An overview of clan GH-H and distantly-related families

E. Ann MacGregor

2 Nicklaus Green, Livingston, West Lothian, U.K., EH54 8RX

Abstract: In 1991 Bernard Henrissat proposed a scheme for classifying glycoside hydrolases, based on amino acid sequence rather than specificity of action. Up to now, nearly one hundred families of glycosidases and transglycosidases have been identified, and some of these have been grouped into clans, where a clan contains families closely related in three-dimensional structure and catalytic mechanism. Clan GH-H consists of families 13, 70 and 77, the so-called α-amylase families, and similarities amongst enzymes of this clan are reviewed. Further, a short history of the α-amylase family up to 1991 is given. Relationships between the enzymes of this clan and families such as 14, 15, 27, 31 and 57 are explored, with emphasis on tertiary structure and active site amino acid residues.

Key words: $(\alpha/\beta)_8$-barrel, clan GH-H, α-amylase family, glycosidases.

Abbreviations: CGTase, cyclodextrin glucanotransferase; TIM, triose phosphate isomerase; RMSD, root mean square deviation.

Introduction

The classification of glycosidases and transglycosidases based on sequence, as proposed by Bernard Henrissat in 1991 (Henrissat, 1991), is now widely accepted and almost invariably cited when such enzymes are discussed. The basic unit of the classification is the family, where all the enzymes in a family have certain well-defined sequence similarities. As more sequences have become available, the original thirty five families have been expanded to nearly one hundred (Henrissat, 1991; Coutinho & Henrissat, 1999). With increasing knowledge of protein tertiary structures, a larger unit, the clan, was defined. A clan contains two or more families that have the same basic three-dimensional fold for their catalytic module. The sequence similarities between one family and another within a clan may, however, be very limited indeed. This reflects the general principle that protein structure is better preserved by evolution than amino acid sequence. Currently fourteen clans (GH-A to GH-N) have been defined for glycosidases and transglycosidases, and the enzymes of the α-amylase families, i.e. 13, 70 and 77 (sometimes known as the α-amylase superfamily), belong to the eighth of these clans, clan GH-H.

Enzymes of clan GH-H; similarities and differences

Of the three families constituting clan GH-H, family 77 is perhaps the least complex, since all the enzymes rec-ognized to date have similar specificities, i.e. are glucanotransferases that break an α-1,4 link between glucose residues and transfer the glycon section of the substrate to an acceptor glucan, with the creation of a new α-1,4 bond. This transfer can be inter- or intra-molecular; in the latter case a cyclic dextrin is formed. The reactions catalysed by family 77 enzymes can, therefore, be written as:

$$G_m + G_n \rightarrow G_{n+x} + G_{m-x}$$

$$G_m \rightarrow \text{cyclic } G_x + G_{m-x}$$
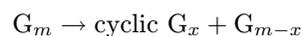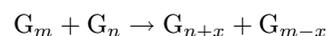
where G_m, G_n, ... etc. represent α-1,4-linked glucans containing m, n, ... etc. glucose residues. The minimum value for x in the second equation varies with the source of the enzyme, but is believed to be at least 17 (Sträter et al., 2002).

Family 70 consists of enzymes, the glucansucrases, that synthesize large polyglucans from sucrose and can form new α-1,2, α-1,3, α-1,4 or α-1,6 links between glucose residues. The reaction catalysed can be written as:

$$\text{sucrose} + G_m \rightarrow \text{fructose} + G_{m+1}$$
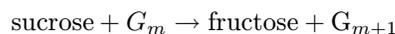
where now G_m and G_{m+1} represent glucans of m and $m + 1$ glucose residues, respectively.

The most complex of the clan GH-H families is family 13, containing hydrolases, transferases and isomerases with over twenty two known specificities. Examples of these, discussed at the Second Symposium on
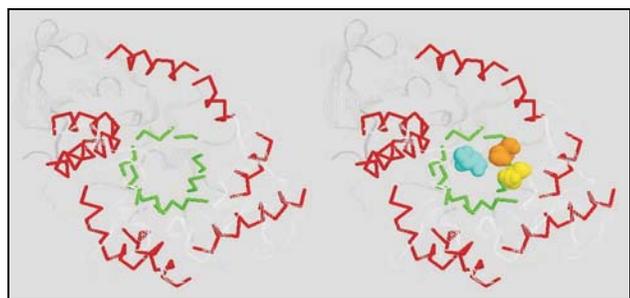
Fig. 1. The $(\alpha/\beta)_8$-barrel typical of clan GH-H enzymes. The β-strands are shown in green and the helices in red. The left-hand image shows the barrel only, while the right-hand image shows, in addition, the three acidic residues necessary for catalysis: nucleophile aspartate in yellow, acid/base glutamic acid in orange and the second aspartic acid in cyan. The structure is that of *A.oryzae* α-amylase (PDB code: 6taa).

the α-Amylase Family (http://imb.savba.sk/∼janecek/ Alamys/alamy_2/), are α-amylase, CGTase, amylosucrase, sucrose isomerase/ isomaltulose synthase, sucrose hydrolase, sucrose phosphorylase, cyclodextrinhydrolyzing enzyme, maltogenic α-amylase, maltogenic amylase, branching enzyme, maltohexaose-producing amylase and pullulanase.

All enzymes of clan GH-H are believed to possess a catalytic domain with the same basic three-dimensional fold, an $(\alpha/\beta)_8$-barrel. This $(\alpha/\beta)_8$-barrel or TIM-barrel – named for the first enzyme, Triose phosphate IsoMerase, in which the structure was found Banner et al., 1975) – consists of eight stretches of β-strand and eight α-helices that alternate along the polypeptide chain. The chain folds so that the β-strands lie approximately parallel to one another, as if on the surface of a cylinder, with the helices lying outside the β-strands (Fig. 1). The C-terminal ends of β-strands are joined, in the three-dimensional structure, to the N-terminal ends of following helices by irregular loops, and loops also link the C-terminal ends of helices to the N-terminal ends of adjacent β-strands. The first set of these loops, however, makes up the active site of the enzyme concerned. This structure has been demonstrated by X-ray crystallography in several enzymes from family 13, of which the first was the α-amylase from *Aspergillus oryzae* (Matsuura et al., 1984) and also in family 77 enzymes, where the first structure determined was that of the amylomaltase of *Thermus aquaticus* (Przylas et al., 2000). For glucansucrases of family 70, however, the presence of an $(\alpha/\beta)_8$-barrel has been predicted by computer-aided techniques (MacGregor et al., 1996), as difficulties in crystallizing the enzymes have precluded structure determination.

The enzymes of clan GH-H are thought to bring about catalysis by a common mechanism that involves an aspartate residue at the C-terminal end of β-strand 4 acting as the nucleophile and a glutamic acid residue at the end of β-strand 5 that plays the role of acid/base. These enzymes appear to require a third acid
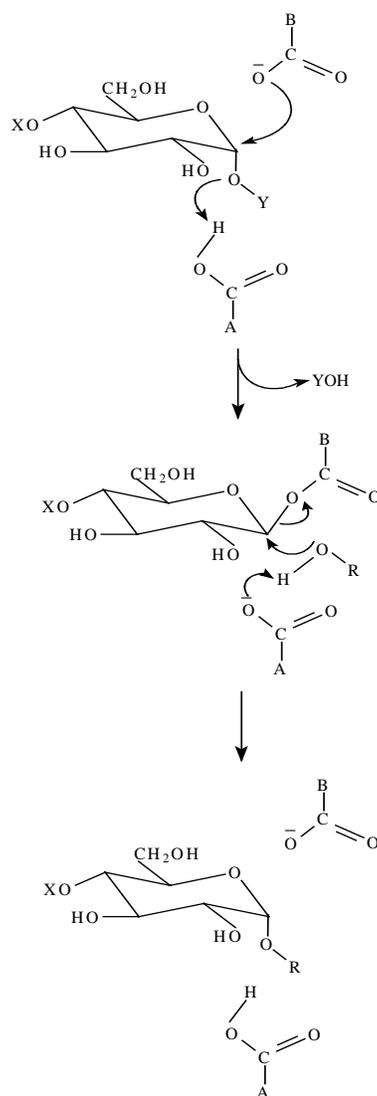

Fig. 2. The proposed mechanism of action of clan GH-H enzymes. X can be H or a chain of glycosidically-linked glucose residues; Y may be fructose, glucose or a chain of glucose residues; R may be H, phosphate, fructose or a chain of glucose residues. A and B represent the remainder of the structures of glutamic acid and aspartic acid, respectively.
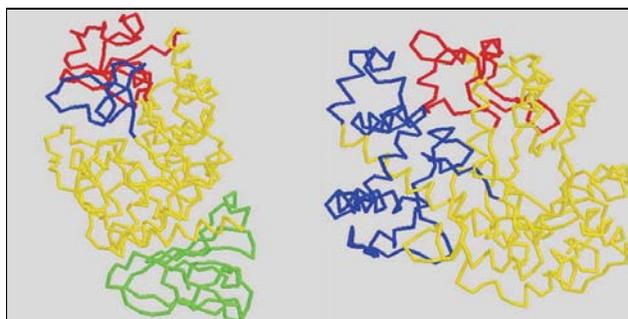

Fig. 3. Comparison of enzymes from families 13 (left-hand image) and 77 (right-hand image). The $(\alpha/\beta)_8$-barrel, domain A, is yellow; loop 3, "domain B", is red; loop 2 is blue and domain C is green. The enzymes are *A. oryzae* α-amylase (PDB code: 6taa) from family 13 and *Thermus aquaticus* amylomaltase (PDB code: 1cwy) from family 77.

residue for catalytic activity, an aspartic acid situated a few residues beyond the C-terminal end of β-strand 7 (Fig. 1). This residue may be important for maintaining the other catalytic acid residues in the correct ionisation state for activity, but is also vital for hydrogen-bonding interaction with the glucose residue of the substrate shown in Fig. 2 (UITDEHAAG et al., 1999; MATSUURA, 2002). Although several conserved sequence regions have been identified in α-amylase family enzymes (see, for example, JANECEK, 2002), it now seems likely that the three acid residues described above are the only ones conserved throughout all enzymes of clan GH-H (MACHOVIC & JANECEK, 2003).

The common mechanism of the clan GH-H enzymes is illustrated in Figure 2. It is believed that the glutamic acid residue at β-strand 5 donates a proton to the glycosidic oxygen, leading to bond cleavage. The aglycon segment of the substrate then leaves the enzyme active site. The glucose residue shown in Fig. 2 is stabilized by hydrogen bonding involving the aspartic acid, located after β-strand 7, and the hydroxyl group at C3 and possibly C2 of the glucose (UITDEHAAG et al., 1999; MATSUURA, 2002). Further stabilization occurs by formation of a covalent β-bond between C1 of the glucose and the nucleophilic aspartate at β-strand 4 (TAO et al., 1989; UITDEHAAG et al., 1999; BARENDS et al., 2004). The enzyme then facilitates approach of the second substrate – at present this is not a well-understood process – and the reaction takes place in reverse, i.e. the transient covalent bond between C1 of the glycon glucose and the nucleophile breaks, the glutamate at β-strand 5 is reprotonated by removal of a proton from the second substrate, and a new α-bond is formed between the glucose shown in Figure 2 and the remainder of the second substrate. Thus, by a two-step mechanism involving inversion of configuration at C1 in the short-lived intermediate, overall configuration is retained.

All enzymes of clan GH-H identified to date require a glucose residue in the substrate as shown in Fig. 2, and reaction takes place at C1 of this residue to give an α-linked product. Initially, this glucose may be part of sucrose, or at the non-reducing end or in the interior of a chain of glycosidically-linked glucose residues. If the reaction catalyzed by the enzyme is hydrolysis, then a new reducing group is formed at C1 of the glucose. Other enzymes may transfer phosphate, however, to give α-glucose-1-phosphate, or fructose to produce sucrose and/or its isomers. More frequently, enzymes of the clan may transfer the glucose of Figure 2 to another mono-, oligo- or polysaccharide of glucose to form a new glucose-glucose link that may have α-1,1, α-1,2, α-1,3, α-1,4 or α-1,6 configuration.

While the clan GH-H enzymes share a common fold for the catalytic domain as well as a common mechanism, and all require an α-linked glucose residue in the substrate, there are important differences between the three families 13, 70 and 77 of the clan. All the en-
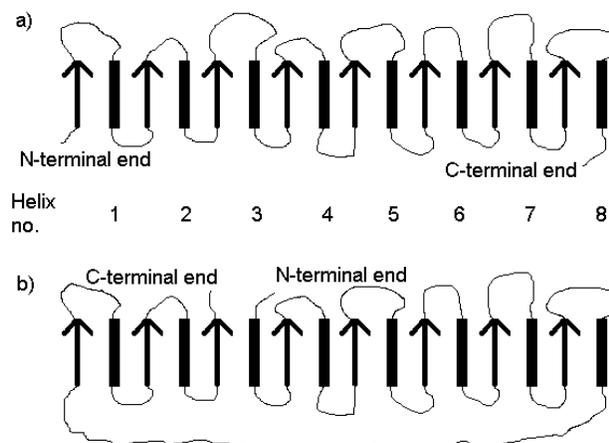

Fig. 4. Comparison of enzymes from families 13/77 and 70. Schematic illustration of (a) the "normal" barrel of families 13 and 77, where arrows represent β-strands and black rectangles represent helices, and (b) the permuted barrel of family 70. Helices are numbered from the N-terminal end of the protein for families 13 and 77. Using the same numbering system for family 70, the order of helices from the N-terminal end is 34567812.

zymes of family 13 have a catalytic module folded as the $(\alpha/\beta)_8$-barrel, followed by a domain consisting of β-strands and irregular loops. Often loop 3, connecting β-strand 3 to helix 3 of the barrel, is long and appears to fold as an independent unit. When this is the case loop 3 can be considered as a domain, and the enzyme is described as having domain A (the $(\alpha/\beta)_8$-barrel excluding loop 3), domain B (loop 3) and domain C (the β-strand domain that lies C-terminal to the barrel). In some enzymes, however, loop 3 is shorter and does not appear to have the characteristics of a domain, i.e. does not constitute an independent folding unit. In such cases, e.g. branching enzyme (ABAD et al., 2002), loop 3 should properly not be called a domain. Although the $(\alpha/\beta)_8$-barrel catalytic module and domain C are found in all family 13 enzymes, some of these enzymes are more complex and may have an additional domain N-terminal to the barrel (the N-domain) as in isoamylase (KATSUYA et al., 1998), or additional β-strand domains (domains D and E) following the C domain, as in CGTases (see, for example, HOFMANN et al., 1989).

The enzymes of family 77 differ from those of family 13 in having, in general, longer loops 2 and 7 in the barrel and a shorter loop 3 (Fig. 3). Domain C is missing from the family 77 enzymes, i.e. the $(\alpha/\beta)_8$-barrel is the C-terminal domain (see for example PRZYLAS et al., 2000).

Since no structure of a family 70 enzyme has been determined, we have to rely on computer-aided predictions (MACGREGOR et al., 1996) to understand the structural difference between families 13, 70 and 77. This is illustrated in Figure 4. If we consider that the acid residues involved in catalysis are situated at β-strands 4, 5 and 7 in family 13 and 77 enzymes and retain this numbering for family 70 glucansucrases, then

in family 13 and 77 proteins, the N-terminal end of the polypeptide chain is linked directly to the N-terminal end of β-strand 1, while the C-terminal end of the chain follows after the C-terminal end of helix 8 of the barrel (see Fig. 4a). In contrast, in family 70 enzymes, the N-terminus of the chain is linked to the N-terminal end of what would be helix 3 in members of families 13 and 77, the C-terminal end of helix 8 is joined by a long loop to the N-terminal end of β-strand 1 and the C-terminal end of β-strand 3 leads to the C-terminal end of the protein (Fig. 4b). This arrangement in family 70 enzymes is a circular permutation of the arrangement in family 13 and 77 enzymes.

Development of the concept of an α-amylase family prior to 1991

While some forty four sequences are listed for family 70 in the CAZy database and around one hundred and twenty sequences for family 77, family 13 is extremely unusual in that over two thousand primary sequences are given (COUTINHO & HENRISSAT, 1999; data obtained from the 7 February 2005 release of CAZy). Many family 13 enzymes are involved in glycogen and starch metabolism and the large number of known sequences may be a reflection of the importance of such metabolism in living organisms. In addition, the number of family 13 sequences, many of bacterial origin, may be connected to the diversity of metabolic pathways in bacteria, although strict symbiotic and parasitic bacteria may lack family 13 enzymes, since they appear to have no ability to metabolise starch/glycogen (HENRISSAT et al., 2002). The concept of a family of enzymes, closely related in structure to α-amylases, already existed in 1991 when Henrissat proposed the sequence-based classification (see below), however, and adoption of this group of enzymes, without subdivision in order to keep the number of families to a minimum, may also have led to the present size of family 13. Future division of the family into sub-families may be possible (OSLANCOVA & JANECEK, 2002) and desirable.

Even fifty years ago it was recognized that α-amylase is not a single enzyme, but rather a family of enzymes whose properties, such as variation of activity with pH, depend on the source of the amylase (see for example the review by GREENWOOD & MILNE, 1968a). These differences became more obvious when quantitative measurements of the amounts of maltodextrins produced at intermediate stages in the amylolysis of amylose showed that the details of action pattern of an α-amylase depend on enzyme source (ROBYT & FRENCH, 1963; GREENWOOD & MILNE, 1968b).

Over thirty years ago it was realized that an enzyme acting on a polysaccharide was likely to possess an active site capable of interacting with more than one monosaccharide residue of the substrate, and a "space" at the active site where one monosaccharide residue could bind became known as a subsite. Calculations based on subsite theory were then applied to α-amylases, to explain differences in action pattern in terms of differences at the enzyme active site (see for example TORGERSON et al., 1979; KONDO et al., 1980). Such calculations, using results of quantitative studies of products and/or kinetics of oligo- or polysaccharide hydrolysis, led to the conclusions that different α-amylases have active sites of different lengths, i.e. have different numbers of subsites. Although similar calculations continue to be carried out on newly-discovered enzymes, it should be noted that the calculations are only possible if it is assumed that all the subsites of an enzyme active site are independent of one another. This is unlikely to be true of any clan GH-H enzyme, and so the fundamental assumption enabling subsite calculations to be carried out is probably erroneous. Hence it is possible that two groups of scientists investigating the numbers of subsites at one enzyme active site may reach different, but legitimate, conclusions if different methodologies are used.

Further developments of the concept of an α-amylase family occurred after amino acid sequences became available for α-amylases and other enzymes that later proved to be related to them. The α-amylase sequences showed a surprising amount of diversity (FRIEDBERG, 1983; NAKAJIMA et al., 1986), but it was soon recognized that they contained four highly-conserved short segments (Fig. 5a) (NAKAJIMA et al., 1986). At first the significance of these segments was not understood, but it became clear from the first detailed three-dimensional structure of an α-amylase (from *A. oryzae*, MATSUURA et al., 1984) that the highly-conserved regions are closely associated with the active site. The structure showed, for the first time, the presence of an $(\alpha/\beta)_8$-barrel fold in an α-amylase, and the four conserved segments make up parts of β-strands 3, 4, 5 and 7 of the barrel. Thus segment 1 contains the C-terminal end of β-strand 3 and a histidine residue (Fig. 5a), common to α-amylases but not all clan GH-H enzymes, important for interaction with the critical glucose residue of the substrate shown in Figure 2. Segment 2 is associated with β-strand 4 and carries the aspartate residue that plays the role of nucleophile during catalysis (Fig. 2), whilst segment 3 makes up β-strand 5 and the glutamic acid residue present here acts as the acid/base. Lastly, segment 4 contains the end of β-strand 7, a histidine residue found in α-amylases, but not universal in clan GH-H enzymes, and the important aspartic acid that is the third acid residue necessary for catalytic activity in clan GH-H (Fig. 5a). Both the histidine residues and the latter aspartic acid are believed to form, during enzyme action, important hydrogen bonds with the glucose residue of the substrate shown in Figure 2.

Given the diversity of α-amylase amino acid sequences, the question arose, after the first structure was determined, as to whether all α-amylases would have the same fold. An important part of the answer

(a)

```
D V V A N H        1

G L R I D T V K H  2

Y C I G E V L D     3

F V E N H D         4
```

(b)

```
G L R I D A V K H  (i)

G W R L D V A N E  (ii)
```

Fig. 5. (a) The four highly-conserved sequence segments of clan GH-H enzymes. Sequences shown are for *A. oryzae* α-amylase (TODA et al., 1982). (b) Segment 2 sequences for (i) a typical α-1,4-linkage-specific enzyme, maltogenic α-amylase from *Bacillus stearothermophilus* (DIDERICHSEN & CHRISTENSEN, 1988) and (ii) a dual-specificity enzyme, active on α-1,4- and α-1,6-linkages, the maltogenic amylase of a *Thermus* strain (KIM et al., 1999). Letters in bold type and underlined show the acidic residues necessary for catalysis; other residues in bold type are conserved in most α-amylases, but not all clan GH-H enzymes, and are important for interaction with the substrate glucose residue shown in Figure 2; italic lettering shows residues important for specificity.

was supplied in 1987 when Richard Haser and his colleagues showed that pig pancreatic α-amylase does indeed have its catalytic module folded as an $(\alpha/\beta)_8$-barrel (BUISSON et al., 1987). Subsequent computer-aided predictions indicated that all α-amylases should possess this basic three-dimensional structure (MAC-GREGOR, 1988; RAIMBAUD et al., 1989), and later X-ray crystallographic studies have shown this to be the case (see for example PUJADAS & PALAU, 2002).

In 1988 Birte Svensson pointed out that sequence similarities between CGTase, a maltase and α-amylases are of such an extent that structural similarities amongst the enzymes should be expected (SVENSSON, 1988). Again, computer-aided studies led to the conclusion that CGTases and the maltase should have the $(\alpha/\beta)_8$-barrel catalytic module (MACGREGOR & SVENSSON, 1989), and an X-ray crystallographic investigation bore this out soon after for a CGTase (HOFMANN et al., 1989). Incidentally, continued work on CGTases has provided not only detailed insights into the complexities of action of CGTases but has also given information on the mechanism of catalysis of clan GH-H enzymes in general (UITDEHAAG et al., 1999).

Thus by 1989 the idea of a family of enzymes related in sequence and structure to α-amylases was firmly established. At the time it was thought that this relationship might extend only to enzymes active on α-1,4-linked glucans, although both Svensson and Kuriki had pointed out limited sequence homologies between branching enzyme, neopullulanase, isoamylase, pullulanase and α-amylases (SVENSSON, 1988; KURIKI & IMANAKA, 1989). Subsequent computer-aided predictions indicated that enzymes, active on α-1,6 bonds

only or on both α-1,4 and α-1,6 linkages between glucose residues, should also possess the $(\alpha/\beta)_8$-barrel fold and belong to the α-amylase family (JESPERSEN et al., 1991). It was at this stage that Henrissat proposed the sequence-based classificiation, and the then-recognized α-amylase family became family 13 (HENRISSAT, 1991).

Interestingly, Birte Svensson demonstrated that the sequence of an amylomaltase of *Streptococcus pneumoniae* contains a region homologous to segment 2 (Fig. 5a) of α-amylases (SVENSSON, 1988), but the full extent of the similarity between this enzyme and those of family 13 was not realized until the three-dimensional structure of another amylomaltase, from *Thermus aquaticus* became available (PRZYLAS et al., 2000) and such amylomaltases were classified in family 77 of clan GH-H.

Nomenclature problems

While Kuriki has emphasized the relationships in specificity between family 13 enzymes (KURIKI, 1992), i.e. within the family there are enzymes capable of making and/or breaking one or both of α-1,4 and α-1,6 glycosidic linkages, the subtle differences between many of the enzymes have often not been fully appreciated. This has resulted in confusing nomenclature within family 13. One such example involves the maltogenic amylases and the maltogenic α-amylases (MACGREGOR et al., 2001; JANECEK et al., 2004). Although in some cases the names have been used interchangeably, the two types of enzyme are distinct from each other and should probably have different Enzyme Commission numbers. While both produce mainly α-maltose by hydrolysis of glucans such as amylose, the maltogenic amylases are dual-specificity enzymes, acting on both α-1,4 and α-1,6 glycosidic linkages. They possess an N-domain preceding the $(\alpha/\beta)_8$-barrel in the three-dimensional structure and, following β-strand 4, usually have an asparagine-glutamic acid dipeptide in the amino acid sequence (Fig. 5b). Both the N-domain and this dipeptide are characterisitic of dual-specificity enzymes and are, in general, not found in α-amylases. In contrast, in the true α-amylases that act only on α-1,4 bonds, the first domain of the three-dimensional structure is domain A, the $(\alpha/\beta)_8$-barrel, while the typical "conserved" segment 2 of the amino acid sequence often has an important lysine-histidine dipeptide beginning three residues after the catalytic nucleophile (Fig. 5a and b). In fact a (lysine or arginine)-(histidine or glycine) dipeptide at this position in a sequence and lack of an N-domain are characteristic of family 13 enzymes able to act on α-1,4 glycosidic links only (MACGREGOR et al., 2001). The residues of this dipeptide are believed to interact, at the enzyme active site, with monosaccharide residues immediately on the aglycon side of the substrate glycosidic bond to be broken (see, for example, MATSUURA et al., 1984) and hence are important in determining specificity.

Table 1. Nucleophile and acid/base residue of related glycosidase families.

Family	Nucleophile	Acid/Base
13, 70, 77	Asp	Glu
14	Glu	Glu
15	Glu	Glu
27	Asp	Asp
31	Asp	Asp
57	Glu	Asp

Other enzymes of family 13 known to act on α-1,6 bonds as well as α-1,4 linkages and containing a dipeptide in segment 2 (Fig.5) other than lysine/arginine-histidine/glycine may well be called amylases, particularly if they catalyse efficient hydrolysis of starch, but should not be called α-amylases.

Distantly-related families

Other families of glycoside hydrolases contain enzymes catalysing hydrolysis of glycosidic linkages between α-linked glucose residues, and thus are related to clan GH-H in specificity. Examples of these are families 14, 15, 31, 57 and even 27 where the majority of enzymes are α-galactosidases but an isomalto-dextranase is known that hydrolyzes isomaltose from an α-1,6-linked dextran (COUTINHO & HENRISSAT, 1999).

The β-amylases and glucoamylases of families 14 and 15, respectively, differ from clan GH-H enzymes in that they bring about inversion of configuration to give β-maltose or β-glucose as the respective products. Family 27, 31 and 57 hydrolases and transferases give retention of α-configuration, like the enzymes of the α-amylase superfamily.

None of these families, however, appear to have the combination of nucleophilic aspartate, acid/base glutamic acid and required second aspartic acid residues found in clan GH-H enzymes as necessary for catalytic activity (Tab. 1). Rather, the families use different combinations of acidic residues for catalysis (COUTINHO & HENRISSAT, 1999).

For the enzymes of families 14, 15, 27, 31 and 57, there is variation in their degree of structural relatedness to family 13 proteins. Some measure of this relatedness can be obtained using MultiProt (SHATSKY et al., 2002) where three-dimensional structures are compared. For two or more proteins, the number of amino acid residues, whose locations are specified by the positions of α-carbon atoms and that can be aligned to a root mean square deviation (RMSD) of approximately two Ångstroms, is calculated. The three-dimensional fold of the α-amylase of *A. oryzae* (478 amino acid residues; PDB code 6taa, SWIFT et al., 1991) was used as "standard" and the structures of representative enzymes from several families compared to it. Results are shown in Table 2. Human myoglobin is included as a protein known to be unrelated to any family 13 α-amylase. Human myoglobin has an all-helical secondary structure, in sharp contrast to the $(\alpha/\beta)_8$-barrel A-domain and β-sheet C-domain of the *A. oryzae* α-amylase. Thus the result for myoglobin can be taken as representative of what may be obtained for proteins not structurally related to family 13 enzymes. As expected, a family 15 glucoamylase, that has an $(\alpha/\alpha)_6$-barrel structure for the catalytic domain, shows about the same degree of unrelatedness to the α-amylase as myoglobin. Family 57 enzymes, however, possess an $(\alpha/\beta)_7$-barrel, rather like an $(\alpha/\beta)_8$-barrel with one strand missing, and hence the structure is more open. Such a fold appears to be slightly more related to the family 13 barrel than the all-helical barrel of family 15, and so a very distant evolutionary relationship between families 13 and 57 cannot be ruled out. Indeed, a possible relationship has already been discussed (JANECEK, 1998).

On the other hand, any enzyme having an $(\alpha/\beta)_8$-barrel (TIM barrel) structure might be expected to be more closely related to the *A. oryzae* α-amylase than the enzymes discussed above, whether or not there is any relationship in specificity (see, for example, JANECEK, 1995; NAGANO et al., 2001, 2002). This is confirmed by the fact that more residues of chicken muscle TIM can be aligned to the α-amylase than those of family 15 or 57 enzymes (Tab. 2). It is obvious, how-

Table 2. MultiProt results for alignment of selected proteins with *A. oryzae* α-amylase.

Protein	GH family	PDB [a]	Alignment [b]	RMSD (Å)
Human myoglobin		2mm1	49	1.9
Aspergillus awamori glucoamylase	15	1glm	57	2.0
Thermococcus litoralis 4-α-glucanotransferase	57	1k1w	89	2.2
Chicken muscle TIM		1tim	107	2.1
Barley β-amylase	14	1b1y	152	2.0
Escherichia coli glycosidase	31	1xsj	203	2.0
Rice α-galactosidase	27	1uas	215	2.1
Thermus aquaticus amylomaltase	77	1cwy	223	1.7
Pseudomonas amyloderamosa isoamylase	13	1bf2	285	1.6
Bacillus circulans CGTase	13	8cgt	393	1.3

[a] Taken from the Protein Data Bank at URL: http://www.rcsb.org/pdb/.
[b] Proteins are arranged in order of increasing alignment length.

ever, that some enzymes possessing $(\alpha/\beta)_8$-barrels are more similar in structure to the *A. oryzae* amylase than others (see the comparison of families 14, 27, 31, 77 and 13 in Table 2). It can be seen that the two family 13 enzymes, as might be expected, are most closely related, with the β-amylase of family 14 being the least related of the β-barrel-containing glycosidases shown here. Surprisingly, almost as many residues of the family 27 α-galactosidase align with the *A. oryzae* amylase as those of the family 77 amylomaltase, a member of clan GH-H like the α-amylase itself. The root mean square deviation is smaller, however, for the amylomaltase than the galactosidase.

At the time of the 2nd Symposium on the Alpha-Amylase Family it was not possible to carry out a direct structural comparison of a family 31 enzyme with the *A. oryzae* α-amylase, for no structure had been published for family 31. Predictions from amino acid sequence indicated, however, that family 31 enzymes should resemble proteins of families 13 and 27 in structure, i.e. possess an $(\alpha/\beta)_8$-barrel domain (RIGDEN, 2002; MACGREGOR, 2004, unpublished results). Since family 27 appears to have a relatively close structural relationship to family 13, we expected a similar situation for family 31.

Further, the enzymes of both families 13 and 27 have an aspartate residue as nucleophile, situated at the C-terminal end of β-strand 4 of the $(\alpha/\beta)_8$-barrel. This is likely to be true also for family 31 enzymes, since aspartic acid residues constitute both the catalytic nucleophile and acid/base of this family (LEE et al., 2001; OKUYAMA et al., 2001). Family 13 and 27 enzymes differ, however, in the nature and position of the acid/base residue. For family 13 it is the glutamic acid residue at the C-terminal end of β-strand 5, while for family 27 it is an aspartic acid residue at the end of β-strand 6.

Recently, the co-ordinates of a family 31 glycosidase, probably a xylosidase, have been deposited in the Protein Data Bank (PDB code 1xsj) and the structure has been described (LOVERING et al., 2005). These show that the enzyme does indeed contain an $(\alpha/\beta)_8$-barrel domain and that the likely catalytic residues are situated at the C-terminal ends of β-strands 4 and 6. Thus the catalytic "machinery" of family 31 enzymes appears to be arranged as in family 27 glycosidases. The MultiProt score for an alignment of the putative enzyme structure with the *A. oryzae* α-amylase shows that the barrels of the two enzymes are almost as closely related as the barrels of family 77 amylomaltase and the α-amylase (Tab. 2).

Enzymes of families 36 and 66, believed to be distantly related to family 13 proteins (RIGDEN, 2002), have not been examined here.

In summary, for the so-called distantly-related families, there is a variation in the degree of their relatedness to clan GH-H enzymes. At one extreme, family 15 enzymes are related in specificity only, by their ability to hydrolyse glucose from non-reducing ends of α-1,4- or α-1,6-linked glucans. α-Glucosidases of family 31, on the other hand, are related not only in specificity, but have an $(\alpha/\beta)_8$-barrel catalytic domain with fairly close structural similarity to that of clan GH-H enzymes.

References

ABAD, M.C., BINDERUP, K., RIOS-STEINER, J., ARNI, R.K., PREISS, J. & GEIGER, J. 2002. J. Biol. Chem. **277:** 42164–42170.

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. & WALEY, S.G. 1975. Nature **255:** 609–614.

BARENDS, T.R.M., KAPER, T., DIJKHUIZEN, L. & DIJKSTRA, B.W. 2004. p. 57. In: JANECEK, S. (ed.) 2nd Symposium on the Alpha-Amylase Family – Program and Abstracts (Smolenice Castle, October 3–7, 2004), Poster P4, ASCO Art & Science, Bratislava.

BUISSON, G., DUEE, E., HASER, R. & PAYAN, F. 1987. EMBO J. **6:** 3909–3916.

COUTINHO, P. & HENRISSAT, B. 1999. Carbohydrate-Active Enzymes server at URL: http://afmb.cnrs-mrs.fr/CAZY/.

DIDERICHSEN, B. & CHRISTENSEN, L. 1988. FEMS Microbiol. Lett. **56:** 53–60.

FRIEDBERG, F. 1983. FEBS Lett. **152:** 139–140.

GREENWOOD, C. T. & MILNE, E.A. 1968a. Adv. Carbohydr. Chem. **23:** 281–366.

GREENWOOD, C. T. & MILNE, E.A. 1968b. Starch/Staerke **20:** 139–150.

HENRISSAT, B. 1991. Biochem. J. **280:** 309–316.

HENRISSAT, B., DELEURY, E. & COUTINHO, P.M. 2002. Trends Genet. **18:** 437–440.

HOFMANN, B.E., BENDER, H. & SCHULZ, G.E. 1989. J. Mol. Biol. **209:** 793–800.

JANECEK, S. 1995. FEBS Lett. **377:** 6–8.

JANECEK, S. 1998. Folia Microbiol. **43:** 123–128.

JANECEK, S. 2002. Biologia, Bratislava **57 (Suppl. 11):** 29–41.

JANECEK, S., SVENSSON, B. & MACGREGOR, E.A. 2004. p. 67. In: JANECEK, S. (ed.) 2nd Symposium on the Alpha-Amylase Family – Program and Abstracts (Smolenice Castle, October 3–7, 2004), Poster P14, ASCO Art & Science, Bratislava.

JESPERSEN. H.M., MACGREGOR, E.A., SIERKS, M.R. & SVENSSON, B. 1991. Biochem J. **280:** 51–55.

KATSUYA, Y., MEZAKI, Y., KUBOTA, M. & MATSUURA, Y. 1998. J. Mol. Biol. **281:** 885–897.

KIM, J.S., CHA, S.S., KIM, H.J., KIM, T.J., HA, N.C., OH, S.T., CHO, H.S., CHO, M.J., KIM, M.J., LEE, H.S., KIM, J.W., CHOI, K.Y., PARK, K.H. & OH, B.H. 1999. J. Biol. Chem. **274:** 26279–26286.

KONDO, H., NAKATANI, H., MATSUNO, R. & HIROMI, K. 1980. J. Biochem. **87:** 1053–1070.

KURIKI, T. 1992. Trends Glycosci. Glycotechnol. **4:** 567–572.

KURIKI, T. & IMANAKA, T. 1989. J. Gen. Microbiol. **135:** 1521–1528.

LEE, S.S., HE, S. & WITHERS, S.G. 2001. Biochem. J. **359:** 381–386.

LOVERING, A.L., LEE, S.S., KIM, Y.W., WITHERS, S.G. & STRYNADKA, N.C. 2005. J. Biol. Chem. **280:** 2105–2115.

MACGREGOR, E.A. 1988. J. Protein Chem. **7:** 399–415.

MACGREGOR, E.A., JANECEK, S. & SVENSSON, B. 2001. Biochim. Biophys. Acta **1546:** 1–20.

MACGREGOR, E.A., JESPERSEN, H. M. & SVENSSON, B. 1996. FEBS Lett. **378:** 263–266.

MACGREGOR, E.A. & SVENSSON, B. 1989. Biochem. J. **259:** 145–152.

MACHOVIC, M. & JANECEK, S. 2003. Biologia, Bratislava **58:** 1127–1132.

Matsuura, Y. 2002. Biologia, Bratislava **57 (Suppl. 11):** 21–27.

Matsuura, Y., Kusunoki, M., Harada, W. & Kakudo, M. 1984. J. Biochem. **95:** 697–702.

Nagano, N., Orengo, C.A. & Thornton, J.M. 2002. J. Mol. Biol. **321:** 741–765.

Nagano, N., Porter, C.T. & Thornton, J.M. 2001. Protein Eng. **14:** 845–855.

Nakajima, R., Imanaka, T. & Aiba, S. 1986. Appl. Microbiol. Biotechnol. **23:** 355–360.

Okuyama, M., Okuno, A., Shimizu, N., Mori, H., Kimura, A. & Chiba, S. 2001. Eur. J. Biochem. **268:** 2270–2280.

Oslancova, A. & Janecek, S. 2002. Cell. Mol. Life Sci. **59:** 1945–1959.

Przylas, I., Tomoo, K., Terada, Y., Takaha, T., Fujii, K., Saenger, W. & Sträter, N. 2000. J. Mol. Biol. **296:** 873–886.

Pujadas, G. & Palau, J. 2002. Biologia, Bratislava **57 (Suppl. 11):** 43–57.

Raimbaud, E., Buleon, A., Perez, S. & Henrissat, B. 1989. Int. J. Biol. Macromol. **11:** 217–225.

Rigden, D.J. 2002. FEBS Lett. **523:** 17–22.

Robyt, J.F. & French, D. 1963. Arch. Biochem. Biophys. **100:** 451–467.

Shatsky, M., Nussinov, R. & Wolfson, H.J. 2002. Lecture Notes Comput. Sci. **2452:** 235–250.

Sträter, N., Przylas, I., Saenger, W., Terada, Y., Fujii, K. & Takaha, T. 2002. Biologia, Bratislava **57 (Suppl. 11):** 93–99.

Svensson, B. 1988. FEBS Lett. **230:** 72–76.

Swift, H.J., Brady, L., Derewenda, Z.S., Dodson, E.J., Dodson, G.G., Turkenburg, J.P. & Wilkinson, A.J. 1991. Acta Crystallogr. **B47:** 535–544.

Tao, B.Y., Reilly, P.J. & Robyt, J.F. 1989. Biochim. Biophys. Acta **995:** 214–220.

Toda, H., Kondo, K. & Narita, K. 1982. Proc. Japan Acad. **58B:** 208–212.

Torgerson, E.M., Brewer, L.C. & Thoma, J.A. 1979. Arch. Biochem. Biophys. **196:** 13–22.

Uitdehaag, J.C.M., Mosi, R., Kalk, K.H., Van Der Veen, B.A., Dijkhuizen, L., Withers, S.G. & Dijkstra, B.W. 1999. Nature Struct. Biol. **6:** 432–436.