

The structure and function of an arabinoxylan-specific xylanase

Márcia A.S. Correia^{1*}, Koushik Mazumder^{3*}, Joana L. A. Brás¹, Susan J. Firbank^{2*}, Yanping Zhu^{2,3*}, Richard J. Lewis², William S. York³, Carlos M. G. A. Fontes^{1**} and Harry J. Gilbert^{2,3**}

¹CIISA - Faculdade de Medicina Veterinária, Universidade Técnica de Lisboa, Avenida da Universidade Técnica, 1300-477 Lisboa, Portugal; ²Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Newcastle upon Tyne NE2 4HH, United Kingdom; ³Complex Carbohydrate Research Center, University of Georgia, 315 Riverbend Road, Athens, Georgia 30602-4712, USA; ⁴Department of Biochemistry & Molecular Biology, University of Georgia, Life Sciences Building, 120 Green Street, Athens, Georgia 30602-7229, USA

Running title: Structure and function of CtXyl5A

*Contributed equally to this work.

**Corresponding authors Email: h.j.gilbert@newcastle.ac.uk and cafontes@fmv.utl.pt.
Address: Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Framlington Place, Newcastle upon Tyne NE2 4HH, U.K.

The enzymatic degradation of plant cell walls plays a central role in the carbon cycle and is of increasing environmental and industrial significance. The enzymes that catalyse this process include xylanases which degrade xylan, a β -1,4 xylose polymer that is decorated with various sugars. Although xylanases efficiently hydrolyse unsubstituted xylans, these enzymes are unable to access highly decorated forms of the polysaccharide, such as arabinoxylans that contain arabinofuranose decorations. Here we show that a *Clostridium thermocellum* enzyme, designated CtXyl5A, hydrolyses arabinoxylans but does not attack unsubstituted xylans. Analysis of the reaction products generated by CtXyl5A showed that all the oligosaccharides contain an O3 arabinose linked to the reducing end xylose. The crystal structure of the

catalytic module (CtGH5) of CtXyl5A, appended to a family 6 non-catalytic carbohydrate binding module (CtCBM6), showed that CtGH5 displays a canonical $(\alpha/\beta)_8$ -barrel fold with the substrate binding cleft running along the surface of the protein. The catalytic apparatus is housed in the centre of the cleft. Adjacent to the -1 subsite is a pocket that could accommodate an L-arabinofuranose linked α -1,3 to the active site xylose, which is likely to function as a key specificity determinant. CtCBM6, which adopts a β sandwich fold, recognizes the termini of xylo- and gluco-configured oligosaccharides, consistent with the pocket topology displayed by the ligand binding site. In contrast to typical modular glycoside hydrolases, there is an extensive hydrophobic interface between CtGH5 and CtCBM6, and thus the two modules

cannot function as independent entities.

The plant cell wall, which is an important biological and industrial resource, consists, primarily, of interlocking polysaccharides (see (1) for review). The biological conversion of the polysaccharides within the plant cell wall to their constituent monosaccharides is central to its biological and industrial exploitation (2,3). An example of this chemical complexity is provided by xylan, which is the major hemicellulosic component of the wall. This polysaccharide comprises a backbone of β -1,4 xylose residues in their pyranose configuration (*Xylp*), which are decorated at O2 with 4-O-methyl-D-glucuronic acid, at O2 and/or O3 with arabinofuranose (*Araf*) residues, while the polysaccharide can also be extensively acetylated. In addition, the *Araf* side chain decorations can also be esterified to ferulic acid that, in some species, provides a chemical link between hemicellulose and lignin, **Figure 1** (1). The precise structure of xylans varies between plant species, tissues and during cellular differentiation (4).

Reflecting the chemical complexity of plant structural polysaccharides, microbial plant cell wall degrading microorganisms express a large number of enzymes, often in excess of 100 biocatalysts, that target specific linkages within these carbohydrate polymers (5-8). The majority of plant cell wall degrading enzymes are glycoside hydrolases, although polysaccharide lyases and carbohydrate esterases also contribute to the catabolic process. These enzymes are grouped into families based on sequence, structural and

catalytic conservation, within the CAZy database (9). As discussed in the accompanying paper, many of these enzymes are appended to non-catalytic carbohydrate binding modules (CBMs) that are also grouped into families on the CAZy database. The xylan backbone is hydrolyzed by xylanases, the majority of which are located in glycoside hydrolase families (GHs) 10 and 11, although they are also present in GH8 and GH30 (10,11). The extensive decoration of the xylan backbone generally restricts the capacity of these enzymes to attack the polysaccharide prior to removal of the side chains (12).

Here we report the biochemical properties and crystal structure of a GH5 enzyme that is appended to a family 6 CBM (*CtCBM6*). The enzyme (defined as *CtXyl5A*) is an arabinoxylan-specific xylanase that utilizes *Araf* decorations, appended to O3 of the *Xylp* bound at the active site, as an essential specificity determinant. The capacity of *CtXyl5A* to also accommodate arabinose side chains in all the other subsites (in addition to the active site) within the substrate binding cleft enables the enzyme to hydrolyze highly decorated arabinoxylans. The functional significance of the specificity of the arabinoxylanase, in the context of the plant cell wall degrading apparatus of the host bacterium, is discussed.

EXPERIMENTAL PROCEDURES

Cloning, expression and purification of components of *CtXyl5A*

DNA encoding *CtGH5*, *CtGH5-CBM6* and *CtCBM6* were amplified using primers, containing *NheI* and *XhoI* restriction sites, which are listed in **supplemental information, Table S1.**

The amplified DNAs were cloned into *NheI/XhoI* restricted pET21a such that the encoded recombinant proteins contained a C-terminal His₆ tag. To express the *C. thermocellum* proteins, *Escherichia coli* strain BL21(DE3), harbouring appropriate recombinant plasmids, was cultured to mid-exponential phase in Luria broth at 37 °C, followed by the addition of isopropyl β-D-galactopyranoside at 1 mM to induce recombinant gene expression, and incubated for a further 5 h at 37 °C. The recombinant proteins were purified to >90 % electrophoretic purity by immobilized metal ion affinity chromatography (IMAC) using TalonTM (Clontech), cobalt-based matrix, and elution with 100 mM imidazole, as described previously (13). When preparing the selenomethionine derivative of CtGH5-CBM6 for crystallography, the proteins were expressed in *E. coli* B834 (DE3), a methionine auxotroph, cultured in media comprising 1 litre SelenoMet Medium BaseTM, 50 ml SelenoMet Nutrient MixTM (Molecular Dimensions) and 4 ml of a 10 mg/ml solution of L-selenomethionine. Recombinant gene expression and protein purification was as described above except that all purification buffers were supplemented with 10 mM β-mercaptoethanol.

Mutagenesis

Site-directed mutagenesis was carried out using the PCR-based QuikChange method (Stratagene) deploying the primers listed in the **supplemental information, Table S1**.

Enzyme assays

CtXyl5A and its derivatives were assayed for enzyme activity using the method of Miller (14) to detect the

release of reducing sugar. The standard assay was carried out in 50 mM sodium phosphate buffer, pH 7.0, and the potential polysaccharide substrate was at 1 mg/ml. The reactions were initiated by the addition of enzyme up to 10 μM and incubated at 60 °C (unless otherwise stated) for up to 16 h. The identification of potential reaction products were also assessed by HPAEC using methodology described previously (15). The capacity of CtGH5 and CtGH5-CBM6 to hydrolyse xylooligosaccharides was assessed by HPAEC using 100 μM of oligosaccharide and 5 μM of protein.

Oligosaccharide analysis

Rye arabinoxylan (5 g) was digested to completion (no further increase in reducing sugar and change in the HPAEC product profile) with 3 μM of CtXyl5A at 60 °C for 48 h. The oligosaccharide products were partially purified by size exclusion chromatography using a Bio-Gel P2 column as described previously (16). The structures of the oligosaccharides were analyzed by NMR, electrospray ionization mass spectrometry (ESI-MS) and HPAEC in combination with selective enzyme treatment. Partially methylated alditol acetate derivatives of the glycosyl residues of the oligosaccharides were prepared and analyzed by gas chromatography electron impact mass spectrometry GC-EIMS.

Preparation of the partially methylated alditol acetates

The mixture of oligosaccharides (~500 μg) was per-O-methylated using the method of Ciucanu and Kerek (17). The per-O-methylated oligosaccharides were hydrolyzed with 2N TFA, reduced and acetylated to generate partially

methylated alditol acetate (PMAA) derivatives (18).

GC-EIMS analysis

PMAA derivatives were analyzed with a Hewlett Packard 5890 gas chromatograph - mass spectrometer. The PMAAs were separated with a SP 2330 column (30 m X 0.25 mm, 0.25 μ m film thickness, Supelco) using the following temperature gradient: 80 °C for 2 min, 80-170°C at 30 °C/min, 170-240°C at 4 °C/min, 240 °C held for 20 min. Samples were ionized by electrons impact at 70eV.

Preparation of per-O-methylated oligoglycosyl alditols

The sample (~500 μ g) was reduced with sodium borohydride to generate oligoglycosyl alditols, which were per-O-methylated as previously described (19).

MALDI-TOF mass spectrometry (MALDI-TOF-MS)

Positive ion MALDI-TOF mass spectra were recorded using an Applied Biosystems Voyager-DE biospectrometry workstation. Samples (1 μ l of a mg/ml solution) were mixed with an equal volume of matrix solution (0.1 M 2,5-dihydroxybenzoic acid and 0.03 M 1-hydroxyisoquinoline in aqueous 50 % MeCN) and dried on MALDI target plate. Typically, spectra from 200 laser shots were summed to generate a mass spectrum.

ESI-MS

The multiple stage ESI mass spectra were recorded in a Thermo Scientific LTQ XL ion trap mass spectrometer. Per-O-methylated oligoglycosyl alditols in methanol were diluted with 50 % acetonitrile-water containing 0.1 % TFA. Samples were infused through a fused silica capillary (150 μ m i.d. X 363 μ m o.d. X ~ 60 cm, Thermo Finnigan, USA) into the source at flow rate of 3 μ l/min using the syringe pump provided with the instrument. The electrospray source was operated at a voltage of 5.0 KV and the capillary heater was set to 275 °C. All the experiments were performed in the positive-ion mode.

NMR spectroscopy

Oligosaccharides (~2 mg) were dissolved in D₂O (0.5 ml, 99.9 %; Cambridge Isotope Laboratories). ¹H NMR spectra were recorded with Varian Inova NMR spectrometer operating at 500 MHz at 298 K. All two dimensional spectra were recorded using standard Varian pulse programs.

Isothermal Titration Calorimetry

The binding of C_tCBM6 to ligands was quantified by isothermal titration calorimetry (ITC), as described previously (20). Titrations were carried out in 50 mM Na/Hepes buffer, pH 7.5, containing 5 mM CaCl₂ at 25 °C. The reaction cell contained protein at 145 μ M, while the syringe contained the monosaccharide or oligosaccharide at 5-15 mM, while polysaccharide, when used as the titrant, was at 3-5 mg/ml. The titrations were analyzed using Microcal Origin version 7.0 software to derive, *n*, *K_a* and Δ H values, while Δ S was calculated using the standard

thermodynamic equation $-RT\ln K_a = \Delta G = \Delta H - T\Delta S$.

Crystallography

Proteins were crystallized using the hanging drop vapour technique at 20 °C with an equal volume (1 µl) of protein and reservoir solution. Native (10 mg/ml) and selenomethionine (3 mg/ml) *CtGH5*-CBM6 crystallised in 16-24 % PEG 3000, 150 mM Na/citrate, pH 5.5. A *CtGH5*-CBM6 construct containing 2 additional methionines, W391M/W397M, was produced to facilitate structure solution by selenomethionine SAD. Crystals were cryoprotected by the inclusion of 25 % glycerol in the crystallization solution and flash frozen in liquid nitrogen. Diffraction data were collected at ID14.4 ESRF, Grenoble, France at the selenium *K* absorption edge to enable structure solution by SAD. The diffraction data were processed in MOSFLM (21) and SCALA (22) and the heavy atom substructure was also solved using SHELXCDE (23) as part of CCP4i, and an initial model was built in Arp/wArp (24), which was completed manually in COOT (25). The complete initial model was used to determine the structure of the wild type protein by molecular replacement and refined at higher resolution from data collected at the Diamond Light Source, UK. The crystal of the reported structure had been soaked in 20 mM “Fraction 1” in an attempt to obtain a structure of the enzyme in complex with carbohydrate although no sugar molecules, other than glycerol, were observed in the electron density.

All structures were refined to convergence using REFMAC5 (26) with manual corrections being applied in COOT (25). The data collection, phasing

and refinement statistics are displayed in **Table S2** and the PDB codes for the protein structures are as follows: 2y8k

RESULTS

Expression and purification of *CtXyl5A*

To investigate the function of the *CtGH5* and *CtCBM6* components of *CtXyl5A*, the modules were expressed as either individual entities or covalently linked. Whilst *CtCBM6* and *CtGH5*-CBM6 were expressed in soluble form at high levels in *E. coli*, *CtGH5* was predominantly insoluble and only a small amount of soluble protein was generated in the enteric bacterium. All three proteins were purified by IMAC to electrophoretic homogeneity.

CtXyl5A is an arabinoxylanase

Screening the capacity of *CtXyl5A* to hydrolyse plant structural polysaccharides revealed that the enzyme was able to degrade rye and wheat arabinoxylan, displayed limited activity against oat spelt xylan, but was unable to act on glucuronoxylan, birch or beechwood xylan, **Table 1**. The enzyme displayed no activity against a range of mannans, pectins, galactans, arabinans, and β-glucans (data not shown). The individual kinetic constants of *CtXyl5A* against rye and wheat arabinoxylan could not be determined as the K_M was greater than the maximum concentration of soluble substrate, however, the catalytic efficiency of the enzyme was similar for both rye and wheat arabinoxylan. The high K_M may reflect weak affinity for the substrate, or the glycosidic bonds targeted by *CtXyl5A* occur rarely in the arabinoxylan substrates. The enzyme displayed trace activity against

xylohexaose with a $k_{\text{cat}}/K_{\text{M}}$ estimated to be $<10^1 \text{ min}^{-1} \text{ M}^{-1}$. These data indicate that *CtXyl5A* hydrolyses arabinoxylans but does not act on xylans that contain few arabinofuranose side chains. This is in sharp contrast to typical xylanases, located mainly in GH10 and GH11, which display a preference for the poorly decorated xylans from birch and beechwood (12). These data show that *CtXyl5A* displays specificity for arabinoxylans and as such is defined as an arabinoxylanase, an activity not previously reported.

Characterization of the reaction products generated by *CtXyl5A* from arabinoxylan

To explore the substrate specificity of *CtXyl5A* in more detail, the reaction products generated by treating rye arabinoxylan with the enzyme were partially purified by size exclusion chromatography to remove high molecular weight polymers. The fractions containing the majority of the products were pooled (designated henceforth as Fraction 1). Fraction 1 was reduced and per-O-methylated and the products were analyzed by MALDI-TOF-MS. The data revealed that the major reaction products were pentose-containing oligosaccharides with degrees of polymerization (DPs) of 3 (m/z 565), 4 (m/z 725) and 5 (m/z 885), respectively, **Figure 2A**. Partially methylated alditol acetate derivatives were then prepared from per-O-methylated Fraction 1 and analyzed by GC-EIMS, **Figure 2B**. This semi-quantitative analysis revealed terminal *Araf* (methylated at O2, O3 and O5), terminal *Xylp* (methylated at O2, O3, and O4), 3-linked *Xylp*, 4-linked *Xylp*, and 3,4-linked *Xylp*. No *Xylp* residues decorated at O2 or at both O2

and O3 were observed. These data indicate that the oligosaccharides consist of a backbone of (1→4) linked *Xylp* residues decorated with *Araf* side chains at O3 of internal or reducing *Xylp* residues (3,4-linked *Xylp*), or at O3 of non-reducing terminal *Xylp* residues (3-linked *Xylp*). Fraction 1 was also treated with *CjAbf51A*, an arabinofuranosidase that releases *Araf* residues from O2 or O3 of singly-branched *Xylp* residues in the xylan backbone (27). HPAEC analysis of the *CjAbf51A* digestion products revealed the presence of arabinose, xylobiose, xylotriose and xylo-tetraose, **Figure 2C**, indicating that the predominant *CtXyl5A* products are xylooligosaccharides in which at least one of the *Xylp* residues bear a mono-*Araf* side chain. By contrast, GH10 and GH11 xylanases generate predominately xylose and xylobiose from wheat arabinoxylan, reflecting a preference for undecorated regions of the polysaccharide (12).

The oligosaccharides in Fraction 1 were analyzed by several 2D NMR methods, including gCOSY, HSQC, TOCSY, and ROESY. These analyses provided scalar and dipolar correlations that allowed the resonances of the most abundant spin systems to be assigned to specific sugar residues (**Table S3**; for a more detailed description of this approach, see, for example, (19,28,29)). Upfield shifts typical of reducing residues (19,28,29) were observed for two C1 resonances (δ 92.4 and 96.6) in the HSQC spectrum of the *CtXyl5A*-generated oligosaccharides, **Figure 3A**. In combination with other 2D NMR data, this allowed these two resonances to be assigned to α -*Xylp* and β -*Xylp* residues at the reducing end of the oligosaccharides. However, the exact ^1H and ^{13}C shifts of these reducing residues

indicate that they are structurally distinct from the unbranched (4-linked) sugars at the reducing end of oligosaccharides, generated by more typical endoxylanases (19,28,29). The data reveal the presence of an *Araf* side chain at O3 (along with a β -Xylp at O4) of the reducing Xylp residues of the *CtXyl5A*-generated oligosaccharides. For example, the C3 resonances of the reducing α -Xylp and β -Xylp units exhibit diagnostic downfield glycosylation shifts (δ_C 77.7 and 77.8), relative to the corresponding unbranched reducing residues produced by more typical endoxylanases (δ_C 71.2 and 73.8). Furthermore, the ROESY spectrum of Fraction 1, **Figure 3B**, revealed strong dipolar interactions between the two most intense α -*Araf* H1 resonances (δ_H 5.342 and 5.391) and the reducing α -Xylp and β -Xylp H3 resonances (δ_H 3.906 and 3.736, respectively), indicating that most of the α -*Araf* residues are linked to O3 of reducing Xylp moieties. The identification of branched, reducing Xylp residues in Fraction 1 is consistent with the detection of 3,4-linked Xylp residues in the partially methylated derivatives, **Figure 2B**. Resonances corresponding to unbranched 4-linked β -Xylp residues at the reducing end of the oligosaccharides (e.g. H1 at δ 4.584, **Figure 3A**) were not detectable in the NMR spectra. Integration of the Xylp and *Araf* H1 resonances in the 1D spectrum of the *CtXyl5A*-generated oligosaccharides, **Figure 3A**, allowed the following quantitative conclusions to be drawn: the oligosaccharides have an average backbone DP of 2.76 and an average overall DP of 4.04; >99 % of the oligosaccharides have an α -L-*Araf* side-chain on O3 of the reducing Xylp residue; approximately 30 % of the

oligosaccharides have a second α -L-*Araf* side chain.

To analyze Fraction 1 by ESI-MSⁿ, the oligosaccharides in this sample were treated with NaBH₄, and the resulting oligoglycosyl alditols were methylated prior to fragmentation. This procedure imparts a distinctive mass label to the newly formed alditol end of the oligosaccharide, facilitating ESI-MSⁿ analysis (19). The data, examples of which are shown in **Figure 4**, provided unambiguous evidence supporting the presence of branched reducing residues in the oligosaccharides in Fraction 1. This conclusion is exemplified by the analysis of the possible tetrasaccharides in Fraction 1. Thus, based on the structure of the polysaccharide substrate, linkage and NMR analysis of Fraction 1, only five different tetrasaccharide structures (Ia, Ib, IIa, IIb, and III) are theoretically possible, **Figure 5**. The ESI-MSⁿ analysis provided information regarding the topology of the oligomers, but did not define the stereochemistry (identity) of the individual pentose residues. Therefore, the terminal pentose residues at the non-reducing end of the main chain in structures Ia, Ib, IIa, IIb, displayed in **Figure 5**, are indicated by the letter P (as the sugar can be either *Araf* or Xylp residues). However, in **Figure 5**, non-terminal backbone residues, and sugars attached to branched backbone units (backbone sugars that are linked at O4 and O3 to other sugars), are known to be Xylp and *Araf*, respectively. Thus, structure I could be (*Araf*)-Xylp-Xylp-Xylol (Ia) or Xylp-Xylp-Xylp-Xylol (Ib) in which (*Araf*) is an arabinose decoration appended to the following xylose residue, while Xylol is the alditol form of the xylose at the reducing end. Structure II could be *Araf*-Xylp-(*Araf*)-

Xylol (IIa) or Xylp-Xylp-(Araf)-Xylol (IIb) and III is Xylp-(Araf)-Xylp-Xylol. The quasimolecular ($M+Na^+$) ion at m/z 725, corresponding to these DP4 structures was selected for MS², **Figure 4A**. The fragmentation pattern is dominated by Y-ions (19,30), which contain the alditol end of the oligomer. The Y-ion (m/z 551) generated by loss of a single terminal pentosyl residue was selected as the precursor for MS³ fragmentation, **Figure 4B**. Comparison of this MS³ spectrum, **Figure 4A,B**, to the theoretical fragmentation pattern for all possible m/z 551 ions, **Figure 5**, indicates that structures I and III are not present, as these would fragment to form ions at m/z 231, which were not observed. This was confirmed by MS⁴ analysis, **Figure 4C,D**, in which MS³ fragment ions at m/z 391 and 377 were selected as precursors. Here, the extremely low abundance of ions at m/z 231 confirms the absence of significant amounts of structures I and III, **Figure 5**. However, all ions predicted for structure II were observed, notably the high-abundance ion at m/z 217, which consists of the alditol residue with two unmethylated hydroxyl groups that were exposed by cleavage of glycosidic bonds during MS² and MS³, **Figure 4**.

When the DP5 oligoglycosyl alditols in Fraction 1 were analyzed by MSⁿ, virtually all of the alditol moieties were branched, **Figures S1 and S2**. ESI-MSⁿ data for the DP5 oligoglycosyl alditols also provide further insight into the extent to which Araf side chains can decorate the xylooligosacchrides produced by CtXyl5A. Notably, MS⁴ of the m/z 537 ion (derived from the alditol pentasaccharide) generates an m/z 363 Y-ion that yields an m/z 217 ion at MS⁵. As shown in the schematic, **Figure S2**,

these species can only be generated if the xylosyl alditol and the adjacent Xylp are both branched. The detection of a m/z 377 ion at MS³, however, demonstrates that the structure Xylp-Xylp-(Araf)-Xylol is also present. Fragmentation of DP3 oligoglycosyl alditols yields an m/z 217 Y-ion at MS³, while only trace amounts of a m/z 231 ion were evident, **Figures S3 and S4**. This again demonstrates that the xylosyl alditol contains a branch and thus the structure of the trisaccharide is predicted to be Xylp-(Araf)-Xylol.

The binding of CtXyl5A to arabinoxylan

The terminal reaction products produced by endo-acting glycoside hydrolases reflects an iterative process in which the products from initial hydrolytic reactions serve as substrates in subsequent rounds of catalysis. Analysis of the structure of the terminal reaction products (which are unable to be further hydrolysed) provides insight into the possible modes of substrate binding to both the negative and positive subsites (see below). The subsite nomenclature of glycoside hydrolases were defined previously by Davies and colleagues (31). Briefly, the scissile bond is positioned between subsites -1 and +1, and subsites that extend towards the non-reducing and reducing ends of the substrate are assigned increasing negative and positive numbers, respectively. The Xylp at the reducing and the non-reducing end of the oligosaccharide products are derived from substrate bound at the -1 and +1 subsites, respectively. As ~ 99 % of the reducing end Xylp residues contain an O3 Araf branch, it is evident that the arabinose decoration of the xylose bound at the -1 subsite is a key specificity determinant of the enzyme.

The detection of terminal *Xylp* (in which O2, O3 and O4 are methylated) and 3-linked *Xylp* residues, both of which occur at the non-reducing end of the oligosaccharide backbone, indicates that a *Xylp* with an *Araf* side chain at O3 can be accommodated in the +1 subsite of *CtXyl5A*, but a side chain in this position is not a specificity determinant. As both (*Araf*)-*Xylp*-(*Araf*)-*Xylol* and *Xylp*-*Xylp*-(*Araf*)-*Xylol* were identified in the tetrasaccharide, an O3-*Araf* side chain is present on some, but not all, of the *Xylp* residues bound in the -2 subsite. Thus, while an O3-*Araf* side chain can be accommodated at the -2 subsite, the arabinose decoration does not define enzyme specificity. The identification of *Xylp*-(*Araf*)-*Xylp*-(*Araf*)-*Xylol* in the pentasaccharide reaction products not only confirms that *Araf* can be present at the -1 and -2 subsites, but also demonstrates that the +2 and +3 (if it exists) subsites can accommodate *Xylp* residues bearing arabinose side chains. It should be noted, however, that *Xylp*-(*Araf*)-*Xylp*-(*Araf*)-*Xylp* is a potential substrate for the enzyme (binding from subsites -2 to +1), suggesting that this molecule is only hydrolyzed very slowly by the enzyme, possibly because it is unable to access the +2 subsites. This is consistent with the absence of *Xylp* or (*Araf*)-*Xylp* in the reaction products; xylose or decorated xylose can only be generated if the substrate is hydrolyzed when it occupies only +1 of the positive subsites of the enzyme. Thus, to summarize, subsites -2 to +2 of *CtXyl5A* can accommodate *Xylp* residues that contain an O3-*Araf* side chain, however, only at the -1 subsite does the arabinose decoration act as an essential specificity determinant.

***Ct*CBM6 specificity**

To investigate whether *Ct*CBM6 is a functional CBM, the capacity of *Ct*GH5-CBM6 to bind to various carbohydrates was assessed by ITC. The data showed that *Ct*GH5-CBM6 bound to celohexaose and cellobiose with similar affinity, **Table 2** (example titrations in **Figure 6**). By contrast, binding to glucose was too low to quantify. The protein also displayed affinity for the reaction products generated by *CtXyl5A* and for undecorated xylooligosaccharides. The protein did not appear to bind to various xylans or to β -1,3- β -1,4-glucans. This indicates that *Ct*GH5-CBM6 recognises the terminal region of these polysaccharides, as the concentration of ligand available to the protein in these polymers, which have DPs >300, would be very low and thus binding would not be detected. It is possible that the catalytic module, rather than CBM6, mediates binding to the xylo- and cello-oligosaccharides. To test this hypothesis, the ligand binding profile of variants of *Ct*GH5-CBM6, in which either Trp424 or Phe478 had been substituted with Ala, was assessed. As discussed below these two aromatic residues are highly conserved in the CBM6 family and comprise the primary binding site in this protein family (32). Both *Ct*GH5-CBM6:W424A and *Ct*GH5-CBM6:F478A, although catalytically active, **Table 1**, displayed no binding to the xylan- and cellulose-derived oligosaccharides, **Table 2**. It is evident, therefore, that the CBM6 component of *Ct*GH5-CBM6 mediates the observed binding to oligosaccharides.

Crystal structure of *Ct*GH5-CBM6

The structure of *Ct*GH5-CBM6 was solved by selenomethionine SAD and

the resulting structure used as a starting model for refinement against native data extending to 1.5 Å resolution, **Table S2** (PDB code 2y8k). The polypeptide chain is visible from Ser37 to Ile516.

CtGH5: As expected, the N-terminal CtGH5 module displays a $(\beta/\alpha)_8$ barrel architecture, although α -helix-8 points away from the barrel and towards CtCBM6 module (discussed below), **Figure 7**. GH5 enzymes are members of clan GH-A in which the two catalytic residues are invariant glutamates presented at the end of β -strands 4 and 7 (33,34). From the structure of CtGH5-CBM6, the catalytic acid-base is likely to be Glu171 (end of β -strand 4) and the catalytic nucleophile Glu279 (end of β -strand 7). The catalytic role of these two residues is confirmed by the observation that the mutants E171A and E279A are inactive, **Table 1**. A narrow V-shaped cleft, approximately 2 Å in length, extends along the full length of the protein and sits over the top of the β -barrel. The dimensions of the cleft, in the centre of which is the catalytic apparatus, suggest that the protein contains ~5 subsites extending from -3 to +2.

An analysis of structural homologues of the CtGH5 component of CtGH5-CBM6 by the DaliLit webserver (http://ekhidna.biocenter.helsinki.fi/dali_server) identified a large number of GH5 and Clan GH-A enzymes that displayed significant structural similarity to CtGH5. The *Pseudoalteromonas haloplanktis* cellulase Cel5G (PDB 1tvn) with a root mean square deviation (rmsd) of 2.8 Å over 253 C α atoms and a Z-score of 24.1, and the *Bacillus agaradhaerens* cellulase BaCel5A (PDB 1qi2) with an rmsd of 2.9 Å over 254 C α

atoms and a Z-score of 23.6, are representative, close structural homologs. The critical -1 subsite, where the transition state is formed, is similar in the arabinoxylanase and the GH5 cellulases. In addition to the two catalytic glutamates CtGH5 contains several key residues that have been identified as “strictly conserved” in family GH5 enzymes (35). These residues in the CtGH5 module, which superimpose with amino acids in the active site of BaCel5A (the cellulase residues are shown in parentheses) are as follows: Asn170 (Asn138), Glu171 (Glu139), Tyr255 (Tyr202), Glu279 (Glu228), and Phe310 (Trp262), **Figure 8A**. The catalytic acid-base, Glu171, makes hydrogen bonds with Asn139 and His253, and these interactions likely contribute to both the position and ionization state of this critical amino acid. Asn170 is highly conserved in clan GH-A glycoside hydrolases and plays an important role in transition state stabilization by making a hydrogen bond with the O2 of the sugar at the -1 subsite (36). The position of the catalytic nucleophile, Glu279, is stabilized through a hydrogen bond with Tyr255, whereas Phe310, based upon comparison with other related hydrolases, is likely to form the sugar-binding hydrophobic platform in subsite -1.

Despite numerous attempts, no structure of CtGH5-CBM6 in complex with its substrate or reaction products has been obtained, in part due to the preference of this protein to crystallize with the N-terminal residues of a symmetry related molecule positioned in the substrate binding cleft, and because co-crystallization experiments did not yield diffracting crystals. Consequently, it is difficult to define precisely the structural

basis for the unusual substrate specificity displayed by the arabinoxylanase. Superimposing *BaCel5A* in complex with 2-deoxy-2-fluoro celotriose with *CtGH5* provides some insight into the specificity displayed by the arabinoxylanase. As discussed above, the catalytic apparatus, the residues that interact with O2 and the endocyclic oxygen of the -1 sugar, and the hydrophobic platform are conserved in *CtGH5*, **Figure 8A**. It is evident, however, that the arabinoxylanase lacks the residues that, in other GH5 enzymes, hydrogen bond with O3 of the active site sugar. For instance, His101 and Tyr66 in *BaCel5A* hydrogen bond with O3 of the -1 Glc, whereas the equivalent residues in *CtGH5* are Gly134 and Cys95, respectively, **Figure 8A**. Indeed, in the -1 subsite of the arabinoxylanase there is a large pocket around the O3 of the superimposed Glc that could accommodate a sugar decoration such as *Araf*, **Figure 8B**. The pocket contains a tyrosine (Tyr92) that may make hydrophobic interactions with the arabinose, and several polar residues, Glu68, Asn135, Asn139 and Asn170 that could make polar contacts with the sugar. Based on the presence of glycerol and water molecules within this region of the enzyme, an *Araf* molecule was modeled into the pocket, **Figure 8C**.

CtCBM6: The structure of the *CtCBM6* module displays a β -sandwich fold typical of other family CBM6 members (32,37,38), **Figure 7**. The twisted pair of β -sheets, which can be viewed as forming an extended barrel, consist of five and four anti-parallel β -strands, respectively. The structure of *CtCBM6* shows strong similarity with numerous CBM6 members. The closest homolog is the CBM6 module (designated

CmCBM6) from the *Cellvibrio mixtus* lichenase *CmLic5A* (PDB 1uz0; rmsd 1.5 Å over 123 C α atoms and a Z-score of 18.1). The major binding site in the CBM6 family is in the loops connecting the two β -sheets. This region, referred to as site A (32,37), may comprise a pocket if terminal sugars are recognized (39), or a cleft for the binding of internal regions of polysaccharides (32). A central feature of site A is a pair of aromatic residues, which bind to the α and β face, respectively, of the terminal sugar (or central sugar in the case of xylan binding modules) and an asparagine, located at the base of the site that makes critical hydrogen bonds with O2, O3 or O4. Specificity is conferred by additional polar and hydrophobic interactions (37). Site A in *CtCBM6* displays a pocket-like topology and contains all the key ligand binding residues present in *CmCBM6* (40), **Figure 9**. The pair of aromatic residues in *CmCBM6*, Trp92 and Tyr33, which straddle the non-reducing, terminal sugar correspond to Phe478 and Trp424, respectively, in *CtCBM6*. Furthermore, Glu20 and Asn121 in *CmCBM6*, which make polar contacts with O3 and O4 of the non-reducing terminal sugar in cello- and xylooligosaccharides, superimpose with Glu411 and Asn507, respectively, in *CtCBM6*. Finally, the amide nitrogen of Tyr33 in *CmCBM6* makes a polar contact with O2 and O3 of the terminal sugar, a contact that is likely to be replicated by that of Trp424 in *CtCBM6*. The structural conservation between site A in *CtCBM6* and *CmCBM6* is consistent with the similar ligand specificities displayed by this binding site in the two proteins, **Table 2** and (39). Thus, both proteins bind to *xylo*- and *gluco*-configured oligosaccharides but do not display affinity for the

corresponding polysaccharides. Thus, the structural similarity between *CmCBM6* and *CtCBM6* is consistent with the view that the *Clostridium* module targets the terminal regions of oligosaccharides. In *CmCBM6* cellooligosaccharides can bind to site A in both orientations, consistent with the targeting of O1/O4, O2 and O3, but not the endocyclic oxygen or O6, which would adopt different positions in the two orientations. It is highly likely, therefore, that *CtCBM6* will also bind to xylo- and cellooligosaccharides in both orientations. Given that the key interactions with the ligand at site A is with the terminal sugar, it is perhaps surprising that *CtCBM6* does not display measurable binding to xylose or glucose. It is possible that the entropic cost of locking the sugar into a pyranose ring conformation may contribute to the weak binding, although it is also possible that the protein makes indirect, water-mediated interactions to the penultimate sugar in the oligosaccharides, as observed in *CmCBM6*-ligand complexes (40).

The linker connecting *CtGH5* with *CtCBM6*: *CtCBM6* is connected to *CtGH5* by a sequence extending from residues Gly336 to Thr373. This linker, which adopts a stable conformation based on its B-factor, makes numerous internal polar contacts and forms hydrogen bonds with β -strand 3 and the loop connecting β -strands 3 and 4 of *CtCBM6*, and α -helices 7 and 8 of *CtGH5*. Furthermore, the C-terminal region of α -helix 8 and the internal region of α -helix 7 make hydrogen bonds with β -strands 3 and 7 of *CtCBM6*. The polar contacts between the two modules are augmented by a large number of apolar interactions

mediated by the linker sequence. The resultant burial of a significant hydrophobic surface, at the interface between *CtGH5* and *CtCBM6*, likely explains why these two modules (or domains) do not fold independently, as occurs in other glycoside hydrolases that contain catalytic modules and CBMs (41). This view is consistent with the observation that *CtCBM6*, when expressed as a discrete entity (Thr373 to Ile516), does not bind to celohexoase or xylohexoase, and *CtGH5* (Asn32 to Thr373) exhibits very low catalytic activity and is considerably more thermolabile than *CtGH5*-CBM6, **Figure S5**.

DISCUSSION

This study reveals a *C. thermocellum* protein that displays arabinoxylanase activity, an activity not previously reported. The vast majority of xylanases are derived from GH10 and GH11 and target the β -1,4-D-xylose polymeric backbone. These enzymes do not generally distinguish between different xylans, although highly decorated forms of the polysaccharide, such as rye arabinoxylan, are poorly degraded as steric constraints restrict enzyme access (12). Indeed, the only other examples of xylanases that utilize side chains as essential specificity determinants are glucuronoxylan specific enzymes from GH30. These enzymes make critical interactions with the 4-O-methyl glucuronic acid (linked α -1,2 to the xylan backbone) that decorates the xylose at the -2 subsite (42). *CtXyl5A* is highly unusual in that its essential Ara f decoration is attached to the xylose positioned in the active site. The only other example of an active site side chain specificity determinant is the α -1,6-Xyl p that decorates the -1 Glc in the

xyloglucan cellobiohydrolase, OXG-RCBH, from *Geotrichum* sp. (43).

The function of CtXyl5A within the context of *C. thermocellum*, which has the genetic capacity to recruit 72 different enzymes into the cellulosomes (44), including seven GH10 and GH11 xylanases, is intriguing. It is likely that the GH10 and GH11 enzymes target xylans that are sparsely decorated with arabinose side chains. By contrast, CtXyl5A most likely hydrolyses xylans where tandem Xylps contain Araf decorations. The recognition of the termini of *xylo*- and *gluco*- configured polymers by CtCBM6, suggests that the arabinoxylanase is targeted to regions of the plant cell wall that is undergoing degradation and is therefore accessible to enzyme attack. Although the primary function of CBMs is to bring their cognate enzymes into close contact with appropriate substrates (45), there is increasing evidence that a subset of these modules, from CBM families 6, 9 and 35, target the termini of polysaccharides and thus may play a similar function to CtCBM6 (37,46,47). In conclusion, CtXyl5A displays a specificity that is complementary to endoxylanases from GH10, GH11 and GH30. As such the enzyme will make a contribution to the toolbox of biocatalysts required to

degrade plant cell walls to their constituent sugars, which can then be used in the biofuel and bioprocessing industries.

Acknowledgments

The European Synchrotron Radiation Facility, the Diamond Light Source, and the staff at both these facilities are thanked for help with data collection. This work was supported by grant PTDC/BIAPRO/69732/2006 from Fundação para a Ciência e Tecnologia (Portugal) and by the individual grant SFRH/BD/16731/2004 (to M.C.) from the same institution. The authors would also like to acknowledge funding provided by The Bioenergy Science Center (BESC), a U.S. Department of Energy Bioenergy Research Center supported by the Office of Biological and Environmental Research in the DOE Office of Science, and DOE Grant no. DE-FG02-96ER20220. We also thank the DOE-funded Center for Plant and Microbial Complex Carbohydrates (grant no. DE-FG02-93ER20097) for supporting critical infrastructure and analytical instrumentation required for this research.

REFERENCES

1. Brett, C. T., and Waldren, K. (1996) *Physiology and Biochemistry of Plant Cell Walls. Topics in Plant Functional Biology*, Chapman and Hall, London
2. Himmel, M. E., and Bayer, E. A. (2009) *Curr Opin Biotechnol* **20**, 316-317
3. Ragauskas, A. J., Williams, C. K., Davison, B. H., Britovsek, G., Cairney, J., Eckert, C. A., Frederick, W. J., Jr., Hallett, J. P., Leak, D. J., Liotta, C. L., Mielenz, J. R., Murphy, R., Templer, R., and Tschaplinski, T. (2006) *Science* **311**, 484-489
4. Ebringerova, A., Hromadkova, Z., and Heinze, T. (2005) *Polysaccharides 1: Structure, Characterization and Use* **186**, 1-67

5. DeBoy, R. T., Mongodin, E. F., Fouts, D. E., Tailford, L. E., Khouri, H., Emerson, J. B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H. J., and Nelson, K. E. (2008) *J Bacteriol* **190**, 5455-5463
6. Weiner, R. M., Taylor, L. E., 2nd, Henrissat, B., Hauser, L., Land, M., Coutinho, P. M., Rancurel, C., Saunders, E. H., Longmire, A. G., Zhang, H., Bayer, E. A., Gilbert, H. J., Larimer, F., Zhulin, I. B., Ekborg, N. A., Lamed, R., Richardson, P. M., Borovok, I., and Hutcheson, S. (2008) *PLoS Genet* **4**, e1000087
7. Berg Miller, M. E., Antonopoulos, D. A., Rincon, M. T., Band, M., Bari, A., Akraiko, T., Hernandez, A., Thimmapuram, J., Henrissat, B., Coutinho, P. M., Borovok, I., Jindou, S., Lamed, R., Flint, H. J., Bayer, E. A., and White, B. A. (2009) *PLoS One* **4**, e6650
8. Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V., and Gordon, J. I. (2003) *Science* **299**, 2074-2076
9. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) *Nucleic Acids Res* **37**, D233-238
10. Gilbert, H. J. (2010) *Plant Physiol* **153**, 444-455
11. Gilbert, H. J., Stalbrand, H., and Brumer, H. (2008) *Curr Opin Plant Biol* **11**, 338-348
12. Pell, G., Taylor, E. J., Gloster, T. M., Turkenburg, J. P., Fontes, C. M., Ferreira, L. M., Nagy, T., Clark, S. J., Davies, G. J., and Gilbert, H. J. (2004) *J Biol Chem* **279**, 9597-9605
13. Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabo, L., McKie, V. A., Gilbert, H. J., and Davies, G. J. (2002) *Proc. Natl. Acad. Sci. U S A* **99**, 14077-14082
14. Miller, G. L. (1959) *Analytical Chemistry* **31**, 426-428
15. Charnock, S. J., Lakey, J. H., Virden, R., Hughes, N., Sinnott, M. L., Hazlewood, G. P., Pickersgill, R., and Gilbert, H. J. (1997) *J Biol Chem* **272**, 2942-2951
16. Proctor, M. R., Taylor, E. J., Nurizzo, D., Turkenburg, J. P., Lloyd, R. M., Vardakou, M., Davies, G. J., and Gilbert, H. J. (2005) *Proc Natl Acad Sci U S A* **102**, 2697-2702
17. Ciucanu, I., and Kerek, F. (1984) *Carbohydr. Res.* **131**, 209-217
18. Carpita, N. C., and Shea, E. M. (1989). in *Analysis of Carbohydrates by GLC and MS* (Biermann, C. J., and McGinnis, G. D. eds.), CRC Press, Boca Raton, FL, USA. pp 157-216
19. Mazumder, K., and York, W. S. *Carbohydr Res* **345**, 2183-2193
20. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J., and Fontes, C. M. (2000) *Biochemistry* **39**, 5013-5021
21. Leslie, A. G. W. (1992) *Joint CCP4 + ESF-EAMCB Newslett. Protein Crystallogr.* **26**
22. Evans, P. (1993) Data reduction, *CCP4 Daresbury Study Weekend: Data Collection and Processing, DL/SCI/R34.*, in *Daresbury Laboratory*, Warrington, UK pp 114-122
23. Sheldrick, G. M. (1990) *Acta Crystallographica Section A* **46**, 467-473
24. Lamzin, V. S., and Wilson, K. S. (1993) *Acta Crystallographica Section D-Biological Crystallography* **49**, 129-147
25. Emsley, P., and Cowtan, K. (2004) *Acta Crystallogr. Biol.* **D60**, 2126-2132

26. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) *Acta Crystallogr. Biol.* **D53**, 240-255
27. Beylot, M. H., McKie, V. A., Voragen, A. G., Doeswijk-Voragen, C. H., and Gilbert, H. J. (2001) *Biochem J* **358**, 607-614
28. Gruppen, H., Hoffmann, R. A., Kormelink, F. J., Voragen, A. G., Kamerling, J. P., and Vliegthart, J. F. (1992) *Carbohydr Res* **233**, 45-64
29. Hoffmann, R. A., Leeftang, B. R., de Barse, M. M., Kamerling, J. P., and Vliegthart, J. F. (1991) *Carbohydr Res* **221**, 63-81
30. Domon, B., and Costello, C. E. (1988) *Biochemistry* **27**, 1534-1543
31. Davies, G. J., Wilson, K. S., and Henrissat, B. (1997) *Biochem J* **321 (Pt 2)**, 557-559
32. Czjzek, M., Bolam, D. N., Mosbah, A., Allouch, J., Fontes, C. M., Ferreira, L. M., Bornet, O., Zamboni, V., Darbon, H., Smith, N. L., Black, G. W., Henrissat, B., and Gilbert, H. J. (2001) *J. Biol. Chem.* **276**, 48580-48587
33. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J. P., and Davies, G. (1995) *Proc Natl Acad Sci U S A* **92**, 7090-7094
34. Jenkins, J., Lo Leggio, L., Harris, G., and Pickersgill, R. (1995) *FEBS Lett* **362**, 281-285
35. Hilge, M., Gloor, S. M., Rypniewski, W., Sauer, O., Heightman, T. D., Zimmermann, W., Winterhalter, K., and Piontek, K. (1998) *Structure* **6**, 1433-1444
36. Williams, S. J., Notenboom, V., Wicki, J., Rose, D. R., and Withers, S. G. (2000) *Journal of the American Chemical Society* **122**, 4229-4230
37. Abbott, D. W., Ficko-Blean, E., van Bueren, A. L., Rogowski, A., Cartmell, A., Coutinho, P. M., Henrissat, B., Gilbert, H. J., and Boraston, A. B. (2009) *Biochemistry* **48**, 10395-10404
38. van Bueren, A. L., Morland, C., Gilbert, H. J., and Boraston, A. B. (2005) *J. Biol. Chem.* **280**, 530-537
39. Henshaw, J. L., Bolam, D. N., Pires, V. M., Czjzek, M., Henrissat, B., Ferreira, L. M., Fontes, C. M., and Gilbert, H. J. (2004) *J. Biol. Chem.* **279**, 21552-21559
40. Pires, V. M., Henshaw, J. L., Prates, J. A., Bolam, D. N., Ferreira, L. M., Fontes, C. M., Henrissat, B., Planas, A., Gilbert, H. J., and Czjzek, M. (2004) *J Biol Chem* **279**, 21560-21568
41. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) *Biochem J* **382**, 769-781
42. Vrsanska, M., Kolenova, K., Puchart, V., and Biely, P. (2007) *FEBS J* **274**, 1666-1677
43. Yaoi, K., Kondo, H., Hiyoshi, A., Noro, N., Sugimoto, H., Tsuda, S., Mitsuishi, Y., and Miyazaki, K. (2007) *J Mol Biol* **370**, 53-62
44. Fontes, C. M., and Gilbert, H. J. (2010) *Annu Rev Biochem* **79**, 655-681
45. Bolam, D. N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M. P., Rixon, J. E., Boraston, A., Hazlewood, G. P., and Gilbert, H. J. (1998) *Biochem. J.* **331**, 775-781
46. Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A., and Kilburn, D. G. (2001) *Biochemistry* **40**, 6240-6247

47. Montanier, C., van Bueren, A. L., Dumon, C., Flint, J. E., Correia, M. A., Prates, J. A., Firbank, S. J., Lewis, R. J., Grondin, G. G., Ghinet, M. G., Gloster, T. M., Herve, C., Knox, J. P., Talbot, B. G., Turkenburg, J. P., Kerovuo, J., Brzezinski, R., Fontes, C. M., Davies, G. J., Boraston, A. B., and Gilbert, H. J. (2009) *Proc Natl Acad Sci U S A* **106**, 3065-3070

Table 1 Catalytic activity of CtXyl5A and its variants

Proteins	K_{cat}/K_M ($\text{min}^{-1} \text{mg}^{-1} \text{ml}$)		
	Rye Arabinoxylan	Wheat Arabinoxylan	Oat spelt xylan
<i>CtXyl5A</i>	1322 ± 357	808 ± 76	15 ± 1.2
<i>CtXyl5A</i> + 2 mM CaCl ₂	2271 ± 274	1658 ± 56	ND ^a
<i>CtXyl5A</i> + 5 mM EDTA	858 ± 102	343 ± 32	ND
<i>CtGH5-CBM6</i>	1012 ± 83	652 ± 21	14 ± 2.2357
W424A (<i>CtGH5-CBM6</i>)	1656 ± 173	728 ± 127	ND
F478A (<i>CtGH5-CBM6</i>)	983 ± 52	713 ± 66	ND
E279A (<i>CtGH5-CBM6</i>)	NA ^b	NA ± 357	ND
E171A (<i>CtGH5-CBM6</i>)	NA	NA	ND
<i>CtGH5</i>	1.1 ± 0.45	0.6 ± 0.07	NA

The enzymes were assayed at 60 °C in 50 mM sodium phosphate buffer, pH 7.0, containing substrate at a concentration of 1 mg ml⁻¹. The reaction was monitored by the release of reducing sugar (14). The catalytic rate could be used to determine K_{cat}/K_M as the substrate concentration was $\ll K_M$ (the rate of reaction was directly proportion to substrate concentration up to 2 mg ml⁻¹). Note that *CtXyl5A* is the full length enzyme while *CtGH5-CBM6* and *CtGH5* are derivatives of the enzymes containing the catalytic module appended to the CBM6 and the catalytic module, respectively.

^aND: Not determined.

^bNA: No activity detected.

Table 2 Binding of CtXyl5A derivatives to polysaccharides and oligosaccharides

<i>CtXyl5A</i> derivative	Ligand	$K_a \times 10^3$ (M^{-1})	ΔG (kcal mol ⁻¹)	ΔH (kcal mol ⁻¹)	$T\Delta S$ (kcal mol ⁻¹)	n
<i>CtGH5-CBM6</i>	Cellobiose	13.7±1.0	-5.6	-3.8±0.11	1.8	1.03±0.13
<i>CtGH5-CBM6</i>	Cellotriose	17.4±3.8	-5.8	-4.7±0.75	1.1	1.02±0.13
<i>CtGH5-CBM6</i>	Cellohexaose	21.0±4.9	-5.9	-3.2±0.44	2.7	1.12±0.12
<i>CtGH5-CBM6</i>	Xylotetraose	3.4±0.2	-4.8	-1.2±0.13	3.6	1.24±0.15
<i>CtGH5-CBM6</i>	Xylotriase	2.5±0.6	-4.6	-1.3±0.03	3.3	1.26±0.13
<i>CtGH5-CBM6</i>	Xylobiose	2.4±0.4	-4.6	-1.2±0.06	3.4	1.1±0.10
<i>CtGH5-CBM6</i>	WAX ^a _ treated with <i>CtXyl5A</i>	9.1±2.1	-5.4	-2.4±0.5	3.0	1.01±0.12
<i>CtGH5-CBM6</i>	WAX treated with <i>CtXyl5A</i> and <i>CjAbf51A</i> ^b	15.1±2.3	-5.7	-4.9±0.45	0.8	1.08±0.06
<i>CtGH5-CBM6</i> :E279A	Cellohexaose	13.4±2.1	-5.6	-4.2±0.98	1.4	1.12±0.14
<i>CtGH5-CBM6</i> :E279A	WAX treated with <i>CjXyn10A</i> ^c	14.2±1.2	-5.7	-3.6±0.83	2.1	1.12±0.12
<i>CtGH5CBM6</i> E279A	No binding to WAX					
<i>CtGH5-CBM6</i> :W424A	No binding to cellohexaose, xyloetraose, xylotriase, WAX treated with <i>CjXyn10A</i> , <i>CtXyl5A</i> or both <i>CtXyl5A</i> and <i>CjAbf51A</i> .					
<i>CtGH5-CBM6</i> :F478A	No binding to cellohexaose, WAX treated with <i>CjXyn10A</i> , <i>CtXyl5A</i> or both <i>CtXyl5A</i> and <i>CjAbf51A</i> .					
<i>CtCBM6</i>	No binding to cellohexaose, xyloetraose, xylotriase, xylobiose, WAX treated with <i>CtXyn5A</i> or <i>CjXyn10A</i>					

The binding of derivatives of *CtXyl5A* to ligands was measured by ITC. The protein was at the 145 μ M in the cell and polysaccharide (3-5 mg/ml) or oligosaccharide (5-15 mM) was in the syringe. ITC was carried out in 50 mM Na/HEPES buffer, pH 7.5, at 25 °C. The concentration of the oligosaccharides generated by the digestion of wheat arabinoxylan (WAX) was fitted to give an n value close to 1.

^aWAX; wheat arabinoxylan

^b*CjXyn10A*; GH10 xylanase from *Cellvibrio japonicus*

^c*CjAbf51A*; GH51 arabinofuranosidase from *Cellvibrio japonicus*

FIGURE LEGENDS

Figure 1 Schematic of xylan

Figure 2 Analysis of the reaction products generated by CtXyl5A from arabinoxylan
Rye arabinoxylan was incubated with CtXyl5A until the reaction was complete and the products purified by size exclusion chromatography. Fraction 1 contained the most abundant oligosaccharides. *Panel A* shows the MALDI-TOF MS analysis of permethylated and NaBH₄ reduced oligosaccharides in Fraction 1. Molecules that contained exclusively pentaose sugars are labelled Pen with the DP in subscript. *Panel B* shows GC-EIMS analysis of Fraction 1. All the hydroxyls of T-Araf and T-Xylp are methylated, while 4-Xylp 3,4-Xylp and 3-Xylp signify the positions of the hydroxyls that are not methylated and were thus involved in a linkage prior to TFA cleavage. *Panel C* shows HPAEC analysis of Fraction 1 treated with the arabinofuranosidase CjAbf51A. Peaks X2, X3 and X4 co-migrate with xylobiose, xylotriose and xylo-tetraose.

Figure 3 NMR analysis of the oligosaccharides generated by CtXyl5A

Panel A depicts a partial HSQC spectrum of Fraction 1 showing upfield shifts of reducing Xylp residues. The arrow indicates the barely detectable H1 resonance of (unbranched) reducing 4-linked residues. *Panel B* is a partial ROESY spectrum of Fraction 1 showing interglycosidic dipolar contacts between the Araf H1 and the reducing Xylp H3 resonances.

Figure 4 ESI-MS of the tetrasaccharides in Fraction 1

The tetrasaccharides in Fraction 1, which contains the most abundant products, were analyzed by ESI-MSⁿ. *Panel A* shows the fragmentation of the *m/z* 725 ion which comprises the tetrasaccharides. *Panel B* shows the fragmentation of the *m/z* 551 ion derived from the *m/z* 725 ion in *Panel A*. *Panel C* and *Panel D* depict the fragmentation pattern of the *m/z* 391 and *m/z* 377 ions, respectively, derived from the *m/z* 551 ion generated in *Panel B*. The masses of Y-ions are indicated unless otherwise stated.

Figure 5 The structure of the tetrasaccharides generated by CtXyl5A

Based on the data displayed in **Figure 4**, the structures of the tetrasaccharides in Fraction 1 were identified. The sugars labelled P can be Araf or Xylp. The data showed that the oligosaccharide ions coloured green were present, while those coloured red were not evident. The solid arrows between oligosaccharides showed the conversion of one oligosaccharide into another, through ESI-MS fragmentation. Dotted arrows between oligosaccharides identified theoretical ESI-MS-mediated oligosaccharide conversions that did not occur in these analyses. The dotted arrow between sugar linkages within the oligosaccharides shows the fragmentation site and the ion identified. Arrows pointing at sugars (but did not link two sugars together) identified hydroxyl groups that were not methylated as they comprised a glycosidic linkage in a parental ion. Xylol is the reducing end xylose that has been reduced to its alditol form by NaBH₄.

Figure 6 Representative ITC data of CtGH5-CBM6 to oligosaccharides

The ligands (10 mM arabinose) in the syringe was titrated into CtGH5-CBM6 (100 μ M) in the cell. The top half of each panel shows the raw ITC heats; the bottom half, the integrated peak areas fitted using a one single binding model by MicroCal Origin software. ITC was carried out in 50 mM Na/HEPES, pH 7.5 at 37 °C.

Figure 7 Crystal structure of CtGH5-CBM6

In the protein cartoon of CtGH5-CBM6 both modules are colour ramped from N-terminus (blue) to C-terminus (red). The loop connecting the two modules is coloured magenta. The two catalytic residues (Glu171 and Glu279) in CtGH5 and the two aromatic amino acids that are conserved in the ligand binding site of family 6 CBMs (Trp424, Phe478) are shown in stick format. The figure, and the other structural figures, was drawn with PyMol (DeLano Scientific; <http://pymol.sourceforge.net/>).

Figure 8 Superimposition of CtGH5 and the cellulase BaCel5A

Panel A shows the superimposition of the residues in the active site (-1 subsite) of BaCel5A (PDB 1qi2; coloured green), which interact with the substrate, with the equivalent amino acids (coloured yellow) in CtGH5. *Panel B* shows the solvent accessible surface of CtGH5 in which 2-deoxy-2-fluoro-cellobiose, derived from BaCel5A, has been superimposed. *Panel C* depicts a model of xylotriose, containing Ara_f appended to O3 of Xyl_p-1, bound to CtGH5. The tetrasaccharide ligand is modelled on the superimposed structure of 2-deoxy-2-fluoro-cellobiose and the glycerol and water molecules in the putative arabinose binding pocket. In *Panel A* and *B* bound ligand is coloured silver (carbons), while the Xyl_p and Ara_f residues in *Panel C* are coloured salmon pink and blue (carbons), respectively.

Figure 9 Superimposition of CtCBM6 and CmCBM6

Panel A shows the superimposition of the residues in the ligand binding site of CmCBM6 (PDB 1uz0; coloured green) with the equivalent amino acids (coloured yellow) in CtCBM6. *Panel B* show the solvent accessible surface of CtCBM6 in complex with cellobiose (superimposed from CmCBM6). Amino acids whose side chains are predicted to contribute to ligand recognition are coloured magenta. In both panels ligand is shown in silver (carbon) stick representation.

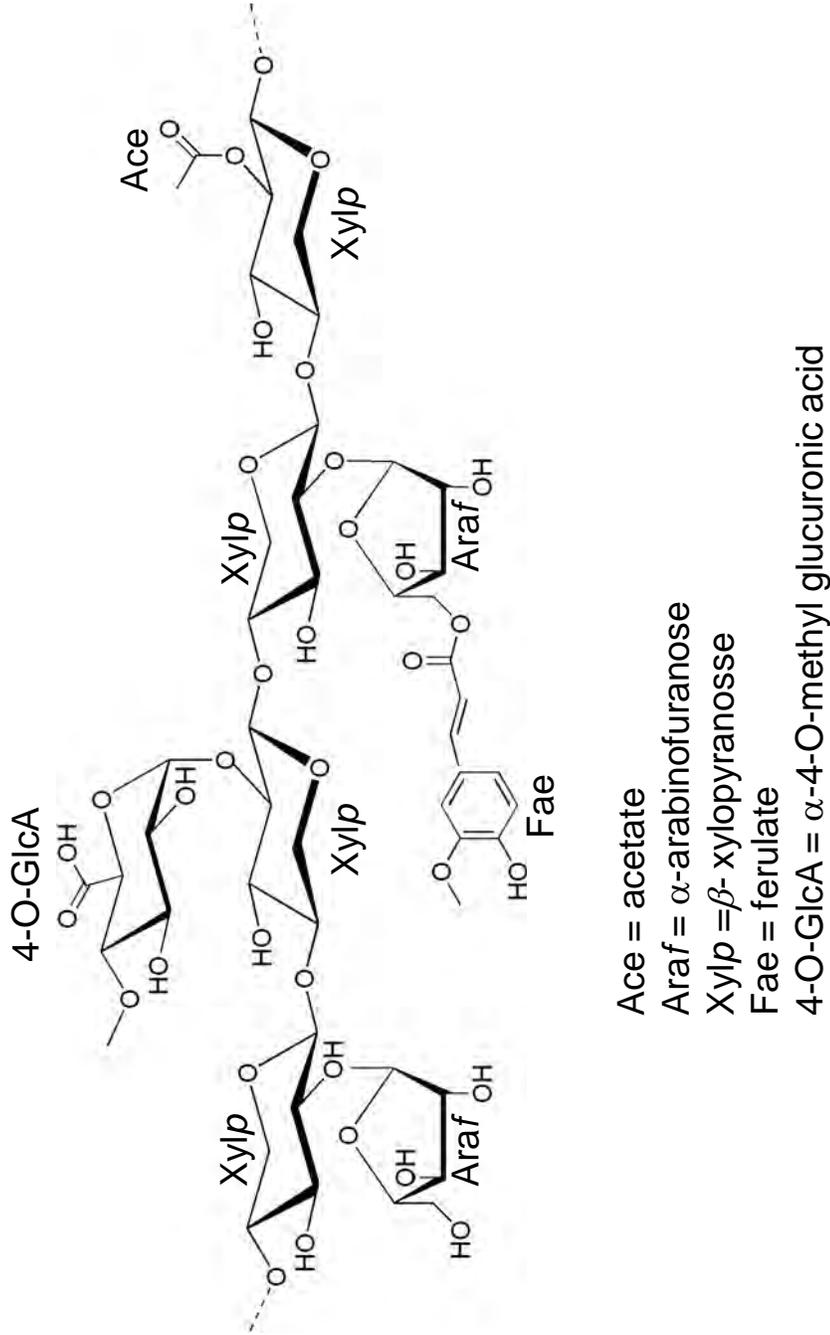


Figure 1

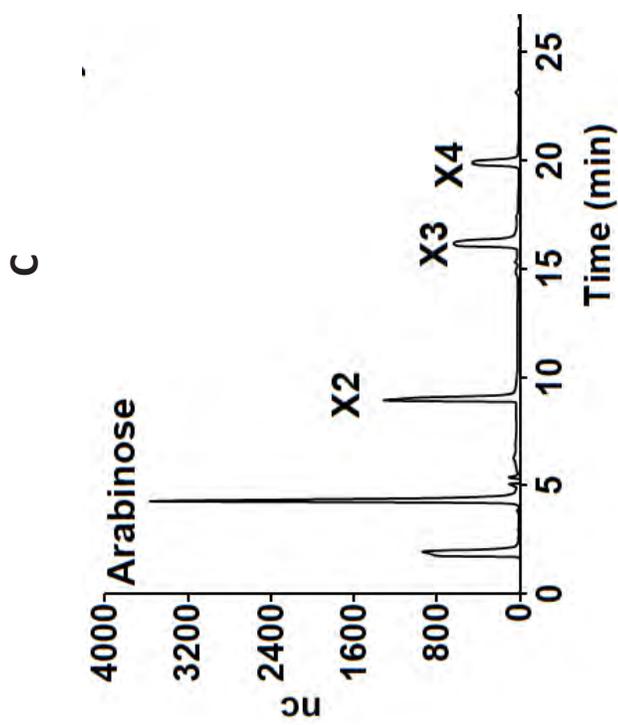
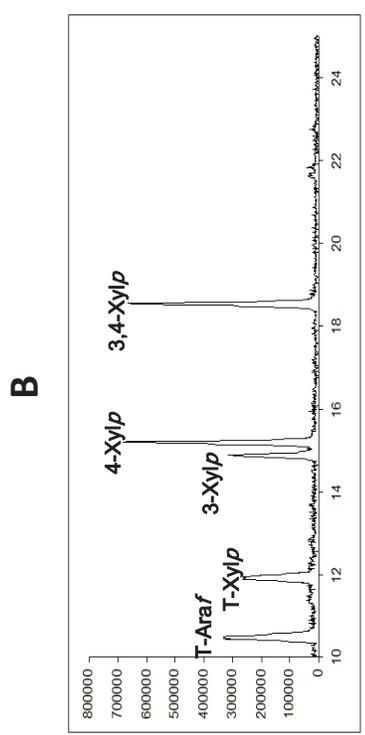
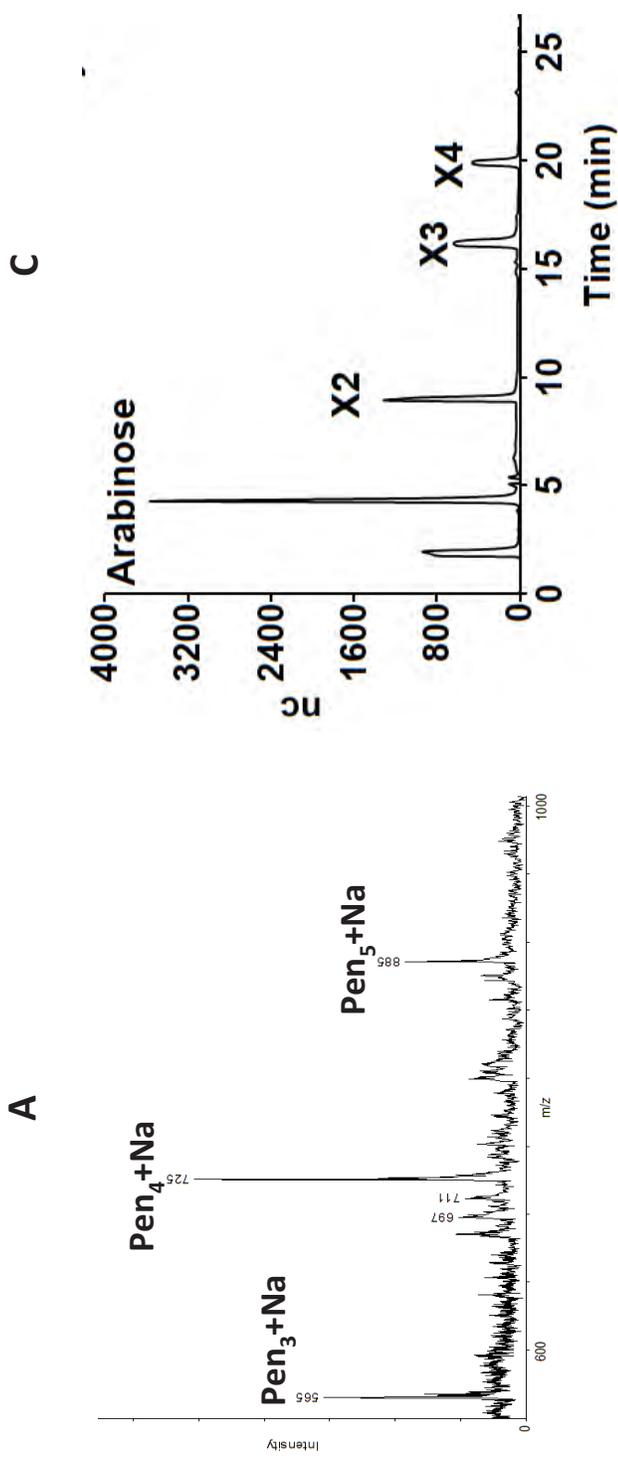


Figure 2

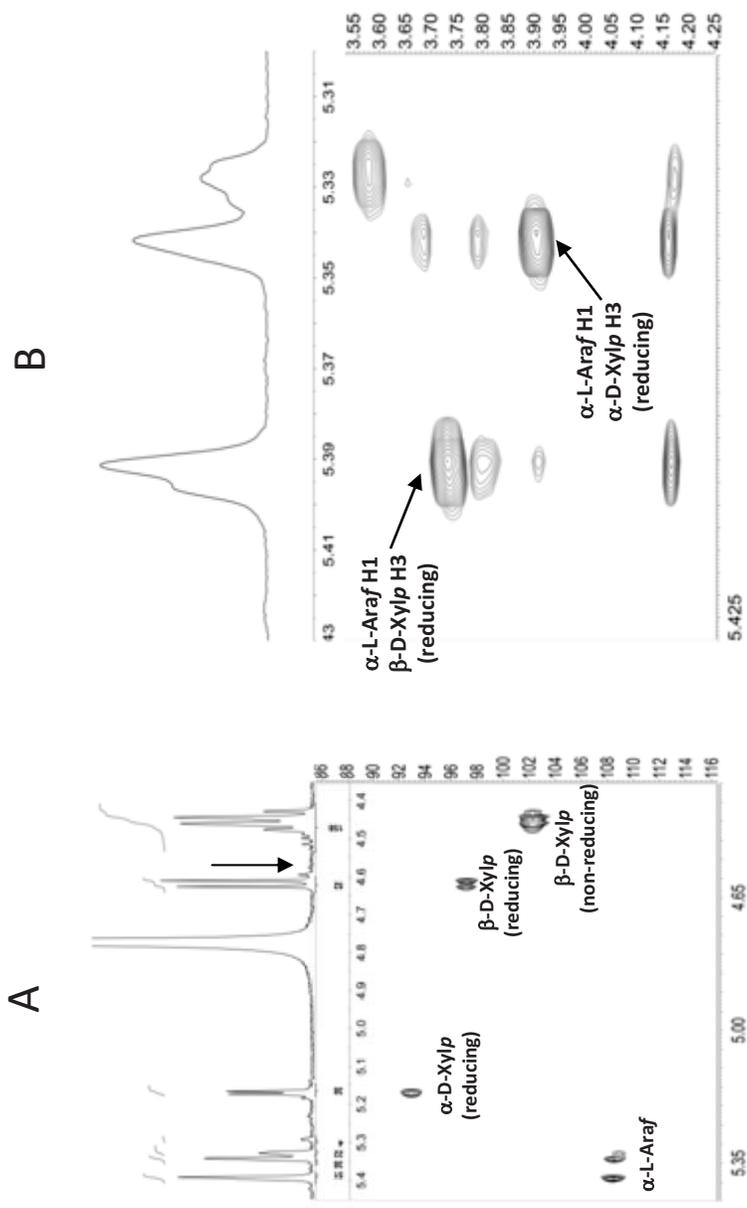


Figure 3

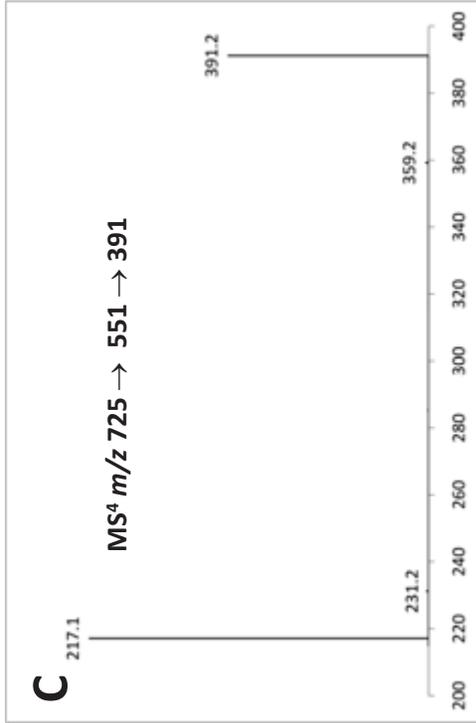
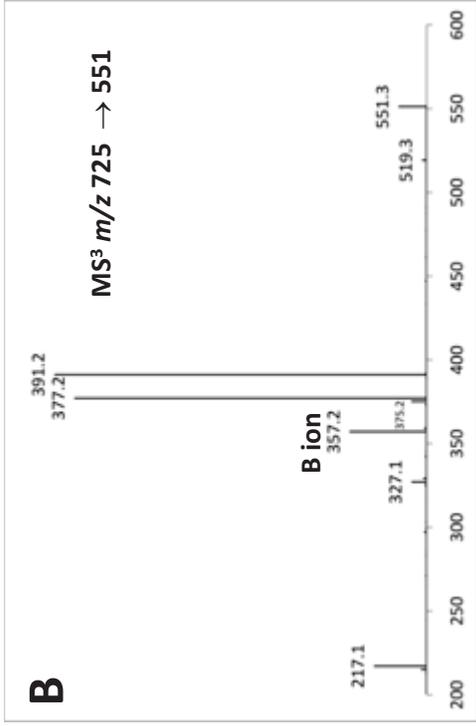
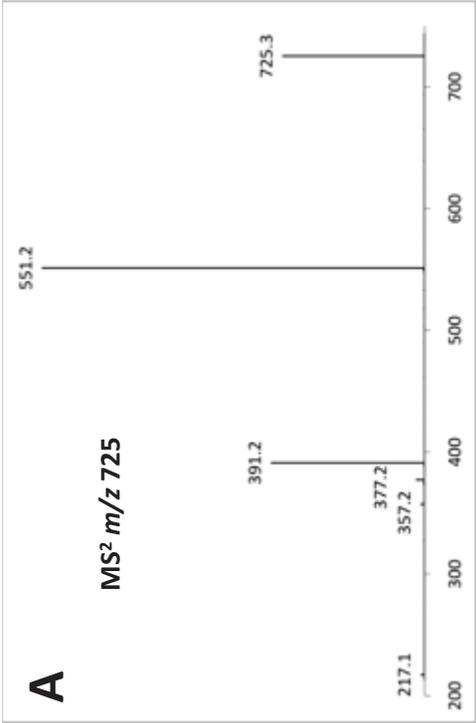


Figure 4

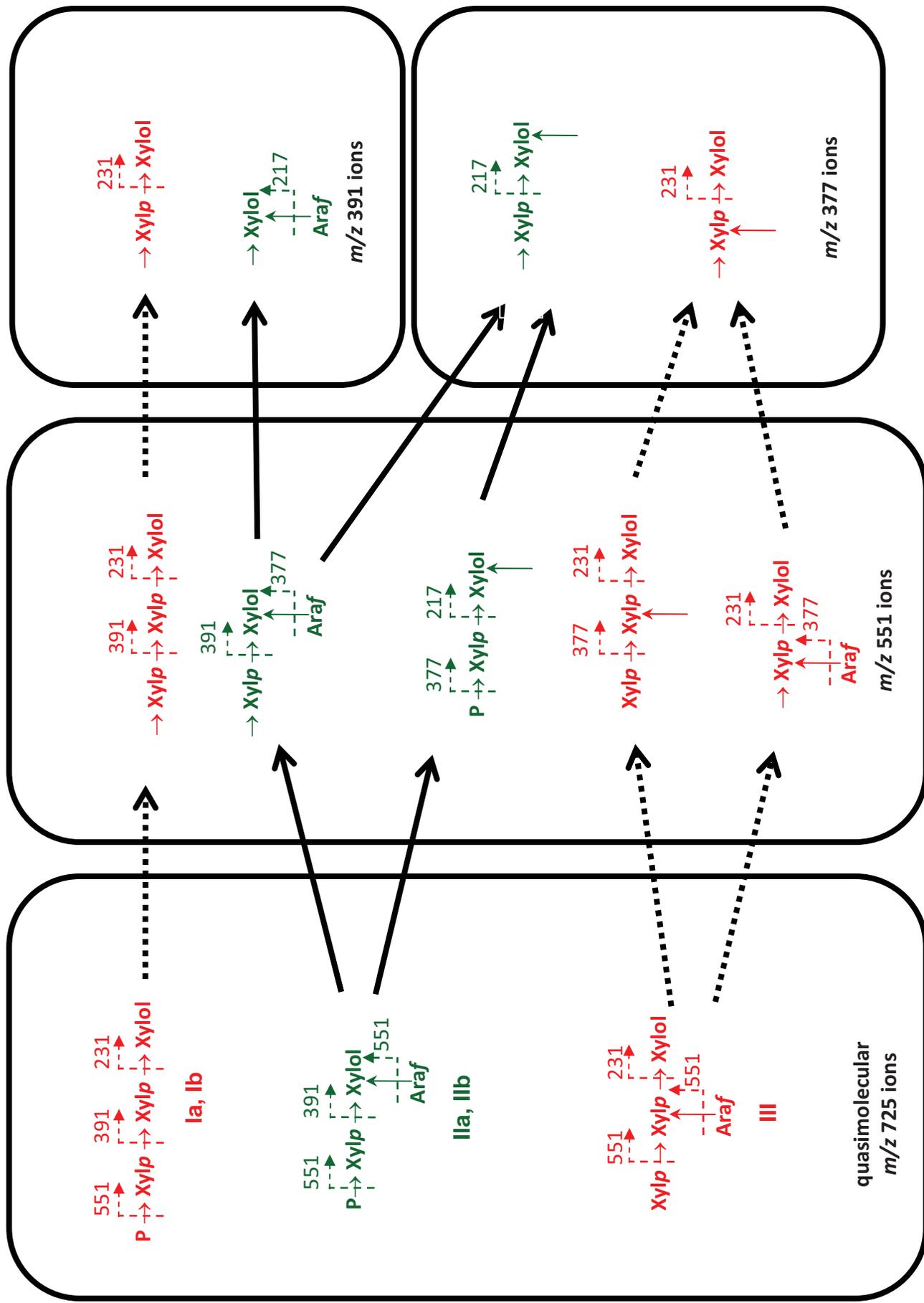


Figure 5

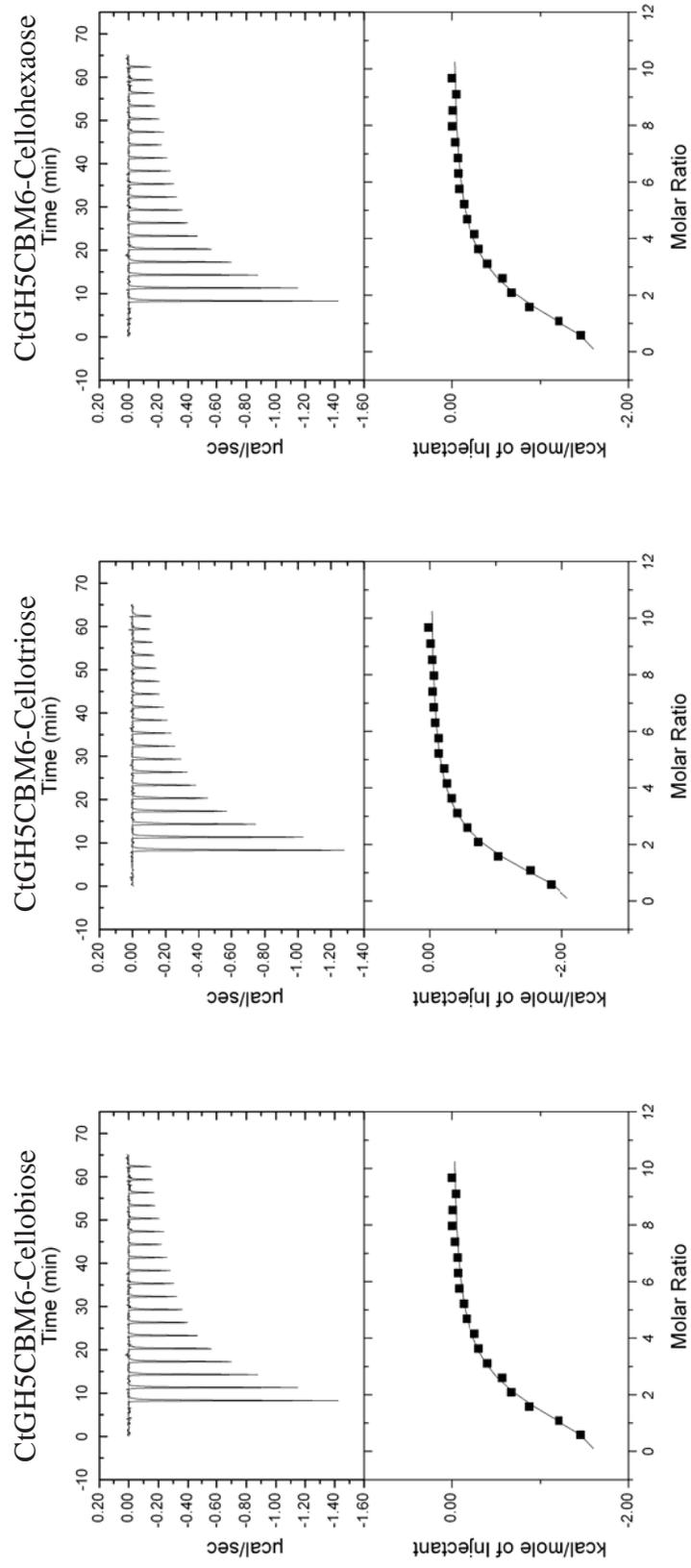


Figure 6

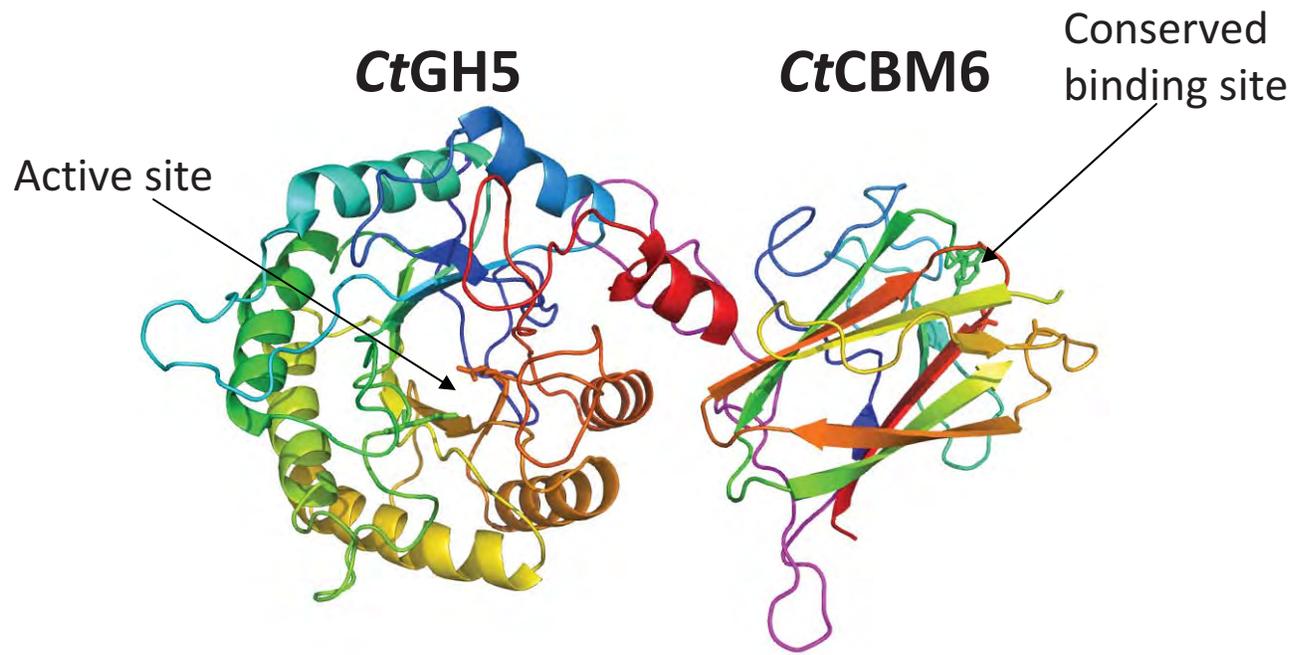
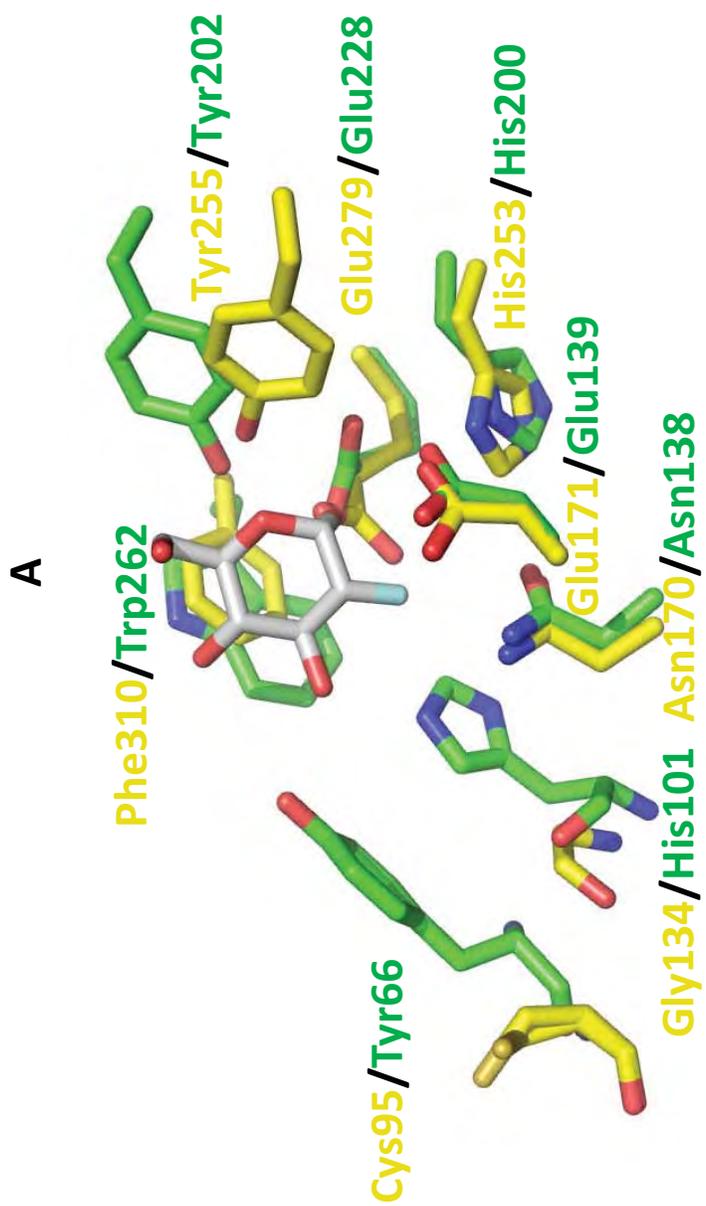
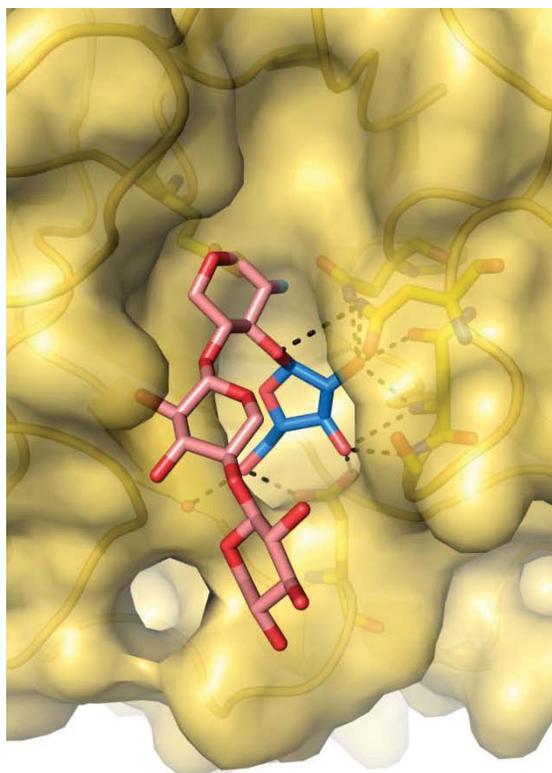


Figure 7



C



B

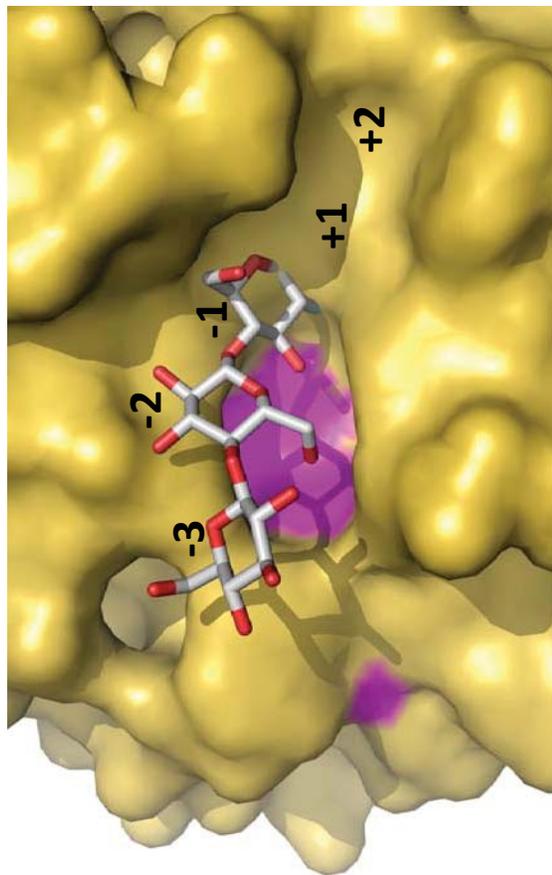


Figure 8

