

The evolution of starch-binding domain

Štefan Janeček^{a,*}, Jozef Ševčík^b

^a Institute of Microbiology, Slovak Academy of Sciences, Štefánikova 3, SK-81434 Bratislava, Slovakia

^b Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, SK-84251 Bratislava, Slovakia

Received 15 March 1999; received in revised form 28 June 1999

Abstract Amylolytic enzymes belonging to three distinct families of glycosidases (13, 14, 15) contain the starch-binding domain (SBD) positioned almost exclusively at the C-terminus. Detailed analysis of all available SBD sequences from 43 different amylases revealed its independent evolutionary behaviour with regard to the catalytic domains. In the evolutionary tree based on sequence alignment of the SBDs, taxonomy is respected so that fungi and actinomycetes form their own separate parts surrounded by bacteria that are also clustered according to taxonomy. The only known N-terminal SBD from *Rhizopus oryzae* glucoamylase is on the longest branch separated from all C-terminal SBDs. The 3-dimensional (3-D) structures of fungal glucoamylase and bacterial CGTase SBDs are compared and used to discuss the interesting SBD evolution.

© 1999 Federation of European Biochemical Societies.

Key words: Amylolytic enzyme; Starch-binding domain; Evolution; Sequence alignment; Structure overlap

1. Introduction

Most of amylolytic enzymes belong to three different families of glycosidases [1]: family 13, α -amylases as well as other enzymes showing amylolytic action (representing about 20 different specificities); family 14, β -amylases; and family 15, glucoamylases. Although these three amylases have related functions (all of them cleave the α -glucosidic bonds of starch), structurally and thus also evolutionarily they constitute their own, independent families with no sequence similarities [2–5]. With regard to the catalytic domains, both α -amylases and β -amylases are (β/α)₈-barrel proteins [6] (with, however, quite different TIM-barrels [7,8]), while the glucoamylases adopt (α/α)₆-barrel fold [9,10]. As far as the mechanism of glycoside bond cleavage is concerned, α -amylases use a retaining mechanism while β -amylases and glucoamylases act by an inverting one [11].

There exists, however, the so-called raw-starch-binding domain (SBD), that is common for all the amylase families although not each member of a family contains the motif. The SBD is always positioned at the C-terminal part of an amylase except for the glucoamylase from *Rhizopus oryzae* which contains the SBD at its N-terminus [12]. While the sequence conservation as well as the basic arrangement of secondary structure elements of the SBD in different amylases has been well recognised [13,14], the opportunity to compare the SBDs from different families and taxonomic groups became possible only recently. Some sequence and structural analyses of SBD focused mainly on glucoamylases were

done by Coutinho and Reilly [4,15,16], but the evolutionary analysis of SBDs from all the three different families of glycosidases is still lacking.

The SBD motif, consisting of several β -strand segments forming an open-sided, distorted β -barrel structure [17,18], is responsible for the ability of an amylase to bind and digest the native raw, granular starch [19,20]. In α -amylases the SBD may govern the enzyme thermostability [21], however, this motif seems to have nothing to do with thermostability in glucoamylases [22]. It has been demonstrated [23] that the SBD independently retains its function even if fused to a protein other than amylase. Very recently, Southall et al. [24] have shown that the SBD from *Aspergillus niger* glucoamylase not only binds to the raw starch, but it also disrupts the structure of starch surface, thereby enhancing the amylolytic rate.

Taking into account all the attributes of these three amylolytic enzyme families, the question of the evolutionary relationships of the individual members of these families that contain the SBD, derived from the sequence-structural analysis of SBDs, should be of great interest. The aim of this work is to contribute to the fundamental questions regarding the evolution of the three different amylase families 13, 14 and 15 by shading more light on the evolution of their raw-starch-binding motif which is present in about 10% of these enzymes.

2. Materials and methods

The SBD motifs were cut from the amino acid sequences of 43 amylases (Table 1) forming three sequence-based families 13, 14 and 15 of glycoside hydrolases. The following enzyme specificities are represented: EC 3.2.1.1, α -amylase; EC 3.2.1.60, maltotetraohydrolase; EC 3.2.1.133, maltogenic amylase; EC 3.2.1.-, maltopentaohydrolase; EC 2.4.1.19, cyclodextrin glycosyltransferase (all from the family 13); EC 3.2.1.2, β -amylase (family 14); and EC 3.2.1.3, glucoamylase (family 15). The sequences were retrieved from the SwissProt [25] and GenBank [26] sequence databases.

Each SBD in the sequences of amylolytic enzymes studied was identified using the sequence of SBD (Cys⁵⁰⁹-Arg⁶¹⁶) from *A. niger* glucoamylase [18] as template. Two sets of SBD sequences were created: (i) the first one comprising all the SBD sequences listed in Table 1 except for the SBD from *R. oryzae* glucoamylase [12]; and (ii) the other set involving the *Rhizopus* N-terminal SBD sequence with several representative SBD sequences from the former set. Both sets of SBD sequences were aligned using the program CLUSTAL W [27] and the computer-produced alignments were slightly manually tuned where applicable taking into account the known elements of secondary structure of SBDs from *Bacillus circulans* strain 251 CGTase [17] as well as *A. niger* glucoamylase [18]. The final alignments served for calculation by the neighbour-joining method [28] of the evolutionary trees. The Phylip format tree output was applied using the bootstrapping procedure [29]; the number of bootstrap trials used was 1000. The trees were drawn with the program TreeView [30].

Two 3-dimensional structures of SBD were retrieved from the Protein Data Bank (PDB) [31] for comparison: the bacterial family 13 CGTase SBD from *B. circulans* strain 251 [17] determined by X-ray crystallography (PDB code: 1CGD) and the fungal family

*Corresponding author. Fax: (421) (7) 393824.
E-mail: sjanecek@ue.savba.sk

15 glucoamylase SBD from *A. niger* [18] determined by NMR (PDB code: 1AC0). The CCP4 software [32] and the program MOLSCRIPT [33] were used for comparison and plotting of the structures, respectively.

3. Results and discussion

3.1. Amino acid sequence alignment

The sequence alignment of all C-terminal SBDs (the N-terminal SBD from *R. oryzae* glucoamylase will be discussed later) is shown in Fig. 1. It is worth mentioning that the consensus residues identified by Svensson et al. [14] are almost fully conserved in the present set of 42 sequences. There are only eight deviations, i.e. 1.7%. These are found (see Fig. 1) at Gly⁵³⁵ (all amino acid numberings throughout the text use the numbers of the glucoamylase from *A. niger*), Pro⁵⁶¹ (two cases), Lys⁵⁷⁸, Trp⁵⁹⁰, and Asn⁵⁹⁵ (three cases).

It is evident that there are two basic groups of SBDs: a bacterial type (including *Bacillus*, *Klebsiella*, *Clostridium* and perhaps also *Pseudomonas*) and a fungal type (including fungi and yeasts). Remarkably the actinomycetes seem to contain more residues characteristic of the fungal type SBD sequence. There are positions where these two types differ from each other: (i) only by a conservative substitution (e.g. Gln⁵³⁹ → Glu, Leu⁵⁶⁹ → Val, Arg⁵⁸¹ → Lys); (ii) more significantly by the change of a type of a residue (e.g. Thr⁵¹⁸ → Arg, Asp⁵⁹³ → Gly); and (iii) more drastically by the complete change of hydrophobic character (e.g. Leu⁵⁵¹ → Pro, Ile⁵⁸² → Lys, Asp⁶¹³ → Val). In all these positions some residues are found that lack their unambiguous affiliation to either fungal or bacterial type of SBD. Most of these 'unusual' residues are present in the sequences of SBDs from *Pseudomonas* and *Cryptococcus* (Fig. 1). The three *Pseudomonas* SBDs belong to the maltooligosaccharide-producing amylases that have been reported to constitute an intermediary group in the evolution of true α -amylases and true CGTases [34]. The yeast *Cryptococcus* α -amylase SBD has been pointed out to be responsible for a remarkable thermostability of this α -amylase [21].

Interestingly, at the C-terminus of the SBD sequence, there is an arginine (Arg⁶¹⁶) in the fungal type versus a glutamine in the bacterial type of SBD (Fig. 1).

3.2. Tertiary structure comparison

Fig. 2 depicts the orientation of selected residues. The side-chains of the two most important tryptophans (Trp⁵⁴³ and Trp⁵⁹⁰) forming the binding site 1 are well conserved in both types (fungal and bacterial) of SBD. The side-chains of the other homologous tryptophan (Trp⁶¹⁵) seem to be not so strictly conserved supporting the observation that the Trp⁶¹⁵ has a primarily structural (not functional) role in the *A. niger* glucoamylase SBD [35]. This is the case of the other residues too, e.g. Gln⁵³⁹ and Arg⁵⁹⁶ (Fig. 2) indicating again the possible existence of two types of SBD (cf. Fig. 1).

For the sake of simplicity, the sequence alignment of the two types of SBD (fungal and bacterial) is shown in Fig. 3 together with the N-terminal SBD from *R. oryzae* glucoamylase. Despite the high degree of sequence identity (similarity) between *Aspergillus* and *Bacillus* SBDs which is about 37% (64%), the strand β 3 of fungal SBD is not present in the bacterial SBD, whereas the strand β 6 of bacterial SBD is not present in the fungal SBD. On the other hand, the sequence of *R. oryzae* SBD exhibits a very low degree of similarity to both representative SBDs and contains a major insertion between the strands β 2 and β 3 [13,14,16] (Fig. 3). Nevertheless, this sequence was demonstrated to be responsible for adsorbing to raw starch and degrading this substrate [12,36].

3.3. Evolutionary relationships

All the sequence features described above are reflected in the evolutionary tree (Fig. 4a). The tree clearly demonstrates that the two types of SBD (fungal and bacterial) can be present. The actinomycetes seem to represent a special group. Although these microorganisms belong to bacteria (prokaryotes), their SBD sequences exhibit a lot of features characteristic of the SBD from fungi (cf. Fig. 1). Therefore, in the evolutionary tree, the SBDs from actinomycetes form their own cluster (Fig. 4a) occupying the position closer to eukaryotic fungi than to their 'evolutionary relatives' prokaryotic bacteria. It is worth mentioning that the sequences of catalytic domains of α -amylases from actinomycetes have also been found more homologous to eukaryotic (animal) counterparts than to bacterial ones [37].

As far as the three SBDs from the genus *Pseudomonas* are concerned, two of them taken from maltotetraohydrolases are clustered together with the SBD from *Klebsiella pneumoniae* CGTase (in agreement with taxonomy). The third *Pseudomonas*-originating SBD from maltopentaohydrolase is positioned near the cluster of actinomycetes. Interestingly, in this case, the SBD sequence behaves similarly like the catalytic domain (α -amylase) which has previously been found to contain the sequence features characteristic of the α -amylases from streptomycetes, insects and animals [34].

In general, however, the taxonomy is respected. This means that the SBD sequences reflect their origin (roughly either fungal or bacterial) rather than the enzyme specificity to which they belong (α -, β - or gluco-amylase). This is seen from the overall arrangement of the tree (Fig. 4a) and can be supported by specific examples: (i) the SBD from *Bacillus* sp. strain TS-23 α -amylase positioned among the SBDs from *B. cereus* and *C. thermosulfurogenes* β -amylases; (ii) more remarkably the SBD from *K. pneumoniae* CGTase placed along with the SBDs from *Pseudomonas* maltotetraohydrolases (and not together with the rest of CGTases SBDs); and (iii) most interestingly the SBD from *Aspergillus kawachii* α -amylase clustered together with the bulk of the SBDs from *Aspergillus* (fungal) glucoamylases.

Fig. 1. Sequence alignment of the C-terminal SBDs from amylolytic enzymes. The abbreviations of the sources of the individual amylolytic enzymes are given in Table 1. The residues characteristic of bacterial type and fungal type of SBD are highlighted in blue and red, respectively, while the residues of 'unusual' character are signified by yellow. The residues belonging to the consensus sequence determined by Svensson et al. [14] are written in magenta. The numbers indicate the start of an SBD in the amino acid sequence of each amylase (mature enzymes are used if known). The asterisks denote the C-terminal end of each enzyme. The vertical arrows above the alignment indicate the positions discussed in text using the numbering of the SBD from *Aspergillus niger* glucoamylase.

| | 518 | 527 | 535 | 543 | 551 | 561 | 569 | 582 | 590 | 595 | 613 |
|-------------|-----|----------------------|----------------------|--------------------|-----------------------|------------------------|---------------------|--------------|-----|-----|-----|
| Bacsp-AMY | 481 | WVAKT-SNVFTVNNATTISG | LVVYVAVNIPELGNWNT | ANAIKGNPS | SYPTWKATIAIPQ | GKAEFKFKRQDAQ | NVTWESG-NRNYTYPFSS | TGSYTASQIVP* | | | |
| Stral-AMY | 437 | GRGTGQTSAGFHVN-ATTA | VENIYVTDQALGNWDP | ARALKDEPA | AYPVWKLDFVAA | GTPQKYLKDAAG | KAVWESGA-NRATATGTTG | ALTINDWRG* | | | |
| Strgr-AMY | 437 | GRGTGQTSAGFHVN-ATTA | VENIYVTDQALGNWDP | ARALKDEPA | AYPVWKLDFVAA | GTPQKYLKDAAG | KAVWESGA-NRATATGTTG | ALTINDWRG* | | | |
| Strfl-AMY | 438 | DDPDPVSSGFAVD-ATT | SVGGDIYVTKRDEPLGHWDP | AGGLQDPA | AYPVWKLDFVAA | GTPQKYLKDAAG | KAVWESGA-NRATATGTTG | ALTINDWRG* | | | |
| Striv-AMY | 440 | TTPPTPAGSENVAT-ATT | VGGNIYVTKRDEPLGHWDP | ASALKQDPA | AYPVWKLDFVAA | GTPQKYLKDAAG | KAVWESGA-NRATATGTTG | ALTINDWRG* | | | |
| Struc-AMY | 467 | GDDCTVTVAEFAEEL-VTT | VGGEVAVVGSIFLGSWQP | AGQVRIRDSG | TYPVWGSAGVDP | PA-GVGEKYLKDAAG | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Aspka-AMY | 512 | TATSTLPLPEEEL-VTT | VGGEVAVVGSIFLGSWQP | SDAVKESADDTSSNPE | WSTVTVSIPV | GTTPEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Cryp-AMY | 503 | PGGCSAGTVFDVY-VQT | CGSSWVLAGNIPOLGNWSP | ANGLNANQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bc-8-CGT | 580 | LFGDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWSTG | STAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bacl-1-CGT | 582 | LSGDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWSTG | AAASIGEPFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bm-1-CGT | 582 | LTADQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWTA | ANAIGEAFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bm-2-CGT | 583 | LFGDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDP | NKAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bacoh-CGT | 571 | LFGDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDP | DQAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bacst-CGT | 575 | LTNDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDT | SKAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bs-1-1-CGT | 571 | LSGNQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDA | DKAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bs-17-1-CGT | 582 | LSGDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDP | NNAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bs-38-2-CGT | 581 | LFGDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDA | DKAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bs-201-CGT | 571 | LSGNQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDA | NNAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bs-663-CGT | 580 | LFGDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWSTG | STAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bs-1011-CGT | 582 | LFGDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDP | NNAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bs-E1-CGT | 571 | LSGNQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDA | DKAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bs-Re-CGT | 582 | LSGDQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDP | NNAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Brebr-CGT | 569 | LSGNQ-VTVSEFVNNAST | TEGNLYLTGNVAELGNWDP | NKAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Klepn-CGT | 525 | ENPTV-QSIFMFCNNGYTI | SGSVYIILGNIPOLGNWDL | TKAVKSET | QYVWGSASLEPS | DLNWKCKRNETNPTANVQWESA | NQVFNNDT | QTTNGSF* | | | |
| Thesp-CGT | 579 | LTGNQ-ICVSEFVNNAST | TEGNLYLTGNVAELGNWDT | SKAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Bacst-MGA | 576 | SGTQT-SVVFVTKSAPPNT | GGKYLITGNIPOLGNWSTG | STAIAGEFNQVYVYDVS | PA-GKQEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Pessa-M4H | 429 | GBGGL-VNMFRCNDGVTQ | CGSSVYAVGNVSELGNWSP | ASAVRFDTS | SYPTWKGSIAIPD | GQNEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Pesep-M4H | 425 | EPGAL-VSVSEFRCNDGAT | QCGSSVYAVGNVSELGNWSP | AAALRFDTS | GYPVWGSASLEPS | DLNWKCKRNETNPTANVQWESA | NQVFNNDT | QTTNGSF* | | | |
| Pesep-M5H | 425 | EPGAL-VSVSEFRCNDGAT | QCGSSVYAVGNVSELGNWSP | AAALRFDTS | GYPVWGSASLEPS | DLNWKCKRNETNPTANVQWESA | NQVFNNDT | QTTNGSF* | | | |
| Bacce-BMY | 414 | LLGVP-LLSILVVKNLPTI | CGFVYITGSRALGSWTQ | NISTSTYDHSN | DWRGNVYVPA | ERNIEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Cloth-BMY | 417 | TPNGT-IPVAFETINNATT | YCGENLYVIGSISOLGNWNT | TYARGASCP | NYPTWITLIP | GEQCKAVKIDSSG | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Amore-GMY | 472 | VSSCQVSIENIN-ATT | YCGENLYVIGSISOLGNWNT | TYARGASCP | NYPTWITLIP | GEQCKAVKIDSSG | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Aspka-GMY | 508 | CTTPTAVAFDLT-ATT | YCGENLYVIGSISOLGNWET | SDGIALSADKYTSSDPL | WVTVVTPA | GESIEKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Aspn-1-GMY | 509 | CTTPTAVAFDLT-ATT | YCGENLYVIGSISOLGNWET | SDGIALSADKYTSSDPL | WVTVVTPA | GESIEKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Aspn-2-GMY | 508 | CTTPTAVAFDLT-ATT | YCGENLYVIGSISOLGNWET | SDGIALSADKYTSSDPL | WVTVVTPA | GESIEKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Corro-GMY | 463 | GGSSGSVEVFDVY-ATT | VGGNIYVTKRDEPLGHWDP | ANGVALSSA | NYPTWITLIP | GEQCKAVKIDSSG | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |
| Humgr-GMY | 485 | CADASEVYVENER-VSA | CGEPIKVVGNVVALGNWDT | SKAVTASAGYKSNDFL | WITVTKATGSAVQKYLKDAAG | NVTWESG-NRATATGTTG | ALTINDWRG* | | | | |
| Neutr-GMY | 491 | CAADHEVLVFNENEK-VTT | SGCTVVKVVGSTARLGNWAP | ASGLTSAKQYSSSNPL | WSTVTVSIPV | GTTPEFKFKRQDAQ | NVTWESG-NRATATGTTG | ALTINDWRG* | | | |

Table 1
Amylolytic enzymes used in the present study containing the starch-binding domain

| EC | Abbreviation | Source | Accession ^a |
|-----------|--------------|---|------------------------|
| Family 13 | | | |
| 3.2.1.1 | Aspka-AMY | <i>Aspergillus kawachii</i> | AB008370gb |
| 3.2.1.1 | Bacsp-AMY | <i>Bacillus</i> sp. strain TS-23 | U22045gb |
| 3.2.1.1 | Crysp-AMY | <i>Cryptococcus</i> sp. strain S-2 | D83540gb |
| 3.2.1.1 | Stral-AMY | <i>Streptomyces albidoflavus</i> | P09794sp |
| 3.2.1.1 | Strgr-AMY | <i>Streptomyces griseus</i> | P30270sp |
| 3.2.1.1 | Strli-AMY | <i>Streptomyces lividans</i> strain TK 24 | Z85949gb |
| 3.2.1.1 | Strvi-AMY | <i>Streptomyces violaceus</i> | P22998sp |
| 3.2.1.1 | Thecu-AMY | <i>Thermomonospora curvata</i> | P29750sp |
| 3.2.1.60 | Psesa-M4H | <i>Pseudomonas saccharophila</i> | P22963sp |
| 3.2.1.60 | Psest-M4H | <i>Pseudomonas stutzeri</i> | P13507sp |
| 3.2.1.133 | Bacst-MGA | <i>Bacillus stearothermophilus</i> | P19531sp |
| 3.2.1.- | Psesp-M5H | <i>Pseudomonas</i> sp. strain KO-8940 | D10769gb |
| 2.4.1.19 | Bc-8-CGT | <i>Bacillus circulans</i> strain 8 | P30920sp |
| 2.4.1.19 | Bc-251-CGT | <i>Bacillus circulans</i> strain 251 | P43379sp |
| 2.4.1.19 | Bacl-1-CGT | <i>Bacillus licheniformis</i> | P14014sp |
| 2.4.1.19 | Bm-1-CGT | <i>Bacillus macerans</i> | P31835sp |
| 2.4.1.19 | Bm-2-CGT | <i>Bacillus macerans</i> strain IFO 3490 | P04830sp |
| 2.4.1.19 | Bacoh-CGT | <i>Bacillus ohbensis</i> | P27036sp |
| 2.4.1.19 | Bacst-CGT | <i>Bacillus stearothermophilus</i> | P31797sp |
| 2.4.1.19 | Bs-1-1-CGT | <i>Bacillus</i> sp. strain 1-1 | P31746sp |
| 2.4.1.19 | Bs-17-1-CGT | <i>Bacillus</i> sp. strain 17-1 | P30921sp |
| 2.4.1.19 | Bs-38-2-CGT | <i>Bacillus</i> sp. strain 38-2 | P09121sp |
| 2.4.1.19 | Bs-201-CGT | <i>Bacillus</i> sp. strain KC201 | D13068gb |
| 2.4.1.19 | Bs-663-CGT | <i>Bacillus</i> sp. strain 6.6.3 | P31747sp |
| 2.4.1.19 | Bs-1011-CGT | <i>Bacillus</i> sp. strain 1011 | P05618sp |
| 2.4.1.19 | Bs-E1-CGT | <i>Bacillus</i> sp. strain E1 | Z34466gb |
| 2.4.1.19 | Bs-Re-CGT | <i>Bacillus</i> sp. strain B1018, re-classified | P17692sp |
| 2.4.1.19 | Brebr-CGT | <i>Brevibacillus brevis</i> | AF011388gb |
| 2.4.1.19 | Cloth-CGT | <i>Clostridium thermosulfurogenes</i> | P26827sp |
| 2.4.1.19 | Klep-1-CGT | <i>Klebsiella pneumoniae</i> | P08704sp |
| 2.4.1.19 | Thesp-CGT | <i>Thermoanaerobacter</i> sp. strain ATCC53627 | Z35484gb |
| Family 14 | | | |
| 3.2.1.2 | Bacce-BMY | <i>Bacillus cereus</i> | P36924sp |
| 3.2.1.2 | Cloth-BMY | <i>Clostridium thermosulfurogenes</i> | P19584sp |
| Family 15 | | | |
| 3.2.1.3 | Amore-GMY | <i>Amorphotheca resiniae</i> | X68143gb |
| 3.2.1.3 | Aspaw-GMY | <i>Aspergillus awamori</i> | K02465gb |
| 3.2.1.3 | Aspka-GMY | <i>Aspergillus kawachii</i> | D00427gb |
| 3.2.1.3 | Aspni-GMY | <i>Aspergillus niger</i> | X00712gb |
| 3.2.1.3 | Aspor-GMY | <i>Aspergillus oryzae</i> | D10698gb |
| 3.2.1.3 | Aspsh-GMY | <i>Aspergillus shiroushimi</i> | D10460gb |
| 3.2.1.3 | Corro-GMY | <i>Corticium rolfsii</i> | D49448gb |
| 3.2.1.3 | Humgr-GMY | <i>Humicola grisea</i> | M89475gb |
| 3.2.1.3 | Neucr-GMY | <i>Neurospora crassa</i> | X67291gb |
| 3.2.1.3 | Rhior-GMY | <i>Rhizopus oryzae</i> | D00049gb |

^aThe accession numbers ending with 'sp' and 'gb' are taken from the Swiss-Prot and GenBank sequence databases, respectively.

close proximity in solution due to considerable flexibility of the linker region. The SBD has been moreover found to be more stable in the whole enzyme molecule in comparison with being isolated [44].

3.4. Conclusions

From the evolutionary point of view, the SBD may be an independent module. Its evolution in the three amylolytic families (13, 14, 15) reflects the evolution of species rather than evolution of the individual amylases. The present forms of SBD may constitute the modern 'descendants' of a domain that might either have been joined to or removed from these proteins during the evolution [14]. Since replacing the SBD in *Bacillus macerans* CGTase with the SBD from *Aspergillus awamori* glucoamylase caused a drastic decrease in activity of the parent enzyme [41], in the light of the present study it would be of interest to know the effect of replacing the SBD in, e.g. *Bacillus* CGTase with the SBD from, e.g. *Aspergillus*

α -amylase, i.e. to replace the SBD from bacterial to fungal type, however, from the same family of glycoside hydrolases.

Acknowledgements: This work was financially supported by grants No. 2/1070/97 and 2/6045/99 from the Slovak Grant Agency for Science and by the Howard Hughes Medical Institute Grant No. 75195-547601. Š.J. thanks The Literary Fund of the Slovak Republic for financial support.

References

- [1] Henrissat, B. (1991) *Biochem. J.* 280, 309–316.
- [2] Janeček, Š. (1997) *Prog. Biophys. Mol. Biol.* 67, 67–97.
- [3] Pujadas, G., Ramírez, F.M., Valero, R. and Palau, J. (1996) *Proteins* 25, 456–472.
- [4] Coutinho, P.M. and Reilly, P.J. (1997) *Proteins* 29, 334–347.
- [5] Sierks, M.R. and Svensson, B. (1994) *Protein Eng.* 7, 1479–1484.
- [6] Janeček, Š. and Bateman, A. (1996) *Biol. Bratisl.* 51, 613–628.
- [7] Matsuura, Y., Kusunoki, M., Harada, W. and Kakudo, M. (1984) *J. Biochem.* 95, 697–702.

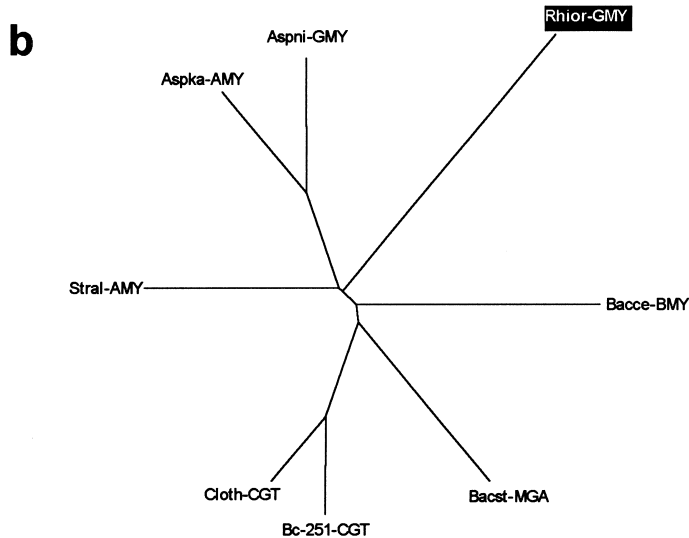
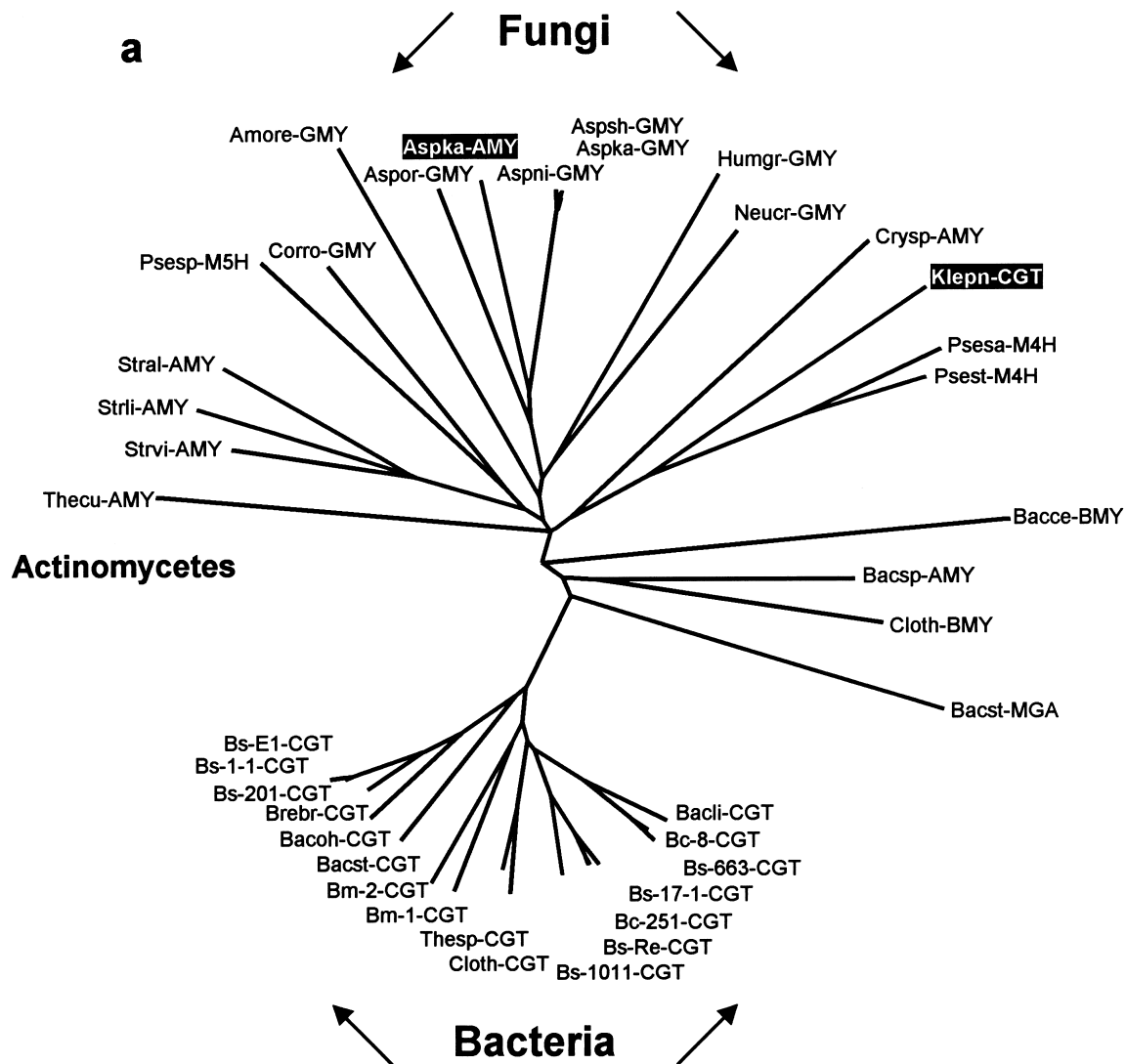


Fig. 4. Evolutionary trees of amyolytic enzymes containing the SBD. a: The tree showing the relationships of all C-terminal SBDs (i.e. all the SBDs from Table 1 except for the SBD from *Rhizopus oryzae* glucoamylase). b: The tree showing the relationships of representative C-terminal SBDs from the above tree (in the same orientation) with the N-terminal SBD from *Rhizopus oryzae* glucoamylase (signified by black rectangle). The abbreviations of the sources of the individual amyolytic enzymes are given in Table 1. The branch lengths are proportional to the divergence of the individual amino acid sequences of the SBD, the sum of the lengths of the branches linking any sources being a measure of the evolutionary distance between them. The two most significant examples in the above tree (Aspka-AMY and Klepn-CGT) that manifest the evolutionary behaviour of SBD (reflecting the evolution of species rather than enzymes) are highlighted by black rectangles.

←

- [8] Mikami, B., Hehre, E.J., Sato, M., Katsube, Y., Hirose, M., Morita, Y. and Sacchettini, J.C. (1993) *Biochemistry* 32, 6836–6845.
- [9] Aleshin, A., Golubev, A., Firsov, L.M. and Honzatko, R.B. (1992) *J. Biol. Chem.* 267, 19291–19298.
- [10] Ševčík, J., Solovicová, A., Hostinová, E., Gašperík, J., Wilson, K.S. and Dauter, Z. (1998) *Acta Cryst. D*54, 854–866.
- [11] Henrissat, B. and Davies, G. (1997) *Curr. Opin. Struct. Biol.* 7, 637–644.
- [12] Ashikari, T., Nakamura, N., Tanaka, Y., Kiuchi, N., ShiBano, Y., Tanaka, T., Amachi, T. and Yoshizumi, H. (1986) *Agric. Biol. Chem.* 50, 957–964.
- [13] Tanaka, Y., Ashikari, T., Nakamura, N., Kiuchi, N., Shibano, Y., Amachi, T. and Yoshizumi, H. (1986) *Agric. Biol. Chem.* 50, 965–969.
- [14] Svensson, B., Jespersen, H., Sierks, M.R. and MacGregor, E.A. (1989) *Biochem. J.* 264, 309–311.
- [15] Coutinho, P.M. and Reilly, P.J. (1994) *Protein Eng.* 7, 393–400.
- [16] Coutinho, P.M. and Reilly, P.J. (1994) *Protein Eng.* 7, 749–760.
- [17] Lawson, C.L., Van Montfort, R., Strokopytov, B., Rozeboom, H.J., Kalk, K.H., De Vries, G.E., Penninga, D., Dijkhuizen, L. and Dijkstra, B.W. (1994) *J. Mol. Biol.* 236, 590–600.
- [18] Sorimachi, K., Jacks, A.J., Le Gal-Coëffet, M.-F., Williamson, G., Archer, D.B. and Williamson, M.P. (1996) *J. Mol. Biol.* 259, 970–987.
- [19] Svensson, B., Pedersen, T.G., Svendsen, I.B., Sakai, T. and Ottesen, M. (1982) *Carlsberg Res. Commun.* 47, 55–69.
- [20] Penninga, D., van der Veen, B.A., Knegt, R.M.A., van Hijum, S.A.F.T., Rozeboom, H.J., Kalk, K.H., Dijkstra, B.W. and Dijkhuizen, L.J. (1996) *Biol. Chem.* 271, 32777–32784.
- [21] Iefuji, H., Chino, M., Kato, M. and Imura, Y. (1996) *Biochem. J.* 318, 989–996.
- [22] Chen, L., Coutinho, P.M., Nikolov, Z. and Ford, C. (1995) *Protein Eng.* 8, 1049–1055.
- [23] Dalmia, B.K., Schütte, K. and Nikolov, Z.L. (1995) *Biotechnol. Bioeng.* 47, 575–584.
- [24] Southall, S.M., Simpson, P.J., Gilbert, H.J., Williamson, G. and Williamson, M.P. (1999) *FEBS Lett.* 447, 58–60.
- [25] Bairoch, A. and Apweiler, R. (1998) *Nucleic Acids Res.* 26, 38–42.
- [26] Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J. and Ouellette, B.F.F. (1998) *Nucleic Acids Res.* 26, 1–7.
- [27] Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) *Nucleic Acids Res.* 22, 4673–4680.
- [28] Saitou, N. and Nei, M. (1987) *Mol. Biol. Evol.* 4, 406–425.
- [29] Felsenstein, J. (1985) *Evolution* 39, 783–791.
- [30] Page, R.D.M. (1996) *Comput. Appl. Biosci.* 12, 357–358.
- [31] Sussman, J.L., Lin, D., Jiang, J., Manning, N.O., Prilusky, J., Ritter, O. and Abola, E.E. (1998) *Acta Crystallogr. D*54, 1078–1084.
- [32] CCP4 Collaborative Computational Project Number 4, *Acta Crystallogr. D*50, 760–763.
- [33] Kraulis, P.J. (1991) *J. Appl. Crystallogr.* 24, 946–950.
- [34] Janeček, Š. (1995) *Biol. Bratisl.* 50, 515–522.
- [35] Williamson, M.P., Le Gal-Coëffet, M.F., Sorimachi, K., Furniss, C.S.M., Archer, D.B. and Williamson, G. (1997) *Biochemistry* 36, 7535–7539.
- [36] Takahashi, T., Kato, K., Ikegami, Y. and Irie, M. (1985) *J. Biochem.* 98, 663–671.
- [37] Janeček, Š. (1994) *Eur. J. Biochem.* 224, 519–524.
- [38] Bui, D.M., Kunze, I., Förster, S., Wartmann, T., Horstmann, C., Manteuffel, R. and Kunze, G. (1996) *Appl. Microbiol. Biotechnol.* 44, 610–619.
- [39] Sorimachi, K., Le Gal-Coëffet, M.F., Williamson, G., Archer, D.B. and Williamson, M.P. (1997) *Structure* 5, 647–661.
- [40] Knegt, R.M.A., Strokopytov, B., Penninga, D., Faber, O.G., Rozeboom, H.J., Kalk, K.H., Dijkhuizen, L. and Dijkstra, B.W. (1995) *J. Biol. Chem.* 270, 29256–29264.
- [41] Chang, H.Y., Irwin, P.M. and Nikolov, Z.L. (1998) *J. Biotechnol.* 65, 191–202.
- [42] Christensen, T., Svensson, B. and Sigurskjold, B.W. (1999) *Biochemistry* 38, 6300–6310.
- [43] Sigurskjold, B.W., Christensen, T., Payre, N., Cottaz, S., Dri-guez, H. and Svensson, B. (1998) *Biochemistry* 37, 10446–10452.
- [44] Tanaka, A., Karita, S., Kosuge, Y., Senoo, K., Obata, H. and Kitamoto, N. (1998) *Biosci. Biotechnol. Biochem.* 62, 2127–2132.