How many conserved sequence regions are there in the α-amylase family?

Štefan Janeček*

Institute of Molecular Biology, Slovak Academy of Sciences, Dubravská cesta 21, SK-84251 Bratislava, Slovakia; tel.: ++ 421 2 5930 7420, fax: ++ 421 2 5930 7416, e-mail: stefan.janecek@savba.sk

Janeček, Š., How many conserved sequence regions are there in the α-amylase family? Biologia, Bratislava, 57/Suppl. 11: 29—41, 2002; ISSN 0006-3088.

The α-amylase family, i.e. the clan GH-H of glycoside hydrolases, is the largest family of glycoside hydrolases, transferases and isomerases comprising nearly 30 different enzyme specificities. One of the most interesting features of this family is that its members contain a few very well-conserved sequence regions despite the overall low sequence similarity. It seems that only 4 amino acid residues may be totally invariantly conserved throughout the family (Arg204 plus the three catalytic residues: Asp206, Glu230 and Asp297; Taka-amylase A numbering). Four conserved sequence regions covering the strands β3, β4, β5, and β7 of the catalytic (β/α)8-barrel domain were identified and used for defining the α-amylase family. The present review is therefore focused on the three additional conserved sequence regions proposed after the basic characteristics of the family have been established. Two of these three regions cover roughly the strands β2 and β8 of the catalytic (β/α)8-barrel and one is located near the C-terminus of domain B (in the β3→α3 connection of the catalytic (β/α)8-barrel). While the four original conserved sequence regions contain the catalytic and substrate-binding residues of the individual members of the family, the three conserved sequence regions later identified are shown to contain amino acid residues connected to a given enzyme specificity. The problems that may arise with correctly identifying the β5-strand catalytic glutamate, positioned in the conserved sequence region III, are discussed and a way that may lead to a correct solution is indicated. In conclusion, it is proposed that the α-amylase enzyme family should be characterised by as many conserved sequence regions as possible.

Key words: alpha-amylase family, conserved sequence regions, sequence similarities, specificity features, evolutionary relationships.

Abbreviations: CGTase, cyclodextrin glucanotransferase.

Introduction

The α-amylase enzyme family was originally recognised as a group of starch hydrolases and related enzymes that exhibit clear sequence similarities and a predicted common supersecondary fold, a parallel (β/α)8-barrel (MacGregor, 1988; Svensson, 1988; MacGregor & Svensson, 1989; Jespersen et al., 1991; 1993). From the beginning it covered several different enzyme speci-
ficities, including cyclodextrin glucanotransferase (CGTase), isoamylase, branching enzyme, neopullulanase and oligo-1,6-glucosidase in addition to α-amylase (Binder et al., 1986; Amemura et al., 1988; Romeo et al., 1988; Kuriki & Imanaka, 1989; Watanabe et al., 1990). The α-amylase family has been known as the family 13 of glycoside hydrolases from 1991 when the sequence-based classification of all glycoside hydrolases, transferases, and isomerases was established (HenriSSat, 1991).

The concept of this group of enzymes as the α-amylase family was proposed a year later (Takata et al., 1992). According to that definition the members of the α-amylase family are enzymes that satisfy the four requirements: (i) they act on α-glucosidic linkages; (ii) they hydrolyse or form by transglycosylation α-glucosidic linkages; (iii) their amino-acid sequence contains four conserved regions; and (iv) they contain Asp, Glu and Asp residues corresponding to the Asp296, Glu230 and Asp297 of Taka-amylase A.

The current situation in the α-amylase family is as follows: (i) the family contains almost 30 different enzyme specificities covering hydrolases, transferases, and isomerases (Svensson et al., 2002); (ii) the family constitutes a clan GH-I of glycoside hydrolases comprising the families 13, 70, and 77 (Coutinho & HenriSSat, 1999); (iii) the enzymes from the family can operate on α-1,1-, α-1,2-, α-1,3- and α-1,5-glucosidic linkages in addition to the α-1,4- and α-1,6-bonds originally considered (MacGregor et al., 2001); (iv) there may be at most only 4 invariant amino acid residues, i.e. 3 catalytic ones plus the Arg in position i−2 with respect to the catalytic nucleophile, the aspartate in the β-strand (Janecek, 2000a); (v) the relationship of the main α-amylase family (clan GH-I) to the “extremophilic” α-amylase family, family 57 of glycoside hydrolases, has still been neither confirmed nor disproved (Janecek, 1998; Isamura et al., 2001a); and (vi) from the points of view and evolution, some mammalian transport proteins and antigens are perhaps related to the enzymatic members of the family (Janecek, 2000b).

Many specialised reports are available focusing on protein engineering of the individual enzymes and certain groups from the α-amylase family, as well as on the possibilities of tailoring their properties to meet specific needs (e.g., Kuriki et al., 1996; Matsui & Svensson, 1997; Binderup & Preiss, 1998; Ibuca et al., 1998; Wind et al., 1998; Saab-Rincon et al., 1999; Beier et al., 2000; Yamamoto et al., 2000; Gottschalk et al., 2001; Mori et al., 2001; Kim et al., 2001; Leemhuis et al., 2002). The findings have been thoroughly discussed in various reviews (e.g., MacGregor, 1993; Svensson, 1994; Janecek, 1997a; Nielsen & Borchert, 2000; Park et al., 2000; van der Veen et al., 2000; Uitdehaag et al., 2002; van der Maarel et al., 2002). In general, however, the residues playing the most important roles are located in the short sequence stretches conserved throughout the family and known as the conserved sequence regions. This review therefore aims to present and summarise the short sequence regions that are best conserved either throughout the α-amylase family or in several closely-related enzyme groups from the family, and to show the relationships between these conserved sequence regions and the individual enzyme specificities.

**From family to clan and from clan to a ... superfamiliy ?**

When the number of sequences recognised as belonging to the α-amylase family was low, i.e. several tens at the beginning of the 1990s (MacGregor & Svensson, 1989; HenriSSat, 1991; Takata et al., 1992), it was clear that, despite the existence of more enzyme specificities, all the sequences constitute a classical enzyme family with members showing divergent evolution (JesperSEN et al., 1993; Janecek, 1994a). As one of the main criteria for deciding whether or not a new sequence belongs to the α-amylase family, the presence of the four conserved sequence regions (e.g., for Taka-amylase A – I: 117DVYANH, II: 202GLRIDAIVKIH, III: 230FVLID, IV: 292FVENHHD) was used as a good criterion (Nakajima et al., 1986).

This basic trend has, in fact, been maintained up to the present day (MacGregor et al., 2001) although the number of sequences has reached almost one thousand (Coutinho & HenriSSat, 1999). According to the HenriSSat classification, however, the original α-amylase family, family 13 of glycoside hydrolases (HenriSSat, 1991), has been changed to a clan, glycoside hydrolases clan GH-II, covering the families 13, 70, and 77 (Bourne & HenriSSat, 2001). While the original family 13 and the family 77 seem to be very closely related to each other (PrzyLas et al., 2000), i.e. they roughly fulfil the attributes of a single family of similar reaction mechanism, fold, and catalytic machinery seen as identical catalytic residues on identical elements of secondary structure, family 70 con-
tains enzymes with a version of the \((\beta/\alpha)_8\)-barrel structure that is a circularly permuted form of the classical \(\alpha\)-amylase-type \((\beta/\alpha)_8\)-barrel (MAC-GREGOR et al., 1996). Although an alternative prediction without circular permutation was presented for the family 70 glucosyltransferases (DEVULAPALLE et al., 1997), the predictions taking into account the possibility of circular permutations seem more reasonable and have been supported by mutagenesis (MACGREGOR et al., 1996; MONCHOIS et al., 1999; 2000; TSAI et al., 2000).

The existence of a single \(\alpha\)-amylase family, the family 13 of glycoside hydrolases, has appeared more “complicated” since 1988 and 1993 when two “unusual” amino acid sequences of \(\alpha\)-amylases were reported: the AmyA from Dictyoglomus thermophilum (FUKUSUMI et al., 1988) and one from Pyrococcus furiosus (LADERMAN et al., 1993). These and also some other mostly extremophilic amylolytic enzymes (DONG et al., 1997b; JEON et al., 1997; ERRA-PUJADA et al., 1999) were found to lack the characteristic conserved sequence regions; this resulted in a new family, family 57 of glycoside hydrolases being established (HENRISSAT & BAIROCH, 1996). Later, however, a distant evolutionary relation between the major family 13 and family 57 was indicated by using HCA (hydrophobic cluster analysis) on the amino-acid sequence of archaeal Methanococcus jannaschii \(\alpha\)-amylase (JANECEK, 1998). At present family 57 contains members mostly from extremophiles (COUTINHO & HENRISSAT, 1999) having \(\alpha\)-amylase, 4-\(\alpha\)-glucanotransferase, amylolipullulanase and even \(\alpha\)-galactosidase (VAN LIESHOUT et al., 2001) specificities in addition to numerous hypothetical proteins coming from various complete-genome sequencing projects. Recent determination of the three-dimensional structure of the 4-\(\alpha\)-glucanotransferase from Thermococcus litoralis, with site-directed mutated glutamate residues, indicated (IMAMURA et al., 2001a,b) that the HCA-based predictions (JANECEK, 1998) may not match the catalytic amino acids and thus may invalidate the postulated close relationship of glycoside hydrolase family 57 to clan GH-H. Despite this recent finding, the common retaining mechanism and a common \((\beta/\alpha)_8\)-barrel fold used by both the clan GH-H and the family 57 preserve the possibility of classifying them into an even larger group than the clan, e.g. a suprafamily representing relatedness at a higher hierarchal level than a clan.

Conserved sequence regions

Study of the \(\alpha\)-amylase family is practically impossible without taking into account the question of the conserved sequence regions. For a considerable time, the sequence similarity has been known to be extremely low (about 10%) even for the \(\alpha\)-amylases alone (i.e. for EC 3.2.1.1) as described from different microorganisms, plants, and animals (NAKAJIMA et al., 1986). Later when the family grew, i.e. when many sequences from various sources and with different enzyme specificities became available, the number of identical residues among the \(\alpha\)-amylase family enzymes had decreased to 8-10 amino acids in 1994 (JANECEK, 1994a; SVENSSON, 1994). At present there might be at most only 4 residues that are totally invariantly conserved throughout the \(\alpha\)-amylase family, i.e. the 3 catalytic residues corresponding to Asp206, Ghu230 and Asp297 in Taka-amylase A plus the 4 arginine equivalent to the Arg204 in Taka-amylase A. The 3 \(\beta\) aspartate (Asp117 in Taka-amylase A) is also extremely well conserved (Fig. 1), however, it has been found substituted at least one time (S. JANECEK, unpublished results): by glutamate in trehalose synthase from Thermus aquaticus (TSUSAKI et al., 1997). Despite this fact, amino-acid sequences of enzymes from the \(\alpha\)-amylase family contain several short conserved stretches, the so-called conserved sequence regions (Fig. 1). Of the four above-mentioned invariantly-conserved residues Asp206 (the catalytic nucleophile) and Ghu230 (the proton donor) together with the second conserved aspartate, Asp 297, and two histidine residues, His122 and His296, were postulated as being crucial for Taka-amylase A (MATSUURA et al., 1984). They form the basis of the four well-known conserved sequence regions in the \(\alpha\)-amylase family enzymes (the regions I, II, III and IV in Fig. 1) and are positioned near the C-termini of strands \(\beta_3\), \(\beta_4\), \(\beta_5\), and \(\beta_7\) of the catalytic \((\beta/\alpha)_8\)-barrel domain (NAKAJIMA et al., 1986; MACGREGOR et al., 2001). It is clear, however, that the two histidines are not conserved throughout the family (Fig. 1).

Although NAKAJIMA et al. (1986) established the four conserved sequence regions of \(\alpha\)-amylases, the authors paid less attention to the remaining part of their alignment of 11 different \(\alpha\)-amylase sequences. This fact was mentioned already by MACGREGOR (1988), who indicated that there could be additional sequence similarities in \(\alpha\)-amylases that were found later in other enzymes belonging to the \(\alpha\)-amylase family by SVENSSON (1988), MACGREGOR & SVENSSON (1989) and
Fig. 1. Conserved sequence regions in the α-amylase family. Sources of the enzymes: α-amylase (Aspergillus oryzae; Toda et al., 1982; Matsuura et al., 1984), cyclodextrin glucanotransferase (Bacillus circulans strain 8; Nitschke et al., 1990; Klein & Schulz, 1991), oligo-1,6-glucosidase (Bacillus cereus; Watanabe et al., 1996; Kizaki et al., 1993), maltotetrahydrolase (Pseudomonas stutzeri; Fujita et al., 1989; Morishita et al., 1997), isoamylase (Pseudomonas amyloferans; Amemura et al., 1988; Katsuya et al., 1998), neopolulolase (Thermoactinomyces vulgaris 95-47; Tonomura et al., 1993; Kamitori et al., 1999), maltogenic amylase (Thermus sp. IM6501; Kim et al., 1999b; Kim et al., 1999a), maltogenic α-amylase (Bacillus steatorrhophilus; Didrichsen & Christiansen, 1988; Dauter et al., 1999), amyloglucosidase (Thermus aquaticus; Otani et al., 1999; Paez et al., 2000), maltoligosyltrehalose hydrolase (Sulfobacillus solfataricus; Kobayashi et al., 1996; Feese et al., 2000), amylolysocarose (Netespera polysaccharae; Potecki De Mentalk et al., 1999; Skov et al., 2001), maltooligosaccharase (Thermotoga maritima; Meissner & Liebl, 1998; Roujinhkova et al., 2001). Colour code: catalytic aspartates and glutamate – blue; functional histidines – yellow; invariant arginine – green; Ca²⁺-binding aspartate – turquoise; conserved residues – grey; non-conserved residues – pink. The first two columns denote the year of three-dimensional structure determination and the EC number of a given member, respectively. Only the members whose structures have already been determined are shown.
Fig. 2. Identifying the \( \beta \)-5-strand catalytic glutamate (Region III) where there might be an ambiguity. (a) The amino-acid sequence of the periplasmic \(-\)amylase from \textit{Xanthomonas campestris} (ABE et al., 1996) with highlighted conserved sequence regions in yellow except for the region III containing the catalytic Glu. The segment 225 \text{EFRQ} (highlighted in pink) was proposed (ABE et al., 1996) as the conserved sequence region III with the 5-strand catalytic glutamate. The start of the mature enzyme is indicated by an arrow. (b) Based on the alignment with closely related sequences and, more importantly, taking into account also the characteristic sequences in the other conserved regions, especially in this particular case in the 5th conserved sequence region (165 \text{MPDLN}, Region V), the catalytic glutamate is proposed rather to be Glu242 (238 \text{YLVGEVSA}, Region III). The glutamates corresponding to Glu225 of \textit{X. campestris} \(-\)amylase in the three other \(-\)amylases are also highlighted in pink. Note that there is no glutamate in the \textit{Bacillus cereus} oligo-1,6-glucosidase corresponding to Glu225 from \textit{X. campestris} enzyme. Sources of the enzymes: BceOGL: \textit{Bacillus cereus} oligo-1,6-glucosidase (WATANABE et al., 1990); XcaAMY: \textit{Xanthomonas campestris} \(-\)amylase (ABE et al., 1996); BmeAMY: \textit{Bacillus megaterium} \(-\)amylase (METZ et al., 1988); DthAMY: \textit{Dictyoglomus thermophilum} AmyC \(-\)amylase (HORTNOUCHI et al., 1988); TmaAMY: \textit{Thermotoga maritima} \(-\)amylase (LIEBL et al., 1997).
is limited experimental support (Fig. 2b). Abe et al. (1999) solved the problem by site-directed mutagenesis on isoamylase, which identified the correct glutamic acid. A similar strategy was applied by, e.g., Devulapalle et al. (1997) to glucosyltransferase, Sarcabala et al. (2000) to amylosucrase and Nakayama et al. (2001) to glycogen debranching enzyme.

**Importance of the additional conserved sequence regions**

The fifth conserved sequence region — in domain B

The fifth conserved sequence region (173 LPDLD in Taka-amylase A; Fig. 1) was proposed first for α-amylases (Janecek, 1992) and later was identified also in the other members of the α-amylase family (Janecek, 1995a). It is positioned near the C-terminus of domain B around the aspartate involved in binding the calcium ion, Asp175 in Taka-amylase A (e.g. Matsura et al., 1984; Qian et al., 1999; Kadziola et al., 1994; Machius et al., 1995). In most cases there are 28-30 residues between this Ca²⁺-binding aspartate and the 4-strand catalytic aspartate, i.e. Asp206 in Taka-amylase A (Janecek, 1992, 1995a).

Although this region seems to be not fully conserved or not easily identifiable in some family members (Janecek, 1997a), in several cases this stretch is very characteristic for a special enzyme specificity (Fig. 3). The most striking example is the oligo-1,6-glucosidase with the template sequence QPDLN (x = P, A, I, W) is found also in α-glucosidase, dextran-glucosidase, trehalose-6-phosphate hydrolase, amyllosucrase, sucrose phosphorylase, trehalose synthase (e.g., Janecek, 1997a, 2000a). There is a characteristic change from QPDLN (the oligo-1,6-glucosidase group) to MPKLN (the neopullulanase group) via the MPDLN belonging to a group of odd members of the α-amylase family (Janecek, 1997). For illustration also the corresponding segment is shown of the human protein responsible for the transport of dibasic and neutral amino acids across cell membranes (amino-acid transporter; Bertrand et al., 1993) containing the oligo-1,6-glucosidase-type of domain B (Janecek et al., 1997). Colour code: Ca²⁺-binding aspartate — blue; lysine playing the role of a calcium — red; oligo-1,6-glucosidase-like glutamine — yellow; neopullulanase-like methionine — green.

### Template:

**Taka-amylase A**

**Oligo-1,6-glucosidase group:**

- Oligo-1,6-glucosidase (3.2.1.10)
- α-Glucosidase (3.2.1.20)
- Dextran glucosidase (3.2.1.70)
- Trehalose-6-P hydrolase (3.2.1.93)
- Amylosucrase (2.4.1.4)
- Sucrose phosphorylase (2.4.1.7)
- Isomaltulose synthase (5.4.99.11)
- Trehalose synthase (5.4.99.16)

**Neopullulanase group:**

- Cyclomaltodextrinase (3.2.1.54)
- Maltoogenic amylase (3.2.1.133)
- Neopullulanase (3.2.1.135)

**Intermediary α-amylases:**

- Bacillus megaterium
- Thermotoga maritima
- Xanthomonas campestris
- Dictyoglomus thermophilum AmyC
- Dictyoglomus thermophilum AmyB

**Non-catalytic relative:**

- Amino acid transporter

**Fig. 3.** The fifth conserved sequence region of the α-amylase family enzymes. The figure focuses on the so-called oligo-1,6-glucosidase and neopullulanase groups of the family. Sources of the enzymes: Taka-amylase A (Toda et al., 1982); oligo-1,6-glucosidase (Watanabe et al., 1990); α-glucosidase (Janda et al., 2000); dextran glucosidase (Russell & Ferretti, 1990); trehalase-6-phosphate hydrolase (Rimmele & Boos, 1994); amyllosucrase (Potocki De Montalk et al., 1999); sucrose phosphorylase (Ferretti et al., 1988); sucrose isomerase (Zhang et al., 2002); trehalose synthase (Tsusuki et al., 1996); cyclodextrinase (Podkorytov & Zeikus, 1992); maltoogenic amylase (Kintz et al., 1999); neopullulanase (Tonozuka et al., 1993); Bacillus megaterium (Metz et al., 1988); Thermotoga maritima (Liebl et al., 1997); Xanthomonas campestris (Abe et al., 1996); Dictyoglomus thermophilum AmyC (Horinouchi et al., 1988); Dictyoglomus thermophilum AmyB (Horinouchi et al., 1988). For illustration the corresponding segment is shown of the human protein responsible for the transport of dibasic and neutral amino acids across cell membranes (amino-acid transporter; Bertrand et al., 1993) containing the oligo-1,6-glucosidase-type of domain B (Janecek et al., 1997). Colour code: Ca²⁺-binding aspartate — blue; lysine playing the role of a calcium — red; oligo-1,6-glucosidase-like glutamine — yellow; neopullulanase-like methionine — green.
The members of the family with characteristic sequences in this region MPDLN (or MPKIN and MPKL; Figs 1.3), such as cyclomaltodextrinases, maltogenic amylases and neopullulanases, have a shorter domain B than the majority and presumably do not bind a calcium ion due to the Asp→Lys substitution (JANECEK et al., 1997). Enzymes with these specificities could also form a group of closely-related members of the α-amylase family (MATZKE et al., 2000; PARK et al., 2000; MACGREGOR et al., 2001; OH et al., 2001), the neopullulanase group (OSLANCOVA & JANECEK, 2002) (Fig. 3). Indeed, it has recently been shown that the longer side-chain of lysine, corresponding to the intermediary sequence MPDLN, or MPKIN and MPKLN, respectively (Fig. 3), should be of great interest for the amylases, cyclomaltodextrinases, and neopullulanase group (OBLANCOVÁ & JANECEK, 2002) seem to contain more α-amylases-like sequences in this region MPKLN (or MPKIN and MPDLN) (Fig. 5). In order to distinguish CGTases from α-amylases the length of this region has to be taken into account, together with special sequence features (Fig. 4). The CGTases usually have eight residues between the Gly and Pro, while the α-amylases have seven (JANECEK et al., 1995). This is, however, not always the case, and several exo-α-amylases, such as maltotriohydrolases and maltotetraohydrolases, also have eight residues between Gly and Pro (KOBAYASHI et al., 1994; FUJITA et al., 1989), while the two known archaeal CGTases appear to have seven (YAMAMOTO et al., 2000; RASHID et al., 2002).

In a wider sense this conserved sequence region may be of help in distinguishing between the true α-amylases, true CGTases, and oligosaccharide-producing amylases (JANECEK, 1996b), such as maltotriohydrolase (Fig. 4). This is possible due to the presence of a glutamine residue preceding the invariant proline at the C-terminus of this region, a feature highly characteristic of CGTases (JANECEK & TOTH, 1994; JANECEK et al., 1995). The α-amylases have a gap in the position corresponding to that glutamine (Gln78 in both BACILLUS CIRCULANS CGTases) (Fig. 4). The other feature that seems to be characteristic of CGTases is the tryptophan (Trp75 in both BACILLUS CIRCULANS CGTases) but it is found also in certain α-amylases. The residue in that position in many α-amylases is be glutamine (Fig. 4), especially in the so-called animal group and/or animal-like α-amylases (JANECEK, 1994b; DA LAGE et al., 1998; SUMITANI et al., 1998; D’AMICO et al., 2000; PUJADAS & PALAU, 2001). The two archaeal THERMOCOCCUS CGTases with shorter β2-strand regions (YAMAMOTO et al., 2000; RASHID et al., 2002) seem to contain more α-amylase-like sequence features also in other conserved sequence regions (JANECEK, unpublished results). Remarkably, the recently sequenced acarviose transferase with specificity very close to a CGTase (HEMKER et al., 2001) has also the shorter version of this region and lacks the glutamine characteristic of CGTase. The predictions concerning the exact enzyme specificity should thus be made carefully in conjunction with experimental biochemical data (WIND et al., 1995; DEL-RIO et al., 1997).

The sixth conserved sequence region – strand β2
As far as the conserved sequence region covering strand β2 of the catalytic (β/α)8-barrel is concerned (Fig. 4), it has been shown to be very helpful in discriminating, for example, the CGTases from α-amylases (JANECEK, 1994a, JANECEK et al., 1995). It was defined first for α-amylases (JANECEK, 1994b) and the efforts were made to show that this β2-strand stretch, flanked in loops by a conserved glycine and proline, could also be evolutionarily important for the (β/α)8-barrel fold (JANECEK, 1995c, 1996a,b, 1997b). The conserved sequence region at strand β2 is thus characterised by a glycine residue followed by seven or eight amino-acid residues and then a conserved proline, although in several cases the Gly and Pro are replaced by other residues (not included in Fig. 4). In order to distinguish CGTases from α-amylases the length of this region has to be taken into account, together with special sequence features (Fig. 4). The CGTases usually have eight residues between the Gly and Pro, while the α-amylases have seven (JANECEK et al., 1995). This is, however, not always the case, and several exo-α-amylases, such as maltotriohydrolases and maltotetraohydrolases, also have eight residues between Gly and Pro (KOBAYASHI et al., 1994; FUJITA et al., 1989), while the two known archaeal CGTases appear to have seven (YAMAMOTO et al., 2000; RASHID et al., 2002).

The seventh conserved sequence region – strand β8
This conserved sequence region (Fig. 5) was identified together with the one covering strand β2 (JANECEK, 1994b). Although it does not contain invariant residues conserved throughout all the α-amylase family members (JANECEK, 1997a;
Fig. 4. The sixth conserved sequence region of the α-amylase family enzymes. The figure focuses on the α-amylases and cyclodextrin glucanotransferases from the family. Sources of the enzymes: α-amylases: *Altemomonas haloplanctis* (Feller et al., 1992), *Bacillus amyloliquefaciens* (Takkinnen et al., 1983), *Bacillus licheniformis* (Yuuki et al., 1985), *Bacillus sp.* KSM-K38 (Hagihara et al., 2001), *Bacillus subtilis* (Yang et al., 1983), *Escherichia coli* (Raha et al., 1992), *Halomonas meridiana* (Coronado et al., 2000), *Lactobacillus amylovorus* (Giraud & Cuny, 1997), *Streptococcus mutans* (Simmons & Russell, 1998), *Streptomyces limosus* (Long et al., 1987), *Thermomonospora curvata* (Petrick et al., 1992), *Thermotoga maritima* (Liebl et al., 1997), *Pyrococcus furiosus* (Dong et al., 1997a), and *Pyrococcus sp.* KOD1 (Tachibana et al., 1996), *Thermococcus hydrothermales* (Levéque et al., 2000), *Thermococcus profundus* (Lee et al., 1996), *Thermococcus sp.* R3 (Jones et al., 1999), *Aspergillus oryzae* (Toda et al., 1982), *Cryptococcus sp.* S-2 (Itofuj et al., 1996), *Saccharomyces fibuligera* (Toda et al., 1987), *Hordeum vulgare* (Rogers & Millman, 1983), *Malus domestica* (Wei et al., 2000), *Drosophila melanogaster* (Boer & Hickey, 1986), *Dermatophagoides pteronyssimus* (Mills et al., 1999), *Panax ginseng* var. *ginseng* (Van Wormhoudt & Sellos, 1996), *Gallus gallus* (Bark et al., 1997), *Struthio camelus* (Kabuto et al., 2000), *Sus scrofa* (Pasero et al., 1988), *Homo sapiens* (Nishi et al., 1986); oligosaccharide-producing amylases: *Bacillus stearothermophilus* (G2) (Didierichen & Christiansen, 1988), *Natronococcus sp.* Ah-36 (G3) (Kobayashi et al., 1994), *Pseudomonas stutzeri* (G4) (Fujita et al., 1989), *Pseudomonas sp.* KO-8940 (G5) (Shida et al., 1992), *Bacillus sp.* #707 (G6) (Tsuchimoto et al., 1988); CGTases: *Bacillus circulans* 8 (Nitschke et al., 1990), *Bacillus circulans* 251 (Lawson et al., 1994), *Bacillus ohbensis* (Sin et al., 1991), *Bacillus sp.* 1011 (Kimura et al., 1987), *Bacillus sp.* A2-5a (Ohdan et al., 2000), *Bacillus sp.* B1018 (Ito et al., 1990), *Bacillus sp.* E1 (Yang et al., 1996), *Bacillus steaithermophilus* (Fujimura et al., 1996), *Brevibacillus brevis* (Kim et al., 1998), *Thermococcus kodakaraensis* sp. ATCC 53627 (Jorgensen et al., 1997), *Thermococcus kodakaraensis* thermosulfurigenes (Bahl et al., 1991), *Klebsiella pneumoniae* (Binder et al., 1986), *Thermococcus kodakaraensis* (Rashid et al., 2002), *Thermococcus sp.* B1001 (Yamamoto et al., 2000). Colour code: conserved glycine and proline at the C- and N-terminal sides of the region − yellow; α-amylase-like features (glutamine and a gap in the positions i−4 and i−1, respectively, in respect to the conserved proline) − red; CGTase-like features (tryptophane and glutamine in the positions i−4 and i−1, respectively, with respect to the conserved proline) − blue.

2000a), it usually starts with a very well-conserved glycine (Gly323 in Taka-amylase A) followed by a proline in the i + 2 position (Fig. 1).
For some specificities it is not easy to identify this region without information derived from the three-dimensional structure (cf. Janecek et al., 2000a; Przygas et al., 2000).

Nevertheless, it has already been demonstrated that this region may contain the features characteristic of archaeal α-amylases (Janecek et al., 1999). Another example is shown in Figure 5 highlighting the features typical for either the “oligo-1,6-glucosidase” or the “neopullulanase” groups that were also discussed above with regard to the fifth conserved sequence region (Fig. 3). These features are mainly the tyrosine and aspartate Tyr454 and Asp456 in neopullulanase in Fig. 5) at the C-terminal side of the region. They could be taken as characteristic residues of the neopullulanase group, while the oligo-1,6-glucosidase-like enzymes should contain glutamine and glutamate, respectively, in the corresponding positions (Fig. 5). Concerning the β8-strand region, amyluscrose, sucrose phosphorylase, and trehalose synthase seem to be rather more like neopullulanase members, although they seem to be clear oligo-1,6-glucosidase-like enzymes with regard to the fifth conserved sequence region (cf. Figs 3, 5). This is not surprising at all if the parts of their sequences between strands β7 and β8 are taken into account; these are different from, and longer than, the equivalent segments of oligo-1,6-glucosidase, α-glucosidase, dextran glucosidase and trehalose-6-phosphate hydrolase (Oslancova & Janecek, 2002). This observation may also be related to the fact that the former members are transferases while the later ones are hydrolases (MacGregor et al., 2001). Finally one can imagine that many possible combinations of the various characteristic motifs will actually be possible and may occur in nature.

The non-catalytic proteins, related to the α-amylase family from the structural and evolutionary points of view (Janecek et al., 1997; Broer & Wagner, 2002), have the segment covering strand β8 of the enzymatic members of the family perfectly conserved in the corresponding positions (Fig. 5). This fact should be of interest because three of the four best-conserved sequence regions with catalytic residues (strands β4, β5 and β7) are less well conserved (Janecek, 2000b).

Conclusion

According to our present knowledge, the α-amylase family consisting of families 13, 70 and 77 of glycoside hydrolases (i.e. clan GH-H) covers hydrolases, transferases, and isomerases (i.e. the enzyme classes 3, 2 and 5). The amino acid transporters and 4F2 heavy-chain antigens (i.e. non-enzymes) are also recognised as relatives of the family. It is possible that the main α-amylase family (clan GH-H) is related to the amylolytic family 57 of glycoside hydrolases, although the hierarchy will very probably be at a higher level (i.e. less strict common criteria) than a clan in terms of its definition. The members of the α-amylase enzyme family act mainly on α-1,4- and α-1,6-glucosidic linkages; however, they also operate on α-1,1-, α-1,2-, α-1,3- and α-1,5-glucosidic bonds. Despite the presence of several conserved sequence regions the invariant amino-acid residues are extremely few, i.e. three catalytic residues in strands β4, β5 and β7 plus the arginine in position i − 2 with regard to the β4-strand catalytic aspartate.

Concerning the number of conserved sequence regions it could be postulated that: (i) each
enzyme specificity and/or even a distinct taxonomic group from the α-amylase family may have its own conserved sequence regions and/or characteristic sequence similarities located in the catalytic (β/α)₇-barrel domain and/or outside this domain; (ii) these regions may discriminate between the α-amylase family members from the point of view of their specificity, structure and evolution; and (iii) all conserved sequence regions may not be common for the entire α-amylase family and they may indicate closer evolutionary relatedness with functionally different enzymes and proteins.

If conserved sequence regions are considered, it can be concluded that the α-amylase family enzymes should be characterised by 4, 5 and in some cases 7 conserved sequence regions, i.e. by as many as possible. So if we go back to the question raised in the title of this review: “How many conserved sequence regions are there in the α-amylase family?” the exact answer is less important than the practical use of the information characteristic of specific regions.

Acknowledgements

This work was supported in part by the VEGA Grant No. 2/2057/22 from the Slovak Grant Agency.

References


Gotttesdiener, K. M., Karpiński, B. A., Lindsten, T., Strominger, J. L., Jones, N. H.,