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

Grégory Arnal, Peter J Stogios, Jathavan Asohan, Mohamed Attia, Tatiana Skarina, Alexander Holm Viborg, Bernard Henrissat, Alexei Savchenko, Harry Brumer

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

Received for publication, June 20, 2019, and in revised form, July 16, 2019 Published, Papers in Press, July 19, 2019, DOI 10.1074/jbc.RA119.009861

Gregory Arnal[‡], Peter J. Stogios[§], Jathavan Asohan[‡], Mohamed A. Attia[‡¶], Tatiana Skarina[§], Alexander Holm Viborg[‡], Bernard Henrissat[‖], Alexei Savchenko[§‡‡1], and Harry Brumer[‡¶§§¶¶2]**

*From the [‡]Michael Smith Laboratories, University of British Columbia, Vancouver, British Columbia V6T 1Z4, Canada, the [§]Department of Chemical Engineering and Applied Chemistry, University of Toronto, Toronto, Ontario M5S 3E5, Canada, the [¶]Department of Chemistry, University of British Columbia, Vancouver, British Columbia V6T 1Z1, Canada, [‖]Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, Aix-Marseille University, 13007 Marseille, France, **INRA, USC1408 Architecture et Fonction des Macromolécules Biologiques (AFMB), Marseille 13007, France, the [‡‡]Department of Microbiology, Immunology, and Infectious Diseases, University of Calgary, Calgary, Alberta T2N 4N1, Canada, the [§§]Department of Biochemistry and Molecular Biology, University of British Columbia, Vancouver, British Columbia V6T 1Z3, Canada, and the [¶¶]Department 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*-dissociative, 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 structure–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 byproducts 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.

[3] The abbreviations used are: XyG, xyloglucan; XyGO, xyloglucan oligosaccharide; GH, glycoside hydrolase; CBM, carbohydrate-binding module; CAZy, carbohydrate active enzyme; HPAEC-PAD, high-performance anion-exchange chromatography coupled with pulsed amperometric detection; HCA, hydrophobic cluster analysis; *p*NP, *p*-nitrophenyl; OXG-RCBH, oligoxyloglucan reducing end-specific cellobiohydrolase; PDB, Protein Data Bank; SelMet, selenomethionine; gDNA, genomic DNA; LIC, ligation-independent cloning; Bistris propane, 1,3-bis[tris(hydroxymethyl)methylamino]propane; SAD, single anomalous dispersion.

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-xyloglucanases, 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 xyloglucanases (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*-xyloglucanases, EC 3.2.1.151) or at the chain end (*exo*-xyloglucanases, EC 3.2.1.150). *Endo*-xyloglucanases 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 *p*NP 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 beechwood 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 xyloglucanase 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 lactoaceticus* 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, *Thermotoga 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 xyloglucanases has been described for a limited number of enzymes. Oligoxyloglucan reducing end-specific cellobiohydrolases (OXG-RCBHs) (EC 3.2.1.150) are *exo*-type xyloglucanases that release Glc_2-based products such as XG or LG from tamarind XyG (17, 49). *Endo*-xyloglucanases can act in a dissociative fashion to generate a

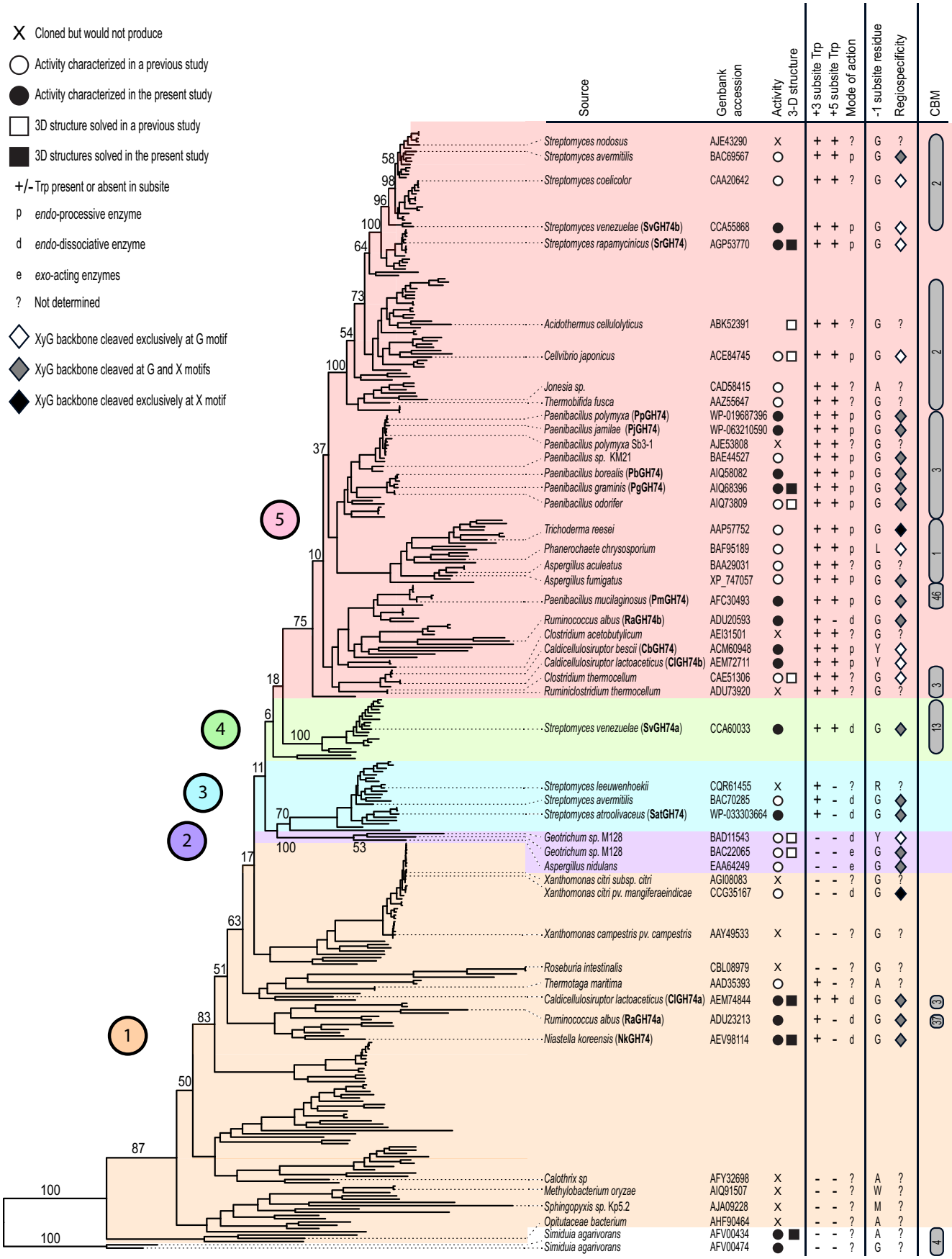

Table 1
Biochemical properties of recombinant GH74 modules on tamarind XyG

Protein	Bacterial strain	Optimum temperature/pH	K_m	k_{cat}
			mg/ml	s^{-1}
*Cl*GH74a	*Caldicellulosiruptor lactoaceticus*	81 °C/pH 5	0.028 ± 0.001	71.8 ± 1.2
*Cl*GH74b	*C. lactoaceticus*	83 °C/pH 6	0.014 ± 0.002	44.9 ± 1.4
*Cb*GH74	*Caldicellulosiruptor bescii*	85 °C/pH 6.5	0.04 ± 0.02	51.2 ± 3.2
*Nk*GH74	*Niastella koreensis*	55 °C/pH 4.5	0.09 ± 0.008	18.1 ± 0.7
*Pg*GH74	*Paenibacillus graminis*	55 °C/pH 6	0.039 ± 0.001	170.2 ± 4.6
*Pb*GH74	*Paenibacillus borealis*	45 °C/pH 6	0.022 ± 0.001	91.8 ± 1.4
*Pp*GH74	*Paenibacillus polymyxa*	54 °C/pH 6	0.095 ± 0.010	26.9 ± 1.4
*Pj*GH74	*Paenibacillus jamilae*	60 °C/pH 6	0.057 ± 0.013	24.4 ± 1.4
*Pm*GH74	*Paenibacillus mucilaginosus*	60 °C/pH 6	0.117 ± 0.011	96.5 ± 2.6
*Ra*GH74a	*Ruminococcus albus*	50 °C/pH 6	0.028 ± 0.003	47.5 ± 2.8
*Ra*GH74b	*R. albus*	45 °C/pH 6	0.314 ± 0.045	37.9 ± 1.9
AFV00434	*Simiduia agarivorans*	NA[a]	NA	NA
AFV00474	*S. agarivorans*	NA	NA	NA
*Sr*GH74	*Streptomyces rapamycinicus*	65 °C/pH 6	0.014 ± 0.001	45.2 ± 2.6
*Sv*GH74a	*Streptomyces venezuelae*	25 °C/pH 6	1.05 ± 0.34	1.0 ± 0.1
*Sv*GH74b	*S. venezuelae*	55 °C/pH 7	0.023 ± 0.003	52.3 ± 2.2
*Sat*GH74	*Streptomyces atroolivaceus*	54 °C/pH 6	0.09 ± 0.01	31.1 ± 1.0

[a] NA, not active on xyloglucan.

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 glucosyl residues (16, 22, 28, 33, 50); however, some cleave exclusively at xylosylated glucosyl units (21, 23), and others can cleave at both G and X motifs (20, 26, 32, 35). To investigate the backbone regiospecificity of GH74 enzymes, the limit digests of tamarind XyG, of XXXGXXXG, and of XXXG were analyzed by HPAEC-PAD (Fig. 3). The details of these analyses for 17 enzymes are discussed below in the context of five GH74 phylogenetic groups that delineate processivity and cleavage regiospecificity (Fig. 1).

Distantly related proteins—Proteins AFV00434 and AFV00474 from the marine bacterium *S. agarivorans* have very low sequence similarity to GH74 members, yet a distant relationship was detected by HCA (48) (data not shown). To investigate the structural basis for their lack of activity on XyG, we solved the tertiary structure of *Sa*AFV00434 (PDB code 6P2K) using SelMet-derivatized protein and single anomalous dispersion (SAD) phasing. This protein shares the canonical GH74 structure (16, 35, 49–52), comprising two seven-bladed β-propeller domains forming a long and wide cleft (Fig. 4*A*). This structure validates the HCA prediction of a distant relationship to GH74 and suggests a superfamily or "clan" (53). Two aspartic acid residues (Asp^{35} and Asp^{419}), located 8.3 Å apart in the putative catalytic site, have an adequate spatial position to catalyze the hydrolysis of the glycosidic bond via an inverting mechanism. However, comparison of the protein backbone in the crystal structure of AFV00434 and Group 5 *Paenibacillus odorifer* GH74 (PDB entry 6MGL) (35) revealed that AFV00434 loops Trp^{121}–Ala^{136}, Ser^{468}–Asn^{476}, and Ser^{706}–Tyr^{716} obstruct the active cleft at subsites −3/−2, +1/+2, and −4/−3, respectively, providing a possible explanation for the lack of polysaccharide hydrolysis (Fig. 4*A*). Furthermore, *Sa*AFV00434 lacks apparent +3, +4, and +5 subsites (*i.e.* aromatic amino acids available for interactions with xyloglucan or other polysaccharides; see below).

Group 1—As defined here by the limits of our ability to discriminate members on the basis of enzyme activity, phylogenetic Group 1 (Fig. 1) encompasses a very sequence-diverse set of enzymes. These 123 bacterial enzymes belong mainly to the phyla *Proteobacteria* and *Firmicutes*, as well as *Cyanobacteria*, but low sequence conservation (identity <50%) results in long branches on the phylogenetic tree (Fig. 1).

Some of the most divergent enzymes we attempted to study could not be recombinantly produced. However, we successfully produced *N. koreensis* GH74, *Ruminococcus albus* GH74a, and *C. lactoaceticus* GH74a. These three enzymes acted 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.

Figure 1. Phylogenetic tree and summary of modes of action of GH74 enzymes. This phylogeny is based on isolated GH74 catalytic module sequences (*i.e.* with CBMs removed). Maximum-likelihood phylogenetic Groups 1–5 are *highlighted* and *numbered*, with bootstrap values indicated at selected nodes. Biochemical properties are indicated as described in the *key*. Data from previous studies comprise BAC69567 and BAC70285 (26), CAA20642 (22), ABK52391 (51), ACE84745 (16) (and this study), CAD58415 (30), AAZ55647 (27), BAE44527 (29, 32), AAP57752 (21), BAF95189 (28), BAA29031 (14), XP_747057 (20), CAE51306 (34, 50), BAD11543 (33, 52), BAC22065 (31, 52), EAA64249 (17), CCG35167 (23), and AAD35393 (19).

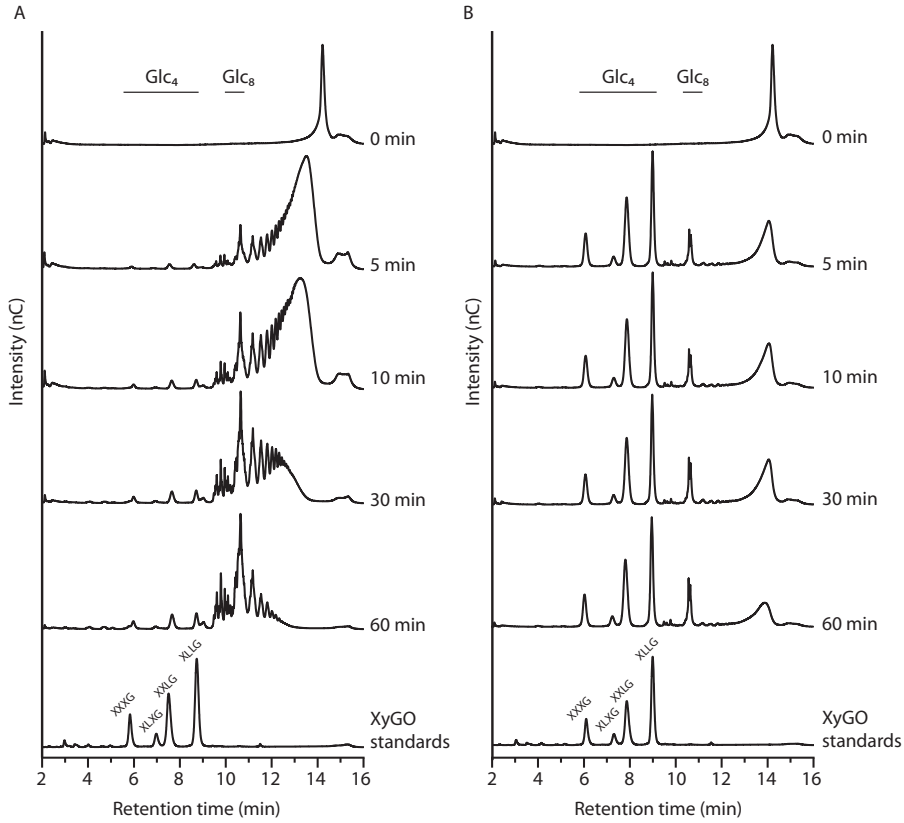

Figure 2. Time-course hydrolysis of tamarind XyG monitored by HPAEC-PAD. *A*, *endo*-dissociative *Nk*GH74; *B*, *endo*-processive *Sr*GH74.

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*B*). These represent the first three-dimensional structures described in Group 1. Vis-à-vis *Sa*AFV00434 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, Trp^{328} and Trp^{329}, 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 (Trp^{328} and Trp^{337}), but instead of being found consecutively in the primary structure, they are interspersed with a loop comprising Ser^{329}–Thr^{336} (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 (Tyr^{122}, Trp^{126}, Trp^{328}, Trp^{329}, and Trp^{375}) and *N. koreensis* GH74 (Tyr^{117}, Phe^{118}, Trp^{328}, Trp^{337}, and Trp^{376}) (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

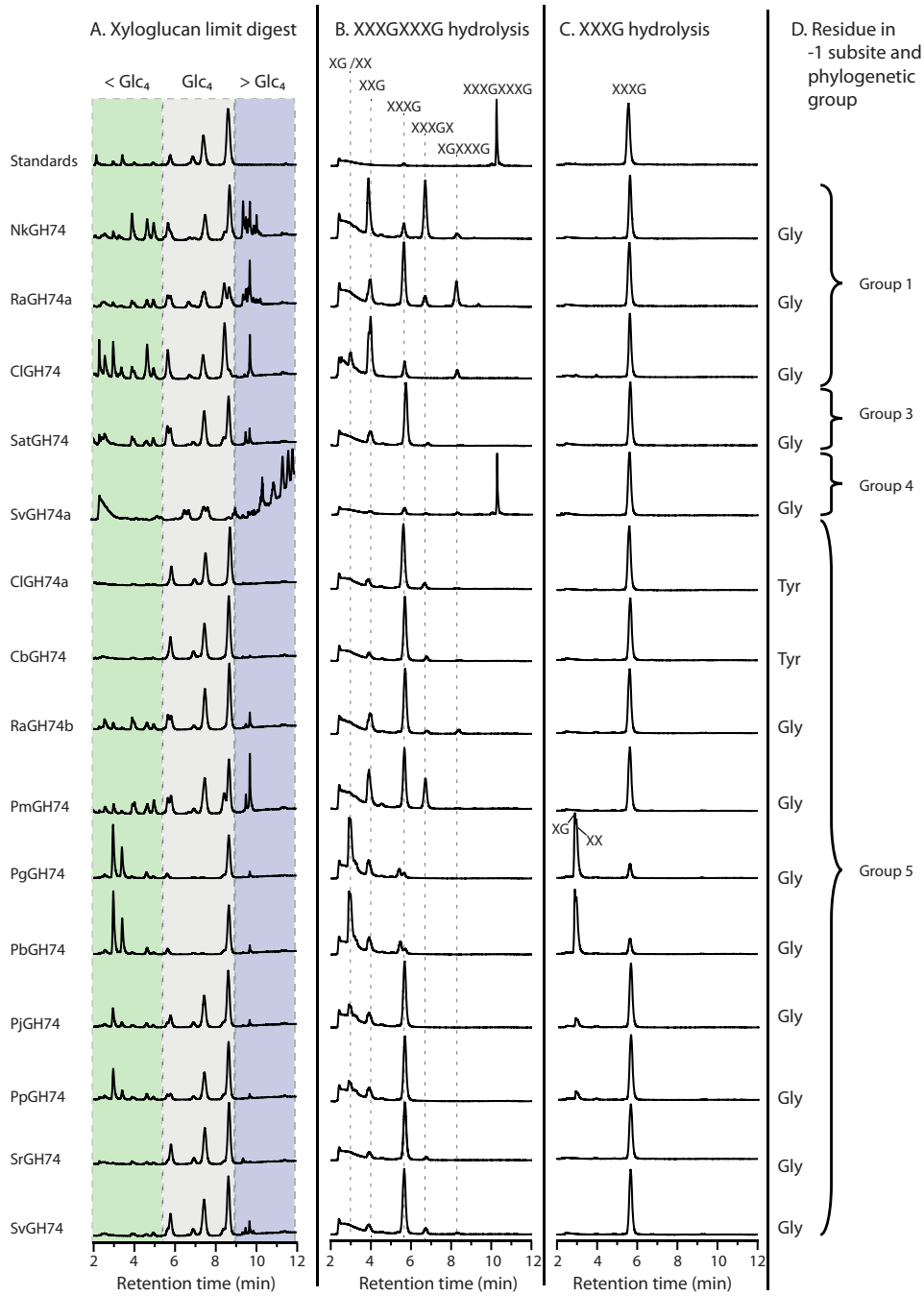

Figure 3. Regiospecificity of recombinant GH74 enzymes. HPAEC-PAD analysis of the limit digest of tamarind xyloglucan polysaccharide (*A*), the limit digest of XXXGXXXG (*B*), and the limit digest of XXXG (*C*) for each enzyme.

Yaoi *et al.* (52) demonstrated that the strict *exo*-activity of the OXG-RCBH enzymes, which results in the production of Glc_2-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 xyloglucanase *P. odorifer* GH74 (35) (see below) (Fig. S5). Within Group 3, *Streptomyces atroolivaceus* GH74 acted as an *endo*-dissociative enzyme (Fig. S4*D*), analogous to the previously characterized

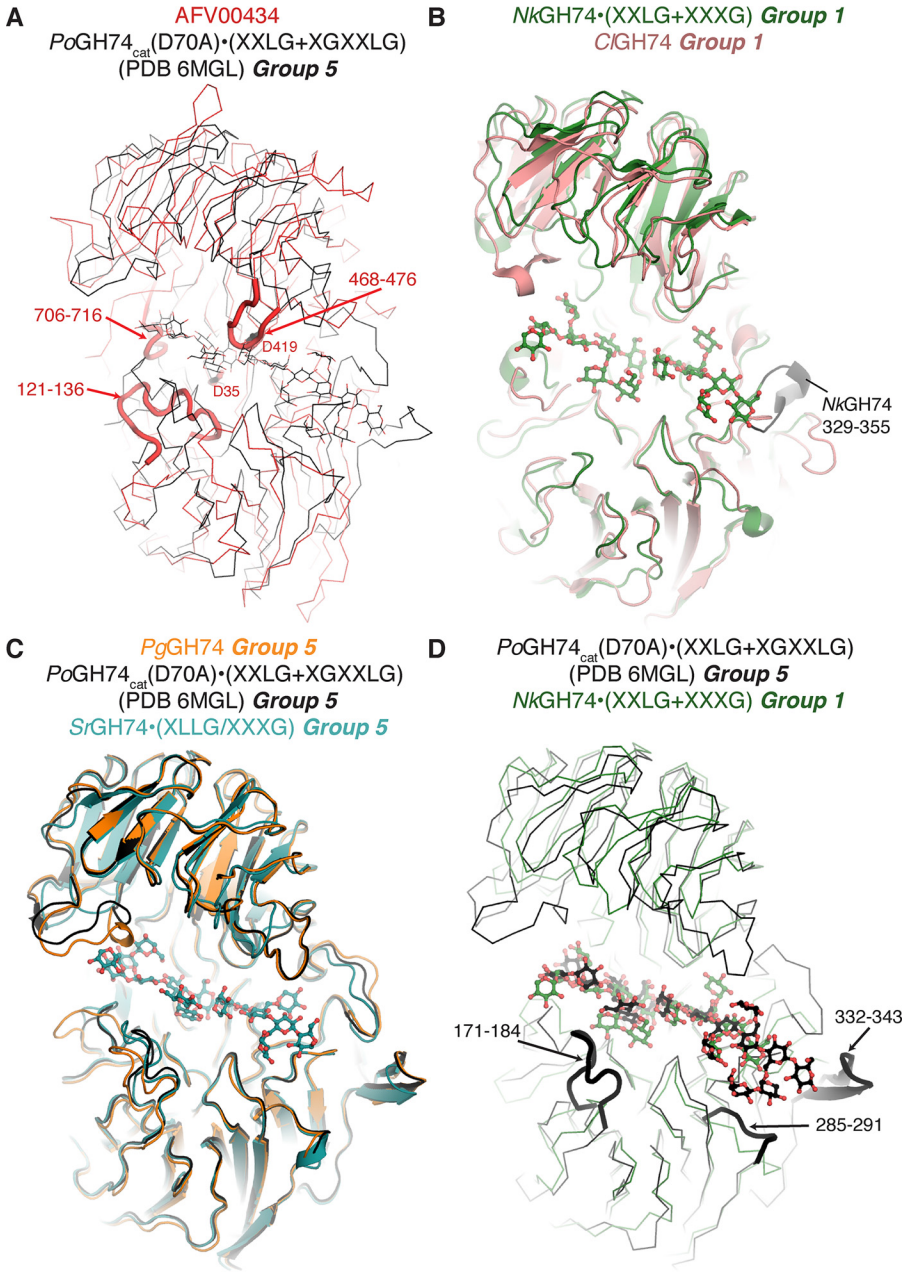

Figure 4. Crystal structures of *S. agarivorans* AFV00434, *N. koreensis* GH74, *C. lactoaceticus* GH74, *P. graminis* GH74, *S. rapamycinicus* GH74, and *P. odorifer* GH74 (PDB code 6MGL). *A*, overlay of crystal structures of AFV00434 (*red*) and Group 5 *Po*GH74·(D70A)·(XXLG + XGXXLG) (PDB code 6MGL) (*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 *Nk*GH74·(XXLG + XXXG) (*green*) and *Cl*GH74a (*pink*) shown in *cartoon representation*. *C*, overlay of crystal structure of Group 5 *Pg*Gh74 (*orange*), *Po*GH74·(D70A)·(XXLG + XGXXLG) (PDB code 6MGL) (*black*), and *Sr*GH74·(XLLG + XXXG) (*cyan*). *D*, overlay of crystal structures of Group 1 *Nk*GH74·(XXLG + XXXG) (*green*) and Group 5 *Po*GH74·(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*.

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. S4*D*). 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.

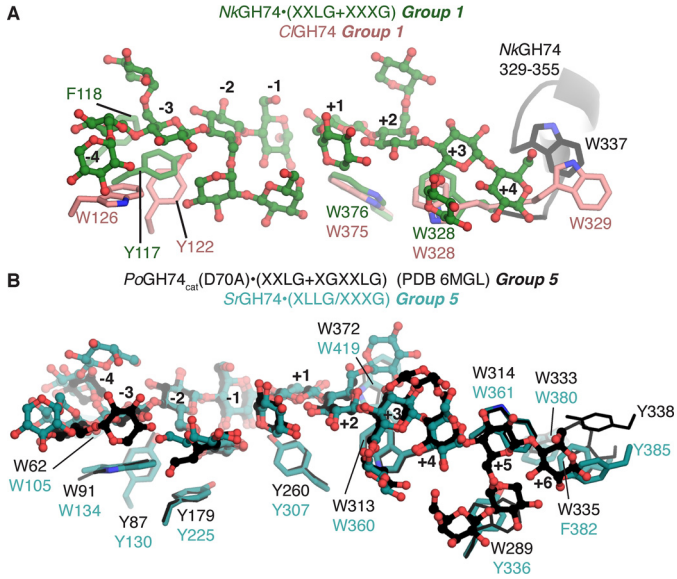

Figure 5. Details of the active sites of the *Nk*GH74·(XXLG + XXXG), *Cl*GH74a, *Po*GH74(D70A)·(XXLG + XGXXLG), and *Sr*GH74·(XLLG + XXXG) complexes. *A*, comparison of *Nk*GH74·(XXLG + XXXG) and *Cl*GH74a, showing that each enzyme possesses five aromatic residues in its active site, including the presence of tryptophan residues at the +3 and +5 subsites. *B*, comparison of *Sr*GH74·(XLLG + XXXG) and *Po*GH74$_{cat}$·(XXLG + XGXXLG) (PDB code 6MGL), showing that the xyloglucan fragments occupy the same position in each enzyme, with small differences in the −4 and −3 subsites. Conservation of the position of all 12 aromatic acids in the active-site clefts of each enzyme is observed.

Nonetheless, time-course hydrolysis of XyG analyzed by HPAEC-PAD analysis clearly indicated that this enzyme acted as an *endo*-dissociative enzyme (Fig. S4*E*) 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. S4*E*).

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*-xyloglucanases 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 xyloglucanase (Fig. S4*H*). 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 xyloglucanases 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. bescii* 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 jamilae* 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 rapamycinicus* 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 XX|XG into a −2 to +2 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 mutant, 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. 4*C*). In particular, these conformational changes caused *P. graminis* GH74 residues Trp^{126}, Tyr^{122}, and Tyr^{214} to rotate out of the active site cleft as compared with their equivalents Trp^{91}, Tyr^{87}, and Tyr^{179} 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. XLLG and XXXG bound to *S. rapamycinicus* GH74 superimposed nearly exactly to XXLG 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. 4*C* and 5*B*).

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 (Tyr^{295}) and *S. rapamycinicus* GH74 (Tyr^{307}) +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 Tyr^{206}–Gly^{215}, Gly^{320}–Tyr^{325}, and Gly^{371}–Ala^{381} provided the scaffold for the insertion of Tyr^{214}, Trp^{325}, and Tyr^{373} in the subsites −2, +5, and +6 in *P. graminis* GH74 (Fig. 4*D*). These loops are conserved across members of Group 5 but are absent in other phylogenetic groups (Fig. S5). Notably, the loop composed of Gly^{371}–Ala^{381} 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 Trp^{406} (+2 subsite) and Tyr^{372} (+6 subsite) contribute to processivity in *P. odorifer* GH74 (35), whereas Trp^{61} (−4 subsite) and Trp^{64} (−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 encoding *C. japonicus* GH74 (a highly efficient, secreted, processive *endo*-xyloglucanase (16) (Fig. S6)) is constitutively expressed at a low level and is not up-regulated in the presence of XyG (9). This regulation contrasts with other highly specific *exo*-glycosidases (GH3, GH31, GH35, and GH95), an *endo*-xyloglucanase (GH5_4), and a transporter (9).

Signal peptide analysis (68) suggests that all of the GH74 enzymes in our study are extracellular, whereas many of those from Group 5 also have CBMs, which is indicative of cell wall targeting. In this context, we might speculate that processivity in GH74 enzymes is independent of XyG type (*i.e.* side-chain composition); processivity appears to be primarily driven by polysaccharide backbone interactions with key active-site aromatic residues, and inspection of the several crystallographic complexes now available reveals little capacity for interaction with distal side-chain residues on xylosyl branches.

Although they are generally associated with saprotrophic organisms (16, 21, 22, 26, 30–33, 35), the extent to which GH74 enzymes might play a role in beneficial plant–microbe interactions remains to be studied (5). We also note that microbial processive glycanases operating on amorphous polysaccharides have been identified among several GH families (69–74), which may also imply a wider deployment of "sensing" (67) enzymes than generally appreciated.

Experimental procedures

Bioinformatic analyses

GH74 protein sequences were extracted from the CAZy database (13) (January 2019), and redundant sequences were excluded using UCLUST (75). In addition to the 22 fungal and 320 bacterial nonredundant GH74 catalytic modules referenced in the CAZy database, previously characterized GH74 modules from *X. citri* pv. *mangiferaeindicae* (GenBank™ accession number CCG35167) (23) and *Aspergillus fumigatus* (GenBank™ XP_747057) (20) and three uncharacterized GH74 modules from *P. polymyxa* Sb3-1, *P. jamilae*, and *S. atroolivaceus* (GenBank™ WP_019687396, WP_063210590, and WP_033303664, respectively) were included. Distantly related sequences were included from the sea bacterium *S. agarivorans* SA1 (GenBank™ accession numbers AFV00434 and AFV00474).

The resulting 342 nonredundant sequences were screened for the presence of a signal peptide using SignalP version 4.0 (68). Modular architecture was inferred from BLASTP analysis (76) and the CAZy database (13). The sequences were aligned with MAFFT G-INS-i (77), and the quality of the alignment was manually inspected in Jalview (78). A maximum-likelihood phylogenetic tree was estimated by RAxML version 8 (79) using File S1, GH74_CatalyticModules_Aligned.mfa as input on the CIPRES gateway (80), using 100 bootstrap replicates and *S. agarivorans* SA1 sequences AFV00434 and AFV00474 as an outgroup. The resulting phylogeny was visualized with FigTree (http://tree.bio.ed.ac.uk/software/figtree/).[4]

Cloning and site-directed mutagenesis

Cloning of target genes was performed as described previously (81). *C. lactoaceticus* (DSM 9545), *C. bescii* (DSM 6725), *N. koreensis* GR20-10 (DSM 17620), *P. graminis* (DSM 15220), *P. borealis* (DSM 13188), *P. polymyxa* (DSM 36), *P. jamilae* (DSM13815), *R. albus* 7 (DSM 20455), *S. agarivorans* SA1 (DSM 21679), *S. rapamycinicus* (DSM 41530), *S. venezuelae* (DSM 40230), and *S. atroolivaceus* (DSM 40137) gDNAs were purchased from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Germany). cDNAs encoding GH74 catalytic modules were PCR-amplified from gDNA using the high-fidelity Q5 DNA polymerase (New England Biolabs) and specific primers (PCR primers are listed in Table S1).

The PCRs were designed such that only the GH74 catalytic module was amplified, thus removing signal peptides and other modules (*e.g.* CBMs), and the sequence was flanked by ligation-independent cloning (LIC) adaptors, following the recommendations given previously (81). LIC was performed in the vector pMCSG53 as described (81) to fuse the recombinant proteins with a N-terminal His_6 tag, with a tobacco etch virus protease cleavage site. Alternatively, LIC was performed in the vector pMCSG-GST or pMCSG69 to fuse the recombinant proteins with an N-terminal GST-His_6 tag or an N-terminal MBP-His_6 tag, respectively (see Table S1).

*Po*GH74cat-G476A, *Po*GH74cat-G476Q, and *Po*GH74cat-G476W were generated using the PCR-based QuikChange II site-directed mutagenesis kit (Agilent Technologies Inc., Santa Clara, CA) in accordance with the manufacturer's instructions and using pMCSG53::*Po*GH74cat as template (35). Similarly, *Cj*GH74-W353A and *Cj*GH74-W354A were generated using pET28a::CjGH74 as a template DNA (16). Primer sequences are provided in Table S2.

Gene expression and protein purification

Constructs were individually transformed into chemically competent *E. coli* BL21 DE3 cells. Colonies were grown on lysogeny broth solid medium supplemented with ampicillin

[4] Please note that the JBC is not responsible for the long-term archiving and maintenance of this site or any other third party hosted site.

(100 μg/ml). 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 at 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 $\times$ *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 $\times$ *g* for 45 min).

Recombinant proteins were purified from the cell-free extract with an Akta 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. Removal of the GST tag for *Ra*GH74a and the MBP tag for *Cl*GH74b and *Cb*GH74 was performed overnight at 4 °C using 1 mg of tobacco etch virus protease per 50 mg of recombinant protein. Untagged proteins were purified using a Ni^{2+} affinity column as described above. The final purification step was performed on a size-exclusion Superdex 200 column eluted with 50 mM sodium phosphate buffer, pH 7.0. Protein concentration was estimated using the Epoch Micro-Volume Spectrophotometer System (BioTek Inc., Winooski, VT) at 280 nm. Molar extinction coefficients used for protein concentration determination were 206,525 $M^{-1}\cdot cm^{-1}$ for *Cl*GH74a, 220,380 $M^{-1}\cdot cm^{-1}$ for *Cl*GH74b, 216,370 $M^{-1}\cdot cm^{-1}$ for *Cb*GH74, 159,085 $M^{-1}\cdot cm^{-1}$ for *Nk*GH74, 210,160 $M^{-1}\cdot cm^{-1}$ for *Pg*GH74, 217,150 $M^{-1}\cdot cm^{-1}$ for *Pb*GH74, 192,170 $M^{-1}\cdot cm^{-1}$ for *Pp*GH74, 190,680 $M^{-1}\cdot cm^{-1}$ for *Pj*GH74, 193,200 $M^{-1}\cdot cm^{-1}$ for *Pm*GH74, 178,955 $M^{-1}\cdot cm^{-1}$ for *Ra*GH74a, 182,380 $M^{-1}\cdot cm^{-1}$ for *Ra*GH74b, 241,100 $M^{-1}\cdot cm^{-1}$ for *Sm*AFV00434, 219,670 $M^{-1}\cdot cm^{-1}$ for *Sm*AFV00474, 208,670 $M^{-1}\cdot cm^{-1}$ for *Sr*GH74, 155,270 $M^{-1}\cdot cm^{-1}$ for *Sv*GH74a, 201,220 $M^{-1}\cdot cm^{-1}$ for *Sv*GH74b, and 181,405 $M^{-1}\cdot cm^{-1}$ for *Sat*GH74. Accurate protein molecular masses were confirmed by intact MS (83).

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). *p*NP-β-D-xylopyranoside and *p*NP-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.

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 *Cl*GH74a, *Cl*GH74b, and *Cb*GH74 or 20 °C for *Sv*GH74a) in a 50 mM concentration of their optimum buffer (citrate or phosphate at the optimum pH; see Fig. S1), using typically 0.1 μg/ml enzyme.

To determine the products released by recombinant GH74 enzymes, tamarind seed XyG was incubated at 37 °C (or 65 °C for *Cl*GH74a, *Cl*GH74b, and *Cb*GH74) 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 *Sv*GH74a). 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 *Sv*GH74a) and 0.1 mg/ml tamarind seed XyG. Limit digestion products of XXXGXXXG and of XXXG were obtained similarly after overnight incubation of 5 μM substrate with 1 μg/ml enzyme (or 10 μg/ml for *Sv*GH74a).

X-ray crystallography

The *Pg*GH74 and AFV00434 proteins were produced as selenomethionine-substituted derivatives using the standard M9 high-yield growth procedure according to the manufacturer's instructions (Shanghai Medicilon) and purified as described above. AFV00434 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: AFV00434 (SelMet), 25 mM zinc acetate, 20% (w/v) PEG 3350, 1 mM magnesium sulfate; AFV00434 (native), 25 mM zinc acetate, 20% (w/v) PEG 3350, 1.5% (w/v) 2-methyl-2,4-pentanediol; *Nk*GH74, 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; *Cl*GH74a, protein + XyGO mixture, 25% PEG 3350 (w/v), 0.1 M Tris pH 8.5; *Pg*GH74, 25% (w/v) PEG

3350, 0.2 M sodium chloride, 0.1 M sodium citrate, pH 5.6, 0.5% (w/v) glycerol; *Sr*GH74, protein + XyGO mixture, 1.6 M ammonium sulfate, 0.1 M sodium chloride, 0.1 M Hepes, 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 *Pg*GH74 SelMet, *Sr*GH74 native, and AFV00434 SelMet and native), beamline 08-ID at the Canadian Macromolecular Crystallography Facility, Canadian Light Source (Saskatoon, Saskatchewan, Canada) (for native *Nk*GH74), or on a Rigaku HF-007 home source with an R-AXIS IV detector (for native *Cl*GH74a). Data for *Pg*GH74 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 *Pg*GH74 was also solved using SAD phasing and Phenix.solve. The structures of *Nk*GH74, *Sr*GH74, and *Cl*GH74 were solved by Molecular Replacement and Phenix.phaser using models constructed by the Phyre2 server (87) onto *Po*GH74 (PDB code 6MGL), a putative xyloglucanase 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 xyloglucan 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. agarivorans* AFV00434, *C. lactoaceticus* 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. rapamycinicus* GH74 in complex with two XyG fragments (XLLG and XXXG), respectively. All X-ray crystallographic statistics are provided in Table S3.

Author contributions—G. A. performed sequence and phylogenetic analysis; cloned, produced, and purified GH74 catalytic modules and site-directed mutants; performed biochemical characterization; produced and purified XyG oligosaccharides XXXGXXXG and XXXG; generated figures; and co-wrote the manuscript. P. S. solved all crystal structures, produced structure figures and data, and co-wrote the manuscript. J. A. cloned, expressed, and purified GH74 catalytic modules and performed biochemical characterization under the supervision of G. A. M. A. produced *Cj*GH74 and site-directed mutants, investigated time-course hydrolysis of XyG by *Cj*GH74 and mutants, and generated Fig. S6. T. S. expressed and purified selenomethionine-substituted derivatives of *Pg*GH74 and AFV00434 and crystallized all proteins. A. H. V. assisted G. A. with phylogenetic analysis. B. H. curated protein sequences and assisted with sequence acquisition. A. S. directed structural studies. H. B. conceived the project, directed research, and revised the article with input from all authors.

Acknowledgments—We thank Nobuhiko Watanabe and Bogulaw 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 Foust, 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 xyloglucan 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. J., and Knox, J. P. (2018) Xyloglucan 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 xyloglucan: a polysaccharide with versatile application potential. *J. Mater. Chem.* **19,** 8528–8536 CrossRef
7. Zhou, Q., Rutland, M. W., Teeri, T. T., and Brumer, H. (2007) Xyloglucan in cellulose modification. *Cellulose* **14,** 625–641 CrossRef
8. Tuomivaara, S. T., Yaoi, K., O'Neill, M. A., and York, W. S. (2015) Generation and structural validation of a library of diverse xyloglucan-derived oligosaccharides, including an update on xyloglucan 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. G., and Brumer, H. (2018) *In vitro* and *in vivo* characterization of three *Cellvibrio japonicus* glycoside hydrolase family 5 members reveals potent xyloglucan 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 xyloglucan. *Curr. Opin. Struct. Biol.* **40,** 43–53 CrossRef Medline
12. Larsbrink, J., Rogers, T. E., Hemsworth, G. R., McKee, L. S., Tauzin, A. S., Spadiut, 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 xyloglucan metabolism in select human gut Bacteroidetes. *Nature* **506,** 498–502 CrossRef Medline
13. Lombard, V., Golaconda Ramulu, H., Drula, 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 *Aspergillus aculeatus*. *Agric. Biol. Chem.* **53,** 1411–1412 CrossRef
15. López-Mondéjar, R., Zühlke, D., Becher, D., Riedel, K., and Baldrian, 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. J., and Brumer, H. (2016) 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.* **283,** 1701–1719 CrossRef Medline
17. Bauer, S., Vasu, P., Mort, A. J., and Somerville, C. R. (2005) Cloning, expression, and characterization of an oligoxyloglucan reducing end-specific xyloglucanobiohydrolase from *Aspergillus nidulans*. *Carbohydr. Res.* **340,** 2590–2597 CrossRef Medline
18. Berezina, O. V., Herlet, J., Rykov, S. V., Kornberger, P., Zavyalov, A., Kozlov, D., Sakhibgaraeva, L., Krestyanova, I., Schwarz, W. H., Zverlov, V. V., Liebl, W., and Yarotsky, S. V. (2017) Thermostable multifunctional GH74 xyloglucanase from *Myceliophthora 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* endoglucanase Cel74 with and without a carbohydrate 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 Piens, K. (2007) An investigation of the substrate specificity of the xyloglucanase Cel74A 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 Hong, S. K. (2012) Identification and characterization of a xyloglucan-specific family 74 glycosyl hydrolase from *Streptomyces coelicolor* A3(2). *Appl. Environ. Microbiol.* **78,** 607–611 CrossRef Medline
23. Feng, T., Yan, K. P., Mikkelsen, M. D., Meyer, A. S., Schols, H. A., Westereng, B., and Mikkelsen, J. D. (2014) Characterisation of a novel endo-xyloglucanase (XcXGHA) from *Xanthomonas* that accommodates a xylosyl-substituted glucose at subsite −1. *Appl. Microbiol. Biotechnol.* **98,** 9667–9679 CrossRef Medline
24. Grishutin, S. G., Gusakov, A. V., Markov, A. V., Ustinov, B. B., Semenova, M. V., and Sinitsyn, A. P. (2004) Specific xyloglucanases as a new class of polysaccharide-degrading enzymes. *Biochim. Biophys. Acta* **1674,** 268–281 CrossRef Medline
25. Hasper, A. A., Dekkers, E., van Mil, M., van de Vondervoort, P. J., and de Graaff, L. H. (2002) EglC, a new endoglucanase from *Aspergillus niger* with major activity towards xyloglucan. *Appl. Environ. Microbiol.* **68,** 1556–1560 CrossRef Medline
26. Ichinose, H., Araki, Y., Michikawa, M., Harazono, K., Yaoi, K., Karita, S., and Kaneko, S. (2012) Characterization of an endo-processive-type xyloglucanase having a β-1,4-glucan-binding module and an endo-type xyloglucanase from *Streptomyces avermitilis*. *Appl. Environ. Microbiol.* **78,** 7939–7945 CrossRef Medline
27. Irwin, D. C., Cheng, M., Xiang, B., Rose, J. K., and Wilson, D. B. (2003) Cloning, expression and characterization of a family-74 xyloglucanase from *Thermobifida fusca*. *Eur. J. Biochem.* **270,** 3083–3091 CrossRef Medline
28. Ishida, T., Yaoi, K., Hiyoshi, A., Igarashi, K., and Samejima, M. (2007) Substrate recognition by glycoside hydrolase family 74 xyloglucanase from the basidiomycete *Phanerochaete chrysosporium*. *FEBS J.* **274,** 5727–5736 CrossRef Medline
29. Matsuzawa, T., Saito, Y., and Yaoi, K. (2014) Key amino acid residues for the endo-processive activity of GH74 xyloglucanase. *FEBS Lett.* **588,** 1731–1738 CrossRef Medline
30. Sianidis, G., Pozidis, C., Becker, F., Vrancken, K., Sjoeholm, C., Karamanou, S., Takamiya-Wik, M., van Mellaert, L., Schaefer, T., Anné, J., and Economou, A. (2006) Functional large-scale production of a novel *Jonesia* sp. xyloglucanase by heterologous secretion from *Streptomyces lividans*. *J. Biotechnol.* **121,** 498–507 CrossRef Medline
31. Yaoi, K., and Mitsuishi, Y. (2002) Purification, characterization, cloning, and expression of a novel xyloglucan-specific glycosidase, oligoxyloglucan reducing end-specific cellobiohydrolase. *J. Biol. Chem.* **277,** 48276–48281 CrossRef Medline
32. Yaoi, K., Nakai, T., Kameda, Y., Hiyoshi, A., and Mitsuishi, Y. (2005) Cloning and characterization of two xyloglucanases from *Paenibacillus* sp. strain KM21. *Appl. Environ. Microbiol.* **71,** 7670–7678 CrossRef Medline
33. Yaoi, K., and Mitsuishi, Y. (2004) Purification, characterization, cDNA cloning, and expression of a xyloglucan endoglucanase from *Geotrichum* sp. M128. *FEBS Lett.* **560,** 45–50 CrossRef Medline
34. Zverlov, V. V., Schantz, N., Schmitt-Kopplin, P., and Schwarz, W. H. (2005) Two new major subunits in the cellulosome of *Clostridium thermocellum*: xyloglucanase Xgh74A and endoxylanase Xyn10D. *Microbiology* **151,** 3395–3401 CrossRef Medline
35. Arnal, G., Stogios, P. J., Asohan, J., Skarina, T., Savchenko, A., and Brumer, H. (2018) Structural enzymology reveals the molecular basis of substrate regiospecificity and processivity of an exemplar bacterial glycoside hydrolase family 74. *Biochem. J.* **475,** 3963–3978 CrossRef Medline
36. Wegmann, U., Louis, P., Goesmann, A., Henrissat, B., Duncan, S. H., and Flint, H. J. (2014) Complete genome of a new Firmicutes species belonging to the dominant human colonic microbiota ("*Ruminococcus bicirculans*") reveals two chromosomes and a selective capacity to utilize plant glucans. *Environ. Microbiol.* **16,** 2879–2890 CrossRef Medline
37. Warnecke, F., Luginbühl, P., Ivanova, N., Ghassemian, M., Richardson, T. H., Stege, J. T., Cayouette, M., McHardy, A. C., Djordjevic, G., Aboushadi, N., Sorek, R., Tringe, S. G., Podar, M., Martin, H. G., Kunin, V., *et al.* (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. *Nature* **450,** 560–565 CrossRef Medline
38. Blumer-Schuette, S. E., Lewis, D. L., and Kelly, R. M. (2010) Phylogenetic, microbiological, and glycoside hydrolase diversities within the extremely thermophilic, plant biomass-degrading genus *Caldicellulosiruptor*. *Appl. Environ. Microbiol.* **76,** 8084–8092 CrossRef Medline
39. Lacerda Júnior, G. V., Noronha, M. F., de Sousa, S. T. P., Cabral, L., Domingos, D. F., Sáber, M. L., de Melo, I. S., and Oliveira, V. M. (2017) Potential of semiarid soil from Caatinga biome as a novel source for mining lignocellulose-degrading enzymes. *FEMS Microbiol. Ecol.* **93,** fiw248 CrossRef Medline
40. DeBoy, R. T., Mongodin, E. F., Fouts, D. E., Tailford, L. E., Khouri, H., Emerson, J. B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H. J., and Nelson, K. E. (2008) Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. *J. Bacteriol.* **190,** 5455–5463 CrossRef Medline
41. Matsuzawa, T., and Yaoi, K. (2016) GH74 xyloglucanases: structures and modes of activity. *Trends Glycosci. Glycotechnol.* **28,** E63–E70 CrossRef Medline
42. Jalak, J., Kurašin, M., Teugjas, H., and Väljamäe, P. (2012) Endo-exo synergism in cellulose hydrolysis revisited. *J. Biol. Chem.* **287,** 28802–28815 CrossRef Medline
43. Hamre, A. G., Lorentzen, S. B., Väljamäe, P., and Sørlie, M. (2014) Enzyme processivity changes with the extent of recalcitrant polysaccharide degradation. *FEBS Lett.* **588,** 4620–4624 CrossRef Medline
44. 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
45. 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 glycoside hydrolase family 18 processive chitinases on crystalline chitin. *Nat. Commun.* **5,** 3975 CrossRef Medline
46. 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 Medline
47. Christensen, S. J., Kari, J., Badino, S. F., Borch, K., and Westh, P. (2018) Rate-limiting step and substrate accessibility of cellobiohydrolase Cel6A from *Trichoderma reesei*. *FEBS J.* **285,** 4482–4493 CrossRef Medline
48. Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J. P. (1989) Cellulase families revealed by hydrophobic cluster analysis. *Gene* **81,** 83–95 CrossRef Medline

49. Yaoi, K., Kondo, H., Noro, N., Suzuki, M., Tsuda, S., and Mitsuishi, Y. (2004) Tandem repeat of a seven-bladed β-propeller domain in oligoxyloglucan reducing-end-specific cellobiohydrolase. *Structure* **12,** 1209–1217 CrossRef Medline
50. 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 *Clostridium thermocellum* xyloglucanase, XGH74A, reveal the structural basis for xyloglucan recognition and degradation. *J. Biol. Chem.* **281,** 24922–24933 CrossRef Medline
51. Alahuhta, M., Adney, W. S., Himmel, M. E., and Lunin, V. V. (2013) Structure of *Acidothermus cellulolyticus* family 74 glycoside hydrolase at 1.82 Å resolution. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **69,** 1335–1338 CrossRef Medline
52. Yaoi, K., Kondo, H., Hiyoshi, A., Noro, N., Sugimoto, H., Tsuda, S., Mitsuishi, Y., and Miyazaki, K. (2007) The structural basis for the exo-mode of action in GH74 oligoxyloglucan reducing end-specific cellobiohydrolase. *J. Mol. Biol.* **370,** 53–62 CrossRef Medline
53. Davies, G. J., and Sinnott, M. L. (2008) Sorting the diverse. *The Biochemist*, **30,** 26–32
54. Yaoi, K., Kondo, H., Hiyoshi, A., Noro, N., Sugimoto, H., Tsuda, S., and Miyazaki, K. (2009) The crystal structure of a xyloglucan-specific endo-β-1,4-glucanase from *Geotrichum* sp. M128 xyloglucanase reveals a key amino acid residue for substrate specificity. *FEBS J.* **276,** 5094–5100 CrossRef Medline
55. Kari, J., Olsen, J., Borch, K., Cruys-Bagger, N., Jensen, K., and Westh, P. (2014) Kinetics of cellobiohydrolase (Cel7A) variants with lowered substrate affinity. *J. Biol. Chem.* **289,** 32459–32468 CrossRef Medline
56. Breyer, W. A., and Matthews, B. W. (2001) A structural basis for processivity. *Protein Sci.* **10,** 1699–1711 CrossRef Medline
57. Asensio, J. L., Ardá, A., Cañada, F. J., and Jiménez-Barbero, J. (2013) Carbohydrate-aromatic interactions. *Acc. Chem. Res.* **46,** 946–954 CrossRef Medline
58. Eklöf, J. M., Ruda, M. C., and Brumer, H. (2012) Distinguishing xyloglucanase activity in endo-β-(1→4)glucanases. *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. Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M., and Henrissat, B. (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of α-amylase-related proteins. *Protein Eng. Des. Sel.* **19,** 555–562 CrossRef Medline
61. St. John, F. J., González, J. M., and Pozharski, E. (2010) Consolidation of glycosyl hydrolase family 30: a dual domain 4/7 hydrolase family consisting of two structurally distinct groups. *FEBS Lett.* **584,** 4435–4441 CrossRef Medline
62. Mewis, K., Lenfant, N., Lombard, V., and Henrissat, B. (2016) Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization. *Appl. Environ. Microbiol.* **82,** 1686–1692 CrossRef Medline
63. Pauly, M., Albersheim, P., Darvill, A., and York, W. S. (1999) Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. *Plant J.* **20,** 629–639 CrossRef Medline
64. Park, Y. B., and Cosgrove, D. J. (2015) Xyloglucan and its interactions with other components of the growing cell wall. *Plant Cell Physiol.* **56,** 180–194 CrossRef Medline
65. Zheng, Y., Wang, X., Chen, Y., Wagner, E., and Cosgrove, D. J. (2018) Xyloglucan in the primary cell wall: assessment by FESEM, selective enzyme digestions and nanogold affinity tags. *Plant J.* **93,** 211–226 CrossRef Medline
66. Grondin, J. M., Tamura, K., Déjean, G., Abbott, D. W., and Brumer, H. (2017) Polysaccharide utilization loci: fueling microbial communities. *J. Bacteriol.* **199,** e00860-16 CrossRef Medline
67. Zhao, Y., Chany, C. J., 2nd, Sims, P. F., and Sinnott, M. L. (1997) Definition of the substrate specificity of the "sensing" xylanase of *Streptomyces cyaneus* using xylooligosaccharide and cellooligosaccharide glycosides of 3,4-dinitrophenol. *J. Biotechnol.* **57,** 181–190 CrossRef Medline
68. 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
69. Rozeboom, H. J., Beldman, G., Schols, H. A., and Dijkstra, B. W. (2013) Crystal structure of endo-xylogalacturonan hydrolase from *Aspergillus tubingensis*. *FEBS J.* **280,** 6061–6069 CrossRef Medline
70. 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 *Reticulitermes speratus* is an endo-processive mannobiohydrolase: heterologous expression and characterization. *Biochem. Biophys. Res. Commun.* **452,** 520–525 CrossRef Medline
71. Ishida, T., Fushinobu, S., Kawai, R., Kitaoka, M., Igarashi, K., and Samejima, M. (2009) Crystal structure of glycoside hydrolase family 55 β-1,3-glucanase from the basidiomycete *Phanerochaete chrysosporium*. *J. Biol. Chem.* **284,** 10100–10109 CrossRef Medline
72. Moroz, O. V., Jensen, P. F., McDonald, S. P., McGregor, N., Blagova, E., Comamala, G., Segura, D. R., Anderson, L., Vasu, S. M., Rao, V. P., Giger, L., Sørensen, T. H., Monrad, R. N., Svendsen, A., Nielsen, J. E., *et al.* (2018) Structural dynamics and catalytic properties of a multimodular xanthanase. *ACS Catalysis* **8,** 6021–6034 CrossRef
73. Arnal, G., Bastien, G., Monties, N., Abot, A., Anton Leberre, V., Bozonnet, S., O'Donohue, M., and Dumon, C. (2015) Investigating the function of an arabinan utilization locus isolated from a termite gut community. *Appl. Environ. Microbiol.* **81,** 31–39 CrossRef Medline
74. Pages, S., Kester, H. C., Visser, J., and Benen, J. A. (2001) Changing a single amino acid residue switches processive and non-processive behavior of *Aspergillus niger* endopolygalacturonase I and II. *J. Biol. Chem.* **276,** 33652–33656 CrossRef Medline
75. Edgar, R. C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26,** 2460–2461 CrossRef Medline
76. Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S., and Madden, T. L. (2008) NCBI BLAST: a better web interface. *Nucleic Acids Res.* **36,** W5–W9 CrossRef Medline
77. Katoh, K., and Standley, D. M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30,** 772–780 CrossRef Medline
78. Waterhouse, A. M., Procter, J. B., Martin, D. M., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25,** 1189–1191 CrossRef Medline
79. Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30,** 1312–1313 CrossRef Medline
80. Miller, M. A. (2010) *Creating the CIPRES Science Gateway for inference of large phylogenetic trees*, pp. 1–8, Institute of Electrical and Electronics Engineers, Piscataway, NJ CrossRef
81. Eschenfeldt, W. H., Lucy, S., Millard, C. S., Joachimiak, A., and Mark, I. D. (2009) A family of LIC vectors for high-throughput cloning and purification of proteins. *Methods Mol. Biol* **498,** 105–115 CrossRef Medline
82. Studier, F. W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 CrossRef Medline
83. Sundqvist, G., Stenvall, M., Berglund, H., Ottosson, J., and Brumer, H. (2007) A general, robust method for the quality control of intact proteins using LC-ESI-MS. *J. Chromatogr. B Analyt. Technol. Biomed. Life Sci.* **852,** 188–194 CrossRef Medline
84. Arnal, G., Attia, M. A., Asohan, J., and Brumer, H. (2017) A low-volume, parallel copper-bicinchoninic acid (BCA) assay for glycoside hydrolases. *Methods Mol. Biol.* **1588,** 3–14 CrossRef Medline
85. Minor, W., Cymborowski, M., Otwinowski, Z., and Chruszcz, M. (2006) HKL-3000: the integration of data reduction and structure solution–from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* **62,** 859–866 CrossRef Medline
86. 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., Oeffner, 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

87. 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
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

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

Gregory Arnal, Peter J. Stogios, Jathavan Asohan, Mohamed A. Attia, Tatiana Skarina, Alexander Holm Viborg, Bernard Henrissat, Alexei Savchenko and Harry Brumer

J. Biol. Chem. 2019, 294:13233-13247.
doi: 10.1074/jbc.RA119.009861 originally published online July 19, 2019