Biochemical and Structural Insights into Xylan Utilization by the Thermophilic Bacterium *Caldanaerobius polysaccharolyticus* 

Yejun Han, Vinayak Agarwal, Dylan Dodd, Jason Kim, Brian Bae, Roderick I. Mackie, Satish K. Nair, and Isaac K. O. Cann

Background: *Caldanaerobius polysaccharolyticus* is a thermophile with a hemicellulose utilization gene cluster. The cluster is induced by xylan. The ligand-binding cleft of XBP1 is optimized for binding xylotriose. This gene cluster encodes all of the proteins required to degrade xylan, transport the fragments, and metabolize them via the pentose-phosphate pathway. This gene cluster could be designed as a cassette to impart a capacity for utilizing hemicellulose.

Hemicellulose is the next most abundant plant cell wall component after cellulose. The abundance of hemicellulose such as xylan suggests that their hydrolysis and conversion to biofuels can improve the economics of bioenergy production. In an effort to understand xylan hydrolysis at high temperatures, we sequenced the genome of the thermophilic bacterium *Caldanaerobius polysaccharolyticus*. Analysis of the partial genome sequence revealed a gene cluster that contained both hydrolytic enzymes and also enzymes key to the pentose-phosphate pathway. The hydrolytic enzymes in the gene cluster were demonstrated to convert products from a large endoxylanase (Xyn10A) predicted to anchor to the surface of the bacterium. We further use structural and calorimetric studies to demonstrate that the end products of Xyn10A hydrolysis of xylan are recognized and bound by XBP1, a putative solute-binding protein, likely for transport into the cell. The XBP1 protein showed preference for xylo-oligosaccharides as follows: xylotriose > xylobiose > xylooligosaccharide. To elucidate the structural basis for the oligosaccharide preference, we solved the co-crystal structure of XBP1 complexed with xylotriose to a 1.8-Å resolution. Analysis of the biochemical data in the context of the co-crystal structure reveals the molecular underpinnings of oligosaccharide length specificity.

The efficient degradation of the polymer has gained increasing interest due to the capacity to convert its monomeric sugars to bioenergy products such as ethanol. Xylan is the most common hemicellulose, and a heterogeneous polysaccharide composed mostly of linear chains of xylose with side chain substitutions. The backbone of xylan is composed of β-1,4-linked D-xylopyranosyl units and may be decorated with 4-O-methyl-D-glucuronyl, 1-arabinofuranosyl, and acetyl substituents. The complete degradation of xylan requires the synergistic activity of several hemicellulolytic enzymes, such as β-1,4-endoxylanase, β-xylosidase, α-glucuronidase, α-1-arabinosidase, and acetylxylan esterase. To facilitate a concerted action of these enzymes for hemicellulose degradation, several microorganisms have evolved gene clusters encoding the different hemicellulolytic enzymes. The transport mechanism for xylan degradation products has been fairly well described in bacteria such as *Streptomyces lividans* and *Geobacillus stearothermophilus*. However, our knowledge in this area of sugar metabolism by bacteria is still limited.

β-Xylosidases and α-glucuronidases are two critical enzymes for xylan hydrolysis. Endoxylanases cleave xylan polysaccharides into xylo-oligosaccharides that may be decorated with methylglucuronic acids and hence impede the effectiveness of the β-xylosidase, the enzyme responsible for cleavage of xylose monomers from xylo-oligosaccharides. Microorganisms have therefore developed the enzymatic activity to remove the methylglucuronic acid side chains. Thus, α-glucuronidases cleave the α-1,2-glycosidic bond between 4-O-methyl α-glucuronic acid and the xlyopyranosyl unit of xylo-oligosaccharides to date, most of the cloned microbial β-xylosidases fall within glycoside hydrolase (GH) families 3, 30, 39, 43, 51, 52.
and 54 (2, 11–14), whereas characterized α-glucuronidases are assigned to either GH family 67 (6, 15, 16) or GH family 115 (17).

With a number of advantages over mesophilic enzymes, thermostable enzymes are especially thought to improve hydrolytic performance and overall economy of the process of biofuel production from the plant cell wall. Thermostable enzymes have thus been gaining increasing attention in the field of biofuels (18, 19).

Caldanaerobius polysaccharolyticus, an anaerobic thermophilic bacterium, was isolated in Illinois (20, 21). Several thermostable hemicellulolytic enzymes have been cloned and characterized from C. polysaccharolyticus (22, 23). Recently, by determining the partial genome sequence of this bacterium, we have identified all of the genes encoding enzymes that will permit reconstitution of a hemicellulolytic enzyme mixture, a highly desirable product in the emerging biofuel industry, from C. polysaccharolyticus. These enzymes include a β-1,4-endoxylanase (Xyn10A), β-xylosidase (Xyl3A), α-glucuronidase (Agu67A), α-L-arabinofuranosidase (Ara51A), and an acetylxyylan esterase (Axe4A).

In this study, Xyl3A and Agu67A appeared to be components of a pentose sugar metabolism cluster and were cloned and characterized from C. polysaccharolyticus. Because the endoxylanase, Xyn10A, is not linked to this gene cluster, we hypothesized that the gene products from the cluster serve to capture nutrients (xylo-oligosaccharides) generated by Xyn10A for further hydrolysis to directly feed them into the pentose-phosphate pathway. Here, we express the recombinant form of each protein and demonstrate their contributions to xylan metabolism by C. polysaccharolyticus. We also identify a membrane-integral ATP-dependent sugar complex that likely transports the end products of xylan degradation into the cell. To determine the chain length preference for this transporter, we carried out biochemical analysis of the solute-binding component (XBP1) of the complex, and we solved the co-crystal structure of this polypeptide in complex with xylotriose. It is anticipated that the clustering of the genes involved in xylan utilization in this thermophilic bacterium will also offer an opportunity to transfer the phenotype to other organisms with tractable genetic systems for further engineering and improvement of xylan utilization.

**EXPERIMENTAL PROCEDURES**

**Materials—**C. polysaccharolyticus (ATCC strain number BAA-17), originally named Thermoanaerobacterium polysaccharolyticum, was isolated from a waste pile of a canning factory in Illinois (20, 21). The pET-46b EK/LIC cloning kit and Perfect Protein Marker™ were purchased from Novagen (San Diego). PicoMaxx high fidelity PCR system, Pfu DNA polymerase, *Escherichia coli* JM109, and BL21-CodonPlus™ (DE3) RIL competent cells were obtained from Stratagene (La Jolla, CA). Restriction enzyme DpnI and 1-kb DNA ladder were purchased from New England Biolabs (Ipswich, MA). The DNeasy Blood and Tissue kit and the QIAprep spin miniprep kit were obtained from Qiagen, Inc. (Valencia, CA). The talon metal affinity resin was from Clontech. Amicon Ultra-15 centrifugal filter units with 30,000- and 50,000-Da molecular mass cutoffs were purchased from Millipore (Billerica, MA). Isopropyl β-D-thiogalactopyranoside, antibiotics, agarose, and sodium citrate were obtained from Fisher.

Xylo-oligosaccharides (xylooligosaccharides), anti-biotics, agarose, and sodium citrate were obtained from Fisher.

Xylo-oligosaccharides (xylooligosaccharides), antibiotics, agarose, and sodium citrate were obtained from Fisher.

Xylo-oligosaccharides (xylooligosaccharides), antibiotics, agarose, and sodium citrate were obtained from Fisher.

Xylo-oligosaccharides (xylooligosaccharides), antibiotics, agarose, and sodium citrate were obtained from Fisher.
Xylan Utilization by Caldanaerobius polysaccharolyticus

pET46 Ek/LIC vector using the Ek/LIC cloning kit (Novagen) and E. coli JM109 as the competent cells by electroporation (Gene Pulser XcelTM from Bio-Rad). All recombinant plasmids (pET46-xyl3A, pET46-agu67A, and pET46-xbp1) were then extracted using QIAprep spin miniprep kit, and the nucleotides were sequenced (W. M. Keck Center for Comparative and Functional Genomics, University of Illinois) to confirm the integrity of the coding sequence. The recombinant plasmids were transformed individually into E. coli BL-21 CodonPlus (DE3) RIL by heat shock and grown overnight at 37 °C on Lysogeny Broth (LB) agar plates supplemented with ampicillin (100 μg/ml) and chloramphenicol (50 μg/ml). A single colony from each plate was picked and pre-cultured at 37 °C for 8 h in LB liquid medium (10 ml) supplemented with ampicillin (100 μg/ml) and chloramphenicol (50 μg/ml). The pre-cultures were then inoculated into fresh LB (1 liter) supplemented with the two antibiotics and cultured at 37 °C with vigorous shaking (225 rpm/min) to an absorbance of 0.3 at 600 nm (A_{600}).

To induce gene expression, isopropyl β-D-thiogalactopyranosidase was added to the culture at a final concentration of 0.1 mM, and the cells were cultured for an additional 16 h at 16 °C. The cells were harvested by centrifugation (4000 × g, 4 °C, 15 min) and resuspended in lysis buffer (30 ml, 50 mM Tris-HCl, 300 mM NaCl, pH 7.0). To release the recombinant proteins, the cell suspension was lysed by two sequential passages through an EmulsiFlex C-3 cell homogenizer (Avestin, Ottawa, Canada). The cell debris was removed by centrifugation at 20,000 × g for 20 min at 4 °C. To decrease the amount of heat-labile E. coli proteins, the supernatant was heated at 65 °C for 30 min and centrifuged at 20,000 × g for 15 min at 4 °C to pellet the denatured proteins. Because each gene was cloned in-frame with a polyhistidine tag encoded by the pET46 Ek/LIC vector, the resulting N-terminal polyhistidine (His_{6})-tagged proteins were loaded onto an immobilized metal ion affinity resin (Talon resin, Novagen) that had been pre-equilibrated with the binding buffer (50 mM Tris-HCl, 300 mM NaCl, pH 7.5). The protein/resin mixture was incubated for 1 h at 4 °C. After washing unbound proteins with 50 column volumes of binding buffer, the proteins that bound to the column were each eluted with 10 column volumes of elution buffer (50 mM Tris-HCl, 300 mM NaCl, 250 mM imidazole, pH 7.5). The purity of the eluted proteins was examined by SDS-PAGE as described by Laemmli (26). After staining with Coomassie Brilliant Blue G-250, the gel was destained with acetic acid/methanol (1:1, v/v).

For Xyn10A, the purified protein fractions were pooled, concentrated, and exchanged into anion exchange binding buffer (50 mM Tris-HCl, pH 7.0) using an Amicon Ultra-15 centrifugal filter unit (50,000 molecular mass cutoff) with three successive concentration and dilution cycles. The concentrated protein was then loaded onto a 5-ml HiTrap Q HP anion exchange column fitted to an AKTA Express chromatography system (GE Healthcare) and eluted with a linear gradient of an elution buffer (50 mM Tris-HCl, 1 mM NaCl, pH 7.0). The absorbance at 280 nm was continuously monitored, and eluted proteins were collected in 0.5-ml fractions and analyzed by SDS-PAGE. The purified fractions were pooled, concentrated, and exchanged into citrate buffer (50 mM sodium citrate, 150 mM NaCl, pH 5.5). The protein was then loaded onto a SuperdexTM 200 HiLoad™ 16/60 size exclusion column and eluted with citrate buffer at a flow rate of 1.2 ml/min. Five hundred microliter fractions were collected and analyzed by SDS-PAGE, and the purified fractions were pooled, concentrated, and exchanged into storage buffer (50 mM Tris-HCl, 150 mM NaCl, pH 7.5) using an Amicon Ultra-15 centrifugal filter unit (50,000 molecular mass cutoff).

Xyl3A, Agu67A, and XBP1 were purified as described for Xyn10A except that a 30,000 MWCO Amicon tube was used for XBP1, and the anion exchange purification step was omitted for the three proteins because two steps were sufficient to obtain pure protein.

The concentrations of the proteins were determined by use of a NanoDrop 1000 from Thermo Scientific (Waltham, MA) according to the UV absorbance at 280 nm and calculated extinction coefficients as follows: Xyn10A, 234,015 M⁻¹ cm⁻¹; Xyl3A, 87,210 M⁻¹ cm⁻¹; Agu67A, 154,590 M⁻¹ cm⁻¹; and XBP1, 53,400 M⁻¹ cm⁻¹ and molecular masses as follows: Xyn10A, 158,357 g mol⁻¹; Xyl3A, 86,760 g mol⁻¹; Agu67A, 79,020 g mol⁻¹; and XBP1, 46,168.5 g mol⁻¹.

Quaternary Structure Determination by Size Exclusion Chromatography—The quaternary structures of Xyl3A and Agu67A were analyzed by size exclusion chromatography using a Superdex 200 10/300 GL size exclusion column. One hundred microliters of Xyl3A (3 mg/ml), Agu67A (3 mg/ml), or a gel filtration standard mixture were loaded onto the column pre-equilibrated with a buffer composed of 50 mM sodium citrate, 150 mM NaCl, pH 5.5. The elution was performed at a flow rate of 0.5 ml/min. A calibration curve of molecular mass versus retention time was constructed with the gel filtration standards, and the apparent molecular masses of the two proteins were calculated by comparison of experimental retention times with the calibration curve.

Hydrolysis of para-Nitrophenol (pNP)-linked Sugars—The hydrolytic activities of the putative α-glucuronidase (Agu67A) and β-xlosidase (Xyl3A) were screened against a panel of pNP-linked substrates by a colorimetric assay using a thermostated Cary 300 UV-visible spectrophotometer (Varian Inc., Palo Alto, CA). The 15 pNP-linked substrates were as follows: pNP-α-L-arabinopyranoside; pNP-α-L-arabinofuranoside; pNP-β-D-fucopyranoside; pNP-α-L-fucopyranoside; pNP-α-D-galactopyranoside; pNP-β-D-galactopyranoside; pNP-β-D-glucopyranoside; pNP-β-D-maltopyranoside; pNP-α-D-maltopyranoside; pNP-α-D-mannopyranoside; pNP-β-D-mannopyranoside; pNP-α-L-rhamnopyranoside; pNP-β-D-xylopyranoside, and pNP-β-D-cellobioside. In the reaction, the pNP-linked substrates (1.0 mM) were incubated with Xyl3A or Agu67A (100 nM) in a citrate buffer (50 mM, pH 5.5) at 65 °C for 30 min, and the rate of pNP released in the reactions was monitored continuously through the absorbance at 400 nm. The extinction coefficient for pNP was determined experimentally by constructing a standard curve of different pNP concentrations at a pH of 5.5 and a temperature of 65 °C and by using the Beer-Lambert Law (A = c, ε, and l, where c is the concentration; ε is the molar extinction coefficient of pNP at 400 nm, and l is the length path of the cuvette). The extinction coefficient of pNP (1636 M⁻¹ cm⁻¹) obtained at pH 5.5 was used to calculate initial velocities. One unit of β-xlosidase activity is the molar extinction coefficient of pNP at 400 nm, and l is the length path of the cuvette). The extinction coefficient of pNP (1636 M⁻¹ cm⁻¹) obtained at pH 5.5 was used to calculate initial velocities. One unit of β-xlosidase activity is the molar extinction coefficient of pNP.
activity was defined as the amount of enzyme that release 1 μmol of pNP from the substrates per min.

**Determination of α-Glucuronidase Activity**—The α-glucuronidase activity was quantified using a colorimetric assay for uronic acids (27). For aldouronic acids, the substrate (6 mg/ml) was incubated with 0.5 μM Agu67A in 100 μl of citrate buffer (50 mM, pH 5.5) at 65 °C for 10 min. The reaction was terminated by adding 150 μl of copper reagent (1.97 mM Na₂SO₄, 0.68 mM NaCl, 0.2 M sodium acetate, 20.8 mM CuSO₄, pH adjusted to 4.8). The mixture was then boiled for 10 min, chilled on ice, and then 150 μl of arsenomolybdate reagent (19.2 mM Na₂H₂AsO₄·7H₂O, 40.5 mM (NH₄)₆Mo₇O₂₄·4H₂O, 788 mM H₂SO₄) was added for color development. The absorbance of the mixture was measured at 600 nm, and a standard curve was constructed using known concentrations of α-glucuronic acid (Sigma).

For polysaccharide substrates, 0.5 μM of the enzyme was incubated with BWX (1%, w/v) or MGX (1%, w/v) (both from Sigma) in citrate buffer (pH 5.5) and incubated at 65 °C for 60 min. The reactions were terminated by heating at 100 °C for 10 min. The reaction mixtures were centrifuged at 17,000 × g for 10 min, and the 4-O-methyl-α-glucuronic acid in the supernatant was determined using the colorimetric assay described above. One unit of α-glucuronidase activity was defined as the amount of enzyme that catalyzes the release of 1 μmol of α-glucuronic acid equivalents per min.

**Hydrolysis of Oligosaccharides by Xyl3A**—To analyze the hydrolytic activity of Xyl3A with oligosaccharides as substrates, xylo-oligosaccharides (xylose, X₁; xylobiose, X₂; xylotriose, X₃; xylotetraose, X₄; xylopentaose; X₅, and xylohexaose, X₆) and cello-oligosaccharides (cellobiose, G₂; cellotriose, G₃; cellotetraose, G₄; and cellopentaose, G₅) were used. Xyl3A (0.5 μM, final concentration) was incubated with each oligosaccharide (10 mg/ml, final concentration) in citrate buffer (50 mM sodium citrate, pH 5.5) in a final reaction volume of 10 μl at 65 °C for 10 h. The control reaction was performed under the same condition except with heat-denatured Xyl3A added as the enzyme. At the end of the reaction, a 20-μl volume of ethanol was added to the hydrolysate, and the mixture was evaporated through a Savant DNA120 SpeedVac® concentrator (Savant; Ramsey, MN). The dried product was resuspended in 2.5 μl of double-distilled H₂O, and 0.5 μl of each sample was spotted on Silica Gel 60 F₂₅₄ TLC plates (Merck). Monomeric xylose (X₁, 3.0 μg) and xylo-oligosaccharides (X₂–X₆) and cello-oligosaccharides (G₂–G₅) were used as standards in the TLC analysis for xylo-oligosaccharides, and glucose (G₁, 3.0 μg) and cello-oligosaccharides (G₂–G₅) were used as standards for cello-oligosaccharides analyses. After drying the plates, the products of the reactions were resolved by one ascent for 4 h with 1-butanol/acetic acid/H₂O (10:5:1, v/v/v) as a mobile phase (28). For visualization, the dried TLC plates were sprayed with a mixture of methanolic ninhydrin (0.05%, w/v) and sulfuric acid (5%, v/v) and then heated at 75 °C for 10 min (29) for color development.

**Determination of Kinetic Parameters for Xyl3A with pNP-linked Sugars as Substrates**—Kinetic parameters of Xyl3A were determined at optimum conditions with pNP-β-D-xylopyranoside (pNPX), pNP-β-D-cellobioside, and pNP-α-D-glucopyranoside (pNPGlu) as substrates. In the reactions, Xyl3A (100 nM, final concentration) was incubated with the pNP-linked substrates at 65 °C in citrate buffer (50 mM sodium citrate, pH 5.5). The substrate concentrations of pNPX and pNPGlu were used in a range of 0.08–10 mM, and the concentration of pNP-β-D-cellobioside ranged from 0.08 to 5 mM. After the substrates were equilibrated to 65 °C in the thermostated Cary 300 UV-visible spectrophotometer, the reactions were initiated by addition of Xyl3A, and the rate of pNP production was evaluated by monitoring the absorbance at 400 nm. The initial velocities of the reactions were calculated with extinction coefficient of 1636 M⁻¹ cm⁻¹ for pNP at pH 5.5. By using GraphPad software (GraphPad version 5.01, San Diego), the initial velocities were plotted against the substrate concentrations, and the Michaelis-Menten constant (Kₘ) and the maximum velocity (Vₘₐₓ) were estimated with a nonlinear curve fit. The kₑₐ was calculated as the quotient of the resulting Vₘₐₓ and the concentration of enzyme used in the reaction.

**Determination of Catalytic Efficiencies for Xyl3A with Xylo-oligosaccharides and Cello-oligosaccharides**—The catalytic constants of Xyl3A for xylo-oligosaccharides (X₁–X₆) and cellobioside ranged from 0.08 to 5 mM. The concentration of pNP-β-D-cellobioside (G₂–G₅) were determined as described previously (30). Briefly, oligosaccharides (30 μM each) were hydrolyzed with Xyl3A in citrate buffer (pH 5.5, 50 mM) at 65 °C with final volume of 500 μl, and the reactions were terminated at 10 min by boiling for another 10 min. The final concentration of Xyl3A was 50 nM for xylo-oligosaccharides (X₂–X₆) and 500 nM for cellobioside (G₂–G₅). The relationship between hydrolysis rate and oligosaccharide substrate concentration (0, 30, and 60 μM) was linear; therefore, the substrate concentration of 30 μM should be well below Kₘ. A linear relationship was also observed between substrate depletion and hydrolysis time (0, 10, and 20 min), so the hydrolytic reactions were terminated at 10 min. The substrate concentrations at the beginning ([Sᵢ]), and termination ([Sᵢ], 10 min) of the reaction were calculated by HPSEC method and used for calculation of the catalytic constants as earlier described (30).

**Substrate Binding Assay Using Isothermal Titration Calorimetry**—The substrate binding activity of XBP1, which constitutes a component of the putative sugar transport system, in the gene cluster was measured at 25 °C using a VP-ITC microcalorimeter (Microcal Inc., Northampton, MA) with different ligands. The ligands (xylose, xylobiose, xylotriose, xylotetraose, glucose, cellobiose, cellotriose, cellotetraose, and aldouronic acids) were dissolved in citrate buffer (50 mM sodium citrate, pH 5.5) to a concentration of 0.5 mM (oligosaccharides) or 0.24 mg/ml (aldouronic acids). The protein XBP1 was diluted to a final concentration of 50 μM using the same buffer. The ligand was injected into the reaction cell containing XBP1 in 28 successive 10-μl aliquots at 300-s intervals and 20 s duration. Nonlinear regression with a single site model (MicroCal Origin) was applied for data analysis, and thermodynamic parameters were calculated using the Gibbs free energy equation \( \Delta G = \Delta H - T\Delta S \), and the relationship, \( \Delta G = -RT\ln K \).

**Structure Determination of the XBP1-Xylotriose Complex**—XBP1 was purified as described above but with an additional size exclusion chromatographic step (using a Superdex™ 200 HiLoad™ 16/60 size exclusion column) in a final buffer com
posed of 20 mM HEPES, pH 7.5, 100 mM KCl. The affinity tag was not removed prior to crystallization. Initial crystallization conditions were obtained by the sparse matrix sampling method using commercial screens. Crystals of the XBP1-xylotriose complex were grown using the hanging vapor drop diffusion method. Briefly, 1 µl of protein at 13.5 mg/ml concentration was incubated with 5 mM xylotriose for 2 h on ice, and the complex was mixed with 1 µl of precipitant solution (30% polyethylene glycol 1500, 100 mM KCl, and 20 mM HEPES, pH 7.5) and equilibrated over a well containing the precipitant solution at 9 °C. Crystals grew within 3 days and were briefly soaked in precipitant solution supplemented with 10% ethylene glycol prior to flash-cooling in liquid nitrogen. Selenomethione-labeled XBP1 was grown as described above, and crystals of SeMet XBP1 were grown under similar conditions.

Flash-cooled crystals of native XBP1 in complex with xylotriose diffracted x-rays to 2.1 Å resolution at an insertion device synchrotron beam line (LS-CAT Sector 21 ID-F, Advanced Photon Source, Argonne, IL). Crystals of selenomethione-labeled XBP1-xylotriose complex diffracted to a slightly higher resolution and subsequently were used for all structural analyses. These crystals occupy space group P21212 with unit cell parameters a = 59.3 Å, b = 150.8 Å, and c = 150.9 Å, with three molecules in the crystallographic asymmetric unit. Although two of the unit cell constants are suspiciously close and indicative of a higher symmetry setting, the data could not be integrated or scaled in any tetragonal space group.

An 8-fold redundant data set was collected to 1.8-Å resolution with an overall Rmerge = 9.4 and I/σ(I) = 8 in the highest resolution shell. All data were indexed and scaled using the HKL2000 package (31). Crystallographic phases were determined by single wavelength anomalous scattering. The heavy atom substructure was determined using HySS (32), and refinement with an overall Rmerge of 22.9 (26.0%).

Refinement:

Resolution 50-1.8 Å (1.86-1.8 Å)
No. of reflections 119,409
Rmerge 22.9/26.0%
No. of atoms Protein 4196 Solvent 998 Xylotriose 56 Average B value Protein 16.8 Solvent 24.6 Xylotriose 7.9 Root mean square deviations Bond angles 1.00 Å Bond lengths 0.007 Å

The protein was cloned and expressed as a recombinant protein bearing a peptide cleavage site between amino acid residues 30 and 31. The recombinant protein was purified to near-homogeneity. The predicted modular architecture for the predicted protein includes a GH10 endoxylanase module flanked on the N terminus by two family 9 endoglucanase domains and on the C terminus by two family 22 CBMs (CBM 22) and on the C terminus by two family 9 CBMs (CBM 9). In addition, there are three surface layer homology (SLH) modules at the C terminus of the protein (Fig. 1A). A SignalP search revealed a predicted signal peptide with a peptidase cleavage site between amino acid residues 30 and 31. The protein was cloned and expressed as a recombinant protein in E. coli and purified to near-homogeneity. The predicted gene, named xyn10A, was used as the reference gene. The primers used for the experiments are listed in supplemental Table S1.

RESULTS

Identification of Xyn10A, a Multimodular Endoxylanase—
To identify genes involved in xylan degradation, we searched the genome of C. polysaccharolyticus for putative endoxylanase genes using a BlastP search. This search revealed xyn10A, a gene predicted to encode a multimodular endoxylanase. The modular architecture for the predicted protein includes a GH10 endoxylanase module flanked on the N terminus by two family 22 CBMs (CBM 22) and on the C terminus by two family 9 CBMs (CBM 9). In addition, there are three surface layer homology (SLH) modules at the C terminus of the protein (Fig. 1A). A SignalP search revealed a predicted signal peptide with a peptidase cleavage site between amino acid residues 30 and 31. The protein was cloned and expressed as a recombinant protein in E. coli and purified to near-homogeneity. The predicted

TABLE 1

| Data collection, phasing, and refinement statistics | Selenomethionine XBP1-xylotriose |
|-----------------------------------------------|-----------------------------------|
| Data collection PDB accession no. | 4G68 |
| Space group | P21212 |
| Cell dimensions | a, b, c | 59.3, 150.8, 150.9 Å |
| Resolution | 50-1.8 Å (1.86-1.8 Å) |
| Rmerge | 9.4 (24.2%) |
| I/σ(I) | 37.9 (8.6) |
| Completeness | 99.8% (98.4%) |
| Total reflections | 998,584 |
| Unique reflections | 126,799 |
| Redundancy | 7.9 (7.2) |
| FOM* | 0.447 |

* Highest resolution shell is shown in parentheses.
* FOM means figure of merit = [P(obs) exp(iθ)]/P(calc) and R-free is the R value for a test set of reflections consisting of a random 5% of the diffraction data not used in refinement.
molecular mass of the purified protein as judged by SDS-PAGE (170 kDa, Fig. 1B) was in agreement with the predicted value based on the amino acid composition (160.1 kDa). Following overnight incubation, of the purified protein with BWX, an increase in reducing ends was detected (Fig. 1C), indicating that the recombinant protein possesses endoxylanase activity. Next, Xyn10A was incubated with BWX, and aliquots were taken at 0, 5, 10, 30, and 60 min and analyzed by thin layer chromatography. By 60 min, a mixture of products that included xylotriose and xylotetraose appeared on the TLC plate (Fig. 1D). Therefore, xyn10A encodes an endoxylanase and the major products of hydrolysis on BWX are short xylo-oligosaccharides.

Identification of a Gene Cluster Targeted toward Xylo-oligosaccharides—Xyn10A is an extracellular endoxylanase that produces mainly xylotriose and xylotetraose from BWX. To utilize the products of hydrolysis of Xyn10A, C. polysaccharolyticus must transport these oligosaccharides into the cell, cleave them into monosaccharides, and metabolize them, most likely through the pentose-phosphate pathway. In studying the genome, we identified an 18-kb region of the genome containing a cluster of genes with predicted roles of transport, hydrolysis, and metabolism of xylo-oligosaccharides (Fig. 2). This cluster includes genes with predicted involvement in oligosaccharide transport (ABC transporter, ORF0542–0545), hydrolysis of branched oligosaccharides (α-glucuronidase, ORF0540, and β-xylosidase, ORF0541), removal of acetyl groups from oligosaccharides (ORF0551), metabolism of xylose through the pentose-phosphate pathway (ORF0542–0545), and a two-component system that could be involved in regulation of these genes at the transcriptional level (ORF0546–0547).

Therefore, this gene cluster includes genes predicted to encode the entire repertoire of proteins required for the transport and metabolism of hydrolytic products of Xyn10A. To test whether genes in this cluster function in the degradation and utilization of xylan, several critical components in the cluster were studied further as described below.

**xyl3A Encodes an Enzyme with Both β-1,4-Xylosidic and β-1,4-Glucosidic Activities—**ORF0541 was predicted to encode a β-xylosidase consisting of three domains as follows: an N-terminal GH 3 domain followed by a C-terminal GH 3 domain, and an FN3-like domain (supplemental Fig. 1). The gene was cloned and expressed in E. coli, and the recombinant N-terminal hexahistidine-tagged protein was purified to near-homogeneity. SDS-PAGE analysis of the purified Xyl3A revealed a single band with a molecular mass of about 86 kDa, which was in agreement with the predicted molecular mass of the hexahistidine fusion protein based on amino acid sequence (88 kDa) (Fig. 3A). The quaternary structure of Xyl3A was further determined through size exclusion chromatography. As shown in Fig. 3B, Xyl3A eluted in a single peak, and the apparent molecular mass was calculated as about 157 kDa. The molecular mass in size exclusion chromatography was nearly two times of that shown in SDS-PAGE, suggesting that Xyl3A exists as a dimer in solution.

The hydrolysis of 15 different pNP-linked sugar substrates by Xyl3A showed that it was most active on pNP-β-D-xylopyranoside (pNPX, specific activity 210 milliunits/mg). Therefore, the temperature and pH optima of Xyl3A were determined with pNPX in a temperature range of 40–75 °C and a pH range of 4.0–6.5. These results revealed a temperature optimum of 65 °C and a pH optimum of 5.5 (data not shown). The catalytic properties of Xyl3A, based on these parameters, indicated that the catalytic efficiency for Xyl3A is higher with pNP-Glu than pNPX (Table 2). However, these are artificial substrates and may not accurately reflect the substrate specificity for natural substrates.

To further test the natural substrate specificity of Xyl3A, the protein (0.5 μM) was incubated with xylo-oligosaccharides (X₃–X₆) and cello-oligosaccharides (G₃–G₆), and the products were resolved by thin layer chromatography. Following overnight incubation of Xyl3A with xylo-oligosaccharides, all of the substrates were converted to xylose (Fig. 3D). However, when Xyl3A was incubated with cello-oligosaccharides at a concen-

---

**FIGURE 1. Properties of Xyn10A, the major endoxylanase in C. polysaccharolyticus.** A modular organization for Xyn10A. NCBi conserved domain database and SignalP version 4.0 server were used for analysis. B, SDS-PAGE analysis of purified recombinant Xyn 10A. Xyn 10A was purified by cobalt affinity chromatography, followed by gel filtration and ion exchange chromatography as described under “Experimental Procedures.” C, hydrolysis of BWX. Xyn10A was incubated with BWX for 15 h and analyzed using a reducing sugar assay. D, Xyn10A was incubated with BWX for the indicated time and then hydrolytic products were analyzed by thin layer chromatography.

**FIGURE 2. Identification of a xylanolytic gene cluster in the genome of C. polysaccharolyticus.** Within the genome sequence of C. polysaccharolyticus, an 18.4-kb region was found that contains genes with predicted roles in degradation, transport, metabolism, and transcriptional response to xylan fragments. Gene annotations were performed by the rapid annotations using subsystems technology (RAST) server (24). Glycoside hydrolase families are shown based on designations from the carbohydrate-active enzymes (CAZY) database (61).
Xylan Utilization by Caldanaerobius polysaccharolyticus

![Figure 3](image-url)

**TABLE 2**

| Substrate                          | $k_{cat}/K_m$ |
|------------------------------------|--------------|
| $p$NP-β-D-xylopyranoside$^{ac}$    | 31 ± 1       |
| $p$NP-β-D-glucopyranoside$^{ac}$   | 51 ± 0.2     |
| $p$NP-β-D-cellobiose$^{ab}$        | 29 ± 0.2     |
| Xylobiose$^{bc}$                   | 62 ± 0.4$^a$|
| Xylotrione$^{ad}$                  | 80 ± 0.3     |
| Xylohexaose$^{ad}$                 | 75 ± 0.4     |
| Xylopentaose$^{ad}$                | 67 ± 0.2     |
| Xylotetraose$^{ad}$                | 46 ± 0.2     |
| Cellobiose$^{ad}$                  | $(2.8 ± 0.6) \times 10^{-4}$ |
| Cellotriose$^{ad}$                 | $(3.3 ± 0.1) \times 10^{-4}$ |
| Cellopentaose$^{ad}$               | $(2.2 ± 0.003) \times 10^{-2}$ |
| Cellohexaose$^{ad}$                | $(2.8 ± 0.2) \times 10^{-2}$ |

$^a$ The experiments were performed in triplicate, and data are reported as means ± S.D.
$^b$ The catalytic efficiencies ($k_{cat}/K_m$) are reported as mmol s$^{-1}$.
$^c$ The catalytic efficiencies ($k_{cat}/K_m$) for $p$NP-β-D-xylopyranoside, $p$NP-β-D-glucopyranoside, and $p$NP-cellobiose were determined as described by Dodd et al. (40).
$^d$ The catalytic efficiencies ($k_{cat}/K_m$) for xyl-oligosaccharides and cello-oligosaccharides were determined as described by Han et al. (30).

To analyze the catalytic activity for Xyl3A with oligosaccharides more quantitatively, an HPAEC-PAD assay was employed as described under “Experimental Procedures.” These experiments showed that the catalytic efficiency ($k_{cat}/K_m$) values for Xyl3A were much lower for cello-oligosaccharides compared with xyl-oligosaccharides (Table 2). The catalytic efficiencies were 5 orders of magnitude lower for cellobiose and cellotriose relative to xylobiose and xylotriose and 3 orders of magnitude lower for cellopentaose, cellopentaose, and cellohexaose relative to the corresponding xyl-oligosaccharides (Table 2). These results clearly show that the dominant catalytic activity for Xyl3A is β-xylosidase activity.

Subsequent studies with polysaccharides, including BWX, oat spelt xylan, MGX, mannan, locust bean gum, guar gum, and glucomannan, revealed detectable activity against BWX, oat spelt xylan, and MGX, with the predominant product being xylose (data not shown). Xyl3A also showed modest activity on CMC with a monosaccharide, likely representing carboxymethylated glucose, identified in the hydrolysate by HPAEC-PAD. However, no activity was detected with mannan, locust bean gum, guar gum, or glucomannan as substrates (data not shown). These results further supported the assignment of Xyl3A as a β-xylosidase.

**Agu67A Encodes an α-Glucuronidase—**4-O-Methylglucuronate groups are common side chains found attached to the backbone xylopyranosyl groups in xylans. Thus, to completely convert xylan to monosaccharides for metabolism, debranching enzymes such as α-glucuronidases are required. ORF5040 was predicted to encode a GH 67 α-glucuronidase and was targeted for further analysis. The domain architecture includes three
conserved domains, including an N-terminal GH 67 domain, a GH 67 middle domain, and a C-terminal GH 67 domain (supplementary Fig. 1). Agu67A was expressed as a hexahistidine fusion protein in E. coli and purified to homogeneity. SDS-PAGE analysis of the purified Agu67A revealed a single band with a molecular mass of about 79 kDa, which was in agreement with the predicted molecular mass of the hexahistidine fusion protein based on amino acid sequence (81 kDa) (Fig. 4A). The quaternary structure of Agu67A was further determined through size exclusion chromatography. As shown in Fig. 4B, Agu67A eluted in a single peak, and the apparent molecular mass was calculated as 158 kDa. The molecular mass in size exclusion chromatography was two times that shown by SDS-PAGE, suggesting that Agu67A exists as a dimer in solution, which is in agreement with other studies on bacterial H9251-glucuronidases (41).

The hydrolytic activity of Agu67A was screened with pNP-linked substrates, aldouronic acids, and polysaccharides. Agu67A showed the highest activity against aldouronic acids, the specific activity reached 154 IU/mg (Fig. 4C). Lower hydrolytic activity was also detected with MGX and BWX as substrates with specific activities of 10.8 and 11.6 IU/mg, respectively. The observable activity with MGX and BWX most likely is due to the hydrolysis of small amounts of aldobioturonic and aldotreiouronic acids present in the polysaccharide mixtures, because HPLC assays of the hydrolysate revealed the appearance of xylose and xylobiose following incubation with Agu67A (data not shown). No activity was observed with any pNP-linked sugars.

The optimum temperature and pH for Agu67A was determined with aldouronic acids as substrates. The enzyme was active in a temperature range of 30–85 °C, and the optimum temperature was 60 °C (data not shown). Analysis of the activity in buffers of pH values ranging from 3.5 to 6.5 revealed a pH optimum of 5.5 (data not shown).

To further evaluate the activity of Agu67A with aldouronic acids, the enzyme was incubated with an aldouronic acid mixture, and the products of hydrolysis were analyzed by HPAEC-PAD. After incubation of Agu67A with aldouronic acids, peaks appeared corresponding to xylobiose, xylotriose, xylotetraose, and xylopentaose, and the peak representing xylose increased in the enzyme-treated mixture as compared with the control reaction (Fig. 4D). This observation indicated that Agu67A cleaved 4-O-methylglucuronic acid from aldobioturonic, aldotreiouronic, aldotreiouronic, and aldopentaouronic acids. Furthermore, the concentration of glucuronic acid equivalents increased from 36 mM in the control reaction to 370 mM in the enzyme-treated reaction (Fig. 4E). These results clearly showed that Agu67A produces xylo-oligosaccharides and glucuronic acid equivalents from a mixture of 4-O-methylglucuronol-sub-

![FIGURE 4. Purification and biochemical characterization of recombinant Agu67A. A, module architecture for Agu67A. B, quaternary structure analysis of Agu67A by gel filtration chromatography. The apparent molecular mass of Agu67A was estimated by comparing the retention time with calibration standards of known molecular weights. mAU, milliabsorbance units. C, hydrolytic activity of Agu67A against aldouronic acids mixture, MGX and BWX. Agu67A was incubated with the three substrates indicated above, and the specific activity was determined by detecting glucuronic acid equivalents as described under "Experimental Procedures." D and E, Agu67A was incubated with a mixture of aldouronic acids, and the products of hydrolysis were analyzed by HPAEC-PAD (D) or by quantifying glucuronic acid equivalents (E). D, peak assignments were made by comparison with a calibration curve made with known concentrations of the corresponding oligosaccharides. B–E, experiments were performed in triplicate, and data are reported as means ± S.D. D, representative HPAEC traces are shown for the control and Agu67A-treated reactions.](image-url)
stituted xylo-oligosaccharides, confirming its activity as an α-glucuronidase.

Synergism of Xyn10A, Xyl13A, and Agu67A in the Hydrolysis of BWX—Given our finding that Xyn10A releases oligosaccharides from BWX, we hypothesized that Agu67A and Xyl13A may facilitate the conversion of oligosaccharides to monosaccharides. To evaluate this hypothesis, we incubated the three enzymes separately and in combination with BWX and analyzed the products of hydrolysis. The three enzymes functioned synergistically to release xylose from BWX (supplemental Fig. 2). These results support our hypothesis that Xyn10A produces branched xylo-oligosaccharides that are subsequently transported into the cell and degraded into monosaccharides by Agu67A and Xyl13A for fermentation by C. polysaccharolyticus.

ORF0548 Encodes a Substrate-binding Component of an ABC Transporter That Is Specific for Xylo-oligosaccharides—Xyn10A is an endoxylanase that produces branched xylo-oligosaccharides. The coding sequence includes a putative signal peptide and three SLH repeats, which indicates that it is secreted outside of the cell and anchored onto the cell wall. The two enzymes that catalyze the hydrolysis of these branched oligosaccharides into monosaccharides do not possess signal peptides and are therefore likely to reside within the cytoplasm of the cell. Therefore, a mechanism must exist whereby branched oligosaccharides are transported across the plasma membrane and into the cell. Within the gene cluster identified in C. polysaccharolyticus, a putative ATP-binding cassette (ABC) importer was found that contains three genes predicted to include Gln-159 to Gly-324 and Asp-378 to the C terminus. The overall structure of the XBP1-xylotriose complex was solved to 1.8-Å resolution by single wavelength anomalous diffraction methods on data collected from crystals of a selenomethionine-labeled protein. The overall structure shows a bilobal fold that is common to other solute-binding components of ABC-type transporters (Fig. 6A). The first domain consists of residues Ile-46 through Asp-158 and residues Gly 325 through the Val-377, and the second domain includes Gln-159 to Gly-324 and Asp-378 to the C terminus. A DALI search of a structure-based comparison against the Protein Data Bank reveals the closest homologs to be the Actinobacteria-specific soluble receptor AcbH from Actinoplanes (PDB code 3O0G; Z-score = 44.4; r.m.s.d. of 2.2 Å over 378 aligned Ca atoms) (42), the glucose/galactose-specific solute-binding protein from Thermus thermophilus (PDB code 2B3B; Z-score = 41.4; r.m.s.d. of 2.4 Å over 373 aligned Ca atoms) (43), and the trehalose/maltose-specific binding protein from Thermococcus litoralis (PDB code 1EU8; Z-score = 39.3;
r.m.s.d. of 2.5 Å over 375 aligned C atoms) (44), among others. The homology of XBP1 to each of these proteins exists only at the structural level, as identity between their primary sequences is below 18%. Additionally, in contrast to XBP1, the structures of each of these proteins have only been determined in complex with monosaccharides.

A DALI comparison of XBP1 against the structures of solute-binding proteins complexed with oligosaccharides identifies the oligogalacturonide-binding protein from Yersinia enterocolitica (PDB code 2UVJ; Z-score 33.0; r.m.s.d. of 3.0 Å over 369 aligned C atoms) (45), the ascarbose-binding protein from Streptomyces glaucescens (PDB code 3JZJ; Z-score 35.9; r.m.s.d. of 2.6 Å over 365 aligned C atoms) (46), and the Streptococcus pneumoniae solute-binding protein in complex with blood group A-trisaccharide (PDB code 2W7Y; Z-score 30.2; r.m.s.d. of 3.1 Å over 355 aligned C atoms) (47). Based on the structure-based classification of solute-binding protein (48), XBP1 belongs to cluster B, which is specific for carbohydrates and branched amino acids and natriuretic peptides.

In the co-crystal structure, the xylotriose ligand is bound in a (roughly) 14-Å cleft that is formed between the two domains (Fig. 6B). An inspection of the binding pocket provides a molecular rationale for the specificity of oligosaccharide chain length. One end of the pocket is capped by a number of aromatic and aliphatic side chains, including Trp-109, Met-211, Trp-212, Ile-329, and Ala-398. The third xylose residue of the trisaccharide is engaged at this site and stacks directly above Trp-208. The second xylose residue is sandwiched between Leu-53 and Phe-287, and the first xylose is packed against Phe-54 and a loop encompassing Thr-55 through Lys-60. This binding pocket is contoured optimally for a trisaccharide substrate; a disaccharide can be accommodated but would not engage in all van der Waals packing interactions, and binding of tetra- or longer oligosaccharides would require movement of the loop encompassing Thr-55 through Lys-60. Hence, the structural data are consistent with the calorimetric analysis that illustrates the preference for a trisaccharide ligand.

The specificity for xylo-oligosaccharides could also be understood in the context of the structural data. In addition to hydrophobic contacts, an extensive set of hydrogen bond interactions stabilizes the interaction between the protein and each of the sugars of the trisaccharide ligand. Polar residues are situated along the length of the ligand-binding cleft where they interact with the hydroxyl groups of the xylotriose. At the site of the first xylose residue (C-1 anomeric carbon), which harbors the reducing end of the xylotrisaccharide, O-3, is within hydrogen bonding distance to Arg-267, and O-1, is within hydrogen bonding distance to the backbone nitrogen of Gly-56. At site 2, both Gln-160 and the indole nitrogen of Trp-109 engage O-3, and Arg-267 interacts with the hemiacetal oxygen O-5. At site 2, both Gln-160 and the indole nitrogen of Trp-109 engage O-3, and Arg-267 interacts with the hemiacetal oxygen O-5.
Xylan Utilization by Caldanaerobius polysaccharolyticus

FIGURE 7. Induction of xylanase genes during growth of C. polysaccharolyticus with BWX as compared with glucose. A, C. polysaccharolyticus was cultured in a defined medium with either BWX or glucose as the major carbohydrate source at 65 °C, and the A_{600nm} values were monitored over time. B, cells were harvested at an A_{600nm} of 0.2 (glucose) or 0.1 (BWX), and RNA was extracted, and quantitative reverse transcription-PCR experiments were performed as described under "Experimental Procedures." Four technical replicates of the Q-PCR were performed for each of three independent biological replicates, and data are reported as means ± S.D.

Of the 13 genes studied, 11 were induced greater than 2-fold (Fig. 7). The most highly induced genes included all three members of the ABC transporter (XBP1, 117-fold; permease, 63-fold; and ATP-binding domain, 42-fold), the putative polysaccharide deacetylase (ORF0550, 98-fold), and the α-glucuronidase (agu67A, 19-fold). Although not induced as strongly as the transporters, esterase, and α-glucuronidase, the endoxylanase gene xyn10A were still induced greater than 5-fold.

These results indicate that these genes are involved in xylan utilization by C. polysaccharolyticus, and furthermore, the bacterium has evolved a transcriptional program that permits the recognition of xylan fragments and subsequently induces the expression of genes involved in its degradation and fermentation.

DISCUSSION

Xyn10A is a multimodular endoxylanase composed of a GH 10 endoxylanase module flanked on the N terminus by a tandem repeat of CBM 22 and on the C terminus by another tandem repeat of CBM 22, followed by three SLH modules (Fig. 1A). This modular organization is conserved among six other bacteria, two of the genus Thermoaerobacter and four of the genus Thermoaerobacterium (supplemental Fig. 3). Beyond these organisms, there are hundreds of homologous proteins (defined as possessing at least one copy of the CBM 22, GH 10, and CBM 9 in the same orientation within a single polypeptide) in the GenBank™ database derived from other bacteria (supplemental Fig. 3). Interestingly, all of the organisms that possess a homolog of Xyn10A are thermophilic or hyperthermophilic bacteria, suggesting that this modular organization imparts an advantage to degrading xylan at elevated temperatures. Studies with XynA from Thermoaerobacterium saccharolyticum (TsXynA), XynA from Thermotoga maritima (TmXynA), and XynC from Paenibacillus barcinonensis (PbXynC) found the N-terminal CBM 22 to be critical for imparting thermostability and thermophilicity to the respective enzymes (50–52). In addition to their thermostabilizing properties, the CBM 22 modules also possess carbohydrate binding activity. A polypeptide composed of both N-terminal CBM 22s of TmXynA bound both soluble xylan as well as mixed linkage β-1,3/β-1,4-glucan but did not bind to crystalline cellulose (52). Furthermore, a polypeptide composed only of the second CBM 22 of TmXynA

3, Lys-90 interacts with O-2_2; Asp-400 engages both O-3_3 and O-4_4 atoms of the sugar in a bidentate manner, and Gln-160 is described under "Experimental Procedures." Four technical replicates of the Q-PCR were performed for each of three independent biological replicates, and data are reported as means ± S.D.
possessed similar binding characteristics to the polypeptide containing both modules (52).

The C-terminal CBM 9 of TmXynA and the noncellulosomal protein XynX from Clostridium thermocellum (CtXynX) possess cellulose binding activities and allow the respective enzymes to bind crystalline cellulose (53, 54). Despite binding to crystalline cellulose, neither of these enzymes have the capacity to degrade this polysaccharide. Although the xylan-binding CBM 22 of these proteins likely aids in juxtaposing substrate and catalyst, the role of the cellulose-binding CBM 9 is less clear. However, in intact plant cell walls, xylans are found in close proximity to cellulose fibers; therefore, one possible function of the CBM 9 could be to aid in the separation and degradation of insoluble xylan fragments closely associated with cellulose fibrils. Although the precise function of CBM 9 in xylan degradation by these organisms remains unclear, the high level of conservation of this modular architecture in these proteins across diverse bacteria clearly suggests that these modules are integral to xylan degradation.

Immunogold labeling and electron microscopy revealed that TmXynX is tethered to the outer membrane (toga) of T. maritima by a hydrophobic stretch of amino acids within the N-terminal signal peptide (55). Although there is a signal peptide within CpXyn10A, there is no significant homology between the signal peptides for the two xylanases nor is there a predicted signal peptidase II cleavage site that might facilitate transfer to a lipid moity. Rather, CpXyn10A is most likely tethered to the surface of the bacterium via the three SLH repeats at the C terminus of the protein. SLH modules recognize and bind to pyruvylated cell wall polysaccharides (56). The enzyme that mediates this modification is encoded by the csaB gene (56), and a homolog of this gene is present within the genome of C. polysaccharolyticus (data not shown), indicating that this mechanism is active in this organism.

Xyl3A is a bifunctional \( \beta \)-xylosidase/\( \beta \)-glucosidase with higher activity with \( p \)NPG relative to \( p \)NPX; however, the catalytic efficiency for xylo-oligosaccharides was several orders of magnitude higher than for cello-oligosaccharides indicating that cleaving debranched xylan fragments is the most likely activity of this protein. In addition to possessing the N-terminal \( (\alpha/\beta)_{8} \) and C-terminal \( \beta \)-sandwich domains characteristic of GH 3 enzymes (57), Xyl3A also has a C-terminal fibronectin repeat 3-like (FnIII-like) domain (supplemental Fig. S1). FnIII domains are thought to have originated in animals and transferred to bacteria (58), where they exist almost exclusively in association with glycoside hydrolase enzymes (59). The recent crystal structure of a three-domain GH 3 \( \beta \)-glucosidase from Thermotoga neapolitana (TnBgl3B) showed that the domain did not make contacts with the active site and adopts a FnIII fold that is distinct from those found in animals as well as those found in other bacterial glycoside hydrolases (60). Domain III of TnBgl3B was annotated as FnIII-like domain, and according to the Pfam database, 3290 of the total 3379 FnIII-like domains are associated with a GH 3 domain. Further-
more, 2949 of the total 6490 GH3 N domains are associated with FnIII-like domains, and the domain organization seen for Xyl3A is the most prevalent for GH 3 proteins in the Pfam database (data not shown). Despite the high abundance of proteins with this domain architecture, the role of this domain in GH3 enzymes is still unknown.

XBP1 is the solute-binding component of an ABC transporter and specifically binds to xylo-oligosaccharides with a preference for xylotriose. Although four sequenced *Thermoanaerobacterium* spp. have uncharacterized homologs of XBP1 ranging in amino acid identity from 74 to 87%, the most closely related protein with demonstrated activity is XynE from *G. stearothermophilus* (52% identity). Similar to XBP1, GsXynE binds preferentially to xylotriose, although it prefers xylo-oligosaccharides to monosaccharide, which is in contrast to XBP1 (8). The gene encoding GsXynE is located within a 39.7-kb gene cluster containing hemicellulose utilization genes (6), and its expression is regulated by a nearby two-component system (8). This arrangement is similar to that for XBP1 in that a putative two-component system lies just upstream of the ABC transporter genes, and it is likely that this system mediates the transcriptional response of these genes in the presence of xylan.

Crystallographic studies of solute-binding proteins illustrate that specificity is mediated by interactions that are largely local to the binding pocket. A comparison of the co-crystal structure of the XBP1-xylotriose complex with other solute-binding proteins illustrates that the binding site for the polysaccharide ligand in XBP1 is optimized for xylotriose. For example, in the glucose-binding protein from *T. thermophilus* (PDB code 2B3B), the binding site is optimized for mono- and disaccharides, and longer oligosaccharides are occluded by the protrusion of a single residue (His-348) at site 3 and a depression of the loop equivalent to Thr-55 through Lys-60 at site 1. The structure of XBP1 illustrates a binding pocket with a contour that is optimized for trisaccharides, which contains polar residues that are suited for interactions with xylose residues at each of the three sites. Our structural and biochemical data further refine the idea that for solute-binding proteins ligand specificity is achieved within the confines of a highly conserved scaffold through modest changes at the binding site.

The gene cluster identified in *C. polysaccharolyticus* appears to be targeted toward the utilization of xylans and specifically 4-O-methylglucuronoxylans as suggested by the presence of an α-glucuronidase gene within the cluster. The biological relevance of this gene cluster to xylan utilization by *C. polysaccharolyticus* is demonstrated by Q-PCR experiments that reveal many of these genes to be induced during growth on BWX relative to glucose. The genes located within this cluster encode the entire repertoire of enzymes required for the following: (a) transport xylan fragments across the cell membrane; (b) cleave branched oligosaccharides to monosaccharides; (c) metabolize xylose through the pentose-phosphate pathway, and (d) coordinate expression of xylanolytic genes in response to environmental availability of this substrate (Fig. 8).

The use of thermophilic organisms capable of fermenting both glucose and xylose in the production of biofuels is highly desirable in this emerging industry. In some instances, attempts are made to engineer an organism that can already use glucose to also metabolize xylan or xylose. The process often involves assembly of individual genes on a cassette to transform into the desired organism. The presence of the genes encoding the hydrolytic enzymes and the key enzymes of the pentose-phosphate pathway in a cluster makes it easier to transfer a co-evolved molecular machinery to confer the xylanolytic phenotype.

**Acknowledgment**—We thank Young-Hwan Moon for technical assistance.

**REFERENCES**

1. Dodd, D., and Cann, I. K. (2009) Enzymatic deconstruction of xylan for biofuel production. Global Change Biol. Bioenergy 1, 2–17
2. Yeoman, C. J., Han, Y., Dodd, D., Schroeder, C. M., Mackie, R. I., and Cann, I. K. (2010) Thermostable enzymes as biocatalysts in the biofuel industry. Adv. Appl. Microbiol. 70, 1–55
3. Baba, T., Shinkle, R., and Nannorni, T. (1994) Identification and characterization of clustered genes for thermostable xylan-degrading enzymes, β-xyloligosaccharides and xylanase, of *Bacillus stearothermophilus* 21. Appl. Environ. Microbiol. 60, 2252–2258
4. Gasparic, A., Martin, J., Daniel, A. S., and Flint, H. J. (1995) A xylan hydrolyase gene cluster in *Prevotella ruminicola* B114. Sequence relationships, synergistic interactions, and oxygen sensitivity of a novel enzyme with exoxylanase and β-(1,4)-xylodisaccharide activities. Appl. Environ. Microbiol. 61, 2958–2964
5. Jun, H. S., Ha, J. K., Malburg, L. M., Jr., Verrinder, G. A., and Forsberg, C. W. (2003) Characteristics of a cluster of xylanase genes in *Fibrobacter succinogenes* 585. Can. J. Microbiol. 49, 171–180
6. Shulami, S., Gat, O., Sonenshein, A. L., and Shoham, Y. (1999) The glucoronid acid utilization gene cluster from *Bacillus stearothermophilus* T-6. J. Bacteriol. 181, 3695–3794
7. Hurbise, Y., Shareeck, F., Klupfel, D., and Morosoli, R. (1995) A cellulase/xylanase-negative mutant of *Streptomyces lividans* 1326 defective in cellulosome and xylolysis uptake is mutated in a gene encoding a protein homologous to ATP-binding proteins. Mol. Microbiol. 17, 367–377
8. Shulami, S., Zaiade, G., Zolotnitsky, G., Langut, Y., Feld, G., Sonenshein, A. L., and Shoham, Y. (2007) A two-component system regulates the expression of an ABC transporter for xylo-oligosaccharides in *Geobacillus stearothermophilus*. Appl. Environ. Microbiol. 73, 874–884
9. Cullen, D., and Kersten, P. (1992) In Applied Molecular Genetics of Fila-mentous Fungi (Kinghorn, J. R., and Turner, G., eds) pp. 100–131. Chapman & Hall, New York
10. Mierzwia, M., Tokarzewka-Zadora, J., Deput, T., Rogalski, I., and Szc-zodrak, J. (2005) Purification and characterization of an extracellular α-li-guronidase from *Phlebia radiata*. Prep. Biochem. Biotechnol. 35, 243–256
11. Czjzek, M., Ben David, A., Bravman, T., Shoham, G., Henrisat, B., and Shoham, Y. (2005) Enzyme-substrate complex structures of a GH39 β-xyl-losidase from *Geobacillus stearothermophilus*. J. Mol. Biol. 353, 838–846
12. Eneskeya, E. V., Ivanen, D. R., Bobrov, K. S., Isaeva-Ivanova, L. S., Shaba-lin, K. A., Savelev, A. N., Golubev, A. M., and Kulminskaya, A. A. (2007) Biochemical and kinetic analysis of the GH3 family beta-xylosidase from *Aspergillus awamori* X-100. Arch. Biochem. Biophys. 457, 225–234
13. Rohman, A., van Oosterwijk, N., Kralj, S., Dijkhuizen, L., Dijkstra, B. W., and Puspaningsih, N. N. (2007) Purification, crystallization, and preliminary x-ray analysis of a thermostable glycoside hydrolase family 43 β-xyloligosaccharides from *Geobacillus thermocumovans* IT-08. Acta Crystalogr. Sect. F. Struct. Biol. Cryst Commun. 63, 932–935
14. Shallow, D., and Shoham, Y. (2003) Microbial hemicellulases. Curr. Opin. Microbiol. 6, 219–228
15. De Wet, B. J., Van Zyl, W. H., and Prior, B. A. (2006) Characterization of the *Aureobasidium pullulans* α-glucuronidase expressed in *Saccharomy-ces cerevisiae*. Enzyme Microb. Technol. 38, 649 –656
16. Nagy, T., Nurizzo, D., Davies, G. J., Biely, P., Lakey, J. H., Bolam, D. N., and
Xylan Utilization by Caldanaerobius polysaccharolyticus

Gilbert, H. J. (2003) The α-glucoronidase, GlcA67A of Cellvibrio japonicus utilizes the carboxylate and methyl groups of aldobionauronic acid as important substrate recognition determinants. J. Biol. Chem. 278, 20286–20292

17. Ryabova, O., Vrsanska, M., Kaneko, S., van Zyl, W. H., and Biely, P. (2009) A novel family of hemicellulolytic α-glucuronidase. FEBS Lett. 583, 1457–1462

18. Maheshwari, R., Bharadwaj, G., and Bhat, M. K. (2000) Thermophilic fungi. Their physiology and enzymes. Microbiol. Mol. Biol. Rev. 64, 461–488

19. Viikari, L., Alapuranen, M., Puranen, T., Vehmaanperä, J., and Siika-Aho, M. (2007) Thermoresistant enzymes in lignocellulose hydrolysis. Biofuels, 2012, 1241–1245

20. Cann, I. K., Stroot, P. G., Mackie, K. R., White, B. A., and Mackie, R. I. (2009) Biochemical analysis of a thermostable α-galactosidase from Thermoaerobacterium polysaccharolyticum sp. nov. and Thermoaerobacterium zea sp. nov., and emedation of the genus Thermoaerobacterium. Int. J. Syst. Evol. Microbiol. 51, 293–302

21. Cann, I. K., Mackie, R. I., and Wielig, J. (2008) Description of Caldanaerobius fijianus gen. nov., sp. nov., an inulin-degrading, ethanol-producing, thermophilic bacterium from a Fijian hot spring sediment, and reclassification of Thermoaerobacterium polysaccharolyticum and Thermoaerobacterium zea as Caldanaerobius polysaccharolyticus comb. nov. and Caldanaerobius zea comb. nov. Int. J. Syst. Evol. Microbiol. 58, 666–670

22. Cann, I. K., Kocherginskaya, S., King, M. R., White, B. A., and Mackie, R. I. (1999) Molecular cloning, sequencing, and expression of a novel multidomain mannanase gene from Thermoaerobacterium polysaccharolyticum. J. Bacteriol. 181, 1643–1651

23. King, M. R., White, B. A., Blaschek, H. P., Chassy, B. M., Mackie, R. I., and Cann, I. K. (2002) Purification and characterization of a thermostable α-galactosidase from Thermoaerobacterium polysaccharolyticum. J. Agric. Food Chem. 50, 5676–5682

24. Aziz, R. K., Bartels, D., Best, A. A., DeJongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kibab, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., McNeil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., and Zagnitko, O. (2008) The RAST Server. Rapid annotations using subsystems technology. BMC Genomics 9, 75

25. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignAP 4.0. Discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786

26. Leemli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685

27. Milner, Y., and Avigad, G. (1967) A copper reagent for the determination of hexuronic acids and certain ketohexoses. Carbohydr. Res. 4, 359–361

28. Kurokawa, J., Hemjinda, E., Arai, T., Kimura, T., Sakka, K., and Ohmiya, K. (2005) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. Acta Crystallogr. D Biol. Crystallogr. 61, 582–601

29. Perrasik, A., Sixma, T. K., Wilson, K. S., and Lamzin, V. S. (1997) wARP. Improvement and extension of crystallographic phases by weighted averaging of multiple-refined dummy atomic models. Acta Crystallogr. D Biol. Crystallogr. 53, 448–455

30. Cowtan, K. (2006) The Buccaneer software for automated model building. J. Appl. Crystallogr. 39, 1022–1011

31. McRee, D. E. (1999) XtalView/Xfit. A versatile program for manipulating atomic coordinates and electron density. J. Struct. Biol. 125, 156–165

32. Murshudov, G. N., Vagin, A. A., and Dodson, E. J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240–255

33. Reid, R. J., Adams, P. D., Arendall, W. B., 3rd, Brunger, A. T., Emsley, P., Joosten, R. P., Kleywegt, G. J., Krissinel, E. B., Lütteke, T., Otwinowski, Z., Perrasik, A., Richardson, J. S., Sheffler, W. H., Smith, J. L., Tickle, I. L., Vriend, G., and Zwart, P. H. (2011) A new generation of crystallographic validation tools for the Protein Data Bank. Structure 19, 1395–1412

34. Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) AQUA and PROCHECK-NMR. Programs for checking the quality of protein structures solved by NMR. J. Biomol. NMR 8, 477–486

35. Dodd, D., Kiyonari, S., Mackie, R. I., and Cann, I. K. (2010) Functional diversity of four glycoside hydrolase family 3 enzymes from the rumen bacterium Prevotella bryantii B14. J. Bacteriol. 192, 2335–2345

36. Shallom, D., Golan, G., Shogam, G., and Shoham, Y. (2004) Effect of dimer dissociation on activity and thermostability of the α-glucuronidase from Geobacillus stearothermophilus. Dissecting the different oligomeric forms of family 67 glycoside hydrolases. J. Bacteriol. 186, 6928–6937

37. Licht, A., Bulut, H., Scheffel, F., Daumke, O., Wehmeier, U. F., Saenger, W., Schneider, E., and Vahedi-Faridi, A. (2011) Crystal structures of the bacterial solute receptor AcbH displaying an exclusive substrate preference for β-D-galactopyranose. J. Mol. Biol. 406, 92–105

38. Cuneo, M. J., Changela, A., Warren, J. J., Beese, L. S., and Hellinga, H. W. (2006) The crystal structure of a thermophilic glucose-binding protein reveals adaptations that interconvert mono- and disaccharide-binding sites. J. Mol. Biol. 362, 259–270

39. Diez, J., Diederichs, K., Greller, G., Holracher, R., Boos, W., and Welte, W. (2001) The crystal structure of a liganded trehalose/maltose-binding protein from the hyperthermophilic archaeon Thermococcus litoralis at 1.85 Å. J. Mol. Biol. 305, 905–915

40. Abbott, D. W., and Boraston, A. B. (2007) Specific recognition of saturated and 4,5-ununsaturated hexuronate sugars by a periplasmic binding protein involved in pectin catabolism. J. Mol. Biol. 369, 759–770

41. Vahedi-Faridi, A., Licht, A., Bulut, H., Scheffel, F., Keller, S., Wehmeier, U. F., Saenger, W., and Schneider, E. (2010) Crystal structures of the solute receptor GacH of Streptomyces glucosaceus in complex with acarbose and an acarbose homolog. Comparison with the acarbose-loaded maltose-binding protein of Salmonella typhimurium. J. Mol. Biol. 397, 709–723

42. Higgins, M. A., Abbott, D. W., Boulanger, M. J., and Boraston, A. B. (2009) Blood group antigen recognition by a solute-binding protein from a serotype 3 strain of Streptococcus pneumoniae. J. Mol. Biol. 388, 299–309

43. Berns, R. P., Smits, S. H., Schmitt, L., Slotboom, D. J., and Poolman, B. (2010) A structural classification of substrate-binding proteins. FEBS Lett. 584, 2606–2617

44. Szabo, I., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2001) Structure of a family 15 carbohydrate-binding module in complex with xyllopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. J. Biol. Chem. 276, 49061–49065

45. Blanco, A., Diaz, P., Zueco, J., Parascandola, P., and Javier Pastor, F. I. (1999) A multidomain xylanase from a Bacillus sp. with a region homologous to thermostabilizing domains of thermophilic enzymes. Microbiology 145, 2163–2170

46. Lee, Y. E., Lowe, S. E., Henriass, B., and Zeikus, J. G. (1993) Characterization of the active site and thermostability regions of endoxylanase from
Xylan Utilization by Caldanaerobius polysaccharolyticus

Thermoanaerobacterium saccharolyticum B6A-RI. J. Bacteriol. 175, 5890–5898

52. Meissner, K., Wassenberg, D., and Liebl, W. (2000) The thermostabilizing domain of the modular xylanase XynA of Thermotoga maritima represents a novel type of binding domain with affinity for soluble xylan and mixed linkage β-1,3/β-1, 4-glucan. Mol. Microbiol. 36, 898–912

53. Selvaraj, T., Kim, S. K., Kim, Y. H., Jeong, Y. S., Kim, Y. J., Phuong, N. D., Jung, K. H., Kim, J., Yun, H. D., and Kim, H. (2010) The role of carbohydrate-binding module (CBM) repeat of a multimodular xylanase (XynX) from Clostridium thermocellum in cellulose and xylan binding. J. Microbiol. 48, 856–861

54. Winterhalter, C., Heinrich, P., Candussio, A., Wich, G., and Liebl, W. (1995) Identification of a novel cellulose-binding domain within the multidomain 120-kDa xylanase XynA of the hyperthermophilic bacterium Thermotoga maritima. Mol. Microbiol. 15, 431–444

55. Liebl, W., Winterhalter, C., Baumeister, W., Armbrrecht, M., and Valdez, M. (2008) Xylanase attachment to the cell wall of the hyperthermophilic bacterium Thermotoga maritima. J. Bacteriol. 190, 1350–1358

56. Mesnage, S., Fontaine, T., Mignot, T., Delepiere, M., Mock, M., and Fouet, A. (2000) Bacterial SLH domain proteins are noncovalently anchored to the cell surface via a conserved mechanism involving wall polysaccharide pyruvylation. EMBO J. 19, 4473–4484

57. Varghese, J. N., Hrmova, M., and Fincher, G. B. (1999) Three-dimensional structure of a barley β-1,3-glucan exohydrolase, a family 3 glycosyl hydrolase. Structure 7, 179–190

58. Bork, P., and Doolittle, R. F. (1992) Proposed acquisition of an animal protein domain by bacteria. Proc. Natl. Acad. Sci. U.S.A. 89, 8990–8994

59. Little, E., Bork, P., and Doolittle, R. F. (1994) Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. J. Mol. Evol. 39, 631–643

60. Pozzo, T., Pasten, J. L., Karlsson, E. N., and Logan, D. T. (2010) Structural and functional analyses of β-glucosidase 3B from Thermotoga neapolitana. A thermostable three-domain representative of glycoside hydrolase 3. J. Mol. Biol. 397, 724–739

61. 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