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 $(\beta/\alpha)_8$-barrel proteins [6] (with, however, quite different TIM-barrels [7,8]), while the glucoamylases adopt $(\alpha/\alpha)_6$-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 (Cys^{509}-Arg^{616}) 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

0014-5793/99/$20.00 © 1999 Federation of European Biochemical Societies. All rights reserved.
PII: S0014-5793(99)00919-9

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 Gly^{535} (all amino acid numberings throughout the text use the numbers of the glucoamylase from *A. niger*), Pro^{561} (two cases), Lys^{578}, Trp^{590}, and Asn^{595} (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. $Gln^{539} \rightarrow Glu$, $Leu^{569} \rightarrow Val$, $Arg^{581} \rightarrow Lys$); (ii) more significantly by the change of a type of a residue (e.g. $Thr^{518} \rightarrow Arg$, $Asp^{593} \rightarrow Gly$); and (iii) more drastically by the complete change of hydropathic character (e.g. $Leu^{551} \rightarrow Pro$, $Ile^{582} \rightarrow Lys$, $Asp^{613} \rightarrow Val$). 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 (Arg^{616}) 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 (Trp^{543} and Trp^{590}) forming the binding site 1 are well conserved in both types (fungal and bacterial) of SBD. The side-chains of the other homologous tryptophan (Trp^{615}) seem to be not so strictly conserved supporting the observation that the Trp^{615} 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. Gln^{539} and Arg^{596} (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 β3 of fungal SBD is not present in the bacterial SBD, whereas the strand β6 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 β2 and β3 [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.

```
Position markers (arrows): 518, 527, 535, 539, 543, 551, 561, 562, 569, 578, 581, 582, 590, 593, 595, 596, 613, 615

Bacsp-AMY    481 WVAKT-SNVTFTVNNATTTSGQNVYVVANIPELGNWNT---ANAIKMNPS----SYPTWKATIALPQ-GKAIEFKFIKKDQAG---NVIWESTS-NRTYTVPFSS------TGSYTASWNVP*
Stral-AMY    437 GPGTGQTSASFHVN-ATTAWGENIYVTGDQAALGNWDP---ARALKLDPA----AYPVWKLDVPLAA-GTPFQYKYLRKDAAG---KAVWESGA-NRTATVGTTG------ALTLNDTWRG*
Strgr-AMY    437 GPGTGQTSASFHVN-ATTAWGENIYVTGDQAALGNWDP---ARALKLDPA----AYPVWKLDVPLAA-GTPFQYKYLRKDAAG---KAVWESGA-NRTATVGTTG------ALTLNDTWRG*
Strli-AMY    438 DPDPVSSGVSFAVD-ATTSWGQDIYVTGNRPGLGHWDP---GGGLQLDPA----AYPVWKRDVELPE-GTTFEYKSLRKADAG---NVTWESGA-NRTATVNTT-------KTTLNDTWRN*
Strvi-AMY    440 TTPPATSGASFNVT-ATTVVGQNIYVTGNRAELGNWAP---ASALKLDPA----TYPVWKLTVGLPA-GTSFEYKYIRKDAAG---NVTWESGA-NRTATVPASG------QLVLNDTFRS*
Thecu-AMY    467 GDDCTTVTARFHAT-VTTWYGQEVAVVGSIPELGSWQP---AQGVRLRTDSG--TYPVWSGAVDLPA-GVGFEYKYVKLNRT-----APWSGSR-AATASPPWMTSGGGC-SQNFYDSWR*
Aspka-AMY    512 TATSTTLPITFEEL-VTTTYGEEVYLSGSISQLGEWHT---SDAVKLSADDYTSSNPEWSVTVSLPV-GTTFEYKFIKVDEGG---SVTWESDP-NREYTVPE-CGSGS--GETVVDTWR*
Crysp-AMY    503 PGGCSAGTVVFDVY-VQTQYGQSVVIAGNIPQLGNWSP---ANGLNLNANQYTASSPKWTGTITGVAPGTTFQWKPIVVTNGN----DNWYPGN-NQQATTGSACSSP---AADIEFTWSS*
Bc-8-CGT     580 LTGDQ-VTVRFVVNNASTTLGQNLYLTGNVAELGNWSTG--STAIGPAFNQVIHQYPTWYYDVSVPA-GKQLEFKFFKK-NGS---TITWESGS-NHTFTTPASG------TATVTVNWQ*
Bc-251-CGT   582 LSGDQ-VSVRFVVNNATTALGQNVYLTGSVSELGNWDP---AKAIGPMYNQVVYQYPNWYYDVSVPA-GKTIEFKFLKK-QGS---TVTWEGGS-NHTFTAPSSG------TATINVNWQP*
Bacli-CGT    580 LSGDQ-VSVRFVINNATTALGENIYLTGNVSELGNWTTG--AASIGPAFNQVIHAYPTWYYDVSVPA-GKQLEFKFFKK-NGA---TITWEGGS-NHTFTTPTSG------TATVTINWQ*
Bm-1-CGT     582 LTADQ-VTVRFKVNNATTALGQNVYLTGNVAELGNWTA---ANAIGPMYNQVEASYPTWYFDVSVPA-NTALQFKFIKV-NGS---TVTWEGGN-NHTFTSPSSG------VATVTVDWQN*
Bm-2-CGT     583 LTGDQ-VTVRFLVNQANTNYGTNVYLVGNAAELGSWDP---NKAIGPMYNQVIAKYPSWYYDVSVPA-GTKLDFKFIKK-GGG---TVTWEGGG-NHTYTTPASG------VGTVTVDWQN*
Bacoh-CGT    571 LTGDQ-VSIRFAVNNATTSLGTNLYMVGNVNELGNWDP---DQAIGPMFNQVMYQYPTWYYDISVPA-EENLEYKFIKKDSSG---NVVWESGN-NHTYTTPATG------TDTVLVDWQ*
Bacst-CGT    575 LTNDQ-VSVRFVVNNATTNLGQNIYIVGNVYELGNWDT---SKAIGPMFNQVVYSYPTWYIDVSVPE-GKTIEFKFIKKDSQG---NVTWESGS-NHVYTTPTNT------TGKIIVDWQN*
Bs-1-1-CGT   571 LSGNQ-VSVRFGVNNATTSPGTNLYIVGNVNELGNWDA---DKAIGPMFNQVMYQYPTWYYDISVPA-GKNLEYKYIKKDQNG---NVVWQSGN-NRTYTSPTTG------TDTVMINW*
Bs-17-1-CGT  582 LSGDQ-VSVRFVVNNATTALGQNVYLAGSVSELGNWDP---AKAIGPLYNQVIYQYPTWYYDVTVPA-GKTIEFKFLKK-QGS---TVTWEGGS-NHTFTAPTSG------TATINVNWQP*
Bs-38-2-CGT  581 LTGDQ-VTVRFVINNATTALGQNVFLTGNVSELGNWDP---NNAIGPMYNQVVYQYPTWYYDVSVPA-GQTIEFKFLKK-QGS---TVTWEGGA-NRTFTTPTSG------TATVNVNWQP*
Bs-201-CGT   571 LSGNQ-VSVRFGVNNATTSPGTNLYIVGNVNELGNWDA---DKAIGLMFNQVMYQYPTWYYDISVPA-GKNLEYKYIKKDQNG---NVVWQSGN-NRTYTSPTTG------TDTVMINW*
Bs-663-CGT   580 LTGDQ-VTVRFVINNASTTLGQNIYLTGNVAELGNWSTG--STAIGPAFNQVIHQYPTWYYDVSVPA-GKELEFKFFKK-NGS---TITWEGGS-NHKFTTPASG------TATVTVNWQ*
Bs-1011-CGT  582 LTGDQ-VTVRFVINNATTALGQNVFLTGNVSELGNWDP---NNAIGPMYNQVVYQYPTWYYDVSVPA-GQTIEFKFLKK-QGS---TVTWEGGA-NRTFTTPTSG------TATVNVNWQP*
Bs-E1-CGT    571 LSGNQ-VSVRFGVNNATTSPGTNLYIVGNVNELGNWDA---DKAIGPMFNQVMYQYPTWYYDISALA-GKNLEYKYIKKDQNG---NVVWQSGN-NRTYTSPTTG------TDTVMINW*
Bs-Re-CGT    582 LSGDQ-VSVRFVVNNATTALGQNLYLTGNVSELGNWDP---AKAIGPMYNQVVYQYPNWYYDVSVPA-GKTIEFKFLKK-QGS---TVTWEGGS-NHTFTAPSSG------TATINVNWQP*
Brebr-CGT    569 LSGNQ-VSVRFAVNNATTNSGTNVYIVGNVSELGNWDP---NKAIGPMFNQVMYKYPTWYYDISVPA-GKNLEYKYIKKDHNG---NVTWQSGN-NRTYTSPATG------TDTVISNW*
Cloth-CGT    579 LTGNQ-ICVRFVVNNASTVYGENVYLTGNVAELGNWDT---SKAIGPMFNQVVYQYPTWYYDVSVPA-GTTIQFKFIKK-NGN---TITWEGGS-NHTYTVPSSS------TGTVIVNWQQ*
Klepn-CGT    525 ENPTV-QSINFTCNNGYTISGQSVYIIGNIPQLGGWDL---TKAVKISPT----QYPQWSASLELPS-DLNVEWKCVKRNETNPTANVEWQSGA-NNQFNSNDT--------QTTNGSF*
Thesp-CGT    579 LTGNQ-VCVRFVVNNATTVWGENVYLTGNVAELGNWDT---SKAIGPMFNQVVYQYPTWYYDVSVPA-GTTIEFKFIKK-NGS---TVTWEGGY-NHVYTTPTSG------TATVIVDWQP*
Bacst-MGA    576 SGTQT-SVVFTVKSAPPTNLGDKIYLTGNIPELGNWSTDTSGAVNNAQGPLLAPNYPDWFYVFSVPA-GKTIQFKFFIKRADG---TIQWENGS-NHVATTPTGA------TGNITVTWQN*
Psesa-M4H    429 GEGGL-VNVNFRCDNGVTQMGDSVYAVGNVSQLGNWSP---ASAVRLTDTS---SYPTWKGSIALPD-GQNVEWKCLIRNEADATLVRQWQSGG-NNQVQAAAG--------ASTSGSF*
Psest-M4H    425 EPGAL-VSVSFRCDNGATQMGDSVYAVGNVSQLGNWSP---AAALRLTDTS---GYPTWKGSIALPA-GQNEEWKCLIRNEANATQVRQWQGGA-NNSLTPSEG--------ATTVGRL*
Psesp-M5H    483 AAGAAVLSLTFNET-ADTVWGQNLFVVGNVGALGNWAP---AAGAAMTWISGSGSTGQWRATVQLPA-DTPVQYKYVKKDGAG---NVVWESGG-NRVVTTPAPGA-----TIAVNDSWK*
Bacce-BMY    414 LLGVP-LLSTIVVKNLPTTIGDTVYITGSRAELGSWTQ---NISTSLYYDSHSN---DWRGNVVLPA-ERNIEFKAFIKSKDGR--VKSWQTIQ-QSWNPVPLK-------TTSHTSSW*
Cloth-BMY    417 TPNGT-IPVTFTINNATTYYGQNVYIVGSTSDLGNWNT---TYARGPASCP---NYPTWTITLNLLP-GEQIQFKAVKIDSSG---NVTWEGGS-NHTYTVPTSG------TGSVTITWQN*
Amore-GMY    472 VTSSCQVSITFNIN-ATTYYGENLYVIGNSSDLGAWNI---ADAYPLSASAYTQDRPLWSAAIPLNA-GEVISYQYVRQEDCD----QPYIYETVNRTLTVPA-CGG---AAVTTDAWMGPVGSSGNC*
Aspaw-GMY    515 CTTPTAVAVTFDLT-ATTTYGENIYLVGSISQLGDWET---SDGIALSADKYTSSDPLWYVTVTLPA-GESFEYKFIRIESDD---SVEWESDP-NREYTVPQACGTS---TATVTDTWR*
Aspka-GMY    508 CTTPTAVAVTFDLT-ATTTYGENIYLVGSISQLGDWET---SDGIALSADKYTSSNPLWYVTVTLPA-GESFEYKFIRVESDD---SVEWESDP-NREYTVPQACGES---TATVTDTWR*
Aspni-GMY    509 CTTPTAVAVTFDLT-ATTTYGENIYLVGSISQLGDWET---SDGIALSADKYTSSDPLWYVTVTLPA-GESFEYKFIRIESDD---SVEWESDP-NREYTVPQACGTS---TATVTDTWR*
Aspor-GMY    479 CQVPTTVSVTFAVK-ATTVYGESIKIVGSISQLGSWNP---SSATALNADSYTTDNPLWTGTINLPA-GQSFEYKFIRVQNG----AVTWESDP-NRKYTVPSTCGVK---SAVQSDVWR*
Aspsh-GMY    508 CTTPTAVAVTFDLT-ATTTYGENIYLVGSISQLGDWET---SDGIALSADKYTSSNPPWYVTVTLPA-GESFEYKFIRVESDD---SVEWESDP-NREYTVPQACGES---TATVTDTWR*
Corro-GMY    463 GGSSGSVEVTFDVY-ATTVYGQNIYITGDVSELGNWTP---ANGVALSSA----NYPTWSATIALPA-DTTIQYKYVNIDGS----TVIWEDAISNREITTPASG------TYTEKDTWDES*
Humgr-GMY    485 CADASEVYVTFNER-VSTAWGETIKVVGNVPALGNWDT---SKAVTLSASGYKSNDPLWSITVPIKATGSAVQYKYIKVGTNG---KITWESDP-NRSITLQTASSAGKCAAQTVNDSWR*
Neucr-GMY    491 CAADHEVLVTFNEK-VTTSYGQTVKVVGSIARLGNWAP---ASGLTLSAKQYSSSNPLWSTTIALPQ-GTSFKYKYVVVNSDG---SVKWENDP-DRSYAVGTDCAS----TATLDDTWR*
```

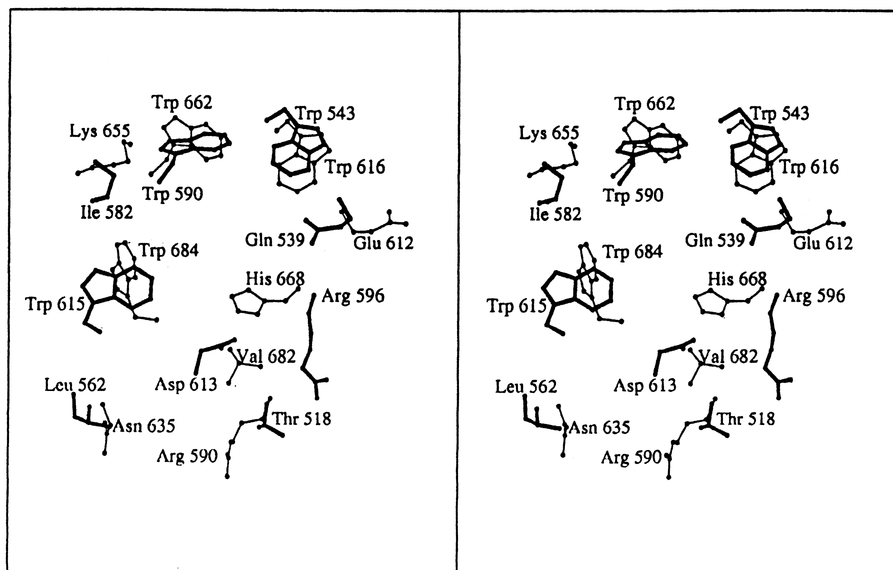

Fig. 2. Similarities and differences between bacterial and fungal type of SBD. The side-chains of the residues of SBD from *Aspergillus niger* glucoamylase (PDB code: 1AC0; thick lines) are overlapped on the side-chains of the residues of SBD from *Bacillus circulans* strain 251 CGTase (PDB code: 1CDG; thin lines). For details, see text.

As far as the relationship of the N-terminal SBD from *R. oryzae* glucoamylase is concerned, the distance tree (Fig. 4b) was constructed using only the 'representative' C-terminal SBD sequences from Fig. 4a. It is evident that *R. oryzae* N-terminal SBD is most distantly related to all present day known SBDs. This 'evolutionary solitude' can be, however, a consequence of lack of data, since *R. oryzae* SBD may represent a newer type SBD at the N-terminus [4,38]. The differences in sequences between the N- and C-terminal SBDs may reflect the possibility that during their molecular evolution the *Rhizopus* and *Aspergillus* glucoamylases obtained their abilities to adsorb to raw starch independently [13].

With regard to the linker region connecting the other constitutive domains (especially the catalytic one) and an SBD, its length varies from a few to several tens of residues [15,17]. Although the SBD may behave as an independent structure-functional module [18,21–23,39–41] with its own evolutionary history (this study), there should be a strong co-operation between the SBD and the other constitutive domains (especially the catalytic one) via the linker region, as documented by studies with the glucoamylase from *A. niger*. It has been, for instance, revealed [42] that the C-terminal part of the linker region has a destabilising effect on the catalytic domain, i.e. the entire structure involving the catalytic domain and the SBD with the full linker is more stable than the catalytic domain containing the linker. It was furthermore demonstrated [43] that the catalytic and starch-binding sites are in

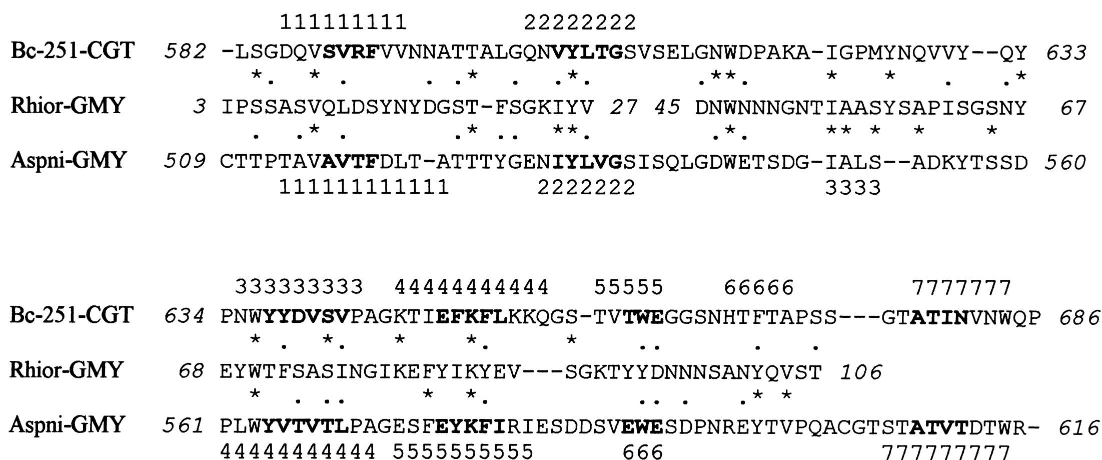

Fig. 3. Comparison of the C-terminal SBD sequences from a bacterial family 13 *Bacillus circulans* CGTase (Bc-251-CGT) and a fungal family 15 *Aspergillus niger* glucoamylase (Aspni-GMY) with the N-terminal SBD sequence from *Rhizopus oryzae* glucoamylase (Rhior-GMY). The seven β-strand segments are indicated by the relevant numbers above and under the alignment blocks. The 27 bolded residues forming the core of the common β-strand segments were used for structural overlap of the two C-terminal SBDs based on least-squares superposition of all 27 C_α-atoms with a root mean square and maximum deviation of 1.13 Å and 1.69 Å, respectively (see Fig. 2). Asterisks and dots above and under the sequence of Rhior-GMY represent the identical and conserved residues between Rhior-GMY and Bc-251-CGT as well as Rhior-GMY and Aspni-GMY, respectively.

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

[a]The 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] Janecek, 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] Janecek, 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.

a

Fungi

Amore-GMY
Aspka-AMY
Aspor-GMY
Aspni-GMY
Aspsh-GMY
Aspka-GMY
Humgr-GMY
Neucr-GMY
Crysp-AMY
Klepn-CGT
Psesa-M4H
Psest-M4H
Psesp-M5H
Corro-GMY
Stral-AMY
Strli-AMY
Strvi-AMY
Thecu-AMY
Actinomycetes
Bacce-BMY
Bacsp-AMY
Cloth-BMY
Bacst-MGA
Bs-E1-CGT
Bs-1-1-CGT
Bs-201-CGT
Brebr-CGT
Bacoh-CGT
Bacst-CGT
Bm-2-CGT
Bm-1-CGT
Thesp-CGT
Cloth-CGT
Bacli-CGT
Bc-8-CGT
Bs-663-CGT
Bs-17-1-CGT
Bc-251-CGT
Bs-Re-CGT
Bs-1011-CGT

Bacteria

b

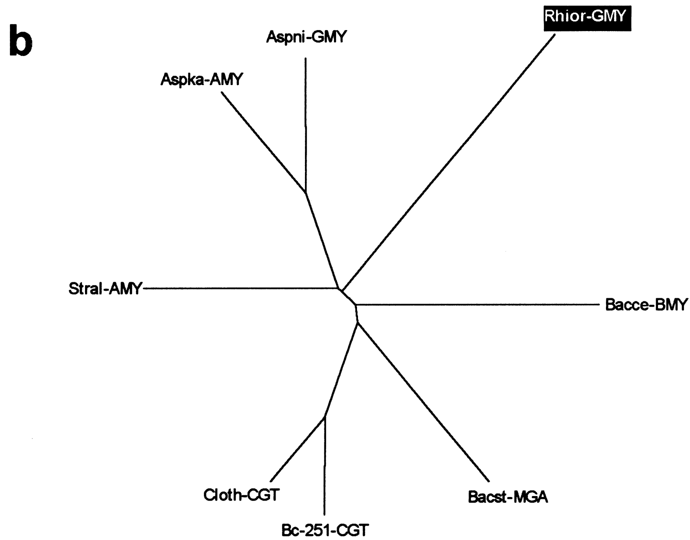

Fig. 4. Evolutionary trees of amylolytic enzymes containing the SBD. a: The tree showing the relationships of all C-terminal SBDs (i.e. all the SBDs from Table 1 except for the SBD from *Rhizopus oryzae* glucoamylase). b: The tree showing the relationships of representative C-terminal SBDs from the above 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 amylolytic enzymes are given in Table 1. The branch lengths are proportional to the divergency 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] Ševcík, J., Solovicová, A., Hostinová, E., Gašperík, J., Wilson, K.S. and Dauter, Z. (1998) Acta Cryst. D54, 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., Knegtel, 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 Iimura, 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. D54, 1078–1084.
[32] CCP4 Collaborative Computational Project Number 4, Acta Crystallogr. D50, 760–763.
[33] Kraulis, P.J. (1991) J. Appl. Crystallogr. 24, 946–950.
[34] Janecek, 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] Janecek, 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] Knegtel, 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., Driguez, 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.