Genetica (2020) 148:77–86
https://doi.org/10.1007/s10709-020-00089-0

ORIGINAL ARTICLE

New groups of protein homologues in the α-amylase family GH57 closely related to α-glucan branching enzymes and 4-α-glucanotransferases

Štefan Janeček[1,2] · Mária Martinovičová[2]

Received: 27 July 2019 / Accepted: 17 February 2020 / Published online: 24 February 2020
© Springer Nature Switzerland AG 2020

Abstract
The glycoside hydrolase family GH57 is known as the second α-amylase family. Its main characteristics are as follows: (i) employing the retaining reaction mechanism; (ii) adopting the $(\beta/\alpha)_7$-barrel (the incomplete TIM-barrel) with succeeding bundle of α-helices as the catalytic domain; (iii) sharing the five conserved sequence regions (CSRs) exhibiting the sequence fingerprints of the individual enzyme specificities; and (iv) using the catalytic machinery consisting of glutamic acid (the catalytic nucleophile) and aspartic acid (the proton donor) positioned at strands β4 (CSR-3) and β7 (CSR-4) of the $(\beta/\alpha)_7$-barrel domain, respectively. Several years ago, a group of hypothetical proteins closely related to the specificity of α-amylase was revealed, the so-called α-amylase-like homologues, the members of which lack either one or even both catalytic residues. The novelty of the present study lies in delivering two additional groups of the "like" proteins that are homologues of α-glucan-branching enzyme (GBE) and 4-α-glucanotransferase (4AGT) specificities. Based on a recently published in silico analysis of more than 1600 family GH57 sequences, 13 GBE-like and 18 4AGT-like proteins from unique sources were collected and analyzed in a detail with respect to their taxonomical origin, sequence and structural features as well as evolutionary relationships. This in silico study could accelerate the efforts leading to experimental revealing the real function of the enzymes-like proteins in the α-amylase family GH57.

Keywords α-Amylase family GH57 · In silico analysis · Catalytic machinery · α-Glucan branching enzyme · 4-α-Glucanotransferase · Evolutionary relatedness

Abbreviations

4AGT	4-α-Glucanotransferase
CSR	Conserved sequence region
GBE	α-Glucan branching enzyme
GH	Glycoside hydrolase
PDB	Protein Data Bank

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10709-020-00089-0) contains supplementary material, which is available to authorized users.

✉ Štefan Janeček
Stefan.Janecek@savba.sk

[1] Laboratory of Protein Evolution, Institute of Molecular Biology, Slovak Academy of Sciences, Dubravska cesta 21, 84551 Bratislava, Slovakia

[2] Department of Biology, Faculty of Natural Sciences, University of SS. Cyril and Methodius, Nam. J. Herdu 2, 91701 Trnava, Slovakia

Introduction

With regard to α-amylase families of glycoside hydrolases (GHs) and the Carbohydrate-Active enZymes (CAZy) database, the family GH13 represents the main, in terms of chronology, the original α-amylase family, whereas the family GH57 is the second and smaller α-amylase family (Janecek et al 2014; Lombard et al. 2014). Established in 1991 (Henrissat 1991) and 1996 (Henrissat and Bairoch 1996), respectively, the former contains almost 77 thousand sequences from all the three domains of life—Bacteria, Archaea and Eucarya, while the latter consists more than 2300 sequences only lacking, in fact, any relevant member from eukaryotes (Lombard et al. 2014).

The main α-amylase family GH13 with about 30 different enzyme specificities belongs to the largest GH families within CAZy (Lombard et al. 2014). In addition to enzymes from hydrolases, transferases and isomerases, it contains also some non-enzymatic proteins involved in amino acid transport (Janecek et al. 1997; Gabrisko and Janecek 2009).

The members of the α-amylase family GH13 (Matsuura et al. 1984; Svensson 1994; Kuriki and Imanaka 1999; MacGregor et al. 2001; Janecek 2002; van der Maarel et al. 2002; Janecek et al. 2014) employ a retaining reaction mechanism, share 4–7 conserved sequence regions (CSRs) and adopt a TIM-barrel domain with the GH13 catalytic machinery, i.e. the Asp206, Glu230 and Asp297 (*Aspergillus oryzae* α-amylase numbering) acting as catalytic nucleophile (CSR-II, strand β4), catalytic proton donor (CSR-III, strand β5) and transition-state stabiliser (CSR-IV, strand β7). Currently, the family is divided into 42 GH13 subfamilies (Stam et al. 2006), for which unique features in their amino acid sequences can be identified to be responsible for their individual specificities (Oslancova and Janecek 2002; Majzlova et al. 2013; Janecek et al. 2015; Kuchtova and Janecek 2016). Recently, a remarkable group of bacterial amylolytic enzymes has been described to possess aberrant catalytic machinery (Sarian et al 2017). The family GH13 forms, together with families GH70 and GH77, the clan GH-H; a remote homology of clan GH-H to family GH31, containing functionally related α-glucosidases, being also already suggested (Janecek et al. 2007).

In comparison with GH13, the second α-amylase family GH57 is a smaller family with only about 2000 members and less than 10 enzyme specificities (Lombard et al. 2014; Martinovicova and Janecek 2018). Although it also employs a retaining reaction mechanism (Palomo et al. 2011), it exhibits its own 5 CSRs and catalytic machinery different from GH13 within an incomplete TIM-barrel catalytic domain fold (Palomo et al. 2011; Imamura et al. 2003; Zona et al. 2004; Janecek and Blesak 2011; Blesak and Janecek 2012, 2013; Martinovicova and Janecek 2018). The individual enzyme specificities also possess unique features within their sequence logos that can be used as sequence-specificity fingerprints (Blesak and Janecek 2012; Martinovicova and Janecek 2018). A previous in silico analysis has suggested (Janecek and Kuchtova 2012) that the third α-amylase family GH119 should share with the family GH57 the CSRs, catalytic machinery and fold of the catalytic domain.

The family GH57 was established in 1996 (Henrissat and Bairoch 1996) based on the existence of two "amylases" originating from thermophilic bacterium *Dictyoglomus thermophilum* (Fukusumi et al. 1988) and hyperthermophilic archaeon *Pyrococcus furiosus* (Laderman et al. 1993b). They, although considered to be real α-amylases, were both sequentially different from sequences of the main α-amylase family GH13 members (Janecek 1998). Actually, they have been re-classified (Laderman et al. 1993a; Nakajima et al. 2004; Kaila et al. 2019) and thus become well-known as 4-α-glucanotransferases (4AGTs). Currently, the α-amylase family GH57 contains, in addition to α-amylase (Janecek et al. 2014), the specificities of 4AGT, amylopullulanase, α-glucan branching enzyme (GBE), dual-specificity of amylopullulanase–cyclomaltodextrinase, α-galactosidase, non-specified amylase, and maltogenic amylases (Comfort et al. 2008; Blesak and Janecek 2013; Jeon et al. 2014; Jung et al. 2014; Park et al. 2014; Martinovicova and Janecek 2018); the last one being still included only among the so-called non-classified sequences of the CAZy database (Lombard et al. 2014). Interestingly, the α-amylase specificity, best represented by the enzyme from *Methanocaldococcus jannaschii* (Bult et al. 1996), may rather be an amylopullulanase (Janecek et al. 2014), since that α-amylase was shown to be able to degrade not only starch but also a pullulan (Kim et al. 2001). With regard to solved tertiary structures, the first one determined by the X-ray crystallography was that of 4AGT from *Thermococcus litoralis* (Imamura et al. 2003), followed by the structures of the GBE from *Thermotoga maritima* (Dickmanns et al. 2006)—a bifunctional enzyme possessing also the α-amylase activity (Blesak and Janecek 2012; Zhang et al. 2019)—and GBEs from *Thermus thermophilus* (Palomo et al. 2011), *Thermococcus kodakarensis* (Santos et al. 2011) and *Pyrococcus horikoshii* (Na et al. 2017) as well as the maltogenic amylase from *Pyrococcus* sp. ST04 (Park et al. 2014). Of more than 2500 sequences available in the family GH57 (Lombard et al. 2014), only 27 members have already been characterised as real enzymes (Martinovicova and Janecek 2018).

As mentioned above, the catalytic domain consists of the so-called incomplete TIM-barrel, which is a $(\beta/\alpha)_7$-barrel domain, succeeded by a bundle of usually 3–4 α-helixes (Imamura et al. 2003; Dickmanns et al. 2006; Palomo et al. 2011; Santos et al. 2011; Park et al. 2014; Na et al. 2017), although some GH57 enzymes possess additional domains (Blesak and Janecek 2012, 2013). The catalytic machinery comprises the catalytic nucleophile—glutamic acid (Glu123 in the 4AGT from *T. litoralis*) and the catalytic proton donor—aspartic acid (Asp214), the former being positioned on the strand β4 (CSR-3) whereas the latter being placed on the strand β7 of the catalytic incomplete TIM-barrel (Imamura et al. 2003; Zona et al. 2004). The 5 CSRs were defined in 2004 (Zona et al. 2004), the first four of which (CSR-1–CSR-4) are located within the $(\beta/\alpha)_7$-barrel on strands β1, β3, β4 and β7, while the last fifth region CSR-5 is positioned within the α-helical bundle that, in fact, forms with the barrel the entire catalytic area of the family GH57 members (Blesak and Janecek 2012).

In the α-amylase family GH57, there has also been revealed a group of hypothetical proteins with sequences very closely related to those of the specificity of α-amylase, the so-called α-amylase-like proteins (Janecek and Blesak 2011). They may represent potential non-enzymatic members, somehow similar to those present in the main α-amylase family GH13, known as the rBAT and 4F2hc-antigen transport proteins that resemble the α-glucosidases from the oligo-1,6-glucosidase subfamily in both sequence

and structure (Janecek et al. 1997; Fort et al. 2007; Gabrisko and Janecek 2009; Chillaron et al. 2010; Lee et al. 2019; Yan et al. 2019). In spite of their function obviously unrelated to amylolysis, the two groups have been assigned even their GH13 subfamily numbers, i.e. the GH13_34 for 4F2hc antigens and GH13_35 from rBAT proteins (Stam et al. 2006; Lombard et al. 2014; Janecek and Gabrisko 2016). For the family GH57 α-amylase-like proteins, a substitution in one or both catalytic residues is typical, although it has still not been revealed whether or not they exhibit any enzymatic and/or a protein function (Janecek and Blesak 2011; Martinovicova and Janecek 2018). From a taxonomic point of view, they originate mainly from the two bacterial genera of *Bacteroides* and *Prevotella* (Janecek and Blesak 2011).

The main goal of the present bioinformatics study was to deliver another story of protein homologues closely related to further two GH57 specificities—GBE and 4AGT. These so-called GBE-like and 4AGT-like proteins lack, similar to α-amylase-like proteins described previously (Janecek and Blesak 2011), one or both catalytic residues. Concerning the sequence, however, they are most similar to their "template" enzyme specificities. This in silico analysis could contribute to a better understanding and a more detailed knowledge of these "like" groups of proteins present in the α-amylase family GH57. It may thus indicate the way leading to predicting their possible functions and, eventually, to shortening the time necessary for experimental revealing their activities.

Materials and methods

Sequence collection and comparison

Sequences were extracted from our recent in silico analysis of the entire family GH57 comparing in total 1602 GH57 members (Martinovicova and Janecek 2018). From that pool of GH57 sequences, 55 GBE-like and 61 4AGT-like proteins that lacked one or both catalytic residues were collected; both groups of hypothetical proteins being completed by four and six experimentally characterised GBEs and 4AGTs, respectively (Fukusumi et al. 1988; Laderman et al. 1993a, b; Jeon et al. 1997; Tachibana et al. 1997; Imamura et al. 2003; Nakajima et al. 2004; Ballschmiter et al. 2006; Dickmanns et al. 2006; Murakami et al. 2006; Labes and Schonheit 2007; Palomo et al. 2011; Santos et al. 2011; Park et al. 2014; Paul et al. 2015; Na et al. 2017; Kaila and Guptasarma 2019; Kaila et al. 2019; Pang et al. 2019; Zhang et al. 2019). In order to make a relevant conclusion, 21 biochemically characterised enzymes representing the remaining specificities from the family GH57 were added to give a preliminary set of 147 studied sequences. For further detailed analysis, the numbers of GBE-like and 4AGT-like proteins were reduced from 55 to 13 and 61 to 18, respectively, in order to take into account only the sequences originated from unique organisms. This means that the identical GBE-like and 4AGT-like proteins from *Treponema pallidum* were eliminated and the final set consisted of 62 sequences (Fig. S1).

All studied sequences were retrieved from GenBank (Benson et al. 2018) and/or UniProt (The UniProt Consortium 2017) sequence databases. The sequence alignment of all 147 GH57 sequences was performed using the program Clustal-Omega (https://www.ebi.ac.uk/Tools/msa/clustalo/; Sievers et al. 2011) with default parameters. The alignment spanned the substantial part of the family GH57 catalytic area including both the incomplete TIM-barrel and the bundle of 3–4 α-helices, exactly from the beginning of the CSR-1 to the end of CSR-5. The alignment was improved with some manual tuning in order to maximize similarities mainly with regard to correct correspondences within the five established CSRs (Zona et al. 2004).

Evolutionary analysis

The evolutionary tree was calculated as a maximum-likelihood tree (Jones et al. 1992) using the MEGA software (https://www.megasoftware.net/; Kumar et al. 2018) applying default programme parameters and bootstrapping procedure (the number of bootstrap trials used was 500). The tree was displayed with the program iTOL (https://itol.embl.de/; Letunic and Bork 2011).

Structure comparison

Three-dimensional structures were retrieved from the Protein Data Bank (PDB; Rose et al. 2015) for representatives of the family GH57, i.e. the GBE from *T. kodakarensis* (PDB code: 3N92; Santos et al. 2011) and the 4AGT from *T. litoralis* (PDB code: 1K1Y; Imamura et al. 2003). For making the three-dimensional models of GBE-like and 4AGT-like proteins, sequences of *Planctomyces* sp. SH-PL62 (UniProt Acc. No.: A0A142YKP4) and *Treponema pedis* (UniProt Acc. No.: S6A1V1) were selected as representatives of the former and latter group, respectively. The structural models were created with the Phyre2 server (https://www.sbg.bio.ic.ac.uk/phyre2/; Kelley and Sternberg 2009). The structures were superimposed using the program MultiProt (https://bioinfo3d.cs.tau.ac.il/MultiProt/; Shatsky et al. 2004) and the overlaps were displayed with the program WebLabViewerLite (Molecular Simulations, Inc.).

Results and discussion

Sequence comparison

The present in silico analysis represents a follow-up study of the original report on the newly identified group of the so-called α-amylase-like proteins from the α-amylase family GH57 that lack one or both catalytic residues although overall they exhibit an unambiguously close sequence similarity to its enzymatic counterpart (Janecek and Blesak 2011). In total, this study delivers 13 and 18 such "like" proteins (Fig. S1) from unique sources found in the evolutionary tree of the entire family GH57 (Martinovicova and Janecek 2018) in a close proximity to specificities of GBEs and 4AGTs, respectively. Interestingly, the taxonomical spectrum of source organisms of these "like" proteins is rather limited to a few genera in both cases, the genus *Treponema* being found dominant (Fig. S1). The alignment of all 62 selected sequences, i.e. 13 GBE-like and 18 4AGT-like proteins as well as 4 and 6 experimentally characterised GBEs and 4AGTs, respectively, completed with 21 remaining characterized GH57, has confirmed the extremely high sequence divergence typical for the family GH57 (Zona et al. 2004; Martinovicova and Janecek 2018), since the sequence section spanning the segment from the beginning of the CSR-1 to the end of CSR-5 reached the consensus length of 857 positions (Fig. S1). It is of note that it covers the so-called GH57 catalytic area (Janecek and Blesak 2011; Blesak and Janecek 2012, 2013) including both the incomplete TIM-barrel and the bundle of 3–4 α-helices, counting 458–463 residues for GBEs and 348–356 residues for 4AGTs. In order to focus on the so-called family GH57 sequence fingerprints (Blesak and Janecek 2013) bearing also the catalytic machinery, Fig. 1 shows the five CSRs well-established in the family (Zona et al. 2004) of all 62 studied sequences. It is evident that in all "like" proteins, there are substitutions present in either or in both catalytic residues.

With regard to GBE-like proteins, the catalytic nucleophile, the glutamic acid in CSR-3 is replaced by serine, tyrosine, phenylalanine, alanine, asparagine and even glycine, whereas the proton donor, the aspartic acid in CSR-4 is exchanged mostly by proline although also serine and asparagines are found there (Figs. 1 and S1). As far as the 4AGT-like proteins are concerned, the two catalytic residues are substituted predominantly by alanine and proline, respectively, but there are also arginine, asparagine, phenylalanine, tyrosine, histidine serine and glycine in the place of the catalytic nucleophile and phenylalanine, histidine, glutamic acid, asparagine, valine, serine and alanine replacing the proton donor (Figs. 1 and S1).

In addition to missing catalytic residues in CSR-3 and CSR-4, the former region seems to have been well conserved but the latter region has obviously undergone a more dramatic change in both "like" proteins. This is especially clear at the end of the CSR-4, which in both cases was defined as the region containing residues of sequence fingerprints of a given GH57 enzyme specificity (Janecek and Blesak 2011; Blesak and Janecek 2012, 2013). Thus in GBE-like proteins, the His-Trp signature unique for GBEs (Blesak and Janecek 2012) is no more seen; the two residues being replaced by various residues mostly by Val-Ala and/or combination of lysine and arginine (Fig. 1). The cysteine residue succeeding the catalytic nucleophile in CSR-3, typical for GBEs (Blesak and Janecek 2012; Martinovicova and Janecek 2018; Zhang et al. 2019) is also absent in GBE-like proteins (Fig. 1a). Concerning the 4AGT-like proteins, the tryptophan at the end of CSR-4 conserved in 4AGTs invariantly (Blesak and Janecek 2012) is substituted by different residues without any special preference; the only exception is represented by the "like" protein originating from *Treponema azonutricium* (Fig. 1).

Obviously, the protein chain segment spanning the two regions bearing the catalytic machinery, i.e. from CSR-3 to CSR-4, is longer for GBE-like proteins in comparison with 4AGT-like proteins (Fig. S1). This feature is, however, consistent with the known domain arrangement and overall sequence-structural organization within this part of both parent GBE and 4AGT enzymes (Imamura et al. 2003; Dickmanns et al. 2006; Palomo et al. 2011; Santos et al. 2011; Blesak and Janecek 2012; Na et al. 2017).

Evolutionary relationships

The evolutionary relationships of both the GBE-like and 4AGT-like proteins within the family GH57 are depicted in the evolutionary tree (Fig. 2). In the most recent and complete in silico analysis of this family comparing 1602 sequences (Martinovicova and Janecek 2018), four groups of the "like" proteins were identified in total—in addition to GBE-like and 4AGT-like proteins described in the present work, there were also the α-amylase-like and amylopullulanase–cyclomaltodextrinase-like proteins. Whereas the former group has already been well-known from previous studies (Janecek and Blesak 2011; Blesak and Janecek 2012, 2013), the latter one consisted of 5 proteins only (Martinovicova and Janecek 2018). Therefore, the main emphasis has been focused on the groups of GBE-like (55 members) and 4AGT-like (61 members) proteins.

Overall view of the evolutionary tree (Fig. 2) clearly indicates the relatedness of selected GBE-like and 4AGT-like proteins to their enzymatic counterparts, i.e. GBEs and 4AGTs. Obviously, the two major subgroups of each of the two "like" protein groups, originating solely from the genus *Treponema* (for GBE-like proteins) or mostly

No.	Protein	CSR-1	CSR-2	CSR-3	CSR-4	CSR-5
1	GBE_AHD18669_B_*Thermotoga_maritima*	HAHLP	GKLEIV	WLA E CG	PF D AELFGHW	AQSSDWAFI
2	GBE_BAD71725_B_*Thermus_thermophilus*	HAHLP	GQVELI	WLP E MA	PY D AELFGHW	LEASDWPFL
3	GBE_BAA30492_A_*Pyrococcus_horikoshii*	HTHIP	GYVEVI	WLP E CA	PY D TELFGHW	LEASDWQFL
4	GBE_BAD85625_A_*Thermococcus_kodakarensis*	HTHIP	GYVEVI	WLP E CA	PY D TELFGHW	LEASDWQFL
5	GBE_like_ACL23066_B_*Chloroflexus_aggregans*	EVHTP	EIVVPL	WLP N RA	PI D TRLWGSR	AEQSDWIDA
6	GBE_like_ABY33308_B_*Chloroflexus_aurantiacus*	EVHVP	GLITPL	WLP G SA	PI D ARLWSRN	AEMSDWLME
7	GBE_like_ACM51503_B_*Chloroflexus_sp_ATCC29364*	EVHVP	GLITPL	WLP G SA	PI D ARLWSRN	AEMSDWLME
8	GBE_like_AMV39746_B_*Planctomyces_sp_SH_PL62*	ELHHP	GAIDLV	WLP F LG	AL S AHDLARA	AQQVDWSYP
9	GBE_like_AEE17013_B_*Treponema_brennaborense*	VAHQG	GNVELL	WLP Y MG	AF D AKILGQS	AQSGDWPSM
10	GBE_like_AEJ19199_B_*Treponema_caldarium*	HAHVP	GKIELL	WLP E AG	AY N ADLFGHR	AQSADWAAI
11	GBE_like_AAS11583_B_*Treponema_denticola*	DAHLP	GYVELL	WIP S MA	II S SDFFGKK	LQSVYWPLY
12	GBE_like_AOF64560_B_*Treponema_pallidum*	DCNLP	GSIELL	YLP E LG	VF P ASLFGVA	CQSLFWPLL
13	GBE_like_AEH40301_B_*Treponema_paraluiscuniculi*	DCNLP	GSIELL	YLP E LG	VF P ASLFGVA	CQSLFWPLL
14	GBE_like_AGT43601_B_*Treponema_pedis*	DAHLP	GYIEIL	WLP A LA	IC S SDFLGKR	MQAFYWPFY
15	GBE_like_AIN94132_B_*Treponema_putidum*	DAHLP	GYVELL	WLP S MA	TI S SDFFGKK	LQSMYWPLY
16	GBE_like_AIW90327_B_*Treponema_sp_OMZ_838*	HAHVP	GCIELL	WLP S MG	VL P LRFLGKT	AQSTDWPLM
17	GBE_like_AEB14373_B_*Treponema_succinifaciens*	EANQG	GIIELI	WLP Y MG	AI P AEFLGQK	AQSGDWPAM
18	4AGT_CAA30735_B_*Dictyoglomus_thermophilum*	HNHQP	GQIEFV	WLA E RV	FD D GEKFGLW	GQANDAYWH
19	4AGT_ABW95829_A_*Archaeoglobus_fulgidus*	HNHQP	GQIEIV	WLT E RV	HD D GEKFGVW	AQCNDAYWH
20	4AGT_AAL80396_A_*Pyrococcus_furiosus*	HNHQP	GQVEIV	WLT E RV	HD D GEKFGIW	AQCNDAYWH
21	4AGT_BAA22062_A_*Thermococcus_kodakarensis*	HNHQP	GQLEIV	WLT E RV	HD D GEKFGVW	AQCNDAYWH
22	4AGT_BAA22063_A_*Thermococcus_litoralis*	HNHQP	GQLEIV	WLT E RV	HD D GEKFGVW	AQCNDAYWH
23	4AGT_ACJ17206_A_*Thermococcus_onnurineus*	HNHQP	GQLEIV	WLT E RV	HD D GEKFGVW	AQCNDAYWH
24	4AGT_like_ADL24659_B_*Fibrobacter_succinogenes*	SPSTS	GVLEFL	FNS S LV	GK S AIALLDI	PAMAPLYFR
25	4AGT_like_AEB12171_B_*Marinithermus_hydrothermalis*	HHHQP	NRVEVL	WLP E RV	PP E GLWVAAV	AQAADAFWP
26	4AGT_like_AEC01764_B_*Sphaerochaeta_coccoides*	YSQIP	EQLEFL	WCY N QI	NI D QLCQGAA	ASGSAAYIR
27	4AGT_like_ADY13165_B_*Sphaerochaeta_globosa*	YSQLS	GKVELL	FCY A QA	NL D QLMQGGI	VSTGTVYLC
28	4AGT_like_AEV28320_B_*Sphaerochaeta_pleomorpha*	YSQIS	GRVELI	WCY G EF	NI D QLCQGDI	IASGSPYVC
29	4AGT_like_AFG37140_B_*Spirochaeta_africana*	YNSLP	RQIEVI	WIP E MV	II P GENASDQ	GQHHSAYWH
30	4AGT_like_ADN02191_B_*Spirochaeta_thermophila*	VLSLP	KQVEIL	YLP F GI	ED F ASWSTLS	GQTGAAYSI
31	4AGT_like_AEF81655_B_*Treponema_azotonutricium*	HGHLP	KQAELL	WIP R GC	AF E SLRTALW	AQDSHLFDH
32	4AGT_like_AEE16801_B_*Treponema_brennaborense*	GNEYV	KQIELF	RLP H DA	EH A AVCCRFT	AQFGGAYVP
33	4AGT_like_AEJ19428_B_*Treponema_caldarium*	HHHVP	KQIELL	WLP G MF	LI N LESVGPE	AQGIDMFYM
34	4AGT_like_AAS11912_B_*Treponema_denticola*	HADYN	KQVEIL	YLP Y FA	AS V VTETSVV	AESGILFNL
35	4AGT_like_AEZ59409_B_*Treponema_pallidum*	KLTGA	KRLELL	FLE A SA	ML P LDCYQKL	AEQGVFFIS
36	4AGT_like_AEH40102_B_*Treponema_paraluiscuniculi*	KLTGA	KRLELL	FLE A SA	ML P LDCYQKL	AEQGVFFIS
37	4AGT_like_AGT44923_B_*Treponema_pedis*	HTDYS	KQIEII	YLP Y FA	AS S VTENSIV	AESGILFNL
38	4AGT_like_AEF86734_B_*Treponema_primitia*	HNHIP	KQVEIL	WIP G AA	QL H PEDPGET	AQGYSLFCD
39	4AGT_like_AIN94328_B_*Treponema_putidum*	HADYN	KQIEIL	YLP Y FA	AS V VTETSVV	AESGILFNL
40	4AGT_like_AIW90418_B_*Treponema_sp_OMZ_838*	QLSIS	KQIELL	FVT A SA	FL P IKTYIKC	AQNAELFAL
41	4AGT_like_AEB14246_B_*Treponema_succinifaciens*	NADLE	RQIEII	TLF G SI	PF S DEFFEPV	AQDGTNYLS
42	AAMY_AAB99631_A_*Methanocaldococcus_jannaschii*	EVHQP	GNVELI	RNT E LI	YM D YETFGEH	LQTSDNLYY
43	AAMY_like_AAO79410_B_*Bacteroides_thetaiotaomicron*	EIHQI	GCCEFL	RNS S LI	FM E LSALGMA	LQASNNFRF
44	AGAL_AAG28455_A_*Pyrococcus_furiosus*	HGNLQ	GLIEIL	WLP E LA	GT D IEFLGYR	AENSDARGW
45	NSA_AEC23345_B_Uncultured_bacterium	HGHFY	GHGNAI	WLP E TA	AT D GETYGHH	MYTSCGWFF
46	MGA_AAL80994_A_*Pyrococcus_furiosus*	HAYQP	TYVEPV	WLP E NV	SS D LESLVAN	ANHSCPRFW
47	MGA_AFK22464_A_*Pyrococcus_sp_ST04*	HAYQP	THVEPT	WLP E NV	SS D LESLVAN	ANHSCPRFW
48	MGA_AFL95073_A_*Thermococcus_cleftensis*	HAYQP	QRFDAV	WLP E AV	AS D LESLLGN	ANHSCPRFW
49	APU_ACK41960_B_*Dictyoglomus_turgidum*	HNHQP	KQIEVT	WPS E GS	IL D GENCWEY	AEGSDWFWW
50	APU_ADN02534_B_*Spirochaeta_thermophila*	HQHQP	GQIEVI	WPA E GS	IL D GENAWEH	AEGSDWFWW
51	APU_ABA33719_A_*Pyrococcus_furiosus*	HQHQP	GNVEVT	WAA E SA	TL D GENPWEH	AEASDWFWW
52	APU_AEH25371_A_*Pyrococcus_yayanosii*	HQHQP	GNVEVT	WAA E SA	TL D GENPWEH	AEASDWFWW
53	APU_AAD28552_A_*Thermococcus_hydrothermalis*	HQHQP	GNVEVT	WAA E SA	TL D GENPVEN	AEASDWFWW
54	APU_BAD85963_A_*Thermococcus_kodakarensis*	HQHQP	GNVEVT	WAA E SA	TL D GENPWEH	AEASDWFWW
55	APU_BAC10983_A_*Thermococcus_litoralis*	HQHQP	GNVEVT	WAA E SA	TL D GENPWEH	AEASDWFWW
56	APU_ACJ03924_A_*Thermococcus_sp_HJ21*	HQHQP	GNVEVT	WAA E SA	TL D GENPWEH	AEASDWFWW
57	APU_CMD_ABW02197_A_*Caldivirga_maquilingensis*	HMHQP	GQLDVL	WTP E MA	AL D GENFIAM	ALDSDFWWA
58	APU_CMD_ACL10679_A_*Desulfurococcus_amylolyticus*	HHHQA	GQIDVL	WTP E MA	AL D GENWMVF	ALDSDYWWA
59	APU_CMD_ABN70497_A_*Staphylothermus_marinus*	HHHQA	GQIDVL	WTP E MA	AL D GENWMSF	ALDSDYWWA
60	APU_CMD_AAY80509_A_*Sulfolobus_acidocaldarius*	NMHQP	GKVEVL	WTP E QA	AF D GENPLIF	AEGSDWTWQ
61	APU_CMD_AAK41420_A_*Sulfolobus_solfataricus*	NMHQP	GKVDVL	WTP E MA	AL D GENPLIF	AEGSDWTWQ
62	APU_CMD_ABL77620_A_*Thermofilum_pendens*	HHHQA	GIVEPL	WTP E MF	AL D GENWMIM	AIDSDFYWY

Fig. 1 Conserved sequence regions of family GH57 members focused on GBE-like and 4AGT-like proteins. The individual groups are distinguished from each other by different colours. The label of the protein source consists of: (i) the abbreviated name of the enzyme/protein; (ii) GenBank accession number; (iii) letter “A” or “B” for archeons and bacteria, respectively; and (iv) the name of the organism. The GH57 enzyme specificities, in addition to GBE (α-glucan branching enzyme) and 4AGT (4-α-glucanotransferase), are abbreviated as follows: *AAMY* α-amylase, *AAMY like* α-amylase-like protein, *AGAL* α-galactosidase, *NSA* non-specified amylase, *MGA* maltogenic amylase, *APU* amylopullulanase, *APU–CMD* amylopullulanase–cyclomaltodextrinase. The catalytic residues, i.e. a glutamic acid in CSR-3 (the catalytic nucleophile) and an aspartic acid in CSR-4 (the proton donor) are boxed. Note that the GBE-like and 4AGT-like proteins lack either one or both catalytic residues. Colour code for the individual residues: Trp, Phe, Tyr: yellow; Asp: blue, Glu: red; Val, Leu, Ile: green; Arg, Lys: cyan; His: brown; Cys: magenta; Gly, Pro: black

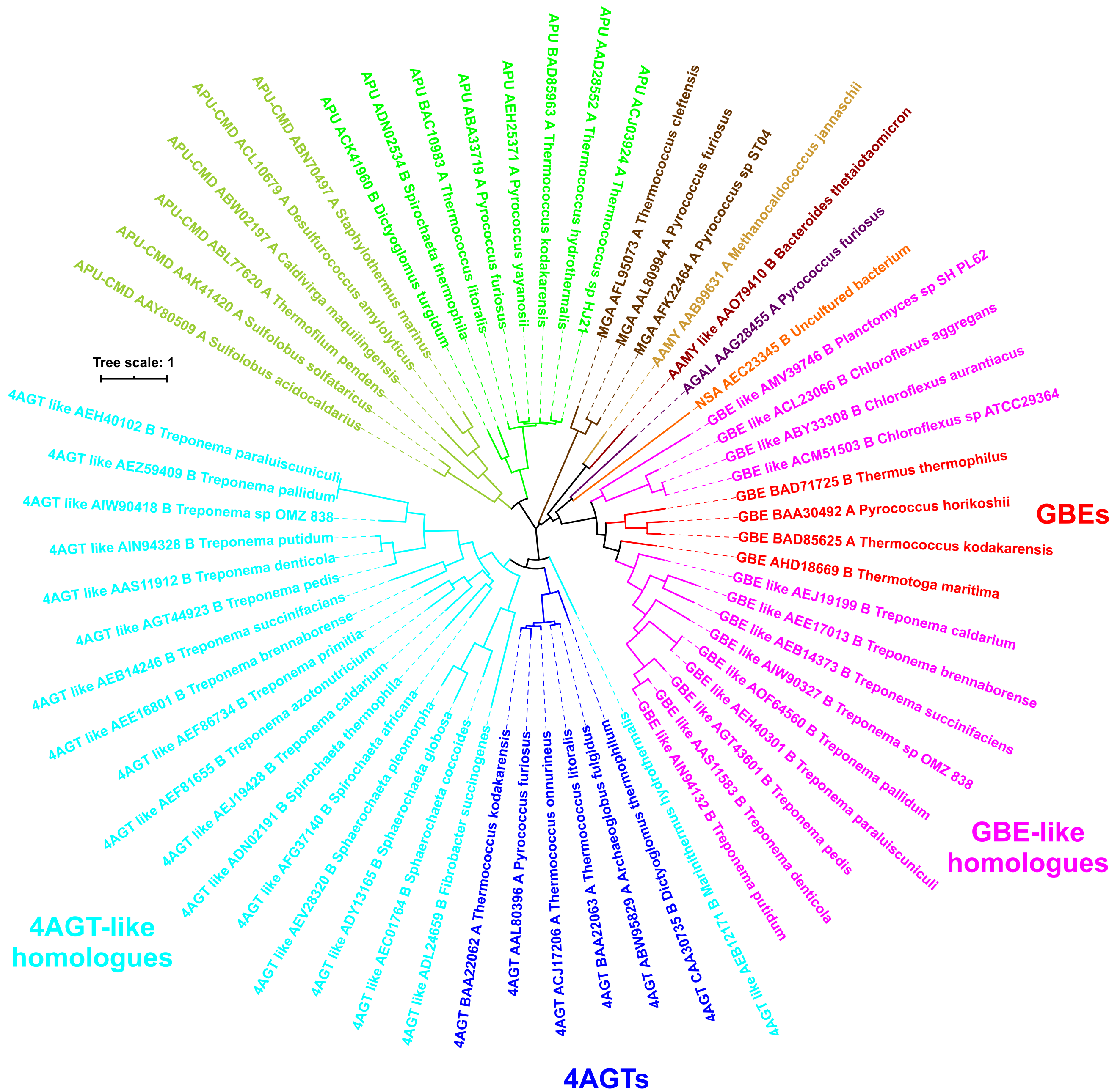

Fig. 2 Evolutionary tree of family GH57 members focused on GBE-like and 4AGT-like proteins. The tree is based on the alignment of sequences spanning the segments from beginning of CSR-1 to the end of CSR-5 (as presented in Fig. S1). The individual groups are distinguished from each other by different colours. The label of the protein source consists of: (i) the abbreviated name of the enzyme/protein; (ii) GenBank accession number; (iii) letter "A" or "B" for archeons and bacteria, respectively; and (iv) the name of the organism. The GH57 enzyme specificities, in addition to GBE (α-glucan branching enzyme) and 4AGT (4-α-glucanotransferase), are abbreviated as follows: *AAMY* α-amylase, *AAMY like* α-amylase-like protein, *AGAL* α-galactosidase, *NSA* non-specified amylase, *MGA* maltogenic amylase, *APU* amylopullulanase, *APU–CMD* amylopullulanase–cyclomaltodextrinase

from *Treponema* accompanied by a few sequences from *Spirochaeta* (for 4AGT-like proteins), form their own taxonomic clusters; the remaining representatives of each

"like" group being placed a bit separately but still together with real GBEs and real 4AGTs (Fig. 2).

It is worth mentioning that all identified members of GBE-like and 4AGT-like proteins are of bacterial origin only, whereas GBEs and 4AGTs have been found in both *Bacteria* and *Archaea* (Martinovicova and Janecek 2018). Even if only the experimentally characterized GBEs and 4AGTs are considered, they originate from prokaryotes as follows (Fig. 2): (i) GBEs—two from bacteria (*T. maritima* and *T. thermophilus*) and two from archaeons (*Pyrococcus horikoshi* and *T. kodakarensis*); and (ii) 4AGTs—one from bacteria (*D. thermophilum*) and four from archaeons (*Archaeoglobus fulgidus*, *P. furiosus*, *T. kodakarensis*, *T. litoralis* and *Thermococcus onnurineus*).

With regard to rather limited spectrum of source organisms of these GBE-like and 4AGT-like proteins (Fig. S1), currently, it is not possible to explain it unambiguously because none of them has been cloned, expressed, and biochemically characterized. It is thus not possible to have a clue concerning, e.g., their properties. The taxonomic spectrum limited to a few organisms—mostly *Treponema*, *Spirochaeta* and *Sphaerochaeta* (all spirochetes)—seems thus to be a puzzle, too. It might be of interest that in the related family GH77 amylomaltases, those originated from borreliae have been revealed as possessing some unique amino acid substitutions in a very specific positions that are otherwise well-conserved in amylomaltases from non-borreliae sources (Godany et al. 2008; Kuchtova and Janecek 2015). In addition, for example, for the specificity of α-amylase, their "like" protein counterparts originate from both Archaea and Bacteria (Janecek and Blesak 2011), whereas, for the GBEs and 4AGTs, their "like" protein homologues have been found only in Bacteria (Fig. 2). Moreover, in the main α-amylase family GH13 (Janecek et al. 2014), which is also a polyspecific GH family (Lombard et al. 2014), there are not known such "like" proteins that would exist as counterparts to various enzyme specificities, like those recognized in the family GH57.

Structure comparison

The structural analysis has been focused on the GBE-like protein from *Planctomyces* sp. SH-PL62 (UniProt Acc. No.: A0A142YKP4) and the 4AGT-like protein from *T. pedis* (UniProt Acc. No.: S6A1V1). Despite the fact that they lack the catalytic machinery of the α-amylase family GH57 (Fig. 1), the models of their three-dimensional structures have been obtained unambiguously according to respective templates—the experimental structures of GBEs (Dickmanns et al. 2006; Palomo et al. 2011; Santos et al. 2011; Na et al. 2017) and 4AGT (Imamura et al. 2003) with adequate (no less than 95%) alignment coverage, respecting also the presence of the additional C-terminal β-stranded domain of *T. litoralis* 4AGT (Imamura et al. 2003). The details from the structural comparison are illustrated in Fig. 3. It is evident that each pair of the GH57 enzyme and its "like" counterpart shares the main structural features. This phenomenon was first observed in the family GH57 for the specificity of α-amylase and its α-amylase-like homologues (Janecek and Blesak 2011) and may be expected to be identified also for other, i.e. remaining enzyme specificities from the family (Martinovicova and Janecek 2018).

Based on the structural superimposition (Fig. 3), it is clear that both studied GBE-like and 4AGT-like proteins do not possess the catalytic residues confirmed for the family GH57. While the real GBE from *T. kodakarensis* (Santos et al. 2011) and 4AGT from *T. litoralis* (Imamura et al. 2003) have their catalytic machineries formed by the pairs of Glu123/Asp214 and Glu183/Asp354, respectively, at corresponding positions there are Phe163/Ser281 and Tyr127/Ser202 found in the structural models of GBE-like protein from *Planctomyces* sp. SH-PL62 (Fig. 3d) and the 4AGT-like protein from *T. pedis* (Fig. 3h), respectively. This observation can be extended to all GBE-like and 4AGT-like proteins of this study, as documented by their sequence alignment (Figs. 1 and S1). Of course, some of the "like" homologues may contain either the catalytic nucleophile or the proton donor, but not both residues simultaneously, i.e. similar to the above-mentioned α-amylase-like homologues (Janecek and Blesak 2011).

Conclusions

The present in silico study delivers two novel groups of hypothetical proteins, the members of the α-amylase family GH57, which are closely related to GBEs and 4AGTs enzyme specificities. Since the putative proteins lack either one or both residues of the family GH57 catalytic machinery, they are suggested to define the so-called GBE-like and 4AGT-like proteins, similar to the α-amylase-like homologues established previously. Although the exact function of these "like" proteins has been not known as yet, the existence of members of GH families with incomplete catalytic machinery is not exceptional. For example, even in the main α-amylase family GH13, there is a well-known group of animal heavy-chains of heteromeric amino acid transport proteins rBAT and 4F2hc antigens lacking the catalytic machinery and amylolytic activity. In each case of the three already revealed family GH57 "like" proteins—α-amylase-like homologues (identified previously) as well as GBE-like and 4AGT-like (the present study)—the experimental evidence concerning their eventual enzyme specificity or at least their exact protein activity would be of a special interest. The presented in silico analysis could accelerate

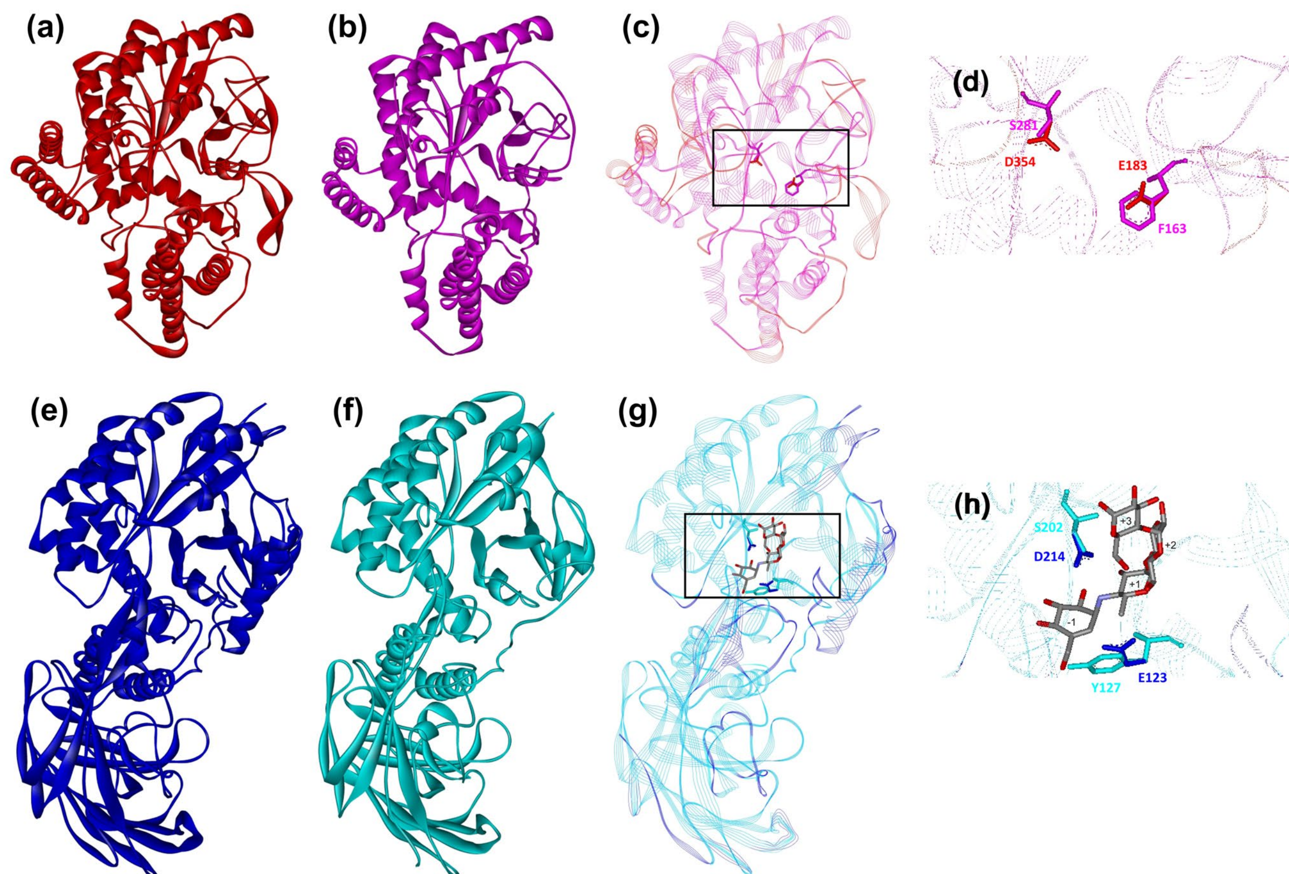

Fig. 3 Structure comparison of family GH57 α-glucan branching enzyme and 4-α-glucanotransferase with their GBE-like and 4AGT-like proteins. **a** *T. kodakarensis* α-glucan branching enzyme (PDB code: 3N92); **b** *Planctomyces* sp. SH-PL62 GBE-like protein (model); **c** their superimposition (444 Cα atoms; RMSD 0.00 Å; residues 4–453 aligned, i.e. 98% coverage, with 24% identity); **d** a close-up focused on catalytic residues of the α-glucan branching enzyme (red) and their eventual correspondences in the GBE-like protein (magenta); **e** *Thermococcus litoralis* 4-α-glucanotransferase (PDB code: 1K1Y); **f** *Treponema pedis* 4AGT-like protein (model); (g) their superimposition (551 Cα atoms; RMSD 0.00 Å; residues 7–605 aligned, i.e. 95% coverage, with 21% identity); **h** a close-up focused on catalytic residues of the 4-α-glucanotransferase (blue) and their eventual correspondences in the 4AGT-like protein (cyan) with bound acarbose occupying the subsites from − 1 to + 3

the efforts focused on experimental studies of these unique proteins.

Acknowledgements This work was financially supported by the Grant No. 2/0146/17 from the Slovak Grant Agency VEGA.

Compliance with ethical standards

Conflict of interest The authors declare no conflict of interests.

References

Ballschmiter M, Fütterer O, Liebl W (2006) Identification and characterization of a novel intracellular alkaline α-amylase from the hyperthermophilic bacterium *Thermotoga maritima* MSB8. Appl Environ Microbiol 72:2206–2211. https://doi.org/10.1128/AEM.72.3.2206-2211.2006

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Ostell J, Pruitt KD, Sayers EW (2018) GenBank. Nucleic Acids Res 46:D41–D47. https://doi.org/10.1093/nar/gkx1094

Blesak K, Janecek S (2012) Sequence fingerprints of enzyme specificities from the glycoside hydrolase family GH57. Extremophiles 16:497–506. https://doi.org/10.1007/s00792-012-0449-9

Blesak K, Janecek S (2013) Two potentially novel amylolytic enzyme specificities in the prokaryotic glycoside hydrolase α-amylase family GH57. Microbiology 159:2584–2593. https://doi.org/10.1099/mic.0.071084-0

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG et al (1996) Complete genome sequence of the methanogenic archaeon *Methanococcus jannaschii*. Science 273:1058–1073. https://doi.org/10.1126/science.273.5278.1058

Chillaron J, Font-Llitjos M, Fort J, Zorzano A, Goldfarb DS, Nunes V, Palacin M (2010) Pathophysiology and treatment of cystinuria. Nat Rev Nephrol 6:424–434. https://doi.org/10.1038/nrneph.2010.69

Comfort DA, Chou CJ, Conners SB, VanFossen AL, Kelly RM (2008) Functional-genomics-based identification and characterization of open reading frames encoding α-glucoside-processing

enzymes in the hyperthermophilic archaeon *Pyrococcus furiosus*. Appl Environ Microbiol 74:1281–1283. https://doi.org/10.1128/AEM.01920-07

Dickmanns A, Ballschmiter M, Liebl W, Ficner R (2006) Structure of the novel α-amylase AmyC from *Thermotoga maritima*. Acta Crystallogr D Biol Crystallogr 62:262–270. https://doi.org/10.1107/S0907444905041363

Fort J, de la Ballina LR, Burghardt HE, Ferrer-Costa C, Turnay J, Ferrer-Orta C, Uson I, Zorzano A, Fernandez-Recio J, Orozco M, Lizarbe MA, Fita I, Palacin M (2007) The structure of human 4F2hc ectodomain provides a model for homodimerization and electrostatic interaction with plasma membrane. J Biol Chem 282:31444–31452. https://doi.org/10.1074/jbc.M704524200

Fukusumi S, Kamizono A, Horinouchi S, Beppu T (1988) Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic thermophile, *Dictyoglomus thermophilum*. Eur J Biochem 174:15–21. https://doi.org/10.1111/j.1432-1033.1988.tb14056.x

Gabrisko M, Janecek S (2009) Looking for the ancestry of the heavy-chain subunits of heteromeric amino acid transporters rBAT and 4F2hc within the GH13 α-amylase family. FEBS J 276:7265–7278. https://doi.org/10.1111/j.1742-4658.2009.07434.x

Godany A, Vidova B, Janecek S (2008) The unique glycoside hydrolase family 77 amylomaltase from *Borrelia burgdorferi* with only catalytic triad conserved. FEMS Microbiol Lett 284:84–91. https://doi.org/10.1111/j.1574-6968.2008.01191.x

Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem J 280:309–316. https://doi.org/10.1042/bj2800309

Henrissat B, Bairoch A (1996) Updating the sequence-based classification of glycosyl hydrolases. Biochem J 316:695–696. https://doi.org/10.1042/bj3160695

Imamura H, Fushinobu S, Yamamoto M, Kumasaka T, Jeon BS, Wakagi T, Matsuzawa H (2003) Crystal structures of 4-α-glucanotransferase from *Thermococcus litoralis* and its complex with an inhibitor. J Biol Chem 278:19378–19386. https://doi.org/10.1074/jbc.M213134200

Janecek S (1998) Sequence of archaeal *Methanococcus jannaschii* α-amylase contains features of families 13 and 57 of glycosyl hydrolases: a trace of their common ancestor? Folia Microbiol 43:123–128. https://doi.org/10.1007/BF02816496

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

Janecek S, Blesak K (2011) Sequence-structural features and evolutionary relationships of family GH57 α-amylases and their putative α-amylase-like homologues. Protein J 30:429–435. https://doi.org/10.1007/s10930-011-9348-7

Janecek S, Gabrisko M (2016) Remarkable evolutionary relatedness among the enzymes and proteins from the α-amylase family. Cell Mol Life Sci 73:2707–2725. https://doi.org/10.1007/s00018-016-2246-6

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. https://doi.org/10.1016/j.febslet.2012.07.020

Janecek S, Svensson B, Henrissat B (1997) Domain evolution in the α-amylase family. J Mol Evol 45:322–331. https://doi.org/10.1007/PL00006236

Janecek S, Svensson B, MacGregor EA (2007) A remote but significant sequence homology between glycoside hydrolase clan GH-H and family GH31. FEBS Lett 581:1261–1268. https://doi.org/10.1016/j.febslet.2007.02.036

Janecek S, Svensson B, MacGregor EA (2014) α-Amylase: an enzyme specificity found in various families of glycoside hydrolases. Cell Mol Life Sci 71:1149–1170. https://doi.org/10.1007/s00018-013-1388-z

Janecek S, Kuchtova A, Petrovicova S (2015) 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. Biologia 70:1284–1294. https://doi.org/10.1515/biolog-2015-0165

Jeon B, Taguchi H, Sakai H, Ohshima T, Wakagi T, Matsuzawa H (1997) 4-α-Glucanotransferase from the hyperthermophilic archaeon *Thermococcus litoralis*—enzyme purification and characterization, and gene cloning, sequencing and expression in *Escherichia coli*. Eur J Biochem 248:171–178. https://doi.org/10.1111/j.1432-1033.1997.00171.x

Jeon EJ, Jung JH, Seo DH, Jung DH, Holden JF, Park CS (2014) Bioinformatic and biochemical analysis of a novel maltose-forming α-amylase of the GH57 family in the hyperthermophilic archaeon *Thermococcus* sp. CL1. Enzyme Microb Technol 60:9–15. https://doi.org/10.1016/j.enzmictec.2014.03.009

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282. https://doi.org/10.1093/bioinformatics/8.3.275

Jung JH, Seo DH, Holden JF, Park CS (2014) Maltose-forming α-amylase from the hyperthermophilic archaeon *Pyrococcus* sp. ST04. Appl Microbiol Biotechnol 98:2121–2131. https://doi.org/10.1007/s00253-013-5068-6

Kaila P, Guptasarma P (2019) An ultra-stable glucanotransferase-cum-exoamylase from the hyperthermophile archaeon *Thermococcus onnurineus*. Arch Biochem Biophys 665:114–121. https://doi.org/10.1016/j.abb.2019.02.017

Kaila P, Mehta GS, Dhaunta N, Guptasarma P (2019) Structure-guided mutational evidence and postulates explaining how a glycohydrolase from *Pyrococcus furiosus* functions simultaneously as an amylase and as a 4-α-glucanotransferase. Biochem Biophys Res Commun 509:892–897. https://doi.org/10.1016/j.bbrc.2019.01.021

Kelley LA, Sternberg MJE (2009) Protein structure prediction on the Web: a case study using the Phyre server. Nat Protoc 4:363–371. https://doi.org/10.1038/nprot.2009.2

Kim JW, Flowers LO, Whiteley M, Peeples TL (2001) Biochemical confirmation and characterization of the family-57-like α-amylase of *Methanococcus jannaschii*. Folia Microbiol 46:467–473. https://doi.org/10.1007/BF02817988

Kuchtova A, Janecek S (2015) *In silico* analysis of family GH77 with focus on amylomaltases from borreliae and disproportionating enzymes DPE2 from plants and bacteria. Biochim Biophys Acta 1854:1260–1268. https://doi.org/10.1016/j.bbapap.2015.05.009

Kuchtova A, Janecek S (2016) Domain evolution in enzymes of the neopullulanase subfamily. Microbiology 162:2099–2115. https://doi.org/10.1099/mic.0.000390

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549. https://doi.org/10.1093/molbev/msy096

Kuriki T, Imanaka T (1999) The concept of the α-amylase family: structural similarity and common catalytic mechanism. J Biosci Bioeng 87:557–565. https://doi.org/10.1016/S1389-1723(99)80114-5

Labes A, Schonheit P (2007) Unusual starch degradation pathway via cyclodextrins in the hyperthermophilic sulfate-reducing archaeon *Archaeoglobus fulgidus* strain 7324. J Bacteriol 189:8901–8913. https://doi.org/10.1128/JB.01136-07

Laderman KA, Asada K, Uemori T, Mukai H, Taguchi Y, Kato I, Anfinsen CB (1993a) α-Amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. Cloning and sequencing of the gene and expression in *Escherichia coli*. J Biol Chem 268:24402–24407

Laderman KA, Davis BR, Krutzsch HC, Lewis MS, Griko YV, Privalov PL, Anfinsen CB (1993b) The purification and characterization of an extremely thermostable α-amylase from the

hyperthermophilic archaebacterium *Pyrococcus furiosus*. J Biol Chem 268:24394–24401

Lee Y, Wiriyasermkul P, Jin C, Quan L, Ohgaki R, Okuda S, Kusakizako T, Nishizawa T, Oda K, Ishitani R, Yokoyama T, Nakane T, Shirouzu M, Endou H, Nagamori S, Kanai Y, Nureki O (2019) Cryo-EM structure of the human L-type amino acid transporter 1 in complex with glycoprotein CD98hc. Nat Struct Mol Biol 26:510–517. https://doi.org/10.1038/s41594-019-0237-7

Letunic I, Bork P (2011) Interactive tree of life v2: online annotation and display of phylogenetic trees made easy. Nucleic Acids Res 39:W475–W478. https://doi.org/10.1093/nar/gkr201

Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B (2014) The carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178

MacGregor EA, 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. https://doi.org/10.1016/s0167-4838(00)00302-2

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. https://doi.org/10.1016/j.carres.2012.11.022

Martinovicova M, Janecek S (2018) In silico analysis of the α-amylase family GH57: eventual subfamilies reflecting enzyme specificities. Biotech 8:307. https://doi.org/10.1007/s13205-018-1325-9

Matsuura Y, Kusunoki M, Harada W, Kakudo M (1984) Structure and possible catalytic residues of Taka-amylase A. J Biochem 95:697–702. https://doi.org/10.1093/oxfordjournals.jbchem.a134659

Murakami T, Kanai T, Takata H, Kuriki T, Imanaka T (2006) A novel branching enzyme of the GH-57 family in the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. J Bacteriol 188:5915–5924. https://doi.org/10.1128/JB.00390-06

Na S, Park M, Jo I, Cha J, Ha NC (2017) Structural basis for the transglycosylase activity of a GH57-type glycogen branching enzyme from *Pyrococcus horikoshii*. Biochem Biophys Res Commun 484:850–856. https://doi.org/10.1016/j.bbrc.2017.02.002

Nakajima M, Imamura H, Shoun H, Horinouchi S, Wakagi T (2004) Transglycosylation activity of *Dictyoglomus thermophilum* amylase A. Biosci Biotechnol Biochem 68:2369–2373. https://doi.org/10.1271/bbb.68.2369

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. https://doi.org/10.1007/PL00012517

Palomo M, Pijning T, Booiman T, Dobruchowska JM, van der Vlist J, Kralj S, Planas A, Loos K, Kamerling JP, Dijkstra BW, van der Maarel MJ, 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. https://doi.org/10.1074/jbc.M110.179515

Pang B, Zhou L, Cui W, Liu Z, Zhou S, Xu J, Zhou Z (2019) A hyperthermostable type II pullulanase from a deep-sea microorganism *Pyrococcus yayanosii* CH1. J Agric Food Chem 67:9611–9617. https://doi.org/10.1021/acs.jafc.9b03376

Paul CJ, Leemhuis H, Dobruchowska JM, Grey C, Önnby L, van Leeuwen SS, Dijkhuizen L, Karlsson EN (2015) A GH57 4-α-glucanotransferase of hyperthermophilic origin with potential for alkyl glycoside production. Appl Microbiol Biotechnol 99:7101–7113. https://doi.org/10.1007/s00253-015-6435-2

Park KH, Jung JH, Park SG, Lee ME, Holden JF, Park CS, Woo EJ (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. https://doi.org/10.1107/S1399004714006567

Rose PW, Prlic A, Bi C, Bluhm W, Christie CH, Dutta S, Green RK et al (2015) The RCSB Protein Data Bank: views of structural biology for basic and applied research and education. Nucleic Acids Res 43:D345–D356. https://doi.org/10.1093/nar/gku1214

Santos CR, Tonoli CC, Trindade DM, Betzel C, Takata H, Kuriki T, Kanai T, Imanaka T, Arni RK, Murakami MT (2011) Structural basis for branching-enzyme activity of glycoside hydrolase family 57: structure and stability studies of a novel branching enzyme from the hyperthermophilic archaeon *Thermococcus kodakaraensis* KOD1. Proteins 79:547–557. https://doi.org/10.1002/prot.22902

Sarian FD, Janecek S, Pijning T, Ihsanawati NZ, Radjasa OK, Dijkhuizen L, Natalia D, van der Maarel MJEC (2017) A new group of glycoside hydrolase family 13 α-amylases with an aberrant catalytic triad. Sci Rep 7:44230. https://doi.org/10.1038/srep44230

Shatsky M, Nussinov R, Wolfson HJ (2004) A method for simultaneous alignment of multiple protein structures. Proteins 56:143–156. https://doi.org/10.1002/prot.10628

Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R et al (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539. https://doi.org/10.1038/msb.2011.75

Stam MR, Danchin EG, Rancurel C, Coutinho PM, 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. https://doi.org/10.1093/protein/gzl044

Svensson B (1994) Protein engineering in the α-amylase family: catalytic mechanism, substrate specificity, and stability. Plant Mol Biol 25:141–157. https://doi.org/10.1007/BF00023233

Tachibana Y, Fujiwara S, Takagi M, Imanaka T (1997) Cloning and expression of the 4-α-glucanotransferase gene from the hyperthermophilic archaeon *Pyrococcus* sp. KOD1, and characterization of the enzyme. J Ferment Bioeng 83:540–548. https://doi.org/10.1016/S0922-338X(97)81134-8

The UniProt Consortium (2017) UniProt: the universal protein knowledgebase. Nucleic Acids Res 45:D158–D169. https://doi.org/10.1093/nar/gkw1099

van der Maarel MJ, van der Veen B, Uitdehaag JC, Leemhuis H, Dijkhuizen L (2002) Properties and applications of starch-converting enzymes of the α-amylase family. J Biotechnol 94:137–155. https://doi.org/10.1016/S0168-1656(01)00407-2

Yan R, Zhao X, Lei J, Zhou Q (2019) Structure of the human LAT1-4F2hc heteromeric amino acid transporter complex. Nature 568:127–130. https://doi.org/10.1038/s41586-019-1011-z

Zhang X, Leemhuis H, Janecek S, Martinovicova M, Pijning T, van der Maarel MJEC (2019) Identification of *Thermotoga maritima* MSB8 GH57 α-amylase AmyC as a glycogen-branching enzyme with high hydrolytic activity. Appl Microbiol Biotechnol 103:6141–6151. https://doi.org/10.1007/s00253-019-09938-1

Zona R, Chang-Pi-Hin F, O'Donohue MJ, Janecek S (2004) Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*. Eur J Biochem 271:2863–2872. https://doi.org/10.1111/j.1432-1033.2004.04144.x

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.