ORIGINAL ARTICLE

Characterization of a furan aldehyde-tolerant \( \beta \)-xylosidase/\( \alpha \)-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, \( \alpha \)-arabinosidase, \( \beta \)-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 \( E. coli \). 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 \( \beta \)-xylosidase and \( \alpha \)-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 \( \beta \)-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 \( \beta \)-1,4-linked \( D \)-xylose units, which may be substituted by different side groups, such as \( D \)-galactose, \( L \)-arabinose, glu- curonic 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, \( \beta \)-1,4-endoxylanase (EC 3.2.1.8), which cleaves the backbone into small oligosaccharides, and \( \beta \)-1,4-xylosidase (EC 3.2.1.37), which cleaves these oligosaccharides into xylose. Next to breaking the side chains of xylan, enzymes like \( \alpha \)-l-arabinosidases, \( \alpha \)-D-glu- curonidases 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, synthesized 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, 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-xylanase/x-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, Xbal/XhoI were used for genes M3030, M1916, M1927, M1989 and M20752; MluI/ XhoI for genes M8239, M8244 and M7068; SalI/ XhoI for M4684; and Sall/ XhoI 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)
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 (pNP-Xyl), p-nitrophenyl α-L-arabinofuranoside (pNP-Ara), p-nitrophenyl α-L-fucopyranoside (pNP-Fuc) and p-nitrophenyl α-D-mannopyranoside (pNP-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 pH21b+ 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 (xylM1989) 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 praligne (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-xylanase 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 pNP-Xyl and pNP-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 pNP-Xyl or pNP-Ara and 20 µl of soluble protein XylM1989 (approximately 1.7 mg ml⁻¹). The kinetic parameters (Kₘ and Vₘₐₓ) of XylM1989 were determined with pNP-Xyl and pNP-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 hydrolysatase 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 pNP-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 Mg²⁺, Ca²⁺ and Mn²⁺ was evaluated with the above parameters. Enzymatic activities were determined from the measured absorbance units using a standard calibration curve. The amount of paranitrophenol (pNP) 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 pNP 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-dinitrosalicylic 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, synthesized and cloned into the pET21b+ expression vector with specific restriction enzyme sites on both sides (Ndel and XhoI). 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 pNP-labelled substrates, that is, pNP-Xyl, pNP-Ara, pNP-Fuc and pNP-Man, to evaluate the putative
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 pNP-Xyl and pNP-Ara, but not with pNP-Man and pNP-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 3.2.1.45) and endo-α-arabinases (EC 3.2.1.99) (Jordon 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 (xynB; 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 bacterial xylosidases and arabinosidases from different Bacteroidetes, specifically belonging to species of Sphingobacterium, Dracobicaterium, 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
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) pNP-Xyl (3.36 U mg⁻¹ of protein) and pNP-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 pNP-Xyl and pNP-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 pNP-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 pNP-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 pNP-β-D-xylopyranoside (pNP-Xyl) (■) and pNP-α-L-arabinopyranoside (pNP-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 $V_{\text{max}}$ values (maximal reaction velocities), using pNP-Xyl and pNP-Ara as the substrates, under optimal conditions. The $K_m$ values of XylM1989 for pNP-Xyl and pNP-Ara were 1.2 and 0.781 mmol l$^{-1}$, and the $V_{\text{max}}$ values 285-71 and 78-12 U mg$^{-1}$ respectively. Given the fact that the $K_m$ and $V_{\text{max}}$ values of protein XylM1989 for pNP-Xyl were higher than those for pNP-Ara, pNP-Xyl was used for further analysis.

Effects of additives on protein XylM1989 activity

The effects of different additives on protein XylM1989 $\beta$-xylosidase activity were assessed (Table 2). Remarkable increases in the $\beta$-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.

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 pNP-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 $\beta$-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 $\beta$-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 $\beta$-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

<table>
<thead>
<tr>
<th>Type of additive</th>
<th>Reagent</th>
<th>% Relative activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Salts</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CaCl$_2$</td>
<td>5 mmol l$^{-1}$</td>
<td>20 mmol l$^{-1}$</td>
</tr>
<tr>
<td>MgCl$_2$</td>
<td>1065-23 ± 74</td>
<td>546-58 ± 39</td>
</tr>
<tr>
<td>MnCl$_2$</td>
<td>1214-18 ± 16</td>
<td>691-11 ± 120</td>
</tr>
<tr>
<td>NiCl$_2$</td>
<td>6-27 ± 1</td>
<td>1-33 ± 0.3</td>
</tr>
<tr>
<td>CuCl$_2$</td>
<td>8-67 ± 1</td>
<td>2-01 ± 1.2</td>
</tr>
<tr>
<td>CuCl$_2$</td>
<td>0-63 ± 0.1</td>
<td>1-94 ± 0.4</td>
</tr>
<tr>
<td>NH$_4$Cl</td>
<td>738-93 ± 8.4</td>
<td>172-88 ± 20</td>
</tr>
<tr>
<td>Glucose</td>
<td>69-32 ± 1</td>
<td>63-61 ± 4.3</td>
</tr>
<tr>
<td>Xylose</td>
<td>20 mmol l$^{-1}$</td>
<td>50 mmol l$^{-1}$</td>
</tr>
<tr>
<td>Arabinose</td>
<td>90-30 ± 1.3</td>
<td>99-28 ± 4.9</td>
</tr>
<tr>
<td>Cellobiose</td>
<td>55-82 ± 2.5</td>
<td>53-71 ± 1.4</td>
</tr>
<tr>
<td>Galactose</td>
<td>164-89 ± 5.4</td>
<td>249-26 ± 3.8</td>
</tr>
<tr>
<td>Salt</td>
<td>80-38 ± 1.8</td>
<td>90-99 ± 12</td>
</tr>
<tr>
<td>NaCl</td>
<td>200 mmol l$^{-1}$</td>
<td>2000 mmol l$^{-1}$</td>
</tr>
<tr>
<td>Chelating agent</td>
<td>96-06 ± 6.1</td>
<td>36-34 ± 0.6</td>
</tr>
<tr>
<td>EDTA</td>
<td>61-66 ± 0.8</td>
<td>30-20 ± 0.6</td>
</tr>
<tr>
<td>Detergents</td>
<td>2%</td>
<td>5%</td>
</tr>
<tr>
<td>SDS</td>
<td>8-68 ± 0.7</td>
<td>6-12 ± 4.0</td>
</tr>
<tr>
<td>Triton</td>
<td>107 ± 5.6</td>
<td>32-70 ± 0.8</td>
</tr>
<tr>
<td>Organic solvents</td>
<td>10%</td>
<td>20%</td>
</tr>
<tr>
<td>Glycerol</td>
<td>101-9 ± 4.7</td>
<td>118-6 ± 24</td>
</tr>
<tr>
<td>Ethanol</td>
<td>26-63 ± 0.1</td>
<td>11-40 ± 0.2</td>
</tr>
<tr>
<td>Methanol</td>
<td>26-47 ± 0.2</td>
<td>16-63 ± 1.9</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>26-10 ± 0.6</td>
<td>12-53 ± 1.9</td>
</tr>
<tr>
<td>DMSO</td>
<td>50-56 ± 1.0</td>
<td>24-55 ± 0.1</td>
</tr>
</tbody>
</table>
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 pNP 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 pNP substrates, protein XylM1989 showed β-D-xylanase 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, xylene 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). 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-xylosidase/α-1-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 Vmax (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 Ca2+, Mg2+ and Mn2+ 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 (Moraës et al. 2012; Ahmed et al. 2013; Santos et al. 2014). In contrast, Cu2+, Ni2+, Co2+ and NH4+ 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 Mn2+, whereas Cu2+, Fe2+, Ag2+ and Zn2+ 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>
<thead>
<tr>
<th>Enzyme</th>
<th>Microbial source</th>
<th>kDa</th>
<th>Substrate</th>
<th>Optimal pH</th>
<th>Optimal T (°C)</th>
<th>Activity (U mg⁻¹)</th>
<th>Kₘ (mmol l⁻¹)</th>
<th>Vmax (μmol min⁻¹ mg⁻¹)</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>XylM1989</td>
<td><em>Sphingobacterium</em> sp.</td>
<td>30</td>
<td>pNP-X/</td>
<td>6.5</td>
<td>50</td>
<td>225</td>
<td>ND</td>
<td>25</td>
<td>This work</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>pNP-A</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td><em>Penicillium herquei</em></td>
<td>37.4</td>
<td>pNP-X</td>
<td>6.5</td>
<td>30</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>Ito et al. (2003)</td>
</tr>
<tr>
<td>Abn2</td>
<td><em>Bacillus subtilis</em></td>
<td>46.0</td>
<td>pNP-A</td>
<td>7.0</td>
<td>50</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>Inácio and de Sá-Nogueira (2008)</td>
</tr>
<tr>
<td>FSUAXH1</td>
<td><em>Fibrobacter succinogenes</em></td>
<td>84.0</td>
<td>pNP-X</td>
<td>7.5</td>
<td>45</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
<td>Yoshida et al. (2010)</td>
</tr>
<tr>
<td>XyLB</td>
<td><em>Aspergillus oryzae</em></td>
<td>37.4</td>
<td>pNP-X</td>
<td>7.0</td>
<td>30</td>
<td>6-1</td>
<td>0.48</td>
<td>42.6</td>
<td>Suzuki et al. (2010)</td>
</tr>
<tr>
<td>PtxXyl43</td>
<td><em>Paecilomyces thermophila</em></td>
<td>52.3</td>
<td>pNP-X</td>
<td>7.0</td>
<td>55</td>
<td>45-4</td>
<td>4.5</td>
<td>90.2</td>
<td>Teng et al. (2011)</td>
</tr>
<tr>
<td>TIXyl43</td>
<td><em>Thermomyces lanuginosus</em></td>
<td>45.0</td>
<td>pNP-X</td>
<td>6.5</td>
<td>55</td>
<td>45-4</td>
<td>3.9</td>
<td>107.6</td>
<td>Chen et al. (2012)</td>
</tr>
<tr>
<td>RuXyn1</td>
<td><em>Prevotella bryantii</em></td>
<td>45.0</td>
<td>pNP-X/</td>
<td>7.0</td>
<td>40</td>
<td>36-3/14-2</td>
<td>3-4/2.23</td>
<td>ND</td>
<td>Zhou et al. (2012)</td>
</tr>
<tr>
<td>Xyl43A</td>
<td><em>Humicola insolens</em></td>
<td>37.0</td>
<td>pNP-X</td>
<td>6.5</td>
<td>50</td>
<td>20-5</td>
<td>12.2</td>
<td>203.8</td>
<td>Yang et al. (2014)</td>
</tr>
<tr>
<td>Xyl43B</td>
<td><em>Humicola insolens</em></td>
<td>62.0</td>
<td>pNPX</td>
<td>7.0</td>
<td>50</td>
<td>1-7</td>
<td>1.29</td>
<td>2.18</td>
<td></td>
</tr>
</tbody>
</table>
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 \(\beta\)-xylanase 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 tolerant xylanases have great potential for use in the biorefining industries, as they would presumably work

### Acknowledgements

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

### Conflict of Interest

The authors declare that they have no conflict of interest.

### References


**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.