A multifunctional GH39 glycoside hydrolase from the anaerobic gut fungus *Orpinomyces* sp. strain C1A

Jessica M Morrison [1] , Mostafa S Elshahed [1] , Noha Youssef Corresp. [1]

[1] Department of Microbiology and Molecular Genetics, Oklahoma State University, Stillwater, OK, USA

Corresponding Author: Noha Youssef
Email address: noha@okstate.edu

Background. The anaerobic gut fungi (phylum Neocallimastigomycota) represent a promising source of novel lignocellulolytic enzymes. Here, we report on the cloning, expression, and characterization of a glycoside hydrolase family 39 (GH39) enzyme (Bgxg1) that is highly transcribed by the anaerobic fungus *Orpinomyces* sp. strain C1A under different growth conditions. This represents the first study of a GH39-family enzyme from the anaerobic fungi.

Methods. Using enzyme activity assays, we performed a biochemical characterization of Bgxg1 on a variety of substrates over a wide range of pH and temperature values to identify the optimal enzyme conditions and the specificity of the enzyme. In addition, substrate competition studies and comparative modeling efforts were completed.

Results. Contrary to the narrow range of activities (β-xylosidase or α-L-iduronidase) observed in previously characterized GH39 enzymes, Bgxg1 is unique in that it is multifunctional, exhibiting strong β-xylosidase, β-glucosidase, β-galactosidase activities (11.5 ± 1.2, 73.4 ± 7.15, and 54.6 ± 2.26 U/mg, respectively) and a weak xylanase activity (10.8 ± 1.25 U/mg), as compared to previously characterized enzymes. Further, Bgxg1 possesses extremely high affinity (as evident by the lowest K_m values), compared to all previously characterized β -glucosidases, β-galactosidases, and xylanases. Physiological characterization revealed that Bgxg1 is active over a wide range of pH (3-8, optimum 6) and temperatures (25-60°C, optimum 39°C), and possesses excellent temperature and thermal stability. Substrate competition assays suggest that all observed activities occur at a single active site. Using comparative modeling and bioinformatics approaches, we putatively identified ten amino acid differences between Bgxg1 and previously biochemically characterized GH39 β-xylosidases that we speculate could impact active site architecture, size, charge, and/or polarity.

Discussion. Collectively, the unique capabilities and multi-functionality of Bgxg1 render it an excellent candidate for inclusion in enzyme cocktails mediating cellulose and hemicellulose saccharification from lignocellulosic biomass.

7 # A multifunctional GH39 glycoside hydrolase from the anaerobic gut
8 ## fungus *Orpinomyces* sp. strain C1A

9 Jessica M. Morrison[1], Mostafa S. Elshahed[1], and Noha H. Youssef[1*]
10
11 [1]Department of Microbiology and Molecular Genetics, Oklahoma State University,
12 Stillwater, OK, USA
13
14 Author Information:

15 Jessica M. Morrison, Ph.D.
16 1110 S. Innovation way, Stillwater, Oklahoma 74074 USA.
17 jessica.morrison@okstate.edu
18
19 Mostafa S. Elshahed, Ph.D.
20 1110 S. Innovation way, Stillwater, Oklahoma 74074 USA.
21 mostafa@okstate.edu
22
23 ***Corresponding author**.
24 Noha H. Youssef, Ph.D.
25 1110 S. Innovation way, Stillwater, Oklahoma 74074 USA.
26 Phone: 1-405-744-1192, Fax. 1-405-744-1112
27 Noha@Okstate.edu
28

30 Keywords: Anaerobic gut fungi, GH39, β-glucosidase, β-xylosidase, β-galactosidase
31 Short Title: Triple oligosaccharide hydrolase activity from a rumen fungus GH39 enzyme

33
34
35
36
37
38

40 **Abstract**

41 **Background.** The anaerobic gut fungi (phylum Neocallimastigomycota) represent a promising

42 source of novel lignocellulolytic enzymes. Here, we report on the cloning, expression, and

43 characterization of a glycoside hydrolase family 39 (GH39) enzyme (Bgxg1) that is highly

44 transcribed by the anaerobic fungus *Orpinomyces* sp. strain C1A under different growth

45 conditions. This represents the first study of a GH39-family enzyme from the anaerobic fungi.

46 **Methods.** Using enzyme activity assays, we performed a biochemical characterization of Bgxg1

47 on a variety of substrates over a wide range of pH and temperature values to identify the optimal

48 enzyme conditions and the specificity of the enzyme. In addition, substrate competition studies

49 and comparative modeling efforts were completed.

50 **Results.** Contrary to the narrow range of activities (β-xylosidase or α-L-iduronidase) observed in

51 previously characterized GH39 enzymes, Bgxg1 is unique in that it is multifunctional, exhibiting

52 strong β-xylosidase, β-glucosidase, β-galactosidase activities (11.5 ± 1.2, 73.4 ± 7.15, and 54.6 ±

53 2.26 U/mg, respectively) and a weak xylanase activity (10.8 ± 1.25 U/mg), as compared to

54 previously characterized enzymes. Further, Bgxg1 possesses extremely high affinity (as evident

55 by the lowest K_m values), compared to all previously characterized β-glucosidases, β-

56 galactosidases, and xylanases. Physiological characterization revealed that Bgxg1 is active over a

57 wide range of pH (3-8, optimum 6) and temperatures (25-60°C, optimum 39°C), and possesses

58 excellent temperature and thermal stability. Substrate competition assays suggest that all

59 observed activities occur at a single active site. Using comparative modeling and bioinformatics

60 approaches, we putatively identified ten amino acid differences between Bgxg1 and previously

61 biochemically characterized GH39 β-xylosidases that we speculate could impact active site

62 architecture, size, charge, and/or polarity.

63   **Discussion.** Collectively, the unique capabilities and multi-functionality of Bgxg1 render it an

64   excellent candidate for inclusion in enzyme cocktails mediating cellulose and hemicellulose

65   saccharification from lignocellulosic biomass.

Introduction

The production of biofuels from lignocellulosic biomass is a global priority, necessitated by the continuous depletion of recoverable fossil fuel reserves, the deleterious impact of fossil fuels on air quality, as well as their contribution to global climate change (Hill et al. 2006; National Research Council 2011; Ragauskas et al. 2006). Lignocellulosic biomass represents a vastly underutilized and largely untapped source of energy, and its mass utilization for biofuel production is one of the goals enacted by the U.S. Congress-implemented Renewable Fuel Standard (RFS), aiming to generate 16 billion gallons of biofuel from lignocellulosic sources by 2022 (National Research Council 2011).

The most frequently used method of biofuel production from lignocellulosic biomass is the enzymatic conversion of cellulose and hemicellulose polymers into sugar monomers/ oligomers that could subsequently be converted into biofuels using dedicated sugar metabolizers (Elshahed 2010; Hill et al. 2006; Kumar et al. 2008). Historically, enzymatic cocktails designed for the breakdown of lignocellulosic biomass focused primarily on cellulose degradation, due to its relative structural simplicity and uniformity across all types of plant biomass. Nevertheless, the hemicellulose components in lignocellulosic biomass should not be ignored, as hemicellulose represents 20-35% of the composition of lignocellulosic biomass (Liu et al. 2008). Unlike cellulose, plant hemicelluloses are structurally more complex, with multiple types of major hemicelluloses (arabinoxylans/ glucuronoarabinoxylans, glucomannans/galactoglucomannans, mixed glucans, and xyloglucans) present in various plants (Scheller & Ulvskov 2010). The most common type of hemicellulose are the arabinoxylans/ glucuronoarabinoxylans that possess a structural backbone of β-1,4-linked xylose units (Scheller & Ulvskov 2010). Xylan degradation requires the consorted action of the endo-acting-β-1,4-xylanases and the oligosaccharide

89   depolymerizing β-xylosidases, among other enzymes (Elshahed 2010; Scheller & Ulvskov

90   2010).

91   The identification and characterization of novel enzymes and enzyme cocktails with

92   superior lignocellulosic biomass saccharification properties (e.g. high substrate affinity and

93   specific activity, activity retention at a wide range of pH and temperatures, and thermal and pH

94   stability) signify essential thrusts in biofuel research. Members of the anaerobic gut fungi

95   (phylum Neocallimastigomycota) represent a promising, and largely untapped, source of

96   biomass-degrading enzymes (Ljungdahl 2008; Wang et al. 2013). Members of the

97   Neocallimastigomycota are found in the herbivorous gut, where they are responsible for the

98   initial colonization and degradation of plant materials ingested by their hosts (Ljungdahl 2008;

99   Wang et al. 2013). While anaerobic gut fungi were initially discovered in sheep, they have since

100 been found in the rumen and alimentary tracks of both ruminant and non-ruminant mammalian

101 and reptilian herbivores (Youssef et al. 2013). The anaerobic gut fungi are excellent biomass

102 degraders, capable of fast, efficient, and simultaneous degradation of the cellulolytic and

103 hemicellulolytic fraction of various plants, including most common lignocellulosic biomass

104 substrates (e.g. Corn Stover, Switchgrass, Sorghum, Energy Cane, and Alfalfa) (Borneman et al.

105 1989; Harhangi et al. 2003; Liggenstoffer et al. 2014; Youssef et al. 2013). Nevertheless, there

106 have been extensive efforts dedicated to bioprospecting novel cellulases and hemicellulases from

107 aerobic fungi (such as *Aspergillus* (Kumar & Ramon 1996; vanPeij et al. 1997), *Trichoderma*

108 (Matsuo & Yasui 1984)), anaerobic prokaryotes (such as *Clostridium* (Bronnenmeier &

109 Staudenbauer 1988) and *Thermoanaerobacterium* (Shao et al. 2011)) and metagenomic sequence

110 data (Brennan et al. 2004; Hess et al. 2011); in comparison to these numerous efforts, the

111   identification, expression, and characterization of such enzymes from anaerobic fungi has not

112   been as well represented in the literature (Borneman et al. 1989; Harhangi et al. 2003).

113   We aim to explore the utility of the anaerobic gut fungus *Orpinomyces* sp. strain C1A

114   (henceforth referred to as C1A) as a novel source of lignocellulolytic enzymes. C1A is an isolate

115   from the feces of an angus steer on cellobiose-switchgrass media (Youssef et al. 2013). Our

116   approach depends on implementing a transcriptomics-guided strategy to identify carbohydrate-

117   active enzymes (CAZyme) transcripts that are highly expressed by C1A when grown on

118   lignocellulosic biomass substrates as candidates for cloning, expression, and characterization.

119   Here, we describe our efforts in cloning, expression, and characterization of one such enzyme: a

120   GH39 transcript bioinformatically annotated as a β-xylosidase, designated Bgxg1, representing

121   the first study of a GH39-family enzyme from anaerobic fungi. Our results document the high

122   affinity, high specific activity, wide pH and temperature ranges, high thermal and pH stability of

123   this enzyme, and novel multiple activities.

124   **Materials and Methods**

125   *Transcriptomics-guided selection of a GH39 enzyme for cloning and characterization.*

126   As a part of an extensive transcriptomic analysis of lignocellulosic biomass degradation by the

127   anaerobic fungal isolate *Orpinomyces* sp. strain C1A (Couger et al. 2015), the most highly

128   transcribed gene annotated as a β-xylosidase was selected for cloning and biochemical

129   characterization. The selected m.21910 transcript (GenBank accession number KT997999) was

130   annotated as member of the GH39 CAZyme family based on the presence of the conserved

131   protein domain pfam01229 (Glyco_hydro_39) family. When strain C1A was grown on different

132   substrates (glucose, Corn Stover, Energy Cane, Switchgrass, and Sorghum), m.21910 constituted

133   58-84% of the transcriptional activity (i.e. normalized FPKM values) of all GH39 transcripts

134 (n=9), and 5.7-18.2% of the transcriptional activities of all C1A genes putatively annotated as β-

135 xylosidases (members of GH39 and GH43, n=41) (Couger et al. 2015). The gene encoding for

136 Bgxg1 protein was previously identified in the genome of strain C1A (GenBank contig accession

137 number ASRE01002650.1, range: 2346-3460, see GCA_000412615.1 for whole genome). The

138 ctg7180000059688.1 gene consists of 1115 bp and no introns (refer to IMG gene ID 2518718918

139 for a visual representation of the gene, https://img.jgi.doe.gov/cgi-

140 bin/m/main.cgi?section=TaxonDetail&page=taxonDetail&taxon_oid=2518645524). The protein

141 product is predicted to be extracellular and non-cellulosomal, based on the presence of a signal

142 peptide, and the absence of a CBM fungal dockerin domain, respectively.

143 *Bgxg1 sequence analysis and phylogeny.*

144 To determine the phylogenetic affiliation of Bgxg1 and the overall topology and global

145 phylogeny of GH39 CAZymes, GH39 β-xylosidase sequences available in CAZY database

146 (http://www.cazy.org/GH39_all.html) (n=1145 total GH39 sequences, retrieved October 28,

147 2015, edited to remove α-iduronidases and duplicates, resulting in n=200 β-xylosidases), in

148 addition to Bgxg1, were aligned using Clustal Omega (Sievers et al. 2011). The generated

149 alignment was used to construct a maximum likelihood tree in RAxML (Stamatakis 2014),

150 which was subsequently visualized and annotated using Mega6 (Sievers et al. 2011; Tamura et

151 al. 2013).

152 *Synthesis, cloning, expression, and purification of Bgxg1 protein.*

153 *bgxg1 gene synthesis and cloning.* A fraction (939 bp, positions 67-1035) of m.21910 transcript

154 was codon optimized for ideal expression in *E. coli* (see Fig. S1 for the alignment of the original

155 gene and codon-optimized gene), and the *bgxg1* insert was synthesized by a commercial provider

156 and inserted into a pET28a(+) plasmid (GenScript, Piscataway, NJ). The plasmid, pET28a(+)-

157 *bgxg1*, harbors kanamycin resistance (*kan*) and *NdeI* and *XhoI* restriction sites for selection and

158 cloning. The pET28a(+)-*bgxg1* plasmid was first transformed into One-Shot Chemically

159 Competent Top10 *E. coli c*ells (Invitrogen, Carlsbad, CA), and the transformants were grown

160 overnight on LB-kanamycin agar (15 µg/mL) for selection. The purified plasmid was

161 electroporated into a protease-deficient BL21(DE3)pLysS *E. coli* strain (Novagen, EMD

162 Millipore, Darmstadt, Germany), possessing an additional chloramphenicol resistance (*cm*)

163 marker, using a single pulse of 1.8 kV in 0.1 cm electrocuvettes. Transformants were grown on

164 LB agar using both kanamycin (15 µg/mL) and chloramphenicol (34 µg/mL) for selection and

165 screened for the presence of correctly sized inserts via colony PCR using T7 forward and reverse

166 primers.

167 *Bgxg1 expression and purification.* Ten milliliters of overnight cultures of BL21(DE3)pLysS *E.*

168 *coli* cells transformed with pET28a(+)-*bgxg1* were used to inoculate 1 L LB broth, containing

169 kanamycin (15 µg/mL) and chloramphenicol (34 µg/mL). The culture was incubated at 37°C

170 with shaking at 200 rpm until an $OD_{600} = 0.6$ was reached. Isopropyl-β-D-thiogalactopyranoside

171 (IPTG, 1mM final concentration) was then added to induce protein production, and the culture

172 was gently shaken at room temperature overnight. Cells were then pelleted by centrifugation

173 (6,000 x g, 10 minutes, 4°C) and the pellets were collected and stored at -20°C.

174   Preliminary small-scale experiments indicated that the protein is expressed in the

175 inclusion body fraction (Fig. S2). Inclusion body extraction was initiated by incubating the

176 cultures in B-Per Cell Lysis Reagent (Thermo Scientific, Grand Island, NY) (10 ml per 500 ml of

177 culture) for 15 minutes at room temperature with gentle shaking to lyse the cells. The

178 homogenate was centrifuged (10,000 x g, 30 minutes, 4°C) and the inclusion body extraction

179 procedure (Grassick et al. 2004) was conducted on the cell pellet as follows: The pellet was

180   resuspended in a urea-based inclusion body extraction buffer (20% glycerol, 8 M urea, 50 mM

181   sodium monobasic phosphate, 500 mM sodium chloride, pH 8.0) for 30 minutes at room

182   temperature with gentle shaking. The homogenate was centrifuged (10,000 x g, 30 minutes, 4°C)

183   and the resultant supernatant containing target inclusion body proteins was subsequently utilized

184   for refolding and purification procedures.

185   Recombinant protein refolding was achieved using slow dialysis as previously described

186   (Grassick et al. 2004). In brief, inclusion body extract was incubated with EDTA (1 mM final

187   concentration) and β-mercaptoethanol (100 mM final concentration) for 2 hours at room

188   temperature with gentle shaking, transferred to dialysis tubing (NMWL: 12,000 – 14,000 Da),

189   and placed for 3 hours into inclusion body exchange buffer (20% glycerol, 8 M urea, 50 mM

190   sodium monobasic phosphate, 500 mM sodium chloride, 1 mM EDTA, pH 8.0) for removal of

191   the β-mercaptoethanol. The buffer was refreshed and dialyzed for an additional 3 hours. The

192   dialysis tubing was then placed into a low-urea refolding buffer (2 M urea, 50 mM sodium

193   monobasic phosphate, 500 mM sodium chloride, 1 mM EDTA, 3 mM reduced glutathione, 0.9

194   mM oxidized glutathione, pH 8.0) and dialyzed overnight, followed by a no-urea refolding buffer

195   (50 mM sodium monobasic phosphate, 500 mM sodium chloride, 1 mM EDTA, 3 mM reduced

196   glutathione, 0.9 mM oxidized glutathione, pH 8.0) for 36 hours.

197   Following dialysis, the contents of the tubing were centrifuged to remove insoluble,

198   precipitated proteins (15,000 x g, 15 minutes, 4°C). The supernatant, containing refolded soluble

199   protein, was then exposed to a nickel-nitriloacetic acid (Ni-NTA, 1:1 ratio) slurry (UBPBio,

200   Aurora, CO), packed in a glass frit column (25 x 200 mm, 98 mL volume Kimble-Chase Kontes

201   Flex Column, Vineland, NJ), and allowed to incubate at 4°C for 1 hour on an orbital shaker.

202   Protein purification followed as detailed previously (Morrison et al. 2012). Samples were

203 concentrated using Amicon Ultra-15 Centrifugal Filter Units (Millipore, NMWL 30 kDa) and

204 protein concentration was determined using a Qubit Fluorimeter (Thermo Scientific) in reference

205 to standard protein concentrations. Protein refolding was checked as activity against PNPX, as

206 described below. An SDS-PAGE gel was run to check protein size and purity, as previously

207 described (Laemmli 1970; Morrison et al. 2012).

208 _Biochemical characterization of Bgxg1 (Enzyme activity assays)._

209 _pH and temperature optima and stability._ The pH range and subsequent pH optimum for Bgxg1

210 was determined by assaying its β-xylosidase activity (described below) at pH 3, 4, 5, 6, 7, 8, 9,

211 and 10, using the following buffer systems: sodium acetate buffer (pH 3.0-6.0), sodium

212 phosphate buffer (pH 7.0-8.0), and glycine buffer (pH 9.0-10). Similarly, the temperature range

213 and subsequent thermal optimum for Bgxg1 was determined by assaying its β-xylosidase activity

214 at 25, 30, 39, 50, and 60°C. In a second and separate study, the stability of Bgxg1 after exposure

215 to pH extremes was determined by assaying its β-xylosidase activity following a one-hour

216 incubation at pH 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, and 13 at 4°C. The following pH buffering systems

217 were used for pH adjustment: sodium acetate buffer (pH 3.0-6.0), sodium phosphate buffer (pH

218 7.0-8.0), glycine buffer (pH 9.0-10), sodium bicarbonate (pH 11.0), and KCl-NaOH (pH 12-13).

219 Similarly, in a separate study, the thermal stability of Bgxg1 was determined by assaying its β-

220 xylosidase activity following a one-hour incubation at 4, 25, 30, 37, 39, 50, 60, and 70ºC. In all

221 cases, 2.2 µg of pure Bgxg1 was used, since this concentration was determined to be optimal in

222 initial testing. Following the one-hour long exposure at the above-described extremes, enzymatic

223 activity was tested at 39°C and pH 6.0, the optimal conditions as determined for this enzyme. All

224 experiments were completed in triplicate, and relative specific activities in relation to the best

225 performing condition (100% activity) were reported.

226   *Enzyme activity assays.* All enzyme assays with Bgxg1 were conducted in pH 6.0 buffer and at

227   39°C, as these conditions were determined to be optimal for Bgxg1. All reagents were purchased

228   from Sigma Aldrich (St. Louis, MO) unless noted otherwise.

229   Endoglucanase, exoglucanase, xylanase, and mannanase activities were determined using

230   a DNS (3,5-dinitrosalicyclic acid)-based assay (Breuil & Saddler 1985), with carboxymethyl

231   cellulose sodium salt (CMC, 1.25% w/v), avicel microcrystalline cellulose (1.25% w/v),

232   beechwood xylan (1.25% w/v), and locust bean gum (0.5% w/v) as substrates, respectively.

233   Glucose, xylose, or mannose were utilized for the generation of a standard curve, dependent on

234   the substrate being tested.

235   Cellobiohydrolase, β-xylosidase, arabinosidase, mannosidase, β-glucosidase, β-

236   galactosidase, and acetyl xylan esterase activities were determined using (10mM) of the *p*-

237   nitrophenol-based (PNP) substrates: *p*-nitrophenyl-β-D-cellobioside (PNPC,), *p*-nitrophenyl-β-D-

238   xylopyranoside (PNPX), *p*-nitrophenyl-β-D-arabinofuranoside (PNPA) *p*-nitrophenyl-β-D-

239   mannoside (PNPM), *p*-nitrophenyl-β-D-glucopyranoside (PNPG), *p*-nitrophenyl-β-D-

240   galactopyranoside (PNPGal), and *p*-nitrophenyl-acetate (PNPAc), respectively (Dashtban et al.

241   2010; Kubicek 1982; Zhang et al. 2009). Assays were conducted in sodium acetate buffer with

242   sodium carbonate (1M) as a stop reagent. The release of PNP was measured colorimetrically at

243   420 nm, following the addition of the stop solution. α-glucuronidase activity was assayed using

244   the Megazyme α-glucuronidase assay kit (Wicklow, Ireland).

245   All experiments were conducted in triplicate. One unit of enzymatic activity (U) was

246   defined as one µmol of products (reducing sugar equivalents in DNS assays, PNP released in

247   PNP substrate-based assays, and aldouronic acid in α-glucuronidase assay) released from the

248   substrate per minute. Specific activity was calculated by determining the units released per mg of

249   enzyme.

250   *Enzyme kinetics.* Standard procedures were used to determine the K_m, V_{max}, and specific activity

251   of Bgxg1 on all substrates described above (Lineweaver & Burk 1934). K_m and V_{max} values were

252   obtained using double-reciprocal Lineweaver-Burke plots, which were used to extrapolate from

253   experimentally-derived values using a constant protein concentration (2.2 µg) and variable PNP-

254   based substrate concentration (0.1 – 100 mM) (Lineweaver & Burk 1934). Given the extinction

255   coefficient of p-nitrophenol (PNP) is 17/mM/cm at 400 nm (Bessey & Love 1952), for a 1 cm

256   path length cuvette and absorbance minimum of 0.010, reliable K_m detection limits in such PNP-

257   based spectrophotometric assays is ≈ 500 nM. Therefore, K_m values < 500 nM are referred to as

258   BDL (below detection limit).

259   *Substrate competition assays.* Competitive inhibition experiments were conducted to determine

260   whether the observed multiple oligosaccharide hydrolase activities are catalyzed via a single or

261   multiple active sites. In such experiments, the effect of cellobiose (as a competitive inhibitor) on

262   the β-xylosidase activity of Bgxg1 was measured by conducting the β-xylosidase assay, using 10

263   mM of PNPX as the substrate, in the presence of different concentrations of cellobiose (0, 10,

264   and 20 mM) and evaluating the impact of cellobiose presence on the release of PNP. Conversely,

265   the effect of xylobiose (as a competitive inhibitor) on the β-glucosidase activity of Bgxg1 was

266   measured by conducting the β-glucosidase assay (using 10 mM of PNPG as the substrate) in the

267   presence of different concentrations of xylobiose (0, 10, and 20 mM), and evaluating the impact

268   of xylobiose presence on the release of PNP. In both experiments, the effect of inhibitor

269   concentration on K_m and V_{max} was evaluated using Lineweaver-Burke plots (Lineweaver & Burk

270   1934). All experiments were conducted in triplicate.

271   Substrate preferences of Bgxg1 were determined by conducting a substrate competition

272   assay, where Bgxg1 (2.2 µg of pure enzyme preparation) was challenged by a mixture of

273   xylobiose (10 mM) and cellobiose (10 mM). The kinetics of xylose and glucose release were

274   compared to the results obtained in control experiments where only one substrate (xylobiose or

275   cellobiose) was utilized. Samples were taken at 0, 1, 5, 10, 15, 30, and 60 minutes for the

276   determination of the glucose and xylose concentrations. Glucose was assayed using PGO

277   Enzyme Preparation Capsules (Sigma-Aldrich, St. Louis, MO) and xylose was assayed using

278   Megazyme Xylose Kit (Wicklow, Ireland). All experiments were conducted in triplicate.

279   *Bgxg1 modeling.*

280   Homology modeling by Iterative Threading ASSEmbly Refinement (I-TASSER) (Roy et al.

281   2010; Yang et al. 2015; Zhang 2008), was conducted to generate a three-dimensional model of

282   Bgxg1 using *Thermoanaerobacterium saccharolyticum* β-xylosidase (PBD entry 1UHV) as a

283   template. PyMOL was used to align the Bgxg1 structural prediction to that of

284   *Thermoanaerobacterium saccharolyticum* (PBD entry 1UHV) to examine and speculate the

285   impact of variations in amino acids residue on the enzyme's active site topology and putative

286   substrate binding capacities (PyMol).

287   **Results**

288   *Bgxg1 phylogenetic affiliation.*

289   Phylogenetic analysis grouped all GH39 sequences into 4 phylogenetically-resolved and

290   bootstrap-supported clades (Classes I-IV in Fig. 1). *Orpinomyces* sp. strain C1A Bgxg1 protein

291   belonged to Class III, forming a well-supported cluster with GH39 proteins from the anaerobic

292   fungus *Piromyces* sp. strain E2, as well as GH39 proteins from the bacterial genera *Clostridium*

293   and *Teredinibacter* (70-74% sequence identities) (Fig. 1). To our knowledge, none of the GH39

294 proteins within this specific cluster, or in the entire Class III GH39, has been biochemically

295 characterized.

296 *Physiological characterization.*

297 SDS-PAGE results show that the Bgxg1 protein is consistent with the predicted size of 42.7 kDa

298 (protein predicted molecular weight is 39.6 KDa + 0.996 kDa linker + 2.101 kDa double

299 histidine tag) (Fig. S3).

300   The thermal and pH ranges and optima were determined by conducting assays at a range

301 of temperatures and pH's, as described above. Bgxg1 exhibited activity in a wide range of pH (3-

302 8) and temperatures (25-60°C), with optimal activity at pH 6 and 39°C (Fig. 2A, 2B). The

303 thermal and pH stabilities of Bgxg1 were examined by conducting activity assays post-stress (pH

304 or thermal)-incubations as described above. Bgxg1 retained more than 80% of its specific

305 activity post-application of pH stress ranging between 6 and 11 (Fig. 2C), and 60% of its specific

306 activity post application of pH stress of 4, 5, and 12 (Fig. 2C). Further, Bgxg1 retained ≥70% of

307 its specific activity across the broad range of temperature stressors applied (4 – 70°C) (Fig. 2D).

308 In addition, exposure to pH stress from 6-11 and temperature stress from 4-70 °C did not

309 produce results that were significantly different from the optimal conditions (p-value >0.05, 95%

310 confidence interval, Fig. 2).

311 *Substrate specificities and kinetics.*

312 To date, all characterized GH39 enzymes exhibit a narrow substrate range (β-xylosidase or α-L-

313 iduronidase) (Table S1). As predicted by sequence analysis, Bgxg1 exhibited β-xylosidase

314 activity (11.5 ± 1.2 U/mg, Table 1), strong as compared to previously reported β-xylosidase

315 activities (Table S1). While having a β-xylosidase activity greater than the majority of the field

316 in Table S1, Bgxg1 does have lower β-xylosidase activity than other anaerobic fungi

317   *Neocallimastix frontalis* (16 U/mg, (Hebraud & Fevre 1990)), *Neocallimastix patriciarum* (30.4

318   U/mg, (Zhu et al. 1994)), *Piromyces communis* (28 U/mg, (Hebraud & Fevre 1988)), and

319   *Sphaeomonas communis* (27 U/mg, (Hebraud & Fevre 1988) (Table 1, S1), however, these are

320   not of the same GH-family. Interestingly, in addition to β-xylosidase activity, Bgxg1 also

321   exhibited strong β-glucosidase (73.4 ± 7.15 U/mg), β-galactosidase (54.6 ± 2.26 U/mg), and

322   weak xylanase (10.8 ± 1.25 U/mg) activities (Table 1), as compared to reported activities from

323   previously characterized enzymes (Tables S2-S4). Our extensive literature review identified 63

324   enzymes that have been biochemically-characterized to have β-glucosidase activity and, of these,

325   only seven have a reported specific activity higher than that of Bgxg1 (Table S2), including

326   *Candida peltata* (108 U/mg, (Saha & Bothast 1996)), *Thermoascus aurantiacus* (232 U/mg,

327   (Tong et al. 1980)), and *Trichoderma reesei* (768 U/mg, (Takashima et al. 1998)). Similarly, we

328   only identified three β-galactosidase with a reported higher activity than Bgxg1 (Table S3),

329   including *Alicyclobacillus acidocaldarius* (229 U/mg, (Yuan et al. 2008)), *Bifidobacterium*

330   *adolescentis* (526 U/mg, (Hinz et al. 2004)), and *Thermus aquaticus* (1750 U/mg, (Ulrich et al.

331   1972)). On the other hand, the xylanase activity of Bgxg1 is relatively weak, with many

332   previously reported xylanases exhibiting a much higher specific activity (Table S4), such as

333   *Aspergillus niger* (19 U/mg, 25 U/mg, 35 U/mg, and 48 U/mg, (John et al. 1979)) and

334   *Trichoderma reesei* (46 U/mg, (Takashima et al. 1998)). Bgxg1 exhibited no detectable

335   exoglucanase, endoglucanase, mannanase, arabinosidase, acetyl xylan esterase,

336   cellobiohydrolase, mannosidase, or α-glucuronidase activities.

337   In addition to its high β-xylosidase, β-glucosidase, and β-galactosidase specific activities,

338   Bgxg1 exhibited remarkably high affinities towards all examined substrates, with K_m values

339 (calculated via extrapolation through Lineweaver-Burke plot) in the low nM range for PNPG and

340 PNPGal, the low µM range for PNPX (Table 2, Table S1-S4).

341 *Substrate competition studies.*

342 Substrate competition studies were conducted using a variable concentration of an unlabeled

343 substrate (acting as an inhibitor) and a fixed concentration of a chromophore (PNP-based)

344 substrate (Table 3). The results strongly suggest the occurrence of cross-substrate competitive

345 inhibition between xylobiose and cellobiose (Table 3), since the presence of increasing

346 concentrations of a single substrate lowers the specific activity and increases the K_m of the

347 enzyme towards the other substrate, whilst not affecting its V_{max} (K_m and V_{max} calculated via

348 extrapolation through Lineweaver-Burke plot). This pattern strongly indicates that a single active

349 site is responsible for the observed activities (Table 3), a conclusion that is in agreement with the

350 lack of identifiable additional domains other than pfam01229 in Bgxg1, as well as with the

351 structural modeling data described below.

352 In single substrate assays, Bgxg1 was capable of converting cellobiose to glucose and

353 xylobiose to xylose at a very fast rate (Fig. 3A, 3B). This reaction occurs more quickly for

354 xylobiose, as a stable maximal xylose concentration is reached after only 1 minute of incubation

355 (Fig. 3B), compared to 15 minutes for glucose release from cellobiose (Fig. 3A). However, the

356 extent of sugar release at the conclusion of the experiment was higher in cellobiose incubations

357 (Fig. 3A) than xylobiose incubations (Fig. 3B). Competition studies using equimolar

358 concentrations of both substrates revealed the preference of Bgxg1 for xylobiose, since a higher

359 proportion of xylose rather than glucose was detected within the first 15 minutes of the

360 incubation (Fig. 3C). Nevertheless, the final concentrations of sugars released after 60 minutes of

361 incubation did not differ when comparing single substrate versus competition experiments (Fig.

362   3A-3C). Similar to the patterns observed in single substrate assays, Bgxg1 reduced a larger

363   amount of cellobiose to glucose than xylobiose to xylose in competition experiments (Fig. 3C),

364   which is consistent with the higher affinity (lower K_m value) of Bgxg1 for PNPG (12.5 nM) over

365   PNPX (4.85 µM) (Table 2).

366   *Structure activity predictions.*

367   The Bgxg1 protein sequence was submitted to I-TASSER for structural prediction by Iterative

368   Threading ASSEmbly Refinement (Roy et al. 2010; Yang et al. 2015; Zhang 2008) utilizing the

369   β-xylosidase originating from *Thermoanaerobacterium saccharolyticum* (PBD entry 1UHV) as a

370   template for model creation (Roy et al. 2010; Yang et al. 2004; Yang et al. 2015; Zhang 2008).

371   Bgxg1 is predicted to have three distinct domains: a catalytic $(\alpha/\beta)_8$ barrel fold domain (position

372   26-307), a small α -helical domain (position 1-25), and a β sandwich domain (position 308-344)

373   (Fig. S4). Overall, the structure is predicted to contain 11 β-sheets (8 in $(\alpha/\beta)_8$-barrel, 3 in β-

374   sandwich), and 10 α-helices (8 in $(\alpha/\beta)_8$-barrel, 2 in α-domain). The catalytic $(\alpha/\beta)_8$-barrel fold

375   domain is predicted to consist of eight parallel β-sheets (β1-β8), and eight parallel α-helices (α1-

376   α8). Consistent with β-xylosidases of *Thermoanaerobacterium saccharolyticum (*1UHV) and

377   *Geobacillus stearothermophilus* (1PX8), the active site pocket of Bgxg1 is predicted to be

378   located on the upper side of the $(\alpha/\beta)_8$-barrel (Czjzek et al. 2005; Yang et al. 2004) (Fig. S4).

379   Alignment and structural predictions identified the conservation of the general acid-base active

380   site residue Glu127 in the C-terminal of β3, as part of the GH39-conserved Asn126-Glu127-

381   Pro128 motif as well as the nucleophilic residue Glu225 in β6 (Fig. 4) (Czjzek et al. 2005; Yang

382   et al. 2004).

383   Using the predicted model we sought to infer structural differences potentially

384   responsible for the observed relaxed substrate specificities in Bgxg1 by investigating the amino

385   acid conservation patterns between Bgxg1 and all structurally and/or biochemically-

386   characterized β-xylosidases. These enzymes are: *Thermoanaerobacterium saccharolyticum* β-

387   xylosidase (Yang et al. 2004), *Geobacillus stearothermophilus* β-xylosidase (Bhalla et al. 2014;

388   Czjzek et al. 2005), and *Bacillus halodurans* C-125 protein BH1068 (Wagschal et al. 2008), all

389   of which belong to Class II (Fig. 1), as well as *Caulobacter crescentus* CcXynB2 (Correa et al.

390   2012), which belongs to Class I (Fig. 1). All of these enzymes have previously been reported to

391   possess β-xylosidase activity (Bhalla et al. 2014; Correa et al. 2012; Czjzek et al. 2005;

392   Wagschal et al. 2008; Yang et al. 2004). We focused on 25 amino acids in two groups: (i) those

393   previously shown to be important for β-xylosidase activity (Czjzek et al. 2005; Yang et al. 2004)

394   [this group includes (in addition to the conserved general acid-base and nucleophilic active sites

395   described above) amino acids providing the tight hydrogen bonding necessary to stabilize the

396   xylosyl-enzyme intermediate formed during the reaction, such as Arg52, His54, Asn159, His228,

397   Tyr230, Glu278, Trp315, Glu322, and Glu323 (locations refer to position in 1UHV)], as well as

398   (ii) those physically interacting with the active site as deduced by the predicted Bgxg1 model

399   (Fig. S3A) [this group includes Val46, Val81, Ile124, Trp125, Gly130, Thr131, Trp132, Phe139,

400   Pro162, Cys163, Tyr164, Ser165, Lys171, His192, Asn242, and Lys247 (locations refer to

401   position in Bgxg1)]. Of these 25 amino acids, 15 differed between Bgxg1 and the four other

402   proteins. Five of these 15 amino acids were not conserved amongst any of the five sequences

403   studied and so were not further investigated (Fig. 4). Therefore, 10 distinct differences (8

404   substitutions and 2 deletions) between Bgxg1 on one hand and the four biochemically-

405   characterized β-xylosidases on the other were identified (Table 4). These differences that are

406   predicted to exist in or around the active site of Bgxg1 would putatively impact the size, charge,

407   and/or polarity within the active site (Table 4, Fig. S4).

408   The expanded substrate specificity observed in this study could be a unique trait in

409   Bgxg1, or it could be specific to all GH39 CAZymes of anaerobic fungi (e.g. Class III-C), or to

410   the entire Class III β-xylosidases. Based on the above speculations about the amino acids

411   potentially responsible for Bgxg1 relaxed specificity, we further investigated the conservation of

412   these 10 amino acid changes (Table 4) within class III of GH39 proteins. Bgxg1 (as well as other

413   GH39 proteins encoded in C1A genome), all three GH39 proteins from the *Piromyces* genome

414   (accession numbers shown in Fig. 1), and all additional sequences from Class III-C belonging to

415   the genera *Clostridium* and *Teredinibacter* were found to encode 9 of the 10 observed amino

416   acid substitutions (Table 4). However, within the broader Class III, little similarity in key amino

417   acids was observed between Bgxg1 sequences and β-xylosidases belonging to Class III-A, III-B,

418   or III-D (Table 4). Collectively, these results putatively suggest that the observed relaxed

419   specificity in Bgxg1 could be exclusive to Class III-C β-xylosidases.

420   **Discussion**

421   In this study, we used a transcriptomics-guided approach to identify, clone, express, and

422   characterize a GH39 protein (Bgxg1) from the anaerobic gut fungus *Orpinomyces* sp. strain

423   C1A. Our results demonstrate that the expressed protein is multifunctional, possessing strong β –

424   xylosidase (11.5 U/mg), β –glucosidase (73.4 U/mg), and β -galactosidase (54.6 U/mg) activities,

425   as well as a weak xylanase activity (10.8 U/mg) (Table 1, 2), as compared to previously

426   characterized enzymes (Tables S1-S4). This novel multi-functionality has not been previously

427   reported in GH39 enzymes (Bhalla et al. 2014), and therefore this work expands on the known

428   activities of GH39 CAZyme family. Further, Bgxg1 retains high levels of activity over a wide

429   range of temperatures (>80% of activity retained between 4-70°C) (Fig. 2D) and pH values

430   (>80% of activity retained between pH 6-11) (Fig. 2C). Though the composition of commercial

431   enzymes cocktails are largely proprietary, the presence of 80-200 different components within a

432   mixture has been previously reported (Banerjee et al. 2010; Van Dyk & Pletschke 2012). It is

433   intuitive to think that the inclusion of such a large number of enzymes represents a large

434   contribution to the cost of production. It is here that Bgxg1 would be beneficial, as the inclusion

435   of a single enzyme, possessing multiple strong activities, would lower the cost of production in

436   biorefineries and therefore would be beneficial to the bottom line.

437   In addition to its relaxed substrate specificity, the enzyme displays strong kinetic

438   properties (high specific activity and affinity) towards its multiple substrates (K_m and V_{max} values

439   calculated via extrapolation through Lineweaver-Burke plot). As a β-xylosidase, Bgxg1 has one

440   of the highest β-xylosidase specific activity among all reported ambient (<50^{0}C) β-xylosidases,

441   but lower than other anaerobic fungi *Neocallimastix frontalis* (16 U/mg, (Hebraud & Fevre

442   1990)), *Neocallimastix patriciarum* (30.4 U/mg, (Zhu et al. 1994)), *Piromyces communis* (28

443   U/mg, (Hebraud & Fevre 1988)), and *Sphaeomonas communis* (27 U/mg, (Hebraud & Fevre

444   1988) (Table 1, S1). Bgxg1 also one of the highest specific activities amongst known GH39 β-

445   xylosidases (Tables 1, S1), with a lower specific activity than the thermophilic

446   *Thermoanaerobacterium saccharolyticum* (53.8 U/mg, (Shao et al. 2011)) and the thermophilic

447   *Geobacillus stearothermophilus* (133 U/mg, (Bhalla et al. 2014)). Compared to other

448   characterized β-glucosidases, Bgxg1 has the highest specific activity for all ambient temperature

449   β-glucosidases, and one of the highest reported specific activities among all β-glucosidase

450   (members of GH1, GH3, GH5, GH9, and GH30 (Cairns & Esen 2010)), regardless of optimal

451   temperature and GH affiliation (Tables 1, S2). Finally, compared to other characterized β-

452   galactosidases, Bgxg1 has the highest specific activity for all ambient temperature β-

453   galactosidases, and one of the highest reported specific activities among all β-galactosidases

454   (members of GH1, GH2, GH35, and GH42 (Skalova et al. 2005)) regardless of optimal

455   temperature and GH affiliation (Tables 1, 2, S4).

456   We reason that the observed kinetics and substrate specificity of Bgxg1 are beneficial for

457   strain C1A and are highly desirable for a saccharolytic enzyme acting within the highly

458   competitive rumen environment, where strain C1A originally existed (*Orpinomyces* sp. strain

459   C1A was isolated from the feces of an angus steer (Youssef et al. 2013)). The high specific

460   activity and high substrate affinity may aid in fast and efficient scavenging of sugars from the

461   surrounding environment, where competition for sugars/oligosaccharide produced by

462   saccharolytic enzymes are intense, and where free sugar levels are permanently low (Garcia-

463   Vallve et al. 2000). We hence speculate that the survival in an anaerobic, eutrophic, and highly

464   competitive environment might be responsible for the acquisition, retention and directed

465   evolution of anaerobic fungal β-xylosidases towards superior kinetics and relaxed specificities.

466   Sequence analysis and structural predictive modeling (Figs. 4 and S4), and substrate

467   competition experiments (Table 3) predict the presence of a single conserved active site within

468   the $(\alpha/\beta)_8$-barrel fold structure typically observed in GH39-family enzymes (Czjzek et al. 2005;

469   Yang et al. 2004) (with the conserved catalytic nucleophile (Glu225) and general acid-base

470   residue (Glu127)) and potentially mediating all observed hydrolytic activities). To provide clues

471   regarding the structural basis of the observed multi-functionality, comparison of amino acid

472   conservation patterns putatively affecting the active site topology between Bgxg1 and

473   biochemically characterized GH39 xylosidases, all four of which display no additional activities

474   beyond β-xylosidase, was undertaken. We identified ten different distinct amino acid changes (8

475   substitutions and 2 deletions) (Table 4, Figs. S4 and 5) in Bgxg1 that putatively affect the

476   polarity (Tyr vs. Val46, Phe vs. Thr131, Tyr vs. Phe139, Ala vs. Cys163, Trp vs. Lys171, Tyr vs.

477 Leu194, and Ala vs. Arg242), constitute significant size changes (Tyr vs. Val46, Phe vs. Thr131,

478 Tyr vs. Leu194, and Ala vs. Arg242), result in the addition of charged moieties or unique

479 functional groups (Asn vs. Asp129, Ala vs. Cys163, Trp vs. Lys171, and Ala vs. Arg242), or

480 result in the deletion of a negatively charged residue, previously determined to be important

481 (Glu322-323 vs. deletion) to the active site (Czjzek et al. 2005). The impact of these speculated

482 changes is unclear, and it remains to be seen if any, all, or a combination of the above differences

483 is responsible for the observed relaxed specificity. However, while all these amino acid changes

484 are speculated to theoretically explain the relaxed specificity of Bgxg1, one such difference is

485 peculiar and deserves special scrutiny; deletions/gaps in the Bgxg1 sequence as opposed to

486 negatively charged glutamic acids in the other four sequences (Table 4, Fig. S4S). GH39

487 enzymes belong to the wider family of β-1,4-retaining hydrolases of clan GH-A e.g. GH1 β-

488 glucosidase and GH5 cellulases. Differences in structure between β 1,4-glucose cleaving

489 enzymes and β 1,4-xylose cleaving enzymes within clan GH-A have been extensively

490 investigated (Czjzek et al. 2005; Czjzek et al. 2001; Ducros et al. 1995; Hovel et al. 2003;

491 Verdoucq et al. 2004). Such studies have demonstrated that, within the active site of β 1,4-

492 glucose cleaving enzymes, a Gln residue (corresponding to position 39 in the enzyme dhurinase

493 of *Sorghum bicolor* (Czjzek et al. 2005; Ducros et al. 1995; Verdoucq et al. 2004)) interacts with

494 the substrate by forming a hydrogen bond with O3 and O4 of the glucose moiety (Czjzek et al.

495 2005; Ducros et al. 1995). On the other hand, β 1,4-xylosidases acting on C5 sugar dimers

496 contain a Glu residue in lieu of Gln (at position 322-323 in *Thermoanaerobacterium*

497 *saccharolyticum,* Fig. 4 and S4, Table 4) that binds to O3 and O4 of the xylose moiety (Czjzek et

498 al. 2005). Interestingly, these Glu residues are aligned with a gap in the sequence of the

499 multifunctional Bgxg1 (Fig. 4), with no apparent occurrence of either Glu or Gln amino acids

500 within the vicinity. Structurally predictive modeling suggests that in lieu of these Glu322-323

501 residues (1UHV numbering) Bgxg1 is predicted to possess Gly-Arg at an approximately

502 sterically-similar location near the active site (Figure S4R-S), representing a significant change

503 from two negatively-charged residues, to an uncharged and positively-charged pair of residues.

504 Since the Glu residues in biochemically characterized β-xylosidases are shown to be important

505 for stabilizing intermediates (Czjzek et al. 2005), the predicted absence of these residues in

506 Bgxg1 and their speculated replacement with Gly-Arg suggests that Bgxg1 might employ a

507 different mechanism for stabilizing its intermediates during the catalytic process; however, this

508 speculation will require further investigation.

509   The ecological relevance, global distribution, and evolutionary patterns of multi-

510 functionality within GH39 β-xylosidases remain to be conclusively determined. Phylogenetic

511 analysis demonstrated the occurrence of nine out of ten amino acids substitutions/deletions in all

512 sequenced members of Class III-C, residues which we speculate to be of importance to the

513 observed multi-functionality of Bgxg1, but as Bgxg1 is the only biochemically-characterized

514 enzyme within Class III, this analysis is purely speculative (Table 4, alignment in Fig. S5). In

515 addition to anaerobic fungal sequences, Class III-C β-xylosidases contain sequences from the

516 genera *Clostridium* and *Teredinibacter* (Fig. 1). Since it has been previously demonstrated that

517 the xylanolytic machinery in anaerobic fungi, including β-xylosidases, has been acquired from

518 bacteria via horizontal gene transfer (Youssef et al. 2013), and speculating that some or all of the

519 amino acids substitutions/deletions in members of class III-C collectively account for the

520 observed multi-functionality (though it is unknown, at this time, whether these GH39 enzymes

521 possess this multi-functionality), therefore we reason that the observed distribution pattern

522 suggests the evolution of relaxed specificity in GH39 β-xylosidases within the domain Bacteria,

523   prior to the acquisition of GH39 β-xylosidases by the anaerobic fungi and that the acquired

524   capability is speculated to be retained in all anaerobic fungal GH39 β-xylosidases.

525   **Conclusions**

526   In conclusion, we have characterized a novel β-xylosidase that represents the first GH39-family

527   enzyme cloned and expressed from anaerobic fungi. The enzyme is multi-functional, capable of

528   hydrolyzing cellobiose, xylobiose, as well as several PNP-glycosides. It also displays high

529   affinity towards various substrates, retains activity over a wide range of temperatures and pHs,

530   and possesses excellent temperature and thermal stability. Structurally predictive modeling

531   identified putative differences which potentially could account for the observed relaxed

532   specificity. Collectively, these capabilities render Bgxg1 an excellent candidate for inclusion in

533   enzyme cocktails mediating cellulose and hemicellulose saccharification from lignocellulosic

534   biomass (Morrison et al. 2016).

535   **Acknowledgements**

536   We thank Dr. Gilbert John (Oklahoma State University) for supplying the *E. coli*

537   BL21(DE3)pLysS cells used in this study. We also thank Dr. Robert Gruninger for helpful

538   discussions.

539

540 **References**

541 Banerjee G, Scott-Craig JS, and Walton JD. 2010. Improving Enzymes for Biomass Conversion:

542   A Basic Research Perspective. *Bioenergy Research* 3:82-92. 10.1007/s12155-009-9067-5

543 Bessey OA, and Love RH. 1952. Preparation and measurement of the purity of the phosphatase

544   reagent, disodium para-nitrophenyl phosphate. *J Biol Chem* 196:175-178.

545 Bhalla A, Bischoff KM, and Sani RK. 2014. Highly thermostable GH39 beta-xylosidase from a

546   *Geobacillus* sp strain WSUCF1. *BMC Biotechnol* 14:963-973. Artn 963 10.1186/S12896-

547   014-0106-8

548 Borneman WS, Akin DE, and Ljungdahl LG. 1989. Fermentation products and plant-cell wall-

549   degrading enzymes produced by monocentric and polycentric anaerobic ruminal fungi.

550   *Appl Environ Microbiol* 55:1066-1073.

551 Brennan Y, Callen WN, Christoffersen L, Dupree P, Goubet F, Healey S, Hernandez M, Keller

552   M, Li K, Palackal N, Sittenfeld A, Tamayo G, Wells S, Hazlewood GP, Mathur EJ, Short

553   JM, Robertson DE, and Steer BA. 2004. Unusual microbial xylanases from insect guts.

554   *Appl Environ Microbiol* 70:3609-3617. Doi 10.1128/Aem.70.6.3609-3617.2004

555 Breuil C, and Saddler JN. 1985. Comparison of the 3,5-dinitrosalicylic acid and Nelson-Somogyi

556   methods of assaying for reducing sugars and determining cellulase activity. *Enzyme*

557   *Microb Tech* 7:327-332. Doi 10.1016/0141-0229(85)90111-5

558 Bronnenmeier K, and Staudenbauer WL. 1988. Purification and properties of an extracellular

559   beta-glucosidase from the cellulolytic thermophile *Clostridium stercorarium*. *Appl*

560   *Microbiol Biot* 28:380-386.

561 Cairns JRK, and Esen A. 2010. Beta-glucosidases. *Cell Mol Life Sci* 67:3389-3405.

562   10.1007/s00018-010-0399-2

563   Correa JM, Graciano L, Abrahao J, Loth EA, Gandra RF, Kadowaki MK, Henn C, and Simao

564   RDG. 2012. Expression and characterization of a GH39 beta-xylosidase II from

565   *Caulobacter crescentus*. *Appl Biochem Biotech* 168:2218-2229. 10.1007/s12010-012-

566   9931-1

567   Couger MB, Youssef NH, Struchtemeyer CG, Liggenstoffer AS, and Elshahed MS. 2015.

568   Transcriptomics analysis of lignocellulosic biomass degradation by the anaerobic fungal

569   isolate *Orpinomyces* sp. strain C1A. *Biotechnol Biofuels* 8:208-224.

570   Czjzek M, Ben David A, Braman T, Shoham G, Henrissat B, and Shoham Y. 2005. Enzyme-

571   substrate complex structures of a GH39 beta-xylosidase from *Geobacillus*

572   *stearothermophilus*. *J Mol Biol* 353:838-846. 10.1016/j.jmb.2005.09.003

573   Czjzek M, Cicek M, Zamboni V, Burmeister WP, Bevan DR, Henrissat B, and Esen A. 2001.

574   Crystal structure of a monocotyledon (maize ZMGlu1) beta-glucosidase and a model of

575   its complex with p-nitrophenyl beta-D-thioglucoside. *Biochem J* 354:37-46. Doi

576   10.1042/0264-6021:3540037

577   Dashtban M, Maki M, Leung KT, Mao CQ, and Qin WS. 2010. Cellulase activities in biomass

578   conversion: measurement methods and comparison. *Crit Rev Biotechnol* 30:302-309.

579   10.3109/07388551.2010.490938

580   Ducros V, Czjzek M, Belaich A, Gaudin C, Fierobe HP, Belaich LP, Davies GJ, and Haser R.

581   1995. Crystal structure of the catalytic domain of a bacterial cellulase belonging to family

582   5. *Structure* 3:939-949. Doi 10.1016/S0969-2126(01)00228-3

583   Elshahed MS. 2010. Microbiological aspects of biofuel production: current status and future

584   directions. *J Adv Res* 1:103-111.

585   Garcia-Vallve S, Romeu A, and Palau J. 2000. Horizontal gene transfer of glycosyl hydrolases of

586   the rumen fungi. *Mol Biol Evol* 17:352-361.

587   Grassick A, Murray PG, Thompson R, Collins CM, Byrnes L, Birrane G, Higgins TM, and

588   Tuohy MG. 2004. Three-dimensional structure of a thermostable native

589   cellobiohydrolase, CBHIB, and molecular characterization of the cel7 gene from the

590   filamentous fungus, *Talaromyces emersonii*. *Eur J Biochem* 271:4495-4506.

591   10.1111/j.1432-1033.2004.04409.x

592   Harhangi HR, Freelove AC, Ubhayasekera W, van Dinther M, Steenbakkers PJ, Akhmanova A,

593   van der Drift C, Jetten MS, Mowbray SL, Gilbert HJ, and Op den Camp HJ. 2003.

594   Cel6A, a major exoglucanase from the cellulosome of the anaerobic fungi *Piromyces* sp.

595   E2 and *Piromyces equi*. *Biochim Biophys Acta* 1628:30-39.

596   Hebraud M, and Fevre M. 1988. Characterization of glycoside and polysaccharide hydrolases

597   secreted by the rumen anaerobic fungi *Neocallimastix frontalis*, *Sphaeromonas communis*

598   and *Piromonas communis*. *J Gen Microbiol* 134:1123-1129.

599   Hebraud M, and Fevre M. 1990. Purification and characterization of an extracellular beta-

600   xylosidase from the rumen anaerobic fungus *Neocallimastix frontalis*. *FEMS Microbiol*

601   *Lett* 72:11-16. DOI 10.1111/j.1574-6968.1990.tb03853.x

602   Hess M, Sczyrba A, Egan R, Kim TW, Chokhawala H, Schroth G, Luo SJ, Clark DS, Chen F,

603   Zhang T, Mackie RI, Pennacchio LA, Tringe SG, Visel A, Woyke T, Wang Z, and Rubin

604   EM. 2011. Metagenomic discovery of biomass-degrading genes and genomes from cow

605   rumen. *Science* 331:463-467. 10.1126/science.1200387

606 Hill J, Nelson E, Tilman D, Polasky S, and Tiffany D. 2006. Environmental, economic, and
607   energetic costs and benefits of biodiesel and ethanol biofuels. *P Natl Acad Sci USA*
608   103:11206-11210. 10.1073/pnas.0604600103

609 Hinz SWA, van den Broek LAM, Beldman G, Vincken JP, and Voragen AGJ. 2004. Beta-
610   galactosidase from *Bifidobacterium adolescentis* DSM20083 prefers beta(1,4)-
611   galactosides over lactose. *Appl Microbiol Biot* 66:276-284. 10.1007/s00253-004-1745-9

612 Hovel K, Shallom D, Niefind K, Belakhov V, Shoham G, Baasov T, Shoham Y, and Schomburg
613   D. 2003. Crystal structure and snapshots along the reaction pathway of a family 51 alpha-
614   L-arabinofuranosidase. *EMBO J* 22:4922-4932. Doi 10.1093/Emboj/Cdg494

615 John M, Schmidt B, and Schmidt J. 1979. Purification and some properties of five endo-1,4-beta-
616   D-xylanases and a beta-D-xylosidase produced by a strain of *Aspergillus niger*. *Can J*
617   *Biochem Cell B* 57:125-134.

618 Kubicek CP. 1982. Beta-glucosidase excretion by *Trichoderma pseudokoningii* - correlation with
619   cell-wall bound beta-1.3-glucanase activities. *Arch Microbiol* 132:349-354. Doi
620   10.1007/Bf00413388

621 Kumar R, Singh S, and Singh OV. 2008. Bioconversion of lignocellulosic biomass: biochemical
622   and molecular perspectives. *J Ind Microbiol Biot* 35:377-391. 10.1007/s10295-008-0327-
623   8

624 Kumar S, and Ramon D. 1996. Purification and regulation of the synthesis of a beta-xylosidase
625   from *Aspergillus nidulans*. *FEMS Microbiol Lett* 135:287-293. DOI 10.1111/j.1574-
626   6968.1996.tb08003.x

627 Laemmli UK. 1970. Cleavage of Structural Proteins during Assembly of Head of Bacteriophage-
628   T4. *Nature* 227:680-&. Doi 10.1038/227680a0

629 Liggenstoffer AS, Youssef NH, Wilkins MR, and Elshahed MS. 2014. Evaluating the utility of

630 hydrothermolysis pretreatment approaches in enhancing lignocellulosic biomass

631 degradation by the anaerobic fungus *Orpinomyces* sp. strain C1A. *J Microbiol Meth*

632 104:43-48. 10.1016/j.mimet.2014.06.010

633 Lineweaver H, and Burk D. 1934. The determination of enzyme dissociation constants. *J Am*

634 *Chem Soc* 56:658-666. Doi 10.1021/Ja01318a036

635 Liu ZL, Saha BC, and Slininger PJ. 2008. Lignocellulosic biomass conversion to ethanol by

636 *Saccharomyces*. In: Harwood CS, Demain AL, and Wall JD, eds. *Bioenergy*. Washington

637 D.C.: ASM Press.

638 Ljungdahl LG. 2008. The cellulase/hemicellulase system of the anaerobic fungus *Orpinomyces*

639 PC-2 and aspects of its applied use. *Ann Ny Acad Sci* 1125:308-321.

640 10.1196/annals.1419.030

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

642 active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42:D490-D495.

643 10.1093/nar/gkt1178

644 Matsuo M, and Yasui T. 1984. Purification and some properties of beta-xylosidase from

645 *Trichoderma viride*. *Agr Biol Chem* 48:1845-1852.

646 Morrison JM, Elshahed MS, and Youssef NH. 2016. A defined enzyme cocktail from the

647 anaerobic fungus *Orpinomyces* sp. strain C1A effectively releases sugars from pretreated

648 corn stover and switchgrass. *Sci Rep* Accepted.

649 Morrison JM, Wright CM, and John GH. 2012. Identification, isolation and characterization of a

650 novel azoreductase from *Clostridium perfringens*. *Anaerobe* 18:229-234.

651 10.1016/j.anaerobe.2011.12.006

652   National Research Council N. 2011. Renewable Fuel Standard: Potential Economic and

653   Environmental Effects of U.S. Biofuel Policy. Washington, D.C.

654   PyMol. The PyMOL Molecular Graphics System. Version 1.7.4 ed: Schrodinger, LLC.

655   Ragauskas AJ, Williams CK, Davison BH, Britovsek G, Cairney J, Eckert CA, Frederick WJ,

656   Hallett JP, Leak DJ, Liotta CL, Mielenz JR, Murphy R, Templer R, and Tschaplinski T.

657   2006. The path forward for biofuels and biomaterials. *Science* 311:484-489.

658   10.1126/science.1114736

659   Roy A, Kucukural A, and Zhang Y. 2010. I-TASSER: a unified platform for automated protein

660   structure and function prediction. *Nat Protoc* 5:725-738. 10.1038/nprot.2010.5

661   Saha BC, and Bothast RJ. 1996. Production, purification, and characterization of a highly

662   glucose-tolerant novel beta-glucosidase from *Candida peltata*. *Appl Environ Microbiol*

663   62:3165-3170.

664   Scheller HV, and Ulvskov P. 2010. Hemicelluloses. *Annu Rev Plant Biol* 61:263-289.

665   10.1146/annurev-arplant-042809-112315

666   Shao WL, Xue YM, Wu AL, Kataeva I, Pei JJ, Wu HW, and Wiegel J. 2011. Characterization of

667   a bovel beta-xylosidase, XylC, from *Thermoanaerobacterium saccharolyticum* JW/SL-

668   YS485. *Appl Environ Microbiol* 77:719-726. 10.1128/Aem.01511-10

669   Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li WZ, Lopez R, McWilliam H, Remmert

670   M, Soding J, Thompson JD, and Higgins DG. 2011. Fast, scalable generation of high-

671   quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol* 7:1-6.

672   Artn 539 10.1038/Msb.2011.75

673   Skalova T, Dohnalek J, Spiwok V, Lipovova P, Vondrackova E, Petrokova H, Duskova J, Strnad

674   H, Kralova B, and Hasek J. 2005. Cold-active beta-galactosidase from *Arthrobacter* sp

675   C2-2 forms compact 660 kDa hexamers: crystal structure at 1.9 angstrom resolution. *J*

676   *Mol Biol* 353:282-294. 10.1016/j.jmb.2005.08.028

677   Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of

678   large phylogenies. *Bioinformatics* 30:1312-1313. 10.1093/bioinformatics/btu033

679   Takashima S, Iikura H, Nakamura A, Hidaka M, Masaki H, and Uozumi T. 1998.

680   Overproduction of recombinant *Trichoderma reesei* cellulases by *Aspergillus oryzae* and

681   their enzymatic properties. *J Biotechnol* 65:163-171. Doi 10.1016/S0168-1656(98)00084-

682   4

683   Tamura K, Stecher G, Peterson D, Filipski A, and Kumar S. 2013. MEGA6: molecular

684   evolutionary genetics analysis version 6.0. *Mol Biol Evol* 30:2725-2729.

685   10.1093/molbev/mst197

686   Tong CC, Cole AL, and Shepherd MG. 1980. Purification and properties of the cellulases from

687   the thermophilic fungus *Thermoascus aurantiacus*. *Biochem J* 191:83-94.

688   Ulrich JT, Temple KL, and Mcfeters GA. 1972. Induction and characterization of beta-

689   galactosidase in an extreme thermophile. *J Bacteriol* 110:691-698.

690   Van Dyk JS, and Pletschke BI. 2012. A review of lignocellulose bioconversion using enzymatic

691   hydrolysis and synergistic cooperation between enzymes-Factors affecting enzymes,

692   conversion and synergy. *Biotechnol Adv* 30:1458-1480.

693   10.1016/j.biotechadv.2012.03.002

694   vanPeij NNME, Brinkmann J, Vrsanska M, Visser J, and deGraaff LH. 1997. Beta-xylosidase

695   activity, encoded by xlnD, is essential for complete hydrolysis of xylan by *Aspergillus*

696   *niger* but not for induction of the xylanolytic enzyme spectrum. *Eur J Biochem* 245:164-

697   173.

698   Verdoucq L, Moriniere J, Bevan DR, Esen A, Vasella A, Henrissat B, and Czjzek M. 2004.

699   Structural determinants of substrate specificity in family 1 beta-glucosidases - Novel

700   insights from the crystal structure of sorghum dhurrinase-1, a plant beta-glucosidase with

701   strict specificity, in complex with its natural substrate. *J Biol Chem* 279:31796-31803.

702   10.1074/jbc.M402918200

703   Wagschal K, Franqui-Espiet D, Lee CC, Robertson GH, and Wong DWS. 2008. Cloning,

704   expression and characterization of a glycoside hydrolase family 39 xylosidase from

705   *Bacillus halodurans* C-125. *Appl Biochem Biotech* 146:69-78. 10.1007/s12010-007-

706   8055-5

707   Wang HC, Chen YC, Huang CT, and Hseu RS. 2013. Cloning and characterization of a

708   thermostable and pH-stable cellobiohydrolase from *Neocallimastix patriciarum* J11.

709   *Protein Expres Purif* 90:153-159. 10.1016/j.pep.2013.06.004

710   Yang JK, Yoon HJ, Ahn HJ, Lee BI, Pedelacq JD, Liong EC, Berendzen J, Laivenieks M, Vieille

711   C, Zeikus GJ, Vocadlo DJ, Withers SG, and Suh SW. 2004. Crystal structure of beta-D-

712   xylosidase from *Thermoanaerobacterium saccharolyticum*, a family 39 glycoside

713   hydrolase. *J Mol Biol* 335:155-165. 10.1016/j.jmb.2003.10.026

714   Yang JY, Yan RX, Roy A, Xu D, Poisson J, and Zhang Y. 2015. The I-TASSER Suite: protein

715   structure and function prediction. *Nat Methods* 12:7-8. 10.1038/nmeth.3213

716   Youssef NH, Couger MB, Struchtemeyer CG, Liggenstoffer AS, Prade RA, Najar FZ, Atiyeh

717   HK, Wilkins MR, and Elshahed MS. 2013. The genome of the anaerobic fungus

718   *Orpinomyces* sp strain C1A reveals the unique evolutionary history of a remarkable plant

719   biomass degrader. *Appl Environ Microbiol* 79:4620-4634. 10.1128/Aem.00821-13

720   Yuan TZ, Yang PL, Wang YR, Meng K, Luo HY, Zhang W, Wu NF, Fan YL, and Yao B. 2008.

721   Heterologous expression of a gene encoding a thermostable beta-galactosidase from

722   *Alicyclobacillus acidocaldarius*. *Biotechnol Lett* 30:343-348. 10.1007/s10529-007-9551-

723   y

724   Zhang Y. 2008. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 9:40-

725   48. Artn 40 10.1186/1471-2105-9-40

726   Zhang YHP, Hong J, and Xinhao Y. 2009. Cellulase Assays. In: Mielenz JR, ed. *Biofuels:*

727   *Methods and Protocols*: Humana Press, 213-231.

728   Zhu H, Cheng KJ, and Forsberg CW. 1994. A truncated beta-xylosidase from the anaerobifungus

729   *Neocallimastix patriciarum* 27. *Can J Microbiol* 40:484-490.

730

731

732

733

734

735

736

737

738

739

740

741

742

743   **Figure Legends**

744   **Figure 1. Phylogenetic analysis of GH39 β-xylosidases, including Bgxg1.** Sequences annotated as GH39 β-

745   xylosidases (n=200 sequences, October 28, 2015) were retrieved from CAZyme databases (Lombard et al. 2014).

746   Genbank accession numbers are shown for reference proteins (due to the unavailability of *Piromyces* proteins in

747   Genbank, those proteins are shown as JGI accession numbers). The Maximum Likelihood tree was generated in

748   RAxML (Stamatakis 2014) using a BLOSUM62 substitution matrix and a GAMMA model of rate heterogeneity.

749   The model estimated an alpha parameter of 2.069. Bootstraps values (100 replicates) are shown for nodes with >50

750   bootstrap support. The sequences were empirically classified into four classes (Classes I-IV), and Class III, to which

751   Bgxg1 is affiliated, is further classified into four distinct lineages (III-A-III-D). The α-iduronidase sequence from

752   *Mus musculus* was utilized as an outgroup. β-xylosidases that were previously characterized biochemically were

753   phylogenetically affiliated with either Class II (*Bacillus halodurans* (BAB04787.1) and *Geobacillus*

754   *stearothermophilus* (ABI49941.1) in bottom Firmicutes wedge, and *Thermoanaerobacterium saccharolyticum*

755   (AAB68820.1) in middle Firmicutes wedge) or Class I (*Caulobacter crescentus* (ACL95907.1), bottom α-

756   Proteobacteria wedge). Bgxg1, from *Orpinomyces* sp. strain C1A, is shown highlighted in yellow.

757

758   **Figure 2. Effect of Temperature and pH on Bgxg1 activity A) Optimal pH, B) Optimal Temperature, C) pH**

759   **Stability, D) Thermal Stability.** All values are presented as relative specific activities, calculated by determining

760   the % activity relative to the highest activity (with the highest activity set at 100%). For A, B, C, and D, PNPX was

761   used as a substrate. Error bars represent standard deviation of triplicate (n=3) samples.

762

763   **Figure 3. Substrate competition and Bgxg1 preference.** Monosaccharides (glucose (■) or xylose (⌀)) release

764   was assayed when Bgxg1 was challenged with 10 mM cellobiose (**A**), 10 mM xylobiose (**B**), or an equimolar

765   mixture of both substrates (**C**). In (**A**), the effect of xylobiose (as a competitive inhibitor) is measured through

766   conducting a β-glucosidase activity assay. In (**B**), the effect of cellobiose (as a competitive inhibitor) is measured

767   through conducting a β-xylosidase activity assay. In (**C**), a competition assay was performed with both cellobiose

768   and xylobiose present, assaying for the presence of glucose or xylose.

769

770 **Figure 4. Alignment of Bgxg1 and the four biochemically-characterized GH39-family enzymes, highlighting**

771 **structural predictions and conservation of or around the active site.** Structural predictions for Bgxg1 sequence

772 were obtained using I-TASSER three-dimensional model (Fig. S3) (Roy et al. 2010; Yang et al. 2015; Zhang 2008).

773 Bgxg1 sequence is compared to those from *Caulobacter crescentus, Thermoanaerobacterium saccharolyticum,*

774 *Geobacillus stearothermophilus,* and *Bacillus halodurans.* α-helices in blue are those within the small α-helical

775 domain, α-helices and β-sheets in green are those within the (α/β)$_8$ barrel, and β-sheets in red are those within the β-

776 sandwich. Red stars (*) represent catalytic residues within the active site. Black stars (*) represent those residues

777 close to the active site, as determined within the Bgxg1 model. Blue stars (*) represent residues noted in the

778 literature to be important for β-xylosidase function (Czjzek et al. 2005; Yang et al. 2004).

779

Table 1

Table 1. Substrate Specificity and Specific Activity of Bgxg1

1   **Table 1. Substrate Specificity and Specific Activity of Bgxg1.**

Substrate[a]	Activity Tested	Specific Activity (U/mg ± SD)
PNPG	β-glucosidase	73.4 ± 7.15
Cellobiose	β-glucosidase	55.1 ± 5.36
PNPGal	β-galactosidase	54.6 ± 2.26
PNPX	β-xylosidase	11.5 ± 1.2
Xylobiose	β-xylosidase	10.9 ± 0.96
Beechwood Xylan	Xylanase	10.8 ± 1.25
Avicel	Exoglucanase	ND[b]
CMC	Endoglucanase	ND[b]
Locust Bean Gum	Mannanase	ND[b]
PNPA	Arabinosidase	ND[b]
PNPAc	Acetyl Xylan Esterase	ND[b]
PNPC	Cellobiohydrolase	ND[b]
PNPM	Mannosidase	ND[b]
Aldouronic acid	α-glucuronidase	ND[b]

2   **a** Abbreviations: PNPC - *p*-nitrophenyl-β-D-cellobioside, PNPX - *p*-nitrophenyl-β-D-xylopyranoside, PNPA - *p*-
3   nitrophenyl-β-D-arabinofuranoside, PNPM - *p*-nitrophenyl-β-D-mannoside, PNPG - *p*-nitrophenyl-β-D-
4   glucopyranoside, PNPGal - *p*-nitrophenyl-β-D-galactopyranoside, PNPAc - *p*-nitrophenyl-acetate.
5   b ND: Not detected.

Enzyme Kinetics for Bgxg1

1   **Table 2. Enzyme Kinetics for Bgxg1.** K_m and V_{max} values were calculated by extrapolation from Lineweaver-Burke
2   plots.

Substrate[a]	Activity Tested	K_m[b]	V_{max} (U/mg)
PNPG	β-glucosidase	BDL[c]	769 ± 18
PNPGal	β-galactosidase	BDL[d]	769 ± 13
PNPX	β-xylosidase	0.00485 mM ± 0.00062	127 ± 8
Beechwood Xylan	Xylanase	0.038 mg/mL ± 0.0039	25.6 ± 10

3   a: Abbreviations: PNPG - p-nitrophenyl-β-D-glucopyranoside, PNPGal - p-nitrophenyl-β-D-galactopyranoside,
4   PNPX - p-nitrophenyl-β-D-xylopyranoside,.
5   b: K_m values are expressed in either mM or mg/mL, depending on the substrate tested. Values are shown ± standard
6   deviation of triplicate samples (n=3).
7   c: BDL: Below detection limit (500 nM). Extrapolated K_m value obtained using Lineweaver-Burke plot was
8   0.0000125 mM ± 0.0000096.
9   d: BDL: Below detection limit (500 nM). Extrapolated K_m value obtained using Lineweaver-Burke plot was
10 0.000214 mM ± 0.000016.

Substrate competition experiments.

2 **Table 3. Substrate competition experiments.** "Activity tested" column refers to the colorimetric substrate tested
3 (PNPX for β-xylosidase, PNPG for β-glucosidase) in the presence of the active site inhibitor (cellobiose or
4 xylobiose, at listed Inhibitor concentrations). Specific activity, K_m, and V_{max} refer to the values calculated for the
5 colorimetric substrate in each experiment.

Activity Tested	Active Site Inhibitor	Inhibitor (mM)	Relative Specific Activity (%)	K_m (mM)	V_{max} (U/mg)
β-xylosidase	Cellobiose	0	100	0.00485	127
		10	78.9	1.438	118
		20	52.1	3.51	129
β-glucosidase	Xylobiose	0	100	0.0000125	769
		10	75.8	0.000235	763
		20	57.2	0.00349	752

Comparison of key amino acids between Bgxg1 and all four biochemically characterized β-xylosidases

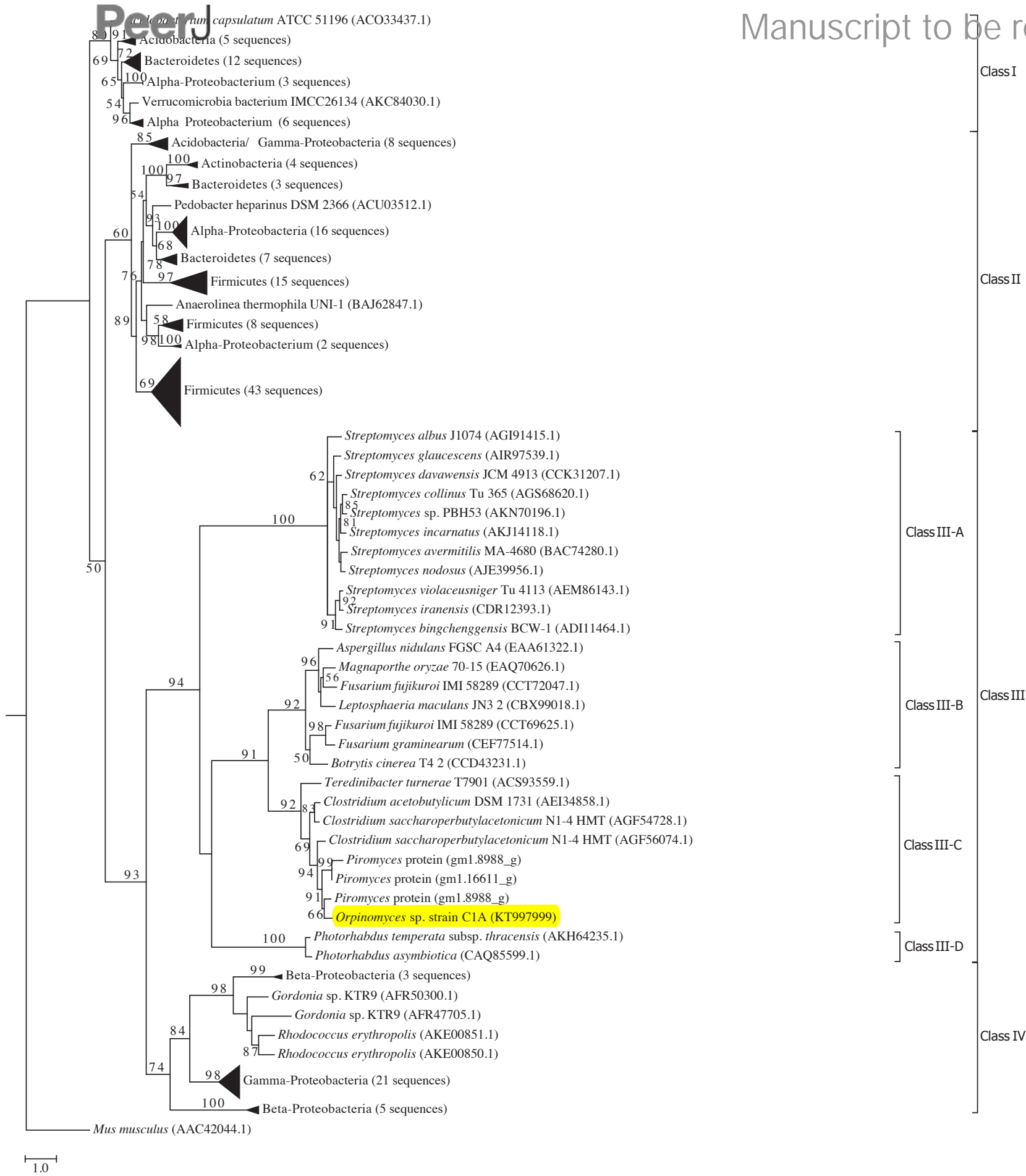
1 **Table 4. Comparison of key amino acids between Bgxg1 and all four biochemically characterized (BC) β-**

2 **xylosidases from *Thermoanaerobacterium saccharolyticum, Bacillus halodurans, Geobacillus***

3 ***stearothermophilus, Caulobacter crescentus*, as well as in Classes III-A, III-B, and III-D[a].**

Pos. [b]	AA in Bgxg1	AA in 4 BC[d]	Significance of Change	Importance of Residue	Class III-A	Class III-B	Class III-C	Class III-D
46	Val	Tyr	Small, nonpolar (Val) vs. Large, polar (Tyr)	Near active site	NC[e]	Ile	Val	NC
129	Asp	Asn	Negative charge (Asp) vs. Neutral charge (Asn)	H-bonding	Lys	Asp	Asp	Asp
131[c]	Thr/NC	Phe	Small, polar (Thr) vs. Large, nonpolar (Phe)	Near active site	NC	NC	NC	NC
139	Phe	Tyr	Large, nonpolar (Phe) vs. Large, polar (Tyr)	Near active site	Tyr	Tyr	Phe	Tyr
163	Cys	Ala	Polar, thiol (Cys) vs. Nonpolar (Ala)	Near active site	Tyr	Ala	Cys	Tyr
171	Lys	Trp	Positive charge (Lys) vs. Nonpolar (Trp)	Near active site	Trp	NC	Lys	Lys
194	Leu	Tyr	Small, nonpolar (Leu) vs. Large, polar (Tyr)	H-bonding	Ser	Ile/Glu	Leu	Tyr
242	Arg	Ala	Positive charge (Arg) vs. Small, nonpolar (Ala)	Near active site	NC	NC	Arg	NC
322-323	-gap-	Glu	Gap vs. Negative charge (Glu)	H-bonding	Arg/Thr/Lys	-gap-	-gap-	-gap-
322-323	-gap-	Glu	Gap vs. Negative charge (Glu)	H-bonding	Gly/-gap-	-gap-	-gap-	-gap-

5 **[a]** No changes were identified in 10 different positions (Arg48, Ile124, Trp125, Asn126, Glu127, Pro128, Trp132,
6 Pro162, His192, Glu225), and 5 positions were variable across all sequences (Val81, Gly130, Tyr164, Ser165,
7 Lys247).
8 **[b]** Pos. (Positions) refer to the position of the amino acid in Bgxg1
9 **[c]** Bgxg1 and all proteins in Class III-C β-xylosidases have identical amino acid sequences in all key positions with
10 one exception (Thr131)
11 **[d]** Sequences identified using the alignment presented in Fig. 5.
12 **[e]** NC = Not conserved

Phylogenetic analysis of GH39 β-xylosidases, including Bgxg1.

Effect of Temperature and pH on Bgxg1 activity A) Optimal pH, B) Optimal Temperature, C) pH Stability, D) Thermal Stability

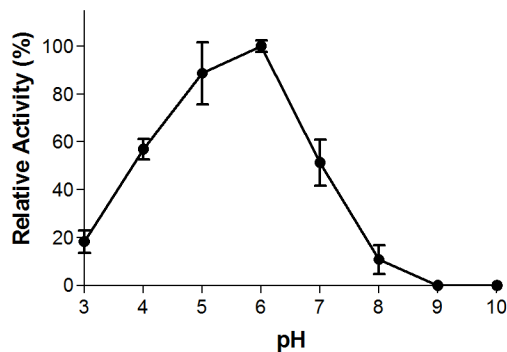
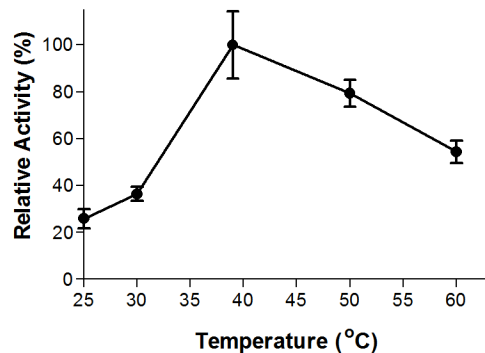
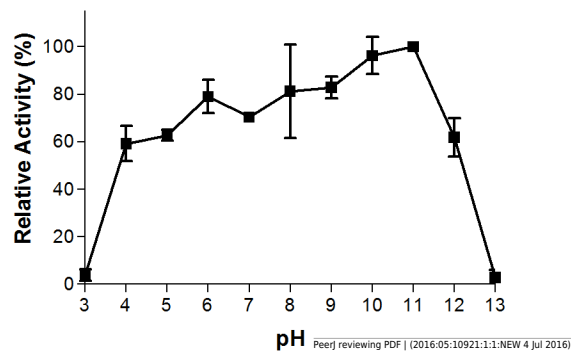
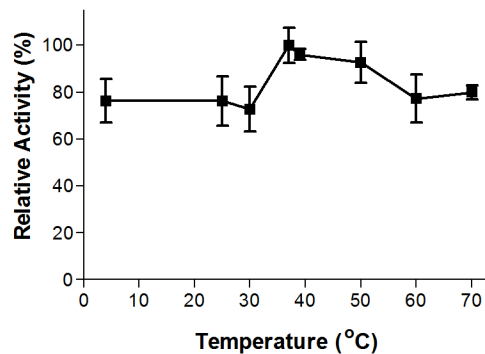
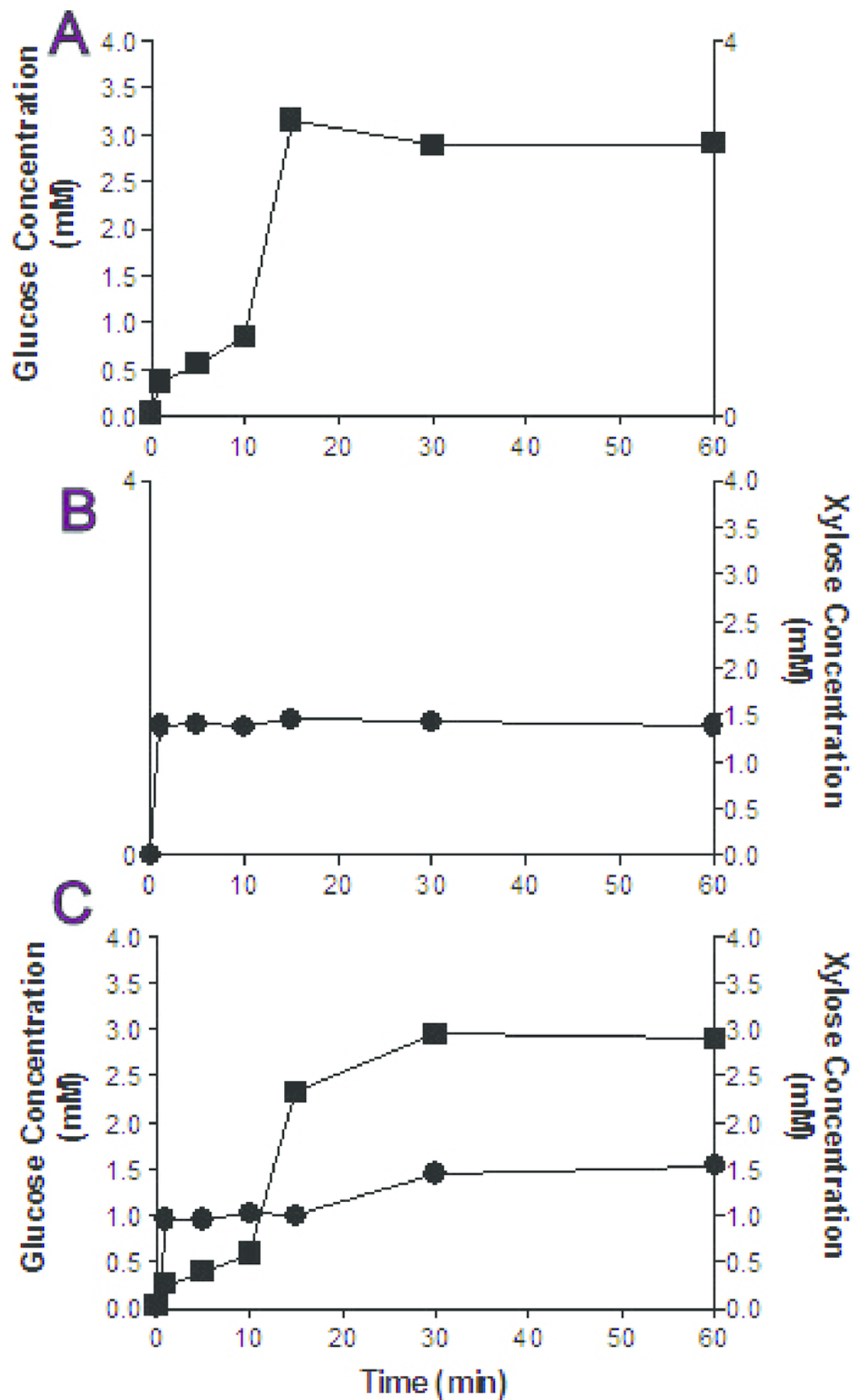



Figure 3

Alignment of Bgxg1 and the four biochemically-characterized GH39-family enzymes, highlighting structural predictions and conservation of or around the active site.

Alignment of Bgxg1 and the four biochemically-characterized GH39-family enzymes, highlighting structural predictions and conservation of or around the active site.

α-1   α-2   α-3   β-1

```
Bgxg1              1   MT......NVLTVE...CNNKL.RRATHCANGSLYGIT..ETTPRDYKNLVDPLHPFVMRN............PARGGNG
Caulobacter        1   MANAGPGARVIDLDLRRAAGPVDRFFDLSIGSDYPGTLIREDSQAQLKTTVDELGFRYIRFHAIFHDVLGTVKVQ.....
Thermoanaerobacterium  1   MIKVRVP.DFSDKFSDRWRYCVGTGRLGLALQKEYIETLKYVKENIDFKYIRGHGLLCDDVGIYRED.VVGD
Geobacillus        1   MKVVNVP.SNGREKFKKNWKFCVGTGRLGLALQKEYLDHLKLVQEKIGFRYIRGHGLLSDDVGIYREV.EIDG
Bacillus           1   MKTVVN.DRSFAYFPKKWKYCIGTGRLGLALQKEYVDHLARLQKELNFQYIRGHGLLHDDIGIYRERKRADG
```

α-4   β-2   α-5

```
Bgxg1             57   NQHPYGDAI............KVARRLADTPGALVSVDLPDMLPGWPYKWPG.......MQNWLNQVKSFI....KDK
Caulobacter       76   DGKIVYDWTKIDQLYDALLAKGIKPFIELGFTPEAMKTSDQ......TIFYWKGNTSHPK.LGPWRDLIDAFVHHLRARY
Thermoanaerobacterium  72   EVKPFYNFTYIDRIFDSFLEIGIRPFVEIGFMPKKLASGTQ......TVFYWEGNVTPPKDYEKWSDLVKAVLHHFISRY
Geobacillus       72   EMKPFYNFTYIDRIVDSYLALNIRPFIEFGFMPKALASGDQ......TVFYWKGNVTPPKDYNKWRDLIVAVVSHFIERY
Bacillus          73   TVEPFYNFTYIDRIFDTFLELNIRPFVEIGFMPKLLASGEQ......TIFDWQGNVTPPKDYDQWKQLIQAVISHFIDRY
```

α-5   β-3   α-6   β-4   α-7   β-5

```
Bgxg1            112   KASGLKNWYGLEIWNEPD..GTWNNSN.GSFEEMWKQTYQAIRQADPNEKIIGPCYSWYTDDKLRNFLKYAKANNCLPDI
Caulobacter      149   GVEEVRTW.FFEVWNEPNLDGFWEKADQAAYFELYDVTARAIKAIDPSLRVGGPATAGA..AWVPEFLAHVKKSGSAVDF
Thermoanaerobacterium 146   GIEEVLKW.PFEIWNEPNLKEFWKDADEKEYFKLYKVTAKAIKEVNENLKVGGPAICGGADYWIEDFLNFCYEENVPVDF
Geobacillus      146   GIEEVRTW.LFEVWNEPNLVNFWKDANKQEYFKLYEVTARAVKSVDPHLQVGGPAICGGSDEWITDFLHFCAERRVPVDF
Bacillus         147   GVEEVTKW.PFEIWNEPNLINFWQHADKKEYFKLYKITARAIKEVHPYIQVGGPAICGGSDEWITDFLQFCHKEEVPVDF
```

β-5   α-8   β-6   α-9

```
Bgxg1            189   ISWHELSGI............DGVSSHLRSYREIEKSLGIPELPISINEYCDAKHELEGQPG..SSARFIG.KF
Caulobacter      226   VTTHTYGVDGGFLDEKGVQDTKLSPSPDAVVGDVRRVREQIEASAFPGLPLYFTEWSTSYTPRDSVHDSYVSAAYIVEKL
Thermoanaerobacterium 225   VSRHAYTSKQG.EYTPHLIYQEIMP.SEYMLNEFKTVREIIKNSHFPNLPFHITEYNTSYSPQNPVHDTPFNAAYIARIL
Geobacillus      225   VSRHAYTSKAPHKKTFEYYYQELEP.PEDMLEQFKTVRALIRQSPFPHLPLHITEYNTSYSPINPIHDTALNAAYIARIL
Bacillus         226   VSRHAYTSAKPHKVTPDYYYQELYE.NTHMLDELKSVKELIQQSPFPNLPFHITEYNTSYSPINPVHDTVLNAAYLARIL
```

α-9   β-7   α-10   β-8   β-9

```
Bgxg1            248   ERYKV.DTAMITWWFV.......PLPGRLGSLLATDTQKGAGWYFYKWYGDMTGDMVYVKPPNDNSNLVDGAACVD
Caulobacter      306   RRVKGLVQAMSYWTYSDLFEEPGPPTAPFQGGFG.LMNPQGIRKPSWFAYKYLNALKGRELVCADDQVFAARDGDRVA..
Thermoanaerobacterium 303   SEGGDYVDSFSYWTFSDVFEERDVPRSQFHGGFG.LVALNMIPKPTFYTFKFFNAMGEEMLYRDEHMLVTRRDDGSVA..
Geobacillus      304   SEGGDYVDSFSYWTFSDVFEEMDVPKALFHGGFG.LVALHSIPKPTFHAFTFFNALGDELLYRDGEMIVTRRKDGSIA..
Bacillus         305   SEAGDIVDSFSYWTFSDVFEEAGVPTAPFHGGFG.LIALHGIAKPTYHLFSFFNQLGEQLLYRDSQMVVTKKQDGSIQ..
```

β-10   β-11

```
Bgxg1            316   TNKEYISFIFGGPNDGTI.RASSIIFQALLDLLPML
Caulobacter      383   ....IVAYAWRQPDQKVSNRPFYTKLHPASDVEPLKVRLTSLKPGRYKLRVRRVGYRRNDAYSAYIDMGSPTTLTESQLQ
Thermoanaerobacterium 380   ....LI..AWNEVMDKTENPDEDYEVE.....IPVRF.......RDVFIKRQLIDEEHGNPWGTWIHMGRPRYPSKEQVN
Geobacillus      381   ....AV..LWNLVMEKGEGFTKEVQLV.....IPVSF.......SAVFIKRQIVNEQYGNAWRVWKQMGRPRFPSRQAVE
Bacillus         382   ....LV..VWNLIMEKGEGLEQTVQIE.....LPTQS.......DAVFIKRKTIDETNGNPWRVWKEMGRPRFPKKNEID
```

```
Bgxg1            351
Caulobacter      459   SLQALTEDRPEIEKALKVSGETVVDLPMRANDVVLIELEPLA
Thermoanaerobacterium 442   TLREVAKPEIMTSQPVANDGYLNLKFKLGKNAVVLYELTERIDESSTYIGLDDSKINGY
Geobacillus      443   TLRQVAQPHVMTEQRRATDGVIHLSIVLSKNEVTLIEIEQVRDETSTYVGLDDGEMTSYSS
Bacillus         444   TLQEVAQPLIRTERRDGLSRTLKLELTLSKNEVSLVEVFPIIDETNTYPGLDDRLIPSY
```