



HAL
open science

Structural and enzymatic characterisation of a Glycoside Hydrolase Family 31 α -xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification

Johan Larsbrink, Atsushi Izumi, Farid Ibatullin, Azadeh Nakhai, Harry J Gilbert, Gideon J Davies, Harry Brumer

► To cite this version:

Johan Larsbrink, Atsushi Izumi, Farid Ibatullin, Azadeh Nakhai, Harry J Gilbert, et al.. Structural and enzymatic characterisation of a Glycoside Hydrolase Family 31 α -xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification. *Biochemical Journal*, 2011, 436 (3), pp.567-580. 10.1042/BJ20110299 . hal-00596265

HAL Id: hal-00596265

<https://hal.science/hal-00596265>

Submitted on 27 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Structural and enzymatic characterisation of a Glycoside Hydrolase Family 31 α -xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification

Johan Larsbrink^{*§}, Atsushi Izumi^{†§}, Farid Ibatullin^{*≈}, Azadeh Nakhai^{*}, Harry J. Gilbert[‡], Gideon J. Davies[†], Harry Brumer^{*}

^{*}Division of Glycoscience, School of Biotechnology, Royal Institute of Technology (KTH), AlbaNova University Centre, 106 91 Stockholm, Sweden.

[†]York Structural Biology Laboratory, Department of Chemistry, The University of York, York YO10 5DD. United Kingdom

[‡]Institute for Cell and Molecular Biosciences, Newcastle University, The Medical School, Framlington Place, Newcastle-upon-Tyne, NE2 4HH, United Kingdom

[§]These authors contributed equally to this study.

[≈] Present address: Petersburg Nuclear Physics Institute, Russian Academy of Science, Molecular and Radiation Biology Division, Gatchina, St. Petersburg 188300, Russia

Corresponding author: Harry Brumer, e-mail harry@biotech.kth.se, phone: +46 8 5537 8367, fax: +46 8 5537 8468

Running title: *Cellvibrio japonicus* CjXyl31A structure/function

Keywords: xyloglucan, plant cell wall, enzymology, hemicellulose, saccharification

Synopsis

The desire for improved methods of biomass conversion to fuels and feedstocks has re-awakened interest in the enzymology of plant cell wall degradation. The complex polysaccharide xyloglucan is abundant in plant matter, where it may account for up to 20% of the total primary cell wall carbohydrates. Despite this, few studies have focussed on xyloglucan saccharification, which requires a consortium of enzymes including *endo*-xyloglucanases, α -xylosidases, β -galactosidases and α -L-fucosidases, among others. Here, we present the characterisation of Xyl31A, a key α -xylosidase in xyloglucan utilisation by the model Gram-negative soil saprophyte *Cellvibrio japonicus*. *Cj*Xyl31A exhibits high regiospecificity for the hydrolysis xylogluco-oligosaccharides, with a particular preference for longer substrates. Crystallographic structures of both the *apo* enzyme and the trapped covalent 5-fluoro- β -xylosyl-enzyme intermediate, together with docking studies with the XXXG heptasaccharide, revealed, for the first time in glycoside hydrolase family 31, the importance of a PA14 domain insert in the recognition of longer oligosaccharides by extension of the active site pocket. The observation that *Cj*Xyl31A was localised to the outer membrane provided support for a biological model of xyloglucan utilisation by *C. japonicus*, in which xylogluco-oligosaccharides generated by the action of a secreted *endo*-xyloglucanase are ultimately degraded in close proximity to the cell surface. Moreover, this study diversifies the toolbox of glycosidases for the specific modification and saccharification of cell wall polymers for biotechnological applications.

Introduction

The enzymatic degradation of cellulose and matrix glycans of the plant cell wall into simple sugars provides energy for organisms from all kingdoms of life, and as such, represents a key aspect of the global carbon cycle [1]. This process is complex, with an implicit requirement for an enormous diversity of monosaccharide- and linkage-specific glycoside hydrolases and carbohydrate lyases [2]. With the rising desire to convert the vast amounts of plant polysaccharides available to liquid fuels and value-added products, significant attention continues to be focussed on improving understanding of the mechanisms of microbial plant cell wall degradation [3].

Cellvibrio japonicus (previously *Pseudomonas fluorescens* subsp. *cellulosa*) is a saprophytic gram-negative bacterium, which was first isolated from Japanese soil [4]. The bacterium has been shown to utilise the major plant polysaccharides cellulose, mannan and xylan and, as such, is a good model for the study of bacterial plant polysaccharide degradation [5]. Recently the genome of *C. japonicus* was sequenced, and not surprisingly, a vast array of carbohydrate degrading enzymes was identified [6]. In contrast with anaerobic species, such as the *Clostridia*, the enzymes of *C. japonicus* that target polysaccharides integral to the plant cell walls do not form large membrane-attached cellulosome complexes [7], but are instead either secreted into the medium or attached to the cell membrane as lipoproteins. Indeed, about one third of the encoded carbohydrate-degrading enzymes have been predicted to be lipoproteins [6], which may suggest that *C. japonicus* has developed a strategy involving the ultimate degradation of plant cell wall-derived oligosaccharides in close proximity to the cell, as has been shown, for example, in the case of xylan utilisation [8, 9].

As a soil saprophyte, an abundant potential nutrient source for *C. japonicus*, apart from cellulose, is the group of non-crystalline cross-linking glycans commonly known as hemicelluloses. Among the most prevalent of these are the xyloglucans (XG), which are present in the primary cell wall of essentially all land plants. For instance, XGs may account for up to 20 % of the total carbohydrate content (dry weight) of the primary cell wall in dicots and non-graminaceous monocots, whereas more limited amounts are found in grass species [10]. XGs are a family of heterogeneous polysaccharides, which are typified by a β -(1 \rightarrow 4)-D-glucan backbone substituted with α -(1 \rightarrow 6)-D-xylopyranosyl moieties at regular intervals. The basic repeating motifs may be further decorated with galactopyranose, arabinofuranose, fucopyranose and O-acetyl moieties in a tissue- and species dependent manner (Fig. 1) [11, 12]. Thus, several enzymatic activities are required to degrade xyloglucan into its constituent monosaccharides. And whereas the endogenous enzymes responsible for the mobilisation of xyloglucans in plants have been previously studied in some detail [13-20], the elucidation of the enzymes responsible for XG degradation by microorganisms such as *C. japonicus* is currently in its infancy [21].

α -(1 \rightarrow 6)-Linked xylopyranose represents the second most abundant sugar in XGs, the cleavage of which is essential for the ultimate liberation of glucose from the polysaccharide backbone. α -Xylosidases have thus far only been identified in Glycoside Hydrolase Family 31 (GH31) (www.cazy.org, [2]), which also contains diverse enzymes active on α -glucosidic bonds, such as α -glucosidases, sucrose-isomaltases, α -glucan lyases and isomaltosyltransglucosidases. Whole-genome analysis of *C. japonicus* has suggested that this organism produces only a single GH31 α -xylosidase, *CjXyl31A* [6], which is thus likely to be a key player in XG degradation by this

organism. We present here the biochemical characterisation of *CjXyl31A* together with the three-dimensional structure of the enzyme. Detailed kinetic data indicated a strict *exo*-specificity for non-reducing-end xylose moieties and increased specificity for extended, non-galactosylated xylogluco-oligosaccharides (XGOs). In harness, structural analysis revealed that this specificity was affected, in part, by a unique PA14 domain insert in the N-terminal domain of *CjXyl31A* that endowed the active site pocket with an extended substrate binding platform. Together, these data provide exquisite molecular insight into the enzymic features that enable *C. japonicus* to utilise the naturally abundant plant cell wall polysaccharide xyloglucan.

Experimental

Curve-fitting was performed using Origin 8 (OriginLab). Ultrapure water was used in all experiments and refers to water purified on a Milli-Q system (Millipore) with a resistivity of $\rho > 18.2 \text{ M}\Omega\cdot\text{cm}$.

General analytical methods.

Mass spectrometry.

For MALDI-TOF, a Biflex IV controlled by FlexControl 2.4 (Bruker Daltonics) was used, with 2,5-dihydroxy benzoic acid used as the matrix. ESI-MS on oligosaccharides was performed as described previously [22]. LC-ESI-MS was used for protein molar mass determination as described previously [23].

NMR spectroscopy.

NMR spectroscopic data were recorded on a Bruker Avance 400 instrument operating at 400.1 MHz for ^1H , using the residual solvent signal as reference. Chemical shifts (δ) are reported in ppm and coupling constants (J) are given in Hertz (Hz).

Native gel electrophoresis.

Oligomerisation of Xyl31A was investigated by gel electrophoresis under native conditions, using a NuPAGE Novex 7 % Tris-Acetate gel (Invitrogen) with 25 mM Tris, 192 mM glycine (pH 8.3) as running buffer. Samples were loaded onto the gel in 50 mM Tris-HCl, 10 % glycerol, 0.05 % bromophenol blue (pH 8.6) and the molecular weights estimated by comparison with a high molecular weight marker for Native-PAGE (HMW Native Marker Kit, GE Healthcare). Proteins were visualised using PageBlue™ Protein Staining Solution (Fermentas).

High Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD).

Oligo- and monosaccharides were analysed on a Dionex ICS-3000 HPLC system operated by Chromelion software version 6.80 (Dionex) using a Dionex Carbopac PA200 column. Solvent A was water, solvent B 1 M sodium hydroxide and solvent C 1 M sodium acetate. Depending on the analytes, the following gradients were employed:

- *Gradient A*: 0 to 5 min, 10 % B, 2 % C; 5 to 12 min, 10 % B and a linear gradient from 2-30 % C; 12 to 12.1 min, 50 % B, 50 % C; 12.1 to 13 min, an exponential gradient of NaOH and NaOAc back to initial conditions; 13 to 17 min, initial conditions.

- *Gradient B*: 0 to 5 min, 10 % B, 0 % C; 5 to 10 min, 10 % B and a linear gradient from 0-30 % C; 10 to 10.1 min, 50 % B, 50 % C; 10.1 to 12.5 min, an exponential gradient of NaOH and NaOAc back to initial conditions; 12.5 to 15 min, initial conditions.
- *Gradient C*: 0 to 4 min, 10 % B, 6 % C; 4 to 17 min, 10 % B and a linear gradient from 0-25 % C; 17 to 17.1 min, 50 % B, 50 % C; 17.1 to 18 min, an exponential gradient of NaOH and NaOAc back to initial conditions; 18 to 22 min, initial conditions.

Enzyme substrates

p-Nitrophenyl (*p*NP) α -D-xylopyranoside, α -D-glucopyranoside, α -D-galactopyranoside and β -L-arabinopyranoside were purchased from Sigma. Tamarind seed xyloglucan and the derived oligosaccharides XXXG (Glc₄Xyl₃; see [24] for oligosaccharide nomenclature), isoprimeverose (X, Xylp- α (1 \rightarrow 6)-Glc_p), and borohydride-reduced xylosyl-cellobiose (XGol, Xylp- α (1 \rightarrow 6)-Glc_p- β (1 \rightarrow 4)-Glc_p) were purchased from Megazyme (catalog numbers: O-X3G4, O-IPRM and O-XCBIR, respectively). XLLG (Glc₄Xyl₃Gal₂) and XXXGXXXG (Glc₈Xyl₆) were prepared from tamarind xyloglucan as described previously [22, 25].

The oligosaccharide XX (Glc₂Xyl₂) was prepared from tamarind XG. 200 g XG was vigorously stirred in 2 l water at 65 °C to obtain a homogeneous mixture. The mixture was cooled to room temperature and ammonium acetate buffer (1 M, pH 4.5) was added until the desired pH was reached. Digestion was performed by addition of 1 g of the xgl1 endoglucanase (Dyadic NL) and the reaction proceeded at room temperature for four days with stirring. The reaction was stopped by heating at 80 °C for 30 minutes. The resulting suspension was centrifuged and the supernatant was collected, concentrated and precipitated using ethanol (4 equivalents v/v). The precipitate was collected and dissolved in water, followed by lyophilisation. A selected amount (4 g) of the XGO mixture was acetylated and purified by column chromatography as described previously [25], to obtain acetylated XX (0.15 g). The structure was verified by NMR (Supplemental Fig. S1) and ESI-MS (*m/z* calculated for C₄₆H₆₂O₃₁Na: 1133.3173; found: 1133.1969). The deacetylation was performed as described previously [25], and the molar mass was verified by ESI-MS (*m/z* calculated for C₂₂H₃₈O₁₉Na: 629.1905; found: 629.2047). Following lyophilization, the resulting powder was dissolved in water for biochemical assays.

Bioinformatic analysis

The phylogenetic tree of GH31 was derived from the protein sequences of biochemically characterised entries in CAZy together with the sequence of Xyl31A. The predicted catalytic domains of the genes were aligned using MUSCLE [26], and the alignment was used to construct the phylogenetic tree using the maximum likelihood method with the program PhyML [27]. The reliability of the tree was analysed by bootstrap analysis of 100 resamplings of the dataset. The tree was displayed using MEGA4 [28].

Cloning of Xyl31A

The open reading frame encoding Xyl31A (GenBank accession number ACE86259.1) was amplified by PCR from genomic DNA of *C. japonicus* Ueda107 using *Phusion* polymerase (Finnzymes) and the following primers (Thermo Fischer Scientific): 5'-CACCATGTTATCGGCACATCAGTG-3' and 5'-TTAACGCGGGCGTTTAATGC-3' with the forward primer incorporating the CACC overhang needed for TOPO cloning. The PCR

fragments were cloned into the pENTR/SD/D-TOPO entry vector (Invitrogen) according to the manufacturer's instructions. Chemically competent TOP10 cells were transformed with the cloning reaction and grown overnight. Plasmid DNA was extracted using the MiniPrep kit (Qiagen), and sequenced (Eurofins MWG Operon) to identify positive clones. The constructs were recombined, using the LR Clonase mix II (Invitrogen), into the pET-DEST42 destination vector (Invitrogen) (providing a C-terminal His₆-tag) according to the manufacturer's instructions.

Recombinant gene expression and protein purification

Plasmids harbouring the Xyl31A gene were transformed into *E. coli* BL21(DE3) by electroporation, and the resulting transformants were grown in Terrific Broth containing 50 µg/ml Ampicillin at 37 °C to an OD₆₀₀ of 0.5-0.8. Gene expression was induced by addition of 0.5 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG), and protein production was carried out at 25 °C overnight. The cells were collected by centrifugation at 3,800 × g for 10 minutes, and were resuspended in Buffer A (20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, 20 mM imidazole). The cells were passed twice through a French Press and centrifuged at 27,000 × g for 45 min. The supernatant was loaded onto a Ni-sepharose fast flow column (GE Healthcare) using an ÄKTA FPLC (GE Healthcare) and washed thoroughly with buffer A. Xyl31A was eluted with a linear gradient of 0-100% of buffer B (20 mM sodium phosphate (pH 7.4), 0.5 M NaCl, 1.0 M imidazole). The eluted protein was washed and concentrated with 50 mM sodium phosphate (pH 7.5) with 10 kDa cutoff Amicon Ultra centrifugal filters (Millipore). The protein was further purified by size exclusion chromatography (SEC) using a HiPrep 26/60 Sephacryl S-300 column (GE Healthcare) using 50 mM sodium phosphate (pH 7.5).

Antibody generation and purification

Antibodies were raised against recombinant Xyl31A in rabbits (AgriSera AB). The immunisation was as follows: immunisation 1: 200 µg antigen (10 mg/ml) + Freund's complete adjuvant (FCA); immunisation 2, 3 and 4 (1, 2 and 3 months later, respectively): 100 µg antigen + Freund's incomplete adjuvant (FIC). The final bleeding was carried out eleven days after immunisation 4.

Xyl31A was coupled to a 1 ml HiTrap NHS-activated HP Column (GE Healthcare) according to the manufacturer's instructions. The column was washed using an ÄKTA FPLC (GE Healthcare) with several column volumes of PBST (PBS (pH 6.9) containing 0.05 % Tween20). To 4.5 ml antiserum containing antibodies against Xyl31A was added 0.5 ml 10x PBS and the solution was applied to the Xyl31A-coupled HP column. The column was washed with several column volumes of PBST. Antibodies bound to the column were eluted with 200 mM glycine, 1 mM EDTA (pH 2.5) in 0.5 ml fractions, to wells containing 50 µl 1M Tris and 75 µl 10x PBS (pH 6.9) to neutralise the eluent. The purified antibodies were passed through a PD10 column packed with His-tagged Albumin Binding Protein coupled to a NHS matrix to remove His-tag specific antibodies.

X-ray crystallography

Initial crystallisation conditions were screened using several kits and needle-shaped crystals were obtained from 20 % (w/v) PEG 4000, 0.1 M MES (pH 6.5), 5 mM NiCl₂. Ultimately, two crystallisation conditions, using the hanging-drop vapor diffusion method, were optimised.

Crystal form 1 grew from protein in 10 mM MES pH 6.5, 0.15 M NaCl with 40 % pentaerythritol propoxylate (5/4 PO/OH), 0.1 M BIS-TRIS (pH 7.0), 5 mM NiSO₄ at 20 °C. Crystal form 2 grew from protein in 10 mM MES (pH 6.5), 0.15 M NaCl, 5 mM NiSO₄ with 25 % PEG monomethyl ether 550 (PEG mME 550), 0.1 M BIS-TRIS (pH 7.0) at 20 °C. The 5-fluoro-β-D-xylopyranosyl-enzyme intermediate was obtained by using crystal form 2 after soaking with approximately 2 mM "5-fluoro-α-D-xylopyranosyl fluoride" (5FαXylF) [29], a kind gift from Professor Stephen Withers, University of British Columbia) for approximately 1 hr.

Crystals of form 1 were cryo-protected using the reservoir solution alone whereas form 2 required elevation of the PEGmME concentration to 40 % PEG mME. Data for form 1 were collected on beamline ID14-2 of the European Synchrotron Radiation Facility (ESRF, Grenoble France) and that for form 2 on ESRF beamline ID29. Data were integrated using the HKL suite and MOSFLM/CCP4 [30]. Data processing and structure refinement statistics are shown in Table 1.

The structure was solved by the molecular replacement method using the MalA structure (*Sulfolobus solfataricus* α-glucosidase, PDB ID 2g3m), which was selected by BALBES [31], and the initial model was constructed using RESOLVE [32] and ARP/wARP [33] by using the 2.1 Å data from crystal form 1. Manual corrections were made using COOT [34]. All other computing used the CCP4 suite [30] unless explicitly stated. The coordinate file and monomer library description for 5/4 PO/OH was prepared using PRODRG [35] and PHENIX [36] whilst those for the trapped 5-fluoro-β-D-xylopyranosyl-enzyme intermediate were generated using the CCP4 suite. B-factors for all structures were refined with PHENIX by using TLS refinement with 5 TLS groups prepared by TLS motion determination web site (<http://skuld.bmsc.washington.edu/~tmsmd/>). The final structures were validated by MolProbity [37] (<http://molprobity.biochem.duke.edu/>). Structure similarity searches were performed with DALI [38].

For docking analysis of the XXXG oligosaccharide, a library file containing oligosaccharide bond and torsion angle information was prepared using the PRODRG server (<http://davapc1.bioch.dundee.ac.uk/prodrp/>) from PDB ID 2cn3 [22]. Subsequently, the non-reducing end xylosyl unit of XXXG was superposed onto the 5-fluoro-β-D-xylopyranosyl group in the glycosyl-enzyme structure using the Superpose Ligands extension in COOT, followed by manual optimisation of glycosidic torsional angles to relieve obvious steric clashes.

Stopped assay for hydrolysis of pNP glycosides

Hydrolysis of pNP glycosides (2 mM) by Xyl31A was performed in 100 µl reactions at 25 °C in 50 mM citrate buffer (pH 6) using typically 2-10 µM enzyme. The reactions were stopped by addition of 100 µl 0.2 M NaCO₃. Released p-nitrophenolate was measured in a Cary50 UV-vis spectrophotometer (Varian) at 410 nm, using 1 cm path length cuvettes (pNP anion extinction coefficient used for calculations was 18,500 M⁻¹·cm⁻¹ [39]).

pH-rate profile

Measurements of the pH-dependence of Xyl31A were carried out using pNP-α-Xyl (2 mM) and the stopped assay. The following buffers were used (50 mM): sodium citrate (pH 3 – 6.5), sodium phosphate (pH 6 – 8), glycylglycine (pH 7.5 – 9) and glycine (pH 9 – 9.5).

Temperature stability

To monitor the temperature stability of Xyl31A, the enzyme was incubated at different temperatures with aliquots taken out at different time points (5, 10, 20, 30, 60, 120, 240 min). The aliquoted enzyme was used in the stopped assay to determine remaining activity.

Specificity for XGO substrates

XGOs (XXXG, XLLG, XX, X) were incubated with Xyl31A (2.5 nM for longer XGOs, 2 μ M for isoprimeverose) in 50 μ l reactions containing 50 mM citrate buffer (pH 6) at 25 °C and the reactions stopped by addition of 1 μ l 5 M NaOH. Product formation was assayed using HPAEC-PAD, with Gradient A used for XXXG, XLLG and XX and Gradient B for X. For the XGol substrate, the BCA reducing sugar assay was used [40, 41]. 250 μ l reactions were made in 50 mM citrate buffer (pH 6) with 20 nM Xyl31A, and stopped by addition of an equal amount of BCA reagent solution. Colour was developed by incubation at 80 °C and measured at 560 nm. Standard curves were made using xylose (Sigma).

XGO regiospecificity determination:

To fully degrade XXXG to cellobiose, Xyl31A was used together with a β -glucosidase (*Clostridium thermocellum* Glucosidase 1A, NZYTech, Lisbon, Portugal). The reactions were carried out in 200 mM citrate buffer (pH 6), at 37 °C for Xyl31A and 55 °C for Glc1A until the reactions had gone to completion. First, Xyl31A was added to hydrolyse the xylose moiety from the non-reducing end, followed by denaturation by boiling. The same procedure was carried out with the β -glucosidase, hydrolysing the now exposed glucose moiety. The procedure was cycled until cellobiose was obtained. Product formation was analysed by ESI-MS.

To further analyse the regiochemistry of Xyl31A, XXXG₂ was used as substrate with subsequent analysis by MALDI-TOF MS.

Cellular localisation of Xyl31A

C. japonicus Ueda107 was grown in 250 ml M9 minimal medium [42] containing tamarind seed XGOs (0.4 % [w/v]) as the sole carbon source at 30 °C for 24 h until an OD₆₀₀ of 1, corresponding to a late exponential phase [43]. The cells were collected by centrifugation at 4,400 \times g for 10 minutes and the supernatant was decanted and concentrated ca. 25-fold using 10 kDa-cutoff Amicon Ultra centrifugal filters (Millipore).

Periplasmatic proteins were collected by osmotic shock. The cells were washed with 10 ml 50 mM Tris-HCl (pH 7.7) and collected by centrifugation at 4,400 \times g for 10 minutes. The pellet was resuspended in 50 ml 30 mM Tris-HCl, 20 % (w/v) sucrose, 1 mM EDTA (pH 8.0) and the cells were incubated in room temperature for 10 minutes. The cells were collected by centrifugation at 4,400 \times g for 15 minutes. 50 ml ice cold 5 mM MgSO₄ was added and the cells were incubated on ice for 10 minutes. The cells were centrifuged at 14,000 \times g for 10 minutes and the supernatant collected.

The cell pellet were resuspended in 50 mM sodium phosphate buffer (pH 7.4) and passed twice through a French Press. The lysate was centrifuged at 5,000 \times g for 10 minutes to remove debris. The supernatant was centrifuged at 100,000 \times g for 1 hour and the supernatant containing the

soluble proteins was collected. The pellet, containing the membrane fraction was resuspended in 100 mM sodium carbonate buffer (pH 9) to remove trapped soluble proteins and/or weakly membrane-associated proteins and centrifuged again at $100,000 \times g$ for 1 hour. The pellet was resuspended in 1 ml 50 mM sodium phosphate buffer (pH 7.4). Western blot was used to visualise the content of Xyl31A in the secreted, periplasmic, soluble intracellular and membrane fractions. The purified anti-Xyl31A antibodies were used as the primary antibody, and anti-rabbit IgG (Sigma) coupled to alkaline phosphatase was used as the secondary antibody, followed by visualisation using SIGMAFAST BCIP/NBT tablets (Sigma) dissolved in water.

Depolymerisation of XG by secreted enzymes of C. japonicus

The secreted protein fraction from the cellular localisation studies, above, was screened for xyloglucanase activity using XG as substrate. The fraction was washed with 50 mM sodium phosphate buffer (pH 7.4) using 10 kDa cutoff Amicon Ultra centrifugal filters (Millipore) to remove traces of XGOs from the growth medium.

Reactions consisted of 10 μ l of the protein fraction (0.04 mg/ml, determined by the Bradford method with a BSA standard curve) added to 145 μ l 1.72 g/l XG in 69 mM ammonium acetate buffer (pH 5.5) and incubation at 37 °C for 10, 20, 30, 120, 275 minutes and 48 hours. The reactions were stopped by boiling for 10 minutes, with subsequent lyophilisation. The samples were re-dissolved in 350 μ l DMSO and heated at 80 °C for 10 minutes. Gel permeation chromatography (GPC) analysis was used to monitor the depolymerisation of XG, as described previously [44] though using two PLgel 10 μ m mixed B columns and a guard column (Agilent) in tandem kept at 70 °C. An 11 point pullulan standard curve (Mw 180-1660000 Da) was used for molecular weight estimation. HPAEC-PAD was used for product analysis of the formed XGOs after the full depolymerisation of XG, using Gradient C.

Results and discussion

Primary structure analysis of Cj Xyl31A and GH31 phylogeny

Previous analysis of the genome of *Cellvibrio japonicus* predicted a vast array of polysaccharide degrading enzymes, of which *CjXyl31A* was annotated as the sole putative α -xylosidase belonging to Glycoside Hydrolase Family 31 (GH31) in the Carbohydrate-Active Enzymes (CAZy) classification [6]. Re-analysis of the primary structure indicated that *CjXyl31A* had the highest protein sequence identity to the *Sulfolobus solfataricus* GH31 α -xylosidase XylS (Genbank accession number CAB99206.1), with values of 27%, 39 %, and 35 % for the N-terminal, catalytic, and C-terminal domains, respectively. Also, a ca. 150 amino acid-long insert in the *CjXyl31A* N-terminal domain, which is not present in biochemically characterised GH31 enzymes, was identified. A BLAST search reveals that this sequence (amino acids 238-384) belongs to the PA14 superfamily (named after the protective antigen of anthrax toxin [45]), which is known to form inserts in a wide variety of enzymes and has been proposed to be involved in substrate binding [46]. Sequence analysis has previously indicated that the PA14 domain is also found in members of glycoside hydrolase families GH2, GH3, GH10, and GH20, in addition to GH31 [46]. However, little is presently known about the tertiary structure and function of PA14 domains in these enzymes. As a lone example, a recent report on the structure-function analysis of the *Kluyveromyces marxianus* β -glucosidase has highlighted that the PA14 domain in that GH3 enzyme occludes the active site pocket, thus conferring a preference for

disaccharide substrates [47]. In *CjXyl31A*, the PA14 domain is likewise positioned close to the active site, but in contrast appears to be involved in the extended recognition of longer xylogluco-oligosaccharides, as discussed below.

To delineate the diverse α -glucoside- and α -xyloside-processing activities in GH31, a phylogenetic tree was constructed using only the $(\beta/\alpha)_8$ catalytic domain of all members of GH31 whose enzyme specificity has previously been determined (data from the CAZy database [2], www.cazy.org), together with *CjXyl31A* (Fig. 2). In this phylogeny, a number of activities toward α -glucosidic bonds stood out clearly as distinct clades, including the α -1,4-glucan lyases, the isomaltosyl- and 6- α -glucosyltransglycosylases, and the sucrase-isomaltases. However, notably, the eight α -xylosidases characterised to-date did not clearly segregate, thus suggesting that α -xylosidase activity has arisen on separate occasions from *gluco*-specific enzymes. In particular, the known plant α -xylosidases (entries BAA99366.1 (*Oryza sativa*), AAD37363.1 (*Arabidopsis thaliana*), CAA10382.2 (*Tropaeolum majus*)) cluster with plant α -glucosidases, while the only other characterised eukaryotic α -xylosidase from *Aspergillus nidulans* (EAA62085.1) is most similar to the bacterial α -xylosidases (Fig. 2). The promiscuity of the *A. thaliana* α -xylosidase/ α -glucosidase [19] further highlights that small differences in protein sequence can modulate substrate discrimination based on the presence or absence of the C-6 hydroxymethyl group on the pyranose ring [48] (see Fig 1B). Indeed, the *E. coli* GH31 α -xylosidase YicI has been converted into an α -glucosidase by site directed mutagenesis [49]. Given the comparatively limited data available to-date, it is clear that phylogenetic groupings do not unambiguously predict catalytic function in GH31, thus motivating further structural and functional characterisation of *CjXyl31A*.

It is worth noting here that an earlier suggestion [29] that bacterial GH31 members invariantly contain the consensus sequence KTDFGE surrounding the conserved catalytic nucleophile (underlined) appears to not be borne out in the analysis of larger datasets. *CjXyl31A*, in particular, contains the sequence WLDAVE, which is also similar to that of the well-characterised archeal α -xylosidase, *SsXylS* (WLDASE, [50], see Supplemental Figure S3). Of the fifteen biochemically characterised bacterial genes used to construct the phylogenetic tree showing in Fig. 2, only *E. coli* YicI (GenBank accession AAC76680.1) and *Lactobacillus pentosus* XylQ (AAC62251.1) contain the exact KTDFGE sequence, while four bacterial isomaltosyltransglucosidases (BAB88401.1, BAD34979.1, BAB88403.1, BAC54957.1) possess the similar KTDFGg consensus sequence; the remainder of the bacterial enzymes in the present analysis contain the consensus sequence WnDmnE.

Heterologous expression and protein purification

CjXyl31A bearing a C-terminal hexahistidine tag was produced in *E. coli* BL21(DE3) cells; the positioning of the purification tag was chosen based on previous report which indicated that the N-terminal portion of GH31 enzymes may influence substrate binding [29]. The recombinant protein was purified by IMAC followed by SEC to remove low molar mass contaminants. SEC data, confirmed by native-PAGE on individual fractions, showed the protein being in mainly monomeric form (ca. 75 %), with minor fractions in dimeric, tetrameric and hexameric forms. The electrophoretically pure (SDS-PAGE, data not shown) pooled fractions from the SEC was used for all subsequent analyses. LC-ESI-MS indicated that the *CjXyl31A* so produced had a molar mass of 113644 ± 3 (calculated 113643.2), corresponding to the His-tagged protein less a

23 residue signal peptide. The LipoP 1.0 Server [51] indicated that this was likely a lipoprotein signal peptide, which would anchor the enzyme to the plasma membrane through the sulfhydryl group of Cys23 after cleavage by Signal Peptidase II [52]. *CjXyl31A* was however produced in good yields (200 mg/l) as a soluble protein in *E. coli*, with Ser24 as the N-terminus after signal peptide processing.

Substrate specificity of the -1 (glycone) subsite

A range of *para*-nitrophenyl (*p*NP) α -glycopyranosides were used to screen the substrate specificity of *CjXyl31A* (Fig. 1B). In keeping with the original prediction of activity [6] and sequence similarity to *SsXylS* (Fig. 2), *CjXyl31A* exhibited the highest activity toward *p*NP α Xyl (Table 1). With the corresponding glucoside, activity was several orders of magnitude lower, and no activity was observed on *p*NP α Gal or *p*NP β -L-Arap substrates. Compared to the *E. coli* α -xylosidase YicI [29], *CjXyl31A* had a six-fold lower k_{cat}/K_m value on *p*NP α Glc (YicI $6.0 \cdot 10^{-3} \text{ mM}^{-1} \text{ s}^{-1}$, *CjXyl31A* $1.2 \cdot 10^{-3} \text{ mM}^{-1} \text{ s}^{-1}$) and a ten-fold higher k_{cat}/K_m value on *p*NP α Xyl (YicI $0.18 \text{ mM}^{-1} \text{ s}^{-1}$, *CjXyl31A* $1.5 \text{ mM}^{-1} \text{ s}^{-1}$), which indicates a significantly higher specificity for xylose (i.e. discrimination against the 6-hydroxymethyl substituent) in the -1 subsite [53] of *CjXyl31A*.

pH dependence and thermostability studies

The pH dependence of catalysis by *CjXyl31A* was assayed using *p*NP α Xyl. A classic bell-shaped pH profile was obtained (Fig. 3), with an optimum at 6.5 and apparent $\text{p}K_a$ values of 4.3 and 8.8, which likely correspond to the ionisation of the catalytic nucleophile and acid/base residues, respectively. The thermal stability of *CjXyl31A* was assayed by incubating the enzyme at different temperatures and assaying remaining activity after defined time intervals (Fig. 3). At 35 °C and 45 °C, *CjXyl31A* exhibits good stability under the solution conditions used, with approximately 15 % activity loss after four hours at both temperatures. At higher temperatures (≥ 55 °C), the protein loses activity rapidly, as might be expected for an enzyme from a mesophile.

Specificity toward xylogluco-oligosaccharides (XGOs)

In nature, α -linked xylose is most prevalent in the plant cell wall polysaccharide xyloglucan [10]. To investigate the potential of *CjXyl31A* to participate in xyloglucan degradation, assays were performed on a small library of xylogluco-oligosaccharides (XGOs) obtained by enzymatic hydrolysis (Fig. 1C). *CjXyl31A* showed a clear preference for longer substrates, with the heptasaccharide XXXG yielding the highest k_{cat}/K_m value, which was similar to the tetrasaccharide XX (Table 1). On the trisaccharide alditol XGol, the k_{cat}/K_m value was about a third of that for XXXG. The disaccharide X (Xyl β - α (1 \rightarrow 6)-Glc β , Fig. 1A) was hydrolysed at a significantly slower rate than the other XGOs and had a 100-fold higher K_m value compared to XXXG. Galactosylation also reduced the specificity of *CjXyl31A* for oligosaccharide substrates; the k_{cat}/K_m value for the nonasaccharide XLLG was 3-fold lower than for XXXG, due to a significantly higher K_m value of the former.

Product analysis by ESI-MS indicated that for all of the XGO substrates, only one xylose residue was released (full data not shown; see e.g. XXXG, Table 2). Similarly, MALDI-TOF MS analysis of the reaction products following incubation of *CjXyl31A* with XXXGXXXG (Fig. 1C) indicated that a single xylosyl unit was cleanly released from the tetradecasaccharide (Fig. 4), presumably from the extreme non-reducing end. The strict non-reducing-end, *exo*-activity of

CjXyl31A was confirmed by sequential incubation of the heptsaccharide XXXG with *CjXyl31A* and *Clostridium thermocellum* *exo*- β -glucosidase Glc1A, incorporating a heat-inactivation step between each enzyme treatment. Thus, XXXG was degraded by *CjXyl31A* to yield only GXXG, which could then be hydrolysed to XXG by *CtGlc1A*, which could again be cleaved by *CjXyl31A* to GXG, and so on, to yield cellobiose after the last xylosidase treatment (Table 2). Together, these results demonstrate that *CjXyl31A* is unable to access internal xylose residues in xyloglucan polymers, even when they are followed by an unbranched glucose residue in the backbone, as in XXXGXXXG. This behaviour is similar to that of previously characterised α -xylosidases from plants and archaea, which require a β -glucosidase for the full degradation of, e.g., the XXXG motif [15, 16, 50]. Analysis of the three-dimensional structure of *CjXyl31A*, reported below in *apo* and substrate-complexed forms, provides a clear rationalisation for this exquisite regiospecificity.

In the context of biotechnological applications, including plant biomass saccharification, *CjXyl31A* displays catalytic constants for the hydrolysis of the artificial substrate *pNP* α Xyl (Table 1) similar to that of other GH31 α -xylosidases. Thus, the k_{cat} and K_{m} values of mesophilic enzymes for *pNP* α Xyl typically fall in the ranges of 0.1 – 0.6 s⁻¹ and 0.1 – 1 mM, respectively (see [29] and references therein), whereas k_{cat} values for the thermophilic α -xylosidase from *Sulfolobus sulfataricus* may approach 3 – 10 s⁻¹ in the temperature range 65 – 85 C [50] (assuming enzyme saturation at [S] = 37 mM used in that study). However, while *CjXyl31A* is more selective for *pNP* α Xyl over *pNP* α Glc than the well-characterized *E. coli* YicI [29] (*vide supra*), the *C. japonicus* enzyme is a much poorer degrader of the minimal disaccharide isoprimeverose (k_{cat} 2.9 s⁻¹, K_{m} 55 mM, Table 1) than either its *E. coli* (k_{cat} 51 s⁻¹, K_{m} 0.55 mM, [54]) or *S. sulfataricus* (specific activity *ca.* 20 s⁻¹ at [S] = 100 mM and 65 C, [50]) homologs. Direct comparison with plant α -xylosidases is hampered by a paucity of detailed kinetic analyses [19, 20]. Thus, the kinetic data (Table 1) suggest that *CjXyl31A* is best suited to technical uses that involve the breakdown of larger xylogluco-oligosaccharides.

Overall three-dimensional structure of *CjXyl31A*

To illuminate the structural bases for catalysis and the potential role of the "PA14" domain, the three-dimensional structure of the xylosidase was solved. Crystals of *CjXyl31A* were obtained in the hexagonal space group P6₃22 with one molecule in the asymmetric unit. The structure was solved by molecular replacement using the structure of the *S. sulfataricus* α -glucosidase MalA (PDB ID 2g3m). The refined structure of the *apo* enzyme, the 5-fluoro- β -xylosyl-enzyme intermediate and the pentaerythritol propoxylate complex were solved at 2.6, 2.5 and 2.3 Å resolution respectively, Table 3. As described above, mass spectrometric analysis indicated that the predicted signal sequence of 24 amino acids was processed during protein expression. In addition, no electron density was visible for the first 20 N-terminal residues of the mature protein in any of the structures. Thirty-one C-terminal residues of the *apo* structure and 32 C-terminal residues of the complex structures were also unresolved and are presumably disordered.

The *CjXyl31A* monomer structure has a complex modular architecture which is difficult to define unambiguously, but which may be considered as five domains consisting of two N-terminal "lobes", a central catalytic domain and two C terminal modules (Fig. 5). The central catalytic domain forms an (β/α)₈-barrel consisting of residues 412-791 and with two significant "inserts" beyond the standard barrel termed insert region-1 (Ins1) at residue 512-556 and insert

region-2 (Ins2) at residue 586-613. The $(\beta/\alpha)_8$ -barrel contains the catalytic apparatus, as will be described subsequently in light of a covalent intermediate complex and by virtue of similarities to related GH31 and Clan GH-D enzymes of known 3-D structure. The N-terminal region clearly forms two essentially discrete lobes. The first, primarily N-terminal, lobe is defined by residues 45-237, which form a β -sandwich domain. The topology of this domain is complicated by the fact that the 3-D structure of the whole domain also involves residues 392-411 when they return from the excursion that forms the second N-terminal domain. This latter domain, also itself a β -sandwich lobe, is an excursion between two of the strands of the N-terminal domain. This inserted domain, defined approximately by residues 238-391, is classified as a "PA14" domain (described previously, and in more detail below). The C-terminal domains follow a more regular topology defined by two discrete sequential β -sandwich domains approximately delineated by residues 793-870 and 871-988, respectively. Interface analysis using PISA [55] suggests that none of the potential interfaces are significant in terms of a stable higher oligomeric complex.

The most striking difference between *CjXyl31A* and other known GH31/Clan GH-D tertiary structures is the unusual N-terminal "PA14" domain, the function of which is not generally known. The PA14 domain derives its name from the N-terminal cleaved domain of anthrax protective antigen [45], with which the corresponding residues in *CjXyl31A* display an rmsd of 2.3 Å over 105 matched C α positions (DALI [38] Z-score = 8.6). The domain is also similar to a number of other sugar-binding domains such as the mannose-binding flocculation protein Flo5 protein [56] (PDB 2xj2, Z score 9.1 with 119 C α overlapping, rmsd 2.5 Å). Most intriguingly, this domain shares high structural similarity with a number of other domains such as observed in a family GH3 β -glucosidase [47] (PDB 3abz, DALI Z-score 13.1, with 132 C α overlapping, rmsd 2.7 Å), the family GH2 β -glucuronidase [57] (PDB 3k46, Z 9.1, 118 C α overlapping, rmsd 3.5 Å), family GH2 β -mannosidases (PDB 2je8, Z-score 7.9, 113 C α overlapping, rmsd 3.2 Å), as well as other similar domains from both β -glucosidase and β -glucuronidases. The PA14 domain observed in *CjXyl31A* is, thus far, unique to GH31 enzymes of known structure - it is not, for example, found in *E. coli* YicI (Fig. 5). Given the proximity of the PA14 domain to the active site pocket of *CjXyl31A*, we speculate this domain play a role in the binding of oligosaccharides, contributing to the preference for longer XGOs, as will be described below.

3-D structure of a trapped covalent glucosyl-enzyme intermediate.

Crystals of "form 2" (see Methods) were soaked in the mother liquor supplemented with approximately 2 mM of "5-fluoro- α -D-xylopyranosyl fluoride" (5F α XylF). Electron density for a covalently linked 5-fluoro- β -xylosyl residue was clearly seen in the -1 subsite with the anomeric carbon atom making a covalent bond to OD2 atom of D582 with 1.4 Å distance (Fig. 5). The ring conformation refined to a 1S_3 (skew boat) conformation, consistent with that observed in structures of α -glycosidases from GH31 and Clan GH-D described previously [29, 58]). D582 and D659 are predicted to act as the catalytic nucleophile and acid/base residues, as previously defined for this sequence family [29]. D470, H740, D659 and R642 all make hydrogen bonds to hydroxyl groups of the 5-fluoro- β -xylosyl ester, whilst F692 and W471 make Van der Waals' contacts with the ligand. A water molecule binding to D659 lies 3.1 Å from the C1 atom of the intermediate and would appear to be perfectly poised to act as the hydrolytic water in the reaction mechanism. Crystal form 1 also yielded a structure with the precipitant pentaerythritol propoxylate bound in the catalytic pocket. The environment of the pocket is almost same as that in *apo* and intermediate structures except that W347 and W471 are flipped away from the inside

of the pocket by ca. 20° and 80°, respectively. Pentaerythritol propoxylate settles between W347 and W542 and has a Van der Waals' contact to W471. Together, these data suggest plasticity of the binding site.

Relation of 3-D structure to observed kinetics and digestion patterns

Previous studies on α -xylosidases from *Tropaeolum majus*, *Arabidopsis thaliana* and *Sulfolobus solfataricus* have shown the selective hydrolysis of one xylose moiety from the non-reducing end of xylogluco-oligosaccharides (XGOs), without observed activity on internal xylose moieties [16, 18, 50]. To date, the three dimensional structure has been determined for five members of GH31, with YicI from *Escherichia coli* being the only α -xylosidase [29]. These five enzymes all display a highly similar fold, with a central catalytic (α/β)₈-barrel flanked by N- and C-terminal β -sheet domains of unknown functions, though the N-terminal domain has been suggested to be involved in substrate binding [29].

In the case of *CjXyl31A*, substrate specificity has two facets: (1) this strictly *exo*-acting enzyme is specific for a terminal, unsubstituted xyloside and (2) this xyloside must lie at the extreme reducing end of a xylogluco-oligosaccharide; there is no activity on the internal xylosides of substrates such as XXXG and XXXGXXXG. The specificity for an unmodified terminal xylose moiety is provided, as with other GH31 enzymes, by the constellation of residues that interact directly with the -1 subsite xyloside. W580, W656 and F692 together provide a steric barrier at the end of the active centre pocket, in which D470, H740, D659 and R642 make hydrogen bonds to the -1 subsite xyloside (described above in light of the xylosyl-enzyme intermediate structure; see also Fig. 5). There is thus no scope, simply in terms of sterics, to accommodate a xyloside with appended Gal or Gal-Fuc moieties, as found in oligosaccharides such as XLLG and XLFG (Fig. 1). The pocket itself, however, is more than merely a blunt-ended -1 subsite. The POCKETFINDER server (<http://www.modelling.leeds.ac.uk/qsitfinder/help.html>) estimates that the catalytic crevice has minimum dimensions of ~11 Å (depth) x 8 Å (width), although the binding surface could certainly be more extended as the mouth of the canyon expands into a larger estuary.

Our attempts to soak/co-crystallize with various XGOs did not yield a complexed structure, which, in the case of soaking, may partially reflect a two-fold axis of crystallographic symmetry that resulted in the active sites of two monomers facing each other. To circumvent this paucity of direct experimental data, a simple manual model of the Michaelis complex between *CjXyl31A* and XXXG was constructed based upon the 3-D structures of a known XGO/*endo*-xyloglucanase complex [22] and the glycosyl-enzyme intermediate described above. As shown in Fig. 6, interactions with the XXXG substrate extend beyond the active site pocket to the extended face provided by the PA14 domain. Another key feature the model is that the O4 atom of the +1 glucoside moiety is also enrobed by a tight surface comprised of, notably, R642 and the loop of residues 582-587, whose distal end begins one of the unusual insertions in the classical (β/α)₈ barrel). Likewise, W542, from the second of the unusual insertions, lies under the potential +1/+2 boundary and may also play a role in discriminating against "endo" attack. Together these residues provide the steric features that prevent binding of a xyloside from an "internal" glucose moiety and demand that only the xylose on the extreme non-reducing end glucoside is accommodated. The data further suggest that if an internal xylose-cleaving enzyme is desired, homologs should be sought with truncations in these regions.

The unique PA14 domain lies on the side of the substrate binding cleft and could, potentially both facilitate binding of longer oligosaccharides and perhaps also play a role in productive binding both of the backbone of the xyloglucan oligosaccharides (W347, for example lies on the floor of the binding canyon such that it may interact with a +2 subsite glucoside) whilst also providing numerous H-bonding residues that may interact productively with appended xyloside moieties. Such proposals are clearly speculative, but it is fascinating to see that other similar PA14 domains in unrelated glycosidases have also been recruited to modulate the oligosaccharide specificity, such as in *Bacteroides thetaiotaomicron* Man2A where the domain serves to block off an otherwise "endo" active centre groove to provide the chain-end specificity demanded by a β -mannosidase [59]. Similarly, the PA14 domain of the family GH3 β -glucosidase has recently been shown to modify the *exo/endo* behaviour of that enzyme [47]. Clearly, our understanding of the potential roles of PA14 domains is only in its infancy, due to the limited number of structure-function studies available presently. However, given the wide distribution of this domain across GH families [46], we anticipate further focussed studies on the importance of PA14 in carbohydrate recognition and catalysis.

The physiological role of CjXyl31A in the scavenging of xyloglucan by Cellvibrio japonicus in the natural environment

The detailed protein structural and enzyme kinetic data presented here lay the initial groundwork for understanding the mechanism by which the soil saprophyte *Cellvibrio japonicus* scavenges xyloglucan in its habitat. As highlighted in the Introduction, complete saccharification of the plant polysaccharide likely involves the concerted action of one or more *endo*-acting xyloglucanases capable of generating shorter oligosaccharide fragments (e.g. EC 3.2.1.151) [21], together with a battery of *exo*-acting glycosidases, including *CjXyl31A*, to release individual monosaccharides for cellular uptake.

Thus far, no xyloglucan-specific *endo*-glucanases have been described from *C. japonicus*, nor is there any general information on the ability of the organism to degrade xyloglucan polymers. Consequently, *C. japonicus* was grown on a minimal medium containing a mixture of tamarind seed XGOs as the sole carbon source, in an attempt to induce the production of xyloglucan-active enzymes. Subsequent analysis of the extracellular medium obtained during the late-exponential growth phase indeed indicated the presence of *endo*-xyloglucanase activity. Specifically, HPSEC and HPAEC-PAD analysis (Supplemental Fig. S3) indicated that this crude preparation was able to hydrolyse tamarind seed xyloglucan to the four component oligosaccharides, XXXG, XLXG, XXLG, and XLLG (see Fig. 1 for structures) by specific cleavage at the unbranched glucosyl unit. While the isolation and further characterisation of the individual enzyme(s) responsible for this activity are beyond the scope of this paper, these results clearly demonstrate that *C. japonicus* has the capacity to generate its own XGOs for further utilisation. It is also noteworthy that the extracellular fraction appears to be essentially devoid of $\beta(1\rightarrow2)$ -galactosidase and $\alpha(1\rightarrow6)$ -xylosidase activities, as evidenced by the peak areas of the component oligosaccharides, which are similar to those obtained via digestion with purified *endo*-xyloglucanases [22, 25, 60].

The observation that *CjXyl31A* is encoded with a predicted lipoprotein signal peptide (*vide supra*) provided a further tantalizing clue about the potential mechanism by which *C. japonicus*

systematically disassembles xyloglucan. To verify the membrane localisation of *CjXyl31A* in the native organism, Western blot analysis was performed on total membrane, soluble intracellular, periplasmic and secreted protein fractions (Fig. 7). The analysis indicates *CjXyl31A* was most abundant in the total membrane fraction, but was also found in significant quantity in the soluble intracellular fraction, and, to a lesser extent, in the periplasm. In keeping with the HPLC data (Supplemental Fig. S3), essentially no *CjXyl31A* was found in the extracellular medium; trace amounts, detectable as a very faint band on the Western blot after prolonged staining, may be attributable to cell lysis late in the exponential growth phase. Localisation of *CjXyl31A* on the outer membrane is consistent with previous studies on membrane-bound glycosidases of *C. japonicus* [8, 9]. Further, the absence of an aspartic acid residue at position +2 – this residue is a serine in *CjXyl31A* – is also suggestive of outer membrane localisation [51].

Conclusion

Taken together, the data present an emerging picture of xyloglucan utilisation by *C. japonicus*, in which soluble, secreted *endo*-xyloglucanase activity liberates oligosaccharides that are further degraded to monosaccharides close to the cell surface (Fig. 8). Notably, this model of step-wise hydrolysis of xylogluco-oligosaccharides by non-reducing-end-specific *exo*-glycosidases also rationalises the low activity that *CjXyl31A* exhibits toward the disaccharide isoprimeverose (Xylp- α (1→6)-Glc, “X” in Fig. 1C), which would not generally be encountered; the shortest oligosaccharide hydrolysed by *CjXyl31A* would be xylosyl-cellobiose (Table 2), the alditol of which (XGol) is cleaved with essentially the same selectivity as the parent oligosaccharides XXXG and XLLG (Table 1). Although the identities of the α (1→2)-L-fucosidase, β (1→2)-galactosidase(s), and β (1→4)-glucosidase necessary for the complete saccharification of XGOs are presently unknown, as is their potential to synergise with each other and with one or more unidentified sugar transporters, this model suggests fertile ground for future study. The detailed enzyme structure-function analysis in the present study has given new insight into the mechanisms of xyloglucan degradation by *C. japonicus*, yet there is still much information missing to fully describe the pathways of polysaccharide utilisation by this soil saprophyte.

Acknowledgements

Work in Stockholm and Newcastle was funded through the European Union WoodWisdom-Net ERA-NET project (www.woodwisdom.net) “FibreSurf – New biotechnological tools for wood fibre surface modification and analysis” by The Swedish Research Council Formas and UK Forestry, respectively. Supplemental funding from Formas via CarboMat – The KTH Advanced Carbohydrate Materials Consortium and the Wallenberg Wood Science Center is also gratefully acknowledged. Work in York was funded by the Biotechnology and Biological Sciences Research Council (BBSRC). GJD is a Royal-Society/Wolfson Research Merit Award recipient. We especially thank Professor Pedro M. Coutinho (Marseille) for his kind help with access to information on biochemically characterised enzymes from the CAZy database. Drs. Gustav Sundqvist and Chunlin Xu are thanked for assistance with mass spectrometric analyses.

References

- 1 Zhao, M. S. and Running, S. W. (2010) Drought-Induced Reduction in Global Terrestrial Net Primary Production from 2000 Through 2009. *Science* **329**, 940-943
- 2 Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res.* **37**, D233-D238
- 3 Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W. and Foust, T. D. (2007) Biomass recalcitrance: Engineering plants and enzymes for biofuels production. *Science* **315**, 804-807
- 4 Humphry, D. R., Black, G. W. and Cummings, S. P. (2003) Reclassification of '*Pseudomonas fluorescens* subsp. *cellulosa*' NCIMB 10462 (Ueda et al. 1952) as *Cellvibrio japonicus* sp. nov. and revival of *Cellvibrio vulgaris* sp. nov., nom. rev. and *Cellvibrio fulvus* sp. nov., nom. rev. *International Journal of Systematic and Evolutionary Microbiology* **53**, 393-400
- 5 Hazlewood, G. P. and Gilbert, H. J. (1998) Structure and function analysis of *Pseudomonas* plant cell wall hydrolases. *Prog. Nucleic Acid Res. Mol. Biol.* **61**, 211-241
- 6 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) Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. *J. Bacteriol.* **190**, 5455-5463
- 7 Fontes, C. and Gilbert, H. J. (2010) Cellulosomes: Highly Efficient Nanomachines Designed to Deconstruct Plant Cell Wall Complex Carbohydrates. In *Annual Review of Biochemistry*, Vol 79, pp. 655-681
- 8 Beylot, M. H., Emami, K., McKie, V. A., Gilbert, H. J. and Pell, G. (2001) *Pseudomonas cellulosa* expresses a single membrane-bound glycoside hydrolase family 51 arabinofuranosidase. *Biochem. J.* **358**, 599-605
- 9 Nagy, T., Emami, K., Fontes, C. M. G. A., Ferreira, L. M. A., Humphry, D. R. and Gilbert, H. J. (2002) The membrane-bound alpha-glucuronidase from *Pseudomonas cellulosa* hydrolyzes 4-O-methyl-D-glucuronoxyloligosaccharides but not 4-O-methyl-D-glucuronoxylan. *J. Bacteriol.* **184**, 4925-4929
- 10 Vogel, J. (2008) Unique aspects of the grass cell wall. *Current Opinion in Plant Biology* **11**, 301-307
- 11 Hoffman, M., Jia, Z. H., Pena, M. J., Cash, M., Harper, A., Blackburn, A. R., Darvill, A. and York, W. S. (2005) Structural analysis of xyloglucans in the primary cell walls of plants in the subclass Asteridae. *Carbohydr. Res.* **340**, 1826-1840
- 12 Hsieh, Y. S. Y. and Harris, P. J. (2009) Xyloglucans of Monocotyledons Have Diverse Structures. *Molecular Plant* **2**, 943-965
- 13 Edwards, M., Dea, I. C. M., Bulpin, P. V. and Reid, J. S. G. (1986) Purification and Properties of a Novel Xyloglucan-Specific Endo-(1-4)-Beta-D-Glucanase from Germinated Nasturtium Seeds (*Tropaeolum majus* L.). *J. Biol. Chem.* **261**, 9489-9494
- 14 Edwards, M., Bowman, Y. J. L., Dea, I. C. M. and Reid, J. S. G. (1988) A Beta-D-Galactosidase from Nasturtium (*Tropaeolum majus* L.) Cotyledons - Purification, Properties, and Demonstration That Xyloglucan Is the Natural Substrate. *J. Biol. Chem.* **263**, 4333-4337
- 15 Crombie, H. J., Chengappa, S., Hellyer, A. and Reid, J. S. G. (1998) A xyloglucan oligosaccharide-active, transglycosylating beta-D-glucosidase from the cotyledons of

- nasturtium (*Tropaeolum majus* L) seedlings - purification, properties and characterization of a cDNA clone. *Plant J.* **15**, 27-38
- 16 Fanutti, C., Gidley, M. J. and Reid, J. S. G. (1991) A Xyloglucan-Oligosaccharide-Specific Alpha-D-Xylosidase or Exo-Oligoxyloglucan-Alpha-Xylohydrolase from Germinated Nasturtium (*Tropaeolum majus* L) Seeds - Purification, Properties and Its Interaction with a Xyloglucan-Specific Endo-(1-4)-Beta-D-Glucanase and Other Hydrolases during Storage-Xyloglucan Mobilization. *Planta* **184**, 137-147
- 17 Leonard, R., Pabst, M., Bondili, J. S., Chambat, G., Veit, C., Strasser, R. and Altmann, F. (2008) Identification of an Arabidopsis gene encoding a GH95 alpha1,2-fucosidase active on xyloglucan oligo- and polysaccharides. *Phytochemistry* **69**, 1983-1988
- 18 Iglesias, N., Abelenda, J. A., Rodino, M., Sampedro, J., Revilla, G. and Zarra, I. (2006) Apoplastic glycosidases active against xyloglucan oligosaccharides of *Arabidopsis thaliana*. *Plant Cell Physiol.* **47**, 55-63
- 19 Sampedro, J., Sieiro, C., Revilla, G., Gonzalez-Villa, T. and Zarra, I. (2001) Cloning and expression pattern of a gene encoding an alpha-xylosidase active against xyloglucan oligosaccharides from Arabidopsis. *Plant Physiol.* **126**, 910-920
- 20 Crombie, H. J., Chengappa, S., Jarman, C., Sidebottom, C. and Reid, J. S. G. (2002) Molecular characterisation of a xyloglucan oligosaccharide-acting alpha-D-xylosidase from nasturtium (*Tropaeolum majus* L.) cotyledons that resembles plant 'apoplastic' alpha-D-glucosidases. *Planta* **214**, 406-413
- 21 Gilbert, H. J., Stalbrand, H. and Brumer, H. (2008) How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Current Opinion in Plant Biology* **11**, 338-348
- 22 Martinez-Fleites, C., Guerreiro, C. I. P. D., Baumann, M. J., Taylor, E. J., Prates, J. A. M., Ferreira, L. M. A., Fontes, C. M. G. A., Brumer, H. and Davies, G. J. (2006) Crystal structures of *Clostridium thermocellum* xyloglucanase, XGH74A, reveal the structural basis for xyloglucan recognition and degradation. *J. Biol. Chem.* **281**, 24922-24933
- 23 Sundqvist, G., Stenvall, M., Berglund, H., Ottosson, J. and Brumer, H. (2007) A general, robust method for the quality control of intact proteins using LC-ESI-MS. *Journal of Chromatography. B, Analytical Technologies in the Biomedical and Life Sciences* **852**, 188-194
- 24 Fry, S. C., York, W. S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J. P., Kato, Y., Lorences, E. P., MacLachlan, G. A., McNeil, M., Mort, A. J., Reid, J. S. G., Seitz, H. U., Selvendran, R. R., Voragen, A. G. J. and White, A. R. (1993) An Unambiguous Nomenclature for Xyloglucan-Derived Oligosaccharides. *Physiol. Plant.* **89**, 1-3
- 25 Greffe, L., Bessueille, L., Bulone, V. and Brumer, H. (2005) Synthesis, preliminary characterization, and application of novel surfactants from highly branched xyloglucan oligosaccharides. *Glycobiology* **15**, 437-445
- 26 Edgar, R. C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *Bmc Bioinformatics* **5**, 1-19
- 27 Guindon, S. and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**, 696-704
- 28 Tamura, K., Dudley, J., Nei, M. and Kumar, S. (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Mol. Biol. Evol.* **24**, 1596-1599

- 29 Lovering, A. L., Lee, S. S., Kim, Y. W., Withers, S. G. and Strynadka, N. C. J. (2005) Mechanistic and structural analysis of a family 31 alpha-glycosidase and its glycosyl-enzyme intermediate. *J. Biol. Chem.* **280**, 2105-2115
- 30 Bailey, S. (1994) The Ccp4 Suite - Programs for Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography* **50**, 760-763
- 31 Long, F., Vagin, A. A., Young, P. and Murshudov, G. N. (2008) BALBES: a molecular-replacement pipeline. *Acta Crystallographica Section D-Biological Crystallography* **64**, 125-132
- 32 Terwilliger, T. C. (2003) SOLVE and RESOLVE: Automated structure solution and density modification. *Macromolecular Crystallography, Pt D* **374**, 22-37
- 33 Perrakis, A., Morris, R. and Lamzin, V. S. (1999) Automated protein model building combined with iterative structure refinement. *Nature Structural Biology* **6**, 458-463
- 34 Emsley, P. and Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallographica Section D-Biological Crystallography* **60**, 2126-2132
- 35 Schuttelkopf, A. W. and van Aalten, D. M. F. (2004) PRODRG: a tool for high-throughput crystallography of protein-ligand complexes. *Acta Crystallographica Section D-Biological Crystallography* **60**, 1355-1363
- 36 Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. and Zwart, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography* **66**, 213-221
- 37 Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. and Richardson, D. C. (2010) MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallographica Section D Biological Crystallography* **66**, 12-21
- 38 Holm, L. and Sander, C. (1993) Protein-Structure Comparison by Alignment of Distance Matrices. *J. Mol. Biol.* **233**, 123-138
- 39 Brumer, H., Sims, P. F. G. and Sinnott, M. L. (1999) Lignocellulose degradation by *Phanerochaete chrysosporium*: purification and characterization of the main alpha-galactosidase. *Biochem. J.* **339**, 43-53
- 40 Mopper, K. and Gindler, E. (1973) A new noncorrosive dye reagent for automatic sugar chromatography. *Anal. Biochem.* **56**, 440-442
- 41 Mcfeeters, R. F. (1980) A Manual Method for Reducing Sugar Determinations with 2,2'-Bicinchoninate Reagent. *Anal. Biochem.* **103**, 302-306
- 42 Miller, J. H. (1975) *Experiments in molecular genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
- 43 Gardner, J. G. and Keating, D. H. (2010) Requirement of the Type II Secretion System for Utilization of Cellulosic Substrates by *Cellvibrio japonicus*. *Appl. Environ. Microbiol.* **76**, 5079-5087
- 44 Brumer, H., Zhou, Q., Baumann, M. J., Carlsson, K. and Teeri, T. T. (2004) Activation of crystalline cellulose surfaces through the chemoenzymatic modification of xyloglucan. *J. Am. Chem. Soc.* **126**, 5715-5721
- 45 Petosa, C., Collier, R. J., Klimpel, K. R., Leppla, S. H. and Liddington, R. C. (1997) Crystal structure of the anthrax toxin protective antigen. *Nature* **385**, 833-838

- 46 Rigden, D. J., Mello, L. V. and Galperin, M. Y. (2004) The PA14 domain, a conserved all-beta domain in bacterial toxins, enzymes, adhesins and signaling molecules. *Trends Biochem. Sci.* **29**, 335-339
- 47 Yoshida, E., Hidaka, M., Fushinobu, S., Koyanagi, T., Minami, H., Tamaki, H., Kitaoka, M., Katayama, T. and Kumagai, H. (2010) Role of a PA14 domain in determining substrate specificity of a glycoside hydrolase family 3 beta-glucosidase from *Kluyveromyces marxianus*. *Biochem. J.* **431**, 39-49
- 48 Ichinose, H., Fujimoto, Z., Honda, M., Harazono, K., Nishimoto, Y., Uzura, A. and Kaneko, S. (2009) A beta-L-Arabinopyranosidase from *Streptomyces avermitilis* Is a Novel Member of Glycoside Hydrolase Family 27. *J. Biol. Chem.* **284**, 25097-25106
- 49 Okuyama, M., Kaneko, A., Mori, H., Chiba, S. and Kimura, A. (2006) Structural elements to convert *Escherichia coli* alpha-xylosidase (YicI) into alpha-glucosidase. *FEBS Lett.* **580**, 2707-2711
- 50 Moracci, M., Ponzano, B. C., Trincone, A., Fusco, S., De Rosa, M., van der Oost, J., Sensen, C. W., Charlebois, R. L. and Rossi, M. (2000) Identification and molecular characterization of the first alpha-xylosidase from an Archaeon. *J. Biol. Chem.* **275**, 22082-22089
- 51 Juncker, A. S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. and Krogh, A. (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* **12**, 1652-1662
- 52 Paetzel, M., Karla, A., Strynadka, N. C. J. and Dalbey, R. E. (2002) Signal peptidases. *Chem. Rev. (Washington, DC, U. S.)* **102**, 4549-4579
- 53 Davies, G. J., Wilson, K. S. and Henrissat, B. (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem. J.* **321**, 557-559
- 54 Okuyama, M., Mori, H., Chiba, S. and Kimura, A. (2004) Overexpression and characterization of two unknown proteins, YicI and YihQ, originated from *Escherichia coli*. *Protein Expression Purif.* **37**, 170-179
- 55 Krissinel, E. and Henrick, K. (2005) Detection of protein assemblies in crystals. *Computational Life Sciences, Proceedings* **3695**, 163-174
- 56 Veelders, M., Bruckner, S., Ott, D., Unverzagt, C., Mosch, H. U. and Essen, L. O. (2010) Structural basis of flocculin-mediated social behavior in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 22511-22516
- 57 Wallace, B. D., Wang, H. W., Lane, K. T., Scott, J. E., Orans, J., Koo, J. S., Venkatesh, M., Jobin, C., Yeh, L. A., Mani, S. and Redinbo, M. R. (2010) Alleviating Cancer Drug Toxicity by Inhibiting a Bacterial Enzyme. *Science* **330**, 831-835
- 58 Guce, A. I., Clark, N. E., Salgado, E. N., Ivanen, D. R., Kulminskaya, A. A., Brumer, H. and Garman, S. C. (2010) Catalytic Mechanism of Human alpha-Galactosidase. *J. Biol. Chem.* **285**, 3625-3632
- 59 Tailford, L. E., Money, V. A., Smith, N. L., Dumon, C., Davies, G. J. and Gilbert, H. J. (2007) Mannose foraging by *Bacteroides thetaiotaomicron* - Structure and specificity of the beta-mannosidase, *BtMan2A*. *J. Biol. Chem.* **282**, 11291-11299
- 60 Gloster, T. M., Ibatullin, F. M., Macauley, K., Eklof, J. M., Roberts, S., Turkenburg, J. P., Bjornvad, M. E., Jorgensen, P. L., Danielsen, S., Johansen, K. S., Borchert, T. V., Wilson, K. S., Brumer, H. and Davies, G. J. (2007) Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. *J. Biol. Chem.* **282**, 19177-19189

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110299

Accepted Manuscript

Table 1. Activity of *CjXyl31A* on various substrates at pH 6.

Substrate ^a	k_{cat} s^{-1}	K_m mM	k_{cat}/K_m $s^{-1} mM^{-1}$	$\Delta\Delta G^\ddagger$ ^b $kJ mol^{-1}$
<i>pNP</i> α Xyl	0.58 ± 0.026	0.39 ± 0.06	1.48	
<i>pNP</i> α Glc	$0.0066 \pm 6 \cdot 10^{-4}$	5.31 ± 1.0	$1.2 \cdot 10^{-3}$	
<i>pNP</i> α Gal	N/D	N/D	N/D	
<i>pNP</i> β -L-Arap	N/D	N/D	N/D	
XXXG	26.7 ± 1.5	0.54 ± 0.073	49.5	0
XLLG	33.6 ± 3.2	2.6 ± 0.42	13.1	3.30
XX	31.2 ± 1.5	0.72 ± 0.076	44.7	0.25
XGol	9.39 ± 0.61	0.70 ± 0.013	13.4	3.23
X (Isoprimeverose)	2.94 ± 0.36	55.0 ± 7.4	0.05	17.1

a. Substrate structures are shown in Fig. 1.

b. $\Delta\Delta G^\ddagger$ was calculated using the formula: $\Delta\Delta G^\ddagger = -RT \ln[(k_{cat}/K_m \text{ XGO}) / (k_{cat}/K_m \text{ XXXG})]$

Table 2. Product analysis following sequential rounds of XXXG hydrolysis by *CjXyl31A* and *CtGlc1A*.

Compound	Expected m/z [M+Na] ⁺	Observed m/z [M+Na] ⁺
XXXG	1085.34	1085.2706
GXXG	953.30	953.2280
XXG	791.24	791.1766
GXG	659.20	659.1823
XG	497.15	497.1376
GG	365.11	365.0771

Table 3. X-ray diffraction and structure refinement statistics for CjXyl31A

Name	<i>apo</i> (PDB ID 2xvg)	5FXyl (PDB ID 2xvl)	5/4 PO/OH (PDB ID 2xvk)
Data collection			
Space group	P6 ₃ 22	P6 ₃ 22	P6 ₃ 22
Cell dimensions			
a, b, c (Å)	156.2, 156.2, 227.8	156.5, 156.5, 227.8	157.65, 157.65, 226.69
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120
Resolution (Å) (outer shell)	87 2.60 (2.74-2.60)	50 - 2.5 (2.54-2.50)	58.5-2.30 (2.42-2.30)
R_{merge}	0.11 (0.81)	0.072 (0.75)	0.172 (0.474)
$I / \sigma I$	14.4 (2.7)	32.2 (2.0)	16.3 (4.5)
Completeness (%)	99.9 (100.0)	99.0 (99.9)	100.0 (100.0)
Redundancy	11.0 (11.3)	10.1 (8.0)	11.3 (7.5)
Refinement			
Resolution (Å)	49.9-2.6	50.0-2.5	54.5-2.3
No. reflections	50969	57181	74143
$R_{\text{work}} / R_{\text{free}}$	0.20/0.26	0.20/0.24	0.145/0.185
No. atoms			
Protein	7578	7575	7583
Ligand	Ni; 2, Cl; 4, SO ₄ ; 4, EDO; 12	Ni; 1, Cl; 4, XYF; 10	Ni; 3, Cl; 7, SO ₄ ; 3, 5/4 PO/OH; 25
Water	103	83	724
B-factors (Å ²)			
Protein* (TLS refinement)	76.1	74.8	29.3
Water	47.3	49.1	33.1
Ligand			
Ni	55.6	45.8	39.7
Cl	62	61.7	42
Other	SO ₄ ; 134.1, EDO; 81.4	XYF; 69.6	SO ₄ ; 92.0, 5/4 PO/OH; 60.5
R.m.s deviations			
Bond lengths (Å)	0.009	0.008	0.007
Bond angles (°)	1.16	1.1	1.04

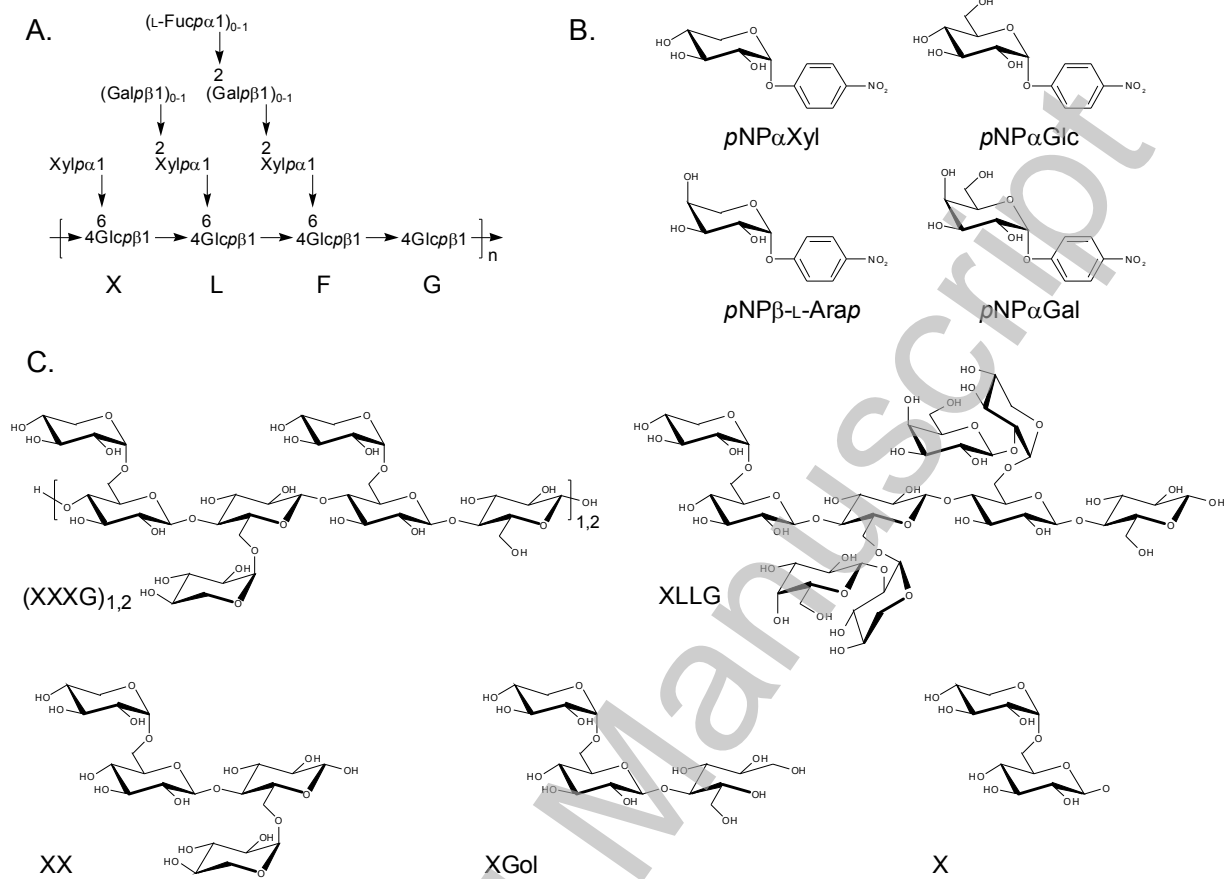


Figure 1 (double column). Structures of xyloglucan and *CjXyl31A* substrates. A. General structure of primary cell wall xyloglucans showing variable branching. B. Aryl glycosides for -1 subsite specificity determination. C. Xylogluco-oligosaccharide used for extended specificity determination. Xyloglucan fragments are named according to [28].

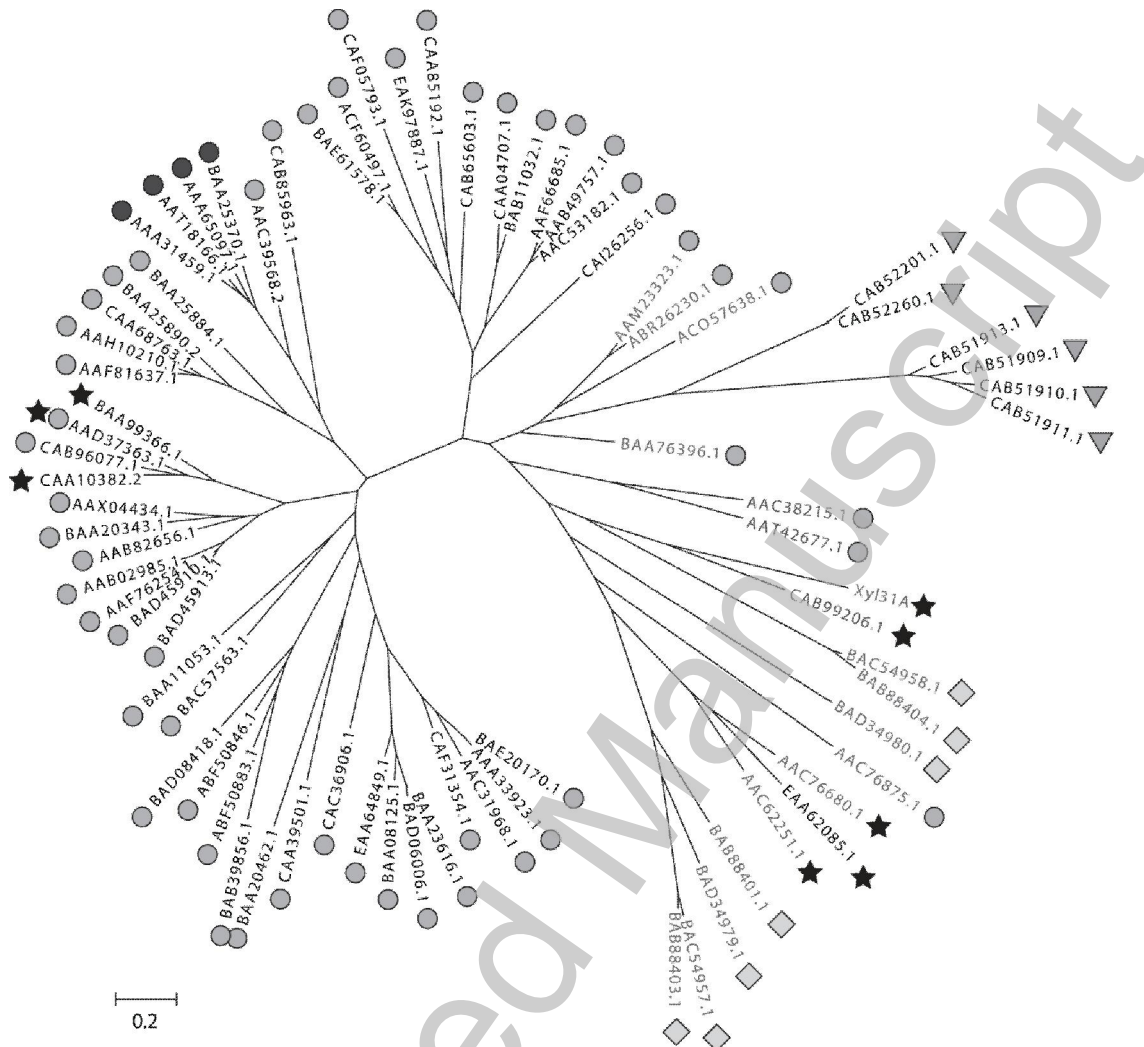


Figure 2 (double column). Phylogenetic tree of all biochemically characterised enzymes of GH31, including *CjXyl31A*. The genes are named by their GenBank accession numbers. Activities are designated as follows: Grey circle – α -glucosidase, Star – α -xylosidase, Diamond – transglycosylase, Triangle – α -glucan lyase and Black circle – sucrase-isomaltase. Accession numbers from eukaryotes are in black, those from bacteria and archaea are in grey.

THIS IS NOT THE VERSION OF RECORD - see doi:10.1042/BJ20110299

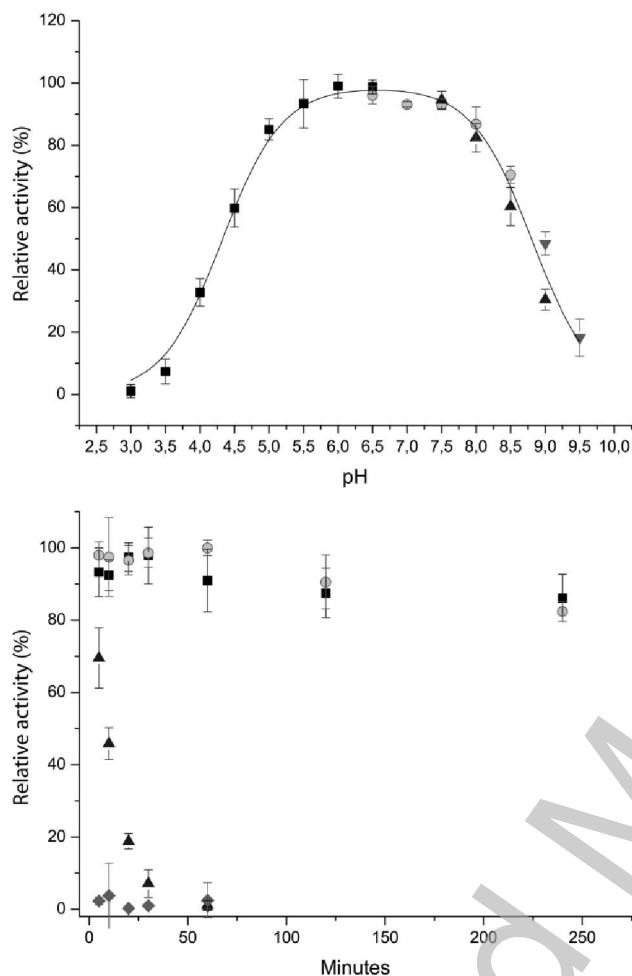


Figure 3 (single column). pH profile and thermostability of *CjXyl31A*. The buffers (50 mM) used in the pH-rate profile were citrate (pH 3-6.5 squares), phosphate (pH 6.5-8.5 circles), glycylglycine (pH 7.5-9 triangles) and glycine (pH 9, 9.5 inverted triangles). For the thermostability study, the enzyme was incubated at different temperatures (35 °C circles, 45 °C squares, 55 °C triangles, 65 °C diamonds), with aliquots of enzyme taken at different time intervals. The remaining activity was used to determine the level of thermal denaturation. 2 mM *p*NP- α -Xyl was used as the substrate for both the pH-rate profile and thermostability measurements.

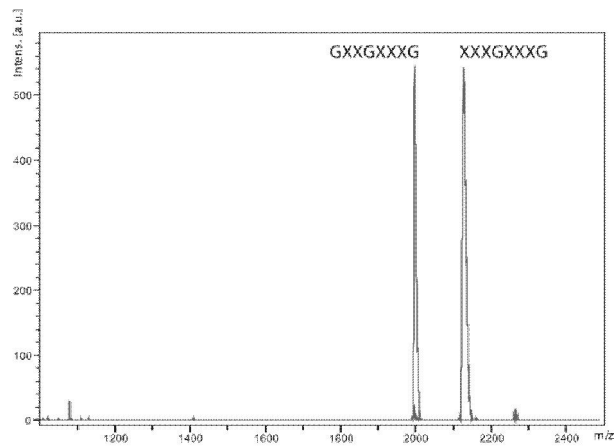


Figure 4 (single column). MALDI-TOF MS analysis of the XXXGXXXG tetradecasaccharide and its hydrolysis product GXXGXXXG obtained by incubation with *CjXyl31A*.

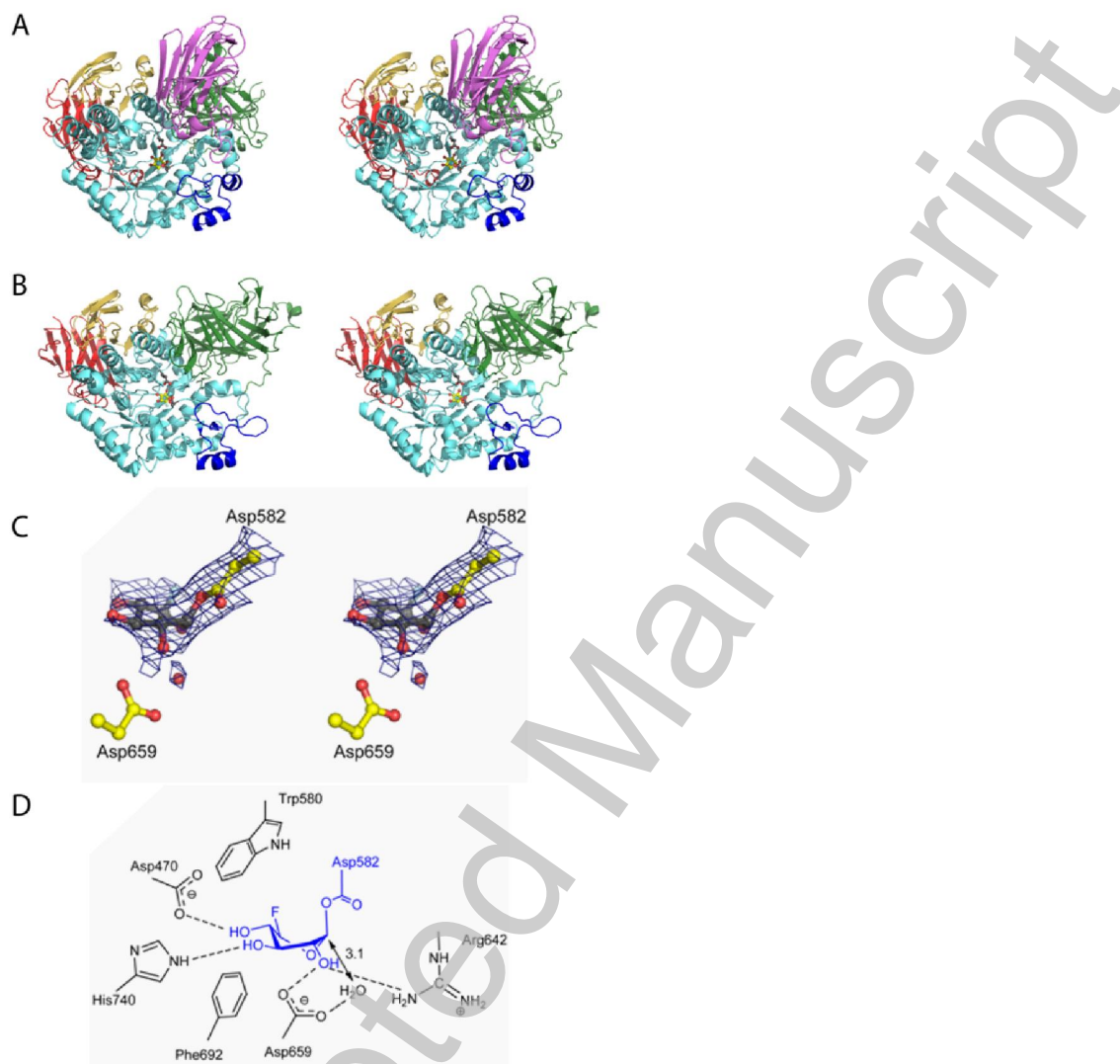


Figure 5 (single column). Structures of the *Cellvibrio japonicus* Xyl31A and *E. coli* YicI 5-fluoroxyl enzyme intermediates. Global structures of *CjXyl31A* (A) and YicI (B), in wall-eyed stereo, highlighting the distinct PA14 insert of *CjXyl31A* (purple) in the N-terminal domain (green). The homologous domains of both enzymes are colored as follows: N-terminal domains, green; catalytic domains, cyan; loop inserts in the catalytic domains, blue; C-terminal domains, yellow and red. The catalytic amino acid residues are shown in ball-and-stick (dark grey) with the nucleophile bound to the 5F- β -Xyl ligand (yellow). C. Electron density map of $2F_o - F_c$ around 5F- β -Xyl, including the nucleophilic water, shown at the 1.7σ level. D. Diagram of the interactions around the 5F- β -Xyl ligand. The arrow indicates the distance in Ångströms between the anomeric carbon of eq-5FaXF and water molecule bound to Asp659. The figure was prepared using Pymol and ChemDraw.

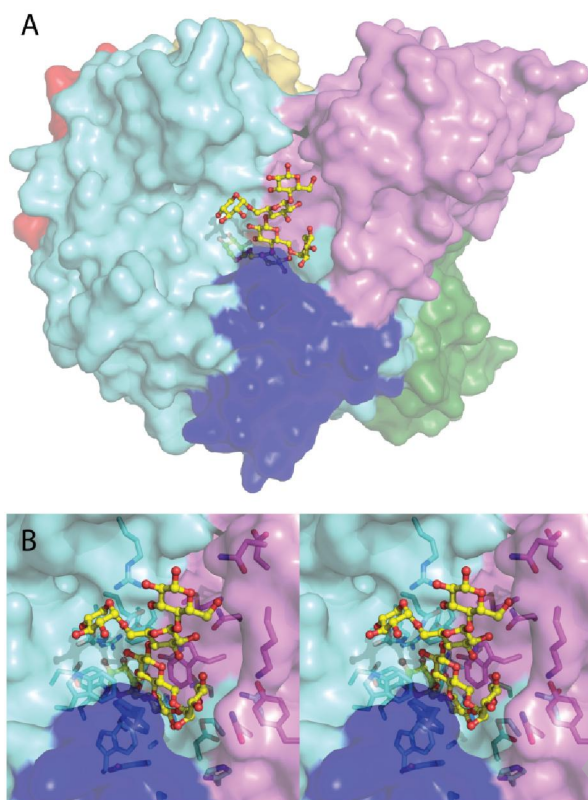


Figure 6 (single column). Molecular surface of *CjXyl31A* with docked heptasaccharide XXXG. A. Global view, B. Detailed view of the active site in wall-eyed stereo. The color scheme is identical to that used in Figure 5: N-terminal domain, green; PA14 domain, pink; catalytic domain, cyan; loop insert in the catalytic domain, blue; C-terminal domains, yellow and red. The figure was prepared using PyMol.

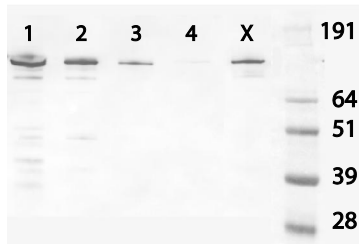


Figure 7 (single column). Western blot analysis of *C. japonicus* fractions using a polyclonal antibody preparation raised against recombinant *CjXyl31A*. The protein contents of the fractions were normalised prior to loading the lanes: 1, membrane; 2, soluble intracellular; 3, periplasmatic; 4, secreted; X, recombinant *CjXyl31A* produced in *E. coli*. The SeeBlue® Plus2 ladder (Invitrogen) was used; note that the phosphorylase marker at 97 kDa faded during the Western protocol.

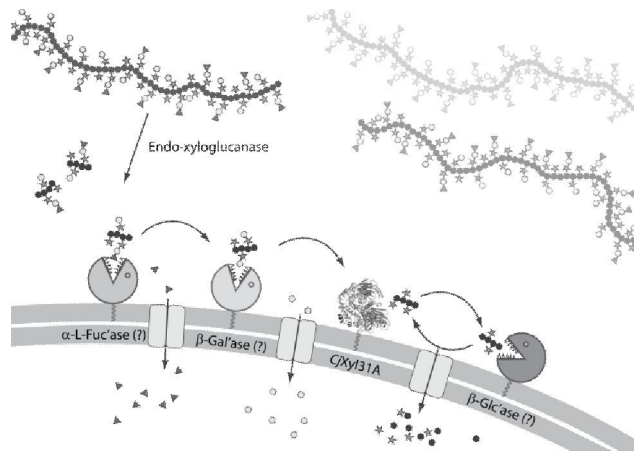


Figure 8 (single column). Working model of xyloglucan utilisation by *C. japonicus*. Secreted *endo*-xyloglucanase activity and outer membrane-anchored *CjXyl31A* demonstrated in the present study are indicated. Hypothetical *exo*-glycosidases, sugar transporters, and their potential order of interaction are indicated with “?”.