DE GRUYTER

Biologia **70**/10: 1284—1294, 2015
Section Cellular and Molecular Biology
DOI: 10.1515/biolog-2015-0165

A novel GH13 subfamily of α-amylases with a pair of tryptophans in the helix α3 of the catalytic TIM-barrel, the LPDlx signature in the conserved sequence region V and a conserved aromatic motif at the C-terminus*

Štefan JANEČEK[1,2]**, Andrea KUCHTOVÁ[1] & Soňa PETROVIČOVÁ[1,2]

[1] *Laboratory of Protein Evolution, Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84551 Bratislava, Slovakia; e-mail: Stefan.Janecek@savba.sk*

[2] *Department of Biology, Faculty of Natural Sciences, University of SS. Cyril and Methodius, SK-91701 Trnava, Slovakia*

Abstract: The α-amylase enzyme specificity has been classified in the Carbohydrate-Active enZyme (CAZy) database into the families GH13, GH57, GH119 and eventually also GH126. α-Amylase is a glycoside hydrolase (GH) that catalyses in an endo-fashion the hydrolysis of the α-1,4-glucosidic linkages in starch and related α-glucans employing the retaining reaction mechanism. The family GH13 is the main α-amylase family with more than 28,000 members and 30 different specificities. The entire family GH13 has already been divided into 40 subfamilies; the α-amylase enzyme specificity being found in the subfamilies GH13_1, 5, 6, 7, 15, 19, 24, 27, 28, 32, 36 and 37. The present *in silico* study delivers a proposal to create a novel GH13 subfamily with the specificity of α-amylase. The proposal is based on a detailed bioinformatics analysis consisting of sequence, structural and evolutionary comparison of experimentally characterized α-amylases from, e.g., *Bacillus aquimaris*, *Anoxybacillus* sp. SK3-4 and DT3-1 and *Geobacillus thermoleovorans*, and hypothetical proteins, accompanied by α-amylases from well-established GH13 subfamilies and by closely related amylolytic enzymes (mainly from the subfamily GH13_31). Three sequence-structural features can be ascribed to the members of the newly proposed GH13 subfamily: (i) the pair of adjacent tryptophan residues positioned between the CSR-V and CSR-II in the helix α3 of the catalytic TIM-barrel; (ii) the sequence LPDlx in their CSR-V; and (iii) a ~30-residue long C-terminal region with a motif of five conserved aromatic residues. From the evolutionary point of view, the novel GH13 α-amylase subfamily is most closely related to fungal and yeast α-amylases classified in the subfamily GH13_1.

Key words: α-amylase; family GH13; GH13 subfamilies; unique sequence features; conserved sequence regions; a pair of adjacent tryptophans; evolutionary relatedness.

Abbreviations: BaqA, *Bacillus aquimaris* α-amylase; CAZy, Carbohydrate-Active enZymes; CSR, conserved sequence region; GH, glycoside hydrolase; PDB, Protein Data Bank.

Introduction

α-Amylase (EC 3.2.1.1) is an endo-type glycoside hydrolase (GH) catalysing with the retaining mechanism the hydrolysis of the α-1,4-glucosidic linkages in starch and related α-glucans (Janecek et al. 2014). Although this enzyme specificity is generally widespread in nature, individual α-amylases produced by various bacterial, archaeal and eukaryotic organisms may differ from each other by their exact substrate preference and product specificity (Svensson 1994; Leveque et al. 2000; Stanley et al. 2005; Kelly et al. 2009; van Zyl et al. 2012; Sharma & Satyanarayana 2013; Li et al. 2014).

Within the sequence-based classification of all GHs (Henrissat 1991) incorporated in the Carbohydrate-Active enZymes (CAZy) database (Cantarel et al. 2009), the α-amylase specificity is present in families GH13, GH57, GH119 and, conditionally, also GH126 (Janecek et al. 2014; Lombard et al. 2014). The family GH13 has been considered to be the main α-amylase family (more than 28,000 members and 30 different specificities), representing, in fact, the clan GH-H formed by families GH70 and GH77 in addition to GH13 (Kuriki & Imanaka 1999; MacGregor et al. 2001; van der Maarel et al. 2002). The family GH57 has been known as the second and smaller α-amylase family (more than 1,400 members and less than 10 specificities) (Zona et al. 2004; Palomo et al. 2011; Blesak & Janecek 2012, 2013; Park et al. 2014) recently shown as closely related with the very small fam-

* Electronic supplementary material. The online version of this article (DOI:10.1515/biolog-2015-0165) contains supplementary material, which is available to authorized users.
** Corresponding author

©2015 Institute of Molecular Biology, Slovak Academy of Sciences

ily GH119 (Janecek & Kuchtova 2012) counting only 12 sequences with one member characterized as an α-amylase (Watanabe et al. 2006; Lombard et al. 2014). With regard to the family GH126, its member characterized as an amylase active on maltooligosaccharides, amylose and glycogen exhibits clear structural similarity to β-glucan-active enzymes employing the inverting mechanism (Ficko-Blean et al. 2011), so the presence of a pure α-amylase specificity in GH126 (Lombard et al. 2014) may be considered as disputable (Janecek et al. 2014).

The α-amylase family GH13 as one of the largest CAZy GH families (Janecek et al. 2014; Lombard et al. 2014) has been in 2006 officially divided by curators into 35 subfamilies (Stam et al. 2006) in order to reflect closer functional similarities and evolutionary relatedness among members of a subfamily. Some polyspecific subfamilies, i.e. oligo-1,6-glucosidase and neopullulanase ones, were proposed earlier (Oslancova & Janecek 2002) based on unique differences in specific conserved sequence regions (CSRs) characteristic for the α-amylase family (Janecek 2002). The subfamily GH13_36 closely related to both subfamilies mentioned above covers interesting α-amylases some of which possess also the activity toward pullulan and cyclodextrins (Majzlova et al. 2013).

A few years ago two closely related α-amylases ASKA and ADTA from two *Anoxybacillus* species were described (Chai et al. 2012). Subsequently, a preliminary bioinformatics analysis of a homologous α-amylase from *Bacillus aquimaris* BaqA revealed the presence of two consecutive tryptophans positioned at the helix $\alpha3$ of the catalytic $(\beta/\alpha)_8$-barrel domain (i.e. the TIM-barrel), a feature discriminating this α-amylase from other well-established GH13 subfamilies with the α-amylase specificity (Puspasari et al. 2013). In fact, the first biochemically characterized α-amylase with such a "double-tryptophan" sequence feature could be the AmyB from *Anaerobranca gottschalkii* (Ballschmiter et al. 2005), but at that time in 2005 there were obviously no clear GH13 homologues available with the same α-amylase specificity. Importantly, the three-dimensional structure has already been solved for a counterpart enzyme from *Geobacillus thermoleovorans* GTA (Mok et al. 2013) that exhibits 100% sequence identity to a previously reported α-amylase from the same organism, strain Pizzo (Finore et al. 2011). Mok et al. (2013) pointed out that GTA possesses at its C-terminal end 5 conserved aromatic residues (phenylalanines and tyrosines), i.e. a motif that might be common for the newly forming group of α-amylases. Furthermore, the specific sequence signature LPDlx, representing the CSR-V positioned in domain B, could be the additional feature of interest (Puspasari et al. 2013; Ranjani et al. 2014). Very recently, an additional α-amylase from *Geobacillus thermoleovorans* GTA-II was found, for which its domain C has been proposed to be responsible for the enzyme adsorption to raw starch (Mehta & Satyanarayana 2014).

Since these six novel and mutually closely related GH13 α-amylases have still not been ascribed to any GH13 subfamily, we have undertaken a bioinformatics study with the main goal to describe this group of α-amylases as a novel GH13 subfamily. To achieve this goal, as many as possible hypothetical homologous α-amylases were retrieved from sequence databases and their sequences were analysed in details in an effort to identify their unique sequence-structural features that would clearly discriminate the entire novel group from all remaining, i.e. not only α-amylase, GH13 subfamilies.

Material and methods

Sequence collection

Sequences were collected based on protein BLAST (Altshul et al. 1990) search against the non-redundant database using the entire amino acid sequence of *Bacillus aquimaris*α-amylase BaqA (Puspasari et al. 2013). In addition to BaqA, the 4-5 more characterized α-amylases – two *Anoxybacillus* α-amylases ASKA and ADTA (Chai et al. 2012), two *Geobacillus thermoleovorans* α-amylases GTA and GTA-II (Finore et al. 2011; Mok et al. 2013; Mehta & Satyanarayana 2014) and eventually also the *Anaerobranca gottschalkii* α-amylase AmyB (Ballschmiter et al. 2005) – have been recognized as potentially forming the new GH13 subfamily. Based on their comparison, the criteria for a sequence assignment to the new GH13 subfamily were as follows: (i) presence of all seven CSRs characteristic of the α-amylase family GH13 (Janecek 2002); (ii) complete catalytic machinery, i.e. the aspartic acid at the strand $\beta4$ (CSR-II), glutamic acid at the strand $\beta5$ (CSR-III) and aspartic acid at the strand $\beta7$ (CSR-IV) corresponding with Asp214, Glu243 and Asp311, respectively, of the BaqA α-amylase (Puspasari et al. 2013); (iii) possessing the two consecutive tryptophans positioned at the helix $\alpha3$ preceding the strand $\beta4$ with the catalytic nucleophile within the catalytic TIM-barrel; (iv) the sequence signature LPDlx in the CSR-V; and (v) a motif of up to five conserved aromatic residues within the ~30-residue long region at the C-terminus.

Those sequences that have not possessed the complete α-amylase family GH13 catalytic machinery were eliminated despite the fact they may contain the two above-mentioned tryptophans. Any fragments, i.e. obviously incomplete sequences, were not taken into the comparison, too. Using these criteria, a set of 101 sequences of the supposedly novel GH13 α-amylase subfamily (Table S1) was obtained.

Several enzymes with non-α-amylase specificity from closely related oligo-1,6-glucosidase and neopullulanase subfamilies (Oslancova & Janecek 2002; Majzlova et al. 2013) were further added as follows: (i) five subfamily GH13_31 α-glucosidases possessing the two tryptophans and the CSR-V in the version QPDLx; (ii) two hypothetical members of subfamilies GH13_29 and GH13_31 along with the currently unclassified cyclomaltodextrinase from *Flavobacterium* sp. No. 192 (Fritzsche et al. 2003) having the two tryptophans and MPDLx as the CSR-V (intermediate character); and (iii) representatives of oligo-1,6-glucosidase (Watanabe et al. 1997), α-glucosidase (Nakao et al. 1994), dextran glucosidase (Hondoh et al. 2008) and sucrose isomerase (Zhang et al. 2003) – all from the subfamily GH13_31 – with QPDLx as the CSR-V but having a phenylalanine in the position corresponding with the first of the two tryptophans.

The entire set was finally completed by selected α-amylases ascribed to the individual well-established GH13

subfamilies, i.e. 1, 5, 6, 7, 15, 19, 24, 27, 28, 32, 36 and 37 (Lombard et al. 2014) that have been used in previous bioinformatics studies (Hostinova et al. 2010; DaLage et al. 2013; Majzlova et al. 2013; Puspasari et al. 2013) so that the final number of studied amylolytic enzymes and proteins was 146 (Table S1).

Evolutionary comparison

All 146 GH13 sequences were retrieved from GenBank (Benson et al. 2014) and UniProt (UniProt Consortium 2014) sequence databases and the set was aligned using the program Clustal-X (Larkin et al. 2007). A subtle manual tuning was done in order to maximize similarities, especially with regard to aligning the individual CSRs. The boundaries of the CSRs were defined based on previous bioinformatics studies (Janecek 2002; Oslancova & Janecek 2002; Da Lage et al. 2004; Majzlova et al. 2013; Puspasari et al. 2013). The evolutionary tree was based on the final alignment of the sequence segment corresponding to 269-residue long region of BaqA α-amylase (Puspasari et al. 2013) spanned almost the entire catalytic $(\beta/\alpha)_8$-barrel domain including the domain B from the beginning of the CSR-VI (strand $\beta 2$; starting with Gly82) to the end of the CSR-VII (strand $\beta 8$; ending with Ser350). The tree was calculated as a Phylip-tree type using the neighbour-joining clustering (Saitou & Nei 1987) and the bootstrapping procedure – the number of bootstrap trials used was 1,000 (Felsenstein 1985) implemented in the Clustal-X package (Larkin et al. 2007). The tree was displayed with the program iTOL (Letunic & Bork 2007).

The sequence logo for the CSRs was created using the WebLogo 3.0 server (http://weblogo.threeplusone.com/; Crooks et al. 2004).

Tertiary structure comparison

Three-dimensional structures were retrieved from the Protein Data Bank (PDB; Berman et al. 2000) for: (i) the α-amylase GTA from *Geobacillus thermoleovorans* (PDB code: 4E2O; Mok et al. 2013) as a representative of the novel α-amylase GH13 subfamily; and (ii) if available, for members of all remaining subfamilies and groups listed in Table S1. In the case there was not a three-dimensional structure available in any of the remaining subfamilies and groups mentioned above (e.g. the α-amylase GH13 subfamilies 19, 27, 32 and 37), structural models for their representatives were created using the Phyre2 server (www.sbg.bio.ic.ac.uk/phyre2/; Kelley & Sternberg 2009). The model was also created for the α-amylase BaqA from *Bacillus aquimaris* (Puspasari et al. 2013) as a leading representative of the novel α-amylase GH13 subfamily.

The individual structures (regardless they represented a real structure or a structural model) were compared to both the real α-amylase GTA from *Geobacillus thermoleovorans* and the model of the α-amylase BaqA from *Bacillus aquimaris*, making their superimposition using the program MultiProt (Shatsky et al. 2004).

Results and discussion

Sequence comparison

The present bioinformatics study delivers a proposal to create a novel GH13 subfamily exhibiting the α-amylase specificity. This subfamily is represented here by six closely related and experimentally characterized α-amylases described in the literature and 95 additional hypothetical proteins caught by BLAST search (Table S1). The former α-amylases, listed chronologically, are as follows: (i) AmyB from *Anaerobranca gottschalkii* (Ballschmiter et al. 2005); (ii) AmyA from *Geobacillus thermoleovorans* subsp. *stromboliensis* – strain Pizzo (Finore et al. 2011); (iii) ASKA and ADTA from *Anoxybacillus* sp. SK3-4 and DT3-1, respectively (Chai et al. 2012); (iv) BaqA from *Bacillus aquimaris* (Puspasari et al. 2013); (v) GTA from *Geobacillus thermoleovorans* (Mok et al. 2013) (identical to that from the strain Pizzo); and (vi) GTA-II from *Geobacillus thermoleovorans* (Mehta & Satyanarayana 2014). The set of studied enzymes was completed (Table S1) by 32 representatives of all individual GH13 subfamilies with the α-amylase specificity, i.e. the subfamilies 1, 5, 6, 7, 15, 19, 24, 27, 28, 32, 36 and 37 (Stam et al. 2006; Lombard et al. 2014), as well as by 13 related amylolytic enzymes exhibiting closely related sequence features especially within the CSRs (Oslancova & Janecek 2002; Majzlova et al. 2013).

Figure 1 shows the seven CSRs that are the best conserved sequence stretches characteristic of the entire α-amylase family (Janecek 2002). As can be seen, the most exclusive sequence feature of the novel GH13 α-amylase subfamily could be the pair of adjacent tryptophan residues Trp201-Trp202 (*Bacillus aquimaris* BaqA α-amylase numbering), positioned between the CSR-V and CSR-II in the helix $\alpha 3$ of the catalytic TIM-barrel since such a feature is not present in any other GH13 subfamily with the α-amylase specificity. It is worth mentioning that the pair of tryptophans cannot be used as a sole sequence marker of the novel α-amylase subfamily (Fig. 1a) since it may be present in some members of the subfamily GH13_31 (Fig. 1b), i.e. of the so-called oligo-1,6-glucosidase subfamily (Oslancova & Janecek 2002; Stam et al. 2006). Thus the members of the novel GH13 α-amylase subfamily should contain the sequence LPDlx in their CSR-V (Fig. 1a), whereas the α-glucosidases from the subfamily GH13_31 possess typically QPDLN (Oslancova & Janecek 2002; Majzlova et al. 2013) as their CSR-V (Fig. 1b). Remarkably, some other related enzymes, currently unassigned to any of the established GH13 subfamilies, represented by the cyclomaltodextrinase from *Flavobacterium* sp. No. 92 (Fritzsche et al. 2003), have the tryptophan pair, too, but these related enzymes differ also from the members of the novel subfamily by their specific CSR-V sequence MPDLN (Fig. 1b). This signature was previously suggested to be a feature of a group of amylolytic enzymes intermediate between true oligo-1,6-glucosidases and true neopullulanases (Oslancova & Janecek 2002) that was later classified as the subfamily GH13_36 (Stam et al. 2006; Lombard et al. 2014). However, the α-amylases from the subfamily GH13_36 that exhibit also, e.g., cyclodextrinase and neopullulanase activities (Majzlova et al. 2013), do not possess the pair of adjacent tryptophan residues (Fig. 1b).

The sequence logo created for all 101 real and hypothetical α-amylases that could constitute a novel GH13 subfamily (Fig. 2) clearly summarizes the residues typically present in the individual positions of each

CSR. It is evident that most residues are conserved almost invariantly, the fact that might be a consequence of, until now, taxonomically rather narrow spectrum of producers (Table S1). Currently, the 4th position in the CSR-I (alanine and valine), the 5th position in the CSR-V (alanine and asparagine), the 8th position in the CSR-II (arginine and lysine), the 1st (phenylalanine and tyrosine), 3rd (leucine and isoleucine) and 8th (mostly ser-

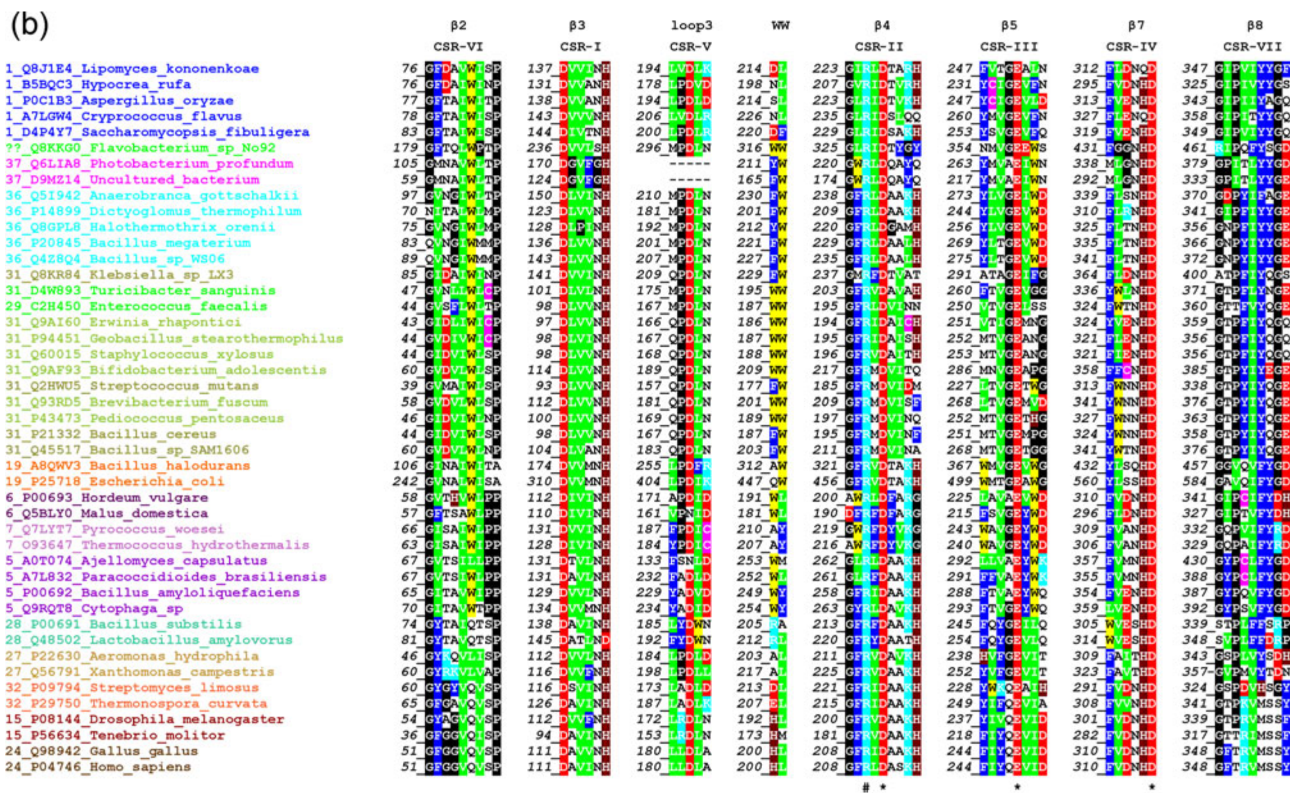

Fig. 1. Sequence alignment of CSRs of studied family GH13 enzymes with focus on the novel α-amylase subfamily. The two consecutive tryptophans characteristic for the novel α-amylase subfamily are also shown. Colour code for the selected residues: W, yellow; F, Y – blue; V, L, I – green; D, E – red; R, K – cyan; H – brown; C – magenta; G, P – black. The catalytic triad is signified by asterisks and the only one additional invariantly conserved position of the arginine in the CSR-II is marked by a hashtag under the alignment. (a) The order from top reflects the relatedness of the new GH13 subfamily α-amylases to the BaqA α-amylase (i.e. their sequence similarity) as delivered by the BLAST search. (b) For the sequences of remaining α-amylases and related amylolytic enzymes, the order corresponds with their arrangement in the evolutionary tree (Fig. 4). The label of the protein source consists of the UniProt (UniParc) accession number and the name of the organism. If there is an additional number at the beginning of the protein source label, it means the number of the GH13 subfamily. The alignment of all 146 enzymes spanning the sequence segment from the beginning of the strand β2 (CSR-VI) to the end of the strand β8 (CSR-VII) is shown in Figure S1.

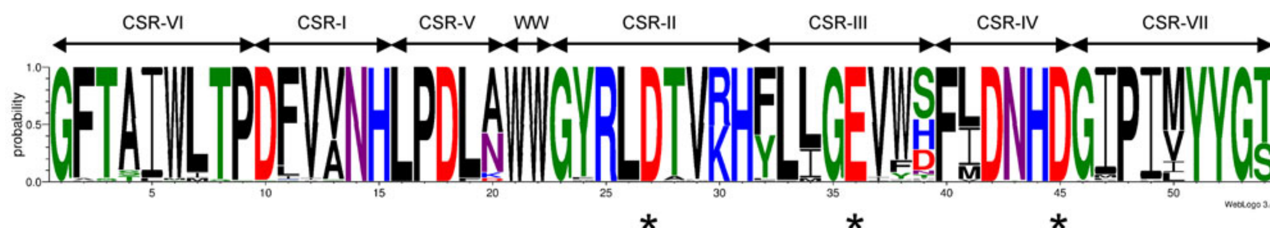

Fig. 2. Sequence logo of the novel GH13 subfamily. CSR-I, residues 10-15; CSR-II, residues 23-31; CSR-III, residues 32-39; CSR-IV, residues 40-45; CSR-V, residues 16-20; CSR-VI, residues 1-9; CSR-VII, residues 46-54. The two adjacent characteristic tryptophans, positioned between the CSR-V and CSR-II, are also shown. The catalytic triad, i.e. the catalytic nucleophile (No. 27, aspartic acid), the proton donor (No. 36, glutamic acid) and the transition-state stabiliser (No. 45, aspartic acid) are indicated by asterisks. The logo is based on 101 sequences of real and hypothetical α-amylases that potentially define the new GH13 subfamily.

ine, histidine and aspartic acid) positions in the CSR-III, the 2nd position in the CSR-IV (leucine, isoleucine and methionine), and the 5th (mostly methionine, valine and isoleucine) and 9th (threonine and serine) positions in the CSR-VII – belong to positions within the CSRs that obviously do not require a strictly invariant amino acid residue (Fig. 2). For example, based on a protein engineering study of the α-amylase from *Anoxybacillus* sp. SK3-4 ASKA it was found that the naturally present alanine at the end of the CSR-V (Fig. 1a) may contribute to the high maltose production of the ASKA (Ranjani et al. 2014). It nevertheless can be awaited – as the number of members of this novel subfamily will increase – the future logo will reveal both the positions evolutionarily conserved and those that are tolerant to changes.

With regard to comparison of additional sequence segments that connect the CSRs, their alignment spanning the region from the beginning of the CSR-VI to the end of the CSR-VII is shown in Figure S1. It covers, in fact, almost the entire catalytic $(\beta/\alpha)_8$-barrel domain (from the strand β2 to the strand β8) includ-

Table 1. Tertiary structure comparison of two α-amylases from the novel subfamily with representatives of the remaining studied family GH13 enzymes.[a]

Family GH13 representatives	G8N704_*Geobacillus_thermoleovorans* (4E2O)		G8IJA7_*Bacillus_aquimaris* (4e2o)	
	C_α	RMSD (Å)	C_α	RMSD (Å)
31_P94451_*Geobacillus_stearothermophilus* (2ze0)	315	1.64	318	1.66
??_Q8KKG0_*Flavobacterium*_sp_No92 (1H3G)	378	1.48	378	1.49
31_P21332_*Bacillus_cereus* (1UOK)	327	1.69	334	1.73
31_Q45517_*Bacillus*_sp_SAM1606 (1uok)	329	1.71	307	1.55
31_Q2HWU5_*Streptococcus_mutans* (2ZID)	312	1.70	314	1.55
31_Q8KR84_*Klebsiella*_sp_LX3 (1M53)	331	1.62	327	1.54
36_Q8GPL8_*Halothermothrix_orenii* (1WZA)	366	1.44	367	1.44
1_P0C1B3_*Aspergillus_oryzae* (2TAA)	379	1.70	379	1.71
5_P00692_*Bacillus_amyloliquefaciens* (3BH4)	291	1.65	292	1.66
6_P00693_*Hordeum_vulgare* (1P6W)	287	1.63	287	1.63
7_Q7LYT7_*Pyrococcus_woesei* (1MXD)	315	1.69	315	1.69
15_P56634_*Tenebrio_molitor* (1JAE)	284	1.64	283	1.63
19_P25718_*Escherichia_coli* (4aee)	314	1.49	315	1.50
24_P04746_*Homo_sapiens* (1HNY)	265	1.58	277	1.63
27_P22630_*Aeromonas_hydrophila* (1jae)	271	1.55	270	1.56
28_P00691_*Bacillus_substilis* (1BAG)	265	1.52	275	1.57
32_P09794_*Streptomyces_limosus* (4jcl)	344	1.47	343	1.46
37_D9MZ14_Uncultured_bacterium (1tcm)	360	1.36	360	1.37

[a] The label of the protein source consists of the UniProt (UniParc) accession number and the name of the organism. If there is an additional number at the beginning of the protein source label, it means the number of the GH13 subfamily. The tertiary structures indicated as PDB codes (in parenthesis) were obtained either as PDB co-ordinates of determined real tertiary structures (capital letters) or model co-ordinates based on real PDB structures (small letters). C_α and RMSD represent the number of superimposed C-alpha atoms and the root-mean-square deviation, respectively.

ing the entire domain B that together, for all members of the novel GH13 subfamily, consist of approximately 270 residues (Table S1). The alignment revealed that the α-amylases proposed here to define a new GH13 subfamily contain, in addition to the abovementioned pair of adjacent tryptophans, several other conserved tryptophan residues, such as Trp103, Trp149, Trp158, Trp177 and Trp224 (*Bacillus aquimaris* BaqA α-amylase numbering). These may have their counterparts mainly among the α-amylases from the subfamilies GH13_1 and GH13_36 as well as among the α-glucosidases from the subfamily GH13_31 (Fig. S1). Since the novel GH13 subfamily could be the group of raw-starch degrading α-amylases (Finore et al. 2011; Puspasari et al. 2011; Mehta & Satyanarayana 2014) without any distinct starch-binding domain (Mok et al. 2013; Puspasari et al. 2013), the aromatic tryptophan positions should be of interest (Janecek et al. 2011; Carvalho et al. 2015). Moreover, Mok et al. (2013) have pointed out the eventual role in raw starch binding and degradation ability of a ∼30-residue long C-terminal region of *Geobacillus thermoleovorans* α-amylase with five aromatic residues (phenylalanines and tyrosines), which might be an additional characteristic feature of the novel GH13 subfamily since that stretch is usually present and well-conserved (Fig. 3). Since it was truncated in the three-dimensional structure, it was only predicted to be an α-helix and its exact role could not be completely elucidated until now (Mok et al. 2013).

Structure comparison

In order to see the closest eventual structural homologues, the solved tertiary structure of *Geobacillus thermoleovorans* α-amylase GTA (Mok et al. 2013) was compared with those – either real structures or their models if real structures were not available – of representatives of all studied groups and/or subfamilies. These data are summarized in Table 1, supported also by data obtained when the structural model of *Bacillus aquimaris* BaqA α-amylase was used. It is obvious that the α-amylase GTA from *Geobacillus thermoleovorans* exhibits the best structural similarity with the α-amylase from *Halothermothrix orenii* from the subfamily GH13_36 (Sivakumar et al. 2006) and currently unassigned cyclomaltodextrinase from *Flavobacterium* sp. No. 92 (Fritzsche et al. 2003). Interesting similarity (Table 1) was revealed also to the α-amylase from uncultured bacterium from the subfamily GH13_37 (Liu et al. 2012), but for this α-amylase only the structure modelled according to the template was used and very probable lack of domain B in the subfamily GH13_37 (Janecek et al. 2014) should also be taken into account. Overall the presented structural comparison may thus indicate also the evolutionary relatedness of the novel proposed GH13 subfamily with the α-amylase subfamily GH13_36, the group represented by the *Flavobacterium* sp. No. 92 cyclomaltodextrinase and, eventually, with the α-amylase subfamily GH13_37.

Evolutionary relationships

The evolutionary relationships among the members of the novel proposed GH13 subfamily as well as those of this subfamily with α-amylases from other GH13 subfamilies and several closely related amylolytic enzymes (mostly α-glucosidases) are depicted in Figure 4. It is clear that currently there are three major clusters reflecting taxonomy consisting of α-amylases from the genera *Geobacillus*, *Anoxybacillus* and mostly *Bacillus*

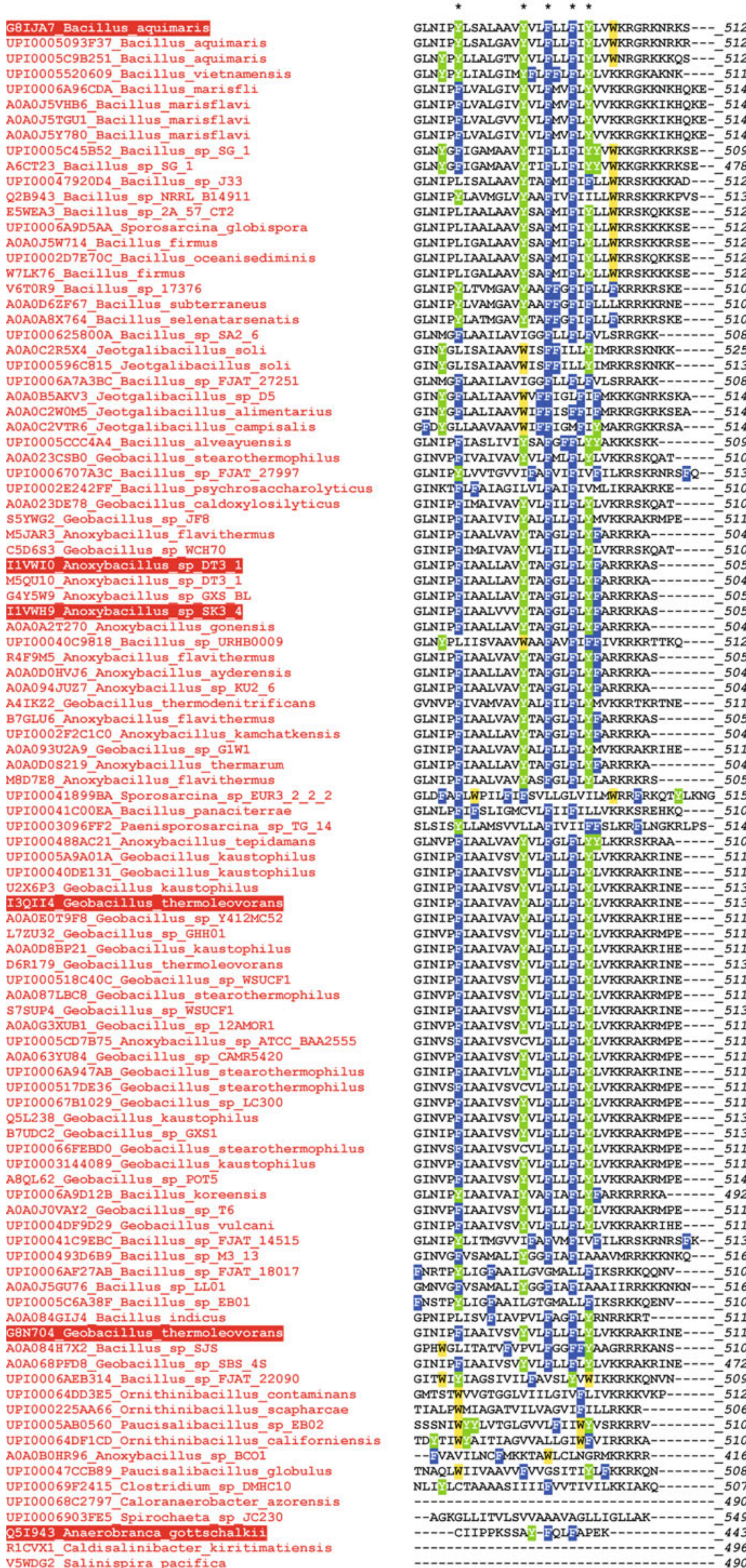

Fig. 3. Sequence alignment of ~30-residue long C-terminal segment of α-amylases of the newly proposed GH13 subfamily. This C-terminus was shown (Mok et al. 2013) to contain five conserved aromatic residues (phenylalanine and/or tyrosine). This feature (signified by asterisks above the alignment) could also be unique for the newly proposed GH13 subfamily of α-amylases. All phenylalanines and tyrosines present in the C-terminal segment are signified, respectively, by blue and green highlighting together with all tryptophan positions (yellow). The label of the protein source consists of the UniProt (UniParc) accession number and the name of the organism. Note that a few putative α-amylases from *Caloranaerobacter azorensis*, *Caldisalinibacter kiritimatiensis* and *Salinispira pacifica*, obviously do not contain such a C-terminus with 5 aromatic residues.

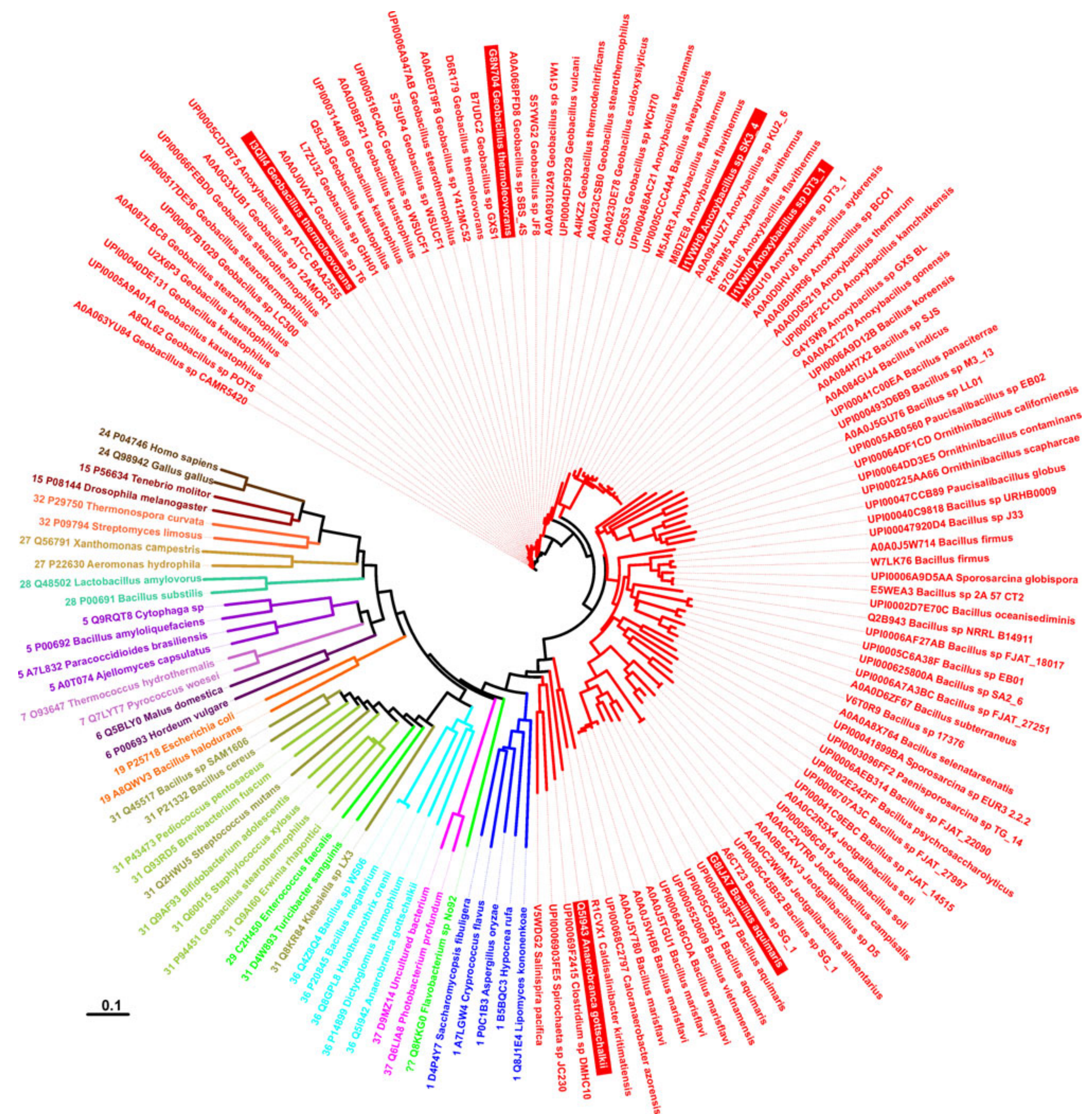

Fig. 4. Evolutionary tree of studied family GH13 enzymes with focus on the novel α-amylase subfamily. The label of the protein source consists of the UniProt (UniParc) accession number and the name of the organism. If there is an additional number at the beginning of the protein source label, it means the number of the GH13 subfamily. The tree is based on the alignment shown in Figure S1.

(covering also *Paucisalibacillus*, *Ornithinibacillus* and *Jeotgalibacillus*). Each of the three named clusters contains at least one real, i.e. biochemically characterised α-amylase (cf. Table S1). Interestingly, the part of the evolutionary tree, where the hypothetical α-amylases from the genus *Bacilus* dominate, covers also those from genera *Sporosarcina* and *Paenisporosarcina*; both, however, belonging to the class of Bacilli of the phylum Firmicutes.

It is worth mentioning that six sequences that, in fact, are neighbouring with remaining α-amylases and related amylolytic enzymes, represented in the tree by fungal and yeast α-amylases from the subfamily GH13_1, may represent some intermediates. In addition to the α-amylase from *Anaerobranca gottchalkii* (Ballschmiter et al. 2005; UniProt accession No.: Q5I943), they are five hypothetical proteins as follows: *Caloranaerobacter azorensis* (UPI00068C2797), *Caldisalinibacter kiritimatiensis* (R1CVX1), *Clostridium* sp. DMHC10 (UPI00069F2415), *Spirochaeta* sp. JC230 (UPI0006903FE5) and *Salinispira pacifica* (V5WDG2). While *Anaerobranca*, *Caloranaerobacter*, *Caldisalinibacter* and *Clostridium* rank all among the class of Clostridia under the phylum Firmicutes, both *Spirochaeta* and *Salinispira* rank among the phylum Spirochaetes. Despite all the six have been included in

the present study (Table S1), none them might necessarily belong to the newly proposed α-amylase subfamily. It is also possible that the number of various genera producing the α-amylases of this novel GH13 subfamily will not dramatically increase and from the taxonomical point of view, the subfamily will remain a bacterial subfamily containing predominantly Firmicutes (Table S1).

The fungal and yeast α-amylases from the subfamily GH13_1, represented by the Taka-amylase A (Matsuura et al. 1984) are the group most closely related to the novel GH13 subfamily. On the other hand, the α-amylases belonging to the so-called "animal" group of α-amylases (Janecek 1994; D'Amico et al. 2000; Da Lage et al. 2004, 2007; Janecek et al. 2014), i.e. those from subfamilies GH13_24 (e.g. mammals), GH13_15 (e.g. insect) and GH13_32 (e.g. actinomycetes) represent the α-amylases most distantly related to those from the newly proposed subfamily. Interestingly, the unclassified cyclomaltodextrinase from *Flavobacterium* sp. No. 92 (Fritzsche et al. 2003) exhibits on the other hand a close relatedness to the new subfamily, followed by the α-amylases from recently established subfamily GH13_37 (Lei et al. 2012). All remaining non-α-amylases from the subfamily GH13_31 (and GH13_29) included in the present study, i.e. the specificities of oligo-1,6-glucosidase (Watanabe et al. 1997), α-glucosidase (Nakao et al. 1994), dextran glucosidase (Hondoh et al. 2008) and sucrose isomerase (Zhang et al. 2003), are found to be clustered together (Fig. 4) and sharing a common branch with the "intermediary" α-amylases from the subfamily GH13_36 (Majzlova et al. 2013).

Conclusions

The amino acid sequences of experimentally characterized α-amylases BaqA from *Bacillus aquimaris*, ASKA and ADTA from *Anoxybacillus* sp. SK3-4 and DT3-1, respectively, GTA and GTA-II both from *Geobacillus thermoleovorans* and eventually also (although less convincingly) the AmyB from *Anaerobranca gottschalkii* were analysed in detail together with their 95 other hypothetical protein homologues available in sequence databases. These α-amylases are proposed to define a novel GH13 subfamily with the α-amylase specificity, in addition to subfamilies 1, 5, 6, 7, 15, 19, 24, 27, 28, 32, 36 and 37, already established in the CAZy database. The novel GH13 subfamily can be characterized by a few exclusive sequence features, such as the pair of adjacent tryptophan residues positioned between the CSR-V and CSR-II in the helix $\alpha3$ of the catalytic TIM-barrel, the sequence signature LPDlx in their CSR-V and a ~30-residue long C-terminal region with a motif of five conserved aromatic residues.

Acknowledgements

This work was supported by the VEGA grant No. 2/150/14.

References

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. 1990. Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Ballschmiter M., Armbrecht M., Ivanova K., Antranikian G. & Liebl W. 2005. AmyA, an α-amylase with β-cyclodextrin-forming activity, and AmyB from the thermoalkaliphilic organism *Anaerobranca gottschalkii*: two α-amylases adapted to their different cellular localizations. Appl. Environ. Microbiol. 71: 3709–3715.

Benson D.A., Clark K., Karsch-Mizrachi I., Lipman D.J., Ostell J. & Sayers E.W. 2014. GenBank. Nucleic Acids Res **42:** D32–D37.

Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N. & Bourne P.E. 2000. The Protein Data Bank. Nucleic Acids Res. **28:** 235–242.

Blesak K. & Janecek S. 2012. Sequence fingerprints of enzyme specificities from the glycoside hydrolase family GH57. Extremophiles **16:** 497–506.

Blesak K. & Janecek S. 2013. Two potentially novel amylolytic enzyme specificities in the prokaryotic glycoside hydrolase α-amylase family GH57. Microbiology **159**: 2584–2593.

Cantarel B.L., Coutinho P.M., Rancurel C., Bernard T., Lombard V. & Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. **37:** D233–D238.

Carvalho C.C., Phan N.N., Chen Y. & Reilly P.J. 2015. Carbohydrate-binding module tribes. Biopolymers 103: 203–214.

Chai Y.Y., Rahman R.N., Illias R.M. & Goh K.M. 2012. Cloning and characterization of two new thermostable and alkalitolerant α-amylases from the *Anoxybacillus* species that produce high levels of maltose. J. Ind. Microbiol. Biotechnol. **39:** 731–741.

Crooks G.E., Hon G., Chandonia J.M. & Brenner S.E. 2004. WebLogo: a sequence logo generator. Genome Res. **14:** 1188–1190.

Da Lage J.L., Binder M., Hua-Van A., Janecek S. & Casane D. 2013. Gene make-up: rapid and massive intron gains after horizontal transfer of a bacterial α-amylase gene to Basidiomycetes. BMC Evolutionary Biology **13:** 40.

Da Lage J.L., Danchin E.G. & Casane D. 2007. Where do animal α-amylases come from? An interkingdom trip. FEBS Lett. **581:** 3927–3935.

Da Lage J.L., Feller G. & Janecek S. 2004. Horizontal gene transfer from Eukarya to bacteria and domain shuffling: the α-amylase model. Cell. Mol. Life Sci. **61:** 97–109.

D'Amico S., Gerday C. & Feller G. 2000. Structural similarities and evolutionary relationships in chloride-dependent α-amylases. Gene **253:** 95–105.

Felsenstein J. 1985. Confidence-limits on phylogenies – an approach using the bootstrap. Evolution **39:** 783–791.

Ficko-Blean E., Stuart C.P. & Boraston A.B. 2011. Structural analysis of CPF_2247, a novel α-amylase from *Clostridium perfringens*. Proteins **79:** 2771–2777.

Finore I., Kasavi C., Poli A., Romano I., Toksoy Oner E., Kirdar B., Dipasquale L., Nicolaus B. & Lama L. 2011. Purification, biochemical characterization and gene sequencing of a thermostable raw starch digesting α-amylase from *Geobacillus thermoleovorans* subsp. *stromboliensis* subsp. nov. World J. Microbiol. Biotechnol. **27:** 2425–2433.

Fritzsche H.B., Schwede T. & Schulz G.E. 2003. Covalent and three-dimensional structure of the cyclodextrinase from *Flavobacterium* sp. no. 92. Eur. J. Biochem. **270:** 2332–2341.

Henrissat B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. **280:** 309–316.

Hondoh H., Saburi W., Mori H., Okuyama M., Nakada T., Matsuura Y. & Kimura A. 2008. Substrate recognition mechanism of α-1,6-glucosidic linkage hydrolyzing enzyme, dextran glucosidase from *Streptococcus mutans*. J. Mol. Biol. **378:** 913–922.

Hostinova E., Janecek S. & Gasperik J. 2010. Gene sequence, bioinformatics and enzymatic characterization of α-amylase from *Saccharomycopsis fibuligera* KZ. Protein J. **29:** 355–364.

Janecek S. 1994. Sequence similarities and evolutionary relationships of microbial, plant and animal α-amylases. Eur. J. Biochem. **224:** 519–524.

Janecek S. 2002. How many conserved sequence regions are there in the α-amylase family? Biologia **57 (Suppl 11):** 29–41.

Janecek S. & Kuchtova A. 2012. In silico identification of catalytic residues and domain fold of the family GH119 sharing the catalytic machinery with the α-amylase family GH57. FEBS Lett. **586:** 3360–3366.

Janecek S., Svensson B. & MacGregor E.A. 2011. Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. *Enzyme Microb. Technol.* **49:** 429–440.

Janecek S., Svensson B. & MacGregor E.A. 2014. α-Amylase – an enzyme specificity found in various families of glycoside hydrolases. Cell. Mol. Life Sci. **71:** 1149–1170.

Kelley L.A. & Sternberg M.J.E. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. **4:** 363–371.

Kelly R.M., Dijkhuizen L. & Leemhuis H. 2009. Starch and α-glucan acting enzymes, modulating their properties by directed evolution. J. Biotechnol. **140:** 184-193.

Kuriki T. & Imanaka T. 1999. The concept of the α-amylase family: structural similarity and common catalytic mechanism. J. Biosci. Bioeng. **87:** 557–565.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. & Higgins D.G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947–2948.

Lei Y., Peng H., Wang Y., Liu Y., Han F., Xiao Y. & Gao Y. 2012. Preferential and rapid degradation of raw rice starch by an α-amylase of glycoside hydrolase subfamily GH13_37. Appl. Microbiol. Biotechnol. **94:** 1577–1584.

Letunic I. & Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics **23:** 127–128.

Leveque E., Janecek S., Belarbi A. & Haye B. 2000. Thermophilic archaeal amylolytic enzymes. Enzyme Microb. Technol. **26:** 2–13.

Li C., Du M., Cheng B., Wang L., Liu X., Ma C., Yang C. & Xu P. 2014. Close relationship of a novel Flavobacteriaceae α-amylase with archaeal α-amylases and good potentials for industrial applications. Biotechnol. Biofuels **7:** 18.

Liu Y., Lei Y., Zhang X., Gao Y., Xiao Y. & Peng H. 2012. Identification and phylogenetic characterization of a new subfamily of α-amylase enzymes from marine microorganisms. Mar. Biotechnol. (NY) **14:** 253–260.

Lombard V., Golaconda Ramulu H., Drula E., Coutinho P.M. & Henrissat B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. **42:** D490–D495.

Long C.M., Virolle M.J., Chang S.Y., Chang S. & Bibb M.J. 1987. α-Amylase gene of *Streptomyces limosus*: nucleotide sequence, expression motifs, and amino acid sequence homology to mammalian and invertebrate α-amylases. J. Bacteriol. **169:** 5745–5754.

MacGregor E.A., Janecek S. & Svensson B. 2001. Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim. Biophys. Acta **1546:** 1–20.

Majzlova K., Pukajova Z. & Janecek S. 2013. Tracing the evolution of the α-amylase subfamily GH13_36 covering the amylolytic enzymes intermediate between oligo-1,6-glucosidases and neopullulanases. Carbohydr. Res. **367:** 48–57.

Matsuura Y., Kusunoki M., Harada W. & Kakudo M. 1984. Structure and possible catalytic residues of Taka-amylase A. J. Biochem. **95:** 697–702.

Mehta D. & Satyanarayana T. 2014. Domain C of thermostable α-amylase of *Geobacillus thermoleovorans* mediates raw starch adsorption. Appl. Microbiol. Biotechnol. **98:** 4503–4519.

Mok S.C., Teh A.H., Saito J.A., Najimudin N. & Alam M. 2013. Crystal structure of a compact α-amylase from *Geobacillus thermoleovorans*. Enzyme Microb. Technol. **53:** 46–54.

Nakao M., Nakayama T., Kakudo A., Inohara M., Harada M., Omura F. & Shibano Y. 1994. Structure and expression of a gene coding for thermostable α-glucosidase with a broad substrate specificity from *Bacillus* sp. SAM1606. Eur J Biochem **220:** 293–300.

Oslancova A. & Janecek S. 2002. Oligo-1,6-glucosidase and neopullulanase enzyme subfamilies from the α-amylase family defined by the fifth conserved sequence region. Cell. Mol. Life Sci. **59:** 1945–1959.

Palomo M., Pijning T., Booiman T., Dobruchowska J.M., van der Vlist J., Kralj S., Planas A., Loos K., Kamerling J.P., Dijkstra B.W., van der Maarel M.J., Dijkhuizen L. & Leemhuis H. 2011. *Thermus thermophilus* glycoside hydrolase family 57 branching enzyme:crystal structure, mechanism of action, and products formed. J. Biol. Chem. **286:** 3520–3530.

Park K.H., Jung J.H., Park S.G., Lee M.E., Holden J.F., Park C.S. & Woo E.J. 2014. Structural features underlying the selective cleavage of a novel exo-type maltose-forming amylase from *Pyrococcus* sp. ST04. Acta Crystallogr. D Biol. Crystallogr. **70:** 1659–1668.

Puspasari F., Nurachman Z., Noer A.S., Radjasa O.K., van der Maarel M.J.E.C. & Natalia D. 2011. Characteristics of raw starch degrading α-amylase from *Bacillus aquimaris* MKSC 6.2 associated with soft coral *Sinularia* sp. Starch – Stärke **63:** 461–467.

Puspasari F., Radjasa O.K., Noer A.S., Nurachman Z., Syah Y.M., van der Maarel M., Dijkhuizen L., Janecek S. & Natalia D. 2013. Raw starch-degrading α-amylase from *Bacillus aquimaris* MKSC 6.2: isolation and expression of the gene, bioinformatics and biochemical characterization of the recombinant enzyme. J. Appl. Microbiol. **114:** 108–120.

Ranjani V., Janecek S., Chai K.P., Shahir S., Noor R., Abdul Rahman Z.R., Chan K.G. & Goh K.M. 2014. Protein engineering of selected residues from conserved sequence regions of a novel *Anoxybacillus* α-amylase. Sci. Rep. **4:** 5850.

Saitou N. & Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Sharma A. & Satyanarayana T. 2013. Microbial acid-stable α-amylases: characteristics, genetic engineering and applications. Process Biochem. **48:** 201–211.

Shatsky M., Nussinov R. & Wolfson H.J. 2004. A method for simultaneous alignment of multiple protein structures. Proteins **56:** 143–156.

Sivakumar N., Li N., Tang J.W., Patel B.K. & Swaminathan K. 2006. Crystal structure of AmyA lacks acidic surface and provide insights into protein stability at poly-extreme condition. FEBS Lett. **580:** 2646–2652.

Stam M.R., Danchin E.G., Rancurel C., Coutinho P.M. & Henrissat B. 2006. Dividing the large glycoside hydrolase family 13 into subfamilies:towards improved functional annotations of α-amylase-related proteins. Protein Eng. Des. Sel. **19:** 555–562.

Stanley D., Farnden K.J.F. & MacRae E.A. 2005. Plant α-amylases: functions and roles in carbohydrate metabolism. Biologia **60 (Suppl.16):** 65–71.

Svensson B. 1994. Protein engineering in the α-amylase family: catalytic mechanism, substrate specificity, and stability. Plant Mol. Biol. **25:** 141–157.

UniProt Consortium 2014. Activities at the Universal Protein Resource (UniProt). Nucleic Acids Res. **42:** D191–D198.

van der Maarel M.J., van der Veen B., Uitdehaag J.C., Leemhuis H. & Dijkhuizen L. 2002. Properties and applications of starch-converting enzymes of the α-amylase family. J. Biotechnol. **94:** 137–155.

van Zyl W.H., Bloom M. & Viktor M.J. 2012. Engineering yeasts for raw starch conversion. Appl. Microbiol. Biotechnol. **95:** 1377–1388.

Watanabe H., Nishimoto T., Kubota M., Chaen H. & Fukuda S. 2006. Cloning, sequencing, and expression of the genes encoding an isocyclomaltooligosaccharide glucanotransferase

and an α-amylase from a *Bacillus circulans* strain. Biosci. Biotechnol. Biochem. **70:** 2690–2702.

Watanabe K., Hata Y., Kizaki H., Katsube Y. & Suzuki Y. 1997. The refined crystal structure of *Bacillus cereus* oligo-1,6-glucosidase at 2.0 Å resolution: structural characterization of proline-substitution sites for protein thermostabilization. J. Mol. Biol. **269:** 142–153.

Zhang D., Li N., Lok S.M., Zhang L.H. & Swaminathan K. 2003. Isomaltulose synthase (PalI) of *Klebsiella* sp. LX3. Crystal structure and implication of mechanism. J. Biol. Chem. **278:** 35428–35434.

Zona R., Chang-Pi-Hin F., O'Donohue M.J. & Janecek S. 2004. Bioinformatics of the family 57 glycoside hydrolases and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*. Eur. J. Biochem. **271:** 2863–2872.

Received September 7, 2015
Accepted October 30, 2015