Novel bacterial enzymes for plant biomass degradation discovered by meta-omics approach
Marutha Muthu, Mukil

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2017

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Characterization of a furan aldehyde-tolerant β-xylosidase/α-arabinosidase obtained through a synthetic metagenomics approach.

Mukil Maruthamuthu*, Diego Javier Jiménez*, and Jan Dirk van Elsas.
*Equal contribution

Abstract

**Aim:** The aim of the study was to characterize ten hemicellulolytic enzymes obtained from a wheat straw-degrading microbial consortium.

**Methods and results:** Based on previous metagenomics analyses, ten 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 *xylM1989*, 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 Ca²⁺, Mg²⁺ and Mn²⁺ 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 [1,2]. 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 [3–5]. Such enzymes are thought to be valuable for diverse industrial (e.g. food, pharma, plastics and biofuels) applications [6].

In spite of the promise of using microorganisms 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 analyze lignocellulose-enriched microbial consortia by DNA-based approaches (as also known as targeted metagenomics) [7]. For example, recently, Jiménez et al. (2016) [8] reported the analysis of three soil-derived microbial consortia cultivated on biologically pretreated plant biomass. They analyzed 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) [9] 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) [10] 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 [11]. 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 metagenomics DNA extracted from the aforementioned consortia - were selected and screened for GH-encoding genes [12]. In the current study, we report the selection of ten such genes on the basis of a combination of criteria. The genes were codon-optimized, synthesized and expressed, after which they were further
Here, we describe the full analysis, placing a focus on a gene for a key furan aldehyde-tolerant \( \beta \)-xylosidase/\( \alpha \)-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 thirteen novel Bacteroidetes-derived hemicellulose utilization loci containing 39 GHs [12]. From the contigs, we selected ten 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, xylloglucan 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 FigureS1 (supporting information), the selected genes and their genomic context are shown.

**Cloning and expression of ten GH-encoding genes recovered from the metagenome assemblages.**

The ten genes were selected based on directed choices with respect to CAZy [13] 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, USA). For codon optimization, we used the OptimumGene \(^{\text{TM}}\) algorithm and the Codon Adaptation Index (CAI). 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, USA) using the manufacturer’s instructions. Following clone selection and purification, plasmid extractions were done for each of the ten cloned genes. Thereafter, the nature of the cloned fragments was checked by restriction fragment analyses. Specifically, *XbaI/XhoI* were used for genes M3030, M1916, M1927, M1989 and M20752; *MluI/XhoI* for genes M8239, M8244 and M7068; *SalI/XhoI* for M4684; and *SacI/XhoI* for M7073 (Figure S2). A single colony of each verified clone was then
introduced into an Erlenmeyer containing 10ml 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 200rpm. The cell pellets were harvested by centrifugation at 10000 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 (6sec on, 15sec off, 30 cycles with amplitude of 10-15 microns) and the lysates centrifuged at 14000g for 10min at 4°C in order to separate the soluble and insoluble protein fractions. Insoluble proteins were washed twice with 750µl of 20 mmol l⁻¹ Tris-HCL (pH 8.0) and solubilized with 2 mol l⁻¹ urea following the freeze-thawing method [13]. Protein concentrations were determined by the Bradford method using bovine serum albumin as the standard. Protein fractions were analyzed by 10% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) [15].

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 20µg (in 20µl) of total proteins per sample. After running the gels at 4°C, they were soaked for 1h in 2.5% of Triton and washed thoroughly in water prior to incubation (1h 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 30min and then de-stained for 2h in 1 mol l⁻¹ NaCl to reveal zones of clearing. Additionally, protein fractions (soluble and insoluble) were recovered and tested for activity using $\text{p}$-nitrophenyl $\beta$-D-xylopyranoside ($\text{pNP-Xyl}$), $\text{p}$-nitrophenyl $\alpha$-L-arabinofuranoside ($\text{pNP-Ara}$), $\text{p}$-nitrophenyl $\alpha$-L-fucopyranoside ($\text{pNP-Fuc}$) and $\text{p}$-nitrophenyl $\alpha$-D-mannopyranoside ($\text{pNP-Man}$). The reaction mixtures consisted of 180µl of 2 mmol l⁻¹ of each $\text{p}$-nitrophenol-glycoside (diluted in 20 mmol l⁻¹ of Tris-HCL pH 7.0) and 20µl of each protein fraction. The mixtures were incubated at 37°C for 30min, 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 color in the reaction plate.
Table 1. Features of the overexpressed selected GH-encoding genes

<table>
<thead>
<tr>
<th>Gene ID</th>
<th>CAZy family</th>
<th>Gene length (bp)</th>
<th>Amino acids</th>
<th>PSI-BLASTp best hit [Taxa] (Accession number)</th>
<th>QC*</th>
<th>Identity</th>
<th>pI**</th>
<th>kDa</th>
</tr>
</thead>
<tbody>
<tr>
<td>M4684</td>
<td>GH92</td>
<td>3,111</td>
<td>1,032</td>
<td>Hypothetical protein [Parabacteroides sp.] (WP_010801039.1)</td>
<td>100%</td>
<td>70%</td>
<td>6.04</td>
<td>117</td>
</tr>
<tr>
<td>M3030</td>
<td>GH92</td>
<td>1,983</td>
<td>657</td>
<td>Alpha-1,2-mannosidase [Sphingobacterium spiritivorum] (WP_002997981.1)</td>
<td>98%</td>
<td>83%</td>
<td>9.24</td>
<td>74.2</td>
</tr>
<tr>
<td>M7068</td>
<td>GH43</td>
<td>996</td>
<td>328</td>
<td>Glycosyl hydrolase family 32 [Paraprevotella clara] (WP_008623026.1)</td>
<td>88%</td>
<td>69%</td>
<td>6.51</td>
<td>37.9</td>
</tr>
<tr>
<td>M7073</td>
<td>GH43</td>
<td>1,935</td>
<td>641</td>
<td>Glycosyl hydrolase [Flavobacterium johnsoniae] (YP_001195445.1)</td>
<td>98%</td>
<td>66%</td>
<td>7.11</td>
<td>73.3</td>
</tr>
<tr>
<td>M1916</td>
<td>GH29</td>
<td>1,641</td>
<td>543</td>
<td>Alpha-1,3/4-fucosidase [Capnocytophaga canimorsus] (YP_004740108.1)</td>
<td>98%</td>
<td>63%</td>
<td>6.53</td>
<td>61.5</td>
</tr>
<tr>
<td>M1927</td>
<td>GH2</td>
<td>3,216</td>
<td>1,068</td>
<td>Beta-galactosidase [Sphingobacterium spiritivorum] (WP_002995396.1)</td>
<td>85%</td>
<td>50%</td>
<td>7.08</td>
<td>120</td>
</tr>
<tr>
<td>M1989</td>
<td>GH43</td>
<td>981</td>
<td>323</td>
<td>Hypothetical protein [Sphingobacterium sp.] (WP_021189556.1)</td>
<td>100%</td>
<td>98%</td>
<td>4.70</td>
<td>37.5</td>
</tr>
<tr>
<td>M20752</td>
<td>GH2</td>
<td>1,293</td>
<td>427</td>
<td>Beta-galactosidase/beta-glucuronidase [Flavobacterium sp.] (WP_007809792.1)</td>
<td>90%</td>
<td>56%</td>
<td>8.63</td>
<td>48.5</td>
</tr>
<tr>
<td>M8239</td>
<td>GH95</td>
<td>1,302</td>
<td>430</td>
<td>Alpha-L-fucosidase [Pedobacter saltans] (YP_004274942.1)</td>
<td>98%</td>
<td>72%</td>
<td>6.74</td>
<td>48</td>
</tr>
<tr>
<td>M8244</td>
<td>GH43</td>
<td>1,605</td>
<td>531</td>
<td>Hypothetical protein [Sphingobacterium sp.] (WP_021189555.1)</td>
<td>55%</td>
<td>99%</td>
<td>5.17</td>
<td>61.1</td>
</tr>
</tbody>
</table>

* query coverage; ** isoelectric point.
Bioinformatics analysis, phylogenetic tree and structural modeling of protein XylM1989

For one protein that was successfully produced into the soluble fraction (gene M1989), the translated gene (xylM1989) was analyzed 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 thirty-five proteins from different origin that belong to CAZy families GH43, GH3 [16] and AA10 (outgroup sequences). For the multiple protein alignment and phylogenetic analyses, the software’s MEGA v6.0 and PRALINE [17,18] were used. In order to detect the catalytic and substrate-binding sites, a XylM1989 protein data bank file was generated using Phyre2 [19]. With this prediction, the closest homolog 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 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 to 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 280µl of 0.5 mmol l⁻¹ of pNP-Xyl or pNP-Ara and 20µl 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 15min. 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 pNP-Xyl (pH 8.0) at 30°C for 30min. 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 para-nitrophenol (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 (500µl) contained 1% of each polysaccharide (diluted in 20
mmol $l^{-1}$ of Tris-HCl pH 8.0) and 150$\mu$L of soluble protein XylM1989 (approximately 1.7 mg ml$^{-1}$). The mixtures were incubated at 30°C for 72h, subsequently the reactions were stopped on ice and centrifuged for 10min at 12000g. The enzymatic activity was determined by measuring the amount of reducing sugars in the supernatant by the 3,5-dinitrosalicilic acid (DNS) method [20]. 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 analyzed 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 ten genes predicted to encode GHs**

From two wheat-straw-degrading microbial consortia, we here selected ten genes encoding predicted GHs for synthesis. All ten genes were codon-optimized for *E. coli*, synthetized and cloned into the pET21b+ expression vector with specific restriction enzyme sites on both sides (NdeI and XhoI). For all genes, the fragment sizes - as measured on agarose gels - were consistent with the predicted sizes of the sequences (Table 1; Figure 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 ten cloned genes, only one protein (gene M1989) occurred in the soluble fraction (~80-90% pure), whereas the remainder was mainly present in inclusion bodies (Figure 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 [13]. 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 six in ten genes produced in *E. coli* had xylanase activity, of which only one, M1989, appeared in the soluble fraction (Figure 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, i.e. 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 (Figure 1c). Indeed, the product of clone M1989 showed dual activity, i.e. with $\beta$NP-Xyl and $\beta$NP-Ara, but not with $\beta$NP-Man and $\beta$NP-Fuc. On the basis of its activity, protein M1989 will be denoted XylM1989. It is the basis of the further results described below.

![Figure 1](image)

**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 $\beta$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.

### 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 $\beta$-xylosidases (EC 3.2.1.37), $\alpha$-arabinosidases (EC 3.2.1.55), galactan 1,3-$\beta$-galactosidases (EC 3.2.1.45) and endo-$\alpha$-arabinases (EC 3.2.1.99) [16,21]. 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 [22]. In addition, protein XylM1989 showed 63% identity with a characterized bifunctional GH43 family xylosidase/arabinosidase (xynB; CAA89208) from *Prevotella bryantii* [23]. 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*, *Dracnobacterium*, *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 (Figure 2). Moreover, it showed 71% identity protein with protein 4MLG (structure of RS223-β-xylosidase) [24]. Based on the predicted 3D structure using protein 4MGL 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) (Figure 3).

![Figure 2. Maximum likelihood phylogenetic analysis of XylM1989 with closest related proteins.](image)

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 monoxygenases (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.
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) (Figure 4a). The activity of XylM1989 decreased, to approximately 20% of the maximal activity, when the temperature was raised from 40 °C to 70 °C for both pNP-Xyl and pNP-Ara. As shown in Figure 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.

Kinetic analysis of the XylM1989 protein - The XylM1989 hydrolytic activity was measured with respect to catalytic properties and by assessment of kinetic parameters, i.e. the \( K_m \) (Michaelis-Menten constant) and \( V_{max} \) 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 mmol l⁻¹ and 0.781 mmol l⁻¹, and the \( V_{max} \) values 285.71 U mg⁻¹ and 78.12 U mg⁻¹, respectively. Given the fact that the \( K_m \) and \( V_{max} \) values of protein XylM1989 for pNP-Xyl were higher than that 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 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 CaCl₂ (10-fold), MgCl₂ (12-fold) and MnCl₂ (7-fold). 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, i.e. 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, i.e. 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) (Figure 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%.

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 uncultured organism (Protein data bank ID: 4MLG).

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 temperature with constant pH 7.0. b) Different pH values at 30°C.
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 Ca²⁺, Mg²⁺ and Mn²⁺) 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 5-fold increased over that of the control (Figure 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 semi-natural 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 glucose levels exceeded that 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, ten GH-encoding genes, retrieved from two wheat straw-degrading microbial consortia [12], were selected for codon optimization, synthesis, cloning, expression and characterization. These genes all originated from different Bacteroidetes, a dominant phylum in wood-degrading communities [25] and were found in different microbial consortia cultivated on agricultural residues [26,27]. 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 [12]. Inspired by recent data [28], we intended to increase the expression levels of such GH-encoding genes (in *E. coli*), using codon swapping by

**Table 2. Effects of different additives on the XylM1989 β-xylanase activity**

<table>
<thead>
<tr>
<th>Type of additive</th>
<th>Reagent</th>
<th>% Relative activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Ions</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CaCl(_2)</td>
<td>1065.23 ± 74</td>
<td>546.58 ± 39</td>
</tr>
<tr>
<td>MgCl(_2)</td>
<td>1214.18 ± 16</td>
<td>691.11 ± 120</td>
</tr>
<tr>
<td>CoCl(_2)</td>
<td>6.27 ± 1.1</td>
<td>1.33 ± 0.3</td>
</tr>
<tr>
<td>NiCl(_2)</td>
<td>8.67 ± 1.5</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>MnCl(_2)</td>
<td>738.93 ± 8.4</td>
<td>172.88 ± 20</td>
</tr>
<tr>
<td>NH(_4)</td>
<td>69.32 ± 1.0</td>
<td>63.61 ± 4.3</td>
</tr>
<tr>
<td><strong>Sugars</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Glucose</td>
<td>69.30 ± 1.3</td>
<td>99.28 ± 4.9</td>
</tr>
<tr>
<td>Xylose</td>
<td>55.82 ± 2.5</td>
<td>53.71 ± 1.4</td>
</tr>
<tr>
<td>Arabinose</td>
<td>164.89 ± 5.4</td>
<td>249.26 ± 3.8</td>
</tr>
<tr>
<td>Cellobiose</td>
<td>82.32 ± 2.0</td>
<td>90.32 ± 14</td>
</tr>
<tr>
<td>Galactose</td>
<td>80.38 ± 1.8</td>
<td>90.99 ± 12</td>
</tr>
<tr>
<td><strong>Salt</strong></td>
<td>200 mmol l(^{-1})</td>
<td>2000 mmol l(^{-1})</td>
</tr>
<tr>
<td>NaCl</td>
<td>96.06 ± 6.1</td>
<td>36.34 ± 0.6</td>
</tr>
<tr>
<td>Chelating agent</td>
<td>20 mmol l(^{-1})</td>
<td>50 mmol l(^{-1})</td>
</tr>
<tr>
<td>EDTA</td>
<td>61.66 ± 0.8</td>
<td>30.20 ± 0.6</td>
</tr>
<tr>
<td><strong>Detergents</strong></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><strong>Organic solvents</strong></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>
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 [29]. In our study, nine of the ten 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 [30,31].

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 (ephemeral) 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 [32,33]. Notably, by using zymography we showed, for the first time, that proteins (M1916 and M8239) annotated as fucosidases have xylanolytic activity (Figure 1b; Table 1). Remarkably, in the presence of the pNP 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 (Figure 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 [22,34]. 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 are involved in xylan degradation, sugar transport and xylose metabolism (Figure 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) [35–42]. 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 hydrolysates has not been adequately addressed [43]. It is worth noting that these GHs constitute key components of enzyme cocktails used for the improved saccharification of lignocellulosic biomass.
Table 3. Comparison of the protein XylM1989 with GH43 family enzymes from other studies

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Microbial Source</th>
<th>kDa</th>
<th>Substrate</th>
<th>pH</th>
<th>Optimal °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 sp.</em></td>
<td>37.5</td>
<td>pNP-X/pNP-A</td>
<td>8.0</td>
<td>20</td>
<td>9.65/1.22</td>
<td>1.2/0.781</td>
<td>285.71/78.12</td>
<td>This work</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>225</td>
<td>*ND</td>
<td>*ND</td>
<td>Ito et al. 2003 [57]</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 Sa-Nogueira 2008 [38]</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 [42]</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 [58]</td>
</tr>
<tr>
<td>PtXyl43</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 [59]</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 [60]</td>
</tr>
<tr>
<td>RuXyn1</td>
<td><em>Prevotella bryantii</em></td>
<td>45.0</td>
<td>pNP-X/pNP-A</td>
<td>7.0</td>
<td>40</td>
<td>36.3/14.2</td>
<td>3.43/2.23</td>
<td>*ND</td>
<td>Zhou et al. 2012 [46]</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 [61]</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>

*ND: not determined.
Targeted synthetic metagenomic arification of plant biomass and production of biofuels [44]. The aforementioned studies showed that some, but not all, GH43 family enzymes have bifunctional activities, in particular \( \beta \)-xylosidase and \( \alpha \)-arabinosidase. For instance, protein XylC, isolated from *Paenibacillus woosongensis*, had dual activity [39], and some proteins encoded by genes isolated from rumen and compost-derived metagenomes also had dual activity [9,45,46]. In contrast, the enzyme His-Xyl43 showed \( \beta \)-xylosidase activity [36], whereas proteins Abn2 and AbnZ2 had only endo-arabinanase activity [38,41]. We speculated, on the basis of the foregoing, that the \( \beta \)-D-xylosidase/\( \alpha \)-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 \( V_{\text{max}} \) (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, e.g. by binding and stabilizing the substrate at the active site, may be invoked [35,47,48]. In contrast, Cu\(^{2+}\), Ni\(^{2+}\), Co\(^{2+}\) and NH\(^{4+}\) strongly inhibited the \( \beta \)-xylosidase activity of XylM1989. Similar to XylM1989, a xylanase produced by a gene isolated from a bovine rumen metagenome showed enhanced \( \beta \)-xylosidase activity with Mn\(^{2+}\), whereas Cu\(^{2+}\), Fe\(^{2+}\), Ag\(^{2+}\) and Zn\(^{2+}\) ions inhibited the activity [49]. 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 [50,51]. These aromatic compounds released in hydrolysates are considered to be major inhibitors of fermentation processes [43]. During plant biomass pretreatment, several released products (including furanic compounds and monosaccharides) can inhibit the activity of the enzymes (or cocktails) used for the subsequently saccharification process. In addition, several studies has been showed a strong inhibition of ethanol production due to the presence of lignocellulosic byproducts [43]. For example, concentration of \( \sim 0.5\% \) (w/v) and \( \sim 0.7\% \) (w/v) of furfural and 5-HMF, respectively, can inhibit the growth rate and production of ethanol by *Issatbenkia orientalis* [52]. 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
1 \textsuperscript{1} (0.022\% w/v) and 5-HMF concentrations up to 0.17 g l\textsuperscript{-1} (0.017\% w/v) \[53\]. Interestingly, protein XylM1989 was tolerant to furfural, as its $\beta$-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 $\beta$-xylosidase/\alpha-arabinosidase that is tolerant to furfural and 5-HMF. Moreover, as was explained before, $\beta$-xylosidases are key in the conversion of xylo-oligosaccharides to xylose as the end-product. However, xylose is one of the major inhibitors of $\beta$-xylosidase activity \[54\]. Thus, xylose-tolerant $\beta$-xylosidases are important in hemicellulose conversion. The sensitivity of most $\beta$-xylosidases to xylose (as tested with fungal-produced xylosidases, such as those from \textit{Arxula adeninivoran}, \textit{Aureobasidium pullulans} and \textit{Trichoderma reesei}) is striking, with $K_i$ (concentration of inhibitor) values for xylose ranging from 2-10 mmol l\textsuperscript{-1} \[55\]. In contrast, \textit{Thermotoga thermarum} Tth xynB3 $\beta$-xylosidases showed high xylose-tolerant activity at 500 mmol l\textsuperscript{-1} \[56\]. Interestingly, our novel xylM1989 protein showed 100\% activity at xylose concentrations of 500 mmol l\textsuperscript{-1}, in the presence of ions (either Ca\textsuperscript{2+}, Mg\textsuperscript{2+} or Mn\textsuperscript{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\textsuperscript{2+}, Mg\textsuperscript{2+} or Mn\textsuperscript{2+}.

Finally, five key features makes that the protein XylM1989 is a good candidate for use in industrial process related with plant biomass saccharification: i) active at alkaline pH, ii) higher reaction speed ($V_{\text{max}}$), iii) tolerance to lignocellulosic hydrolysates-derived inhibitors, iv) tolerance to high concentration of xylose in presence of Ca\textsuperscript{2+}, Mg\textsuperscript{2+} and Mn\textsuperscript{2+} and v) activity and release of sugars from complex polysaccharides such as a XB, OX and ARB. The 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.

\section*{Acknowledgements}
This work was supported by the BE-Basic foundation (\url{http://www.be-basic.org}). We thank H. Ruijsseenaars and R. van Kranenburg for scientific support.

\section*{References}


Supplementary figures

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

Supplementary Figure S3. Genomic context of the XyIM1989 in comparison with a genomic fragment from Sphingobacterium TN19. ABCT means ABC transporters.
Supplementary Figure S1. Genomic context of the selected GHs-encoding genes that were subsequently used for codon optimization, synthesis, cloning and expression. Abbreviations: Two-component system proteins (TCSPs), ABC transporters (ABCT), TolC-dependent receptors (TBR), transketolases (TKT), xylanase isomerase (XI) and xylulose kinase (XKN) genes.