

(37) Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., and Schwede, T. (2018) SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res. 46*, W296–W303.

(38) Lovell, S. C., Davis, I. W., Arendall, W. B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S., and Richardson, D. C. (2003) Structure validation by Cα geometry: ϕ,ψ and Cβ deviation. *Proteins Struct. Funct. Bioinforma. 50*, 437–450.

(39) Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph. 14*, 33–38.

(40) Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kalé, L., and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 1781–1802.

(41) Huang, J., and MacKerell, A. D. (2013) CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem. 34*, 2135–2145.

(42) Kracher, D., Andlar, M., Furtmüller, P. G., and Ludwig, R. (2018) Active-site copper reduction promotes substrate binding of fungal lytic polysaccharide monooxygenase and reduces stability. *J. Biol. Chem. 293*, 1676–1687.

(43) Loose, J. S. M., Arntzen, M. Ø., Bissaro, B., Ludwig, R., Eijsink, V. G. H., and Vaaje-Kolstad, G. (2018) Multipoint Precision Binding of Substrate Protects Lytic Polysaccharide Monooxygenases from Self-Destructive Off-Pathway Processes. *Biochemistry 57*, 4114–4124.

(44) Kuusk, S., Bissaro, B., Kuusk, P., Forsberg, Z., Eijsink, V. G. H. H., Sørlie, M., and Väljamäe, P. (2018) Kinetics of H_2O_2 -driven degradation of chitin by a bacterial lytic polysaccharide monooxygenase. *J. Biol. Chem. 293*, 523–531.

(45) Eijsink, V. G. H., Petrovic, D., Forsberg, Z., Mekasha, S., Røhr, Å. K., Várnai, A., Bissaro, B., and Vaaje-Kolstad, G. (2019) On the functional characterization of lytic polysaccharide monooxygenases (LPMOs). *Biotechnol. Biofuels 12*, No. 58, 1-16.

(46) Frandsen, K. E. H., Tovborg, M., Jørgensen, C. I., Spodsberg, N., Rosso, M. N., Hemsworth, G. R., Garman, E. F., Grime, G. W., Poulsen, J. C. N., Batth, T. S., Miyauchi, S., Lipzen, A., Daum, C., Grigoriev, I. V., Johansen, K. S., Henrissat, B., Berrin, J. G., and Leggio, L. Lo. (2019) Insights into an unusual Auxiliary Activity 9 family member lacking the histidine brace motif of lytic polysaccharide monooxygenases. *J. Biol. Chem. 294*, 17117–17130.

(47) Muderspach, S. J., Tandrup, T., Frandsen, K. E. H., Santoni, G., Poulsen, J. N., and Leggio, L. Lo. (2019) Further structural studies of the lytic polysaccharide monooxygenase AoAA13 belonging to the starch-active AA13 family. *Amylase 3*, 41–54.

(48) Courtade, G., Wimmer, R., Røhr, Å. K., Preims, M., Felice, A. K. G., Dimarogona, M., Vaaje-Kolstad, G., Sørlie, M., Sandgren, M., Ludwig, R., Eijsink, V. G. H., and Aachmann, F. L. (2016) Interactions of a fungal lytic polysaccharide monooxygenase with β-glucan substrates and cellobiose dehydrogenase. *Proc. Natl. Acad. Sci. 113*, 5922–5927.

(49) Karnaouri, A., Antonopoulou, I., Zerva, A., Dimarogona, M., Topakas, E., Rova, U., and Christakopoulos, P. (2019) Thermophilic enzyme systems for efficient conversion of lignocellulose to valuable products: Structural insights and future perspectives for esterases and oxidative catalysts. *Bioresour. Technol. 279*, 362–372.

(50) Lo Leggio, L., Weihe, C. D., Poulsen, J.-C. 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. *Carbohydr. Res. 469*, 55–59.

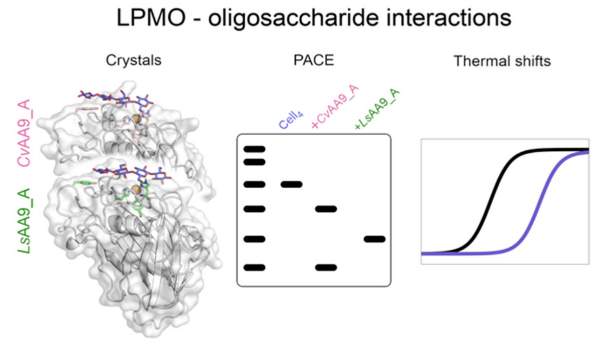
(51) Singh, R. K., Blossom, B. M., Russo, D. A., van Oort, B., Croce, R., Jensen, P. E., Felby, C., and Bjerrum, M. J. (2019) Thermal unfolding and refolding of a lytic polysaccharide monooxygenase from Thermoascus aurantiacus. *RSC Adv. 9*, 29734–29742.

(52) Zhang, R., Liu, Y., Zhang, Y., Feng, D., Hou, S., Guo, W., Niu, K., Jiang, Y., Han, L., Sindhu, L., and Fang, X. (2019) Identification of a thermostable fungal lytic polysaccharide monooxygenase and evaluation of its effect on lignocellulosic degradation. *Appl. Microbiol. Biotechnol. 103*, 5739–5750.

(53) 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., and Sandgren, M. (2017) High-resolution structure of a lytic polysaccharide monooxygenase from Hypocrea jecorina reveals a predicted linker as an integral part of the catalytic domain. *J. Biol. Chem. 292*, 19099–19109.

(54) Fauré, R., Courtin, C. M., Delcour, J. A., Dumon, C., Faulds, C. B., Fincher, G. B., Fort, S., Fry, S. C., Halila, S., Kabel, M. A., Pouvreau, L., Quemener, B., Rivet, A., Saulnier, L., Schols, H. A., Driguez, H., and O'Donohue, M. J. (2009) A Brief and Informationally Rich Naming System for Oligosaccharide Motifs of Heteroxylans Found in Plant Cell Walls. *Aust. J. Chem. 62*, 533–537.

FOR TABLE OF CONTENTS ONLY



Supporting Information:

Oligosaccharide binding and thermostability of two related AA9 Lytic Polysaccharide Monooxygenases

*Tobias Tandrup[1], Theodora Tryfona[2], Kristian Erik Høpfner Frandsen[1,3], Katja Salomon Johansen[4], Paul Dupree[2], and Leila Lo Leggio[1]**

[1]Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100-DK, Copenhagen, Denmark. [2]Department of Biochemistry, University of Cambridge, Cambridge, CB2 1QW, UK. [3]INRAE, Aix-Marseille Université, Biodiversité et Biotechnologie Fongiques (BBF), Marseille, France. [4]Department for Geosciences and Natural resource management, University of Copenhagen, 1958-DK, Frederiksberg, Denmark.

Corresponding Author: *Leila Lo Leggio (leila@chem.ku.dk)

SUPPLEMENTARY TABLES

Glycosidic torsion angles		***Cv*AA9_A**	***Cv*AA9_A**	***Cv*AA9_A**	***Ls*AA9_A**
		Cell$_4$	**Cell$_5$**	**Cell$_6$**	**Cell$_4$**
Subsite +2/+3 (°)	Φ	-	-	-74.56	
	Ψ	-	-	128.19	
Subsite +1/+2 (°)	Φ	-	-	-85.48	-91.65
	Ψ	-	-	87.27	88.43
Subsite -1/+1 (°)	Φ	(A) -82.40 (B) -97.39	-97.14	-101.22	-90.68
	Ψ	(A) 94.37 (B) 89.14	99.99	105.39	96.99
Subsite -1/-2 (°)	Φ	(A) -75.79 (B) -70.03 (C) -66.37	-82.32 (-87.35)	-74.50	-78.75
	Ψ	(A) 96.16 (B) 105.57 (C) 103.77	100.83 (98.75)	98.34	101.20
Subsite -2/-3 (°)	Φ	(A) -78.86 (B) -88.05 (C) -89.51	-76.64	-88.54	-
	Ψ	(A) 103.27 (B) 98.28 (C) 93.93	101.51	99.37	-
Values from model (°)	φ	-98.75			
	ψ	90.91			
Definitions	φ	$O_{5'}$-$C_{1'}$-O_4-C_4			
	ψ	$C_{1'}$-O_4-C_4-C_3			

Table S1. Torsion angles in bound cellooligosaccharides of presented complexes. Model values from cellulose microfibril in http://polysac3db.cermav.cnrs.fr[1].

Subsite	**-4**	**-3**	**-2**	**-1**	**+1**	**+2**	**+3**
*Cv*AA9_A H-bonds*	NA	1	2	1	(2)	2(1)	1(1)
		Asp155	Glu153, Arg164	Asp76, Ser78		Thr28, Arg67	Arg3 (+crystal contacts)
*Ls*AA9_A H-bonds*	0	0	3	1	(1)	3(1)	NA
	(+ crystal contacts)		Glu148, Arg159	Ser77		Asn28, His66, Asn67	
*Cv*AA9_A buried area §	NA	84.2 Å^2	114.9 Å^2	103.6 Å^2	103.5 Å^2	99.0 Å^2	101.1 Å^2
*Ls*AA9_A buried area §	66.8 Å^2	72.0 Å^2	118.7 Å^2	109.1 Å^2	106.9 Å^2	87.2 Å^2	NA

* number of direct H-bonds between ligand and protein within 3.4 Å cutoff, identified in PyMOL through the ligand site preset. Numbers in parenthesis are the number of H-bonds between ligand and protein mediated by a single water molecule. Residues given are the ones interacting directly with the ligand. Interactions arising from crystal contacts are not considered.

§ Area of buried accessible surface on complexation. The calculation is carried out at each subsite with an intact glucosyl unit (thus the sum of all subsites is larger than the surface area buried by the oligosaccharide)

Table S2. Analysis of individual subsites interactions based on Cell$_6$ complexes with *Ls*AA9_A (PDB 5ACI, [2]) and *Cv*AA9_A (this manuscript).

Ligand	**Cell$_3$**	**Cell$_4$**	**Cell$_5$**	**Cell$_6$**	**Xyl$_5$**
*Ls*AA9_A					
PDB code	5ACF	6YDG	5NLS	5ACI	5NLN
Buried area (Å^2)	222	286	376	386	313
Total number of H-bonds	8	9	10	9	7
*Cv*AA9_A					
PDB code	-	6YDC	6YDD*	6YDE	-
Buried area in Å^2		271/298[#]		421	
Total number of H-bonds		8		11	

Table S3. Analysis of total buried surface for different complexes. # Buried protein surface is given for chains A and B (where Cell$_4$ binding is clearly visible) * not calculated since only 2 and 4 glucosyl units are clearly bound to different protein chains

		*Cv*AA9_A
LPMO		57.7 ± 0.9
+10 mM	Cell$_3$	58.4 ± 0.8
	Cell$_4$	59.1 ± 2.2
	Cell$_5$	59.5 ± 1.2
	Cell$_6$	59.4 ± 0.2
	Mal$_5$	58.4 ± 0.1
	Man$_6$	58.4 ± 0.2
	G4G3G4G	57.6 ± 0.6
	Xyl$_5$	58.6 ± 0.4
	XXXG	58.0 ± 0.1
	Chi$_5$	58.1 ± 0.5
+100 mM	Cell$_3$	59.2 ± 0.1
	Cell$_4$	61.4 ± 0.2
	Xyl$_5$	58.6 ± 0.2
	Mal$_5$	58.2 ± 0.4

Table S4. Points of thermal inflection (T_i) for *Cv*AA9_A in the absence and presence of oligosaccharides. Inflection temperatures extracted from DSF unfolding curves *Cv*AA9_A. T_i measurements are listed in °C. Measurements were performed using 0.2 mg/ml LPMO in 20 mM Na-acetate pH 5.5.

LPMO	*Ls*AA9_A	*Cv*AA9_A
Sequence length	235	236*
Sequence modelled in structure	1-235	1-224 (1-227 in some chains in 5NLT)
Charged residues [%]		
(Arg, Asp, Glu, His, MeHis, Lys)	16.6	25.0
Polar residues [%]		
(Asn, Cys, Gln, Gly, Ser, Thr, Tyr)	38.7	33.3
Hydrophobic residues [%]		
(Ala, Ile, Leu, Met, Phe, Pro, Trp, Val)	44.7	41.7
Aromatic residues [%]		
(His, MeHis, Phe, Trp, Tyr)	12.8	13.1
Number of Pro / Gly residues	20 / 18	20 / 20
Accessible surface area [Å^2]		
Total	9890	10217
Charged residues	2381 (24%)	4423 (43%)
Polar residues	4375 (44%)	2776 (27%)
Hydrophobic residues	3134 (32%)	3018 (30%)
Secondary structure content [%]		
α-helices	4.7	6.7
β-sheet	29.4	29.0
3_{10}-helices	2.6	0
No. of hydrogen bonds §		
Main chain – Main chain	112 (0.47)	103 (0.43)
Main chain – Side chain	69 (0.29)	54 (0.22)
Side chain – Side chain	56 (0.23)	38 (0.16)
Residues in possible ion-pairs (4 Å cutoff)		
His/Lys/Arg	11	12
Asp/Glu	13	15
PDB ID	5ACH	6YDF

Table S5. Structural features of *Ls*AA9_A published in [2] and *Cv*AA9_A (this manuscript).

* Sequence length determined by mass spectrometry. Construct length is 252 residues, thus proteolysis must have occurred during crystallization.

§ Number in parenthesis is average number of hydrogen bonds pr residue.

SUPPLEMENTARY FIGURES

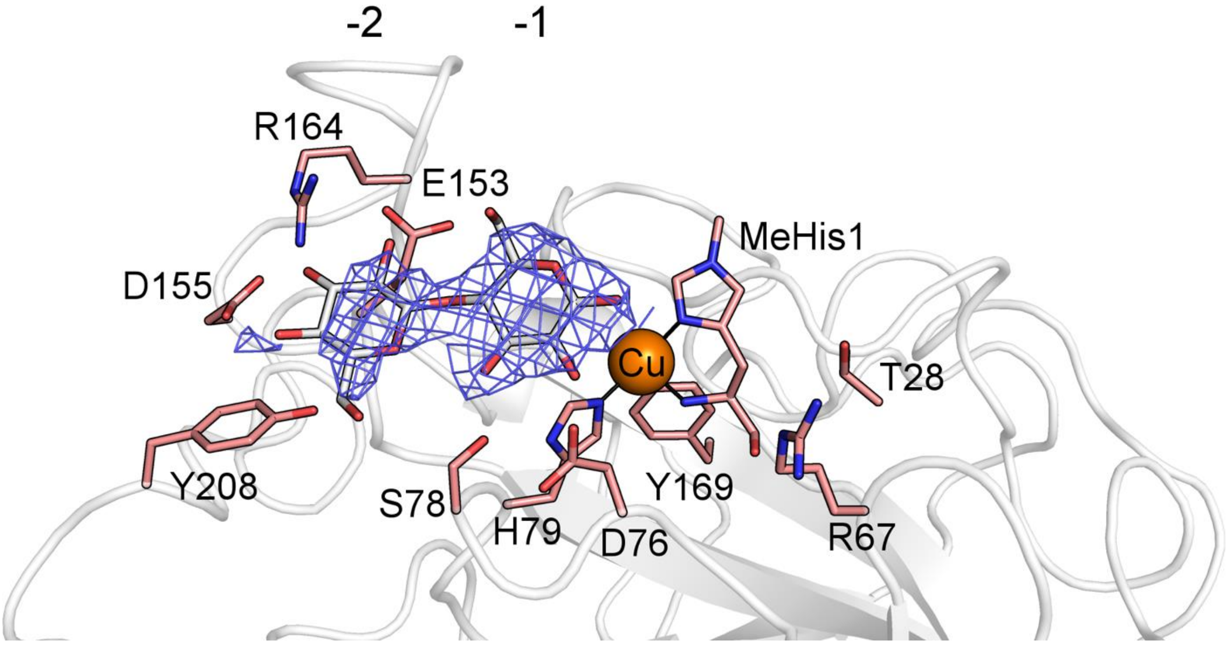


Figure S1. Active site surface of molecule B in *Ls*AA9_A-Cell$_5$. Only two glucosyl units could be modeled into the available electron density (final 2F$_O$-F$_C$ is shown as blue mesh at 1.0 σ contour level).

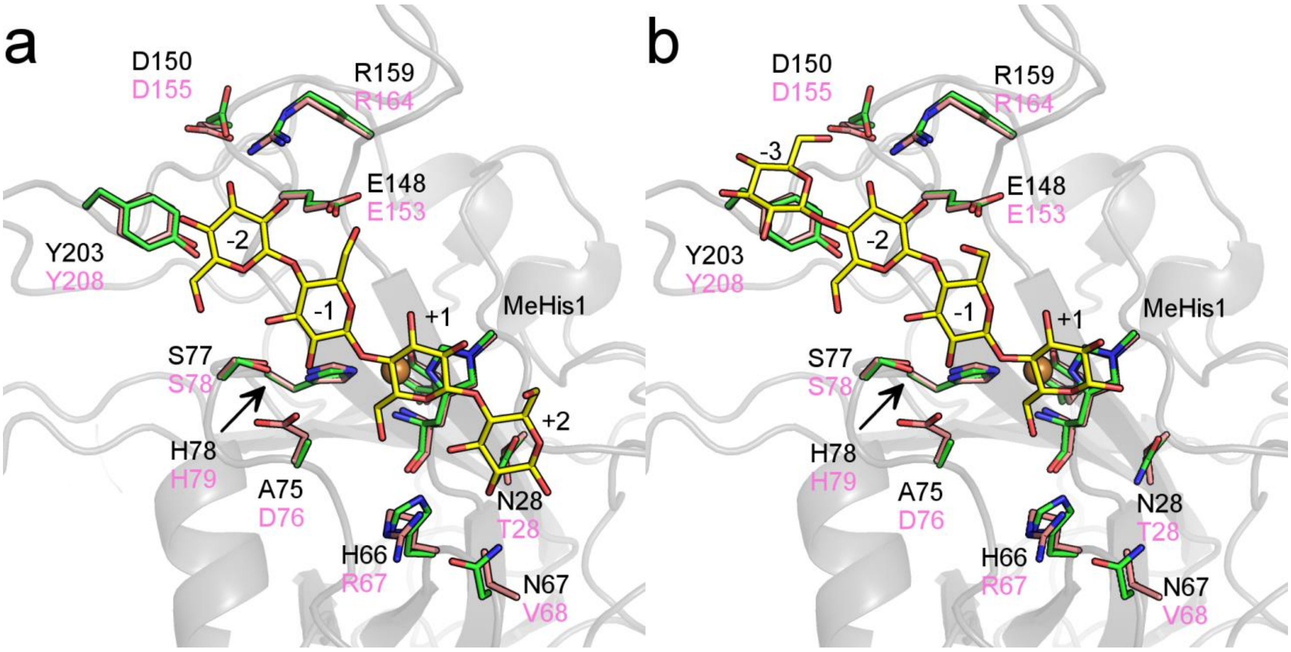


Figure S2. Superposition made using chain A of *Cv*AA9_A-Cell$_4$ and *Ls*AA9_A-Cell$_4$. The superposition was made using 1371 atoms with the PyMOL command *super*. (a) *Ls*AA9_A-Cell$_4$ showing cellotetraose bound in subsites -2 to +2. (b) *Cv*AA9_A-Cell$_4$ showing cellotetraose bound in subsites -3 to +1. Residues and residue names are shown in green and black respectively for *Ls*AA9_A, while they are shown in pink and magenta for *Cv*AA9_A.

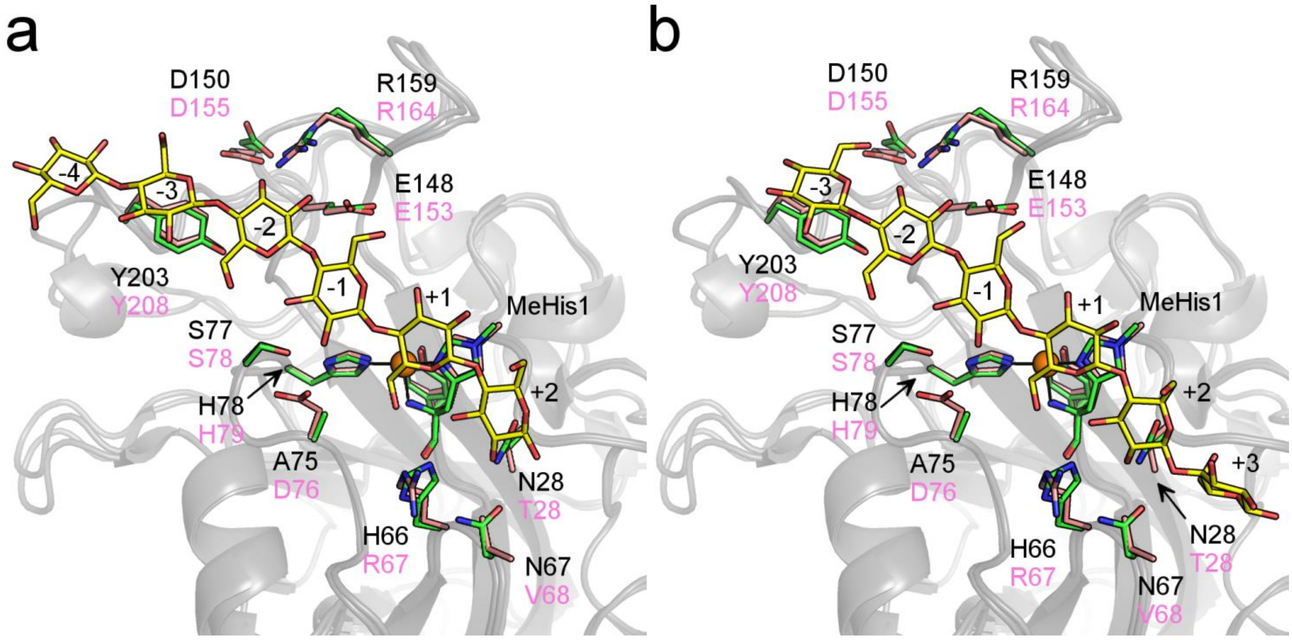


Figure S3. Superposition made using chain A of *Cv*AA9_A-Cell$_6$ and *Ls*AA9_A-Cell$_6$. The superposition was made using 1389 atoms with the PyMOL command *super*. (a) *Ls*AA9_A-Cell$_6$ showing cellohexaose bound in subsites -4 to +2. (b) *Cv*AA9_A-Cell$_6$ showing cellohexaose bound in subsites -3 to +3. Residues and residue names are shown in green and black respectively for *Ls*AA9_A, while they are shown in pink and magenta for *Cv*AA9_A.

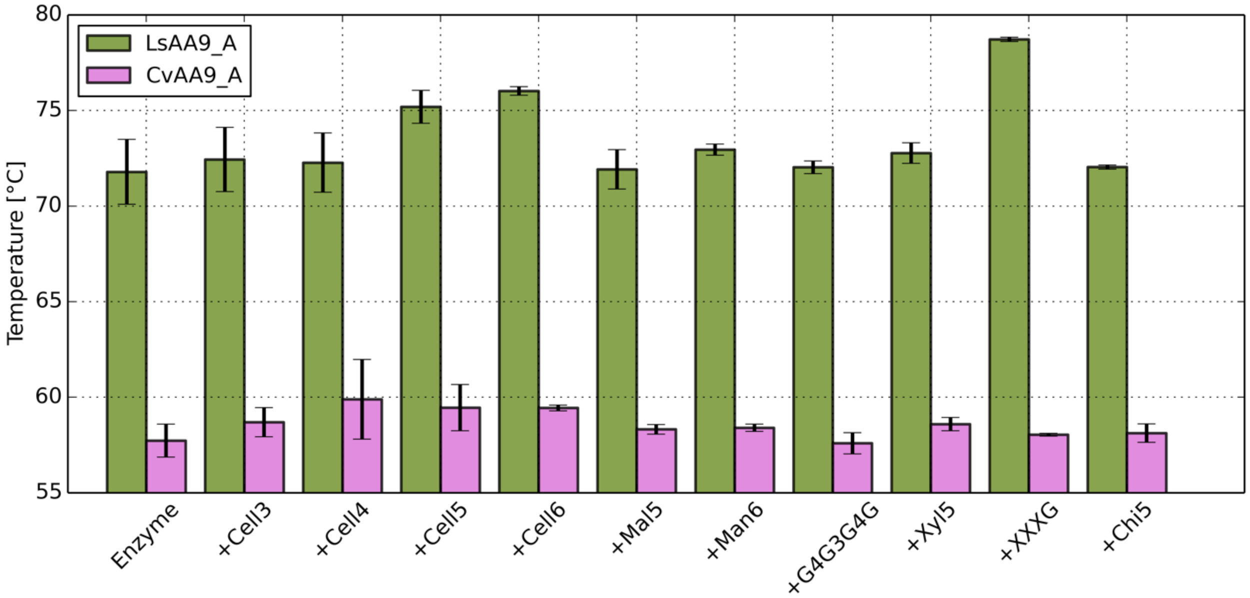


Figure S4. Inflection temperatures (T_i) of Cu-loaded *Ls*AA9_A and *Cv*AA9_A in solution without/with 10 mM oligosaccharides. Graph is based on values presented in Table 2 and Table S4. Cellopentaose ($Cell_5$), Cellohexaose ($Cell_6$) and xyloglucan oligosaccharide (X_3Glc_4) significantly increases the T_i for *Ls*AA9_A, while shifts in T_is are less conclusive for *Cv*AA9_A. Error bars indicate standard deviation.

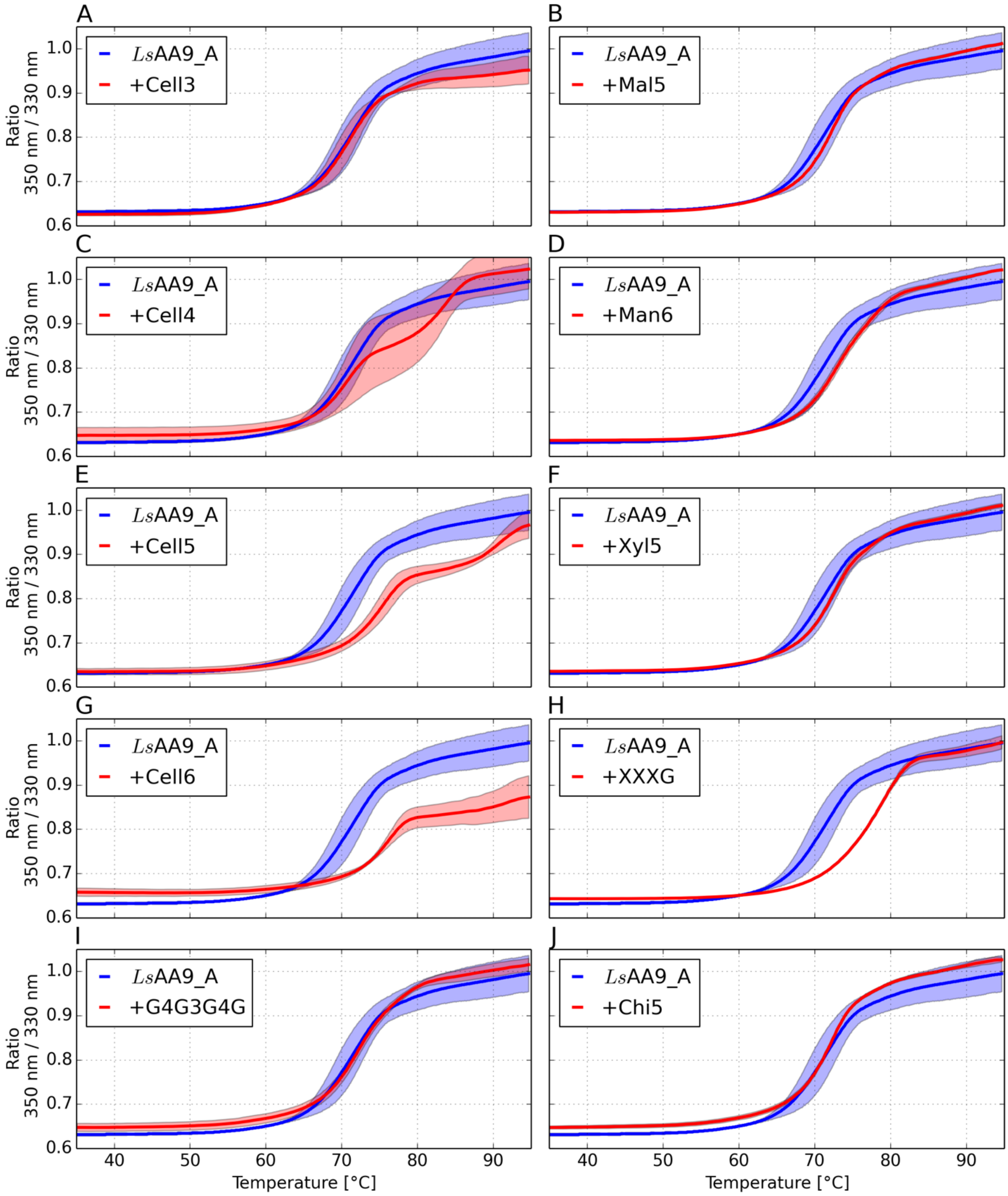


Figure S5. Unfolding curves of *Ls*AA9_A in solution with 10 mM oligosaccharides. Panels (A-J) with individual oligosaccharides as labeled. Transparent colored areas represent the standard deviation.

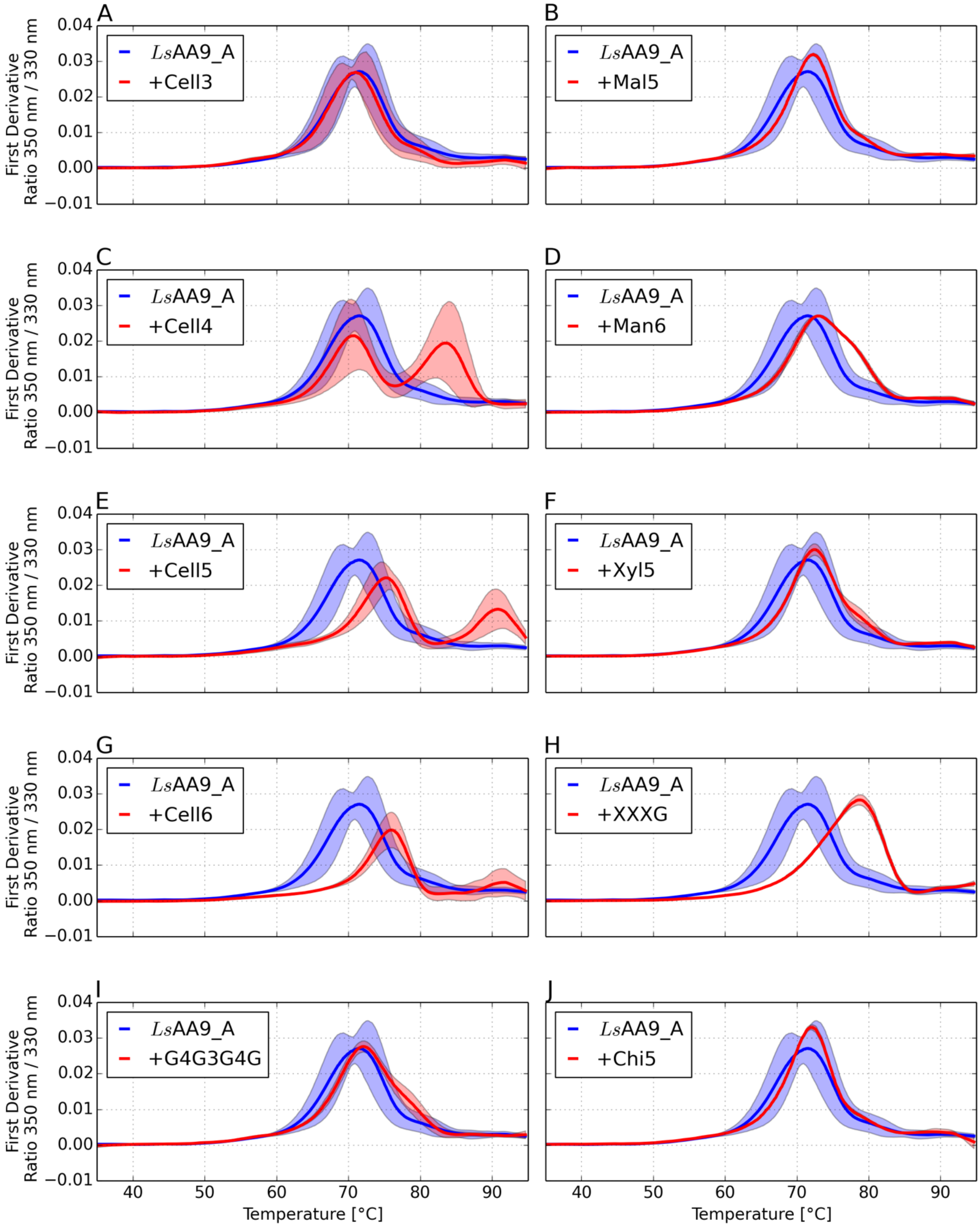


Figure S6. First derivative of unfolding curves of *Ls*AA9_A in solution with 10 mM oligosaccharides. Panels (A-J) with individual oligosaccharides as labeled. Transparent colored areas represent the standard deviation.

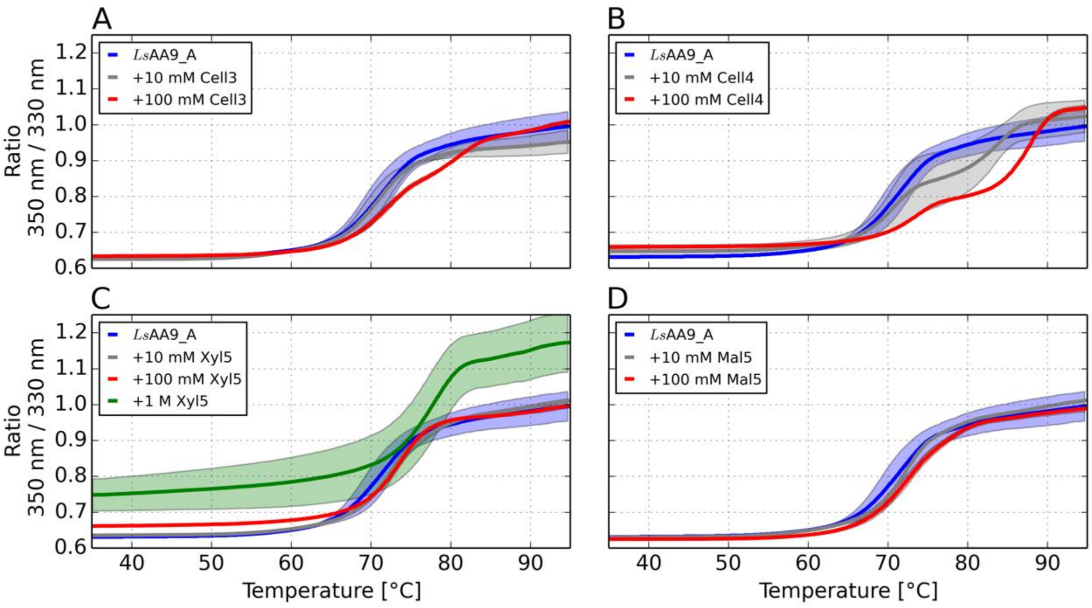


Figure S7. Unfolding curves of *Ls*AA9_A in solution with 10 mM, 100 mM or 1 M oligosaccharides as indicated in the legends. *Ls*AA9_A with $Cell_3$ or $Cell_4$ result in biphasic unfolding curves. With 10 mM or 100 mM Xyl_5 the unfolding curves for *Ls*AA9_A appear unchanged from the unfolding curve of the native enzyme. At 1 M Xyl_5 the unfolding curve of *Ls*AA9_A has shifted towards a higher T_i. Transparent colored areas represent the standard deviation.

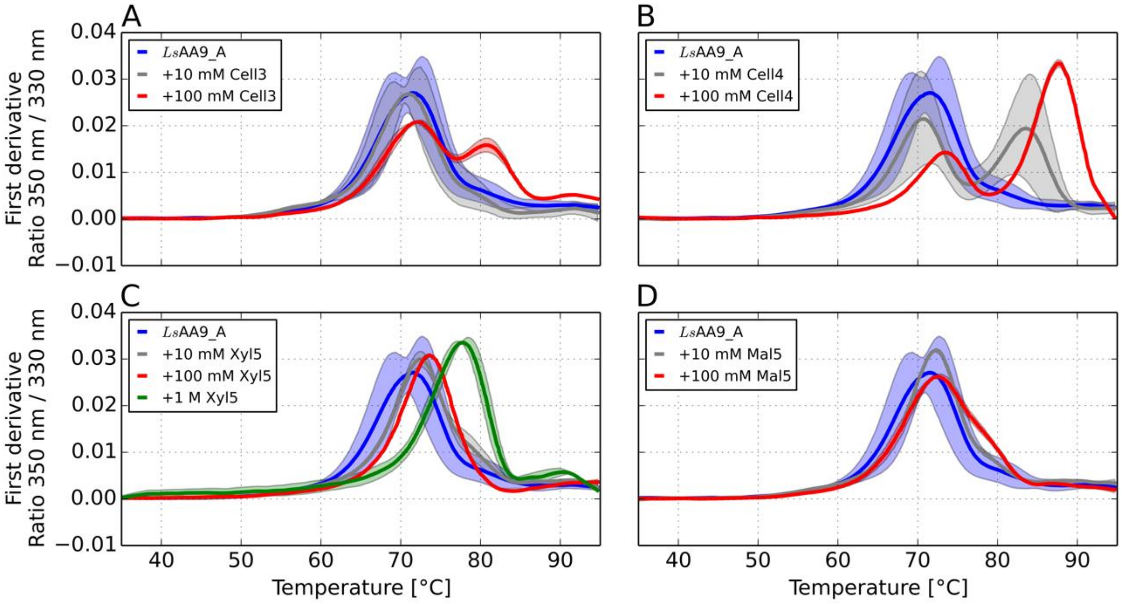


Figure S8. First derivatives of unfolding curves of *Ls*AA9_A in solution with 10 mM, 100 mM or 1 M oligosaccharides as indicated in the legends.

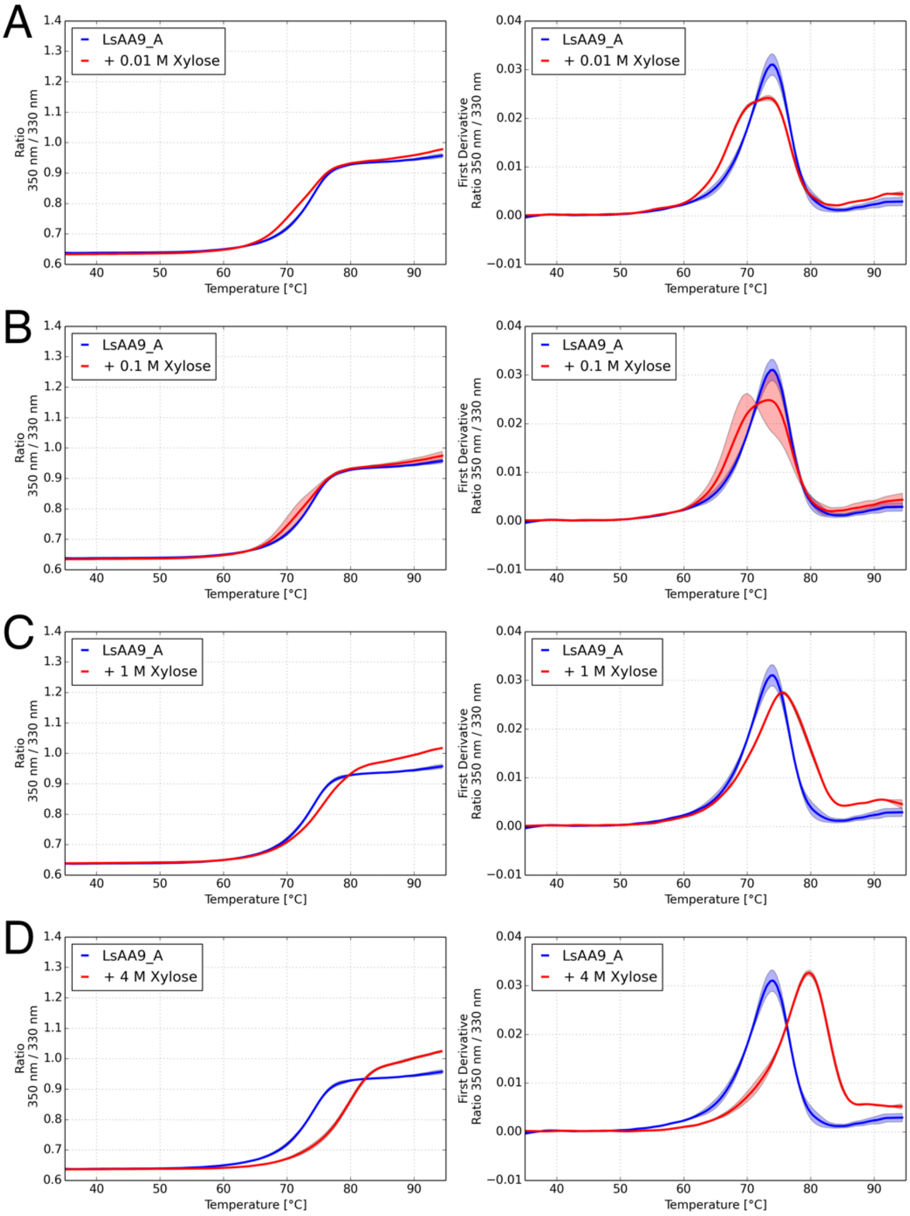


Figure S9. Unfolding curve of Cu-loaded *Ls*AA9_A in solution with xylose in increasing concentrations. At concentrations up to 0.1 M xylose the Ti is not significantly changed, but slight differences are found in the shape of the unfolding curves. At concentrations of 1-4 M xylose there is a clear thermal shift, indicative of non-specific interactions having an effect on the thermal unfolding at these concentrations.

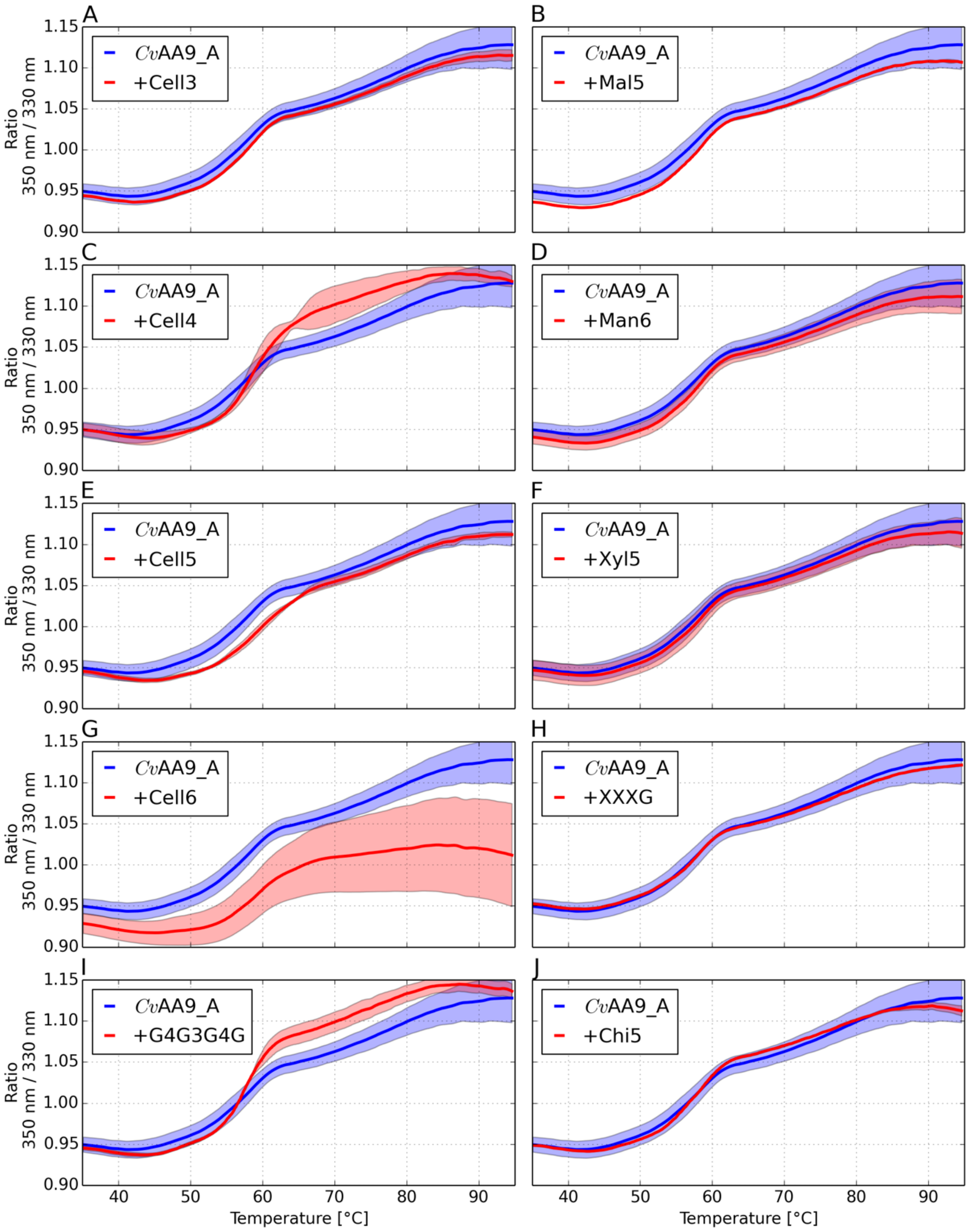


Figure S10. Unfolding curves of *Cv*AA9_A in solution with 10 mM oligosaccharides. Panels (A-J) with individual oligosaccharides as labeled. Transparent colored areas represent the standard deviation.

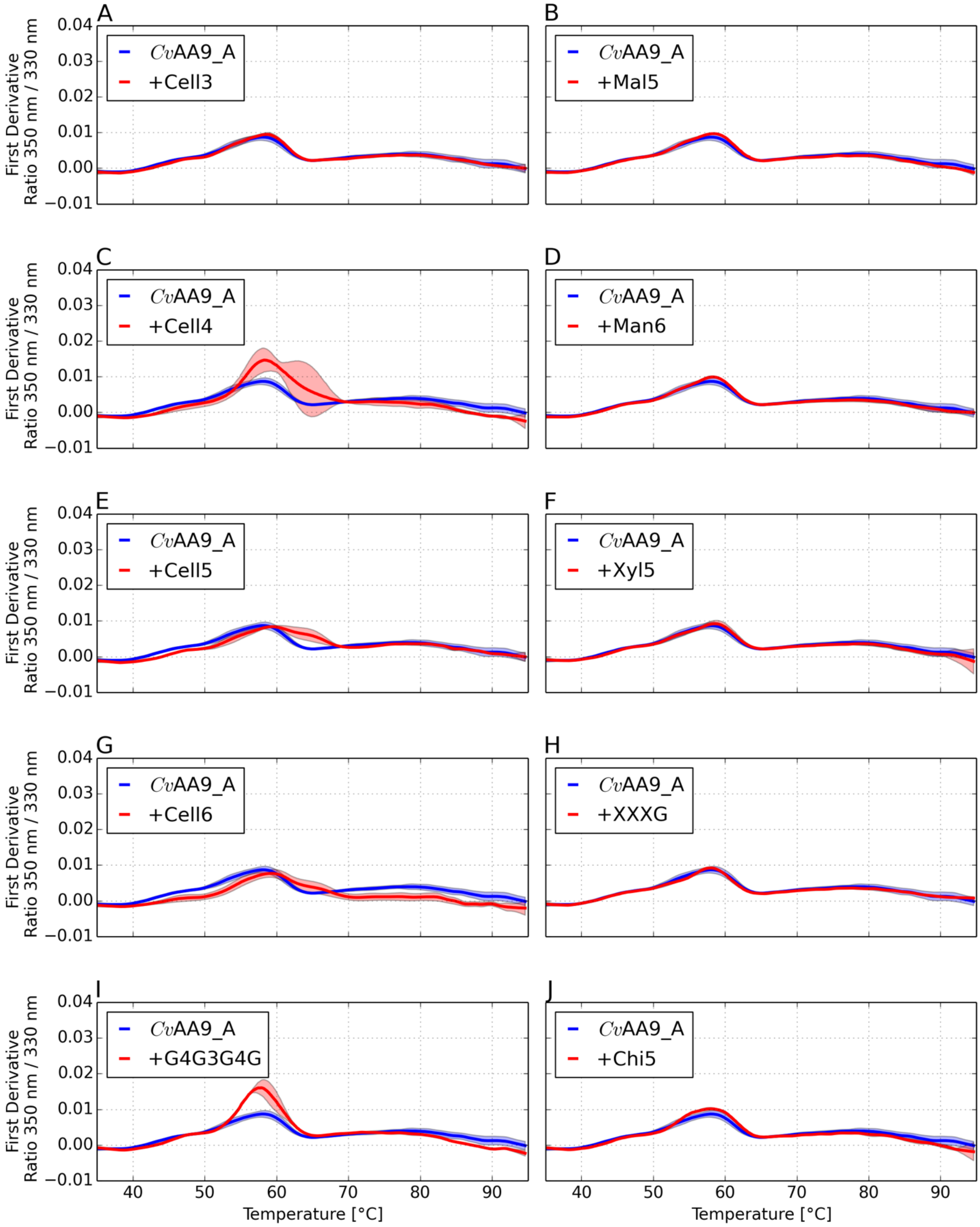


Figure S11. First derivative of unfolding curves of *Cv*AA9_A in solution with 10 mM oligosaccharides. Panels (A-J) with individual oligosaccharides as labeled. Transparent colored areas represent the standard deviation.

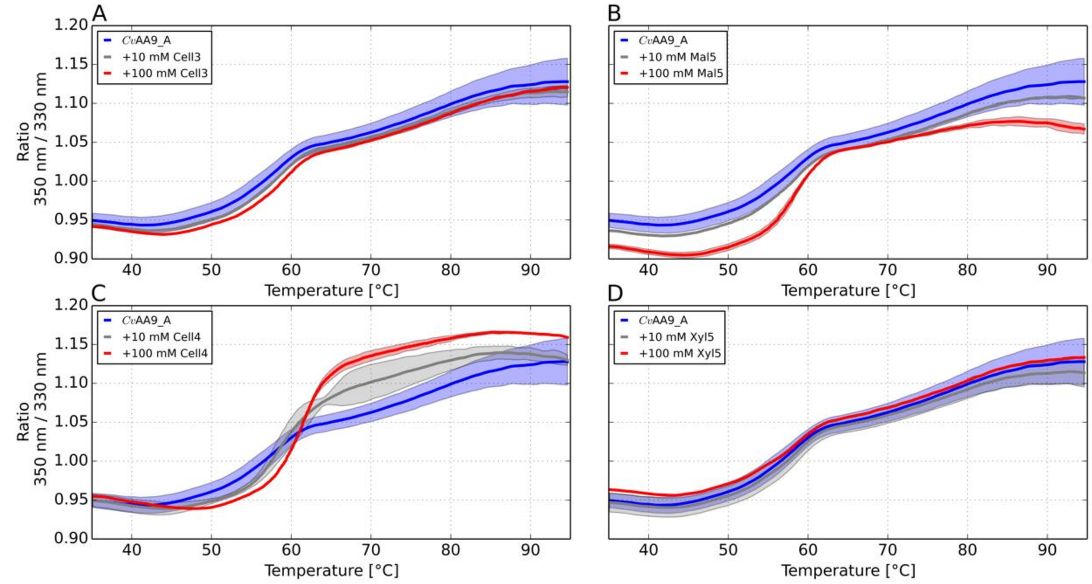


Figure S12. Unfolding curves of *Cv*AA9_A in solution with 10 mM or 100 mM oligosaccharides as indicated in the legends. Transparent colored areas represent the standard deviation.

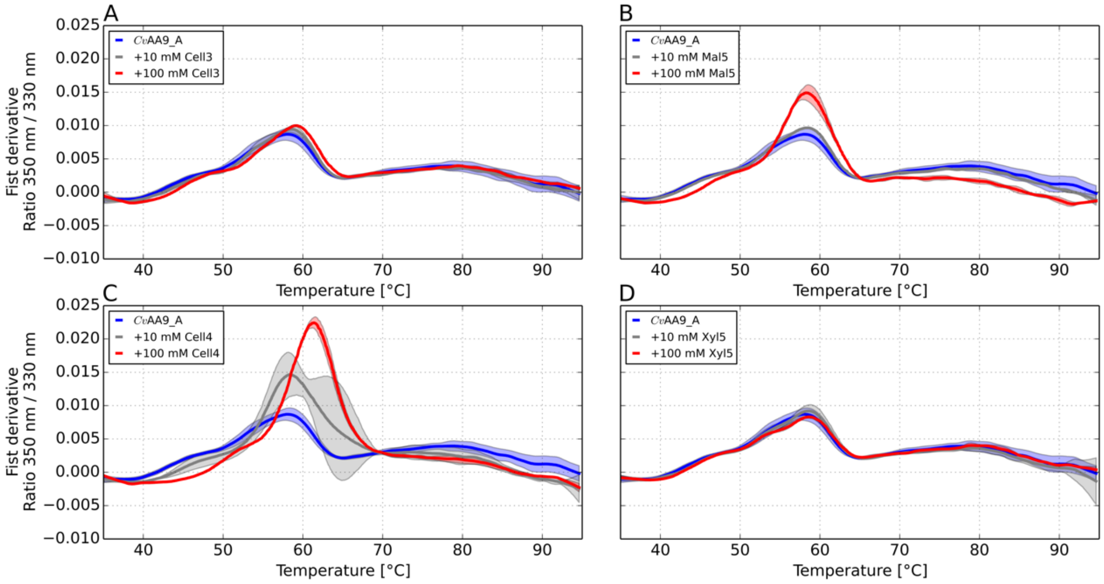


Figure S13. First derivative of unfolding curves of *Cv*AA9_A in solution with 10 mM or 100 mM oligosaccharides as indicated in the legends. Transparent colored areas represent the standard deviation.

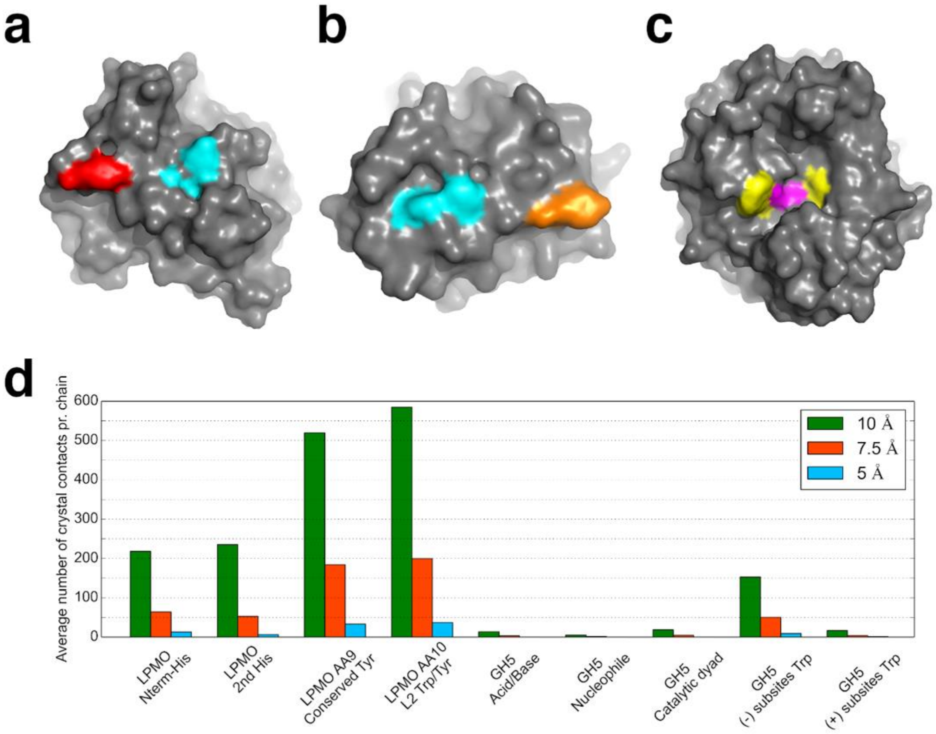


Figure S14. Crystal contacts in LPMOs and GH5. (a) *Ls*AA9_A (PDB 5ACH) active site surface with the conserved Tyr (in AA9s) near the minus subsites (red). (b) *Sm*AA10_A (PDB 2BEM) active site surface with extended L2 loop harboring the Tyr/Trp position (orange). For both (a) and (b) the His-brace residues are in cyan (c) *Ta*GH5_A (PDB 1H1N) active site surface showing a cleft around the catalytic Glu dyad (magenta) and the two substrate interacting Trp (yellow). (d) average number of crystal contacts pr. chain observed within 5 Å, 7.5 Å and 10 Å of the above mentioned residues.

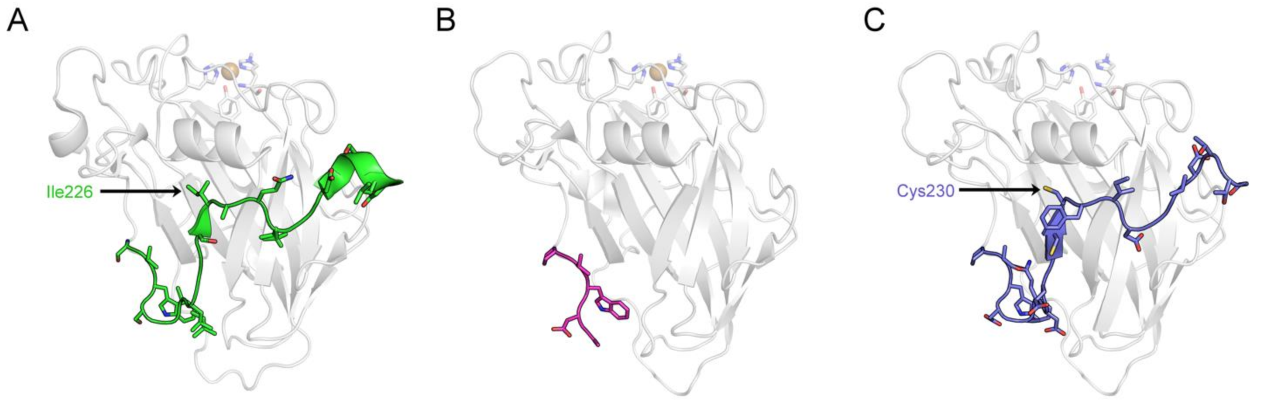


Figure S15. C-terminus of *Ls*AA9_A (PDB 5ACH), *Cv*AA9_A (*P*2$_1$, PDB 6YDF) and *Cv*AA9_A (extended C-terminal used in molecular dynamics simulations). (A) For *Ls*AA9_A residue numbers 216-235 of the C-terminal are highlighted in green. Indicated with an arrow is the buried Ile226. (B) C-terminal residues 220-224 are highlighted in magenta. (C) *Cv*AA9_A with an extended C-terminal modeled using *Ls*AA9_A (PDB 5ACH) as a template. Highlighted in blue is residue number 220-236. The equivalent position to Ile226 in (A) is Cys230 show in (C). The elongated C-terminus of *Cv*AA9_A was modeled to compare C-terminal flexibility between *Cv*AA9_A and *Ls*AA9_A in molecular dynamics simulations.

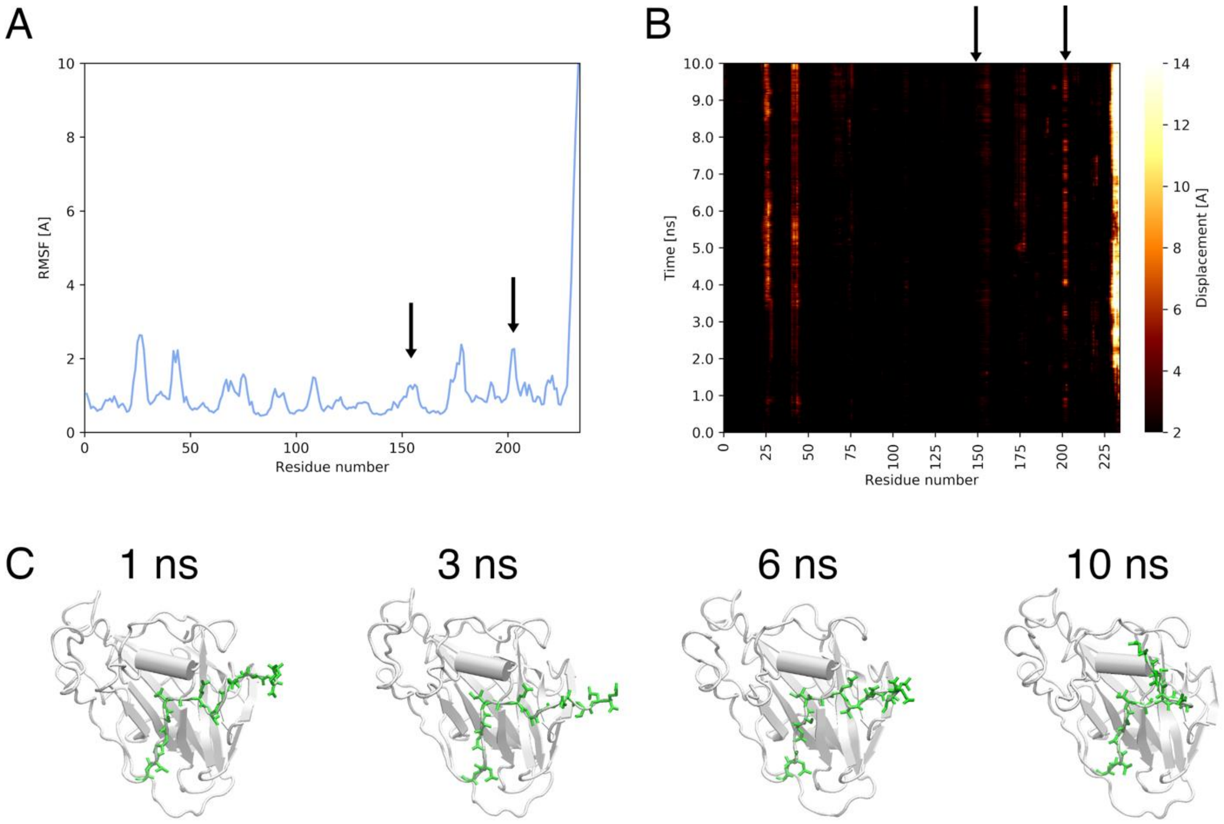


Figure S16. Molecular dynamics simulation of *Ls*AA9_A at 400 K. (A) RMSF of *Ls*AA9_A over the 10 ns simulation. (B) is the displacement of each amino acid Cα shown as a heat map over the course of the simulation. (C) panel displays modeled residues 220-235 over the course of the simulation. The C-terminal segment from residue 227-235 exhibits large fluctuations in coordinates, while the segment 220-226 largely retains the initial conformation, as found from panels (A) and (B). Arrows indicate the L7 and LC loops.

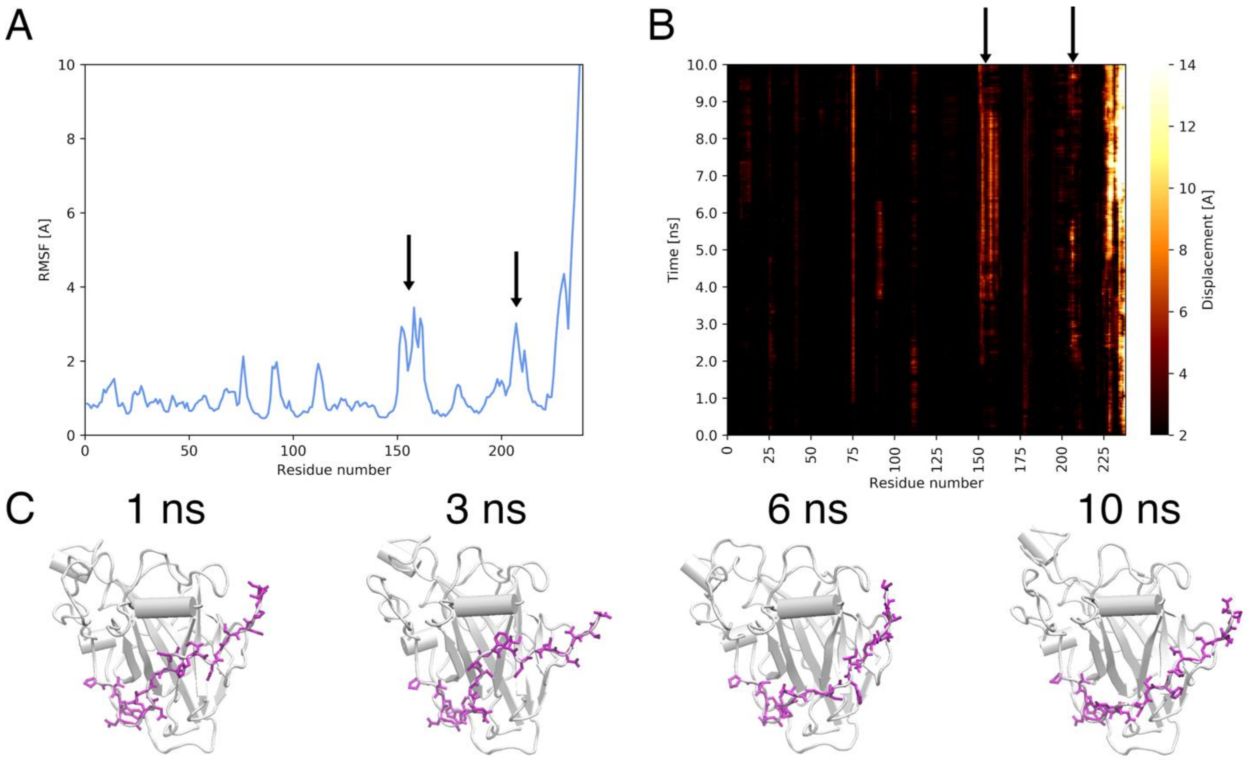


Figure S17. Molecular dynamics simulation of *Cv*AA9_A 400 K. (A) RMSF of *Cv*AA9_A over the 10 ns simulation. (B) is the displacement of each amino acid Cα shown as a heat map over the course of the simulation. (C) panel displays modeled residues 220-236 over the course of the simulation. The modeled C-terminal part of *Cv*AA9_A exhibits large fluctuations, as found from panels (A) and (B). Arrows indicate the L7 and LC loops.

REFERENCES

(1) Pérez, S., and Samain, D. (2010) Structure and Engineering of Celluloses, in *Advances in Carbohydrate Chemistry and Biochemistry*, pp 25–116.

(2) 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., and Walton, P. H. (2016) The molecular basis of polysaccharide cleavage by lytic polysaccharide monooxygenases. *Nat. Chem. Biol. 12*, 298–303.