Eur. J. Biochem. **271**, 2863–2872 (2004) © FEBS 2004 doi:10.1111/j.1432-1033.2004.04144.x

Bioinformatics of the glycoside hydrolase family 57 and identification of catalytic residues in amylopullulanase from *Thermococcus hydrothermalis*

Richard Zona[1,*], Florent Chang-Pi-Hin[2,*], Michael J. O'Donohue[2] and Štefan Janeček[1]

[1]*Institute of Molecular Biology, member of the Centre of Excellence for Molecular Medicine, Slovak Academy of Sciences, Bratislava, Slovakia;* [2]*Institut National de la Recherche Agronomique, UMR FARE, Reims, France*

Fifty-nine amino acid sequences belonging to family 57 (GH-57) of the glycoside hydrolases were collected using the CAZy server, Pfam database and BLAST tools. Owing to the sequence heterogeneity of the GH-57 members, sequence alignments were performed using mainly manual methods. Likewise, five conserved regions were identified, which are postulated to be GH-57 consensus motifs. In the 659 amino acid-long 4-α-glucanotransferase from *Thermococcus litoralis*, these motifs correspond to 13_HQP (region I), 76_GQLEIV (region II), 120_WLTERV (region III), 212_HDDGEKFGVW (region IV), and 350_AQCNDA YWH (region V). The third and fourth conserved regions contain the catalytic nucleophile and the proton donor, respectively. Based on our sequence alignment, residues Glu291 and Asp394 were proposed as the nucleophile and proton donor, respectively, in a GH-57 amylopullulanase from *Thermococcus hydrothermalis*. To validate this prediction, site-directed mutagenesis was performed. The results of this work reveal that both residues are critical for the pullulanolytic and amylolytic activities of the amylopullulanase. Therefore, these data support the prediction and strongly suggest that the bifunctionality of the amylopullulanase is determined by a single catalytic centre. Despite this positive validation, our alignment also reveals that certain GH-57 members do not possess the Glu and Asp corresponding to the predicted GH-57 catalytic residues. However, the sequences concerned by this anomaly encode putative proteins for which no biochemical or enzymatic data are yet available. Finally, the evolutionary trees generated for GH-57 reveal that the entire family can be divided into several subfamilies that may reflect the different enzyme specificities.

Keywords: amylopullulanase; catalytic residues; conserved sequence region; glycoside hydrolase family 57; site-directed mutagenesis.

Amylolytic enzymes form a large group of enzymes acting on starch and related oligo- and polysaccharides. The majority of these enzymes have been grouped into the α-amylase family [1] that in the sequence-based classification of glycoside hydrolases [2] constitutes the clan GH-H covering three glycoside hydrolase families (GH-13, 70 and 77). All members of clan GH-H are multidomain proteins that exhibit a catalytic $(\beta/\alpha)_8$-barrel fold (TIM barrel), use a common catalytic machinery, and employ a retaining mechanism for α-glycosidic bond cleavage [3]. GH-13 is the main family [1] and contains almost 30 enzyme specificities, including cyclodextrin glucanotransferase, oligo-1,6-glucosidase, neopullulanase, amylosucrase, etc., in addition to α-amylase. Recently, several closely related members of GH-13 were grouped into subfamilies [4]. GH-70 consists of glucan-synthesizing glucosyltransferases, which display a circularly permuted form of the catalytic $(\beta/\alpha)_8$-barrel domain [5]. GH-77 covers amylomaltases (4-α-glucanotransferases) that lack domain C, which succeeds the catalytic $(\beta/\alpha)_8$-barrel in GH-13 members [6]. The characteristic feature common to the entire clan GH-H is the existence of between four and seven conserved sequence motifs [7].

Two other types of amylolytic enzymes – β-amylase and glucoamylase – are classified in families GH-14 and GH-15, respectively [8]. Members of both families employ an inverting mechanism for glucosidic bond cleavage [9]. From a structural point of view, β-amylase adopts a $(\beta/\alpha)_8$-barrel architecture [10], while the glucoamylase belongs to the $(\alpha/\alpha)_6$-barrel proteins [11]. Finally, family GH-31 also contains some enzymes that display α-glucosidase and glucoamylase activities [12]. Like those of the clan GH-H, GH-31 members employ the retaining mechanism; however, no 3D structure is available at present [2].

More than 15 years ago the sequence of a heat-stable α-amylase from a thermophilic bacterium, *Dictyoglomus*

Correspondence to Š. Janeček, Institute of Molecular Biology, Member of the Centre of Excellence for Molecular Medicine, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84551 Bratislava 45, Slovakia.
Fax: + 421 25930 7416, Tel.: + 421 25930 7420,
E-mail: Stefan.Janecek@savba.sk

Abbreviations: GH-57, glycoside hydrolase family 57.

Enzymes: pullulanase (EC 3.2.1.41), 4-α-glucanotransferase (EC 2.4.1.25), α-amylase (EC 3.2.1.1).

**Note*: These authors contributed equally to this work.

Note: a website is available at http://imb.savba.sk/~janecek/Papers/GH-57/

(Received 28 January 2004, revised 10 March 2004, accepted 2 April 2004)

thermophilum, was published [13]. Despite the fact that this sequence encodes an α-amylase, its analysis did not reveal any detectable similarities with known sequences of GH-13. Later, a similar sequence encoding an α-amylase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, was determined [14]. Together, these two sequences became the basis for a new amylolytic family, GH-57 [15]. The main reason for establishing GH-57 was the fact that these two α-amylases lack the conserved sequence regions characteristic of typical GH-13 α-amylases [7].

Significantly, GH-57 is mainly composed of thermostable enzymes from extremophiles, which exhibit α-amylase, 4-α-glucanotransferase, amylopullulanase, and α-galactosidase specificities [2]. At least one half of the family is formed by ORFs coding for putative proteins of uncharacterized activity and specificity. A striking feature of GH-57 is the sequence and length diversity of the individual members. Indeed, certain GH-57 enzymes can be less than 400 residues in length, while others can be composed of over 1500 residues. Consequently, GH-57 sequences cannot be aligned using routine alignment programs. Moreover, the structural information for GH-57 is very poor. To date, only one structure, which was recently released, has been determined [16]. The structural data for the GH-57 4-α-glucanotransferase from *Thermococcoccus litoralis* has revealed a $(\beta/\alpha)_7$-barrel fold (i.e. an incomplete TIM barrel) and two acidic residues, Glu123 and Asp214, which appear to define the catalytic centre of the enzyme. Importantly, the distance between the pair of oxygen atoms of Glu123 and Asp214 is appropriate for retaining enzymes (less than 7 Å) [16], thus confirming that GH-57 employs a retaining mechanism for α-glycosidic bond cleavage [9]. Despite this important advancement in the study of GH-57, no detailed alignment of the complete sequences of GH-57 members has yet been accomplished. To date, only partial or selected sequences have been compared [17–19]. An alignment of GH-57 members is available in the Pfam database (entry PF03065) [20]. However, as this alignment is focused on the ≈ 300 N-terminal amino acid residues only, by taking into account the previously discussed diversity of GH-57 sequences its value may be considered to be limited.

Previously, we have isolated and characterized the sequence (*apu*) encoding a hyperthermostable amylopullulanase from *Thermococcus hydrothermalis* AL662. The analysis of this sequence revealed that the encoded enzyme is a member of GH-57 [21,22]. The cloning and expression of *apu* in *Escherichia coli* has led to the production of a C-terminally truncated protein (designated ThApuΔ2), which nevertheless exhibits full catalytic functionality when compared with wild-type amylopullulanase [23,24]. Importantly, despite truncation, ThApuΔ2 displays wild-type physicochemical characteristics and, like the parent enzyme, is able to hydrolyse α-1,4-glucosidic bonds in substrates such as amylose and α-1,6-glucosidic bonds in pullulan. More recently, using recombinant ThApuΔ2 as an experimental model, we have attempted to explore the molecular basis of its catalytic activity, to provide new understanding concerning its bifunctionality and to establish links between this GH-57 amylopullulanase and other non pullulan-degrading GH-57 and GH-13 amylolytic enzymes (F. Chang-Pi-Hin, L. Greffe, H. Driguez & M. J. O'Donohue, unpublished data).

Therefore, in attempt to provide the first elements towards the understanding of the functionality of the potentially valuable, heat stable GH-57 enzymes, especially that of the *T. hydrothermalis* amylopullulanase, the present work has focused on a detailed analysis of all the available complete GH-57 amino acid sequences. This study was performed with a view to achieving several goals, specifically (a) to identify homologous regions common to the whole family, (b) to reveal the invariant and/or strongly conserved residues that could be functional determinants in these enzymes and to verify their functional relevance by site-directed mutagenesis, (c) to define the subfamilies of the GH-57, reflecting the sequence similarities and/or differences, and (d) to draw an evolutionary picture, as complete as possible, of this diversified family of glycoside hydrolases.

Materials and methods

Bioinformatics studies

GH-57 enzymes included in the present study are listed in Table 1. To collect the sequences, the CAZy server and Pfam database were used. The sequences were retrieved from GenBank [25] and UniProt [26]. The coordinates of the 3D structure of *T. litoralis* 4-α-glucanotransferase was retrieved from the Protein Data Bank [27] under the PDB code 1K1W [16].

Owing to the aforementioned sequence-diversity problem, alignment of the GH-57 family was carried out manually. Partial and pairwise alignments were performed using the program CLUSTALW [28]. The method used for building the evolutionary trees was the neighbour-joining method [29]. The Phylip format tree output was applied using the bootstrapping procedure [30]; 1000 bootstrap trials were used. The trees were drawn using the TREEVIEW program [31]. In order to detect new GH-57 members within the incomplete genome sequencing projects, which are not yet present in CAZy, the BLAST routine [32] was applied using known GH-57 members as templates.

Site-directed mutagenesis, mutant protein preparation and initial analysis

Mutation of residues Glu291 and Asp394 was performed using the QuikChange site-directed mutagenesis kit (Stratagene), the plasmid pAPUΔ2 [22,23] and appropriate oligonucleotides (only forward primers are shown and the mutated codon is underlined): Glu291Ala (5′-CGG ATGGGCGGCTGCGAGCGCCCTCAAGAC-3′) and Asp394Ala (5′-GTGGTCACGCTCGCCGGCGAGAAC CCGTGGGAG-3′).

After mutagenesis and verification by DNA sequencing using a MEGABACE 1000 automated sequencing system and DYEnamic™ ET dye terminator technology (Amersham Biosciences, Saclay, France), the plasmid-borne mutated genes were expressed in *E. coli* JM109 DE3 cells and mutated proteins were purified as previously described [23]. In order to verify overall correct folding, the secondary structures of each mutant protein were examined by CD using a Jobin-Yvon CD 6 spectrophotopolarimeter (Jobin Yvon S.A.S., Longjumeau, France).

Table 1. The proteins from the family GH-57 used in the present study. ND, not determined. The two GH-57 members, the 4-α-glucanotransferase with known three-dimensional structure and the amylopullulanase mutated in this study, are highlighted in bold. Domain of life, either Archaea (A) or Bacteria (B), is given in parentheses under Microorganism. The abbreviations consist of the UniProt Accession numbers [26] and UniProt species code (http://www.expasy.org/cgi-bin/speclist). The only exception is the patented α-galactosidase (GenPept: AAE28307.1) available in the UniProt archive (UniParc) under the Accession number UPI000014BAB4. The GenPept protein identification numbers are from GenBank [25].

Enzyme (hypothetical protein)	EC	Microorganism	Abbreviation	GenPept	Length
ALR2450	ND	*Anabaena* sp. PCC7120 (B)	Q8YUA2_ANASP	BAB74149.1	529
ALR1310	ND	*Anabaena* sp. PCC7120 (B)	Q8YXA5_ANASP	BAB73267.1	744
ALR0627	ND	*Anabaena* sp. PCC7120 (B)	Q8YZ60_ANASP	BAB72585.1	907
AQ_720	ND	*Aquifex aeolicus* VF5 (B)	O66934_AQUAE	AAC06900.1	477
BH1415	ND	*Bacillus halodurans* C-125 (B)	Q9KD04_BACHD	BAB05134.1	923
BT4305 (α-amylase)	ND	*Bacteroides thetaiotaomicron* VPI-5482 (B)	Q89ZS1_BACTN	AAO79410.1	460
CAC2414	ND	*Clostridium acetobutylicum* ATCC824 (B)	Q97GF3 °CLOAB	AAK80369.1	527
α-Amylase (amyA)	3.2.1.1	*Dictyoglomus thermophilum* (B)	P09961_DICTH	CAA30735.1	686
Gll1326	ND	*Gloeobacter violaceus* PCC 7421 (B)	Q7NL00_GLOVI	BAC89267.1	729
MJ1611 (α-amylase)	ND	*Methanococcus jannaschii* (A)	Q59006_METJA	AAB99631.1	467
MA4053 (α-amylase)	ND	*Methanosarcina acetivorans* C2A (A)	Q8TIT8_METAC	AAM07401.1	378
MA4052 (α-amylase)	ND	*Methanosarcina acetivorans* C2A (A)	Q8TIT9_METAC	AAM07400.1	396
MM0861 (α-amylase)	ND	*Methanosarcina mazei* Goe1 (A)	Q8PYK0_METMA	AAM30557.1	378
MM0862 (α-amylase)	ND	*Methanosarcina mazei* Goe1 (A)	Q8PYJ9_METMA	AAM30558.1	398
ML1714	ND	*Mycobacterium leprae* TN (B)	Q9CBR4_MYCLE	CAC30667.1	522
RV3031	ND	*Mycobacterium tuberculosis* H37Rv (B)	O53278_MYCTU	AAK47445.1	526
NE2031	ND	*Nitrosomonas europaea* ATCC 19718 (B)	Q82T87_NITEU	CAD85942.1	573
NE2032 (AmyA)	ND	*Nitrosomonas europaea* ATCC 19718 (B)	Q82T86_NITEU	CAD85943.1	670
PG1683	ND	*Porphyromonas gingivalis* W83 (B)	Q7MU72_PORGI	AAQ66699.1	428
PAE3428	ND	*Pyrobaculum aerophilum* IM2 (A)	Q8ZT57_PYRAE	AAL64906.1	457
PAE1048	ND	*Pyrobaculum aerophilum* IM2 (A)	Q8ZXX1_PYRAE	AAL63225.1	471
PAE3454 (pullulanase)	ND	*Pyrobaculum aerophilum* IM2 (A)	Q8ZT36_PYRAE	AAL64927.1	999
PAB0644	ND	*Pyrococcus abyssi* GE5 (A)	Q9V038_PYRAB	CAB49868.1	597
PAB1857	ND	*Pyrococcus abyssi* GE5 (A)	Q9V0M7_PYRAB	CAB49676.1	602
PAB0118 (amyA)	ND	*Pyrococcus abyssi* GE5 (A)	Q9V298_PYRAB	CAB49100.1	655
PAB0122 (amylopullulanase)	ND	*Pyrococcus abyssi* GE5 (A)	Q9V294_PYRAB	CAB49104.1	1362
α-Galactosidase (galA; PF0444)	3.2.1.22	*Pyrococcus furiosus* DSM3638 (A)	Q9HHB5_PYRFU	AAG28455.1	364
PF0870	ND	*Pyrococcus furiosus* DSM3638 (A)	Q8U2G5_PYRFU	AAL80994.1	597
PF1393	ND	*Pyrococcus furiosus* DSM3638 (A)	Q8U136_PYRFU	AAL81517.1	632
α-Amylase	3.2.1.1	*Pyrococcus furiosus* DSM3638 (A)	P49067_PYRFU	AAA72035.1	649
PF0272 (α-amylase)	ND	*Pyrococcus furiosus* DSM3638 (A)	P49067_PYRFU	AAL80396.1	656
Amylopullulanase	3.2.1.1/41	*Pyrococcus furiosus* DSM3638 (A)	O30772_PYRFU	AAB71229.1	853
PF1935 (amylopullulanase)	ND	*Pyrococcus furiosus* DSM3638 (A)	Q8TZQ1_PYRFU	AAL82059.1	985
PH0368	ND	*Pyrococcus horikoshi* OT3 (A)	O58106_PYRHO	BAA29442.1	364
PH1386	ND	*Pyrococcus horikoshi* OT3 (A)	O50094_PYRHO	BAA30492.1	560
PH1023	ND	*Pyrococcus horikoshi* OT3 (A)	O58774_PYRHO	BAA30120.1	598
PH0193 (α-amylase)	3.2.1.1	*Pyrococcus horikoshi* OT3 (A)	O57932_PYRHO	BAA29262.1	633
4-α-Glucanotransferase	2.4.1.25	*Pyrococcus kodakaraensis* (A)	O32450_PYRKO	BAA22062.1	653
RB2160	ND	*Rhodopirellula baltica* (B)	Q7UWA6_RHOBA	CAD72460.1	719
SO3268	ND	*Shewanella oneidensis* MR-1 (B)	Q8EC76_SHEON	AAN56266.1	638
SSO0988 (α-amylase)	ND	*Sulfolobus solfataricus* P2 (A)	Q97ZD2_SULSO	AAK41260.1	447
SSO1172	ND	*Sulfolobus solfataricus* P2 (A)	Q97YY0_SULSO	AAK41420.1	902
ST0817	ND	*Sulfolobus tokodaii* 7 (A)	Q973T0_SULTO	BAB65830.1	443
ST1102	ND	*Sulfolobus tokodaii* 7 (A)	Q972N0_SULTO	BAB66135.1	895
TLL1974	ND	*Synechococcus elongatus* BP-1 (B)	Q8DHI5_SYNEL	BAC09526.1	529
TLL1277	ND	*Synechococcus elongatus* BP-1 (B)	Q8DJE8_SYNEL	BAC08829.1	785
TLR2270	ND	*Synechococcus elongatus* BP-1 (B)	Q8DGP5_SYNEL	BAC09822.1	852
SLL0735	ND	*Synechocystis* sp. PCC6803 (B)	P74630_SYNY3	BAA18743.1	529
SLR0337	ND	*Synechocystis* sp. PCC6803 (B)	Q55545_SYNY3	BAA10043.1	729
TTE1931	ND	*Thermoanaerobacter tengcongensis* MB4 (B)	Q8R8R4_THETN	AAM25110.1	875
Amylopullulanase	**3.2.1.1/41**	***Thermococcus hydrothermalis* (A)**	**Q9Y8I8_THEHY**	**AAD28552.1**	**1310**
4-α-Glucanotransferase	**2.4.1.25**	***Thermococcus litoralis* (A)**	**O32462_THELI**	**BAA22063.1**	**659**
Amylopullulanase	3.2.1.1/41	*Thermococcus litoralis* (A)	Q8NKS8_THELI	BAC10983.1	1089
TA0339	ND	*Thermoplasma acidophilum* DSM1728 (A)	Q9HL91_THEAC	CAC11483.1	380

Table 1. (Continued).

Enzyme (hypothetical protein)	EC	Microorganism	Abbreviation	GenPept	Length
TA0129	ND	*Thermoplasma acidophilum* DSM1728 (A)	Q9HLU6_THEAC	CAC11276.1	1641
TVG0421416 (α-amylase)	ND	*Thermoplasma volcanium* GSS1 (A)	Q97BM4_THEVO	BAB59573.1	378
TP0358	ND	*Treponema palidum* (B)	O83377_TREPA	AAC65344.1	526
TP0147 (α-amylase)	ND	*Treponema palidum* (B)	O83182_TREPA	AAC65134.1	619
α-Galactosidase (patent)	ND	Unknown prokaryote (?)	UNKP	AAE28307.1	346

Enzyme assay

Owing to the extremely low activity of the mutants, measurement of mutant enzyme-catalysed hydrolysis was performed in the presence of sodium azide using 2-chloro-4-nitrophenyl-α-D-maltotriose as the substrate. The initial rate of 2-chloro-4-nitrophenol release was monitored by spectrophotometry at 401 nm. For this method, 180 μL of 2-chloro-4-nitrophenyl-α-D-maltotriose (1.25 mM in 50 mM sodium acetate, 5 mM $CaCl_2$, 0.55 M sodium azide, pH 5.5) was preincubated at 80 °C for 10 min before adding 20 μL of enzyme solution. Afterwards, aliquots of the reaction (25 μL) were removed at regular intervals for spectrophotometric analysis. Free 2-chloro-4-nitrophenol was quantified after the addition of 975 μL of Na_2CO_3 (50 mM). One unit (U) of activity was defined as the quantity of enzyme necessary to release 1 μmol of 2-chloro-4-nitrophenol per min under the assay conditions, using 2-chloro-4-nitrophenol as the standard. For the determination of kinetic parameters, K_m and V_{max}, substrate concentration was varied over the range and the measured initial velocities were analysed using SIGMAPLOT equipped with the kinetic module 1.0 (SPSS Science, Paris, France).

Results and Discussion

Sequence comparison

This study presents results from the first detailed comparison and alignment of all available and complete amino acid sequences of GH-57 members. With regard to the origin of GH-57 enzymes, our data support the view that most members are derived from microorganisms belonging to either the Bacteria domain (24 members of 59) or, most frequently, the Archaea domain (Table 1). Importantly, a substantial proportion of the GH-57 members were isolated from hyperthermophilic microorganisms. The extreme sequence diversity in GH-57 is well illustrated by the sequence lengths, which vary from 346 to 1641 amino acid residues (Table 1). In an effort to prepare the most representative and complete sample of GH-57, the final set of 59 sequences (Table 1) was collected according to the information at CAZy [2] and Pfam [20]. Although the Pfam database (entry PF03065) [20] already provides an alignment of GH-57 members, which allows the generation of an evolutionary tree, our alignment is much more extensive, because the vast majority of the aligned sequences are complete. Therefore, our alignment provides an almost complete picture of GH-57.

In our alignment, in certain cases the extra N- and C-terminal ends were omitted. In the case of Q9Y8I8_THEHY, the excised sequence corresponds to three regions that were originally described as a SLH-like domain (SLD2), a threonine-rich region and a putative transmembrane domain [22]. Interestingly, the 3D structure of a protein domain, which is clearly homologous to SLD1 and -2 of Q9Y8I8_THEHY, was described in the GH-15 glucodextranase from *Arthrobacter globiformis* [33].

Although the sequence similarity between some members of the GH-57 may be high, it was very difficult to find corresponding sequence segments throughout the whole family. This problem can be attributed not only to the previously mentioned sequence diversity, but also to a lack of relevant information concerning structure–function relationships. However, for practical purposes, as the 3D structure of the 4-α-glucanotransferase from *T. litoralis* [34] is now available [16], we considered this enzyme to be a paradigm for GH-57. On the basis of our study, we propose that five short sequence motifs are conserved in all GH-57 members (Fig. 1).

Our more extensive alignment shows that several groups of closely related GH-57 members can be identified. These groups might correspond to GH-57 subfamilies. The evolutionary trees that are described in detail below support this supposition. The mutual relatedness of the individual subfamily members can be seen not only in the complete alignment, but also from the inspection of the five conserved sequence regions (Fig. 1).

The first conserved sequence motif (region I, consensus sequence His-Gln-Pro), although short, is strongly conserved throughout the family. With reference to *T. litoralis* 4-α-glucanotransferase, this motif is positioned near the C-terminus of the first β-strand of the catalytic $(\beta/\alpha)_7$-barrel [16]. Interestingly, the three shortest GH-57 members, which include the *P. furiosus* α-galactosidase (Q9HHB5_PYRFU), exhibit the noncanonical sequence 7_His-Gly-Asn (Q9HHB5_PYRFU numbering) in place of the consensus sequence His-Gln-Pro. However, these sequences also posses invariant Gln11 and Pro16 residues further along (analogous to residues Gln14 and Pro15 in O32462_THELI and to residues Gln16 and Pro17 in Q9Y8I8_THEHY) that might correspond to the Gln-Pro dipeptide. Importantly, together with the Glu79 (O32462_THELI numbering) from region II, His13 constitutes one of the two best-conserved residues in that region of GH-57 sequence which precedes the catalytic nucleophile, Glu123, in *T. litoralis* 4-α-glucanotransferase. Considering the extremely high level of diversity in GH-57, these two residues will be obvious candidates for future site-directed mutagenesis studies. The second motif

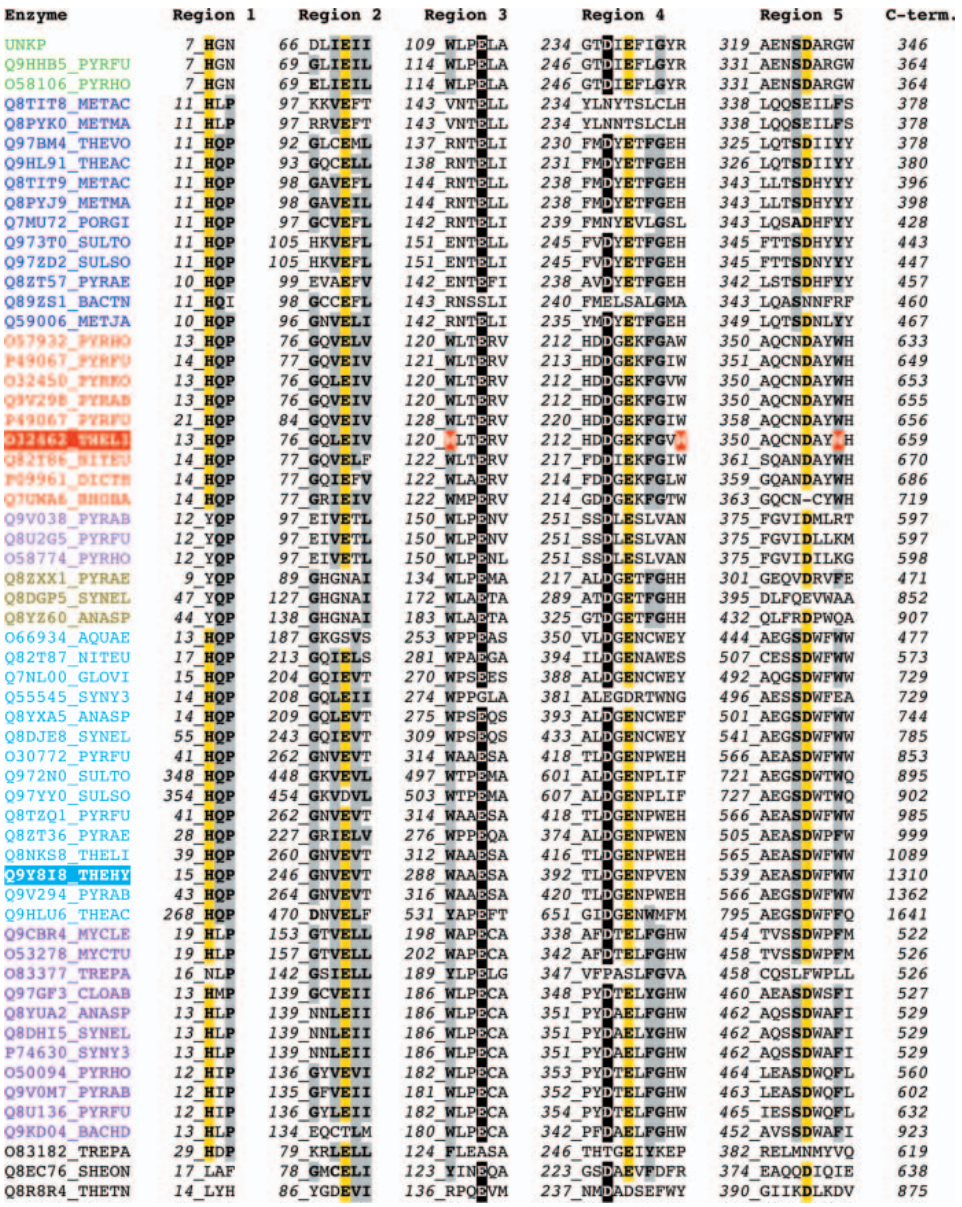

Fig. 1. Conserved sequence regions in the family GH-57. Abbreviations used for the GH-57 members are listed in Table 1. Most sequences are arranged into the seven subfamilies, with only three being more or less independent members (coloured black). Within a given subfamily, members are ordered according to increasing sequence length and, in the case of equal lengths, alphabetically. The division is based on the evolutionary trees (Fig. 4). For the amylopullulanase from *Thermococcus hydrothermalis* (Q9Y8I8_THEHY), the numbering of the mature enzyme is used [23]. The two GH-57 catalytic residues – Glu291 and Asp394 (Q9Y8I8_THEHY) – are highlighted in black. The four potentially important residues – His15, Glu249, Glu396 and Asp543 in Q9Y8I8_THEHY – are highlighted in yellow. Based on inspection of the 3D structure, the three additional aromatic residues – Trp120, Trp221 and Trp357 in O32462_THELI (highlighted in red) – could be of experimental interest, too. The residues conserved at least at 50% level are highlighted in grey.

(region II), which forms the third β-strand (β3) of the $(\beta/\alpha)_7$-barrel, belongs to the best-conserved regions in all members (Fig. 1). However, remarkably Glu79 (analogous to Glu249 in Q9Y8I8_THEHY) has no equivalent in six sequences, of which three are closely related (Q8ZXX1_PYRAE, Q8DGP5_SYNEL and Q8YZ60_ANASP) and very probably constitute a GH-57 subfamily. Intriguingly, examination of the crystal structure of the *T. litoralis* 4-α-glucanotransferase complexed with acarbose [16], does not allow a role to be assigned to Glu79. In contrast, His13 has been found to be involved in the subsite-1 [16], together with the other His residue occupying position *i-2* with respect to His13 (data not shown).

On the basis of comparison with *T. litoralis* 4-α-glucanotransferase, the conserved sequence regions III and IV should contain the two catalytic residues, Glu123 (identified as a catalytic nucleophile) [35] and Asp214 (proposed as a proton donor) [16]. Structurally, both of these residues are located near the C-termini of the strands β4 and β7 of the catalytic $(\beta/\alpha)_7$-barrel [16]. However, these residues have no

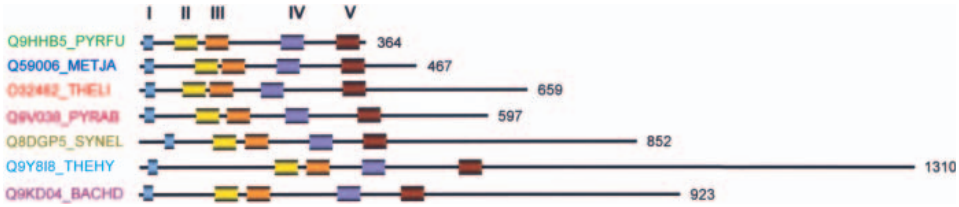

Fig. 2. Schematic view of conserved sequence regions in GH-57 representatives. Seven sequences, which are representative members of the seven GH-57 subfamilies, are used to illustrate the conserved regions. The individual conserved sequence regions are shown as rectangles, as follows: I, blue; II, yellow; III, orange; IV, violet; V, brown. The sequence lengths of the seven representatives are also indicated. The abbreviated member names are defined in Table 1.

equivalents in some GH-57 members: Ser, Gly or Ala in Q89ZS1_BACTN, Q55545_SYNY3 and O83182_TREPA replaces Glu123, respectively, while Asp214 is even more variable. It is substituted three times with Asn (Q8TIT8_METAC, Q8PYK0_METMA and Q7MU72_PORGI), twice with Glu (Q89ZS1_BACTN and Q55545_SYNY3) and once with Pro (O83377_TREPA) or Thr (O83182_TREPA). These observations could be explained by the fact that, at the present time, all of these GH-57 members are only hypothetical proteins for which no enzyme activity has been demonstrated.

The fifth conserved sequence region (region V) (Fig. 1) belongs to a structural motif that includes a three-helix bundle which participates in the active site cleft at the C-terminus of the $(\beta/\alpha)_7$-barrel of the *T. litoralis* 4-α-glucanotransferase [16]. It contains a well-conserved aspartate residue, Asp354 (O32462_THELI numbering; analogous to Glu543 in Q9Y8I8_THEHY), which has been shown to interact with the two active-site water molecules [16]. According to our alignment, this residue possesses no equivalent in seven GH-57 members (Fig. 1), all of the seven being hypothetical proteins.

Recently, in order to identify the residues responsible for catalysis, site-directed mutagenesis was performed on a GH-57 α-galactosidase from *P. furiosus* [36]. This protein is among the shortest members of GH-57 and exhibits an unusual specificity towards galactosidic bonds. The alignment and mutagenesis strategy employed by van Lieshout *et al.* [36] allowed the identification of Glu117 as the catalytic nucleophile (analogous to residue Glu123 in O32462_THELI and to residue Glu291 in Q9Y8I8_THEHY), which is in good agreement with the alignment presented in this work (Fig. 1). However, with regard to the catalytic acid-base, in our opinion these authors misaligned the succeeding parts of the GH-57 sequences and therefore falsely identified Glu193 as the best candidate. Upon mutagenesis, this error was confirmed, as the corresponding Glu193Ala displayed significant residual activity [36]. This is not surprising, because according to the Henrissat's classification criteria [8], all members of a glycoside hydrolase family should have identical catalytic machinery. Therefore, one would expect that, like *T. litoralis* 4-α-glucanotransferase, in all GH-57 members the catalytic acid-base should be an aspartate residue. Accordingly, in our alignment, Asp248 (analogous to residue Asp214 in O32462_THELI and to residue Asp394 in Q9Y8I8_THEHY) is predicted to play the role of proton donor in the *P. furiosus* α-galactosidase (Q9HHB5_PYRFU; Fig. 1). Importantly, this example of the *P. furiosus* α-galactosidase highlights the difficulties associated with the alignment of sequences that display substantial length variation and sequential diversity. Such differences are clearly illustrated by the distances between the individual conserved sequence regions (Fig. 2), e.g. the III-to-IV insertion in *P. furiosus* α-galactosidase or the I-to-II insertion in *T. hydrothermalis* amylopullulanase, in comparison to the corresponding distances in *T. litoralis* 4-α-glucanotransferase (Fig. 2).

In order to see how the five conserved sequence regions, and especially the proposed potentially functional residues (His13, Glu79, Glu216 and Asp354), are arranged in the structure of a GH-57 member, Fig. 3 was prepared using the X-ray coordinates of the 4-α-glucanotransferase from *T. litoralis*. It is evident that at least three of the four residues, corresponding to His13, Glu216 and Asp354 of *T. litoralis* 4-α-glucanotransferase, might play a functional role in GH-57. Concerning the Glu79, its side-chain is oriented far from the catalytic (active) centre, but its functional meaningless has to be verified experimentally. The fact that this residue is conserved in 90% of GH-57 members (Fig. 1) is worth mentioning. Based on the inspection of the structure (Fig. 3), we concluded that also the three aromatic residues, corresponding to Trp120, Trp221 and Trp357 of *T. litoralis* 4-α-glucanotransferase (Fig. 1), should be involved in our future site-directed mutagenesis studies.

To provide experimental support for our alignment data, we chose the *T. hydrothermalis* amylopullulanase as a candidate for structure/function studies by site-directed mutagenesis. In agreement with the alignment, we propose that in this enzyme Glu291 and Asp394 are the catalytic nucleophile and proton donor, respectively. Additionally, we propose that His15, Glu249, Glu396 and Asp543 will prove to be important residues (Fig. 1).

Site-directed mutagenesis

With regard to our prediction concerning the catalytic residues in *T. hydrothermalis* amylopullulanase, the residues Glu291 and Asp394 were substituted by alanine. These mutations led to the abolition of detectable activity towards both pullulan and amylose in both Petri dish tests and reducing sugar assays (data not shown). Similarly, no activity was detected in the presence of the more reactive substrate, 2-chloro-4-nitrophenyl-α-D-maltotriose. Consequently, in order to measure hydrolyses catalysed by the mutant enzymes, nucleophilic azide ions were included in

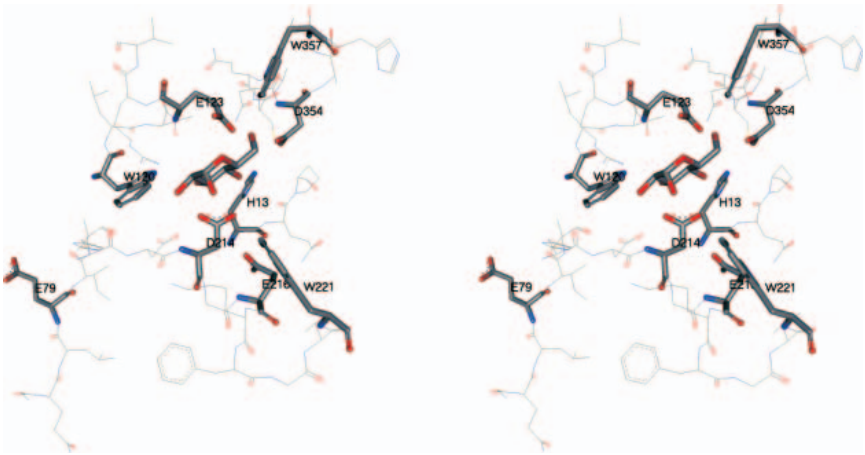

Fig. 3. Active site of the 4-α-glucanotransferase from *Thermococcus litoralis*. The segments of the five conserved sequence regions identified in this study are shown with highlighted catalytic residues (E123, catalytic nucleophile; and D214, proton donor) as well as the residues H13, E79, E216 and D354, proposed as important for the GH-57 members. The residues of *T. litoralis* 4-α-glucanotransferase (O32462_THELI) correspond to the residues of *T. hydrothermalis* amylopullulanase (Q9Y8I8_THEHY), as follows: Glu123 (Glu291), Asp214 (Asp394), His13 (His15), Glu79 (Glu249), Glu216 (Glu396) and Asp354 (Asp543). Also, the three tryptophans (W120, W221 and W357, highlighted), as well as the residues in the corresponding positions in other GH-57 members, could be of interest. The glucose molecule (in the middle) is also shown. The PDB X-ray coordinates, 1K1W, were used [16]. The figure was created using the WEBLAB VIEWERLITE 4.0 (Molecular Simulations, Inc.).

the reaction medium [37,38]. Likewise, it was possible to detect low, but measurable, activities for both mutants (Table 2). Even in the presence of azide, V_{max} values for both mutant enzymes were 10^3-fold lower than that of the wild-type enzyme. However, with regard to the K_m values, Asp394Ala displayed a nearly wild-type value, whereas Glu291Ala displayed reduced substrate affinity. These results indicate that Glu291 and Asp394 are both critical for the hydrolytic activities of *T. hydrothermalis* amylopullulanase and, in contrast to previously described data [24], support the notion of a single active site responsible for both amylolytic and pullulanolytic activities. Additionally, it is noteworthy that although substitution of either residue abolished hydrolytic activity, CD spectra indicated that both mutant enzymes were correctly folded. This conclusion is also supported by the fact that the reactivation of the enzymes could be achieved by the addition of an external nucleophile to the reaction medium. Gratifyingly, in *T. hydrothermalis* amylopullulanase, the identification (by site-directed mutagenesis) of Glu291 and Asp394 as the catalytic pair (based on our sequence comparison; Fig. 1) is in good agreement with the known catalytic residues of *T. litoralis* 4-α-glucanotransferase [16,35]. Finally, our results fulfil the original Henrissat's criteria concerning the conservation of the catalytic machinery [8].

Table 2. Kinetic parameters for 2-chloro-4-nitrophenyl-α-D-maltotriose hydrolysis catalysed by ThApuΔ2 and mutant derivatives.

Enzyme	V_{max} (IU)	K_M (mM)
Th-ApuΔ2[a]	45 652 ± 1428	0.75 ± 0.02
Glu291Ala	53.69 ± 7.7	3.21 ± 0.7
Asp394Ala	84.55 ± 5.5	0.92 ± 0.11

[a] Measured in the absence of azide.

Evolutionary relationships

In order to draw the present-day evolutionary picture of the family GH-57, several evolutionary trees were constructed. Figure 4 shows two trees. The first (Fig. 4A) is based on the complete alignment of sequences with the gaps included for the calculation, whereas the second (Fig. 4B) is based on the conserved sequence regions. As can be seen from the clustering of the family members in the trees, the entire present-day GH-57 can be divided into seven subfamilies, plus three more or less independent members (O83182_TREPA, Q8EC76_SHEON and Q8R8R4_THETN). At present, these three members can be considered as independent because new GH-57 members with sequences closely related to them may emerge in the future. It is also highly probable that in the future further subfamilies will be identified, owing to the appearance of new members or by subdivision of the existing subfamilies. Indeed, there are several GH-57 candidates in the unfinished sequencing genome projects (as revealed by BLAST) – both from Archaea and bacteria: *Ferroplasma acidarmanus* (GenPept accession number: ZP_00000807.1, length: 377), *Methanosarcina barkeri* (ZP_00079232.1, 378; ZP_00079233.1, 398), *Cytophaga hutchinsonii* (ZP_00116896.1, 397), *Geobacter metallireducens* (ZP_00080528.1, 659; ZP_00082306.1, 740), and *Nostoc punctiforme* (ZP_00108689.1, 742). Likewise, the possibility that certain members will be separated (e.g. Q8TIT8_METAC and Q8PYK0_METMA – blue; Q97YY0_SULSO and Q972N0_SULTO – turquoise; O83377_TREPA – violet), leading to the establishment of new subfamilies, cannot be excluded. Moreover, the fusion of other subfamilies to form larger ones can be expected.

With regard to enzyme specificities that characterize the individual GH-57 subfamilies, several subfamilies are exclusively composed of hypothetical proteins. Therefore, at present it is impossible to form any conclusions for

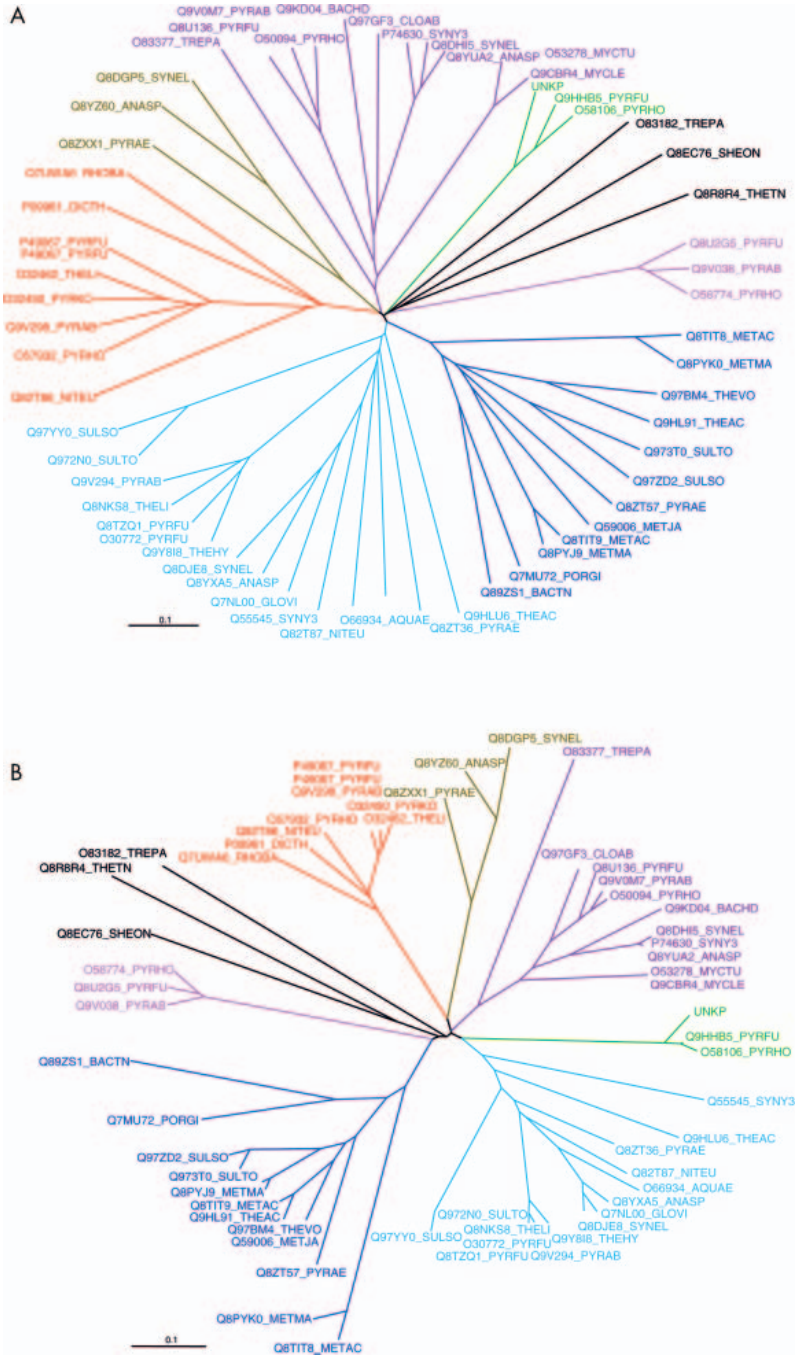

Fig. 4. Evolutionary trees of the family GH-57. The trees are based on (A) complete alignment including the gaps, and (B) conserved sequence regions. Branch lengths are proportional to sequence divergence. The seven subfamilies are colour coded, with only three being more or less independent members (coloured black). The abbreviated member names are defined in Table 1.

these. On the other hand, three subfamilies contain experimentally characterized enzymes (Table 1), such as α-galactosidase (Q9HHB5_PYRFU; green), α-amylase and 4-α-glucanotransferase (P49067_PYRFU, P09961_DICTH, O32450_PYRKO and O32462_THELI; red), and amylopullulanase (O30772_PYRFU, Q8NKS8_THELI and Q9Y8I8_THEHY; turquoise). As the α-galactosidase from *P. furiosus* exhibits neither amylase nor amylopullulanase activity [39], this subfamily could be a pure α-galactosidase subfamily. As for the subfamily containing both α-amylases and 4-α-glucanotransferases, the latter specificity was unambiguously demonstrated for the enzymes from *T. litoralis* [34] and *P. kodakaraensis* [18]. Interestingly the α-amylase from *P. furiosus* [40] also displayed 4-α-glucanotransferase activity. Unfortunately, the biochemical information available for the *D. thermophilum*

enzyme [13] does not permit an unambiguous conclusion to be reached and leaves open the question of the presence of the α-amylase specificity in this subfamily. With regard to the amylopullulanase-containing subfamily, both amylolytic and pullulanolytic activities were confirmed for the amylopullulanases from *P. furiosus* [17] and *T. hydrothermalis* [23]. However, both the data presented here and \the unpublished data of F. Chang-Pi-Hin, L. Greffe, H. Driguez & M. J. O'Donohue, unpublished results), concerning the characterization of the active site of the *T. hydrothermalis* enzyme, clearly demonstrate that both activities are defined by a unique active site. Therefore, these enzymes can be considered to be true amylopullulanases and not bifunctional, dual-domain α-amylase pullulanases.

Finally, it is noteworthy that the evolutionary relatedness of the individual GH-57 subfamilies can be inferred from the trees (Fig. 4; see also Supplementary material). When comparing the arrangement in the trees, subtle modifications and rearrangements can be found, i.e. those concerning either the relationships within a subfamily or the relatedness between the subfamilies (Fig. 4). Importantly, the overall integrity of all subfamilies was saved in all trees, including the Pfam-tree, based on simplified alignment of ≈ 300 N-terminal amino acid residues. Therefore, together with the proposed conserved sequence regions (Fig. 1), our alignment constitutes a valid base for the identification of other functional residues in both the present and future GH-57 members.

Acknowledgements

The authors wish to thank both the Slovak Grant Agency for Science (VEGA grant no. 2/2057/24) and Europol'Agro (Conseil Général de la Marne) for financial support. Mr Rolland Monserret (IBCP-Lyon, France) is thanked for the CD analyses and Mrs Béatrice Hermant for her skilful technical assistance.

References

1. MacGregor, E.A., Janeček, Š. & Svensson, B. (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. *Biochim. Biophys. Acta* **1546**, 1–20.
2. Coutinho, P.M. & Henrissat, B. (1999) Carbohydrate-active enzymes: an integrated database approach. In *Recent Advances in Carbohydrate Bioengineering* (Gilbert, H.J., Davies, G., Henrissat, B. & Svensson, B., eds), pp. 3–12. The Royal Society of Chemistry, Cambridge, UK.
3. Matsuura, Y. (2002) A possible mechanism of catalysis involving three essential residues in the enzymes of α-amylase family. *Biologia, Bratislava* **57** (Suppl. II), 21–27.
4. Oslancová, A. & Janeček, Š. (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.
5. MacGregor, E.A., Jespersen, H.M. & Svensson. B. (1996) A circularly permuted α-amylase-type α/β-barrel structure in glucan-synthesizing glucosyltransferases. *FEBS Lett.* **378**, 263–266.
6. Przylas, I., Tomoo, K., Terada, Y., Takaha, T., Fujii, K., Saenger, W. & Sträter, N. (2000) Crystal structure of amylomaltase from *Thermus aquaticus*, a glycosyltransferase catalysing the production of large cyclic glucans. *J. Mol. Biol.* **296**, 873–886.
7. Janeček, Š. (2002) How many conserved sequence regions are there in the α-amylase family? *Biologia, Bratislava* **57** (Suppl. II), 29–41.
8. Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem. J.* **280**, 309–316.
9. Rye, C.S. & Withers, S.G. (2000) Glycosidase mechanisms. *Curr. Opin. Chem. Biol.* **4**, 573–580.
10. Mikami, B., Hehre, E.J., Sato, M., Katsube, Y., Hirose, M., Morita, Y. & Sacchettini, J.C. (1993) The 2.0-Å resolution structure of soybean β-amylase complexed with α-cyclodextrin. *Biochemistry* **32**, 6836–6845.
11. Aleshin, A., Golubev, A., Firsov, L.M. & Honzatko, R.B. (1992) Crystal structure of glucoamylase from *Aspergillus awamori* var. X100–2.2-Å resolution. *J. Biol. Chem.* **267**, 19291–19298.
12. Frandsen, T.P. & Svensson, B. (1998) Plant α-glucosidases of the glycoside hydrolase family 31. Molecular properties, substrate specificity, reaction mechanism, and comparison with family members of different origin. *Plant Mol. Biol.* **37**, 1–13.
13. 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.
14. Laderman, K.A., Asada, K., Uemori, T., Mukai, H., Taguchi, Y., Kato, I. & Anfinsen, C.B. (1993) 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.
15. Henrissat, B. & Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. *Biochem. J.* **316**, 695–696.
16. Imamura, H., Fushinobu, S., Yamamoto, M., Kumasaka, T., Jeon, B.S., 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.
17. Dong, G., Vieille, C. & Zeikus, J.G. (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.
18. 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.
19. Janeček, Š. (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.
20. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. & Sonnhammer, E.L. (2002) The Pfam protein families database. *Nucleic Acids Res.* **30**, 276–280.
21. Gantelet, H. & Duchiron, F. (1998) Purification and properties of a thermoactive and thermostable pullulanase from *Thermoccocus hydrothermalis*, a hyperthermophilic archaeon isolated from deep sea hydrothermal vent. *Appl. Microbiol. Biotechnol.* **49**, 770–777.
22. Erra-Pujada, M., Debeire, P., Duchiron, F. & O'Donohue, M.J. (1999) The type II pullulanase of *Thermococcus hydrothermalis*: molecular characterization of the gene and expression of the catalytic domain. *J. Bacteriol.* **181**, 3284–3287.
23. Erra-Pujada, M., Chang-Pi-Hin, F., Debeire, P., Duchiron, F. & O'Donohue, M.J. (2001) Purification and properties of the catalytic domain of the thermostable pullulanase type II from *Thermococcus hydrothermalis*. *Biotechnol. Lett.* **23**, 1273–1277.
24. Chang-Pi-Hin, F., Erra-Pujada, M., Dauchez, M., Debeire, P., Duchiron, F. & O'Donohue, M.J. (2002) Expression and characterization of the catalytic domain of an archaeal family 57 pullulanase type II. *Biologia, Bratislava* **57** (Suppl. II), 155–162.
25. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. & Wheeler, D.L. (2004) GenBank: update. *Nucleic Acids Res.* **32**, D23–D26.

26. Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., Martin, M.J., Natale, D.A., O'Donovan, C., Redaschi, N. & Yeh, L.S. (2004) UniProt: the Universal Protein knowledgebase. *Nucleic Acids Res.* **32**, D115–D119.
27. Bourne, P.E., Addess, K.J., Bluhm, W.F., Chen, L., Deshpande, N., Feng, Z., Fleri, W., Green, R., Merino-Ott, J.C., Townsend-Merino, W., Weissig, H., Westbrook, J. & Berman, H.M. (2004) The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res.* **32**, D223–D225.
28. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
29. Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
30. Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
31. Page, R.D. (1996) TREEVIEW: an application to display phylogenetic trees on personal computers. *Comput. Applic. Biosci.* **12**, 357–358.
32. Altschul, S.F., Stephen, F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402.
33. Mizuno, M., Tonozuka, T., Suzuki, S., Uotsu-Tomita, R., Kamitori, S., Nishikawa, A. & Sakano, Y. (2004) Structural insights into substrate specificity and function of glucodextranase. *J. Biol. Chem.* **279**, 10575–10583.
34. Jeon, B.S., 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.
35. Imamura, H., Fushinobu, S., Jeon, B.S., Wakagi, T. & Matsuzawa, H. (2001) Identification of the catalytic residue of *Thermococcus litoralis* 4-α-glucanotransferase through mechanism-based labeling. *Biochemistry* **40**, 12400–12406.
36. van Lieshout, J.F.T., Verhees, C.H., Ettema, T.J.G., van der Sar, S., Imamura, H., Matsuzawa, H., van der Oost, J. & de Vos, W.M. (2003) Identification and molecular characterization of a novel type of α-galactosidase from *Pyrococcus furiosus*. *Biocatal. Biotransform.* **21**, 243–252.
37. MacLeod, A.M., Tull, D., Rupitz, K., Warren, R.A. & Withers, S.G. (1996) Mechanistic consequences of mutation of active site carboxylates in a retaining β-1,4-glycanase from *Cellulomonas fimi*. *Biochemistry* **35**, 13165–13172.
38. Cobucci-Ponzano, B., Trincone, A., Giordano, A., Rossi, M. & Moracci, M. (2003) Identification of the catalytic nucleophile of the family 29 α-L-fucosidase from *Sulfolobus solfataricus* via chemical rescue of an inactive mutant. *Biochemistry* **42**, 9525–9531.
39. van Lieshout, J.F.T., Verhees, C.H., van der Oost, J. & de Vos, W.M. (2001) Identification of active-site residues of an archaeal α-galactosidase, a unique member of glycosyl hydrolase family 57. In *1st Symposium on the α-Amylase Family* (Janeček, Š., ed.), p. 26. ASCO Art & Science, Bratislava.
40. Laderman, K.A., Davis, B.R., Krutzsch, H.C., Lewis, M.S., Griko, Y.V., Privalov, P.L. & Anfinsen, C.B. (1993) The purification and characterization of an extremely thermostable α-amylase from the hyperthermophilic archaebacterium *Pyrococcus furiosus*. *J. Biol. Chem.* **268**, 24394–24401.

Supplementary material

The following material is available from http://blackwellpublishing.com/products/journals/suppmat/EJB/EJB4144/EJB4144sm.htm

Table S1. The enzymes and proteins from the family GH-57 used in the present study (extended coloured version from the manuscript with active links to Accession Numbers in the sequence databases).

Fig. S1. Alignment of GH-57 sequences.

Fig. S2. A tree based on alignment of GH-57 sequences (gaps excluded).

Fig. S3. Pfam tree (our version of the Pfam tree; Pfam entry: PF03065; July 2003).