Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

The continuing expansion of CAZymes and their families

Marie-Line Garron[1,2] and Bernard Henrissat[1,2]

Carbohydrate-active enzymes (CAZymes) catalyze the assembly and breakdown of glycans and glycoconjugates. Some have been discovered, studied and exploited for numerous applications long ago. For instance, amylase and invertase were isolated in the second half of the 19th century and lysozyme was the first enzyme whose 3-D structure was determined. In spite of this early start, the number of families of carbohydrate-active enzymes continues to grow steadily in the early 21st century. This review examines the CAZyme families reported during the last two years and posits that the current expansion will continue in the future, progressively uncovering the massive diversity of glycans whose breakdown requires a large diversity of bespoke enzymes.

Addresses
[1] Architecture et Fonction des Macromolécules Biologiques (AFMB), CNRS, Université Aix-Marseille, 163 Avenue de Luminy, 13288, Marseille, France
[2] INRA, USC 1408 AFMB, 13288, Marseille, France

Corresponding author:
Henrissat, Bernard (bernard.henrissat@afmb.univ-mrs.fr)

Current Opinion in Chemical Biology 2019, **53**:82–87

This review comes from a themed issue on **Mechanistic biology**

Edited by **Hermen S Overkleeft** and **David J Vocadlo**

https://doi.org/10.1016/j.cbpa.2019.08.004

Introduction

Carbohydrate oligomers and polymers represent the most abundant photosynthesis-fixed carbon on the planet, and these biomolecules are utilized by almost all living organisms whether autotrophic or heterotrophic for energy, structure, and a plethora of intra and intercellular signalling cascades including host-pathogen interactions. Carbohydrate-active enzymes (CAZymes) designate several classes of enzymes that assemble or breakdown oligosaccharides and polysaccharides. Because of the multiple roles of carbohydrates in nature, these enzymes find numerous applications in health, nutrition, and biotechnology. This review covers the recently described families of CAZymes that cleave glycosidic bonds by hydrolysis, phosphorolysis, oxidation or β-elimination, viz. those families that degrade glycans. Their sequence-based classification started in 1991 for the glycoside hydrolases (GHs) [1] and was subsequently extended to other CAZyme classes. The classification of CAZymes is presented since 1998 in the continuously updated CAZy database (www.cazy.org; [2]). For degradative enzymes, the classification currently describes a total of 215 families for 680 000 sequences (Table 1), a number that increases exponentially due to systematic genome sequencing. Figure 1 presents the growth of the largest category of degradative CAZymes, the GHs. Strikingly the number of GHs with 3-D structures grows much slower and currently represents less than 0.25% of the number of GH sequences present in the CAZy database. Because the CAZy database only lists proteins that have been released as finished entries in the daily releases of GenBank [2], the ratio of structures/sequences is even much smaller if one considers the number of GH sequences present in the non-redundant protein sequence databank of the NCBI.

The CAZy families group together enzymes that can have different specificity but share a common fold, a common catalytic machinery and the same mechanism, providing useful predictive power on the orientation of the glycosidic bond cleaved and potential transglycosylation side-reaction [3]. In 1995 Davies and Henrissat had (correctly) predicted that the 52 GH families that were known in 1995 would soon see their 3-D structure determined [4]. What was not anticipated was that the number of families would continue to grow by about four new GH families per year (Figure 2), progressively uncovering a massive enzyme diversity that matches the diversity of the carbohydrate structures they must break down. The exploration of the polysaccharide utilization loci (PULs) that occur in bacteria of the Bacteroidetes phylum [5] and the discovery of the enzymatic oxidative cleavage of polysaccharides [6] have largely fuelled the discovery of novel degradative CAZymes during the last two years. During the period covered by the present review, families GH137-156, PL 27-29 and AA14-15 have been added to the CAZy database (Table 2). Families GH150, GH151, GH152 and PL28 are not included here as they were described before the reviewing period and their addition to CAZy was simply overdue. Finally, family GH155 which was originally reported as a new family of α-L-arabinofuranosidases involved in arabinogalactan degradation has recently been deleted.

PULs for the discovery of new activities and new CAZyme families

A spectacular project, which allowed to identify seven new GH families, was performed by Ndeh *et al.* in the Gilbert laboratory on the depolymerisation of type II rhamnogalacturonan (RG-II) [7••]. RG-II, a pectin

Table 1

The families of degradative enzymes in the CAZy database as of January 2019

Category	Families	Number of sequences in families[c]	Number of PDB structures in families
Glycoside hydrolases (GHs)[a]	GH1 – GH156	593 529	1391
Polysaccharide lyases (PLs)	PL1 – PL29	17 249	77
Lytic polysaccharide mono-oxygenases (LPMOs)[b]	AA9, AA10, AA11, AA13 – AA15	5166	39

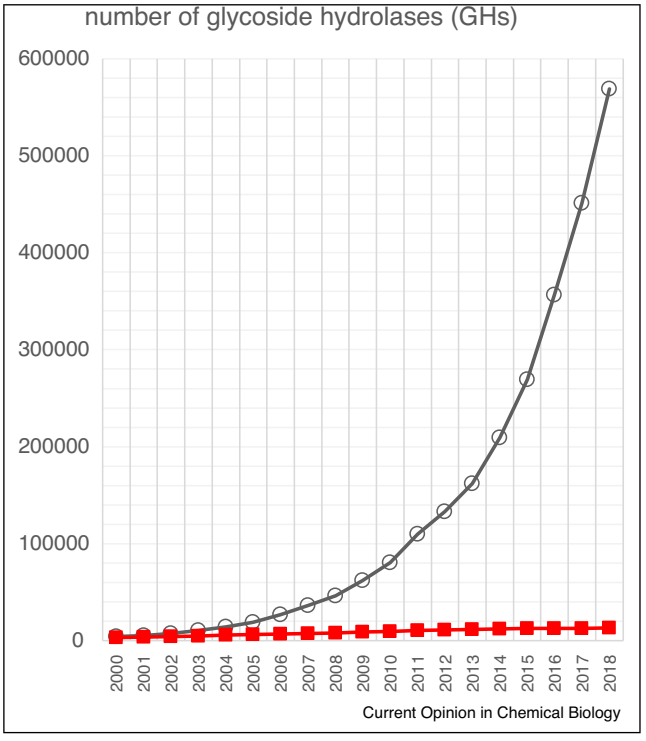
[a] Glycoside phosphorylases are found either in GH or in GT families depending on their evolutionary history.
[b] LPMOs are found in the auxiliary activity category of CAZy that also lists a number of lignolytic enzymes.
[c] As of January 2019; note that the CAZy database only lists proteins that have been released as finished entries in the daily releases of GenBank [2].

component, is a highly complex polysaccharide whose complete degradation requires the cleavage of 21 different glycosidic bond. Ndeh *et al.* have identified three PULs study, some are specific of rare linkages mainly found in RG-II as GH140 (apiosidase), GH141 (α-2-*O*-methyl-L-fucosidase) and GH143 (2-keto-3-Deoxy-D-Lyxo-heptulosaric acid hydrolase) whereas the other new families cleave more 'common' linkages. Thus, families GH137 and GH142 contain β-L-arabinofuranosidases while families GH138 and GH141 contain α-galacturonidases and α-L-fucosidases, respectively. In addition to the new families, previously known families were enriched by newly discovered activities. Thus, two novel activities were found for family GH2 enzymes (β-D-galacturonidase and α-arabinopyranosidase) and a GH127 family member was characterized as an aceric acid hydrolase.

In another piece of work from the Gilbert laboratory, Luis *et al.* have identified four PULs involved in the depolymerization of two other pectin components: homogalacturonan and type I rhamnogalacturonan [8•]. The characterization of the enzymes encoded by these PULs revealed two new GH families: GH146 and GH147, based on the characterization of a novel β-L-arabinofuranosidase and of a novel β-galactosidase, both involved in the breakdown of RG-I. Interestingly, GH147 was found in *Bacteroides ovatus* but is not conserved in the equivalent PUL of *Bacteroides thetaiotaomicron*, where the β-galactosidase activity may be compensated by another β-galactosidase of family GH2 found in the Gal-PUL [8•].

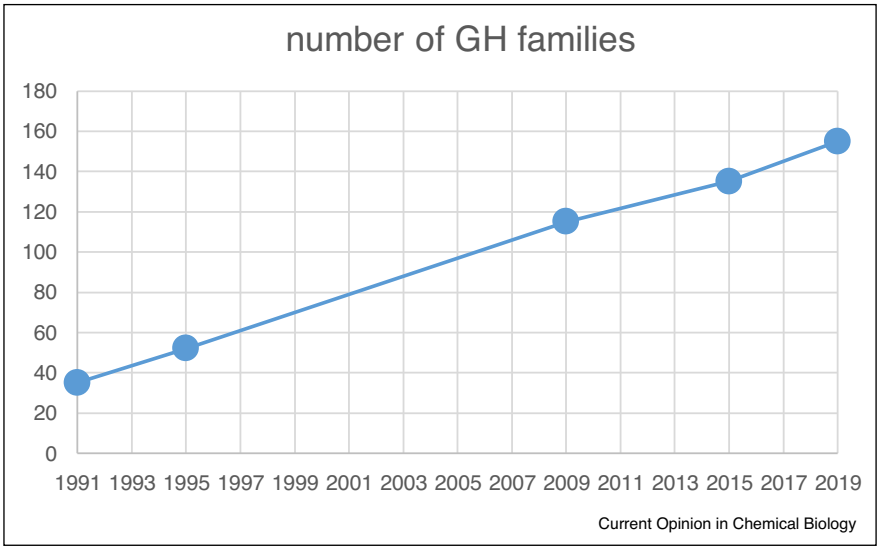
The detailed investigation of the degradation of the glycan part of arabinogalactan proteins (AGPs) has allowed to discover three novel CAZymes families, GH145, GH154 and PL27 [9,10•,11]. These families catalyze the depolymerisation of the specific side chains found in gum arabic AGP (GA-AGP). The non-reducing end of the side chains is made of Rha *p*-α1,4-GlcA which is cleaved by the GH145 α-L-rhamnosidase or by the PL27 L-rhamnose-α-1,4-D-glucuronate lyase, depending of the upregulated PUL. In PULs coding for the GH145 enzyme, the degradation of the GA-AGP decoration is completed by a member of family GH154, which is a novel β-glucuronidase family. Three-dimensional structures for representatives of families GH145 and PL27 have been solved and all show folds commonly found in CAZymes families, such as β-propeller for family GH145 and $(\alpha/\alpha)_6$ barrel for family PL27. Interestingly, despite a classical β-propeller fold, the structures of the GH145 enzymes reveal an unusual active site. Munoz-Munoz *et al.* have found that the active site is not located on the anterior face as expected from the β-propeller fold, but on the opposite face [9]. Site-directed mutagenesis confirmed this unusual active site location and a singular catalytic machinery involving a histidine residue was proposed. Surprisingly, the 'normal' (anterior) side of the β-propeller bears the highest residue conservation in the family and may well have another function. The

Figure 1



Growth of the number of GHs in the CAZy database. Black: number of sequences; red: number of biochemically characterized entries.

Figure 2



Growth of the number of GH families in the CAZy database.

latter is unknown but strong resemblance to family PL25 suggests a polysaccharide lyase activity [9].

The number of polysaccharides cleavable by PLs is restricted by the necessary presence of a uronic acid that can undergo the characteristic β-elimination mechanism of PLs. In addition to the PL27 family described above, family PL29 was characterized by Ndeh *et al.* in a *B. thetaiotaomicron* PUL involved in chondroitin sulfate degradation [12]. The enzymatic characterization shows that PL29, from *B. thetaiotaomicron*, is able to cleave chondroitin sulfate, but also dermatan sulfate and hyaluronic acid. Such a broad activity against various glycosaminoglycans was already observed in family PL8 [13].

New GH families found by other methods

The founding member of family GH149 family was identified after a proteomic investigation of the photosynthetic eukaryote *Euglena gracilis* [14]. Recombinant protein production and functional characterization demonstrated that the enzyme cleaves β-1,3-glucans by phosphorolysis. Distant sequence similarities and the conservation of important amino acids allowed to group family GH149 with family GH94, another phosphorylase family, in clan GH-Q (Table 2). The phylogenetic study performed by Kuhaudomlarp showed that family GH149 members are sometimes encountered in bacterial genomes in gene clusters. In particular, in the *Bacteroidetes* phylum, the family GH149 member is found in a conserved PUL that also encodes enzymes of families GH16, GH17 and GH30. This enzyme composition suggests that this PUL targets complex β-1,3-glucans [14].

Families GH148 and GH156 were discovered by functional metagenomics studies aimed at identifying novel thermostable enzymes which could be useful for applications [15,16•]. Family GH148 was createad after the characterization of a novel β-1,3-glucanase acting on β-1,3-1,4-glucans. Distant relatedness and conserved catalytic residues suggest that family GH148 belongs to clan GH-A, the largest structural clan in the CAZy database (Table 2). Interestingly, the analysis of family GH148 sequences shows that the (β/α)8 catalytic domain is interrupted by an inserted CBM4 domain. An insertion is always found at this position in family GH148 sequences although the size and the composition of the insert varies. Insertions in (β/α)8 barrels have been observed in other glycosidase families such as families GH13 and GH42. Family GH156 comprises exosialidases that act with an inverting mechanism that makes them stand out in comparison to all other sialidase families which all operate by a retaining mechanism. The cloned fosmid DNA sequence that served to isolate the first GH156 enzyme shows that the sialidase ORF is part of a gene cluster that suggests coordinated action with family GH20 (β-hexosaminidase) and GH29 (α-L-fucosidase) enzymes [16•].

Several Gram negative bacteria secrete polysaccharides which can serve as biofilms or virulence factors. The quest of enzymes that target these polysaccharides led to the discovery of families GH144 and GH153. Abe *et al.* characterized the first member of family GH144 and showed that the enzyme from *Chitinophaga arvensicola* cleaves β-1,2-glycan by an inverting mechanism [17•].

Table 2

The 21 families of degradative CAZymes reported during the last two years

Family	Mechanism	Activities	Fold	PDB	Clan
GH137	n.d.	β-L-arabinofuranosidase	β propeller (5 blades)	5MT2 [7••]	
GH138	Retaining	α-galacturonidase	$(\beta/\alpha)_8$	6HZF [30]	
GH139	n.d.	α-2-*O*-Me-L-fucosidase			
GH140	Retaining	Apiosidase	$(\beta/\alpha)_8$	5MSY [7••]	
GH141	n.d.	α-L-fucosidase, xylanase	β-helix	5MQP	
GH142	n.d.	β-L-arabinofuranosidase	$(\alpha/\alpha)_6$	5MQR [7••]	
GH143	Retaining (inferred)	2-keto-3-deoxy-D-lyxo-heptulosaric acid hydrolase	β propeller (5 blades)	5MQR [7••]	
GH144	Inverting	Endo-β-1,2-glucanase	$(\alpha/\alpha)_6$	5GZK [17•]	
GH145	Retaining	L-Rha-α-1,4-GlcA α-L-rhamnohydrolase	β propeller (7 blades)	5MVH [9]	
GH146	Retaining	β-L-arabinofuranosidase	$(\alpha/\alpha)_6$	5OPJ [8•]	GH-P
GH147	Retaining (inferred)	β-galactosidase	$(\beta/\alpha)_8$ (inferred)		GH-A (inferred)
GH148	Retaining (inferred)	β-1,3-glucanase	$(\beta/\alpha)_8$ (inferred)		GH-A (inferred)
GH149	Inverting	β-1,3-glucan phosphorylase	$(\alpha/\alpha)_6$	6HQ6 [31]	GH-Q
GH153	n.d.	poly-β-1,6-D-glucosamine hydrolase	$(\beta/\alpha)_8$	6AU1 [18•]	
GH154	n.d.	β-glucuronidase			
GH156	Inverting	Exo-α-sialidase			
PL27	β-elimination	L-rhamnose-α-1,4-D-glucuronate lyase	$(\alpha/\alpha)_6$	5NO8 [11]	
PL29	β-elimination	Hyaluronate lyase chondroitin-sulfate ABC endolyase dermatan sulfate lyase			
AA14	C-1 oxidation	Lytic xylan monooxygenase	Antiparallel β-sandwich	5NO7 [22•]	
AA15	C-1 oxidation	Lytic cellulose monooxygenase lytic chitin monooxygenase	Antiparallel β-sandwich	5MSZ [23•]	

In nature β-1,2-glucan can be observed either in linear or cyclic forms. Detailed characterization of the GH144 enzyme from *C. arvensicola* showed that it is able to cleave the two glycans forms. This observation was supported by the crystallographic structure that showed a large active site allowing the accommodation of at least 7 β-glucose residues. Little *et al.* have characterized two members of the new GH153 family, one from *Bortella bronchiseptica* and the other from *Escherichia coli* [18•]. Their phylogenetic analysis shows that GH153 enzymes are always bimodular and with a deacetylase domain on the same polypeptide. These bimodular enzymes are able to depolymerize poly-β-1,6-*N*-acetyl-D-glucosamine (PNAG) which is a major component of biofilms.

Novel lytic polysaccharide mono-oxygenase (LPMO) families

The oxidative cleavage of polysaccharides was first described by Vaaje-Kolstad *et al.* on chitin [6]. Quinlan and colleagues have shown that this was catalyzed by copper-dependent oxidative enzymes collectively called LPMOs [19,20]. The discovery of the oxidative cleavage of polysaccharides by enzymes has generated a lot of excitement as their surface exposed catalytic machinery of LPMOs can function without requiring the isolation of a single polymer chain. Thus some LPMOs are thought to be able to initiate crystalline cellulose cleavage and thereby explain why they boost the activity of lignocellulolytic enzyme cocktails [21].

Initially described to act on cellulose, chitin and starch, the action of LPMOs on xylan was rarely described until the discovery of family AA14 [22•]. Two AA14 enzymes from basidiomycete fungus *Pycnoporus coccineus* were functionally characterized as copper-dependent LPMOs that cleave xylan with oxidation at C-1. The three-dimensional structure shows that AA14 adopts the same fold as seen in other LPMO families, suggesting that all LPMOs may have a single common ancestor. While the LPMOs were thought to be restricted to the microbial world (fungi, bacteria and viruses), Sabbadin and colleagues have found a novel LPMO family, AA15, in the arthropod *Thermobia domestica* [23•]. Two enzymes from *T. domestica* were characterised as copper-dependent LPMOs that oxidatively cleave cellulose and chitin by C-1 oxidation. The phylogenetic study of family AA15 shows that it is widespread in marine and terrestrial invertebrates (crustaceans, molluscs, insects, millipedes, and spiders) as well as in groups of algae and oomycetes, not only expanding considerably the taxa where LPMOs can be found, but suggesting that animal LPMOs could be involved in the remodelling of endogenous chitin scaffolds during development and metamorphosis.

How long will the discovery of novel CAZyme families continue and why are there so many families of CAZymes?

The diversity of monosaccharides, the large variety of possible linkages, the modification of carbohydrates by sulfation or acetylation, the length and the form adopted by the polysaccharides (linear, circular, more or less decorated . . .) provide glycans with the widest diversity of all biomolecules. This broad diversity of substrates is correlated by the large number of functions covered by

CAZyme families. Even if some folds are found more frequently than others, CAZyme families display an important structural diversity and frequently reveal surprises such as the unexpected location of the active site in family GH145 and its uncommon architecture. Moreover, because CAZy families often group together enzymes that act on different substrates, the direct prediction of specificity based on family assignment can be very difficult and error-prone. Regrettably, generalist genome annotators are not aware of this and, as a result, general sequence databases are polluted by large amounts of wrong predicted functions [24,25]. The subdivision of large multifunctional families into smaller subfamilies has shown that the CAZy subfamilies display a much narrower set of activities, making them particularly powerful for functional prediction [26–28]. The subfamilies thus show that CAZymes carry a strong functional signal that makes them more powerful than other classes of enzymes to derive functional information from genomes and metagenomics datasets. Unfortunately, not all CAZyme subfamilies have a biochemically characterized representative and many families are waiting to be divided in stable subfamilies — a challenging task due to the number of sequences to analyze and to the necessary stability in subfamily definitions. The discovery of CAZyme families is essential to this long-term goal as knowledge-based functional predictions require that all CAZyme families are identified. Thus for CAZymes, bioinformatics and wet biochemistry form a virtuous partnership where the former can help choose CAZymes to study and the latter ultimately improve functional assignments in genomes and metagenomes.

The identification of the PULs in the *Bacteroidetes* phylum has boosted the recent discovery of new CAZyme families. This ingenious gene organization allows the bacterium to attack each linkage and totally disassemble a complex polysaccharide in a highly coordinated process. The particular genetic organization of the PULs simplifies their bioinformatics prediction, as shown in the PULBD extension of the CAZy database (http://www.cazy.org/PULDB/) [29]. The thousands of uncharacterized PULs in this resource are expected to help discover novel activities and CAZyme families in the future. We are confident that the sequence databases probably contain dozens of seeds for new families awaiting discovery and characterization. The continuously growing share of CAZymes in genomes slowly uncovers their importance as one of the fast developing fields of the early 21st century.

Conflict of interest statement

Nothing declared.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Henrissat B: **A classification of glycosyl hydrolases based on amino acid sequence similarities**. *Biochem J* 1991, **280**:309-316.

2. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B: **The carbohydrate-active enzymes database (CAZy) in 2013**. *Nucleic Acids Res* 2014, **42**:D490-D495.

3. Henrissat B, Davies G: **Structural and sequence-based classification of glycoside hydrolases**. *Curr Opin Struct Biol* 1997, **7**:637-644.

4. Davies G, Henrissat B: **Structures and mechanisms of glycosyl hydrolases**. *Struct Lond Engl 1993* 1995, **3**:853-859.

5. Martens EC, Koropatkin NM, Smith TJ, Gordon JI: **Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm**. *J Biol Chem* 2009, **284**:24673-24677.

6. Vaaje-Kolstad G, Westereng B, Horn SJ, Liu Z, Zhai H, Sørlie M, Eijsink VGH: **An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides**. *Science* 2010, **330**:219-222.

7. •• Ndeh D, Rogowski A, Cartmell A, Luis AS, Baslé A, Gray J, Venditto I, Briggs J, Zhang X, Labourel A *et al.*: **Complex pectin metabolism by gut bacteria reveals novel catalytic functions**. *Nature* 2017, **544**:65-70

In this seminal paper Ndeh *et al.* disclosed a whopping 7 new GH families during a tour de force that succeeded in dissecting completely the enzymatic degradation apparatus of *B. thetaiotaomicron* for type II rhamnogacalturonan, a component of pectin.

8. • Luis AS, Briggs J, Zhang X, Farnell B, Ndeh D, Labourel A, Baslé A, Cartmell A, Terrapon N, Stott K *et al.*: **Dietary pectic glycans are degraded by coordinated enzyme pathways in human colonic Bacteroides**. *Nat Microbiol* 2018, **3**:210-219

This work reports the β-L-arabinofuranosidase activity of BT0349 (founding member of family GH146) and the β-galactosidase activity of BACOVA_05493 on pectic galactan (founding member of family GH147).

9. Munoz-Munoz J, Cartmell A, Terrapon N, Henrissat B, Gilbert HJ: **Unusual active site location and catalytic apparatus in a glycoside hydrolase family**. *Proc Natl Acad Sci U S A* 2017, **114**:4936-4941.

10. • Cartmell A, Muñoz-Muñoz J, Briggs JA, Ndeh DA, Lowe EC, Baslé A, Terrapon N, Stott K, Heunis T, Gray J *et al.*: **A surface endogalactanase in *Bacteroides thetaiotaomicron* confers keystone status for arabinogalactan degradation**. *Nat Microbiol* 2018, **3**:1314-1326

This paper reports that *Bacteroides thetaiotaomicron* BT3677 is a β-glucuronidase, the founding member of GH154 and that BT3679 has α-L-arabinofuranosidase activity and establishes family GH155.

11. Munoz-Munoz J, Cartmell A, Terrapon N, Baslé A, Henrissat B, Gilbert HJ: **An evolutionarily distinct family of polysaccharide lyases removes rhamnose capping of complex arabinogalactan proteins**. *J Biol Chem* 2017, **292**:13271-13283.

12. Ndeh D, Munoz Munoz J, Cartmell A, Bulmer D, Wills C, Henrissat B, Gray J: **The human gut microbe *Bacteroides thetaiotaomicron* encodes the founding member of a novel glycosaminoglycan-degrading polysaccharide lyase family PL29**. *J Biol Chem* 2018, **293**:17906-17916.

13. Garron M-L, Cygler M: **Uronic polysaccharide degrading enzymes**. *Curr Opin Struct Biol* 2014, **28**:87-95.

14. Kuhaudomlarp S, Patron NJ, Henrissat B, Rejzek M, Saalbach G, Field RA: **Identification of *Euglena gracilis* β-1,3-glucan phosphorylase and establishment of a new glycoside hydrolase (GH) family GH149**. *J Biol Chem* 2018, **293**:2865-2876.

15. Angelov A, Pham VTT, Übelacker M, Brady S, Leis B, Pill N, Brolle J, Mechelke M, Moerch M, Henrissat B *et al.*: **A metagenome-derived thermostable β-glucanase with an unusual module architecture which defines the new glycoside hydrolase family GH148**. *Sci Rep* 2017, **7**:17306.

16. • Chuzel L, Ganatra MB, Rapp E, Henrissat B, Taron CH: **Functional metagenomics identifies an exosialidase with an inverting**

catalytic mechanism that defines a new glycoside hydrolase family (GH156). *J Biol Chem* 2018, **293**:18138-18150
Chuzel *et al.* have identified and characterized as an inverting exo-sialidase the founding member of family GH156 via a functional metagenomic screen of DNA from a thermal spring.

17. • Abe K, Nakajima M, Yamashita T, Matsunaga H, Kamisuki S, Nihira T, Takahashi Y, Sugimoto N, Miyanaga A, Nakai H *et al.*: **Biochemical and structural analyses of a bacterial endo-β-1,2-glucanase reveal a new glycoside hydrolase family**. *J Biol Chem* 2017, **292**:7487-7506
The authors have shown the endo-b-1,2-glucanase activity of a Chitinophaga pinensis enzyme, founding member of family GH144.

18. • Little DJ, Pfoh R, Le Mauff F, Bamford NC, Notte C, Baker P, Guragain M, Robinson H, Pier GB, Nitz M *et al.*: **PgaB orthologues contain a glycoside hydrolase domain that cleaves deacetylated poly-β(1,6)-*N*-acetylglucosamine and can disrupt bacterial biofilms**. *PLoS Pathog* 2018, **14**:e1006998
Little *et al.* have shown that *E. coli* PgaB has two domains with respectively poly-β-1,6-*N*-acetyl-D-glucosamine *N*-deacetylase and poly-β-1,6-D-glucosamine hydrolase; the latter is the founding member of family GH153, invariably associated to the deacetylase domain.

19. Quinlan RJ, Sweeney MD, Lo Leggio L, Otten H, Poulsen J-CN, Johansen KS, Krogh KBRM, Jørgensen CI, Tovborg M, Anthonsen A *et al.*: **Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components**. *Proc Natl Acad Sci U S A* 2011, **108**:15079-15084.

20. Beeson WT, Phillips CM, Cate JHD, Marletta MA: **Oxidative cleavage of cellulose by fungal copper-dependent polysaccharide monooxygenases**. *J Am Chem Soc* 2012, **134**:890-892.

21. Harris PV, Welner D, McFarland KC, Re E, Navarro Poulsen J-C, Brown K, Salbo R, Ding H, Vlasenko E, Merino S *et al.*: **Stimulation of lignocellulosic biomass hydrolysis by proteins of glycoside hydrolase family 61: structure and function of a large, enigmatic family**. *Biochemistry (Mosc)* 2010, **49**:3305-3316.

22. • Couturier M, Ladevèze S, Sulzenbacher G, Ciano L, Fanuel M, Moreau C, Villares A, Cathala B, Chaspoul F, Frandsen KE *et al.*: **Lytic xylan oxidases from wood-decay fungi unlock biomass degradation**. *Nat Chem Biol* 2018, **14**:306-310
In this paper, Couturier *et al.* show that family AA14 proteins are copper-dependent lytic polysaccharide monooxygenases (LPMOs); cleavage of xylan with oxidation of C-1 is demonstrated for two AA14 LPMOs from Trametes coccinea (*Pycnoporus coccineus*).

23. • Sabbadin F, Hemsworth GR, Ciano L, Henrissat B, Dupree P, Tryfona T, Marques RDS, Sweeney ST, Besser K, Elias L *et al.*: **An ancient family of lytic polysaccharide monooxygenases with roles in arthropod development and biomass digestion**. *Nat Commun* 2018, **9**:756
Sabbadin *et al.* have shown the copper-dependent LPMO activity on cellulose and chitin of two AA15 enzymes from *Thermobia domestica*. These are the first LPMOs from animal origin; they are speculated to play a role in remodelling endogenous chitin scaffolds during development and metamorphosis.

24. Jones CE, Brown AL, Baumann U: **Estimating the annotation error rate of curated GO database sequence annotations**. *BMC Bioinformatics* 2007, **8**:170.

25. Schnoes AM, Brown SD, Dodevski I, Babbitt PC: **Annotation error in public databases: misannotation of molecular function in enzyme superfamilies**. *PLoS Comput Biol* 2009, **5**:e1000605.

26. Stam MR, Danchin EGJ, Rancurel C, Coutinho PM, Henrissat B: **Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins**. *Protein Eng Des Sel PEDS* 2006, **19**:555-562.

27. Aspeborg H, Coutinho PM, Wang Y, Brumer H, Henrissat B: **Evolution, substrate specificity and subfamily classification of glycoside hydrolase family 5 (GH5)**. *BMC Evol Biol* 2012, **12**:186.

28. Mewis K, Lenfant N, Lombard V, Henrissat B: **Dividing the large glycoside hydrolase family 43 into subfamilies: a motivation for detailed enzyme characterization**. *Appl Environ Microbiol* 2016, **82**:1686-1692.

29. Terrapon N, Lombard V, Drula É, Lapébie P, Al-Masaudi S, Gilbert HJ, Henrissat B: **PULDB: the expanded database of polysaccharide utilization loci**. *Nucleic Acids Res* 2018, **46**:D677-D683.

30. Labourel A, Baslé A, Munoz-Munoz J, Ndeh D, Booth S, Nepogodiev SA, Field RA, Cartmell A: **Structural and functional analyses of glycoside hydrolase 138 enzymes targeting chain A galacturonic acid in the complex pectin rhamnogalacturonan II**. *J Biol Chem* 2019, **294**:7711-7721.

31. Kuhaudomlarp S, Stevenson CEM, Lawson DM, Field RA: **The structure of a GH149 β-(1→3) glucan phosphorylase reveals a new surface oligosaccharide binding site and additional domains that are absent in the disaccharide-specific GH94 glucose-β-(1→3)-glucose (laminaribiose) phosphorylase**. *Proteins* 2019, **87**:885-892 http://dx.doi.org/10.1002/prot.25745.