A hierarchical classification of polysaccharide lyases for glycogenomics

Vincent Lombard, Thomas Bernard, Corinne Rancurel, Harry Brumer, Pedro M Coutinho, Bernard Henrissat

A hierarchical classification of polysaccharide lyases for glycogenomics

V. Lombard*, T. Bernard*†, C. Rancurel*, H Brumer[‡], P.M. Coutinho* & B. Henrissat*[1]

*Architecture et Fonction des Macromolécules Biologiques, UMR6098, CNRS, Université de la Méditerranée, Université de Provence, Case 932, 163 Avenue de Luminy, 13288 Marseille cedex 9, France
[‡]School of Biotechnology, Royal Institute of Technology (KTH), AlbaNova University Centre, 106 91 Stockholm, Sweden
† Present address: Biométrie et Biologie Évolutive, UMR CNRS 5558, UCB Lyon 1, Bât. Grégor Mendel, 43 bd du 11 novembre 1918, 69622 Villeurbanne cedex, France
[1]To whom correspondence should be addressed: bernard.henrissat@afmb.univ-mrs.fr

Abstract:

Carbohydrate-active enzymes face huge substrate diversity in a highly selective manner with only a limited number of available folds. They are therefore subjected to multiple divergent and convergent evolutionary events. This and their frequent modularity render their functional annotation in genomes difficult in a number of cases. A classification of polysaccharide lyases (the enzymes that cleave polysaccharides using an elimination instead of a hydrolytic mechanism) is presented thoroughly for the first time. Based on the analysis of a large panel of experimentally characterized polysaccharide lyases, we examined the correlation of various enzyme properties with the three levels of the classification: fold, families and subfamilies. The resulting hierarchical classification, which should help annotate relevant genes in genomic efforts, is available and constantly updated at the Carbohydrate-Active Enzymes Database (www.cazy.org).

Keywords: Polysaccharide lyases, modular structure, catalytic mechanism, enzyme families, functional annotation

Introduction

Polysaccharide Lyases (PLs) are a group of enzymes (EC 4.2.2.-) that cleave uronic acid-containing polysaccharides via a β-elimination mechanism to generate an unsaturated hexenuronic acid residue and a new reducing end at the point of cleavage (Figure 1) [1]. PLs are ubiquitous in Nature, having been identified in organisms ranging from bacteriophages and Archaea to Eubacteria and higher eukaryotes, such as fungi, algae, plants, and mammals [2]. For all these organisms, polysaccharide lyases represent a complimentary mechanistic strategy to the glycoside hydrolases (GHs; EC 3.2.1.-) [3] for the breakdown of C-6 carboxylated polysaccharides (Figure 1), with the contrasting feature that PL-catalyzed cleavage occurs without intervention of a water molecule. PLs are implicit in diverse biochemical processes, including biomass degradation, tissue matrix recycling, and pathogenesis [2, 4-9]. Moreover, the widespread use of polyuronic acids in the food and

medical sectors makes PLs attractive as specific catalysts for the modification of substrates such as pectins, alginates, and heparins in biotechnological applications [2, 10-12].

The catalytic mechanism employed by PLs (Figure 2) can be broadly described as consisting of three events: (1) abstraction of the C-5 proton on the sugar ring of a uronic acid or ester by a basic amino acid sidechain, (2) stabilization of the resulting anion by charge delocalization into the C-6 carbonyl group, and (3) lytic cleavage of the O-4:C-4 bond, facilitated by proton donation from a catalytic acid, to yield a hexenuronic acid (or ester) moiety at the newly-formed non-reducing chain end [1, 13]. Depending on the monosaccharide composition of the substrate and its conformation in the PL active site, the proton removed from C-5 and departing oxygen on C-4 may lie either *syn* or *anti* to each other. This, in turn, imposes certain requirements on the position of active site groups and the possibilities for a stepwise or concerted elimination reaction (Figure 2A & B). Polysaccharide recognition in PLs is often dependent on the interaction of tightly held divalent cations (often Ca^{++}), or positively charged amino acid sidechains, with uronic acid groups in the substrate. Such cations may play an additional role in stabilizing the transient anion in the reaction pathway. The extent to which these molecular events are concerted, as well as the nature and individual contributions of the catalytic groups, in the mechanisms of specific enzymes have not been fully clarified, although significant advances have been made in a few cases (see [14, 15] and refs. therein). Detailed structural information on the catalytic modules of PLs is available in [16].

In common with GHs, PLs frequently have multi-modular structures, in which the catalytic module can be appended to a variable number of ancillary modules such as Carbohydrate Binding Modules (CBMs) [17, 18], other catalytic modules, or modules with other functions (see below). Interestingly, many non-catalytic modules borne by PLs may also be appended to GHs. Following a full dissection of their modular organization, we have grouped the PLs into amino acid sequence-based families to provide a framework for structural and mechanistic studies. Here we present a hierarchical classification of PLs including sub-families, families and clans/superfamilies and we discuss the value of these levels for genome mining and functional prediction. This classification is implemented in the Carbohydrate-Active Enzymes (CAZy) Database (www.cazy.org). [19].

Experimental

Included and excluded enzyme classes

For the purpose of this family classification, the scope of the term Polysaccharide Lyase is restricted to those enzymes which operate according to the general mechanisms described in Figures 1 and 2, to produce a terminal hexenuronic acid moiety by β-elimination. This is a clear distinction from the broader NC-IUBMB classification of carbon-oxygen lyases acting on polysaccharides into EC 4.2.2.- (http://www.chem.qmul.ac.uk/iubmb/enzyme/). In

particular, the following enzymes are not included PL classification described in CAZy, as they are structurally and mechanistically more similar to the Glycoside Hydrolases (GHs):

- exo-α-(1,4)-D-glucan lyases (EC 4.2.2.13) cleave malto-oligosaccharides to produce 1,5-anhydro-D-fructose without the intervention of a water molecule. These enzymes are structurally similar to GH31 α-glucoside hydrolases, with which they are currently classified. Analogous to other GH31 enzymes, the first step in the catalytic mechanism involves the formation of a covalent glycosyl-enzyme intermediate. However, in α-glucan lyases this intermediate decomposes through a *syn*-elimination mechanism, rather than hydrolysis [13, 20].

- Lytic transglycosylases (LT) cleave the β-(1,4)-glycosidic bond between the *N*-acetylmuramic acid and the *N*-acetylglucosamine residues of peptideglycan via a substrate participation mechanism, with no intervention of water, to yield a 1,6-anhydro sugar derivative [21]. LTs are structurally and mechanistically closely related to lysozymes and are currently classified in GH families GH23, GH102, GH103, and GH104 [22].

- Levan and inulin fructotransferases (EC 4.2.2.16, EC 4.2.217, EC 4.2.2.18) cleave fructo-oligosaccharides by intra-molecular attack to yield various anhydro-fructodisaccharides. These enzymes are presently classified into GH91, along with a sequence-similar enzyme that hydrolyzes the α-D-fructofuranose β-D-fructofuranose 1,2':2,3'-dianhydride (DFA III) product of the EC 4.2.2.18 inulin fructotransferase [23, 24]. As such, mechanistic commonality with GHs (and loosely with LTs) is predicted.

Family and subfamily groupings

The PL families were first built by searching sequence homologues of experimentally characterized enzymes. To avoid the creation of large number of families, distant homologues were assigned to existing families. These families have been presented online in the CAZy database since its launch in 1998 with the occasional creation of novel families subsequent to the experimental characterization of PLs with no or insufficient similarity to known families. Within families, subfamilies have been defined by procedures similar to that described for the large glycoside hydrolase family GH13, which is comprised of a diversity of starch-active enzymes of similar structure [25].

Briefly, in each family, the sequences were edited to isolate the catalytic domains to avoid interference from the presence or absence of additional modules. The catalytic domains were then subject to a multiple sequence alignment using MUSCLE [26] and a distance matrix was created using the BLOSUM62 [27] substitution model. The distance matrix was then used as input for an automatic analysis based on the SECATOR algorithm [28], which proposes the breakdown of the family into a number of subfamilies, based on a reconstructed phylogenetic tree. The robustness of the subfamilies was tested by a resampling approach whereby a proportion of the sequences were randomly removed from

the sample. The clustering procedure was iterated typically 10,000 times with random variations of the parameters of the automatic partitioning algorithm. The percentage of sequences removed from the sample was also picked randomly from 5 to 30% at each iteration. Sequences found in the same cluster over 80% of the time were assigned to the same subfamily. Finally, only subfamilies containing more than five members were retained in order to define significant subfamilies. Unassigned sequences will be subjected to a new round of analysis as more sequences become available.

Results and Discussion

Modular structure of PLs

Carbohydrate-active enzymes are frequently comprised of a modular structure, in which a catalytic module carries one or more ancillary modules [29]. Polysaccharide lyases are no exception and there is a large variety of multi-modular PLs (Figure 3). Perhaps the most common situation is the occurrence of one or more carbohydrate-binding modules in tandem with the catalytic PL module. However, other arrangements have been observed, such as the addition of domains that promote binding to other macromolecules, including SLH domains for cell attachment [30] or dockerin modules for cellulosome assembly [31]. Some PLs may even be arranged with an additional PL module or a complementary carbohydrate esterase (CE) module, as well as domains whose function is presently unknown (termed "X" modules, Figure 3). The number of possible combinations of domains is in principle unlimited, and their presence poses a specific challenge for sequence-based family grouping and annotation. Whole genome annotations are particularly prone to false identification (and subsequent misleading functional annotation) due to spurious hits on ancillary modules common to two distinct proteins. Consequently, a systematic excision of the ancillary modules was performed prior to all sequence alignments of PLs, and indeed this approach is the principal *modus operandi* of the CAZy classification [19, 32].

Families and folds

In April 1999 there were about one hundred PL sequences arranged in nine families [33]. Since then, the number of PL sequences has increased approx. 20-fold, essentially due to whole genome-sequencing projects. Thanks to the biochemical characterization of many novel PLs, the number of PL families has progressively grown over the years to reach 21 in 2010. The corresponding eleven years of structural biology have vastly expanded knowledge of the three-dimensional structures of PLs (for a thorough review on 3D-structure: function relationships of PLs see [16]): Whereas only one of the initial nine families of PLs had a structural representative in 1999, the fold of only two (PL12 and PL17) out of the 21 current PL families remain to be determined (Figure 4).

PLs show a large variety of fold types, ranging from β-helices to α/α barrels (Figure 4). The abundance of PL folds indicates that PLs have been invented more than once during evolution, from totally different scaffolds. The most extreme example of the convergent

evolution of PLs is perhaps with PL1 and PL10 pectate lyases, in which the different folds carry an identically poised catalytic machinery that performs the same reaction on the same substrate [34]. The plasticity of the active site of polysaccharide lyases to accommodate a variety of substrates is reminiscent of that of glycoside hydrolases [35]. Interestingly most of the PL folds are also found in GH families, an indication of possible common evolutionary origins between the two enzyme classes.

In addition to being well-characterized at the 3-D level, examination of the CAZy database (www.cazy.org) shows that more than 10% of the PLs in the database have been biochemically (kinetically) characterized, which is the highest proportion among all the classes of carbohydrate-active enzymes described in CAZy (GHs, GTs, PLs, and CEs). This wealth of biochemical data indicates that most PL families group enzymes with diverse substrate specificities (Table 1). This situation has been previously observed for other CAZyme classes, especially the glycoside hydrolases [32] and glycosyltransferases [36]. One likely explanation for this phenomenon is that the number of available protein folds is considerably smaller than the number of carbohydrate structures and hence Nature has adventitiously tuned existing scaffolds for exquisite substrate specificity.

Less immediately apparent, however, are the structural similarities among the various substrates processed by individual family members. As an example, family PL8 can be considered polyspecific, as it groups together three different enzyme activities: hyaluronate lyase (EC 4.2.2.1), xanthan lyase (EC 4.2.2.12) and chondroitin AC lyase (EC 4.2.2.5). Here, the common names of these polysaccharides belie the fact that these three types of enzymes act at the *same* position on the *same* sugar, i.e., they cleave the C-O bond at position 4 of unsubstituted glucuronic acid in the backbone (Figure 5). What differentiates the three substrates is the substituent attached to the 4-oxygen of the glucuronic acid, a situation similar to, for instance, glycoside hydrolases that exhibit aglycon specificity [37]. The three enzymes therefore can – and in the case of PL8 *do* – utilize an identical catalytic machinery to cleave their respective substrates.

Subfamilies

The functional prediction (viz. substrate specificity) of thousands of putative carbohydrate active enzymes derived from genome data is highly desirable, but requires a direct, unequivocal relationship between sequence groupings and substrate specificity. Because the sequence-based families of PLs generally do not correlate with the fine substrate specificity as described above, we have examined the definition of subfamilies to assess whether functional grouping and prediction could be improved. A similar approach was previously applied to the large polyspecific GH13 family of α-amylase-related enzymes, in which most of the sequence-derived subfamilies were indeed found to correspond to a single enzyme activity [25].

With the sequence data available to date, we were able to break down the 21 PL families into a total of 41 subfamilies covering 72% of all sequences analyzed (Table 1). The sequences that could not be assigned to subfamilies will most likely generate new subfamilies as more sequences become available in the future. The subfamilies are identified by an Arabic numeral following the family identifier; for instance, PL5_1 designates subfamily 1 within family PL5. As shown in Table 1, the vast majority of subfamilies have at least one representative that has been characterized with respect to substrate specificity; only 7 subfamilies are lacking an experimentally characterized member. Depending on the subfamily, the cumulated biochemical characterization data varies from low (e.g., subfamilies PL3_1 & PL4_2) to high (e.g., PL1_5, PL1_6, & PL5_1). These variations can have a profound effect on any subsequent functional predictions based on subfamily membership, since reliability obviously depends (i) on the number of characterized enzymes per subfamily and (ii) on how detailed and reliably each characterization was performed.

We observe that of the 41 subfamilies identified here, 37 (90%) appear monospecific, thus indicating that the subfamilies correlate with substrate specificity significantly better than the family level. Only three subfamilies remained apparently polyspecific (i.e. grouping enzymes with different EC numbers): PL1_5, PL9_1 & PL14_3. These three subfamilies were further inspected to identify the origin of their polyspecificity. In the case of subfamilies PL1_5 and PL9_1, the apparent polyspecificity is due to the presence of both endo-acting (EC 4.2.2.2) and exo-acting (EC 4.2.2.9) polygalacturonate lyases. These two types of enzymes have exactly the same substrate specificity and differ only in the degree of polymerization of the released products. As with other types of carbohydrate-cleaving enzymes, the basis of endo- vs. exo-activity within a family is typically due to subtle details in the three-dimensional structure of the enzymes, and rigidly distinguishing the two activities can be tricky [38]. In the case of subfamily PL14_3, the apparent polyspecificity is associated with the presence of both poly- and oligo-alginate lyases (EC 4.2.2.3 and 4.2.2.-, respectively). Here again, the difference is subtle: the bond cleaved is identical, and the difference in the definitions of the activities pertains to the degree of polymerization of the substrate. It may well be that such a difference is not biologically significant or, if it is, sequence data alone will perhaps never be able to sort one from the other.

Occurrence of PLs in genomes

We entered the genomic era about 15 years ago and the current pace of genome release is on the order of 1 to 2 per day. Next generation sequencing will boost this flow of sequence data even further. Our analyses of more than 1300 genomes from diverse organisms, ranging from Archaea to higher plants and animals, show that the amount of PLs is usually low and consistently less than that of GHs (representing 3 to 5% of the number of GHs). The most likely explanation for this observation is that the substrates of PLs – polysaccharides containing uronic acids – represent just a small proportion of all carbohydrate polymers. The organisms that have the largest number of PLs share a common focus: the plant cell wall.

The genomes of both plants and microorganisms that feed on living or dead plant tissue (phytopathogens or saprophytes, respectively) typically encode large numbers of PLs (Table 2). The abundance of PLs in plants is due to the emergence of large multigene CAZy families [39] and the biological importance of pectin in plant development [40]. The pectic network contributes to the structural integrity of the plant cell wall, and as such, it is an obvious target for phytopathogens and symbionts (including bacteria, fungi, oomycetes, and nematodes) to gain access via an arsenal of pectinolytic enzymes. And, because it is far more digestible than cellulose and lignin, pectin is also a delicacy for most saprophytic organisms, which draw nutrients from decaying plant material.

Recommendations for large scale sequence annotation

Next generation sequencing machines will deliver ever more sequences, whose utility largely depends on our ability to correlate them with molecular functions. The hierarchical classification that we advocate here, based on *fold*, *family*, and *subfamily*, provides a convenient way to produce the best possible functional assignments that take into account distance with experimentally characterized enzymes.

Fold: At the most general end of the spectrum, very distant similarity (such as that resulting from PSI-BLAST analyses or the use of degenerate HMMs) should be used only to assign a protein to a folding class and not to a function. For example, Stam et al. have shown that a PSI-BLAST search starting with a β-helical polygalacturonase (EC 3.2.1.15) of family GH28 retrieved β-helical pectate lyases of family PL1 and dextranases of family GH49 after only two iterations [41], despite the fact that these two enzymes employ distinctly different catalytic mechanisms. Although this may reflect ancient evolutionary events, the detection of such distant similarities is of little use when it comes to anticipating a molecular function.

Family: The next level is the assignment to a family. This is typically reflected by significant BLAST scores over the entire length of the catalytic module (*not the entire protein*, which may contact multiply ancillary modules, *vide supra*). Here, similarity is sufficient to predict a global "polysaccharide lyase" function, especially if the catalytic residues are conserved in the sequence under consideration. Even though the PL families are often polyfunctional, commonalities between the various substrates known to be cleaved by family members can guide experimental design to determine the actual specificity of novel enzymes.

Subfamily: Finally, the most fine-grained annotation is reached at the other end of the spectrum when a sequence can be assigned to one of the defined PL subfamilies. Two cases will arise:

(i) The subfamily to which the sequence can be assigned contains one or several experimentally characterized members (and, the more, the better). Here the function of the query protein can reasonably be assigned, for instance "putative hyaluronate lyase".

(ii) The query protein belongs to a non-characterized subfamily, or does not belong to any defined subfamily. Here the precise substrate cannot be predicted with confidence, and the best possible annotation is simply "putative polysaccharide lyase".

One consequence of the above hierarchy is that functional predictions should be dynamic, varying as biochemical data accumulates in the various subfamilies. Additionally, we suggest that an EC number should only be assigned to the query protein and included in public databases when and only when the precise substrate specificity has been established experimentally, to avoid unchecked propagation of erroneous assignments. In general, we advocate a conservative approach to functional assignment based on sequence analysis, guided by the mantra that no annotation is better than a misleading annotation.

Acknowledgements

We are grateful to Novozymes for support. H.B. acknowledges funding from the Swedish Research Council Formas, the Swedish Research Council, *Vetenskapsrådet*, and the Wallenberg Wood Science Center.

Numbered references

1 Yip, V. L. and Withers, S. G. (2006) Breakdown of oligosaccharides by the process of elimination. Curr. Opin. Chem. Biol. **10**, 147-155

2 Sutherland, I. W. (1995) Polysaccharide Lyases. FEMS Microbiol. Rev. **16**, 323-347

3 Yip, V. L. Y. and Withers, S. G. (2004) Nature's many mechanisms for the degradation of oligosaccharides. Org. Biomol. Chem. **2**, 2707-2713

4 Abbott, D. W. and Boraston, A. B. (2008) Structural biology of pectin degradation by Enterobacteriaceae. Microbiol. Mol. Biol. Rev. **72**, 301-316

5 Gacesa, P. (1987) Alginate-modifying enzymes - a proposed unified mechanism of action for the lyases and epimerases. FEBS Lett. **212**, 199-202

6 Girish, K. S. and Kemparaju, K. (2007) The magic glue hyaluronan and its eraser hyaluronidase: A biological overview. Life Sci. **80**, 1921-1943

7 Herron, S. R., Benen, J. A. E., Scavetta, R. D., Visser, J. and Jurnak, F. (2000) Structure and function of pectic enzymes: Virulence factors of plant pathogens. Proc. Natl. Acad. Sci. U. S. A. **97**, 8762-8769

8 Vorwerk, S., Somerville, S. and Somerville, C. (2004) The role of plant cell wall polysaccharide composition in disease resistance. Trends Plant Sci. **9**, 203-209

9 Linhardt, R. J., Galliher, P. M. and Cooney, C. L. (1986) Polysaccharide lyases. Appl. Biochem. Biotechnol. **12**, 135-176

10 Wong, T. Y., Preston, L. A. and Schiller, N. L. (2000) Alginate lyase: Review of major sources and enzyme characteristics, structure-function analysis, biological roles, and applications. Annu. Rev. Microbiol. **54**, 289-340

11 Hatch, R. A. and Schiller, N. L. (1998) Alginate lyase promotes diffusion of aminoglycosides through the extracellular polysaccharide of mucoid *Pseudomonas aeruginosa*. Antimicrob. Agents Chemother. **42**, 974-977

12 Cescutti, P., Scussolin, S., Herasimenka, Y., Impallomeni, G., Bicego, M. and Rizzo, R. (2006) First report of a lyase for cepacian, the polysaccharide produced by *Burkholderia cepacia* complex bacteria. Biochem. Biophys. Res. Commun. **339**, 821-826

13 Yip, V. L. Y., Varrot, A., Davies, G. J., Rajan, S. S., Yang, X. J., Thompson, J., Anderson, W. F. and Withers, S. G. (2004) An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 beta-glycosidase from *Thermotoga maritima*. J. Am. Chem. Soc. **126**, 8354-8355

14 Rye, C. S., Matte, A., Cygler, M. and Withers, S. G. (2006) An atypical approach identifies TYR234 as the key base catalyst in chondroitin AC lyase. ChemBioChem. **7**, 631-637

15 Shaya, D., Hahn, B. S., Bjerkan, T. M., Kim, W. S., Park, N. Y., Sim, J. S., Kim, Y. S. and Cygler, M. (2008) Composite active site of chondroitin lyase ABC accepting both epimers of uronic acid. Glycobiology. **18**, 270-277

16 Garron, M. and Cygler, M. (2010) Structural and mechanistic classification of polysaccharide lyases. Glycobiology doi:10.1093/glycob/cwq122

17 Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. **382**, 769-781

18 Guillen, D., Sanchez, S. and Rodriguez-Sanoja, R. Carbohydrate-binding domains: multiplicity of biological roles. Appl. Microbiol. Biotechnol. **85**, 1241-1249

19 Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. **37**, D233-D238

20 Lee, S. S., Yu, S. and Withers, S. G. (2003) Detailed dissection of a new mechanism for glycoside cleavage: alpha-1,4-glucan lyase. Biochemistry. **42**, 13081-13090

21 Holtje, J. V., Mirelman, D., Sharon, N. and Schwarz, U. (1975) Novel type of murein transglycosylase in *Escherichia coli*. J. Bacteriol. **124**, 1067-1076

22 Blackburn, N. T. and Clarke, A. J. (2001) Identification of four families of peptidoglycan lytic transglycosylases. J. Mol. Evol. **52**, 78-84

23 Saito, K., Sumita, Y., Nagasaka, Y., Tomita, F. and Yokota, A. (2003) Molecular cloning of the gene encoding the di-D-fructofuranose 1,2 ': 2,3 ' dianhydride hydrolysis enzyme (DFA IIIase) from *Arthrobacter* sp H65-7. J. Biosci. Bioeng. **95**, 538-540
24 Sakurai, H., Yokota, A., Sumita, Y., Mori, Y., Matsui, H. and Tomita, F. (1997) Metabolism of DFA III by *Arthrobacter* sp. H65-7: Purification and properties of a DFA III hydrolysis enzyme (DFA IIIase). Biosci. Biotechnol. Biochem. **61**, 989-993
25 Stam, M. R., Danchin, E. G. J., Rancurel, C., Coutinho, P. M. and Henrissat, B. (2006) Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. Protein Eng. Des. Sel. **19**, 555-562
26 Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. **32**, 1792-1797
27 Henikoff, S. and Henikoff, J. G. (1992) Amino-acid substitution matrices from protein blocks. Proc. Natl. Acad. Sci. U. S. A. **89**, 10915-10919
28 Wicker, N., Perrin, G. R., Thierry, J. C. and Poch, O. (2001) Secator: A program for inferring protein subfamilies from phylogenetic trees. Mol. Biol. Evol. **18**, 1435-1441
29 Gilbert, H. J. The biochemistry and structural biology of plant cell wall deconstruction. Plant Physiol. **153**, 444-455
30 Fouet, A. and Mesnage, S. (2002) *Bacillus anthracis* cell envelope components. Curr Top Microbiol Immunol. . **271**, 87-113
31 Fontes, C. M. G. A. and Gilbert, H. J. (2010) Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. **79**, 655-681
32 Henrissat, B. and Davies, G. J. (2000) Glycoside hydrolases and glycosyltransferases. Families, modules, and implications for genomics. Plant Physiol. **124**, 1515-1519
33 Coutinho, P. M. and Henrissat, B. (1999) Carbohydrate-active enzymes: an integrated database approach. In Recent advances in carbohydrate bioengineering. pp. 3-12, The Royal Society of Chemistry, Cambridge
34 Charnock, S. J., Brown, I. E., Turkenburg, J. P., Black, G. W. and Davies, G. J. (2002) Convergent evolution sheds light on the anti-beta-elimination mechanism common to family 1 and 10 polysaccharide lyases. Proc. Natl. Acad. Sci. U. S. A. **99**, 12067-12072
35 Davies, G. and Henrissat, B. (1995) Structures and mechanisms of glycosyl hydrolases. Structure. **3**, 853-859
36 Coutinho, P. M., Deleury, E., Davies, G. J. and Henrissat, B. (2003) An evolving hierarchical family classification for glycosyltransferases. J. Mol. Biol. **328**, 307-317
37 Czjzek, M., Cicek, M., Zamboni, V., Bevan, D. R., Henrissat, B. and Esen, A. (2000) The mechanism of substrate (aglycone) specificity in beta-glucosidases is revealed by crystal structures of mutant maize beta-glucosidase-DIMBOA, -DIMBOAGlc, and -dhurrin complexes. Proc. Natl. Acad. Sci. U. S. A. **97**, 13555-13560
38 Armand, S., Drouillard, S., Schulein, M., Henrissat, B. and Driguez, H. (1997) A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases. J. Biol. Chem. **272**, 2709-2713
39 Coutinho, P. M., Stam, M., Blanc, E. and Henrissat, B. (2003) Why are there so many carbohydrate-active enzyme-related genes in plants? Trends Plant Sci. **8**, 563-565
40 Caffall, K. H. and Mohnen, D. (2009) The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. Carbohydr. Res. **344**, 1879-1900
41 Stam, M. R., Blanc, E., Coutinho, P. M. and Henrissat, B. (2005) Evolutionary and mechanistic relationships between glycosidases acting on alpha- and beta-bonds. Carbohydr. Res. **340**, 2728-2734
42 Haas, B. J., Kamoun, S., Zody, M. C., Jiang, R. H. Y., Handsaker, R. E., Cano, L. M., Grabherr, M., Kodira, C. D., Raffaele, S., Torto-Alalibo, T., Bozkurt, T. O., Ah-Fong, A. M. V., Alvarado, L., Anderson, V. L., Armstrong, M. R., Avrova, A., Baxter, L., Beynon, J., Boevink, P. C., Bollmann, S. R., Bos, J. I. B., Bulone, V., Cai, G. H., Cakir, C., Carrington, J. C., Chawner, M., Conti, L., Costanzo, S., Ewan, R., Fahlgren, N., Fischbach, M. A., Fugelstad, J., Gilroy, E. M., Gnerre, S., Green, P. J., Grenville-Briggs, L. J., Griffith, J., Grunwald, N. J., Horn, K., Horner, N. R., Hu, C. H., Huitema, E., Jeong, D. H., Jones, A. M. E., Jones, J. D. G., Jones, R. W., Karlsson, E. K., Kunjeti, S. G., Lamour, K., Liu, Z. Y., Ma, L. J., MacLean,

D., Chibucos, M. C., McDonald, H., McWalters, J., Meijer, H. J. G., Morgan, W., Morris, P. F., Munro, C. A., O'Neill, K., Ospina-Giraldo, M., Pinzon, A., Pritchard, L., Ramsahoye, B., Ren, Q. H., Restrepo, S., Roy, S., Sadanandom, A., Savidor, A., Schornack, S., Schwartz, D. C., Schumann, U. D., Schwessinger, B., Seyer, L., Sharpe, T., Silvar, C., Song, J., Studholme, D. J., Sykes, S., Thines, M., van de Vondervoort, P. J. I., Phuntumart, V., Wawra, S., Weide, R., Win, J., Young, C., Zhou, S. G., Fry, W., Meyers, B. C., van West, P., Ristaino, J., Govers, F., Birch, P. R. J., Whisson, S. C., Judelson, H. S. and Nusbaum, C. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature. **461**, 393-398

43 Tyler, B. M., Tripathy, S., Zhang, X. M., Dehal, P., Jiang, R. H. Y., Aerts, A., Arredondo, F. D., Baxter, L., Bensasson, D., Beynon, J. L., Chapman, J., Damasceno, C. M. B., Dorrance, A. E., Dou, D. L., Dickerman, A. W., Dubchak, I. L., Garbelotto, M., Gijzen, M., Gordon, S. G., Govers, F., Grunwald, N. J., Huang, W., Ivors, K. L., Jones, R. W., Kamoun, S., Krampis, K., Lamour, K. H., Lee, M. K., McDonald, W. H., Medina, M., Meijer, H. J. G., Nordberg, E. K., Maclean, D. J., Ospina-Giraldo, M. D., Morris, P. F., Phuntumart, V., Putnam, N. H., Rash, S., Rose, J. K. C., Sakihama, Y., Salamov, A. A., Savidor, A., Scheuring, C. F., Smith, B. M., Sobral, B. W. S., Terry, A., Torto-Alalibo, T. A., Win, J., Xu, Z. Y., Zhang, H. B., Grigoriev, I. V., Rokhsar, D. S. and Boore, J. L. (2006) *Phytophthora* genome sequences uncover evolutionary origins and mechanisms of pathogenesis. Science. **313**, 1261-1266

44 Tuskan, G. A., DiFazio, S., Jansson, S., Bohlmann, J., Grigoriev, I., Hellsten, U., Putnam, N., Ralph, S., Rombauts, S., Salamov, A., Schein, J., Sterck, L., Aerts, A., Bhalerao, R. R., Bhalerao, R. P., Blaudez, D., Boerjan, W., Brun, A., Brunner, A., Busov, V., Campbell, M., Carlson, J., Chalot, M., Chapman, J., Chen, G. L., Cooper, D., Coutinho, P. M., Couturier, J., Covert, S., Cronk, Q., Cunningham, R., Davis, J., Degroeve, S., Dejardin, A., Depamphilis, C., Detter, J., Dirks, B., Dubchak, I., Duplessis, S., Ehlting, J., Ellis, B., Gendler, K., Goodstein, D., Gribskov, M., Grimwood, J., Groover, A., Gunter, L., Hamberger, B., Heinze, B., Helariutta, Y., Henrissat, B., Holligan, D., Holt, R., Huang, W., Islam-Faridi, N., Jones, S., Jones-Rhoades, M., Jorgensen, R., Joshi, C., Kangasjarvi, J., Karlsson, J., Kelleher, C., Kirkpatrick, R., Kirst, M., Kohler, A., Kalluri, U., Larimer, F., Leebens-Mack, J., Leple, J. C., Locascio, P., Lou, Y., Lucas, S., Martin, F., Montanini, B., Napoli, C., Nelson, D. R., Nelson, C., Nieminen, K., Nilsson, O., Pereda, V., Peter, G., Philippe, R., Pilate, G., Poliakov, A., Razumovskaya, J., Richardson, P., Rinaldi, C., Ritland, K., Rouze, P., Ryaboy, D., Schmutz, J., Schrader, J., Segerman, B., Shin, H., Siddiqui, A., Sterky, F., Terry, A., Tsai, C. J., Uberbacher, E., Unneberg, P., Vahala, J., Wall, K., Wessler, S., Yang, G., Yin, T., Douglas, C., Marra, M., Sandberg, G., Van de Peer, Y. and Rokhsar, D. (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science. **313**, 1596-1604

45 Alonso, J. M., Stepanova, A. N., Leisse, T. J., Kim, C. J., Chen, H. M., Shinn, P., Stevenson, D. K., Zimmerman, J., Barajas, P., Cheuk, R., Gadrinab, C., Heller, C., Jeske, A., Koesema, E., Meyers, C. C., Parker, H., Prednis, L., Ansari, Y., Choy, N., Deen, H., Geralt, M., Hazari, N., Hom, E., Karnes, M., Mulholland, C., Ndubaku, R., Schmidt, I., Guzman, P., Aguilar-Henonin, L., Schmid, M., Weigel, D., Carter, D. E., Marchand, T., Risseeuw, E., Brogden, D., Zeko, A., Crosby, W. L., Berry, C. C. and Ecker, J. R. (2003) Genome-wide Insertional mutagenesis of *Arabidopsis thaliana*. Science. **301**, 653-657

46 Coleman, J. J., Rounsley, S. D., Rodriguez-Carres, M., Kuo, A., Wasmann, C. C., Grimwood, J., Schmutz, J., Taga, M., White, G. J., Zhou, S. G., Schwartz, D. C., Freitag, M., Ma, L. J., Danchin, E. G. J., Henrissat, B., Coutinho, P. M., Nelson, D. R., Straney, D., Napoli, C. A., Barker, B. M., Gribskov, M., Rep, M., Kroken, S., Molnar, I., Rensing, C., Kennell, J. C., Zamora, J., Farman, M. L., Selker, E. U., Salamov, A., Shapiro, H., Pangilinan, J., Lindquist, E., Lamers, C., Grigoriev, I. V., Geiser, D. M., Covert, S. F., Temporini, E. and VanEtten, H. D. (2009) The Genome of *Nectria haematococca*: Contribution of Supernumerary Chromosomes to Gene Expansion. PLoS Genet. **5**, e1000618

47 Weiner, R. M., Taylor, L. E., Henrissat, B., Hauser, L., Land, M., Coutinho, P. M., Rancurel, C., Saunders, E. H., Longmire, A. G., Zhang, H. T., Bayer, E. A., Gilbert, H. J., Larimer, F., Zhulin, I. B., Ekborg, N. A., Lamed, R., Richardson, P. M., Borovok, I. and Hutcheson, S. (2008) Complete genome sequence of the complex carbohydrate-degrading marine bacterium, *Saccharophagus degradans* strain 2-40(T). PLoS Genet. **4**, 13

48 McCarter, J. P., Mitreva, M. D., Martin, J., Dante, M., Wylie, T., Rao, U., Pape, D., Bowers, Y., Theising, B., Murphy, C. V., Kloek, A. P., Chiapelli, B. J., Clifton, S. W., Bird, D. M. and Waterston, R. H.

(2003) Analysis and functional classification of transcripts from the nematode *Meloidogyne incognita*. Genome Biol. **4**

49 Machida, M., Asai, K., Sano, M., Tanaka, T., Kumagai, T., Terai, G., Kusumoto, K. I., Arima, T., Akita, O., Kashiwagi, Y., Abe, K., Gomi, K., Horiuchi, H., Kitamoto, K., Kobayashi, T., Takeuchi, M., Denning, D. W., Galagan, J. E., Nierman, W. C., Yu, J. J., Archer, D. B., Bennett, J. W., Bhatnagar, D., Cleveland, T. E., Fedorova, N. D., Gotoh, O., Horikawa, H., Hosoyama, A., Ichinomiya, M., Igarashi, R., Iwashita, K., Juvvadi, P. R., Kato, M., Kato, Y., Kin, T., Kokubun, A., Maeda, H., Maeyama, N., Maruyama, J., Nagasaki, H., Nakajima, T., Oda, K., Okada, K., Paulsen, I., Sakamoto, K., Sawano, T., Takahashi, M., Takase, K., Terabayashi, Y., Wortman, J. R., Yamada, O., Yamagata, Y., Anazawa, H., Hata, Y., Koide, Y., Komori, T., Koyama, Y., Minetoki, T., Suharnan, S., Tanaka, A., Isono, K., Kuhara, S., Ogasawara, N. and Kikuchi, H. (2005) Genome sequencing and analysis of *Aspergillus oryzae*. Nature. **438**, 1157-1161

50 Galagan, J. E., Calvo, S. E., Cuomo, C., Ma, L. J., Wortman, J. R., Batzoglou, S., Lee, S. I., Basturkmen, M., Spevak, C. C., Clutterbuck, J., Kapitonov, V., Jurka, J., Scazzocchio, C., Farman, M., Butler, J., Purcell, S., Harris, S., Braus, G. H., Draht, O., Busch, S., D'Enfert, C., Bouchier, C., Goldman, G. H., Bell-Pedersen, D., Griffiths-Jones, S., Doonan, J. H., Yu, J., Vienken, K., Pain, A., Freitag, M., Selker, E. U., Archer, D. B., Penalva, M. A., Oakley, B. R., Momany, M., Tanaka, T., Kumagai, T., Asai, K., Machida, M., Nierman, W. C., Denning, D. W., Caddick, M., Hynes, M., Paoletti, M., Fischer, R., Miller, B., Dyer, P., Sachs, M. S., Osmani, S. A. and Birren, B. W. (2005) Sequencing of *Aspergillus nidulans* and comparative analysis with *A.fumigatus* and *A.oryzae*. Nature. **438**, 1105-1115

51 Han, C., Spring, S., Lapidus, A., Glavina Del Rio, T., Tice, H., Copeland, A., Cheng, J.-F., Lucas, S., Chen, F., Nolan, M., Bruce, D., Goodwin, L., Pitluck, S., Ivanova, N., Mavromatis, K., Mikhailova, N., Pati, A., Chen, A., Palaniappan, K., Land, M., Hauser, L., Chang, Y.-J., Jeffries, C. D., Saunders, E., Chertkov, O., Brettin, T., Goeker, M., Rohde, M., Bristow, J., Eisen, J. A., Markowitz, V., Hugenholtz, P., Kyrpides, N. C., Klenk, H.-P. and Detter, J. C. (2009) Complete genome sequence of *Pedobacter heparinus* type strain (HIM 762-3T). Stand Genomic Sci. **1**, 54-62

52 Goff, S. A., Ricke, D., Lan, T. H., Presting, G., Wang, R. L., Dunn, M., Glazebrook, J., Sessions, A., Oeller, P., Varma, H., Hadley, D., Hutchinson, D., Martin, C., Katagiri, F., Lange, B. M., Moughamer, T., Xia, Y., Budworth, P., Zhong, J. P., Miguel, T., Paszkowski, U., Zhang, S. P., Colbert, M., Sun, W. L., Chen, L. L., Cooper, B., Park, S., Wood, T. C., Mao, L., Quail, P., Wing, R., Dean, R., Yu, Y. S., Zharkikh, A., Shen, R., Sahasrabudhe, S., Thomas, A., Cannings, R., Gutin, A., Pruss, D., Reid, J., Tavtigian, S., Mitchell, J., Eldredge, G., Scholl, T., Miller, R. M., Bhatnagar, S., Adey, N., Rubano, T., Tusneem, N., Robinson, R., Feldhaus, J., Macalma, T., Oliphant, A. and Briggs, S. (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp *japonica*). Science. **296**, 92-100

53 Xu, J., Bjursell, M. K., Himrod, J., Deng, S., Carmichael, L. K., Chiang, H. C., Hooper, L. V. and Gordon, J. I. (2003) A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. Science. **299**, 2074-2076

54 Deboy, R. T., Mongodin, E. F., Fouts, D. E., Tailford, L. E., Khouri, H., Emerson, J. B., Mohamoud, Y., Watkins, K., Henrissat, B., Gilbert, H. J. and Nelson, K. E. (2008) Insights into plant cell wall degradation from the genome sequence of the soil bacterium *Cellvibrio japonicus*. J. Bacteriol. **190**, 5455-5463

55 Bell, K. S., Sebaihia, M., Pritchard, L., Holden, M. T. G., Hyman, L. J., Holeva, M. C., Thomson, N. R., Bentley, S. D., Churcher, L. J. C., Mungall, K., Atkin, R., Bason, N., Brooks, K., Chillingworth, T., Clark, K., Doggett, J., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Norbertczak, H., Ormond, D., Price, C., Quail, M. A., Sanders, M., Walker, D., Whitehead, S., Salmond, G. P. C., Birch, P. R. J., Parkhill, J. and Toth, I. K. (2004) Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp *atroseptica* and characterization of virulence factors. Proc. Natl. Acad. Sci. U. S. A. **101**, 11105-11110

56 Land, M., Lapidus, A., Mayilraj, S., Chen, F., Copeland, A., Glavina Del Rio, T., Nolan, M., Lucas, S., Tice, H., Cheng, J.-F., Chertkov, O., Bruce, D., Goodwin, L., Pitluck, S., Rohde, M., Goeker, M., Pati, A., Ivanova, N., Mavromatis, K., Chen, A., Palaniappan, K., Hauser, L., Chang, Y.-J., Jeffries, C. D., Brettin, T., Detter, J. C., Han, C., Chain, P., Tindall, B. J., Bristow, J., Eisen, J. A., Markowitz, V.,

Hugenholtz, P., Kyrpides, N. C. and Klenk, H.-P. (2009) Complete genome sequence of *Actinosynnema mirum* type strain (101T). Stand Genomic Sci. **1**, 46-53

Tables and Figures

Table 1: **Activities in PL families and subfamilies.**

Family	Taxonomical range†	Sub Family	Known activities	Characterized enzymes ‡
PL1	A,B,E,U			
		1	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	4/173
		2	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	1/53
		3	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	10/38
		4	pectin lyase (EC 4.2.2.10)	18/40
		5	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	16/38
			exo-polygalacturonate lyase (EC 4.2.2.9)	4/38
		6	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	19/66
		7	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	7/22
		8	pectin lyase (EC 4.2.2.10)	4/22
PL2	A,B			
		1	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	4/21
		2	exo-polygalacturonate lyase (EC 4.2.2.9)	2/23
PL3	B,E			
		1	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	3/137
		2	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	16/115
		3	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	1/8
PL4	B,E			
		1	rhamnogalacturonan lyase (EC 4.2.2.-)	2/21
		2		0/15
		3	rhamnogalacturonan lyase (EC 4.2.2.-)	1/7
		4	rhamnogalacturonan lyase (EC 4.2.2.-)	1/6
PL5	B			
		1	poly(b-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	9/27
PL6	B		poly(b-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	1/21
			chondroitin-sulfate-ABC endolyase (EC 4.2.2.4)	1/21
PL7	B,E			
		1	poly(b-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	1/20
		2		0/7
		3	poly(a-L-guluronate) lyase (EC 4.2.2.11)	2/8
		4		0/5
		5	poly(a-L-guluronate) lyase (EC 4.2.2.11)	4/17
PL8	A,B			
		1	hyaluronate lyase (EC 4.2.2.1)	11/85
		2	chondroitin-sulfate-ABC endolyase (EC 4.2.2.20)	3/30
		3	chondroitin AC lyase (EC 4.2.2.5)	3/7
PL9	B,E			
		1	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	6/66
			exo-polygalacturonate lyase (EC 4.2.2.9)	2/66
		2		0/18
PL10	B			
		1	endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	5/47
PL11	A,B,E			
		1	rhamnogalacturonan lyase (EC 4.2.2.-)	4/64
PL12	A,B			

		1		0/35
		2	heparin-sulfate lyase (EC 4.2.2.8)	2/10
PL13	B		heparin lyase (EC 4.2.2.7)	2/7
PL14	B,E,V			
		1	glucuronate lyase (EC 4.2.2.-)	1/8
		2		0/6
		3	poly(b-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	1/13
			exo-oligo-alginate lyase (EC 4.2.2.-)	1/13
PL15	B			
		1	oligo-alginate lyase (EC 4.2.2.-)	3/5
PL16	B,V			
		1		0/14
		2	hyaluronate lyase (EC 4.2.2.1)	2/8
		3	hyaluronate lyase (EC 4.2.2.1)	1/10
PL17	B			
		1		0/14
		2	poly(b-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	1/15
PL18	B		poly(b-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	3/6
			poly(a-L-guluronate) lyase (EC 4.2.2.11)	3/6
PL20	B,E		endo-b-1,4-glucuronan lyase (EC 4.2.2.14)	1/11
PL21	B		heparin lyase (EC 4.2.2.7)	1/9
			heparin-sulfate lyase (EC 4.2.2.8)	1/9
PL22	A,B		oligogalacturonate lyase(EC 4.2.2.6)	1/47

† A=Archea, B=Bacteria, E=Eukaryota, U=Unclassified, V=Virus.

‡ Number of characterized enzymes/total number of enzymes in subfamily. An enzyme was considered characterized if we could identify published direct evidence for its activity.

Biochemical Journal Immediate Publication. Published on 07 Oct 2010 as manuscript BJ20101185

Table 2: **List of fully sequenced organisms with the highest number of PLs**.

Organism	Reference	Number of PLs*	Description
Phytophthora infestans T30-4	[42]	67	Phytopathogenic oomycete
Phytophthora sojae	[43]	54	Phytopathogenic oomycete
Phytophthora ramorum	[43]	49	Phytopathogenic oomycete
Populus trichocarpa	[44]	39	Plant
Arabidopsis thaliana	[45]	34	Plant
Nectria haematococca mpVI	[46]	33	Phytopathogenic fungus
Saccharophagus degradans 2-40	[47]	32	Marine saprophytic bacterium
Meloidogyne incognita	[48]	30	Phytopathogenic nematode
Aspergillus oryzae RIB40	[49]	23	Phytopathogenic fungus
Aspergillus nidulans FGSC A4	[50]	21	Phytopathogenic fungus
Pedobacter heparinus DSM 2366	[51]	18	Soil bacterium
Oryza sativa Japonica Group	[52]	16	Plant
Dickeya zeae Ech1591	JGI-DOE CP001655	16	Phytopathogenic bacterium
Bacteroides thetaiotaomicron VPI-5482	[53]	16	Human gut bacterium
Cellvibrio japonicus Ueda107	[54]	14	Soil saprophytic bacterium
Streptomyces scabiei 87.22	Wellcome Trust Sanger Institute FN554889	13	Soil bacterium
Pectobacterium carotovorum subsp. *carotovorum* PC1	[55]	13	Phytopathogenic bacterium
Pectobacterium atrosepticum SCRI1043	[55]	13	Phytopathogenic bacterium
Dickeya dadantii Ech586 (*Erwinia chrysanthemi*)	JGI-DOE CP001836	13	Phytopathogenic bacterium
Actinosynnema mirum DSM 43827	[56]	13	Environmental bacterium isolated from grass

*numbers derived from the CAZy database at www.cazy.org

Figure 1. Comparison of the products of polysaccharide lyases (PL) and glycoside hydrolases (GH), exemplified by polygalacturonate (pectate) cleavage. Both enzyme classes generate a new reducing chain end (light grey). GHs cleave the glycosidic bond (C-1':O-4) by addition of water, maintaining the 4-OH group at the new non-reducing chain end. PLs, in contrast, generate a hexeneuronic acid moiety (HexA, 4-deoxy-hex-4-eneuronic acid) at the new non-reducing end by eliminative cleavage of the O-4:C-4 bond.

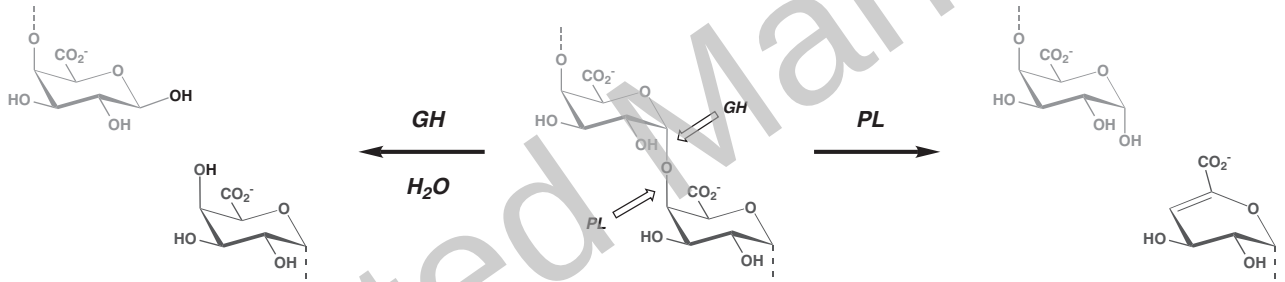
Figure 2. General mechanisms of polysaccharide lyases (PL). A, syn-Elimination, as in chondroitin lyase; B, anti-Elimination, as in a- (1,4)-polygalacturonan (pectate) lyase. In both, polysaccharides are cleaved to produce a 4-deoxy-hex-4-eneuronic acid moiety at the newly-formed non-reducing end of the chain; due to loss of the asymmetric center at C-4, gluco- or galacto-configured substrates yield essentially the same product (depending on the stereochemistry at C-1). As a prelude to chain scission, the C-5 proton adjacent to the carbonyl group is abstracted by a suitably poised basic amino acid sidechain (**B:**). Departure of the glycosidic oxygen is likely to be facilitated by proton donation from a catalytic acid (**B:H**). Coordinating and charge-stabilizing cations, Ca^{++} or a positively charged amino-acid sidechain, are also a common feature of PL actives sites.

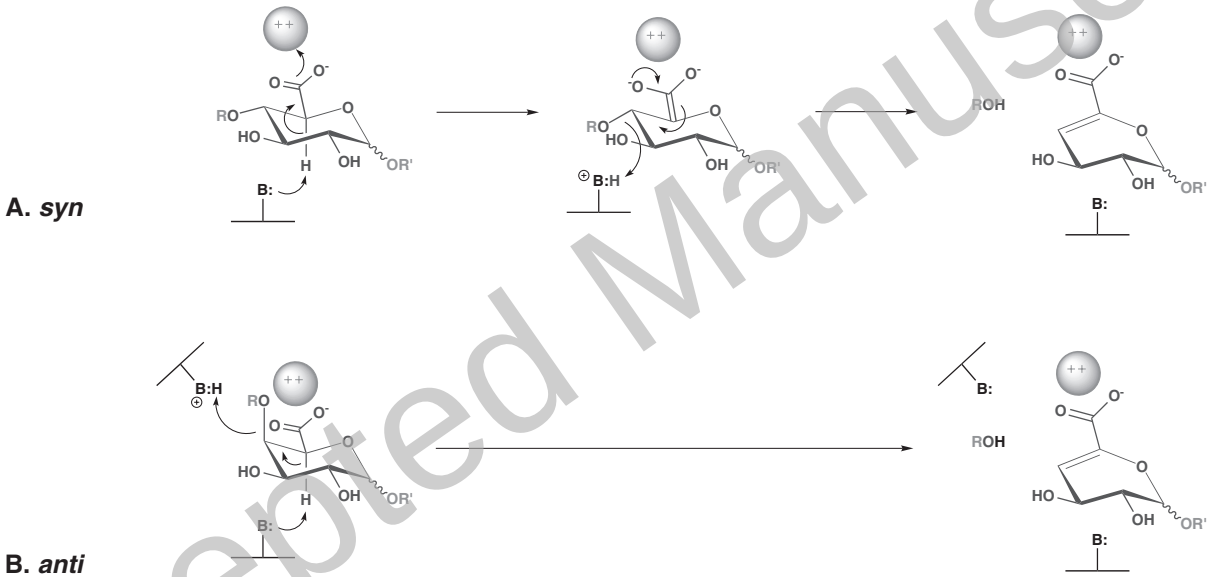
Figure 3: **Examples of modular PLs.** PL: polysaccharide lyase module, CBM: carbohydrate-binding module and CE: carbohydrate esterase module. Other modules include dockerins (DOC), S-layer homology domains (SLH), fibronectin type 3 domains (FN3) and conserved domains of unknown function (X). Unassigned regions are in grey. GenBank accession numbers are indicated for each protein.

Figure 4: **Folds and structures of PL families with known three-dimensional structures.** Representative structures for each family are shown organized by fold. The following Protein Databank (PDB) entries are depicted: PL1(PDB:2QY1), PL3(1EE6), PL6(1OFM), PL9(1RU4), PL11(2ZUY), PL22(3C5M), PL5(1HV6), PL8(1OJM), PL15(3AO0), PL21(2FUQ),PL2(2V8K), PL10(1GXN), PL4(INKG), PL7(1UAI), PL13(3IKW), PL14(3AON), PL18(1J1T), PL20(2ZZJ), PL16(2YW0). Within the β-propeller grouping, PL11 and PL22 members are comprised of 8-bladed and 7-bladed β-propellers, respectively.

Figure 5: Structures of polysaccharide lyase family 8 (PL8) substrates. A, hyaluronan; B, chondroitin (R = R' = H) and chondroitin sulphates A (R = SO_3^-, R' = H) and C (R = H, R' = SO_3^-); C, xanthan. The common glucuronic acid residue is in dark grey, and the scissile bond is identified with an arrow. All monosaccharides not explicitly drawn are D-sugars in the pyranose ring form.

Biochemical Journal Immediate Publication. Published on 07 Oct 2010 as manuscript BJ20101185





Paenibacillus sp. JDR-2, ACT01837.1

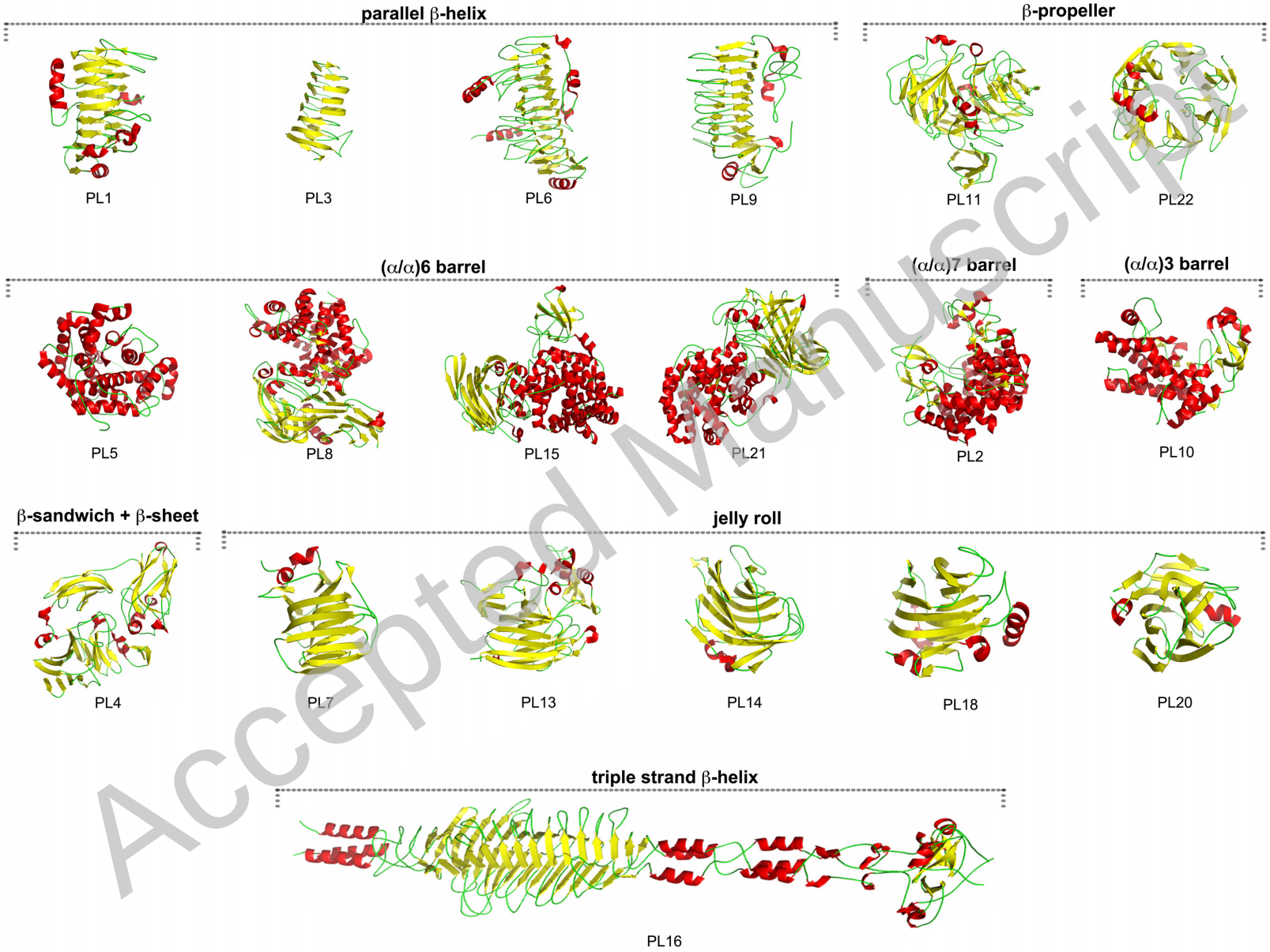
Saccharophagus degradans 2-40, pectin/pectate lyase, ABD80205.1

Bacillus sp. P-358, BAB90989.1

Pseudoalteromonas haloplanktis ANT/505, pectate lyase, AAF86344.2

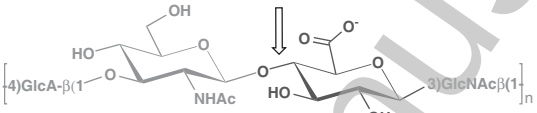
Clostridium thermocellum ATCC 27405, ABN51485.1

Biochemical Journal Immediate Publication. Published on 07 Oct 2010 as manuscript BJ20101185

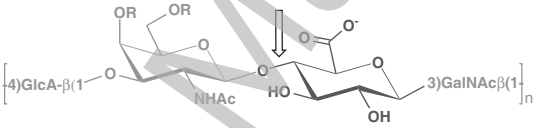


Biochemical Journal Immediate Publication. Published on 07 Oct 2010 as manuscript BJ20101185

A. hyaluronan



B. chondroitin (sulfate A & C)



C. xanthan