

Archive ouverte UNIGE

https://archive-ouverte.unige.ch

Article scientifique Article

e 1996

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

Updating the sequence-based classification of glycosyl hydrolases

Henrissat, Bernard; Bairoch, Amos Marc

How to cite

HENRISSAT, Bernard, BAIROCH, Amos Marc. Updating the sequence-based classification of glycosyl hydrolases. In: Biochemical journal, 1996, vol. 316 (Pt 2), p. 695–696. doi: 10.1042/bj3160695

This publication URL:https://archive-ouverte.unige.ch/unige:36909Publication DOI:10.1042/bj3160695

© This document is protected by copyright. Please refer to copyright holder(s) for terms of use.



Updating the sequence-based classification of glycosyl hydrolases

A classification of glycosyl hydrolases based on amino-acidsequence similarities was proposed in this Journal a few years ago [1]. This classification originated from the analysis of ~ 300 sequences and their grouping into 35 families designated 1–35. Because such a classification is necessarily sensitive to the sample, it was anticipated that it was incomplete and that new families would be determined when additional sequences would become available. When the number of glycosyl hydrolase sequences reached ~ 480, ten additional families (designated 36–45) could be defined and were added to the classification [2]. There are at present over 950 sequences of glycosyl hydrolases in the databanks (EMBL/GenBank and SWISS-PROT). Their analysis shows that the vast majority of the ~ 470 additional sequences that have become available since the last update could be classified in the existing families. However, several sequences not fitting the existing families allow the definition of new families (designated 46–57) (Table 1). When the several present genome sequencing projects have reached completion, the number of

Table 1 New families in the classification of glycosyl hydrolases

 Family	Enzyme	Organism	SWISS-PROT	EMBL/GenBank
46	Chitosanase	Bacillus circulans MH-K1	P33673	D10624
46	Chitosanase	Streptomyces sp. N174	P33665	L07779
46	Chitosanase	Nocardioides sp.	P48846	L40408
47 47 47 47 47 47 47 47 47	 ~-Mannosidase Open reading frame 	Drosophila melanogaster Human Mouse Mouse Penicillium citrinum Rabbit Saccharomyces cerevisiae Caenorhabditis elegans	P33908 P39098 P45700 P45701 P32906	X82641 X74837 U03458 U04299 D45839 U04301 M63598; Z49631 Z47073
48 48 48 48	Cellulase CelS Cellulase CelF Open reading frame Cellobiohydrolase B	Clostridium thermocellum Clostridium cellulolyticum Caldocellum saccharolyticum Cellulomonas fimi	P38686 P37698 P22534	S56455 U30321 M36063 L38827
49	Dextranase	Arthrobacter sp.	P39652	D00834
49	Dextranase	Penicillium minioluteum	P48845	L41562
50	Agarase A	<i>Vibrio</i> sp.	P48839	D14721
50	Agarase B	<i>Vibrio</i> sp.	P48840	D21202
51 51	Arabinofuranosidase A Arabinofuranosidase	Aspergillus niger Streptomyces lividans	P42254	L29005 U04630
52 52	eta-Xylosidase eta -Xylosidase	Bacillus stearothermophilus 236 Bacillus stearothermophilus 21	P45704 P45702	U15984 D28121
53	Galactanase 1	Aspergillus aculeatus	P48842	L34599
53	Open reading frame	Bacillus polymyxa	P48843	L03425
53	Galactanase	Pseudomonas fluorescens	P48841	X91885
54	Arabinofuranosidase B	Aspergillus niger	P42255	X74777
54	Arabinofuranosidase/xylanase	Trichoderma koningii	P48792	U38661
55 55	Exo-1,3- eta -glucanase Endo-1,3- eta -glucanase	Cochliobolus carbonum Trichoderma harzianum	P49426	L48994 X84085
56	Hyaluronidase	Apis mellifera	Q08169	L10710
56	Hyaluronidase	Cavia porcellus	P23613	X56332
56	Hyaluronidase	Dolichovespula maculata	P49 371	L34548
56	Hyaluronidase	Human	P38567	L13781
56	Hyaluronidase	Macaca fascicularis	P38568	L13780
56	Hyaluronidase	Mouse	P48794	U33958
56	Hyaluronidase	Rabbit	P38566	U09183
56	Hyaluronidase	Vespula vulgaris	P49370	L43562
57	α -Amylase 1	Dictyoglomus thermophilum	P09961	X07896
57	α -Amylase	Pyrococcus furiosus	P49067	L22346

696	BJ	Letters
-----	----	---------

GUX2 TRIRE (P07987)

Family 6
Description: Endoglucanases (EC 3.2.1.4) and cellobiohydrolases (EC 3.2.1.91).
PROSITE: PDOC00563
3D structure status: Available
Reaction stereochemical outcome: Inverted anomeric configuration
Catalytic nucleophile/base: Asp (experimental)
Catalytic proton donor: Asp (experimental)
Clan: None
Known taxonomic range: eukaryotae, prokaryotae.
Note: formerly known as cellulase family B.
GUNA_CELFI (P07984), GUNB_FUSOX (P46236), GUNA_MICBI (P26414),
GUNA_STRHA (P33682), GUN1_STRSQ (P13933), GUN2_THEFU (P26222),

Figure 1 Example of a section of the glycosyl hydrolase classification document

The text in **bold** denotes hypertext links to other electronic servers (ENZYME, PROSITE and SWISS-PROT).

glycosyl hydrolase sequences will probably increase dramatically. There are two major problems with keeping the classification: (i) how to make it available *in toto*, and (ii) how to disclose the new families when they are discovered. One way is the publication in scientific journals of papers whose interest decreases as they become progressively similar to stamp collections. Another way is the use of more adapted media such as electronic databases.

We are happy to announce that a permanently updated version of the classification is now available through the ExPASy WWW server [3] at the URL: 'http://expasy.hcuge.ch/cgi-bin/ lists?glycosid.txt'. For each family of glycosyl hydrolases, a section of the document exists (Figure 1) that briefly lists the main enzymes that belong to this family. This section includes links to the relevant SWISS-PROT [4] protein-sequence entries. Links are also provided to the relevant EC numbers in the ENZYME [5] nomenclature database as well as to PROSITE [6] entries (which currently exist for more than half of the known glycosyl hydrolase families). This electronic classification should answer the need for rapid updates and availability *in toto* or family by family and allow users to navigate seamlessly between various types of network resources providing information on these enzymes.

There are two major mechanisms for glycosyl hydrolases, leading to overall retention or inversion of the stereochemistry at the cleavage point [7]. The mechanism appears to be conserved within each family [8]. The following families have been found to act with a retaining mechanism : 1, 2, 5, 7, 10, 11, 12, 13, 16, 17, 22, 30, 31, 32, 33, 34, 35, 39 and 42 (the mechanism of families 30, 35 and 42 was inferred from sequence similarities [9]). The inverting mechanism prevails in families 6, 8, 9, 14, 15, 19, 24, 37, 43, 44, 45, 46, 47 and 48. The electronic classification indicates the type of mechanism for each family where it is known.

The three-dimensional structure is now known for at least one member of families 1, 2, 5, 6, 7, 9, 10, 11, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 33, 34, 45 and 46 (for a review, see [10]). The availability of a three-dimensional structure is indicated in the electronic version of the classification.

A 'clan' is a group of families that are thought to have a common ancestry and are recognized by significant similarities in tertiary structure together with conservation of the catalytic

Table 2 Clan grouping of glycosyl hydrolase families

Clan	Families grouped	
GH-A	1, 2, 5, 10, 17, 30, 35, 39, 42	[9,11]
GH-B	7, 16	[12]
GH-C	11, 12	[13]
GH-D	27, 36	[14,15]
GH-E	33, 34	[16]

residues and catalytic mechanism. The growing number of threedimensional structures solved for glycosyl hydrolases and/or improved sequence comparison strategies have revealed the relationship between some glycosyl hydrolases families which can be grouped in clans (Table 2).

The grouping into clans is also indicated in the electronic classification. Families 19, 22, 23 and 24 have been proposed to be related on the basis of weak local three-dimensional similarities [17]. However, since family 22 acts with a retaining mechanism while families 19 and 24 use the inverting mechanism, and since the mechanism of family 23 is not yet known, we feel it is safer to consider these families as independent until a stronger evidence for their grouping is available.

Bernard HENRISSAT*‡ and Amos BAIROCH†

*Centre de Recherches sur les Macromolécules Végétales§, C.N.R.S., BP 53, 38041 Grenoble Cédex, France, and †Medical Biochemistry Department, Centre Médical Universitaire, CH-1211 Geneva 4, Switzerland

- ‡ To whom correspondence should be addressed.
- § Affiliated with the Université Joseph Fourier, Grenoble, France.
- 1 Henrissat, B. (1991) Biochem. J. 280, 309-316
- 2 Henrissat, B. and Bairoch, A. (1993) Biochem. J. 293, 781-788
- 3 Appel R. D., Bairoch A. and Hochstrasser D. F. (1994) Trends Biochem. Sci. 19, 258-260
- 4 Bairoch A. and Apweiler R. (1996) Nucleic Acids Res. 24, 21-25
- 5 Bairoch A. (1996) Nucleic Acids Res. 24, 221-222
- 6 Bairoch A., Bucher P. and Hofmann K. (1996) Nucleic Acids Res. 24, 189-196
- 7 Sinnott, M. L. (1990) Chem. Rev. 90, 1171–1202
- 8 Gebler, J. C., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Béguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, Jr., R. C., Warren, R. A. J. and Withers, S. G. (1992) J. Biol. Chem. **267**, 12559–12561
- 9 Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P. and Davies, G. (1995) Proc. Natl. Acad. Sci. U.S.A. 92, 7090–7094
- 10 Davies, G. and Henrissat, B. (1995) Structure 3, 853-859
- 11 Jenkins, J., Lo Leggio, L., Harris, G. and Pickersgill, R. (1995) FEBS Lett. 362, 281–285
- 12 Divne, C., Stahlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T. and Jones, T. A. (1994) Science 265, 524–528
- 13 Törrönen, A., Kubicek, C. P. and Henrissat, B. (1993) FEBS Lett. 321, 135-139
- 14 Dagnall, B. H., Paulsen, I. T. and Saier, Jr., M. H. (1995) Biochem. J. 311, 349-350
- 15 Romeu, A. and Henrissat, B. (1995) Biochem. J. **311**, 350–351
- 16 Crennel, S. J., Garman, E. F., Laver, W. G., Vimr, E. R. and Taylor, G. L. (1993) Proc. Natl. Acad. Sci. U.S.A. 90, 9852–9856
- 17 Holm, L. and Sander, C. (1994) FEBS Lett. 340, 129-132

Received 11 March 1996