Proc. Natl. Acad. Sci. USA
Vol. 92, pp. 7090–7094, July 1995
Evolution

Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases

BERNARD HENRISSAT*†, ISABELLE CALLEBAUT‡, SYLVIE FABREGA§, PIERRE LEHN§, JEAN-PAUL MORNON‡, AND GIDEON DAVIES¶

*Centre de Recherches sur les Macromolécules Végétales, Centre National de la Recherche Scientifique, Unité Propre de Recherche 5301, BP 53, F-38041 Grenoble Cedex 9, France; ‡Systèmes Moléculaires et Biologie Structurale, Laboratoire de Minéralogie-Cristallographie, Centre National de la Recherche Scientifique, Unité de Recherche Associée 09, Universités Paris VI-Paris VII, T16, case 115, 4 place Jussieu, F-75252 Paris Cedex 5, France; §Hôpital R. Debré, Institut National de la Santé et de la Recherche Médicale U120, 48 Bd Sérurier, F-75019 Paris, France; ¶Department of Chemistry, University of York, Heslington, York YO1 5DD, United Kingdom

Communicated by Russell F. Doolittle, University of California at San Diego, La Jolla, CA, March 20, 1995 (received for review November 29, 1994)

ABSTRACT **The regions surrounding the catalytic amino acids previously identified in a few "retaining" *O*-glycosyl hydrolases (EC 3.2.1) have been analyzed by hydrophobic cluster analysis and have been used to define sequence motifs. These motifs have been found in more than 150 glycosyl hydrolase sequences representing at least eight established protein families that act on a large variety of substrates. This allows the localization and the precise role of the catalytic residues (nucleophile and acid catalyst) to be predicted for each of these enzymes, including several lysosomal glycosidases. An identical arrangement of the catalytic nucleophile was also found for *S*-glycosyl hydrolases (myrosinases; EC 3.2.3.1) for which the acid catalyst is lacking. A $(\beta/\alpha)_8$ barrel structure has been reported for two of the eight families of proteins that have been grouped. It is suggested that the six other families also share this fold at their catalytic domain. These enzymes illustrate how evolutionary events led to a wide diversification of substrate specificity with a similar disposition of identical catalytic residues onto the same ancestral $(\beta/\alpha)_8$ barrel structure.**

Glycosyl hydrolases (EC 3.2.1–3.2.3) are a widespread group of enzymes hydrolyzing the glycosidic bond between two or more carbohydrates or between a carbohydrate and a noncarbohydrate moiety. The biological functions of oligo- and polysaccharides being extremely varied, glycosyl hydrolases intervene in many essential steps of life: hydrolysis of structural or storage polysaccharides, defense against pathogens, penetration of certain pathogens into cells, turnover of cell surface carbohydrates, etc. Heritable deficiencies in glycosyl hydrolases—for example, lactose intolerance (1) or the large group of lysosomal storage diseases (2)—are among the most frequent genetically based syndromes in man.

The extensive diversity in stereochemistry of carbohydrates and the astronomical number of their possible combinations [there are $>10^{12}$ possible oligosaccharide isomers for a reducing hexasaccharide (3)] are paralleled by an immense variety of enzymes designed for their selective hydrolysis. The number of protein folds has been estimated to be not more than a few thousand (4, 5). This raises the question of the origin of the specificity of glycosyl hydrolases and suggests that glycosyl hydrolases of different substrate specificity could well have similar folds.

Glycosyl hydrolases have been recently grouped and classified in families on the basis of amino acid sequence similarities (6, 7). When the sequences of two or more glycosyl hydrolases could be aligned over an entire domain, they were assigned to the same family. While cellulases belonging to different glycosyl hydrolase families have been found to have different folds (8, 9), some enzymes from different families have related folds (10). The three-dimensional structures of two plant β-glucanases with distinct substrate specificities have recently been found to be strongly related (11). Given that the fold of proteins is better conserved than the sequence, it is possible that several glycosyl hydrolase families share similar folds. A time-consuming way to verify this assertion would be to solve at least one three-dimensional structure in each family. We present here another approach for detecting possible folding similarities between glycosyl hydrolase families. The strategy provides a significant improvement of the sensitivity of sequence comparison methods by (*i*) restraining the analysis only to enzymes operating with the same molecular mechanism and (*ii*) focusing on the regions subjected to the most intense conservation pressure—i.e., those carrying the catalytic residues.

Glycosyl hydrolases function by using one of two general mechanisms leading to either overall retention or inversion of the anomeric configuration at the hydrolysis site (12, 13). In both mechanisms, two residues participate directly in catalysis: a nucleophile and a proton donor. In retaining enzymes, the two residues are disposed on the two opposite sides of the glycosidic bond to hydrolyze and are separated by a distance of ≈5.5 Å. In inverting enzymes, this distance is larger (≈9.5 Å) to accommodate a water molecule in addition to the substrate (14, 15). Successful sequence comparison of catalytic regions, therefore, requires knowledge of both the stereochemistry of hydrolysis and the role of the catalytic residues. A precise identification of the catalytic nucleophile in a number of "retaining" glycosyl hydrolases is now available through the elegant work of Withers and collaborators (for a review see ref. 14). The use of 2-fluoro glycosides has thus enabled the identification of the catalytic nucleophile in several retaining glycosyl hydrolases belonging to different families. In one instance, the location of the proton donor has been inferred from detailed kinetic studies of mutant proteins (16).

We have chosen hydrophobic cluster analysis (HCA) (17, 18) as the sequence comparison method for this work because of its known sensitivity at low sequence identity level and its ability to significantly detect secondary structure elements in proteins (19).

METHODS AND DATA

The HCA method (17) is based primarily on the basic rules underlying the folding of globular proteins (hydrophilic sur-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: HCA, hydrophobic cluster analysis; TIM-barrel, $(\beta/\alpha)_8$ barrel structure first found in triose-phosphate isomerase.

†To whom reprint requests should be addressed.

face vs. hydrophobic core). Instead of involving the maximization of sequence similarity scores—a scheme whose reliability decreases strongly as sequence identity becomes lower than 20%—the HCA method uses a bidimensional plot in which the amino acid sequence of a protein is displayed as an unrolled and duplicated longitudinal cut of a cylinder, where the amino acid residues follow an α-helical pattern. The duplication of the helical net allows the full sequence environment of each amino acid to be represented. On this representation, the clusters of contiguous hydrophobic residues (V, I, L, M, F, W, Y) have been shown to correspond significantly to secondary structure elements in globular proteins (19). The segmentation of a protein into successive secondary structure elements becomes visible along the horizontal axis of the diagram, while the sequence itself can be read on an almost vertical axis. The analysis then involves the comparison of cluster shape [for instance large horizontal clusters correspond predominantly to α-helices and short vertical ones to strands (17, 18)] and cluster distribution between several plots in order to find correspondences.

Sequences were taken from the Swiss-Prot protein sequence data bank and were converted into Postscript HCA plots by using the program HCA-PLOT V3.0 (Doriane, Le Chesnay, France). To facilitate visual inspection of the plots, some amino acids are represented by symbols: ★ for proline, ◆ for glycine, ⊡ for serine, and □ for threonine. Clusters of contiguous hydrophobic residues are drawn automatically by the program. Subsequent graphical manipulations of the plots were done with ISLANDDRAW V3.0 (Island Graphics, Hoofddorp, The Netherlands). The significance of the proposed correspondences was assessed with the statistical significance test of MACAW (20) by using the BLOSUM62 matrix (21).

RESULTS AND DISCUSSION

HCA plots have been produced for different retaining glycosyl hydrolases whose nucleophile has been identified unambiguously (Table 1). These enzymes are representatives of glycosyl hydrolase families 1, 2, 5, 10, and 30. Examination of the HCA plots around the nucleophile (noted B^- in Fig. 1) in these proteins reveals that the nucleophile is an invariant glutamic acid located immediately after a hydrophobic cluster. For the two enzymes whose three-dimensional structures are known (14, 22), this cluster corresponds to a β-strand (Fig. 1, II and III). For the other enzymes in Fig. 1, whose three-dimensional structures are not known, the vertical shape of the corresponding clusters is compatible with a β-strand (17, 18).

Endoglucanase C from *C. thermocellum* (Fig. 1, IV) belongs to glycosyl hydrolase family 5. In this family, there are only two invariant glutamic acid residues. One corresponds to the experimentally identified nucleophile (23), and, therefore, the other (noted AH in Fig. 1) is likely to be the proton donor. Mutations at this residue have strongly suggested that it is indeed the second catalytic residue (24, 25). This putative proton donor glutamic acid is located upstream of the nucleophile and appears on the HCA plots after two clusters strongly indicative of an α-helix and a β-strand (Fig. 1, IV). This glutamic acid residue always appears after an invariant asparagine. In the *Cell. fimi* β-glycanase (family 10), the proton donor has been recently identified (16) and is also a glutamic acid residue, upstream of the nucleophile, preceded by an invariant asparagine and located downstream of a large cluster corresponding to a helix followed by a strand (Fig. 1, II). Such a conserved glutamic acid residue with these features is also present in the enzymes representing the other families (1, 2, and 30) (Fig. 1, I, III, and V). It should be noted that Gräbnitz *et al.* (26) had already suggested a possible similarity between family 1 and family 5 enzymes around this residue. The equivalent residue in β-galactosidase Z of *E. coli* (Glu-461) has long been suspected to be catalytic (27); in the case of human glucocerebrosidase, the identified residue (Glu-274) constitutes a direct prediction of the active site acid catalyst. A pseudogene for human glucocerebrosidase has recently been identified in non-Gaucher disease cells (28). The two catalytic residues are both present in the encoded sequence suggesting that the protein could be functional.

The above observations were used to define "loose" motifs describing the regions around the acid catalyst and the nucleophile as follows: the first motif (acid catalyst) contains two consecutive hydrophobic clusters compatible with a helix

Table 1. Predicted catalytic machinery in various families of glycosyl hydrolases

Family	Enzyme	Source	EC no.	Swiss-Prot accession no.	Acid catalyst (Glu)	Nucleophile (Glu)	Code in Figs. 1 and 2	Substrate in Fig. 3
1	β-Glucosidase	*Agrobacterium* sp.	3.2.1.21	P12614	171	359*	I	1
	Lactase phlorizin hydrolase†	Human	3.2.1.62/108	P09848	1065	1273	Not shown	1/4
	6-Phospho β-glucosidase	*Escherichia coli*	3.2.1.86	P11988	172	361	Not shown	3
	6-Phospho β-galactosidase	*Staphylococcus aureus*	3.2.1.85	P11175	160	375	Not shown	5
	Myrosinase	*Sinapis alba*	3.2.3.1	P29092	None	426	IX	9
	β-Galactosidase	*Sulfolobus solfataricus*	3.2.1.23	P14288	209	389	Not shown	4
2	β-Galactosidase Z	*E. coli*	3.2.1.23	P00722	461	537*	III	4
	β-Glucuronidase	Human	3.2.1.31	P08236	451	540	Not shown	2
5	Endoglucanase C	*Clostridium thermocellum*	3.2.1.4	P07985	140	280*	IV	11
	β-Mannanase	*Caldocellum saccharolyticum*	3.2.1.78	P22533	162	257	Not shown	6
	Exo-1,3-β-glucanase	*Saccharomyces cerevisiae*	3.2.1.58	P23776	232	334	Not shown	1
	Cellodextrinase	*Pseudomonas fluorescens*	3.2.1.-	P27033	503	653	Not shown	11
10	β-Glycanase (Cex)	*Cellulomonas fimi*	3.2.1.8	P07986	168*	274*	II	10
30	Glucocerebrosidase	Human	3.2.1.45	P04062	274	379*	V	1
35	β-Galactosidase	Human	3.2.1.23	P16278	188	268	VI	4
39	β-Xylosidase	*Cal. saccharolyticum*	3.2.1.37	P23552	163	275	Not shown	7
	α-L-Iduronidase	Human	3.2.1.76	P35475	182	299	X	12
42	β-Galactosidase	*Bacillus stearothermophilus*	3.2.1.23	P19668	148	303	VII	4
NC	Agarase	*Pseudomonas atlantica*	3.2.1.81	P13734	200	322	VIII	8

Only one example of each EC number found in each family is displayed. These families contain a total of more than 150 glycosyl hydrolases representing to date ≈20% of all glycosyl hydrolases with known sequences. The numbering of the residues includes initiation methionine and signal peptide except for β-galactosidase Z of *E. coli*, where the numbering of the mature peptide is given. NC, sequence not classified.

*Residues unequivocally identified both in location and in role.

†Domain III of preprolactase phlorizin hydrolase.

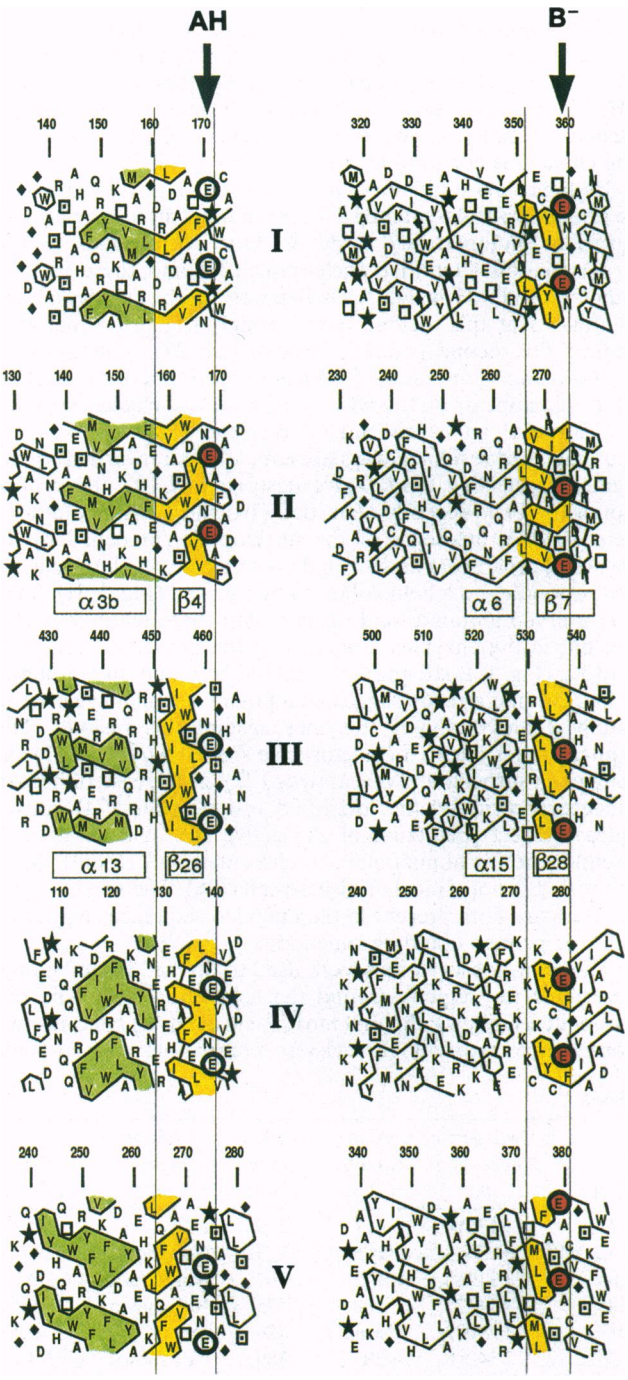

FIG. 1. Partial HCA plots of glycosyl hydrolases whose catalytic nucleophile has been experimentally identified. Enzyme codes (I–V) are defined in Table 1. The secondary structure elements found in the three-dimensional structure of II and III (14, 22) are shown as horizontal boxes. Clusters corresponding to strands have been colored yellow, and those corresponding to helices have been colored green. Vertical lines delineate the boundaries of the β-strand elements carrying the catalytic residues and correspond to the region whose significance was assessed. The nucleophilic Glu residue (B^-) and the acid catalyst Glu residue (AH) have been circled, and those that have been experimentally determined have been colored red. EC numbers, family numbers, and Swiss-Prot accession numbers are given in Table 1.

(large horizontal cluster) and a strand (short vertical cluster), respectively, followed by an Asn-Glu dipeptide; the second motif contains a glutamic acid residue preceded by a short vertical hydrophobic cluster indicative of a strand. Further selection criteria were that the two glutamic acid residues should be invariant in each of the protein families and that the acid catalyst should precede the nucleophile in the sequence.

These two motifs have then been searched for in other glycosyl hydrolase sequences. Striking similarities in the HCA plots of these two regions have been found for a number of enzymes representing a large range of substrate specificities (Fig. 2, VI–X and Table 1). The multiple sequence alignments upstream of each catalytic residues—i.e., the regions between the vertical lines in Figs. 1 and 2—have been found to be

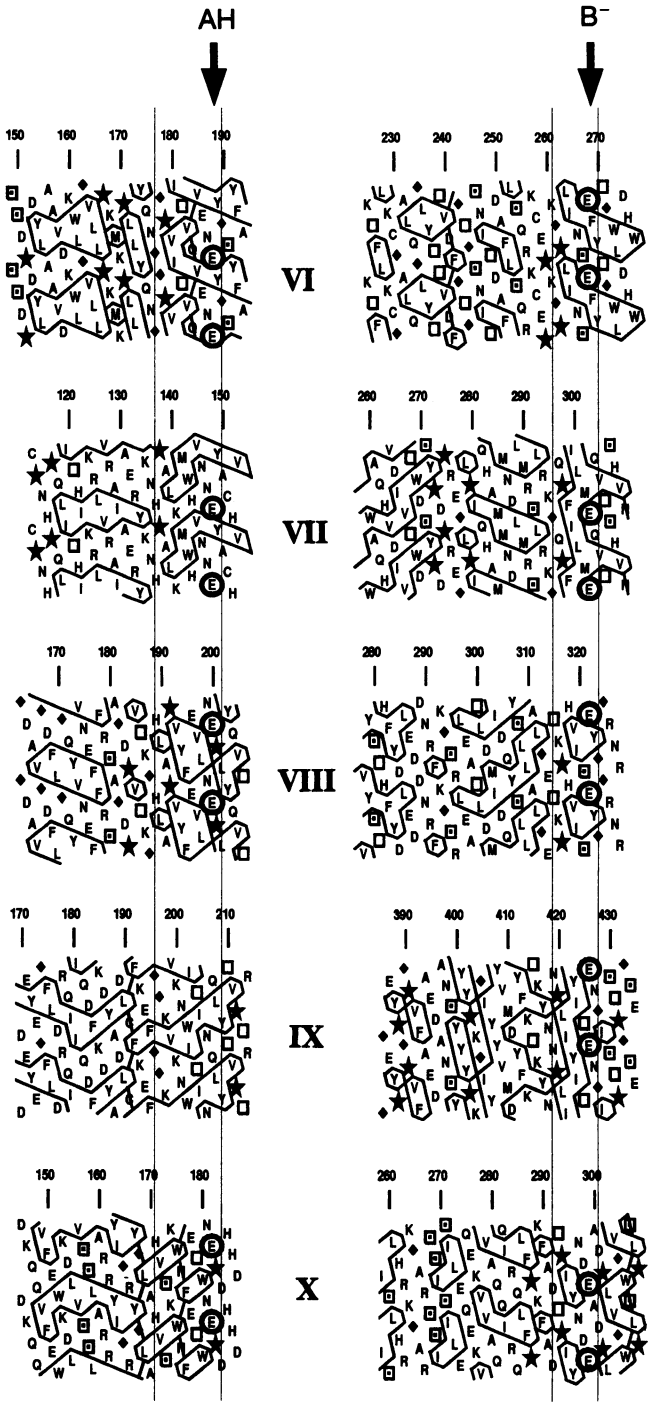

FIG. 2. Partial HCA plots of selected glycosyl hydrolases showing similarities with the catalytic regions shown in Fig. 1. Enzyme codes (VI–X) are defined in Table 1. The nucleophilic (B^-) and acid catalyst (AH) Glu residues assigned by similarity have been circled. EC numbers, family numbers, and Swiss-Prot accession numbers are given in Table 1.

individually significant (probability of occurrence by chance of only 10^{-15} and 5.3×10^{-7} for the strands carrying the acid catalyst and the nucleophile, respectively). The probability for both segments to occur by chance only is even lower. These results suggest a common ancestry and that the disordered $(\beta/\alpha)_8$ barrel structure first found in triose-phosphate isomerase (TIM-barrel) found in glycosyl hydrolases belonging to families 2 and 10 (14, 22, 29) could also prevail in the catalytic domain of members of families 1, 5, 30, 35, 39, 42, and, in an enzyme still unclassified, the agarase of *P. atlantica.*

This grouping also allows the location and role of the two catalytic residues of several biomedically important enzymes to be predicted, such as for human β-glucuronidase, β-galactosidase, and α-L-iduronidase (Table 1). For human glucocerebrosidase and lactase phlorizin hydrolase, the location of the nucleophile was already known either directly or by similarity. The present work allows the assignment of the location of the acid catalyst (Table 1).

Besides *O*-glycosyl hydrolases, family 1 also contains myrosinases which hydrolyze *S*-glycosidic bonds. Examination of the *Sin. alba* myrosinase in the regions corresponding to the catalytic residues of the other members of this family reveals that the nucleophile is indeed present, but the acid catalyst is missing (Fig. 2, IX). This observation suggests that the nucleophile alone is sufficient for the enzymatic hydrolysis of *S*-glycosidic bonds.

The active site nucleophile of retaining glycosyl hydrolases can be readily identified through the use of Withers' reagents. In some instances, however, the synthesis of a suitable 2-fluoro glycoside might prove tedious or impossible. In addition, there is at present no reagent allowing the straightforward identification of the acid catalyst. Our work makes use of existing assignments to predict the location and precise function of the two catalytic residues in ≈150 enzymes presently found in glycosyl hydrolase families 1, 2, 5, 10, 30, 35, 39, and 42.

Table 1 shows that in the enzymes that were analyzed, the number of residues between the proton donor and the nucleophile varies from ≈75 to more than 200. These variations could originate from evolutionary events leading to loops of widely different lengths between the secondary structure elements or leading to the insertion of one or more extra domains. Indeed, many glycosyl hydrolases have been found to have a multiple domain architecture with a catalytic domain attached to one or several noncatalytic domains (30). Once the precise location of the catalytic residues is known, delineating the boundaries of the β/α barrel catalytic domain in each family becomes possible, and this may reveal whether or not additional domains exist.

Although the sequences from one family to another cannot be easily aligned, except around the catalytic residues, the families grouped here are probably related by common ancestry and by fold. The combination of HCA and catalytic information can detect structural similarities at sequence divergence levels usually revealed after three-dimensional structure determination. Only three residues are indeed invariant in the present grouping. This achievement reflects the intense evolutionary constraints on catalytic residues in glycosyl hydrolases and the sensitivity of the HCA method.

The large variety of substrates hydrolyzed by the enzymes that have been grouped here is reflected by the 16 EC numbers represented (Table 1). One should, however, note the stereochemical similarity of all the substrates around the scissile bond (Fig. 3). Indeed, all the substrates can be similarly hydrolyzed by a mechanism with the acid catalyst and the nucleophile located above and below the sugar plane, respectively. In other words, all these substrates can react with the same catalytic machinery oriented in a similar fashion with respect to the plane of the sugar ring. There are other families of retaining glycosyl hydrolases, however, that have totally different folds. Glycosyl hydrolases thus provide good examples to study both divergent and convergent evolution at the molecular level.

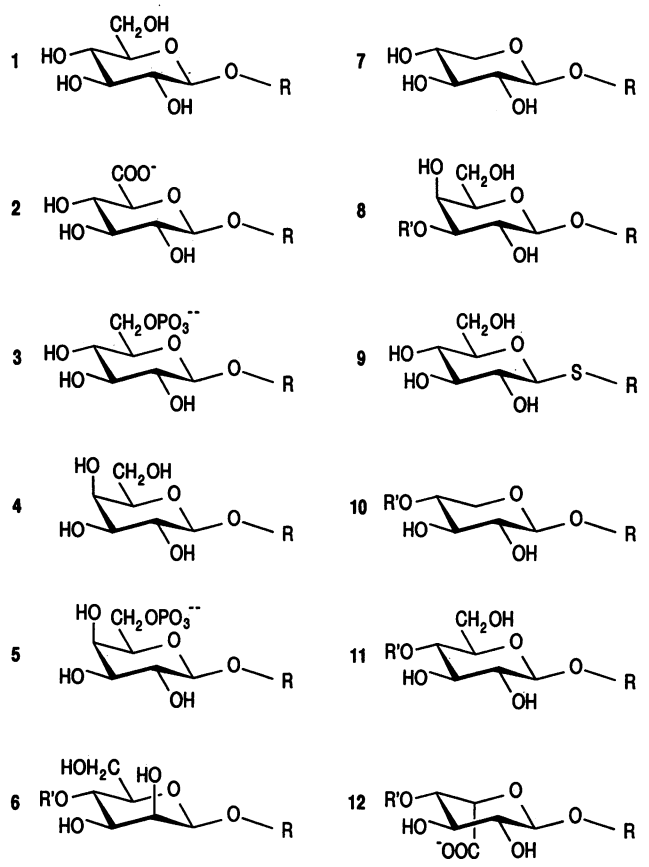

FIG. 3. Chemical structure of the substrates of the enzymes of this work at the hydrolysis site. —OR (except in 9 where it is —SR) represents the glycosidic bond that is hydrolyzed. 1, β-D-Glucoside; 2, β-D-glucuronide; 3, 6-phospho β-D-glucoside; 4, β-D-galactoside; 5, 6-phospho β-D-galactoside; 6, mannan (β-D-mannoside with R′ = β-D-mannosyl)$_n$; 7, β-D-xyloside; 8, agarose (β-D-galactoside with R′ = 3,6-anhydro-α-L-galactosyl)$_n$; 9, thio glucoside; 10, xylan (β-D-xyloside with R′ = β-D-xylosyl)$_n$; 11, cellulose (β-D-glucoside with R′ = β-D-glucosyl)$_n$; and 12, dermatan sulfate (α-L-iduronate with R′ = 2-acetamido-2-deoxy-β-D-galactosyl 4-sulfate)$_n$.

There are glycosyl hydrolases belonging to other families than those grouped here and whose three-dimensional structures are also TIM-barrels: family 13 α-amylases, family 14 β-amylases, family 17 β-1,3-1,4-glucanases and β-1,3-glucanases, and family 18 chitinases. Families 13, 14, and 18 enzymes were not found by our motif search for one or several of the following reasons: (*i*) their catalytic machinery is not composed of two glutamic acid residues, (*ii*) the invariant asparagine residue preceding the acid catalyst is lacking, and (*iii*) the mechanism is "inverting." These negative results further strengthen the idea of a common ancestry and fold for the eight families grouped here. The case of family 17 β-1,3-1,4-glucanases and β-1,3-glucanases is more intriguing, since the two motifs were indeed found in all the sequences but one. In these enzymes, the nucleophile has been previously identified by active-site labeling with an epoxyalkyl disaccharide and the proton donor on the basis of chemical modification with carbodiimide (31). Our motif analysis places the nucleophile exactly at the position determined by active-site labeling. However, our analysis places the proton donor at Glu-93 and Glu-94 of the barley β-1,3-1,4-glucanases and β-1,3-glucanases, respectively. Glu-288, which was proposed to be catalytic in both enzymes (31), is not totally invariant and appears to be at a distance from the nucleophile substantially larger than that expected for retaining glycohydrolases. On the other hand, the

residue picked by our motif analysis is on strand β4 at a distance of ≈5.5 Å from the nucleophile, a distance corresponding exactly to that found in retaining glycosyl hydrolases.

This work demonstrates the remarkable versatility of the TIM-barrel fold, which evolution has finely tuned to generate many different substrate specificities around the same catalytic reaction with a similar disposition of identical catalytic residues on the same ancestral structure. This versatility suggests an explanation for the large overrepresentation of the TIM-barrel in protein folds (5) and that a wide range of artificial enzymes could be designed and engineered on the basis of this fold.

Note. After this manuscript was submitted, a grouping of glycosyl hydrolases families similar to ours, albeit without including families 30 and 35, was proposed (32).

We are indebted to S. Armand and P. Lamotte (Grenoble) for their help in sequence retrieval and the production of HCA plots. This work was supported by a Conception de Macromolécules Assistée par Ordinateur research grant from Organibio and from the Ministère de la Recherche. We also acknowledge the support from the Association Vaincre les Maladies Lysosomales and from the European Community Biotechnology Programme.

1. Semenza, G. & Auricchio, S. (1989) in *The Metabolic Basis of Inherited Diseases*, eds. Scrives, C. R., Beaudet, A. L. & Valle, V. S. (McGraw-Hill, New York), pp. 2975–2997.
2. Neufeld, E. F. (1991) *Annu. Rev. Biochem.* **60,** 257–280.
3. Laine, R. A. (1994) *Glycobiology* **4,** 759–767.
4. Chothia, C. (1992) *Nature (London)* **357,** 543–544.
5. Orengo, C., Jones, D. T. & Thornton, J. M. (1994) *Nature (London)* **372,** 631–634.
6. Henrissat, B. (1991) *Biochem. J.* **280,** 309–316.
7. Henrissat, B. & Bairoch, A. (1993) *Biochem. J.* **293,** 781–788.
8. Davies, G. J., Dodson, G. G., Hubbard, R. E., Tolley, S. P., Dauter, Z., Wilson, K. S., Hjort, C., Mikkelsen, J. M., Rasmussen, G. & Schülein, M. (1993) *Nature (London)* **365,** 362–364.
9. Divne, C., Ståhlberg, J., Reinikainen, T., Ruohonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T. & Jones, T. A. (1994) *Science* **265,** 524–528.
10. Holm, L. & Sander, C. (1994) *FEBS Lett.* **340,** 129–132.
11. Varghese, J. N., Garrett, T. P. J., Colman, P. M., Chen, L., Høj, P. B. & Fincher, G. B. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 2785–2789.
12. Koshland, D. E. (1953) *Biol. Rev. Cambridge Philos. Soc.* **28,** 416–436.
13. Sinnott, M. L. (1990) *Chem. Rev.* **90,** 1171–1202.
14. White, A., Withers, S. G., Gilkes, N. R. & Rose, D. R. (1994) *Biochemistry* **33,** 12546–12552.
15. McCarter, J. D. & Withers, S. G. (1994) *Curr. Opin. Struct. Biol.* **4,** 885–892.
16. MacLeod, A. M., Lindhorst, T., Withers, S. G. & Warren, R. A. J. (1994) *Biochemistry* **33,** 6371–6376.
17. Gaboriaud, C., Bissery, V., Benchetrit, T. & Mornon, J. P. (1987) *FEBS Lett.* **224,** 149–155.
18. Lemesle-Varloot, L., Henrissat, B., Gaboriaud, C., Bissery, V., Morgat, A. & Mornon, J. P. (1990) *Biochimie* **72,** 555–574.
19. Woodcock, S., Mornon, J.-P. & Henrissat, B. (1992) *Protein Eng.* **5,** 629–635.
20. Schuler, G. D., Altschul, S. F. & Lipman, D. J. (1991) *Proteins* **9,** 180–190.
21. Henikoff, S. & Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* **89,** 10915–10919.
22. Jacobson, R. H., Zhang, X.-J., DuBose, R. F. & Matthews, B. W. (1994) *Nature (London)* **369,** 761–766.
23. Wang, Q., Tull, D., Meinke, A., Gilkes, N. R., Warren, R. A. J., Aebersold, R. & Withers, S. G. (1993) *J. Biol. Chem.* **268,** 14096–14102.
24. Baird, S. D., Hefford, M. A., Johnson, D. A., Sung, W. L., Yaguchi, M. & Seligy, V. L. (1990) *Biochem. Biophys. Res. Commun.* **169,** 1035–1039.
25. Py, B., Bortoli-German, I., Haiech, J., Chippaux, M. & Barras, F. (1991) *Protein Eng.* **4,** 325–333.
26. Gräbnitz, F., Seiss, M., Rücknagel, K. P. & Staudenbauer, W. L. (1991) *Eur. J. Biochem.* **200,** 301–309.
27. Cupples, C. G., Miller, J. H. & Huber, R. E. (1990) *J. Biol. Chem.* **265,** 5512–5518.
28. Imai, K., Nakamura, M., Yamada, M., Asano, A., Yokoyama, S., Tsuji, S. & Ginns, E. I. (1993) *Gene* **136,** 365–368.
29. Derewenda, U., Swenson, L., Green, R., Wei, Y., Morosoli, R., Sharek, F., Kuepfel, D. & Derewenda, Z. S. (1994) *J. Biol. Chem.* **269,** 20811–20814.
30. Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C. & Warren, R. A. J. (1991) *Microbiol. Rev.* **55,** 303–315.
31. Chen, L., Fincher, G. B. & Høj, P. B. (1993) *J. Biol. Chem.* **268,** 13318–13326.
32. Jenkins, J., Lo Leggio, L., Harris, G. & Pickersgill, R. (1995) *FEBS Lett.* **362,** 281–285.