1

2

3 **High-throughput generation of product profiles for**

4 **arabinoxylan-active enzymes from metagenomes**

5 Metagenome-derived AX-active enzymes product profiles

6

7

8 **Maria João Maurício da Fonseca[a], Zachary Armstrong[b,*], Stephen G. Withers[b],**

9 **Yves Briers[a,#]**

10 [a] Department of Biotechnology, Ghent University, Ghent, Belgium.

11 [b] Department of Chemistry, University of British Columbia, Vancouver, Canada.

12

13

14

15

16 [#] Corresponding author: yves.briers@ugent.be

17 *Current affiliation: Department of Chemistry, The University of York, York, United Kingdom.

18

19

20

22

1

Abstract

Metagenomics is an exciting alternative to seek for carbohydrate-active enzymes from a range of sources. Typically, metagenomics reveals dozens of putative catalysts that require functional characterization for further application in industrial processes. High-throughput screening methods compatible with adequate natural substrates are crucial for an accurate functional elucidation of substrate preferences. Based on DNA sequencer-aided fluorophore-assisted carbohydrate electrophoresis (DSA-FACE) analysis of enzymatic reaction products, we generated product profiles to consequently infer substrate cleavage positions, resulting in the generation of enzymatic degradation maps. Product profiles were produced in high-throughput for arabinoxylan (AX)-active enzymes belonging to the glycoside hydrolase families GH43 (subfamilies 2 ($MG43_2$), 7 ($MG43_7$) and 28 ($MG43_{28}$)) and GH8 (MG8) starting from twelve (arabino)xylo-oligosaccharides. These enzymes were discovered through functional metagenomic studies of faeces from the North American beaver (*Castor canadensis*). This work shows how enzyme loading alters the product profiles produced by all enzymes studied and gives insight into AX degradation patterns revealing sequential substrate preferences of AX-active enzymes.

Importance

Arabinoxylan is mainly found in the hemicellulosic fractions of rice straw, corn cobs and rice husk. Converting arabinoxylan into (arabino)xylo-oligosaccharides as added value products that can be applied in food, feed, and cosmetics presents a sustainable and economic alternative for the biorefinery industries. An efficient and profitable AX degradation requires a set of enzymes with particular characteristics. Therefore, enzyme discovery and study of substrate preferences is of utmost importance. Beavers, as consumers of woody biomass are a promising source of a repertoire of enzymes able to deconstruct hemicelluloses into soluble oligosaccharides. High-

2

47 throughput analysis of oligosaccharide profiles produced by these enzymes will assist in the

48 selection of the most appropriate enzymes for the biorefinery.

Introduction

49

50 Metagenomic studies of the microbial communities associated with plant cell wall degraders

51 reveal a large number of gene sequences coding for potential carbohydrate-active enzymes

52 (CAZymes) (1). Accurate functional analysis of new CAZymes must accompany this continuous

53 discovery at the genomic level. Such newly functionally validated enzymes can be applied in the

54 biorefinery industry and/or for further protein engineering (2-5). Glycoside hydrolases (GHs) that

55 cleave polysaccharide main chains and/or substituents can have complex substrate preferences,

56 often showing multi-substrate specificities (6-9). These substrate preferences are often also

57 dependent on main chain substituents. Consequently, the degradation of complex carbohydrate

58 polymers often requires synergistic or combinatorial action of multiple enzymes. Tedious

59 techniques, long analysis times, demanding hands-on assays, specialised equipment and lack of

60 appropriate representative substrates contribute to the existing gap between enzyme discovery

61 and functional characterization (10). Accordingly, high-throughput (HT) techniques that can deal

62 with a large number of metagenome-derived putative enzymes and that can give an insight into

63 the substrate specificities of unannotated enzymes in a relatively short time are required (11).

64 DNA sequencer-aided fluorophore-assisted carbohydrate electrophoresis (DSA-FACE) offers an

65 interesting approach to study the substrate specificities of CAZymes, primarily due to the

66 possibility to use substrates that represent the natural carbohydrates instead of artificial aryl

67 glycoside substrates like *p*-nitrophenyl- and 4-methylumbelliferyl-derivatives (12). In fact, the

68 latter substrate derivatives may mask the real enzymatic substrate specificity due to, for example,

69 steric differences in comparison to the natural oligosaccharides. DSA-FACE has shown

70 outstanding oligosaccharide resolution and sensitivity (a detection limit ranging from 38 to 55

71 pM for the substrates studied), short hands-on time and analysis time (13). In addition, DSA-

4

72   FACE allows analysis of the substrate specificities in HT when using multiple parallel capillaries

73   as in standard capillary sequencing devices.

74   In this work we focus on arabinoxylan (AX) which is a hemicellulosic polysaccharide that may

75   contain a range of substitutions including α-L-arabinofuranosyl, α-D-glucuronic acid, 4-O-

76   methyl-α-D-glucuronic acid, α-D-galactopyranose and ferulic acid residues, depending on the

77   source (14). Deconstruction of AX by GHs leads to useful sugars for the production of

78   bioethanol, food and nutraceutical added-value products such as xylitol (15-18) and prebiotics

79   (19-23). GHs with diverse substrate specificities for the degradation of complex AX structures

80   are being discovered continuously and annotated in the various protein databases (24). The

81   carbohydrate-active enzymes database (CAZy) classifies GHs into families according to amino

82   acid sequence similarity (25). Although structural similarity often correlates with enzyme

83   substrate specificity, the CAZy family division cannot always be used to predict enzyme

84   substrate specificity because of the different specificities assigned per GH family. Additionally,

85   up to now only approximately 1% of the annotated GHs in CAZy have been experimentally

86   characterized (25). Due to the abundance of GH43 members and the large substrate specificity

87   variety found within this family, the GH43 family was further divided into subfamilies on the

88   basis of sequence analyses, suggesting that the correlation between functional annotation of the

89   enzyme and subfamily assignment is more accurate (26). The GH43 family subdivision and

90   analysis relies on computational and experimental data, which are mainly based on synthetic

91   substrates such as *p*-nitrophenyl (*p*NP) monosaccharides. To obtain a more accurate

92   understanding of the substrate specificity of AX-active enzymes, we have recently used DSA-

93   FACE to analyze the hydrolysates of AX-active enzymes with natural representative

94   (arabino)xylo-oligosaccharides ((A)XOS) (13). In this work, DSA-FACE is used to elucidate

95   substrate specificities of enzymes derived from metagenomic studies on the North American

5

96 beaver (*Castor canadensis*) fecal microbiome (27). Three genes belonging to uncharacterized or

97 poorly characterized subfamilies 2, 7 and 28 of GH43 and one from GH8 that were identified in

98 active fosmids were shown to be active on AX by preliminary functional screening tests with aryl

99 glycosides and/or high-performance anion-exchange chromatography with pulsed amperometric

100 detection (HPAEC-PAD) (27) (Figure 1). To gain further insight into AX degradation patterns by

101 the aforementioned enzymes, we introduce DSA-FACE product profiles and associated

102 sequential degradation maps for a convenient representation of the activity(ies) of each enzyme

103 on twelve (A)XOS substrates. By implementing DSA-FACE product profiles, we reveal the

104 preferred substrate cleavage sites by a $GH43_{28}$ member, the modular specificity of a $GH43_2$ –

105 GH8 enzyme and the dependence of the enzymatic activity of a $GH43_7$ member on the activity of

106 the aforementioned $GH43_2$ – GH8 enzyme.

107

108 **Results**

109 *DSA-FACE allows rapid evaluation of product profiles produced by metagenome-derived AX-*

110 *active enzymes*

111 In this study, we have used DSA-FACE to set up product profiles for three newly-discovered

112 enzymes of the GH43 family active on AX. These enzymes were identified through a preceding

113 metagenomic analysis of the North American beaver (*Castor canadensis*) fecal microbiome (27).

114 The first selected enzyme ($MG43_{28}$) contains a GH43 subfamily 28 domain. The second enzyme

115 ($MG43_2$-8) is a modular enzyme composed of a GH43 subfamily 2 and a GH8 domain. To

116 differentiate the specificities of each domain, we have also set up product profiles produced by

117 two mutated variants, in which the respective domains are inactivated by mutagenesis of the

118 catalytic acid residue (27), resulting in $MG43_2$ and MG8, respectively (Table 1). The third

119 enzyme $MG43_7$ comprises a single domain assigned to GH43 subfamily 7. These enzymes were

6

120 selected because of the limited characterization for these subfamilies. An initial analysis of these

121 enzymes with chromogenic substrates and HPAEC-PAD confirmed that they have AX-acting

122 activities as summarized in Table 1. The initial low expression yields of MG43$_2$-8, MG43$_2$ and

123 MG43$_7$ were optimized by variation of expression strains, growth temperature, induction and

124 purification protocols (Table 2). Eluted fractions from the different expressions/purifications

125 containing the desired protein were pooled for further analysis.

126 DSA-FACE product profiles are a qualitative representation of the carbohydrates present after

127 enzymatic reactions with different (A)XOS (Figure 2A). Twelve oligosaccharides (5 XOS, 7

128 AXOS) are used in these enzymatic reactions. Reaction hydrolysates are then analyzed with

129 DSA-FACE. The output of DSA-FACE are electropherograms that need to be interpreted by

130 referencing to (A)XOS standards to reveal the identity of the resulting products. Peak areas are

131 quantified and normalized to calculate the relative conversion of each substrate and relative

132 proportion of the products. This has been exemplified in Figure 2B for a previously characterized

133 α-L-arabinofuranosidase from *Bifidobacterium adolescentis* (*Ba*AXH-d3) (13). Alpha-L-

134 arabinofuranosidases are classified into arabinoxylan arabinofuranohydrolases (AXH) that

135 hydrolyze the *O*-2 and/or *O*-3 arabinofuranosyl monomers from the doubly substituted xyloses

136 (AXH-d2, AXH-d3, AXH-d2,3) or from the mono substituted xyloses (AXH-m2, AXH-m3,

137 AXH-m2,3). To simplify the representation of enzymatic substrate preferences, we introduce

138 here product profiles with a fixed color code instead of electropherograms as the final DSA-

139 FACE outcome. Substrate conversions are easily observed in the product profiles by a color

140 change. Accordingly, there is only a color change in the case of A^{2+3}XX and XA^{2+3}XX for

141 *Ba*AXH-d3. Next to the product profiles, degradation maps highlighting cleavage positions for

142 the different (A)XOS tested are elaborated to assist in evaluating enzyme substrate preferences.

143

7

144 *Different product profiles for increasing MG43$_{28}$ concentrations*

145 Purified MG43$_{28}$ was tested against the panel of five XOS and seven AXOS using four different

146 enzyme concentrations (0.3 – 1 – 6 – 32 μM). The product profiles of MG43$_{28}$ show diverse

147 hydrolytic products which change depending on the enzyme concentration (Figure 3). At the

148 lowest concentration tested (0.3 μM) MG43$_{28}$ completely removes an *O*-2 arabinose from a non-

149 reducing end singly substituted xylose, i.e. A^{2}XX to X$_3$ (AXH-m2 activity). In addition, at

150 concentrations at or above 0.3 μM MG43$_{28}$ both *O*-2 and *O*-3 arabinoses are partially removed

151 from the non-reducing end doubly substituted xylose of A^{2+3}XX, resulting in X$_3$ (AXH-d2,3

152 activity), but higher concentrations (6 μM) are needed for a full conversion. The internal *O*-2

153 arabinose from XA^{2}XX is only removed from 1 μM and even at the highest concentration studied

154 (32 μM) MG43$_{28}$ is not able to remove *O*-2 and *O*-3 arabinoses from XA^{2+3}XX, indicating that

155 the internal single/double arabinose substitutions are less accessible for hydrolysis. Removal of

156 the *O*-3 arabinoses from A^{3}XX and XA^{3}XX (AXH-m3 activity), resulting in X$_3$ and X$_4$,

157 respectively, is observed from 1 μM MG43$_{28}$. MG43$_{28}$ thus exerts diverse arabinofuranosidase

158 specificities with the best conversion of A^{2}XX. Xylanolytic activity is visible at concentrations at

159 or above 1 μM as well, with a preference for the longest xylo-oligosaccharide tested (X$_6$). Also,

160 AXOS with removed arabinoses are further partially degraded (e.g. XA^{3}XX is converted to X$_2$

161 and X$_4$).

162

163 *The product profiles for MG43$_2$-8 are the sum of the product profiles of its respective domains*

164 Enzymatic reactions with purified MG43$_2$, MG8 and MG43$_2$-8 were performed. The product

165 profiles in Figure 4 demonstrate that MG43$_2$ is a β-xylosidase and MG8 is a reducing end xylose-

166 releasing exo-oligoxylanase (Rex). At a concentration of 3 μM, MG43$_2$ partially converts XOS

167 into X$_2$. It seems there are more structural hindrances or product inhibition in comparison to MG8

8

168   based on the slower observed degradation rate. MG43$_2$ does hydrolyze the non-reducing xylose

169   monomers from XA^{2}XX, XA^{3}XX and XA^{2+3}XX (Figure 4A) but not the non-reducing arabinose

170   substituted xylose monomers (A^{2}XX, A^{3}XX and A^{2+3}XX). At a concentration of 3 μM MG8 fully

171   converts X$_3$, X$_4$ and X$_5$ to X$_2$, but we did not observe visible conversion of X$_2$ to X. However, X$_6$

172   seems to be more slowly converted to X$_2$ due to possible hindrances in accommodating long XOS

173   into MG8 catalytic subsites or due to product inhibition, indicating a preference for smaller XOS.

174   As typical for Rex enzymes, MG8 requires two non-substituted xyloses from the reducing end to

175   hydrolyse the reducing end xylose (28, 29). This is observed when the reducing end xylose

176   monomer is hydrolysed from A^{2}XX, A^{3}XX, A^{2+3}XX, XA^{2}XX, XA^{3}XX and XA^{2+3}XX but not

177   from A^{3}X. We also evaluated different concentrations for MG8 (0.2, 0.8, 3 and 17 μM) (Figure

178   S1). Notably, at the minimum MG8 concentration tested (0.2 μM) only X$_3$ to X$_6$ were partially

179   hydrolysed into smaller dp XOS, showing the preference for XOS over AXOS by MG8. At

180   concentrations from 0.8 to 17 μM, MG8 shows the same product profiles as the ones performed

181   with 3 μM enzyme (Figure 4).

182   At concentrations of 3 μM MG43$_2$-8 displays the sum of both β-xylosidase and Rex activities

183   (Figure 4). When we evaluated MG43$_2$-8 at a higher concentration (17 μM), MG43$_2$-8 also

184   exhibited α-L-arabinofuranosidase activity when hydrolyzing the mixture of XA^{2}XX and XA^{3}XX

185   into A^{2}X, A^{3}X and X$_2$ (Figure S2). This is likely performed by the MG43$_2$ domain since 17 μM

186   MG8 did not show α-L-arabinofuranosidase activity on any of the substrates tested and 3 μM

187   MG43$_2$ showed a small amount of α-L-arabinofuranosidase activity on A^{3}X but not on A^{3}XX or

188   XA^{3}XX (Figure 4).

189

190   *Product profiles of MG43$_7$ change with increasing MG43$_7$ concentration and in the presence of*

191   *MG43$_2$-8/MG43$_2$/MG8*

9

192   MG43$_7$ activity was only detected at the highest concentration tested (38 µM). MG43$_7$ showed a

193   xylanase activity on X$_4$, X$_5$ and X$_6$ (Figure S3). Yet, when enzymatic reactions were performed

194   with 8 µM MG43$_7$ in the presence of 3 µM MG43$_2$-8, A^{3}X is converted to X$_2$, showing an

195   additional AXH-m3 activity, which does not happen when enzymatic reactions are performed

196   with the same concentrations of MG43$_2$-8 or MG43$_7$ alone (Figure 5). X$_2$ and X$_3$ can also be

197   observed after enzymatic reactions with 8 µM MG43$_7$ in the presence of 3 µM MG43$_2$-8 and

198   A^{3}XX or XA^{3}XX, again showing additional AXH-m3 activity. Notably, O-2 arabinofuranosyl

199   substitutions are not a substrate since A^{2}XX is not further hydrolyzed to X$_2$. When enzymatic

200   reactions were performed with elevated concentrations (38 µM MG43$_7$ in the presence of 17 µM

201   MG43$_2$-8) and A^{3}X, the end product was again X$_2$, showing no further hydrolysis, but when the

202   same concentrations were tested against A^{2}XX, A^{2}X and X$_2$ appeared as reaction products

203   (Figure S2).

204   To discriminate whether the additional AXH-m3 or AXH-m2 activity that appears when

205   combining MG43$_7$ and MG43$_2$-8, comes from either MG43$_7$ or MG43$_2$-8, we investigated again

206   the activity of MG43$_7$ in the presence of the derivatives of MG43$_2$-8 in which one domain was

207   inactivated by mutation (MG43$_2$ and MG8). When A^{3}X reacted with 8 µM MG43$_7$ and 3 µM

208   MG43$_2$ (Figure S4), approximately 50% A^{3}X was converted to X$_2$. The same reaction but in

209   combination with 3 µM MG8 also resulted in a minor fraction of X$_2$ (Figure S1). These data

210   indicate that the AXH-m3 activity does indeed result from MG43$_7$. This finds further support as

211   also AXH-m3 activity is detected in both cases when a mixture of A^{2}XX/A^{3}XX reacts with

212   MG43$_7$ in the presence of either MG43$_2$ (Figure S4) or MG8 (Figure S1). Likely MG43$_7$ has a

213   preference for non-reducing O-3 substituted xyloses since 8 µM MG43$_7$ and 3 µM MG8 do not

214   hydrolyze XA^{3}XX further into X$_2$. These findings are consistent with the previously identified

215   arabinofuranosidase activity of MG43$_7$ on A^{3}X detected by HPAEC-PAD (27). In sum, the scans

10

216 indicate that MG43$_7$ shows AXH-m3 activity on small O-3 arabinose substituted AXOS (A^{3}X

217 and A^{3}XX) in the presence of MG43$_2$-8, and xylanase activity at elevated concentrations.

218

219 **Discussion**

220 Functional metagenomic studies of the beaver fecal microbiome revealed enzymes from

221 subfamilies 2, 7 and 28 of the GH43 CAZy family. Whereas subfamily 2 has two characterized α-

222 L-arabinofuranosidases from *Chitinophaga pinensis* DSM2588 and *Mucilaginibacter mallensis*

223 MP1X4 (3, 30), subfamilies 7 and 28 have no characterized enzymes to date. Activity on CMU-X

224 indicated that MG43$_{28}$ is a β-xylosidase (27). However, the MG43$_{28}$ product profiles reveal that

225 MG43$_{28}$ is able to hydrolyze all (A)XOS substrates except XA^{2+3}XX, showing xylanase, AXH-

226 m2,3 and AXH-d2,3 activities. We cannot confirm if the observed xylanase activity is due to

227 endo- or sequential exo-xylanase activity. If MG43$_{28}$ acts as a β-xylosidase it would be expected

228 that X$_2$ is also degraded to X, which does not seem to happen even at the highest concentration

229 tested. Though, monomeric xylose cannot be detected by DSA-FACE. Similarly, GH43

230 *Pc*Axy43A from *Paenibacillus curdlanolyticus* (GH43 subfamily 35) is also unable to hydrolyze

231 XA^{2+3}XX and shows endo-xylanase, β-xylosidase, AXH-d2,3 and AXH-m2,3 activities in a

232 single catalytic domain (31), but the presence of both exo- and endo-activity in a single enzyme

233 can be considered unusual. Ara 1 isolated from barley malt also has both AXH-m2,3 and AXH-

234 d2,3 activities and a four times higher enzyme concentration is needed for conversion of

235 XA^{2+3}XX into X$_4$ than for conversion of A^{2+3}XX (32). Besides a N-terminal MG43$_{28}$ GH43

236 domain, a C-terminal discoidin domain has been identified using the Conserved Domain

237 database. This discoidin domain has putative lectin-like properties, binding carbohydrates (33).

238 To unravel how MG43$_{28}$ deals with such a variety of substrates, the influence of this C-terminal

239 domain on the observed multiple activities may be investigated. A blastp analysis of MG43$_{28}$

11

240   against all GH43 CAZy family characterized sequences (182 characterized sequences out of

241   16250) revealed an enzyme from *Belliella baltica* DSM 15883 (accession number: AFL85801.1)

242   as the most similar sequence (E-value 5×10^{-10} with 45% query cover and 25% percentage

243   identity). This enzyme is annotated in GH43 subfamily 31 of the CAZy database and was

244   identified as a β-D-galactofuranosidase in the study of (3). The endo-1,4-β-xylanase from an

245   uncultured bacterium URE4 (accession number: ACM91046.1) shows the maximum query cover

246   of 85% (E-value: 1×10^{-4}, percentage identity: 23%). This enzyme is annotated in subfamily 29 of

247   the CAZy database. These relatively low similarities spread over different subfamilies emphasize

248   the need for detailed analyses of the substrate specificity as done here for $MG43_{28}$.

249   The substrate preferences of $MG43_2$-8, $MG43_2$, MG8, and $MG43_7$ in the presence of $MG43_2$-8

250   were previously analyzed by HPAEC-PAD upon enzymatic reactions with 0.5 µM purified

251   enzyme and 4 mM A^3X, A^2XX and a mixture of XA^2XX and XA^3XX (27). At the conditions

252   tested $MG43_2$, MG8, and $MG43_7$ showed β-xylosidase, Rex and AXH-m3 activities, respectively.

253   Due to the limited HT capacity of HPAEC-PAD, a restricted number of substrates were tested,

254   omitting doubly substituted XOS, for example. The minor α-L-arabinofuranosidase activity of the

255   $MG43_2$ β-xylosidase against A^3X indicates $MG43_2$ shows both β-xylosidase and AXH-m3

256   activities at concentrations higher than 3 µM. Bifunctional β-xylosidase and α-L-

257   arabinofuranosidase activities have already been reported before in the GH43 CAZy family but

258   not yet in subfamily 2. It seems these enzymes can accommodate both xylose and arabinose units

259   in their active sites not only due to obvious structural similarities between arabinose and xylose

260   sugars, but also due to rotations on the α-arabinose linkage to xylose that can resemble a β-xylose

261   linkage in the main chain (34, 35). Accordingly, it can be questioned whether $MG43_2$ is actually

262   bifunctional or misrecognizes the substrate, which can be observed at an elevated enzyme

263   concentration.

12

264 Xylanolytic activity of MG43$_2$-8 and MG43$_2$ against X$_2$ remains ambiguous. DSA-FACE is not

265 able to detect xylose and can thus not detect possible X$_2$ degradation. Previous HPAEC-PAD

266 analyses show that X$_2$ was not hydrolyzed by MG43$_2$-8 and MG43$_2$ (27). In accordance with

267 these results, MG43$_2$-8 and MG43$_2$ were also not active against chromogenic pNP-X. However,

268 MG43$_2$-8 and MG43$_2$ showed activity with fluorogenic CMU-X, suggesting X$_2$ xylanolytic

269 activity. This discrepancy may be explained by the higher sensitivity when using fluorogenic

270 substrates. In sum, if MG43$_2$-8 and MG43$_2$ can hydrolyze X$_2$, it will be at maximum with a low

271 activity. This contrasts to many GH43 β-xylosidases that digest X$_2$ (36). Yet, β-xylosidases such

272 as XylB from *Bifidobacterium adolescentis* that prefer longer dp XOS over X$_2$ have also been

273 reported (37).

274 DSA-FACE demonstrated a strict Rex substrate specificity for MG8 and showed complete

275 substrate conversion at the maximum concentration tested. Up to now there are only four GH8

276 Rex enzymes characterized in the CAZy database, including enzymes from *Bacillus halodurans*

277 (38), *Bifidobacterium adolescentis* (39), *Bacteroides intestinalis* (40) and *Paenibacillus*

278 *barcinonensis* (29). MG8 shows a typical Rex activity (as the characterized Rex enzymes listed

279 above): MG8 does not hydrolyze pNP-X, is active on XOS with dp 3 to 6 and has a preference

280 for short dp XOS (41). Similar to Rex8A from *Paenibacillus barcinonensis*, which was the first

281 one tested against branched oligosaccharides (MeGlcA decorated xylooligomers), MG8 is able to

282 hydrolyze the reducing end xylose of branched AX-oligosaccharides. Notably, the *rex8A* gene

283 from *Bacteroides intestinalis* is located downstream a *xyl3A* gene, which encodes a β-xylosidase.

284 The X$_2$ generated by Rex8A is therefore hydrolyzed by the Xyl3A β-xylosidase. This is not the

285 case for MG43$_2$-8 as both MG43$_2$ and MG8 cannot efficiently hydrolyze X$_2$ as shown here and

286 before with HPAEC-PAD. β-Xylosidases such as MG43$_2$, and Rex such as MG8 which have low

287 or no X$_2$ hydrolytic activity are interesting for the incomplete degradation of AX into X$_2$. X$_2$ has

13

288 been demonstrated to be the most efficient prebiotic in terms of promoting a higher growth of

289 *Bifidobacterium* and *Lactobacillus* strains among the xylose polymers and to present an increased

290 sweetness power in comparison to sucrose (42).

291 MG43$_7$ shows a unique substrate specificity pattern, requiring the presence of MG43$_2$-

292 8/MG43$_2$/MG8 for activity. In fact, the MG43$_2$-8 and MG43$_7$ coding sequences were identified in

293 the same operon, already suggesting a natural synergy between these two enzymes. Notably,

294 MG43$_7$ only handles *O*-3 arabinofuranosyl substitutions of rather small AXOS. Previously a

295 GH43$_{18}$ metagenome-derived enzyme also showed a single preference for A^{3}X from the (A)XOS

296 studied (13). At higher enzyme concentrations MG43$_7$ shows xylanase (either endo- or exo-

297 xylanase activity as discussed for MG43$_{28}$) activity on higher dp XOS and AXH-d2 activity on

298 internal arabinose substituted xyloses. Further investigation should be made to understand such

299 particular substrate recognition by MG43$_7$ and GH43$_{18}$.

300 It is worth noting that our study provides detailed insights into substrate preferences but not into

301 kinetics. Overnight reactions were performed, but often incomplete conversions were observed.

302 Similar observations of incomplete conversions were observed before with AX-acting enzymes

303 (31, 43-45). This may either indicate low rates, enzyme death and/or product inhibition. The

304 latter is less likely to be an issue in natural systems where other enzymes further convert the

305 product from the first reaction. Yet, product inhibition is highly relevant in industrial applications

306 where high substrate concentrations are used. Consequently, enzymes are either selected based on

307 low product inhibition levels (46, 47), or their crystal structure is determined to unravel the

308 structural basis of product inhibition, giving rise to opportunities for protein engineering to

309 release or reduce product inhibition (48).

310 In conclusion, DSA-FACE enables a HT analysis of enzymatic substrate preferences of AXOS-

311 acting enzymes in a relatively short experimental and analysis time. Thanks to the HT nature of

14

312 the approach, by performing enzymatic reactions at different enzyme concentrations, different

313 (A)XOS structures can be ranked as preferred substrates and sequential enzymatic cleavages can

314 be determined. This approach allowed us to create degradation maps for five metagenome-

315 derived enzymes for twelve different (A)XOS substrates. The knowledge of the exact substrate

316 preferences is undoubtedly essential to achieve either desired hydrolysis products (e.g. prebiotics)

317 or to come to a full hydrolysis. Finally, given the variety, promiscuity and flexible substrate

318 preferences of the majority of carbohydrate active enzymes, DSA-FACE may be explored for

319 other activities rather than (arabino)xylanolytic activities.

320

321 **Materials and methods**

322 *Expression and purification of metagenome-derived enzymes*

323 pET28 plasmids containing the enzyme DNA sequences were obtained as described previously

324 (27). Chemically competent *Escherichia coli* TOP10 and *E. coli* BL21 (DE3), *E. coli* BL21

325 CodonPlus (DE3) or *E. coli* ArcticExpress strains prepared according to the rubidium chloride

326 method were transformed with these plasmids.

327 Table 2 gives an overview of the enzyme expression conditions obtained after preceding

328 optimization steps. For optimal aeration, Erlenmeyer flasks exceeding at least four times the

329 expression volume were used. The different expression hosts were grown at indicated

330 temperatures in lysogeny broth (LB) with appropriate antibiotics until reaching an OD_{600} of

331 approximately 0.6 followed by isopropyl β-D-1-thiogalactopyranoside (IPTG) induction. LBE-

332 5052 auto-induction medium consisted of 1% tryptone, 0.5% yeast extract, 40 mM K_2HPO_4, 10

333 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4, 2 mM Mg_2SO_4, 0.5% glycerol, 0.05% glucose,

334 0.2% lactose, 50 μg/mL kanamycin and trace metals mix (50 μM $FeCl_3$, 20 μM $CaCl_2$, 10 μM

335 $MnCl_2$, 10 μM $ZnSO_4$, 2 μM $CoCl_2$, 2 μM $CuCl_2$, and 2 μM $NiCl_2$).

15

336   Cells were harvested by centrifugation at 3100 x g for 30 min at 4 °C. Pellets were then

337   suspended in 1/25 of the original volume in equilibration buffer for metal affinity

338   chromatography (see below) and 1 mg/mL lysozyme and incubated on ice for 30 min. After three

339   freeze-thaw cycles, sonication was performed on ice (3 x 30 s with 30 s interval, 40% amplitude).

340   Cell debris were removed by centrifugation at 20000 x g for 30 min at 4 °C and resulting

341   supernatants were clarified by filtration with a 0.45 µm filter. Purifications by metal affinity

342   chromatography were performed either with His GraviTrap columns (GE Healthcare) or HisPur

343   Ni-NTA Superflow agarose (Thermo Fisher Scientific). The manufacturer's protocols were

344   followed in both cases with the exception for the latter that the sample-resin incubation time was

345   extended to 1 h and buffers used were modified (20 mM sodium phosphate 500 mM NaCl 20

346   mM imidazole pH 7.4 as equilibration buffer, 20 mM sodium phosphate 500 mM NaCl 50 mM

347   imidazole pH 7.4 as wash buffer and 20 mM sodium phosphate 500 mM NaCl 500 mM

348   imidazole pH 7.4 as elution buffer).

349   Eluted samples were diluted with reducing sample buffer, boiled for 5 min and analyzed by 12%

350   SDS-Page (Roti®-Mark standard from Carl Roth was used). Fractions containing the protein of

351   interest were then dialyzed against 20 mM HEPES-NaOH buffer pH 7.0 and 300 mM NaCl,

352   pooled and concentrated with Vivaspin concentrators when necessary. Dialysis was done with

353   Slide-A-Lyzer™ MINI Dialysis Devices, 3.5K MWCO (Thermo Fisher Scientific) or with

354   SERVAPOR® dialysis tubing, MWCO 12000-14000 RC, diameter 16 mm. Protein

355   concentrations were measured with the Abs280nm app of the DeNovix DS-11 series

356   spectrophotometer. Extinction coefficients were calculated with the ProtParam tool (ExPASy).

357

358

359

360 *Enzymatic reactions of metagenome-derived enzymes with (A)XOS*

361 Metagenome-derived enzymes were tested against (A)XOS (Figure 6) supplied by Megazyme

362 (Megazyme International Ireland, Bray, Ireland), which have a minimum purity of 95% except

363 for the mixture of A^2XX and A^3XX, which has a minimum purity of 90%, and for $XA^{2+3}XX$

364 which has a minimum purity of 85%. Enzymatic reactions with a total volume of 100 µL (or in

365 50 µL to achieve desired enzyme concentration when there was only a limited enzyme volume

366 available) in a 96-well plate contained 0.2-38 µM enzyme, 10 µM (A)XOS, 50 mM HEPES-

367 NaOH 50 mM NaCl pH 7.0. Mineral oil (30-50 µl) was used to avoid evaporation from the 96-

368 well plate during enzymatic reaction (Figure 7). Substrate and enzyme blanks, where enzyme and

369 substrate (respectively) were replaced by the corresponding buffer, have been added. Some

370 repetitions of reactions were performed in a 1.5 mL Eppendorf for reasons of simplicity.

371 Enzymatic reactions were incubated at 37 °C and 750 rpm in a Thermomixer comfort

372 (Eppendorf). The number of replicates done per enzyme/substrate combination is given in Table

373 S1. After 22 h, reactions were stopped by incubation at 80 °C for 30 min.

374

375 *Analysis of enzymatic reaction hydrolysates by DSA-FACE*

376 Reaction hydrolysates were diluted 10-fold with ultrapure water and 10 µL were lyophilized.

377 Carbohydrates present in the lyophilized fraction were then derivatized with 8-aminopyrene-

378 1,3,6-trisulfonic acid trisodium salt (APTS) by reductive amination as in (13). Afterwards,

379 samples were quenched by diluting the reactions 200-fold with ultrapure water. Ten µL of

380 derivatized hydrolysate was analyzed by the Applied Biosystems[TM] 3130 Genetic Analyzer with

381 36 cm capillaries filled with Applied Biosystems[TM] POP-7[TM] polymer as in (13) (Figure 7).

382 Through DSA-FACE electropherograms the carbohydrates before and after enzymatic reactions

383 are identified by comparison to standards. Xylose and arabinose monomers are not detected by

17

384 DSA-FACE as they fall into the DSA-FACE noise region due to their high electrophoretic

385 mobility. Since A^2X, $A^{2+3}X$, XA^2X, XA^3X and $XA^{2+3}X$ standards are not commercially available,

386 they were identified by comparison between the electrophoretic mobilities of the hydrolysates,

387 the electrophoretic mobilities of the available standards and based on spiking experiments (Figure

388 S5 and Figure S6). Previously it was seen that AXOS with dp z present an electrophoretic

389 mobility between XOS with dp z-1 and z showing an increased electrophoretic mobility in

390 comparison with XOS with the same dp. For example, A^2XX and A^3XX are therefore expected to

391 have an electrophoretic mobility in between X_3 and X_4 (13).

392

393 *DSA-FACE product profiles*

394 DSA-FACE product profiles were made with the excel graph function. Peak areas were collected

395 with the GeneMapper® Software Version 4.0. DSA-FACE peak area reproducibility is dependent

396 on the amount of labeled carbohydrate injected in each run, which may vary due to the

397 electrokinetic injection mechanism of the 3130 Genetic Analyzer. Intrinsic carbohydrate

398 electrophoretic mobilities affect the amount of sample injected by the electrokinetic mechanism

399 (49). Therefore peak areas are corrected by dividing the hydrolysate peak areas by the peak area

400 of the blank with same (A)XOS structure. When this AXOS was not one of the standard AXOS,

401 the peak area of an AXOS with the same dp is taken. The average of the corrected peak areas is

402 then taken for the DSA-FACE product profiles. To normalize all peak areas obtained for the

403 same enzyme but different enzyme concentrations and substrates, the largest peak area (or the

404 largest sum of the carbohydrate peak areas when more peaks are present in an hydrolysate) is

405 taken as the maximum amount of carbohydrate possibly found in a hydrolysate. All product

406 profiles revealed by DSA-FACE are summarized in Table S2.

407

408   *Data availability*

409   GenBank accession numbers for the enzyme DNA sequences 12_H03-13 (MG43$_2$-8), 12_H03-12

410   (MG43$_7$), and 12_J03-18 (MG43$_{28}$) (Table 1) are MT603581, MT603582 and MT603583,

411   respectively.

412

413   **Acknowledgments**

416

417   **References**

418   1.   Hess M, Sczyrba A, Egan R, Kim T-W, Chokhawala H, Schroth G, Luo S, Clark DS,

419   Chen F, Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z,

420   Rubin EM. 2011. Metagenomic discovery of biomass-degrading genes and genomes from

421   cow rumen. Science 331:463-467.

422   2.   Armstrong Z, Mewis K, Strachan C, Hallam SJ. 2015. Biocatalysts for biomass

423   deconstruction from environmental genomics. Current Opinion in Chemical Biology

424   29:18-25.

425   3.   Helbert W, Poulet L, Drouillard S, Mathieu S, Loiodice M, Couturier M, Lombard V,

426   Terrapon N, Turchetto J, Vincentelli R, Henrissat B. 2019. Discovery of novel

427   carbohydrate-active enzymes through the rational exploration of the protein sequences

428   space. Proceedings of the National Academy of Sciences of the United States of America

429   116:10184-10185.

19

430 4. Chaturvedi T, Torres AI, Stephanopoulos G, Thomsen MH, Schmidt JE. 2020.

431   Developing process designs for biorefineries-definitions, categories, and unit operations.

432   Energies 13:22.

433 5. Villota EM, Dai Z, Lu Y, Yang B. 2020. Enzymes for cellulosic biomass hydrolysis and

434   saccharification, p 283-326. *In* Vertès AA, Qureshi N, Blaschek HP, Yukawa H (ed),

435   Green energy to sustainability: strategies for global industries. John Wiley & Sons Ltd.

436 6. Khandeparker R, Numan MT. 2008. Bifunctional xylanases and their potential use in

437   biotechnology. Journal of industrial microbiology&biotechnology 35:635-644.

438 7. Lee K-T, Toushik SH, Baek J-Y, Kim J-E, Lee J-S, Kim K-S. 2018. Metagenomic mining

439   and functional characterization of a novel KG51 bifunctional cellulase/hemicellulase from

440   black goat rumen. Journal of agricultural and food chemistry 66:9034-9041.

441 8. Yang W, Bai Y, Yang P, Luo H, Huang H, Meng K, Shi P, Wang Y, Yao B. 2015. A

442   novel bifunctional GH51 exo-α-L-arabinofuranosidase/endo-xylanase from

443   Alicyclobacillus sp. A4 with significant biomass-degrading capacity. Biotechnology for

444   Biofuels 8:197.

445 9. Malgas S, Mafa MS, Mkabayi L, Pletschke BI. 2019. A mini review of xylanolytic

446   enzymes with regards to their synergistic interactions during hetero-xylan degradation.

447   World Journal of Microbiology and Biotechnology 35:187.

448 10. Kračun SK, Schückel J, Westereng B, Thygesen LG, Monrad RN, Eijsink VGH, Willats

449   WGT. 2015. A new generation of versatile chromogenic substrates for high-throughput

450   analysis of biomass-degrading enzymes. Biotechnology for Biofuels 8:70.

451 11. Borsenberger V, Ferreira F, Pollet A, Dornez E, Desrousseaux M-L, Massou S, Courtin

452   CM, O'Donohue MJ, Fauré R. 2012. A versatile and colorful screening tool for the

453   identification of arabinofuranose-acting enzymes. Chembiochem 13:1885-1888.

454   12.   Chen H-m, Armstrong Z, Hallam SJ, Withers SG. 2016. Synthesis and evaluation of a

455   series of 6-chloro-4-methylumbelliferyl glycosides as fluorogenic reagents for screening

456   metagenomic libraries for glycosidase activity. Carbohydrate Research 421:33-39.

457   13.   Fonseca MJMd, Jurak E, Kataja K, Master ER, Berrin J-G, Stals I, Desmet T, Landschoot

458   AV, Briers Y. 2018. Analysis of the substrate specificity of α-L-arabinofuranosidases by

459   DNA sequencer-aided fluorophore-assisted carbohydrate electrophoresis. Applied

460   Microbiology and Biotechnology 23:10091-10102.

461   14.   Scheller HV, Ulvskov P. 2010. Hemicelluloses. Annual Review of Plant Biology 61:263-

462   289.

463   15.   Li Z, Qu H, Li C, Zhou X. 2013. Direct and efficient xylitol production from xylan by

464   *Saccharomyces cerevisiae* through transcriptional level and fermentation processing

465   optimizations. Bioresource Technology 149:413-419.

466   16.   Poletto P, Pereira GN, Monteiro CRM, Pereira MAF, Bordignon SE, Oliveira Dd. 2020.

467   Xylooligosaccharides: Transforming the lignocellulosic biomasses into valuable 5-carbon

468   sugar prebiotics. Process Biochemistry 91:352-363.

469   17.   Vázquez MJ, Alonso JL, Dominguez H, Parajó JC. 2001. Xylooligosaccharides:

470   manufacture and applications. Trends in Food Science & Technology 11:387-393.

471   18.   Dumon C, Song L, Bozonnet S, Fauré R, O'Donohue MJ. 2012. Progress and future

472   prospects for pentose-specific biocatalysts in biorefining. Process Biochemistry 47:346-

473   357.

474   19.   Wilkens C, Andersen S, Dumon C, Berrin J-g, Svensson B. 2017. GH62

475   arabinofuranosidases: Structure, function and applications. Biotechnology Advances

476   35:792-804.

477 20. Biely P, Singh S, Puchart V. 2016. Towards enzymatic breakdown of complex plant xylan

478   structures: State of the art. Biotechnology Advances 34:1260-1274.

479 21. Deutschmann R, Dekker RFH. 2012. From plant biomass to bio-based chemicals: Latest

480   developments in xylan research. Biotechnology Advances 30:1627-1640.

481 22. Kabel MA, Kortenoeven L, Schols HA, Voragen AGJ. 2002. In vitro fermentability of

482   differently substituted xylo-oligosaccharides. Journal of agricultural and food chemistry

483   50:6205-6210.

484 23. Laere KMJV, Hartemink R, Bosveld M, Schols HA, Voragen AGJ. 2000. Fermentation of

485   plant cell wall derived polysaccharides and their corresponding oligosaccharides by

486   intestinal bacteria. Journal of agricultural and food chemistry 48:1644-1652.

487 24. Holck J, Djajadi DT, Brask J, Pilgaard B, Krogh KBRM, Meyer AS, Lange L, Wilkens C.

488   2019. Novel xylanolytic triple domain enzyme targeted at feruloylated arabinoxylan

489   degradation. Enzyme and Microbial Technology 129:109353.

490 25. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B. 2014. The carbohydrate-

491   active enzymes database (CAZy) in 2013. Nucleic acids research 42:490-495.

492 26. Mewis K, Lenfant N, Lombard V, Henrissat B. 2016. Dividing the large glycoside

493   hydrolase family 43 into subfamilies: A motivation for detailed enzyme characterization.

494   Applied and Environmental Microbiology 82:1686-1692.

495 27. Armstrong Z, Mewis K, Liu F, Melanie CM-l, Ming H, Kevin C, Stephen M, Hallam SJ.

496   2018. Metagenomics reveals functional synergy and novel polysaccharide utilization loci

497   in the *Castor canadensis* fecal microbiome. The ISME Journal 12:2757-2769.

498 28. Jiménez-Ortega E, Valenzuela S, Ramírez-Escudero M, Pastor FJ, Sanz-Aparicio J. 2020.

499   Structural analysis of the reducing-end xylose-releasing exo-oligoxylanase Rex8A from

500   *Paenibacillus barcinonensis* BP-23 deciphers its molecular specificity. The FEBS Journal.

22

501 29. Valenzuela SV, Lopez S, Biely P, Sanz-Aparicio J, Pastor FIJ. 2016. The glycoside

502   hydrolase family 8 reducing-end xylose-releasing exo-oligoxylanase Rex8A from

503   *Paenibacillus barcinonensis* BP-23 is active on branched xylooligosaccharides. Applied

504   and Environmental Microbiology 82:5116 –5124.

505 30. Rio TGD, Abt B, Spring S, Lapidus A, Nolan M, Tice H, Copeland A, Cheng J-F, Chen

506   F, Bruce D, Goodwin L, Pitluck S, Ivanova N, Mavromatis K, Mikhailova N, Pati A,

507   Chen A, Palaniappan K, Land M, Hauser L, Chang Y-J, Jeffries CD, Chain P, Saunders E,

508   Detter JC, Brettin T, Rohde M, Göker M, Bristow J, Eisen JA, Markowitz V, Hugenholtz

509   P, Kyrpides NC, Klenk H-P, Lucas S. 2010. Complete genome sequence of *Chitinophaga*

510   *pinensis* type strain (UQM 2034T). Standards in Genomic Sciences 2:87-95.

511 31. Teeravivattanakit T, Baramee S, Phitsuwan P, Waeonukul R, Pason P, Tachaapaikoon C,

512   Sakka K, Ratanakhanokchai K. 2016. Novel trifunctional xylanolytic enzyme Axy43A

513   from *Paenibacillus curdlanolyticus* strain B-6 exhibiting endo-xylanase, beta-D-

514   xylosidase, and arabinoxylan arabinofuranohydrolase activities. Applied and

515   Environmental Microbiology 82:6942-6951.

516 32. Broberg A, Duus J, Thomsen KK, Ferre H. 2000. A novel type of arabinoxylan

517   arabinofuranohydrolase isolated from germinated barley analysis of substrate preference

518   and specificity by nano-probe NMR. European Journal of Biochemistry 267:6633-6641.

519 33. Mathieu SV, Aragão KS, Imberty A, Varrot A. 2010. Discoidin I from *Dictyostelium*

520   *discoideum* and interactions with oligosaccharides: specificity, affinity, crystal structures

521   and comparison with Discoidin II. Journal of Molecular Biology 400:540-554.

522 34. Vincent P, Shareck F, Dupont C, Morosoli R, Kluepfel D. 1997. New α-L-

523   arabinofuranosidase produced by *Streptomyces lividans*: cloning and DNA sequence of

524   the abfB gene and characterization of the enzyme. Biochemical Journal 322:845-852.

23

525 35. Adelsberger H, Hertel C, Glawischnig E, Zverlov VV, Schwarz WH. 2004. Enzyme

526 system of *Clostridium stercorarium* for hydrolysis of arabinoxylan: reconstitution of the

527 *in vivo* system from recombinant enzymes. Microbiology 150:2257-2266.

528 36. Saha BC. 2003. Hemicellulose bioconversion. Industrial Biotechnology Journal 30:279-

529 291.

530 37. Lagaert S, Pollet A, Delcour JA, Lavigne R, Courtin CM, Volckaert G. 2011.

531 Characterization of two β-xylosidases from *Bifidobacterium adolescentis* and their

532 contribution to the hydrolysis of prebiotic xylooligosaccharides. Applied Microbiology

533 and Biotechnology 92:1179-1185.

534 38. Honda Y, Kitaoka M. 2004. A family 8 glycoside hydrolase from *Bacillus halodurans* C-

535 125 (BH2105) is a reducing end xylose-releasing exo-oligoxylanase. The Journal of

536 biological chemistry 279:55097–55103.

537 39. Lagaert S, Campenhout SV, Pollet A, Bourgois TM, Delcour JA, Courtin CM, Volckaert

538 G. 2007. Recombinant expression and characterization of a reducing-end xylose-releasing

539 exo-oligoxylanase from *Bifidobacterium adolescentis*. Applied and Environmental

540 Microbiology 73:5374-5377.

541 40. Hong P-Y, Iakiviak M, Dodd D, Zhang M, Mackie RI, Canna I. 2014. Two new xylanases

542 with different substrate specificities from the human gut bacterium *Bacteroides*

543 *intestinalis* DSM 17393. Applied and Environmental Microbiology 80:2084-2093.

544 41. Lagaert S, Pollet A, Courtin CM, Volckaert G. 2014. β-Xylosidases and α-L-

545 arabinofuranosidases: accessory enzymes for arabinoxylan degradation. Biotechnology

546 Advances 32:316-332.
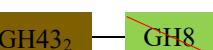
547   42.   Amorim C, Silvério SC, Prather KLJ, Rodrigues LR. 2019. From lignocellulosic residues

548   to market: Production and commercial potential of xylooligosaccharides. Biotechnology

549   Advances 37:107397.

550   43.   Wongratpanya K, Imjongjairak S, Waeonukul R, Sornyotha S, Phitsuwan P, Pason P,

551   Nimchua T, Tachaapaikoon C, Ratanakhanokchai K. 2015. Multifunctional properties of

552   glycoside hydrolase family 43 from *Paenibacillus curdlanolyticus* strain B-6 including

553   exo-beta-xylosidase, endoxylanase, and alpha-L-arabinofuranosidase activities.

554   Bioresources 10:2492-2505.

555   44.   Broeker J, Mechelke M, Baudrexl M, Mennerich D, Hornburg D, Mann M, Schwarz WH,

556   Liebl W, Zverlov VV. 2018. The hemicellulose-degrading enzyme system of the

557   thermophilic bacterium *Clostridium stercorarium*: comparative characterisation and

558   addition of new hemicellulolytic glycoside hydrolases. Biotechnology for Biofuels

559   11:229.

560   45.   Wang Y, Sakka M, Yagi H, Kaneko S, Katsuzaki H, Kunitake E, Kimura T, Sakka K.

561   2018. *Ruminiclostridium josui* Abf62A-Axe6A: A tri-functional xylanolytic enzyme

562   exhibiting α-L-arabinofuranosidase, endoxylanase, and acetylxylan esterase activities.

563   Enzyme and Microbial Technology 117:1-8.

564   46.   Bachmann SL, McCarthy AJ. 1991. Purification and cooperative activity of enzymes

565   constituting the xylan-degrading system of Thermomonospora fusca. Applied and

566   Environmental Microbiology 57:2121-2130.

567   47.   Huang Y, Zheng X, Pilgaard B, Holck J, Muschiol J, Li S, Lange L. 2019. Identification

568   and characterization of GH11 xylanase and GH43 xylosidase from the chytridiomycetous

569   fungus, *Rhizophlyctis rosea*. Applied Microbiology and Biotechnology 103:777-791.

570   48.   Rohman A, Oosterwijk Nv, Puspaningsih NNT, Dijkstra BW. 2018. Structural basis of

571   product inhibition by arabinose and xylose of the thermostable GH43 β-1,4-xylosidase

572   from *Geobacillus thermoleovorans* IT-08. PloS ONE 13:e0196358.

573   49.   Schaeper JP, Sepaniak MJ. 2000. Parameters affecting reproducibility in capillary

574   electrophoresis. Electrophoresis 21:1421-1429.

575   50.   Fauré R, Courtin CCM, Delcour DJA, Dumon DC, Faulds CCB, Geoffrey EB, Fort FS,

576   Fry GSC, Halila HS, Kabel GMA, Pouvreau L, Bernard I, Rivet JA, Saulnier GL, Schols

577   JHA, Driguez IH, A MJOD. 2009. A brief and informationally rich naming system for

578   oligosaccharide motifs of heteroxylans found in plant cell walls. Australian Journal of

579   Chemistry 62:533-537.

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594   **Tables**

595

596   **Table 1 - Metagenomic AX-active enzymes analyzed in this study.** The CAZy family (and

597   subfamily in subscript) is given for the modules that constitute each enzyme. Domains

598   inactivated by mutagenesis are indicated with a red line. Activity tests on aryl glycosides p-

599   nitrophenyl β-D-xylopyranoside (pNP-X), 4-methylumbelliferyl β-D-xylopyranoside (MU-X), 6-

600   chloro-4-methylumbelliferyl   β-D-xylopyranoside   (CMU-X),   p-nitrophenyl   α-L-

601   arabinofuranoside (pNP-Ara) and 4-methylumbelliferyl α-L-arabinofuranoside (MU-Ara) and

602   HPAEC-PAD analysis using A^3X, A^2XX, XA^3XX and XA^2XX as substrates were performed by

603   (27). CBM = Carbohydrate binding module. Rex = reducing end xylose-releasing exo-

604   oligoxylanase.

Protein name used in this study	Protein name used in (27)	Enzyme modularity	Activity on aryl glycosides	Activity detected by HPAEC-PAD	Observations
$MG43_{28}$	12_J03-18	GH43$_{28}$	CMU-X	Not tested	-
$MG43_2$-8	12_H03-13	GH43$_2$ GH8	CMU-X	Rex and β-xylosidase	Does not hydrolyse X_2
$MG43_2$	12_H03-13_E507A	GH43$_2$ GH8	CMU-X	β-xylosidase	Does not hydrolyse X_2
MG8	12_H03-13_E209A	GH43$_2$ GH8	Not detected	Rex	-
$MG43_7$	12_H03-12	GH43$_7$ CBM13	Not detected	Cleaves O-3-arabinose decorations from A^3X	Only active in the presence of $MG43_2$-8

605

27

606 **Table 2 - Expression conditions of the metagenome-derived enzymes studied in this work.**

Enzyme	Expression host	Host growth temperature	Induction type	Purification method
MG43$_{28}$	*E. coli* BL21 (DE3)	24 h at 37 ºC, 250 rpm on LBE50-52 auto-induction medium		HisGraviTrap
MG43$_2$-8	*E. coli* BL21 CodonPlus (DE3)	37 ºC, 250 rpm	1 mM IPTG at 16 ºC for 18 to 20 hours, 250 rpm	HisPur Ni-NTA Superflow agarose (250/100 µL resin)
	E. coli ArcticExpress	30 ºC, 250 rpm	1 mM IPTG at 16 ºC for 24 hours, 250 rpm	HisPur Ni-NTA Superflow agarose (100 µL resin)
MG43$_2$	*E. coli* BL21 CodonPlus (DE3)	30 ºC/37 ºC, 250 rpm	1 mM IPTG at 16 ºC for 18 to 20 hours, 250 rpm	HisPur Ni-NTA Superflow agarose (500/100 µL resin)
MG43$_7$		18 h at 30 ºC, 250 rpm on LBE50-52 auto-induction medium		HisGraviTrap
		30 ºC, 250 rpm	1 mM IPTG at 16 ºC for 18 hours, 250 rpm	HisPur Ni-NTA Superflow agarose (500 µL resin)
MG8	*E. coli* BL21 (DE3)	37 ºC, 250 rpm	1 mM IPTG at 16 ºC for 18 hours, 250 rpm	HisPur Ni-NTA Superflow agarose (500 µL resin)

607

608

609

610

611

612

613

614

615

616

28

617   **Figures**

618

A Environmental metagenomic libraries

Beaver faecal sample collection

DNA extraction and sequencing

Large insert DNA libraries
4500 clones

Parallel expression and preliminary functional screening with aryl glycosides
51 positive clones

Positive clones sequencing and gene (cluster of genes) identification
51 clones sequenced, 135 GH genes from 28 GH families identified

B AX-active enzymes characterization

27 GH43 genes from 10 subfamilies were identified in active fosmids. One of them is a modular enzyme comprising a GH43 – subfamily 2 domain and a GH8 domain. Two other GH43 genes belong to the uncharacterized subfamilies 7 and 28.

Genes cloning, mutagenesis, protein expression and purification
- One $GH43_7$ and one $GH43_{28}$ gene and the modular $GH43_2$ – GH8
- Two mutants: $GH43_2$ and GH8 alone

Preliminary functional screening with aryl glycosides

Substrate specificity characterization by HPAEC-PAD
$GH43_7$, $GH43_{28}$, $GH43_2$ – GH8 and mutants $GH43_2$ and GH8 are active on AX

635   **Figure 1 – Preceding functional screening of putative enzymes derived from metagenomics**

636   **on beaver fecal samples.** A) Upon environmental sample collection and gDNA extraction, a

637   metagenomic DNA library of 4500 clones suitable for heterologous expression was constructed.

638   These clones were expressed and checked for active hits by high-throughput preliminary

639   functional screening methods. Fifty one active hits were sequenced and 135 putative glycoside

640   hydrolases (GHs) from 28 GH families were identified by in silico analysis. B) Three GH43

29

641   genes, one of which is modular with an additional GH8 domain, and two mutants thereof were

642   characterized by enzymatic activity tests with aryl glycosides and by HPAEC-PAD using

643   representative arabinoxylan oligosaccharides (27).

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

660

661

662

663

664
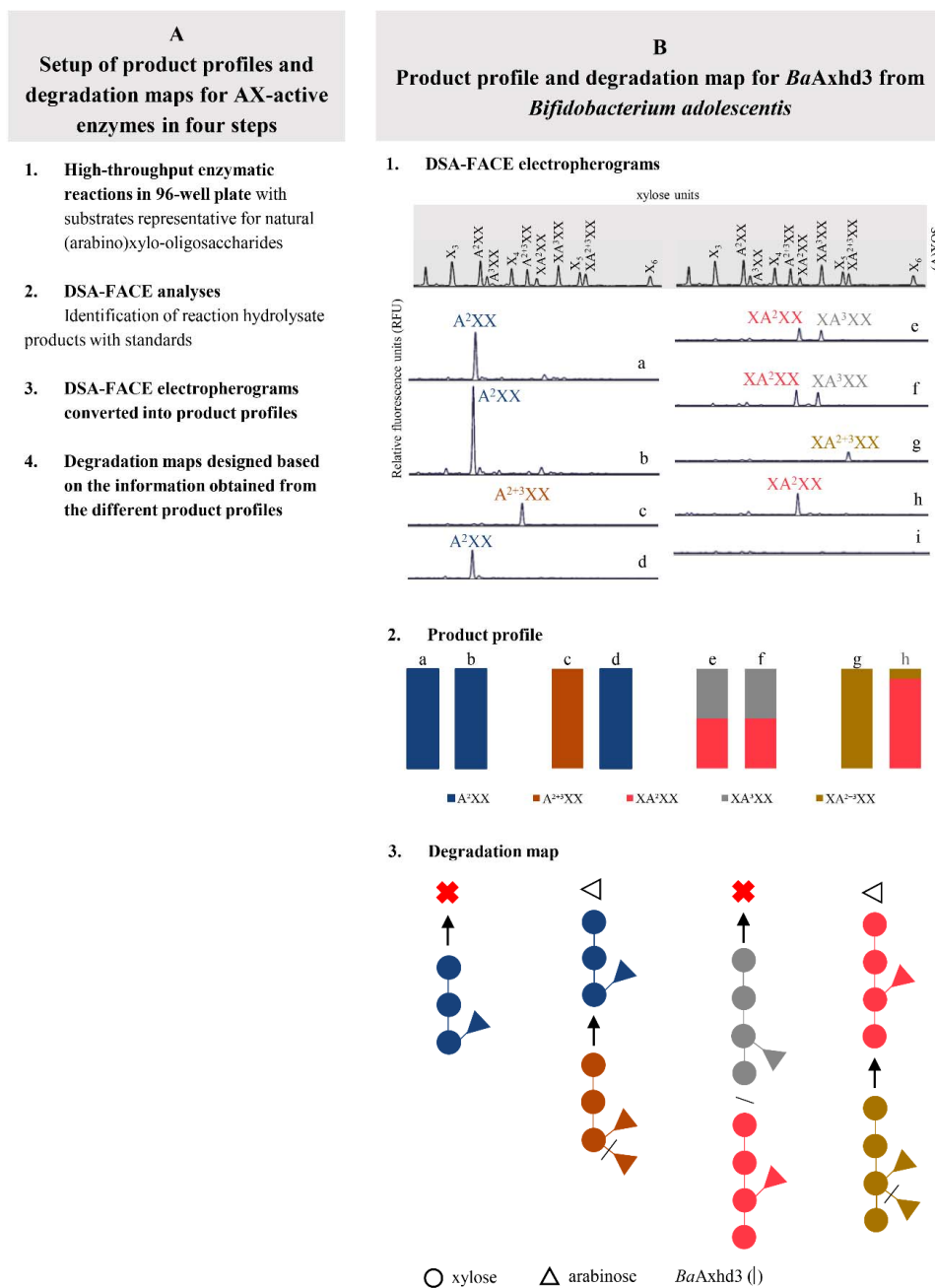
665 **Figure 2 – Product profiles and degradation maps.** (A) General approach to establish DSA-

666 FACE product profiles and a corresponding degradation map in four steps. (B) Product profiles

31

667 for *Ba*Axhd3 from *Bifidobacterium adolescentis* after reaction with A^{2}XX, A^{2+3}XX, XA^{2}XX,

668 XA^{3}XX and XA^{2+3}XX. Electropherograms a, c, e and g show the substrate blanks, whereas

669 electropherograms b, d, f and h show the corresponding hydrolysates upon enzymatic reaction

670 with *Ba*Axhd3. The peaks are compared to standards for carbohydrate peaks identification (1). A

671 qualitative interpretation of the electropherograms is then displayed on a product profile (bars are

672 labeled with the letters of corresponding electropherograms) (2). Substrate conversions are easily

673 observed by a color change. The first bar corresponds to the substrate blank followed by bar(s)

674 showing colors corresponding to the (A)XOS found upon enzymatic reaction. A degradation map

675 is obtained from the different product profiles for *Ba*Axhd3 (3).

676

677

678

679

680

681

682

683

684

685

686

687

A

B
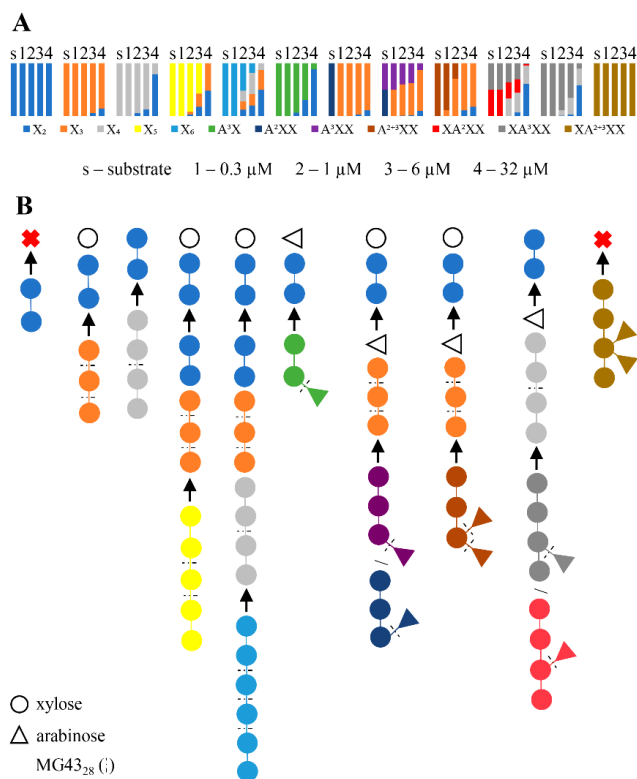
Figure 3 - Product profiles starting from 12 (A)XOS and degradation map of MG43$_{28}$. The product profiles (A) show the hydrolysis products obtained after 22 h of enzymatic reactions with 0.3, 1, 6 and 32 µM MG43$_{28}$ (1, 2, 3 and 4 respectively). The (A)XOS used as substrates for the enzymatic reactions are identified as 's'. A degradation map is given in B with a schematic representation of the (A)XOS structures used as substrates and the ones obtained as hydrolysis products, using corresponding colors. Based on hydrolysis products obtained, cleavage positions for MG43$_{28}$ are indicated with ⫶.

33

711

712

713

714

715

716

717

718

719

720

721

722

723

724
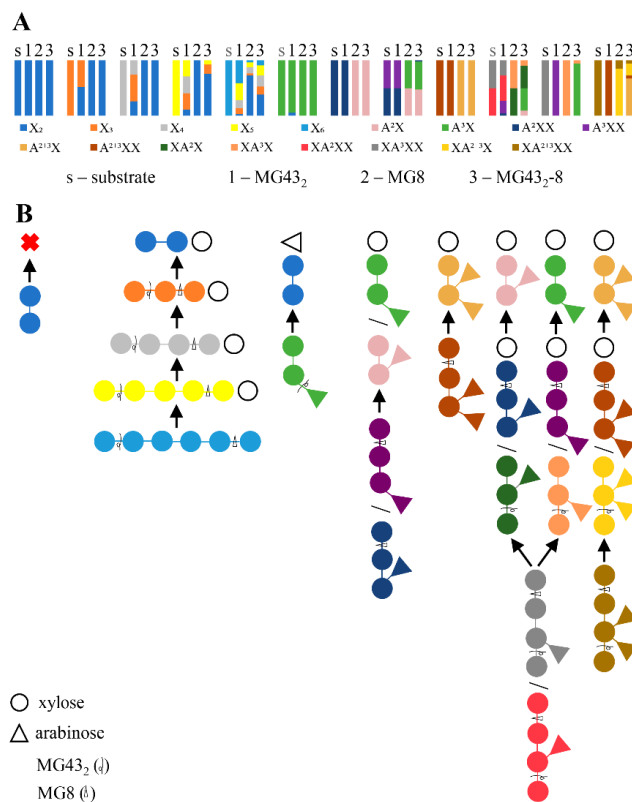
725 **Figure 4 - Product profiles of MG43$_2$, MG8 and MG43$_2$-8.** The product profiles in A show the

726 hydrolysis products obtained after 22 h of enzymatic reactions with 3 μM MG43$_2$ (1), MG8 (2)

727 and MG43$_2$-8 (3). The (A)XOS used as substrates for the enzymatic reactions are identified as

728 's'. A degradation map is given in B with a schematic representation of the (A)XOS structures

729 used as substrates and the ones obtained as hydrolysis products, using corresponding colors.

730 Based on hydrolysis products obtained, cleavage positions for MG43$_2$ and MG8 are indicated

731 with ∤ and ∆ , respectively.

732

34

733
734
735
736
737
738
739
740
741
742
743
744
745
746
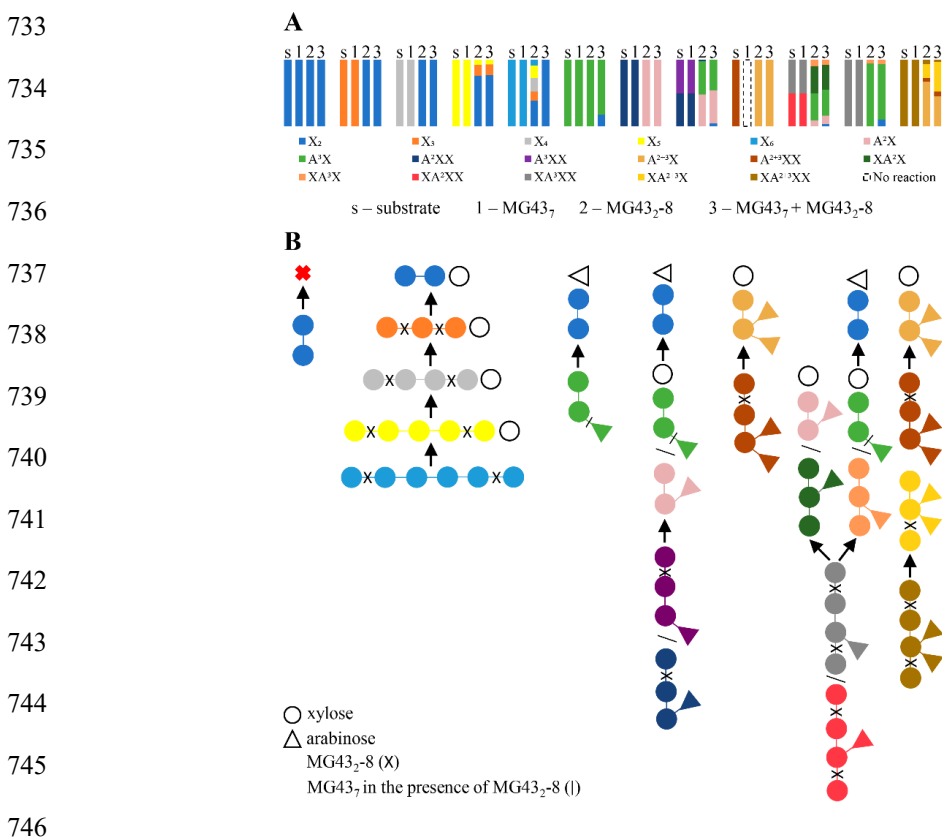
747 **Figure 5 - Product profiles of MG43$_7$, MG43$_2$-8, and MG43$_7$ in the presence of MG43$_2$-8.** The

748 product profiles (A) show the hydrolysis products obtained after 22 h of enzymatic reactions with

749 8 μM MG43$_7$ (1), 3 μM MG43$_2$-8 (2) and 8 μM MG43$_7$ in the presence of 3 μM MG43$_2$-8 (3).

750 The (A)XOS used as substrates for the enzymatic reactions are identified as 's'. The dotted line

751 means there was no reaction performed to test the hydrolysis of A^{2+3}XX by MG43$_7$. A

752 degradation map is given in B with a schematic representation of the (A)XOS structures used as

753 substrates and the ones obtained as hydrolysis products. Based on hydrolysis products obtained,

754 cleavage positions for MG43$_7$ and MG43$_2$-8 are indicated with $^|$ and x, respectively.

755

35

X₂ to X₆   **A²⁺³XX**

A³X   **XA²XX**

A²XX   **XA³XX**

A³XX   **XA²⁺³XX**
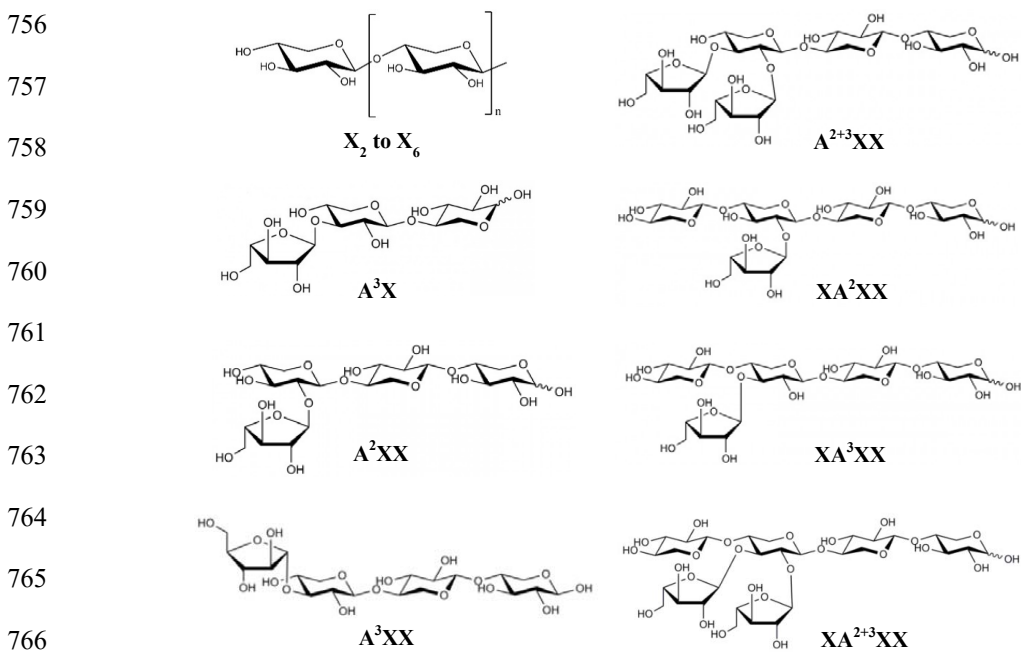
Figure 6 - Twelve different (arabino)xylo-oligosaccharides ((A)XOS) used as substrates in the enzymatic reactions and as standards for the DSA-FACE analysis. AXOS are named according to nomenclature proposed by (50).

A. Enzymatic reactions in 96-well plate
(10 μM substrate)
6 enzymes, 5 XOS, 7 AXOS, 12 substrate blanks, 6 enzyme blanks

enzyme + (A)XOS + buffer

22h

B. Lyophilization of hydrolysates

2h

C. Carbohydrates derivatization

APTS

18h

D. DSA-FACE

~14h
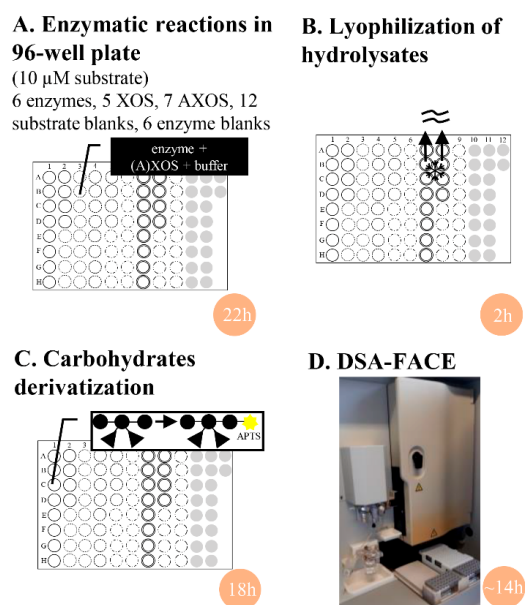
Figure 7 - Protocol for high-throughput study of substrate specificities of arabinoxylan-active enzymes by DSA-FACE. Putative AX-active enzymes are incubated with (A)XOS for 22 hours (A). Six enzymes were tested against 12 (A)XOS, including 12 substrate blanks and 6 enzyme blanks (90 samples in total). Reaction hydrolysates are then diluted with ultrapure water and lyophilized (B). Afterwards, reductive amination reactions are performed to derivatize the carbohydrates at their reducing end with the negatively charged and fluorescent APTS (C). Ten microliters of derivatized reaction hydrolysate are analyzed by DSA-FACE (D). All steps are done in a 96-well plate and 90 samples are analyzed in approximately 14 hours.

A

Environmental metagenomic libraries
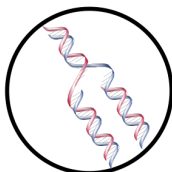
Beaver faecal sample collection

DNA extraction and sequencing

Large insert DNA libraries
4500 clones
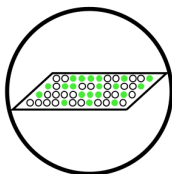
Parallel expression and preliminary functional screening with aryl glycosides
51 positive clones

ATTCGATAGG
CCATACTAGT
CCACACTTGT
ATCCACACTT
GCCAGCATA

Positive clones sequencing and gene (cluster of genes) identification
51 clones sequenced, 135 GH genes from 28 GH families identified

B

AX-active enzymes characterization

27 GH43 genes from 10 subfamilies were identified in active fosmids. One of them is a modular enzyme comprising a GH43 – subfamily 2 domain and a GH8 domain. Two other GH43 genes belong to the uncharacterized subfamilies 7 and 28.

Genes cloning, mutagenesis, protein expression and purification

- One $GH43_7$ and one $GH43_{28}$ gene and the modular $GH43_2$ – GH8
- Two mutants: $GH43_2$ and GH8 alone

Preliminary functional screening with aryl glycosides

Substrate specificity characterization by HPAEC-PAD
$GH43_7$, $GH43_{28}$, $GH43_2$ – GH8 and mutants $GH43_2$ and GH8 are active on AX

A
Setup of product profiles and degradation maps for AX-active enzymes in four steps

1. **High-throughput enzymatic reactions in 96-well plate** with substrates representative for natural (arabino)xylo-oligosaccharides

2. **DSA-FACE analyses** Identification of reaction hydrolysate products with standards

3. **DSA-FACE electropherograms converted into product profiles**

4. **Degradation maps designed based on the information obtained from the different product profiles**

B
Product profile and degradation map for *Ba*Axhd3 from *Bifidobacterium adolescentis*

1. **DSA-FACE electropherograms**
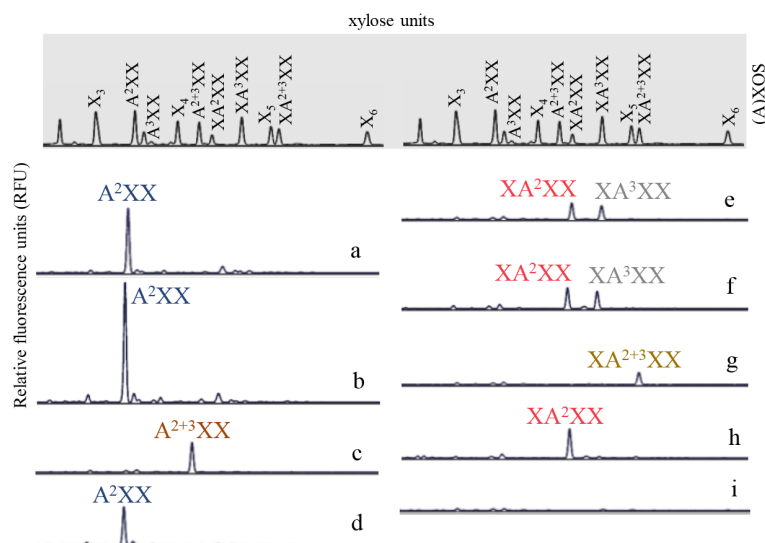
2. **Product profile**

3. **Degradation map**

○ xylose   △ arabinose   *Ba*Axhd3 (∣)

A

s – substrate   1 – 0.3 μM   2 – 1 μM   3 – 6 μM   4 – 32 μM

X₂   X₃   X₄   X₅   X₆   A³X   A²XX   A³XX   A²⁺³XX   XA²XX   XA³XX   XA²⁺³XX

B

○ xylose
△ arabinose
MG43₂₈ (¦)

A

s 1 2 3 (repeated across panels)

■ X_2 ■ X_3 ■ X_4 ■ X_5 ■ X_6 ■ A^2X ■ A^3X ■ A^2XX ■ A^3XX

■ $A^{2+3}X$ ■ $A^{2+3}XX$ ■ XA^2X ■ XA^3X ■ XA^2XX ■ XA^3XX ■ $XA^{2+3}X$ ■ $XA^{2+3}XX$

s – substrate   1 – MG43$_2$ 2 – MG8   3 – MG43$_2$-8

B

○ xylose

△ arabinose

MG43$_2$ ( )

MG8 ( )

A

Legend:
- X_2
- X_3
- X_4
- X_5
- X_6
- A^2X
- A^3X
- A^2XX
- A^3XX
- $A^{2+3}X$
- $A^{2+3}XX$
- XA^2X
- XA^3X
- XA^2XX
- XA^3XX
- $XA^{2+3}X$
- $XA^{2+3}XX$
- No reaction

s – substrate   1 – MG43$_7$ 2 – MG43$_2$-8   3 – MG43$_7$ + MG43$_2$-8

B

○ xylose
△ arabinose
MG43$_2$-8 (X)
MG43$_7$ in the presence of MG43$_2$-8 (|)

X_2 to X_6

$A^{2+3}XX$

A^3X

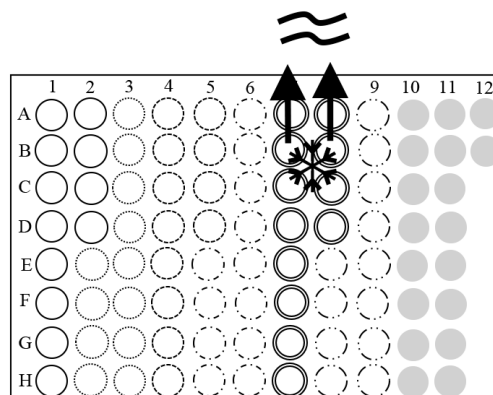XA^2XX

A^2XX

XA^3XX

A^3XX

$XA^{2+3}XX$

A. Enzymatic reactions in 96-well plate

(10 µM substrate)
6 enzymes, 5 XOS, 7 AXOS, 12 substrate blanks, 6 enzyme blanks
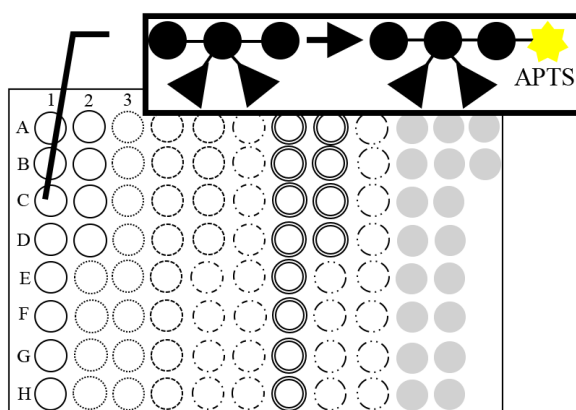
enzyme + (A)XOS + buffer

22h

B. Lyophilization of hydrolysates

2h

C. Carbohydrates derivatization

APTS

18h

D. DSA-FACE

~14h