ISSN 0006-2979, Biochemistry (Moscow), 2011, Vol. 76, No. 6, pp. 622-635. © Pleiades Publishing, Ltd., 2011.
Original Russian Text © D. G. Naumoff, 2011, published in Biokhimiya, 2011, Vol. 76, No. 6, pp. 764-780.

REVIEW

Hierarchical Classification of Glycoside Hydrolases

D. G. Naumoff[1,2]

[1]*S. N. Winogradsky Institute of Microbiology, Russian Academy of Sciences, pr. 60-letiya Oktyabrya 7/2, 117312 Moscow, Russia*
[2]*State Institute for Genetics and Selection of Industrial Microorganisms, Pervyi Dorozhnyi Proezd 1, 117545 Moscow, Russia; E-mail: daniil_naumoff@yahoo.com*

Received January 21, 2011

Abstract—This review deals with structural and functional features of glycoside hydrolases, a widespread group of enzymes present in almost all living organisms. Their catalytic domains are grouped into 120 amino acid sequence-based families in the international classification of the carbohydrate-active enzymes (CAZy database). At a higher hierarchical level some of these families are combined in 14 clans. Enzymes of the same clan have common evolutionary origin of their genes and share the most important functional characteristics such as composition of the active center, anomeric configuration of cleaved glycosidic bonds, and molecular mechanism of the catalyzed reaction (either inverting, or retaining). There are now extensive data in the literature concerning the relationship between glycoside hydrolase families belonging to different clans and/or included in none of them, as well as information on phylogenetic protein relationship within particular families. Summarizing these data allows us to propose a multilevel hierarchical classification of glycoside hydrolases and their homologs. It is shown that almost the whole variety of the enzyme catalytic domains can be brought into six main folds, large groups of proteins having the same three-dimensional structure and the supposed common evolutionary origin.

DOI: 10.1134/S0006297911060022

Key words: glycoside hydrolase, CAZy, protein hierarchical classification, protein evolution, protein family, clan, superfamily, subfamily, search for homologs, TIM-barrel, SCOP, Pfam, PSI-BLAST

"The establishing of a system for classification and nomenclature can be seen as a clear indication of the 'coming of age' of a branch of science."

Alan J. Barrett, 1994 [1]

Glycoside hydrolases (glycosidases or carbohydrases, EC 3.2.1) comprise a widespread group of enzymes catalyzing glycolytic cleavage of *O*-glycosidic bond. They are attributed to catabolic enzymes of carbohydrate metabolism; during many decades they have been the objects of various biochemical investigations [2-10]. Glycoside hydrolase genes are found in almost all living organisms, the only exception being some Archaeans and some unicellular parasitic eukaryotes [11, 12]. Some viruses also encode glycosidases. About 1% of all known genes code glycosidases and their close homologs [12]. The share of these genes in the genome significantly varies depending on the organism's taxonomic position [11]. The set of glycoside hydrolases in microorganisms is much dependent on their ecological niche: glycosidase genes often undergo duplications (they can be present in the form of several paralogs [7, 13-17]), elimination (the set of genes can differ even in strains of the same species [18]), and horizontal transfer (the glycosidase phylogenetic trees usually differ greatly from those of the host organisms [14-16, 19, 20]). Evolutionarily closely related glycosidases often differ in substrate specificity [9, 11, 21, 22], while enzymes with very different enzyme activities as well as proteins devoid of enzyme activity are found among glycosidase homologs [11, 23]. Small changes in the primary structure of glycoside hydrolases are able to change their substrate specificity [24]. Many glycoside hydrolases have complex domain structure, and different domains of the same protein often have independent evolutionary history. Two and more homologous [25-27] or non-homologous [7, 28, 29] catalytic domains can be present in the same protein.

Glycoside hydrolases and their genes have become model objects for various fields in biology: investigations of yeast lysates containing invertase (EC 3.2.1.26) once

laid the foundations of enzymology [30]; the *Escherichia coli lac* operon, containing β-galactosidase (EC 3.2.1.23) gene *lacZ* that had been widely adopted for constructing cloning vectors became a kind of standard for gene expression regulation in bacteria [31]; hen's egg lysozyme (EC 3.2.1.17) [32] became one of the first proteins with experimentally determined tertiary structure (PDB, 1HEW).

The two most widespread organic substances in living organisms are cellulose and chitin [2, 33], which makes carbohydrates dominant by mass in living matter on the Earth. Some glycosidic bonds, in particular those between two glucose residues in some polysaccharides, are the chemically most stable covalent bonds between monomers in natural biopolymers [34]. Glycoside hydrolases cleaving such bonds accelerate the corresponding reaction by 17 orders of magnitude compared to spontaneous hydrolysis, which makes them one of most efficient known catalysts [34].

TRADITIONAL (ENZYMATIC) CLASSIFICATION OF GLYCOSIDE HYDROLASES

Glycoside hydrolases in a broad sense of the word (EC 3.2) combine *O*-, *N*-, and *S*-glycoside hydrolases [4, 8, 35]. However, *N*-glycoside hydrolases (EC 3.2.2) are mainly enzymes of nucleic acid metabolism and are not considered in this review. The only known *S*-glycoside hydrolase (EC 3.2.3)—thioglucosidase (EC 3.2.3.1)—was reclassified in 2001 as an *O*-glycoside hydrolase with enzymatic number EC 3.2.1.147 [36].

The great variety of known enzymatic activities of glycoside hydrolases is due to tremendous number of their natural substrates, di-, oligo-, and polysaccharides, and their derivatives. The number of theoretically possible carbohydrate structures is indeed astronomical. Thus, low molecular weight reducing hexasaccharide $C_{36}H_{62}O_{31}$ can have over 10^{12} branched and linear isomers [37]. Traditional biochemical nomenclature of glycosidases is based exclusively on their substrate specificity, and only sometimes (in the case of amylases EC 3.2.1.1 and EC 3.2.1.2) it takes into account molecular mechanism of the catalyzed reaction [35]. Moreover, only the main, most preferable substrate of any enzyme is considered. The advantage of this nomenclature is relative simplicity and long-term experience in its use, which provided for its standardization. Enzymatic activities from EC 3.2.1.1 to EC 3.2.1.165 have now been registered, and the registration of some numbers has been cancelled for various reasons [36].

Many glycosidases exhibit broad substrate specificity [3, 21], making difficult their classification. Often glycosidases along with glycoside hydrolase (EC 3.2.1) exhibit also transglycosidase activity (EC 2.4.1) [2-4, 21, 38-40]. This formally makes it possible to consider them as members of a different enzyme class, i.e. transferases. For quite a number of glycosyltransferases (more exactly, for transglycosidases) significant structural similarity was found (at the level of primary and/or tertiary structures) with some glycoside hydrolases [11]. However, these data indicative of common evolutionary origin and similarity of catalytic mechanism in enzymes of these two groups are not taken into consideration by the biochemical classification [35, 36].

Glycosidases catalyze hydrolysis of a glycosidic bond via two different mechanisms resulting either in retaining (via a double-displacement) or in inverting of the substrate anomeric configuration during reaction [4-6, 19, 41-44]. Each enzyme exhibits one of them independently of the substrate choice, and enzymatic hydrolysis of a glycosidic bond is always stereospecific. However, there are cases when hydrolysis of some low molecular weight synthetic substrates, carbohydrate derivatives or analogs, follows a mechanism different from the natural substrate hydrolysis, or hydrolysis of synthetic substrates (such as glycosyl fluorides) of both (α and β) anomeric configurations can happen [3-5, 39, 44, 45]. Knowledge of stereochemistry of hydrolysis reactions makes it possible to forecast the enzyme's ability for catalysis of a side reaction such as transfer of glycosyl residues between substrate molecules (transglycosylation). Such ability is peculiar only to glycosidases acting with retention of the substrate anomeric configuration in the reaction product [3-7, 19, 46].

Extensive experimental material has been accumulated during investigations of the mechanism of action and specificity of different glycoside hydrolases. It was shown that glycosidases exhibit absolute specificity in nearly all cases to only two elements of substrate structure, to configuration of cleaved glycoside bond (α or β) and to the size of the hemiacetal/ketal cycle of the cleaved monosaccharide residue (six-membered pyranose or five-membered furanose cycles). As to the rest, their specificity is characterized by the large-scale width [5].

For glycosidases affecting pyranose residues [4, 46-48], it was proposed to use along with the reaction mechanism (resulting either in retaining or inversion of anomeric configuration), the axial (as a rule, α) or equatorial (as a rule, β) direction of the hydrolyzed glycosidic bond, which makes it possible to distinguish four enzyme classes. A similar system of classification is also applicable to furanosidases. At the same time, glycosidases are subdivided to *syn*- and *anti*-protonators depending on spatial arrangement of the enzyme catalytic residues relative to substrate [49]. Stereochemistry of the hydrolysis reaction is caused by mutual arrangement of catalytically important groups in the enzyme active center. As a rule, enzymatic hydrolysis of a glycosidic bond needs the presence of two catalytically important amino acid residues in the active center. In the overwhelming majority of cases, carboxyl-containing residues (Asp and/or Glu) fulfill this role. Carboxyls of these two residues are usually localized at a distance of about 5-6 or 9-10 Å in the case of glycosi-

dases acting with retention and inversion of the substrate anomeric configuration, respectively. In the latter case, the longer distance is due to the necessity of simultaneous presence of substrate and a water molecule in the enzyme active center during the hydrolysis reaction [3, 4, 41-44, 49-52].

Glycosidases that hydrolyze oligo- and polysaccharides can be subdivided to two groups according to the character of substrate cleavage, to exo- and endo-type enzymes. Despite the existence of some intermediate cases, on the whole such division is quite successful [5, 7, 46].

CAZy CLASSIFICATION: GLYCOSIDE HYDROLASE FAMILIES AND CLANS

The accumulation of data on glycoside hydrolase primary structures has made it possible to elaborate a principally new system for classification of these enzymes based on comparison of amino acid sequences of their catalytic domains. This approach was first used in analysis of β-1,4-glycosidases, which made it possible to combine them in eight families [53-55]. Then all known glycoside hydrolase sequences were studied (over 300), and 36 families of protein homologs were revealed [56]. During two decades that passed since that time, the number of glycosidases with known amino acid sequences increased at least 100-fold (Table 1). The bulk of new sequences belong to the previously known families, while the others form a number of new ones [11, 29, 46, 47, 57-59]. In 1998 the glycoside hydrolase classification was extended to enzymes exhibiting transglycosidase activity. Now 120 glycosidase families (GH1-GH125) are known. The families GH21, GH40, GH41, GH60, and GH69 were omitted because more careful investigations showed that enzymes of these families were devoid of glycosidase activities [11, 47, 60]. The current classification is available on the internet at the CAZy site [11].

Table 1. Progress in glycoside hydrolase classification

Year (reference)	Number of proteins	Number of families (those in clans)	Number of clans
1989 [55]	21	6 (0)	–
1991 [56]	322	36 (0)	–
1993 [57]	482	45 (0)	–
1995 [41, 61]	...	52 (17)	5
1996 [58]	~950	57 (17)	5
1997 [46, 60]	~1500	62 (19)	5
1998 [11]	~2200	70 (26)	8
1999 [11]	>2800	77 (...)	10
2000 [11]	...	82 (33)	10
2001 [11]	...	85 (37)	11
2002 [11]	~8000	87 (40)	12
2003 [11]	...	91 (43)	13
2004 [11]	...	95 (46)	14
2005 [11]	>20 000	101 (46)	14
2006 [11]	~30 000	106 (46)	14
2007 [11]	>30 000	110 (46)	14
2008 [11]	>40 000	114 (49)	14
2009 [11]	>59 000	115 (50)	14
2010 [11]	~82 000	118 (50)	14
2011 [11]	~86 000	123 (50)	14

Note: For 1989, data on classification are shown only for β-1,4-glycosidases [55]. Statistics for 2011 include data of February 15, 2011. Three dots mean absence of data, hyphen means absence of classification.

The criterion for any protein, or more exactly, any protein domain to be attributed to some glycoside hydrolase family is the presence in its sequence of an extended region (no less than 100 amino acid residues) homologous to the sequence of at least one already known member of this family [56]. The use of this criterion resulted in inclusion into the glycoside hydrolase families of a number of enzymes devoid of glycosidase activities [23], as well as of numerous hypothetical proteins with unknown enzymatic properties [11]. The portion of uncharacterized proteins among members of the glycosidase families continuously increases due to the high rates of new genome sequencing. It was suggested to subdivide the glycosidase families with high number of known members to subfamilies or subtypes according to the level of similarity of their amino acid sequences [53, 61, 62].

At the present time, in most glycoside hydrolase families for at least one member the molecular mechanism of the catalyzed reaction is known: with inversion or retention of the anomeric configuration of the substrate [9, 11, 41, 43, 46, 49, 58, 63]. In almost all known cases, enzymes of the same family have the same mechanism [41, 43, 63], the only exception being the GH23 and GH97 families that combine retaining and inverting glycoside hydrolases [64]. Within each glycosidase family, the anomeric configuration of the cleaved glycosidic bond (α or β), or more exactly its axial or equatorial orientation, is also conservative [43, 46]. The only known exception is the GH4 glycosidase family: it includes both enzymes cleaving an α-glucosidic bond and those cleaving a β-glucosidic bond [11, 52, 65].

A higher conservation of protein tertiary structures compared to primary structures suggests a possible similarity between 3D structures of some glycosidases belonging to different families. There are quite a number of such cases, first of all concerning families containing proteins with experimentally determined tertiary structures. In some other cases, the supposed similarity between glycosidase three-dimensional structures can be judged on

the basis of remote but statistically reliable similarity of amino acid sequences of these proteins. It was supposed to combine in superfamilies the families of glycosidases that have or appear to have related tertiary structures [61]. The number of superfamilies gradually increased. This was due to three reasons. First, the number of glycosidases with known 3D structures continuously increases. Second, new sequences exhibiting statistically reliable similarity with members of two families at once (here the rule "friends of my friends are my friends" is acting [61]) appear in databases. Third, due to increased sensitivity of methods of bioinformatic sequence analysis, similarities between different glycosidase families are revealed.

Generally accepted classification of glycoside hydrolases (CAZy) based on sequence homology considers the similarity between proteins of just those families that are characterized by the same mechanism of glycosidic bond hydrolysis [11, 46, 58, 66]. By analogy with protease classification [1, 67], such families are combined in 14 presently known clans (from GH-A to GH-N; Table 2). It is assumed that all glycosidases belonging to the families of the same clan have common evolutionary origin, similar three-dimensional structure, conservative arrangement of catalytic residues (i.e. in homologous positions), and the same mechanism of glycosidic bond hydrolysis [42, 46, 47, 58-60, 63, 66].

Although one glycosidase family usually does not include enzymes with different mechanisms of glycosidic bond hydrolysis (except families GH23 and GH97), it was shown that the hydrolysis mechanism can be changed to the opposite one by replacement of just a single amino acid residue. Replacement of a nucleophilic carbonyl-containing residue (Asp/Glu) in the enzyme active center by a non-nucleophilic residue (Ala or Gly) in glycosidases that hydrolyze substrate with retention of its anomeric configuration results in inactivation of the enzyme. However, addition to the reaction medium of nucleophile molecules (azide, acetate, or formate) results in inversion, compared to the wild-type enzyme, of the mechanism of hydrolysis of 2,4-dinitrophenyl glycosides: the product of such reaction is glycosyl-azide (or acetate, or formate) with inverted anomeric configuration [44, 51, 68, 69].

Substitution of the Thr26 residue in the bacteriophage T4 lysozyme (GH24 glycosidase family) active center by a nucleophilic residue (Glu or His) altered the catalyzed reaction. It then followed the double-displace-

Table 2. Glycoside hydrolase clans

Clan	Families (GH)	Anomeric configuration	Three-dimensional structure	Reference
GH-A	1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86, 113	retained (eq.)	$(\beta/\alpha)_8$-barrel	7, 10, 11, 25, 40, 46, 48, 57, 59, 66, 88, 94, 95, 96, 99, 150, 151
GH-B	7, 16	retained (eq.)	β-sandwich	152
GH-C	11, 12	retained (eq.)	β-sandwich	19, 153
GH-D	27, 31, 36	retained (ax.)	$(\beta/\alpha)_8$-barrel	16, 56, 61, 75, 79, 103-106, 154, 155
GH-E	33, 34, 83, 93	retained (eq.)	6-bladed β-propeller	11, 156
GH-F	43, 62	inverted (eq.)	5-bladed β-propeller	78, 130-132
GH-G	37, 63	inverted (ax.)	$(\alpha/\alpha)_6$-barrel	99
GH-H	13, 70, 77	retained (ax.)	$(\beta/\alpha)_8$-barrel	79, 108, 111, 157-162
GH-I	24, 46, 80	inverted (eq.)	α + β-lysozyme	11, 140-142
GH-J	32, 68	retained (β-furanoside)	5-bladed β-propeller	57, 78, 104, 129-132, 163, 164
GH-K	18, 20, 85	retained (eq.)	$(\beta/\alpha)_8$-barrel	11, 48, 88, 95
GH-L	15, 65, 125	inverted (ax.)	$(\alpha/\alpha)_6$-barrel	11, 49, 99
GH-M	8, 48	inverted (eq.)	$(\alpha/\alpha)_6$-barrel	99
GH-N	28, 49	inverted (ax.)	$(\beta)_3$-solenoid	95, 99, 144, 145

Note: The type of substrate glycosidic bond (equatorial or axial) and its change during hydrolysis is shown in the column "Anomeric configuration".

ment pathway resulting in final retention of the substrate β-anomeric configuration in the product [52, 70, 71].

All glycoside hydrolase clans contain at least one family that includes glycosidases with known 3D structures (Table 2). Isolation of each of these clans as well as membership of concrete families in them is usually beyond doubt. Many families remain not included into the clans, and no new clan has been established during the last seven years (Table 1). Thus, at the present time the CAZy classification is two-level (family and clan) [11]. Subfamily level has been distinguished only in the case of the GH13 and GH30 families. Already over two years ago (on-line publication of October 5, 2008 [12]) a report appeared concerning subdivision of GH1, GH2, and GH5 families into subfamilies, but this has not been realized in the database [11]. Moreover, in non-peer-reviewed publications of CAZy authors in 2005-2006 there have been reports concerning distinguishing subfamilies into 19 glycoside hydrolase families [72, 73]. However, only in the case of GH1, GH7, GH13, and GH15 families have data been published allowing readers to discriminate at least to some extent members of different subfamilies [73].

HIERARCHICAL CLASSIFICATION OF GLYCOSIDE HYDROLASES

In 1988 the similarity of amino acid sequences of glycoside hydrolases later included in families (GH13 and GH31), belonging to different clans (GH-H and GH-D, respectively) was shown for the first time [74]. Later a relationship of GH13 and GH31 was confirmed [60, 75-77], and even expediency of their combination in the same clan was discussed [60]. In 1990, already before creation of the universal classification of all glycoside hydrolases [56], it was suggested to distinguish subfamilies within four of eight families described by that time [53]. In 2000 we proposed for the first time a hierarchical (subfamily/family/clan/superfamily) glycoside hydrolase classification on the example of the β-fructosidase superfamily [78]. In 2004 a similar classification was proposed for α-galactosidase and α-glucosidase superfamilies [79], while in 2006 it was extended to all glycoside hydrolases with catalytic domain in the form of a $(\beta/\alpha)_8$ barrel, with simultaneous increase up to seven in the number of levels in the classification [80]. Upper levels in this classification are "fold" and "group of similar folds" in which only uniformity of 3D structures is required from catalytic domains along with supposed common evolutionary origin. The "fold" level in our classification approximately corresponds to the "clan" level in the Pfam classification [81] and to the "superfamily" level in SCOP classification [82]. In a recent work [83] it has been suggested to combine fifty families of glycoside hydrolase catalytic domains into seven C-families (C1-C7).

Glycosidases with $(\beta/\alpha)_8$ Form Catalytic Domain

A domain with this type of 3D structure (PDB, 1TIM) was found for the first time in triose phosphate isomerase [84] (EC 5.3.1.1), and for this reason it is often called the TIM-barrel. This type of structure is one of the most often found in proteins [85-87], and it is the most frequent catalytic domain structure type in enzymes [86]. About half of known glycoside hydrolase families have such structures [11, 88]. Among all families of functionally characterized proteins with $(\beta/\alpha)_8$ barrel three-dimensional structure, the overwhelming majority are enzymes [89, 90]. Members of about half of these families exhibit hydrolase activities, mainly *O*-glycoside hydrolase [88, 89]. It should be noted that most known glycoside hydrolase activities were detected in enzymes with $(\beta/\alpha)_8$ barrel catalytic domain structure [88].

It is assumed that all domains with this type of three-dimensional structure have a common origin [17, 90]: the gene encoding their common evolutionary precursor supposedly emerged via duplication of the gene that encoded the $(\beta/\alpha)_4$ structure [85, 86, 91, 92], or even via two successive duplications of the gene encoding the $(\beta/\alpha)_2$ structure [93]. This allows one to consider all $(\beta/\alpha)_8$ domains of glycoside hydrolases as evolutionarily related, although to all appearances they do not form a monophyletic group. The most divergent are three glycoside hydrolase families (GH6, GH38, and GH57) having unusual incomplete $(\beta/\alpha)_7$ structure. In the hierarchical classification proposed by us [80] these three families are classified as a separate fold, whereas most other families are grouped in four main types within a single fold. The same type domains of different families find each other (or are expected to find in cases where it is still not checked) in most combinations upon iterative screening using the PSI-BLAST program. In the SCOP classification, glycosidases with $(\beta/\alpha)_8$- and $(\beta/\alpha)_7$-barrel type catalytic domains are classified, respectively, as "TIM beta/alpha-barrel" fold ("(Trans)glycosidase" superfamily) and "7-stranded beta/alpha-barrel" fold (superfamilies "Glycosyl hydrolase family 6, cellulases" and "Glycoside hydrolase/deacetylase") [82], whereas in the Pfam classification they belong to clans CL0058 and CL0158 [81]. In classification of C-families they are combined together in the C3 family within which 27 subgroups are distinguished [83]. There were doubts in the literature concerning common evolutionary origin of $(\beta/\alpha)_8$ and $(\beta/\alpha)_7$ glycosidases [89]. In the CAZy classification [11] some $(\beta/\alpha)_8$ glycosidase families are combined in four clans (GH-A, GH-D, GH-H, and GH-K); all of them contain only enzymes that retain anomeric substrate configuration during the catalyzed reaction (Table 2). Nagano et al. [48] in 2001 combined nine glycosidase families (including members of three different clans) with experimentally determined at that time $(\beta/\alpha)_8$-type 3D structures in six functional subgroups

(F1-F6) and at the higher hierarchical level in four structural groups (S1-S4).

Type I classical $(\beta/\alpha)_8$ glycosidases. This type includes the clan GH-A that now combines 18 glycoside hydrolase families (Table 2). The corresponding enzymes use substrates with equatorially oriented glycosidic bonds and retain the substrate anomeric configuration in the product. The GH-A clan is also known as the 4/7 superfamily of glycoside hydrolases [94, 95] by the arrangement of two catalytically important Glu residues, proton donor/acceptor and nucleophile, at the end of the $(\beta/\alpha)_8$ barrel fourth and seventh β-strands, respectively. Several other glycoside hydrolase families (GH14, GH29, GH44, and GH71) were also included in this type [80]. The GH14 family was included according to its high similarity with GH5 [48], GH10 [40, 94, 95], and GH42 [81, 88] families, while families GH29, GH44, and GH71 exhibited similarity with the GH2 and GH5 families [88]. Proteins of family GH44 exhibit similarity with members of the GH30 family [96]. The type I classical $(\beta/\alpha)_8$ glycosidases can also include family GH98 on the basis of its similarity with GH1 and GH5 [97] and GH99 family by similarity with GH71 (unpublished data). It should be noted that enzymes of GH14, GH71, and GH98 glycosidase families change the substrate anomeric configuration during hydrolysis. The clan GH-A families (GH1, GH2, GH5, GH10, and GH17) were included in the functional subgroup F2, and GH14 family was included in subgroup F3. These two subgroups form structural group S2 [48]. Thus, type I classical $(\beta/\alpha)_8$ glycosidases approximately correspond to group S2.

For GH1 family three independent subdivisions to subfamilies were proposed: according to one proposal, five subfamilies were distinguished [25], while according to two others fourteen subfamilies were distinguished [62, 73]. The GH2 family was shown to consist of several isolated groups of proteins [40]. The GH5 family was first separated into five subfamilies [53], and later their number was increased to eight [98]. In the GH30 family two protein groups consisting of eight subgroups were distinguished [96], which are now considered in the CAZy classification as eight subfamilies [11]. Comparison of glycosidases from the GH30 and GH59 families revealed their high similarity [25], which makes it possible to consider these two families as subfamilies of one family GH30/59. Recently the border between the GH5 and GH30 families was changed by transfer of some subfamilies from GH5 to GH30 [11, 96].

Iterative screening of amino acid sequence database using the GH5 family proteins made it possible to trace their evolutionary relationships with members of the GH13 and GH14 glycoside hydrolase families [48], using the GH14 family proteins revealed their relationship with proteins of the GH35, GH42, and some other families of the GH-A clan [99], whereas screening using proteins of the GH50 family found relationship with proteins of the GH42 family [88].

Type II classical $(\beta/\alpha)_8$ glycosidases. This type includes α-galactosidase (clan GH-D; Table 2) and α-glucosidase (clan GH-H) superfamilies as well as glycoside hydrolase families GH66, GH97, GH101, and GH114 [80, 100]. This type of enzymes affect axially oriented glycosidic bond, and almost all of them (the only exception is the GH97 family) retain anomeric configuration of the substrate during hydrolysis [11]. On the basis of homology, this type can also include a variety of families of enzymatically uncharacterized proteins (hypothetical glycosidases) [75, 80, 101, 102]. The problem of necessary combination of GH-D and GH-H clans at a higher hierarchical level was discussed in work [77]. The GH13 family was included into functional subgroup F1 forming structural subgroup S1 [48]. Since this was the only family of classical $(\beta/\alpha)_8$ glycosidases for which 3D structure was already known at that time, we can conditionally identify type II with group S1.

α-Galactosidase superfamily (clan GH-D). The term "α-galactosidase superfamily" was first proposed in 2001 (formerly the term "DAG family" was used [103]) to join glycoside hydrolase families GH27 and GH36 [104] that comprised clan GH-D [11] and included most of the then known α-galactosidases (EC 3.2.1.22). At that time GH27 and GH36 families were sufficiently clearly differentiated from each other [103]. Then it was possible to distinguish within the GH36 family two subfamilies [104] (later called GH36A and GH36B [105]). In 2002 [106] it was suggested to include families GH31 and aGalT (now GH36C) [105] into this superfamily. The aGalT (from α-galactosyltransferase) family was composed [106] of a group of mainly plant origin proteins for which α-galactosyltransferase activities (EC 2.4.1.67 and EC 2.4.1.82) were demonstrated. After publication of *Sulfolobus solfataricus* genomic sequence (GenBank, AE006641), in summer 2002 the protein encoded by this archaean (GenPept, AAK43227), together with a number of other members of aGalT family, was included in CAZy classification into GH36 family [11]. However, according to the level of similarity of amino acid sequences, proteins of the aGalT family are approximately equidistant from GH27 and GH36 glycoside hydrolase families, having no more than 20% identical amino acid residues. Such expanded interpretation of the GH36 family later caused incorporation into it of an even greater number of only distantly related proteins. For example, sequencing of the α-*N*-acetylgalactosaminidase (EC 3.2.1.49) gene from *Clostridium perfringens* [107] resulted in inclusion into this family of an additional group of homologs (now GH36D [105]). After that, in CAZy classifications [11] practically all proteins of the GH-D clan, except artificially isolated but relatively homogeneous family GH27, were included into the GH36 family. Already in 2003 the absence of monophyletic status was established in proteins classified at that time in GH36 family, and its subdivision to four subfamilies (GH36A-GH36D) was suggest-

ed [105]. In 2004 [79] it was suggested to consider these subfamilies as separate families within the α-galactosidase superfamily (clan GH-D). Now eleven families (GH36A-GH36K; Table 3) are distinguished on the basis of the former GH36 family [16, 75, 100, 101], but only some of them contain proteins with experimentally identified enzymatic activities. At the beginning, only three subfamilies were distinguished within the GH27 family [16, 79], but later their number increased to six [100]. Within the GH31 family 38 subfamilies were distinguished in advance [75, 100], but some of them will probably be combined into larger ones, because they have no monophyletic status. About three years ago in the CAZy classification the GH31 family was included in GH-D clan along with families GH27 and GH36 [11]. In the Pfam classification three families (PF01055 (GH31), PF02065 (GH27 and partially GH36), and PF05691 (GH36C)) correspond to clan GH-D [81].

Iterative screening of amino acid sequence database using α-galactosidase superfamily proteins revealed their evolutionary connections with a number of proteins from different families such as GH5, GH13, GH66, GH97, GH101, GH114, COG1306, COG1649, and COG2115 [17, 75, 76, 79, 80, 100]. It should be noted that COG2115 is the xyloseisomerase family (EC 5.3.1.5).

α-Glucosidase superfamily (clan GH-H). The term α-glucosidase superfamily was first suggested in 2004 to combine glycoside hydrolase families GH13, GH70, and GH77 forming the GH-H clan [79]. As follows from databases of orthologous microbial (COG) and eukaryotic (KOG) protein groups, in both cases proteins of the GH13 family are combined into four clusters COG0296, COG0366, COG1523, COG3280, KOG0470, KOG0471, KOG2212, and KOG3625 [108]. Family GH13 is one of largest glycoside hydrolase families (>9000 proteins [11]). It includes enzymes with 22 types of enzymatic activities (including hydrolases, transferases, and synthases) as well as amino acid transporters [11]. It is one of two glycoside hydrolase families for which subdivision into subfamilies is given on the CAZy site: 36 subfamilies (from GH13_1 to GH13_36) combine some proteins from the GH13 family [11]. Initially, in 2006 it was proposed to distinguish 35 subfamilies [109], but later their number increased to 40 [73]. Comparison of proteins from the 36 subfamilies pre-

Table 3. Subdivision of GH36 family into 11 new families (GH36A-GH36K)

Protein family	Member					Total number of proteins	
	Number in GenPept	Organism	Length	Fragment	Annotation	GenPept	CAZy
GH36A	AAG49420.1	*Geobacillus stearothermophilus*	729	325-626	α-galactosidase AgaA	795	351
GH36B	CAA04514.1	*Thermotoga maritima*	552	180-482	α-galactosidase	104	50
GH36C	AAD02832.1	*Cucumis sativus*	784	198-568	raffinose synthase	281	127
GH36D	AAM55479.1	*Clostridium perfringens*	629	213-538	α-*N*-acetylgalactosaminidase	64	26
GH36E	EFB00895.1	*Victivallis vadensis*	654	268-556	hypothetical protein	20	1
GH36F	EDM35914.1	*Pedobacter* sp. BAL39	684	278-590	α-galactosidase	29	8
GH36G	ABF39669.1	*Candidatus Koribacter versatilis*	684	279-593	glycoside hydrolase, clan GH-D	9	3 + 1
GH36H	BAC74977.1	*Streptomyces avermitilis*	610	196-494	hypothetical protein	15	3
GH36I	ABW01612.1	*Caldivirga maquilingensis*	708	214-530	hypothetical protein	15	0
GH36J	ADB29787.1	*Kribbella flavida*	569	256-521	hypothetical protein	1	0
GH36K	CBL19648.1	*Ruminococcus* sp. SR1/5	561	201-500	α-galactosidase	6	3

Note: Length means total number of amino acid residues in a precursor protein. Fragment shows ordinal numbers of the first and last amino acid residues corresponding to the hypothetical catalytic domain. Total number of proteins means the number of non-identical amino acid sequences in the databases containing this family domain. GenPept means the number of proteins found in the GenPept (NCBI) database using search by homology versus non-redundant protein sequences, carried out on January 13, 2011. CAZy means the number of proteins from preceding columns shown in the CAZy database for the same date [11]. In CAZy all of them are classified in the GH36 family. The only exception is one of four proteins of the GH36G family (GenPept, CAR57390.1), that is in the list of unclassified glycoside hydrolases (known also as an arbitrary family GH0).

sented on the CAZy site [11] revealed [108] dissimilarity between levels of distinctions among proteins of different subfamilies. Three pairs of families (GH13_15 and GH13_24, GH13_20 and GH13_21, as well as GH13_29 and GH13_31) were very similar to each other and can be considered as three subfamilies: GH13_15/24 (or KOG2212), GH13_20/21, and GH13_29/31. Subfamilies GH13_25 and GH13_33 were evolutionarily more distant from the other subfamilies of this family than the GH70 family, which made possible to consider GH13_25 (KOG3625 or GL3R2899 in the Génolevures database) and GH13_33 as independent families within the α-glucosidase superfamily. It was suggested to distinguish within the GH13 family 10 more subfamilies: from GH13_A to GH13_J [108]. The GL3C0220 family from the Génolevures database of yeast genomes can be considered as the eleventh subfamily [14, 110]. Eleven subfamilies were distinguished in the GH77 family [100], and the GH70 family subdivision to subfamilies was recognized as unsuitable due to its high homogeneity [111].

Iterative screening of a database of amino acid sequences using proteins of GH13 family revealed their evolutionary connections with a number of proteins from different families: GH31, GH36D, GH70, COG1649, COG2342, GHL4, and GHL5 [101, 108]. By the level of amino acid sequence similarities [102] the GHL5 family of experimentally uncharacterized prokaryotic proteins can be included into the α-glucosidase superfamily along with glycoside hydrolase families GH13, GH13_25, GH13_33, GH70, and GH77.

GH97 family. This is a unique family that combines retaining α-galactosidases (EC 3.2.1.22) and inverting α-glucosidases (EC 3.2.1.20) [11, 64, 69, 112, 113]. It became possible to distinguish five subfamilies within the GH97 family [17, 114]. Iterative screening of amino acid sequence database using the GH97 protein family revealed [17] their evolutionary connections with a number of proteins from different glycoside hydrolase families such as GH13, GH20, GH27, GH31, and GH36, as well as with proteins of the COG0535 family (hypothetical Fe-S oxidoreductase family). Analysis of 3D structures showed that glycosidases of the GH27, GH36, GH13, GH31, GH84, and GH5 families have ten structures best resembling the GH97 domain [112].

GH101 family. Catalytic domains of this family have distorted $(\beta/\alpha)_8$ barrel structure [102]. Six subfamilies are distinguished within the GH101 family [102, 115]. Iterative amino acid sequence database screening using proteins of this family revealed their evolutionary relationships with a number of proteins from different families such as GH13, GH20, GH27, GH29, GH31, GH36, GH66, GH70, GH97, COG1306, COG1649, GHL1, GHL2, GHL3, GHL4, and GHL5 [101, 102, 115]. The GH13 family is evolutionarily the closest glycoside hydrolase family to GH101 [102, 115]. On the basis of amino acid sequence similarity, the GH101 family can be combined in one superfamily with three families of experimentally uncharacterized proteins GHL1-GHL3.

Families GH114 and *COG2342.* Phylogenetic analysis showed that it is impossible to subdivide the GH114 (or COG3868) family to clearly isolated subfamilies [101, 116]. On the basis of similarity of their amino acid sequences, families GH114 and COG2342 can be combined in one superfamily. In the Pfam classification they are considered as one family PF03537 (DUF297) [81]. The amino acid sequence database iterative screening using proteins of the GH114 family revealed their evolutionary relationships with a number of proteins from different families such as GH13, GH18, GH20, GH27, GH29, GH31, GH35, GH36A, GH36B, GH36E, GH36F, GH36G, GH36H, GH36J, GH36K, GH66, COG1306, COG1649, COG2342, as well as from an additional 13 GHL families (GHL3-GHL15) [101]. The use of proteins of COG2342 for database iterative screening revealed their evolutionary connections with proteins of families GH5, GH13, GH20, GH27, GH29, GH31, GH35, GH36A, GH36B, GH36F, GH36G, GH36H, GH42, GH66, GH97, GH101, GH114, COG1306, COG1649, GHL3, and GHL4 [116].

COG1649 family. Four main subfamilies were distinguished within family COG1649 (DUF187) [80]. Evolutionary connections with proteins of the GH13, GH20, GH31, and GH36 glycoside hydrolase families were found [80], which made it possible to consider them as a family of hypothetical glycosidases.

GHL1-GHL15 families. Fifteen new families GHL1-GHL15 of hypothetical proteins were found during iterative search of homologs of glycosidases from families GH13 [108], GH101 [102], and GH114 [101] using the PSI-BLAST program. The use of members of these 15 new families, in turn, made it possible to trace their evolutionary connections both between some of them and with proteins of the GH5, GH13, GH13_33, GH17, GH18, GH20, GH27, GH29, GH31, GH35, GH36A, GH36B, GH36C, GH36D, GH36E, GH36F, GH36G, GH36H, GH36J, GH36K, GH39, GH42, GH53, GH66, GH97, GH101, GH107, GH112, GH114, DUF3111 (PF11308 [81]), COG1082, COG1306, COG1649, and COG2342 families (unpublished).

Type III classical $(\beta/\alpha)_8$ glycosidases. This type includes GH-K clan (Table 2) and glycoside hydrolase families GH25 and GH84 [80]. The GH25 family was included on the basis of its high similarity with the GH18 family [82], and the GH84 family with GH20 [88]. Glycosidases of the GH18 family belong to functional subgroups F4 and F5 forming structural group S3, while enzymes of the GH20 family are included in subgroup F6 (group S4) [48]. A detailed hierarchical classification was suggested for proteins of the GH18 family, which is based on subdivision into three large subfamilies [27]. Eukaryotic proteins of the GH20 family form three subfamilies [117].

Type IV classical $(\beta/\alpha)_8$ glycosidases. This type includes glycoside hydrolase families GH67 and GH89 [80] on the basis of data presented in work [88]. Three subfamilies were distinguished within the GH67 family [118].

Other families of classical $(\beta/\alpha)_8$ glycosidases. The GH3 and GH56 glycoside hydrolase families with $(\beta/\alpha)_8$ catalytic domain structure are not classified in any of the four above-mentioned types [80]. The GH3 family was subdivided to three subfamilies, four clusters were distinguished within one of them, and eight protein subclusters within three of these clusters [119].

Catalytic domains of the GH112 family have structure of a partially broken $(\beta/\alpha)_8$ barrel, and structurally they are most closely related to domains of the GH42 glycoside hydrolase family [120]. Remote similarity between proteins of the GH29, GHL4, GHL6, and GHL11 families and the GH107 family glycosidases (unpublished) was detected, which suggests that the catalytic domain of the GH107 family has the $(\beta/\alpha)_8$ barrel structure. The predicted secondary structure of β-*N*-acetylgalactosaminidase of *Paenibacillus* sp. suggests that the catalytic domain of this enzyme, belonging to the recently described family GH123, also has tertiary structure in the form of $(\beta/\alpha)_8$ barrel [121].

It is reasonable to consider within the framework of glycoside hydrolase classification the PF11790 family of enzymatically uncharacterized proteins, which is included in clan CL0058 within the Pfam classification and is closest to the GH39 family [81].

Glycosidase families with incomplete $(\beta/\alpha)_7$ structure. As mentioned above, the glycosidase catalytic domains having this unusual variant of three-dimensional structure are classified in the SCOP and Pfam databases independently of glycosidase domains with classical $(\beta/\alpha)_8$ barrel. In the SCOP, domains of the GH6, GH38, and GH57 families are classified in the 7-stranded beta/alpha-barrel fold [82], while in the Pfam classification only GH38 and GH57 are entered into clan CL0158, whereas family GH6 (PF01341) is not included in any existing clan [81]. Probably, this group of three families can also include the recently distinguished family GH119 whose close relationship to GH57 is suggested [11].

The PSI-BLAST program revealed the relationship of the GH38 and GH57 families; the catalytically important nucleophilic residue is in homologous position in the corresponding proteins [122]. There are also data concerning homology of the GH13 and GH57 glycosidase families [123], which points to common evolutionary origin of $(\beta/\alpha)_7$ and $(\beta/\alpha)_8$ barrels.

The II and III α-mannosidase families [124] can be considered as two subfamilies within the GH38 family. Seven main subfamilies were distinguished within with the GH57 family [125]. Family GH6 was subdivided to eight subfamilies [126].

Similarity between 3D structures of GH6 family glycosidases and xyloseisomerases was demonstrated [40].

Glycosidases with β-Propeller Fold

Domains with this type of 3D structure can be divided into five variants depending on the number of blades in the β-propeller (from four to eight). In the SCOP classification these five variants are considered as five different folds [82]. It is assumed that all or almost all β-propellers, independently of the number of blades, are homologs [127]. The gene of their common evolutionary precursor encoded a protein corresponding to one blade of β-propeller [127, 128]. Among glycoside hydrolase catalytic domains there are β-propellers with five, six, and seven blades. Almost all of them form a compact (supposedly monophyletic) group among all β-propellers, and glycosidases with equal number of blades in the propeller are usually more similar to each other [127]. In the CAZy classification [11] they are included in 11 families, some of which form three clans (GH-E, GH-F, and GH-J; Table 2). In the classification of the C-families they are combined in the C7 family [83].

Five-bladed β-propellers (furanosidase superfamily). This variant of β-propellers is very rare and appears almost only among catalytic domains of glycoside hydrolases. Depending on the used mechanism, glycosidases of this group belong to GH-F (inverting mechanism) and GH-J (retaining mechanism) clans. Historically this is the first glycoside hydrolase group for which hierarchical classification was proposed [78].

Comparison of amino acid sequences of β-fructosidases (GH32 family) and β-fructosyltransferases (later classified as the GH68 family) revealed numerous sites of local similarity, which made it possible to combine them in the β-fructosidase superfamily [129]. The latter was recognized by the CAZy classification [11] where both these families formed clan GH-J (Table 2). Multiple alignment of amino acid sequences of all proteins of the GH68 family showed that the most extended conservative region is the L13 sequence fragment that contains a proton donor from the enzyme active center [130]. The use of a generalized (i.e. consensus) amino acid sequence of β-fructosyltransferase L13 fragment allowed to detect its homology with the region of sequences of glycoside hydrolase family GH43 by the blastp program [131]. By that time, the GH43 and GH62 families were already combined in clan GH-F [11], for which reasons it was first suggested that there is an evolutionary relationship between two different glycosidase clans (GH-F and GH-J). The use of GH43 family proteins as a query for the iterative screening of amino acid sequence database confirmed their homology with GH32, GH62, and GH68 glycoside hydrolase families as well as with experimentally uncharacterized proteins from the new GHLP family (later called COG2152 and DUF377) [78, 130, 131]. These data suggest the inclusion of the GH-F clan families into the β-fructosidase superfamily. Keeping in mind that the common property of all enzyme families of this

superfamily is the ability to cleave the substrate's monosaccharide residues in furanosidase form (α-*L*-arabinofuranosidase in the case of GH-F clan families and β-*D*-fructofuranosidase in the case GH-J clan families), it was renamed to the furanosidase superfamily [130]. On the basis of homology, this superfamily also includes the GHLP family [78, 130, 131]. In the Pfam classification this superfamily corresponds to clan CL0143 [81], while in the SCOP classification it corresponds to the "Arabinanase/levansucrase/invertase" superfamily that also includes family DUF1861 [82]. The latter is evolutionary more close to the GHLP family (unpublished).

Four, seven, and two main subfamilies were distinguished, respectively, within the GH32, GH43, and GH68 families [130, 132]. However, the two most divergent proteins of the GH43 family were not classified to subfamilies. Among them there is the experimentally uncharacterized protein from bacterium *Streptomyces coelicolor* (GenPept, CAB61805.1), which was later included into the recently formed glycoside hydrolase family GH117 [11], but in the Pfam classification the GH43 and GH117 families continue to be considered as one family (PF04616) [81].

Six- and seven-bladed β-propellers. Six-bladed β-propellers form catalytic domains of clan GH-E glycoside hydrolases combining four families GH33, GH34, GH83, and GH93 (Table 2). The same structure is characteristic of glycosidases of the GH58 family. Seven-bladed β-propellers were found in the GH74 family. In the Pfam classification all these families are considered as clan CL0434 [81]. Comparison of 3D structures of isolated blades of β-propellers reveals their high similarity in glycosidases of families GH33, GH58, and GH74 [127], thus confirming the validity of combining in one group of six- and seven-bladed β-propellers. In the SCOP classification, glycoside hydrolases with the six- and seven-bladed β-propeller structure were classified, respectively, in "Sialidases" and "Oligoxyloglucan reducing end-specific cellobiohydrolase" superfamilies [82].

Glycosidases with Catalytic Domain in the Form of β-Sandwich (β-Jelly Roll Fold)

This type of 3D structure is represented in glycoside hydrolases of two clans, GH-B and GH-C, both of which combine enzymes active against equatorially oriented glycosidic bonds with retaining of their anomeric configuration in the hydrolysis products (Table 2). In the Pfam classification families of these two clans are included in clan CL0004 [81], in the SCOP classification they belong to Concanavalin A-like lectins/glucanases superfamily [82], while in the C-family classification they belong to two families, C4 and C5 [83]. The relationship of clans GH-B and GH-C was demonstrated by comparison of 3D structures of catalytic domains of the GH11 and GH16 families [133]. The catalytic domain in the form of β-sandwich was also found in the GH54 family [134]. The detailed hierarchical classification was suggested for the GH16 protein family [133, 135], and several subfamilies were distinguished in each of the GH7 and GH12 families [22]. An alternative classification distinguishes in the GH7 family just two subfamilies [73].

Glycosidases with Catalytic Domain in the Form of $(\alpha/\alpha)_6$ Barrel

This type of 3D structure is represented in glycoside hydrolases of three clans, GH-G, GH-L, and GH-M (Table 2). All corresponding enzymes invert the substrate anomeric configuration in the hydrolysis products. The same 3D structure is characteristic of glycosidase families GH9, GH47, GH76, GH78, GH88, GH92, GH94 (formerly GT36), GH95, GH100, GH105, and GH116, as well as polysaccharide lyases of the PL10 family. In the Pfam classification clan CL0059 corresponds to them [81], while in the C-family classification the C6 family corresponds to them [83]. In the SCOP classification glycosidases of this group are classified in two superfamilies, "Six-hairpin glycosidases" and "Seven-hairpin glycosidases", within "alpha/alpha toroid" fold [82]. It should be noted that the catalytic domain of the GH47 family proteins has a modified $(\alpha/\alpha)_7$ barrel structure. Phylogenetic analysis distinguishes four protein groups within this family [136]. Three subfamilies were distinguished within the GH8 family [137], four subfamilies within GH15 [73], and three structural subclasses of plant proteins were distinguished within the GH9 family [138].

Amino acid sequence database iterative screening using proteins of the GH78 family revealed their evolutionary connections with members of glycoside hydrolase families GH94, GH63, GH15, GH37, GH95, and GH65. A less statistically reliable relationship was also found with proteins of the GH92, GH9, GH8, and GH48 families [99].

Glycosidases with Lysozyme Type Catalytic Domain

Lysozymes (EC 3.2.1.17) comprise a vast group of glycoside hydrolases with a peculiar type of 3D structure also characteristic of chitinases (EC 3.2.1.14), chitosanases (EC 3.2.1.132), and lactalbumins. In the Pfam classification clan CL0037 [81], in the SCOP classification the Lysozyme-like superfamily [82], and in the C-family classification the C1 family [83] correspond to them.

Catalytic domains of this group belong to several glycoside hydrolase families like GH19, GH22, GH23, GH24, GH46, GH73, GH80, GH103, GH104, GH108, and GH124. This group includes both types of enzymes,

those that invert (GH19, GH24, GH46, and GH124) and those that retain (GH22) equatorial orientation of the hydrolyzed glycosidic bond in the reaction products [11]. A special case is the GH23 family that contains retaining peptidoglycan lytic transglycosylases and inverting lysozymes [64].

Lysozymes were historically the first glycoside hydrolase group, on the example of which there was demonstrated the possibility of evolutionary relationship of enzymes with different mechanism of the catalyzed reaction [71, 139-141].

It should be noted that families GH24, GH46, and GH80 form clan GH-I (Table 2). Moreover, families GH46 and GH80 are so similar to each other that amino acid sequences of proteins included form a continuum. This suggested consideration of GH46 and GH80 as a united family [142]. High 3D structure similarity was shown for proteins of the GH19 and GH23 families, which raised a question concerning combination of these two families at a higher hierarchical level [140]. Two subfamilies were distinguished in each family of GH22 and GH24 [141].

Functionally poorly studied proteins of the COG5526 family structurally resemble lysozymes and are probably involved in bacterial cell wall degradation [143].

Glycosidases with Catalytic Domain in the Form of Right-Handed $(\beta)_3$ Solenoid

It is assumed that all protein domains with this type of three-dimensional structure have common evolutionary origin [95, 144]. In the Pfam classification clan CL0268 [81] corresponds to them, in the SCOP classification it is Pectin lyase-like family [82], and in the C-family classification it is the C2 family [83]. Glycoside hydrolases of the GH28, GH49, GH55, GH82, GH87, GH90, and GH91 (formerly PL19) families have $(\beta)_3$ solenoid structure. The GH28 and GH49 families form clan GH-N (Table 2). In the SCOP classification the corresponding superfamily also contains members of polysaccharide lyase families PL1, PL3, PL6, and PL9 as well as of carbohydrate esterase family CE8 [82]. Probably, glycosidases of the GH120 family also have right-handed $(\beta)_3$ solenoid structure because they exhibit similarity with members of the DUF1565 family of clan CL0268 (unpublished data).

Iterative screening of amino acid sequence database using the GH28 family proteins traced evolutionary connections with glycosidases of the GH49, GH55, GH82, and GH87 families as well as polysaccharide lyases of the PL1 family [99, 145]; proteins of the GH82 family revealed evolutionary connections with glycosidases of the GH28, GH55, and GH87 families [145].

Phylogenetic analysis distinguished several protein groups within the GH28 glycoside hydrolase family [146].

Other Glycoside Hydrolase Families

Six main folds and fold groups of glycoside hydrolase catalytic domains have been considered above. However, some glycosidases have other types of catalytic domains. Thus, in the Pfam classification [81] there are two glycosidase-containing clans, CL0063 and CL0199, most families of which do not contain enzymes exhibiting glycoside hydrolase activities.

Clan CL0063 includes the GH4 (PF02056) and GH109 (PF01408) families characterized by a mechanism using NAD^+ as cofactor [11], which is unusual for glycoside hydrolases. The GH4 family is unique because some of its glycosidases use substrates with axially oriented, while others use substrates with equatorially oriented glycosidic bond [11]. Phylogenetic analysis distinguished in the GH4 family four monophyletic protein groups characterized by distinct enzymatic activities [147].

Clan CL0199 contains families GH45 (PF02015) and GH102 (PF06725). Five subfamilies were distinguished in the GH45 family [148].

It should be noted in conclusion that almost all families of glycoside hydrolase catalytic domains were successfully combined in a small number of folds. Most likely, within each of them all domains have common evolutionary origin. However, several families remained unclassified. Most of them are recently described families for which there are no experimental data concerning 3D structure of catalytic domains. Further investigations of these families will reveal their position in the glycoside hydrolase hierarchical classification. Thus, recently a preliminary report has appeared concerning crystallization of a protein containing a domain of the GH115 glycoside hydrolase family [149]. This suggests that soon data concerning its 3D structure will appear that are necessary for classification.

When this paper was being refereed, two new glycoside hydrolase families, GH124 and GH125, appeared in the CAZy database. The GH124 family is closely related to GH23, while family GH125 is included in clan GH-L [11].

REFERENCES

1. Barrett, A. J. (1994) *Meth. Enzymol.*, **244**, 1-15.
2. Kochetkov, N. K., Bochkov, A. F., Dmitriev, B. A., Usov, A. I., Chizhov, O. S., and Shibaev, V. N. (1967) *Carbohydrate Chemistry* [in Russian], Khimiya, Moscow.
3. Krasikov, V. V., Karelov, D. V., and Firsov, L. M. (2001) *Biochemistry (Moscow)*, **66**, 267-281.
4. Sinnott, M. L. (1990) *Chem. Rev.*, **90**, 1171-1202.
5. Khorlin, A. Ya. (1974) in *Structure and Function of Enzyme Active Centers* [in Russian], Nauka, Moscow, pp. 39-69.
6. Khorlin, A. Ya. (1988) in *Chemical Encyclopedia* [in Russian], Vol. I, Sovetskaya Entsiklopediya, Moscow, pp. 575-576.

7. Rabinovich, M. L., Melnik, M. S., and Bolobova, A. V. (2002) *Appl. Biochem. Microbiol. (Moscow)*, **38**, 355-373.
8. Wiederschain, G. Ya. (1980) *Biological Bases of Glycosidoses* [in Russian], Meditsina, Moscow.
9. Rabinovich, M. L., Melnik, M. S., and Bolobova, A. V. (2002) *Biochemistry (Moscow)*, **67**, 850-871.
10. Rabinovich, M. L., and Melnik, M. S. (2000) *Uspekhi Biol. Khim.*, **40**, 205-266.
11. *Carbohydrate-Active Enzymes server* (2011) (http://www.cazy.org/).
12. Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009) *Nucleic Acids Res.*, **37**, D233-D238.
13. Hazlewood, G. P., and Gilbert, H. J. (1998) *Biochem. Soc. Trans.*, **26**, 185-190.
14. Naumoff, D. G. (2010) in *Reports of III Int. Conf. "Mathematical Biology and Bioinformatics"*, Pushchino, October 10-15, 2010, MAX Press Publishing House, Moscow, pp. 139-140.
15. Parrent, J. L., James, T. Y., Vasaitis, R., and Taylor, A. F. S. (2009) *BMC Evol. Biol.*, **9**, Art.148.
16. Naumoff, D. G. (2004) *Mol. Biol. (Moscow)*, **38**, 388-399.
17. Naumoff, D. G. (2005) *BMC Genomics*, **6**, Art.112.
18. Turakainen, H., Aho, S., and Korhola, M. (1993) *Appl. Environ. Microbiol.*, **59**, 2622-2630.
19. Tomme, P., Warren, R. A. J., and Gilkes, N. R. (1995) *Adv. Microb. Physiol.*, **37**, 1-81.
20. Garcia-Vallve, S., Romeu, A., and Palau, J. (2000) *Mol. Biol. Evol.*, **17**, 352-361.
21. Tymowska-Lalanne, Z., and Kreis, M. (1998) *Adv. Botan. Res.*, **28**, 71-117.
22. Vlasenko, E., Schulein, M., Cherry, J., and Xu, F. (2010) *Bioresour. Technol.*, **101**, 2405-2411.
23. Coutinho, P. M., Stam, M., Blanc, E., and Henrissat, B. (2003) *Trends Plant Sci.*, **8**, 563-565.
24. Andrews, S. R., Charnock, S. J., Lakey, J. H., Davies, G. J., Claeyssens, M., Nerinckx, W., Underwood, M., Sinnott, M. L., Warren, R. A., and Gilbert, H. J. (2000) *J. Biol. Chem.*, **275**, 23027-23033.
25. Mian, I. S. (1998) *Blood Cells Mol. Dis.*, **24**, 83-100.
26. Naumoff, D. G. (2007) *Mol. Biol. (Moscow)*, **41**, 962-973.
27. Karlsson, M., and Stenlid, J. (2009) *J. Mol. Microbiol. Biotechnol.*, **16**, 208-223.
28. Park, B. H., Karpinets, T. V., Syed, M. H., Leuze, M. R., and Uberbacher, E. C. (2010) *Glycobiology*, **20**, 1574-1584.
29. Coutinho, P. M., and Henrissat, B. (1999) in *Genetics, Biochemistry and Ecology of Cellulose Degradation* (Ohmiya, K., Hayashi, K., Sakka, K., Kobayashi, Y., Karita, S., and Kimura, T., eds.) University Publishers Co., Tokyo, pp. 15-23.
30. Buchner, E. (1897) *Ber. Dt. Chem. Ges.*, **30**, 117-124 (cited by English translation: http://bip.cnrs-mrs.fr/bip10/buchner0.htm).
31. Jacob, F., and Monod, J. (1961) *J. Mol. Biol.*, **3**, 318-356.
32. Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C., and Sarma, V. R. (1965) *Nature*, **206**, 757-761.
33. Seidl, V., Huemer, B., Seiboth, B., and Kubicek, C. P. (2005) *FEBS J.*, **272**, 5923-5939.
34. Wolfenden, R., Lu, X., and Young, G. (1998) *J. Am. Chem. Soc.*, **120**, 6814-6815.
35. *Enzyme Nomenclature. Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes* (1992) Academic Press, San Diego.
36. *ExplorEnz – The Enzyme Database* (2011) (http://www.enzyme-database.org/index.php).
37. Laine, R. A. (1994) *Glycobiology*, **4**, 759-767.
38. Watt, G. M., Lowden, P. A. S., and Flitsch, S. L. (1997) *Curr. Opin. Struct. Biol.*, **7**, 652-660.
39. Sinnott, M. L. (1998) *Biochem. Soc. Trans.*, **26**, 160-164.
40. Juers, D. H., Huber, R. E., and Matthews, B. W. (1999) *Protein Sci.*, **8**, 122-136.
41. Davies, G., and Henrissat, B. (1995) *Structure*, **3**, 853-859.
42. Davies, G. J. (1998) *Biochem. Soc. Trans.,* **26**, 167-173.
43. McCarter, J. D., and Withers, S. G. (1994) *Curr. Opin. Struct. Biol.*, **4**, 885-892.
44. Wang, Q., Graham, R. W., Trimbur, D., Warren, R. A. J., and Withers, S. G. (1994) *J. Am. Chem. Soc.*, **116**, 11594-11595.
45. Kasumi, T., Tsumuraya, Y., Brewer, C. F., Kersters-Hilderson, H., Claeyssens, M., and Hehre, E. J. (1987) *Biochemistry*, **26**, 3010-3016.
46. Henrissat, B., and Davies, G. (1997) *Curr. Opin. Struct. Biol.*, **7**, 637-644.
47. Coutinho, P. M., and Henrissat, B. (1999) in *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H. J., Davies, G., Henrissat, B., and Svensson, B., eds.) The Royal Society of Chemistry, Cambridge, UK, pp. 3-12.
48. Nagano, N., Porter, C. T., and Thornton, J. M. (2001) *Protein Eng.*, **14**, 845-855.
49. Vasella, A., Davies, G. J., and Bohm, M. (2002) *Curr. Opin. Chem. Biol.*, **6**, 619-629.
50. White, A., and Rose, D. R. (1997) *Curr. Opin. Struct. Biol.*, **7**, 645-651.
51. Zechel, D. L., and Withers, S. G. (2000) *Acc. Chem. Res.*, **33**, 11-18.
52. Rye, C. S., and Withers, S. G. (2000) *Curr. Opin. Chem. Biol.*, **4**, 573-580.
53. Beguin, P. (1990) *Ann. Rev. Microbiol.,* **44**, 219-248.
54. Gilkes, N. R., Henrissat, B., Kilburn, D. G., Miller, R. C., Jr., and Warren, R. A. J. (1991) *Microbiol. Rev.*, **55**, 303-315.
55. Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J.-P. (1989) *Gene*, **81**, 83-95.
56. Henrissat, B. (1991) *Biochem. J.*, **280**, 309-316.
57. Henrissat, B., and Bairoch, A. (1993) *Biochem. J.*, **293**, 781-788.
58. Henrissat, B., and Bairoch, A. (1996) *Biochem. J.*, **316**, 695-696.
59. Himmel, M. E., Karplus, P. A., Sakon, J., Adney, W. S., Baker, J. O., and Thomas, S. R. (1997) *Appl. Biochem. Biotech.*, **63-65**, 315-325.
60. Henrissat, B. (1998) *Biochem. Soc. Trans.*, **26**, 153-156.
61. Henrissat, B., and Romeu, A. (1995) *Biochem. J.*, **311**, 350-351.
62. Marques, A. R., Coutinho, P. M., Videira, P., Fialho, A. M., and Sa-Correia, I. (2003) *Biochem. J.*, **370**, 793-804.
63. Gebler, J., Gilkes, N. R., Claeyssens, M., Wilson, D. B., Beguin, P., Wakarchuk, W. W., Kilburn, D. G., Miller, R. C., Jr., Warren, R. A. J., and Withers, S. G. (1992) *J. Biol. Chem.*, **267**, 12559-12561.
64. Gloster, T. M., Turkenburg, J. P., Potts, J. R., Henrissat, B., and Davies, G. J. (2008) *Chem. Biol.*, **15**, 1058-1067.
65. Thompson, J., Ruvinov, S. B., Freedberg, D. I., and Hall, B. G. (1999) *J. Bacteriol.*, **181**, 7339-7345.

66. Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.-P., and Davies, G. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 7090-7094.
67. Rawlings, N. D., and Barrett, A. J. (1993) *Biochem. J.*, **290**, 205-218.
68. Birsan, C., Johnson, P., Joshi, M., MacLeod, A., McIntosh, L., Monem, V., Nitz, M., Rose, D. R., Tull, D., Wakarchuck, W. W., Wang, Q., Warren, R. A. J., White, A., and Withers, S. G. (1998) *Biochem. Soc. Trans.*, **26**, 156-160.
69. Okuyama, M., Kitamura, M., Hondoh, H., Kang, M. S., Mori, H., Kimura, A., Tanaka, I., and Yao, M. (2009) *J. Mol. Biol.*, **392**, 1232-1241.
70. Kuroki, R., Weaver, L. H., and Matthews, B. W. (1993) *Science*, **262**, 2030-2033.
71. Kuroki, R., Weaver, L. H., and Matthews, B. W. (1995) *Nat. Struct. Biol.*, **2**, 1007-1011.
72. Stam, M., Bernard, T., Rancurel, C., Coutinho, P. M., and Henrissat, B. (2005) in *Journees Ouvertes Biologique Informatique Mathematiques (JOBIM'2005)*, July 6-8, 2005, Lyon (http://pbil.univ-lyon1.fr/events/jobim2005/proceedings/P133Stam.pdf).
73. Stam, M. (2006) *Evolution et Prediction des Activites des Glycoside Hydrolases*: Ph. D. thesis, Universite Aix-Marseille 1, Marseille, France (http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.105.8057&rep=rep1&type=pdf).
74. Svensson, B. (1988) *FEBS Lett.*, **230**, 72-76.
75. Naumoff, D. G., and Carreras, M. (2009) *Mol. Biol. (Moscow)*, **43**, 652-664.
76. Rigden, D. J. (2002) *FEBS Lett.*, **523**, 17-22.
77. Janecek, S., Svensson, B., and MacGregor, E. A. (2007) *FEBS Lett.*, **581**, 1261-1268.
78. Naumoff, D. G. (2000) in *Abst. Fourth Int. Fructan Symp. "Fructan 2000"*, August 16-20, 2000, Arolla, Switzerland, P.1.4 (http://www.kokkinias.com/fructan/admin/abstracts/pdf/naumoff.pdf).
79. Naumoff, D. G. (2004) in *Proc. Fourth Int. Conf. on Bioinformatics of Genome Regulation and Structure*, July 25-30, 2004, Vol. 1, Novosibirsk, Russia, pp. 315-318 (http://www.bionet.nsc.ru/meeting/bgrs_proceedings/papers/2004/BGRS_2004_V1_079.pdf).
80. Naumoff, D. G. (2006) in *Proc. Fifth Int. Conf. on Bioinformatics of Genome Regulation and Structure*, July 16-22, 2006, Vol. 1, Novosibirsk, Russia, pp. 294-298 (http://www.bionet.nsc.ru/meeting/bgrs_proceedings/papers/2006/BGRS_2006_V1_067.pdf).
81. Finn, R. D., Mistry, J., Tate, J., Coggill, P., Heger, A., Pollington, J. E., Gavin, O. L., Gunasekaran, P., Ceric, G., Forslund, K., Holm, L., Sonnhammer, E. L., Eddy, S. R., and Bateman, A. (2010) *Nucleic Acids Res.*, **38**, D211-D222.
82. Andreeva, A., Howorth, D., Chandonia, J. M., Brenner, S. E., Hubbard, T. J., Chothia, C., and Murzin, A. G. (2008) *Nucleic Acids Res.*, **36**, D419-D425.
83. Roka, P., Czanik, P., Ponyi, T., and Fulop, L. (2008) *Bull. Szent Istvan Univ.*, **1**, 5-10.
84. Banner, D. W., Bloomer, A. C., Petsko, G. A., Phillips, D. C., Pogson, C. I., Wilson, I. A., Corran, P. H., Furth, A. J., Milman, J. D., Offord, R. E., Priddle, J. D., and Waley, S. G. (1975) *Nature*, **255**, 609-614.
85. Hocker, B., Jurgens, C., Wilmanns, M., and Sterner, R. (2001) *Curr. Opin. Biotechnol.*, **12**, 376-381.
86. Henn-Sax, M., Hocker, B., Wilmanns, M., and Sterner, R. (2001) *Biol. Chem.*, **382**, 1315-1320.
87. Orengo, C. A., Jones, D. T., and Thornton, J. M. (1994) *Nature*, **372**, 631-634.
88. Rigden, D. J., Jedrzejas, M. J., and de Mello, L. V. (2003) *FEBS Lett.*, **544**, 103-111.
89. Nagano, N., Orengo, C. A., and Thornton, J. M. (2002) *J. Mol. Biol.*, **321**, 741-765.
90. Farber, G. K., and Petsko, G. A. (1990) *Trends Biochem. Sci.*, **15**, 228-234.
91. Hocker, B., Schmidt, S., and Sterner, R. (2002) *FEBS Lett.*, **510**, 133-135.
92. Lang, D., Thoma, R., Henn-Sax, M., Sterner, R., and Wilmanns, M. (2000) *Science*, **289**, 1546-1550.
93. Richter, M., Bosnali, M., Carstensen, L., Seitz, T., Durchschlag, H., Blanquart, S., Merkl, R., and Sterner, R. (2010) *J. Mol. Biol.*, **398**, 763-773.
94. Jenkins, J., Leggio, L. L., Harris, G., and Pickersgill, R. (1995) *FEBS Lett.*, **362**, 281-285.
95. Pickersgill, R., Harris, G., Leggio, L. L., Mayans, O., and Jenkins, J. (1998) *Biochem. Soc. Trans.*, **26**, 190-198.
96. St John, F. J., Gonzalez, J. M., and Pozharski, E. (2010) *FEBS Lett.*, **584**, 4435-4441.
97. Rigden, D. J. (2005) *FEBS Lett.*, **579**, 5466-5472.
98. Lo Leggio, L., and Larsen, S. (2002) *FEBS Lett.*, **523**, 103-108.
99. Stam, M. R., Blanc, E., Coutinho, P. M., and Henrissat, B. (2005) *Carbohyd. Res.*, **340**, 2728-2734.
100. Naumoff, D. G. (2008) in *Materials of the Fifth Cong. of Yu. A. Ovchinnikov Soc. Russ. Biotechnologists*, December 2-4, 2008, Moscow, pp. 116-117 (http://bioros.tmweb.ru/Vcongress/Naumov.pdf).
101. Naumoff, D. G., and Stepushchenko, O. O. (2011) *Mol. Biol. (Moscow)*, **45**, in press.
102. Naumoff, D. (2010) *J. Bioinform. Comput. Biol.*, **8**, 437-451.
103. Dagnall, B. H., Paulsen, I. T., and Saier, M. H., Jr. (1995) *Biochem. J.*, **311**, 349-350.
104. Naumoff, D. G. (2001) *Glycoconj. J.*, **18**, 109.
105. Naumoff, D. G. (2003) in *Program Abst. 5th Carbohydrate Bioengineering Meet.*, April 6-9, 2003, Groningen, The Netherlands, p. 81.
106. Naumoff, D. G. (2002) in *Int. Summer School "From Genome to Life: Structural, Functional and Evolutionary Approaches"*, July 15-27, 2002, Cargese, Corsica, France, p. 40 (http://www-archbac.u-psud.fr/Meetings/cargese2002/abstracts/NAUMOFF.html).
107. Calcutt, M. J., Hsieh, H.-Y., Chapman, L. F., and Smith, D. S. (2002) *FEMS Microbiol. Lett.*, **214**, 77-80.
108. Gizatullina, D. I., and Naumoff, D. G. (2009) in *Proc. Int. Moscow Conf. on Computational Molecular Biology*, July 20-23, 2009, Moscow, Russia, pp. 249-250 (http://mccmb.belozersky.msu.ru/2009/MCCMB09_Proceedings.pdf).
109. Stam, M. R., Danchin, E. G., Rancurel, C., Coutinho, P. M., and Henrissat, B. (2006) *Protein Eng. Des. Sel.*, **19**, 555-562.
110. Naumoff, D. G., and Naumov, G. I. (2010) in *Reports of III Int. Conf. "Mathematical Biology and Bioinformatics"*, Pushchino, October 10-15, 2010, MAX Press Publishing House, Moscow, pp. 135-136.
111. Gizatullina, D. I., and Naumoff, D. G. (2008) in *The Sixth Int. Conf. on Bioinformatics of Genome Regulation and*

Structure, June 22-28, 2008, Novosibirsk, Russia, p. 82 (http://www.bionet.nsc.ru/meeting/bgrs2008/BGRS2008_Proceedings.pdf).
112. Kitamura, M., Okuyama, M., Tanzawa, F., Mori, H., Kitago, Y., Watanabe, N., Kimura, A., Tanaka, I., and Yao, M. (2008) *J. Biol. Chem.*, **283**, 36328-36337.
113. Hidaka, M. (2010) *Trends Glycosci. Glycotechnol.*, **22**, 41-44.
114. Naumoff, D. G. (2005) *FEBS J.*, **272**, 94-95.
115. Naumoff, D. G. (2009) *Glycoconj. J.*, **26**, 847.
116. Naumoff, D., and Stepuschenko, O. (2010) *FEBS J.*, **277**, 233-234.
117. Intra, J., Pavesi, G., and Horner, D. S. (2008) *BMC Evol. Biol.*, **8**, Art. 214.
118. Shallom, D., Golan, G., Shoham, G., and Shoham, Y. (2004) *J. Bacteriol.*, **186**, 6928-6937.
119. Cournoyer, B., and Faure, D. (2003) *J. Mol. Microbiol. Biotechnol.*, **5**, 190-198.
120. Hidaka, M., Nishimoto, M., Kitaoka, M., Wakagi, T., Shoun, H., and Fushinobu, S. (2009) *J. Biol. Chem.*, **284**, 7273-7283.
121. Sumida, T., Fujimoto, K., and Ito, M. (2011) *J. Biol. Chem.*, **286**, 14065-14072.
122. Imamura, H., Fushinobu, S., Jeon, B.-S., Wakagi, T., and Matsuzawa, H. (2001) *Biochemistry*, **40**, 12400-12406.
123. Janecek, S. (1998) *Folia Microbiol.*, **43**, 123-128.
124. Gonzalez, D. S., and Jordan, I. K. (2000) *Mol. Biol. Evol.*, **17**, 292-300.
125. Zona, R., Chang-Pi-Hin, F., O'Donohue, M. J., and Janecek, S. (2004) *Eur. J. Biochem.*, **271**, 2863-2872.
126. Mertz, B., Kuczenski, R. S., Larsen, R. T., Hill, A. D., and Reilly, P. J. (2005) *Biopolymers*, **79**, 197-206.
127. Chaudhuri, I., Soding, J., and Lupas, A. N. (2008) *Proteins*, **71**, 795-803.
128. Chaudhuri, I., Coles, M., Martin, J., and Lupas, A. N. (2005) *FEBS J.*, **272**, 79.
129. Naumov, D. G., and Doroshenko, V. G. (1998) *Mol. Biol. (Moscow)*, **32**, 761-766.
130. Naumoff, D. G. (2001) *Proteins*, **42**, 66-76.
131. Naumoff, D. G. (2000) in *Biosphere and the Mankind. Materials of Conf. in Commemoration of Centenary of N. V. Timofeev-Ressovsky*, September 20-21, Obninsk, pp. 116-122.
132. Pons, T., Naumoff, D. G., Martinez-Fleites, C., and Hernandez, L. (2004) *Proteins*, **54**, 424-432.
133. Strohmeier, M., Hrmova, M., Fischer, M., Harvey, A. J., Fincher, G. B., and Pleiss, J. (2004) *Protein Sci.*, **13**, 3200-3213.
134. Fushinobu, S., Hidaka, M., Miyanaga, A., and Imamura, H. (2007) *J. Appl. Glycosci.*, **54**, 95-102.
135. *The Family GH16 Glycoside Hydrolase Database* (2011) (http://www.ghdb.uni-stuttgart.de).
136. Jordan, I. K., Bishop, G. R., and Gonzalez, D. S. (2001) *Bioinformatics*, **17**, 965-976.
137. Adachi, W., Sakihama, Y., Shimizu, S., Sunami, T., Fukazawa, T., Suzuki, M., Yatsunami, R., Nakamura, S., and Takenaka, A. (2004) *J. Mol. Biol.*, **343**, 785-795.
138. Urbanowicz, B. R., Bennett, A. B., Del Campillo, E., Catala, C., Hayashi, T., Henrissat, B., Hofte, H., McQueen-Mason, S. J., Patterson, S. E., Shoseyov, O., Teeri, T. T., and Rose, J. K. (2007) *Plant Physiol.*, **144**, 1693-1696.
139. Holm, L., and Sander, C. (1994) *FEBS Lett.*, **340**, 129-132.
140. Monzingo, A. F., Marcotte, E. M., Hart, P. J., and Robertus, J. D. (1996) *Nat. Struct. Biol.*, **3**, 133-140.
141. Wohlkonig, A., Huet, J., Looze, Y., and Wintjens, R. (2010) *PLoS One*, **5**, e15388.
142. Tremblay, H., Blanchard, J., and Brzezinski, R. (2000) *Can. J. Microbiol.*, **46**, 952-955.
143. Pei, J., and Grishin, N. V. (2005) *Protein Sci.*, **14**, 2574-2581.
144. Jenkins, J., Mayans, O., and Pickersgill, R. (1998) *J. Struct. Biol.*, **122**, 236-246.
145. Rigden, D. J., and Franco, O. L. (2002) *FEBS Lett.*, **530**, 225-232.
146. Markovic, O., and Janecek, S. (2001) *Protein Eng.*, **14**, 615-631.
147. Berezina, O. V., Lunina, N. A., Zverlov, V. V., Naumoff, D. G., Liebl, W., and Velikodvorskaya, G. A. (2003) *Mol. Biol. (Moscow)*, **37**, 678-685.
148. Igarashi, K., Ishida, T., Hori, C., and Samejima, M. (2008) *Appl. Environ. Microbiol.*, **74**, 5628-5634.
149. Fujimoto, Z., Ichinose, H., Biely, P., and Kaneko, S. (2011) *Acta Crystallogr.*, **67**, 68-71.
150. Moller, P. L., Jorgensen, F., Hansen, O. C., Madsen, S. M., and Stougaard, P. (2001) *Appl. Environ. Microbiol.*, **67**, 2276-2283.
151. Henrissat, B., Teeri, T. T., and Warren, R. A. J. (1998) *FEBS Lett.*, **425**, 352-354.
152. Divne, C., Stahlberg, J., Reinikainen, T., Ruchonen, L., Pettersson, G., Knowles, J. K. C., Teeri, T. T., and Jones, T. A. (1994) *Science*, **265**, 524-528.
153. Torronen, A., Kubicek, C. P., and Henrissat, B. (1993) *FEBS Lett.*, **321**, 135-139.
154. Margolles-Clark, E., Tenkanen, M., Luonteri, E., and Penttila, M. (1996) *Eur. J. Biochem.*, **240**, 104-111.
155. Liebl, W., Wagner, B., and Schellhase, J. (1998) *System. Appl. Microbiol.*, **21**, 1-11.
156. Crennell, S. J., Garman, E. F., Laver, W. G., Vimr, E. R., and Taylor, G. L. (1993) *Proc. Natl. Acad. Sci. USA*, **90**, 9852-9856.
157. Ferretti, J. J., Gilpin, M. L., and Russell, R. R. B. (1987) *J. Bacteriol.*, **169**, 4271-4278.
158. Rojas, A., Garcia-Vallve, S., Palau, J., and Romeu, A. (1999) *Biologia*, **54**, 255-277.
159. MacGregor, E. A., Jespersen, H. M., and Svensson, B. (1996) *FEBS Lett.*, **378**, 263-266.
160. MacGregor, E. A., Janecek, S., and Svensson, B. (2001) *Biochim. Biophys. Acta*, **1546**, 1-20.
161. Mooser, G., Hefta, S. A., Paxton, R. J., Shively, J. E., and Lee, T. D. (1991) *J. Biol. Chem.*, **266**, 8916-8922.
162. Janecek, S. (2005) *Biologia (Bratislava)*, **60** (Suppl. 16), 177-184.
163. Pons, T., Olmea, O., Chinea, G., Beldarrain, A., Marquez, G., Acosta, N., Rodriguez, L., and Valencia, A. (1998) *Proteins*, **33**, 383-395.
164. Sprenger, N., Bortlik, K., Brandt, A., Boller, T., and Wiemken, A. (1995) *Proc. Natl. Acad. Sci. USA*, **92**, 11652-11656.