Review

Relationship of sequence and structure to specificity in the \(\alpha\)-amylase family of enzymes

E. Ann MacGregor a, Štefan Janeček b, Birte Svensson c,*

a Department of Chemistry, University of Manitoba, Winnipeg, Manitoba R3T 2N2, Canada
b Institute of Microbiology, Slovak Academy of Sciences, Stefaníkova 3, SK-81434 Bratislava, Slovakia
c Department of Chemistry, Carlsberg Laboratory, Gamle Carlsberg Vej 10, DK-2500 Copenhagen Valby, Denmark

Received 19 October 2000; accepted 20 December 2000

Abstract

The hydrolases and transferases that constitute the \(\alpha\)-amylase family are multidomain proteins, but each has a catalytic domain in the form of a \((\beta/\alpha)_8\)-barrel, with the active site being at the C-terminal end of the barrel \(\beta\)-strands. Although the enzymes are believed to share the same catalytic acids and a common mechanism of action, they have been assigned to three separate families – 13, 70 and 77 – in the classification scheme for glycoside hydrolases and transferases that is based on amino acid sequence similarities. Each enzyme has one glutamic acid and two aspartic acid residues necessary for activity, while most enzymes of the family also contain two histidine residues critical for transition state stabilisation. These five residues occur in four short sequences conserved throughout the family, and within such sequences some key amino acid residues are related to enzyme specificity. A table is given showing motifs distinctive for each specificity as extracted from 316 sequences, which should aid in identifying the enzyme from primary structure information. Where appropriate, existing problems with identification of some enzymes of the family are pointed out. For enzymes of known three-dimensional structure, action is discussed in terms of molecular architecture. The sequence–specificity and structure–specificity relationships described may provide useful pointers for rational protein engineering. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: \(\alpha\)-Amylase family; Glycoside hydrolase family 13; Glycoside hydrolase family 70; Glycoside hydrolase family 77; Sequence motif; Specificity; \(\alpha\)-1,4 Glucoside bond; \(\alpha\)-1,6 Glucoside bond

1. Introduction

Comparisons of amino acid sequences of glycoside hydrolases and transglycosylases have allowed a classification scheme to be established for these enzymes, based on structure rather than specificity [1–3], i.e., the enzymes have been grouped into more than 80 families, where the members of one family share a common three-dimensional structure and mechanism, and have from a few to many sequence similarities. Families 13, 70 and 77 in this classification contain structurally and functionally related enzymes catalysing hydrolysis or transglycosylation of \(\alpha\)-linked glucans, with retention of anomeric configuration. Many of these enzymes act on starch, and one of the most important starch-degrading enzymes, \(\alpha\)-amylase, is also the most widely-studied member of family 13. Hence this group of enzymes, here con-
sidered to consist of families 13, 70 and 77, is often known as the α-amylase family of enzymes, but is, in fact, comprised of enzymes with almost 30 different specificities. In some cases, specificities of different enzymes in the family overlap and this, coupled with amino acid sequence similarities, has led to confusion in identification, particularly of family 13 enzymes. In this review, we attempt to clarify the relationship between sequence and specificity, and discuss, where information is available, key structural features that contribute to specificity.

2. Domain architecture

The enzymes are multidomain proteins, but share a common catalytic domain in the form of a (β/α)₈-barrel, i.e., a barrel of eight parallel β-strands surrounded by eight helices, the so-called domain A (Fig. 1). This structure has been demonstrated by X-ray crystallography in several enzymes of the α-amylase family (Table 1), although in one instance only seven of the eight helices in the barrel fold are present [16]. In addition, studies of amino acid sequence similarities have led to the prediction that many other enzymes belong to this family and have a similar catalytic domain (Table 1). Usually, the loops that link β-strands to the adjacent helices carry amino acid residues of the active site; some of these loops may be long enough to be considered as domains in their own right. Thus, in most cases where the structure has been determined by crystallography, a large loop between the third β-strand and third helix is discussed as a separate domain, domain B. This loop has an irregular structure that varies from enzyme to enzyme, and, it has been argued, should not always be considered a separate domain but, in some cases, should be thought of as part of a structural unit containing other loops [16]. Similarities in domain B amongst members of sub-groups of the α-amylase family have, however, been found, e.g., various glucosidases resemble each other, while a relationship can be demonstrated between enzymes such as neopullulanase and cyclomaltooligosaccharide [31,32].

In one set of enzymes, the glucan sucrases (often called glucosyltransferases) and dextran sucrases, similarities to α-amylase family proteins can be dem-onstrated in the amino acid sequences of the catalytic domains, yet well-recognised sequence segments appear in a different order. It has been proposed that a circular permutation of the barrel elements occurred during evolution to give enzymes where the first element of the barrel is a helix equivalent to helix 3 of an α-amylase barrel and the last element is a strand equivalent to strand 3 of α-amylases, i.e., instead of B1-H1-B2-H2.....B8-H8 as in α-amylases (where B and H represent strand and helix, respectively), we have in glucan sucrases H3-B4-H4.....B8-H8-B1-H1- B2-H2-B3 [36]. In the enzyme classification scheme based on structure, such enzymes belong to family 70 and are said to form a ‘clan’ GH-H with family 13 and family 77 proteins [35].

In most, but not all, α-amylases, domain A occurs at the N-terminal end of the protein. However, several other members of the family have a distinct domain, domain N, preceding the catalytic domain. The role of this domain is as yet uncertain, and is most probably not identical in all enzymes. In iso-amylase and malto-oligosyltrehalose trehalohydrolase, domain N is close to the loop linking β-strand 3 of the barrel to helix 3, and in the former enzyme, amino acid residues of domain N may form part of the active site [16,20]. (Unfortunately, this domain has been labelled domain E in malto-oligosyltrehalose
trehalohydrolase [20], in disagreement with earlier nomenclature discussed below.) In contrast, it has been suggested in an ‘α-amylase II’ from *Thermoactinomyces vulgaris* that domain N is too far from the active site to participate in enzyme activity, and the domain does indeed appear to be located in a different position, relative to the rest of the molecule, from that found in isoamylase [16,19]. In the crystal structure of the closely related maltogenic ‘amyrase’ (also referred to as cyclodextrinase or CDase) from *Thermus* sp., the domain N of one monomer of a dimer has been suggested to be involved in the enzyme specificity through its interaction with the active site region of the other monomer [18]. Such a situation may also hold for the ‘α-amylase II’, since it is found by X-ray crystallography to exist as a similar dimer, and its domain N occupies a position analogous to that of the equivalent domain in the maltogenic ‘amylase’ [19]. The malto-oligosyltrehalose trehalohydrolase appears to function as a covalent dimer, although dimerisation is not necessary for activity or specificity [20].

The vast majority of members of the enzyme family have another domain, domain C, following the catalytic β/β-barrel. This domain is made up of β-strands and is thought to stabilise the catalytic domain by shielding hydrophobic residues of domain A from the solvent. It has also been suggested that domain C may aid in substrate binding [17,23]. Some amylomaltases, however, appear to lack this

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Structural information on the α-amylase family of enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzyme</td>
<td>EC Number</td>
</tr>
<tr>
<td>(a) Enzymes of known three-dimensional structure</td>
<td></td>
</tr>
<tr>
<td>α-Amylase</td>
<td>3.2.1.1</td>
</tr>
<tr>
<td>Oligo-1,6-glucosidase</td>
<td>3.2.1.10</td>
</tr>
<tr>
<td>Maltotetraohydrolase</td>
<td>3.2.1.60</td>
</tr>
<tr>
<td>Isoamylase</td>
<td>3.2.1.68</td>
</tr>
<tr>
<td>Maltogenic amylase&lt;sup&gt;a&lt;/sup&gt;</td>
<td>3.2.1.133</td>
</tr>
<tr>
<td>Neopullulanase&lt;sup&gt;b&lt;/sup&gt;</td>
<td>3.2.1.135</td>
</tr>
<tr>
<td>Malto-oligosyltrehalose trehalohydrolase</td>
<td>3.2.1.141</td>
</tr>
<tr>
<td>Amylosucrase</td>
<td>2.4.1.4</td>
</tr>
<tr>
<td>Cyclodextrin glucanotransferase</td>
<td>2.4.1.19</td>
</tr>
<tr>
<td>Amylomaltase (4-Glucanotransferase)</td>
<td>2.4.1.25</td>
</tr>
<tr>
<td>(b) Enzymes predicted to belong to the α-amylase family</td>
<td></td>
</tr>
<tr>
<td>α-Glucosidases&lt;sup&gt;c&lt;/sup&gt;</td>
<td>3.2.1.10</td>
</tr>
<tr>
<td>Pullulanase (Limit dextrinase)</td>
<td>3.2.1.41</td>
</tr>
<tr>
<td>Amylopullulanase</td>
<td>3.2.1.141</td>
</tr>
<tr>
<td>Cyclomaltoastarchinase&lt;sup&gt;a,b&lt;/sup&gt;</td>
<td>3.2.1.54</td>
</tr>
<tr>
<td>Dextran glucosidase</td>
<td>3.2.1.70</td>
</tr>
<tr>
<td>Trehalose-6-phosphate hydrolase</td>
<td>3.2.1.93</td>
</tr>
<tr>
<td>Maltohexaohydrolase</td>
<td>3.2.1.98</td>
</tr>
<tr>
<td>Maltotriobhydrase</td>
<td>3.2.1.116</td>
</tr>
<tr>
<td>Maltopentaohydrolase</td>
<td>3.2.1.</td>
</tr>
<tr>
<td>Sucrose phosphorylase</td>
<td>2.4.1.7</td>
</tr>
<tr>
<td>Branching enzyme</td>
<td>2.4.1.18</td>
</tr>
<tr>
<td>Glucan debranching enzyme</td>
<td>2.4.1.25/3.2.1.33</td>
</tr>
<tr>
<td>Maltosyl transferase</td>
<td>2.4.1.1</td>
</tr>
<tr>
<td>Dextran sucrose/Alternansucrase</td>
<td>2.4.1.5/2.4.1.140</td>
</tr>
<tr>
<td>Maltooligosyl trehalose synthase&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5.4.99.15</td>
</tr>
<tr>
<td>Trehalose synthase&lt;sup&gt;d&lt;/sup&gt;</td>
<td>5.4.99.16</td>
</tr>
</tbody>
</table>

<sup>a</sup>α-Amylase-type (AM-type), see Section 5.1 [17]. Neopullulanase-type (N-type), see Section 5.3 [18].
<sup>b</sup>Neopullulanase, resembles closely the maltogenic amylase (N-type), see Section 5.3: referred to also as cyclodextrinase [18].
<sup>c</sup>Only α-glucosidases belonging to family 13 are considered here; other α-glucosidases belong to family 31.
<sup>d</sup>These are considered to be isomerases.
domain; it cannot therefore be necessary for activity or stability in all cases [27].

Cyclodextrin glucanotransferases and a second type of maltogenic ‘amylase’, from Bacillus stea-
thermophilus and quite distinct from that of Thermus sp., have additional β-sheet domains, domains D and E, after domain C [17,22–26]. While no role has been assigned to domain D, domain E occurs in a variety of other starch-degrading enzymes, not all from family 13, i.e., glucoamylases (family 15) and a few β-
amylases (family 14) and α-amylases [29]. This domain is believed to be important for binding granular starch and has also been proposed to provide an extension to the active site in cyclodextrin glucano-
transferases [22,29].

Much of the specificity of each enzyme is, however, thought to be defined by amino acid residues of domains A and B, and characteristic differences in these domains will be discussed here.

3. Catalytic mechanism and substrate binding at the active site

Throughout the family, the enzymes are believed to have a similar mechanism of action, and so the catalytic amino acid residues are thought to be common to all the enzymes [37]. Anomeric configuration is retained when the substrate is converted to product, i.e., the enzymes act on α-linkages in glucans or glucosides and yield α-linked products. The reaction is believed to proceed by a double displacement mechanism (Fig. 2). During the first displacement an acid group on the enzyme protonates the glyco-
sidic oxygen, bringing about scission of the C1-O bond and transient formation of an oxocarbenium ion-like transition state [38–40]. A nucleophilic acid group of the protein attacks at the sugar anomeric centre to give a β-glycosyl enzyme intermediate, while the aglycone of the substrate leaves the active site. In hydrolysis, during the second displacement, the process just described is essentially reversed by attack, at the anomeric centre, by a water molecule activated by the carboxylate form of the former pro-	on donor (Fig. 2). This second stage of the reaction proceeds via an ion-like transition state, as before, to yield a product with α-anomeric configuration and reprotonation of the original acid group. Transglycosylation can occur if the attacking group in the second displacement of the reaction is a free hydrox-
yl of a sugar residue rather than water [40,41].

In Taka-amylase A, the first protein of this family to be examined by X-ray crystallography, three acidic residues, i.e., one glutamic and two aspartic acids were found at the centre of the active site [8], and subsequent mutational studies have shown that these residues are essential for catalysis [33,37]. The glutamic acid residue is now believed to be the proton donor, while the first of the two conserved as-
dpartic acids appearing in the amino acid sequence of an α-amylase family member is thought to act as the nucleophile. The role of the second aspartic acid is less certain, but it has been suggested to be involved in stabilising the oxocarbenium ion-like transition state and also in maintaining the glutamic acid in the correct state of protonation for activity [40].

The active site of an enzyme belonging to the α-
amylase family is considered to be made up of a number of subsites, each subsite being capable of interaction with one glucose residue of the substrate (Fig. 3). The subsites themselves are composed of side chains of amino acid residues situated on loops in the enzyme structure that connect the C-terminal ends of β-strands to the N-terminal ends of the ad-
jacent helices of the (β/α)₈-barrel of the catalytic domain. In the case of malto-oligosyltrehalose trehalo-
dyrolase, however, an α-L loop, i.e., joining a helix
to the next β-strand of the barrel, is unusually long
and may contribute to the specificity of the active site
[20]. Because the architecture of the β-α loops varies
from enzyme to enzyme, the number and nature of
the subsites at the active site also varies, and is char-
acteristic of a particular enzyme. Different enzymes
of the α-amylase family act on different substrates
and the only shared feature of the substrates of the
different enzymes is an α-linked glucose residue that
should bind at subsite −1. Hence strong similarities
should be found in the residues making up this sub-
site.

In Taka-amylase A the proton donor (Glu230), the
catalytic nucleophile (Asp206), the second conserved
aspartic acid (Asp297), and two histidine residues
(His122 and His296) were postulated as being impor-
tant at subsite −1 and this has been confirmed by
investigations on mutated proteins, although the His
residues are not invariant throughout the family
[33,36,37]. These five residues occur near the ends
of strands 3, 4, 5 and 7 of the β/α-barrel and are
found in four short sequences, long-recognised as
being conserved in α-amylase family enzymes. Typi-
cal examples of these sequences are shown in Fig. 4.

Further, less obvious conservations of short se-
quence segments can be found at strands 2 and 8
and in loop 3 between strand 3 and helix 3 (i.e.,
domain B of many enzymes) [33]. The conservations
may be related to maintenance of structure while
variations within these segments may be character-
istic of certain enzyme specificities [33]. In many en-
zymes of the α-amylase family, the conserved se-
quence at strand 2 is characterised by a glycine
residue followed eight or nine amino acid residues
later by proline, although this is not invariant. It
has been suggested that this region can be used to
distinguish cyclodextrin glucanotransferases from α-
amylases, since the former enzymes usually have
eight residues between the Gly and Pro, while the
latter enzymes have seven [42]. This is, however,
not always the case, and several exo-α-amylases
have eight residues between Gly and Pro [43–45],
and one CGTase appears to have seven [46].

4. Conserved sequences and specificity

Different enzymes of the family have different spe-
cificities; some are active only on α-1,4 glycosidic
bonds between glucose residues, others on α-1,6
bonds exclusively, some on both bond types, yet
others can cleave sucrose, while there are related en-
zymes that can hydrolyse or form the inter-glucose
link in trehalose (Fig. 5). For all of the enzymes,
activity involves binding a glucose residue of the sub-
strate at subsite −1, while the nature of the portion
of the substrate binding at subsites +1 and +2 varies
with the specificity of the enzyme, particularly with
respect to the type of bond to be formed or de-
stroyed, e.g., whether α-1,4 or α-1,6. Thus the amino
acid residues making up subsites +1 and +2 of an
enzyme may be expected to vary with the specificity
of the enzyme.

From structural studies with mutated enzymes and
substrates or inhibitors, it appears that amino acids
of conserved sequence segments II and III (Fig. 4)
generally constitute subsites +1 and +2 [17,40,47–54].
In particular these acids are 209, 210, 231, 232 and
possibly 233 and 234 (Taka-amylase A numbering
will be used throughout), and occur on loops 4 and
5 of the (β/α)₈ barrel [8]. Thus residues at these po-
sitions may be characteristic for specificity with re-
spect to bond-type. Of 175 sequences of enzymes

| Short conserved amino acid sequences of α-amylase family enzymes: a, for Taka-amylase A; b, a generalised set of sequences. X, usually a hydrophobic residue; B, usually a hydrophilic residue; Z, a residue important for specificity. |
|---|---|---|
| I | 116 VDVVANH 122 | II | 202 GLRIDTVKH 210 |
| III | 226 YCIGEVLDO 237 | IV | 292 FVENHD 297 |
| More generally | | |
| I | XDXXXNH | II | GXRXXXXX |
| III | XXX(G/A)EZZZ | IV | XBBHBD |

Fig. 4. Short conserved amino acid sequences of α-amylase family enzymes: a, for Taka-amylase A; b, a generalised set of sequences. X, usually a hydrophobic residue; B, usually a hydrophilic residue; Z, a residue important for specificity.
from families 13 and 77 that are specific for α-1,4 glucosidic bonds, 160 or 91% have Lys or Arg at the position equivalent to 209 of Taka-amylase A, while these two residues are conspicuously absent from this position in enzymes with specificity for other α-glucosidic bonds. Further, 170 or 97% of the α-1,4-bond-specific enzymes have His or Gly at the position equivalent to 210. Where Gly is present at 210, an aromatic residue is found at 207 and seems to replace, in the three-dimensional structure, the His ring otherwise found at 210 in α-1,4-bond-specific enzymes [9,15]. Amongst the α-amylases, all the archaeal and plant enzymes have Gly at position 210 [55], but the significance of this is not understood. A view of a substrate analogue bound at the active site of an α-amylase is shown in Fig. 6.

In the case of enzymes acting on α-1,6-linkages only, the situation is less clear-cut. In 28 sequences, the most common amino acid residues at 209 are Gly or Ser found in 20 or 71% of these proteins. Gly is also found at this position, however, in some α-glucosidases that appear not to be active on α-1,6-linkages [47,56] and in sucrose phosphorylases that catalyse the phosphorolysis of sucrose [57–60]. In addition, 26 or 93% of the enzymes that are specific for α-1,6-bonds have a hydrophobic residue, Val, Leu or Met, at position 207.

At position 210, in α-1,6-link-specific enzymes, 24 or 86% of the proteins have a hydrophobic or aromatic amino acid. Enzymes that can act on both α-1,4 and α-1,6 bonds at a single active site, such as neopullulanases and some amylopullulanases, have mainly Asn at 209 and Glu at 210 (in 16 out of 22, or 68% of the enzymes) but other possibilities are known.

Enzymes acting on sucrose can be considered in two groups - the glucan sucrases with a permuted β-barrel, and the sucrose phosphorylases and amylosucrase with a ‘normal’ barrel. In the first group, all of the sequences examined have Asp–Asn at positions equivalent to 209–210 of Taka-amylase A, while in the second group, all have Gly or Ala at 209, followed by an aromatic Phe or Tyr at 210. Thus there is more than one architecture possible, in the sucrose-binding enzymes, for subsite +1, where the fructose might be expected to bind.

Most of the glucosidases (24 out of 31, or 77%), whether oligo-1,6-glucosidase, α-glucosidase, dextran glucosidase or trehalose 6-phosphate hydrolase, have a large hydrophobic or aromatic residue at the position equivalent to 210 of Taka-amylase A. This might allow for interaction with a variety of sugar or even non-sugar residues at subsite +1.

At positions 231 and 232, the most obviously distinctive sequence is the Trp-Tyr/Phe that occurs in cyclodextrin glucanotransferases (CGTases) and one maltogenic ‘amylase’, but none of the other family 13 enzymes (Table 2). The latter enzyme lacks other key features of the CGTases (see below) and has no cyclising activity [17,61].

In some, but not all, enzymes with marked transferase activity an acidic residue, either Asp or Glu is

Fig. 5. Four possible substrates for members of clan GH-H. A: maltose, B: isomaltose, C: sucrose and D: trehalose.

Fig. 6. Schematic diagram of the interactions of acarbose with Taka-amylase A. Hydrogen bonds are shown as dotted lines and water molecules as spheres. (Reproduced from [54] with permission.)
found at 231. Such an arrangement of two contiguous acidic residues (231 following the catalytic Glu at 230) is not found in hydrolases of the family (see Table 2).

Janecek et al. have already pointed out that residues characteristic of a particular type of enzyme can be located at other positions in the amino acid sequence, e.g., an Ile at position 203 is found in cyclodextrin glucanotransferases [42], but not in other 1,4-bond-specific members of the family, notably α-amylases and the above-mentioned maltogenic ‘amylase’.

A close examination of over 300 amino acid sequences of α-amylase family members (listed in [62]) has enabled a table to be drawn up that shows sequences characteristic of known specificities of enzymes within this family (Table 2). In several cases, insufficient structural information is available for it to be possible to offer an explanation of the relationship between these sequences and enzyme specificity.

5. Structure–specificity relationships in individual enzymes

5.1. Enzymes specific for α-1,4 glycosidic bonds

5.1.1. α-Amylases

α-Amylases are generally considered to be endo-acting enzymes. There are enzymes, however, that hydrolyse starch polysaccharides to products with the α-anomeric configuration and are believed to act preferentially at one end of a polysaccharide chain to give primarily one size of small oligosaccharide, i.e., are exo- rather than endo-acting hydrolases.

Some of these enzymes now have Enzyme Commission numbers distinct from that of α-amylase, e.g., maltogenic ‘amylase’ (3.2.1.133), maltotriohydrolase (3.2.1.116), maltotetraohydrolase (3.2.1.60), and maltooctahydrolase (3.2.1.98) producing α-maltose, maltotriose, maltotetraose and maltohexaose, respectively. With the exception of some so-called maltogenic ‘amylases’, these enzymes cannot be distinguished from endo-acting α-amylases on the basis of amino acid residues in conserved sequence regions II and III (Fig. 4). This is to be expected, since the exo-activity is conferred, at least in the case of the maltotetraohydrolase where the structure is known, by residues on loop 3, while residues in conserved regions II and III form the beginnings of loops 4 and 5 of the β-barrel [63]. At this stage in our understanding of structure-function relationships in family 13 enzymes it is not possible to predict which residues of the other exo-acting members of the family ensure exo- rather than endo-activity, although they are likely to be in loops 1, 2, 3, 7 and 8, since these loops contribute to binding glucose residues of the substrate on the glycone side of the bond on which the enzyme acts [6,48,63,64].

Consideration of the maltogenic ‘amylases’ illustrates some of the problems of confusion and nomenclature that have arisen in this family. In fact, the enzymes classed as maltogenic ‘amylases’ under EC number 3.2.1.133 appear to be of two distinct types, one which has the characteristic signature of α-1,4-linkage-specific enzymes of Lys–His at positions 209–210 [65,66] and the other with the Asn–Glu characteristic of enzymes active on both α-1,4 and α-1,6 bonds [67–70] (Table 2). There is evidence that enzymes of the latter type are indeed active on both types of linkage [68,71] – hence, in this case, the name amylase, usually reserved for 1,4-bond-specific enzymes, is misleading. The differences between the two types of maltogenic ‘amylase’ extend much further than simple signature sequences. The B. stearothermophilus enzyme of the first type (hereafter called type AM, i.e., amylase-like) has been studied by X-ray crystallography, and the overall structure has been found to be remarkably similar to that of a cyclodextrin glucanotransferase, with five domains, the first being the catalytic (β/α)-barrel [17]. The authors suggest that the enzyme attacks by an endo-mechanism, but that the final product is mainly maltose. Indeed they consider that the enzyme should be classified as an α-amylase. The structure shows key differences in loop 3 of the barrel between this enzyme and a CGTase (see below). Such differences make it less likely that the maltogenic ‘amylase’ AM could hydrolyse a hexa-, hepta- or octa-saccharide from the non-reducing end of a substrate, as would be required for cyclodextrin formation, or that cyclisation could take place. A second maltogenic ‘amylase’ (from B. subtilis) has the Lys–His characteristic of an α-1,4-linkage-specific enzyme and is considered here as an AM-type enzyme. It has an amino acid sequence long enough and similar enough to the B. stearothermophilus enzyme that it too may
Table 2
Sequences at \( \beta \rightarrow \alpha 4 \) and 5 of domain A characteristic for enzymes in clan GH-H of different specificity

<table>
<thead>
<tr>
<th>Enzyme</th>
<th>Residues at position number (Taka-amylase A numbering)</th>
<th>No. of sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enzymes acting on sucrose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amylosucrase</td>
<td>A V A F</td>
<td>1</td>
</tr>
<tr>
<td>Glucan sucrase</td>
<td>A V D N M/D/A</td>
<td>14</td>
</tr>
<tr>
<td>Sucrose phosphorylase</td>
<td>A/V G Y/F</td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>A F A Y</td>
<td>2</td>
</tr>
<tr>
<td>Enzymes acting on trehalose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maltooligosyl trehalose synthase</td>
<td>H P/T/L/I D G</td>
<td>8</td>
</tr>
<tr>
<td>Trehalose-6-phosphate hydrolyase</td>
<td>L V I/V N L S V/L</td>
<td>2</td>
</tr>
<tr>
<td>Trehalose synthase</td>
<td>A I/V P Y N M/Q</td>
<td>5</td>
</tr>
<tr>
<td>Enzymes acting on ( \alpha 1,4- ) and ( \alpha 1,6 )-bonds</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Branching enzyme</td>
<td>A/G I/V/T A/T/V S/N E/D</td>
<td>26</td>
</tr>
<tr>
<td>Glycogen debranching enzyme</td>
<td>Y/N C H S</td>
<td>3</td>
</tr>
<tr>
<td>Neopullulanase</td>
<td>A A M/Q H/Y</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>T H P Y</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>V P D C</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>V A N E</td>
<td>1</td>
</tr>
<tr>
<td>Amylopullulanase</td>
<td>V A/E N E</td>
<td>No K/L 4</td>
</tr>
<tr>
<td>Malto-1,6-glucosidase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cyclomaltodextrinase(^b)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>V A N E</td>
<td>No K/L 4</td>
</tr>
<tr>
<td>( \alpha )-Glucosidases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oligo-1,6-glucosidase</td>
<td>V I G/N F/M/S</td>
<td>5</td>
</tr>
<tr>
<td></td>
<td>V L W L</td>
<td>2</td>
</tr>
<tr>
<td>Dextran glucosidase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pullulanase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Limit dextrinase)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isoamylase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Enzymes acting on ( \alpha 1,6 )-bonds only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maltooligosyltrehalose trehalohydrolase</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Malto-1,4-glucosidase</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
be a 5-domain protein [65]. Curiously, the sequence is 97% identical, over the first three domains, to that of a *B. subtilis* \(\alpha\)-amylase of known three-dimensional structure [6]. The \(\alpha\)-amylase is, however, truncated with respect to the maltogenic 'amylase' and lacks the last 192 residues of the latter enzyme, i.e., possible domains D and E. Such truncations in *B. subtilis* \(\alpha\)-amylase (resulting from base deletions in the gene) have been observed previously [72] and are not believed to affect specificity on soluble substrates, but decrease activity on starch granules. The enzyme produces maltose as a major, but not sole, product of starch hydrolysis. We suggest that the *B. subtilis* maltogenic 'amylase' also may be considered as an \(\alpha\)-amylase.

X-ray crystallographic studies of one of the second type of maltogenic 'amylase' (called N-type, i.e., neopullulanase-like) [18] shows a different domain organisation, with the enzyme having a domain N before the catalytic \(\beta\)-barrel, and lacking domains D and E. Again the enzyme gives maltose as the major, but not only, product of starch hydrolysis. In addition, the enzyme is shown to be active on both \(\alpha\)-1,4 and \(\alpha\)-1,6-linkages between glucose residues, and its amino acid sequence is closely related to those of another maltogenic 'amylase' [68] and a neopullulanase [73] known to have dual bond-type specificity (see later).

Thus, at present, the name maltogenic amylases and EC number 3.2.1.133 are being used for two distinct enzymes, neither of which produce exclusively maltose from starch.

Further confusion arises from the description of two enzymes from *Thermoactinomyces vulgaris* as 'pullulan-hydrolysing' \(\alpha\)-amylases, TVAI and TVAII. Specificity studies on both enzymes show that they can hydrolyse \(\alpha\)-1,4 glucosidic linkages, giving, for example, panose from pullulan. Both enzymes, however, can hydrolyse cyclodextrins and \(\alpha\)-1,6 linkages in small oligosaccharides, i.e., they show activities characteristic of neopullulanases. In addition, the amino acid sequences of the enzymes show a marked similarity to a neopullulanase and a cyclodextrinase [74]. The structure of TVAII has been determined by X-ray crystallography and shows the domain N that is rare in \(\alpha\)-amylases, but common in dual bond-specificity enzymes [19]. For this review, we have considered both TVAI and TVAII as neopullulanases.

The true \(\alpha\)-amylases act only on \(\alpha\)-1,4-glucosidic linkages, bringing about hydrolysis of starch polysaccharides or transglycosylation with small oligosaccharides that are poor substrates for the enzymes [75–77]. While there are exo-\(\alpha\)-amylases producing mainly one size of oligosaccharide from the non-reducing end of a long \(\alpha\)-1,4-linked glucan chain, the endo-\(\alpha\)-amylases also give greater yields of some oligosaccharides than others by preferential action near chain ends. This action can be the combined result of the number of subsites at the active site and the
tendency to multiple attack, i.e., the scission of more than one glucosidic bond during a single enzyme-substrate encounter. Enzymes giving high yields of maltose and maltooltriose during amylose hydrolysis, e.g., α-amylases from mammals, some insects and bacteria, are generally considered to have 5 important subsites at the active site (−3 to +2 of Fig. 3) and in several cases this has been confirmed by X-ray crystallography of enzyme-oligosaccharide or enzyme-carbohydrate inhibitor complexes [4,6,10, 13,49,50,78]. In some enzymes a sixth subsite at +3 is believed to be present [17]. The ability to bind long substrates with five or six contiguous internal glucose residues interacting directly with the enzyme gives these enzymes endo- rather than exo-character. In addition, some mammalian α-amylases are believed capable of high levels of multiple attack, accentuating the production of small oligosaccharides such as maltose or maltooltriose early in amylose hydrolysis [79]. This multiple attack may be facilitated by the presence, in mammalian α-amylases, of a movable loop, within loop 7 of the β-barrel, that can interact with the substrate at subsites −2 and +2 [13,49,50,80] and has been observed in two distinct conformations by X-ray crystallography. Differences both in the level of multiple attack and in the detailed action on oligosaccharides among the highly conserved mammalian α-amylases are believed to be due to a small number of amino acid substitutions at the active site [13].

Other endo-acting α-amylases give larger oligosaccharides as products of starch degradation, mainly because scission of α-1,4-linkages close to the non-reducing end of a substrate molecule is difficult for these enzymes. Such α-amylases, notably from Aspergillus oryzae, barley and Bacillus licheniformis, have more extended active sites where the architecture of loops 1, 2, 3, 7 and 8 of the β-barrel is different for each enzyme and unlike that of the mammalian amylases [5,8,9,11–13,54,80,81]. These are the loops critical for interaction with substrate on the glycone side of the bond to be broken. Modelling of a complex between barley α-amylase and an extended amylose fragment suggests, for example, that a tyrosine on loop 3 of the barrel may be important for interaction with a glucose residue at subsite −6 [51], hence ensuring that barley α-amylases produce maltodextrins smaller than maltohexaose from amylopectin (i.e., from non-reducing substrate chain ends) only with great difficulty [82].

Other α-amylases are known that appear to lack several of the short conserved sequences of family 13 enzymes. Such amylases are found in thermophilic bacteria and archaea, and form family 57 of the classification scheme for glycoside hydrolases. A possible distant sequence relationship between the two families has been discussed [83].

5.1.2. Cyclodextrin glucanotransferases

Another subgroup of enzymes of the α-amylase family that have been widely studied are the cyclodextrin glucanotransferases, and several three-dimensional structures have been determined, for free enzymes, mutated enzymes and complexes with oligosaccharides [22–26,40,41,48,52,53,84]. As stated earlier, these enzymes have five domains, the first domain at the N-terminal end of the protein being the catalytic (β/α)8-barrel, while the last domain, E, may carry amino acid residues important for guiding a long substrate to the active site [85]. These enzymes are capable of bringing about some hydrolysis of glucan chains, but differ from α-amylases primarily in their ability to produce cyclodextrins by cyclising an oligosaccharide of 6–8 glucose residues cut from the non-reducing end of a chain. Mutation studies indicate that a Phe or Tyr residue (amino acid 195 in the CGTase of B. circulans [22,23]) in loop 3 of the β-barrel, that is lacking in loop 3 of α-amylases, is important for cyclisation activity in CGTases [86]. This critical residue is absent from the maltogenic ‘amylase’ AM of B. stearothermophilus that resembles a CGTase in structure [17,66]. Detailed structural studies of CGTase with bound oligosaccharides or cyclodextrin have led to the conclusion that when a long substrate is bound, the enzyme active site adopts a conformation similar to that of an α-amylase where a Lys residue (209 of Fig. 4) is important for substrate binding at subsite +2 (Fig. 3). After bond scission, the glycone part of the substrate – 6, 7 or 8 glucose residues long, with the size preference depending on the source of the CGTase – remains covalently bound to an Asp residue (206 of Fig. 4) [87], and a conformational shift takes place in the enzyme that forces the non-reducing-end glucose of the substrate into subsite +1, while the penultimate glucose occupies subsite +2. The substrate, however,
is now further away from the Lys at subsite +2 and residues specific for CGTase become more important for substrate binding at this subsite, e.g., a phenylalanine of loop 3 (residue 183 of *B. circulans* CGTase [22,23]) and another on loop 5, equivalent to residue 232 of Taka-amylase A [40,41,48,53]. Cyclisation can then take place by glucosidic bond formation between C1 and C4 of the glucose residues at subsites −1 and +1, respectively – essentially a reversal of bond scission. The first of these phenylalanine residues is missing from the maltogenic ‘amylase’ AM, and both are absent from α-amylases, presumably rendering these enzymes unable to form cyclodextrins [66]. Mutational studies indicate that critical residues in loop 3 may influence the size of the cyclodextrins produced by the CGTase, i.e., whether predominantly α-, β- or γ-cyclodextrin [48], although other residues can have an effect [88].

5.1.3. 4-α-Glucanotransferases

Other enzymes, in addition to cyclodextrin glucanotransferases, are primarily transferases, but without cyclising activity, e.g., maltosyl transferase and 4-glucanotransferases. They act on linear α-1,4-linked glucans, removing a fragment from one and transferring it to another. Although the structure of one of these 4-glucanotransferases has been determined, the relationship between structure and activity is not yet understood [27]. The enzyme clearly, however, belongs to the α-amylase family in terms of its structure and putative catalytic amino acid residues, and it seems an unnecessary complication that the enzyme is not included in family 13 of the glycoside hydrolase classification scheme, but has been assigned to a new family, family 77, even if that family is considered as part of the α-amylase ‘clan’ GH-H [35]. Several, but not all of these transferases, share an unusual feature at sequence I of Fig. 4, i.e., His 122 forming an important part of subsite −1 is not conserved, but may be replaced by Thr in maltosyl transferase or Pro, Val and Ser in 4-glucanotransferases. It is not yet possible to say whether this is relevant to transfer activity.

5.2. Enzymes specific for α-1,6 glycosidic bonds

At present, structures of oligosaccharide complexes with α-1,6-linkage-specific enzymes have not been obtained. In general, however, these enzymes have relatively large residues, Leu, Val or Gln following the first catalytic Asp residue (206 of Fig. 4) in the sequence. In both known structures of α-1,6-bond-specific enzymes, isoamylase and oligo-1,6-glucosidase, this residue could interfere with binding of a 4-linked glucose on the aglycone side of the bond to be broken and may help to determine bond specificity [14,16]. In α-amylases, when a large residue follows the Asp, e.g., phenylalanine in barley α-amylase, the side chain of this residue is closer to the position of the common His210 of several 1,4-bond-specific enzymes and helps rather than hinders binding at subsite +1 [64]. Part of the active site of the oligo-1,6-glucosidase appears to be in the form of a pocket which may be able to bind the single 1,6-linked glucose residue to be split from the substrate [14], while that of the isoamylase is a shallow groove more characteristic of the active site of a polysaccharide endo-hydrolase [16]. Isoamylase easily removes long branches from amylopectin, but the common plant debranching enzymes, known as limit dextrinases, are better adapted to the hydrolysis of α-1,6-linkages in small oligosaccharides resulting from the hydrolysis of starch by α- and β-amylases during seed germination. These plant debranching enzymes resemble most closely, in sequence, the bacterial pullulanases (30–40% identity) and are now often called plant pullulanases. The structural basis for the specificity differences between isoamylases, limit dextrinases and bacterial pullulanases is not yet understood.

5.3. Enzymes acting on both α-1,4 and α-1,6 glycosidic bonds

Enzymes acting on both α-1,4- and α-1,6-linkages may do so in a very specific way, e.g., branching enzymes, that usually rupture an α-1,4-bond and make a 1,6-linkage, create branches in glycogen or amylopectin by removing an oligosaccharide from the non-reducing end of a straight 1,4-linked glucan chain and reattaching the oligosaccharide as a branch. Two major groups of starch branching enzymes are found in plants (see for example [89]), with the A group (also known as SBE-II) having a much longer domain N than the B group (SBE-I). In general
the A-type enzymes are more active on amylopectin than amylose, while the opposite is true for the B-type branching enzymes. In addition, the former enzymes seem to transfer shorter chains to form branches than the latter [90]. It is not yet known whether domain N can form an extension of the active site and hence be important for specificity, but in the maize enzymes both the N-terminal and C-terminal regions of the proteins seem to exert an influence [90,91].

As pointed out earlier, the catalytic glutamic acid of branching enzymes is usually followed, in the amino acid sequence, by a second acidic residue, either glutamic or aspartic acid. Mutation of this second acidic amino acid in Escherichia coli branching enzyme showed that it is important, but not essential, for activity and has an effect on specificity [92].

Glycogen debranching enzymes are believed to have two active sites, one for breaking and making 1,4-linkages and the other for removal of a 1,6-linked glucose residue, to release β-glucose. The first active site is thought to resemble that of an α-amylase-family enzyme. Indeed the aspartic acid residue acting as the catalytic nucleophile has been identified for the rabbit enzyme [93]. The second site, in contrast, forms a product with inversion of anomeric configuration and so must differ substantially from a family 13 active site. Its amino acid residues have not yet been identified.

Neopullulanases, certain maltogenic ‘amylases’ (N-type) and most amylopullulanases, however, can hydrolyse and synthesise both α-1,4- and α-1,6-linkages at a single active site. Unfortunately a large protein with two active sites, one specific for α-1,4-bonds and the other for α-1,6-bonds, has also been called an amylopullulanase, leading to confusion with the enzymes possessing a single active site [94–96]. Further confusion has arisen from the practice of calling pullulanases, that are α-1,6-linkage-specific, ‘pullulanase type I’ enzymes and amylopullulanases, with a single active site that processes α-1,4- and α-1,6-bonds, ‘pullulanase type II’ enzymes [97].

Where enzymes can act on more than one type of linkage, there must be some flexibility at the active site, particularly at subsite +1, since a 4-linked glucose and a 6-linked glucose cannot fit in an identical way into an active site. In many ‘dual-specificity’ enzymes, Asp, Asn or Gln is found at position 209 (Fig. 4) and an Asp or Glu at 210. Mutation of these residues has been shown to alter bond specificity [98,99] and relative amounts of hydrolysis versus transglycosylation [100], but the effect on enzyme structure has not yet been determined.

Cyclomaltodextrinases are usually considered to act on α-1,4-linkages only, but there is a report that they can hydrolyse other α-bonds [101]. They have been included here with dual bond-type specific enzymes because of close sequence similarities with enzymes such as neopullulanases. Cyclodextrinases are defined as enzymes that hydrolyse cyclodextrins and some linear maltodextrins quickly, but have little activity on starch. Some maltogenic ‘amylases’ (N-type) also can hydrolyse cyclodextrins, as well as starch, [68,71,102] and so the specificities of the two types of enzymes are not distinctly different. Several maltogenic ‘amylases’ N, notably those with sequence VANE at positions 207–210 (Taka-amylase A numbering, Fig. 4) can also act on α-1,6 bonds in addition to α-1,4-linkages [68,71,102,103]. Neopullulanases are defined as enzymes hydrolysing primarily α-1,4-bonds of pullulan to give panose, but are also known to be active on α-1,6-linkages. Since maltogenic ‘amylases’ N have been shown to give panose from pullulan [68,71,102,103], again there is an overlap in specificity between neopullulanases and the maltogenic ‘amylases’ N that act on more than one type of linkage. In addition, the amino-acid sequences of some maltogenic ‘amylases’ N and neopullulanases are very similar: 86% identity for example between a Bacillus stearothermophilus neopullulanase and a maltogenic ‘amylase’ from Thermus species [70,73] and 99% identity between a putative neopullulanase from Bacillus subtilis and a maltogenic ‘amylase’ from the same organism [69,104]. In this latter case it seems very likely that the two sequences correspond to isozymes of the same enzyme. As has been pointed out earlier in this review, structures of a maltogenic ‘amylase’ N from Thermus sp. and an ‘α-amylase II’, here considered to be a neopullulanase, from Thermomonospora vulgaris appear very similar [18,19]. Thus it may not always be possible to distinguish between maltogenic ‘amylases’ (N-type) and neopullulanases, and the names may have been assigned simply on the basis of the first substrates studied.

In contrast, the cyclodextrinases seem to be less
crose and K gerose or isomaltose [47]. Again, there must be some structural features.

Furthermore, it has been postulated that some enzymes of a group of cyclodextrinases, maltogenic ‘amylases’ and neopolullanases have been misnamed [106]. More recently it has been suggested that, since neopolullanases, maltogenic ‘amylases’ (N-type) and cyclomaltodextrinases all hydrolyse cyclodextrins and share marked sequence similarities, they may be considered a cyclodextrin-metabolising subfamily of the α-amylase enzyme family [32].

Interestingly, a pullulanase from Bacillus flavocaldarius has many sequence features resembling those of a neopolullanase [32,108], but appears to lack the domain N usually found in pullullanases and neopolullanases.

The amylopullullanases are much larger proteins, in general, than neopolullanases, cyclodextrinases and N-type maltogenic ‘amylases’, e.g., molecular masses of amylopullullanases are well over 100 000 (ca. 140 000–180 000) while the other enzymes have molecular masses in the range 60 000–80 000. Amylopullullanases can bring about hydrolysis of α-1,4-linkages in amylose and α-1,6-linkages in pullulan [109], but it is not yet possible to relate this activity to structural features.

5.4. α-Glucosidases

α-Glucosidases are enzymes that remove an α-linked glucose, usually from the non-reducing end of a chain, although at least one enzyme appears to remove glucose from the reducing end of a chain [110]. The α-glucosidases that act at non-reducing chain ends may not be specific for the aglycone part of the substrate – for example, an α-glucosidase from Candida albicans can hydrolyse maltose, sucrose and α-methyl maltoside, but not trehalose, nigerose or isomaltozose [47]. Again, there must be some flexibility at the active site for binding substrate at subsite +1.

Some confusion seems to have arisen, however, when classifying an enzyme as an α-glucosidase or as an oligo-1,6-glucosidase, specific for hydrolysing an α-1,6-linked glucose residue from a substrate. An ‘oligo-1,6-glucosidase’ from Bacillus sp. F5 has been shown to be active on both α-1,4- and α-1,6-linkages [111], as has an ‘α-glucosidase’ from Bacillus sp. SAM 1606 [112]. It is interesting that both of these enzymes have the sequence Gly-Gly at positions equivalent to 232–233 of Taka-amyase A, i.e., in conserved sequence III of Fig. 4 [111,113], while glucosidases that hydrolyse either α-1,4- or α-1,6-bonds (but not both) do not. Studies of mutations of the first of these glycine residues in the ‘α-glucosidase’ have shown that this residue has a marked influence on specificity [114]. Further, a Bacillus flavocaldarius oligo-1,6-glucosidase, for which the specificity has been studied [115,116], is 98% identical in sequence to a Thermus caldophilus ‘α-glucosidase’ [117]. It seems unlikely, given such a high sequence similarity, that the specificities of the two enzymes are substantially different, and so it is probable that the ‘α-glucosidase’ in this case will prove to be an oligo-1,6-glucosidase.

The glucosidases of family 13, i.e., oligo-1,6-glucosidase, α-glucosidase, dextran glucosidase and trehalose 6-phosphate hydrolase, show some strong sequence similarities [118], and may be considered a sub-group of the α-amylase family. Thus any prediction of specificity based on sequence comparisons should be carried out with great care.

5.5. Enzymes acting on sucrose

Enzymes, other than cyclodextrin glucanotransferases, function as transferases, but some transfer glucose from a sucrose molecule to a growing polysaccharide chain. These are amyllosucrase that synthesises an α-1,4-bond, and glucan sucrase (also known as glucosyl transferase). In the case of the second enzyme, however, the new bond formed may be an α-1,2-, α-1,3-, or α-1,6-linkage. The basis of this specificity is not yet understood.

While it has been suggested that secondary structure elements of the (β/α)8-barrel of glucan sucrases are permuted [36], the contrary argument has been
made that the β-barrel elements of these enzymes appear in the same order as in α-amylases [119]. The sequences on which the latter suggestions were based, however, are very variable in the glucan sucrases and, while occurring in one such enzyme, are not found in others. In contrast, the sequences on which the postulate of circular permutation was based are common to all the glucan sucrases [36]. At present, no explanation can be offered as to why amylosucrase should function with a ‘normal’ β-barrel while glucan sucrases require a permuted barrel. As in some other transferases, one of the conserved histidine residues making up subsite −1, i.e., His122 of Taka-amylase A (Fig. 4), is replaced, by a glutamine residue in glucan sucrases. Its relevance to activity is not known, although mutational studies indicate an involvement in specificity with acceptor molecules [120].

Even less is known of sucrose phosphorylase, which might have been expected, from the reaction catalysed, to resemble a phosphorylase more than an α-amylase, yet the sequence bears the hallmarks of an α-amylase family enzyme.

6. Relevance to protein engineering

The large diversity of specificity and the different types of reaction catalysed by enzymes in glycoside hydrolase families 13, 70, and 77 – or clan GH-H – invite rational engineering of the enzyme specificity. Early mutational analyses investigated structure/function relationships [37] and protein engineering moreover addressed important industrial goals such as improvement of thermostability or changing the pH activity dependence [121–123]. Similarly, modification of the product specificity has major industrial relevance. One issue in the case of cyclodextrin glucanotransferases has been to engineer the product composition to have higher contents of a given cyclodextrin and transglycosylation reactions [74,98,125,126]. At an early point an α-amylase was altered to enhance transglycosylation, probably by slow product release which allowed attack of a new substrate molecule [127,128]. Random mutation at a localised substrate binding region of the polypeptide chain, e.g., in β→α loops 4 [129] and 7 [130,131] in barley α-amylase shifted the position of the preferred productive binding mode of the oligosaccharide substrates. In addition mutant enzymes were obtained with altered substrate specificity, including increased specific activity towards insoluble starch [130,131]. Several examples are reported on the usefulness of chimaeic enzymes from this family, including chimera of related Bacillus α-amylases [132,133], barley α-amylase isoamylases [134,135] and starch branching enzyme isoamylases from maize endosperm [90,91]. These chimaeic enzymes all have modulated specificity – typically intermediate between the two or similar to one of the parent enzymes – and their properties enable tentative correlation of structural features with specificity variation. Cyclodextrin glucanotransferase specificity was rationally engineered in an α-amylase of B. stearothermophilus which was closely related to a CGTase in sequence from the same organism [136]. Recently the mutation of a glutamine – normally found as a transition state stabilising histidine at subsite −1 in family 13 – to histidine in glucansucrase from Streptococcus downei (a circularly permuted family 70 enzyme) altered the acceptor reaction to produce new oligosaccharides [120]. Interestingly the reverse mutation of histidine to glutamine in the related amylosucrase from Neisseria polysaccharea – which is a non-circularly permuted family 13 member – was the superior of a series of replacements and gave an enzyme with enhanced sensitivity to glycogen activation [137].

7. Conclusions and recommendations

Enzymes of the α-amylase family can bring about scission and synthesis of α-1,4-, α-1,6- and less commonly α-1,2- and α-1,3-glucosidic linkages, as well as act on sucrose and trehalose. The enzyme active sites may be considered to be composed of subsites, each capable of interacting with one monosaccharide residue. All enzymes of the family require an α-linked
glucose residue in the substrate to interact with the glycone-binding subsite adjacent to the catalytic acids. Enzymes of different specificities, however, act on substrates containing different aglycones; hence the aglycone-binding subsites and the amino acid residues that constitute them are characteristic of particular specificities. The table relating specificity to short sequences that are likely to contain such amino acids has been compiled using some 300 representative enzyme sequences. New primary structures may be obtained that do not conform to any of the signatures given here. In such a case, we would recommend detailed examination of the specificity of the protein involved. Indeed it is possible that enzymes may be found, in future, with a specificity that combines the specificities of more than one known member of family 13, or that is intermediate between the specificities of known enzymes.

Although short sequences that are characteristic of certain specificities have been defined here, in many cases it is not possible to offer a rational explanation for these sequence-specificity relationships. More work is required on three-dimensional structures, particularly of enzyme–carbohydrate complexes, before the relationships will become clear.

It is unfortunate that some enzymes, clearly demonstrated to be active on both $\alpha$-1,4- and $\alpha$-1,6-glucosidic linkages are described in the literature as $\alpha$-amylases, which by Enzyme Commission definition, catalyse hydrolysis of $\alpha$-1,4-bonds only. We recommend that the name $\alpha$-amylase be restricted to those enzymes that fit the definition.

Where a single name, e.g., maltogenic amylase is used for what are demonstrably two different types of enzymes, we recommend that the name be used only for the type of enzyme that can be considered as a specialised $\alpha$-amylase and is inactive on $\alpha$-1,6-glucosidic linkages. Where enzymes are shown to act on both $\alpha$-1,4- and $\alpha$-1,6-glucosidic linkages and produce maltose from starch, it is likely they can also hydrolyse cyclohextrins and pullulans [32]. In this case, the preferred substrate should be identified – this is more likely to be cyclodextrin or pullulan rather than starch [32] – and the enzyme named accordingly, i.e., either cyclodextrinase or neopullulanase.

For the enzymes that catalyse hydrolysis of $\alpha$-1,4-linkages in amylose and $\alpha$-1,6-linkages in pullulan, we recommend the proteins with a single active site be called amylopullulanases, while the enzymes with two active sites, each active site being specific for one bond type, could be called $\alpha$-amylase-pullulanases.

**Acknowledgements**

M.T. Jensen and J.G. Olsen are thanked for help with Figs. 1, 3 and 6. E.A. MacG. wishes to thank the University of Manitoba for the provision of office space and computer facilities. S.J. thanks the Slovak Grant Agency for Science (VEGA Grant No. 2/6045/99), FEBS for a short-term fellowship, and the Slovak literary Fund for financial support. B.S. was supported by the EU 4th Framework Programme (BIO4-CT98-0022).

**References**


[19] E.A. MacGregor, B. Svensson, A circularly permuted α-amylase-type α/β-barrel structure in glucan-syn-


[61] Tables of enzymes studied are given at URL: http://nic.svba.sk/~unikstef/bba


K. Binderup, J. Preiss, Glutamate-459 is important for factors determining α-amylase and cyclodextrin glycosyltransferase specificity in the cyclodextrin glycosyltransferase from Thermomanaerobacterium thermosulfurigenes EM1, Eur.

E.A. MacGregor et al. / Biochimica et Biophysica Acta 1546 (2001) 1–20
quence and molecular structure of an alkaline amylopol-
uluranase from *Bacillus* that hydrolyses α-1,4 and α-1,6 link-
ages in polysaccharides at different active sites, J. Biol.
[95] K. Ara, K. Igarashi, K. Saecki, S. Ito, An alkaline amylopo-
ullanase from alkalophilic *Bacillus* sp. KSM 1378; kinetic
evidence for two independent active sites for the α-1,4 and
α-1,6 hydrolytic reactions, Biosci. Biotechnol. Biochem. 59
[96] K. Ara, K. Igarashi, H. Hagihara, K. Sawada, T. Kobaya-
shi, S. Ito, Separation of functional domains for the α-1,4 and
α-1,6 hydrolytic activities of a *Bacillus* amylullanu-
lnase by limited proteolysis with papain, Biosci. Biotechnol.
[97] C.H. Kim, Specific detection of pullulanase type I in poly-
332.
the active center of *Bacillus steaarrowthophilus* neopullu-
[99] A. Ibuka, T. Tonozuka, H. Matsuzawa, H. Sakai, Conver-
sion of neopullulanase-α-amylase from *Thermoactinomyces
vulgaris* R-47 into an amylullanulase-type enzyme, J. Bio-
Lee, T.W. Moon, J.W. Kim, B.H. Oh, K.H. Park, Role of the
glutamate 332 residue in the transglycosylation activity of
6773–6780.
[101] S. Kitahata, S. Okoda, Hydrolytic action on various malto-
dextrins at different active sites, J. Biol. Chem. 271 (1996)
24075–24083.
[102] S. Mathupala, B.C. Saha, J.G. Zeikus, Substrate competi-
tion and specificity at the active site of amylullanululanase
from *Clostridium thermohydrosulphuricum*, Biochim.
[103] R. Peist, C. Schneider-Frenenius, W. Boos, The Mal-T de-
pendent and malZ-encoded maltodextrin glucosidase of *Es-
cherichia coli* can be converted into a dextrinyl transferase
by a single mutation, J. Biol. Chem. 271 (1996) 10681–
10689.
[104] M. Yamamoto, K. Horikoshi, Nucleotide sequence of al-
kalophilic *Bacillus* oligo-1,6-glucosidase gene and the prop-
erties of the gene product by *Escherichia coli* HB101, Den-
pun Kagaku 37 (1990) 137–144.
Ikemoto, S. Kobayashi, S. Shibano, Purification and char-
acterization of a *Bacillus* sp. SAM 1606 thermo-stable α-
glucosidase with transglycosylation activity, Appl. Micro-
[106] M. Nakao, T. Nakayama, A. Kakudo, M. Inohara, M.
Harada, F. Omura, Y. Shibano, Structure and expression of
a gene coding for a thermostable α-glucosidase with a broad
substrate specificity from *Bacillus* sp. SAM 1606,
Ueda, Y. Shibano, Altering substrate specificity of *Bacillus
sp.* SAM 1606 α-glucosidase by comparative site-specific
and characterization of extremely thermostable exo-oligo-
1,6-glucosidase from a caldoactive *Bacillus* sp. KP1228,
[109] S. Kashiwabara, Y. Matsuki, T. Kishimoto, Y. Suzuki,
Clustered proline residues around the active-site cleft in
thermostable oligo-1,6-glucosidase of *Bacillus flavocal-
1093–1102.
[110] O. Nashiru, S.Y. Lee, D.S. Lee, GenBank data base, acces-
sion no. AF096282 (1998).
[111] Y. Takii, K. Takahashi, K. Yamamoto, Y. Sogabe, Y. Su-
zuki, *Bacillus steaarrowthophilus* ATCC 12016 α-glucosidase
specific for α-1,4 bonds of malto-oligosaccharides and α-glucans
shows high amino acid sequence similarities to seven α-1,4-
glucosidases with different substrate specificity, Appl.
Mooser, Knowledge-based model of a glucosyl-transferase
from the oral bacterial group of mutants streptococci, Pro-
[113] V. Monchois, M. Vignon, P.-C. Escalier, B. Svensson,
R.R.B. Russell, Involvement of Gin937 of *Streptococcus


[129] I. Matsui, B. Svensson, Improved activity and modulated action pattern obtained by random mutagenesis at the fourth β-α loop involved in substrate binding to the catalytic β(βα)-barrel domain of barley α-amylase 1, J. Biol. Chem. 272 (1997) 22456–22463.


