Substrate specificity, regiospecificity, and processivity in glycoside hydrolase family 74

Gregory Arnal1, Peter J. Stogios5, Jathavan Asohan6, Mohamed A. Attia4, Tatiana Skarina5, Alexander Holm Viborg7, Bernard Henriassat89, Alexei Savchenko910, and Harry Brumer11

From the 1Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada, the 5Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario M5S 3E5, Canada, the 6Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada, the 9Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada, and the 11Department of Botany, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada

Edited by Gerald W. Hart

Glycoside hydrolase family 74 (GH74) is a historically important family of endo-β-glucanases. On the basis of early reports of detectable activity on cellulose and soluble cellulose derivatives, GH74 was originally considered to be a “cellulase” family, although more recent studies have generally indicated a high specificity toward the ubiquitous plant cell wall matrix glycan xyloglucan. Previous studies have indicated that GH74 xyloglucanases differ in backbone cleavage regiospecificities and can adopt three distinct hydrolytic modes of action: exo, endo-disassociative, and endo-processive. To improve functional predictions within GH74, here we coupled in-depth biochemical characterization of 17 recombinant proteins with structural biology—based investigations in the context of a comprehensive molecular phylogeny, including all previously characterized family members. Elucidation of four new GH74 tertiary structures, as well as one distantly related dual seven-bladed β-propeller protein from a marine bacterium, highlighted key structural–function relationships along protein evolutionary trajectories. We could define five phylogenetic groups, which delineated the mode of action and the regiospecificity of GH74 members. At the extremes, a major group of enzymes diverged to hydrolyze the backbone of xyloglucan nonspecifically with a dissociative mode of action and relaxed backbone regiospecificity. In contrast, a sister group of GH74 enzymes has evolved a large hydrophobic platform comprising 10 subsites, which facilitates processivity. Overall, the findings of our study refine our understanding of catalysis in GH74, providing a framework for future experimentation as well as for bioinformatics predictions of sequences emerging from (meta)genomic studies.

Terrestrial plants harbor ~80% of the biomass on Earth, some 450 gigatons of carbon, in the form of lignocellulose (cell walls comprised of cellulose, matrix glycans, lignin, and other polymers) (1). Although terrestrial biomass represents an attractive renewable resource for the production of fuels, chemicals, and materials for human consumption, the controlled degradation of lignocellulose, whether (thermo)chemical or enzymatic, is hindered by its heterogeneous composition and complex organization (2). Hence, significant efforts have been made to identify enzymes able to efficiently modify and deconstruct this complex material.

Xyloglucans (XyGs)3 comprise a prominent family of cell wall matrix glycans (hemicelluloses). XyGs are ubiquitous in land plants, in which they constitute up to 20% of the dry weight of cell walls (3, 4). Notably, XyGs are secreted by roots of diverse plant species and are therefore likely to actively influence rhizobiota (5). XyGs are also found as storage polysaccharides comprising ~50% of the mass of some seeds (e.g. tamarind and nasturtium) and therefore represent important agricultural by-products with applications in the food, biomaterial, and medical sectors (6, 7). XyGs have a β-1,4-linked glucosyl backbone (“G” unit), some of which are decorated with an α-(1,6)-d-xylosyl residue (together comprising an “X” unit; nomenclature according to Ref. 8). Generally, three of four contiguous glucosyl units are xylosylated, forming repeating (XXXG)n-type

This work was supported by the Natural Sciences and Engineering Research Council of Canada (via a Discovery Grant to H. B. and a Strategic Partnership Grant for Networks to H. B. and A. S. for the Industrial Biocatalysis Network), the Canada Foundation for Innovation, and the British Columbia Knowledge Development Fund. The authors declare that they have no conflicts of interest with the contents of this article.

This article was selected as one of our Editors’ Picks.

This article contains File S1, Tables S1–S3, and Figs. S1–S8.

The atomic coordinates and structure factors (codes 6P2K, 6P2L, 6P2M, 6P2N, and 6P2O) have been deposited in the Protein Data Bank (http://wwpdb.org/).

1 To whom correspondence may be addressed: 3330 Health Sciences Centre, University of Calgary, Calgary, Alberta T2N 1N4, Canada. Tel.: 403-210-7980; E-mail: alexei.savchenko@ucalgary.ca.
2 To whom correspondence may be addressed: Michael Smith Laboratories, University of British Columbia, 2185 East Mall, Vancouver, British Columbia V6T 1Z4, Canada. Tel.: 604-827-3738; E-mail: brumer@msl.ubc.ca.

© 2019 Arnal et al. Published under exclusive license by The American Society for Biochemistry and Molecular Biology, Inc.
EDITORS’ PICK: GH74 structure–function analysis

XyGs. Depending on the plant tissue, the xylosyl branches may be further substituted with a variety of other saccharides (4). Therefore, total saccharification of XyGs requires the concerted action of several side chain–debranching and backbone-cleaving enzymes (9–12).

Endo-xylglycanases, which cleave the XyG backbone, are found in glycoside hydrolase (GH) families GH5, GH9, GH12, GH16, GH44, and GH74 (11). Of these, GH74 currently comprises ~500 members, ranking it among the smaller GH families. GH74 is further distinguished from these other poly-specific families by a nearly singular specificity for XyG (13). The first GH74 enzyme to be biochemically characterized, from Aspergillus aculeatus, was described in 1989 as an “avicelase” (Avicel® is a brand of microcrystalline cellulose) (14). As a result, GH74 is sometimes myopically referred to as a cellulase family, and its members are often annotated as such in (meta)-genomics studies (15). However, numerous studies since the turn of this century have shown that many GH74 enzymes are in fact highly specific xylglycanases (16–35). The biological importance of this family is underscored by (meta)genomic studies, which have revealed the ubiquity of GH74 members in diverse ecological niches, including soil, termite and human guts, and hot springs (36–40).

GH74 members present subtle structural variation and modes of action, which have been reviewed recently (41). Briefly, backbone hydrolysis can occur either in the middle of the polysaccharide chain (endo-xylglycanases, EC 3.2.1.151) or at the chain end (exo-xylglycanases, EC 3.2.1.150). Endo-xylglycanases can be further delineated into endo-dissociative enzymes, which hydrolyze the backbone and immediately release both new chain ends, and endo-processive enzymes, which perform multiple hydrolytic events, releasing short oligosaccharides before disengaging. The ability of some GH74 enzymes (26, 28, 29, 35) to act processively on soluble XyG is notable, considering that this mode of action is more commonly associated with GHs acting on crystalline cellulose or chitin (42–47).

To unify disparate studies on GH74 members and resolve gaps in our current understanding of the distribution of the distinct modes of action in the family, molecular phylogeny was coupled with detailed enzymology to elucidate the substrate specificity, backbone cleavage regiospecificity, and processivity of 17 recombinant GH74 proteins in the present study. The determination of crystal structures of four GH74s and one distantly related dual seven-bladed β-propeller protein, together with analysis of existing GH74 structures, highlighted key structure–activity relationships across this family. Overall, this study refines our understanding of catalysis in GH74 and reveals the evolutionary trajectory of this enzyme family from dissociative toward processive modes of action.

Results
Production and biochemical characterization

A molecular phylogeny using isolated GH74 catalytic modules from the CAZy database (13) was generated to guide protein production, enzymology, and structural biology. Previously characterized GH74 enzymes that were absent from the CAZy database were also included in this analysis (GenBank™ accession numbers CCG35167 (23) and XP_747057 (20)). In addition, two proteins (GenBank accession numbers AFV00434 and AFV00474) encoded in the Simiduia agarivorans genome, which are distantly related to GH74 enzymes based on hydrophobic cluster analysis (HCA) (48), were included as an outgroup. Thirty candidates were selected across the phylogenetic tree, of which 17 proteins were successfully recombinantly produced and purified (Fig. 1).

Proteins were first screened for activity on a range of substrates, including polysaccharides and pNP substrates. The recombinant AFV00434 and AFV00474 proteins from S. agarivorans were not active on the range of substrates tested, including XyG (data not shown). All other recombinant GH74 modules showed a strict preference for tamarind XyG. No endo-mannanase activity toward konjac glucomannan, no endo-xylanase activity toward wheat flour arabinoxylan and beechnow wood xylan, and no endo-glucanase activity using CM-cellulose were observed. Endo-glucanase activity on HE-cellulose and on barley β-glucan was generally estimated to be less than 1% compared with xylglycanase activity (data not shown). As such, we did not perform further biochemical characterization on these substrates. Overall, these results, together with previous studies (16–35), suggest that GH74 enzymes are, in general, very specific for XyG.

To further investigate the biochemical properties of GH74 enzymes, optimum pH (Fig. S1) and temperature (Fig. S2) ranges were evaluated using XyG as a substrate (Table 1). Generally, recombinant enzymes were active at pH values ranging from 5 to 8, with optimum activities observed around pH 6 (except for Niastella koreensis GH74, which displayed maximum activity at pH 4.5). The highest activities were observed at temperatures ranging from 45 to 65 °C for most recombinant enzymes, except for the thermophilic Caldicellulosiruptor lactoacticus GH74a and GH74b and Caldicellulosiruptor bescii GH74, whose highest activities were recorded at 80 °C. Michaelis–Menten analysis confirmed the high specificity of the GH74 catalytic domains for XyG, with $K_m$ and $k_{cat}$ values generally ranging from 0.02 to 0.31 mg/ml and from 18.1 to 170.2 s$^{-1}$, respectively (Table 1 and Fig. S3). These values are in the same range as previously characterized GH74 enzymes (16, 22, 29, 35). Exceptionally, recombinant Streptomyces venezuelae GH74b was very unstable and precipitated rapidly in solution, which did not allow accurate kinetic characterization.

Among previously characterized GH74 enzymes, Thermo-toga maritima Xeg74 showed higher activity for mixed linkage barley β-glucan than tamarind XyG (19), which constitutes an anomaly in this family. Unfortunately, we were unable to recombinantly produce T. maritima Xeg74 to verify this finding independently.

Regiospecificity and processivity of GH74 members

The mode of action of GH74 xylglycanases has been described for a limited number of enzymes. Oligoxylglucan reducing end-specific celllobiohydrolases (OXG-RCBHs) (EC 3.2.1.150) are exo-type xylglycanases that release Glc$_2$-based products such as XG or LG from tamarind XyG (17, 49). Endo-xylglycanases can act in a dissociative fashion to generate a
Cloned but would not produce

Activity characterized in a previous study

Activity characterized in the present study

3D structure solved in a previous study

3D structures solved in the present study

+/− Trp present or absent in subsite

p endo-processing enzyme

d endo-dissociative enzyme

e exo-acting enzymes

? Not determined

□ XyG backbone cleaved exclusively at G motif

□ XyG backbone cleaved at G and X motifs

□ XyG backbone cleaved exclusively at X motif
wide distribution of product chain lengths (23, 26, 33), or they can act in a processive fashion to rapidly release small xyloglucan oligosaccharides (XyGOs) at early stages of XyG hydrolysis (20, 26, 28, 29, 35). To investigate the mode of action of our recombinant GH74 enzymes, we analyzed the time-course hydrolysis of tamarind XyG at early stages of the reaction by HPAEC-PAD (Fig. 2 and Fig. S4).

Also, most previously characterized GH74 endo-xyloglucanases release XXXG-type XyGOs via the exclusive hydrolysis of XyG at unbranched glucosey residues (16, 22, 28, 33, 35). To investigate the backbone regiospecificity of GH74 enzymes, the limit digest of tamarind XyG, of XXXGXXXG, could not be recombinantly produced. However, we successfully produced N. koreensis GH74, Ruminococcus albus GH74a, and C. lactoaceticus GH74a. These three enzymes act as endo-dissociative xyloglucanases (Fig. 2 and Fig. S4 (A, B, and C)) like the previously characterized Xanthomonas citri pv. mangiferaeindicae GH74 (23), which also belongs to the phylogenetic Group 1. Notably, this group also contains T. maritima Cel74, which was previously shown to be 4 times more active on barley β-glucan than on tamarind XyG (19); this difference of specificity is not easily rationalized in light of the phylogenetic relationship with C. lactoaceticus GH74a (Fig. 1).

Recent reports showed that two tryptophan residues found in the +3 and +5 subsites in the active site cleft are necessary for the processivity of GH74 enzymes (32, 35) and are conserved in all previously reported endo-processive xyloglucanases in this family (26, 28) (see below; Group 5). Very few sequences from the Group 1 enzymes possess one or both +3 and +5 subsite Trp residues (15 and 4%, respectively), consistent with the lack of processivity observed in our examples.

Table 1
Biochemical properties of recombinant GH74 modules on tamarind XyG

| Protein        | Bacterial strain       | Optimum temperature/pH | $K_u$ (mg/ml) | $K_a$ ($s^{-1}$) |
|---------------|------------------------|-------------------------|---------------|-----------------|
| CjGH74a       | Caldicellulosiruptor    | 81°C/pH 5               | 0.028 ± 0.001 | 71.8 ± 1.2      |
| CjGH74b       | C. lactoaceticus        | 83°C/pH 6               | 0.014 ± 0.002 | 44.9 ± 1.4      |
| CdGH74        | Caldicellulosiruptor    | 85°C/pH 6.5             | 0.04 ± 0.02   | 51.2 ± 3.2      |
| NkGH74        | Nistella koreensis      | 55°C/pH 4.5             | 0.09 ± 0.008  | 18.1 ± 0.7      |
| PgGH74        | Paenibacillus graminis  | 55°C/pH 6               | 0.039 ± 0.001 | 170.2 ± 4.6     |
| PhGH74        | Paenibacillus borealis  | 45°C/pH 6               | 0.022 ± 0.001 | 91.8 ± 1.4      |
| PyGH74        | Paenibacillus polyxylus | 54°C/pH 6               | 0.095 ± 0.010 | 26.9 ± 1.4      |
| PyGH74        | Paenibacillus jamaicai  | 60°C/pH 6               | 0.057 ± 0.013 | 24.4 ± 1.4      |
| Ps1174        | Paenibacillus mucilaginosus | 60°C/pH 6              | 0.117 ± 0.011 | 96.5 ± 2.6      |
| RaGH74a       | Ruminococcus albus      | 50°C/pH 6               | 0.028 ± 0.003 | 47.5 ± 2.8      |
| RaGH74b       | R. albus                | 45°C/pH 6               | 0.314 ± 0.045 | 37.9 ± 1.9      |
| AFV00434      | Simidiasia agarivorans  | NA*                     | NA*           | NA*             |
| AFV00474      | S. agarivorans          | NA                      | NA            | NA              |
| SrGH74        | Streptomyces ramapycinicus | 65°C/pH 6             | 0.014 ± 0.001 | 45.2 ± 2.6      |
| SvGH74a       | Streptomyces venezueae  | 25°C/pH 6               | 1.05 ± 0.34   | 1.0 ± 0.1       |
| SvGH74b       | S. venezueae            | 55°C/pH 7               | 0.023 ± 0.003 | 52.3 ± 2.2      |
| SrGH74        | Streptomyces atroolvoceus | 54°C/pH 6             | 0.09 ± 0.01   | 31.1 ± 1.0      |

* NA, not active on xyloglucan.
Remarkably, C. lactoaceticus GH74a has both positive-subsite Trp residues (Fig. S5) but nonetheless acted as a dissociative enzyme. Thus, the presence of this pair of Trp residues is necessary but not sufficient for processivity in GH74.

N. koreensis GH74, R. albus GH74a, and C. lactoaceticus GH74a all had relaxed regiospecificity and were thus able to cleave the backbone of XyG at both xylosylated (X) and unbranched glucosyl (G) units (Fig. 3). In contrast, X. citri pv. mangiferaeindicae GH74 cleaved specifically after X motifs (23). These four enzymes have a Gly residue in the –1 subsite, as do 60% of enzymes from Group 1. This residue has been shown to be responsible for the ability of previously characterized GH74 endo-xyloglucanases to cleave at X units (35, 54).

However, some Group 1 members have an Ala (20%), a Trp (10%), or a Gln (7%) residue in the corresponding position, suggesting that some Group 1 members may have a strict preference for XyG hydrolysis at G units.

To investigate the structural determinants for the mode of action of enzymes belonging to Group 1, we solved the tertiary structure of C. lactoaceticus GH74a in complex with the XyG fragment LLG (PDB code 6P2M), and of N. koreensis GH74 in complex with two XyG fragments, XXLG and XXXG (PDB code 6P2L; Fig. 4). These represent the first three-dimensional structures described in Group 1. Vis-à-vis SaAFV00434 in the distantly related sister clade (Fig. 1), these structures reveal a broad, active-site cleft poised to accept the highly branched XyG polysaccharide chain. The structure of C. lactoaceticus GH74a clearly demonstrates the positioning of consecutive Trp residues, Trp328 and Trp329, comprising the +3 and +5 subsites (Fig. 5 and Fig. S5). Remarkably, the N. koreensis GH74 active-site cleft also harbors two Trp residues in homologous +3 and +5 subsite positions (Trp328 and Trp337), but instead of being found consecutively in the primary structure, they are interspersed with a loop comprising Ser329–Thr336 (Fig. 5 and Fig. S5).

Active-site aromatic residues, in particular tryptophan residues, are important for substrate recognition and processivity in glycoside hydrolases (55–57). Across the active-site cleft, we found only five hydrophobic residues positioned to interact with the XyG backbone from the –4 to the +5 subsite in C. lactoaceticus GH74a (Tyr122, Trp126, Trp328, Trp329, and Trp375) and N. koreensis GH74 (Tyr117, Phe118, Trp328, Trp329, and Trp376) (Fig. 5). In comparison, the active-site cleft of the processive xyloglucanase P. odorifer GH74 (PDB code 6MLG) of Group 5 (see below) is lined with 12 aromatic residues (35), which create a large hydrophobic platform extending from the –4 to the +6 subsites (Fig. 5). C. lactoaceticus GH74a and N. koreensis GH74 completely lack a corresponding +6 subsite. Overall, these results suggest that Group 1 comprises enzymes with the first sequence features allowing for dissociative endo-xyloglucanase activity but that the limited number of hydrophobic interactions in their active cleft does not enable processivity.

Group 2—Group 2 specifically segregates the fungal endo-xyloglucanase Geotrichum sp. XEG74 (EC 3.2.1.151) and two OXG-RCBHs (EC 3.2.1.150) from Geotrichum sp. (49) and Aspergillus nidulans (17). This small clade comprises previously characterized enzymes. In particular, seminal work by...
Yaoi et al. (52) demonstrated that the strict exo-activity of the OXG-RCBH enzymes, which results in the production of Glc₄-based products (e.g. XG and LG), is dictated by the presence of an 11-amino acid loop that blocks one end of the active-site cleft. Our analysis of the current CAZy database, which contains only GenBank™-deposited sequences (13), indicated that only Geotrichum sp. and A. nidulans OXG-RCBHs possess this “exo-loop.” However, Damasio et al. (20) found 19 additional putative OXG-RCBHs from the analysis of 293 Eurotiomycete and Ascomycete genomes, reinforcing the observation that OXG-RCBHs enzymes form their own evolutionarily divergent clade within GH74.

**Group 3**—Group 3 is currently comprised of 22 bacterial enzymes belonging to the genus *Streptomyces* as well as one *Proteobacteria* enzyme. All enzymes from Group 3 carry the Trp residue in subsite +3, which is found in some members of Group 1 but is ubiquitous in Groups 4 and 5. At the same time, Group 3 members lack the +5 subsite Trp found in Groups 4 and 5 (Fig. 1). In addition, enzymes from Group 3 have also acquired hydrophobic residues in the −4 and −3 subsites that are conserved in the Group 5 processive xylanase *P. odorifer* GH74 (35) (see below) (Fig. S5). Within Group 3, *Streptomyces atroolivaceus* GH74 acted as an endo-dissociative enzyme (Fig. S4D), analogous to the previously characterized...
Streptomyces avertimilis GH74b (26). Both enzymes were able to cleave XyG backbone at both G and X motifs, yet with a clear preference for the unbranched G unit (Fig. 3) (26), reflective of the presence of a Gly residue in the −1 subsite.

Despite the lack of three-dimensional structural representatives from phylogenetic Group 3, sequence analysis indicates the presence of hydrophobic residues in subsites −4, −3, +2, and +3 in the active cleft of these enzymes (Fig. S5). As in Group 1, these, and especially the limited aromatic platform in the positive subsites, are apparently insufficient to enable processivity (Fig. S4D). The current data indicate that Group 3 members are endo-dissociative enzymes that preferentially hydrolyze the XyG backbone at unbranched glucosyl units (Fig. 3).

Group 4—Group 4 is comprised of 19 bacterial enzymes belonging to the family Streptomycetaceae. Sequence alignment indicates that Group 4 members have retained the all active-site aromatic residues characteristic of Group 3 and additionally acquired the +5 subsite Trp residue found in Group 5 members (Fig. S5). Unfortunately, instability of recombinant S. venezuelae GH74a precluded detailed enzymology.

Figure 4. Crystal structures of S. agarivorans AFV00434, N. koreensis GH74, C. lactoaceticus GH74, P. graminis GH74, S. rapamycinicus GH74, and P. odorifer GH74 (PDB code 6MGL). A, overlay of crystal structures of AFV00434 (red) and Group 5 PoGH74−(D70A)+(XXLG + XGXXLG) (PDB code 6MGL) (black) in ribbon configuration; putative catalytic residues of AFV00434 are indicated in a stick configuration, and loops impeding xyloglucan accommodation in the active site of AFV00434 are represented in cartoon representation. B, overlay of crystal structure of Group 1 NkGH74−(XXLG + XXXG) (green) and C1GH74a (pink) shown in cartoon representation. C, overlay of crystal structure of Group 5 PgGH74 (orange), PoGH74−(D70A)+(XXLG + XGXXLG) (PDB code 6MGL) (black), and SrGH74−(XXLG + XXXG) (cyan). D, overlay of crystal structures of Group 1 NkGH74−(XXLG + XXXG) (green) and Group 5 PoGH74−(D70A)+(XXLG + XGXXLG) (PDB code 6MGL) (black). The loop additions/extensions allowing the position of aromatic residues in the −2, +5, and +6 subsites in the active site of Group 5 enzymes are shown in cartoon representation.
Nonetheless, time-course hydrolysis of XyG analyzed by HPAEC-PAD analysis clearly indicated that this enzyme acted as an endo-dissociative enzyme (Fig. S4E) and hydrolyzed the polysaccharide backbone at both X and G units (Fig. 3). The presence of a conserved Gly in subsite −1 of all enzymes from Group 4 is consistent with this relaxed regiospecificity (Fig. S5). However, the presence of an extended positive subsite platform is insufficient to support processivity (Fig. S4E).

**Group 5**—Group 5 is comprised of 173 bacterial and fungal enzymes that form a monophyletic group supported by a high bootstrap value of 75. Notably, most GH74 catalytic modules of this group are appended to a carbohydrate-binding module (CBM) (16, 26, 35), whereas CBMs are generally absent in enzymes from other phylogenetic groups (Fig. 1). Nearly all enzymes from Group 5 (166 of 173) contain the subsite +3/+5 Trp pair, which constitute an extended substrate-binding platform also observed in Group 4 (Fig. 1). This platform appears to be a prerequisite for processivity, as all presently (Fig. S4, F, G, and I–O) and previously characterized processive GH74 endo-xylanoglanases belong to Group 5 (20, 24, 26, 28, 29, 35). Indeed, previous work on Group 5 members *P. odorifer* GH74 (35) and *Paenibacillus* sp. strain KM21 (29) used site-directed mutagenesis to define the critical role of both +3 and +5 aromatic residues in processivity. Further, *R. albus* GH74b is a rare instance of a natural variant in this phylogenetic group, in which the conserved +5 subsite Trp has been substituted with Ala. Accordingly, *R. albus* GH74b is an endo-dissociative xylanoglanacase (Fig. S4F). Thus, both Trp residues are not sufficient (as in Group 4); they are nonetheless necessary for processivity (as in Group 5).

These observations prompted us to reevaluate our previous analysis of *Cellvibrio japonicus* GH74, in which we described this Group 5 enzyme as endo-dissociative (16). However, the presence of the pair of +3/+5 subsite Trp residues (Trp^{353} and Trp^{354}) in this enzyme predicts an endo-processive mode of action. A more refined time-course analysis of XyG degradation showed that the WT *C. japonicus* GH74 had an endo-processive mode of action, consistent with its active-site composition and placement in Group 5, whereas the subsite variants W353A and W354A acted as endo-dissociative enzymes (Fig. S6), analogous to homologous mutants (29, 35).

As in other phylogenetic groups, the residue occupying the −1 subsite in the active cleft of GH74 xylanoglanases affects the backbone cleavage regiospecificity of Group 5 enzymes, yet it is not the only determinant. The vast majority (90%) of enzymes from Group 5 have a Gly residue in subsite −1, whereas the remainder have either a Tyr, a Leu, an Ala, or an Arg residue in this position. Among this latter group, the previously characterized *Phanerochaete chrysosporium* Xgh74B has a Leu in the −1 subsite (28), whereas *C. besseii* GH74 and *C. lactoaceticus* GH74 have a Tyr here (Fig. 1 and Fig. S5). These three enzymes showed a strict specificity for XyG backbone hydrolysis at unbranched G units (Fig. 3).

Among enzymes with a Gly residue in the −1 subsite, the data were more equivocal. Whereas the regiospecificities of *R. albus* GH74b and *Paenibacillus mucilaginosus* GH74 are relaxed, *Paenibacillus graminis* GH74 and *Paenibacillus borealis* GH74 were the only Group 5 enzymes that could efficiently hydrolyze XXXG to XX + XG. On the other hand, *Paenibacillus jambilae* GH74 and *Paenibacillus polymyxa* GH74 showed a clear, but not exclusive, preference for cleavage at G units and a propensity to hydrolyze XXXG. Last, *Streptomyces raphyunicins* GH74 and *S. venezuelae* GH74 strictly cleave XyG backbone at the unbranched glucosyl unit (Fig. 3).

To further investigate the determinants for the cleavage pattern of GH74 enzymes, we used *P. odorifer* GH74 (35) as a platform for site-directed mutagenesis. This enzyme shares over 90% sequence identity with *P. graminis* GH74 and *P. borealis* GH74 and likewise hydrolyzes XXXG to XX + XG (Fig. 3). *P. odorifer* GH74 has a mobile loop (Asn^{642}–Ala^{651}) that is conserved in *P. graminis* GH74 and *P. borealis* GH74 (Fig. S5). In the closed conformation, this loop protrudes into the active site, covering subsite −4 and hindering subsite −3 (35). Thus, we first eliminated the possibility that this loop might force XGXG into a −2 to −1 binding mode in these enzymes, thereby promoting hydrolysis between two X units (as indicated here with the vertical bar). Indeed, the *P. odorifer* GH74 deletion variant ΔAsn^{642}–Ala^{651} behaved like the WT enzyme (Fig. S7).

Hence, we investigated the role of the residue found in the −1 subsite in the active site cleft of *P. odorifer* GH74. In a previous study, we showed that a G476Y mutation in the −1 subsite switched the mode of action to exclusively cleave the XyG backbone at the G unit (35). Analogously, here we produced three single-point mutations representing the other amino acid variants found in the −1 subsite of GH74 enzymes (viz. G476A, G476W, and G476Q). Like the G476Y variant, G476A, G476W, and G476Q variants all showed strict specificity for XyG hydro-
lysis at the G motif (Fig. S7). Thus, even the relatively small methyl side chain of the Ala residue hinders the accommodation of a xylose side chain in the subsite −1 and shifts the register of XyG backbone hydrolysis to the canonical unbranched G unit (11, 58).

We solved the crystal structure of Group 5 members P. graminis GH74 (PDB code 6P2N) and S. rapamycinicus GH74 in complex with two XyG fragments (XLLG and XXXG) (PDB code 6P2O), thereby increasing the number of Group 5 tertiary structures from four to six (16, 35, 50, 51). Using P. odorifer GH74 as a reference, P. graminis GH74 and S. rapamycinicus GH74 are similar in overall conformation, with notable deviations localized to loops impinging on the −4, −3, and −2 subsites (i.e. P. graminis GH74 residues 643–653, 121–128, and 209–217 versus P. odorifer GH74 residues 607–618, 86–94, and 174–182 and S. rapamycinicus GH74 residues 643–646, 129–137, and 218–227) (Fig. 4C). In particular, these conformational changes caused P. graminis GH74 residues Trp126, Tyr122, and Tyr214 to rotate out of the active site cleft as compared with their equivalents Trp91, Tyr87, and Tyr179 in P. odorifer GH74 (Fig. S8). The absence of a bound xyloglucan ligand in our structure of P. graminis GH74 may explain these conformational movements and likely reflects inherent flexibility in this region. XLGG and XXXG bound to S. rapamycinicus GH74 superimposed nearly exactly to XLGLG and XGXXLG bound to P. odorifer GH74 (PDB code 6MGL), with the exception of small changes in the positions of the sugars in the −3 and −4 subsites (Figs. 4C and 5B).

A striking feature of P. odorifer GH74 was the presence of 12 aromatic residues that lined the active-site cleft of the enzyme, which formed a large hydrophobic platform that extended from the −4 to the +6 subsites (35). Consistent with the conserved binding position of the xyloglucan fragments noted above, these residues are conserved in S. rapamycinicus GH74 (Fig. 5) as well as in P. graminis GH74 and nearly all enzymes from Group 5 (Fig. S5 and File S1 (GH74_CatalyticModules_Aligned.mfa)). In comparison, Group 1 members N. koreensis GH74 and C. lactoaceticus GH74a only have up to five of these active-site cleft aromatic residues (Fig. 5).

As might be expected, sequence analysis revealed that the acquisition of some of these key aromatic residues by Group 5 enzymes occurred through single point mutations. For instance, a Tyr residue is found in the P. graminis GH74 (Tyr299) and S. rapamycinicus GH74 (Tyr307) +1 subsites, whereas an Asn or a Ser occupies the corresponding positions in C. lactoaceticus GH74a and N. koreensis GH74, respectively (Fig. S5). However, loop extensions have also played a major role in building the hydrophobic platform. In particular, loops Tyr206–Gly213, Gly320–Tyr325, and Gly371–Ala381 provided the scaffold for the insertion of Tyr214, Trp325, and Tyr373 in the subsites −2, +5, and +6 in P. graminis GH74 (Fig. 4D). These loops are conserved across members of Group 5 but are absent in other phylogenetic groups (Fig. S5). Notably, the loop composed of Gly371–Ala381 added the +6 subsite, which is found only in Group 5. The insertion of these aromatic residues created a network of stacking interactions with the XyG backbone that contribute to the processivity of GH74 enzymes. For example, residues Trp406 (+2 subsite) and Tyr372 (+6 subsite) contribute to processivity in P. odorifer GH74 (35), whereas Trp61 (−4 subsite) and Trp64 (−3 subsite) contribute to the processivity of Paenibacillus KM.21 XEG74 (29), beyond the essential requirement of Trp residues in subsites +3 and +5 in these enzymes. Most of these auxiliary aromatic residues are conserved in Group 5 enzymes but are not found in the other phylogenetic groups.

Discussion

Enzymes from the same GH family share a common structural fold and catalytic mechanism (13, 53). However, many CAZyme families harbor members with diverse specificities (poly-specific families), which makes functional annotation challenging due a general lack of detailed biochemical characterization (13). For a handful of larger GH families examined to date, phylogeny-based subfamily classification has enabled further refinement of activities into monospecific clades in some cases (59–62). Thus, phylogenies highlight different structural trajectories within GH families that correlate with conserved sequence residues and substrate specificities. Not least, such delineation guides functional and structural analyses toward the characterization of enzymes significantly divergent from those previously studied and thus can resolve knowledge gaps.

Through the largest systematic experimental analysis to date, this study provides a broad overview of structure–function relationships in GH74. Enzymes from this family have evolved a unique tertiary structure comprising a large cleft to accommodate the highly branched XyG chain. From this scaffold, we observe different evolutionary trajectories that delineate the mode of action and backbone cleavage regiospecificity. Notably, GH74 is sister to a group of distantly related, dual seven-bladed β-propeller proteins, of which we were able to solve the first tertiary structure, but for which we were unable to find polysaccharide hydrolase activity.

Across the GH74 phylogeny, the characterized members of the diverse Group 1 generally evidence a relaxed backbone cleavage specificity, with the ability to hydrolyze at X or G units through an endo–dissociative (i.e. nonprocessive) mode of action. Although we were only able to observe strict XyG specificity in the examples we characterized, the observation that T. maritima Cel74 is 4 times more active on β-glucan than on XyG (19) might imply that broader specificity exists among the sequence-diverse Group 1 members. At the same time, the C. lactoaceticus GH74a in a closely related sister clade was a strict xyloglucanase (Fig. 1). Regrettably, we were unable to reproduce T. maritima Cel74 to explore this further, but certainly functional characterization of additional Group 1 members, including from completely uncharacterized major clades (Fig. 1), is warranted.

Phylogenetic Groups 3 and 4 are individually dominated by single genera or phyla and therefore may simply reflect speciation and not functional evolution. Nonetheless, characterized members of these clades possess unique constellations of active-site residues (as well as CBM modularity) (Fig. 1). In particular, the stepwise gain of key active-site aromatic residues, which are necessary for processivity in Group 5 enzymes, may suggest that these group represent extant evolutionary intermediates. However, generally low bootstrap values for many clades preclude definitive conclusions from being drawn.
in this regard. Most distinctly, members of Group 5 have evolved a large hydrophobic platform of 10 subsites through a series of point mutations and loop insertions, which engender a processive mode of action.

The biological basis of the molecular selection for processivity across a wide range of Group 5 members is not immediately intuited. Processivity is generally considered to be advantageous for enzymes acting on crystalline substrates such as cellulose or chitin, where initial chain engagement is thought to be rate-limiting (42–47). However, this would not be expected for soluble polysaccharides, such as XyG, especially under dilute assay conditions in vitro. In the plant cell wall, XyG associates with crystalline cellulose microfibrils and other matrix glycans in an amorphous, hydrated state (63–65).

Hence, we hypothesize that processivity in GH74 may be utilized in the context of substrate sensing, in which the initial, rapid release of short, highly diffusible XyG oligosaccharides acts as a signal to up-regulate the production of cognate enzymes (66, 67). In contrast, classical endo-dissociative activity predominantly generates large polysaccharide fragments during early stages of attack, which would remain associated with the cell wall. Supporting this proposal, recent transcriptomics analysis revealed that the gene which would remain associated with the cell wall.

...
Isolated colonies of the transformed *E. coli* cells were inoculated in lysogeny broth medium containing ampicillin (100 μg/ml) and grown overnight at 37 °C with rotary shaking at 200 rpm. Precultures were used to inoculate ZYP5052 autoinducing medium (82) containing ampicillin (100 μg/ml). Cultures were grown at 37 °C for 4.5 h and transferred to 16 °C for overnight incubation with rotary shaking at 200 rpm until reaching an A_{600 nm} of approximately 11. Cultures were then centrifuged at 4500 × g for 30 min, and pellets were resuspended in 50 mM sodium phosphate buffer, pH 7.4, 500 mM NaCl, 20 mM imidazole, and the suspension was frozen at −20 °C. Frozen cells were thawed and lysed by the addition of lysozyme (0.5 mg/ml) and benzonase (25 units) followed by incubation at 37 °C for 1 h. In addition, cells were disrupted by sonication, and the cell-free extract was separated by centrifugation at 4 °C (14,500 × g for 45 min).

Recombinant proteins were purified from the cell-free extract with an Acta Purifier FPLC system using a Ni^{2+} affinity column. A gradient up to 100% elution buffer (50 mM sodium phosphate buffer, pH 7.4, 500 mM NaCl, 500 mM imidazole) was applied. The purity of the recombinant proteins was determined by SDS-PAGE and staining with Coomassie Brilliant Blue. Pure fractions were pooled, concentrated, and buffer-exchanged against 50 mM sodium phosphate buffer, pH 7.0.

To determine Michaelis–Menten parameters of recombinant proteins for XyG, different concentrations of substrate solutions were used over the range 0.02–2 mg/ml. The reactions were performed at 37 °C (or 65 °C for thermostable enzymes *CIGH74a*, *CIGH74b*, and *ClBGH74* or 20 °C for *SvGH74a*) in a 50 mM concentration of their optimum buffer (citrate or phosphate at the optimum pH; see Fig. S1), using tamarind seed XyG at a concentration of 0.5 mg/ml and appropriate concentration of recombinant protein (typically around 0.5 μg/ml) at temperatures ranging from 25 to 98 °C.

To determine the products released by recombinant GH74 enzymes, tamarind seed XyG was incubated at 37 °C (or 65 °C for *CIGH74a*, *CIGH74b*, and *ClBGH74*) in a 50 mM concentration of their optimum buffer (citrate or phosphate at the optimum pH; see Fig. S1) at a concentration of 0.5 mg/ml in the presence of 0.1 μg/ml enzyme (or 1 μg/ml for *SvGH74a*). After various incubation times (0, 5, 10, 30, and 60 min), 100 μl of the reaction were sampled and transferred into 100 μl of boiling water for 15 min. The reaction solution was then analyzed by HPAEC-PAD. Limit digestion products were obtained similarly after 72 h using 10 μg/ml enzyme (or 100 μg/ml for *SvGH74a*) and 0.1 mg/ml tamarind seed XyG. Limit digestion products of XXXGXXG and of XXXG were obtained similarly after overnight incubation of 5 μM substrate with 1 μg/ml enzyme (or 10 μg/ml for *SvGH74a*).

**Carbohydrate sources**

Tamarind seed xyloglucan, konjac glucomannan, barley β-glucan, wheat flour arabinoxylan, and beechwood xylan were obtained from Megazyme (Bray, Ireland). Hydroxyethyl-cellulose was purchased from Amresco (Solon, OH) and carboxymethyl cellulose from Acros Organics (Morris Plains, NJ). pNP-β-D-xlyopyranoside and pNP-β-D-glucopyranoside were obtained from Sigma-Aldrich. A mixture of XyGOS (XXXG, XLXG, XXLG, and XLLG), XXXG, and XXXGXXXG were prepared from tamarind seed XyG as described previously (58).

**Carbohydrate analytics**

HPAEC-PAD and MALDI-TOF MS were performed exactly as described previously (35).

**Enzyme kinetics and product analysis**

For all enzyme assays on polysaccharides, the activity was determined using the BCA assay as described previously (84). Substrate specificity was determined in 50 mM sodium phosphate buffer, pH 7.0, using 0.5 mg/ml substrate and 1 μg/ml enzyme overnight at 37 °C. The optimum pH was established in 50 mM citrate buffer, pH 3.0, 4.0, 5.0, 5.5, and 6.0, or 50 mM sodium phosphate buffer, pH 6.0, 6.5, 7.0, and 8.0. The optimum temperature was determined in a 50 mM concentration of the optimum buffer (citrate or phosphate at the optimum pH; see Fig. S1), using tamarind seed XyG at a concentration of 0.5 mg/ml and appropriate concentration of recombinant protein (typically around 0.5 μg/ml) at temperatures ranging from 25 to 98 °C.

For all enzyme assays, the activity was determined using the BCA assay as described previously (84). Substrate specificity was determined in 50 mM sodium phosphate buffer, pH 7.0, using 0.5 mg/ml substrate and 1 μg/ml enzyme overnight at 37 °C. The optimum pH was established in 50 mM citrate buffer, pH 3.0, 4.0, 5.0, 5.5, and 6.0, or 50 mM sodium phosphate buffer, pH 6.0, 6.5, 7.0, and 8.0. The optimum temperature was determined in a 50 mM concentration of the optimum buffer (citrate or phosphate at the optimum pH; see Fig. S1), using tamarind seed XyG at a concentration of 0.5 mg/ml and appropriate concentration of recombinant protein (typically around 0.5 μg/ml) at temperatures ranging from 25 to 98 °C.

**X-ray crystallography**

The *PgGH74* and *AVF00434* proteins were produced as selenomethionine-substituted derivatives using the standard M9 high-yield growth procedure according to the manufacturer’s instructions (Shanghai Mediclon) and purified as described above. AVF00434 was also purified as the native protein to obtain higher-resolution crystals. All other proteins were purified as native proteins for crystallography. All crystals were grown using the sitting-drop method at 22 °C. The following protein and reservoir solutions were utilized for crystal growth: AVF00434 (SelMet), 25 mM zinc acetate, 20% (w/v) PEG 3350, 1 mM magnesium sulfate; AVF00434 (native), 25 mM zinc acetate, 20% (w/v) PEG 3350, 1.5% (w/v) 2-methyl-2,4-pentanediol; *NkGH74*, protein + XyGO mixture (i.e. XXXG, XLXG, XXLG, and XLLG), 1 M ammonium sulfate, 1 M sodium chloride, 0.1 M Bistris propane, pH 7; *CIGH74a*, protein + XyGO mixture, 25% PEG 3350 (w/v), 0.1 M Tris pH 8.5; *PmGH74*, 25% (w/v) PEG 14002, 1.5% (w/v) 2-methyl-2,4-pentanediol; *ClBGH74a*, protein + XyGO mixture, 25% PEG 3350 (w/v), 0.1 M Tris pH 8.5; *PmGH74*, 25% (w/v) PEG
3350, 0.2 M sodium chloride, 0.1 M sodium citrate, pH 5.6, 0.5% (w/v) glycerol; SrGH74, protein + XyGO mixture, 1.6 M ammonium sulfate, 0.1 M sodium chloride, 0.1 M Heps, pH 7.5. Crystals were cryoprotected with glycerol, PEG 200, or paratone oil before flash freezing in a liquid nitrogen stream.

X-ray diffraction data were collected at beamline 19-ID/BM of the Structural Biology Center, Advanced Photon Source, Argonne National Laboratory (Argonne, IL) for PgGH74 SelMet, SrGH74 native, and AFV00434 SelMet and native, beamline 08-ID at the Canadian Macromolecular Crystallography Facility, Canadian Light Source (Saskatoon, Saskatchewan, Canada) for native NkGH74a), or on a Rigaku HF-007 home source with an R-AXIS IV detector (for native CjGH74a). Data for PgGH74 and AFV00434 SelMet crystals were collected at the selenomethionine absorption peak wavelength. X-ray diffraction data were reduced using HKL-3000 (85).

The structure of AFV00434 SelMet was solved using SAD phasing using Phenix.solve (86) and Phenix.autobuild; subsequent refinement was completed using higher-resolution crystals of AFV00434 native protein using this initial model. The structure of PgGH74 was also solved using SAD phasing and Phenix.solve. The structures of NkGH74, SrGH74, and CjGH74 were solved by Molecular Replacement and Phenix.phaser using models constructed by the Phyre2 server (87) onto PoGH74 (PDB code 6MGL), a putative xylolucanase from Streptomyces sp. SirexAA-E (PDB code 5JWZ), and C. japonicus GH74 (PDB code 5FKQ), respectively.

Phenix.autobuild, Phenix.refine, and Coot (88) were used for refinement and model building. The presence of xylolucan was readily apparent in $F_o - F_c$ maps after resolving the positions of the protein atoms. All B-factors were refined, and TLS parameterization was included in the final rounds of refinement. All geometry was verified using the Phenix and the wwPDB server, and structures were deposited to the Protein Data bank with accession numbers 6P2K, 6P2M, 6P2L, 6P2N, and 6P2O for S. agarivorus AFV00434, C. lac-toaceticus GH74a in complex with the XyG fragment LLG, N. koreensis GH74 in complex with two XyG fragments (XXLG and XXXG), and P. graminis GH74 and S. rapa-mycinicus GH74 in complex with two XyG fragments (XLG and XXXG), respectively. All X-ray crystallographic statistics are provided in Table S3.

Acknowledgments—We thank Nobuhiko Watanabe and Bogulav Nocek for synchrotron diffraction data collection and/or structure solution for NkGH74, SrGH74, and AFV00434. NkGH74 and SrGH74 data were collected at the Canadian Macromolecular Crystallography Facility (89). For AFV00434, structural work presented in this paper was performed at Argonne National Laboratory, Structural Biology Center at the Advanced Photon Source. Argonne is operated by U. Chicago Argonne, LLC, for the United States Department of Energy, Office of Biological and Environmental Research under contract DE-AC02-06CH11357. Waters Corp. is gratefully acknowledged for the provision of the intact protein LC-MS system used in the present study. We thank Dr. Julie Grondin for critically reading and editing the manuscript.

References

1. Bar-On, Y. M., Phillips, R., and Milo, R. (2018) The biomass distribution on Earth. Proc. Natl. Acad. Sci. U.S.A. 115, 6506–6511 CrossRef Medline

2. Himmel, M. E., Ding, S. Y., Johnson, D. K., Adney, W. S., Nimlos, M. R., Brady, J. W., and Fout, T. D. (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. Science 315, 804–807 CrossRef Medline

3. Scheller, H. V., and Ulvskov, P. (2010) Hemicelluloses. Annu. Rev. Plant Biol. 61, 263–289 CrossRef Medline

4. Schultink, A., Liu, L., Zhu, L., and Pauly, M. (2014) Structural diversity and function of xylolucan sidechain substituents. Plants (Basel) 3, 526–542 CrossRef Medline

5. Galloway, A. F., Pedersen, M. J., Merry, B., Marcus, S. E., Blacker, J., Benning, L. G., Field, K. L., and Knox, J. P. (2018) Xylolucan is released by plants and promotes soil particle aggregation. New Phytol. 217, 1128–1136 CrossRef Medline

6. Mishra, A., and Malhotra, A. V. (2009) Tamarind xylolucan: a polysaccharide with versatile application potential. J. Mater. Chem. 19, 8528–8536 CrossRef Medline

7. Zhou, Q., Rutland, M. W., Teeri, T. T., and Brumer, H. (2007) Xylolucan in cellulose modification. Cellulose 14, 625–641 CrossRef Medline

8. Tuomivaara, S. T., Yaoi, K., O’Neill, M. A., and York, W. S. (2015) Generation and structural validation of a library of diverse xylolucan-derived oligosaccharides, including an update on xylolucan nomenclature. Carbohydr. Res. 402, 56–66 CrossRef Medline

9. Attia, M. A., Nelson, C. E., Offen, W. A., Jain, N., Davies, G. J., Gardner, J. J., and Brumer, H. (2018) In vitro and in vivo characterization of three Cellvibrio japonicus glycoside hydrolase family 5 members reveals potent xylolucan backbone-cleaving functions. Biotechnol. Biofuels 11, 45 CrossRef Medline

10. Nelson, C. E., Attia, M. A., Rogowski, A., Morland, C., Brumer, H., and Gardner, J. G. (2017) Comprehensive functional characterization of the glycoside hydrolase family 3 enzymes from Cellvibrio japonicus reveals unique metabolic roles in biomass saccharification. Environ. Microbiol. 19, 5025–5039 CrossRef Medline

11. Attia, M. A., and Brumer, H. (2016) Recent structural insights into the enzymology of the ubiquitous plant cell wall glycan xylolucan. Curr. Opin. Struct. Biol. 40, 43–53 CrossRef Medline

12. Larsbrink, J., Rogers, T. E., Hemsworth, G. R., McKee, L. S., Tazuin, A. S., Spadut, O., Klinter, S., Pudlo, N. A., Urs, K., Koropatkin, N. M., Creagh, A. L., Haynes, C. A., Kelly, A. G., Cederholm, S. N., Davies, G. J., et al. (2014) A discrete genetic locus confers xylolucan metabolism in select human gut Bacteroidetes. Nature 506, 498–502 CrossRef Medline

13. Lombard, V., Golaconda Ramulu, H., Druli, E., Coutinho, P. M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. 42, D490–D495 CrossRef Medline

14. Arai, M., Sakamoto, R., and Murao, S. (1989) Different action by two avicelases from Cellvibrio japonicus on various cellulose types and enzymes for biofuels production. J. Indust. Environ. Microbiol. 31, 56–66 CrossRef Medline

15. Montreal-Méndez, R., Zühlik, D., Becher, D., Riedel, K., and Baldran, P. (2016) Cellulose and hemicellulose decomposition by forest soil bacteria
proceeds by the action of structurally variable enzymatic systems. *Sci. Rep.* **6**, 25279 *CrossRef Medline*

16. Attia, M., Stepper, J., Davies, G. I., and Brumer, H. (2016) Functional and structural characterization of a potent GH74 endo-xyloglucanase from the soil saprophyte *Cellvibrion japonicus* unravels the first step of xyloglucan degradation. *FEBS J.* **283**, 1701–1719 *CrossRef Medline*

17. Bauer, S., Vasu, P., Mort, A. J., and Somerville, C. R. (2005) Cloning, expression, and characterization of an oligosaccharin reducing end-specific xyloglucanohydrolase from *Aspergillus nidulans*. *Carbohydr. Res.* **340**, 2590–2597 *CrossRef Medline*

18. Berezina, O. V., Herlet, J., Rykov, S. V., Kornberger, P., Zavyalov, A., Kožlov, D., Sahibgaravae, L., Krestyana, I., Schwarz, W. H., Zverlov, V. V., Liebl, W., and Yarotsky, S. V. (2017) Thermolabile multifunctional GH74 xyloglucanase from *Mycelocellulithora thermophila*—high-level expression in *Pichia pastoris* and characterization of the recombinant protein. *Appl. Microbiol. Biotechnol.* **101**, 5653–5666 *CrossRef Medline*

19. Chhabra, S. R., and Kelly, R. M. (2002) Biochemical characterization of *Thermotoga maritima* endogluccanase Cell74 with and without a carbohydate binding module (CBM). *FEBS Lett.* **531**, 375–380 *CrossRef Medline*

20. Damasio, A. R., Rubio, M. V., Gonçalves, T. A., Persinoti, G. F., Segato, F., Prade, R. A., Contesini, F. J., de Souza, A. P., Buckeridge, M. S., and Squina, F. M. (2017) Xyloglucan breakdown by endo-xyloglucanase family 74 from *Aspergillus fumigatus*. *Appl. Microbiol. Biotechnol.* **101**, 2893–2903 *CrossRef Medline*

21. Desmet, T., Cantaert, T., Gualfetti, P., Nerinckx, W., Gross, L., Mitchinson, C., and Piams, K. (2007) An investigation of the substrate specificity of the xyloglucanase Cell74A from *Hypocrea jecorina*. *FEBS J.* **274**, 356–363 *CrossRef Medline*

22. Enkhbaatar, B., Temuujin, U., Lim, J. H., Chi, W. J., Chang, Y. K., and Shahab, N., Sorek, R., Tringe, S. G., Podar, M., Martin, H. G., Kunin, V., and *et al.* (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450**, 560–565 *CrossRef Medline*

23. Blumer-Schuette, S. E., Lewis, D. L., and Kelly, R. M. (2010) Phylogenetic, microbiological, and glycoside hydrolase diversity within the extremophilic, plant biomass-degrading genus *Caldicellulosiruptor*. *Appl. Environ. Microbiol.* **76**, 8084–8092 *CrossRef Medline*

24. Lacerda Júnior, G. V., Noronha, M. F., de Sousa, S. T. P., Cabral, L., Domingos, D. F., Sá, M. L., de Melo, I. S., and Oliveira, V. M. (2017) Potential of semiarid soil from Caatinga biome as a novel source for lignocellulose-degrading enzymes. *FEMS Microbiol. Ecol.* **93**, fiw248 *CrossRef Medline*

25. DeBoy, R. T., Mongodin, E. F., Fouts, D. E., Talford, L. E., Khouri, H., Emerson, J. B., Mohamoud, Y., Watkins, K., Henriassat, B., Gilbert, H. J., and Nelson, K. E. (2008) Insights into plant cell wall degradation from the genome sequence of the bacterium *Cellvibrion japonicus*. *J. Bacteriol.* **190**, 5455–5463 *CrossRef Medline*

26. Matsuura, T., and Yaoi, K. (2016) GH74 xyloglucanases: structures and modes of activity. *Trends Glycobiol. Glycotechnol.* **28**, E63–E70 *CrossRef Medline*

27. Jakal, J., Kurasín, M., Teugjas, H., and Válčík, P. (2012) Endo-xyloglucanases: Current thinking about the extent of recalcitrant polysaccharide degradation. *FEBS Lett.* **588**, 4620–4624 *CrossRef Medline*

28. Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttilä, M., Ando, T., and Samejima, M. (2011) Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* **333**, 1279–1282 *CrossRef Medline*

29. Igarashi, K., Uchihashi, T., Uchiyama, T., Sugimoto, H., Wada, M., Suzuki, K., Sakuda, S., Ando, T., Watanabe, T., and Samejima, M. (2014) Two-way traffic of glycosidase hydrolase family 18 processive chitinases on crystalline chitin. *Nat. Commun.* **5**, 9755 *CrossRef Medline*

30. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (2014) Release of the catalytic domain and two cellulose binding modules of a family-74 xyloglucanase. *FEBS Lett.* **588**, 200208–200215 *CrossRef Medline*

31. Igarashi, K., Uchihashi, T., Koivula, A., Wada, M., Kimura, S., Okamoto, T., Penttilä, M., Ando, T., and Samejima, M. (2011) Traffic jams reduce hydrolytic efficiency of cellulase on cellulose surface. *Science* **333**, 1279–1282 *CrossRef Medline*

32. Igarashi, K., Uchihashi, T., Uchiyama, T., Sugimoto, H., Wada, M., Suzuki, K., Sakuda, S., Ando, T., Watanabe, T., and Samejima, M. (2014) Two-way traffic of glycosidase hydrolase family 18 processive chitinases on crystalline chitin. *Nat. Commun.* **5**, 9755 *CrossRef Medline*

33. Irwin, D., Shin, D. H., Zhang, S., Barr, B. K., Sakon, J., Karplus, P. A., and Wilson, D. B. (1998) Roles of the catalytic domain and two cellulose binding domains of *Thermomonospora fusca* E4 in cellulose hydrolysis. *J. Bacteriol.* **180**, 1709–1714 *CrossRef Medline*

34. Christensen, S. I., Kari, I., Bardino, S. F., Borch, K., and Westph, P. (2018) Rate-limiting step and substrate accessibility of cellulohydrolase Cel6A from *Trichoderma reesei*. *FEBS J.* **285**, 4482–4493 *CrossRef Medline*

35. Henriassat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J. P. (1989) Cellulase families revealed by phodobacteric analysis. *Gene* **81**, 83–95 *CrossRef Medline*
53. Davies, G. J., and Sinnott, M. L. (2008) Sorting the diverse. J. Biol. Chem. 283, 5094–5100 CrossRef Medline

54. Yaoi, K., Kondo, H., Hiyoshi, A., Noro, N., Sugimoto, H., Tsuda, S., and Miyazaki, K. (2007) The structural basis for the exo-mode of action in GH74 oligoxygenase reducing end-specific cellobiohydrolase. J. Mol. Biol. 370, 53–62 CrossRef Medline

55. Kari, J., Olsen, J. B., Cruys-Bagger, N., Jensen, K., and Westh, P. (2006) Crystal structures of Clostridium thermocellum xylooligocanse, XGH74A, reveal the structural basis for xylooligocanse recognition and degradation. J. Biol. Chem. 281, 24922–24933 CrossRef Medline

56. Martinez-Fleites, C., Guerreiro, C. I., Baumann, M. J., Taylor, E. J., Prates, J. A., Ferreira, L. M., Fontes, C. M., Brumer, H., and Davies, G. J. (2006) Crystal structures of Acidothermus cellulolyticus family 74 glycoside hydrolase at 1.82 Å resolution. Acta Crystallogr. Sect. F Struc. Biol. Cryst. Commun. 69, 1335–1338 CrossRef Medline

57. Geotrichum candidum xyloglucanase reveals a key amino acid residue for substrate specificity. FEBS J. 276, 5094–5100 CrossRef Medline

58. Eklöf, J. M., Ruda, M. C., and Brumer, H. (2012) Distinguishing xylooligocanse activity in endo-β-(1→4)galactans. Methods Enzymol. 510, 97–120 CrossRef Medline

59. Aspeborg, H., Coutinho, P. M., Wang, Y., Brumer, H., 3rd, and Henrissat, B. (2012) Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5). BMC Evol. Biol. 12, 186 CrossRef Medline

60. St. John, F. J., González, J. M., and Pozharski, E. (2010) Consolidation of cell wall. assessment by FESEM, selective enzyme digestion and nanogold affinity tags. Plant J. 63, 211–226 CrossRef Medline

61. Davies, G. J., and Sinnott, M. L. (2008) Sorting the diverse. The Biochemist, 30, 26–32

62. Petersen, T. N., Brunak, S., von Heijne, G., and Nielsen, H. (2011) SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat. Methods 8, 785–786 CrossRef Medline

63. Rozeboom, H. J., Boldman, G., Schols, H. A., and Dijkstra, B. W. (2013) Crystal structure of endo-xyloglacturonan hydrolase from Aspergillus tubingensis. FEBS J. 280, 6061–6069 CrossRef Medline

64. Tsukagoshi, H., Nakamura, A., Ishida, T., Otagiri, M., Moriya, S., Samejima, M., Igarashi, K., Kitamoto, K., and Arioka, M. (2014) The GH26 β-mannanase RsMan26H from a symbiotic protist of the termite Reticillitenes speratus is an endo-processive mannanohydrolase: heterologous expression and characterization. Biochem. Biophys. Res. Commun. 452, 520–525 CrossRef Medline

65. Zheng, Y., Wang, X., Chen, Y., Wagner, E., and Cosgrove, D. J. (2018) Structural dynamics and catalytic properties of a multimodular xantha-

66. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., and Brumer, H. (2014) Crystal structures of the substrate specificity of the “sensing” xylanase of Streptomyces cyaneus using xylooligosaccharide and cellobio-oligosaccharide glycosides of 3,4-dinitrophenol. J. Biotechnol. 57, 181–190 CrossRef Medline

67. Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W.,
McCoy, A. J., Moriarty, N. W., Oefner, R., Read, R. J., Richardson, D. C., et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213–221 CrossRef Medline

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., and Sternberg, M. J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. 10, 845–858 CrossRef Medline

Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501 CrossRef Medline

Grochulski, P., Fodje, M. N., Gorin, J., Labiuk, S. L., and Berg, R. (2011) Beamline 08ID-1, the prime beamline of the Canadian Macromolecular Crystallography Facility. J. Synchrotron Radiat. 18, 681–684 CrossRef Medline

EDITORS’ PICK: GH74 structure–function analysis

88. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501 CrossRef Medline
89. Grochulski, P., Fodje, M. N., Gorin, J., Labiuk, S. L., and Berg, R. (2011) Beamline 08ID-1, the prime beamline of the Canadian Macromolecular Crystallography Facility. J. Synchrotron Radiat. 18, 681–684 CrossRef Medline