Lawrence Berkeley National Laboratory
LBL Publications

Title
A structural and kinetic survey of GH5_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis

Authors

Glasgow, Evan M
Kemna, Elias I
Bingman, Craig A
et al.

A structural and kinetic survey of GH5_4 endoglucanases reveals determinants of broad substrate specificity and opportunities for biomass hydrolysis

Evan M. Glasgow[1,2], Elias I. Kemna[1], Craig A. Bingman[1,2], Nicole Ing[3,4], Kai Deng[3,4], Christopher M. Bianchetti[1,2], Taichi E. Takasuka[5], Trent R. Northen[3,6], and Brian G. Fox[1,2,*]

From the [1]Department of Biochemistry, University of Wisconsin, Madison, Wisconsin, USA, the [2]Great Lakes Bioenergy Research Center, Madison, Wisconsin, USA, the [3]Joint BioEnergy Institute, Emeryville, California, USA, [4]Sandia National Laboratories, Livermore, California, USA, [5]Hokkaido University, Sapporo, Japan, and [6]Lawrence Berkeley National Laboratory, Berkeley, California, USA

Edited by Gerald W. Hart

Broad-specificity glycoside hydrolases (GHs) contribute to plant biomass hydrolysis by degrading a diverse range of polysaccharides, making them useful catalysts for renewable energy and biocommodity production. Discovery of new GHs with improved kinetic parameters or more tolerant substrate-binding sites could increase the efficiency of renewable bioenergy production even further. GH5 has over 50 subfamilies exhibiting selectivities for reaction with β-(1,4)–linked oligo- and polysaccharides. Among these, subfamily 4 (GH5_4) contains numerous broad-selectivity endoglucanases that hydrolyze cellulose, xyloglucan, and mixed-linkage glucans. We previously surveyed the whole subfamily and found over 100 new broad-specificity endoglucanases, although the structural origins of broad specificity remained unclear. A mechanistic understanding of GH5_4 substrate specificity would help inform the best protein design strategies and the most appropriate industrial application of broad-specificity endoglucanases. Here we report structures of 10 new GH5_4 enzymes from cellulolytic microbes and characterize their substrate selectivity using normalized reducing sugar assays and MS. We found that GH5_4 enzymes have the highest catalytic efficiency for hydrolysis of xyloglucan, glucomannan, and soluble β-glucans, with opportunistic secondary reactions on cellulose, mannan, and xylan. The positions of key aromatic residues determine the overall reaction rate and breadth of substrate tolerance, and they contribute to differences in oligosaccharide cleavage patterns. Our new composite model identifies several critical structural features that confer broad specificity and may be readily engineered into existing industrial enzymes. We demonstrate that GH5_4 endoglucanases can have broad specificity without sacrificing high activity, making them a valuable addition to the biomass deconstruction toolset.

Sustainable, biological solutions to the growing climate and energy crises have been the subject of increasing interest in the last 2 decades. In contrast to biofuel production from edible polysaccharides, such as starch, recent efforts have focused on next-generation bioenergy, with higher-energy fuels derived from inedible lignocellulosic biomass found in a variety of abundant plant materials.

Enzyme hydrolysis of biomass releases sugars, which can be converted into a growing range of fuels and commodities. Enzymes, however, are a major operational expense in the cellulosic bioenergy process (1), owing to the difficulty of hydrolyzing cellulose relative to starch, prompting search and design efforts for more effective enzyme mixtures. Broad-specificity glycoside hydrolases (GHs) may replace many specialized enzymes with fewer, more flexible catalysts, increasing sugar yield and reducing enzyme variability between feedstocks (2). One such family rich in broad-specificity cellulases is GH5.

GH5 members adopt the $(\alpha/\beta)_8$-barrel fold and display high structural similarity, despite having low (sometimes <10%) sequence identity (3–7). Variability of function is thought to occur through differences in the loop regions connecting the β-strands to the α-helices, particularly at the C-terminal side of the barrel where the active site is built (8, 9). Functionally relevant structural differences also occur at locations distal to the active site (9–12), where loops may contain disulfide bonds (12, 13), form short β-sheets or helices (10–12), or even encompass auxiliary domains (14).

Because GH5 is one of the largest and most catalytically diverse GH families, a classification scheme was devised that separates the family into over 50 subfamilies based on global analysis of sequences, biochemical data, and structures (7). Subfamily 4 (GH5_4) was noted for being particularly enriched in broad-specificity β-(1,4)-endoglucanases (enzymes that cleave in the middle of long glucan chains, like cellulose).

As of writing, the Protein Data Bank (PDB) contains 42 entries for 16 GH5_4 enzymes (Table S1), also summarized in the Carbohydrate-Active Enzyme (CAZy) database (15), making GH5_4 the second-best structurally characterized subfamily after GH5_2. Many efforts have been directed toward understanding how the sequence and structure of a particular GH5_4 enzyme dictate its substrate selectivity (9, 10, 12, 13, 16–19).

For example, crystal structures with linear (12, 16, 18–20) or branched (9–11, 16, 17) oligosaccharides bound to the positive (16, 19) or negative (10, 12, 16–20) subsites (positive being oriented toward the reducing end of the chain) or spanning the entire cleft (9, 11) show how sugar chains are coordinated in

17752 *J. Biol. Chem.* (2020) 295(51) 17752–17769

Published in the U.S.A.

the active site. Some GH5_4 enzymes prefer highly branched xyloglucan (XG), whereas others hydrolyze only linear polysaccharides (see Fig. S1 for polysaccharide structures), and this has been rationalized by differences in the width and depth of the binding cleft, as well as side pockets to accommodate branched sugars or enforce their positioning (11, 12, 17, 18, 20). The ability of some enzymes (6, 7, 11, 18) to hydrolyze both β-(1,4) and β-(1,3) linkages with similar efficiency occurs when a β-(1,4) linkage of the heterogeneous polysaccharide can be placed between the −2 and −1 subsites. Additional studies with oligosaccharides have revealed subsite-by-subsite binding energies and rate constants for hydrolysis of simple oligosaccharides (9–12, 16–20).

Despite the numerous structural models available, prediction of selectivity by sequence and structure homology remains a significant challenge (6, 21, 22). This challenge is exacerbated by inconsistent assay methodologies and end-point activity measurements that offer limited kinetic insight. Understanding subtle differences in structure that lead to variations in substrate selectivity and catalysis requires examination of closely related proteins, consistency in experimental methods, and a combination of experimental approaches.

We recently carried out a normalized catalytic analysis of over 240 members of the GH5_4 subfamily and found that 55% are active on lichenan, mannan, and xylan, and many have high activity on XG (6). Other researchers have shown that some also hydrolyze additional mixed-linkage glucans, like barley β-glucan (BBG), as well as glucomannan (GM) (8, 9, 11, 18, 23–27) (Table S1). These analyses helped guide the selection of enzymes for additional catalytic studies and efforts to solve crystal structures.

In this work, we report the crystal structures of 10 new enzymes that sample the GH5_4 phylogenetic space from two of its three distinct clades. In combination with normalized catalytic studies, this work provides a fuller picture of the evolutionary changes and structural features contributing to the breadth of substrate selectivity and hydrolysis reactions observed in GH5_4.

Results

Structural features of GH5_4 active-site clefts

We sought to structurally characterize enzymes from each of the three major GH5_4 clades, and crystallization trials were conducted for ~20 enzymes. Of these, 10 structures were successfully solved: two from clade 2 and eight from clade 3. Enzymes from clade 1 tended to be more difficult to purify and crystallize due to poor solubility, which may be related to their intimate dependence on CBM46 modules (6, 18), which were removed prior to expression.

All 10 structures adopt the expected $(\alpha/\beta)_8$, or TIM-barrel, fold (Fig. 1A), with the active-site Glu residues positioned at the C-terminal ends of strand 4 and strand 7. The ~5-Å distance between the side chains of these residues is consistent with the retaining mechanism of glycoside hydrolases in GH5 (28). It should be noted that the full polypeptide sequences of some of these proteins include carbohydrate-binding modules, dockerin domains, and additional enzyme domains (Fig. 1C), but the

experimental design for this study included only the GH5_4 enzyme domains.

Four of the enzymes originate from cellulose-degrading environmental bacteria. 4IM4 is derived from *Hungateiclostridium thermocellum*, a thermophilic anaerobe isolated from decaying cotton bales (29). Other source organisms (*Hungateiclostridium cellulolyticum* and *Clostridium cellulovorans*) are found on decaying grass and wood (30, 31), whereas another (*Clostridium acetobutylicum*) is found in soil and water (32). The remaining six source organisms dwell in the digestive tracts of herbivorous mammals (or chickens, for *Bacteroides salanitronis*), playing critical roles in aiding digestion of plant feedstocks (33–38).

Table 1 briefly summarizes data collection and refinement statistics for the 10 structures reported here. A fuller set of data are provided in Table S2. As a set, the structures have similar resolution, completeness, and *R*-values. The structures are distributed across the GH5_4 phylogenetic tree (Fig. 1D), with new structures added to subclades 2A and 3F and additional structures added to subclades 2B, 3B, and 3G (6). There is ~30–70% sequence identity in this set of structures.

After considerable effort and limited success in obtaining structures with bound oligosaccharides (Table S2, PDB 6WQV with cellotriose bound in the negative subsites), we relied on alignment of structures previously solved by Dos Santos *et al.* (16) (PDB entry 4W88) and Gloster *et al.* (17) (PDB entry 2JEQ) to model the binding of XG fragments. Of particular utility, 4W88 provides a model for a XG variant having a backbone of three glucose with branches of one xylose and one xylose-galactose in the positive subsites, whereas the 2JEQ structure provides a model having a backbone of four glucose with branches of one xylose- and one xylose-galactose in the negative subsites. Alignment of our new structures with these previous exemplars gave average root mean square deviation for alignments of the protein backbone of <1 Å and essentially no steric clashes between aligned enzymes and modeled polysaccharides. This lends great confidence to the predictive modeling of XG fragments into the new GH5_4 active sites.

Six features of the binding clefts were explored for contributions to substrate selectivity: 1) the dimensions of the clamp surrounding the hydrolysis site, dictated chiefly by loop lengths; the position of residues that interact with 2) negative-subsite and 3) positive-subsite sugars; 4) interactions with branch sugars; and the presence of protein-xylose interactions 5) near the +1 subsite and 6) near the −1 subsite.

Clamp-forming loops—The loops connecting the C termini of the interior β-strands to the N termini of the subsequent helices form the walls of the substrate-binding cleft and provide contact surfaces for substrate interaction. Several of these loops are varied in length, residue composition, and the presence of secondary structure elements (Table 2). Changes in these loops are the primary determinants of the diversity of active-site shapes found within GH5_4.

The cross-sections of the 10 enzymes shown in Fig. 2 highlight the substrate-binding cleft, which is bounded by walls built mainly from loops 4 and 6. These loops form a U-shaped clamp into which the glycan chain settles for hydrolysis. Specifically, loops 4 and 6 contribute to the "left-hand" and "right-

SASBMB

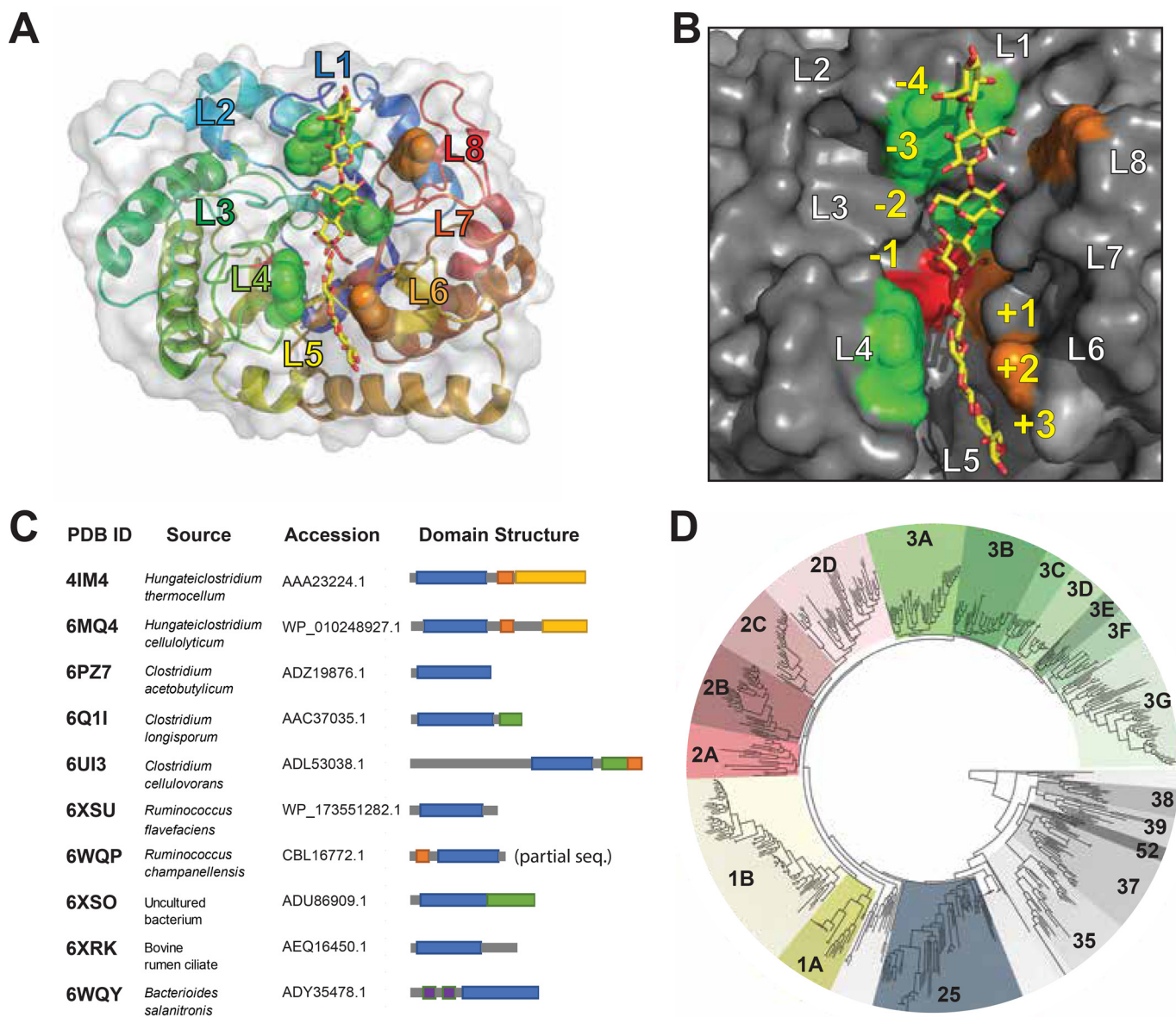*J. Biol. Chem.* (2020) 295(51) 17752–17769 **17753**

Figure 1. Structural, biological, and phylogenetic overview of GH5_4 enzymes. *A*, tertiary structure and the location of polysaccharide binding modeled from glucan chain of XG fragments in the 2JEQ (17) and 4W88 (16) structures. Aromatic residues in the binding cleft are displayed as *spheres*. Loops are *labeled* and *color-coded*. *B*, close-up of the binding cleft showing surface interactions with substrate chain: Trp (*green*), Tyr (*orange*), and catalytic Glu residues (*red*). The *numbering* of the sugar-binding subsites is indicated. *C*, source organism, accession code, and native domain structure of the 10 enzymes in this study. Domain annotations were obtained from NCBI, UniProt, and CAZy. *Blue*, GH5_4 domain; *green*, CBM; *yellow*, carbohydrate esterase; *orange*, dockerin; *purple*, BACON domain. *D*, phylogenetic tree of clades and subclades of GH5_4 (*yellow*, *red*, and *green wedges*) and closely related subfamilies (*numbered gray wedges*).

hand" walls of the clamp, respectively, and largely determine its width. Aromatic residues in these loops interact with the polysaccharide chain (Trp (*green*) and Tyr (*orange*) in Fig. 2). Fig. 2 also shows the location of the Glu residues that serve as the nucleophile (*red*) and proton donor (*blue*) for the retaining reaction.

Seven of the eight clade 3 proteins (4IM4, 6MQ4, 6PZ7, 6Q1I, 6UI3, 6XSU, and 6WQP) possess a deep cleft with mostly vertical clamp walls of constant width. Among these seven enzymes, loop 4 and loop 6 show up to 78% identity (Table S3). Loop 4 contains 18 residues (except for 14 residues in 6UI3), loop 6 contains 18–21 residues, and similar short secondary

structural elements are found in each (Table 2). Fig. 3 (*A and B*) shows the surface representations of these loops and the negative and positive subsites for 4IM4, which is representative of the active-site clefts of the seven enzymes mentioned above. As described below, different members from this group have the highest catalytic efficiencies for each of the polysaccharides tested (Table 3).

Although 6XSO is in clade 3, its loop 4 and loop 6 are significantly longer and shorter, respectively, than the other clade 3 proteins. Its clamp is incomplete because it lacks an aromatic residue in loop 6 to interact with the substrate (Figs. 2 and 3 (*C and D*)). 6XSO has middle to low activity with all substrates (Table 3).

Table 1
Data collection and refinement statistics

PDB identifier	4IM4	6MQ4	6PZ7	6Q1I	6UI3	6SXU	6XSO	6WQP	6XRK	6WQY	6WQV
GH5_4 subclade	3B	3B	3B	3B	3B	3F	3G	3G	2A	2B	3G
Space group	P 1 2$_1$ 1	P 3$_1$	P 2$_1$2$_1$2$_1$	P 1	P 2$_1$2$_1$2$_1$	P 1 2$_1$ 1	P 2$_1$2$_1$2$_1$	P 1 2$_1$ 1	P 2 2$_1$2$_1$	P 2$_1$2$_1$2$_1$	P 1
Data collection											
Unique reflections	92,255	63,716	74,845	128,382	85,209	63,716	140,736	94,298	141,220	165,260	224235
Completeness (%)	99.5	99.98	96.54	94.83	95.68	99.99	99.96	97.26	99.97	99.19	93.20
Mean I/s	3.6	22.1	19.11	20.86	29.24	22.1	19.45	8.31	18.36	22.62	17.00
Wilson B-factor	14.6576	11.56	9.83	10.73	10.14		21.76	17.19	20.02	11.23	15.41
R merge	0.193	0.05344	0.0649	0.04018	0.04056	0.05344	0.07386	0.1073	0.07724	0.05084	0.0463
Refinement											
Reflections	92,137	63,714	74,834	128,302	85,197	63,716	140,701	94,067	141,185	165,244	224186
Resolution (Å)	2.42	1.4	1.25	1.35	1.3	1.4	1.5	1.6	1.42	1.05	1.45
R-Value work	0.209	0.1216	0.1254	0.1116	0.1263	0.1209	0.1458	0.2029	0.1447	0.1389	0.1511
R-Value free	0.264	0.1379	0.1487	0.1318	0.1412	0.1377	0.1664	0.2422	0.1674	0.1430	0.1681

Table 2
Loop lengths (in amino acids) and intraloop secondary structure elements of GH5_4 enzymes
Secondary structure elements were determined by DSSP (80, 81). a, α-helix; b, β-ladder; c, β-turn; d, 3$_{10}$ helix.

| Enzyme	L1	L2	L3	L4	L5	L6	L7	L8
4IM4	18c	17cd	15c	18c	8d	19ac	7	24bc
6MQ4	26cd	17c	12c	18c	7d	18ac	7	24bc
6PZ7	19c	17cd	11c	18c	7d	18ac	7	23bc
6Q1I	30ac	15cd	11c	18c	7d	21ac	7	24c
6UI3	19c	16c	14d	14c	7d	20ac	7	23c
6XSU	26bcd	15cd	16cd	18c	7d	20ac	7	28bcd
6XSO	31ac	15cd	19c	24cd	16cd	8c	10c	16cd
6WQP	30ac	17cd	11c	18c	7d	18ac	7	23bc
6XRK	30ac	15cd	18cd	14bc	5d	35abc	9	23bc
6WQY	23bc	17c	20acd	20cd	5d	32acd	7	24c
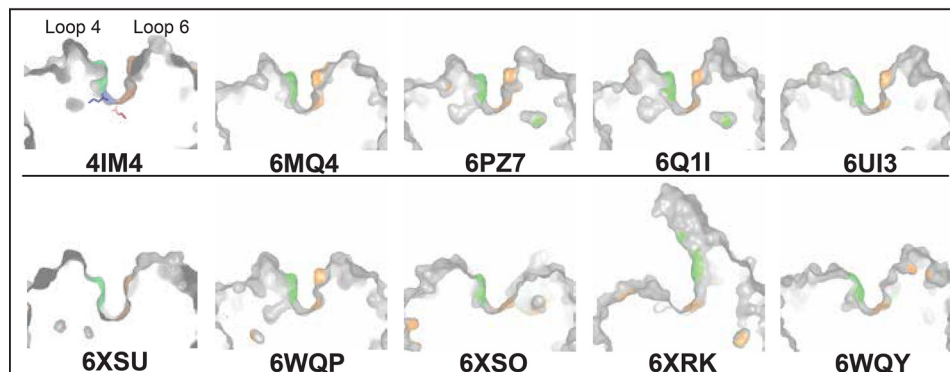
Figure 2. Cross-sections of GH5_4 enzymes showing the active-site cleft at the approximate position of the active site. The view is from the positive sugar-binding subsites toward the negative subsites. The variety of cleft shapes owes mainly to length differences in loops 4 (*left*) and 6 (*right*). The locations of the conserved catalytic Glu nucleophile (*red*) and Glu proton donor (*blue*) are shown for 4IM4, and the locations of aromatic residues (Trp in *green* and Tyr in *orange*) are highlighted in each structure.

The two structures from clade 2 feature greater loop diversity than clade 3. In 6XRK and 6WQY, loop 6 is substantially longer than the average of 21 residues, at 35 and 32 residues, respectively. Both structures contain an insertion in loop 6, but at different relative locations in the loop. The insertion occurs before Trp-254 in 6XRK and after Trp-269 in 6WQY (position of the aligned Trp residues shown by *red arrow* in Fig. S3*A*). In 6WQY, the loop folds back onto itself and makes a stabilizing contact with helix 7.

The two clade 2 enzymes appear to lack complete clamp motifs, as they are missing either one or both of the aromatic residues in the walls of the cleft. This contributes to their overall more open active-site shape (Fig. 2). 6XRK lacks aromatic residues on the left wall provided by loop 4 (Fig. 3*E*), whereas the two Trp residues mentioned above provided by loop 6 extend upward from the cleft and could feasibly provide distal contacts with a bent polysaccharide chain (Fig. 3*F*).

Furthermore, loop 6 has a unique two-strand antiparallel β-sheet that extends outward an additional 20–25 Å from the globular protein core in 6XRK (Fig. 2). This configuration is supported by crystal packing, as the extended loop makes contacts with its counterpart in the asymmetric unit. However, in the monomer expected in solution, loop 6 in 6XRK has two solvent-exposed Trp residues (Fig. 3*F*, *green*) that may potentially contact positive-subsite sugars and so act as a rudimentary CBM and contribute to reactions with XG. Elastic network
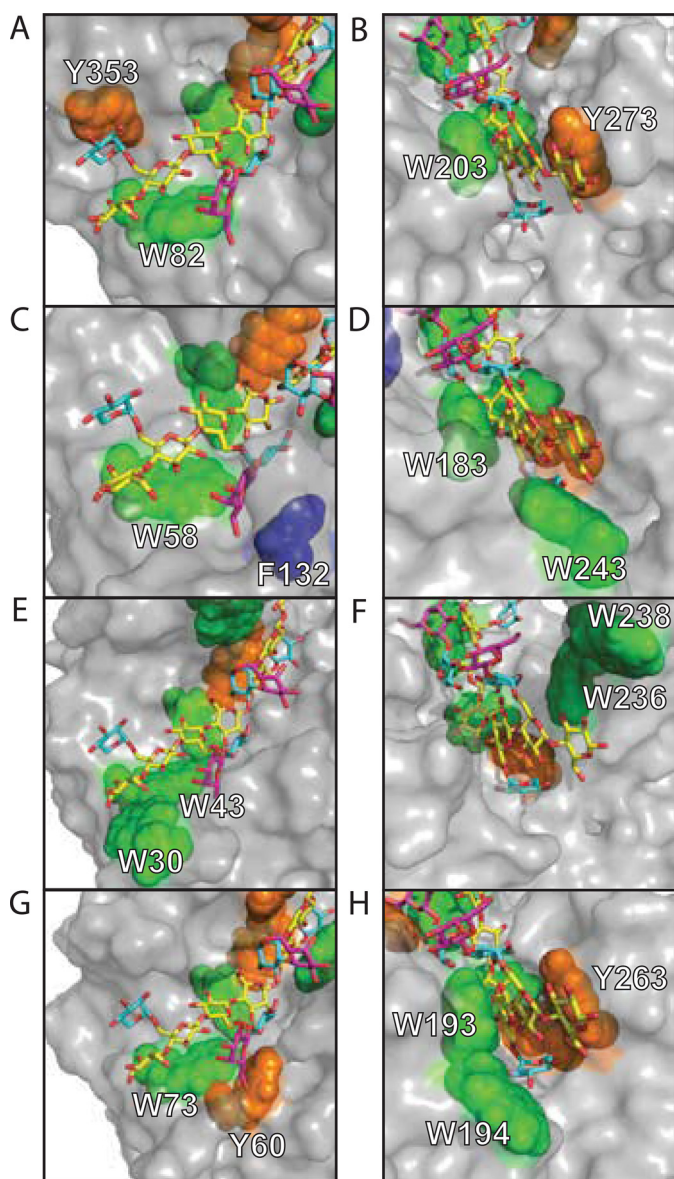
Figure 3. Aromatic contacts for XG binding modeled by alignment with PDB 4W88 (16) and 2JEQ (17). *Left panels*, negative subsites; *right panels*, positive subsites. *A* and *B*, 4IM4; *C* and *D*, 6XSO; *E* and *F*, 6XRK; *G* and *H*, 6XSU. The sugar colors are as follows: glucose (*yellow*), xylose (*cyan*), and galactose (*magenta*). The aromatic residue colors are as follows: Trp (*green*), Tyr (*orange*), and Phe (*blue*).

modeling using the *elNémo* server (39) revealed several possible low-frequency normal modes, which are springlike motions resulting from protein flexibility. The modeling predicted large dynamic motions of the extended loop 6 in 6XRK that might sample different binding interactions with polysaccharides. Similar motions were not predicted for the other clade 2 enzyme, 6WQY, as its own elongated loop 6 was folded back against the protein and stabilized by numerous interresidue contacts.

Interactions with negative-subsite sugars—Loop 1 varies from 18 to 31 residues in length and determines the overall size of the "loading platform" on each enzyme. This is a mostly flat surface, starting at the edge of the protein containing all of the negative subsites and funneling toward the active site of the substrate-binding cleft (see Fig. 1).

Table 3
Catalytic efficiencies (k_{cat}/K_m) determined for GH5_4 reactions with soluble and partially soluble substrates

Mannan and PASC are not included because efficiencies could not be determined on these insoluble substrates. k_{cat}/K_m (liters·g^{-1}·min^{-1}). Errors are the S.D. of three experimental replicates. XG, xyloglucan; GM, glucomannan; L, lichenan; X, xylan; ND, not determined. Individual k_{cat} and K_m parameters and their pairwise p values correlating reactions of an enzyme pairs of substrates are presented in Tables S4 and S5.

	XG	GM	L	X
4IM4	140 ± 7.7	300 ± 22	160 ± 15	3.2 ± 0.14
6MQ4	55 ± 2.4	240 ± 21	140 ± 20	23 ± 1.0
6PZ7	110 ± 5.0	170 ± 13	140 ± 4.7	16 ± 0.76
6Q1I	140 ± 5.2	310 ± 21	170 ± 14	6.3 ± 0.18
6UI3	120 ± 18	310 ± 12	ND	ND
6XSU	93 ± 4.9	340 ± 36	170 ± 6.6	12 ± 0.43
6WQP	130 ± 5.8	250 ± 26	170 ± 3.9	8.2 ± 0.59
6XSO	120 ± 3.2	110 ± 3.1	87 ± 13	2.3 ± 0.061
6XRK	140 ± 14	69 ± 3.8	ND	ND
6WQY	110 ± 42.6	49 ± 4.8	ND	ND

The shortest loop 1 belongs to 4IM4, and correspondingly, so does the smallest platform (Fig. 3*A*). The four enzymes that have the longest loop 1 (Table 2, 30 or 31 residues) are 6XSO (Fig. 3*C*), 6XRK (Fig. 3*E*), 6Q1I, and 6WQP. Their loading platforms extend out ~10 Å farther than that of 4IM4. These and the other enzymes, such as 6XSU (Fig. 3*G*), with loop 1 greater than 25 residues contain a small α-helical segment (~1.5 turn) that sits just below the platform-extending loop, in a shelf-and-bracket motif, providing a stable, compacted support for the extended surface.

Interactions with positive-subsite sugars—Loops 4 and 6 contribute additional aromatic residues—beyond those that define the clamp—that interact with sugar moieties in different ways, from +1 out to +3, +4, or potentially +5 subsites in different enzymes. 6XSU has an additional Trp after the critical Trp (Trp-193) in the +1 site, which doubles the length of the aromatic surface of the left wall (Fig. 3*H*). Access to this putative stabilizing protein-sugar interaction would only be available to a polysaccharide chain that maintained a linear configuration as it exited the substrate-binding cleft. Other examples of linear-only positive-subsite interactions include a Tyr on 6MQ4 and a Trp on 6WQP, near the edge of the protein that might conceivably interact with a strictly linear polysaccharide chain at +5. 6WQY also possesses an analogous extra Trp, but a Gly is inserted between the two Trp residues, and so the extended surface is altered; 6WQY has lower catalytic efficiency compared with 6XSU (Table 3).

Branch sugar interactions—Loop 3 is adjacent to the left wall–forming loop 4 and is poised to present side chains that can directly interact with sugars preceding the hydrolysis site. In other GH5_4 structures (*e.g.* 4V2X (18), 3ZMR (40), 5OYC (10), and 4W8A (16, 41)), an aromatic residue in loop 3 stacks against the primary xylose branch at −2; none of the structures presented here has this binding motif.

6XSO possesses a Lys-Asp dipeptide in loop 3 that could form a hydrogen bond network with two galactose branches of XG (Fig. 4). Phe-132 is sheltered beneath the β-turn in loop 3 and is optimally oriented to stack against a secondary galactose branch at the −2 position (Fig. 3*C*, *blue*).

Loop 4 provides additional favorable interactions with branch sugar hydroxyl groups. For seven of the structures
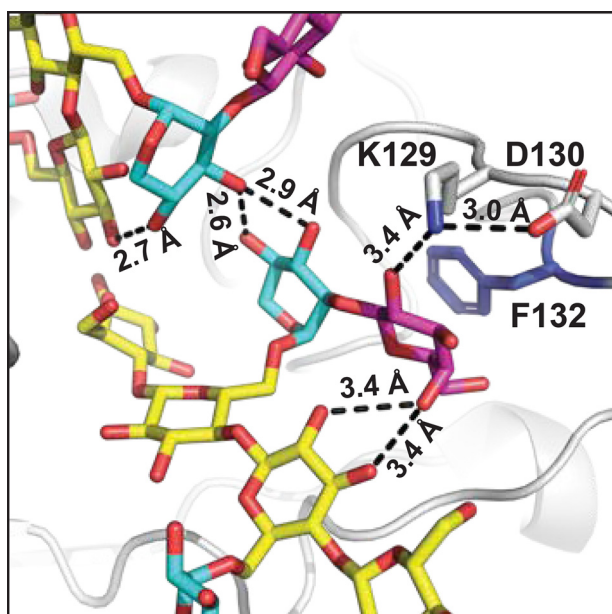
Figure 4. Interactions of loop 3 in 6XSO predicted from the XG ligands provided by alignment with 2JEQ. Potential polar contacts between side chains of Lys-129 and Asp-130 and to galactose branch in XG are shown. A network of possible intersugar hydrogen bonds is also indicated. The sugar colors are as follows: glucose (*yellow*), xylose (*cyan*), and galactose (*magenta*). Distances are reported in Å.

presented here (all except 6XSO, 6XRK, and 6WQY), the position immediately preceding the clamp-forming Trp in loop 4 (*e.g.* Trp-203 in 4IM4) is occupied by either Asp or Glu, whose side chain comes within hydrogen-bonding distance of the modeled galactose branch off the +2 position.

Loop 8 also provides additional potential interactions with branch sugars. 4IM4, 6MQ4, and 6Q1I possess a Tyr (*e.g.* Tyr-353 in 4IM4; Fig. 3*A*) oriented parallel to the modeled xylose branch at the −3 position, suggesting a close planar aromatic-carbohydrate interaction. This interaction appears to be specifically optimized to the geometry of XG. Other structures have small, hydrophobic side chains or Ser or Thr in this position and so offer different possibilities for interaction with a branch sugar. However, 6XSO has a diminished loop 8, and no protein residues appear to be within contact distance of the xylose branch at −3 (Fig. 3*C*).

Xylose branch interactions at the +1 subsite—A xylose branch at +1 rests at the floor of the cavity and helps orient the +1 glucose, which becomes the leaving group after nucleophilic attack. The glucan chain in the modeled XG is deflected sharply away from the cleft axis after the site of chain scission, with the deeper placement of the xylose branch excluding the glucan backbone from the active-site cleft (see Fig. 3, *B, D, F*, and *H*).

In addition to creating intrapolysaccharide hydrogen bonds with the glucan chain, the +1 xylose branch interacts closely with polar residues from loop 5 and the bottom of loop 6 indicated by the position of the modeled xylose. The only enzyme with no predicted polar contacts with this sugar is 4IM4, although a similar pocket shape and other interactions are present and likely position the +1 xylose branch and glucan chain exiting the active-site cleft.

6MQ4, 6PZ7, 6Q1I, and 6WQP contain a Ser in loop 5 that potentially forms a hydrogen bond with OH-2 or OH-3 of the branch xylose, whereas 6XSU and 6UI3 have a Ser-*X*-Asp triad with greater potential for polar contacts with these sugar hydroxyl groups. 6WQY has a Thr-Asn motif oriented favorably to form polar contacts, as does 6XRK. In addition, 6WQY possesses an Asp at the bottom of loop 6 that may coordinate with OH-4 on the xylose. 6XSO, despite having a loop 5 approximately twice the length of the other enzymes, offers only one potential polar contact, a suboptimally oriented Asn at position 238.

In summary, no favorable aromatic contacts are available to the +1 xylose, but several polar contacts might occur. In any case, all 10 GH5_4 structures show a distinct concave region that could accommodate a +1 xylose, and all 10 enzymes are able to hydrolyze XG (see below).

Xylose branch interactions at the −1 subsite—Interactions with the α-(1,6)-xylose branch sugar on the glucose in the −1 subsite are hypothesized to play key roles in determining catalytic selectivity (16). Many GH74 *exo*-XGases feature a pocket where an aromatic residue contacts this xylose, allowing cleavage of the substrate at branched sugars, but this feature is rarer in the GH74 *endo*-XGases (42). In GH5, such a pocket is also rare, and none of the 10 present GH5_4 structures has either a Trp or Tyr in the correct position to interact with a −1 xylose branch. 6XRK and 6WQY, the structures from clade 2, possess the most open clefts and thus potentially offer the highest tolerance to a branch xylose at the −1 subsite.

Catalytic efficiency with soluble polysaccharides

The high solubility of XG and GM allowed determination of k_{cat} and apparent K_m (Table S4), and the semisoluble characteristic of lichenan and xylan allowed these determinations for a subset of enzymes. Experimental errors were ~1–20% for k_{cat} and ~3–30% for K_m. The high viscosity of XG and GM solutions and the heterogeneity of lichenan and xylan suspensions were likely sources of experimental error for these substrates.

In summary, kinetic analysis revealed that 1) the enzymes clustered into several groups based on catalytic efficiency for different substrates and 2) despite the an apparent convergence in k_{cat}/K_m, individual k_{cat} and K_m values varied widely.

Three groups of XGase catalytic efficiency emerged (Table 3). 4IM4, 6XRK, 6Q1I, 6WQP, 6UI3, and 6XSO had a group average k_{cat}/K_m of ~130 liters·min^{-1}·g^{-1}. None of the enzymes within this group had statistically significantly different catalytic efficiency values, yet the individual k_{cat} and K_m values varied widely. 6PZ7 and 6WQY represented another grouping of average catalytic efficiencies (~110 liters·min^{-1}·g^{-1}), whereas 6XSU and 6MQ4 had distinctly lower efficiencies.

The enzymes separated into two groups for GMase catalytic efficiency (Table 3), with 4IM4, 6Q1I, 6UI3, and 6XSU averaging ~320 liters·min^{-1}·g^{-1} and 6MQ4 and 6WQP averaging ~250 liters·min^{-1}·g^{-1}. The remaining enzymes had lower, distributed catalytic efficiencies with GM and were not grouped.

All of the lichenan-saturable enzymes had similar catalytic efficiencies, except for 6XSO (Table 3), which was about half as efficient (k_{cat}/K_m of 87 compared with the group average of 157

ASBMB

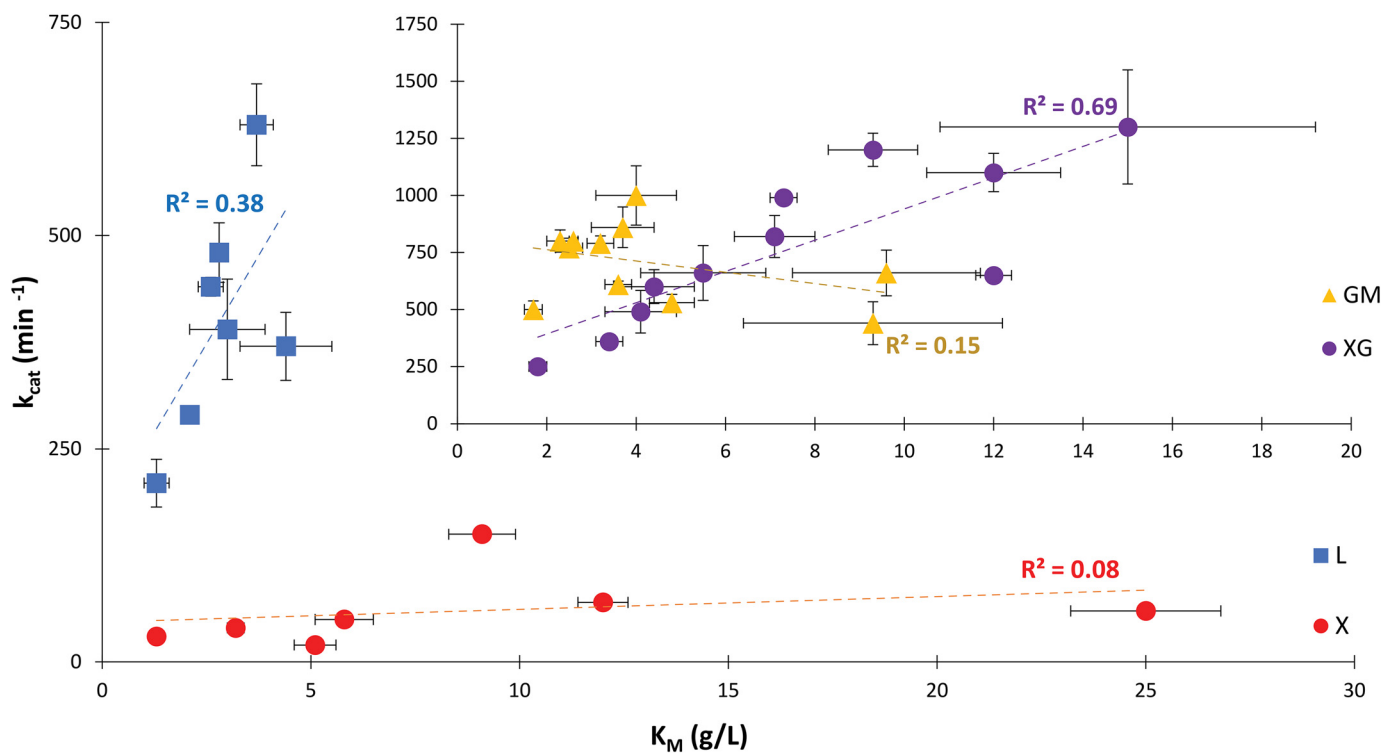*J. Biol. Chem.* (2020) 295(51) 17752–17769 **17757**

Figure 5. Catalytic efficiency shown as a plot of turnover numbers (k_{cat}) and Michaelis constants (K_m) for reaction of GH5_4 with GM (*yellow triangles*), XG (*purple circles*), lichenan (*L, blue squares*), and xylan (*X, red circles*). The slope of the imaginary line drawn from the origin out to each point yields the catalytic efficiency. *Error bars*, one S.D. of three simultaneous experimental replicates. *Dashed lines*, linear correlation functions for each substrate, with R^2 indicated.

liters·min^{-1}·g^{-1}). No grouping occurred for the seven enzymes saturable on beechwood xylan, as each enzyme had a statistically unique catalytic efficiency, ranging from 6MQ4 (23 liters·min^{-1}·g^{-1}) to 6XSO (2 liters·min^{-1}·g^{-1}).

Strikingly, k_{cat} and K_m for XG show a roughly linear correlation (Fig. 5; R^2 = 0.69), indicating a tendency for the catalytic efficiencies in GH5_4 to converge for this substrate. Lichenan parameters were correlated to a lesser extent than XG (R^2 = 0.38), and a weak and inverse correlation was observed with GM (R^2 = 0.15) due to the occurrence of several efficiency clusters. It should be noted that if the single lowest-performing enzyme was removed for XG and lichenan (6MQ4 and 6XSO, respectively), the correlation coefficient rose to 0.88. Kinetic parameters for xylan were not significantly correlated (R^2 = 0.08), and although K_m ranged from 1.3 to 25 g/liter, k_{cat} did not surpass 150 min^{-1}.

Reaction with insoluble polysaccharides

When the present enzymes were tested on phosphoric acid–swollen cellulose (PASC) and mannan, a linear, saturable relationship between [S] and k_{cat} was not observed, so the turnover number measured at 10 g/liter substrate was used for comparison (Fig. 6). The activities on these insoluble substrates were about an order of magnitude slower than on XG, GM, and lichenan (with xylanase activities typically in between).

End products of reaction with oligosaccharides

Nanostructure-initiator MS, or NIMS (43), was used to determine the products released from hexose and pentose oli-
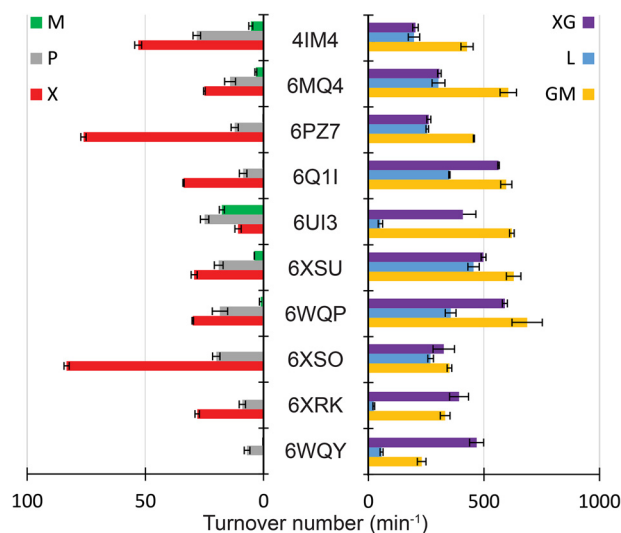

Figure 6. Turnover numbers of GH5_4 measured at 10 mg/ml substrate. Substrates were as follows: mannan (*M*), phosphoric acid–swollen cellulose (*P*), beechwood xylan (*X*); tamarind XG, lichenan (*L*), and konjac GM. *Error bars*, one S.D. over three simultaneous experimental replicates.

gosaccharides, which have sufficient length to span across the active-site cleft and revealed unique patterns for occupancy of the negative and positive binding sites leading to catalysis (Table 4). Overall, reactivities observed for the oligosaccharide substrates matched those observed for full-length cellulose and mannan, but not for xylan (compare Table 4 and Fig. 6). All 10 enzymes hydrolyze PASC and C6, and five (4IM4, 6MQ4, 6UI3,

Table 4

Concentration (mM) of products detected by NIMS from hydrolysis of cellohexaose, mannohexaose, and xylohexaose by GH5_4 enzymes

Reactions showing only C6, M6, or X6 indicate a lack of substrate cleavage. Values are the average of three simultaneous experimental replicates. Blank cells indicate that <0.01 mM product was observed.

Enzyme	Cellohexaose						Mannohexaose						Xylohexaose					
	C1	C2	C3	C4	C5	C6	M1	M2	M3	M4	M5	M6	X1	X2	X3	X4	X5	X6
4IM4	0.03	0.37	0.13	0.01					0.03			0.62		0.02	0.17	0.01		0.11
6MQ4	0.02	0.42	0.19					0.02	0.19	0.02		0.64		0.03	0.29	0.03	0	0.04
6PZ7	0.04	0.37	0.27									0.81		0.03	0.21	0.02		0.05
6Q1I	0.05	0.28	0.26									0.57		0.01	0.08	0.01		0.19
6UI3	0.02	0.42	0.12					0.17	0.69	0.21		0.81						0.28
6XSU	0.12	0.26	0.16					0.09	0.2	0.1		0.49		0.03	0.11	0.02		0.13
6XSO	0.11	0.13	0.29	0.03								0.81		0.01	0.05	0.01		0.11
6WQP	0.09	0.45	0.17					0.01	0.04	0.01		0.64			0.03			0.28
6XRK	0.06	0.09	0.15	0.11	0.07							0.81						0.21
6WQY	0.02	0.09	0.24	0.1								0.54						0.25
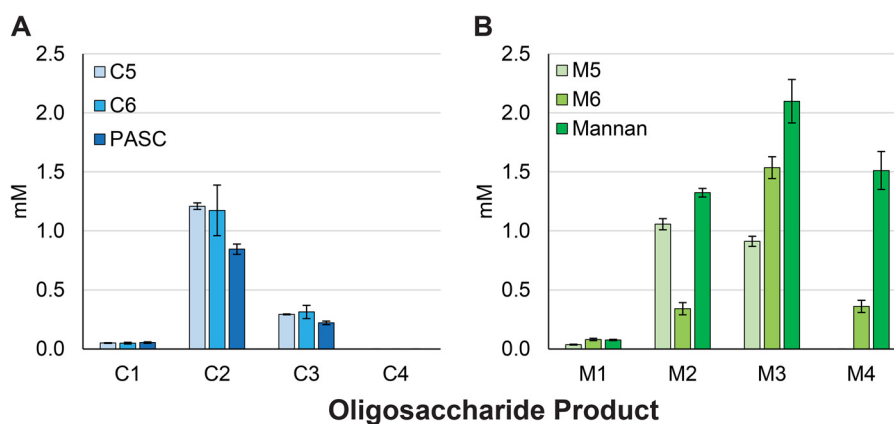
Figure 7. End products of oligo- and polysaccharide hydrolysis by 6UI3 as determined by quantitative NIMS. *A,* cellulose-based substrates. *B,* mannose-based substrates. *Error bars,* one S.D. of three simultaneous experimental replicates.

6XSU, and 6WQP) also hydrolyze mannan and M6. Whereas nine enzymes (all but 6WQY) hydrolyze xylan, two of these nine (6UI3 and 6XRK) do not hydrolyze xylohexaose (X6).

The two enzymes from clade 2, 6WQY and 6XRK, hydrolyzed C6 into a breadth of products that only modestly favored C3. Moreover, 6XSO from clade 3G also favored C3 among several products, suggesting low selectivity for occupation of sugar-binding sites spanning the site of catalysis.

The remaining seven enzymes from clade 3 favored C2 (~50–75% of total products) without accumulation of C4 or C5, implying processivity of hydrolysis of intermediate products, as was previously shown with 4IM4 (44). Within clade 3, further distinctions were identified, as all clade 3B enzymes showed minimal β-(1,4)-glucosidase activity (<10% of total products as glucose), whereas the single 3F enzyme (6XSU) and two 3G enzymes (6XSO and 6WQP) released >10% glucose.

Although M6 and X6 were not completely hydrolyzed in these experiments, there was less variation in their hydrolysis products relative to C6. Both M6 and X6 were preferentially hydrolyzed into the trisaccharide. Except for 4IM4 and 6WQP, lesser but equivalent amounts of the di- and tetrasaccharides were observed, suggesting that asymmetric binding across the catalytic site was less favorable for M6 and X6 than for C6, but still allowed.

The dual cellulase-mannanase 6UI3 offered an opportunity to further explore cleavage specificity with cellulose and man-nan oligosaccharides (Fig. 7). For C5, C6, and PASC, C2 was the dominant product, with lesser amounts of C3 and C1 and no C4 or C5 observed (Fig. 7A). The C6 results imply preferred asymmetric binding across the catalytic site and also the processive hydrolysis of C4. In contrast, for M6 and mannan, M3 was the dominant product, and near-equimolar amounts of M2 and M4 were observed. The M6 results imply a contrasting outcome: preferred symmetric binding of M6 across the catalytic site and no processive hydrolysis of M4. For M5, near-equimolar amounts of M3 and M2 were observed along with no processive hydrolysis of M3 (Fig. 7B).

Binding affinity to polysaccharides

Affinity gel electrophoresis (Fig. S2) was used to estimate K_D for XG, GM, and low-molecular weight xylan. In general, when binding interactions could be observed, the estimated K_D value was 2–4 orders of magnitude lower than the corresponding K_m. For example, 6PZ7 has a K_D of ~0.01 g/liter on XG but a K_m of 3.4 ± 0.3 g/liter. The binding of polysaccharides to 6WQY and 6XSO could not be tracked by electrophoresis, likely due to mismatch between the enzyme pI (6XSO, pI = 6.44; 6WQY, pI = 4.60) and the buffer needed for the electrophoresis.

None of the active enzymes showed changes in gel mobility in the presence of low-molecular weight xylan, which may have been caused by in-gel hydrolysis. After creation of Glu-to-Gln mutations of the catalytic Glu residues for five enzymes (4IM4,

6PZ7, 6UI3, 6XSO, and 6WQY), only the mutated 4IM4 showed detectable binding to soluble xylan with K_D ~0.03 g/liter.

The blue fluorescent protein mKalama1 was fused to 6UI3, and a simple pulldown assay was used to determine the K_D on insoluble substrates PASC (0.53 ± 0.03 g/liter) and high-molecular weight beechwood xylan (0.89 ± 0.09 g/liter). Although the K_D values suggest tight binding, the catalytic activity increased linearly with increasing substrate concentration up to 10 g/liter, emphasizing the mismatch between apparent binding and effective binding leading to catalysis. No binding was detected on insoluble β-(1,4)-mannan.

Discussion

Biological occurrence of GH5_4 enzymes

GH5_4 is primarily a bacterial subfamily, although the occasional occurrence of GH5_4 sequences in fungal and protozoan genomes may be attributed to horizontal gene transfer from bacteria. Many organisms using GH5_4 enzymes are rumen-dwellers, and microbial communities in animal digestive tracts have been noted for high frequencies of glycoside hydrolase horizontal gene transfer (45).

GH5_4 polypeptides also occur frequently in cellulosomes (6, 46), the large, extracellular nanomachines anchored to the outer membranes of certain anaerobic bacteria and fungi. Four of the present enzymes are naturally incorporated into cellulosomes, as evidenced by their fusion to dockerin domains (Fig. 1C). Cellulosomal organisms may be either ruminal or terrestrial. Two of the present cellulosomal GH5_4 proteins (4IM4 and 6MQ4) also possess carbohydrate esterase domains, which help debranch xylan and break linkages to lignin in plant biomass (47).

There appears to be no correlation between the habitat of the host organism and a preferred domain arrangement or GH5_4 clade. This may be a testament to the inherent catalytic flexibility of GH5_4 enzymes, which allows incorporation of the catalytic domains into a wide variety of biomolecular frameworks for optimal nutrient acquisition.

Functional diversity via loop length

GH5 enzymes adopt the $(\alpha/\beta)_8$-barrel fold, where loops connecting the outer α-helices to the core β-strands contribute to active-site shape. We examined the structural arrangements of these loops in greater detail and identified several features that contribute to substrate selectivity across GH5_4.

Loop contributions to the active-site shape

In all cases but one (6XRK), loop 4 forms the left wall of the active-site cleft and provides a critical Trp residue that stacks against the +1 sugar in the active site, positioning it for catalysis. Loop 6 forms the right wall of the active-site cleft and supplies additional aromatic residues to position the sugar chain perpendicular to the floor of the active site. Loop 6 also has the greatest diversity in length, ranging from the 8-residue loop in 6XSO (offset by a larger-than-average loop 4) to the 35-residue loop containing two β strands in 6XRK (offset by a smaller-than-average loop 4). The size and conformation of these two

loops determine the width of the cleft immediately following the site of catalysis—the pinch point where the +1 sugar is clamped into position for hydrolysis.

Enzymes with a narrower cleft tend to have lower K_m for GM than XG, with the exception of 6PZ7 and 4IM4, which have similar K_m for both substrates, suggesting the linear GM can more easily settle across the active site when space is limited. Again, with the exception of 4IM4 and 6PZ7, a narrow cleft also appears to correlate generally with a higher k_{cat} on all substrates. This makes sense, considering that accurate placement of the +1 sugar may help force the scissile glycoside bond into proper position. Conversely, a narrower cleft may interfere with the initial chain-loading step for complex substrates like XG, raising K_m.

Other loops impact the shape of the binding cleft, especially loops 1, 3, 5, and 8.

Loop 1 determines the length of the "loading platform" and the number of potential interactions with negative-subsite sugars. It is not clear whether these extended platforms provide a defined fourth negative subsite, but 4IM4 and the other proteins with a shortened loop 1 can accommodate no more than three.

Loop 3 provides one interesting case, where in 6XSO, a β-turn harboring a Lys-Asp pair projects in toward the cleft, providing potential close polar contacts with several branch sugars in the modeled substrate (Fig. 4). Loop 3 also supplies a Phe side chain oriented to contact a galactose branch in XG, but the functional relevance of these potentially specific interactions requires further study.

Loop 5, with the shortest average length of all the loops, provides some potential stabilizing interactions at the +1 subsite, where a xylose branch was observed to bind in an oligosaccharide co-crystallization study (16). Although this xylose binding interaction may enable XGase activity in GH5_4, the potential strength of the interaction did not appear to correlate with differences in either K_m or k_{cat}.

Finally, loop 8 extends the boundary of the right wall formed by loop 6. 4IM4, 6MQ4, and 6Q1I possess a Tyr that stacks optimally with a xylose branch at the −3 position, but the presence of this interaction did not correlate with a lower K_m or higher k_{cat} on XG in this data set.

In summary, the contributions of each of these auxiliary loops to substrate selectivity may be incremental compared with cleft width. However, the combination of multiple interactions likely contributes to the fine-tuning enzyme selectivity. Additional studies of the contributions of the above-mentioned four loops may help clarify their contributions to branch sugar interactions.

Aromatic surfaces guide substrate selectivity

Interactions between electronegative π-systems of aromatic side chains and electropositive C–H bonds in sugar rings are frequent and important (48). The number and location of enthalpically driven π-CH interactions, especially involving Trp, are key determinants of protein affinity for carbohydrate ligands (49), and most sugars display anisotropy between their faces; that is, one side of the ring is more electropositive and interacts more often with aromatic side chains. Glucose, however, is a special case because all ring hydroxyls are oriented

equatorially (48), meaning both faces have high potential for π-CH interactions. GH5_4-binding clefts appear to take advantage of this property when interacting with glucan substrates by clamping them with aromatic residues on both sides in the vicinity of the active site, but not in every case (Fig. 3).

Aromatic surfaces may help guide a bound substrate chain to the *exit*, and GH5_4 proteins use two strategies: linear and bent. Several enzymes with high catalytic efficiency for reaction with GM have additional Trp and Tyr residues oriented along the cleft axis, which may keep a linear substrate in a linear exit conformation. 6XSU exemplifies this strategy best, because it has two exposed Trp side chains comprising a continuous surface along the left wall toward the exit, encouraging a linear conformation (Fig. 3*D*). This makes 6XSU the most efficient GMase, but a poor XGase when compared with the other enzymes.

6WQY and 6XRK are among the most efficient XGases and the poorest GMases, but probably for different reasons. 6WQY possesses two Trp residues in the left wall, which is critical for GMase activity and appears to guide a linear exit; the right wall, however, lacks a properly positioned aromatic side chain (Fig. 2). A linear exit is thus encouraged by the left wall but not enforced by a clamp.

In contrast, 6XRK lacks a left wall but presents aromatic surfaces on the right wall to guide a bent glucan exit conformation. This makes it an excellent XGase, because the +1 xylose branch predisposes the substrate to a bent exit, but the lack of a clamp renders it a poor GMase. This emphasizes an important finding that in GH5_4, high efficiency on complex substrates such as XG and on linear substrates such as GM are not mutually exclusive, so long as both a clamp and an optional bent exit (or at least an unenforced linear exit) are available.

A structure recently published by Venditto *et al.* (18) of Cel5B, a GH5_4 enzyme from *Bacillus halodurans* (PDB entry 4V2X), shows a clear case where multiple strategies are simultaneously employed to achieve broad substrate selectivity. The enzyme bears two additional C-terminal domains, an Ig-like linker domain followed by CBM46. The latter is unique to clade 1 proteins GH5_4 and presents an additional Trp to the right wall of the cleft, continuous with loop 6. Using the XG oligosaccharide ligand from 4W88 as a model, the angle of exit of the glucan chain points almost directly to this residue, which could interact favorably with a glucose at +4. In our structure of 6XRK, the position of the loop 6 Trp-238 aligns closely with Trp-501 on the CBM46 of Cel5B (distance between the Trp CG atoms of these two residues in the two structures is 4.7 Å after alignment of the central TIM barrels). This evolved convergence of structures suggests a common purpose for the Trp as a positionally specific supplement to substrate binding.

In addition, Cel5B possesses a well-defined left wall with a Trp contacting the +1 glucose, which our results indicate is a requirement for high catalytic efficiency on linear substrates; however, this Trp is missing in 6XRK. Indeed, Cel5B reacts with both BBG and XG and is ∼5-fold more reactive with BBG. However, upon deletion of CBM46, k_{cat} for BBG is detrimentally affected. This implies that in the case of Cel5B, the left wall is necessary but insufficient for high linear substrate efficiency and that the extended right wall provided by CBM46 plays a

critical role in keeping the chain in a catalytically competent position. The overall cleft is relatively wide, which is expected to reduce catalytic efficiency with linear substrates (as in the case of 6XSO), but by taking advantage of the bent exit normally used by XG, BBG can form sufficient stabilizing interactions to remain in the cleft long enough for nucleophilic attack to occur.

Bending of glucan chains to fit the particular shape of GH5 active sites has been reported previously (9), and although frequently observed in lichenases that hydrolyze both β-(1,3) and β-(1,4) linkages, a "kinked" β-(1,3) linkage need not be a specific requirement for bending, given the flexibility of β-(1,4) polysaccharides in solution. Additionally, the pyranose rings of sugars need not be centered directly over the aromatic plane of Trp and Tyr residues to benefit from stabilizing forces (50), supporting a model of dynamic interaction between a flexible glucan chain and a protein surface "painted" with π-electron density.

In summary, our data combined with those of others support a mechanistic framework for dual branched-linear substrate selectivity in GH5_4 along five general principles: 1) high catalytic efficiency with linear polysaccharides requires, at minimum, π-CH interactions from specifically positioned aromatic residues on the left side of the +1 sugar; 2) additional interactions with positive-subsite sugars enhance catalysis with XG, but must be paired with a bent exit; 3) linear substrates may use either bent or linear pathways to exit the channel, and at least one must be available for high catalytic efficiency; 4) XG will use the bent pathway to exit the channel, if available; and 5) a narrow, two-sided clamp enhances hydrolysis but can reduce catalytic efficiency with XG (via reduced saturability) in the absence of additional stabilizing interactions.

Catalytic efficiency and substrate solubility

GH5_4 has received attention in recent years as a source of multifunctional "cellulases," but this term is ambiguous and may obscure the true function of individual members of this subfamily. Our present study indicates that GH5_4 is best characterized as a β-glucanase subfamily with some members that exhibit a breadth of reactivity on mixed-linkage glucans, XGs, and hetero-β-(1,4)-glucans, plus some less frequent and lower reactivity on amorphous cellulose and insoluble β-(1,4)-polysaccharides.

We observed convergence of catalytic efficiency (k_{cat}/K_m) for XG and GM and to a lesser extent for the mixed-linkage glucan lichenan, but not for xylan, mannan, or PASC. GH5_4 may have reached an efficiency ceiling for XG and GM, so these may be the "true" substrates of the subfamily.

The lack of convergence for the other substrates suggests that either 1) there is less selection pressure for improvement on insoluble substrates in the enzymes' native context, or 2) improvement on both classes of substrate is mutually exclusive because of incompatible protein structural requirements (*i.e.* flexible and hydrated polymers require a different active-site shape than rigid and crystalline ones).

In every case where k_{cat} could be measured for xylan, it was lower than for XG, GM, and lichenan, although K_m was occasionally lower for xylan than other substrates. GH5_4 enzymes
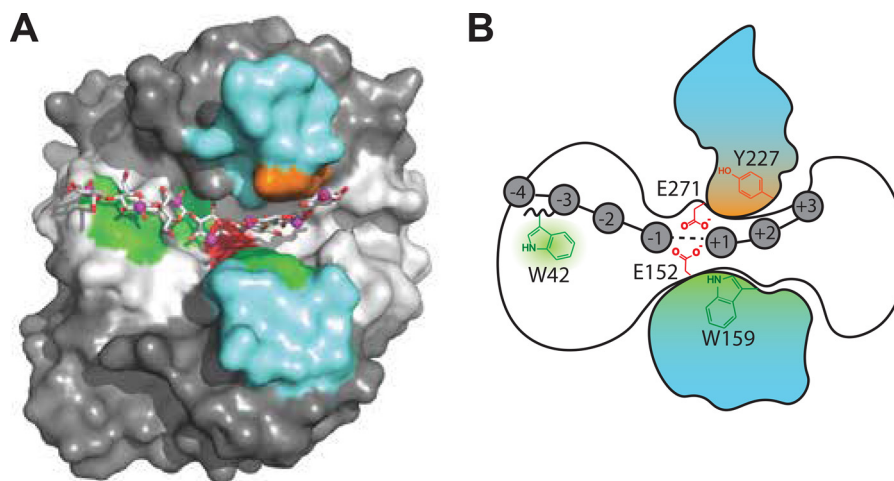
Figure 8. Model of cellulose and mannan cleavage specificity in the 6UI3 active-site cleft. *A*, structure of 6UI3 with *β*-(1,4)-glucan chain of XG modeled in to approximate the binding of cellohexaose. C-2 carbon atoms are shown as *magenta spheres* to highlight the location of stereochemical difference between glucose and mannose. Clamp-forming loop 4 and loop 6 are highlighted in *blue*. *Green*, Trp; *orange*, Tyr. *B*, schematic of 6UI3-binding cleft, highlighting sugar subsites and key binding (Trp-42, Trp-159, Tyr-227) and catalytic (Glu-152, Glu-271) residues. Trp-42 is hypothesized to play an important role in the registry shift from cellulose to mannan binding at the −3 subsite.

may thus be able to accept *β*-(1,4)–linked pentose cell-wall polysaccharides, like xylan, but are more optimally evolved to hydrolyze hexose substrates, which better fill the space in the active-site cleft.

Activities on PASC were lower still, and mannan hydrolysis occurred above background for only five of the 10 proteins. We attribute the low reactivity on these substrates to their near total insolubility in water (and thus poor enzyme accessibility), although in the case of mannan, we considered the possible contribution of the axial hydroxyl at C2 in mannose (see next section). That GM was one of the most reactive substrates suggests that *β*-(1,4)-mannosyl linkages may be as reactive as *β*-(1,4)-glucosyl linkages when hydrated, especially considering that the GM substrate was ~60% mannose by weight.

6UI3 was the most active mannanase in our original screen (6), yet it did not bind mannan in our pulldown assay (but it did bind GM), further suggesting that mannanase activity is limited by low accessibility of the substrate.

Sugar subsite occupancy and selectivity

The dual cellulase-mannanase 6UI3 offered an opportunity use NIMS to further explore the cleavage selectivity of two distinct oligosaccharides. Fig. 8 provides a schematic of binding in the negative and positive sugar subsites that provides insight into the origin of the different product distributions observed with C6 and M6.

For M6 and M5 hydrolysis, the following reactions are relevant.

$$M6 \rightarrow M3 + M3$$

$$M6 \rightarrow M4 + M2$$

$$M5 \rightarrow M3 + M2$$

Reactions 1–3

The results of Table 4 and Fig. 7 indicate that symmetric cleavage of M6 is preferred (Reaction 1, ~65% of hydrolysis products), which must occur by occupancy of −3 to +3 sites (Fig. 8*B*). A secondary, asymmetric cleavage of M6 (Reaction 2) could utilize either the −4 to +2 or the −2 to +4 sites, leading to the observed nearly equal proportions of M2 and M4. Furthermore, M4 is not further hydrolyzed, so 6UI3 requires occupation of at least five sugar subsites for its mannan hydrolysis reaction.

For M5, equimolar amounts of M3 and M2 were observed, consistent with Reaction 3. *β*-(1,4)-Mannosidase activity, removal of a single mannose residue, was not observed because neither M1 nor M4 was detected in the M5 reactions above trace amounts. The product distribution from M5 is therefore consistent with occupation of the −3 to +2 (or, less likely, the −2 to +3 sites), leading to the observed equal proportions of M2 and M3.

For C6 hydrolysis, the following equations are relevant.

$$C6 \rightarrow C3 + C3$$

$$C6 \rightarrow C4 + C2$$

$$C4 \rightarrow C2 + C2$$

Reactions 4–6

The results of Table 4 and Fig. 7 indicate that C6 reacts in two ways fundamentally different from M6. First, the lack of any C4 product indicates that cellulose hydrolysis only requires occupancy of the −2 to +2 sites (Fig. 8), one fewer negative site than mannan. The requirement for occupation of four sugar-binding sites means that C3 is not further processed into C2, and this is supported by the negligible amount of C1 observed.

Second, although symmetric binding leads to the accumulation of C3, 6UI3 seems to preferentially use an asymmetric binding mode, which can be rationalized as follows. Assuming no reaction of C3, if Reaction 4 and the combination of

Reactions 5 and 6 were equally favorable, the expected C2:C3 ratio would to be ~1.5. However, Table 4 shows that the actual ratio of C2:C3 is ~3.5. Consequently, an asymmetric binding of C6 into either (or both) of the −2 to +4 or the −4 to +2 binding sites is required to account for the observed product distribution.

Fig. 8 provides insight into the possible origin of the differences in cellulose and mannan hydrolysis by 6UI3. Overall, the discrimination appears to reside at the −3 site, because occupancy of this site is required for mannan hydrolysis but not for cellulose hydrolysis. This makes sense structurally, because the axial -OH at C2 in mannose points (favorably) away from the protein at −1 and −3 sites but (unfavorably) toward the protein at the −2 site (Fig. 8*A*).

Binding of additional sugars before the −1 site is critical for GH5 activity (11); however, the axial OH in mannose likely disrupts binding at the −2 site. Therefore, we propose that binding at the −3 site is required to compensate. The mannose sugars bound in the −1 and the −3 sites may also have higher affinity for the conserved aromatic residues than glucose because of the increased electropositivity of mannose contacts with the aromatic residues, potentially compensating for the disrupted binding at the −2 site (48).

The Trp residue at the −3 site is conserved in GH5_4, and all of the enzymes that were active with M6 yielded a similar product distribution to 6UI3. However, not all GH5_4 enzymes hydrolyze mannan. Thus, the interaction of mannose and Trp at the −3 site helps to explain product distributions, but not the presence or absence of mannan reactivity.

Pentose versus hexose polysaccharides

The similarity of product distributions for M6 and X6 (Table 4) suggests similar requirements for occupation of the sugar subsites. However, the physical rationale for this requirement is likely not the same, because unlike mannose, xylose has relatively symmetric electropositivity on each of its faces. Instead, the lack of C6-OH on xylose may affect the positioning of the −1 and +1 sugar. For the −1 sugar, a C6-OH may provide leverage to insert the pyranose ring deep in the active site, whereas at the +1 sugar, this functional group could contact the cleft surface directly, angling the leaving sugar upward to assist the positioning of the scissile bond. These putative interactions are absent in xylose, which may explain the need for more stabilizing interactions at the distal −3 site to compensate.

Saturation of binding and catalysis

Whereas K_D is commonly expressed as the ratio of k_{off}/k_{on} for substrate binding to free enzyme, K_m has been conceptualized as the ratio of substrate capture rate to product release rate (51), with [S] = K_m being the substrate concentration at which these two rates are equal. At [S] < K_m, activity is limited by the capture rate, whereas at [S] > K_m, total velocity is often limited by product release. The low values of K_D relative to K_m in this study suggest a rapid association of substrate with free enzyme, but despite the ability to saturate binding at low concentration, catalysis is not saturated until much higher concentrations.

One possible explanation for this discrepancy is the distinction between binding and *productive binding*; that is, whereas nearly 100% of the enzyme may be associated with a substrate well before reaching [S] = K_m, only a certain proportion of these *E*S pairs may be adequately poised to initiate the first step of catalysis, which is glycosylation of the enzyme via nucleophilic attack of the catalytic Glu. Tight, nonproductive binding of cellulases to polysaccharides and the aromatic polymer lignin has been widely observed and modeled (52–56), and given the modular nature and promiscuity of GH5 active sites, as well as the structural similarity among polysaccharides, it is straightforward to envision how an enzyme could interact favorably with a glycan chain without achieving the proximity to the catalytic residues required for hydrolysis. An inverse relationship exists between substrate affinity and hydrolysis rate in many cellulases (56), which may stem from such nonproductive binding.

A second possibility may be related to release of products from the active site. Given that the structure of the substrate and products are similar and that the majority of the protein-ligand interactions are unchanged from the beginning to the end of catalysis, it may take additional time, post-deglycosylation, for the two products to dissociate. Short oligosaccharides are known to be competitive inhibitors of some cellulases (57–59), supporting the notion that hydrolysis products retain appreciable affinity for the active site. High substrate availability could help to speed product dissociation, because an uncut substrate could cooperatively anchor into both the negative and positive subsites, whereas the two products only access one anchor each.

The reason K_D is up to several orders of magnitude smaller than K_m in all cases we tested may be a combination of the factors above. Whereas nonproductive binding may seem like a disadvantage, the ability to interact with multiple polysaccharides sharing broadly similar characteristics (*i.e.* β-(1,4)–linked main chains) is the fundamental requirement for broad selectivity, a trait that has obvious evolutionary advantages in a complex or changing nutritional environment.

The wide range of K_m and k_{cat} on XG, lichenan, and GM (Table 2), along with the colinear clustering of k_{cat}/K_m (Fig. 5), also hints at evolutionary adaptation to diverse environments. Individual GH5_4 enzymes may, for example, be adapted to lower substrate concentrations (lower K_m), but their increased affinity comes at the cost of dampened reaction velocity (lower k_{cat}). Whether this trade-off is imposed by the physical limits of mass action on the substrate or by a fitness ceiling in the local protein sequence landscape is the subject of ongoing investigation in many laboratories.

Conclusions

In this work, our standardized comparative examination of GH5_4 structure and function revealed complex molecular mechanisms for determining substrate selectivity. Polysaccharide recognition is mediated by aromatic residues at several key points in the binding cleft, some of which sit at considerable distance from the site of catalysis. GH5_4 widely accommodates features of branched XG, whereas individual members

may be further optimized for cleavage of more linear glucans by favoring different chain conformations in the positive subsites. Whereas the GH5_4 fold is well-conserved, the shapes of GH5_4-binding clefts may show radical differences, as various structural strategies have evolved to supply critical protein-carbohydrate interactions. Diagnostic glucan and mannan oligosaccharides are cleaved into products of different sizes, suggesting that the negative subsites play an important role in enabling broad spectrum activity. Furthermore, the ability of GH5_4 proteins to tolerate so much sequence diversity (7, 15) and to operate on such a wide array of substrates highlights the potential of GH5_4 as a starting point for efforts to create custom glycoside hydrolases for industrial bioenergy applications (60–62).

Experimental procedures

Gene sequence sources and PCR

DNA sequences encoding GH5_4 domains were amplified from a library of pEU cell-free expression plasmids synthesized in previous work (6). Insertion into isopropyl 1-thio-β-D-galactopyranoside–inducible overexpression vector pVP67K or into arabinose-inducible pBAD-mKalama1 was confirmed by colony PCR and Sanger sequencing.

Site-directed mutagenesis

The catalytic nucleophile was identified in each sequence by protein sequence alignment (Clustal Omega), using 4IM4 as a reference. Mutagenic primers, of ~40 base pairs in length, were designed to mutate the nucleophilic Glu to a noncatalytic Gln. Site-directed mutagenesis was performed by typical PCR protocols, and mutagenesis was confirmed by Sanger sequencing.

Construction of fluorescent fusion proteins

The plasmid pBAD-mKalama1 was obtained as a gift from Robert Campbell (Addgene 14892). GH5_4 catalytically inactive gene sequences were cloned between the N-terminal His$_6$ tag and the fluorescent mKalama sequence by Gibson assembly. Insertions were confirmed by PCR and Sanger sequencing.

Protein expression and purification

Rosetta (Millipore-Sigma) cells were transformed with sequence-confirmed WT or catalytically inactivated plasmids (pVP67K, N-terminal His$_8$ tobacco etch virus–cleavable tag) and plated on selective medium. Protein expression was performed using autoinduction (63–65). Starter cultures were grown in non-inducing medium overnight and used to inoculate autoinduction medium, and cells were grown at 25 °C for 24 h. Cells were harvested by centrifugation and lysed using lysozyme and sonication. Lysate was centrifuged at 20,000 rpm for 45 min, and the supernatant was clarified with a 0.4-μm PES filter. Filtered supernatant was separated by nickel affinity chromatography on an Akta Start FPLC. Purified protein was desalted and concentrated to 10–20 mg/ml.

Expression of mKalama1 fusion proteins was performed by arabinose induction in LB medium (66), but purification was performed the same as above.

X-ray crystallography

Crystallization experiments were conducted in the Collaborative Crystallography Core in the Department of Biochemistry at the University of Wisconsin (Madison, WI, USA). Unless otherwise specified, the following procedures were followed for all targets. Crystallization experiments were set with a TTP Labtech Mosquito robot in MRC SD-2 plates at 20 °C (297 K). General screens used in this study were Hampton Research IndexHT, Molecular Dimensions JCSG+ (67), Morpheus (68), and PACT premier (69). Diffraction data were reduced with XDS (70), scaled with XSCALE (71), solved by molecular replacement with Phaser in the Phenix suite of programs (72), automatically rebuilt with phenix.autobuild (73), iteratively rebuilt in Coot (74), refined using phenix.refine (75), and validated using MOLPROBITY. Diffraction data were collected using beamlines at the Advanced Photon Source (APS) Argonne National Laboratory at 100 K.

4IM4 crystals grew from protein solution (20 mg/ml protein, 50 mM NaCl, 10 mM MOPS, pH 7.0) mixed in a 1:1 ratio with reservoir solution (9% PEG 20,000, 18% MMPEG 550, 0.3 M diethyleneglycol, 0.3 M triethylene glycol, 0.3 M tetraethylene glycol, 0.3 M pentaethylene glycol, 100 mM MES/imidazole buffer, pH 6.5.) The crystal was cryoprotected by supplementing the reservoir with 15% ethylene glycol. Data were collected at LS-CAT 21ID-G at 0.97857 Å on June 10, 2012, using a Rayonix MX300 detector. Data were reduced and scaled with HKL-2000 (76). The structure was phased with 3NDY.

6MQ4 crystals were obtained from 1 mM protein incubated with 5 mM cellohexaose for 4 h at room temperature. The mixture (of protein and cellohexose) was screened using a TTP Labtech Mosquito, using MRC SD-2 plates. Droplets composed of 200 nl of protein-substrate mixture and 200 nl of reservoir were equilibrated against 50 μl reservoir. An exceptional crystal formed directly from the JCSG+ screen, condition B6, 40% ethanol, 10% PEG 1000, 0.1 M phosphate citrate buffer, pH 4.2. Crystals were exposed to vapor from 45% ethanol for 30 s prior to plunge cooling in liquid nitrogen. Data were collected on August 23, 2015 at LS-CAT 21ID-F, 0.97857 Å (540 frames, 0.5°, 90 mm) using a Rayonix MX300 CCD detector. The structure was phased by MR using 2WAB.

6PZ7 protein was crystallized in SD2 plates set by a Mosquito crystallization robot. The hit came from PACT Premiere Screen A7, 200 mM NaCl, 100 mM sodium acetate, pH 5.0, 25% PEG 6000. Crystals were cryopreserved in reservoir solution supplemented with 15% ethylene glycol. Data were collected at LS-CAT 21ID-D on October 10, 2013 at 1.07805 Å (180 frames, 1°, 225 mm) using a Rayonix MX300 CCD detector. The structure was phased by MR using 3NDZ.

6Q1I crystals were set with a Mosquito into SD2 plates. The condition that provided the crystals for this data set came from Morpheus Screen D9, consisting of 10% alcohols, 50% Tris/Bicine, pH 8.5, and 15% P550MME_P20K. Data were collected at LS-CAT 21ID-D on October 13, 2013 at 0.978566 Å using a Rayonix MX300 CCD detector. The structure was phased by MR using 3NDY.

6UI3 crystals were grown by hanging drop vapor diffusion in a VDXm plate against 26% PEG 3350 0.1 M BisTris, pH 5.5, and

cryoprotected with reservoir solution supplemented with 15% ethylene glycol. Data were collected on October 13, 2013 at LS-CAT 21ID-D, 0.9786 Å (360 frames, 0.5°, 1.07805 Å), using a Rayonix MX300 CCD detector.

6SXU crystals were grown by hanging drop vapor diffusion from a 40 mg/ml protein solution in 10 mM MOPS, pH 7.0, 50 mM NaCl, mixed 1:1 with a reservoir solution of 31% PEG 3350, 0.2 M NaCl, 0.1 M BisTris, pH 5.5. Data were collected on August 15, 2015 at LS-CAT 21ID-F at 0.97856 Å (900 frames, 0.3°, 120 mm) using a Rayonix MX300 detector. Data were reduced with xia2-dials using default parameters (77, 78) The structure was phased with 1EDG.

6XSO crystals grew directly from a general screen, JCSG+ (G8). 200 nl of protein at 37 mg/ml, 10 mM MOPS buffer, pH 7.0, 50 mM NaCl were mixed with 200 nl of 0.15 M DL-malic acid, pH 7.0, 20% PEG 3350. Data were collected March 5, 2017 at GM/CA 23ID-B at 0.7749 Å (3600 frames, 0.1°, 250 mm) using an Eiger 16M pixel-array detector. The structure was phased with 3VDH.

6WQP crystals formed from Hampton IndexHT(C9) 0.5% Jeffamine ED-2001, 100 mM HEPES, pH 7.0, 1.1 M sodium malonate, pH 7. Samples were cryoprotected in reservoir supplemented with 20% ethylene glycol. Data were collected at LS-CAT 21ID-D on October 13, 2013 at 0.9786 Å (200 frames, 1°, 200 mm) using a Rayonix MX300 CCD detector. The structure was phased by MR using 4IM4.

6XRK crystals were obtained from protein solution at 20 mg/ml in 50 mM NaCl and 5 mM MOPS, pH 7.0, with 200 nl of protein mixed with 100 nl of reservoir consisting of 22% PEG 3350, 0.1 M MgCl$_2$, 0.1 M BisTris, pH 5.5. Crystals were cryoprotected by supplementing the reservoir solution to 35% PEG 3350. Data were collected on March 5, 2017 at GM/CA 23-ID-B at 0.7749 Å (3600 frames, 0.1°, 250 mm) using an Eiger 16M pixel-array detector. The structure was phased with 2JEP.

6WQY crystals were obtained from protein solution at 59 mg/ml in 50 mM NaCl and 5 mM MOPS, pH 7.0, by mixing 200 nl of protein solution with 130 nl of reservoir consisting of 16% PEG 3350, 0.2 M MgCl$_2$, 0.1 M sodium acetate buffer, pH 5.0. Crystals were cryoprotected by supplementing the reservoir solution to 35% PEG 3350. Data were collected on October 20, 2016 at GM/CA 23-ID-B at 0.7749 Å (3600 frames, 0.1°, 200 mm) using an Eiger 16M pixel-array detector. The structure was phased with 4YHE.

6WQV crystals were obtained by soaking 6WQP crystals in 200 mM cellotriose and cryoprotected with 20% ethylene glycol. Data were collected at LS-CAT 21ID-G on April 4, 2014 at 0.97856 Å (400 frames, 1°, 200 mm) using a Rayonix MX300 CCD detector. The structure was phased with 6WQP.

Elastic network modeling

The atomic coordinates for 6XRK and 6WQY were submitted to the *elNémo* server (39) (RRID:SCR_019137). Using the default settings (NMODE = 5, DQMIN = −100, DQMAX = 100, DQSTEP = 20), the five largest low-frequency modes were calculated and visualized as animations in PyMOL.

Enzyme kinetic analysis

All substrates were obtained from Megazyme, except beech-wood xylan (Sigma–Aldrich) and PASC (prepared in-house from Avicel from Sigma–Aldrich). Enzyme activities on polysaccharides were determined using the BCA reducing-sugar assay. Assays were conducted in triplicate with enzyme concentration of 1 μM and substrate concentrations up to 10 g/liter, in a 200-μl total volume of 0.05 M phosphate buffer, pH 6. Reactions were incubated at 30 °C, with shaking at 900 rpm. XG, GM, and lichenan reactions were incubated for 10 min, whereas PASC, mannan, and xylan reactions were incubated for 60 min. The 96-well plates were spun at 3000 × g for 30 min at 4 °C, pelleting the undigested substrate. The supernatant (5 μl) was added to 100 μl of BCA working reagent, and plates were incubated at 80 °C for 15 min. Sample absorbance was measured at 562 nm and converted to μmol/min using a glucose standard curve.

Data were fit to the Michaelis–Menten equation using the Solver function in Microsoft Excel (GRG nonlinear least-squares error minimization with k_{cat} and K_m as parameters).

End-product analysis by NIMS

To determine the concentrations of oligosaccharide enzyme digest products, reactions were carried out at 30 °C, pH 6 (sodium phosphate, 50 mM) for 2 h using 1 μM enzyme. Substrate concentrations were 0.5 g/liter for oligosaccharides, 5 g/liter for mannan, and 2.5 g/liter for PASC. Analysis was performed by NIMS, as described in detail elsewhere (43, 79). All oligosaccharides were obtained from Megazyme.

Insoluble polysaccharide binding assay

Enzyme binding to insoluble polysaccharides was determined using mKalama1 fluorescence. Assays were conducted in triplicate using an enzyme concentration of 2 μM, PASC concentrations up to 5 g/liter, and insoluble xylan concentrations up to 50 mg/ml in a 200-μl total volume with 0.05 M phosphate buffer, pH 6. The enzyme-substrate reactions were incubated at 4 °C, with shaking at 900 rpm for 60 or 90 min for insoluble xylan or PASC, respectively. Samples were spun at 3000 × g for 30 min (for PASC) or 20,000 × g for 15 min (for xylan) at 4 °C to pellet the bound enzymes along with the substrate. The fluorescence of the supernatant (100-μl nominal volume) was measured using excitation at 385 nm and emission at 456 nm.

Affinity gel electrophoresis

To prepare the 7.5% resolving gel, 3.75 ml of 30% acrylamide, 3.75 ml of 1.5 M Tris, pH 8.8, and 7.5 ml of water were mixed in a vacuum flask and degassed. Polysaccharide affinity gels were prepared by substituting a portion of the water with polysaccharide substrate solution. TEMED (7.5 μl) and ammonium persulfate (75 μl) were added, and the solution was pipetted into gel molds, with 1 ml of isopropyl alcohol added to the top. To prepare the 4% stacking gel, 2 ml of 30% acrylamide, 3.8 ml of 0.5 M Tris, pH 6.8, and 9 ml of water were mixed in a vacuum flask and degassed. TEMED (15 μl) and ammonium persulfate

(75 μl) were added, and when the resolving gel was solidified, the surface was rinsed of isopropyl alcohol, and the stacking gel solution and combs were added.

To determine K_D, native electrophoresis was conducted in a pH 8.3 Tris-glycine running buffer at 200 V for 60 min at 4 °C with enzyme concentration of 0.5 g/liter and substrate concentrations up to 0.1 g/liter. Gels were stained in Coomassie dye and imaged on a light box. R_f was determined for each band relative to BSA, a protein standard that did not bind polysaccharides, and K_D was estimated using normalized R_f curves with respect to ligand concentration.

Data availability

The structures presented in this paper have all been deposited in the PDB under accession numbers 6MQ4, 6PZ7, 6Q1I, 6UI3, 6XSU, 6XSO, 6WQP, 6XRK, 6WQY, 6WQV, and 4IM4. All remaining data are contained in the paper.

Author contributions—E. M. G., E. I. K., C. M. B., T. E. T., T. R. N., and B. G. F. conceptualization; E. M. G., E. I. K., C. A. B., N. L. I., K. D., C. M. B., T. E. T., T. R. N., and B. G. F. data curation; E. M. G., E. I. K., C. A. B., N. L. I., K. D., C. M. B., T. E. T., T. R. N., and B. G. F. formal analysis; E. M. G., T. E. T., T. R. N., and B. G. F. supervision; E. M. G., E. I. K., C. A. B., N. L. I., K. D., C. M. B., T. E. T., and B. G. F. visualization; E. M. G., E. I. K., C. A. B., N. L. I., K. D., C. M. B., T. E. T., T. R. N., and B. G. F. methodology; E. M. G., E. I. K., C. A. B., N. L. I., T. E. T., T. R. N., and B. G. F. writing-original draft; E. M. G., E. I. K., C. A. B., N. L. I., K. D., C. M. B., T. E. T., T. R. N., and B. G. F. writing-review and editing; E. I. K., C. A. B., N. L. I., K. D., C. M. B., and T. E. T. resources; E. I. K., C. A. B., N. L. I., K. D., C. M. B., and T. E. T. software; T. E. T., T. R. N., and B. G. F. funding acquisition; T. E. T., T. R. N., and B. G. F. investigation; T. E. T., T. R. N., and B. G. F. project administration.

Conflict of interest—The authors declare that they have no conflicts of interest with the contents of this article.

Abbreviations—The abbreviations used are: GH, glycoside hydrolase; PDB, Protein Data Bank; CAZy, Carbohydrate-Active Enzyme; C1, glucose; C2, cellobiose; C3, cellotriose; C4, cellotetraose; C5, cellopentaose; C6, cellohexaose; M1, mannose; M2, mannobiose; M3, mannotriose; M4, mannotetraose; M5, mannopentaose; M6, mannohexaose; X1, xylose; X2, xylobiose; X3, xylotriose; X4, xylotetraose; X5, xylopentaose; X6, xylohexaose; mannan, β-(1,4)-d-mannan; MR, molecular replacement; Bicine, *N,N*-bis(2-hydroxyethyl) glycine; XG, xyloglucan; BBG, barley β-glucan; GM, glucomannan; GMase, glucomannanase; PASC, phosphoric acid–swollen cellulose; NIMS, nanostructure-initiator MS; BisTris, 2-[bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol; TEMED, *N,N*, *N',N'*-tetramethylethylenediamine.

References

1. Chandel, A. K., Garlapati, V. K., Singh, A. K., Antunes, F. A. F., and da Silva, S. S. (2018) The path forward for lignocellulose biorefineries: bottlenecks, solutions, and perspective on commercialization. *Bioresour. Technol.* **264,** 370–381 CrossRef Medline
2. Oke, M. A., Annuar, M. S. M., and Simarani, K. (2016) Mixed feedstock approach to lignocellulosic ethanol production-prospects and limitations. *Bioenerg. Res.* **9,** 1189–1203 CrossRef
3. Nagano, N., Orengo, C. A., and Thornton, J. M. (2002) One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions. *J. Mol. Biol.* **321,** 741–765 CrossRef Medline
4. Goldman, A. D., Beatty, J. T., and Landweber, L. F. (2016) The TIM barrel architecture facilitated the early evolution of protein-mediated metabolism. *J. Mol. Evol.* **82,** 17–26 CrossRef Medline
5. Omelchenko, M. V., Galperin, M. Y., Wolf, Y. I., and Koonin, E. V. (2010) Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. *Biol. Direct* **5,** 31 CrossRef Medline
6. Glasgow, E. M., Vander Meulen, K. A., Takasuka, T. E., Bianchetti, C. M., Bergeman, L. F., Deutsch, S., and Fox, B. G. (2019) Extent and origins of functional diversity in a subfamily of glycoside hydrolases. *J. Mol. Biol.* **431,** 1217–1233 CrossRef Medline
7. 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
8. Bianchetti, C. M., Brumm, P., Smith, R. W., Dyer, K., Hura, G. L., Rutkoski, T. J., and Phillips, G. N., Jr. (2013) Structure, dynamics, and specificity of endoglucanase D from *Clostridium cellulovorans. J. Mol. Biol.* **425,** 4267–4285 CrossRef Medline
9. Meng, D. D., Liu, X., Dong, S., Wang, Y. F., Ma, X. Q., Zhou, H., Wang, X., Yao, L. S., Feng, Y., and Li, F. L. (2017) Structural insights into the substrate specificity of a glycoside hydrolase family 5 lichenase from *Caldicellulosiruptor* sp. F32. *Biochem. J.* **474,** 3373–3389 CrossRef Medline

10. 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

11. McGregor, N., Morar, M., Fenger, T. H., Stogios, P., Lenfant, N., Yin, V., Xu, X., Evdokimova, E., Cui, H., Henrissat, B., Savchenko, A., and Brumer, H. (2016) Structure-function analysis of a mixed-linkage β-glucanase/xyloglucanase from the key ruminal Bacteroidetes *Prevotella bryantii* B₁4. *J. Biol. Chem.* **291,** 1175–1197 CrossRef Medline

12. Tseng, C. W., Ko, T. P., Guo, R. T., Huang, J. W., Wang, H. C., Huang, C. H., Cheng, Y. S., Wang, A. H., and Liu, J. R. (2011) Substrate binding of a GH5 endoglucanase from the ruminal fungus *Piromyces rhizinflata*. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **67,** 1189–1194 CrossRef Medline

13. Lo Leggio, L., and Larsen, S. (2002) The 1.62 Å structure of *Thermoascus aurantiacus* endoglucanase: completing the structural picture of subfamilies in glycoside hydrolase family 5. *FEBS Lett.* **523,** 103–108 CrossRef Medline

14. Badieyan, S., Bevan, D. R., and Zhang, C. (2012) Study and design of stability in GH5 cellulases. *Biotechnol. Bioeng.* **109,** 31–44 CrossRef Medline

15. 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

16. Dos Santos, C. R., Cordeiro, R. L., Wong, D. W., and Murakami, M. T. (2015) Structural basis for xyloglucan specificity and α-D-Xylp(1 → 6)-D-Glcp recognition at the −1 subsite within the GH5 family. *Biochemistry* **54,** 1930–1942 CrossRef Medline

17. Gloster, T. M., Ibatullin, F. M., Macauley, K., Eklöf, J. M., Roberts, S., Turkenburg, J. P., Bjørnvad, M. E., Jørgensen, P. L., Danielsen, S., Johansen, K. S., Borchert, T. V., Wilson, K. S., Brumer, H., and Davies, G. J. (2007) Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. *J. Biol. Chem.* **282,** 19177–19189 CrossRef Medline

18. Venditto, I., Najmudin, S., Luis, A. S., Ferreira, L. M., Sakka, K., Knox, J. P., Gilbert, H. J., and Fontes, C. M. (2015) Family 46 carbohydrate-binding modules contribute to the enzymatic hydrolysis of xyloglucan and β-1,3-1,4-glucans through distinct mechanisms. *J. Biol. Chem.* **290,** 10572–10586 CrossRef Medline

19. Wu, T. H., Huang, C. H., Ko, T. P., Lai, H. L., Ma, Y., Chen, C. C., Cheng, Y. S., Liu, J. R., and Guo, R. T. (2011) Diverse substrate recognition mechanism revealed by *Thermotoga maritima* Cel5A structures in complex with cellotetraose, cellobiose and mannotriose. *Biochim. Biophys. Acta* **1814,** 1832–1840 CrossRef Medline

20. Yuan, S. F., Wu, T. H., Lee, H. L., Hsieh, H. Y., Lin, W. L., Yang, B., Chang, C. K., Li, Q., Gao, J., Huang, C. H., Ho, M. C., Guo, R. T., and Liang, P. H. (2015) Biochemical characterization and structural analysis of a bifunctional cellulase/xylanase from *Clostridium thermocellum*. *J. Biol. Chem.* **290,** 5739–5748 CrossRef Medline

21. Chen, Z., Friedland, G. D., Pereira, J. H., Reveco, S. A., Chan, R., Park, J. I., Thelen, M. P., Adams, P. D., Arkin, A. P., Keasling, J. D., Blanch, H. W., Simmons, B. A., Sale, K. L., Chivian, D., and Chhabra, S. R. (2012) Tracing determinants of dual substrate specificity in glycoside hydrolase family 5. *J. Biol. Chem.* **287,** 25335–25343 CrossRef Medline

22. Couturier, M., Féliu, J., Bozonnet, S., Roussel, A., and Berrin, J. G. (2013) Molecular engineering of fungal GH5 and GH26 β-(1,4)-mannanases toward improvement of enzyme activity. *PLoS ONE* **8,** e79800 CrossRef Medline

23. Dorival, J., Ruppert, S., Gunnoo, M., Orłowski, A., Chapelais-Baron, M., Dabin, J., Labourel, A., Thompson, D., Michel, G., Czjzek, M., and Genicot, S. (2018) The laterally acquired GH5 ZgEngAGH5_4 from the marine bacterium *Zobellia galactanivorans* is dedicated to hemicellulose hydrolysis. *Biochem. J.* **475,** 3609–3628 CrossRef Medline

24. Ducros, V., Czjzek, M., Belaich, A., Gaudin, C., Fierobe, H. P., Belaich, J. P., Davies, G. J., and Haser, R. (1995) Crystal structure of the catalytic domain of a bacterial cellulase belonging to family 5. *Structure* **3,** 939–949 CrossRef Medline

25. Fierobe, H. P., Gaudin, C., Belaich, A., Loutfi, M., Faure, E., Bagnara, C., Baty, D., and Belaich, J. P. (1991) Characterization of endoglucanase A from *Clostridium cellulolyticum*. *J. Bacteriol.* **173,** 7956–7962 CrossRef Medline

26. Liberato, M. V., Silveira, R. L., Prates, É. T., de Araujo, E. A., Pellegrini, V. O. A., Camilo, C. M., Kadowaki, M. A., Neto, M. D O., Popov, A., Skaf, M. S., and Polikarpov, I. (2016) Molecular characterization of a family 5 glycoside hydrolase suggests an induced-fit enzymatic mechanism. *Sci. Rep.* **6,** 36428 CrossRef Medline

27. Naas, A. E., MacKenzie, A. K., Dalhus, B., Eijsink, V. G. H., and Pope, P. B. (2015) Structural features of a Bacteroidetes-affiliated cellulase linked with a polysaccharide utilization locus. *Sci. Rep.* **5,** 11666 CrossRef Medline

28. Davies, G., and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. *Structure* **3,** 853–859 CrossRef Medline

29. Freier, D., Mothershed, C. P., and Wiegel, J. (1988) Characterization of *Clostridium thermocellum* Jw20. *Appl. Environ. Microb.* **54,** 204–211 CrossRef Medline

30. Desvaux, M. (2005) *Clostridium cellulolyticum*: model organism of mesophilic cellulolytic clostridia. *FEMS Microbiol. Rev.* **29,** 741–764 CrossRef Medline

31. Tamaru, Y., Miyake, H., Kuroda, K., Nakanishi, A., Kawade, Y., Yamamoto, K., Uemura, M., Fujita, Y., Doi, R. H., and Ueda, M. (2010) Genome sequence of the cellulosome-producing mesophilic organism *Clostridium cellulovorans* 743B. *J. Bacteriol.* **192,** 901–902 CrossRef Medline

32. Nölling, J., Breton, G., Omelchenko, M. V., Makarova, K. S., Zeng, Q. D., Gibson, R., Lee, H. M., Dubois, J., Qiu, D. Y., Hitti, J., Wolf, Y. I., Tatusov, R. L., Sabathe, F., Doucette-Stamm, L., Soucaille, P., *et al.* (2001) Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* **183,** 4823–4838 CrossRef Medline

33. Varel, V. H. (1989) Reisolation and characterization of *Clostridium longisporum*, a ruminal sporeforming cellulolytic anaerobe. *Arch. Microbiol.* **152,** 209–214 CrossRef Medline

34. Julliand, V., de Vaux, A., Millet, L., and Fonty, G. (1999) Identification of *Ruminococcus flavefaciens* as the predominant cellulolytic bacterial species of the equine cecum. *Appl. Environ. Microb.* **65,** 3738–3741 CrossRef Medline

35. Lan, P. T., Sakamoto, M., Sakata, S., and Benno, Y. (2006) *Bacteroides barnesiae* sp. nov., *Bacteroides salanitronis* sp. nov., and *Bacteroides gallinarum* sp. nov., isolated from chicken caecum. *Int. J. Syst. Evol. Microbiol.* **56,** 2853–2859 CrossRef Medline

36. Findley, S. D., Mormile, M. R., Sommer-Hurley, A., Zhang, X. C., Tipton, P., Arnett, K., Porter, J. H., Kerley, M., and Stacey, G. (2011) Activity-based metagenomic screening and biochemical characterization of bovine ruminal protozoan glycoside hydrolases. *Appl. Environ. Microb.* **77,** 8106–8113 CrossRef Medline

37. Chassard, C., Delmas, E., Robert, C., Lawson, P. A., and Bernalier-Donadille, A. (2012) *Ruminococcus champanellensis* sp nov., a cellulose-degrading bacterium from human gut microbiota. *Int. J. Syst. Evol. Micr.* **62,** 138–143 CrossRef Medline

38. Dai, X., Zhu, Y. X., Luo, Y. F., Song, L., Liu, D., Liu, L., Chen, F. R., Wang, M., Li, J. B., Zeng, X. W., Dong, Z. Y., Hu, S. N., Li, L. Y., Xu, J., Huang, L., *et al.* (2012) Metagenomic insights into the fibrolytic microbiome in yak rumen. *PLoS ONE* **7,** e40430 CrossRef Medline

39. Suhre, K., and Sanejouand, Y. H. (2004) ElNemo: a normal mode web server for protein movement analysis and the generation of templates for molecular replacement. *Nucleic Acids Res.* **32,** W610–W614 CrossRef Medline

40. 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

41. Wong, D. D., Chan, V. J., McCormack, A. A., and Batt, S. B. (2010) A novel xyloglucan-specific endo-β-1,4-glucanase: biochemical properties and

inhibition studies. *Appl. Microbiol. Biotechnol.* **86,** 1463–1471 CrossRef Medline

42. Matsuzawa, T., and Yaoi, K. (2016) GH74 xyloglucanases: structures and modes of activity. *Trends Glycosci. Glyc.* **28,** E63–E70 CrossRef

43. Deng, K., Takasuka, T. E., Heins, R., Cheng, X., Bergeman, L. F., Shi, J., Aschenbrener, R., Deutsch, S., Singh, S., Sale, K. L., Simmons, B. A., Adams, P. D., Singh, A. K., Fox, B. G., and Northen, T. R. (2014) Rapid kinetic characterization of glycosyl hydrolases based on oxime derivatization and nanostructure-initiator mass spectrometry (NIMS). *ACS Chem. Biol.* **9,** 1470–1479 CrossRef Medline

44. Deng, K., Takasuka, T. E., Bianchetti, C. M., Bergeman, L. F., Adams, P. D., Northen, T. R., and Fox, B. G. (2015) Use of nanostructure-initiator mass spectrometry to deduce selectivity of reaction in glycoside hydrolases. *Front. Bioeng. Biotechnol.* **3,** 165 CrossRef Medline

45. Garcia-Vallvé, S., Romeu, A., and Palau, J. (2000) Horizontal gene transfer of glycosyl hydrolases of the rumen fungi. *Mol. Biol. Evol.* **17,** 352–361 CrossRef Medline

46. Venditto, I., Luis, A. S., Rydahl, M., Schückel, J., Fernandes, V. O., Vidal-Melgosa, S., Bule, P., Goyal, A., Pires, V. M., Dourado, C. G., Ferreira, L. M., Coutinho, P. M., Henrissat, B., Knox, J. P., Baslé, A., *et al.* (2016) Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 7136–7141 CrossRef Medline

47. Raman, B., Pan, C., Hurst, G. B., Rodriguez, M., Jr., McKeown, C. K., Lankford, P. K., Samatova, N. F., and Mielenz, J. R. (2009) Impact of pretreated Switchgrass and biomass carbohydrates on *Clostridium thermocellum* ATCC 27405 cellulosome composition: a quantitative proteomic analysis. *PLoS ONE* **4,** e5271 CrossRef Medline

48. Hudson, K. L., Bartlett, G. J., Diehl, R. C., Agirre, J., Gallagher, T., Kiessling, L. L., and Woolfson, D. N. (2015) Carbohydrate-aromatic interactions in proteins. *J. Am. Chem. Soc.* **137,** 15152–15160 CrossRef Medline

49. Tatko, C. (2008) Sugars stack up. *Nat. Chem. Biol.* **4,** 586–587 CrossRef Medline

50. Asensio, J. L., Ardá, A., Cañada, F. J., and Jiménez-Barbero, J. (2013) Carbohydrate-aromatic interactions. *Acc. Chem. Res.* **46,** 946–954 CrossRef Medline

51. Northrop, D. B. (1998) On the meaning of *Km* and *V/K* in enzyme kinetics. *J. Chem. Educ.* **75,** 1153–1157 CrossRef

52. Himmel, M. E., Ruth, M. F., and Wyman, C. E. (1999) Cellulase for commodity products from cellulosic biomass. *Curr. Opin. Biotechnol.* **10,** 358–364 CrossRef Medline

53. Rahikainen, J. L., Evans, J. D., Mikander, S., Kalliola, A., Puranen, T., Tamminen, T., Marjamaa, K., and Kruus, K. (2013) Cellulase-lignin interactions-the role of carbohydrate-binding module and pH in nonproductive binding. *Enzyme Microb. Technol.* **53,** 315–321 CrossRef Medline

54. Koivula, A., Kinnari, T., Harjunpää, V., Ruohonen, L., Teleman, A., Drakenberg, T., Rouvinen, J., Jones, T. A., and Teeri, T. T. (1998) Tryptophan 272: an essential determinant of crystalline cellulose degradation by *Trichoderma reesei* cellobiohydrolase Cel6A. *FEBS Lett.* **429,** 341–346 CrossRef Medline

55. Harjunpaa, V., Teleman, A., Koivula, A., Ruohonen, L., Teeri, T. T., Teleman, O., and Drakenberg, T. (1996) Cello-oligosaccharide hydrolysis by cellobiohydrolase II from *Trichoderma reesei*: association and rate constants derived from an analysis of progress curves. *Eur. J. Biochem.* **240,** 584–591 CrossRef Medline

56. Gao, D., Chundawat, S. P., Sethi, A., Balan, V., Gnanakaran, S., and Dale, B. E. (2013) Increased enzyme binding to substrate is not necessary for more efficient cellulose hydrolysis. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 10922–10927 CrossRef Medline

57. Xue, S., Uppugundla, N., Bowman, M. J., Cavalier, D., Da Costa Sousa, L., E Dale, B., and Balan, V. (2015) Sugar loss and enzyme inhibition due to oligosaccharide accumulation during high solids-loading enzymatic hydrolysis. *Biotechnol. Biofuels* **8,** 195 CrossRef Medline

58. Andrić, P., Meyer, A. S., Jensen, P. A., and Dam-Johansen, K. (2010) Reactor design for minimizing product inhibition during enzymatic lignocellulose hydrolysis: I. Significance and mechanism of cellobiose and glucose inhibition on cellulolytic enzymes. *Biotechnol. Adv.* **28,** 308–324 CrossRef Medline

59. Fort, S., Varrot, A., Schülein, M., Cottaz, S., Driguez, H., and Davies, G. J. (2001) Mixed-linkage cellooligosaccharides: a new class of glycoside hydrolase inhibitors. *Chembiochem* **2,** 319–325 CrossRef Medline

60. Annamalai, N., Rajeswari, M. V., and Balasubramanian, T. (2016) Endo-1,4-β-glucanases: Role, applications and recent developments. In *Microbial Enzymes in Bioconversions of Biomass* (Gupta, V. K., ed) pp. 37–45, Springer International Publishing, Chem, Switzerland

61. Peng, X., Su, H., Mi, S., and Han, Y. (2016) A multifunctional thermophilic glycoside hydrolase from *Caldicellulosiruptor owensensis* with potential applications in production of biofuels and biochemicals. *Biotechnol. Biofuels* **9,** 98 CrossRef Medline

62. Srivastava, N., Srivastava, M., Mishra, P. K., Gupta, V. K., Molina, G., Rodriguez-Couto, S., Manikanta, A., and Ramteke, P. W. (2018) Applications of fungal cellulases in biofuel production: advances and limitations. *Renew. Sust. Energ. Rev.* **82,** 2379–2386 CrossRef

63. Studier, F. W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expr. Purif.* **41,** 207–234 CrossRef Medline

64. Blommel, P. G., Becker, K. J., Duvnjak, P., and Fox, B. G. (2007) Enhanced bacterial protein expression during auto-induction obtained by alteration of lac repressor dosage and medium composition. *Biotechnol. Progr.* **23,** 585–598 CrossRef Medline

65. Fox, B. G., and Blommel, P. G. (2009) Autoinduction of protein expression. *Curr. Protoc. Protein Sci.* Chapter 5, Unit 5.23

66. Ai, H. W., Shaner, N. C., Cheng, Z., Tsien, R. Y., and Campbell, R. E. (2007) Exploration of new chromophore structures leads to the identification of improved blue fluorescent proteins. *Biochemistry* **46,** 5904–5910 CrossRef Medline

67. Page, R., Grzechnik, S. K., Canaves, J. M., Spraggon, G., Kreusch, A., Kuhn, P., Stevens, R. C., and Lesley, S. A. (2003) Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the *Thermotoga maritima* proteome. *Acta Crystallogr. D Biol. Crystallogr.* **59,** 1028–1037 CrossRef Medline

68. Gorrec, F. (2016) Protein crystallization screens developed at the MRC Laboratory of Molecular Biology. *Drug Discov. Today* **21,** 819–825 CrossRef Medline

69. Newman, J., Egan, D., Walter, T. S., Meged, R., Berry, I., Ben Jelloul, M., Sussman, J. L., Stuart, D. I., and Perrakis, A. (2005) Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG+ strategy. *Acta Crystallogr. D Biol. Crystallogr.* **61,** 1426–1431 CrossRef Medline

70. Kabsch, W. (2010) XDS. *Acta Crystallogr D Biol. Crystallogr.* **66,** 125–132 CrossRef Medline

71. Diederichs, K., McSweeney, S., and Ravelli, R. B. G. (2003) Zero-dose extrapolation as part of macromolecular synchrotron data reduction. *Acta Crystallogr. D Biol. Crystallogr.* **59,** 903–909 CrossRef Medline

72. Liebschner, D., Afonine, P. V., Baker, M. L., Bunkóczi, G., Chen, V. B., Croll, T. I., Hintze, B., Hung, L. W., Jain, S., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Poon, B. K., Prisant, M. G., Read, R. J., *et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* **75,** 861–877 CrossRef Medline

73. Terwilliger, T. (2004) SOLVE and RESOLVE: automated structure solution, density modification, and model building. *J. Synchrotron Radiat.* **11,** 49–52 CrossRef Medline

74. 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

75. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68,** 352–367 CrossRef Medline

76. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276,** 307–326

77. Winter, G. (2010) xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Crystallogr.* **43,** 186–190 CrossRef

78. Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K., and Evans, G. (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D Struct. Biol.* **74,** 85–97 CrossRef Medline

79. Deng, K., Guenther, J. M., Gao, J., Bowen, B. P., Tran, H., Reyes-Ortiz, V., Cheng, X., Sathitsuksanoh, N., Heins, R., Takasuka, T. E., Bergeman, L. F., Geertz-Hansen, H., Deutsch, S., Loque, D., Sale, K. L., *et al.* (2015) Development of a high throughput platform for screening glycoside hydrolases based on oxime-NIMS. *Front. Bioeng. Biotechnol.* **3,** 153 CrossRef Medline

80. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22,** 2577–2637 CrossRef Medline

81. Joosten, R. P., Te Beek, T. A., Krieger, E., Hekkelman, M. L., Hooft, R. W., Schneider, R., Sander, C., and Vriend, G. (2011) A series of PDB related databases for everyday needs. *Nucleic Acids Res.* 39, D411–D419 CrossRef Medline

ASBMB

J. Biol. Chem. (2020) 295(51) 17752–17769 **17769**