Characterization of a furan aldehyde-tolerant β-xylosidase/α-arabinosidase obtained through a synthetic metagenomics approach

Maruthamuthu, M.; Jimenez, D. J.; van Elsas, J. D.

Published in:
Journal of Applied Microbiology

DOI:
10.1111/jam.13484

Publication date:
2017

Journal of Applied Microbiology

Journal of Applied Microbiology ISSN 1364-5072

ORIGINAL ARTICLE

Characterization of a furan aldehyde-tolerant β-xylosidase/α-arabinosidase obtained through a synthetic metagenomics approach

M. Maruthamuthu*, D.J. Jiménez* and J.D. van Elsas

Cluster of Microbial Ecology, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

Keywords
furfural, hemicellulose, microbial consortium, synthetic metagenomics, xylose, α-arabinosidase, β-xylosidase.

Correspondence
Mukil Maruthamuthu and Diego Javier Jiménez, Cluster of Microbial Ecology, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Nijenborgh 7, 9747AG Groningen, The Netherlands.
E-mails: m.marutha.muthu@rug.nl; d.j.jimenez.avella@rug.nl

*Equal first authorship.

2017/0365: received 22 February 2017, revised 26 April 2017 and accepted 3 May 2017

doi:10.1111/jam.13484

Abstract

Aims: The aim of the study was to characterize 10 hemicellulolytic enzymes obtained from a wheat straw-degrading microbial consortium.
Methods and Results: Based on previous metagenomics analyses, 10 glycosyl hydrolases were selected, codon-optimized, synthetized, cloned and expressed in *Escherichia coli*. Nine of the overexpressed recombinant proteins accumulated in cellular inclusion bodies, whereas one, a 37·5-kDa protein encoded by gene *xyl*M1989, was found in the soluble fractions. The resulting protein, denoted XylM1989, showed β-xylosidase and α-arabinosidase activities. It fell in the GH43 family and resembled a *Sphingobacterium* sp. protein. The XylM1989 showed optimum activity at 20°C and pH 8·0. Interestingly, it kept approximately 80% of its β-xylosidase activity in the presence of 0·5% (w/v) furfural and 0·1% (w/v) 5-hydroxymethylfurfural. Additionally, the presence of Ca^{2+}, Mg^{2+} and Mn^{2+} ions increased the enzymatic activity and conferred complete tolerance to 500 mmol l^{-1} of xylose. Protein XylM1989 is also able to release sugars from complex polysaccharides.
Conclusion: We report the characterization of a novel bifunctional hemicellulolytic enzyme obtained through a targeted synthetic metagenomics approach.
Significance and Impact of the Study: The properties of XylM1989 turn this protein into a promising enzyme that could be useful for the efficient saccharification of plant biomass.

Introduction

Plant-derived lignocellulose represents an abundantly available and renewable energy source. Lignocellulose comprises cellulose, hemicellulose and lignin moieties. Hemicellulose consists of hetero-polymers that are composed of pentoses and hexoses. In this fraction, xylan is the major component, constituting nearly one third of all renewable carbon in nature. Xylan (or arabinoxylan) is composed of β-1,4-linked D-xylose units, which may be substituted by different side groups, such as D-galactose, L-arabinose, glucuronic acid, acetyl, feruloyl and *p*-coumaroyl residues (Saha 2003; De Souza 2013). In the enzyme-mediated catalysis that is required for hemicellulose degradation, microbial glycoside hydrolases (GHs) are key enzymes. These cleave the glycosidic linkages between carbohydrate residues, allowing to produce sugars that are released. Xylan can be degraded through the action of a set of different GHs. For instance, β-1,4-endoxylanase (EC 3.2.1.8), which cleaves the backbone into small oligosaccharides, and β-1,4-xylosidase (EC 3.2.1.37), which cleaves these oligosaccharides into xylose. Next to breaking the side chains of xylan, enzymes like α-L-arabinosidases, α-D-glucuronidases and acetyl esterases can play vital roles (Barker *et al.* 2010; Bokhari *et al.* 2010; Van den Brink and de Vries 2011). Such enzymes are thought to be valuable for diverse industrial (e.g. food, pharma, plastics and biofuels) applications (Gao *et al.* 2011).

In spite of the promise of using micro-organisms from natural settings as sources of novel GHs, the discovery of

novel enzymes or activities has been hampered by problems of unculturability. Thus, recent research has set out to analyse lignocellulose-enriched microbial consortia by DNA-based approaches (also known as targeted metagenomics) (D'haeseleer *et al.* 2013). For example, recently Jiménez *et al.* (2016) reported the analysis of three soil-derived microbial consortia cultivated on biologically pretreated plant biomass. They analysed the microbial structure, GH profile and extracellular enzymatic activities. Moreover, in a process known as 'synthetic metagenomics', GH-encoding genes can be custom-synthesized and codon-optimized, after which their efficient expression can be achieved in a suitable host. In this respect, Dougherty *et al.* (2012) identified, synthetized and expressed a total of 19 GHs originating from the metagenome of a switchgrass-adapted compost community. In the same way, Gladden *et al.* (2014) discovered 18 active GHs that were tolerant to 10% of 1-ethyl-3-methylimidazolium acetate (ionic liquid used in the pretreatment of plant biomass).

In previous work, we developed two wheat straw-degrading microbial consortia derived from forest soil, in which substantially enriched (hemi)cellulolytic genes and activities were found (Jiménez *et al.* 2014). In order to explore these consortia further, here we performed a targeted synthetic metagenomics approach. Thirteen large contigs—produced previously on the basis of shotgun sequencing of metagenomic DNA extracted from the aforementioned consortia—were selected and screened for GH-encoding genes (Jiménez *et al.* 2015). In the current study, we report the selection of 10 such genes on the basis of a combination of criteria. The genes were codon-optimized, synthesized and expressed, after which they were further tested. Here, we describe the full analysis, placing a focus on a gene for a key furan aldehyde-tolerant β-xylosidase/α-arabinosidase (CAZy family GH43) enzyme that is proposed for biorefining processes, especially the saccharification of pretreated plant biomass.

Materials and methods

Identification and selection of GHs from a wheat straw-degrading consortial metagenome

Previous analyses of contigs constructed following shotgun metagenomics and sequencing of DNA from two wheat straw-degrading microbial consortia identified 13 novel Bacteroidetes-derived hemicellulose utilization loci containing 39 GHs (Jiménez *et al.* 2015). From the contigs, we selected 10 predicted GH-encoding genes on the basis of the following criteria: (i) genes encode highly enriched GH families compared with the original soil inoculum; (ii) GHs are predicted to allow deconstruction of xylan, xyloglucan and galacto(gluco)mannan; (iii) GHs are flanked by genes for membrane transporters and genes involved in sugar metabolism (i.e. coherent genomic context); (iv) predicted GHs have low amino acid identity (e.g. <80%) compared to proteins in databases; (v) GHs contain identifiable start and stop codons and a complete intact reading frame. In Fig. S1 (supporting information), the selected genes and their genomic contexts are shown.

Cloning and expression of 10 GH-encoding genes recovered from the metagenome assemblages

The 10 genes were selected based on directed choices with respect to CAZy (Lombard *et al.* 2014) family allocation. Specifically, we selected genes for families GH92, GH43, GH2, GH95 and GH29, as follows: M4684 and M3030 (GH92); M7068, M7073, M1989 and M8244 (GH43); M1927 and M20752 (GH2); M1916 (GH95) and M8239 (GH29) (Table 1). All genes were codon-optimized for expression in *Escherichia coli*, synthesized and cloned into the pET21b+ vector, with the help of a commercial partner (GenScript, Piscatawy, NJ). For codon optimization, we used the OptimumGene™ algorithm and the Codon Adaptation Index. Additionally, other major codon usage biases, such as premature Poly-A sites, GC contents, internal chi-sites and ribosome-binding sites, repeat sequences (direct, reverse and dyad repeats), as well as restriction sites that might interfere with cloning, were changed. The expression clones were introduced into *E. coli* BL21(DE3) competent cells (Invitrogen, Carlsbad, CA) using the manufacturer's instructions. Following clone selection and purification, plasmid extractions were done for each of the 10 cloned genes. Thereafter, the nature of the cloned fragments was checked by restriction fragment analyses. Specifically, *Xba*1/*Xho*1 were used for genes M3030, M1916, M1927, M1989 and M20752; *Mlu*1/ *Xho*1 for genes M8239, M8244 and M7068; *Sal*1/ *Xho*1 for M4684; and *Sac*1/ *Xho*1 for M7073 (Fig. S2). A single colony of each verified clone was then introduced into an Erlenmeyer containing 10 ml of fresh LB medium containing Overnight Express™ Autoinduction System 1 reagents (Novagen, Darmstadt, Germany) and ampicillin (100 μg ml^{-1}). The bacterial cultures were incubated overnight at 37°C with constant shaking at 200 rev min^{-1}. The cell pellets were harvested by centrifugation at 10 000 ***g*** for 10 min and resuspended in 2 ml of lysis buffer (20 mmol l^{-1} Tris-HCl pH 7·5, 100 mmol l^{-1} NaCl, 1 mmol l^{-1} EDTA, 0·1% Triton, 5 mmol l^{-1} CHAPS and a mini tablet of protease inhibitor—Roche, Mannheim, Germany—to 50 ml). Subsequently, the resuspended cells were sonicated on ice (6 s on, 15 s off, 30 cycles with amplitude of 10–15 microns)

Table 1 Features of the selected GH-encoding genes

Gene ID	CAZy family	Gene length (bp)	Amino acids	PSI-BLASTP best hit [Taxon] (accession number)	QC* (%)	Identity (%)	pI†	kDa
M4684	GH92	3111	1032	Hypothetical protein [*Parabacteroides* sp.] (WP_010801039.1)	100	70	6·04	117
M3030	GH92	1983	657	Alpha-1,2-mannosidase [*Sphingobacterium spiritivorum*] (WP_002997981.1)	98	83	9·24	74·2
M7068	GH43	996	328	Glycosyl hydrolase family 32 [*Paraprevotella clara*] (WP_008623026.1)	88	69	6·51	37·9
M7073	GH43	1935	641	Glycosyl hydrolase [*Flavobacterium johnsoniae*] (YP_001195445.1)	98	66	7·11	73·3
M1916	GH29	1641	543	Alpha-1,3/4-fucosidase [*Capnocytophaga canimorsus*] (YP_004740108.1)	98	63	6·53	61·5
M1927	GH2	3216	1068	Beta-galactosidase [*Sphingobacterium spiritivorum*] (WP_002995396.1)	85	50	7·08	120
M1989	GH43	981	323	Hypothetical protein [*Sphingobacterium* sp.] (WP_021189556.1)	100	98	4·70	37·5
M20752	GH2	1293	427	Beta-galactosidase/beta-glucuronidase [*Flavobacterium* sp.] (WP_007809792.1)	90	56	8·63	48·5
M8239	GH95	1302	430	Alpha-L-fucosidase [*Pedobacter saltans*] (YP_004274942.1)	98	72	6·74	48
M8244	GH43	1605	531	Hypothetical protein [*Sphingobacterium* sp.] (WP_021189555.1)	55	99	5·17	61·1

*****Query coverage.
†Isoelectric point.

and the lysates centrifuged at 14 000 ***g*** for 10 min at 4°C in order to separate the soluble and insoluble protein fractions. Insoluble proteins were washed twice with 750 μl of 20 mmol l^{-1} Tris-HCl (pH 8·0) and solubilized with 2 mol l^{-1} urea following the freeze–thawing method (Qi *et al.* 2015). Protein concentrations were determined by the Bradford method using bovine serum albumin as the standard. Protein fractions were analysed by 10% sodium dodecyl sulphate–polyacrylamide gel electrophoresis (Sambrook and Russell 2001).

Zymographic analysis and detection of enzymatic activity using para-nitrophenol-glycosides

Zymograms were used to detect xylanase activity on SDS polyacrylamide gels (4% stacking, 10% resolving gels) containing 0·2% of xylan from beechwood (Sigma-Aldrich, Zwijndrecht, the Netherlands). Each well was loaded with 20 μg (in 20 μl) of total proteins per sample. After running the gels at 4°C, they were soaked for 1 h in 2·5% of Triton and washed thoroughly in water prior to incubation (1 h at 30°C) in 50 mmol l^{-1} of sodium citrate buffer pH 6·0. The gel was stained with 0·1% Congo red for 30 min and then de-stained for 2 h in 1 mol l^{-1} NaCl to reveal zones of clearing. Additionally, protein fractions (soluble and insoluble) were recovered and tested for activity using *p*-nitrophenyl β-D-xylopyranoside (*p*NP-Xyl), *p*-nitrophenyl α-L-arabinofuranoside (*p*NP-Ara), *p*-nitrophenyl α-L-fucopyranoside (*p*NP-Fuc) and *p*-nitrophenyl α-D-mannopyranoside (*p*NP-Man). The reaction mixtures consisted of 180 μl of 2 mmol l^{-1} of each *p*-nitrophenol-glycoside (diluted in 20 mmol l^{-1} of Tris-HCl pH 7·0) and 20 μl of each protein fraction. The mixtures were incubated at 37°C for 30 min, after which the reactions were stopped on ice. Three negative controls were used for all assays: (i) reaction mixture without substrate; (ii) reaction mixture using the protein fractions from *E. coli* BL21(DE3) transformed with pET21b+ vector; and (iii) reaction mixture without proteins. Activity was detected by the presence of yellow colour in the reaction plate.

Bioinformatics analysis, phylogenetic tree and structural modelling of protein XylM1989

For one protein that was successfully produced into the soluble fraction (gene M1989), the translated gene (*xyl*M1989) was analysed based on BLASTP searches against the NCBI nonredundant protein database. In addition, catalytic domains were identified using the pFam database. The protein XylM1989 was aligned by ClustalW against 35 proteins from different origin that belong to CAZy families GH43, GH3 (Lagaert *et al.* 2014) and AA10 (outgroup sequences). For the multiple protein alignment and phylogenetic analyses, the software's MEGA ver. 6.0 and PRALINE (Simossis and Heringa 2005; Tamura *et al.* 2013) were used. In order to detect the catalytic and substrate-binding sites, a XylM1989 protein data

bank file was generated using Phyre2 (Kelley and Sternberg 2009). With this prediction, the closest homologue of XylM1989 was a β-xylosidase protein, 4MLG (Protein data bank ID), from an uncultivable bacterium (EC 3.2.1.37, GH43 family). This protein was used as a template for structural predictions using the PyMOL platform (http://www.pymol.org).

Biochemical properties of protein XylM1989

The optimum temperature for activity was determined in the range 10–70°C using *p*NP-Xyl and *p*NP-Ara (at pH 7·0). The pH optimum was determined in a pH range from 3·0 to 10·0 (at 30°C) using the following buffers: 50 mmol l^{-1} sodium citrate (pH 3·0–6·0), 50 mmol l^{-1} Tris-HCl (pH 7·0–9·0) and 50 mmol l^{-1} glycine-NaOH (pH 10·0). The reaction mixture consisted of 280 μl of 0·5 mmol l^{-1} of *p*NP-Xyl or *p*NP-Ara and 20 μl of soluble protein XylM1989 (approximately 1·7 mg ml^{-1}). The kinetic parameters (K_m and V_{max}) of XylM1989 were determined with *p*NP-Xyl and *p*NP-Ara concentrations ranging from 0 to 50 mmol l^{-1} in 20 mmol l^{-1} Tris-HCl (pH 8·0) at 30°C for 15 min. The data were plotted according to the Lineweaver–Burk method (double reciprocal plot). The effects of lignocellulosic hydrolysate inhibitors (furfural, 5-hydroxymethylfurfural and acetic acid) and chemical additives (ions, sugars, NaCl, EDTA, detergents and organic solvents), at different concentrations, on the activity of the XylM1989 protein were evaluated with *p*NP-Xyl (pH 8·0) at 30°C for 30 min. Additionally, the effect of xylose (ranging from 0 to 1000 mmol l^{-1}) in the presence of 5 mmol l^{-1} of Mg^{2+}, Ca^{2+} and Mn^{2+} was evaluated with the above parameters. Enzymatic activities were determined from the measured absorbance units using a standard calibration curve. The amount of paranitrophenol (*p*NP) liberated was measured by absorbance at 410 nm. One unit (U) of enzyme activity was defined as the activity required for the formation of 1 μmol of *p*NP per min under the above conditions.

Activity of the XylM1989 protein on complex polysaccharides

The enzymatic activity of XylM1989 was evaluated on three complex polysaccharides (xylan from beechwood, oat spelt xylan and soluble arabinoxylan). The reaction mixtures (500 μl) contained 1% of each polysaccharide (diluted in 20 mmol l^{-1} of Tris-HCl pH 8·0) and 150 μl of soluble protein XylM1989 (approximately 1·7 mg ml^{-1}). The mixtures were incubated at 30°C for 72 h, subsequently the reactions were stopped on ice and centrifuged for 10 min at 12 000 ***g***. The enzymatic activity was determined by measuring the amount of reducing sugars in the supernatant by the 3,5-dinitrosalicilic acid method (Miller 1959). A standard calibration curve was used, as previously constructed with different concentrations of xylose. In addition, the types of sugars and their concentrations, released by the enzymatic reaction, were analysed by high-performance liquid chromatography (HPLC). Two negative controls were set up: (i) reaction without substrate and (ii) reaction without protein.

Results

Synthesis, cloning, expression and enzymatic analysis of 10 genes predicted to encode GHs

From two wheat straw-degrading microbial consortia, we here selected 10 genes encoding predicted GHs for synthesis. All 10 genes were codon-optimized for *E. coli*, synthetized and cloned into the pET21b+ expression vector with specific restriction enzyme sites on both sides (*Nde*1 and *Xho*1). For all genes, the fragment sizes—as measured on agarose gels—were consistent with the predicted sizes of the sequences (Table 1; Fig. S2). Then, using overnight cultures (in LB medium) of each selected purified clone, gene expression was induced. The data revealed that, among the cultures from the 10 cloned genes, only one protein (gene M1989) occurred in the soluble fraction (~80–90% pure), whereas the remainder was mainly present in inclusion bodies (Fig. 1a). To enhance solubility, we applied solubilization methods to the latter, so as to recover and refold each aggregated protein into its native state. The nine inclusion bodies were thus isolated, purified and then treated with 2 mol l^{-1} urea following the freeze–thawing method. Fractions were diluted (1-, 10- and 100-fold) into PBS buffer in order to decrease the urea concentration and improve the refolding of the protein (Qi *et al.* 2015). Unfortunately, the final suspensions did not show any enzymatic activities, suggesting persistent misfolding or aggregation of the nine proteins (data not shown). However, zymogram analysis with beechwood xylan revealed that the 'insoluble' protein fractions of genes M7068, M1916, M20752, M8239 and M8244 represented a clear zone with enzymatic activity. Moreover, the (soluble) M1989 cell lysate also had activity. Thus, the products of 6 in 10 genes produced in *E. coli* had xylanase activity, of which only one, M1989, appeared in the soluble fraction (Fig. 1b).

Testing for different enzymatic activities

Based on the prediction of the activities of the gene products by CAZy database annotation (Table 1), we selected four *p*NP-labelled substrates, that is, *p*NP-Xyl, *p*NP-Ara, *p*NP-Fuc and *p*NP-Man, to evaluate the putative

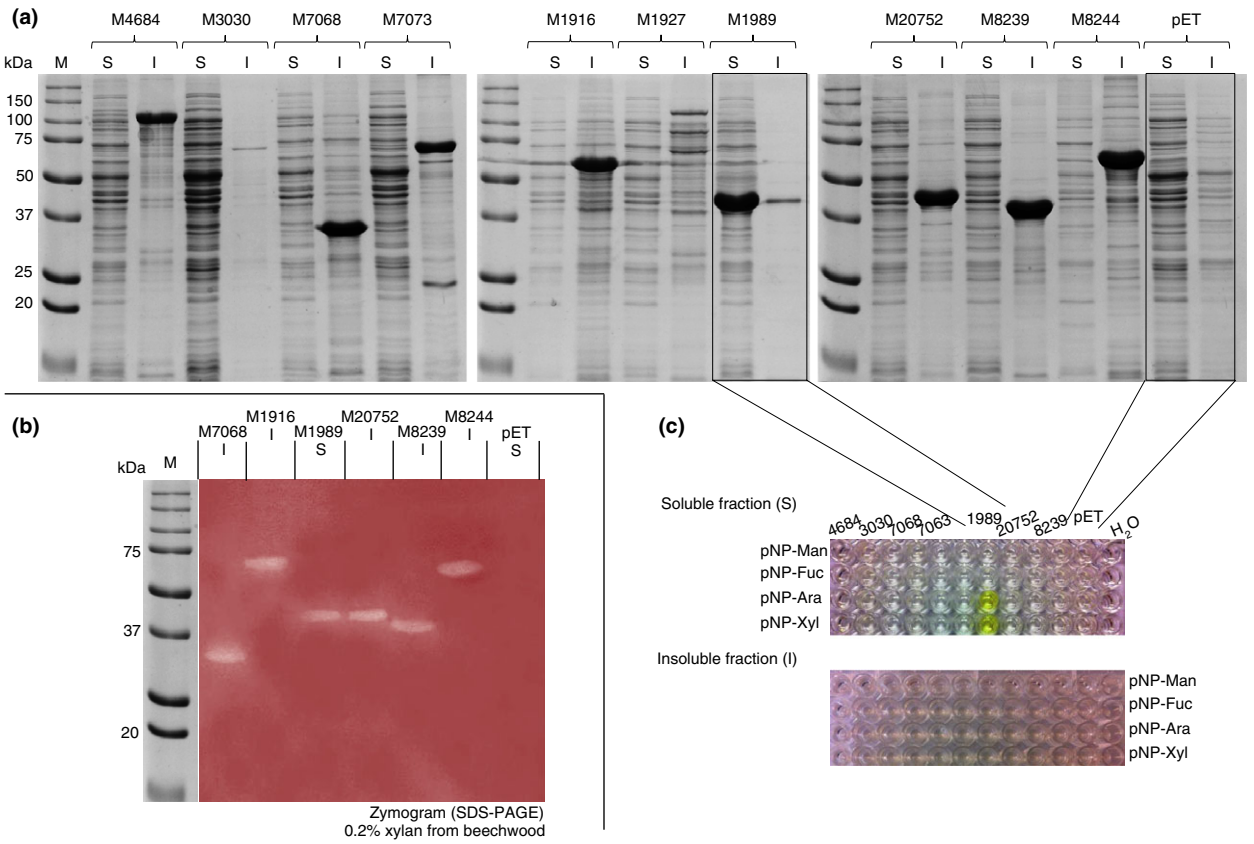

Figure 1 (a) SDS-PAGE of the overexpressed GHs in *Escherichia coli* Bl21(DE3) cells. (b) Zymogram analysis of six GHs that showed hydrolytic activity and one negative control (soluble protein fraction from *E. coli* BL21(DE3) transformed with pET21b+ vector; lane pET S). (c) Detection of enzymatic activities of the overexpressed GHs by using *p*NP-labelled substrates. Supernant fractions for the cell extracts are denoted as S, and for the insoluble pellets of the lysates are denoted as I. M, protein marker. [Colour figure can be viewed at wileyonlinelibrary.com]

enzymatic activities (using both insoluble and soluble fractions). However, with the exception of the lysate of clone M1989 (soluble fraction), none of the lysates showed hydrolytic activity with any of the selected substrates (Fig. 1c). Indeed, the product of clone M1989 showed dual activity, that is, with *p*NP-Xyl and *p*NP-Ara, but not with *p*NP-Man and *p*NP-Fuc. On the basis of its activity, protein M1989 will be denoted XylM1989. It is the basis of the further results described below.

Analysis of the XylM1989 protein—CAZy family, phylogeny and structural prediction

The protein XylM1989 has a calculated isoelectric point (pI) of 6·16 and a molecular weight of 37·5 kDa. In addition, XylM1989 was predicted to belong to the GH43 family, which contains mostly β-xylosidases (EC 3.2.1.37), α-arabinosidases (EC 3.2.1.55), galactan 1,3-β-galactosidases (EC 3.2.145) and endo-α-arabinases (EC 3.2.1.99) (Jordan *et al.* 2013; Lagaert *et al.* 2014). Based on the BLASTP analysis, the amino acid sequence of XylM1989 showed 95% identity (100% coverage) with an uncharacterized GH43 family protein (ACX30651) encoded by a chromosomal segment of *Sphingobacterium* sp. TN19 (Zhou *et al.* 2010). In addition, protein XylM1989 showed 63% identity with a characterized bifunctional GH43 family xylosidase/arabinosidase (*xyn*B; CAA89208) from *Prevotella bryantii* (Gasparic *et al.* 1995). The more detailed phylogenetic analyses (including different types of family GH3 and GH43 enzymes) further showed that protein XylM1989 clustered with uncharacterized family GH43 xylosidases and arabinosidases from different Bacteroidetes, specifically belonging to species of *Sphingobacterium*, *Draconibacterium*, *Proteiniphilum*, *Dysgonomonas* and *Chryseobacterium*. XylM1989 revealed a relatively low degree of similarity with characterized family GH43 bacterial xylosidases (EC 3.2.1.37) next to fungal endo-arabinases (Fig. 2). Moreover, it showed 71% identity with protein 4MLG (structure of RS223-β-xylosidase) (Jordan *et al.* 2015). Based on the predicted 3D structure using protein 4MLG as the template and multiple alignments with phylogenetically closer

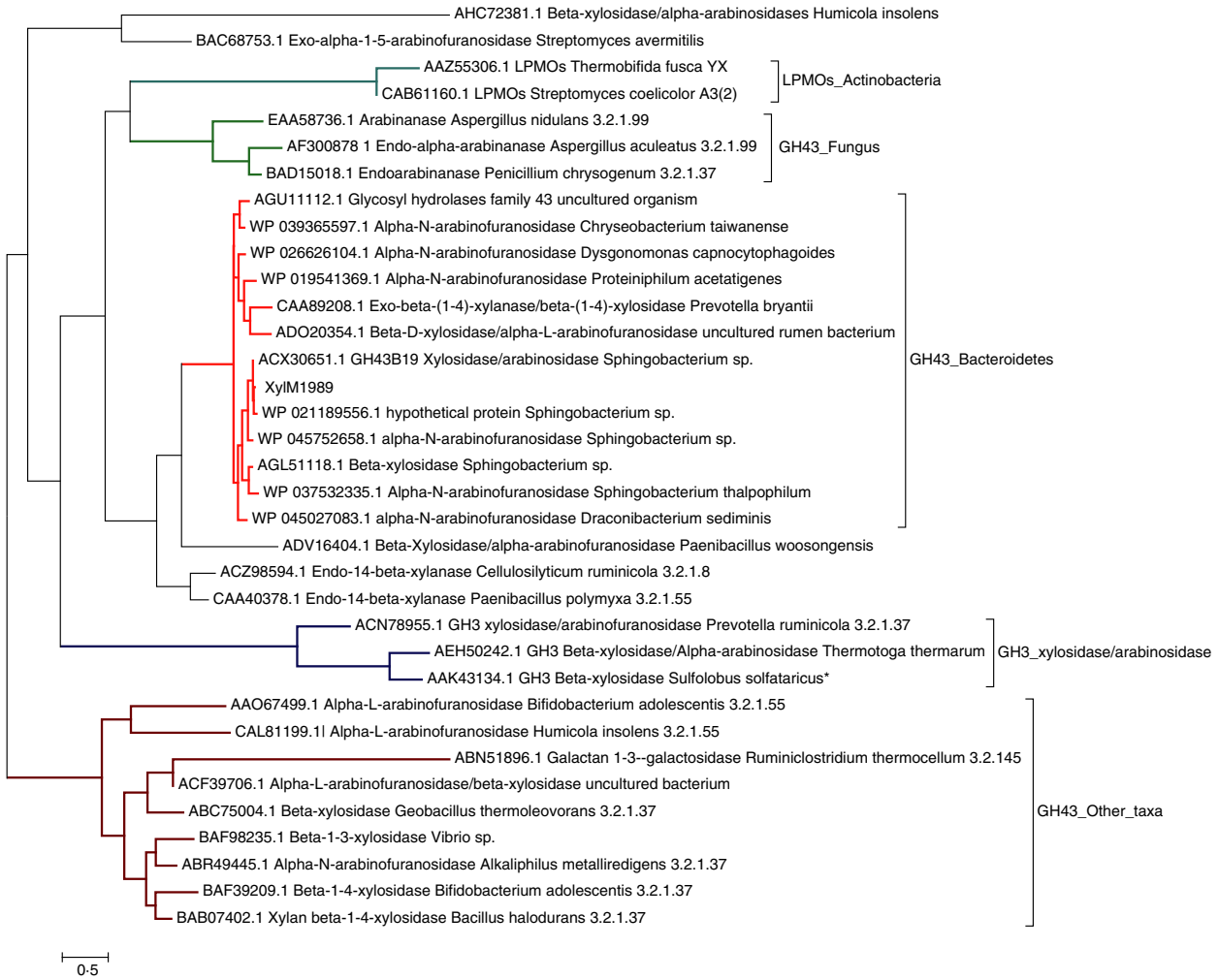

Figure 2 Maximum likelihood phylogenetic analysis of XylM1989 with closest related proteins. The amino acid sequences of the other GH43 and GH3 enzymes were obtained from published data (Lagaert *et al.* 2014). Two Actinobacteria-derived lytic polysaccharide monooxygenases (LPMOs) were used as outgroup (CAZy family AA10). The tree is drawn to scale, with branch lengths measured conform the number of substitutions (amino acids) per site. All positions containing gaps and missing data were eliminated and the evolutionary analyses were conducted in MEGA ver. 6.0. [Colour figure can be viewed at wileyonlinelibrary.com]

proteins, we identified a catalytic triad (Asp15—Asp135—Glu222) and a substrate-binding site (Trp83—Ile134—Thr271) (Fig. 3).

Biochemical characterization of protein XylM1989

Effects of temperature and pH on protein XylM1989 activity

Based on the finding that protein XylM1989 had *β*-xylosidase and *α*-arabinosidase activities, these activities were characterized with respect to temperature and pH ranges. Protein XylM1989 exhibited maximal activities at 20°C in the presence of 0·5 mmol l^{-1} of (buffer pH 7·0) *p*NP-Xyl (3·36 U mg^{-1} of protein) and *p*NP-Ara (0·65 U mg^{-1} of protein) (Fig. 4a). The activity of XylM1989 decreased, to approximately 20% of the maximal activity, when the temperature was raised from 40 to 70°C for both *p*NP-Xyl and *p*NP-Ara. As shown in Fig. 4b, this activity was then assessed at pH values between 3·0 and 10·0. In this analysis, the maximal *β*-xylosidase activity was reached at pH 8·0 (30°C, 0·5 mmol l^{-1} of *p*NP-Xyl). The activity, overall, remained at ~70% of the maximum at pH 9·0, whereas it was completely lost at pH 10·0. Moreover, activity on *p*NP-Ara was also maximal at pH 8·0. Indeed, the latter activity was still at 60% at pH 9·0, suggesting that the protein is considerably active under slightly to strongly alkaline conditions.

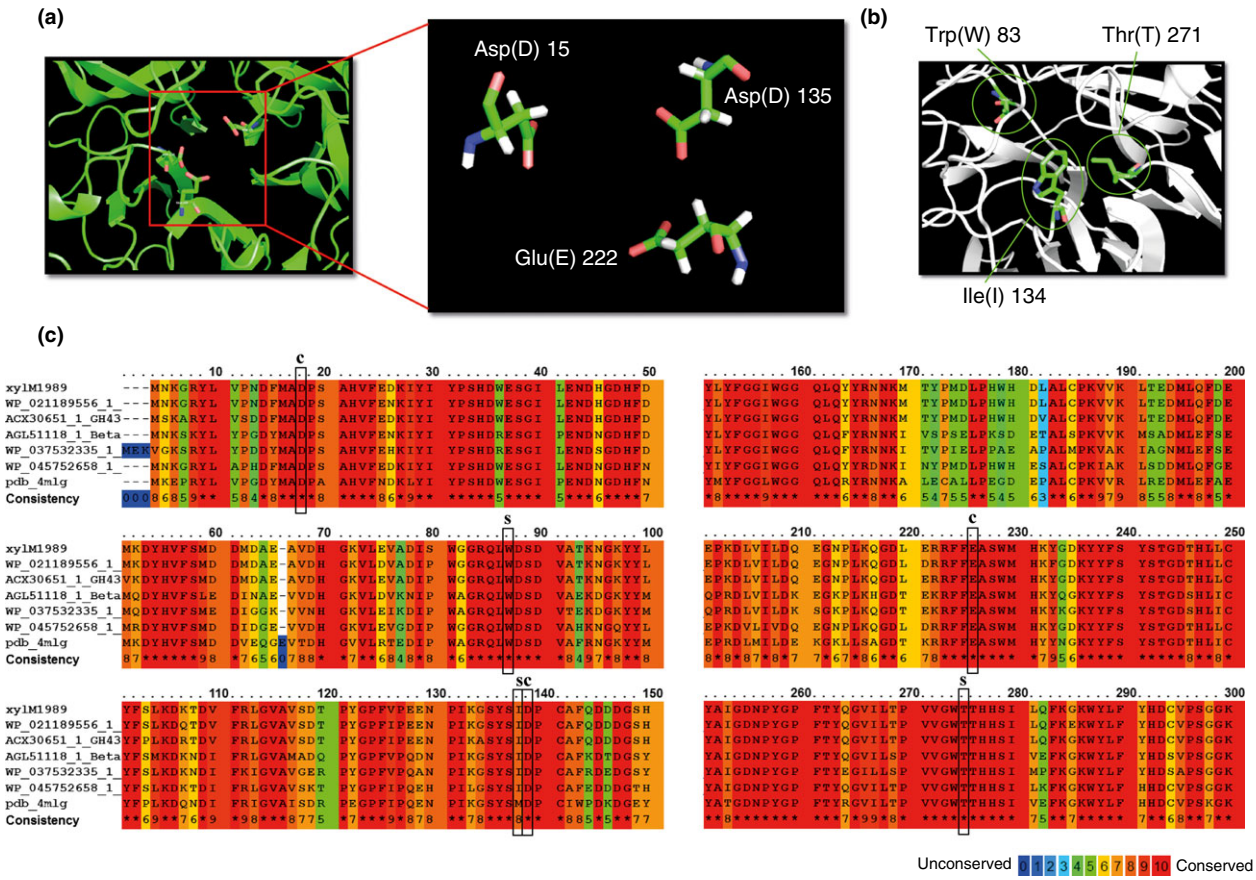

Figure 3 (a) *In silico* 3D structure prediction of the enzyme XylM1989, generated by molecular modelling, showing the catalytic and (b) substrate-binding sites. (c) Conserved blocks in the deduced amino acid sequences of XylM1989. Highly conserved residues are red shaded. Residues of the catalytic triad and substrate-binding pocket are denoted by C and S respectively. The aligned GH43 sequences came from *Sphingobacterium*-related organism (WP_021189556.1; ACX30651.1; WP_045752658.1; WP_037532335.1; AGL51118.1) and an uncultured organism (Protein data bank ID: 4MLG). [Colour figure can be viewed at wileyonlinelibrary.com]

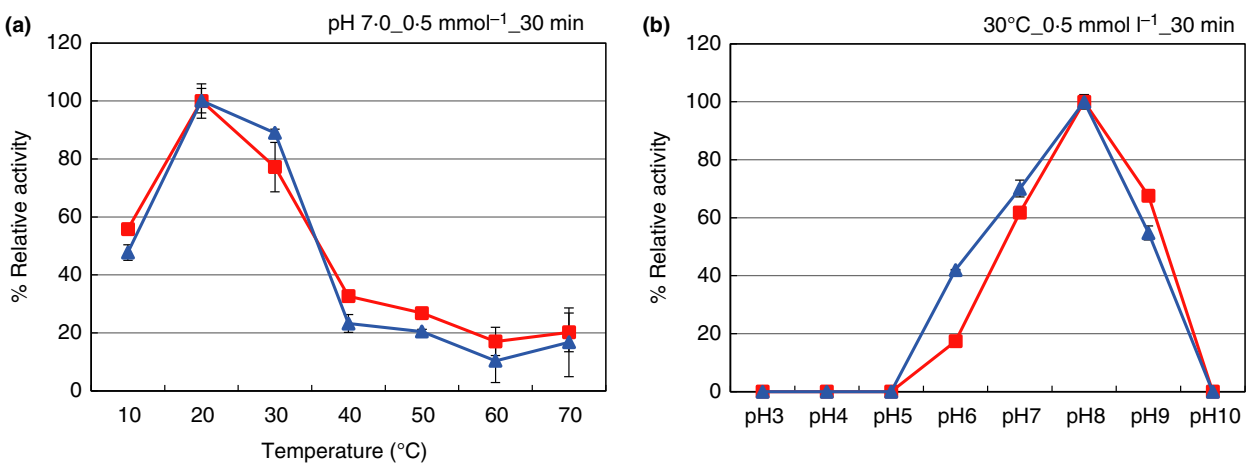

Figure 4 Relative enzymatic activity of the XylM1989 protein measured against *p*NP-*β*-D-xylopyranoside (*p*NP-Xyl) (■) and *p*NP-*α*-L-arabinopyranoside (*p*NP-Ara) (▲). (a) Different temperatures with constant pH 7·0. (b) Different pH values at 30°C. [Colour figure can be viewed at wileyonlinelibrary.com]

Kinetic analysis of the XylM1989 protein

The XylM1989 hydrolytic activity was measured with respect to catalytic properties and by assessment of kinetic parameters, that is, the K_m (Michaelis-Menten constant) and V_{max} values (maximal reaction velocities), using *p*NP-Xyl and *p*NP-Ara as the substrates, under optimal conditions. The K_m values of XylM1989 for *p*NP-Xyl and *p*NP-Ara were 1·2 and 0·781 mmol l^{-1}, and the V_{max} values 285·71 and 78·12 U mg^{-1} respectively. Given the fact that the K_m and V_{max} values of protein XylM1989 for *p*NP-Xyl were higher than those for *p*NP-Ara, *p*NP-Xyl was used for further analysis.

Effects of additives on protein XylM1989 activity

The effects of different additives on protein XylM1989 β-xylosidase activity were assessed (Table 2). Remarkable increases in the β-xylosidase activity were observed in the presence of 5 mmol l^{-1} of $CaCl_2$ (10-fold), $MgCl_2$ (12-fold) and $MnCl_2$ (sevenfold). In addition, a 2·5-fold enhanced activity was observed with 50 mmol l^{-1} L-arabinose, but this activity decreased by 47% in the presence of xylose. Furthermore, the activity decreased by 60% with 50 mmol l^{-1} of EDTA. It increased slightly (118·6 ± 24) upon addition of 20% glycerol. Concurrently, the activity dropped by 75% in the presence of all organic solvents, that is, ethanol, methanol and isopropanol. Moreover, the activity decreased by 50% in the presence of 10% DMSO.

Table 2 Effects of different additives on the XylM1989 β-xylosidase activity

Type of additive	Reagent	% Relative activity	
Salts		5 mmol l^{-1}	20 mmol l^{-1}
	$CaCl_2$	1065·23 ± 74	546·58 ± 39
	$MgCl_2$	1214·18 ± 16	691·11 ± 120
	$CoCl_2$	6·27 ± 1·1	1·33 ± 0·3
	$NiCl_2$	8·67 ± 1·5	2·01 ± 1·2
	$CuCl_2$	0·63 ± 0·1	1·94 ± 0·4
	$MnCl_2$	738·93 ± 8·4	172·88 ± 20
	NH_4Cl	69·32 ± 1·0	63·61 ± 4·3
Sugars		20 mmol l^{-1}	50 mmol l^{-1}
	Glucose	69·30 ± 1·3	99·28 ± 4·9
	Xylose	55·82 ± 2·5	53·71 ± 1·4
	Arabinose	164·89 ± 5·4	249·26 ± 3·8
	Cellobiose	82·32 ± 2·0	90·32 ± 14
	Galactose	80·38 ± 1·8	90·99 ± 12
Salt		200 mmol l^{-1}	2000 mmol l^{-1}
	NaCl	96·06 ± 6·1	36·34 ± 0·6
Chelating agent		20 mmol l^{-1}	50 mmol l^{-1}
	EDTA	61·66 ± 0·8	30·20 ± 0·6
Detergents		2%	5%
	SDS	8·68 ± 0·7	6·12 ± 4·0
	Triton	107 ± 5·6	32·70 ± 0·8
Organic solvents		10%	20%
	Glycerol	101·9 ± 4·7	118·6 ± 24
	Ethanol	26·63 ± 0·1	11·40 ± 0·2
	Methanol	26·47 ± 0·2	16·63 ± 1·9
	Isopropanol	26·10 ± 0·6	12·53 ± 1·9
	DMSO	50·56 ± 1·0	24·55 ± 0·1

Effects of plant biomass hydrolysate-derived compounds on protein XylM1989 activity

Three lignocellulosic hydrolysate-derived compounds with potential inhibitory activity, that is, furfural, 5-hydroxymethylfurfural (5-HMF) and acetic acid, were tested for their effects on the activity of protein XylM1989 with *p*NP-Xyl (0·5 mmol l^{-1}; pH 8·0 at 30°C) (Fig. 5a). Interestingly, in the presence of 0·3% (w/v) acetic acid, inhibition was high and the relative β-xylosidase activity was nearly zero. Moreover, it was also strongly blocked by 0·7% (w/v) 5-HMF (~80% inhibition), whereas low levels (0·05–0·1% w/v) of 5-HMF showed only 15–20% of inhibition. The presence of 0·7% (w/v) furfural also resulted in around 60% of inhibition. However, at lower concentrations (0·05–0·5% w/v of furfural) the activity inhibition remained at 20–40%.

Effect of xylose on protein XylM1989 activity

The β-xylosidase activity of XylM1989 was inhibited by addition of xylose (Table 2). However, the activity increased approximately 3·5-fold in the presence of 50 mmol l^{-1} xylose with ions (5 mmol l^{-1} of Ca^{2+}, Mg^{2+} and Mn^{2+}) over the control (without xylose and ions). In the presence of each of the three bivalent cations, without xylose, the xylM1989 β-xylosidase activity was fivefold increased over that of the control (Fig. 5b). At 200 mmol l^{-1} xylose (with ions), protein XylM1989 still showed an activity of 80% over that of the control without xylose and ions. Finally, the enzyme activity dropped to around 50% of that of the control at high concentrations of xylose (700 and 1000 mmol l^{-1}).

Activity of protein XylM1989 on seminatural substrates

The hydrolytic activity of protein XylM1989 on 1% of xylan from beechwood (XB), oat spelt xylan (OX), and arabinoxylan (ARB) was tested, using 5 mmol l^{-1} of Mg^{2+}. Interestingly, 68·73 ± 3·52 mg of sugars g^{-1} of polysaccharide were released from XB and 62·28 ± 6·96 mg g^{-1} of polysaccharide from OX, whereas only 14·33 ± 1·64 mg g^{-1} of polysaccharide was produced from ARB. Based on HPLC data (not shown), the most abundant sugars released in the reaction with XB were (listed in order of estimated quantity): xylose>glucose>galactose>xylobiose. Regarding the OX and ARB reactions, we observed that the xylose and

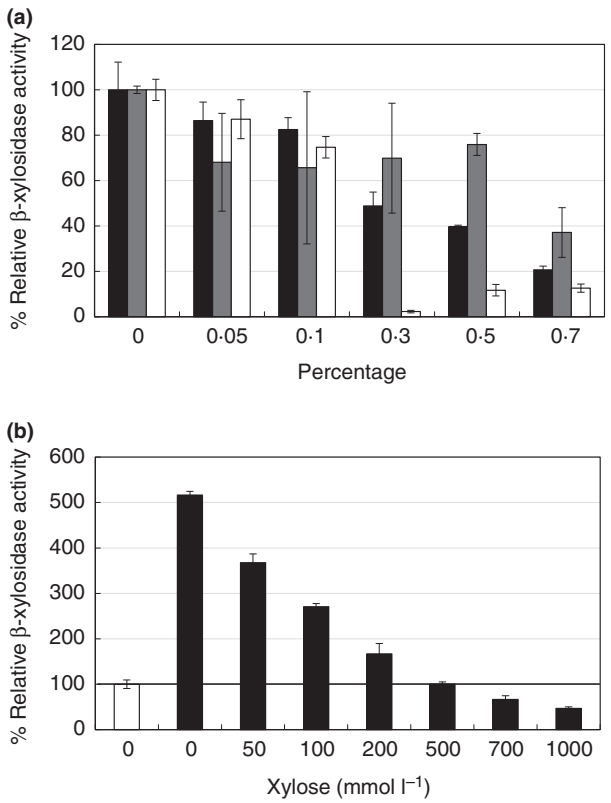

Figure 5 Relative *β*-xylosidase activity of the protein XylM1989 at (a) different concentrations (0–0·7%)% w/v) of furfural (■), 5-hydroxymethylfurfural (5-HMF) (▩) and acetic acid (□); and (b) different concentrations (0–1000 mmol l^{-1}) of xylose in the presence of 5 mmol l^{-1} ions (Ca^{2+}, Mg^{2+} and Mn^{2+}) (□ without ions; ■ with 5 mmol l^{-1} ions (Ca^{2+}, Mg^{2+} and Mn^{2+}).

glucose levels exceeded those of galactose, arabinose and xylobiose. On the basis of these results, we posit that protein XylM1989 works avidly on xylan from beechwood, oat spelt xylan and arabinoxylan, releasing sugars in accordance with the specifics of these substrates.

Discussion

In this study, 10 GH-encoding genes, retrieved from two wheat straw-degrading microbial consortia (Jiménez *et al.* 2015), were selected for codon optimization, synthesis, cloning, expression and characterization. These genes all originated from different Bacteroidetes, a dominant phylum in wood-degrading communities (Van der Lelie *et al.* 2012) and were found in different microbial consortia cultivated on agricultural residues (Brossi *et al.* 2015; Cortes-Tolalpa *et al.* 2016). The microbial origin, the genomic context and the annotation of the genes for these selected GHs all suggest that they play an important role in plant biomass degradation (Jiménez *et al.* 2015). Inspired by recent data (Elena *et al.* 2014), we intended to increase the expression levels of such GH-encoding genes (in *E. coli*), using codon swapping by a commercial routine. However, such codon optimization comes with potential drawbacks that are related to the fast depletion of the precise cognate tRNAs in the host organism, which can incite translational errors. An exact 'smart mix' of major and minor codons appears to be necessary in each case, which may be gene and host dependent. Indeed, the feasibility of any protocol for heterologous protein expression is often not theoretically foreseeable (Kurland and Gallant 1996). In our study, 9 of the 10 selected GHs were found in inclusion bodies inside the *E. coli* host, which was likely due to overexpression. Such inclusion body location is actually an advantage in the industrial production of high quantities of proteins. However, due to misfolding and aggregation, the included proteins may become inactive. To tackle this problem, recently different types of expression host (e.g. *E. coli* origami; Novagen), protein refolding kits and new protocols have been developed (Vallejo and Rinas 2004; Yamaguchi and Miyazaki 2014). In our study, the finding that the products of five genes (in the insoluble fractions) apparently had xylanase activity, whereas no activity could be detected with any of the *p*NP substrates, indicated that protein aggregates may have been differentially dissociated and thus (ephemerally) active. For instance, in the zymography assay, the removal of SDS by Triton may refold the protein correctly in the gel with subsequent evidence of activity (Peterson *et al.* 2011; Vandooren *et al.* 2013). Notably, by using zymography we showed, for the first time, that proteins (M1916 and M8239) annotated as fucosidases have xylanolytic activity (Fig. 1b; Table 1). Remarkably, in the presence of the *p*NP substrates, protein XylM1989 showed *β*-D-xylosidase activity and α-L-arabinosidase activity. Given the problems with the insolubility of the other proteins, we henceforth only studied the soluble protein XylM1989 in detail. Interestingly, the genomic context of the XylM1989-encoding gene resembled that of the gene for an uncharacterized enzyme (GH43B19) located in a 37·5-kb chromosome fragment of *Sphingobacterium* sp. TN19 (Fig. S3). This strain had been isolated from the gut of *Batocera horsfieldi* larvae, a beetle which develops in woody tissues. In the chromosome fragment, genes for three xylanolytic enzymes have been characterized (XynB19, GH43A19, XynA19), suggesting their importance in hemicellulose depolymerization (Zhou *et al.* 2009, 2010). We here found a shared synteny (with our contig_248) of genes encoding transketolases, xylulokinases, xylose isomerases and ABC transporters. These findings suggest that the flanking genes to the gene encoding XylM1989 produce proteins that are involved in xylan degradation, sugar transport and xylose metabolism (Fig. S3).

In earlier work, several β-xylosidases and α-arabinosidases of CAZy family GH43 were recovered and characterized from various organisms (e.g. *Bacillus*, *Fibrobacter*, *Thermobifida*, *Clostridium*, *Thermotoga*, *Enterobacter* and *Paenibacillus*) (Inácio and de Sá-Nogueira 2008; Kim and Yoon 2010; Yoshida *et al.* 2010; Fekete and Kiss 2012; Ahmed *et al.* 2013; Campos *et al.* 2014; Shi *et al.* 2014; Wang *et al.* 2014). However, this was often accompanied by an incomplete characterization which did not allow the understanding of their full potential. For instance, the tolerance to high levels of (inhibitor) compounds derived from lignocellulosic hydrolysate has not been adequately addressed (Van der Pol *et al.* 2014). It is worth noting that these GHs constitute key components of enzyme cocktails used for the improved saccharification of plant biomass and production of biofuels (Machado *et al.* 2015). The aforementioned studies showed that some, but not all, GH43 family enzymes have bifunctional activities, in particular β-xylosidase and α-arabinosidase. For instance, protein XylC, isolated from *Paenibacillus woosongensis*, had dual activity (Kim and Yoon 2010), and some proteins encoded by genes isolated from rumen and compost-derived metagenomes also had dual activity (Dougherty *et al.* 2012; Zhou *et al.* 2012; Matsuzawa *et al.* 2015). In contrast, the enzyme His-Xyl43 showed β-xylosidase activity (Campos *et al.* 2014), whereas proteins Abn2 and AbnZ2 had only endo-arabinase activity (Inácio and de Sá-Nogueira 2008; Wang *et al.* 2014). We speculated, on the basis of the foregoing, that the β-D-xylosidase/α-L-arabinosidase activity of our XylM1989 enzyme may constitute a key asset in wheat straw biomass saccharification, with specific involvement in xylan and arabinan degradation. Indeed, compared to previously described family GH43 enzymes, it showed a raised V_{max} (also compared with fungal enzymes) (Table 3). Thus, XylM1989 has a higher reaction speed, enabling a faster substrate processing when the protein is saturated with the substrate.

Regarding the biochemical characterization, the three bivalent cations Ca^{2+}, Mg^{2+} and Mn^{2+} clearly enhanced the activity of XylM1989, thus indicating that such cations are important as enzyme cofactors. A role in the enzymatic reaction, for example, by binding and stabilizing the substrate at the active site, may be invoked (Moraïs *et al.* 2012; Ahmed *et al.* 2013; Santos *et al.* 2014). In contrast, Cu^{2+}, Ni^{2+}, Co^{2+} and NH^{4+} strongly inhibited the β-xylosidase activity of XylM1989. Similar to XylM1989, a xylanase produced by a gene isolated from a bovine rumen metagenome showed enhanced β-xylosidase activity with Mn^{2+}, whereas Cu^{2+}, Fe^{2+}, Ag^{2+} and Zn^{2+} ions inhibited the activity (Gong *et al.* 2013). Helper molecules like sugars and ions can control enzyme activities by 'setting' proteins 'on' and 'off' in response to environmental changes. Thus, feedback inhibition may occur due to allosteric regulation, in which molecules bind to the catalytic site of enzymes, altering their structural shape and changing the protein to an active or inactive form. The fact that XylM1989 activity was slightly stimulated by L-arabinose may relate to the binding of this sugar to the substrate-binding site, thus enabling the XylM1989 catalytic residues to react effectively.

Furfural and 5-HMF are major byproducts from the pretreatment of lignocellulosic materials (Garrote *et al.* 2004; Klinke *et al.* 2004). These aromatic compounds

Table 3 Comparison of the protein XylM1989 with GH43 family enzymes from other studies

Enzyme	Microbial source	kDa	Substrate	Optimal pH	Optimal T (°C)	Activity (U mg^{-1})	K_m (mmol l^{-1})	V_{max} (μmol min^{-1} mg^{-1})	References
XylM1989	*Sphingobacterium* sp.	37·5	*p*NP-X/ *p*NP-A	8·0	20	9·65/1·22	1·2/0·781	285·71/78·12	This work
S2	*Penicillium herquei*	37·4	*p*NP-X	6·5	30	225	ND	ND	Ito *et al.* (2003)
Abn2	*Bacillus subtilis*	46·0	*p*NP-A	7·0	50	ND	ND	ND	Inácio and de Sá-Nogueira (2008)
FSUAXH1	*Fibrobacter succinogenes*	84·0	*p*NP-X	7·5	45	ND	ND	ND	Yoshida *et al.* (2010)
XylB	*Aspergillus oryzae*	37·4	*p*NP-X	7·0	30	6·1	0·48	42·6	Suzuki *et al.* (2010)
PtXyl43	*Paecilomyces thermophila*	52·3	*p*NP-X	7·0	55	45·4	4·5	90·2	Teng *et al.* (2011)
TlXyl43	*Thermomyces lanuginosus*	45·0	*p*NP-X	6·5	55	45·4	3·9	107·6	Chen *et al.* (2012)
RuXyn1	*Prevotella bryantii*	45·0	*p*NP-X/ *p*NP-A	7·0	40	36·3/14·2	3·43/2·23	ND	Zhou *et al.* (2012)
Xyl43A	*Humicola insolens*	37·0	*p*NP-X	6·5	50	20·5	12·2	203·8	Yang *et al.* (2014)
Xyl43B	*Humicola insolens*	62·0	*p*NPX	7·0	50	1·7	1·29	2·18	

ND, not determined.

released in hydrolysates are considered to be major inhibitors of fermentation processes (Van der Pol *et al.* 2014). During plant biomass pretreatment, several released products (including furanic compounds and monosaccharides) can inhibit the activity of the enzymes (or cocktails) used for the subsequent saccharification process. In addition, several studies have shown a strong inhibition of ethanol production due to the presence of lignocellulosic byproducts (Van der Pol *et al.* 2014). For example, concentrations of ~0·5% (w/v) and ~0·7% (w/v) of furfural and 5-HMF, respectively, can inhibit the growth and production of ethanol by *Issatchenkia orientalis* (Kwon *et al.* 2011). The concentration of furfural and 5-HMF in plant biomass hydrolysates depends on the pretreatment conditions and the feedstock. However, analyses of pretreated corn stover, poplar and pine materials showed furfural concentrations up to 0·22 g l^{-1} (0·022% w/v) and 5-HMF concentrations up to 0·17 g l^{-1} (0·017% w/v) (Du *et al.* 2010). Interestingly, protein XylM1989 was tolerant to furfural, as its β-xylosidase activity was still at approximately 80% at 0·5% (w/v) furfural. It also showed some (restricted) tolerance to 0·3% (w/v) 5-HMF (50%). We posit here that such furan-tolerant xylanases have great potential for use in the biorefining industries, as they would presumably work well in the presence of expected levels of furfurals and related compounds. As far as we know, this is the first report of a β-xylosidase/α-arabinosidase that is tolerant to furfural and 5-HMF. Moreover, as was explained before, β-xylosidases are key in the conversion of xylo-oligosaccharides to xylose as the end-product. However, xylose is one of the major inhibitors of β-xylosidase activity (Yan *et al.* 2008) Thus, xylose-tolerant β-xylosidases are important in hemicellulose conversion. The sensitivity of most β-xylosidases to xylose (as tested with fungal-produced xylosidases, such as those from *Arxula adeninivoran, Aureobasidium pullulans* and *Trichoderma reesei*) is striking, with *Ki* (concentration of inhibitor) values for xylose ranging from 2 to 10 mmol l^{-1} (Zanoelo *et al.* 2004). In contrast, *Thermotoga thermarum* Tth xynB3 β-xylosidase showed high xylose-tolerant activity at 500 mmol l^{-1} (Shi *et al.* 2013). Interestingly, our novel xylM1989 protein showed 100% activity at xylose concentrations of 500 mmol l^{-1}, in the presence of ions (either Ca^{2+}, Mg^{2+} or Mn^{2+}). Strikingly, the XylM1989 activity in the presence of xylose (with ions) was even enhanced by relatively low levels of xylose. In conclusion, our enzyme is inhibited by xylose, but this inhibition is alleviated by the presence of Ca^{2+}, Mg^{2+} or Mn^{2+}.

Finally, five key features makes that the protein XylM1989 is a good candidate for use in industrial processes related with plant biomass saccharification: (i) active at alkaline pH, (ii) high reaction speed (V_{max}), (iii) tolerance to lignocellulosic hydrolysate-derived inhibitors, (iv) tolerance to high concentration of xylose in the presence of Ca^{2+}, Mg^{2+} and Mn^{2+} and (v) activity and release of sugars from complex polysaccharides such as XB, OX and ARB. These properties of XylM1989 turn this enzyme into a promising puzzle part for the design of enzyme cocktails useful for the saccharification of (pretreated) plant biomass.

Acknowledgements

This work was supported by the BE-Basic foundation (http://www.be-basic.org). We thank H. Ruijssenaars and R. van Kranenburg for scientific support.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

Ahmed, S., Luis, A.S., Bras, J.L., Ghosh, A., Gautam, S., Gupta, M.N., Fontes, C.M. and Goyal, A. (2013) A novel α-L-arabinofuranosidase of family 43 glycoside hydrolase (Ct43Araf) from *Clostridium thermocellum*. *PLoS ONE* **8**, e73575.

Barker, I.J., Petersen, L. and Reilly, P.J. (2010) Mechanism of xylobiose hydrolysis by GH43 β-xylosidase. *J Phys Chem B* **114**, 15389–15393.

Bokhari, S.A.I., Latif, F., Akhtar, M.W. and Rajoka, M.I. (2010) Characterization of a β-xylosidase produced by a mutant derivative of *Humicola lanuginosa* in solid state fermentation. *Ann Microbiol* **60**, 21–29.

Brossi, M.J., Jiménez, D.J., Cortes-Tolalpa, L. and van Elsas, J.D. (2015) Soil-derived microbial consortia enriched with different plant biomass reveal distinct players acting in lignocellulose degradation. *Microb Ecol* **71**, 616–627.

Campos, E., Negro Alvarez, M.J., Sabarís di Lorenzo, G., Gonzalez, S., Rorig, M., Talia, P., Grasso, D.H., Sáez, F. *et al.* (2014) Purification and characterization of a GH43 β-xylosidase from *Enterobacter* sp. identified and cloned from forest soil bacteria. *Microbiol Res* **169**, 213–220.

Chen, Z., Jia, H., Yang, Y., Yan, Q., Jiang, Z. and Teng, C. (2012) Secretory expression of a β-xylosidase gene from *Thermomyces lanuginosus* in *Escherichia coli* and characterization of its recombinant enzyme. *Lett Appl Microbiol* **55**, 330–337.

Cortes-Tolalpa, L., Jiménez, D.J., de Lima Brossi, M.J., Salles, J.F. and van Elsas, J.D. (2016) Different inocula produce distinctive microbial consortia with similar lignocellulose degradation capacity. *Appl Microbiol Biotechnol* **100**, 7713–7725.

De Souza, W.R. (2013) Microbial degradation of lignocellulosic biomass. In *Sustainable Degradation of*

Lignocellulosic Biomass Techniques, Applications and Commercialization ed. Chandel, A. and Da Silva, S. pp. 208–209. Croatia: InTech.

D'haeseleer, P., Gladden, J.M., Allgaier, M., Chain, P.S., Tringe, S.G., Malfatti, S.A., Aldrich, J.T., Nicora, C.D. *et al.* (2013) Proteogenomic analysis of a thermophilic bacterial consortium adapted to deconstruct switchgrass. *PLoS ONE* **8**, e68465.

Dougherty, M.J., D'haeseleer, P., Hazen, T.C., Simmons, B.A., Adams, P.D. and Hadi, M.Z. (2012) Glycoside hydrolases from a targeted compost metagenome, activity-screening and functional characterization. *BMC Biotechnol* **12**, 38.

Du, B., Sharma, L.N., Becker, C., Chen, S.F., Mowery, R.A., van Walsum, G.P. and Chambliss, C.K. (2010) Effect of varying feedstock-pretreatment chemistry combinations on the formation and accumulation of potentially inhibitory degradation products in biomass hydrolysates. *Biotechnol Bioeng* **107**, 430–440.

Elena, C., Ravasi, P., Castelli, M.E., Peirú, S. and Menzella, H.G. (2014) Expression of codon optimized genes in microbial systems: current industrial applications and perspectives. *Front Microbiol* **5**, 21.

Fekete, C.A. and Kiss, L. (2012) Purification and characterization of a recombinant β-D-xylosidase from *Thermobifida fusca* TM51. *Protein J* **31**, 641–650.

Gao, D., Uppugundla, N., Chundawat, S.P., Yu, X., Hermanson, S., Gowda, K., Brumm, P., Mead, D. *et al.* (2011) Hemicellulases and auxiliary enzymes for improved conversion of lignocellulosic biomass to monosaccharides. *Biotechnol Biofuels* **4**, 5.

Garrote, G., Cruz, J.M., Moure, A., Domínguez, H. and Parajó, J.C. (2004) Antioxidant activity of byproducts from the hydrolytic processing of selected lignocellulosic materials. *Trends Food Sci Technol* **15**, 191–200.

Gasparic, A., Martin, J., Daniel, A.S. and Flint, H.J. (1995) A xylan hydrolase gene cluster in *Prevotella ruminicola* B(1)4: sequence relationships, synergistic interactions, and oxygen sensitivity of a novel enzyme with exoxylanase and beta-(1,4)-xylosidase activities. *Appl Environ Microbiol* **61**, 2958–2964.

Gladden, J.M., Park, J.I., Bergmann, J., Reyes-Ortiz, V., D'haeseleer, P., Quirino, B.F., Sale, K.L., Simmons, B.A. *et al.* (2014) Discovery and characterization of ionic liquid-tolerant thermophilic cellulases from a switchgrass-adapted microbial community. *Biotechnol Biofuels* **7**, 15.

Gong, X., Gruniniger, R.J., Forster, R.J., Teather, R.M. and McAllister, T.A. (2013) Biochemical analysis of a highly specific, pH stable xylanase gene identified from a bovine rumen-derived metagenomic library. *Appl Microbiol Biotechnol* **97**, 2423–2431.

Inácio, J.M. and de Sá-Nogueira, I. (2008) Characterization of abn2 (yxiA), encoding a *Bacillus subtilis* GH43 arabinanase, Abn2, and its role in arabino-polysaccharide degradation. *J Bacteriol* **190**, 4272–4280.

Ito, T., Yokoyama, E., Sato, H., Ujita, M., Funaguma, T., Furukawa, K. and Hara, A. (2003) Xylosidases associated with the cell surface of *Penicillium herquei* IFO 4674. *J Biosci Bioeng* **96**, 354–359.

Jiménez, D.J., Dini-Andreote, F. and van Elsas, J.D. (2014) Metataxonomic profiling and prediction of functional behaviour of wheat straw degrading microbial consortia. *Biotechnol Biofuels* **7**, 92.

Jiménez, D.J., Chaves-Moreno, D. and van Elsas, J.D. (2015) Unveiling the metabolic potential of two soil-derived microbial consortia selected on wheat straw. *Sci Rep* **5**, 13845.

Jiménez, D.J., de Lima Brossi, M.J., Schückel, J., Kračun, S.K., Willats, W.G. and van Elsas, J.D. (2016) Characterization of three plant biomass-degrading microbial consortia by metagenomics- and metasecretomics-based approaches. *Appl Microbiol Biotechnol* **100**, 10463–10477.

Jordan, D.B., Wagschal, K., Grigorescu, A.A. and Braker, J.D. (2013) Highly active β-xylosidases of glycoside hydrolase family 43 operating on natural and artificial substrates. *Appl Microbiol Biotechnol* **97**, 4415–4428.

Jordan, D.B., Braker, J.D., Wagschal, K., Lee, C.C., Chan, V.J., Dubrovska, I., Anderson, S. and Wawrzak, Z. (2015) X-ray crystal structure of divalent metal-activated β-xylosidase, RS223BX. *Appl Biochem Biotechnol* **177**, 637–648.

Kelley, L.A. and Sternberg, M.J.E. (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363–371.

Kim, Y.A. and Yoon, K.H. (2010) Characterization of a *Paenibacillus woosongensis* beta-Xylosidase/alpha-Arabinofuranosidase produced by recombinant *Escherichia coli*. *J Microbiol Biotechnol* **20**, 1711–1716.

Klinke, H.B., Thomsen, A.B. and Ahring, B.K. (2004) Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl Microbiol Biotechnol* **66**, 10–26.

Kurland, C. and Gallant, J. (1996) Errors of heterologous protein expression. *Curr Opin Biotechnol* **7**, 489–493.

Kwon, Y.J., Ma, A.Z., Li, Q., Wang, F., Zhuang, G.Q. and Liu, C.Z. (2011) Effect of lignocellulosic inhibitory compounds on growth and ethanol fermentation of newly-isolated thermotolerant *Issatchenkia orientalis*. *Bioresour Technol* **102**, 8099–8104.

Lagaert, S., Pollet, A., Courtin, C.M. and Volckaert, G. (2014) β-Xylosidases and α-l-arabinofuranosidases: accessory enzymes for arabinoxylan degradation. *Biotechnol Adv* **32**, 316–332.

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

Machado, C.B., Citadini, A.P., Goldbeck, R., de Lima, E.A., Figueiredo, F.L., da Silva, T.M., Hoffmam, Z.B., de Sousa, A.S. *et al.* (2015) Increased biomass saccharification by supplementation of a commercial enzyme cocktail with

endo-arabinanase from *Bacillus licheniformis*. *Biotechnol Lett* **37**, 1455–1462.

Matsuzawa, T., Kaneko, S. and Yaoi, K. (2015) Screening, identification, and characterization of a GH43 family β-xylosidase/α-arabinofuranosidase from a compost microbial metagenome. *Appl Microbiol Biotechnol* **99**, 8943–8954.

Miller, G.L. (1959) Use of dinitrosalicylic acid reagent for determination of reducing sugar. *Anal Chem* **31**, 426–428.

Moraïs, S., Salama-Alber, O., Barak, Y., Hadar, Y., Wilson, D.B., Lamed, R., Shoham, Y. and Bayer, E.A. (2012) Functional association of catalytic and ancillary modules dictates enzymatic activity in glycoside hydrolase family 43 β-xylosidase. *J Biol Chem* **287**, 9213–9221.

Peterson, R., Grinyer, J. and Nevalainen, H. (2011) Extracellular hydrolase profiles of fungi isolated from koala faeces invite biotechnological interest. *Mycol Prog* **10**, 207–218.

Qi, X., Sun, Y. and Xiong, S. (2015) A single freeze-thawing cycle for highly efficient solubilization of inclusion body proteins and its refolding into bioactive form. *Microb Cell Fact* **14**, 24.

Saha, B.C. (2003) Hemicellulose bioconversion. *J Ind Microbiol Biotechnol* **30**, 279–291.

Sambrook, J. and Russell, D.W. (2001) *Molecular Cloning. A laboratory Manual* (3rd edn). Cold Spring Harbor: Cold Spring Harbor Laboratory.

Santos, C.R., Polo, C.C., Costa, M.C.M.F., Nascimento, A.F.Z., Meza, A.N., Cota, J., Hoffmam, Z.B., Honorato, R.V. *et al.* (2014) Mechanistic strategies for catalysis adopted by evolutionary distinct family 43 arabinanases. *J Biol Chem* **289**, 7362–7373.

Shi, H., Li, X., Gu, H., Zhang, Y., Huang, Y., Wang, L. and Wang, F. (2013) Biochemical properties of a novel thermostable and highly xylose-tolerant β-xylosidase/α-arabinosidase from *Thermotoga thermarum*. *Biotechnol Biofuels* **6**, 27.

Shi, H., Ding, H., Huang, Y., Wang, L., Zhang, Y., Li, X. and Wang, F. (2014) Expression and characterization of a GH43 endo-arabinanase from *Thermotoga thermarum*. *BMC Biotechnol* **14**, 35.

Simossis, V.A. and Heringa, J. (2005) PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. *Nucleic Acids Res* **33**, W289–W294.

Suzuki, S., Fukuoka, M., Ookuchi, H., Sano, M., Ozeki, K., Nagayoshi, E., Takii, Y., Matsushita, M. *et al.* (2010) Characterization of *Aspergillus oryzae* glycoside hydrolase family 43 beta-xylosidase expressed in *Escherichia coli*. *J Biosci Bioeng* **109**, 115–117.

Tamura, K., Stecher, G., Peterson, D., Filipski, A. and Kumar, S. (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol Biol Evol* **30**, 2725–2729.

Teng, C., Jia, H., Yan, Q., Zhou, P. and Jiang, Z. (2011) High-level expression of extracellular secretion of a β-xylosidase gene from *Paecilomyces thermophila* in *Escherichia coli*. *Bioresour Technol* **102**, 1822–1830.

Vallejo, L. and Rinas, U. (2004) Strategies for the recovery of active proteins through refolding of bacterial inclusion body proteins. *Microb Cell Fact* **3**, 11.

Van den Brink, J. and de Vries, R.P. (2011) Fungal enzyme sets for plant polysaccharide degradation. *Appl Microbiol Biotechnol* **91**, 1477–1492.

Van der Lelie, D., Taghavi, S., McCorkle, S.M., Li, L.L., Malfatti, S.A., Monteleone, D., Donohoe, B.S., Ding, S.Y. *et al.* (2012) The metagenome of an anaerobic microbial community decomposing poplar wood chips. *PLoS ONE* **7**, e36740.

Van der Pol, E.C., Bakker, R.R., Baets, P. and Eggink, G. (2014) By-products resulting from lignocellulose pretreatment and their inhibitory effect on fermentations for (bio)chemicals and fuels. *Appl Microbiol Biotechnol* **98**, 9579–9593.

Vandooren, J., Geurts, N., Martens, E., Van den Steen, P.E. and Opdenakker, G. (2013) Zymography methods for visualizing hydrolytic enzymes. *Nat Methods* **10**, 211–220.

Wang, S., Yang, Y., Zhang, J., Sun, J., Matsukawa, S., Xie, J. and Wei, D. (2014) Characterization of abnZ2 (yxiA1) and abnZ3 (yxiA3) in *Paenibacillus polymyxa*, encoding two novel endo-1,5-α-l-arabinanases. *Bioresour Bioprocess* **1**, 14.

Yamaguchi, H. and Miyazaki, M. (2014) Refolding techniques for recovering biologically active recombinant proteins from inclusion bodies. *Biomolecules* **4**, 235–251.

Yan, Q.J., Wang, L., Jiang, Z.Q., Yang, S.Q., Zhu, H.F. and Li, L.T. (2008) A xylose-tolerant beta-xylosidase from *Paecilomyces thermophila*: characterization and its co-action with the endogenous xylanase. *Bioresour Technol* **99**, 5402–5410.

Yang, X., Shi, P., Huang, H., Luo, H., Wang, Y., Zhang, W. and Yao, B. (2014) Two xylose-tolerant GH43 bifunctional β-xylosidase/α-arabinosidases and one GH11 xylanase from *Humicola insolens* and their synergy in the degradation of xylan. *Food Chem* **148**, 381–387.

Yoshida, S., Hespen, C.W., Beverly, R.L., Mackie, R.I. and Cann, I.K. (2010) Domain analysis of a modular alpha-L-Arabinofuranosidase with a unique carbohydrate binding strategy from the fiber-degrading bacterium *Fibrobacter succinogenes* S85. *J Bacteriol* **192**, 5424–5436.

Zanoelo, F.F., Polizeli Md Mde, L., Terenzi, H.F. and Jorge, J.A. (2004) Purification and biochemical properties of a thermostable xylose-tolerant beta- D-xylosidase from *Scytalidium thermophilum*. *J Ind Microbiol Biotechnol* **31**, 170–176.

Zhou, J., Huang, H., Meng, K., Shi, P., Wang, Y., Luo, H., Yang, P., Bai, Y. *et al.* (2009) Molecular and biochemical characterization of a novel xylanase from the symbiotic

Sphingobacterium sp. TN19. *Appl Microbiol Biotechnol* **85**, 323–333.

Zhou, J., Meng, K., Yang, P., Shi, P., Wang, Y., Luo, H. and Yao, B. (2010) Characterization of a chromosomal segment showing xylanolytic activity from the symbiotic *Sphingobacterium* sp. TN19. *World J Microbiol Biotechnol* **26**, 761–765.

Zhou, J., Bao, L., Chang, L., Zhou, Y. and Lu, H. (2012) Biochemical and kinetic characterization of GH43 β-D-xylosidase/α-L-arabinofuranosidase and GH30 α-L-arabinofuranosidase/β-D -xylosidase from rumen metagenome. *J Ind Microbiol Biotechnol* **39**, 143–152.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Genomic context of the selected GHs-encoding genes that were subsequently used for codon optimization, synthesis, cloning and expression.

Figure S2. Restriction profile of the GHs-encoding genes cloned in the pET21b(+) expression vector.

Figure S3. Genomic context of the XylM1989 in comparison with a genomic fragment from *Sphingobacterium* sp. TN19. ABCT means ABC transporters.

