

Sequence fingerprints of enzyme specificities from the glycoside hydrolase family GH57

Karol Blesák · Štefan Janeček

Received: 17 February 2012 / Accepted: 2 April 2012 / Published online: 22 April 2012
© Springer 2012

Abstract The glycoside hydrolase family 57 (GH57) contains five well-established enzyme specificities: α -amylase, amylopullulanase, branching enzyme, 4- α -glucanotransferase and α -galactosidase. Around 700 GH57 members originate from *Bacteria* and *Archaea*, a substantial number being produced by thermophiles. An intriguing feature of family GH57 is that only slightly more than 2 % of its members (i.e., less than 20 enzymes) have already been biochemically characterized. The main goal of the present bioinformatics study was to retrieve from databases, and analyze in detail, sequences having clear features of the five GH57 enzyme specificities mentioned above. Of the 367 GH57 sequences, 56 were evaluated as α -amylases, 99 as amylopullulanases, 158 as branching enzymes, 46 as 4- α -glucanotransferases and 8 as α -galactosidases. Based on the analysis of collected sequences, sequence logos were created for each specificity and unique sequence features were identified within the logos. These features were proposed to define the so-called sequence fingerprints of GH57 enzyme specificities. Domain arrangements characteristic of the individual enzyme specificities as well as evolutionary relationships within the family GH57 are also discussed. The results of this study could find use in rational protein design of family

GH57 amylolytic enzymes and also in the possibility of assigning a GH57 specificity to a hypothetical GH57 member prior to its biochemical characterization.

Keywords Alpha-amylase · Amylopullulanase · Bioinformatics analysis · Enzyme specificity · Evolutionary relatedness · Family GH57 · Sequence fingerprints

Abbreviations

CAZy Carbohydrate-Active enZYme
CSR Conserved sequence regions
GH Glycoside hydrolase

Introduction

The glycoside hydrolase (GH) family 57 was established in 1996 (Henrissat and Bairoch 1996). It was based on the fact that amino acid sequences of two supposed α -amylases did not exhibit similarities to the α -amylases known at that time and already classified in the well-known α -amylase family GH13 (Henrissat 1991; Janecek 1994; Svensson 1994; Kuriki and Imanaka 1999; MacGregor et al. 2001; van der Maarel et al. 2002). The first two GH57 members originated from thermophilic prokaryotes—one from bacterium *Dictyoglomus thermophilum* (Fukusumi et al. 1988) and the other from archaeon *Pyrococcus furiosus* (Laderman et al. 1993a). Although both enzymes are actually 4- α -glucanotransferases (Laderman et al. 1993b; Nakajima et al. 2004), the family GH57 was considered to be the second α -amylase family, i.e., a smaller one and distantly related to the main α -amylase family GH13 (Janecek 2005; MacGregor 2005), especially after the finding of branching enzyme specificity in the family GH57 (Murakami et al.

Communicated by H. Atomi.

Electronic supplementary material The online version of this article (doi:10.1007/s00792-012-0449-9) contains supplementary material, which is available to authorized users.

K. Blesák · Š. Janeček (✉)
Laboratory of Protein Evolution, Institute of Molecular Biology,
Slovak Academy of Sciences, Dúbravská cesta 21,
84551 Bratislava, Slovakia
e-mail: Stefan.Janecek@savba.sk

2006). At present, within the carbohydrate-active enzyme (CAZy) database classification (Cantarel et al. 2009), the family contains around 700 members, exclusively from prokaryotes, many of which are (hyper)-thermophilic *Archaea* that also produce typical GH13 α -amylases (Jorgensen et al. 1997; Janecek et al. 1999; Linden et al. 2003). Extremostable α -amylases and related starch hydrolases are highly desired from an industrial point of view (Sunna et al. 1997; Leveque et al. 2000; Bertoldo and Antranikian 2002).

Because of concentration on genome sequencing projects, fewer than 20 GH57 members have already been biochemically characterized (Janecek and Blesak 2011). In fact, five defined enzyme specificities have been classified within the family GH57 (Janecek 2010): α -amylase (EC 3.2.1.1; hydrolysis of α -1,4-glucosidic linkages in starch and related α -glucans), amylopullulanase (EC 3.2.1.1/41; hydrolysis of both α -1,4 and α -1,6 linkages in starch, pullulan and other related α -glucans), branching enzyme (EC 2.4.1.18; formation of α -1,6-branching points in glycogen and amylopectin), 4- α -glucanotransferase (EC 2.4.1.25; disproportionation of α -1,4-glucosidic linkages in α -glucans), and α -galactosidase (EC 3.2.1.22; release of galactose from melibiose and raffinose). However, based on evolutionary comparison (Zona et al. 2004; Janecek 2005) and preliminary biochemical studies (Comfort et al. 2008; Wang et al. 2011) one may expect that additional specificities will be confirmed in the future.

From the structural point of view, all GH57 members contain a (β/α)₇-barrel as their catalytic domain. The enzymes have two catalytic residues, equivalent to Glu123 of *Thermococcus litoralis* 4- α -glucanotransferase at strand β ₄ and Asp214 at strand β ₇ of the barrel. These act as the catalytic nucleophile and proton donor, respectively (Imamura et al. 2001, 2003). The retaining mechanism is employed as evidenced by ¹H-NMR analysis of the product mixture obtained by incubation of *Thermus thermophilus* branching enzyme with amylose. This confirmed the α -anomeric configuration of the 1,6-glucosidic bond formed (Palomo et al. 2011).

The sequences of family GH57 members vary a great deal, in general, in both length and sequence—from less than 400 to more than 1,300 residues, with many long insertions and even different domains characteristic of the individual specificities. In spite of this, five conserved sequence regions (CSRs) were proposed in 2004, based on the alignment of 59 GH57 amino acid sequences (Zona et al. 2004). Since the number of GH57 sequences has increased more than tenfold from that time, it makes sense to re-evaluate the CSRs in order to generalize their importance as sequence fingerprints for individual enzyme specificities. This is of special importance if the fact is taken into account that the vast majority of GH57

sequences are for putative proteins. Thus assigning a specificity, based on the presence/absence of unambiguous sequence features supported by the wealth of available sequence data, could be highly desirable. It is worth mentioning that the family GH57 contains not only many hypothetical enzymes (i.e., as yet biochemically uncharacterized proteins), but almost one half of the more than 100 GH57 members exhibiting clear α -amylase sequence features represent proteins lacking one or both catalytic residues (Janecek and Blesak 2011).

The main goal of the present bioinformatics study was the in silico analysis of as many GH57 members as possible that exhibit clear sequence features of the five well-established GH57 enzyme specificities. In total, 367 sequences were collected and analyzed in detail with the yield of sequence logos for their CSRs. The logos can define the so-called GH57 sequence fingerprints for the individual enzyme specificities. They may be useful especially as unambiguous identifiers for a given specificity for GH57 putative proteins as well as in rational protein design of these industrially important amylolytic enzymes.

Materials and methods

Sequence collection

Sequences were collected based on basic protein BLAST (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschul et al. 1990) searches using the complete sequences of 14 experimentally characterized GH57 enzymes: α -amylase from *Methanocaldococcus jannaschii* (Bult et al. 1996; Kim et al. 2001; Li and Peebles 2004), amylopullulanases from *Pyrococcus furiosus* (Dong et al. 1997; Kang et al. 2005), *Thermococcus hydrothermalis* (Erra-Pujada et al. 1999), *Thermococcus litoralis* (Imamura et al. 2004) and *Thermococcus siculi* (Jiao et al. 2011), branching enzymes from *Thermococcus kodakaraensis* (Murakami et al. 2006; Santos et al. 2011), *Thermotoga maritima* (Ballschmitter et al. 2006; Dickmanns et al. 2006) and *Thermus thermophilus* (Palomo et al. 2011), 4- α -glucanotransferases from *Archaeoglobus fulgidus* (Labes and Schönheit 2007), *Dictyoglomus thermophilum* (Fukusumi et al. 1988; Nakajima et al. 2004), *T. kodakaraensis* (Tachibana et al. 1997, 2000), *T. litoralis* (Jeon et al. 1997) and *P. furiosus* (Laderman et al. 1993a, b), and α -galactosidase from *P. furiosus* (van Lieshout et al. 2003).

The initial set consisting of more than a thousand sequences was reduced by several rounds of refining in an effort to focus attention on potentially real enzymes; i.e., those exhibiting clear sequence features characteristic of a given enzyme specificity (Zona et al. 2004). Almost 400 of the GH57 proteins were then divided into the 5 potential

Table 1 Summary of the GH57 enzyme specificities used in the present study

Enzyme	Number	Archaea	Bacteria	Length ^a	GH57 domain ^b
α -Amylase	56	47	9	~440	~400
Amylopullulanase	99	18	81	~780	~550
Branching enzyme	158	9	149	~550	~550
4- α -Glucanotransferase	46	14	32	~670	~390
α -Galactosidase	8	8	0	~360	~360

For details, see the Supplementary Table S1

^a Average length of a protein

^b This consists of the catalytic (β/α)₇-barrel and the succeeding α -helical region

GH57 aforementioned enzyme specificities (in each case a sequence had to possess both catalytic residues). This BLAST-derived set was further completed by sequences not caught by BLAST but present in the CAZy database (Cantarel et al. 2009) and also with regard to previous bioinformatics analysis (Zona et al. 2004). The final set (Table 1) thus covered 367 proteins as follows: 56 α -amylases, 99 amylopullulanases, 158 branching enzymes, 46 4- α -glucanotransferases and 8 α -galactosidases (details concerning all collected sequences are listed in Table S1).

Sequence analysis

Domain arrangement of selected representatives of the five enzyme specificities were completed based on: (1) structural information available in the literature (Imamura et al. 2001, 2003; Dickmanns et al. 2006; Palomo et al. 2011; Santos et al. 2011); (2) alignment of 14 biochemically characterized GH57 members using the program Clustal-W2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) (Larkin et al. 2007); (3) BLAST (Altschul et al. 1990) results concerning identification of conserved domains; (4) data from the Pfam database (<http://www.sanger.ac.uk/resources/databases/pfam.html>) (Punta et al. 2012); and (5) predictions of both secondary and tertiary structures obtained from the PHYRE server (<http://www.sbg.bio.ic.ac.uk/~phyre/>) (Kelley and Sternberg 2009).

For each enzyme specificity, i.e., for 56 α -amylases, 99 amylopullulanases, 158 branching enzymes, 46 4- α -glucanotransferases and 8 α -galactosidases, a sequence logo was created using the WebLogo 3.0 server (<http://weblogo.berkeley.edu/>) (Crooks et al. 2004).

Evolutionary comparison

Most of the 367 sequences were retrieved from the UniProt knowledge database (The UniProt Consortium 2012), while a few (Table S1) were obtained from GenBank (Benson et al. 2012). The alignment covered the aforementioned catalytic (β/α)₇-barrel and the succeeding

α -helical regions that are characteristic of GH57 enzymes, i.e., the C-terminal stretches of the sequences were not used, except in the cases of the α -galactosidases and a few α -amylases (for details, see Table S1).

The alignment was performed using the program Clustal-W2 (Larkin et al. 2007). A manual tuning was done in order to maximize similarities. Three evolutionary trees were prepared based on the alignment of five CSRs and complete alignment including and excluding the positions with gaps. The trees were calculated as a Phylip-tree type using the neighbor-joining clustering (Saitou and Nei 1987) and the bootstrapping procedure (Felsenstein 1985) (the number of bootstrap trials used was 1,000) implemented in the Clustal-X package (Larkin et al. 2007). The trees were displayed with the program TreeView (Page 1996).

Results and discussion

Domain arrangement and sequence comparison

At present there are only five clearly defined enzyme specificities in the family GH57 (Cantarel et al. 2009; Janecek 2010; Janecek and Blesak 2011). In addition to 4- α -glucanotransferase and branching enzyme, for which three-dimensional structures are available (Imamura et al. 2003; Dickmanns et al. 2006; Palomo et al. 2011; Santos et al. 2011), these are α -amylase (Kim et al. 2001; Li and Peebles 2004; Janecek and Blesak 2011), amylopullulanase (Dong et al. 1997; Erra-Pujada et al. 1999, 2001; Zona et al. 2004; Kang et al. 2005) and α -galactosidase (van Lieshout et al. 2003). As already indicated in the first thorough in silico analysis of the family GH57 (Zona et al. 2004), novel enzyme specificities as well as new GH57 groups or subfamilies can be expected in the future due to accumulation of more sequence and biochemical data. Thus, two interesting GH57 amylolytic enzymes have been described: one from *P. furiosus* (Comfort et al. 2008) and the other from an uncultured bacterium (Wang et al. 2011), which do not exhibit the sequence features of the five

well-established GH57 enzyme specificities. Since these two have been only partially biochemically characterized, they were not included in the present study and their analysis will be described elsewhere.

With regard to origin (Table 1), α -amylases (56 sequences) come mostly from *Archaea*, whereas both amylopullulanases (99) and branching enzymes (158) originate mainly from *Bacteria*. While 4- α -glucanotransferases (46) are roughly divided as one-third from *Archaea* and two-thirds from *Bacteria*, all 8 sequences of α -galactosidases are exclusively from *Archaea* (for details, see Table S1).

Although the catalytic $(\beta/\alpha)_7$ -barrel domain contains both GH57 catalytic residues, it is very probable that the $(\beta/\alpha)_7$ -barrel alone is not enough for the enzyme activity as evidenced by loss of enzyme activity by the deletion of the α -helical domain (called also domain C) succeeding the barrel in the branching enzyme from *T. thermophilus* (Palomo et al. 2011). Therefore, both the $(\beta/\alpha)_7$ -barrel and the succeeding α -helical region (including a three-helix bundle) are essential for correct functioning of a GH57 enzymatic member and may be considered to constitute the GH57 catalytic area (Erra-Pujada et al. 2001; Imamura et al. 2003; Palomo et al. 2011). This domain arrangement is characteristic for α -amylase and α -galactosidase, whereas the enzymes possessing the remaining three specificities contain some additional domains (Fig. 1a). The GH57 α -amylases seem to exist without a signal peptide since there is no information about it for the only characterized representative from *M. jannaschii* (Kim et al. 2001) and the CSR-1 is typically positioned very close to the protein N-terminus (Fig. S1). It is worth mentioning that domain C (the α -helical region) in α -amylases may usually be ~50–100 residues longer than in all other specificities (Fig. 1a). It is thus possible that the enzymes with α -amylase specificity may contain an extra region at the C-terminus in addition to the typical $(\beta/\alpha)_7$ -barrel and the three-helix bundle. Since this unique extra region has no counterparts in enzymes with non- α -amylase specificities, it was eliminated from all sequence comparison (Table S1; Fig. 1b).

With regard to amylopullulanases, most of them possess a signal peptide that precedes directly the catalytic $(\beta/\alpha)_7$ -barrel (Dong et al. 1997; Erra-Pujada et al. 1999; Jiao et al. 2011). Importantly there are several domains in the C-terminal part of amylopullulanases that are probably connected to the α -helical region via a linker (Fig. 1a). The β -strand domain may correspond to the C-terminal domain of 4- α -glucanotransferases (Imamura et al. 2003). In contrast, the two so-called SLD domains representing the surface layer motif-bearing domains (Erra-Pujada et al. 1999; Zona and Janecek 2005), a threonine-rich region positioned at the very C-terminus and the α -helical region within the catalytic barrel seem to be unique to GH57 amylopullulanases (Fig. 1a).

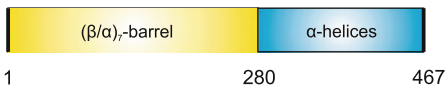
Fig. 1 a Domain arrangement of the five GH57 specificities. The catalytic GH57 region is formed from a $(\beta/\alpha)_7$ incomplete TIM-barrel domain (yellow) and the succeeding α -helical domain (blue). The remaining domains and segments are characteristic of the individual specificities (for details, see text), e.g., SP signal peptide, SLD SLH-like motifs, TRR threonine-rich region, HhH helix–hairpin–helix motif. **b** Sequence alignment of catalytic domains of biochemically characterized GH57 enzymes. α -Helices and β -strands (predicted by the Phyre server; Kelley and Sternberg 2009) are colored in red and green, respectively. The loop important for branching enzymes (Palomo et al. 2011) located between the CSR-3 and CSR-4 is highlighted in yellow. AAMY_Mccja α -amylase from *Methanocaldococcus jannaschii*; APU_Pycfu, APU_Thchy, APU_Thcli and APU_Thcsi amylopullulanases from *Pyrococcus furiosus*, *Thermococcus hydrothermalis*, *Thermococcus litoralis* and *Thermococcus siculi*, respectively; BE_Thcko, BE_Theth and BE_Thima branching enzymes from *Thermococcus kodakaraensis*, *Thermus thermophilus* and *Thermotoga maritima*, respectively; 4AGT_Arcfu, 4AGT_Pycfu, 4AGT_Thcko, 4AGT_Thcli, 4AGT_Dicth 4- α -glucanotransferases from *Archaeoglobus fulgidus*, *P. furiosus*, *T. kodakaraensis*, *T. litoralis* and *Dictyoglomus thermophilum*, respectively; and AGAL_Pycfu α -galactosidase from *P. furiosus*. CSRs are emphasized by rectangles and catalytic residues (CSR-3-glutamate catalytic nucleophile and CSR-4-aspartate proton donor) are indicated by asterisks. CSRs 1–4 are located in the $(\beta/\alpha)_7$ -barrel domain, whereas the CSR-5 is positioned in the α -helical part of the GH57 catalytic region (color figure online)

Domain composition for branching enzymes, based on the *T. kodakaraensis* branching enzyme structure (Santos et al. 2011), might not reflect all branching enzymes available, since for example the enzyme from *T. thermophilus* (Palomo et al. 2011) consists of only the $(\beta/\alpha)_7$ -barrel and the succeeding helical region. It is noteworthy that the structure of the *T. kodakaraensis* branching enzyme was solved for the GH57 catalytic domain together with the adjacent linker (Santos et al. 2011), i.e., for ~560 residues only. The α -helical segment between the two linkers was predicted by the PHYRE server (Kelley and Sternberg 2009). The two C-terminal copies of the helix–hairpin–helix motif can also be found in other enzymes and proteins and probably play a role in nucleic acid binding (Murakami et al. 2006). In branching enzymes there are two α -helical inserts within the catalytic $(\beta/\alpha)_7$ -barrel, the first one named domain B (Palomo et al. 2011) may correspond positionally to domain B in amylopullulanases and the second one (B') seems to be unique to branching enzymes (Fig. 1a).

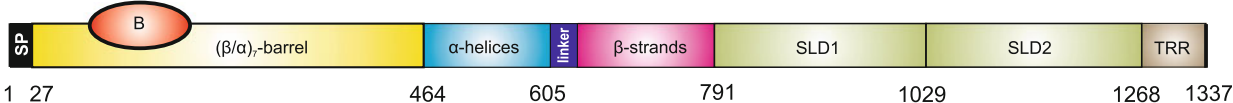
The alignment of 14 biochemically characterized GH57 members (Fig. 1b) was carried out using the segments of sequence that include the catalytic $(\beta/\alpha)_7$ -barrel plus the succeeding three-helix bundle characteristic of GH57 enzymes (Imamura et al. 2003; Dickmanns et al. 2006; Palomo et al. 2011; Santos et al. 2011). This was done in order to demonstrate the presence of CSRs typical for a given enzyme specificity and their positions in the sequences as well as secondary structure elements. The corresponding alignment of all 367 studied sequences

(a)

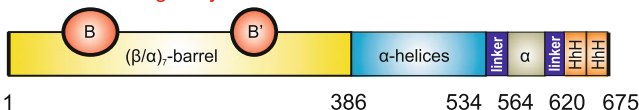
GH57 α -amylase from *Methanocaldococcus jannaschii*



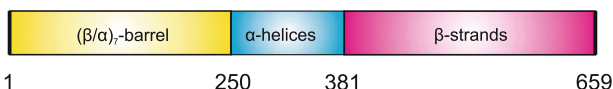
GH57 amylopullulanase from *Thermococcus hydrothermalis*



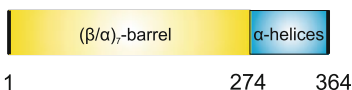
GH57 branching enzyme from *Thermococcus kodakaraensis*



GH57 4- α -glucanotransferase from *Thermococcus litoralis*



GH57 α -galactosidase from *Pyrococcus furiosus*



(b)

	CSR-1	CSR-2	CSR-3	CSR-4	CSR-5
AMY_MocJa	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
AMY_PyeFu	27 EKRPKLVITLWIQIQ	27 EKRPKLVITLWIQIQ	27 EKRPKLVITLWIQIQ	27 EKRPKLVITLWIQIQ	27 EKRPKLVITLWIQIQ
AMY_ThySu	28 EKRPKLVITLWIQIQ	28 EKRPKLVITLWIQIQ	28 EKRPKLVITLWIQIQ	28 EKRPKLVITLWIQIQ	28 EKRPKLVITLWIQIQ
AMY_TheLi	25 EKRPKLVITLWIQIQ	25 EKRPKLVITLWIQIQ	25 EKRPKLVITLWIQIQ	25 EKRPKLVITLWIQIQ	25 EKRPKLVITLWIQIQ
AMY_TheSi	33 EKRPKLVITLWIQIQ	33 EKRPKLVITLWIQIQ	33 EKRPKLVITLWIQIQ	33 EKRPKLVITLWIQIQ	33 EKRPKLVITLWIQIQ
BE_TheCo	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
BE_TheTh	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
BE_TheMa	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
4AGT_ArcFu	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
4AGT_PyeFu	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
4AGT_TheCo	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
4AGT_TheLi	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
4AGT_TheTh	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
4AGT_DicTh	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP
AGA_L_PyeFu	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP	1 ---MIIITNPIRIVQIP

(Table S1) can be found in the Supplementary material (Fig. S1). As is clear (Fig. 1b), all five GH57 specificities are very similar in substantial parts of their sequences, especially with regard to the presence of the secondary structure elements (α -helices and β -strands). There are also some differences among them (Fig. S1), including the domain arrangement of the representatives of the individual specificities (Fig. 1a).

The first 4 CSRs are positioned within the $(\beta/\alpha)_7$ -barrel on strands $\beta 1$ (CSR-1), $\beta 3$ (CSR-2), $\beta 4$ (CSR-3), and $\beta 7$ (CSR-4), while the last CSR-5 is located on the second α -helix of the three-helix bundle (Figs. 1b, S1). It is worth mentioning that the GH57 CSRs were originally described by Zona et al. (2004), but at that time only 59 sequences were available. Moreover, the 59 sequences also included those with a substitution in one or both catalytic residues as well as novel potential enzyme specificities that had not been characterized at the time. Subsequently, the specificity of branching enzyme was revealed in 2006 (Murakami et al. 2006). As far as the CSRs are concerned, it is possible to say that, after analysis of 367 sequences, they

have remained as originally proposed (Zona et al. 2004), except for the CSR-1. This region was refined here because, as demonstrated by three-dimensional structures of *T. litoralis* 4- α -glucanotransferase (Imamura et al. 2003) and *T. thermophilus* branching enzyme (Palomo et al. 2011), both histidines (e.g., CSR-1: 9_HAHLP for BE_Theth; Fig. 1b) are involved in substrate binding; now the CSR-1 covers 5 residues instead of 3.

Sequence fingerprints and evolutionary relationships

Although the GH57 CSRs were defined previously (Zona et al. 2004) and were found to apply also for the current situation, the importance of the present study is that it is based on a larger number of sequences, enabling creation of so-called sequence logos for individual enzyme specificities (Fig. 2). Thus, the present study includes 56 α -amylases, 99 amylopullulanases, 158 branching enzymes, 46 4- α -glucanotransferases and 8 α -galactosidases, in comparison with 8, 14, 10, 9 and 2, respectively, in the study of Zona et al. (2004) in 2004.

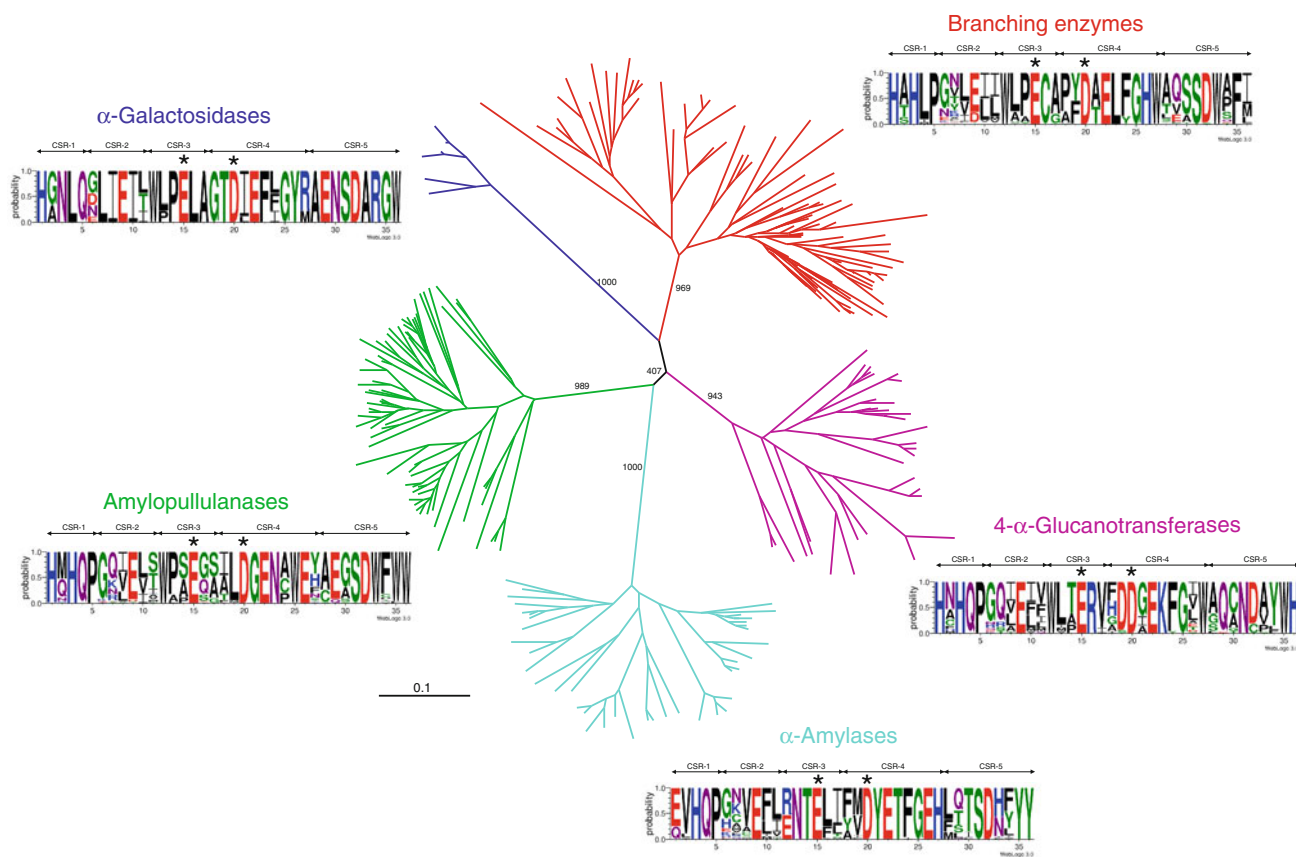


Fig. 2 Evolutionary tree and sequence logos of the five GH57 specificities. The analyzed set contains 367 GH57 enzymes and proteins and covers 56 α -amylases (cyan), 99 amylopullulanases (green), 8 α -galactosidases (blue), 158 branching enzymes (red) and 46 4- α -glucanotransferases (magenta). The tree is based on the

alignment of the five CSRs (36 residues). Sequence logos (created using the WebLogo 3.0 server; Crooks et al. 2004): CSR-1 residues 1–5, CSR-2 residues 6–11, CSR-3 residues 12–17 (No. 15, glutamate, is the catalytic nucleophile), CSR-4 residues 18–27 (No. 20, aspartate, is the proton donor), CSR-5 residues 28–36 (color figure online)

Despite the fact that sequence logos of the five GH57 specificities are mostly similar to each other, every specificity exhibits its own characteristic sequence features (Fig. 2). Thus in the α -amylase sequence, logo positions 1, 12, 13, 21, 27 and 35–36 are unique for this specificity. The positions 1 (mostly glutamate; CSR-1) and 12 (arginine or glutamate; CSR-3) are characterized by the lack of histidine and tryptophan, respectively, present invariably in these positions in all four remaining specificities. The invariant presence of asparagine and tyrosine in positions 13 (CSR-3) and 21 (CSR-4), respectively, is also exclusive to the α -amylases, since in the other specificities there are different residues that are, moreover, not so strictly conserved. Similarly, there is an invariant histidine in position 27 (CSR-4), although a corresponding histidine can also be found in some representatives of amylopullulanases. Of note is the presence of a histidine in this position in a recently published GH57 sequence from an uncultured bacterium (Wang et al. 2011), but this unspecified amylase was not used in the present study and it may establish a novel GH57 specificity (group) in the future. The two adjacent tyrosines at the end of the logo (positions 35–36; CSR-5) represent the most typical GH57 α -amylase signature because none of the 311 sequences of the remaining four specificities contains a tyrosine in either position (Fig. 2).

It is very important to say that the last two positions in the GH57 sequence logo (35–36; CSR-5) belong to a sequence fingerprint that best distinguishes the individual enzyme specificities from each other. There are usually two tryptophans in amylopullulanases, although the first one is not always conserved and is replaced by a phenylalanine in a few cases. This position (35) in α -galactosidases is invariably occupied by a glycine succeeded by an invariant tryptophan. In 4- α -glucanotransferases there are tryptophan and histidine residues in these positions, again completely conserved. In branching enzyme, the first position (35) is occupied by a totally conserved phenylalanine, whereas the residue at position 36 is not conserved, but is usually a hydrophobic non-aromatic residue (Fig. 2). Furthermore, both branching enzymes and amylopullulanases possess another invariant aromatic residue, tryptophan at position 33 (CSR-5) that is not present in any of the remaining specificities.

Amylopullulanases have one more invariant tryptophan (position 25; CSR-4). Interestingly, it is not only unique for amylopullulanases, but all the four remaining specificities (i.e., α -amylases, branching enzymes, 4- α -glucanotransferases and α -galactosidases) have invariant glycine in the corresponding position. The presence of an invariant arginine in position 16 (CSR-3) is exclusively unique for 4- α -glucanotransferases. Remarkably, in 158 of 159 compared branching enzyme sequences, there is a cysteine in position 16 (CSR-3), and only *T. thermophilus* branching enzyme has the cysteine substituted by a methionine Met158

(Palomo et al. 2011). Both aforementioned specificities possess a tryptophan in position 27 (CSR-4), while the 4- α -glucanotransferase contains an invariant asparagine in position 31 (CSR-5) that, in all remaining specificities, is exclusively occupied by a serine. As far as the α -galactosidase is concerned, there is a characteristic three-residue-long stretch NLQ starting at the position 3 in CSR-1, although a hydrophobic residue in the position 4 (mostly leucine) is found also in branching enzymes. Finally, position 23 (CSR-4) deserves special attention since all the five GH57 specificities possess an invariant residue in that position that discriminates them from each other as follows: α -amylases—threonine, amylopullulanases—asparagine, branching enzymes—leucine, 4- α -glucanotransferases—lysine, and α -galactosidases—phenylalanine.

It is worth mentioning that of the positions typical for the individual GH57 enzyme specificities described above, some of them were unambiguously recognized as essential or at least important for their function. The roles were proven for 4- α -glucanotransferase from *T. litoralis* (Imamura et al. 2003) and branching enzymes from *T. thermophilus* (Palomo et al. 2011) and *T. kodakaraensis* (Santos et al. 2011). Their three-dimensional structures solved in complex with acarbose (Imamura et al. 2003) or with modeled maltotriose (Palomo et al. 2011) revealed the roles various residues play in substrate binding sites to help the catalytic machinery carry out the enzymatic activity. Thus, His11 (position 1, CSR-1; 4- α -glucanotransferase from *T. litoralis* numbering) was identified as involved in the donor -1 subsite (for the subsites nomenclature, see Davies et al. (1997)) as well as both His13 (position 3, CSR-1) and Trp357 (position 35, CSR-5) although only indirectly via a water molecule (Imamura et al. 2003; Palomo et al. 2011). On the other hand, positions 16 and 27, i.e., Arg124 and Trp221 in the 4- α -glucanotransferase (Imamura et al. 2003), both function at the acceptor subsite $+1$ (Imamura et al. 2003; Palomo et al. 2011). The latter residue has already been identified as contributing to transglycosylation activity of *P. furiosus* 4- α -glucanotransferase (Tang et al. 2006). Note that enzymes with various GH57 specificities often possess highly specific residues in all these important positions (Fig. 2), and this information can be used to predict specificity. For example, the presence of an almost unique cysteine in the CSR-3, a clear branching enzyme sequence feature, makes it possible to propose that the amylolytic enzyme AmyC from *T. maritima* (Dickmanns et al. 2006), originally described as an “ α -amylase” (Ballschmitter et al. 2006), may also have branching enzyme activity (Fig. 1b). Branching enzymes should moreover contain, between the CSR-3 and CSR-4, a flexible loop (235_PYGEEALG in *T. thermophilus* branching enzyme) believed essential for branching activity, because the Y236A mutant lost all branching activity

and acquired an increased hydrolytic activity (Palomo et al. 2011). In the branching enzymes studied here the Tyr236 is neither conserved invariantly, nor is it always replaced by an aromatic residue (Fig. S1). It could therefore be a sequence-structural feature that discriminates potential GH57 branching enzyme subgroups (subfamilies) from each other.

It should be pointed out that residues at specific positions in the sequence logos can be considered as sequence fingerprints of individual enzyme specificities and their mutual exchange can be applied in an effort to modify enzyme substrate/product specificity and/or even to improve enzyme efficiency in a way similar to that already described for amylolytic hydrolases/transferases from the main α -amylase family GH13 (Leemhuis et al. 2002, 2003a, b, 2004; Kelly et al. 2007).

The uniqueness of every specificity was clearly documented in the evolutionary trees (Fig. 2, S2). Although the specificities may contain some groups of more or less closely related enzymes, all identified GH57 members belonging to a given specificity should be positioned on a common branch. This is best made evident in the tree based on the alignment of CSRs (Fig. 2) although the two additional trees based on the alignment of the whole GH57 catalytic part, i.e., the catalytic (β/α)₇-barrel and the succeeding helical region, deliver, in fact, comparable arrangements whether positions with gaps were included or excluded (Fig. S2). Overall, it is clear that amylolytic hydrolases (α -amylase and amylopullulanase) and transferases (branching enzyme and 4- α -glucanotransferase) go together, while the evolutionary relationship of the α -galactosidases to the other GH57 enzymes is more complex: in the CSR-based tree it clusters with branching enzyme (Fig. 2), whereas in both catalytic-region-based trees it is moved towards α -amylase (Fig. S2). In the main α -amylase family GH13 (with more than 30 different enzyme specificities) α -amylase, amylopullulanase, branching enzyme and 4- α -glucanotransferase belong to separate clusters/subfamilies (Janecek 1994, 1997; Stam et al. 2006, Janecek et al. 2007); but there is no α -galactosidase specificity in the family GH13 (Cantarel et al. 2009). It therefore makes little sense to try to compare strictly the evolutionary relationships within the two families GH13 and GH57. It should nevertheless be clear that one should expect the division of GH57 specificity clusters depicted in the evolutionary trees (Figs. 2, S2) to correspond with GH57 subfamilies in the future.

Conclusions

The present in silico study focused on five well-established GH57 enzyme specificities, namely α -amylase, amylopullulanase, branching enzyme, 4- α -glucanotransferase and α -galactosidase. Based on a detailed analysis of 367

sequences, unique specificity features were identified in their sequence logos and discussed as the GH57 sequence fingerprints. In addition, a domain arrangement characteristic for the individual specificities was proposed together with a description of their basic evolutionary relationships. The results of this study could find use in the possibility of assigning a GH57 specificity to a hypothetical GH57 member prior to its biochemical characterization. The other no less significant achievement of the present study is the opportunity to utilize the results in rational protein design of GH57 amylolytic enzymes in an effort to prepare these industrially important enzymes with tailored properties.

Acknowledgments This work was supported by the Slovak Research and Development Agency under the contract No. LPP-0417-09 and by the VEGA grant No. 2/0148/11. We would like to thank Dr. E. Ann McGregor (Livingston, West Lothian, UK) for her critical reading of the manuscript and language corrections.

References

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215:403–410
- Ballschmitter 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
- Benson DA, Karsch-Mizrachi I, Clark K, Lipman DJ, Ostell J, Sayers EW (2012) GenBank. *Nucleic Acids Res* 40(Database issue): D48–D53
- Bertoldo C, Antranikian G (2002) Starch-hydrolyzing enzymes from thermophilic archaea and bacteria. *Curr Opin Chem Biol* 6: 151–160
- Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, Kerlavage AR, Dougherty BA, Tomb JF, Adams MD, Reich CI, Overbeek R, Kirkness EF, Weinstock KG, Merrick JM, Glodek A, Scott JL, Geoghagen NS, Venter JC (1996) Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273:1058–1073
- Cantarel BL, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B (2009) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. *Nucleic Acids Res* 37(Database issue):D233–D238
- Comfort DA, Chou CJ, Connors 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
- Crooks GE, Hon G, Chandonia JM, Brenner SE (2004) WebLogo: a sequence logo generator. *Genome Res* 14:1188–1190
- Davies GJ, Wilson KS, Henrissat B (1997) Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* 321:557–559
- Dickmanns A, Ballschmitter M, Liebl W, Ficner R (2006) Structure of the novel α -amylase AmyC from *Thermotoga maritima*. *Acta Crystallogr D Biol Crystallogr* 62:262–270
- Dong G, Vieille C, Zeikus JG (1997) Cloning, sequencing, and expression of the gene encoding amylopullulanase from *Pyrococcus furiosus* and biochemical characterization of the recombinant enzyme. *Appl Environ Microbiol* 63:3577–3584

- Erra-Pujada M, Debeire P, Duchiron F, O'Donohue MJ (1999) The type II pullulanase of *Thermococcus hydrothermalis*: molecular characterization of the gene and expression of the catalytic domain. *J Bacteriol* 181:3284–3287
- Erra-Pujada M, Chang-Pi-Hin F, Debeire P, Duchiron F, O'Donohue MJ (2001) Purification and properties of the catalytic domain of the thermostable pullulanase type II from *Thermococcus hydrothermalis*. *Biotechnol Lett* 23:1273–1277
- Felsenstein J (1985) Confidence-limits on phylogenies—an approach using the bootstrap. *Evolution* 39:783–791
- 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
- Henrissat B (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 280:309–316
- Henrissat B, Bairoch A (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* 316:695–696
- Imamura H, Fushinobu S, Jeon BS, Wakagi T, Matsuzawa H (2001) Identification of the catalytic residue of *Thermococcus litoralis* 4- α -glucanotransferase through mechanism-based labeling. *Biochemistry* 40:12400–12406
- 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
- Imamura H, Jeon BS, Wakagi T (2004) Molecular evolution of the ATPase subunit of three archaeal sugar ABC transporters. *Biochem Biophys Res Commun* 319:230–234
- Janecek S (1994) Parallel β/α -barrels of α -amylase, cyclodextrin glycosyltransferase and oligo-1,6-glucosidase versus the barrel of β -amylase: evolutionary distance is a reflection of unrelated sequences. *FEBS Lett* 353:119–123
- Janecek S (1997) α -Amylase family: molecular biology and evolution. *Prog Biophys Mol Biol* 67:67–97
- Janecek S (2005) Amylolytic families of glycoside hydrolases: focus on the family GH-57. *Biologia* 60(Suppl 16):177–184
- Janecek S (2010) Glycoside hydrolase family 57. In *CAZyedia*. (<http://www.cazyedia.org/>). Accessed 18 January 2012
- 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
- Janecek S, Leveque E, Belarbi A, Haye B (1999) Close evolutionary relatedness of α -amylases from Archaea and plants. *J Mol Evol* 48:421–426
- 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
- Jeon BS, 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
- Jiao YL, Wang SJ, Lv MS, Xu JL, Fang YW, Liu S (2011) A GH57 family amylopullulanase from deep-sea *Thermococcus siculi*: expression of the gene and characterization of the recombinant enzyme. *Curr Microbiol* 62:222–228
- Jorgensen S, Vorgias CE, Antranikian G (1997) Cloning, sequencing, characterization, and expression of an extracellular α -amylase from the hyperthermophilic archaeon *Pyrococcus furiosus* in *Escherichia coli* and *Bacillus subtilis*. *J Biol Chem* 272:16335–16342
- Kang S, Vieille C, Zeikus JG (2005) Identification of *Pyrococcus furiosus* amylopullulanase catalytic residues. *Appl Microbiol Biotechnol* 66:408–413
- Kelley LA, Sternberg MJ (2009) Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* 4:363–371
- Kelly RM, Leemhuis H, Dijkhuizen L (2007) Conversion of a cyclodextrin glucanotransferase into an α -amylase: assessment of directed evolution strategies. *Biochemistry* 46:11216–11222
- 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
- Kuriki T, Imanaka T (1999) The concept of the α -amylase family: structural similarity and common catalytic mechanism. *J Biosci Bioeng* 87:557–565
- 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
- Laderman KA, Asada K, Uemori T, Mukai H, Taguchi Y, Kato I, Anfinsen CB (1993a) α -Amylase from the hyperthermophilic archaeobacterium *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 archaeobacterium *Pyrococcus furiosus*. *J Biol Chem* 268:24394–243401
- Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947–2948
- Leemhuis H, Dijkstra BW, Dijkhuizen L (2002) Mutations converting cyclodextrin glycosyltransferase from a transglycosylase into a starch hydrolase. *FEBS Lett* 514:189–192
- Leemhuis H, Kragh KM, Dijkstra BW, Dijkhuizen L (2003a) Engineering cyclodextrin glycosyltransferase into a starch hydrolase with a high exo-specificity. *J Biotechnol* 103:203–212
- Leemhuis H, Rozeboom HJ, Wilbrink M, Euverink GJ, Dijkstra BW, Dijkhuizen L (2003b) Conversion of cyclodextrin glycosyltransferase into a starch hydrolase by directed evolution: the role of alanine 230 in acceptor subsite +1. *Biochemistry* 42:7518–7526
- Leemhuis H, Wehmeier UF, Dijkhuizen L (2004) Single amino acid mutations interchange the reaction specificities of cyclodextrin glycosyltransferase and the acarbose-modifying enzyme acarboxyl transferase. *Biochemistry* 43:13204–13213
- Leveque E, Janecek S, Belarbi A, Haye B (2000) Thermophilic archaeal amylolytic enzymes. *Enzyme Microb Technol* 26:2–13
- Li M, Peeples TL (2004) Purification of hyperthermophilic archaeal amylolytic enzyme (MJA1) using thermoseparating aqueous two-phase systems. *J Chromatogr B* 807:69–74
- Linden A, Mayans O, Meyer-Klaucke W, Antranikian G, Wilmanns M (2003) Differential regulation of a hyperthermophilic α -amylase with a novel (Ca,Zn) two-metal center by zinc. *J Biol Chem* 278:9875–9884
- MacGregor EA (2005) An overview of clan GH-H and distantly-related families. *Biologia* 60(Suppl 16):5–12
- 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
- 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
- Nakajima M, Imamura H, Shoun H, Horinouchi S, Wakagi T (2004) Transglycosylation activity of *Dictyoglomus thermophilum* amylase A. *Biosci Biotechnol Biochem* 68:2369–2373
- Page RD (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* 12:357–358

- 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
- Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD (2012) The Pfam protein families database. *Nucleic Acids Res* 40(Database issue):D290–D301
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425
- 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
- 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
- Sunna A, Moracci M, Rossi M, Antranikian G (1997) Glycosyl hydrolases from hyperthermophiles. *Extremophiles* 1:2–13
- Svensson B (1994) Protein engineering in the α -amylase family: catalytic mechanism, substrate specificity, and stability. *Plant Mol Biol* 25:141–157
- 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
- Tachibana Y, Takaha T, Fujiwara S, Takagi M, Imanaka T (2000) Acceptor specificity of 4- α -glucanotransferase from *Pyrococcus kodakaraensis* KOD1, and synthesis of cycloamylose. *J Biosci Bioeng* 90:406–409
- Tang SY, Yang SJ, Cha H, Woo EJ, Park C, Park KH (2006) Contribution of W229 to the transglycosylation activity of 4- α -glucanotransferase from *Pyrococcus furiosus*. *Biochim Biophys Acta* 1764:1633–1638
- The UniProt Consortium (2012) Reorganizing the protein space at the Universal Protein Resource (UniProt). *Nucleic Acids Res* 40(Database issue):D71–D75
- 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
- van Lieshout JFT, Verhees CH, van der Oost J, de Vos WM, Ettema TJG, van der Sar S, Imamura H, Matsuzawa H (2003) Identification and molecular characterization of a novel type of α -galactosidase from *Pyrococcus furiosus*. *Biocatal Biotransform* 21:243–252
- Wang H, Gong Y, Xie W, Xiao W, Wang J, Zheng Y, Hu J, Liu Z (2011) Identification and characterization of a novel thermostable *gh-57* gene from metagenomic fosmid library of the Juan de Fuca Ridge hydrothermal vent. *Appl Biochem Biotechnol* 164:1323–1338
- Zona R, Janecek S (2005) Relationships between SLH motifs from different glycoside hydrolase families. *Biologia* 60(Suppl 16):115–121
- 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