Eur. J. Biochem. **270**, 635–645 (2003) © FEBS 2003 doi:10.1046/j.1432-1033.2003.03404.x

Relation between domain evolution, specificity, and taxonomy of the α-amylase family members containing a C-terminal starch-binding domain

Štefan Janeček[1], Birte Svensson[2] and E. Ann MacGregor[3]

[1]*Institute of Molecular Biology, Slovak Academy of Sciences, Bratislava, Slovakia;* [2]*Department of Chemistry, Carlsberg Laboratory, Copenhagen Valby, Denmark;* [3]*Department of Chemistry, University of Manitoba, Winnipeg, Canada*

The α-amylase family (glycoside hydrolase family 13; GH 13) contains enzymes with approximately 30 specificities. Six types of enzyme from the family can possess a C-terminal starch-binding domain (SBD): α-amylase, maltotetraohydrolase, maltopentaohydrolase, maltogenic α-amylase, acarviose transferase, and cyclodextrin glucanotransferase (CGTase). Such enzymes are multidomain proteins and those that contain an SBD consist of four or five domains, the former enzymes being mainly hydrolases and the latter mainly transglycosidases. The individual domains are labelled A [the catalytic $(\beta/\alpha)_8$-barrel], B, C, D and E (SBD), but D is lacking from the four-domain enzymes. Evolutionary trees were constructed for domains A, B, C and E and compared with the 'complete-sequence tree'. The trees for domains A and B and the complete-sequence tree were very similar and contain two main groups of enzymes, an amylase group and a CGTase group. The tree for domain C changed substantially, the separation between the amylase and CGTase groups being shortened, and a new border line being suggested to include the *Klebsiella* and *Nostoc* CGTases (both four-domain proteins) with the four-domain amylases. In the 'SBD tree' the border between hydrolases (mainly α-amylases) and transglycosidases (principally CGTases) was not readily defined, because maltogenic α-amylase, acarviose transferase, and the archaeal CGTase clustered together at a distance from the main CGTase cluster. Moreover the four-domain CGTases were rooted in the amylase group, reflecting sequence relationships for the SBD. It appears that with respect to the SBD, evolution in GH 13 shows a transition in the segment of the proteins C-terminal to the catalytic $(\beta/\alpha)_8$-barrel (domain A).

Keywords: α-amylase family; glycoside hydrolase family 13; starch-binding domain; evolutionary tree; domain evolution.

The α-amylase family (glycoside hydrolase family 13, with close relatives in families 70 and 77) consists at present of enzymes of almost 30 different specificities comprising hydrolases, transglycosidases and isomerases [1]. All of these contain a catalytic $(\beta/\alpha)_8$-barrel domain first recognized in Taka-amylase A, an α-amylase from *Aspergillus oryzae* [2]. This fold was confirmed by crystallography for other specificities, such as cyclodextrin glucanotransferase (CGTase) [3], oligo-1,6-glucosidase [4], maltotetraohydrolase [5], isoamylase [6], neopullulanase [7], maltogenic α-amylase [8], maltogenic amylase [9], amylomaltase [10], glycosyltrehalose trehalohydrolase [11], amylosucrase [12], maltosyltransferase [13], cyclomaltodextrinase [14], 4-α-glucanotransferase [15], and branching enzyme [16]. Structure determinations of family members with yet other specificities are in progress (e.g. [17,18]). Furthermore, prediction of the presence of this $(\beta/\alpha)_8$-barrel fold in other family members has been carried out using unambiguous sequence similarities, particularly at well-known conserved sequence motifs [19–22].

In the sequence-based classification of glycoside hydrolases [23] family 13 (most typically α-amylases) forms the GH-H group together with glycoside hydrolase families (GHs) 70 (glucan sucrase-type glycosyltransferases) and 77 (amylomaltases). These enzymes are multidomain proteins that contain several characteristic domains in addition to domain A, the catalytic $(\beta/\alpha)_8$-barrel [1]. Most of them possess a domain B that protrudes from the barrel between the third β-strand and third α-helix and varies greatly in length, sequence and tertiary structure [20,24]. Domain C, which immediately succeeds the catalytic barrel, is essentially a β-sandwich structure (e.g. [2–5]), characteristic for GH 13 members, but missing in GH 77, as shown by the structure of amylomaltase from *Thermus aquaticus* [10]. Domain C is, moreover, lacking in its common form in

Correspondence to Š. Janeček, Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84551 Bratislava, Slovakia.
Fax: + 421 2 5930 7416, Tel.: + 421 2 5930 7420,
E-mail: stefan.janecek@savba.sk

Abbreviations: CBM, carbohydrate-binding module family; CGTase, cyclodextrin glucanotransferase; GH, glycoside hydrolase family; SBD, starch-binding domain.

Enzymes: α-amylase (EC 3.2.1.1); maltotetraohydrolase (3.2.1.60); maltopentaohydrolase (EC 3.2.1.-); maltogenic α-amylase (EC 3.2.1.133); cyclodextrin glucanotransferase (EC 2.4.1.19); acarviose transferase (EC 2.4.1.-).

(Received 17 September 2002, revised 18 November 2002, accepted 28 November 2002)

GH 70 where the glycosyltransferases have a circularly permuted catalytic $(\beta/\alpha)_8$-barrel [21]. Several GH 13 members contain one or more N-terminal domains preceding the barrel [19]; such domains have occasionally been named domain N although they are not all structurally related. Finally, a group of enzymes in GH 13 contain one or two additional all-β domains, D and E, at the C-terminal end, following the above-mentioned domain C. If the enzymes possess both domains D and E, they do not normally contain an N-domain, and are thus five-domain proteins possessing the catalytic $(\beta/\alpha)_8$-barrel (domain A) and the four domains B, C, D and E. In the case of four-domain proteins without an N-domain, only domain E (but not domain D) is present. It should be noted that the function of domain D is as yet unknown [19,22]. Domain E, however, was recognized early and has attracted much attention due to its raw starch-binding function (e.g [25–32]), which facilitates degradation of starch granules by the enzymes containing such a domain. Throughout this paper, domain E is referred to as SBD, the starch-binding domain.

In a classification of carbohydrate-binding modules, this starch-binding domain is considered to belong to family 20 (CBM 20) [33], and is central to the present study. It is worth mentioning here that amylolytic enzymes containing a completely different kind of starch-binding site [34,35] or a second type of SBD consisting of some sequence repeats of unknown structure [36,37] are outside the scope of this work. The SBD of the present study, CBM 20, is well-known as domain E in CGTases [3,38–41]. It occurs, however, not only in some enzymes of the GH 13 α-amylase family but also in certain β-amylases (GH 14), and in the vast majority of glucoamylases (GH 15), despite the fact that while GH 13 enzymes bring about retention of configuration, both β-amylases and glucoamylases are inverting enzymes and possess catalytic domains that differ from the $(\beta/\alpha)_8$-barrel characteristic of the α-amylase family [2,42,43]. This 'classical' SBD motif consists of seven β-strand segments forming an open-sided distorted β-barrel, as demonstrated by the crystal structures of CGTases from *Bacillus circulans* strains 8 and 251 [3,27], *Bacillus stearothermophilus* [38], *Thermoanaerobacterium thermosulfurogenes* [40], *Bacillus* sp. strain 1011 [41], and β-amylase from *Bacillus cereus* [44], and the NMR solution structure of the isolated recombinant SBD of glucoamylase from *Aspergillus niger* [28].

The SBD is present in approximately 10% of amylolytic enzymes from GHs 13, 14 and 15 [26,30]. In the α-amylase family, this module has been recognized in enzymes having six of the almost 30 specificities: α-amylase, maltotetraohydrolase, maltopentaohydrolase, maltogenic α-amylase, CGTase, and the acarviose transferase (which has, however, been assigned the same EC number as CGTase). While the first three enzymes are four-domain proteins, the latter three have five domains, with the SBD being the C-terminal domain in all cases. Furthermore, the presence of the SBD in an amylolytic enzyme is closely connected with the enzyme origin. Only microorganisms, in particular filamentous fungi, Gram-positive bacteria (Firmicutes), Proteobacteria of the γ-subdivision, actinomycetes and Archaea are known to produce α-amylase family members containing an SBD. Some species, e.g. among aspergilli or streptomycetes, produce GH 13 enzymes with an SBD, and others without this domain. Interestingly, certain mammalian proteins such as laforin [45,46] and genethonin [47], having functions completely unrelated to starch hydrolysis, were found very recently to exhibit unambiguous sequence similarity to an SBD, suggesting a more universal role for this domain.

The present work analyses and compares sequences of the individual domains of all GH 13 members containing an SBD. It is documented by their evolutionary trees that overall the SBD sequences are evolutionarily related according to the taxonomy of the organisms, while the accompanying catalytic and other domains when analysed in the full length sequence, respect the enzyme specificity. Detailed analysis of evolutionary trees calculated for individual domains also reveals that a transition occurs in parts of the proteins which are C-terminal to domain A, discriminating the various GH 13 hydrolases from the transglycosidases having four and five domains, respectively.

Materials and methods

All amino acid sequences of the enzymes studied in this work are listed in Table 1. Most of the sequences were retrieved from the SwissProt database and its supplement TrEMBL [87]. In a few cases, the GenBank [88] was used (Table 1).

BLAST [89] was used for performing the searches in the molecular biology databases (using the default parameters) to retrieve for comparison all the relevant enzymes from the α-amylase family having a C-terminal SBD. As query, the entire sequence of the SBD from *B. circulans* strain 251 CGTase (610 SGDQVSVRFV VNNATTALGQ NVYLTGSVSE LGNWDPAKAI GPMYNQVVYQ YP NWYYDVSV PAGKTIEFKF LKKQGSTVTW EGGS NHTFTA PSSGTATINV NWQP 713) [39] was used.

Published three-dimensional structures of representatives of GH 13 were used as templates that served as definition criteria for individual domains of enzymes listed in Table 1. These were the α-amylase from *A. oryzae* [2], CGTases from *B. circulans* strain 8 and strain 251 [3,27,39], maltotetraohydrolase from *Pseudomonas stutzeri* [5], and maltogenic α-amylase from *B. stearothermophilus* [8]. Some structural information was extracted also from the SwissProt database [87] and from sequence-oriented studies focused on the GH 13 enzymes published previously [19–22,24–26,30].

All sequence alignments were performed using the program CLUSTAL W [90] and then manually tuned where required. The method used for building the evolutionary trees was the neighbour-joining method [91] with the Phylip format tree output implemented in the CLUSTAL W package. The trees were drawn with the program TREEVIEW [92].

The three-dimensional structure of *Bacillus circulans* strain 251 CGTase was retrieved from the Protein Data Bank [93] under the PDB code 1CDG [39]. The protein structure was displayed using the program WEBLABVIEWERLITE (Molecular Simulations, Inc.).

Table 1. The enzymes from the α-amylase family used in the present study.

Enzyme	Source	Abbr.	SwissProt[a]	Reference
α-Amylase	*Aspergillus nidulans*	Aspnd	Q9UV09	Unpublished
	Aspergillus kawachii	Aspka	P13296	[49]
	Bacillus sp. TS-23	Bacsp	Q59222	[50]
	Cryptococcus sp. S2	Crcsp	Q92394	[51]
	Streptomyces albidoflavus	Stral	P09794	[52]
	Streptomyces griseus	Strgr	P30270	[53]
	Streptomyces lividans TK21	Strli21	O86876	Unpublished
	Streptomyces lividans TK24	Strli24	P97179	[55]
	Streptomyces venezuelae	Strve	P22998	[56]
	Thermomonospora curvata	Thscu	P29750	[57]
Maltotetrao-hydrolase	*Pseudomonas saccharophila*	Psesa	P22963	[58]
	Pseudomonas stutzeri	Psest	P13507	[59]
Maltopentao-hydrolase	*Pseudomonas* sp. KO-8940	Psesp	Q52516	[60]
Maltogenic α-amylase	*Bacillus stearothermophilus*	Bacst	P19531	[61]
Acarviose transferase	*Actinoplanes* sp. SE50	Actsp	Q9K5L5	[62]
Cyclodextrin glucanotransferase	*Bacillus* sp. 1–1	Bac11	P31746	[63]
	Bacillus sp. 17–1	Bac17	P30921	[64]
	Bacillus sp. 38–2	Bac38	P09121	[65]
	Bacillus sp. 6.6.3	Bac663	P31747	Unpublished
	Bacillus sp. 1011	Bac1011	P05618	[67]
	Bacillus sp. A2–5a	BacA2	O82984	[68]
	Bacillus sp. B1018	Bac1018	P17692	[69]
	Bacillus sp. E-1	BacE1	Z34466*	[70]
	Bacillus sp. KC201	BacKC	Q59239	[71]
	Bacillus brevis	Bacbr	O30565	[72]
	Bacillus circulans 8	Bacci8	P30920	[73]
	Bacillus circulans 251	Bacci251	P43379	[39]
	Bacillus circulans A11	BacciA	Q9F5W3	Unpublished
	Bacillus clarkii	Baccl	AB082929*	[75]
	Bacillus licheniformis	Bacli	P14014	[76]
	Bacillus macerans IB7	BacmaIB7	O52766	Unpublished
	Bacillus macerans IFO 3490	BacIFO	P04830	[78]
	Bacillus ohbensis	Bacoh	P27036	[79]
	Bacillus stearothermophilus ET1	Bacst1	Q9ZAQ0	[80]
	Bacillus stearothermophilus no. 2	Bacst2	P31797	[81]
	Klebsiella pneumoniae	Klepn	P08704	[82]
	Nostoc sp. PCC 9229	Nossp	AF497477*	[83]
	Thermoanaerobacter sp. ATCC53627	Thbsp	Z35484*	[84]
	Thermoanaerobacter thermosulfurogenes	Thbth	P26827	[85]
	Thermococcus sp. B1001	Thcsp	Q9UWN2	[86]

[a] The accession numbers with * are the numbers from GenBank.

Results and discussion

Domain arrangement and linkers

The initial analysis of 40 amino acid sequences of GH 13 members having the 'classical' SBD (Table 1) revealed that, in fact, there are two groups of these enzymes, which are the five-domain proteins (mostly CGTases, i.e. transglycosidases) and the four-domain proteins (mostly hydrolases). A few exceptions, however, are observed. The maltogenic α-amylase from *B. stearothermophilus* is clearly a hydrolase, yet contains five domains as shown by sequence studies [1,20,94] and its three-dimensional structure [8]. In contrast, the two CGTases from *Klebsiella pneumoniae* [82] and *Nostoc* sp. PCC 9229 [83] lack almost all of a typical domain D, a fact that differentiates them from the CGTases produced by bacilli.

The structural arrangement of domains in a five-domain member of GH 13 is presented in Fig. 1. No three-dimensional structure has been determined for a complete four-domain member, although structures are available for several α-amylases that consist of domains A, B and C only [2,95–99]. It should be noted, however, that crystals have been obtained for the four-domain maltotetrao-hydrolase from *P. stutzeri*, but the SBD was found by X-ray crystallography to be in a disordered state [100]. Figure 1, in addition to illustrating the arrangement of all domains in the five-domain members of the GH 13, can be taken as an approximation of the first three domains in the four-domain members. It also shows the typical

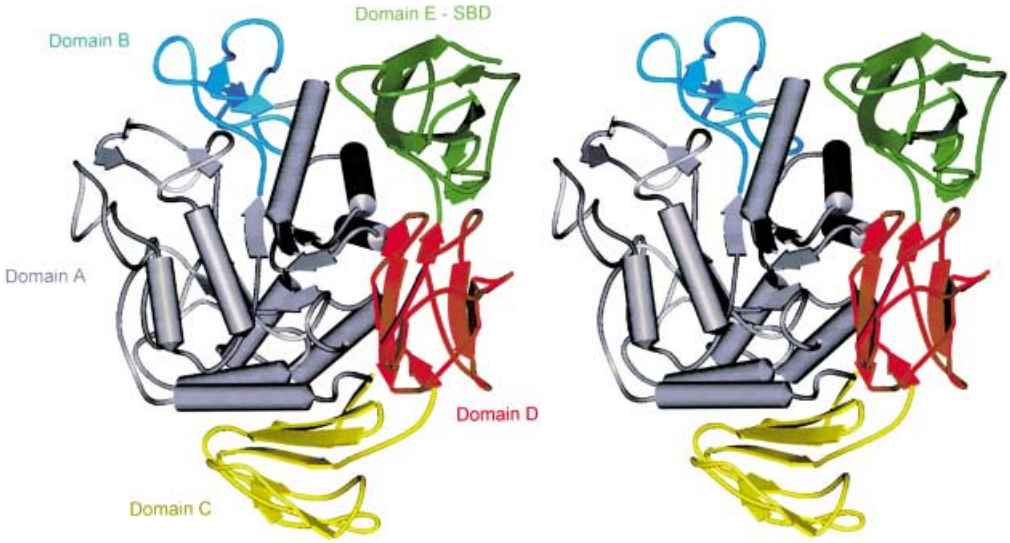

Fig. 1. Stereo view of a CGTase as an example of a five-domain member of the α-amylase family having the C-terminal SBD.

A	Aspka AAM	497	GGSGNTTTTTTAATSTSKATTSSSSSSAAATTSSSCTATSTT
	Aspnd AAM	495	GLTNPTSKTTTATTTSTTTCASATAT
	Crcsp AAM	494	TTYTTASPPPGGCS
	Bacsp AAM	505	GGSVS
	Stral AAM	432	SGGGT
	Strgr AAM	432	SGGGT
	Strli21 AAM	467	DGGDP
	Strli24 AAM	467	DGGDP
	Strve AAM	432	GSTPT
	Thscu AAM	456	GTPPGG
	Psesa M4H	418	SGSGDGGGNDGGEGG
	Psest M4H	418	SGTGSGGGEPGA
	Psesp M5H	483	AAGAA
B	Aspni GAM	468	SIVATGGTTTTATPTGSGSVTSTSKTTATASKTSTSTSSTS
	Horre GAM	471	NVTSS
	Humgr GAM	460	SKQTPNPSAAPSPSPYPTA

Fig. 2. Linkers connecting the SBD to a preceding domain in amylolytic enzymes. (A) Probable linkers connecting domains C and E in the four-domain GH 13 members of this study. AAM, α-amylase; M4H, maltotetraohydrolase; M5H, maltopentaohydrolase. Other abbreviations (Aspka, Aspnd, etc.) are explained in Table 1. (B) For comparison, linkers from GH 15 glucoamylases published in [101] are shown. Aspni GAM, *Aspergillus niger* glucoamylase; Horre GAM, *Hormoconis resinae* glucoamylase; Humgr GAM, *Humicola grisea* glucoamylase.

structure of an SBD as its basic features seem well-conserved [3,8,22,26–30,38–41,44].

While in the five-domain CGTases, the maltogenic α-amylase, and most probably the acarviose transferase, the SBD immediately follows the preceding domain D (Fig. 1), a linker sequence is likely to be necessary in the four-domain proteins to connect domain C to the SBD. Possible linker sequences for the α-amylases and maltotetrao- and maltopentao-hydrolases are shown in Fig. 2A. These sequences vary in length from 5–40 amino acid residues. While the linker in the maltopentaohydrolase is shown as five residues long, uncertainty exists here because the preceding sequence segment, which should correspond to domain C, does not match domain C of any of the other GH 13 sequences reported to date, being unusually high in arginine (37 out of 124 residues). The linkers in all cases are characteristically rich in glycine, serine, threonine and proline (Fig. 2).

For comparison, in glucoamylases (GH 15) the SBD is separated from the catalytic domain by a linker (Fig. 2B) of varying length from a few to more than 50 amino acid residues [101], the longest linker of 68 residues being found in *A. niger* glucoamylase G1 [102]. It should be noted that there is a strong resemblance between the linkers of *Aspergillus* α-amylase and *Aspergillus* glucoamylases, indicating that taxonomy rather than the specificity may play a major role in linker design. These longer linkers should be flexible, while the shorter linkers, particularly those containing proline, may be more rigid.

Evolutionary trees

The differences in the modular organization of the enzymes studied here (Table 1) are clearly reflected in their evolutionary tree (Fig. 3A) calculated using the complete amino acid sequences including the SBD. Unambiguously there is an 'amylase group' and a 'CGTase group' in the tree covering at present the hydrolases (four-domain GH 13 members) and transglycosidases (five-domain members), respectively. The two CGTases, probably lacking domain D

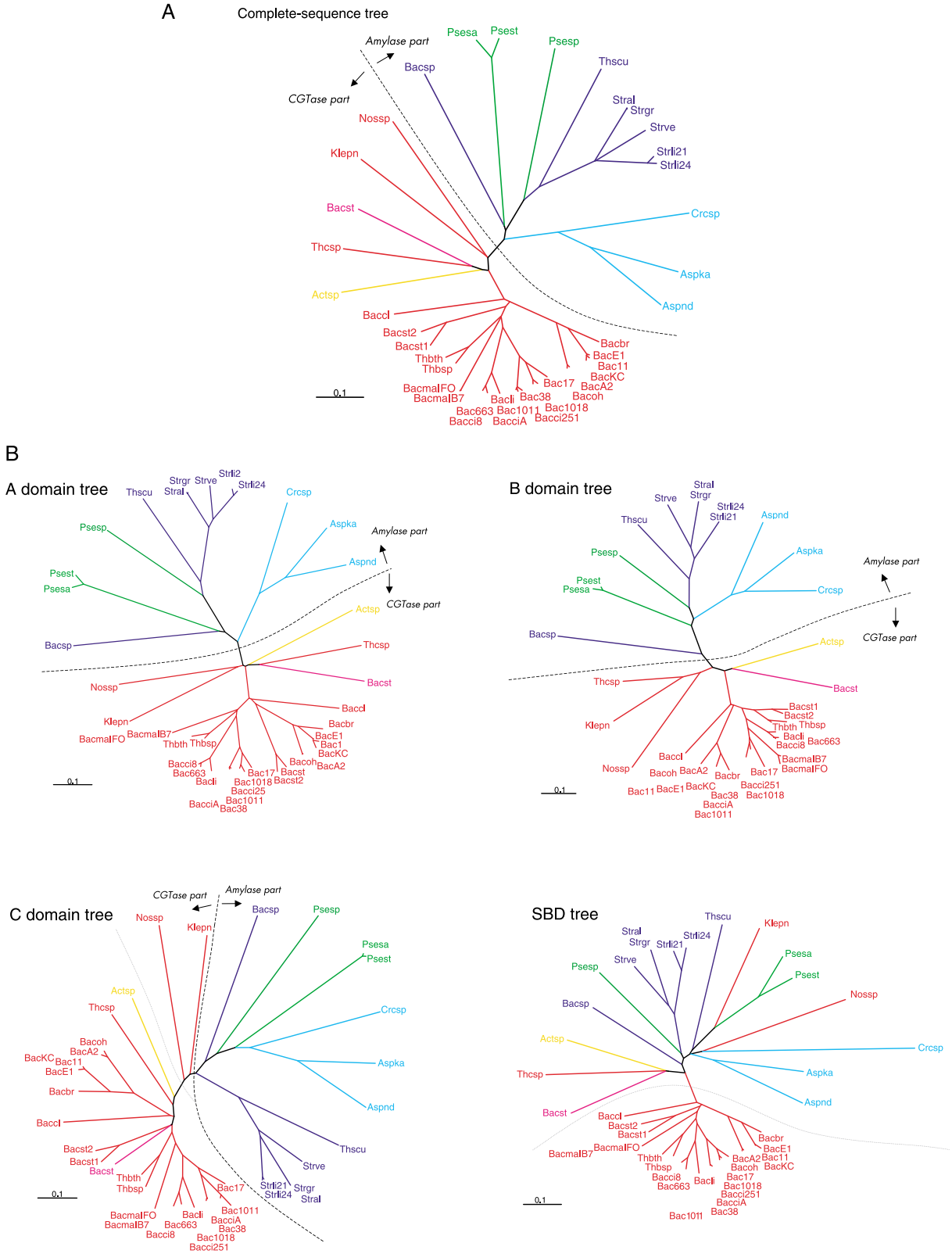

Fig. 3. The evolutionary trees. (A) 'Complete-sequence tree' and (B) trees calculated for individual domains A, B, C and E (SBD). The abbreviations are explained in Table 1. Colour code: red, CGTases; yellow, acarviose transferase; pink, maltogenic α-amylase; blue, α-amylases from *Bacillus* and actinomycetes; light blue, α-amylases from fungi and yeast; green, maltotetraohydrolases and maltopentaohydrolase. A thick dashed line separates the amylase group from the CGTase group, while the thin dotted line indicates the change of the border between the two parts in the C domain tree and the SBD tree.

completely (*K. pneumoniae*; Klepn, and *Nostoc* sp. PCC9229; Nossp), are on branches adjacent to each other and close to the border that separates the two major parts of the tree. Note that the *B. stearothermophilus* maltogenic α-amylase (Bacst) is placed in the 'CGTase group' of the tree. This is, however, not surprising as the enzyme has approximately 60% sequence identity with *Bacillus* CGTases [1,8,20,61,94] and was recently successfully converted by protein engineering into a CGTase [103]. Nevertheless, the unique features discriminating it from the highly similar *Bacillus* CGTases are demonstrated by its appearance in a different cluster (Fig. 3A) together with the only representatives of archaeal CGTases (*Thermococcus* sp. B1001; Thcsp) and acarviose transferases (*Actinoplanes* sp. SE50; Actsp).

Several groups of closely related sequences can be found in both parts of the tree, e.g. the α-amylases produced by streptomycetes or fungi, and the CGTases from the genera *Bacillus* and *Thermoanaerobacter* (Fig. 3A). The α-amylase from *Bacillus* sp. TS23 (Bacsp) is on a long branch, indicating that another bacterial group could emerge in the future as more sequences become available. In the 'amylase group' of the tree the amino acid sequence of the maltopentaohydrolase from *Pseudomonas* sp. KO-8940 (Psesp) is more similar to the sequences of α-amylases originating from the streptomycetes than the two *Pseudomonas* maltotetraohydrolases (Psesa and Psest) (Fig. 3A). It is worth mentioning that the positions of the malto-oligosaccharide-producing amylases in the tree shown in Fig. 3A (the complete-sequence tree) are in agreement with those found in the evolutionary tree built on the alignment of short conserved sequence regions extracted only from domains A and B [20,94]. Both of these trees, i.e. the complete-sequence tree and the tree based on short conserved sequences from domains A and B, respect enzyme specificity.

In order to improve our understanding of evolutionary relationships among the GH 13 four- and five-domain members, partial evolutionary trees were constructed (Fig. 3B) based on the alignments of the individual domains A, B, C and E (i.e. the SBD). A tree was not constructed on the D domain because, as mentioned above, the four-domain amylases and the two CGTases from *Klebsiella* and *Nostoc* lack this domain.

The tree for domain A, i.e. of the catalytic $(\beta/\alpha)_8$-barrel, looks very much like the complete-sequence tree shown in Fig. 3A. In the amylase group of the A domain tree, the α-amylase from *Bacillus* sp. TS-23 (Bacsp) is clustered again together with the two *Pseudomonas* maltotetraohydrolases, although it still preserves its own long branch. In the CGTase group of this tree there are no dramatic changes. This essentially shared arrangement of the two trees obviously reflects the fact that domain A constitutes a substantial part, representing more than 50% of the consensus sequence length, of the final alignment. Moreover, the domain A contains most of the functionally important residues which are conserved in the short sequence motifs [1,19–22,94,104–110].

The tree for domain B is also quite similar to the full-length tree, albeit with a few small changes. In the amylase group of the tree, fungal α-amylases have joined the region of α-amylases from streptomycetes and the *Pseudomonas* malto-oligosaccharide-producing amylases to form a more compact large 'amylase' cluster. The α-amylase from *Bacillus* sp. TS-23 (Bacsp) maintains its own long branch, but approaches the border between the two groups. In the CGTase group, the major change concerns the archaeal CGTase from *Thermococcus* sp. B1001 (Thcsp) that leaves the maltogenic α-amylase (Bacst) and joins the *Klebsiella* CGTase.

In general it should be pointed out that the overall arrangement of the trees constructed for domains A and B (Fig. 3B) are similar to each other and in good agreement with the complete-sequence tree (Fig. 3A). In the group of CGTases produced by the genera *Bacillus* and *Thermoanaerobacter* (the large compact clusters in the CGTases group of the trees), the longest separated branch is occupied by the CGTase from *Bacillus clarkii* (Baccl) [75], indicating that this CGTase is at present the most distantly related CGTase from that group.

The arrangement and clustering of the individual enzymes and enzyme specificities are substantially changed in the C domain tree (Fig. 3B) compared to the two partial trees discussed above and the complete-sequence tree. The C domain tree suggests that a transition occurs in sequence segments C-terminal to domain A such that the amylase/CGTase distinction is altered slightly. Several lines of evidence support this: (a) the distance separating the 'hydrolase' part of the tree from the 'transglycosidase' part has been dramatically shortened in the 'C domain tree'; (b) the two CGTases lacking domain D (Klepn and Nossp) branch off closer to the four-domain GH 13 members, suggesting a new border-line between the two parts of the tree; (c) the *Bacillus stearothermophilus* maltogenic α-amylase (Bacst) is now rooted deeply in the cluster of *Bacillus* and *Thermoanaerobacter* CGTases; (d) this entire large CGTase cluster is joined to the rest by a clearly shorter branch; (e) α-amylases from streptomycetes move closer to the border.

Some of the findings resulting from the C domain tree are not surprising and simply reflect the obvious differences seen in sequences and structures. For example, the 'isolated' position of the maltopentaohydrolase from *Pseudomonas* sp. KO-8940 is based on its C-domain [60] which is unlike other GH 13 domain C sequences. Further, the three-dimensional structure of domain C of maltotetraohydrolase from *P. stutzeri* is reported [5] to resemble that of barley α-amylase [111], a three-domain protein lacking the C-terminal SBD. In all known cases of GH 13 enzymes, domain C is a β-sheet structure [2–9,11–15,38–41,95–99,111], although the length of this domain is variable within the family.

The final partial tree, the 'SBD tree', lacks the character of a tree consisting of two groups, i.e. the amylase group and a CGTase group. It was originally reported [30] for the evolutionary relationships of the SBDs originating from the three families GH 13, GH 14, and GH 15 that their evolutionary tree reflects taxonomy rather than the enzyme specificity. In this study focused on the GH 13 members the two four-domain CGTases (Klepn and Nossp) are rooted obviously in the amylase group of the SBD tree that could involve also the cluster of acarviose transferase (Actsp), archaeal CGTase (Thcsp) and the maltogenic α-amylase (Bacst) due to its longer branch separating it from the

Table 2. Sequence identity (similarity) in percentage for SBD of the two CGTases lacking domain D and selected GH 13 members.

Species	*Klebsiella pneumoniae*	*Nostoc* sp. PCC 9229
Bacillus circulans strain 251 CGT	27.8 (47.2)	24.3 (36.0)
Klebsiella pneumoniae CGT	–	15.2 (33.9)
Nostoc sp. PCC 9229 CGT	15.2 (33.9)	–
Thermococcus sp. B1001 CGT	16.8 (35.4)	15.5 (28.5)
Actinoplanes sp. SE50 ACT	20.9 (40.0)	18.8 (31.3)
Bacillus stearothermophilus MAA	22.3 (42.0)	23.5 (37.4)
Aspergillus kawachii AAM	27.0 (46.0)	21.6 (33.6)
Bacillus sp. TS-23 AAM	30.8 (53.3)	22.9 (43.1)
Streptomyces griseus AAM	25.2 (46.2)	27.9 (43.2)
Pseudomonas stutzeri M4H	42.6 (62.4)	18.0 (36.0)
Pseudomonas sp. KO-8940 M5H	26.4 (50.9)	18.4 (37.7)

compact cluster of *Bacillus* and *Thermoanaerobacter* CGTases (Fig. 3B). Thus for the SBD tree there is not an obvious border between the hydrolases and transglycosidases, but rather there may be one between the compact cluster of *Bacillus* and *Thermoanaerobacter* CGTases and the remaining enzymes. The positions of the two CGTases from *K. pneumoniae* and *Nostoc* sp. PCC-9229 are, however, in agreement with the values of amino acid sequence identity and similarity of their SBD to SBDs from other sources (Table 2). This is evident, for example, for the SBD from *Klebsiella* CGTase that exhibits more than 42% identity to the SBD from *Pseudomonas* maltotetraohydrolase (compare Table 2 and the SBD tree in Fig. 3B). This value is almost 15% higher than that for *B. circulans* strain 251 CGTase representing the CGTases from bacilli. With regard to the *Nostoc* CGTase, it matches best the α-amylase from *Streptomyces griseus*, a representative of the α-amylases produced by streptomycetes. The positioning of a *Nostoc* CGTase in the assumed amylase group of the SBD tree (Fig. 3B) very probably reflects rather the values of sequence similarities (see these values for *Bacillus* sp. TS-23 and *S. griseus* α-amylases vs. that for *B. circulans* strain 251 CGTase in Table 2). Overall the SBD from *Nostoc* sp. PCC 9229 CGTase exhibits a low degree of both sequence identity and similarity to the SBDs from all sources studied here (Table 1), a fact reflected in its long branch in the SBD tree. For comparison, the values of sequence identity for the SBD from *B. circulans* strain 251 CGTase with the SBDs from the CGTases from *Thermococcus* sp. B1001, *Bacillus ohbensis*, and *Thermoanaerobacter thermosulfurogenes* are 37.3%, 63.8% and 74.0%, respectively. Even for the acarviose transferase and the maltogenic α-amylase SBDs compared to the *B. circulans* strain 251 CGTase SBD these values are 39.8% and 45.0%, respectively.

Conclusions

When SBD-containing GH 13 members are analysed, a change in the evolutionary trees from a specificity-determined relationship at the N-terminal part of the enzymes to one influenced more by taxonomy at the C-terminal part of the same enzymes (Figs 3 and 4) can be seen in the present study. The four- and five-domain members of GH 13 can be

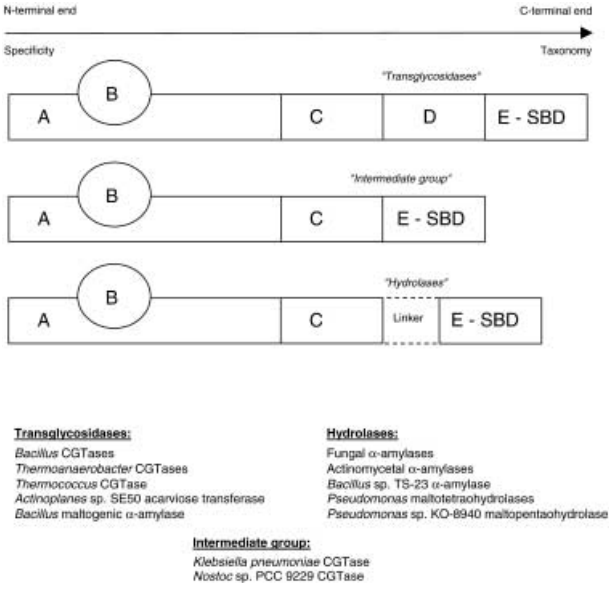

Fig. 4. The proposed relationship between four- and five-domain GH 13 members. It is indicated that there might be a change in domain evolution from specificity to taxonomy when moving from the N-terminal to the C-terminal end of a sequence for this particular group of enzymes.

referred to generally as the SBD-containing hydrolases (mainly α-amylases, but generally classified as EC 3.2.1.x) and transglycosidases (mainly CGTases, but classified as EC 2.4.1.x), and with a noticeable small intermediate group comprising at present the CGTases from *K. pneumoniae* [82] and *Nostoc* sp. PCC 9229 [83] (Fig. 4). The fact that SBD occurs in GH 13, GH 14, and GH 15 [26] supports the idea that there has been a separate evolution of this domain [30]. This together with the findings of the present study indicates a separate evolution of the domains C and E compared to the domains A and B.

The recent introduction by gene fusion of a *Bacillus* CGTase SBD into a *Bacillus subtilis* α-amylase [112] and of the fungal SBD including a linker segment of glucoamylase from *A. niger* to the barley α-amylase 1 [113,114] promoted the α-amylase activity towards starch granules by two- to threefold. The conversion of a CGTase from a transglycosidase into a starch hydrolase was also demonstrated recently [115]. This work, taken together with the results of the present study, as well as with many theoretical and experimental results on sequence and structure similarities between amylases and CGTases [19,20,22,26,30,94,116–118], their phylogenies [20,94,106,119–121], and a novel SBD in an archaeal CGTase [122] can shed more light on, in general, the relations between protein evolution and taxonomy of species [123] and, in particular, the evolution of these industrially important glycoside hydrolases with possible exploitation for their development with enhanced performance.

Acknowledgements

This work was financially supported in part by the VEGA grant no. 2/2057/22 from the Slovak Grant Agency for Science and the EMBO Short-Term Fellowship to ŠJ.

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. Matsuura, Y., Kusunoki, M., Harada, W. & Kakudo, M. (1984) Structure and possible catalytic residues of Taka-amylase A. *J. Biochem.* **95**, 697–702.
3. Klein, C. & Schulz, G.E. (1991) Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. *J. Mol. Biol.* **217**, 737–750.
4. Kizaki, H., Hata, Y., Watanabe, K., Katsube, Y. & Suzuki, Y. (1993) Polypeptide folding of *Bacillus cereus* ATCC7064 oligo-1,6-glucosidase revealed by 3.0 Å resolution X-ray analysis. *J. Biochem.* **113**, 646–649.
5. Morishita, Y., Hasegawa, K., Matsuura, Y., Katsube, Y., Kubota, M. & Sakai, S. (1997) Crystal structure of a maltotetraose-forming exo-amylase from *Pseudomonas stutzeri*. *J. Mol. Biol.* **267**, 661–672.
6. Katsuya, Y., Mezaki, Y., Kubota, M. & Matsuura, Y. (1998) Three-dimensional structure of *Pseudomonas* isoamylase at 2.2 Å resolution. *J. Mol. Biol.* **281**, 885–897.
7. Kamitori, S., Kondo, S., Okuyama, K., Yokota, T., Shimura, Y., Tonozuka, T. & Sakano, Y. (1999) Crystal structure of *Thermoactinomyces vulgaris* R.-47 α-amylase II (TVAII) hydrolyzing cyclodextrins and pullulan at 2.6 Å resolution. *J. Mol. Biol.* **287**, 907–921.
8. Dauter, Z., Dauter, M., Brzozowski, A.M., Christensen, S., Borchert, T.V., Beier, L., Wilson, K.S. & Davies, G.J. (1999) X-ray structure of Novamyl, the five-domain 'maltogenic' α-amylase from *Bacillus stearothermophilus*: maltose and acarbose complexes at 1.7 Å resolution. *Biochemistry* **38**, 8385–8392.
9. Kim, J.S., Cha, S.S., Kim, H.J., Kim, T.J., Ha, N.C., Oh, S.T., Cho, H.S., Cho, M.J., Kim, M.J., Lee, H.S., Kim, J.W., Choi, K.Y., Park, K.H. & Oh, B.H. (1999) Crystal structure of a maltogenic amylase provides insights into a catalytic versatility. *J. Biol. Chem.* **274**, 26279–26286.
10. Przylas, I., Tomoo, K., Terada, Y., Takaha, T., Fujii, K., Saenger, W. & 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.
11. Feese, M.D., Kato, Y., Tamada, T., Kato, M., Komeda, T., Miura, Y., Hirose, M., Hondo, K., Kobayashi, K. & Kuroki, R. (2000) Crystal structure of glycosyltrehalose trehalohydrolase from the hyperthermophilic archaeum *Sulfolobus solfataricus*. *J. Mol. Biol.* **301**, 451–464.
12. Skov, L.K., Mirza, O., Henriksen, A., De Montalk, G.P., Remaud-Simeon, M., Sarcabal, P., Willemot, R.M., Monsan, P. & Gajhede, M. (2001) Amylosucrase, a glucan-synthesizing enzyme from the α-amylase family. *J. Biol. Chem.* **276**, 25273–25278.
13. Roujeinikova, A., Raasch, C., Burke, J., Baker, P.J., Liebl, W. & Rice, D.W. (2001) The crystal structure of *Thermotoga maritima* maltosyltransferase and its implications for the molecular basis of the novel transfer specificity. *J. Mol. Biol.* **312**, 119–131.
14. 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. & Park, K.H. (2002) Cyclomaltodextrinase, neopullulanase, and maltogenic amylase are nearly indistinguishable from each other. *J. Biol. Chem.* **277**, 21891–21897.
15. Roujeinikova, A., Raasch, C., Sedelnikova, S., Liebl, W. & Rice, D.W. (2002) Crystal structure of *Thermotoga maritima* 4-α-glucanotransferase and its acarbose complex: implications for substrate specificity and catalysis. *J. Mol. Biol.* **321**, 149–162.
16. Abad, M.C., Binderup, K., Rios-Steiner, J., Arni, R.K., Preiss, J. & Geiger, J.H. (2002) The X-ray crystallographic structure of *Escherichia coli* branching enzyme. *J. Biol. Chem.* **277**, 42164–42170.
17. Kobayashi, M., Kubota, M. & Matsuura, Y. (1999) Crystallization and improvement of crystal quality for X-ray diffraction of maltooligosyl trehalose synthase by reductive methylation of lysine residues. *Acta Crystallogr.* **D55**, 931–933.
18. Lebbink, J.H.G., Bertoldo, C., Tibbelin, G., Andersen, J.T., Duffner, F., Antranikian, G. & Ladenstein, R. (2000) Crystallization and preliminary X-ray crystallographic studies of the thermoactive pullulanase type I, hydrolyzing α-1,6 glycosidic linkages, from *Fervidobacterium pennivorans* Ven5. *Acta Crystallogr.* **D56**, 1470–1472.
19. Jespersen, H.M., MacGregor, E.A., Sierks, M.R. & Svensson, B. (1991) Comparison of the domain-level organization of starch hydrolases and related enzymes. *Biochem. J.* **80**, 51–55.
20. Jespersen, H.M., MacGregor, E.A., Henrissat, B., Sierks, M.R. & 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.
21. 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.
22. Janeček, Š. (2002) How many conserved sequence regions are there in the α-amylase family? *Biologia, Bratislava* **57** (Suppl. 11), 29–41.
23. 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.
24. Janeček, Š., Svensson, B. & Henrissat, B. (1997) Domain evolution in the α-amylase family. *J. Mol. Evol.* **45**, 322–331.
25. Tanaka, Y., Ashikari, T., Nakamura, N., Kiuchi, N., Shibano, Y., Amachi, T. & Yoshizumi, H. (1986) Comparison of amino acid sequences of three glucoamylases and their structure-function relationships. *Agric. Biol. Chem.* **50**, 965–969.
26. Svensson, B., Jespersen, H., Sierks, M.R. & MacGregor, E.A. (1989) Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. *Biochem. J.* **264**, 309–311.
27. Penninga, D., van der Veen, B.A., Knegtel, R.M.A., van Hijum, S.A.F.T., Rozeboom, H.J., Kalk, K.H., Dijkstra, B.W. & Dijkhuizen, L. (1996) The raw starch binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. *J. Biol. Chem.* **271**, 32777–32784.
28. Sorimachi, K., Le Gal-Coeffet, M.F., Williamson, G., Archer, D.B. & Williamson, M.P. (1997) Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin. *Structure* **5**, 647–661.
29. Southall, S.M., Simpson, P.J., Gilbert, H.J., Williamson, G. & Williamson, M.P. (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. *FEBS Lett.* **447**, 58–60.
30. Janeček, Š. & Ševčík, J. (1999) The evolution of starch-binding domain. *FEBS Lett.* **456**, 119–125.
31. Sauer, J., Sigurskjold, B.W., Christensen, U., Frandsen, T.P., Mirgorodskaya, E., Harrison, M., Roepstorff, P. & Svensson, B. (2000) Glucoamylase: structure/function relationships, and protein engineering. *Biochim. Biophys. Acta* **1543**, 275–293.
32. Giardina, T., Gunning, A.P., Juge, N., Faulds, C.B., Furniss, C.S., Svensson, B., Morris, V.J. & Williamson, G. (2001) Both binding sites of the starch-binding domain of *Aspergillus niger* glucoamylase are essential for inducing a conformational change in amylose. *J. Mol. Biol.* **313**, 1149–1159.
33. Coutinho, P.M. & Henrissat, B. (1999) The modular structure of cellulases and other carbohydrate-active enzymes: an integrated

database approach. In *Genetics, Biochemistry and Ecology of Cellulose Degradation* (Ohmiya, K., Hayashi, K., Sakka, K., Kobayashi, Y., Karita, S. & Kimura, T., eds), pp. 15–23. Uni Publishers Co, Tokyo, Japan.
34. Søgaard, M., Kadziola, A., Haser, R. & Svensson, B. (1993) Site-directed mutagenesis of histidine 93, aspartic acid 180, glutamic acid 205, histidine 290, and aspartic acid 291 at the active site and tryptophan 279 at the raw starch binding site in barley α-amylase 1. *J. Biol. Chem.* **268**, 22480–22484.
35. Tibbot, B.K., Wong, D.W.S. & Robertson, G.H. (2002) Studies on the C-terminal region of barley α-amylase 1 with emphasis on raw starch-binding. *Biologia, Bratislava* **57** (Suppl. 11), 229–238.
36. Rodriguez Sanoja, R., Morlon-Guyot, J., Jore, J., Pintado, J., Juge, N. & Guyot, J.P. (2000) Comparative characterization of complete and truncated forms of *Lactobacillus amylovorus* -amylase and role of the C-terminal direct repeats in raw-starch binding. *Appl. Envir. Microbiol.* **66**, 3350–3356.
37. Sumitani, J.I., Tottori, T., Kawaguchi, T. & Arai, M. (2000) New type of starch-binding domain: the direct repeat motif in the C-terminal region of *Bacillus* sp, 195 α-amylase contributes to starch binding and raw starch degrading. *Biochem. J.* **350**, 477–484.
38. Kubota, M., Matsuura, Y., Sakai, S. & Katsube, Y. (1991) Molecular structure of *B. stearothermophilus* cyclodextrin glucanotransferase and analysis of substrate binding site. *Denpun Kagaku* **38**, 141–146.
39. Lawson, C.L., van Montfort, R., Strokopytov, B., Rozeboom, H.J., Kalk, K.H., de Vries, G.E., Penninga, D., Dijkhuizen, L. & Dijkstra, B.W. (1994) Nucleotide sequence and X-ray structure of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251 in a maltose-dependent crystal form. *J. Mol. Biol.* **236**, 590–600.
40. Knegtel, R.M., Wind, R.D., Rozeboom, H.J., Kalk, K.H., Buitelaar, R.M., Dijkhuizen, L. & Dijkstra, B.W. (1996) Crystal structure at 2.3 Å resolution and revised nucleotide sequence of the thermostable cyclodextrin glycosyltransferase from *Thermonanaerobacterium thermosulfurigenes* EM1. *J. Mol. Biol.* **256**, 611–622.
41. Harata, K., Haga, K., Nakamura, A., Aoyagi, M. & Yamane, K. (1996) X-Ray structure of cyclodextrin glucanotransferase from alkalophilic *Bacillus* sp. 1011. Comparison of two independent molecules at 1.8 Å resolution. *Acta Crystallogr.* **D52**, 1136–1145.
42. 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.
43. 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.
44. Mikami, B., Adachi, M., Kage, T., Sarikaya, E., Nanmori, T., Shinke, R. & Utsumi, S. (1999) Structure of raw starch-digesting *Bacillus cereus* β-amylase complexed with maltose. *Biochemistry* **38**, 7050–7061.
45. Minassian, B.A., Ianzano, L., Meloche, M., Andermann, E., Rouleau, G.A., Delgado-Escueta, A.V. & Scherer, S.W. (2000) Mutation spectrum and predicted function of laforin in Lafora's progressive myoclonus epilepsy. *Neurology* **55**, 341–346.
46. Wang, J., Stuckey, J.A., Wishart, M.J. & Dixon, J.E. (2002) A unique carbohydrate binding domain targets the Lafora disease phosphatase to glycogen. *J. Biol. Chem.* **277**, 2377–2380.
47. Janeček, Š. (2002) A motif of a microbial starch-binding domain found in human genethonin. *Bioinformatics* **18**, 1534–1537.
48. Reference withdrawn.
49. Kaneko, A., Sudo, S., Sakamoto, Y., Tamura, G., Ishikawa, T. & Ohba, T. (1996) Molecular cloning and determination of the nucleotide-sequence of a gene encoding an acid-stable α-amylase from *Aspergillus kawachii*. *J. Ferment. Bioeng.* **81**, 292–298.
50. Lin, L.L., Hsu, W.H. & Chu, W.S. (1997) A gene encoding for an α-amylase from thermophilic *Bacillus* sp. strain TS-23 and its expression in *Escherichia coli*. *J. Appl. Microbiol.* **82**, 325–334.
51. Iefuji, H., Chino, M., Kato, M. & Iimura, Y. (1996) Raw-starch-digesting and thermostable α-amylase from the yeast *Cryptococcus* sp. S-2: purification, characterization, cloning, and sequencing. *Biochem. J.* **318**, 989–996.
52. Long, C.M., Virolle, M.J., Chang, S.Y., Chang, S. & Bibb, M.J. (1987) α-Amylase gene of *Streptomyces limosus*: nucleotide sequence, expression motifs, and amino acid sequence homology to mammalian and invertebrate α-amylases. *J. Bacteriol.* **169**, 5745–5754.
53. Vigal, T., Gil, J.A., Daza, A., Garcia-Gonzalez, M.D. & Martin, J.F. (1991) Cloning, characterization and expression of an α-amylase gene from *Streptomyces griseus* IMRU3570. *Mol. General Genet.* **225**, 278–288.
54. Reference withdrawn.
55. Yin, X.H., Gagnat, J., Gerbaud, C., Guerineau, M. & Virolle, M.J. (1997) Cloning and characterization of a new α-amylase gene from *Streptomyces lividans* TK24. *Gene* **197**, 37–45.
56. Virolle, M.J., Long, C.M., Chang, S. & Bibb, M.J. (1988) Cloning, characterisation and regulation of an α-amylase gene from *Streptomyces venezuelae*. *Gene* **74**, 321–334.
57. Petricek, M., Tichy, P. & Kuncova, M. (1992) Characterization of the α-amylase-encoding gene from *Thermomonospora curvata*. *Gene* **112**, 77–83.
58. Zhou, J., Baba, T., Takano, T., Kobayashi, S. & Arai, Y. (1989) Nucleotide sequence of the maltotetraohydrolase gene from *Pseudomonas saccharophila*. *FEBS Lett.* **255**, 37–41.
59. Fujita, M., Torigoe, K., Nakada, T., Tsusaki, K., Kubota, M., Sakai, S. & Tsujisaka, Y. (1989) Cloning and nucleotide sequence of the gene (amyP) for maltotetraose-forming amylase from *Pseudomonas stutzeri* MO-19. *J. Bacteriol.* **171**, 1333–1339.
60. Shida, O., Takano, T., Takagi, H., Kadowaki, K. & Kobayashi, S. (1992) Cloning and nucleotide sequence of the maltopentaose-forming amylase gene from *Pseudomonas* sp. KO-8940. *Biosci. Biotechn Biochem.* **56**, 76–80.
61. Diderichsen, B. & Christiansen, L. (1988) Cloning of a maltogenic α-amylase from *Bacillus stearothermophilus*. *FEMS Microbiol. Lett.* **56**, 53–60.
62. Hemker, M., Stratmann, A., Goeke, K., Schroder, W., Lenz, J., Piepersberg, W. & Pape, H. (2001) Identification, cloning, expression, and characterization of the extracellular acarbose-modifying glycosyltransferase, AcbD, from *Actinoplanes* sp. strain SE50. *J. Bacteriol.* **183**, 4484–4492.
63. Schmid, G., Englbrecht, A. & Schmid, D. (1988) Cloning and nucleotide sequence of a cyclodextrin glycosyltransferase gene from the alkalophilic *Bacillus* 1–1. In *Proceedings of the Fourth International Symposium on Cyclodextrins* (Huber, O. & Szejtli, J., eds), pp. 71–76. Kluwer Academic Publishers, Dordrecht, Germany/Boston, USA.
64. Kaneko, T., Song, K.B., Hamamoto, T., Kudo, T. & Horikoshi, K. (1989) Construction of a chimeric series of *Bacillus* cyclomaltodextrin glucanotransferases and analysis of the thermal stabilities and pH optima of the enzymes. *J. General Microbiol.* **135**, 3447–3457.
65. Kaneko, T., Hamamoto, T. & Horikoshi, K. (1988) Molecular cloning and nucleotide sequence of the cyclomaltodextrin glucanotransferase gene from the alkalophilic *Bacillus* sp. strain, 38–2. *J. General Microbiol.* **134**, 97–105.
66. Reference withdrawn.
67. Kimura, K., Kataoka, S., Ishii, Y., Takano, T. & Yamane, K. (1987) Nucleotide sequence of the β-cyclodextrin glucanotransferase gene of alkalophilic *Bacillus* sp. strain 1011 and similarity

of its amino acid sequence to those of α-amylases. *J. Bacteriol.* **169**, 4399–4402.
68. Ohdan, K., Kuriki, T., Takata, H. & Okada, S. (2000) Cloning of the cyclodextrin glucanotransferase gene from alkalophilic *Bacillus* sp. A2–5a and analysis of the raw starch-binding domain. *Appl. Microbiol. Biotechnol.* **53**, 430–434.
69. Itkor, P., Tsukagoshi, N. & Udaka, S. (1990) Nucleotide sequence of the raw-starch-digesting amylase gene from *Bacillus* sp. B1018 and its strong homology to the cyclodextrin glucanotransferase genes. *Biochem. Biophys. Res. Commun.* **166**, 630–636.
70. Yong, J., Choi, J., Kang, H., Park, C., Park, K. & Choi, Y. (1996) Molecular cloning of CGTase gene from alkalophilic *Bacillus* sp. E-1 and its overexpression in *E. coli. Biotechnol. Lett.* **18**, 1223–1228.
71. Kitamoto, N., Kimura, T., Kito, Y. & Ohmiya, K. (1992) Cloning and sequencing of the gene encoding cyclodextrin glucanotransferase from *Bacillus* sp. K.C.201. *J. Ferment. Bioeng.* **74**, 345–351.
72. Kim, M.H., Sohn, C.B. & Oh, T.K. (1998) Cloning and sequencing of a cyclodextrin glycosyltransferase gene from *Brevibacillus brevis* CD162 and its expression in *Escherichia coli*. *FEMS Microbiol. Lett.* **164**, 411–418.
73. Nitschke, L., Heeger, K., Bender, H. & Schulz, G.E. (1990) Molecular cloning, nucleotide sequence and expression in *Escherichia coli* of the β-cyclodextrin glycosyltransferase gene from *Bacillus circulans* strain 8. *Appl. Microbiol. Biotechnol.* **33**, 542–546.
74. Reference withdrawn.
75. Takada, M., Nakagawa, Y. & Yamamoto, M. (2003) Biochemical and genetic analyses of a novel γ-cyclodextrin glucanotransferase from an alkalophilic *Bacillus clarkii* 7364. *J. Biochem.* **133**, in press.
76. Hill, D.E., Aldape, R. & Rozzell, J.D. (1990) Nucleotide sequence of a cyclodextrin glucosyltransferase gene, cgtA, from *Bacillus licheniformis*. *Nucleic Acids Res.* **18**, 199.
77. Reference withdrawn.
78. Takano, T., Fukuda, M., Monma, M., Kobayashi, S., Kainuma, K. & Yamane, K. (1986) Molecular cloning, DNA nucleotide sequencing, and expression in *Bacillus subtilis* cells of the *Bacillus macerans* cyclodextrin glucanotransferase gene. *J. Bacteriol.* **166**, 1118–1122.
79. Sin, K.A., Nakamura, A., Kobayashi, K., Masaki, H. & Uozumi, T. (1991) Cloning and sequencing of a cyclodextrin glucanotransferase gene from *Bacillus ohbensis* and its expression in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **35**, 600–605.
80. Chung, H.J., Yoon, S.H., Lee, M.J., Kim, M.J., Kweon, K.S., Lee, I.W., Kim, J.W., Oh, B.H., Lee, H.S., Spiridonova, V.A. & Park, K.H. (1998) Characterization of a thermostable cyclodextrin glucanotransferase isolated from *Bacillus stearothermophilus* ET1. *J. Agric. Food Chem.* **46**, 952–959.
81. Fujiwara, S., Kanemoto, M., Kim, B., Lejeune, A., Sakaguchi, K. & Imanaka, T. (1992) Cyclization characteristics of cyclodextrin glucanotransferase are conferred by the NH_2-terminal region of the enzyme. *Appl. Environ. Microbiol.* **58**, 4016–4025.
82. Binder, F., Huber, O. & Boeck, A. (1986) Cyclodextrin-glycosyltransferase from *Klebsiella pneumoniae* M5a1: cloning, nucleotide sequence and expression. *Gene* **47**, 269–277.
83. Wouters, J., Bergman, B. & Janson, S. (2003) Cloning and expression of a putative cyclodextrin glucosyltransferase from the symbiotically competent cyanobacterium *Nostoc* sp. PCC 9229. *FEMS Microbiol. Lett.* in press.
84. Joergensen, S.T., Tangney, M., Starnes, R.L., Amemiya, K. & Joergensen, P.L. (1997) Cloning and nucleotide sequence of a thermostable cyclodextrin glycosyltransferase gene from *Thermoanaerobacter* sp. ATCC 53627 and its expression in *Escherichia coli*. *Biotechnol. Lett.* **19**, 1027–1031.
85. Bahl, H., Burchhardt, G., Spreinat, A., Haeckel, K., Wienecke, A., Schmidt, B. & Antranikian, G. (1991) α-Amylase of *Clostridium thermosulfurogenes* EM1: nucleotide sequence of the gene, processing of the enzyme, and comparison of other α-amylases. *Appl. Environ. Microbiol.* **57**, 1554–1559.
86. Yamamoto, T., Shiraki, K., Fujiwara, S., Takagi, M., Fukui, K. & Imanaka, T. (1999) In vitro heat effect on functional and conformational changes of cyclodextrin glucanotransferase from hyperthermophilic archaea. *Biochem. Biophys. Res. Commun.* **265**, 57–6177.
87. Bairoch, A. & Apweiler, R. (2000) The SWISS-PROT protein sequence database and its Supplement TrEMBL in 2000. *Nucleic Acids Res.* **28**, 45–48.
88. Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J., Rapp, B.A. & Wheeler, D.L. (2002) GenBank. *Nucleic Acids Res.* **30**, 17–20.
89. 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.
90. Thompson, J.D., Higgins, D.G. & Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment trough sequence weighting, position specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680.
91. Saitou, N. & Nei, M. (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
92. Page, R.D. (1996) TreeView: an application to display phylogenetic trees on personal computers. *Comput. Applic. Biosci.* **12**, 357–358.
93. 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. & Zardecki, C. (2002) The protein data bank. *Acta Crystallogr.* **D58**, 899–907.
94. Janeček, Š. (1995) Tracing the evolutionary lineages among α-amylases and cyclodextrin glycosyltransferases: the question of so-called 'intermediary' enzymes. *Biologia, Bratislava* **50**, 515–522.
95. Brady, R.L., Brzozowski, A.M., Derewenda, Z.S., Dodson, E.J. & Dodson, G.G. (1991) Solution of the structure of *Aspergillus niger* acid α-amylase by combined molecular replacement and multiple isomorphous replacement methods. *Acta Crystallogr.* **B47**, 527–535.
96. Machius, M., Wiegand, G. & Huber, R. (1995) Crystal structure of calcium-depleted *Bacillus licheniformis* α-amylase at 2.2 Å resolution. *J. Mol. Biol.* **246**, 545–559.
97. Aghajari, N., Feller, G., Gerday, C. & Haser, R. (1998) Crystal structures of the psychrophilic α-amylase from *Alteromonas haloplanctis* in its native form and complexed with an inhibitor. *Protein Sci.* **7**, 564–572.
98. Fujimoto, Z., Takase, K., Doui, N., Momma, M., Matsumoto, T. & Mizuno, H. (1998) Crystal structure of a catalytic-site mutant α-amylase from *Bacillus subtilis* complexed with maltopentaose. *J. Mol. Biol.* **277**, 393–407.
99. Suvd, D., Fujimoto, Z., Takase, K., Matsumura, M. & Mizuno, H. (2001) Crystal structure of *Bacillus stearothermophilus* α-amylase: possible factors determining the thermostability. *J. Biochem.* **129**, 461–468.
100. Mezaki, Y., Katsuya, Y., Kubota, M. & Matsuura, Y. (2001) Crystallization and structural analysis of intact maltotetraose-forming exo-amylase from *Pseudomonas stutzeri*. *Biosci. Biotechnol. Biochem.* **65**, 222–225.
101. Sauer, J., Christensen, T., Frandsen, T.P., Mirgorodskaya, E., McGuire, K.A., Driguez, H., Roepstorff, P., Sigurskjold, B.W. &

Svensson, B. (2001) Stability and function of interdomain linker variants of glucoamylase 1 from *Aspergillus niger*. *Biochemistry* **40**, 9336–9346.
102. Svensson, B., Larsen, K., Svendsen, I. & Boel, E. (1983) The complete amino acid sequence of the glycoprotein, glucoamylase G1, from *Aspergillus niger*. *Carlsberg Res. Commun.* **48**, 529–544.
103. Beier, L., Svendsen, A., Andersen, C., Frandsen, T.P., Borchert, T.V. & Cherry, J.R. (2000) Conversion of the maltogenic α-amylase Novamyl into a CGTase. *Protein Eng.* **13**, 509–513.
104. MacGregor, E.A. (1993) Relationships between structure and activity in the α-amylase family of starch-metabolising enzymes. *Starch* **7**, 232–237.
105. Svensson, B. (1994) Protein engineering in the α-amylase family: catalytic mechanism, substrate specificity, and stability. *Plant Mol. Biol.* **25**, 141–157.
106. Janeček, Š. (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.
107. Nielsen, J.E. & Borchert, T.V. (2000) Protein engineering of bacterial α-amylases. *Biochim. Biophys. Acta* **1543**, 253–274.
108. van der Veen, B.A., Uitdehaag, J.C., Dijkstra, B.W. & Dijkhuizen, L. (2000) Engineering of cyclodextrin glycosyltransferase reaction and product specificity. *Biochim. Biophys. Acta* **1543**, 336–360.
109. van der Maarel, M.J., van der Veen, B., Uitdehaag, J.C., Leemhuis, H. & Dijkhuizen, L. (2002) Properties and applications of starch-converting enzymes of the α-amylase family. *J. Biotechnol.* **94**, 137–155.
110. Uitdehaag, J.C.M., van der Veen, B.A., Dijkhuizen, L. & Dijkstra, B.W. (2002) Catalytic mechanism and product specificity of cyclodextrin glycosyltransferase, a prototypical transglycosylase from the α-amylase family. *Enzyme Microb. Technol.* **30**, 295–304.
111. Kadziola, A., Abe, J., Svensson, B. & Haser, R. (1994) Crystal and molecular structure of barley α-amylase. *J. Mol. Biol.* **239**, 104–121.
112. Ohdan, K., Kuriki, T., Takata, H., Kaneko, H. & Okada, S. (2000) Introduction of raw starch-binding domains into *Bacillus subtilis* α-amylase by fusion with the starch-binding domain of *Bacillus* cyclomaltodextrin glucanotransferase. *Appl. Environ. Microbiol.* **66**, 3058–3064.
113. Svensson, B., Tovborg Jensen, M., Mori, H., Bak-Jensen, K.S., Bønsager, B., Nielsen, P.K., Kramhøft, B., Prætorius-Ibba, M., Nøhr, J., Juge, N., Greffe, L., Williamson, G. & Driguez, H. (2002) Fascinating facets of function and structure of amylolytic enzymes of glycoside hydrolase family 13. *Biologia, Bratislava* **57** (Suppl. 11), 5–19.
114. Juge, N., Le Gal-Coëffet, M.F., Furniss, C.S.M., Gunning, A.P., Kramhøft, B., Morris, V.J., Williamson, G. & Svensson, B. (2002) The starch binding domain of glucoamylase from *Aspergillus niger*: overview of its structure, function, and role in raw-starch hydrolysis. *Biologia, Bratislava* **57** (Suppl. 11), 239–245.
115. Leemhuis, H., Dijkstra, B.W. & Dijkhuizen, L. (2002) Mutations converting cyclodextrin glycosyltransferase from a transglycosylase into a starch hydrolase. *FEBS Lett.* **514**, 189–192.
116. MacGregor, E.A. & Svensson, B. (1989) A super-secondary structure predicted to be common to several α-1,4-D-glucan-cleaving enzymes. *Biochem. J.* **259**, 145–152.
117. Janeček, Š., MacGregor, E.A. & Svensson, B. (1995) Characteristic differences in the primary structure allow discrimination of cyclodextrin glucanotransferases from α-amylases. *Biochem. J.* **305**, 685–686.
118. Lo, H.F., Lin, L.L., Chiang, W.Y., Chie, M.C., Hsu, W.H. & Chang, C.T. (2002) Deletion analysis of the C-terminal region of the α-amylase of *Bacillus* sp. strain TS-23. *Arch. Microbiol.* **178**, 115–123.
119. del-Rio, G., Morett, E. & Soberon, X. (1997) Did cyclodextrin glycosyltransferases evolve from α-amylases? *FEBS Lett.* **416**, 221–224.
120. Bikbulatova, S.M., Chemeris, A.V., Usanov, N.G. & Vakhitov, V.A. (2000) Establishment of the phylogenetic relationship between the microbial producers of cyclodextrin glucanotransferases using their complete amino acid sequences. *Mikrobiologiya* **69**, 686–693.
121. Pujadas, G. & Palau, J. (2001) Evolution of α-amylases: architectural features and key residues in the stabilization of the $(\beta/\alpha)_8$ scaffold. *Mol. Biol. Evol.* **18**, 38–54.
122. Rashid, N., Cornista, J., Ezaki, S., Fukui, T., Atomi, H. & Imanaka, T. (2002) Characterization of an archaeal cyclodextrin glucanotransferase with a novel C-terminal domain. *J. Bacteriol.* **184**, 777–784.
123. Pace, N.R. (1997) A molecular view of microbial diversity and the biosphere. *Science* **276**, 734–740.