A scheme for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants

B. Henrissat[a], T.T. Teeri[b], R.A.J. Warren[c,*]

[a] *Centre de Recherches sur les Macromolécules Végétales, C.N.R.S., P.O. Box 53, F-38041 Grenoble Cedex 9, France*
[b] *Royal Institute of Technology, Department of Biochemistry, S-10044 Stockholm, Sweden*
[c] *Department of Microbiology and Immunology, University of British Columbia, Vancouver, BC V6T 1Z3, Canada*

Received 29 January 1998

Abstract A scheme is proposed for designating enzymes that hydrolyse the polysaccharides in the cell walls of plants. These enzymes are predominantly β-1,4-glycanases. The scheme is based on the classification of the catalytic domains of glycoside hydrolases into families of related amino acid sequences. The new designation for an enzyme indicates its family and, because all members of a family have these characteristics in common, its three-dimensional fold and stereospecificity of hydrolysis. The scheme is intended to simplify comparison of the systems of enzymes produced by different microorganisms for the hydrolysis of plant cell walls.

Key words: Enzyme nomenclature; Structure; Glycosyl hydrolase

1. Introduction

O-Glycosyl hydrolases (EC 3.2.1.-) are a widespread group of enzymes which hydrolyse the glycosidic bond between two or more carbohydrates or between a carbohydrate and a non-carbohydrate moiety. Many of these are enzymes produced by microorganisms for the degradation of the polysaccharides in the cell walls of plants. Several hundred hydrolases from a variety of microorganisms have been characterised by cloning and sequencing the genes encoding them, deducing their amino acid sequences, determining their enzymatic properties, and, increasingly, determining their three-dimensional structures [1–5]. The proteins and genes were named randomly as they were identified, and given the appropriate designations according to the IUB-MB enzyme nomenclature. However, the IUB-MB enzyme nomenclature of glycosyl hydrolases is based on their substrate specificity and occasionally on their molecular mechanism; such a classification does not reflect (and was not intended to reflect) the structural features of these enzymes. This did not present a problem at first, when too few enzymes had been characterised to allow comparison of the hydrolase systems produced by different microorganisms. Now, it is clear that the systems from widely different microorganisms share similarities that are not apparent from the designations of their enzymes [3–5].

A classification of glycosyl hydrolases into families based on amino acid sequence similarities was proposed a few years ago [6–9]. Because there is a direct relationship between the amino acid sequence and the folding of an enzyme, such a classification is expected to: (i) reflect the structural features of these enzymes better than substrate specificity alone; (ii) help to reveal the evolutionary relationships between these enzymes; and (iii) provide a convenient tool to derive mechanistic information from the protein sequence data [7,8] (an updated list of the glycosyl hydrolase families can be found at the Expasy server http://www.expasy.ch/cgi-bin/lists?glycosid.txt). Many glycoside hydrolases are modular, comprising a catalytic domain (CD) and one or more ancillary domains [1,10]. The CDs catalyse hydrolysis with either retention or inversion of the configuration at the anomeric centre of the substrate [11–14]. In addition to revealing overall structural relationships of glycosyl hydrolases, the hydrophobic cluster analysis (HCA) and amino acid sequence alignment of the catalytic domains allow prediction of the stereospecific outcome and the key amino acid residues involved in the catalytic mechanism of the reaction [6–9,15]. Thus, the CDs in a family have the same three-dimensional fold and exhibit the same stereospecificity of hydrolysis; for example, retaining in family 5 and inverting in family 6.

The cell walls of plants contain several insoluble β-1,4-linked polysaccharides. The enzyme systems produced by microorganisms for their hydrolysis are complex; they usually comprise hydrolases from several families, and there may be multiple enzymes hydrolysing each polysaccharide. The current nomenclature for these enzymes is not systematic and does not allow easy comparison of similar enzymes from different organisms. This paper proposes a scheme for standard designations, based on the CD families, for enzymes hydrolysing plant cell walls.

2. Gene and protein

In accordance with standard practice in bacterial genetics, the genes and their products are designated by three letters. If the enzymes are named according to their preferred substrates, then the designations would be as given in Table 1. Note that there must be strict correspondence between the letters for gene and protein.

3. Families and designations

The designation shows to which family an enzyme belongs; the family indicates the stereospecificity indirectly because it is the same for all members of a family. An enzyme from family 5 of glycoside hydrolases will be Cel5 or Man5, etc., depending on its preferred substrate. If an organism produces multiple enzymes from a family, these will be designated Cel5A, Cel5B, etc., with the letters after the family number corresponding to the order in which the enzymes were first reported. If an enzyme contains more than two catalytic do-

*Corresponding author.
E-mail: rajw@unixg.ubc.ca

PII S0014-5793(98)00265-8

Table 1
Acronyms for genes and encoded enzymes

Enzyme	Gene	Protein	EC designations
Cellulase	*cel*	Cel	EC 3.2.1.4; EC 3.2.1.91
Xylanase	*xyn*	Xyn	EC 3.2.1.8
Mannanase	*man*	Man	EC 3.2.1.78
Lichenase	*lic*	Lic	EC 3.2.1.73; EC 3.2.1.58
Laminarinase	*lam*	Lam	EC 3.2.1.39

mains, the designation will so indicate. For example, endoglucanase CelA from *Caldocellulosiruptor saccharolyticum* will be Cel9A-Cel48A, written in the conventional sense from the amino- to the carboxyl-terminus.

The scheme does not distinguish between endo- and exo-acting enzymes, etc. Such distinctions are not absolute; rather, an enzyme has a predominantly exo- or endoglycanolytic mode of action [3,5,14]. Particular enzymes will be referred to as endoglucanase Cel5A, cellobiohydrolase Cel6A, cellodextrinase Cel3, and so on.

Ideally, for facile comparison of different enzyme systems, designations should reflect as many similarities as possible. In the past, enzymes from a microorganism were given letters to indicate the order in which they were first reported: endoglucanases CenA, CenB, CenC and CenD from *Cellulomonas fimi*, for example. This will still apply to enzymes from the same family: CenB and CenC will become endoglucanases Cel9A and Cel9B, respectively.

Application of the proposed scheme to four microorganisms is presented in Table 2; it makes clear at once the similarities and differences between their hydrolytic enzyme systems. Two similar enzymes from different organisms are differentiated by indicating the organism of origin. For example, the Cel9B enzymes from *Cellulomonas fimi* and *Thermomonospora fusca* would be Cf Cel9B and Tf Cel9B, respectively. The proposed nomenclature would reduce significantly the jargon currently prevailing in literature of glycoside hydrolases.

4. Active sites

As the three-dimensional structures of more and more glycoside hydrolases are solved, it becomes imperative that a single system be used to designate the sugar-binding subsites within the active sites of the enzymes. This should be the −3, −2, −1, +1, +2, +3, system, where −1 is the site for the glycone and +1 is the site for the aglycone moiety of the bond being hydrolysed [16–18].

5. Clans

Because the folds of proteins are better conserved than their sequences, some of the families of CDs discussed here can be grouped into 'clans' when (i) new sequences are found to be related to more than one family; or (ii) the sensitivity of sequence comparison methods is increased; or (iii) structural determinations demonstrate the resemblance between members of different families [9,19]. For example, families 1, 2, 5, 10, 17, 26, 30, 35, 39, 42 and 53 of retaining glycoside hydrolases form clan GH-A, in which the CDs are $(\beta/\alpha)_8$ barrels with the catalytic glutamates near the carboxy-terminal ends of β-strands 4 and 7 [15,20]. Other clans are becom-

Table 2
Designations for enzymes from *Trichoderma reesei, Cellulomonas fimi, Clostridium thermocellum* and *Thermomonospora fusca*

Organism	Enzyme	Current designation	New designation
T. reesei	cellobiohydrolase I	CBHI	Cel7A
	cellobiohydrolase II	CBHII	Cel6A
	endoglucanase I	EGI	Cel7B
	endoglucanase II	EGII	Cel5A
	endoglucanase III	EGIII	Cel12A
	endoglucanase V	EGV	Cel45A
C. fimi	endoglucanase A	CenA	Cel6A
	endoglucanase B	CenB	Cel9A
	endoglucanase C	CenC	Cel9B
	endoglucanase D	CenD	Cel5A
	cellobiohydrolase A	CbhA	Cel6B
	cellobiohydrolase B	CbhB	Cel48A
	xylanase B	Cex	Xyn10A
C. thermocellum	endoglucanase B	CelB	Cel5A
	endoglucanase C	CelC	Cel5B
	endoglucanase C	CelC	Cel5C
	endoglucanase G	Cel G	Cel5D
	endoglucanase H	CelH	Cel26A-Cel5E
	endoglucanase A	CelA	Cel8A
	endoglucanase D	CelD	Cel9A
	endoglucanase	CelF	Cel9B
	endoglucanase I	CelI	Cel9C
	endoglucanase J	CelJ	Cel9D-Cel44A
	lichenase B	LicB	Lic16A
	cellobiohydrolase S	CelS	Cel48A
	endoglucanase M	CelM	Cel60A
T. fusca	endoglucanase	E1	Cel9B
	endoglucanase	E2	Cel6A
	cellobiohydrolase	E3	Cel6B
	endoglucanase	E4	Cel9A
	endoglucanase	E5	Cel5A
	cellobiohydrolase	E6	Cel48A

Table 3
Module arrangements in enzymes from *Cellulomonas fimi*

Enzyme	Module arragement
Endoglucanase Cel5A	CD5/FN3/FN3/CBD2
Endoglucanase Cel6A	CBD2/CD6
Cellobiohydrolase Cel6B	CD6/FN3/FN3/FN3/CBD2
Endoglucanase Cel9A	CD9/CBD3/FN3/FN3/FN3/CBD2
Endoglucanase Cel9B	CBD4/CBD4/CD9/?/?
Cellobiohydrolase Cel48A	CD48/FN3/FN3/FN3/CBD2
Xylanase Xyn10A	CD10/CBD2
Xylanase Xyn10B	NodB/TST/CD10/CBD9/CBD9/?
Xylanase Xyn11A	CD11/CBD2/NodB/CBD2

CD: catalytic domain with family number; CBD: cellulose-binding domain with family number; FN3: fibronectin type III-like repeat; TST: thermostabilising domain; NodB: acetylxylan esterase domain related to NodB from *Rhizobium* spp.; ?: domain of unknown function.

ing apparent. However, inclusion in its designation of the clan to which an enzyme belongs is felt to be unnecessary. As with its stereospecificity, its clan will follow from the family designation.

Although all of the enzymes in a clan can be said to belong to a single super-family, alignment of the amino acid sequences of the enzymes in clan GH-A clearly differentiates them into the smaller families. This is illustrated by comparing the consensus amino acid sequences around the catalytic glutamates in different families in the clan (not shown). Use of the smaller families affords more detailed comparison of the enzyme systems from different microorganisms (Table 2).

6. Modular structures

Many of the hydrolases are modular proteins. All of them contain a CD. The commonest ancillary module is a cellulose-binding domain (CBD). CBDs are also classified into families of related amino acid sequences [1,21]. In instances where a cartoon of a modular structure is inappropriate, a written designation of the modular organisation would be useful (Table 3).

7. Conclusion

Acceptance of the scheme proposed here will require changing the acronyms for all the β-1,4-glycanases described to date, but the advantages of a uniform nomenclature should overcome objections to this. New enzymes would be designated according to the new scheme when first described. The scheme will reduce much of the confusion surrounding this important group of enzymes.

Acknowledgements: We are grateful to Sharon Shoemaker for organising a meeting on the nomenclature of glycoside hydrolases at Fallen Leaf Lake, California in the spring of 1994, thereby initiating the process of renaming the enzymes. The scheme proposed here is the outcome of that meeting. Valuable suggestions were made by many colleagues, especially Pierre Béguin, Gideon Davis, Doug Eveleigh, Martin Schülein and David Wilson.

References

[1] Tomme, P., Warren, R.A.J. and Gilkes, N.R. (1995) Adv. Microb. Physiol. 37, 81.
[2] Saddler, J.N. and Penner, M. (Eds.) (1995) Enzymatic Degradation of Insoluble Carbohydrates, Vol. 618, Am. Chem. Soc. Symp. Ser., Washington, DC.
[3] Warren, R.A.J. (1996) Annu. Rev. Mirobiol. 50, 183–212.
[4] Davies, G. and Henrissat, B. (1995) Structure 3, 853–859.
[5] Teeri, T.T. (1997) Trends Biotechnol. 15, 160–167.
[6] Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L. and Mornon, J.P. (1989) Gene 81, 83–95.
[7] Henrissat, B. (1991) Biochem. J. 280, 309–316.
[8] Henrissat, B. and Bairoch, A. (1993) Biochem. J. 293, 781–788.
[9] Henrissat, B. and Bairoch, A. (1996) Biochem. J. 316, 695–696.
[10] Bork, P., Downing, A.K., Kieffer, B. and Campbell, I.D. (1996) Q. Rev. Biophys. 29, 119–167.
[11] Koshland, D.E. (1953) Biol. Rev. 28, 416–436.
[12] Sinnott, M. (1990) Chem. Rev. 90, 1171–1202.
[13] Withers, S.G. (1995) Pure Appl. Chem. 67, 1673–1682.
[14] Davies, G., Sinnott, M. and Withers, S.G. (1998) in: Comprehensive Biological Catalysis (Sinnott, M., Ed.), pp. 119–208, Academic Press, New York.
[15] Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P. and Davies, G. (1995) Proc. Natl. Acad. Sci. USA 92, 7090–7094.
[16] Sugunama, T., Matsuno, R., Ohnishi, M. and Hiromi, K. (1978) J. Biochem. (Tokyo) 84, 293–316.
[17] Biely, P., Vrsanska, M. and Claeyssens, M. (1991) Eur. J. Biochem. 200, 157–163.
[18] Davies, G.J., Wilson, K. and Henrissat, B. (1997) Biochem. J. 321, 557–559.
[19] Henrissat, B. and Davies, G.J. (1997) Curr. Opin. Struct. Biol. 7, 637–644.
[20] Jenkins, J., Lo Leggio, L., Harris, G. and Pickersgill, R. (1995) FEBS Lett. 362, 281–285.
[21] Tomme, P., Warren, R.A.J., Miller, R.C. Jr., Kilburn, D.G. and Gilkes, N.R. (1995) in: Enzymatic Degradation of Insoluble Carbohydrates (Saddler, J.N. and Penner, M.H., Eds.), Vol. 618, pp. 142–163, Am. Chem. Soc. Symp. Ser., Washington, DC.