Two potentially novel amylolytic enzyme specificities in the prokaryotic glycoside hydrolase α-amylase family GH57

Karol Blesák and Štefan Janeček

Laboratory of Protein Evolution, Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84551 Bratislava, Slovakia

Glycoside hydrolase (GH) family 57 consists of more than 900 proteins from Archaea (roughly one-quarter) and Bacteria (roughly three-quarters), mostly from thermophiles. Fewer than 20 GH57 members have already been biochemically characterized as real, (almost exclusively) amylolytic enzymes. In addition to a recently described dual-specificity amylolpullulanase–cyclomaltodextrinase, five enzyme specificities have been well established in the family – α-amylase, amylopullulanase, branching enzyme, 4-α-glucanotransferase and 2-α-galactosidase – plus a group of the so-called α-amylase-like homologues probably without the enzyme activity. A (β/α)8-barrel succeeded by a bundle of a few α-helices forming the catalytic domain, and five conserved sequence regions (CSRs), are the main characteristics of family GH57. The main goal of the present bioinformatics study was to describe two novel groups within family GH57 that represent potential non-specified amylases (127 sequences mostly from Bacteria) and maltogenic amylases (12 sequences from Archaea). These were collected from sequence databases based on an indication of their biochemical characterization. Although both the non-specified amylases and the maltogenic amylases share the in silico identified catalytic machinery and predicted fold with the experimentally determined GH57 members, the two novel groups may define new GH57 subfamilies. They are distinguishable from the other, previously recognized, subfamilies by specific sequence features present especially in their CSRs (the so-called sequence fingerprints), also reflecting their own evolutionary histories.

INTRODUCTION

Amylolytic enzymes are of great importance since starch, their main substrate, belongs to one of the most widespread biopolymer sources of energy on Earth. In addition to plants and animals (Pujadas et al., 1996; Coutinho & Reilly, 1997; Janeček, 1997; Da Lage et al., 2007; Smith et al., 2005; Ball et al., 2011), fungi and prokaryotic microorganisms are known to produce various amylases and related starch hydrolases which may differ from each other in their exact substrate preferences and product specificities (Janeček et al., 2011).

α-Amylase (EC 3.2.1.1) is one of the best known and studied amylolytic enzymes. It catalyses the hydrolysis of the internal α-1,4-glicosidic bonds in starch and related poly- and oligosaccharides (Janeček et al., 2013). Based on their amino acid sequence similarities, α-amylases have been classified into three different glycoside hydrolase (GH) families in the Carbohydrate-Active enZyme (CAZy) server (Cantarel et al., 2009): (i) family GH13 – the main α-amylase family, established more than 20 years ago (Henrissat, 1991; Jespersen et al., 1991; Takata et al., 1992), with more than 14 000 sequences and ~30 different enzyme specificities from Bacteria, Archaea and Eukarya, forming together with families GH70 and GH77 the α-glucan-active clan GH-H (MacGregor et al., 2001); (ii) family GH57 – the second and smaller α-amylase family, established in 1996 (Henrissat & Bairoch, 1996), containing a few amylolytic specificities from prokaryotes only, often from extremophiles (Janeček & Blesášk, 2011; Blesášk & Janeček, 2012); and (iii) family GH119 – created after the study by Watanabe et al. (2006) describing the protein product of the igtZ gene as an α-amylase, and recently indicated to be very closely related to family GH57 (Janeček & Kuchtová, 2012). A recently published structure of the amylase from Clostridium perfringens indicated the presence of α-amylase specificity even in family GH126 (Ficko-Blean et al., 2011), but owing to its obvious homology with inverting β-glucan-active hydrolases more biochemical characterization is required to confirm the exact enzyme specificity (Janeček et al., 2013).
Originally family GH57 consisted of only two supposed ‘z’-amyloses’ (Henrissat & Bairoch, 1996) whose amino acid sequences were obviously not easily alignable with those of typical GH13 members (Janecˇek, 1998). The two proteins were the enzyme from Dictyoglomus thermophilum (Fukusumi et al., 1988) and the one from Pyrococcus furiosus (Laderman et al., 1993a). Whereas it was more than 15 years before the former ‘amylose’ was re-evaluated as a glucanotransferase (Nakajima et al., 2004), the latter enzyme was immediately found to exhibit the transference activity at the time of its discovery (Laderman et al., 1993b). Now, both these enzymes are known as 4-‘z’-glucanotransferases (EC 2.4.1.25) (Cantarel et al., 2009), i.e. one of the five well-established enzyme specificities in family GH57, the remaining four specificities being the z-amyrase, amylpullulanase (EC 3.2.1.41), branching enzyme (EC 2.4.1.18) and z-galactosidase (EC 3.2.1.22) (Janecˇek & Blesa´ k, 2011; Blesa´k & Janecˇek, 2012).

For the 4-‘z’-glucanotransferase from Thermococcus litoralis (Imamura et al., 2003) and branching enzymes from Thermus thermophilus (Palomo et al., 2011) and Thermococcus kodakaraensis (Santos et al., 2011) the three-dimensional structures were also determined. The studies revealed a (β/α)7-barrel, i.e. an incomplete TIM-barrel fold as a catalytic domain, succeeded by a bundle of three or four α-helices with two catalytic residues, Glu123 as the catalytic nucleophile and Asp214 as the proton donor (Thermococcus litoralis 4-‘z’-glucanotransferase numbering; Imamura et al., 2003). For family GH57, five conserved sequence regions (CSRs) were defined by Zona et al. (2004). The fact that the fifth CSR with functionally essential residues (Palomo et al., 2011) is positioned in the z-helical bundle succeeding the incomplete TIM-barrel domain indicated that the GH57 catalytic site is formed by both the barrel and the helical bundle (Blesa´k & Janecˇek, 2012). Interestingly, the family may also contain members with mixed enzyme specificity. One of these might be the AmyC from Thermotoga maritima, for which the three-dimensional structure has also been solved (Dickmanns et al., 2006). Although it was described as an z-amyrase (Ballschmiter et al., 2006), its sequence-structural features clearly suggest branching enzyme activity (Blesa´k & Janecˇek, 2012). Another example is a recently described amylpullulanase from Staphylothermus marinus that also exhibits cyclodextrinase activity (Li et al., 2013). It is worth mentioning that, of the current more than 900 members of family GH57, fewer than 20 proteins have thus far been biochemically characterized (Cantarel et al., 2009). Furthermore, for example, there are more than 100 GH57 members exhibiting clear sequence features of the z-amyrase from Methanocaldococcus jannaschii (Kim et al., 2001), but about half of them are proteins that lack one or both catalytic residues (Janecˇek & Blesa´k, 2011). Any progress concerning family GH57 enzyme specificities is thus highly desired.

In an effort to contribute further to the knowledge of how many and which enzyme specificities actually belong to the z-amyrase family GH57, the present bioinformatics study was undertaken with a focus on two partially characterized GH57 members: a non-specified amylose from an uncultured bacterium (Wang et al., 2011) and a potential maltogenic amylose from P. furiosus (PF0870; Comfort et al., 2008). To this end, amino acid sequences of 139 GH57 proteins were collected from sequence databases. Of these sequences, 127, mostly from Bacteria (115 sequences), have been described as belonging to the new group of non-specified amylose, whereas 12 originating solely from Archaea have been assigned the specificity of maltogenic amylose. Interestingly, within CSR-1 both these novel GH57 specificities contain a tyrosine residue invariably conserved just in the position where a GH57 functionally important histidine dominates. Mutual evolutionary relationships of the two novel GH57 groups as well as their relatedness to the remaining five well-established GH57 are also presented. The results of this study may be helpful in experimental approaches aimed at elucidating the structure/function relationships and rational protein design of all GH57 amyloytic enzymes.

### METHODS

**Sequence collection.** Sequences were collected based on a BLAST (Altschul et al., 1990) search using the complete sequences of the non-specified amylose from an uncultured bacterium (UniProt accession. no. F4ZW71; Wang et al., 2011) and the potential maltogenic amylose from P. furiosus (UniProt accession no. Q8U2G5; Comfort et al., 2008). The BLAST results were complemented by the known GH57 sequences from the CAZy database (Cantarel et al., 2009) that contained a tyrosine residue instead of the otherwise conserved histidine in CSR-1 (Palomo et al., 2011; Blesa´k & Janecˇek, 2012). All the selected potential sequences had to possess the GH57 characteristic signatures, such as both catalytic residues, five CSRs and an obvious (β/α)7-barrel domain. Thus 139 GH57 sequences were obtained, which were clearly divided into 127 prokaryotic non-specified new amyloses, the majority (115) originating from Bacteria, and 12 archaeal maltogenic amyloses (Table 1). For as complete a comparison as possible, all the biochemically characterized GH57 enzymes were included too (Table S1, available in Microbiology Online).

**Sequence and evolutionary comparison.** All sequences (Table S1) were retrieved from the UniProt knowledge database (UniProt Consortium, 2012) and/or GenBank (Benson et al., 2012). Domain arrangements of the non-specified amylose from the uncultured bacterium and of P. furiosus maltogenic amylose as the representatives

**Table 1. Summary of proteins from both GH57 subfamilies used in the present study**

A total of 139 proteins were obtained after several rounds of analysis from more than 2000 BLAST results. The data were retrieved from the BLAST searches using the sequences of 17 biochemically characterized GH57 enzymes as queries (Blesa´k & Janecˇek, 2012). The set was completed using the CAZy server (Cantarel et al., 2009).

<table>
<thead>
<tr>
<th>Enzyme Number</th>
<th>Enzyme Description</th>
<th>Archaea</th>
<th>Bacteria</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-specified amyloses</td>
<td>127</td>
<td>12</td>
<td>115</td>
</tr>
<tr>
<td>Maltogenic amyloses</td>
<td>12</td>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>139</td>
<td>24</td>
<td>115</td>
</tr>
</tbody>
</table>
of the two novel GH57 groups (enzyme specificities) were completed utilizing the available structural information (Imamura et al., 2003; Dickmanns et al., 2006; Palomo et al., 2011; Santos et al., 2011), previous sequence alignment of biochemically characterized GH57 members (Blesák & Janeček, 2012) and predictions of both secondary and tertiary structures obtained from the Phyre-2 server (http://www.ebi.ac.uk/Tools/msa/clustalw2/) (Larkin et al., 2007). In order to maximize similarities, the alignment was manually tuned taking into account the previous bioinformatics studies (Zona et al., 2004; Janeček & Blesák, 2011; Blesák & Janeček, 2012).

For the CSRs of both novel family GH57 groups, i.e. the non-specified amylase (127 sequences) and the maltogenic amylase (12 sequences), a sequence logo was created using the WebLogo 3.0 server (http://weblogo.berkeley.edu/) (Crooks et al., 2004).

Two evolutionary trees were prepared: (i) one for all 139 sequences of non-specified amylases and maltogenic amylases based on the alignment of their catalytic (β/α)-barrel and the succeeding α-helical bundle (including the gaps); and (ii) one based on the alignment of the GH57 CSRs only, for the above-mentioned 139 sequences, 15 selected characterized GH57 enzymes and one α-amylase-like protein from Bacteroides thetaiotaomicron (Table S1). Both trees were calculated as a PHYLIP-tree type using the neighbour-joining clustering (Saitou & Nei, 1987) and the bootstrapping procedure (Felsenstein, 1985) (the number of bootstrap trials used was 1000) implemented in the CLUSTAL_X package (Larkin et al., 2007). The trees were displayed with the program TreeView (Page, 1996).

Tertiary structure comparison. Three-dimensional structure was retrieved from the Protein Data Bank (PDB) (Deshpande et al., 2005) for the family GH57 representative, i.e. Thermococcus litoralis 4-α-glucanotransferase (PDB code 1KY; Imamura et al., 2003). A three-dimensional model for the non-specified amylase from an uncultured bacterium (UniProt accession no. F4ZW71; Wang et al., 2011) was created with the Phyre-2 server (Kelley & Sternberg, 2009) using the experimentally determined branching enzyme structure from Thermococcus kodakarenensis (PDB code 3N92; Santos et al., 2011) and the model of the maltogenic amylase from P. furiosus (UniProt accession no. Q0U2G5; Comfort et al., 2008) was then obtained from the SwissModel server (Arnold et al., 2006) using the modelled structure of the non-specified amylase from an uncultured bacterium as template. The structural models were superimposed using the program MULTIPROJ (Shatsky et al., 2004) and the structures were visualized with the program WebLabViewerLite (Molecular Simulations). Topology diagrams of the modelled structures of both the non-specified amylase from an uncultured bacterium and the maltogenic amylase from P. furiosus were generated by the program PDBsum (Laskowski et al., 1997) and their Ramachandran plots including the scores related to model accuracy were produced by the PROCHECK program (Laskowski et al., 1993).

RESULTS AND DISCUSSION

Sequence comparison

The present bioinformatics study delivers a comparison of 139 amino acid sequences of two novel, still potential enzyme specificities of family GH57 (Table S1). The large group (127 sequences) covers a currently non-specified amylase (115 bacterial and 12 archaeal members), whereas the smaller group (12 sequences, solely from Archaea) may represent the specificity of a maltogenic amylase. It is worth mentioning that assigning the specificities is based on a partial biochemical characterization only. Thus in the former group of non-specified amylases, the novel amylase from an uncultured bacterium was shown to hydrolyse soluble starch (Wang et al., 2011), but the exact enzyme specificity and the reaction product composition still need to be determined. As far as the group of maltogenic amylases is concerned, the GH57 protein encoded by PF0870 from the P. furiosus genome was found to produce maltose from short α-glucans, an observation that led authors Comfort et al. (2008) to name the enzyme β-amylase. Since β-amylase, from its definition, is an inverting enzyme, the protein product of the PF0870 gene should rather be a maltogenic amylase because the members of family GH57 employ the retaining mechanism (Palomo et al., 2011). This issue, however, can unambiguously be solved only if the anomeric configuration of the maltose liberated from α-glucans by P. furiosus PF0870 is determined.

The non-specified amylase from the uncultured bacterium (Table S1, protein no. 1, AMY_Uncba) and the probable maltogenic amylase from P. furiosus (Table S1, no. 128, MGA_Pycfu) are therefore representatives of two novel GH57 groups (subfamilies). The length of maltogenic amylases is, in general, less than 600 residues, ranging from 575 to 599, whereas that of non-specified amylases varies more substantially, from 460 (no. 97) to 1085 (no. 105), the most frequent length being around 800–900 residues. Larger variability in length among the non-specified amylases very probably reflects a larger sample of sequences (127 in comparison with 12) spanning also a wider spectrum of prokaryotic taxa (Table S1). The extra-long putative non-specified amylase from Isosphaera pallida (no. 75; UniProt accession no. E8R431) with 1842 residues may also contain an amylolytic enzyme from the main α-amylase family GH13 succeeding the N-terminally positioned family GH57 catalytic domain (Punta et al., 2012). However, as for the other GH57 enzyme specificities (Janeček & Blesák, 2011; Blesák & Janeček, 2012), the catalytic domains for both non-specific amylases and maltogenic amylases seem to be formed by the N-terminal ~500 residues (Fig. 1), covering the catalytic (β/α)-barrel (containing the first four CSRs) and the essential bundle of a few α-helices (containing CSR-5) (Imamura et al., 2003; Dickmanns et al., 2006; Palomo et al., 2011; Santos et al., 2011). Based on the alignment with experimentally characterized GH57 members (Fig. 1), the candidate residues for catalytic nucleophile and proton donor were Glu153 and Asp253 in MGA_Pycfu, which are invariably conserved in all 127 non-specified amylases and 12 maltogenic amylases, respectively (Fig. S1). Both new GH57 groups are very similar in substantial parts of their sequences, especially with regard to the presence of secondary structure elements (α-helices and β-strands),
although some differences, reflecting most probably differences in their specificity, can also be found among them (Fig. S1).

In order to identify unique sequence features characteristic of each of the two novel GH57 enzyme groups, sequence logos covering the five GH57 CSRs were created (Fig. 2). One of the most exclusive residues is the invariant tyrosine in position 3 (CSR-1). Although it is present in both specificities, i.e. this tyrosine cannot be used for discriminating the two specificities from each other, currently no other GH57 specificity contains a tyrosine in the corresponding position (Janeček & Blesák, 2012). With the exception of α-galactosidase (with an asparagine), all remaining known amylolytic enzyme specificities from family GH57 have in that position an invariant histidine (Blesák & Janeček, 2012), which is present even in the α-amylase-like proteins that lack one or both catalytic residues (Janeček & Blesák, 2011).

Despite sharing the above-mentioned tyrosine in position 3, each of the two specificities exhibits its own characteristic sequence features. Thus, in the logo of the non-specified amylase (Fig. 2a), positions 26 and 27 occupied by two adjacent histidines (end of CSR-4) and aromatic positions 34–36 with an invariant tryptophan followed by two phenylalanines (end of CSR-5) are unique for this specificity. Although branching enzymes possess an invariant histidine in position 26 (Blesák & Janeček, 2012) and α-amylase contains one in position 27 (Janeček & Blesák, 2011), the presence of two invariant adjacent histidines in both positions is unique for non-specified amylases (Fig. 2a) represented by the amylase from an uncultured bacterium (Wang et al., 2011). Similarly unique is the aromatic end of CSR-5 (tripetide WFF observed invariantly in 127 sequences; Fig. S1), which again is present only in the non-specified amylase (Fig. 2a). The end of the logo (i.e. the end of CSR-5) is also unique to the maltogenic amylase since the invariant sequence RWF (Fig. 2b) is not present in any of the five well-established GH57 specificities (Blesák & Janeček, 2012). Position 21 (CSR-4) can furthermore discriminate the maltogenic amylase from a non-specified amylase because the former enzyme has in that position an invariant leucine, whereas the latter possesses an invariant glycine (Fig. 2).

**Structure comparison**

To support the observations from sequence comparison, three-dimensional structures of representatives of the two novel GH57 specificities were modelled and compared with each other as well as with the experimentally determined structure of 4-α-glucanotransferase from Thermococcus litoralis (Fig. 3a, b). Both the non-specified amylase and maltogenic amylase adopt the structure of a typical family GH57 catalytic (β/α)7-barrel with succeeding α-helical bundle as known from solved crystal structures (Imamura et al., 2003; Dickmanns et al., 2006; Palomo et al., 2011; Santos et al., 2011), although the latter enzyme seems to contain more additional β-strand segments (Fig. 3a). The best templates for modelling the structure of the non-specified amylase from an uncultured bacterium (Wang et al., 2011) using the Phyre-2 server (Kelley & Sternberg, 2009) were the structures of branching enzymes, of which the one from Thermococcus kodakaraensis (PDB code 3N92; Santos et al., 2011) was chosen for the model. With regard to modelling the P. furiosus maltogenic amylase structure (Comfort et al., 2008), efforts to obtain the model of both segments constituting the catalytic part of a family GH57 member (i.e. the incomplete TIM-barrel and the α-helical bundle) using the Phyre-2 server failed, but this model was successfully constructed by the SwissModel server (Arnold et al., 2006) with the non-specified amylase modelled structure as template. The main aim, why the three-dimensional structures of both the representative non-specified amylase and maltogenic amylase were modelled, was to get support for identification of all five GH57 CSRs based on sequence comparison (Fig. 2). It was thus more important than to have both models as complete as possible to have them more accurate (but rather incomplete). It is therefore worth mentioning that topology diagrams of both the non-specified amylase (Fig. 3c) and maltogenic amylase (Fig. 3d) may not reflect absolutely the real situation in their tertiary structures; however, the trend of alternating β-strands and α-helices of predicted N-terminal catalytic (β/α)7-barrel as well as α-helical segments at the C-terminal end is obviously seen.

For the model of the non-specified amylase from an uncultured bacterium, of the total 391 number of residues, there were 342 non-glycine and non-proline residues. Of these, 278, i.e. 81.3 %, were found in the most favoured regions of the Ramachandran plot (Fig. S2a). However, if the additional allowed regions were also included, a further 51 residues (14.9 %) were added, thus representing more than 96.2 % of the model. Concerning the 430 non-glycine and non-proline residues of the model of the maltogenic amylase from P. furiosus (there were 476 residues in total), 348 residues were found in the most favoured regions of the Ramachandran plot (Fig. S2b) and 409 residues were positioned in both the most favoured and additional allowed regions, i.e. 80.9 % and 95.1 %, respectively. Although a good-quality model should have over 90 % in the most favoured regions (Laskowski et al., 1993), for the aim of this part of the present study, i.e. getting support for identification of all five GH57 CSRs based on sequence comparison, the quality of both models was considered satisfactory.

Based on structure superposition of both modelled novel GH57 enzymes with the 4-α-glucanotransferase from Thermococcus litoralis (PDB code 1K1Y; Imamura et al., 2003) it is clear that the catalytic machinery consisting of a glutamic acid in strand β3 (catalytic nucleophile; in CSR-3) and an aspartic acid in strand β4 (proton donor; in CSR-4) is perfectly conserved in both the non-specified amylase and the maltogenic amylase (Fig. 3b). The two pairs of
proposed catalytic residues, i.e. Glu154 and Asp263 in AMY_Uncba, non-specified amylase from uncultured bacterium (Wang et al., 2011); MGA_Pycfu, maltogenic amylase from *P. furiosus* (Comfort et al., 2008); AAMY_Mccja, α-amylase from *M. jannaschii* (Kim et al., 2001); PAMY_Bctth, putative α-amylase-like protein from *B. thetaiotaomicron* (Janeček & Blesák, 2011); APU_Pycfu, APU_Thchy, APU_Thcli and APU_THcsi, amylolpullulanases from *P. furiosus* (Dong et al., 1997), *Thermococcus hydrothermalis* (Erra-Pujada et al., 1999), *Thermococcus litoralis* (Imamura et al., 2004) and *Thermococcus siculi* (Jiao et al., 2011), respectively; APCD_Sttm, amylolpullulanase–cycloamitetodextrinase from *S. marinus* (Li et al., 2013); BE_Thck, BE_Thtm and BE_Theth, branching enzymes from *Thermococcus kodakaraensis* (Murakami et al., 2006), *Thermotoga maritima* (Ballschmiter et al., 2006) and *Thermus thermophilus* (Palomo et al., 2011), respectively; 4AGT_Arcfu, 4AGT_Dicth, 4AGT_Pycfu, 4AGT_Thck, 4AGT_Thcli, 4-α-glucanotransferases from *Archaeoglobus fulgidus* (Labes & Schönheit, 2007), *D. thermophilum* (Fukusumi et al., 1988), *P. furiosus* (Laderman et al., 1993a), *Thermococcus kodakaraensis* (Tachibana et al., 1997) and *Thermococcus litoralis* (Jeon et al., 1997), respectively; AGAL_Pycfu, α-galactosidase from *P. furiosus* (van Lieshout et al., 2003). CSRs are emphasized by rectangles and catalytic residues (CSR-3 glutamate, catalytic nucleophile; CSR-4 aspartate, proton donor) are indicated by asterisks. CSR-1–4 are located in the (β/α)7-barrel domain (yellow lane above the alignment), whereas CSR-5 is positioned in the α-helical part (cyan lane above the alignment) of the GH57 catalytic region. The complete alignment of all 139 sequences of non-specified amylases and potential maltogenic amylases is shown in Fig. S1.

Two novel family GH57 amylolytic enzyme specificities

Fig. 1. Sequence alignment of catalytic domains of selected representatives of family GH57. α-Helices and β-strands (predicted by the Phyre-2 server; Kelley & Sternberg, 2009) are coloured in red and green, respectively. Abbreviations: AMY_Uncba, non-specified amylase from uncultured bacterium (Wang et al., 2011); MGA_Pycfu, maltogenic amylase from *P. furiosus* (Comfort et al., 2008); AAMY_Mccja, α-amylase from *M. jannaschii* (Kim et al., 2001); PAMY_Bctth, putative α-amylase-like protein from *B. thetaiotaomicron* (Janeček & Blesák, 2011); APU_Pycfu, APU_Thcly, APU_Thcli and APU_Thcsi, amylolpullulanases from *P. furiosus* (Dong et al., 1997), *Thermococcus hydrothermalis* (Erra-Pujada et al., 1999), *Thermococcus litoralis* (Imamura et al., 2004) and *Thermococcus siculi* (Jiao et al., 2011), respectively; APCD_Sttm, amylolpullulanase–cycloamitetodextrinase from *S. marinus* (Li et al., 2013); BE_Thck, BE_Thtm and BE_Theth, branching enzymes from *Thermococcus kodakaraensis* (Murakami et al., 2006), *Thermotoga maritima* (Ballschmiter et al., 2006) and *Thermus thermophilus* (Palomo et al., 2011), respectively; 4AGT_Arcfu, 4AGT_Dicth, 4AGT_Pycfu, 4AGT_Thck, 4AGT_Thcli, 4-α-glucanotransferases from *Archaeoglobus fulgidus* (Labes & Schönheit, 2007), *D. thermophilum* (Fukusumi et al., 1988), *P. furiosus* (Laderman et al., 1993a), *Thermococcus kodakaraensis* (Tachibana et al., 1997) and *Thermococcus litoralis* (Jeon et al., 1997), respectively; AGAL_Pycfu, α-galactosidase from *P. furiosus* (van Lieshout et al., 2003). CSRs are emphasized by rectangles and catalytic residues (CSR-3 glutamate, catalytic nucleophile; CSR-4 aspartate, proton donor) are indicated by asterisks. CSR-1–4 are located in the (β/α)7-barrel domain (yellow lane above the alignment), whereas CSR-5 is positioned in the α-helical part (cyan lane above the alignment) of the GH57 catalytic region. The complete alignment of all 139 sequences of non-specified amylases and potential maltogenic amylases is shown in Fig. S1.

Fig. 2. Sequence logos of family GH57 non-specified amylases (a) and maltogenic amylases (b). CSR-1, residues 1–5; CSR-2, residues 6–11; CSR-3, residues 12–17; CSR-4, residues 18–27; CSR-5, residues 28–36. The catalytic nucleophile (no. 15, glutamate) and the proton donor (no. 20, aspartate) in both specificities are indicated by asterisks. The logos are based on sequences of 127 non-specified amylases and 12 maltogenic amylases.
Evolutionary relationships

The evolutionary relationships between the non-specified amylases and maltogenic amylases as well as their relatedness to the remaining well-established specificities of family GH57 are illustrated in Fig. 4 (and in Fig. S3). It is clear that both novel GH57 specificities (or groups) are well separated from each other (Fig. 4a), an observation that supports the reliability of the division of the 139 collected sequences (Table S1) into two groups. The 129 sequences of potential non-specified amylases are grouped into several larger bacterial clusters, although three (Table S1, nos 116–118) of the 12 archaeal sequences have been placed among the Bacteria (Fig. S3a). Note that all three ‘exceptions’ are longer non-specified amylases with ~800...
residues as compared with the less than 500-residue-long remaining nine archaeal representatives (nos 119–127). As far as the 12 maltogenic amylases are concerned, in the trees based on the alignment of catalytic \((b/\alpha)7\)-barrels with a succeeding \(\alpha\)-helical bundle (Fig. S3) they were divided into two groups, whereas in the trees based on the alignment of CSRs only (36 residues; Fig. 4) they form three distinct clusters regardless of whether other GH57 specificities are also taken into account. While in the former trees (Fig. S3) the putative maltogenic amylases from \textit{Thermococcus barophilus} (Table S1, no. 133) and \textit{Thermococcus sibiricus} (no. 137) are positioned among all five maltogenic amylases from the genus \textit{Pyrococcus}, in the latter trees (Fig. 4) the two \textit{Thermococcus} sequences define their own cluster.

What may, however, be of special interest is the observation that both non-specified amylases and maltogenic amylases are most closely related to each other when other GH57 specificities are also included (Figs 4b and S3b). Their own clusters are on adjacent branches, similar to relatedness observed previously for GH57 \(\alpha\)-amyloses and their \(\alpha\)-amylase-like homologues (Janeček & Blesáš, 2011). It is worth mentioning that the tree shown in Fig. 4b represents currently the most complete evolutionary picture of the GH57 family, since the recently described dual-specificity amylopullulanase–cyclomaltodextrinase from \textit{S. marinus} (Li et al., 2013), closely related to typical GH57 amylopullulanases, was included in the comparison.

**Conclusions**

A partially characterized open reading frame from the \textit{P. furiosus} genome (PF0870) exhibiting the ability to release maltose from \(\alpha\)-glucans (Comfort et al., 2008) and an amylolytic enzyme from an uncultured bacterium able to hydrolyse soluble starch (Wang et al., 2011) served as a starting point in collecting two groups of homologous sequences that may define novel enzyme specificities (i.e. also subfamilies) in family GH57: maltogenic amylases and non-specified amylases. Whereas only 12 sequences originating solely from \textit{Archaea} were obtained for maltogenic amylases, 127 sequences were assigned the specificity of

---

**Fig. 4.** Evolutionary trees of family GH57 enzymes. (a) The tree of GH57 non-specified amylases and maltogenic amylases. Archaeal and bacterial non-specified amylases are coloured cyan and blue, respectively, whereas all maltogenic amylases – originating only from \textit{Archaea} – are in red. (b) The tree of family GH57 characterized representatives with all 139 GH57 non-specified amylases and maltogenic amylases. Both trees are based on the alignment of five CSRs (36 residues). For the sake of simplicity, only the branches are shown. All the enzymes depicted in the trees are based on sequence comparison of catalytic \((b/\alpha)7\)-barrels with succeeding \(\alpha\)-helical bundle presented in Fig. S3. Details of all enzymes and proteins compared in both trees as well as their colour codes are given in Table S1.
non-specified amylase, with 115 and 12 members from Bacteria and Archaea, respectively. This indicates that the specificity of a GH57 maltogenic amylase has probably only evolved in Archaea, whereas the GH57 non-specified amylase dominates in Bacteria, although the archaeal counterparts can be found too. All 139 putative proteins should be enzymically active GH57 members since they possess the GH57 family catalytic machinery appropriately positioned within the catalytic (β/α)7-barrel domain and contain all GH57 characteristic CSRs, with CSR-5 located in the α-helical bundle succeeding the barrel. Both novel specificities can be distinguished from the remaining GH57 enzymes by the presence of an invariant tyrosine corresponding with an otherwise conserved histidine (position 3 in CSR-1) involved in the subsite −1 (Imamura et al., 2003; Palomo et al., 2011). The evolutionary analysis of all currently known family GH57 enzyme specificities has demonstrated that the non-specified amylases and maltogenic amylases analysed in this study are most closely related to each other but simultaneously they have to be treated as independent enzyme entities.

ACKNOWLEDGEMENTS

This work was supported by the Slovak Research and Development Agency under contract no. LPP-0417-09 and by VEGA grant no. 2/0148/11.

REFERENCES


Edited by: Servé Kengen