FEBS Letters 581 (2007) 1261–1268

A remote but significant sequence homology between glycoside hydrolase clan GH-H and family GH31

Štefan Janeček[a,*], Birte Svensson[b], E. Ann MacGregor[c]

[a] Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84551 Bratislava, Slovakia
[b] Enzyme and Protein Chemistry, BioCentrum-DTU, Technical University of Denmark, Søltofts Plads, DK-2800 Kgs. Lyngby, Denmark
[c] 2 Nicklaus Green, Livingston, EH54 8RX, West Lothian, UK

Received 11 December 2006; revised 2 February 2007; accepted 19 February 2007

Available online 28 February 2007

Edited by Hans Eklund

Abstract **Although both the α-amylase super-family, i.e. the glycoside hydrolase (GH) clan GH-H (the GH families 13, 70 and 77), and family GH31 share some characteristics, their different catalytic machinery prevents classification of GH31 in clan GH-H. A significant but remote evolutionary relatedness is, however, proposed for clan GH-H with GH31. A sequence alignment, based on the idea that residues equivalent in the primordial catalytic GH-H/GH31 $(\beta/\alpha)_8$-barrel may not be found in the present-day GH-H and GH31 structures at strictly equivalent positions, shows remote sequence homologies covering β3, β4, β7 and β8 of the GH-H and GH31 $(\beta/\alpha)_8$-barrels. Structure comparison of GH13 α-amylase and GH31 α-xylosidase guided alignment of GH-H and GH31 members for construction of evolutionary trees. The closest sequence relationship displayed by GH31 is to GH77 of clan GH-H.**

Keywords: α-Amylase; α-Xylosidase; Glycoside hydrolase family 13; Glycoside hydrolase family 31; Structural alignment; Remote homology; Evolutionary tree

1. Introduction

Many proteins and enzymes form structural families, often without obvious functional similarity. On the other hand, functionally related proteins and enzymes exist that clearly do not share evolutionary homology. Glycoside hydrolases (GHs) can be very good examples of the above-mentioned phenomena. Because of the great variety of naturally occurring saccharides, a large number of carbohydrate-metabolising enzymes have evolved. A sequence-based classification system of GHs [1] available at the CAZy (*C*arbohydrate-*A*ctive en*Zy*mes) web-site [2] is independent of EC numbers given to members of any GH family. The individual GH families are defined solely by similarities in primary structures which, in turn, are found to reflect reaction mechanism, catalytic machinery and fold, i.e. evolution. Of the more than 100 GH families defined at present in CAZy, the typical starch hydrolases and related enzymes are found in families 13, 14, 15, 31, 57, 70 and 77 [3–5]. Family GH13 is well known as the α-amylase family and, together with GH70 and GH77, forms clan GH-H [6–10]. This analysis focuses on clan GH-H and family GH31.

Family GH13 is one of the largest GH families, with more than 2500 sequences [2]. The enzymes have almost 30 different specificities, such as cyclodextrin glucanotransferase, α-glucosidase, isoamylase, neopullulanase, branching enzyme, etc., in addition to α-amylase [3–10]. Further, three-dimensional structures are known for enzymes representing approximately 20 of these specificities. The main structural feature is the central catalytic $(\beta/\alpha)_8$-barrel domain (i.e. a TIM-barrel fold), in most cases having a distinct domain (called domain B) protruding from the barrel in the place of loop 3 (β3 → α3 connection) [11,12]. Family GH77 enzymes contain this $(\beta/\alpha)_8$-barrel, but lack domain C found C-terminal to the barrel in GH13 enzymes (see, for example [13]). Family GH70, in contrast, is believed to possess a circularly permuted version of the GH13-type $(\beta/\alpha)_8$-barrel [14]. Enzymes of the entire clan GH-H are characterised by several (from 4 to 7) conserved sequence regions [15,16] and a common catalytic machinery involving an aspartate in strand β4, a glutamate in strand β5, and an aspartate after strand β7 that are essential for activity [3–11,17]. These three constitute the only invariant residues in GH-H [18], aspartate at the β4-strand acting as catalytic nucleophile and the β5-strand glutamate as general acid/base catalyst [19].

GH13 and GH31 share a retaining catalytic mechanism and enzymes with α-glucosidase specificity [2,6,20]. At the sequence level, however, similarity between the two families is not obvious, but remote evolutionary relatedness between GH13 and GH31 was indicated by results achieved with iterative database searches [21] and fold recognition threading methods [22]. The former study identified conservation of the catalytic nucleophile aspartate at β4-strand in both families, while the latter revealed structural similarities between the catalytic domains. These observations were confirmed when determination of structures [23,24] showed that basic characteristics of GH31 are shared with GH13, i.e. a $(\beta/\alpha)_8$-barrel catalytic domain carrying an excursion at the position of loop 3. In GH31, however, two aspartate residues situated at strands β4 and β6 act as catalytic nucleophile and general acid/base catalyst [23–25]. Thus families GH13 and GH31 share the catalytic aspartate at β4, whereas the acid/base is glutamate at β5 in GH13 and aspartate at β6 in GH31. Since the catalytic machinery is not conserved, it is not possible to group the two families in a common clan according to the GH clan definition [26].

*Corresponding author. Fax: +421 2 5930 7416.
E-mail address: Stefan.Janecek@savba.sk (Š. Janeček).

Abbreviations: GH, glycoside hydrolase; TIM, triose phosphate isomerase

0014-5793/$32.00
doi:10.1016/j.febslet.2007.02.036

The main goal of this article is to demonstrate a very remote but significant homology between GH13 and GH31, based on a bioinformatics analysis of the two GH families. It is proposed that both families might have evolved from a common ancestor. In addition, some evolutionary relationships within clan GH-H are also investigated.

2. Materials and methods

The amino acid sequences representing the individual enzyme specificities of both the α-amylase "super-family" (i.e. the clan GH-H) and family GH31 used in the present study are listed in Table 1. To include a specific sequence in the analysed set, two main criteria were applied: (i) the protein should represent a biochemically well-characterised enzyme specificity; and (ii) members with solved three-dimensional structure were preferred. Sequences and structures were retrieved from the SwissProt [27] and PDB [28] databases, respectively. Three-dimensional structures were displayed with the program WebLabViewerLite 4.0 (Accelrys Ltd., Cambridge, UK; http://www.accelrys.com/).

Multiprot [29] structural alignments of the following pairs of enzymes were carried out: GH13 α-amylase from *Aspergillus oryzae* (PDB code 7TAA) [30] with chicken triose phosphate isomerase (PDB code 1TIM) [31] or GH31 α-xylosidase from *Escherichia coli* (PDB code 1XSJ) [23] or GH31 α-glucosidase from *Sulfolobus solfataricus* (PDB code 2G3M) [24], and the chicken triose phosphate isomerase with the *E. coli* α-xylosidase.

Sequence alignment of the GH-H and GH31 members listed in Table 1 was based on the Multiprot structural alignment of representatives of the two GH families: GH13 α-amylase from *A. oryzae* [30] and *E. coli* α-xylosidase of GH31 [23]. The remaining specificities from both GH-H and GH31 were added to this structural alignment of GH13 α-amylase and GH31 α-xylosidase, based mainly on the authors' prior experience and background in bioinformatics analysis of these enzymes and using the information retrieved from the Pfam database [32] and obtained by Blast searches [33]. For routine aligning, the program ClustalW [34] was used. Information already available on the conserved sequence regions in clan GH-H [14,16] was also used. The GH-H/GH31 alignment spanned the $(\beta/\alpha)_8$-barrel domain (i.e. from β1 to α8) including the domain B between β3 and α3. For the two GH70 members, domain B was omitted since no three-dimensional structure is available and there is large uncertainty in that region due to circular permutation in the $(\beta/\alpha)_8$-barrel [14].

The evolutionary tree was calculated on the European Bioinformatics Institute's server for ClustalW (http://www.ebi.ac.uk/clustalw/) as Phylip-tree type [35] using the complete final alignment (from β1 to α8). Gaps in the alignment were excluded. The tree was displayed with the program TreeView [36].

Table 1
List of enzymes from clan GH-H and family GH31 used in the present study

Family	EC	Enzyme	Abbreviation	Source	SwissProt	PDB
GH13	3.2.1.1	α-Amylase	GH13-Aspor-AAMY	*Aspergillus oryzae*	P0C1B3	7taa
	2.4.1.19	Cyclodextrin glucanotransferase	GH13-Bacci-CGT	*Bacillus circulans* No.8	P30920	1cgt
	3.2.1.10	Oligo-1,6-glucosidase	GH13-Bacce-OGLU	*Bacillus cereus*	P21332	1uok
	3.2.1.60	Maltotetraohydrolase	GH13-Psest-M4H	*Pseudomonas stutzeri*	P13507	2amg
	3.2.1.68	Isoamylase	GH13-Pseam-ISA	*Pseudomonas amyloderamosa*	P10342	1bf2
	3.2.1.133	Maltogenic amylase	GH13-Thesp-MGA	*Thermus* sp. IM6501	O69007	1sma
	3.2.1.133	Maltogenic α-amylase	GH13-Bacst-MGAA	*Bacillus stearothermophilus*	P19531	1qho
	3.2.1.141	Maltooligosyltrehalose hydrolase	GH13-Sulso-MOTH	*Sulfolobus solfataricus*	Q55088	1eha
	2.4.1.4	Amylosucrase	GH13-Neipo-AMSU	*Neisseria polysaccharea*	Q9ZEU2	1g5a
	2.4.1.-	Maltosyltransferase	GH13-Thtma-MT	*Thermotoga maritima*	O33838	1gju
	2.4.1.25	4-α-Glucanotransferase	GH13-Thtma-4AGT	*Thermotoga maritima*	P80099	1lwh
	2.4.1.18	Glucan branching enzyme	GH13-Ascco-BE	*Escherichia coli*	P07762	1m7x
	3.2.1.54	Cyclomaltodextrinase	GH13-Bacsp-CMD	*Bacillus* sp. I-5	Q59226	1ea9
	3.2.1.135	Neopullulanase	GH13-Bacst-NPU	*Bacillus stearothermophilus*	P38940	1j0h
	5.4.99.11	Isomaltulose synthase	GH13-Klesp-ISMS	*Klebsiella* sp. LX3	Q8KR84	1m53
	5.4.99.15	Maltooligosyltrehalose synthase	GH13-Sulac-MOTS	*Sulfolobus acidocaldarius*	Q53688	1iv8
	2.4.1.7	Sucrose phosphorylase	GH13-Bifad-SPH	*Bifidobacterium adolescentis*	Q84HQ2	1r7a
	3.2.1.98	Maltohexaohydrolase	GH13-Bacsp-M6H	*Bacillus* sp. 707	P19571	1wp6
	3.2.1.41	Pullulanase	GH13-Klepn-PUL	*Klebsiella pneumoniae*	P07811	2fhf
	3.2.1.20	α-Glucosidase	GH13-Sacca-AGLU	*Saccharomyces carlsbergensis*	P07265	–
	3.2.1.1/41	Amylopullulanase	GH13-Thbsa-APU	*Thermoanaerobacter saccharolyticum*	P36905	–
	3.2.1.70	Dextran glucosidase	GH13-Stcmu-DGLU	*Streptococcus mutans*	Q99040	–
	3.2.1.93	Trehalose-6-phosphate hydrolase	GH13-Bacsu-T6PH	*Bacillus subtilis*	P39795	–
	3.2.1.116	Maltotriohydrolase	GH13-Nacam-M3H	*Natronococcus amylolyticus*	Q60224	–
	3.2.1.-	Maltopentaohydrolase	GH13-Psesp-M5H	*Pseudomonas* sp. KO-8940	Q52516	–
	3.2.1.-	Sucrose hydrolase	GH13-Xanax-SH	*Xanthomonas axonopodis*	Q6UVM5	–
	2.4.1.25/3.2.1.33	Glucan debranching enzyme	GH13-Orycu-GDE	*Oryctolagus cuniculus*	P35574	–
	5.4.99.16	Trehalose synthase	GH13-Pimsp-TS	*Pimelobacter* sp. R48	P72235	–
GH70	2.4.1.7	Glucosyltransferase	GH70-Stcso-GTF	*Streptococcus sobrinus*	P11001	–
	2.4.1.141	Alternan sucrase	GH70-Leume-ALSU	*Leuconostoc mesenteroides*	Q9RE05	–
GH77	2.4.1.25	4-α-Glucanotransferase	GH77-Theaq-4AGT	*Thermus aquaticus*	O87172	1esw
	2.4.1.25	4-α-Glucanotransferase	GH77-Borbu-4AGT	*Borrelia burgdorferi*	O51188	–
GH31	3.2.1.-	α-Xylosidase	GH31-Escco-AXYL	*Escherichia coli*	P31434	1xsj
	3.2.1.20	α-Glucosidase	GH31-Sulso-AGLU	*Sulfolobus solfataricus*	O59645	2g3m
	3.2.1.48/3.2.1.10	Sucrase-isomaltase	SUI	*Oryctolagus cuniculus*	P07768	–
	3.2.1.48	Sucrase subunit	GH31-Orycu-SUIs			
	3.2.1.10	Isomaltase subunit	GH31-Orycu-SUIi			
	2.4.1.-	3-α-Isomaltosyltransferase	GH31-Spogl-3IMT	*Sporosarcina globispora*	Q84IQ3	–
	2.4.1.-	6-α-Glucosyltransferase	GH31-Spogl-6GTF	*Sporosarcina globispora*	Q84IQ2	–
	4.2.2.13	α-1,4-Glucan lyase	GH31-Grale-AGLY	*Gracilariopsis lemaneiformis*	Q9STC1	–

3. Results and discussion

3.1. Structure-based comparison

The first available three-dimensional structures of GH31 enzymes [23,24] allowed the comparison of GH13 and GH31 in detail (Fig. 1). The alignment of the three-dimensional structures of the α-amylase from *A. oryzae* [30] representing GH13 and α-xylosidase from *E. coli* [23] representing GH31, using MULTIPROT [29], resulted in 202 corresponding residues spanning all three domains A, B and C of the GH13 α-amylase (root-mean-square deviation 1.89 Å between the C_α atoms). While the 11 correspondences within 129 positions for domains C are probably not significant, the remaining 191 within the consensus length of 437 residues of the catalytic $(\beta/\alpha)_8$-barrels emphasize important similarity (Table 2 and Fig. 1 a,b).

This alignment is valid beyond simply aligning any two $(\beta/\alpha)_8$-barrel proteins since the archetypal TIM-barrel protein, chicken triose phosphate isomerase [31], gave alignment lengths of 94 and 116 at root-mean-square deviations of 2.01 Å and 1.91 Å with the *A. oryzae* α-amylase and *E. coli* α-xylosidase, respectively. Furthermore, MULTIPROT alignment of the α-amylase with GH31 *S. solfataricus* α-glucosidase gave a structural alignment of 191 residues at a root-mean-square deviation of 1.90 Å, i.e. the barrels of GH31 enzymes are more similar to those of GH13 than to TIM barrels in general.

Although the catalytic machineries differ between the two GH families, remarkable similarity is observed for the side-chain orientation of the GH31 β6-aspartate (proton donor) [23–25] and the GH13 β7-aspartate (transition-state stabiliser) [19,37], and for the positioning of the respective β-strands (Fig. 1c). In addition to the structural alignment of the nucleophiles (β4-aspartate), there is also a significant correspondence between the α-xylosidase Arg466 at β5 and the α-amylase acid-base catalyst (β5-Glu230). Lovering et al. [23] showed that Arg466 is likely to interact directly with substrate, and conservation of this arginine throughout GH31 [23–25,38] indicates its importance at the active site. Its role is, however, still unknown. Thus three critical residues of both GH13 and GH31 appear to be closely aligned (Fig. 1c).

All data from the structure-based sequence comparison are summarised in Table 2. The higher number of both sequence identities and structurally equivalent residues was seen for the first half of the $(\beta/\alpha)_8$-barrel domain and was also found in the analogous *A. oryzae* α-amylase/*S. solfataricus* α-glucosidase comparison (data not shown). In the case of several glycoside hydrolases, the N-terminal $(\beta/\alpha)_4$-half is more conserved, since various GH families can be aligned, at least to some extent, in the region that roughly spans their N-terminal $(\beta/\alpha)_4$-halves. This was first demonstrated for GH families 27, 31, 36 and 66 [21] that may share a common evolutionary origin with family GH13, and all contain the aspartate at β4 as catalytic nucleophile [21,23,24]. It appears (Table 2) that, for GH13 and GH31, the N-terminal $(\beta/\alpha)_4$-half is better conserved than its C-terminal counterpart. Two six-way MULTIPROT comparisons of the two available family 31 structures with four structures from either family 13 or family 77 also indicated better conservation of structure in the N-terminal half than in the C-terminal half barrel (data not shown). Recently a theory of half-TIM-barrel fold evolution [39–42] has been proposed, according to which the $(\beta/\alpha)_8$-barrel fold may have evolved by tandem duplication and fusion from an ancestral half $(\beta/\alpha)_4$-barrel [40]. Such a possibility was experimentally documented for TIM-barrel-fold enzymes involved in histidine and tryptophan biosynthesis pathways [41,42]. Since a satisfactory alignment of the N-terminal $(\beta/\alpha)_4$-half barrel with the C-terminal $(\beta/\alpha)_4$-half barrel of GH-H and GH31 enzymes cannot readily be achieved, because the two halves are too unlike each other, the "half-barrel" duplication theory is not likely to apply for these glycoside hydrolases. Divergent evolution of N-terminal and C-terminal half barrels has on the other hand been suggested for phosphoinositide-specific phospholipases [43] where different enzymes have greater resemblance between N-terminal than C-terminal half-barrels, and this may also apply to GH-H and GH31 enzymes.

3.2. Remote sequence homologies

One of the main aims of this article is to examine the possibility of the present-day families GH13 (or clan GH-H) and GH31 sharing a common ancestor. The divergence, however, is too large for the representatives existing today to be considered as members of the same GH clan [26], although they adopt a similar structural fold (TIM-barrel with domain B) and employ a retaining reaction mechanism catalysed, however, by non-identical catalytic machineries.

To identify residues that reflect remote homology between GH13 (or clan GH-H) and GH31, it is necessary to accept that residues equivalent in primordial versions of the GH13/GH31 barrel may not be in structurally strictly equivalent positions in the present-day GH13 and GH31 $(\beta/\alpha)_8$-barrels. This means that some functionally important and conserved residues in GH13 (or clan GH-H) may still have their counterparts in GH31 and vice versa, but the remote homologies do not necessarily lend themselves to structure-based sequence comparison (Fig. 1b). Moreover, due to very large divergence, original functions of important residues from the primordial GH13/GH31 $(\beta/\alpha)_8$-barrel may no longer be preserved in both clan GH-H and GH31 enzymes.

The proposed remote sequence homologies between representatives of clan GH-H and family GH31 are illustrated in Fig. 2. They cover two stretches within the N-terminal $(\beta/\alpha)_4$-half barrel (around β3 and β4) and two shorter stretches in the C-terminal $(\beta/\alpha)_4$-half barrel (around β7 and β8). Some N-terminal homologies were already demonstrated in a previous bioinformatics analysis [21], especially near β4 with the catalytic nucleophile (Asp206 in GH13 vs. Asp416 in GH31).

The novelty of the approach presented here is that in order to maximise sequence similarity (identity) the correspondences derived from three-dimensional structural positions are not taken strictly into account. Thus, for example, the aspartates near the C-terminus of the β3 (Fig. 2; a residue important for structural integrity in GH13) are not structurally fully equivalent (cf. Fig. 1b). The two remote homologies in the C-terminal parts (Fig. 2) are even more prominent examples of such subtle sequence adjustment. The aspartate near the C-terminus of β7 is: (i) in GH-H an invariant catalytic site residue [3–11]; and (ii) in GH31 an invariant and important residue [23–25]. Structurally, however, the two aspartates do not correspond to each other (Fig. 1b), whereas in the $(\beta/\alpha)_8$-barrel topology an aspartate, structurally related to that at β7 of GH13, is found at β6 in GH31 (Fig. 1c). Similarly, aligned residues around strand β8 are not structural equivalents in GH-H

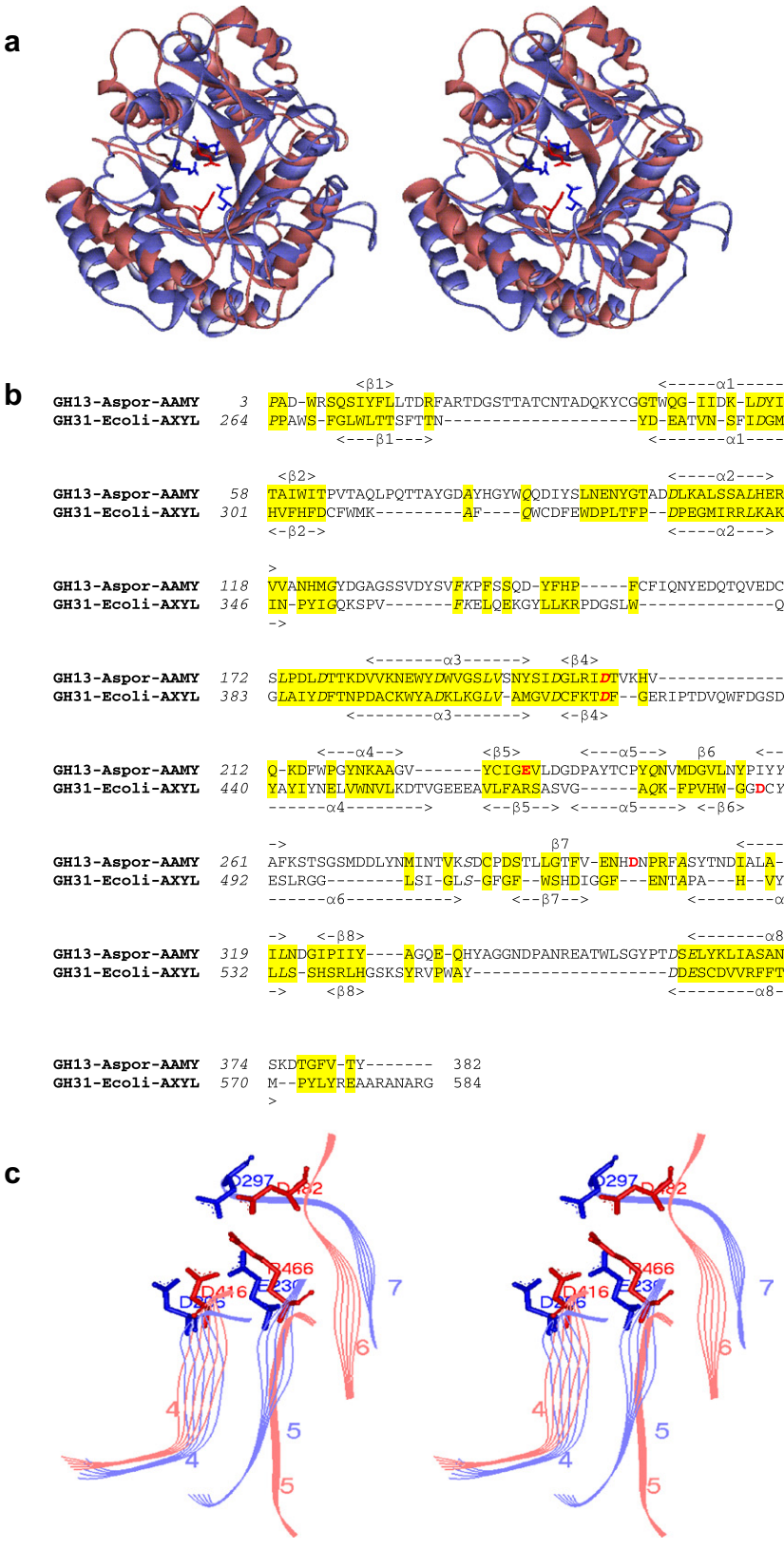

Fig. 1. Structure-based comparison of the GH13 α-amylase from *A. oryzae* and GH31 α-xylosidase from *E. coli*. (a) Superimposed structures of the $(\beta/\alpha)_8$-barrels (α-amylase in blue; α-xylosidase in red) with highlighted catalytic residues. The barrel β-strands are oriented anti-clockwise from the N-terminus to the C-terminus. (b) Structure-based sequence alignment (corresponding residues are highlighted in yellow). Identities are shown in italics and catalytic residues are in red bold font. (c) Overlay of the catalytic residues with corresponding β-strand segments (α-amylase in blue; α-xylosidase in red). In addition to catalytic residues β4-Asp206, β5-Glu230 and β7-Asp297 (α-amylase) and β4-Asp416 and β6-Asp482 (α-xylosidase), GH31 β5-Arg466 corresponding in position to the GH13 acid/base catalyst is shown.

Table 2
Analysis of structure-based alignment of GH13 α-amylase and GH31 α-xylosidase

Comparison (length)[a]	Identities (%)	Correspondences (%)[c]
$(\beta/\alpha)_8$-barrel (437)[b]	23 (5.3)	191 (43.7)
β1 → α4 part (259)	16 (6.2)	119 (46.0)
β5 → α8 part (178)	7 (3.9)	72 (40.5)
Domain C (129)	2 (1.5)	11 (8.5)

[a]Length means the length of the alignment including the gaps.
[b]The $(\beta/\alpha)_8$-barrel here also involves the domain B inserted between strand β3 and helix α3.
[c]Correspondences are those residues in the proteins studied, where the C_α atoms are aligned by MULTIPROT within an average root-mean-square deviation of 1.89 Å.

and GH31, although both segments belong to conserved sequence regions in GH-H [16] and GH31 [38]. The conserved aromatic residues (Fig. 2) appear to support further the homology. Fig. 2 shows not only the alignment of closely related proteins (within clan GH-H or within family GH31), but also the remote homologies between families of proteins (clan GH-H and family GH31) that are different in terms of their catalytic machineries. It can be seen that the aromatic residues are concentrated within very short, well-defined regions (Fig. 2) representing the remote homologies, supporting the idea that their primordial TIM-barrels shared a common ancestor.

3.3. Evolutionary relationships

The structure-based alignment of two GH13 and GH31 proteins (Fig. 1b) enabled us, for the first time, to align a substantial part of the sequences from the catalytic domains of all specificities (see Table 1) from clan GH-H and family GH31 (alignment not shown). The consensus length of the final alignment was 442 positions including 191 correspondences from the structural alignment, resulting in 131 aligned amino acid residues after deleting all sites with a gap in any sequence.

The evolutionary tree shown in Fig. 3 is based on the final alignment that spans roughly the $(\beta/\alpha)_8$-barrel including the domains B. The basic information obtained from the tree is that GH31 has retained its own independence and GH77 of clan GH-H is the closest relative to GH31. It is worth mentioning that the GH77 4-α-glucanotransferases from *Borrelia* [44,45] are the only representatives of the entire α-amylase family that contain lysine instead of the otherwise invariant arginine at two positions in the sequence before the catalytic nucleophile β4 aspartate [18]. In GH31 a conserved tryptophan predominates at the corresponding position [46,47]. A lysine residue is, however, found occasionally (Fig. 2). Both this lysine and the corresponding arginine of GH-H are known to interact directly with substrate [23,37], but equivalent information is not yet available on the GH31 tryptophan. Several pairwise Multiprot comparisons of two GH31 structures with four GH77 structures and four GH13 structures have shown that, in general, the GH31 enzymes align better with GH13 enzymes

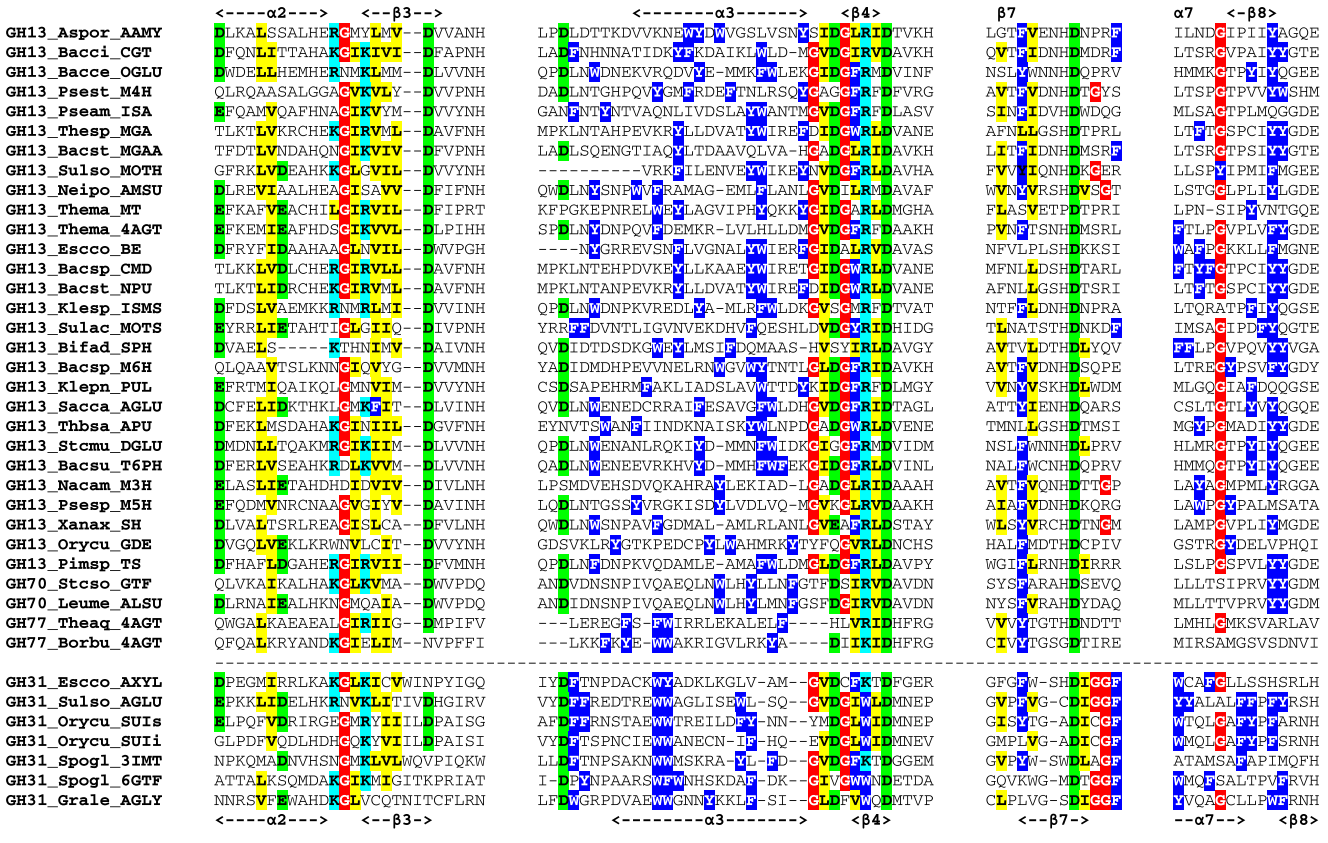

Fig. 2. Regions of remote sequence homology between clan GH-H (families GH13, GH70, GH77) and GH31 enzymes. Some of the residues are shifted relative to each other compared with structure-based alignment (Fig. 1b) to emphasize possible remote homologies (i.e. to achieve more correspondence) between clan GH-H and GH31. Known secondary structure elements for GH13 *A. oryzae* α-amylase and GH31 *E. coli* α-xylosidase are indicated above and below the alignment, respectively. Colour code for selected residues: aspartate, glutamate – green; valine, leucine, isoleucine – yellow; arginine, lysine – turquoise; glycine – red; phenylalanine, tyrosine, tryptophan – blue.

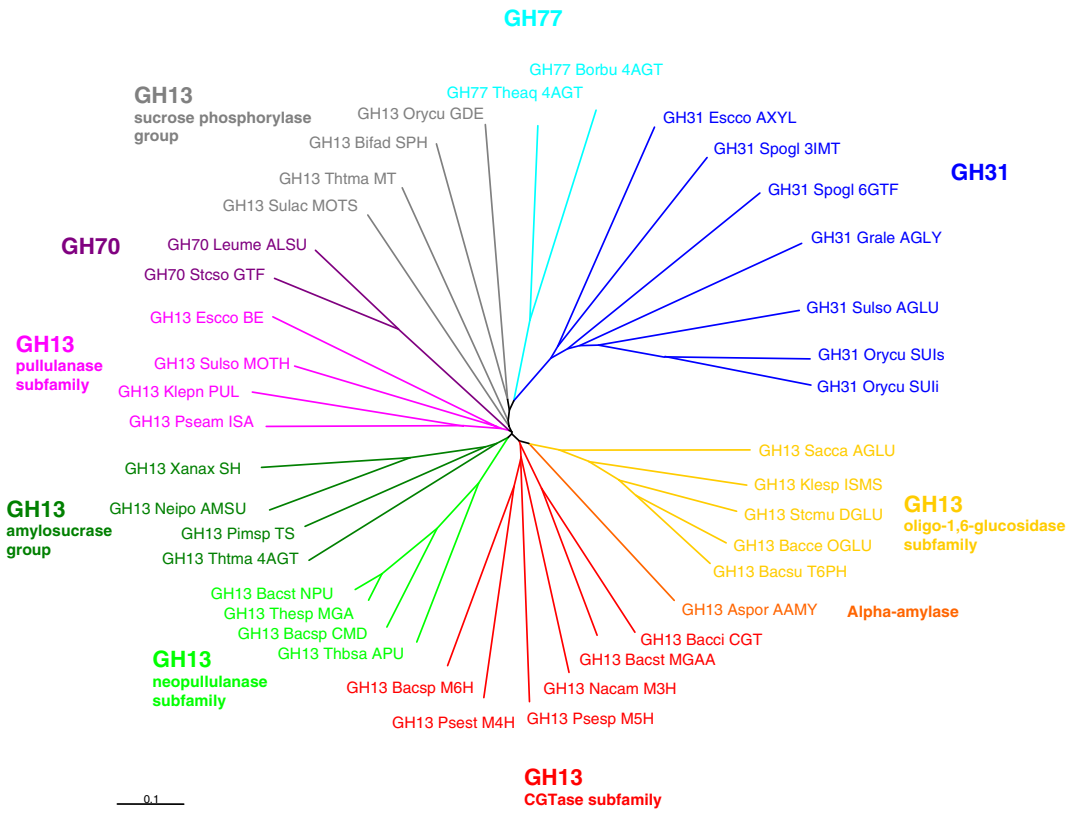

Fig. 3. Evolutionary tree of clan GH-H and family GH31 (blue). Families GH31, 70 and 77 and subfamilies of GH13 are all shown in different colours. The tree was based on an alignment spanning, in both families, the catalytic $(\beta/\alpha)_8$-barrel including domain B (consensus length 442 positions) and calculated with exclusion of the gaps (131 aligned amino acid residues).

than GH77 enzymes (data not shown). It is evident that although the family GH31 shows up as more closely related to GH77 on the evolutionary tree (Fig. 3) i.e. in terms of sequence, this does not seem to be the case for structure. It thus should be emphasised that there is a difference in relatedness of sequence and structure between GH13/77 and GH31 enzymes.

The tree (Fig. 3) reflects the similarities and differences between clan GH-H and family GH31, and in addition contributes several novel findings to the evolutionary relationships known previously within the α-amylase family [10,48–51]: (i) the circularly permuted GH70 family (glucosyltransferase and alternansucrase) is most closely related to the pullulanase subfamily of GH13 represented by pullulanase, isoamylase, maltooligosyl trehalose hydrolase and branching enzyme; (ii) the oligo-1,6-glucosidase subfamily (oligo-1,6-glucosidase, α-glucosidase, dextran glucosidase, trehalose-6-phosphate hydrolase and isomaltulose synthase) is closest to the α-amylase and in a wider sense to the CGTase subfamily (CGTase and maltooligosaccharide-producing amylases); (iii) the neopullulanase subfamily (neopullulanase, cyclomaltodextrinase and maltogenic amylase) that may also contain amylopullulanase borders on a more diverse amylosucrase group including sucrose hydrolase, amylosucrase, trehalose synthase and 4-α-glucanotransferase. The remaining four GH13 specificities, labelled as the sucrose phosphorylase group (maltooligosyl trehalose synthase, maltosyltransferase, sucrose phosphorylase and glucan debranching enzyme) are either on independent or long branches (Fig. 3). It is worth mentioning that Stam et al. [10] have recently described a more detailed and exhaustive division of the GH13 family into subfamilies that shows some agreement with results presented here (Fig. 3). The two studies were, however, based on different numbers of domains of the GH13 enzymes.

4. Conclusions

The $(\beta/\alpha)_8$-barrel of GH31 enzymes has been shown to be more closely related to the barrel of clan GH-H members than to the archetypal TIM-barrel, the relationship being closer for the N-terminal $(\beta/\alpha)_4$-half-barrel than for the C-terminal half. This suggests probable separate evolution of the two half barrels in GH31 and clan GH-H enzymes. The relationship between GH31 and GH-H members extends further to the positioning, but not always the nature, of three critical residues at the active site, and also to remote sequence homologies in both the N- and C-terminal $(\beta/\alpha)_4$-half-barrels. The ideas presented here suggest the possibility of a further level of similarity of glycoside hydrolases beyond that already proposed [26]. This would apply to the relationship between GH31 and enzymes of the GH-H clan, which is less close than within a clan, but includes families with related tertiary structure, partial identity in catalytic machinery and remote sequence homologies.

Acknowledgements: This work was supported in part by the VEGA Grant No. 2/5067/26 from the Slovak Grant Agency for Science (to Š.J.) and the Danish Natural Science Research Council (to B.S.).

References

[1] Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 280, 309–316.
[2] Coutinho, P.M. and Henrissat, B. (1999) Carbohydrate-Active Enzymes server at URL: http://afmb.cnrs-mrs.fr/CAZY/.
[3] Svensson, B., Jensen, M.T., Mori, H., Bak-Jensen, K.S., Bonsager, B., Nielsen, P.K., Kramhoft, B., Praetorius-Ibba, M., Nohr, J., Juge, N., Greffe, L., Williamson, G. and Driguez, H. (2002) Fascinating facets of function and structure of amylolytic enzymes of glycoside hydrolase family 13. Biologia 57 (Suppl. 11), 5–19.
[4] MacGregor, E.A. (2005) An overview of clan GH-H and distantly related families. Biologia 60 (Suppl. 16), 5–12.
[5] Janecek, S. (2005) Amylolytic families of glycoside hydrolases: focus on the family GH-57. Biologia 60 (Suppl. 16), 177–184.
[6] MacGregor, E.A., Janecek, S. and Svensson, B. (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim. Biophys. Acta 1546, 1–20.
[7] van der Maarel, M.J., van der Veen, B., Uitdehaag, J.C., Leemhuis, H. and Dijkhuizen, L. (2002) Properties and applications of starch-converting enzymes of the α-amylase family. J. Biotechnol. 94, 137–155.
[8] Linden, A. and Wilmanns, M. (2004) Adaptation of class-13 α-amylases to diverse living conditions. ChemBioChem 5, 231–239.
[9] Kuriki, T., Takata, H., Yanase, M., Ohdan, K., Fujii, K., Terada, Y., Takaha, T., Hondoh, H., Matsuura, Y. and Imanaka, T. (2006) The concept of the α-amylase family: a rational tool for interconverting glucanohydrolases/glucanotransferases, and their specificities. J. Appl. Glycosci. 53, 155–161.
[10] Stam, M.R., Danchin, E.G., Rancurel, C., Coutinho, P.M. and 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.
[11] Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M. (1984) Structure and possible catalytic residues of Taka-amylase A. J. Biochem. 95, 697–702.
[12] Janecek, S., Svensson, B. and Henrissat, B. (1997) Domain evolution in the α-amylase family. J. Mol. Evol. 45, 322–331.
[13] Przylas, I., Tomoo, K., Terada, Y., Takaha, T., Fujii, K., Saenger, W. and Strater, N. (2000) Crystal structure of amylomaltase from *Thermus aquaticus*, a glycosyltransferase catalysing the production of large cyclic glucans. J. Mol. Biol. 296, 873–886.
[14] MacGregor, E.A., Jespersen, H.M. and Svensson, B. (1996) A circularly permuted α-amylase-type α/β-barrel structure in glucan-synthesizing glucosyltransferases. FEBS Lett. 378, 263–266.
[15] Nakajima, R., Imanaka, T. and Aiba, S. (1986) Comparison of amino acid sequences of eleven different α-amylases. Appl. Microbiol. Biotechnol. 23, 355–360.
[16] Janecek, S. (2002) How many conserved sequence regions are there in the α-amylase family? Biologia 57 (Suppl. 11), 29–41.
[17] van Hijum, S.A., Kralj, S., Ozimek, L.K., Dijkhuizen, L. and van Geel-Schutten, I.G. (2006) Structure–function relationships of glucansucrase and fructansucrase enzymes from lactic acid bacteria. Microbiol. Mol. Biol. Rev. 70, 157–176.
[18] Machovic, M. and Janecek, S. (2003) The invariant residues in the α-amylase family: just the catalytic triad. Biologia 58, 1127–1132.
[19] Matsuura, Y. (2002) A possible mechanism of catalysis involving three essential residues in the enzymes of α-amylase family. Biologia 57 (Suppl. 11), 21–27.
[20] Frandsen, T.P. and 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.
[21] Rigden, D.J. (2002) Iterative database searches demonstrate that glycoside hydrolase families 27, 31, 36 and 66 share a common evolutionary origin with family 13. FEBS Lett. 523, 17–22.
[22] Ernst, H.A., Lo Leggio, L., Yu, S., Finnie, C., Svensson, B. and Larsen, S. (2005) Probing the structure of glucan lyases by sequence analysis, circular dichroism and proteolysis. Biologia 60 (Suppl. 16), 149–159.
[23] Lovering, A.L., Lee, S.S., Kim, Y.W., Withers, S.G. and Strynadka, N.C.J. (2005) Mechanistic and structural analysis of a family 31 α-glycosidase and its glycosyl-enzyme intermediate. J. Biol. Chem. 280, 2105–2115.
[24] Ernst, H.A., Lo Leggio, L., Willemoes, M., Leonard, G., Blum, P. and Larsen, S. (2006) Structure of the *Sulfolobus solfataricus* α-glucosidase: implications for domain conservation and substrate recognition in GH31. J. Mol. Biol. 358, 1106–1124.
[25] Okuyama, M., Okuno, A., Shimizu, N., Mori, H., Kimura, A. and Chiba, S. (2001) Carboxyl group of residue Asp647 as possible proton donor in catalytic reaction of α-glucosidase from *Schizosaccharomyces pombe*. Eur. J. Biochem. 268, 2270–2280.
[26] Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316, 695–696.
[27] Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. and Schneider, M. (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31, 365–370.
[28] Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K., Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D. and Zardecki, C. (2002) The protein data bank. Acta Crystallogr. D 58, 899–907.
[29] Shatsky, M., Nussinov, R. and Wolfson, H.J. (2004) A method for simultaneous alignment of multiple protein structures. Proteins 56, 143–156.
[30] Brzozowski, A.M. and Davies, G.J. (1997) Structure of the *Aspergillus oryzae* α-amylase complexed with the inhibitor acarbose at 2.0 Å resolution. Biochemistry 36, 10837–10845.
[31] Banner, D.W., Bloomer, A., Petsko, G.A., Phillips, D.C. and Wilson, I.A. (1976) Atomic coordinates for triose phosphate isomerase from chicken muscle. Biochem. Biophys. Res. Commun. 72, 146–155.
[32] Bateman, A., Birney, E., Cerruti, L., Durbin, R., Etwiller, L., Eddy, S.R., Griffiths-Jones, S., Howe, K.L., Marshall, M. and Sonnhammer, E.L. (2002) The Pfam protein families database. Nucleic Acids Res. 30, 276–280.
[33] Altschul, S.F., Stephen, F., Madden, T.L., Schaeffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25, 3389–3402.
[34] Thompson, J.D., Higgins, D.G. and 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.
[35] Felsenstein, J. (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
[36] Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput. Appl. Biosci. 12, 357–358.
[37] Uitdehaag, J.C.M., Mosi, R., Kalk, K.H., van der Veen, B.A., Dijkhuizen, L., Withers, S.G. and Dijkstra, B.W. (1999) X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family. Nature Struct. Biol. 6, 432–436.
[38] Kashiwabara, S., Azuma, S., Tsuduki, M. and Suzuki, Y. (2000) The primary structure of the subunit in *Bacillus thermoamyloliquefaciens* KP1071 molecular weight 540,000 homohexameric α-glucosidase II belonging to the glycosyl hydrolase family 31. Biosci. Biotechnol. Biochem. 64, 1379–1393.
[39] Lang, D., Thoma, R., Henn-Sax, M., Sterner, R. and Wilmanns, M. (2000) Structural evidence for evolution of the β/α barrel scaffold by gene duplication and fusion. Science 289, 1546–1550.
[40] Henn-Sax, M., Hocker, B., Wilmanns, M. and Sterner, R. (2001) Divergent evolution of $(\beta\alpha)_8$-barrel enzymes. Biol. Chem. 382, 1315–1320.
[41] Hocker, B., Claren, J. and Sterner, R. (2004) Mimicking enzyme evolution by generating new $(\beta\alpha)_8$-barrels from $(\beta\alpha)_4$-half-barrels. Proc. Natl. Acad. Sci. USA 101, 16448–16453.
[42] Soberon, X., Fuentes-Gallego, P. and Saab-Rincon, G. (2004) In vivo fragment complementation of a $(\beta/\alpha)_8$ barrel protein: generation of variability by recombination. FEBS Lett. 560, 167–172.
[43] Heinz, D.W., Essen, L.O. and Williams, R.L. (1998) Structural and mechanistic comparison of prokaryotic and eukaryotic

phosphoinositide-specific phospholipases C. J. Mol. Biol. 275, 635–650.

[44] Fraser, C.M., Casjens, S., Huang, W.M., Sutton, G.G., Clayton, R.A., Lathigra, R., White, O., Ketchum, K.A., Dodson, R., Hickey, E.K., Gwinn, M., Dougherty, B., Tomb, J.F., Fleischmann, R.D., Richardson, D., Peterson, J., Kerlavage, A.R., Quackenbush, J., Salzberg, S., Hanson, M., van Vugt, R., Palmer, N., Adams, M.D., Gocayne, J.D., Weidman, J., Utterback, T., Watthey, L., McDonald, L., Artiach, P., Bowman, C., Garland, S., Fujii, C., Cotton, M.D., Horst, K., Roberts, K., Hatch, B., Smith, H.O. and Venter, J.C. (1997) Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. Nature 390, 580–586.

[45] Glockner, G., Lehmann, R., Romualdi, A., Pradella, S., Schulte-Spechtel, U., Schilhabel, M., Wilske, B., Suhnel, J. and Platzer, M. (2004) Comparative analysis of the *Borrelia garinii* genome. Nucleic Acids Res. 32, 6038–6046.

[46] Nakai, H., Okuyama, M., Kim, Y.M., Saburi, W., Wongchawalit, J., Mori, H., Chiba, S. and Kimura, A. (2005) Molecular analysis of α-glucosidase belonging to GH-family 31. Biologia 60 (Suppl. 16), 131–135.

[47] Okuyama, M., Kaneko, A., Mori, H., Chiba, S. and Kimura, A. (2006) Structural elements to convert *Escherichia coli* α-xylosidase (YicI) into α-glucosidase. FEBS Lett. 580, 2707–2711.

[48] Jespersen, H.M., MacGregor, E.A., Henrissat, B., Sierks, M.R. and Svensson, B. (1993) Starch- and glycogen-debranching and branching enzymes: prediction of structural features of the catalytic $(\beta/\alpha)_8$-barrel domain and evolutionary relationship to other amylolytic enzymes. J. Protein Chem. 12, 791–805.

[49] Janecek, S. (1994) Parallel βα-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.

[50] Lee, H.S., Kim, M.S., Cho, H.S., Kim, J.I., Kim, T.J., Choi, J.H., Park, C., Lee, H.S., Oh, B.H. and Park, K.H. (2002) Cyclomaltodextrinase, neopullulanase, and maltogenic amylase are nearly indistinguishable from each other. J. Biol. Chem. 277, 21891–21897.

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