

This is a repository copy of *In vitro and in vivo characterization of three Cellvibrio japonicus Glycoside Hydrolase Family 5 members reveals potent xyloglucan backbone-cleaving functions.*

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/126900/>

Version: Accepted Version

Article:

Attia, Mohamed A., Nelson, Cassandra E., Offen, Wendy A. orcid.org/0000-0002-2758-4531 et al. (4 more authors) (2018) In vitro and in vivo characterization of three Cellvibrio japonicus Glycoside Hydrolase Family 5 members reveals potent xyloglucan backbone-cleaving functions. *Biotechnology for biofuels*. 45. ISSN 1754-6834

<https://doi.org/10.1186/s13068-018-1039-6>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

[Click here to view linked References](#)

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

***In vitro and in vivo* characterization of three *Cellvibrio japonicus* Glycoside Hydrolase
Family 5 members reveals potent xyloglucan backbone-cleaving functions**

Mohamed A. Attia^{1,2}, Cassandra E. Nelson³, Wendy A. Offen⁴, Namrata Jain^{1,2}, Gideon J.
Davies⁴, Jeffrey G. Gardner³, and Harry Brumer^{1,2,5,6,*}

5 ¹Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC,
V6T 1Z4, Canada.

²Department of Chemistry, University of British Columbia, 2036 Main Mall, Vancouver, British
Columbia V6T 1Z1, Canada.

10 ³Department of Biological Sciences, University of Maryland, Baltimore County, Baltimore MD
21250, USA.

⁴Department of Chemistry, University of York, Heslington, York YO10 5DD, UK.

⁵Department of Biochemistry and Molecular Biology, University of British Columbia, 2350
Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada.

15 ⁶Department of Botany, University of British Columbia, 6270 University Blvd., Vancouver,
British Columbia V6T 1Z4, Canada

*To whom correspondence should be addressed: E_mail brumer@msl.ubc.ca; Tel. (+1)
6048273738; Fax (+1) 6048222114.

20 ***Dedication:*** *This article is dedicated to Cellvibrio japonicus vanguard Prof. Harry J. Gilbert on
the occasion of his retirement.*

Abstract

Background: Xyloglucan (XyG) is a ubiquitous and fundamental polysaccharide of plant cell walls. Due to its structural complexity, XyG requires a combination of backbone-cleaving and sidechain-debranching enzymes for complete deconstruction into its component monosaccharides. The soil saprophyte *Cellvibrio japonicus* has emerged as a genetically tractable model system to study biomass saccharification, in part due to an innate capacity to utilize a wide range of plant polysaccharides for growth. Whereas the downstream debranching enzymes of the xyloglucan utilization system of *C. japonicus* have been functionally characterized, the requisite backbone-cleaving *endo*-xyloglucanases were unresolved.

Results: Combined bioinformatic and transcriptomic analyses implicated three Glycoside Hydrolase Family 5 Subfamily 4 (GH5_4) members, with distinct modular organization, as potential keystone *endo*-xyloglucanases in *C. japonicus*. Detailed biochemical and enzymatic characterization of the GH5_4 modules of all three recombinant proteins confirmed particularly high specificities for the XyG polysaccharide versus a panel of other cell wall glycans, including mixed-linkage beta-glucan and cellulose. Moreover, product analysis demonstrated that all three enzymes generated XyG oligosaccharides required for subsequent saccharification by known exo-glycosidases. Crystallographic analysis of GH5D, which was the only GH5_4 member specifically and highly upregulated during growth on XyG, in free, product-complex, and active-site affinity-labelled forms revealed the molecular basis for the exquisite XyG specificity among these GH5_4 enzymes. Strikingly, exhaustive reverse-genetic analysis of all three GH5_4 members and a previously biochemically characterized GH74 member failed to reveal a growth defect, thereby indicating functional compensation *in vivo*, both among members of this cohort and by other, yet unidentified, xyloglucanases in *C. japonicus*. Our systems-based analysis

1
2
3
4 indicates distinct substrate-sensing (GH74, GH5E, GH5F) and attack-mounting (GH5D)
5
6 functions for the endo-xyloglucanases characterized here.
7

8
9 **Conclusions:** Through a multi-faceted, molecular systems-based approach, this study provides a
10 new insight into the saccharification pathway of xyloglucan utilization system of *C. japonicus*.
11

12
13
14 50 The detailed structural-functional characterization of three distinct GH5_4 endo-xyloglucanases
15 will inform future bioinformatics predictions across species, and provides new CAZymes with
16 defined specificity that may be harnessed in industrial and other biotechnological applications.
17
18
19
20
21

22 23 **Keywords**

24
25
26 Xyloglucan, saccharification, glycoside hydrolase, *Cellvibrio japonicus*, saprophyte
27
28

29 55
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Background

Renewable plant biomass is envisioned as a promising alternative to fossil petroleum for the production of liquid fuels and high-value chemicals [1, 2]. Plant cell walls are, however, chemically and structurally complex in nature and require harsh thermo-chemical treatment to yield fermentable sugars. Such processes often generate undesirable by-products that inhibit subsequent microbial conversion [3]. In light of their ability to catalyze the degradation of recalcitrant plant cell walls under ambient conditions, enzymes from saprophytic micro-organisms constitute an attractive palette of biocatalysts for improved biomass saccharification [4]. The discovery and characterization of new enzymes from saprophytes is thus central to advancing biotechnology and, not least, underpins fundamental understanding of the biological roles of these micro-organisms in the global carbon cycle.

The Gram-negative bacterium, *Cellvibrio japonicus* Ueda107 (formerly, *Pseudomonas fluorescens* subsp. *cellulosa*) has emerged as a model saprophytic micro-organism with a demonstrated ability to utilize nearly all plant cell wall polysaccharides, including cellulose, xylans, mannans, arabinans, and pectins [5, 6]. Indeed, sequencing of the *C. japonicus* genome in 2008 revealed vast array of carbohydrate-active enzymes (CAZymes [7]) predicted to be involved in plant cell wall saccharification [8]. The recent development of genome editing techniques for *C. japonicus* has further advanced the biology and bioengineering of this bacterium in biomass conversion [9-13].

The xyloglucans (XyG) comprise an important family of cell wall matrix polysaccharides, which are ubiquitous and abundant across the plant kingdom [14, 15]. In dicots, XyGs may constitute up to 25% of the primary cell wall dry-weight, with lower amounts found in conifers

1
2
3
4 (10%) and grasses (<5%) [16, 17]. Structurally, XyGs have brush-like architectures built upon a
5
6 linear, cellulosic $\beta(1\rightarrow4)$ -D-glucan backbone that is extensively branched with $\alpha(1\rightarrow6)$ -
7
8
9 80 xylopyranosyl residues at regular intervals. Further elaboration of these branch points with
10
11 diverse monosaccharides and acetyl groups is dependent on the species and tissue of origin [18,
12
13 19]; presently *ca.* 20 distinct sidechain saccharide compositions are known [20, 21]. The
14
15 structure of the canonical dicot (fucogalacto)xyloglucan is shown in Fig. 1A. Due to this
16
17 structural complexity, complete XyG saccharification requires the concerted action of numerous
18
19
20
21 85 backbone-cleaving *endo*-xyloglucanases and side-chain-cleaving *exo*-glycosidases [22, 23].
22
23

24
25 As part of our ongoing effort to elucidate the xyloglucan (XyG) utilization system of *C.*
26
27 *japonicus*, we functionally characterized a multi-gene XyG utilization locus (XyGUL) in the *C.*
28
29 *japonicus* genome via a combination of genetics, enzymology, and structural biology. This
30
31 XyGUL encodes the three *exo*-glycosidases required for (fucogalacto)xyloglucan sidechain
32
33
34 90 cleavage (a GH95 α -L-fucosidase, a GH35 β -galactosidase, and a GH31 α -xylosidase) together
35
36 with a predicted TonB-dependent transporter (TBDT) (Fig. 1B & C) [13, 24, 25]. A highly
37
38 specific β -glucosidase, Bgl3D, which is encoded elsewhere in the genome, works in concert with
39
40 the *exo*-glycosidases of the XyGUL to effect the complete saccharification of XyG
41
42 oligosaccharides (XyGOs) in the periplasm (Fig. 1B & 1C) [26]. Noting that this locus likewise
43
44
45
46 95 lacked an associated *endo*-xyloglucanase, we also provided biochemical and structural evidence
47
48 that the lone, secreted *C. japonicus* GH74 member (Fig. 1B & C) could efficiently generate the
49
50 Glc₄-based XyGOs required by the downstream *exo*-glycosidases [27].
51
52
53
54

55 As we now show, genetic deletion of this GH74 *endo*-xyloglucanase did not, however,
56
57 impede the growth of *C. japonicus* on the polysaccharide, which suggested the involvement of
58
59
60 100 additional, unidentified *endo*-xyloglucanases. Hence, we also explored the *in vitro* and *in vivo*
61
62
63
64
65

1
2
3
4 function of three candidate *endo*-xyloglucanases from GH5 subfamily 4 (GH5_4) [28], guided
5
6 by bioinformatic and transcriptomic analyses. Utilizing a combination of reverse genetics,
7
8 enzymology, and structural biology, the present study provides a new insight into the upstream
9
10 deconstruction of XyG.
11
12
13
14
15

16 105 **Results and Discussion**

17 18 19 20 **Transcriptomic analysis reveals a potential keystone *endo*-xyloglucanase** 21 22 23 **from Glycoside Hydrolase (GH) Family 5, subfamily 4.** 24

25 We previously showed via quantitative PCR (qPCR) that the *C. japonicus* gene cluster
26
27 containing *xyl31A* (CJA_2706), *bgl35A* (CJA_2707), CJA_2709, and *afc95A* (CJA_2710) (Fig.
28
29 1B), was up-regulated during growth on xyloglucan-containing medium [13]. Biochemical
30 110 characterization confirmed that *xyl31A*, *bgl35A*, and *afc95A* encode a XyGO-specific GH31 α -
31
32 xylosidase, GH35 β -galactosidase, and GH95 α -L-fucosidase, respectively, while CJA_2709 was
33
34 predicted to encode a TonB-dependent transporter (TBDT) [13, 24, 25]. To aid identification of
35
36 potential *C. japonicus* *endo*-xyloglucanases acting upstream of these enzymes, a comprehensive
37
38 expression analysis via RNAseq was performed in the present study. Samples were collected
39
40 from both exponentially growing and stationary phase cells grown on glucose or xyloglucan as
41
42 115 the sole carbon source to allow for analyses of gene expression based on early-stage substrate
43
44 detection (Additional file 1: Figure S1), late-stage substrate detection (Additional file 1: Figure
45
46 S2A), or growth rate (Additional file 1: Figure S2B).
47
48
49
50
51
52
53
54

55 120 During exponential growth, there were 27 CAZyme-encoding genes significantly up-
56
57 regulated on XyG, including the four genes of the *C. japonicus* XyG cluster, which corroborated
58
59
60
61

1
2
3
4 previous qPCR results (Additional file 1: Table S1). Notably, CJA_3010, which encodes a GH5
5
6 subfamily 4 (GH5_4) member previously annotated as *cel5D*, was the highest upregulated gene
7
8 followed by the XyG cluster genes CJA_2709 (encoding a predicted TBDT) and CJA_2706
9
10
11 125 (*xyl31A*, encoding a GH31 α -xylosidase) [8]. Among the large and functionally diverse GH5
12
13 family [28], subfamily 4 is the only subfamily known to contain predominant *endo*-
14
15 xyloglucanases [23], which suggested a keystone role for this enzyme in xyloglucan utilization
16
17 by *C. japonicus*. Notably, CJA_2477 (previously annotated as *gly74* [8]; Fig. 1B) was not
18
19 significantly up-regulated during growth on XyG, despite the encoded GH74 *endo*-
20
21
22
23 130 xyloglucanase being previously shown to have high, specific activity for this polysaccharide
24
25 [27]. Instead, CJA_2477 appeared to be constitutively expressed at a low level (RPKM levels in
26
27 the 100-200 range), as were 14 other predicted CAZyme-encoding genes (Additional file 1:
28
29 Table S2).
30
31
32
33

34
35 The remaining CAZyme genes up-regulated during exponential growth are predicted to
36
37 135 have roles in the degradation of a diverse set of polysaccharides, which suggests that there is
38
39 complex cross-regulation of expression. As xyloglucan is unlikely to be encountered alone
40
41 during the saprophytic growth habit of *C. japonicus*, these results are suggestive of xyloglucan
42
43 degradation being one component of a sophisticated plant cell wall degradation response.
44
45 Congruently, when comparing the stationary phase *C. japonicus* cells growing on XyG and
46
47
48 glucose, only two genes of the XyG cluster, *bgl35A* and *afc95A* were still up-regulated on XyG,
49 140
50 together with 33 additional predicted and confirmed hemicellulase- and pectinase-encoding
51
52 genes (Additional file 1: Figure S2A, Additional file 1: Table S3). Additionally, when comparing
53
54 the exponential phase to the stationary phase for xyloglucan-grown cells, we observed that there
55
56
57
58 was a growth-phase-dependent response manifested as a significant shift in the suite of expressed
59
60
61
62
63
64
65

1
2
3
4 145 CAZyme genes (Additional file 1: Figure S2B, Additional file 1: Table S4). Specifically, these
5
6 differentially expressed CAZyme genes were not predicted to be XyG-specific, based on their
7
8 CAZy family membership, which suggested they are part of regulatory circuit that responds
9
10 generally to polysaccharides. Similar growth-phase-dependent responses have been previously
11
12 observed during cellulose utilization by *C. japonicus* [12] and *Clostridium thermocellum* (now
13
14
15
16 150 *Ruminiclostridium thermocellum*) [29, 30].

19 20 **Bioinformatic analysis and recombinant production of GH5_4 members from** 21 22 ***C. japonicus***

23
24
25 Spurred-on by the implication of the GH5 subfamily 4 (GH5_4) member encoded by
26
27 CJA_3010 in xyloglucan utilization by *C. japonicus*, we searched the genome for potential
28
29 homologs. *C. japonicus* encodes 15 GH5 members, of which only three belong to subfamily 4
30 155 ([8] see <http://www.cazy.org/b776.html>): The aforementioned CJA_3010 (GenBank
31
32 ACE84905.1, previously annotated as *cel5D* [8]), CJA_3337 (GenBank ACE83841.1, previously
33
34 annotated as *cel5E* [8]), and CJA_2959 (GenBank ACE86198.1, previously annotated as *cel5F*
35
36 [8]). Protein sequence analysis revealed that each of these gene products had a unique, multi-
37
38 modular architecture that suggested the possibility of distinct cellular localization and biological
39
40 160 function (Additional file 1: Figure S3). Considering the lack of demonstrable activity on
41
42 cellulose and high activity on xyloglucan (*vide infra*), the corresponding encoded enzymes are
43
44 referred to as *CjGH5D*, *CjGH5E*, and *CjGH5F* hereafter.
45
46
47
48
49
50
51
52

53 The highly up-regulated CJA_3010 encodes a signal peptidase II lipoprotein signal peptide
54
55 165 (predicted by LipoP 1.0 [31]), followed by a serine-rich linker and a GH5_4 catalytic module,
56
57 and was thus predicted to be anchored extracellularly in the outer membrane by N-terminal
58
59
60
61
62
63
64
65

1
2
3
4 cysteine lipidation (Additional file 1: Figure S3). CJA_3337 encodes an N-terminal signal
5
6 peptide (predicted by SignalP 4.0 [32]) and two carbohydrate-binding modules (CBMs [33]),
7
8 CBM2 and CBM10, in train with a GH5_4 catalytic module (Additional file 1: Figure S3).
9
10
11 170 CJA_2959 encodes a signal peptide (predicted by SignalP 4.0 [32] a Fibronectin type 3 (FN3)
12
13 domain, an undefined region, and a C-terminal GH5_4 catalytic module (Additional file 1:
14
15 Figure S3). The presence of signal peptides, and CBMs in the case of *CjGH5E*, is indicative of
16
17 extracellular secretion of both *CjGH5E* and *CjGH5F*.
18
19
20
21

22 Amino acid alignment of the catalytic modules of *CjGH5D*, *CjGH5E*, and *CjGH5F*

23
24 175 demonstrate conservation of the two catalytic glutamate residues, but low to moderate overall
25
26 sequence conservation (26 to 45% identity). Notably, alignment with endo-xyloglucanases from
27
28 *Bacteroides ovatus* [22], *Paenibacillus pabuli* [34], and a rumen metagenome [35] suggests that
29
30 the *C. japonicus* proteins are members of Subfamily 4 (Additional file 1: Figure S4, Additional
31
32 file 1: Table S5). GH5_4 is one of the largest GH5 subfamilies and contains, in addition to
33
34
35 180 specific endo-xyloglucanases, promiscuous endo- β (1,4)-glucanases, strict cellulases, and mixed-
36
37 linkage endo- β (1,3)/ β (1,4)glucanases (reviewed in [23, 28]). As such, we undertook the
38
39 recombinant production and enzymological characterization of the three *C. japonicus* GH5_4
40
41 members to precisely define their catalytic activities in the context of potential biological
42
43 function.
44
45
46
47
48
49

50 185 Our initial attempts to produce the full-length, multi-modular proteins recombinantly in *E.*
51
52 *coli* by replacement of the native signal peptides with an N-terminal hexahistidine (His₆)
53
54 purification tag were consistently unsuccessful: Intact protein mass spectrometry revealed
55
56 proteolytic instability of His₆-SRL-GH5D, while His₆-CBM2-CBM10-GH5E and His₆-FN3-
57
58 GH5F had very poor production yields (data not shown). In contrast, His₆-GH5D (Additional
59
60
61
62
63
64
65

1
2
3
4 190 file 1: Figure S3A) was produced as a stable, intact, active protein (calculated mass, 44222.2 Da;
5
6 observed by ESI-MS, 44222.6 Da) in excellent yield (150 mg L⁻¹). Likewise, our attempts to
7
8 produce the individual catalytic modules of *CjGH5E* and *CjGH5F* as N-terminally His₆-tagged
9
10 constructs (Additional file 1: Figure S3B & C) were successful (*His₆-CjGH5E* calculated mass,
11
12 41367.1 Da; observed by ESI-MS, 41370.1 Da, *His₆-CjGH5F* calculated mass, 40253.8 Da;
13
14
15
16 195 observed by ESI-MS 40253.9 Da) with approximate production yields of 14 and 9 mg L⁻¹,
17
18 respectively.
19
20
21

22 ***CjGH5_4* enzymes are highly efficient, specific endo-xyloglucanases.**

23
24
25 Informed by the subfamily membership of the three GH5_4 members, we anticipated that
26
27 these enzymes might exhibit significant *endo*-hydrolytic activity towards XyG. Hence, this
28
29 polysaccharide was used to determine pH and temperature optima. *CjGH5D*, *CjGH5E*, and
30 200 *CjGH5F* each exhibited approximately bell-shaped pH profiles, with the highest activity
31
32 achieved in 50 mM phosphate buffer (pH 7.5 in case of *CjGH5D* and *CjGH5E*, and pH 7.0 in
33
34 case of *CjGH5F*; Additional file 1: Figure S5). When the 3 enzymes were incubated with XyG at
35
36 different temperatures over the course of 10 minutes, optimum temperature was identified as 50
37
38 °C (*CjGH5D* and *CjGH5F*) and 55 °C (*CjGH5E*) (Additional file 1: Figure S5). To determine
39
40 substrate specificity of the three GH5_4 members, a panel of nine soluble polysaccharide
41
42 205 substrates were screened under these optimal conditions. Indeed *CjGH5D*, *CjGH5E*, and
43
44 *CjGH5F* all displayed high specific activity toward XyG (Table 1). No detectable activity
45
46 toward barley mixed-linkage 1,3/1,4-β-glucan, guar galactomannan, konjac glucomannan,
47
48 beechwood xylan, wheat flour arabinoxylan, or xanthan for any of the three enzymes was
49
50 observed. *CjGH5D* appeared to strictly require the branched XyG structure, while *CjGH5E*
51
52 demonstrated trace activities against the artificial 1,4-β-glucans hydroxyethylcellulose (HEC)
53
54 210
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 and carboxymethylcellulose (CMC) at the highest tested substrate concentration (2 mg mL⁻¹);
5
6 specific activities were 200-1500-fold less than XyG, respectively (Table 1). Similarly, CjGH5F
7
8
9 215 was able to hydrolyze HEC with an 800-fold lower specific activity than XyG, while no activity
10
11 towards CMC was detected. Michaelis-Menten analysis for XyG further underscored the high
12
13 XyG specificity of the three enzymes: remarkably low K_m values were observed and high k_{cat}
14
15 values recapitulated those previously observed for predominant *endo*-xyloglucanases, including
16
17 CjGH74 [22, 27, 36, 37] (Table 1, Additional file 1: Figure S6).
18
19
20
21

22 220 Time-course analyses of native XyG polysaccharide hydrolysis products by HPAEC-PAD
23
24 revealed that all three GH5_4 enzymes generated products of intermediate retention time in the
25
26 early stages of the reactions, with no significant generation of the Glc₄-based XXXG, XLXG,
27
28 XXLG, and XLLG limit-digestion products (Additional file 1: Figure S7, Additional file 1:
29
30 Figure S8; *cf.* Fig. 1). These results indicate that the three enzymes hydrolyze XyG through a
31
32 dissociative, rather than processive [38] mechanism, and are thus canonical *endo*-xyloglucanases
33
34 225 (EC 3.2.1.151; *cf.* EC 3.2.1.150, EC 3.2.1.155). The limit-digestion products further revealed
35
36 that all *C. japonicus* GH5_4 enzymes specifically catalyze hydrolysis at the anomeric position of
37
38 the unbranched glucose residues of the (galacto)XyG polysaccharide chain (Fig. 1). This
39
40 cleavage pattern is typical for many GH5 [22, 34, 37], GH9 [39, 40], GH12 [34, 41-44], GH16
41
42 [45] and GH74 [46-50] *endo*-xyloglucanases, although certain GH5 [35, 36], GH7 [51], GH44
43
44 230 [40, 52] and GH74 [53-55] members preferentially hydrolyze the XyG backbone between
45
46 branched glucosyl residues. The canonical XXXG-type XyGOs produced by CjGH5D,
47
48 CjGH5E, and CjGH5F are direct substrates for the *exo*-glycosidases of the XyG gene cluster
49
50 [13].
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 235 With knowledge of the cleavage specificity of the GH5_4 members, we determined kinetic
5
6 parameters for the hydrolysis of a panel of chromogenic oligosaccharides to reveal the
7
8 contribution of side chain substitution on substrate recognition and catalysis (Table 2). All three
9
10 enzymes were only weakly active on 2-chloro-4-nitrophenyl cellotrioside (GGG- β -CNP) and 2-
11
12 chloro-4-nitrophenyl cellotetraoside (GGGG- β -CNP), with meagre increases in k_{cat}/K_m values
13
14 arising from the addition of potential -4 subsite binding for the cellotetraoside (Table 2,
15
16 240 Additional file 1: Figure S9) GH subsite nomenclature according to [56]. Strikingly, the addition
17
18 of three $\alpha(1\rightarrow6)$ -xylopyranosyl residues to the glucan backbone resulted in significant increases
19
20 in catalytic efficiency for all GH5_4 members, which was manifested as 65-, 700-, and 150-fold
21
22 higher k_{cat}/K_m values for XXXG- β -CNP vis-à-vis GGGG- β -CNP with CjGH5D, CjGH5E, and
23
24 CjGH5F, respectively (Table 2, Additional file 1: Figure S9). These values correspond to 11, 18,
25
26 245 and 13 kJ/mol, respectively, of additional transition state stabilization in the formation of the
27
28 covalent glycosyl-enzyme in these anomeric-configuration-retaining GH5 enzymes (calculated
29
30 using the formula: $\Delta\Delta G^\ddagger = -RT \ln[(k_{cat}/K_m \text{ XXXG}) / (k_{cat}/K_m \text{ GGGG})]$) [57]. With XLLG- β -CNP,
31
32 the specificity constants (k_{cat}/K_m) were only increased 1.5- to 5-fold for the three *endo*-
33
34 xyloglucanases, thereby indicating that extending $\beta(1\rightarrow2)$ -galactopyranosyl residues (Fig. 1)
35
36 250 have little additional effect on catalysis (Table 2).
37
38
39
40
41
42
43
44
45
46

47 Covalent labeling of CjGH5D with an active-site-directed inhibitor

48
49 Active-site affinity-based inhibitors are important tools for the detailed kinetic analysis of
50
51 GH enzymes [58]. In particular, *N*-bromoacetyl-glycosylamine derivatives of xyloglucan
52
53 255 oligosaccharides have been previously demonstrated to be specific active-site affinity labels for
54
55 *endo*-xyloglucanases [59] [36]. A time- and concentration-dependent inactivation of the enzyme
56
57 CjGH5D was observed upon incubation with XXXG-NHCOCH₂Br, which followed pseudo-
58
59
60
61
62
63
64
65

1
2
3
4 first-order kinetics (Fig. 2). The dissociation constant K_i and the irreversible inactivation constant
5
6 k_i for towards *CjGH5D* were 1.78 ± 0.17 mM and 0.17 ± 0.01 min⁻¹, respectively, resulting in a
7
8
9 260 k_i/K_i value (9.3×10^{-2} mM⁻¹.min⁻¹) that was comparable to that previously observed for a
10
11 *Prevotella bryantii* GH5_4 member (*PbGH5A*) [36]. Notably, intact protein mass spectrometry
12
13 of *CjGH5D* following incubation with the inhibitor indicated covalent labelling with 1:1
14
15 stoichiometry and no over-labelling of the enzyme (Additional file 1: Figure S10).
16
17
18
19

20 *CjGH5D* crystallography

21
22 265 A tertiary structure of the catalytic domain of *CjGH5D* was determined at 1.6 Å resolution
23
24 in uncomplexed “apo” form by X-ray crystallography and molecular replacement with *BoGH5A*
25
26 (pdb: 3ZMR, [22]). The overall structure of *CjGH5D* (residues Gly96 to Gln468) is an (β/α)₈
27
28 barrel as is typical for GH5 family members (Fig. 3A). Despite sequence identities in the 25-40%
29
30 range, the structure is similar to the catalytic domains of many GH5 enzymes, most of which are
31
32 annotated as xyloglucanases, glucanases and lichenases, with typical alignment values of
33
34 270 approximately 310 residues aligning with an rmsd of 1.3 Å [60]. For example, the structure used
35
36 for molecular replacement, *BoGH5A*, overlaps with an r.m.s.d. of 1.1 Å over 332 equivalent $C\alpha$
37
38 atoms with 40% identity. Minor differences are observed between the two structures in loops at
39
40 the end of core helices. *BoGH5A* has an extra loop Val170-Gly180 (residues equivalent to
41
42 Ile137-Gly138 in *CjGH5D*) which enables the formation of a hydrogen bond to the -4'-xylosyl
43
44 residue of ligand XXXG (between N Val182 and the sugar ring O atom, *vide infra*).
45
46 275
47
48
49
50
51
52

53 A 1.9 Å-resolution product complex of *CjGH5D* was obtained by soaking crystals with a
54
55 mixture of Glc₁₂-based XyGOs of variable sidechain galactosylation. Here, we anticipated that
56
57 the substrate mixture would be hydrolyzed and that the enzyme would selectively bind the
58
59 oligosaccharide for which it had the best affinity. Commensurate with limit-digest analysis, we
60 280
61
62
63
64
65

1
2
3
4 observed a Glc₄-based oligosaccharide backbone spanning the -4 to -1 subsites for both
5
6 molecules in the asymmetric unit: GXLG in molecule A (with glucose in the -4 subsite and the -
7
8 3'-xylosyl group modelled at occupancies of 0.5 and 0.7 respectively) and GXXG in molecule B
9
10 (here, there was insufficient electron density in the F_o-F_c difference map to allow unambiguous
11
12 modelling of a galactose on the -2'-xylosyl unit). In the -1 subsite, the glucosyl residue interacts
13
14 285 with the catalytic acid base Glu255 (via O1), and nucleophile Glu390 via O2. In addition, O3 is
15
16 with the catalytic acid base Glu255 (via O1), and nucleophile Glu390 via O2. In addition, O3 is
17
18 hydrogen bonded to His208. In molecule B, the equivalent glucose also hydrogen bonds via O2
19
20 to Asn254 and His208 (Fig. 3B & C).
21
22
23

24
25 A second oligosaccharide complex was obtained at 2.1 Å resolution by soaking *CjGH5D*
26
27 290 crystals with the *N*-bromoacetyl affinity label XXXG-NHCOCH₂Br, in which the reagent had
28
29 indeed reacted through attack of the catalytic general acid/base sidechain to displace the bromide
30
31 nucleofuge. In molecule A of the asymmetric unit there is electron density for GXXG-
32
33 NHCOCH₂-*CjGH5D*, whilst in molecule B, XXXG-NHCOCH₂-*CjGH5D* is modeled, but with
34
35 the -3' - and -4'-xylosyl sugars modelled at half occupancy. The carboxyl oxygen of the *N*-acetyl
36
37 moiety forms a hydrogen bond with His323. There are hydrogen bonds between this subsite -1
38
39 295 sugar and the catalytic nucleophile Glu390, and also to His208 and Asn254 (Fig. 3B & D). These
40
41 are similar to interactions observed in the structure of an analogous XXXG-NHCOCH₂-
42
43 *PbGH5A* complex structure (pdb: 5D9P, [36]). Glucose in the -2 subsite is hydrogen bonded via
44
45 O3 to ND2 Asn132 and via O2 to NE1 Trp432; this latter interaction is notably long (ca. 3.2 Å),
46
47
48
49
50
51 300 which may reflect the positioning of the tryptophan as the -1 subsite stacking residue. The
52
53 equivalent Asn/Trp interactions are also seen in related enzymes: Asn28 and Trp324 in the
54
55 XXXG-NHCOCH₂-*PbGH5A* complex (pdb: 5D9P) and Asn165 and Trp472 in the *BoGH5A*-
56
57 XXXG complex (PDB 3ZMR). In addition to Trp432, Trp143 provides aromatic stacking
58
59
60
61
62
63
64
65

1
2
3
4 interactions with the glucose in the -3 subsite (homologous to Trp324 and Trp48 in *PbGH5A*,
5
6 305 and Trp472 and Trp185 in *BoGH5A*, respectively), while Trp209 lies against the -2'-xylosyl
7
8 residue (as does the equivalent Trp252 in *BoGH5A*). This pattern of conserved/ highly invariant
9
10 residues interacting with the xyloglucan chain presumably accounts for the fact that despite
11
12 sharing amino acid identity as low as 30%, these enzymes are all tailored for xyloglucan as a
13
14 substrate (Fig. 4). None of these three GH5D structures exhibits direct interactions with glucosyl
15
16 units in the -3 and -4 subsites with the protein. The -3'-xylosyl unit is tethered by two hydrogen
17
18 310 bonds between O3 and O4 and Asp438, which hold the sugar perpendicular to the orientation of
19
20 the equivalent xylose in the XXXG-NHCOCH₂-*PbGH5A* and *BoGH5A*:XXXG complexes (in
21
22 the latter, the xylose lies parallel to the side chain of Tyr476).
23
24
25
26
27
28

29
30 The covalent adduct formation through the reactivity of the *N*-bromoacetyl reagent is
31
32 315 fascinating given that in the structures observed here, the attack is made by the acid-base
33
34 Glu255, as opposed to the enzymatic nucleophile of the enzymatic reaction, Glu390. This latter
35
36 residue is poised for nucleophilic attack at the anomeric carbon, C1, of the -1 subsite glucoside.
37
38 However, Glu390 is too distant (6-7 Å) from the reactive carbon of the *N*-bromoacetyl moiety,
39
40 and has impossible geometry and steric hindrance, to permit nucleophilic interception. The
41
42 reactive group, however is located in the +1 subsite – some 3.8 Å from C1 – thus can be
43
44 320 fortuitously attacked by the acid/base which is in almost ideal position for S_N2 attack on the
45
46 reactive carbon to displace the bromide. Such a reaction is facilitated, either prior, or subsequent
47
48 to attack by rotation around the CB-CG bond, which leaves the side-chain in a different rotamer
49
50 after the reaction relative to its “normal” position in unreacted complexes (Figure 3D).
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 325 **Mutational analyses of *C. japonicus* GH5_4 genes indicate a complex mode of**
5
6
7 **action for the initial stages of xyloglucan degradation.**
8
9

10
11 Facilitated by knowledge of their broadly similar catalytic properties, we embarked on a
12
13 comprehensive reverse-genetic analysis in an attempt to delineate the biological functions of the
14
15 individual GH5_4 and GH74 endo-xyloglucanases, using recently developed in-frame gene
16
17 deletion techniques [9].
18 330

19
20
21 In-frame deletion mutants were first generated in the XyG gene cluster encoding the three
22
23 *exo*-glycosidases and the TBDT (Fig. 1B) to provide benchmark controls for subsequent analysis
24
25 of *endo*-xyloglucanase deletion mutants. Recapitulating our previous work using insertional
26
27 mutants [13], an in-frame $\Delta xyI31A$ (α -xylosidase) mutant was unable to grow on XyG due to an
28
29 inability to remove non-reducing-terminal xylosyl residues as the first essential step in XyGO
30
31 335
32
33 saccharification (Fig. 5A *cf.* Fig. 1). A ΔCJA_2709 (TBDT) single mutant strain had a
34
35 significant growth defect, presumably resulting from a decreased ability to uptake extracellularly
36
37 produced XyGOs into the periplasm. The deletion of *bgl35A* also attenuated growth, due to an
38
39 inability of the strain to access the full complement of sidechain monosaccharides. As expected,
40
41 growth of the $\Delta afc95A$ (α -L-fucosidase) mutant on tamarind (galacto)xyloglucan was identical to
42
43 the wild-type strain, because this readily available substrate lacks the terminal fucosyl residues
44 340
45
46 typically found in dicot primary cell wall XyG (Fig. 1A). Moreover, all XyG gene cluster mutant
47
48 strains grew similarly to wild type in glucose containing medium (Additional file 1: Figure S11).
49
50
51
52
53

54
55 With these control experiments complete, we next analyzed the effect of deleting the
56
57 345 individual GH5_4- and GH74-encoding genes. Despite original indications by RNAseq analysis
58
59 of a potential lead role for *CjGH5D* in XyG utilization, in-frame deletion of *CJA_3010*
60
61

1
2
3
4 surprisingly did not elicit a statistically significant growth defect (Fig. 5B). Likewise, strains
5
6 containing single in-frame deletions of CJA_3337, CJA_2959, and CJA_2477 grew identically
7
8 to the wild-type strain. Moreover, comprehensive combinatorial mutagenesis did not yield a
9
10 strain with a substantial growth defect for any combination of double, triple, or quadruple
11 350 mutants (Fig. 5C & D). Furthermore, the growth traits (maximum OD and growth rate) of both
12
13 the wild-type and the quadruple deletion mutant were similar when reduced (0.25%) or limiting
14
15 (0.125%) concentrations of XyG were used (Additional file 1: Figure S12 and Table S6).
16
17 Collectively, these results suggest that despite their individual high activities and specificities
18
19 toward XyG, as defined by the biochemical and structural analyses described above, the GH5_4
20
21 and GH74 members are not the only enzymes encoded by *C. japonicus* with sufficient
22
23 xyloglucanase activity to support growth.
24 355
25
26
27
28
29
30
31

32 CjGH5D is predicted to be attached to the exterior face of the outer membrane by N-
33
34 terminal lipidation, while CjGH5E, CjGH5F, and CjGH74 are likely secreted enzymes (*vide*
35
36 *supra*). Thus, we hypothesized that other secreted enzymes, with either predominant or side
37 360 hydrolytic activities toward XyG, may be enabling growth of the quadruple mutant. Deletion of
38
39 the Type-Two Secretion System (T2SS) in the Δgsp mutant has been previously shown to
40
41 abolish the ability of *C. japonicus* to secrete cellulases [11], and constitutes a powerful tool to
42
43 restrict extracellular secretion of CAZymes in general. Interestingly, introduction of the
44
45 CJA_3010 deletion into the Δgsp background resulted in only limited growth attenuation on
46
47 xyloglucan (Additional file 1: Figure S13). With the T2SS extracellular secretion pathway
48
49 365 disabled, the ability of the $\Delta gsp \Delta CJA_3010$ strain to grow on XyG strongly suggests the
50
51 presence of other membrane-bound XGases that effect XyG depolymerization in a
52
53 physiologically relevant manner. These data sharply contrast observations for the human gut
54
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 370 symbiont *Bacteroides ovatus*, for which deletion of a single GH5_4 member from the XyGUL
5
6
7 resulted in complete loss of growth on XyG [22].
8
9

10
11 Predominant xyloglucanase activity has been demonstrated previously in members of
12
13 CAZyme families GH5, GH7, GH9, GH12, GH16, GH44, and GH74, and potentially may
14
15 constitute a side activity in other *endo*- β (1,4)glucanases (cellulases) [23, 61]. Examination of *C.*
16
17 375 *japonicus* genome indicates the presence of multiple GH5 ($n = 15$), GH9 ($n = 3$), and GH16 ($n =$
18
19 9) encoding genes, in addition to the single GH74 member ([8]; for a summary table, see
20
21 <http://www.cazy.org/b776.html>). Further, Deboy, *et al.* [8] predicted that there are
22
23 approximately 45 membrane-bound CAZymes. Although it constitutes a significant undertaking
24
25 that is beyond the scope of the present study, our future investigations will focus on scrutinizing
26
27 these additional CAZymes in the context of XyG utilization by *C. japonicus*.
28
29
30 380

31 32 33 34 35 **Conclusions** 36 37

38
39 We previously proposed a model of XyG utilization by *C. japonicus*, in which an
40
41 extracellular *endo*-xyloglucanase mediates degradation of the polysaccharide to XyGOs for
42
43 uptake via the TBDT, followed by complete hydrolysis to monosaccharides in the periplasm by
44
45 385 the *exo*-glycosidases encoded by the XyG gene cluster [13]. Our present study, combining
46
47 biochemical and reverse-genetic analyses, reveals that the number of actors in the initial cleavage
48
49 event is significantly greater than originally anticipated by bioinformatics. We propose that the
50
51 existence of so many extracellular *endo*-xyloglucanases of apparently overlapping biochemical
52
53 function can be explained by a physiological interplay of secreted reconnaissance enzymes and
54
55
56
57 390 cell-surface-bound, proximal XyG degraders (Fig. 1B).
58
59
60
61
62
63
64
65

1
2
3
4 Thus, the secreted GH74 and two secreted GH5_4 enzymes may act as highly mobile,
5
6 primary “unravellers” of the plant cell, liberating XyG fragments from the lignocellulose matrix
7
8 (Fig. 1C). Indeed, the concept of “sensing” polysaccharidases playing a lead role in generating
9
10 inducers has been previously proposed [62]. As plant cell wall polysaccharide degradation
11
12 advances, more intimate contact between the bacterial cell surface and the substrate may ensue,
13
14 395 engaging the outer-membrane-bound CjGH5D and a more efficient interplay between XyG
15
16 backbone hydrolysis and direct TBDT-mediated uptake of the oligosaccharide products. The
17
18 coordinated capture, hydrolysis, and uptake of partially hydrolyzed polysaccharides as a
19
20 successful competitive strategy has considerable precedent in the Polysaccharide Utilization Loci
21
22
23
24
25
26 400 of the Bacteroidetes [63]. Moreover, the need to initiate cell wall “unravelling” has been
27
28 suggested to explain why saprophytes such as *C. thermocellum*, which are unable to utilize
29
30 xyloglucan or xylan for growth, contain *endo*-xyloglucanases and *endo*-xylanases within their
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

405 **Methods**

8 **Transcriptomic analysis**

11 RNAseq sampling and analysis was performed as previously described [9, 12]. Briefly, *C.*
12 *japonicus* cultures were grown in 500 mL flasks at 30 °C shaking at 200 RPM. OD₆₀₀ was
13 measured every hour to monitor growth and samples were taken during exponential and
14
15 stationary phase. Within two minutes of sampling metabolism was stopped using a
16
17
18 410 phenol/ethanol solution (5%/ 95%). The samples were immediately pelleted by centrifugation at
19
20
21
22
23 8000 *g* at 4 °C for five minutes. The supernatant was discarded and cell pellets were then flash
24
25 frozen using a dry ice/ethanol bath and stored at -80 °C. RNA extraction, library preparation, and
26
27 sequencing was performed by GeneWIZ (South Plainfield, NJ). Illumina HiSeq2500 was
28
29
30 415 performed as 50 bp single-reads with at least 10 million reads generated per sample. The raw
31
32 data have been submitted to the NCBI Gene Expression Omnibus (Accession GSE109594).
33
34
35

36 **Bioinformatic analysis**

38 The full-length proteins encoded by ORFs CJA_3010 (*CjGH5D*), CJA_3337 (*CjCBM2-*
39
40
41
42
43
44 420 *CBM10-GH5E*) and CJA_2959 (*CjFN3-GH5F*) in *C. japonicus* genome were screened for the
45
46 presence of a signal peptide using SignalP 4.0 [32] and LipoP 1.0 [65]. The modular architecture
47
48 of the three enzymes was obtained from BLASTP analysis and additional alignment with
49
50 representative GH and CBM modules from the CAZy Database [7] using ClustalW [66].
51

52 **Cloning of cDNA encoding protein modules**

54 cDNA encoding the full-length enzymes *CjSRL-GH5D*, *CjCBM2-CBM10-GH5E* and
55
56
57 425 *CjFN3-GH5F*, in addition to the catalytic domains *CjGH5D*, *CjGH5E* and *CjGH5F* were PCR
58
59 amplified from *C. japonicus* genomic DNA; all constructs were designed such that the native
60
61

1
2
3
4 predicted signal peptide was removed (PCR primers are listed in Additional file 1: Table S7).

5
6 The amplified *CjSRL-GH5D*, *CjCBM2-CBM10-GH5E*, *CjFN3-GH5F*, *CjGH5D* and *CjGH5F*
7
8 products were double-digested with *NheI* and *XhoI*, gel purified and ligated to the respective
9
10 sites of *pET28a* to fuse an N-terminal 6x His-Tag. The amplified *CjGH5E* product was ligated in
11 430 an *SspI* linearized pMCSG53 vector using Ligation Independent Cloning (LIC) strategy [67].
12
13
14 Successful cloning was confirmed by PCR and plasmid DNA sequencing. Q5 high fidelity DNA
15
16 polymerase was used for all the PCR amplifications.
17
18
19
20
21

22 **Gene expression and protein purification**

23
24 435 Constructs were individually transformed into the chemically competent *E. coli* Rosetta DE3
25
26 cells. Colonies were grown on LB solid media containing kanamycin (50 $\mu\text{g mL}^{-1}$) and
27
28 chloramphenicol (30 $\mu\text{g mL}^{-1}$) [*CjSRL-GH5D*, *CjCBM2-CBM10-GH5E*, *CjFN3-GH5F*,
29
30 *CjGH5D* and *CjGH5F*], or containing ampicillin (50 $\mu\text{g mL}^{-1}$) and chloramphenicol (30 $\mu\text{g mL}^{-1}$)
31
32 [*CjGH5E*]. One colony of the transformed *E. coli* cells was inoculated in 5 mL of LB medium
33
34 containing the same antibiotics and grown overnight at 37 °C (200 rpm). The whole overnight
35
36 440 culture was used to inoculate 500 mL of TB liquid medium containing the proper antibiotics.
37
38
39 Cultures were grown at 37 °C (200 rpm) until $D_{600} = 0.6$. Overexpression was induced by adding
40
41 IPTG to a final concentration of 0.1 mM. After induction, cultures were grown overnight at 16
42
43 °C (200 rpm). Cultures were then centrifuged and pellets were resuspended in 5 mL of *E. coli*
44
45 445 lysis buffer containing 20 mM HEPES, pH 7.0, 500 mM NaCl, 40 mM imidazole, 5% glycerol, 1
46
47 mM DTT and 1 mM PMSF. Cells were then disrupted by sonication and the clear supernatant
48
49
50 was separated by centrifugation at 4 °C (4400 rpm for 45 minutes). Recombinant proteins were
51
52 purified from the clear soluble lysates using a Ni^{+2} - affinity column utilizing a gradient elution
53
54 up to 100% elution buffer containing 20 mM HEPES, pH 7.0, 100 mM NaCl, 500 mM
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4 450 imidazole, and 5% glycerol in an FPLC system. Purity of the recombinant proteins was
5
6 determined by visualizing the protein contents of the fractions on SDS-PAGE. Pure fractions
7
8 were pooled, concentrated, and buffer exchanged against 50 mM phosphate buffer (pH 7.0)
9
10 containing 10% glycerol. Protein concentrations were then determined using Epoch Micro-
11
12 Volume Spectrophotometer System (BioTek®,USA) at 280 nm, and identities of the expressed
13
14 proteins were confirmed by intact mass spectrometry [68]. Purified proteins were then aliquoted
15
16 455 and stored at -80 °C until needed.
17
18
19
20
21

22 Carbohydrate sources

23
24 Tamarind seed XyG, konjac glucomannan (KGM), barley β -glucan (BBG), wheat flour
25
26 arabinoxylan, and beechwood xylan were purchased from Megazyme® (Bray, Ireland).
27
28
29 460 Hydroxyethylcellulose (HEC) was purchased from Amresco® (Solon, USA). Carboxymethyl
30
31 cellulose was purchased from Acros Organics (New Jersey, USA). Guar gum was purchased
32
33 from Sigma Aldrich® (St. Louise, USA). Xanthan gum was purchased from Spectrum® (New
34
35 Brunswick, USA). 2-Chloro-4-nitrophenyl (CNP)- β -D-celotrioside (GGG- β -CNP) and CNP- β -
36
37 D-celotetraoside (GGGG- β -CNP) were purchased from Megazyme®. XXXG- β -CNP and
38
39 465 XLLG- β -CNP were prepared as previously described [69]. Glc₄-based XyGOs (XXXG, XLXG,
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

52 Carbohydrate analytics

53
54
55 470 High Performance Anion-Exchange Chromatography with Pulsed Amperometric Detection
56
57 (HPAEC-PAD) was performed on a Dionex ICS-5000 DC HPLC system operated by the
58
59 Chromeleon software version 7 (Dionex) using a Dionex Carbopac PA200 column. Solvent A
60
61
62
63
64
65

1
2
3
4 was double-distilled water, solvent B was 1 M sodium hydroxide (NaOH), and solvent C was 1
5
6 M sodium acetate (NaOAc). The gradient used was: 0–4 min, 10% solvent B and 2.5% solvent
7
8 C; 4–24 min, 10% B and a linear gradient from 2.5–25% C; 24–24.1 min, 50% B and 50% C;
9 475 24.1 – 25 min, an exponential gradient of NaOH and NaOAc back to initial conditions; and 25–
10
11 31 min, initial conditions.
12
13
14
15
16

17 Matrix-assisted laser desorption ionization-time of flight (MALDI-TOF) was performed on a
18
19 Bruker Daltonics Autoflex System (Billerica, USA). The matrix, 2,5-dihydroxy benzoic acid,
20
21 480 was dissolved in 50% methanol in water to a final concentration of 10 mg mL⁻¹. Oligosaccharide
22
23 samples were mixed 1:1 (v/v) with the matrix solution. One µl of this solution was placed on a
24
25 Bruker MTP 384 ground steel MALDI plate and left to air dry for two hours prior to analysis.
26
27
28
29

30 Enzyme kinetic analysis 31

32 All enzyme activities toward polysaccharides were determined using a bicinchoninic acid
33
34 485 (BCA) reducing-sugar assay [70]. The effect of temperature on xyloglucanase activity was
35
36 determined by incubating the recombinant catalytic domain: *CjGH5D* (0.098 µg), *CjGH5E*
37
38 (0.086 µg), *CjGH5F* (0.017 µg) with tamarind seed xyloglucan at a final concentration of 1 mg
39
40 mL⁻¹. Citrate buffer (pH 6, *CjGH5D* and *CjGH5F*) or phosphate buffer (pH 7.5, *CjGH5E*) was
41
42 used to a final concentration of 50 mM in a total reaction volume of 200 µL. Reaction mixtures
43
44 490 were incubated for 10 minutes at temperatures ranging from 25 °C to 80 °C prior to the BCA
45
46 assay. To determine the pH-rate profile, the same XyG concentration was incubated with the
47
48 same enzyme amounts, except for *CjGH5D* (0.049 µg), for 10 minutes at 50 °C (*CjGH5D*,
49
50 *CjGH5F*), or 55 °C (*CjGH5E*), with 50 mM final concentration of the following buffers: citrate
51
52 (pH 3-6.5), phosphate (pH 6.5-8), and glycine (pH 8.5-9).
53
54
55
56
57
58
59
60
61
62

1
2
3
4 495 For qualitative activity assessment against the other polysaccharide substrates, 1 μg of each
5
6 recombinant enzyme was added to XyG, HEC, CMC, BBG, KGM, wheat flour arabinoxylan,
7
8 beechwood xylan, xanthan gum, and guar gum to a final concentration of 2 mg mL^{-1} in 200 μL
9
10 reaction volumes containing 50 mM phosphate buffer (pH 7.5: *CjGH5D* and *CjGH5E*, or pH 7:
11
12 *CjGH5F*). Mixtures were then incubated at 50 $^{\circ}\text{C}$ (*CjGH5D* and *CjGH5F*) or 55 $^{\circ}\text{C}$ (*CjGH5E*) for
13
14
15
16 500 10 minutes before the generated reducing ends were detected using BCA assay.
17
18
19

20 To determine specific activity values of *CjGH5* enzymes toward XyG, final concentration of
21
22 0.75, 2.59, and 0.71 nM of the recombinant purified catalytic modules *CjGH5D*, *CjGH5E*, and
23
24 *CjGH5F*, respectively, was incubated with tamarind seed XyG (1 mg mL^{-1}) in 200 μL reaction
25
26 mixtures containing 50 mM phosphate buffer (pH 7.5: *CjGH5D* and *CjGH5E*, or pH 7:
27
28 *CjGH5F*). Likewise, specific activity values of *CjGH5* enzymes toward HEC were obtained by
29
30 505 *CjGH5E* and *CjGH5F* at a final concentration of 1.04 and 0.65 μM , respectively, with
31
32 incubating *CjGH5E* and *CjGH5F* at a final concentration of 1.04 and 0.65 μM , respectively, with
33
34 2 mg mL^{-1} HEC in 200 μl reaction mixtures containing 50 mM phosphate buffer (pH 7.5:
35
36 *CjGH5E* or pH 7: *CjGH5F*). For specific activity toward CMC, final concentration of 1.04 μM
37
38 of the purified catalytic module *CjGH5E* was incubated with CMC (2 mg mL^{-1}) in 200 μL
39
40
41 510 reaction volume containing 50 mM phosphate buffer (pH 7.5). All reaction mixtures were
42
43 incubated at 50 $^{\circ}\text{C}$ (*CjGH5D* and *CjGH5F*) or 55 $^{\circ}\text{C}$ (*CjGH5E*) for 10 minutes prior to the BCA
44
45
46
47 assay and all assays were performed in triplicates.
48
49

50 To determine Michaelis-Menten parameters for XyG, eight different concentrations of XyG
51
52 solutions were used over the range 0.025 to 1 mg mL^{-1} . The recombinant enzyme *CjGH5D*
53
54 515 (0.007 μg), *CjGH5E* (0.022 μg), and *CjGH5F* (0.006 μg) was individually incubated with each
55
56
57 XyG concentration at 50 $^{\circ}\text{C}$ (*CjGH5D* and *CjGH5F*) or 55 $^{\circ}\text{C}$ (*CjGH5E*) for 10 min in a 200 μl
58
59
60
61
62
63
64
65

1
2
3
4 final reaction mixture containing 50 mM phosphate buffer (pH 7.5: *CjGH5D* and *CjGH5E*, or pH
5
6
7 7: *CjGH5F*). K_m and k_{cat} values were determined by non-linear fitting of the Michaelis-Menten
8
9 equation to the data in Sigmaplot® (Systat software Inc.)
10

11
12
13 520 To identify Michaelis-Menten constants for the chromogenic substrates, different dilution
14 series were established to give final concentration ranges of 0.0625- 8 mM (CNP- β -GGG),
15
16 0.0625- 8 mM (CNP- β -GGGG), 0.002-4 mM (XXXG- β -CNP), and 0.002- 2 mM (XLLG- β -
17
18 CNP). Substrate mixtures (225 μ L) containing 50 mM phosphate buffer with the optimum pH of
19
20 the enzyme were pre-incubated for 10 minutes at the optimum temperature of the enzyme (*vide*
21
22 *supra*). 25 μ L of 10X *CjGH5D* (to give 30-3800 nM final concentration according to the tested
23
24 525 substrate), *CjGH5E* (3- 260 nM), and *CjGH5F* (4- 650 nM) was added to the substrate mixtures
25
26 before the release of the aglycone was continuously monitored by measuring the change in
27
28 absorbance at 405 nM for 2 minutes in a Cary50 UV-visible spectrophotometer (Varian). CNP
29
30 molar extinction coefficients were determined to be 17288 M⁻¹cm⁻¹ in 50 mM phosphate buffer
31
32 pH 7 and 17741 M⁻¹cm⁻¹ in 50 mM phosphate buffer pH 7.5.
33
34
35
36
37 530
38
39
40

41 Enzyme product analysis

42
43 To determine the limit-digest products of the *CjGH5s*, 5 μ g of each recombinant enzyme
44
45 was incubated with tamarind seed XyG at final concentration of 0.25 mg mL⁻¹ for 7 hours (40
46
47 °C) in a 200 μ L reaction mixture that contained 50 mM phosphate buffer of the optimum pH of
48
49 the tested enzyme (pH 7.5: *CjGH5D* and *CjGH5E*, or pH 7: *CjGH5F*). The reaction mixture was
50
51 535 then diluted 5 times prior to product analysis by HPAEC-PAD. To determine the mode of action
52
53 of the enzyme, 0.01 μ g of *CjGH5s* was incubated at 40 °C with 1 mg mL⁻¹ final concentration of
54
55 tamarind seed XyG in 200 μ L reaction volumes containing the same buffers used in limit-digest
56
57
58
59
60
61
62
63
64
65

1
2
3
4 analysis. The reaction was stopped at different time points by adding 100 μL of NH_4OH .

5
6
7 540 Reaction mixtures were then diluted 2 times with water prior to product analysis by HPAEC-
8
9 PAD.

10 11 12 **Inhibition kinetics and active-site labeling**

13
14
15 Inhibition kinetic parameters were determined as previously described [59]. Briefly, a final
16
17 concentration of 0.23 μM of CjGH5D was incubated with a series of different concentrations
18
19 (0.5-16 mM) of XXXG-NHCOCH₂Br at 40 °C in 20 mM phosphate buffer (pH 7.5) for up to 90
20 545 minutes. BSA to a final concentration of 0.1 mg. mL⁻¹ was added to the inhibition mix to prevent
21
22 the non-specific loss of activity. Small samples (10 μL) of the incubate were periodically diluted
23
24 1:100 in 20 mM phosphate buffer (pH 7.5), and 100 μL of the diluted incubate was added to 100
25
26 μL of the pre-incubated substrate XXXG-CNP at 40 °C (0.1 mM final substrate concentration in
27
28 the assay). Residual activities of the enzyme was determined by measuring the rate of the release
29
30 of the chromophore 2-chloro-4-nitrophenolate [69] at 405 nm in Agilent Cary 60 UV-Vis
31
32 550 Spectrophotometer. Initial-rate kinetics were measured in the strictly linear range of the enzyme.
33
34 Equations 1 and 2 were used to determine K_i and k_i values by non-linear regression curve-fitting
35
36 using OriginPro 2015 software as previously described [59].
37
38
39
40
41
42
43

44 555
$$V = V_0 \exp(-k_{\text{app}}t) + y_{\text{offset}} \quad (1)$$

45
46
47
48
$$k_{\text{app}} = \frac{k_i [\text{I}]}{K_i + [\text{I}]} \quad (2)$$

49
50
51
52
53
54
55
56
57
58
59
60
61

1
2
3
4 Intact protein masses were determined on a Waters Xevo Q-TOF with a nanoACQUITY
5
6 560 UPLC system, according to the method previously published [68], with 2.5 mM inhibitor and
7
8 4.52 μ M of enzyme.
9

10 11 12 **Construction of *C. japonicus* mutants and growth conditions** 13 14

15 In-frame deletion mutants were made as previously described [9]. Briefly, 500bp regions up-
16
17 and down-stream of the genes of interest were amplified by PCR (CJA_3010, CJA_3337, and
18
19 CJA_2959) or synthesized by GeneWIZ (South Plainfield, NJ) (CJA_2477) and assembled into
20 565 pK18*mobsacB* by the method of Gibson, *et al.* [71]. Deletions were confirmed by PCR. For a
21
22 complete list of primers used see Additional file 1: Table S8. Cultures were grown at 30 °C with
23
24 200 RPM shaking in MOPS minimal media containing 0.25% (w:v) glucose or 0.5% (w:v)
25
26 tamarind seed xyloglucan (Megazyme) as the sole carbon source in 18 mm test tubes or in 96
27
28 well flat bottom polystyrene plates (Corning). Growth was measured using a Spec20D+
29
30 spectrophotometer (Thermo Scientific) or a Tecan Plate reader (Tecan, Switzerland). All
31
32 570 experiments were performed in biological triplicate. Statistical analysis was performed using
33
34 Graphpad Prism 6 software package (La Jolla, CA) where appropriate.
35
36
37
38
39
40
41
42

43 **Crystallization, X-ray crystallography and structure solution** 44

45 575 *CjGH5D* was crystallized in sitting drops by the Vapour Diffusion method, using protein at
46
47 21 mg mL⁻¹ in 50 mM sodium citrate pH 6.5, 10 % glycerol over a well solution comprised of
48
49 1.9 M ammonium sulfate, 0.1 M 2-(N-morpholino)ethanesulfonic acid (MES) pH 5.5. The drop
50
51 consisted of 0.5 μ l enzyme, 0.1 μ l seed stock and 0.4 μ l well solution, and the seed stock was
52
53 prepared by vortexing crystals, grown in 1.4 M ammonium sulfate, 0.1 M MES pH 5.5, 1 %
54
55 (w/v) polyethylene glycol 1,000, in an Eppendorf tube with a polystyrene bead. Crystals were
56
57
58 580
59
60
61
62
63
64
65

1
2
3
4 harvested into liquid nitrogen using nylon CryoLoops™ (Hampton Research). A non-ligand
5
6 complexed “apo” dataset was collected from a crystal after immersing for a few minutes in a
7
8 cryoprotectant solution, comprised of the mother liquor supplemented with 20 % (v/v) glycerol.
9
10
11 Data were collected at Diamond beamline I04, and processed using *DIALS* [72], and scaled using
12
13
14 585 *AIMLESS* [73] to 1.6 Å. The space group is P2₁2₁2₁ with unit cell dimensions 55.0, 96.4, 159.0
15
16 Å, and there are 2 molecules in the asymmetric unit.
17
18
19

20 The structure was solved by molecular replacement using *Phaser* [74] using residues 138 to
21
22 500 of PDB entry 3zmr as the search model [22], which align to residues 106 to 464 of CjGH5D,
23
24 with which they share 38% identity (using the program *lalign* from the FASTA package [75]).
25
26
27 590 The structure was built automatically using *Buccaneer* [76] and refined using cycles of manual
28
29 model rebuilding using *Coot* [77] followed by refinement with *REFMAC* [78], including cycles
30
31 using anisotropic B-factor refinement. In addition to the two protein chains, there are 20
32
33 molecules of glycerol, 5 sulfate ions and 3 molecules of PEG (introduced from the seed stock
34
35 solution). Data collection and refinement statistics for all structures are given in Additional file
36
37
38
39 595 1: Table S9.
40
41
42

43 A crystal of CjGH5D, grown as above over a well solution comprised of 2.3 M ammonium
44
45 sulfate, 0.1 M MES pH 5.5, was soaked for 27.5 hours in 1.85 M ammonium sulfate, 0.1 M MES
46
47 pH 5.5 with 4.5 mM XXXG-NHCOCH₂Br, and fished directly into liquid nitrogen. Data were
48
49 collected at Diamond beamline I03, and processed using *DIALS* [72]. After scaling with
50
51
52 600 *AIMLESS* [73] the data were cut off at a resolution of 2.1 Å, as although the X-ray images
53
54 showed significant spot smearing, the R_{merge} and CC_{1/2} values were good (6.2% overall, 54.7% in
55
56 the outer shell and 0.998 overall, 0.933 in the outer shell respectively).
57
58
59
60
61
62
63
64
65

1
2
3
4 Crystals grown under similar conditions, over a well containing 1.6 M ammonium sulfate,
5
6 0.1 M MES pH 5.5, were soaked in the presence of a mixture of Glc₁₂-based XyGOs (produced
7
8
9 605 as described in [46] at a concentration of 5 mM for 5 hours, before fishing into liquid nitrogen
10
11 via a cryoprotectant solution, as for the apo crystal. A dataset was collected at Diamond
12
13 beamline I04 and processed using *DIALS* and scaled using *AIMLESS* to 1.9 Å.
14
15

16
17 Both ligand structures were solved initially using the apo structure as a model for *REFMAC*,
18
19 and the ligand was placed after the protein chain had been rebuilt (using cycles of Coot
20
21
22 610 interspersed with refinement in *REFMAC*) and some water molecules added. All models were
23
24 validated using *MolProbity* [79] and the sugar conformations of the ligand in the complex
25
26 structures were checked using Privateer [80]. Problems with diffraction anisotropy in both ligand
27
28 datasets limited the possibility of refining the structures to R/Rfree lower than 0.22/0.28 and
29
30 0.23/0.30 for the complexes with XXXG-NHCOCH₂ and GXLG (produced after the Glc₁₂ soak)
31
32
33
34 615 respectively.
35
36
37
38

39 **Additional file**

40
41
42 Additional file 1. Supporting Tables 1-9 and Supporting Figures 1-13.
43
44

45 **Abbreviations**

46
47
48
49 XyG, tamarind seed xyloglucan; XyGO, xylogluco-oligosaccharides; CAZy, carbohydrate active
50
51 620 enzymes; GH5, glycoside hydrolase family 5; CBM, carbohydrate binding module; SRL, serine
52
53 rich linker; HPAEC-PAD, high performance anion exchange chromatography- pulsed
54
55 amperometric detector; MALDI-TOF, matrix assisted laser desorption ionization- time of flight;
56
57 KGM, konjac glucomannan; BBG, barley β-glucan; HEC, hydroxyethylcellulose; CNP 2-chloro-
58
59
60
61
62
63
64
65

1
2
3
4 4-nitrophenyl; BCA, biconchonic acid; TBDT, TonB-dependent transporter; XyGUL,
5
6 625 xyloglucan utilization locus; MES, 0.1 M 2-(N-morpholino)ethanesulfonic acid.
7
8
9

10 11 **Authors' contributions**

12
13
14 MA performed the bioinformatic analysis, recombinant protein production, and biochemical
15
16 characterization. CN performed the RNAseq analysis and all mutational analysis. NJ synthesized
17
18 the active-site affinity label and, together with MA, performed inhibition and protein MS
19
20 630 analysis. JGG designed the RNAseq and mutational analysis experiments, and assisted in
21
22 analyzing the RNAseq and mutational data. WAO solved X-ray structures which were analysed
23
24 under supervision of GJD. HB devised the overall study and supervised research. All authors
25
26 contributed to writing the article.
27
28
29
30
31
32

33 **Author details**

34
35
36 635 ¹Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, BC,
37
38 V6T 1Z4, Canada. ²Department of Chemistry, University of British Columbia, 2036 Main Mall,
39
40 Vancouver, British Columbia V6T 1Z1, Canada. ³Department of Biological Sciences,
41
42 University of Maryland, Baltimore County, Baltimore MD 21250, USA. ⁴Department of
43
44 Chemistry, University of York, Heslington, York YO10 5DD, UK. ⁵Department of
45
46 640 Biochemistry and Molecular Biology, University of British Columbia, 2350 Health Sciences
47
48 Mall, Vancouver, British Columbia V6T 1Z3, Canada. ⁶Department of Botany, University of
49
50 British Columbia, 6270 University Blvd., Vancouver, British Columbia V6T 1Z4, Canada
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Acknowledgments

We thank Diamond Light Source (Harwell, UK) for access to beamlines I03 and I04 (proposal mx9948) that contributed to the results presented here. Additionally, we thank E.C. Monge (Gardner Lab, UMBC) for assistance with *C. japonicus* growth experiments.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The structural datasets generated during the current study are available in the RCSB protein repository under the following PDB IDs: *CjGH5D*: 5OYC, XXXG-NHCOCH₂-*CjGH5D*: 5OYD, *CjGH5D*-GXLG: 5OYE (<https://www.rcsb.org/>). The RNAseq datasets generated during the current study are available in the NCBI Gene Expression Omnibus (GEO, <https://www.ncbi.nlm.nih.gov/geo/>) repository under accession GSE109594. Other datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethical approval and consent to participate

Not applicable.

Funding

Work at the University of British Columbia was generously supported by the Natural Sciences and Engineering Research Council of Canada (Discovery Grant and Strategic Network Grant for the “NSERC Industrial Biocatalysis Network”, <http://www.ibnet.ca/>), the Michael Smith Laboratories (faculty start-up and post-graduate funding), the Canada Foundation for Innovation, and the British Columbia Knowledge Development Fund. Work at UMBC was supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research under Award Number DE-SC0014183. Additional support to CEN came from a NIGMS Initiative for Maximizing Student Development under Award Number R25-GM55036. H.B. and J.G.G. thank the UMBC Eminent Scholar program for catalyzing the ongoing collaboration between the UMBC and UBC groups. GJD is supported through the Royal Society "Ken Murray" Research Professorship.

Disclaimer

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views

1
2
3
4 and opinions of authors expressed herein do not necessarily state or reflect those of the United
5
6 States Government or any agency thereof.
7
8
9

10 11 **References**

- 12
13
14 685 1. Parajuli R, Dalgaard T, Jorgensen U, Adamsen APS, Knudsen MT, Birkved M, Gylling
15 M, Schjorring JK: **Biorefining in the prevailing energy and materials crisis: a review**
16 **of sustainable pathways for biorefinery value chains and sustainability assessment**
17 **methodologies.** *Renewable & Sustainable Energy Reviews* 2015, **43**:244-263.
18
19 2. Tuck CO, Perez E, Horvath IT, Sheldon RA, Poliakoff M: **Valorization of Biomass:**
20 690 **Deriving More Value from Waste.** *Science* 2012, **337**:695-699.
21
22 3. Horn SJ, Vaaje-Kolstad G, Westereng B, Eijsink VGH: **Novel enzymes for the**
23 **degradation of cellulose.** *Biotechnology for Biofuels* 2012, **5**.
24
25 4. Li LL, McCorkle SR, Monchy S, Taghavi S, van der Lelie D: **Bioprospecting**
26 695 **metagenomes: glycosyl hydrolases for converting biomass.** *Biotechnology for Biofuels*
27 2009, **2**.
28
29 5. Gardner JG: **Polysaccharide degradation systems of the saprophytic bacterium**
30 ***Cellvibrio japonicus*.** *World Journal of Microbiology & Biotechnology* 2016, **32**:121.
31
32 6. Hazlewood GP, Gilbert HJ: **Structure and function analysis of *Pseudomonas* plant cell**
33 **wall hydrolases.** *Biochemical Society Transactions* 1998, **26**:185-190.
34
35 700 7. Lombard V, Ramulu HG, Drula E, Coutinho PM, Henrissat B: **The carbohydrate-active**
36 **enzymes database (CAZy) in 2013.** *Nucleic Acids Research* 2014, **42**:D490-D495.
37
38 8. Deboy RT, Mongodin EF, Fouts DE, Tailford LE, Khouri H, Emerson JB, Mohamoud Y,
39 705 **Watkins K, Henrissat B, Gilbert HJ, Nelson KE: Insights into plant cell wall**
40 **degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*.**
41 *Journal of Bacteriology* 2008, **190**:5455-5463.
42
43 9. Nelson CE, Gardner JG: **In-frame deletions allow functional characterization of**
44 **complex cellulose degradation phenotypes in *Cellvibrio japonicus*.** *Applied and*
45 710 *Environmental Microbiology* 2015, **81**:5968-5975.
46
47 10. Gardner JG, Keating DH: **Genetic and functional genomic approaches for the study of**
48 **plant cell wall degradation in *Cellvibrio japonicus*.** *Methods in Enzymology* 2012,
49 **510**:331-347.
50
51 11. Gardner JG, Keating DH: **Requirement of the Type II Secretion System for**
52 **Utilization of Cellulosic Substrates by *Cellvibrio japonicus*.** *Applied and*
53 *Environmental Microbiology* 2010, **76**:5079-5087.
54
55 715 12. Gardner JG, Crouch L, Labourel A, Forsberg Z, Bukhman YV, Vaaje-Kolstad G, Gilbert
56 HJ, Keating DH: **Systems biology defines the biological significance of redox-active**
57 **proteins during cellulose degradation in an aerobic bacterium.** *Molecular*
58 *Microbiology* 2014, **94**:1121-1133.
59
60 13. Larsbrink J, Thompson AJ, Lundqvist M, Gardner JG, Davies GJ, Brumer H: **A complex**
61 **gene locus enables xyloglucan utilization in the model saprophyte *Cellvibrio***
62 ***japonicus*.** *Mol Microbiol* 2014, **94**:418-433.
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

14. Carpita N, McCann M: **The Cell Wall**. In *Biochemistry and Molecular Biology of Plants*. Edited by Buchanan BB, Gruissem W, Jones RL. Somerset, NJ: John Wiley & Sons; 2000: 55-108

725 15. Fangel JU, Ulvskov P, Knox JP, Mikkelsen MD, Harholt J, Popper ZA, Willats WGT: **Cell wall evolution and diversity**. *Frontiers in Plant Science* 2012, **3**.

16. Vogel J: **Unique aspects of the grass cell wall**. *Current Opinion in Plant Biology* 2008, **11**:301-307.

17. Scheller HV, Ulvskov P: **Hemicelluloses**. *Annual Review of Plant Biology* 2010, **61**:263-289.

730 18. Hsieh YSY, Harris PJ: **Xyloglucans of monocotyledons have diverse structures**. *Molecular Plant* 2009, **2**:943-965.

19. Hoffman M, Jia ZH, Pena MJ, Cash M, Harper A, Blackburn AR, Darvill A, York WS: **Structural analysis of xyloglucans in the primary cell walls of plants in the subclass Asteridae**. *Carbohydrate Research* 2005, **340**:1826-1840.

735 20. Tuomivaara ST, Yaoi K, O'Neill MA, York WS: **Generation and structural validation of a library of diverse xyloglucan-derived oligosaccharides, including an update on xyloglucan nomenclature**. *Carbohydrate Research* 2015, **402**:56-66.

740 21. Pauly M, Keegstra K: **Biosynthesis of the Plant Cell Wall Matrix Polysaccharide Xyloglucan**. *Annu Rev Plant Biol* 2016, **67**:235-259.

22. Larsbrink J, Rogers TE, Hemsworth GR, McKee LS, Tauzin AS, Spadiut O, Klintner S, Pudlo NA, Urs K, Koropatkin NM, et al: **A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes**. *Nature* 2014, **506**:498-502.

745 23. Attia MA, Brumer H: **Recent structural insights into the enzymology of the ubiquitous plant cell wall glycan xyloglucan**. *Curr Opin Struct Biol* 2016, **40**:43-53.

24. Larsbrink J, Izumi A, Ibatullin FM, Nakhai A, Gilbert HJ, Davies GJ, Brumer H: **Structural and enzymatic characterization of a glycoside hydrolase family 31 alpha-xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification**. *Biochemical Journal* 2011, **436**:567-580.

750 25. Silipo A, Larsbrink J, Marchetti R, Lanzetta R, Brumer H, Molinaro A: **NMR Spectroscopic analysis reveals extensive binding interactions of complex xyloglucan oligosaccharides with the *Cellvibrio japonicus* glycoside hydrolase family 31 alpha-xylosidase**. *Chemistry-A European Journal* 2012, **18**:13395-13404.

755 26. Nelson CE, Attia MA, Rogowski A, Morland C, Brumer H, Gardner JG: **Comprehensive functional characterization of the Glycoside Hydrolase Family 3 enzymes from *Cellvibrio japonicus* reveals unique metabolic roles in biomass saccharification**. *Environ Microbiol* 2017.

27. Attia M, Stepper J, Davies GJ, Brumer H: **Functional and structural characterization of a potent GH74 endo-xyloglucanase from the soil saprophyte *Cellvibrio japonicus* unravels the first step of xyloglucan degradation**. *FEBS J* 2016, **283**:1701-1719.

760 28. Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B: **Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5)**. *Bmc Evolutionary Biology* 2012, **12**.

765 29. Raman B, McKeown CK, Rodriguez M, Brown SD, Mielenz JR: **Transcriptomic analysis of *Clostridium thermocellum* ATCC 27405 cellulose fermentation**. *Bmc Microbiology* 2011, **11**.

- 1
2
3
4
5
6
7
8 770
9
10
11
12
13
14 775
15
16
17
18
19
20 780
21
22
23
24
25
26 785
27
28
29
30
31
32 790
33
34
35
36
37
38 795
39
40
41
42
43
44
45 800
46
47
48
49
50
51 805
52
53
54
55
56
57 810
58
59
60
61
62
63
64
65
30. Riederer A, Takasuka TE, Makino S, Stevenson DM, Bukhman YV, Elsen NL, Fox BG: **Global Gene Expression Patterns in *Clostridium thermocellum* as Determined by Microarray Analysis of Chemostat Cultures on Cellulose or Cellobiose.** *Applied and Environmental Microbiology* 2011, **77**:1243-1253.
 31. Paetzel M, Karla A, Strynadka NCJ, Dalbey RE: **Signal peptidases.** *Chemical Reviews* 2002, **102**:4549-4579.
 32. Petersen TN, Brunak S, von Heijne G, Nielsen H: **SignalP 4.0: discriminating signal peptides from transmembrane regions.** *Nature Methods* 2011, **8**:785-786.
 33. Boraston AB, Bolam DN, Gilbert HJ, Davies GJ: **Carbohydrate-binding modules: fine-tuning polysaccharide recognition.** *Biochemical Journal* 2004, **382**:769-781.
 34. Gloster TM, Ibatullin FM, Macauley K, Eklof JM, Roberts S, Turkenburg JP, Bjornvad ME, Jorgensen PL, Danielsen S, Johansen KS, et al: **Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12.** *Journal of Biological Chemistry* 2007, **282**:19177-19189.
 35. dos Santos CR, Cordeiro RL, Wong DWS, Murakami MT: **Structural Basis for Xyloglucan Specificity and alpha-D-Xylp(1 -> 6)-D-Glcp Recognition at the-1 Subsite within the GH5 Family.** *Biochemistry* 2015, **54**:1930-1942.
 36. McGregor N, Morar M, Fenger TH, Stogios P, Lenfant N, Yin V, Xu XH, Evdokimova E, Cui H, Henrissat B, et al: **Structure-Function Analysis of a Mixed-linkage beta-Glucanase/Xyloglucanase from the Key Ruminant Bacteroidetes *Prevotella bryantii* B(1)4.** *Journal of Biological Chemistry* 2016, **291**:1175-1197.
 37. Wong D, Chan VJ, McCormack AA, Batt SB: **A novel xyloglucan-specific endo-beta-1,4-glucanase: biochemical properties and inhibition studies.** *Applied Microbiology and Biotechnology* 2010, **86**:1463-1471.
 38. Matsuzawa T, Saito Y, Yaoi K: **Key amino acid residues for the endo-processive activity of GH74 xyloglucanase.** *Febs Letters* 2014, **588**:1731-1738.
 39. Ravachol J, Borne R, Tardif C, de Philip P, Fierobe H-P: **Characterization of All Family-9 Glycoside Hydrolases Synthesized by the Cellulosome-producing Bacterium *Clostridium cellulolyticum*.** *Journal of Biological Chemistry* 2014, **289**:7335-7348.
 40. Ravachol J, de Philip P, Borne R, Mansuelle P, Maté MJ, Perret S, Fierobe HP: **Mechanisms involved in xyloglucan catabolism by the cellulosome-producing bacterium *Ruminiclostridium cellulolyticum*.** *Sci Rep* 2016, **6**:22770.
 41. Song S, Tang YB, Yang SQ, Yan QJ, Zhou P, Jiang ZQ: **Characterization of two novel family 12 xyloglucanases from the thermophilic *Rhizomucor miehei*.** *Applied Microbiology and Biotechnology* 2013, **97**:10013-10024.
 42. Damasio ARL, Ribeiro LFC, Ribeiro LF, Furtado GP, Segato F, Almeida FBR, Crivellari AC, Buckeridge MS, Souza TACB, Murakami MT, et al: **Functional characterization and oligomerization of a recombinant xyloglucan-specific endo-beta-1,4-glucanase (GH12) from *Aspergillus niveus*.** *Biochimica Et Biophysica Acta-Proteins and Proteomics* 2012, **1824**:461-467.
 43. Master ER, Zheng Y, Storms R, Tsang A, Powlowski J: **A xyloglucan-specific family 12 glycosyl hydrolase from *Aspergillus niger*: recombinant expression, purification and characterization.** *Biochemical Journal* 2008, **411**:161-170.

- 1
2
3
4 44. Powlowski J, Mahajan S, Schapira M, Master ER: **Substrate recognition and hydrolysis by a fungal xyloglucan-specific family 12 hydrolase.** *Carbohydrate Research* 2009, **344**:1175-1179.
- 5
6
7
8 45. Eklof JM, Shojania S, Okon M, McIntosh LP, Brumer H: **Structure-Function Analysis of a Broad Specificity Populus trichocarpa Endo-beta-glucanase Reveals an Evolutionary Link between Bacterial Licheninases and Plant XTH Gene Products.** *Journal of Biological Chemistry* 2013, **288**:15786-15799.
- 9 815
10
11
12
13 46. Martinez-Fleites C, Guerreiro C, Baumann MJ, Taylor EJ, Prates JAM, Ferreira LMA, Fontes C, Brumer H, Davies GJ: **Crystal structures of Clostridium thermocellum xyloglucanase, XGH74A, reveal the structural basis for xyloglucan recognition and degradation.** *Journal of Biological Chemistry* 2006, **281**:24922-24933.
- 14
15 820
16
17
18 47. Yaoi K, Nakai T, Kameda Y, Hiyoshi A, Mitsuishi Y: **Cloning and characterization of two xyloglucanases from Paenibacillus sp. strain KM21.** *Applied and Environmental Microbiology* 2005, **71**:7670-7678.
- 19
20
21 825
22
23
24
25 48. Yaoi K, Kondo H, Hiyoshi A, Noro N, Sugimoto H, Tsuda S, Miyazaki K: **The crystal structure of a xyloglucan-specific endo-beta-1,4-glucanase from Geotrichum sp. M128 xyloglucanase reveals a key amino acid residue for substrate specificity.** *Febs Journal* 2009, **276**:5094-5100.
- 26
27
28 830
29
30
31
32
33 49. Yaoi K, Mitsuishi Y: **Purification, characterization, cDNA cloning, and expression of a xyloglucan endoglucanase from Geotrichum sp. M128.** *Febs Letters* 2004, **560**:45-50.
- 34
35
36
37
38 50. Enkhbaatar B, Temuujin U, Lim JH, Chi WJ, Chang YK, Hong SK: **Identification and Characterization of a Xyloglucan-Specific Family 74 Glycosyl Hydrolase from Streptomyces coelicolor A3(2).** *Applied and Environmental Microbiology* 2012, **78**:607-611.
- 39
40 835
41
42
43
44
45 51. Desmet T, Cantaert T, Gualfetti P, Nerinckx W, Gross L, Mitchinson C, Piens K: **An investigation of the substrate specificity of the xyloglucanase Cel74A from Hypocrea jecorina.** *FEBS Journal* 2007, **274**:356-363.
- 46
47
48
49
50 52. Ariza A, Eklof JM, Spadiut O, Offen WA, Roberts SM, Besenmatter W, Friis EP, Skjot M, Wilson KS, Brumer H, Davies G: **Structure and Activity of Paenibacillus polymyxa Xyloglucanase from Glycoside Hydrolase Family 44.** *Journal of Biological Chemistry* 2011, **286**:33890-33900.
- 51
52 840
53
54
55
56 53. Feng T, Yan K-P, Mikkelsen MD, Meyer AS, Schols HA, Westereng B, Mikkelsen JD: **Characterisation of a novel endo-xyloglucanase (XcXGHA) from Xanthomonas that accommodates a xylosyl-substituted glucose at subsite-1.** *Applied Microbiology and Biotechnology* 2014, **98**:9667-9679.
- 57
58 845
59
60
61
62 54. Yaoi K, Mitsuishi Y: **Purification, characterization, cloning, and expression of a novel xyloglucan-specific glycosidase, oligoxyloglucan reducing end-specific cellobiohydrolase.** *Journal of Biological Chemistry* 2002, **277**:48276-48281.
- 63
64
65 55. Ichinose H, Araki Y, Michikawa M, Harazono K, Yaoi K, Karita S, Kaneko S: **Characterization of an endo-processive-type xyloglucanase having a beta-1,4-glucan-Binding Module and an endo-type xyloglucanase from Streptomyces avermitilis.** *Applied and Environmental Microbiology* 2012, **78**:7939-7945.
- 66
67
68 56. Davies GJ, Wilson KS, Henrissat B: **Nomenclature for sugar-binding subsites in glycosyl hydrolases.** *Biochemical Journal* 1997, **321**:557-559.

- 1
2
3
4 57. Barras F, Bortoligerman I, Bauzan M, Rouvier J, Gey C, Heyraud A, Henrissat B: **Stereochemistry of the hydrolysis reaction catalyzed by endoglucanase Z from *Erwinia chrysanthemi***. *Febs Letters* 1992, **300**:145-148.
58. Gloster TM, Vocadlo DJ: **Developing inhibitors of glycan processing enzymes as tools for enabling glycobiology**. *Nature Chemical Biology* 2012, **8**:683-694.
- 860 59. Fenger TH, Brumer H: **Synthesis and Analysis of Specific Covalent Inhibitors of endo-Xyloglucanases**. *Chembiochem* 2015, **16**:575-583.
60. Krissinel E, Henrick K: **Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions**. *Acta Crystallographica Section D-Biological Crystallography* 2004, **60**:2256-2268.
- 865 61. Eklof JM, Ruda MC, Brumer H: **Distinguishing xyloglucanase activity in endo-beta(1 - > 4)glucanases**. *Methods in Enzymology* 2012, **510**:97-120.
62. Zhao YX, Chany CJ, Sims PFG, Sinnott ML: **Definition of the substrate specificity of the 'sensing' xylanase of *Streptomyces cyaneus* using xylooligosaccharide and celooligosaccharide glycosides of 3,4-dinitrophenol**. *Journal of Biotechnology* 1997, **57**:181-190.
- 870 63. Grondin JM, Tamura K, Déjean G, Abbott DW, Brumer H: **Polysaccharide Utilization Loci: Fuelling microbial communities**. *J Bacteriol* 2017.
64. Raman B, Pan C, Hurst GB, Rodriguez M, McKeown CK, Lankford PK, Samatova NF, Mielenz JR: **Impact of Pretreated Switchgrass and Biomass Carbohydrates on *Clostridium thermocellum* ATCC 27405 Cellulosome Composition: A Quantitative Proteomic Analysis**. *Plos One* 2009, **4**.
- 875 65. Rahman O, Cummings SP, Harrington DJ, Sutcliffe IC: **Methods for the bioinformatic identification of bacterial lipoproteins encoded in the genomes of Gram-positive bacteria**. *World Journal of Microbiology & Biotechnology* 2008, **24**:2377-2382.
- 880 66. Thompson JD, Higgins DG, Gibson TJ: **CLUSTAL-W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. *Nucleic Acids Research* 1994, **22**:4673-4680.
67. Eschenfeldt WH, Lucy S, Millard CS, Joachimiak A, Mark ID: **A family of LIC vectors for high-throughput cloning and purification of proteins**. *Methods Mol Biol* 2009, **498**:105-115.
- 885 68. Sundqvist G, Stenvall M, Berglund H, Ottosson J, Brumer H: **A general, robust method for the quality control of intact proteins using LC-ESI-MS**. *Journal of Chromatography B-Analytical Technologies in the Biomedical and Life Sciences* 2007, **852**:188-194.
- 890 69. Ibatullin FM, Baumann MJ, Greffe L, Brumer H: **Kinetic analyses of retaining endo-(xylo)glucanases from plant and microbial sources using new chromogenic xylogluco-oligosaccharide aryl glycosides**. *Biochemistry* 2008, **47**:7762-7769.
- 895 70. McFeeters RF: **A manual method for reducing sugar determinations with 2,2'-bicinchoninate reagent**. *Analytical Biochemistry* 1980, **103**:302-306.
71. Gibson DG, Young L, Chuang RY, Venter JC, Hutchison CA, Smith HO: **Enzymatic assembly of DNA molecules up to several hundred kilobases**. *Nature Methods* 2009, **6**:343-U341.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

900 72. Waterman DG, Winter G, Gildea RJ, Parkhurst JM, Brewster AS, Sauter NK, Evans G: **Diffraction-geometry refinement in the DIALS framework.** *Acta Crystallographica Section D, Structural Biology* 2016, **72**:558-575.

73. Evans PR, Murshudov GN: **How good are my data and what is the resolution?** *Acta Crystallographica Section D* 2013, **69**:1204-1214.

905 74. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ: **Phaser crystallographic software.** *Journal of Applied Crystallography* 2007, **40**:658-674.

75. Pearson WR: **Flexible Sequence Similarity Searching with the FASTA3 Program Package.** In *Bioinformatics Methods and Protocols. Volume 132.* Edited by Stephen Misener SAK. Totowa, NJ: Humana Press; 1999: 185-219

910 76. Cowtan K: **The Buccaneer software for automated model building. 1. Tracing protein chains.** *Acta Crystallographica Section D-Biological Crystallography* 2006, **62**:1002-1011.

77. Emsley P, Lohkamp B, Scott WG, Cowtan K: **Features and development of Coot.** *Acta Crystallographica Section D-Biological Crystallography* 2010, **66**:486-501.

915 78. Murshudov GN, Skubak P, Lebedev AA, Pannu NS, Steiner RA, Nicholls RA, Winn MD, Long F, Vagin AA: **REFMAC5 for the refinement of macromolecular crystal structures.** *Acta Crystallographica Section D-Biological Crystallography* 2011, **67**:355-367.

79. Chen VB, Arendall WB, Headd JJ, Keedy DA, Immormino RM, Kapral GJ, Murray LW, Richardson JS, Richardson DC: **MolProbity: all-atom structure validation for macromolecular crystallography.** *Acta Crystallographica Section D-Biological Crystallography* 2010, **66**:12-21.

920 80. Agirre J, Iglesias-Fernandez J, Rovira C, Davies GJ, Wilson KS, Cowtan KD: **Privateer: software for the conformational validation of carbohydrate structures.** *Nat Struct Mol Biol* 2015, **22**:833-834.

925 81. McNicholas S, Potterton E, Wilson KS, Noble MEM: **Presenting your structures: the CCP4mg molecular-graphics software.** *Acta Crystallographica Section D-Biological Crystallography* 2011, **67**:386-394.

930 **Supporting information**

Additional file 1: Supporting tables

Table S1. List of genes up-regulated during exponential growth on xyloglucan compared to glucose.

Table S2. List of genes with low-level constitutive expression on xyloglucan.

935 **Table S3.** List of genes up-regulated during stationary phase on xyloglucan compared to glucose.

1
2
3
4 **Table S4.** List of genes up-regulated during stationary phase compared to exponential phase on
5
6 xyloglucan.
7

8
9
10 **Table S5.** Identity matrix between *CjGH5_4* enzymes and other specific previously
11
12 940 characterized endo-xyloglucanases.
13

14 **Table S6.** Maximum growth and growth rate of *C. japonicus* strains using reduced or limiting
15
16 concentrations of XyG.
17

18
19
20 **Table S7.** Primer sequences used for recombinant protein production.
21

22 **Table S8.** Primers used for generation of in-frame deletion mutants.
23

24
25 945 **Table S9.** X-ray data collection and refinement statistics for *CjGH5D*.
26

27 28 **Additional file 1: Supporting figures** 29

30
31
32 **Figure S1. Volcano plots summarizing the RNAseq data for a comparative analysis of *C.***
33
34 ***japonicus* exponential phase cells grown on either glucose or xyloglucan.** The volcano plot
35
36 represents a comparison between exponentially growing cells (glucose vs xyloglucan). Each gray
37
38 circle denotes a single gene, and the blue-filled circles indicate up-regulated CAZyme genes. The
39 950 complete list of up-regulated CAZyme genes can be found in Additional file 1: Table S1. Fold
40
41 change in gene expression (\log_2 scale) is plotted on the x-axis and p-value ($-\log_{10}$ scale) is plotted
42
43 on the y-axis. For orientation on the x-axis (fold change), positive values indicate genes that are
44
45 up-regulated when grown using xyloglucan as the sole carbon source. The red dashed lines
46
47 indicate significance cut-off values (2-fold for gene expression and p-value of 0.01).
48
49
50
51 955
52

53
54
55 **Figure S2. Volcano plots summarizing transcriptomic analysis of *C. japonicus* cells grown**
56
57 **on either glucose or xyloglucan.** The volcano plots represent comparisons between **A)**
58
59 stationary phase cells (glucose vs xyloglucan), or **B)** xyloglucan grown cells in either exponential
60
61

1
2
3
4 or stationary phase. Each gray circle denotes a single gene, and the blue-filled circles indicate up-
5
6 regulated CAZyme genes. The complete list of up-regulated CAZyme genes for panel A can be
7 960 found in Additional file 1: Table S3, and for panel B in Additional file 1: Table S4. The fold
8
9 change (\log_2 scale) is plotted on the x-axis and the p-value ($-\log_{10}$ scale) is plotted on the y-axis.
10
11 For orientation on the x-axis, positive values indicate genes that are up-regulated when grown
12
13 using xyloglucan as the sole carbon source for panel A, and genes up-regulated during stationary
14
15 phase for panel B. The red dashed lines indicate the significance cut-off values (2-fold for gene
16
17 expression and p-value of 0.01).
18
19 965
20
21
22
23
24

25 **Figure S3. Modular architecture of the native *CjGH5_4* enzymes with the different**

26 **expression constructs used in the current study. A)** The locus CJA_3010 (GenBank
27
28 ACE84905.1) encodes a signal peptide, a serine rich linker, and a GH5 catalytic domain
29
30 (*CjGH5D*). **B)** CJA_3337 (GenBank ACE83841.1) encodes a signal peptide, two carbohydrate
31
32 970 binding modules (CBM2 and CBM10), and a GH5 catalytic domain (*CjGH5E*). **C)** CJA_2959
33
34 (GenBank ACE86198.1) encodes a signal peptide, an FN3 domain and a GH5 catalytic domain
35
36 (*CjGH5F*). All expression constructs were designed to produce 6x His-Tag at the N-terminus of
37
38 the recombinant protein.
39
40
41
42
43
44

45 975 **Figure S4. Amino acid sequence alignment showing regions of structural similarity between**
46
47 **the catalytic domains of *Cellvibrio japonicus* GH5_4 enzymes (*CjGH5D*, *CjGH5E*, and**
48
49 ***CjGH5F*) and other members exhibiting endo-xyloglucanase activity belonging to the same**
50
51 **family.**
52
53

54
55 *Cellvibrio japonicus* GH5s (*CjGH5D*, Accession: ACE84905.1; *CjGH5E*, Accession:
56
57 980 ACE83841.1; and *CjGH5F*, Accession: ACE86198.1); *Bacteroides ovatus* GH5 (*BoGH5A*,
58
59 Accession: WP_004298445.1); *Prevotella bryantii* B14 GH5 (*PbGH5A*, Accession:
60
61

1
2
3
4 EFI71705.1); *Paenibacillus Pabuli* GH5 (*PpGH5*, PDB: 2JEP_A); ruminal metagenomic GH5s
5
6 (XEG5A, Accession: ACZ54907.1; and XEG5B, Accession: ADB44000.1). Secondary
7
8 structural elements are shown for *CjGH5D*, with η referring to 310-helices, and α to α -helices
9
10 (displayed as small and medium squiggles respectively), and β -strands shown as arrows, with TT
11 985 (and TTT representing strict β -turns and strict α -turns respectively). Alignment was created using
12
13 and TTT representing strict β -turns and strict α -turns respectively. Alignment was created using
14
15 Clustal Omega and Esript 3.0. Catalytic residues are marked with red asterisks, *CjGH5D* active
16
17 site residues interacting with the ligands (GXLG and XXXG-NHCOCH₂Br) via hydrogen bond
18
19 formation and stacking interactions are marked with blue and green asterisks, respectively.
20
21
22
23
24

25 990 **Figure S5. pH and temperature profiles of *CjGH5_4* enzymes with tamarind seed XyG as a**
26
27 **substrate. A) *CjGH5D*. B) *CjGH5E*. C) *CjGH5F*.** Left panels are pH rate profiles while right
28
29 panels are temperature profiles. Black squares, citrate buffer; red circles, phosphate buffer; and
30
31 blue triangle, glycine buffer. Error bars represent standard error of the mean for 3 replicates.
32
33
34
35 Lines were drawn to guide the eye with no physical significance.
36
37

38 995 **Figure S6. Michaelis-Menten kinetics of *CjGH5_4* enzymes on tamarind seed XyG. A)**
39
40 ***CjGH5D*. B) *CjGH5E*. C) *CjGH5F*.** Error bars represent standard error based on 3 replicates.
41
42
43

44 **Figure S7. HPAEC-PAD analysis of the hydrolysis time course and limit-digest of *CjGH5_4***
45
46 **enzyme-xyloglucan degradation products. A) *CjGH5D*. B) *CjGH5E*. C) *CjGH5F*.**
47
48
49

50 **Figure S8. MALDI-TOF analysis of the limit digest products of *CjGH5_4* enzymes upon**
51
52 1000 **incubation with tamarind seed XyG. A) *CjGH5D*. B) *CjGH5E*. C) *CjGH5F*.** The observed
53
54 molecular masses of the major 3 peaks were 1085.21, 1247.29 and 1409.37; which correspond to
55
56 [M+Na]⁺ of XXXG (calculated: 1085.9), XLXG/XXLG (calculated: 1248.05), and XLLG
57
58 (calculated: 1410.19), respectively.
59
60
61
62
63
64
65

1
2
3
4 **Figure S9. Michaelis-Menten kinetics of CjGH5_4 enzymes on a panel of chromogenic**
5
6
7 1005 **(xylo)gluco-oligosaccharide glycosides. A-D) CjGH5D. E-H) CjGH5E. I-L) CjGH5F.** Error
8
9 bars represent standard errors of the mean for 2 replicates. Only one replicate was done on
10
11 XLLG- β -CNP due to limited availability of the substrate.
12
13

14
15 **Figure S10. Intact protein mass spectrometry of CjGH5D with the XXXG-NHCOCH₂Br**
16
17 **inhibitor. A) CjGH5D negative control with no inhibitor. B) CjGH5D incubated with 2.5 mM of**
18
19 the inhibitor for 3 hours at 37 °C. The peak at 44226.5 Da corresponds to CjGH5D (calculated
20 1010 44222.2 Da) and 45330.5 Da to CjGH5D-inhibitor covalent adduct (calculated: 45325.2 Da).
21
22 Peak at 44403.1 Da (and correspondingly at 45509.5 Da) is attributed to the post translationally
23
24 modified protein due to the N-gluconylation of the His-tag of the recombinant protein.
25
26
27
28
29

30
31 **Figure S11. Growth analysis control experiments for in-frame deletions mutants of the**
32
33 1015 **GH5_4, and GH74 genes on xyloglucan.** Cultures were grown for 24 hours at 30 °C with high
34
35 aeration (200 RPM) in MOPS defined media supplemented with 0.5% (w:v) glucose as the sole
36
37 carbon source. Graphs represent the average of three biological replicates and error bars
38
39 represent the standard deviation, **A) XyGUL mutants, B) single, double, triple, and quadruple**
40
41 **deletion mutants of the GH5_4 and GH74 genes. The figure can be read as follows: *cel5D***
42
43 **(CJA_3010) encodes CjGH5D, *cel5E* (CJA_3337) encodes CjGH5E, *cel5F* (CJA_2959) encodes**
44
45 1020 **CjGH5F, and *gly74A* (CJA_2477) encodes CjGH74A).**
46
47
48
49
50

51 **Figure S12. Growth analysis of *C. japonicus* strains using reduced or limiting**
52
53 **concentrations of XyG. Wild type, a Δ *xyI31A* strain, and the quadruple mutant strain were**
54
55 **grown using A) 0.2% glucose, B) 0.5% xyloglucan, C) 0.25% xyloglucan, or D) 0.125%**
56
57
58
59
60
61

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1025 xyloglucan as the sole carbon source. Standard deviation from biological triplicate experiments
is shown. Maximum OD and growth rate calculations can be found in Table S6.

Figure S13. Growth analysis of Δ CJA_3010 and Δ gsp mutant strains when using glucose or xyloglucan. Cultures were grown in 18 mm test tubes at 30 °C with shaking at 200 RPM using MOPS minimal media supplemented with **A**) 0.25% (w:v) glucose or **B**) 0.5% (w:v) xyloglucan as the sole carbon source. Open circles represent wild type, Δ gsp is represented by closed squares, Δ CJA_3010 is represented by open triangles, and Δ CJA_3010 Δ gsp is represented by inverted closed triangles. Graphs depict the average of biological triplicate experiments, and the error bars represent the standard deviation. The *cel5D* gene (CJA_3010) encodes CjGH5D.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1035 **Tables**

Table 1. Activity of *CjGH5_4* enzymes against different polysaccharide substrates^a.

Enzyme	Substrate	<i>K_m</i>	<i>k_{cat}</i>	Specific activity
Catalytic domains		mg. mL⁻¹	sec⁻¹	μmol/ min. mg
<i>CjGH5D</i>	XyG	<0.025	30.3 ± 0.4	43.3 ± 1.9
<i>CjGH5E</i>	XyG	0.020 ± 0.002	10.3 ± 0.1	15.1 ± 0.1
	hydroxyethylcellulose (HEC)	ND ^b	ND ^b	0.070 ± 0.004
	carboxymethylcellulose (CMC)	ND ^b	ND ^b	0.010 ± 0.002
<i>CjGH5F</i>	XyG	0.040 ± 0.003	52.4 ± 0.8	74.8 ± 4.1
	hydroxyethylcellulose (HEC)	ND ^b	ND ^b	0.090 ± 0.003

^aAssays conducted at pH 7.5 (*CjGH5D* and *CjGH5E*) or pH 7 (*CjGH5F*). Recombinant enzymes were incubated at 50 °C (*CjGH5D* and *CjGH5F*) or 55 °C (*CjGH5E*) with the different tested substrates.

^bNot determined due to poor specific activity.

Table 2. Kinetic parameters of *CjGH5_4* enzymes for (xylo)gluco-oligosaccharide glycosides

Enzyme	Substrate	K_m	k_{cat}	k_{cat}/K_m
catalytic domains		mM	min⁻¹	min⁻¹. mM⁻¹
<i>CjGH5D</i>	GGG-CNP	ND ^a	ND ^a	2.21 ± 0.05
	GGGG-CNP	ND ^a	ND ^a	5.36 ± 0.07
	XXXG-CNP	0.81 ± 0.10	281 ± 12	347 ± 45
	XLLG-CNP	0.18 ± 0.02	162 ± 4	900 ± 103
<i>CjGH5E</i>	GGG-CNP	11.8 ± 0.6	191 ± 7	16.2 ± 1.0
	GGGG-CNP	5.02 ± 0.35	180 ± 7	35.9 ± 2.8
	XXXG-CNP	0.010 ± 0.001	254 ± 4	(25.4 ± 2.6) x 10 ³
	XLLG-CNP	0.010 ± 0.001	332 ± 9	(33.2 ± 3.4) x 10 ³
<i>CjGH5F</i>	GGG-CNP	ND ^a	ND ^a	6.45 ± 0.30
	GGGG-CNP	ND ^a	ND ^a	16.5 ± 1.0
	XXXG-CNP	0.07 ± 0.01	169 ± 4	(2.41 ± 0.35) x 10 ³
	XLLG-CNP	0.030 ± 0.002	393 ± 8	(13.1 ± 0.9) x 10 ³

^anot determined due to limited availability of substrate.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1045 **Figure legends**

Fig. 1. Xyloglucan (XyG) and the xyloglucan utilization system in *C. japonicus*. **A)** Structure of dicot XXXG- type fucogalacto-XyG. XyG substructure nomenclature is according to [20]. **B)** *C. japonicus* genes involved in XyG utilization. Genes encoding backbone-cleaving endo-xyloglucanases (GH5 and GH74) are indicated in navy blue, genes encoding side-chain-cleaving exo-glycosidases (GH35 β -galactosidases; GH31 α -xylosidases and GH95 α -L-fucosidase) are in cyan, and the TonB dependent transporter (TBDT) is shown in green. **C)** Spatial model of XyG utilization in *C. japonicus*.

Fig. 2. Inhibition kinetics of *CjGH5D* with XXXG-NHCOCH₂Br. **A)** Initial-rate enzyme activity over time (single determinations). **B)** Pseudo-first-order rate constants (k_{app}) obtained from the fitted curves shown in panel A. Bars represent errors in k_{app} values from curve-fitting. The 95% confidence interval is indicated (pink band) for the fitted curve (solid line).

Fig. 3. Three-dimensional structure of *CjGH5D* in complex with XXXG-NHCOCH₂Br and XyGOs. **A)** Cartoon representation of the secondary structure of *CjGH5D* colour ramped from the N-terminus (blue) to the C-terminus (Red). The two ligands XXXG-NHCOCH₂Br and GXLG are overlaid in the active site cleft and shown in green and magenta sticks, respectively. **B)** A close-up view of the active site cleft with the overlaid ligands XXXG-NHCOCH₂Br in green and XXLG in magenta showing different amino acids interacting with the carbohydrate ligands. **C)** $2F_o - F_c$ (σ_A /maximum likelihood weighted) electron density contoured in blue around GXLG in the *CjGH5D*-XXLG complex (left panel) and the chemical structure of the corresponding ligand (Right panel). Insufficient electron density was observed for the -4' xylosyl residue to allow modelling, therefore it is shown in grey. **D)** $2F_o - F_c$ electron density at 1σ

1
2
3
4 (approx. $0.2 \text{ e}^-/\text{\AA}^3$) contoured in blue around XXXG-NHCOCH₂ moiety in the CjGH5D-XXXG-
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

(approx. $0.2 \text{ e}^-/\text{\AA}^3$) contoured in blue around XXXG-NHCOCH₂ moiety in the CjGH5D-XXXG-NHCOCH₂Br complex (Left panel) and chemical structure of the corresponding ligand (Right panel). The bromide leaving group is shown in grey.

1070 **Fig. 4. Divergent (wall-eyed) stereo surface representation of CjGH5D-GXLG showing**
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

regions of sequence conservation. Surfaces of conserved and non-conserved residues, shown in purple at reduced opacity and sea-green, respectively, were calculated from an amino acid sequence alignment of GH5 domains of CjGH5D, CjGH5E, CjGH5F and five additional GH5 members showing E.C. 3.2.1.151 activity (Additional file 1: Figure S4). Figure was generated using CCP4MG [81].

1080 **Fig. 5. Growth analysis of in-frame deletions of GH5_4, and GH74 mutant strains on**
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

xyloglucan. **A)** Control experiment with XyGUL in-frame deletion mutant strains. **B)** Single, **C)** double, **D)** triple and quadruple deletion mutants were made with the GH5_4 and GH74 genes; CJA_3010 encodes CjGH5D, CJA_3337 encodes CjGH5E, CJA_2959 encodes CjGH5F, and CJA_2477 encodes CjGH74. Graphs represent the average of three biological replicates and error bars represent the standard deviation. All strains grew like wild-type when grown with MOPS-glucose defined medium (Additional file 1: Figure S12).









