Cell. Mol. Life Sci. 63 (2006) 2710–2724
1420-682X/06/232710-15
DOI 10.1007/s00018-006-6246-9

Cellular and Molecular Life Sciences

Review

Starch-binding domains in the post-genome era

M. Machovič and Š. Janeček*

Institute of Molecular Biology, Slovak Academy of Sciences, Dúbravská cesta 21, 84551 Bratislava (Slovakia), Fax: +421 2 5930 7416; e-mail: Stefan.Janecek@savba.sk

Received 25 May 2006; received after revision 26 June 2006; accepted 3 August 2006
Online First 29 September 2006

Abstract. Starch belongs to the most abundant biopolymers on Earth. As a source of energy, starch is degraded by a large number of various amylolytic enzymes. However, only about 10% of them are capable of binding and degrading raw starch. These enzymes usually possess a distinct sequence-structural module, the so-called starch-binding domain (SBD). In general, all carbohydrate-binding modules (CBMs) have been classified into the CBM families. In this sequence-based classification the individual types of SBDs have been placed into seven CBM families: CBM20, CBM21, CBM25, CBM26, CBM34, CBM41 and CBM45. The family CBM20, known also as a classical C-terminal SBD of microbial amylases, is the most thoroughly studied. The three-dimensional structures have already been determined by X-ray crystallography or nuclear magnetic resonance for SBDs from five CBM families (20, 25, 26, 34 and 41), and the structure of the CBM21 has been modelled. Despite differences among the amino acid sequences, the fold of a distorted β-barrel seems to be conserved together with a similar way of substrate binding (mainly stacking interactions between aromatic residues and glucose rings). SBDs have recently been discovered in many non-amylolytic proteins. These may, for example, have regulatory functions in starch metabolism in plants or glycogen metabolism in mammals. SBDs have also found practical uses.

Keywords. Starch-binding domain, carbohydrate-binding module families, amylase, raw starch degradation, CBM clan, evolutionary relatedness.

Starch-binding domain: a historical overview

In general, the starch-binding domain (SBD), is usually a distinct sequence-structural module that improves the efficiency of an amylolytic enzyme so that the amylase can bind and digest raw starch. Because the motif was first recognised in amylases and thus revealed to cope with raw starch, it was named the raw (granular) starch-binding site [1–4]. Nowadays, due to occurrence of SBDs in a wide spectrum of often non-amylolytic enzymes, it has become reasonable to expect more variable function in specific cases. However, at least at present the SBD still seems to have something to do with polysaccharides related to starch [5]. Nevertheless, the pure starch-binding and degrading function appears to be reserved for microorganisms [6].

Approximately 10% of amylolytic enzymes contain a distinct SBD. It should be noted that there are a few amylolytic enzymes capable of binding and digesting raw starch without a specialized functional domain in their sequence and structure [7–9]. However, these enzymes are outside the scope of the present article and are only briefly mentioned here. The two plant α-amylase isozymes from barley might be the best example. Barley α-amylase contains a surface-binding site consisting of two critically oriented tryptophan residues within the catalytic domain [7, 10, 11]. A second surface site, the so-called sugar tongs surface-binding site (a sugar molecule entrapped by a tyrosine) recently discovered and confirmed in the C-terminal domain [12–14], seems to be unique to the barley isozyme AMY1. A similar tyrosine capturing feature was

* Corresponding author.

also observed in the surface binding site of *Saccharomycopsis fibuligera* raw-starch degrading glucoamylase [15, 16].

The pioneering sequence alignment (Fig. 1) by Svensson et al. [17] in 1989 was significant in galvanising research into SBD structure-function studies. At that time, all the SBDs were known as C-terminally positioned modules except for one located N-terminally in the glucoamylase from *Rhizopus oryzae* [18]. Moreover, neither the classification of glycoside hydrolases [19] nor that of carbohydrate-binding modules [20], which would have made comparison easier, were established. Importantly, Svensson et al. [17] not only aligned the SBD sequences originating from various amylolytic enzymes (basically from α-amylases, β-amylases and glucoamylases), but they also took into account the N-terminal SBD from *R. oryzae* glucoamylase. In fact, the consensus SBD residues they identified (Fig. 1) have held until now.

From the evolutionary point of view, the most insightful observation was made in 1999 by Janecek and Sevcik [21] who demonstrated using the evolutionary tree (Fig. 2) that SBD behaves independently with regard to the catalytic domains of α-amylases, β-amylases and glucoamylases. In other words, the evolution of SBD studied reflected the evolution of species rather than evolution of the individual amylases. In the evolutionary tree based on the alignment of more than 40 SBD sequences, fungi and actinomycetes were clustered separately, surrounded by other bacteria that were also grouped according to their taxonomy. SBD from *Aspergillus kawachii* α-amylase (Fig. 2) clustered together with the bulk of the *Aspergillus*-originated SBDs from glucoamylases has been the most convincing example [21].

Starch-binding domains as CBM families

As mentioned above, for many years SBD was recognised as a C-terminal sequence motif of about 10% of amylases, except for the N-terminal case of *R. oryzae* glucoamylase. The second example of an N-terminal SBD was reported in 1995 by Steyn et al. [22] who described the corresponding motif at the N-terminus of *Lipomyces kononenkoae* α-amylase. Soon after, the SBD of this type was found in glucoamylase from *Arxula adeninivorans* [23]. The situation started to change dramatically in 2000 when Sumitani et al. [24] reported a raw-starch degrading α-amylase from *Bacillus* sp. strain 195. But the starch-binding and degrading function in this α-amylase was ascribed to a tandem C-terminal repeat, with no sequence similarity to previously known 'classical' C-terminal SBDs [24].

At present individual SBDs are best characterised within the sequence-based classification of all carbohydrate-binding modules (CBMs) as the so-called CBM families [20]. This classification has emerged from the well-established classification of catalytic domains that classified α-amylases, β-amylases and glucoamylases into the three respective glycoside hydrolase (GH) families GH13, GH14 and GH15, respectively [19]. Thus, all present-

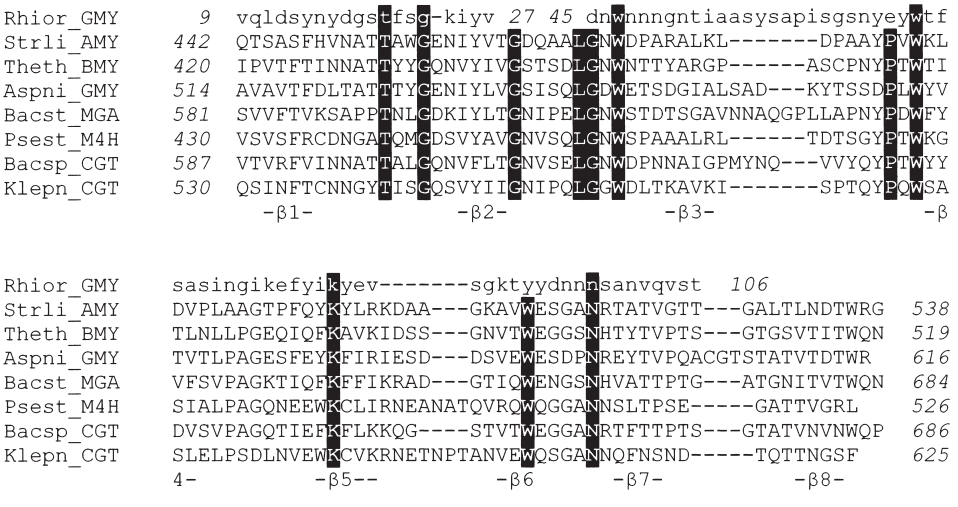

Figure 1. Starch-binding domain of amylolytic enzymes. This is the pioneering alignment of C-terminally positioned SBD motifs belonging to the CBM20 family together with the N-terminal SBD of *Rhizopus oryzae* glucoamylase later classified into CBM21. Sources and abbreviations of the enzymes: Rhior_GMY, glucoamylase from *Rhizopus oryzae* (SwissProt [111] accession number P07683); Strli_AMY, α-amylase from *Streptomyces limosus* (P09794); Theth_BMY, β-amylase from *Thermoanaerobacter thermosulfurogenes* (P19584); Aspni_GMY, glucoamylase from *Aspergillus niger* (P04064); Bacst_MGA, maltogenic α-amylase from *Bacillus stearothermophilus* (P19531); Psest_M4H, maltotetraohydrolase from *Pseudomonas stutzeri* (P13507); Bacsp_CGT, cyclodextrin glucanotransferase from *Bacillus* sp. 1011 (P05618); Klepn_CGT, cyclodextrin glucanotransferase from *Klebsiella pneumoniae* (P08704). Gaps are indicated by dashes. The consensus residues are highlighted by inversion. The sequence of SBD from *R. oryzae* is shown in lower-case letters. Secondary structure elements are indicated under the alignment blocks. Adapted from [17].

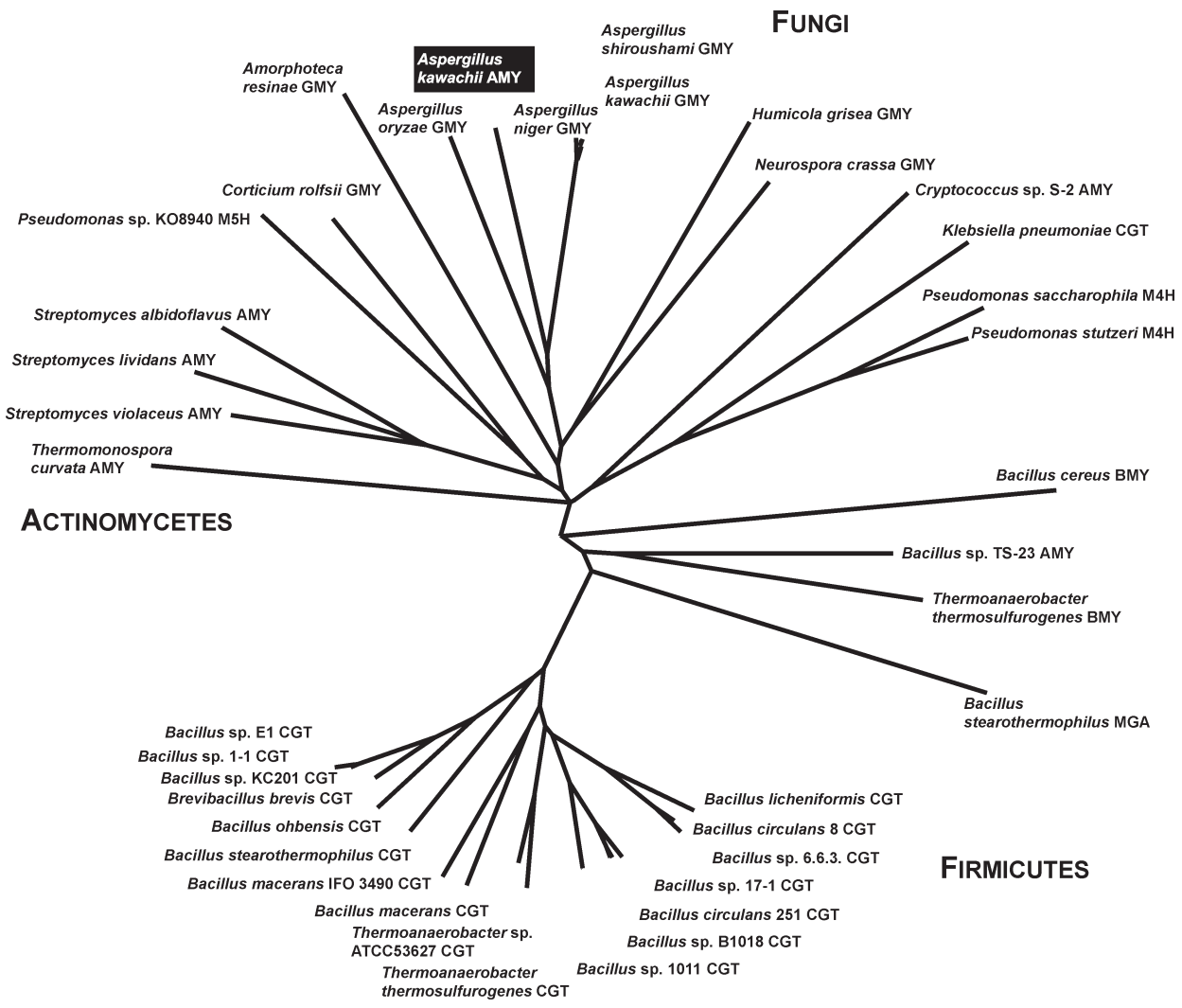

Figure 2. Evolutionary tree of amylolytic enzymes containing the SBD. This is the tree illustrating the evolutionary feature of the CBM20 SBDs reflecting the evolution of species rather than evolution of the individual amylase specificities. The enzymes sources are abbreviated as follows: AMY, α-amylase; CGT, cyclodextrin glucanotransferase; MGA, maltogenic α-amylase; M4H, maltotetraohydrolase; M5H, maltopentaohydrolase; BMY, β-amylase; GMY, glucoamylase. The branch lengths are proportional to the divergence of the SBD sequences. The most striking example (the SBD from *Aspergillus*-originated α-amylase placed among the SBDs *Aspergillus*-originated glucoamylases), which manifests the remarkable evolutionary behaviour of SBD, is highlighted by a black rectangle. Adapted from [21].

day SBDs have been divided into seven CBM families [20, 25]: CBM20, CBM21, CBM25, CBM26, CBM34, CBM41 and CBM45. The entire system is accessible at the CAZy Web server (http://www.cazy.org/CAZY/).

Like the CBM families, also reflect the individual SBD families chronology. This means that the family with the lowest number (CBM20) represents the 'oldest' SBD, i.e. the 'classical' best-known C-terminally positioned SBD, whereas the N-terminal SBDs represented by the SBD from *R. oryzae* glucoamylase were grouped into the family CBM21. The arrangement of all seven CBM families of SBDs with regard to the enzyme specificities of their catalytic domains is schematically illustrated in Figure 3.

The CBM20 family

CBM20 is the most well studied SBD family. The granular raw starch-binding function has been demonstrated in several cases, such as in glucoamylase from *Aspergillus niger* [26, 27], cyclodextrin glucanotransferase (CGTase) from *Bacillus circulans* strain 251 [28, 29], maltogenic α-amylase from *Bacillus stearothermophilus* [30] and β-amylase from *Bacillus cereus* var. *mycoides* [31, 32]. It is possible to note that all real SBD modules, i.e. the SBDs that bind and degrade raw starch, are positioned C-terminally with regard to the catalytic domain [5]. These have been found most frequently in the three GH families: GH13 – the α-amylase family, GH14 – β-amylases and GH15 – glucoamylases. They are predominantly of microbial origin [5]. The α-amylase family (clan GH-

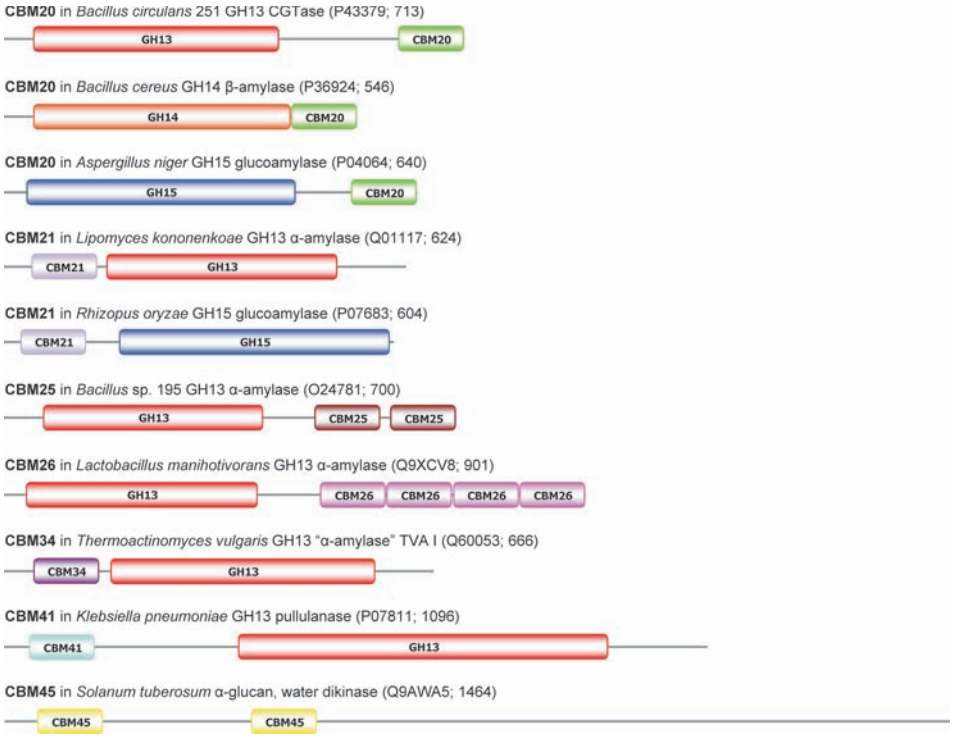

Figure 3. Arrangement of the SBD modules in CBM families. The position and size of both CBMs and GH domains in the individual representatives correlates with the size of the real proteins. The SwissProt accession numbers and the lengths of the enzymes are given in parentheses.

H), which consists of three GH families, 13, 70 and 77 [33], covers most enzyme specificities with SBD [34], such as α-amylase (EC 3.2.1.1), CGTase (EC 2.4.1.19), maltotetraohydrolase (EC 3.2.1.60), maltopentaohydrolase (EC 3.2.1.-), maltogenic α-amylase (EC 3.2.1.133) and amylopullulanase (EC 3.2.1.1/41), plus the GH77 plant-originating 4-α-glucanotransferase (EC 2.4.1.25). The CBM20 motif was recently also found in the GH31 6-α-glucosyltransferase (EC 2.4.1.-) from *Arthrobacter globiformis* [35].

Sequentially similar motifs have also been recognised in many non-amylolytic proteins. They have been added to CBM20 because part of their sequence exhibits similarity to SBD. Since many of these proteins come from genome sequencing projects, their functions are still unknown. However, some of them may possess a significant function related to that of genuine SBD, e.g. α-glucan, water dikinase [36], laforin [37], genethonin-1 [38], carbohydrate esterases and many others [5].

Generally, whereas SBD modules of the CBM20 type are positioned C-terminally in GHs, in non-amylolytic proteins SBD-like modules can be positioned at both the C- and N-termini or even inside the sequence [5]. In any case the modules consist of ~100-amino acid residues (Fig. 1).

The binding sites in CBM20

The basic function of SBD is to bind to raw starch [27, 29, 39]. In general, being a natural part of an amylase, the SBD fulfils several particular roles, such as (i) enabling the enzyme molecule to interact with the insoluble substrate in solution; (ii) delivering the substrate to the active site in the catalytic domain; and (iii) disrupting the surface of the starch granule [6].

The structure-function relationships of the CBM20 motif are quite well understood. Most knowledge is based on nuclear magnetic resonance (NMR) studies of *A. niger* glucoamylase SBD [26, 27] and the X-ray crystallography of *B. circulans* strain 251 CGTase [28, 29]. The CBM20 structures have also been solved for the motifs of CGTases from *Bacillus stearothermophilus* [40], *Bacillus circulans* strain 8 [41], *Bacillus* sp. strain 1011 [42] and *Thermoanaerobacter thermosulfurogenes* [43], maltogenic α-amylase from *B. stearothermophilus* [30] and β-amylase from *B. cereus* var. *mycoides* [31, 32].

The three-dimensional structure of CBM20 SBD of *A. niger* glucoamylase (i.e. of a GH15 amylase) has been determined in solution by NMR in both a free state [26] and in a complex with β-cyclodextrin, a cyclic analogue of starch [27]. The SBD forms an open-sided, distorted, β-barrel structure [26]. The overall topology shows eight β-strands (Fig. 4a) arranged into two major β-sheets [44]. One is a five-stranded antiparallel β-sheet, while the sec-

ond sheet consists of one parallel and one antiparallel β-strand pair. The N- and C-termini are at opposite ends of the longest axis of the molecule [26]. The structure of the SBD bound to β-cyclodextrin showed the presence of the two binding sites (Fig. 4a), which differ from each other functionally as well as structurally [27]. Binding site 1 is small, accessible and acts as the initial starch recognition site. In comparison with free SBD, the structure of site 1 remains almost unchanged in the SBD-β-cyclodextrin complex. It is composed of two dominant tryptophans, Trp^{543} and Trp^{590} (Fig. 1), which are essential for binding. They form a compact, rigid and surface-exposed hydrophobic site with an inter-ring spacing appropriate for binding to α-1,4-linked glucoses [27]. On the other hand, binding site 2 is more extended and flexible, since upon binding to β-cyclodextrin, it undergoes a significant conformational rearrangement. The stacking interaction at this site between SBD and the substrate is also governed by hydrophobic effects from two aromatic rings: Tyr^{527} and Tyr^{556}. Of the two tyrosines, the former appears to be very flexible, whereas the latter seems to be reasonably static [27]. The third well-conserved tryptophan, Trp^{563} (Fig. 1), has been classified as a buried residue with negligible surface accessibility. Although incapable of direct interaction with β-cyclodextrin, it was found in close contact with many residues in and around binding site 2 [27]. Finally, although the fourth tryptophan, Trp^{615} (Fig. 1), is not so strictly conserved in CBM20 [21], it may have an essential structural (i.e. not functional) role in SBD from *A. niger* glucoamylase [45].

Two analogous binding sites were also identified in the SBD X-ray structure of maltose-dependent CGTase from *B. circulans* strain 251 [29], i.e. of a GH13 amylolytic enzyme. Binding site 1, involving two tryptophans, Trp^{616}

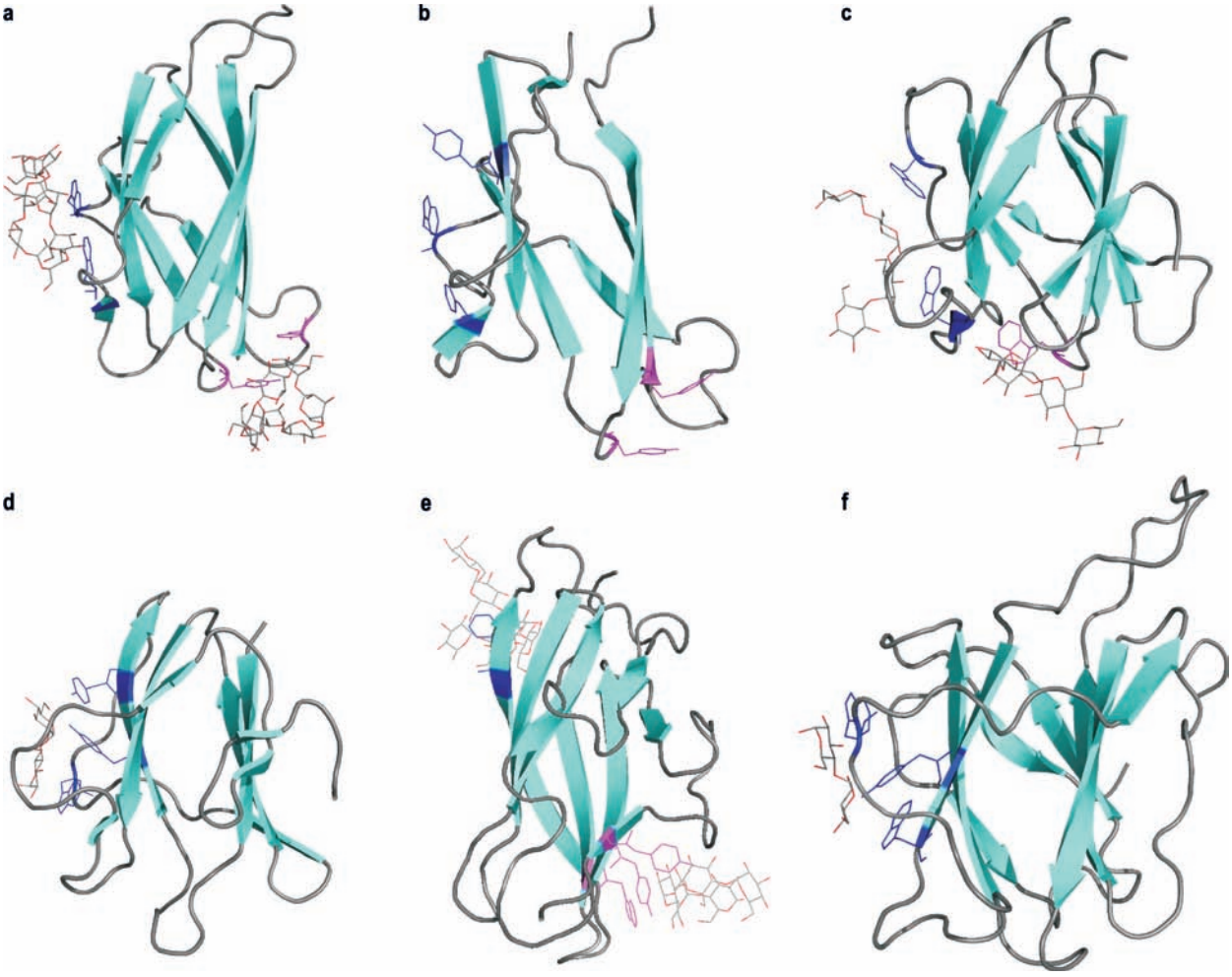

Figure 4. The structural features of SBD from the individual CBM families. Wherever possible, structures are oriented similarly. The tryptophan and tyrosine residues involved in stacking interactions with substrate in binding site 1 are coloured blue and in binding site 2, if present, by magenta. (*a*) The CBM20 of *Aspergillus niger* glucoamylase in complex with β-cyclodextrin [27]; PDB code [112]: 1ac0. (*b*) The CBM21 of *Rhizopus oryzae* glucoamylase [60]. (*c*) The CBM25 of *Bacillus halodurans* maltohexaose-forming amylase [64]; PDB code: 2c3w. (*d*) The CBM26 of *Bacillus halodurans* maltohexaose-forming amylase [64]; PDB code: 2c3h. (*e*) The CBM 34 of *Thermoactinomyces vulgaris* 'α-amylase' TVA I [74]; PDB code: 1uh4. (*f*) The CBM41 of *Klebsiella pneumoniae* pullulanase [81]; PDB code: 2fhf. The structures were displayed with the program Pymol [113].

and Trp^{662} (corresponding to Trp^{543} and Trp^{590} of *A. niger* SBD), is most important for raw starch-binding capacity, whereas binding site 2 (Tyr^{633}) has an important role in guiding linear starch chains to the active site [29]. The situation should be similar in the SBD of β-amylase, i.e. of a GH14 amylase, especially in binding site 1 (Trp^{449} and Trp^{495} in *B. cereus* β-amylase) [31, 32, 46, 47]. However, binding studies of *A. niger* glucoamylase SBD with amylose suggest that both binding sites might be essential for correct CBM20 functioning [48].

In an effort to elucidate SBD boundaries and the contribution of individual SBD segments to raw starch binding and degrading, several authors undertook deletion analyses of SBDs. Goto et al. [49] prepared mutant glucoamylases from *Aspergillus awamori* var. *kawachi*, in which the SBD was truncated by deleting the C-terminal segments according to the position of the four tryptophan residues. They revealed that the sequence around Trp^{562} (i.e. Trp^{563} in *A. niger* glucoamylase) is essential for digestion of raw starch, whereas the sequence around Trp^{589} (Trp^{590}) contributes to the adsorption to raw starch [49]. Chen et al. [50] studied mutant glucoamylase from *A. awamori* containing extensive deletions at the C-terminus of or within the SBD. They concluded that any deletion mutations in SBD result in diminution or loss of raw starch binding and digesting ability of the glucoamylase [50]. Using a similar approach with the α-amylase from *Bacillus* sp. strain TS23, Lo et al. [51] observed in SBD-deletion mutants a diminution in raw starch-binding ability but without significant effect on raw starch-degrading activity. Mutation of Trp^{545} and Trp^{588} (equivalent to Trp^{616} and Trp^{662} in *B. circulans* CGTase) in the same α-amylase confirmed that the two tryptophans are also important for this SBD [52].

The CBM21 family

There has been substantially less information about CBM21 in comparison with CBM20. The key member of the CBM21 family is the SBD from *R. oryzae* glucoamylase [18]. The SBD function was ascribed to the N-terminal part of the enzyme based on comparison of multiple forms (resulting probably from limited proteolysis) differing in the ability to adsorb and degrade raw starch [18, 53], supported by sequence comparison with other glucoamylases [54]. Although at present five amylolytic enzymes are known to contain this type of N-terminally positioned SBD [5], the unambiguous evidence of actual SBD function was demonstrated only for *R. oryzae* glucoamylase [18]. Of the four additional amylases, two are α-amylases from *Lipomyces kononenkoae* [22] and *Lipomyces starkeyi* [55] and two are glucoamylases from *Arxula adeninivornas* [23] and *Mucor circinelloides* [56]. Of these the N-terminal CBM21 motif of the *L. kononenkoae* α-amylase LKA1 was recently shown to be responsible for raw starch binding [57].

A substantial part of the CBM21 family is formed by the group of regulatory subunits of Ser/Thr-specific protein phosphatases [5]. These regulatory subunits direct the protein phosphatase to glycogen [58]. Bork et al. [59] described their sequence similarities to SBDs of both CBM20 and CBM21, but more pronounced similarities have been found with members of the CBM21 family [5].

The CBM21 SBD is ~100 amino acid residues long. In amylases the SBD module is located exclusively N-terminally, whereas in regulatory subunits of Ser/Thr-specific protein phosphatases and other putative proteins its exact position may vary; the C-terminal positioning seems nevertheless to dominate [5].

To date, no three-dimensional structure has been determined for a CBM21 member [20]; however, Chou et al. [60] recently elucidated the two raw starch-binding sites in SBD from *R. oryzae* glucoamylase based on the modelling the structure and site-directed mutagenesis. They not only revealed that the overall fold of the CBM21 should be closely similar to that of the CBM20, but they also identified two raw-starch binding sites analogous to those present in CBM20 (Fig. 4b). Binding site 1 (responsible mainly for binding) involves the residues Trp^{47} and Tyr^{93} (corresponding to Trp^{543} and Trp^{590} of *A. niger* glucoamylase; Fig. 1), whereas the key residue of binding site 2 (responsible mainly for facilitating binding) was identified as Tyr^{32} [60]. These results have clearly confirmed the pioneering sequence alignment of SBDs (Fig. 1) [17] and strongly support the recent proposal [5] to group the CBM20 and CBM21 families into a common CBM clan.

The CBM20 and CBM21 clan

The idea of a common evolutionary origin of the two best-known types of SBD, i.e. classical C-terminal CBM20 SBD from *A. niger* glucoamylase [61] and *B. circulans* CGTase [62] and the less-characterised N-terminal CBM21 SBD from *R. oryzae* glucoamylase [18], was already indicated in 1989 by Svensson et al. [17] who aligned the sequences together. Later, this idea was reinforced by finding a motif similar to the N-terminal SBD in the group of eukaryotic (also mammalian) regulatory subunits of Ser/Thr-specific protein phosphatases [59] that target the protein phosphatases to glycogen. The possibility that C- and N-terminal SBDs are related was taken up again Janecek and Sevcik [21]; however, due to the lack of relevant CBM21 amylase-originating sequence data, no insightful conclusions could be made at that time. Moreover, when in 1999 Coutinho and Henrissat [20] established the sequence-based classification of

all carbohydrate-binding modules into the CBM families, the two SBDs were placed into two different CBM families: CBM20 (*A. niger* glucoamylase, *B. circulans* CGTase, etc.) and CBM21 (*R. oryzae* glucoamylase).

Nevertheless, in the last decade the complete genome sequencing projects have offered many novel proteins (often not amylolytic and often only putative ones) that exhibit unambiguous sequence similarities with SBDs of either CBM20 or CBM21 type. Thus laforin (a dual Tyr-Thr/Ser-specific protein phosphatase involved in the Lafora type of epilepsy) [37], genethonin-1 (a skeletal muscle protein of yet unknown function) [38] and the α-glucan, water dikinase (a plant-specific glucan phosphorylase) [36], can be mentioned as perhaps the most remarkable examples in addition to the regulatory subunits of Ser/Thr-specific protein phosphatases mentioned above. It has also become clear that the original idea [17, 21] of the CBM20 module being at the C-terminus and the CBM21 module at the N-terminus of a protein is no longer applicable [5]. This was especially due to the SBDs discovered in non-amylolytic proteins.

The two CBM families CBM20 and CBM21 are finally now suggested to share a common evolutionary origin based on a rigorous bioinformatics analysis of 125 CBM20 members and 56 CBM21 members [5]. The best-conserved sequence features (Fig. 5) cover the regions around the two tryptophans (Trp^{543} and Trp^{590} of the *A. niger* glucoamylase CBM20) that constitutes the starch-binding site 1 [27]. On the evolutionary tree the two CBM families retain their own independence [5], the SBDs from the GH13 bacterial amylopullulanases being revealed as candidates for evolutionary intermediates between the two CBM families (Fig. 6). An enzyme clan within the classification of catalytic modules consists of a group of enzyme families with a common ancestry, similar tertiary structure and conserved catalytic machinery and reaction mechanisms [63]. A CBM clan proposed by Machovic et al. [5] contains CBM families having a common evolutionary origin, similar tertiary structure and similar binding site residues and mode of carbohydrate binding. The idea that the CBM20 and CBM21 families should be grouped into a common CBM clan is strongly been supported by the recent rational modelling of the *R. oryzae* CBM21 three-dimensional structure (cf. Fig. 4a,b) and identifying its two raw starch-binding sites as analogous to those present in the CBM20 family [60].

The CBM25, CBM26, CBM34, CBM41 and CBM45 families

Currently there are five CBM families with the SBD function in addition to the above-mentioned CBM20 and CBM21 [20]. These are CBM25, CBM26, CBM34, CBM41 and CBM45. The schematic arrangement of their representatives in a real protein is illustrated in Figure 3. It is worth mentioning that although classifying these SBD modules into the individual CBM families has mainly been based on the differences between their amino acid sequences, the basic features of their three-dimensional structures as well as the rough architecture of the raw starch-binding site(s) seem to be adopted in a similar manner (Fig. 4).

The CBM25 and CBM26 families

The above-mentioned possibility was also pointed out by Boraston et al. [64] who recently solved the structures of both CBM25 (Fig. 4c) and CBM26 (Fig. 4d), which are naturally present in the maltohexaose-forming amylase from *Bacillus halodurans* [65]. Of these two structurally related CBM families [20], CBM25 was established based on revealing a novel type of SBD in 1998 in the α-amylase from a *Bacillus* sp. strain 195 [24]. This α-amylase contains CBM25 in two copies (Fig. 3), but CBM25 may also be present in a single copy, e.g. in the β-amylase from *Bacillus circulans* [66]. The CBM26 family relies especially on the SBD motifs recognised in the α-amylases from lactobacilli [67]. The SBD function has also been demonstrated in the maltotriose-forming amylase from *Streptococcus bovis* [68]. The SBD modules of CBM26 are mostly organized in tandem repeats [69–71] (Fig. 3). Both CBM25 and CBM26 are small families; at present they count no more than 20 members each [20]. Despite the low degree of sequence identity between CBM25 and CBM26 (~15%), the overall topology looks very similar [64]. Both motifs adopt extremely related β-sandwich folds – 10 β-strands for CBM25 and 9 β-strands for CBM26 (Fig. 4c,d); strand β3 of CBM25 has no equivalent in CBM26. These folds are closely similar to that recognised for CBM20 (Fig. 4a) and even CBM21 (Fig. 4b). Based on the structure of CBM25 and CBM26 complexed with maltotetraose and maltose, respectively, the former very probably contains two binding sites, whereas the latter possesses only one site for binding raw starch [64]. Of the residues forming the main binding platform, only the central amino acid residue, His^{882} in CBM25 and Tyr^{788} in CBM26 (*B. halodurans* maltohexaose-forming amylase numbering), is conserved at the structural and sequence level. The other two residues involved in the binding site – Trp^{890} and Trp^{930} of CBM25 and Trp^{801} and Tyr^{790} of CBM26 – overlap in space but do not correspond with each other in sequence [64].

The CBM34 family

The CBM34 family belongs to the large CBM families (~100 CAZy entries) [20]. From the point of view of a

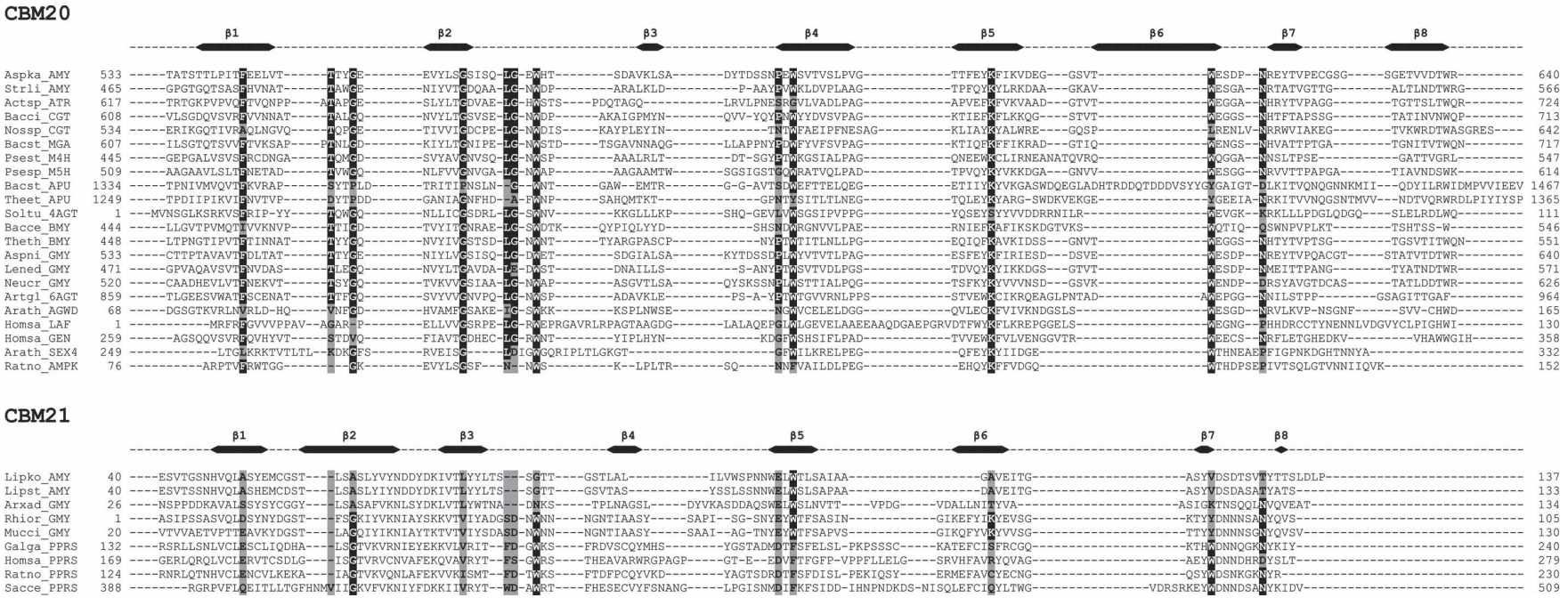

Figure 5. Amino acid sequence alignment of selected representatives of CBM20 and CBM21. The alignment reflects the proposal [5] that these two families constitute a common CBM clan. The consensus residues highly conserved in CBM20 are highlighted by black-and-white inversion. Substitutions or deletions in these positions are highlighted in grey. The β-strands shown above the alignment blocks are derived from the NMR structure of *Aspergillus niger* glucoamylase CBM20 [27] and the modelled structure of *Rhizopus oryzae* glucoamylase CBM21 [60]. Abbreviations of enzymes and protein sources (in capitals): AMY, α-amylase; ATR, acarviose transferase; CGT, cyclodextrin glucanotransferase; MGA, maltogenic α-amylase; M4H, maltotetraohydrolase; M5H, maltopentaohydrolase; APU, amylopullulanase; 4AGT, 4-α-glucanotransferase; BMY, β-amylase; GMY, glucoamylase; 6AGT, 6-α-glucosyltransferase; AGWD, α-glucan, water dikinase; LAF, laforin; GEN, genethonin-1; SEX4, starch excess 4 protein; AMPK, β-subunit of AMP-activated protein kinase; PPRS, regulatory subunit of protein phosphatase. The organisms are abbreviated as follows: Aspka, *Aspergillus kawachii*; Strli, *Streptomyces limosus*; Actsp, *Actinoplanes* sp. 50/100; Bacci, *Bacillus circulans*; Nossp, *Nostoc* sp. PCC 7120; Bacst, *Bacillus stearothermophilus*; Psest, *Pseudomonas stutzeri*; Psesp, *Pseudomonas* sp. KO-8940; Theet, *Thermoanaerobacter ethanolicus*; Soltu, *Solanum tuberosum*; Bacce, *Bacillus cereus*; Theth, *Thermoanaerobacter thermosulfurogenes*; Aspni, *Aspergillus niger*; Lened, *Lentinula edodes*; Neucr, *Neurospora crassa*; Artgl, *Arthrobacter globiformis*; Arath, *Arabidopsis thaliana*; Homsa, *Homo sapiens*; Ratno, *Rattus norvegicus*; Lipko, *Lipomyces kononenkoae*; Lipst, *Lipomyces starkeyi*; Arxad, *Arxula adeninivorans*; Rhior, *Rhizopus oryzae*; Mucci, *Mucor circinelloides*; Galga, *Gallus gallus*; Sacce, *Saccharomyces cerevisiae*. The SBD of the starch excess 4 protein (Arath_SEX4) and the glycogen-binding domain of the β-subunit of AMP-activated protein kinase (Ratno_AMPK) were added to CBM20 members due to their unambiguous sequence similarities, although they still have not been classified into the CAZy system [20]. The sequences were aligned with the program ClustalW [114]. The sequences are numbered including the signal peptides, if any.

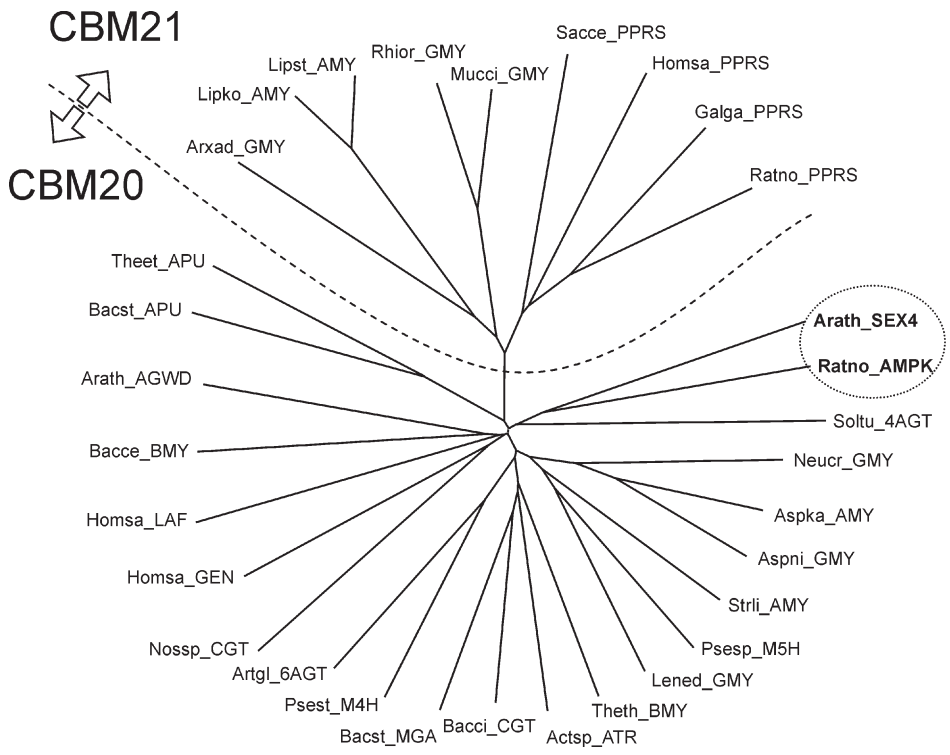

Figure 6. Evolutionary tree of selected representatives of CBM20 and CBM21. The tree is based on the alignment shown in Figure 5, where all the abbreviations are also explained. The SBD-like sequences from GH13 amylopullulanases were recently identified as the candidates for the intermediate members between the two CBM families [5]. Based on the most recent observations [85, 100] the SBD of the starch excess 4 protein (Arath_SEX4) and the glycogen-binding domain of the β-subunit of AMP-activated protein kinase (Ratno_AMPK) should also be of interest. The tree was calculated on the EBI server (http://www.ebi.ac.uk/) within the ClustalW package as a Phylip tree type and displayed with the program TreeView [115].

catalytic domain, this family covers the N-terminal domains of neopullulanase, maltogenic amylase and cyclomaltodextrinase [72]. These are all closely related members of the α-amylase family GH13 grouped into the so-called neopullulanase subfamily [73]. The raw starch-binding function has been demonstrated in the case of the N-terminal domain of *Thermoactinomyces vulgaris* 'α-amylase' TVA I [74]. Both TVA I and TVA II are a typical α-amylases [75]. Since they exhibit the properties of neopullulanase, maltogenic amylase and cyclomaltodextrinase, it is difficult to describe them succinctly [76]. Domain N of TVA I acts as an anchor in the catalytic reaction of the enzyme; function as a pullulan-binding domain has also been suggested [77]. The CBM34 fold (Fig. 4e) is very similar to CBM20, CBM21 and both CBM25 and CBM26 because it is also a distorted β-barrel structure consisting of nine β-strand segments [78]. The crucial amino acid residue involved in the raw starch binding in TVA I, Trp^{65} [74], clearly has no counterpart in TVA II despite the otherwise pronounced sequence similarity between the two respective domains N. The absence of that tryptophan could be one of the factors responsible for the isolated position of domain N in the structure of TVA II, which makes possible critical participation in the formation of a TVA II dimer [78].

The CBM41 family

The CBM41 family was defined in 2004 based on the finding [79] that the very N-terminal domain of the GH13 pullulanase from *Thermotoga maritima* [80] binds tightly to α-glucans. The family (more than 50 entries) covers mostly bacterial pullulanases from the α-amylase family GH13 and many uncharacterised proteins [20]. Interestingly, a substantial number of these pullulanases are from human pathogens [79]. The recent structural study [81], however, has revealed an even more remarkable fact: similarity of the sugar-binding site of CBM41 to that of CBM20. Despite the differences between the amino acid sequences of CBM20 and CBM41 [20], the overall structure of CBM41 (Fig. 4f) is again a distorted antiparallel β-sandwich fold (eight-stranded in this case) with two tryptophans, Trp^{80} and Trp^{95} (*Klebsiella pneumoniae* pullulanase numbering [82]), making stacking interactions with the glucose rings of the substrate [81]. The position of the tryptophan residues is different in CBM41 and CBM20; moreover, the additional CBM41 binding residue, Tyr^{78}, has no equivalent in the CBM20 structure [81].

The CBM45 family

The CBM45 family belongs to the most recently established CBM families [20]. Until now, all its mem-

bers have originated from eukaryotic proteins from the plant kingdom as the N-terminal modules of plastidial α-amylases and α-glucan, water dikinases. The experimental evidence was delivered by Mikkelsen et al. [83], who revealed in the potato α-glucan, water dikinase that the N-terminally positioned motif of this enzyme is specific for plastidial α-glucan degradation. This type of SBD usually occurs as tandem repeats containing three conserved tryptophans, two of which (Trp^{62} and Trp^{117}) were already confirmed to be responsible for carbohydrate binding [83]. Unfortunately, the three-dimensional structure of a CBM45 motif has not yet been determined.

SBD-related homologues in plants and mammals

Nowadays, in the post-genome era when complete sequences of a number of genomes has become available, SBDs or at least the sequences exhibiting similarities to SBD have also been recognised in enzymes and proteins that are not necessarily amylases. Dual-specific protein phosphatases should deserve special attention because these proteins are involved in various important physiological processes in plants and mammals [84]. It is worth mentioning that in plants these processes concern starch metabolism, whereas in mammals they participate in the metabolism of glycogen [85]. The presence of an SBD motif in the sequences of protein phosphatases reflects their regulatory function since they are involved in polysaccharide metabolism indirectly via modulation of activity of degradative enzymes (i.e. also amylases), such as isoamylase, β-amylase and disproportionating enzyme [85]. In particular, the initial steps of starch degradation at the granule surface are regulated, and the modulation concerns mainly phosphorylation [85].

SEX4 protein

The starch excess 4 (*SEX4*) locus encodes a putative dual-specificity protein phosphatase [86]. Genes coding for highly similar proteins have been known to be evolutionarily conserved in higher plants, such as tomato, rice and maize. Recently, Niittyla et al. [85] and Kerk et al. [87] independently reported on the presence of a CBM20 motif at the C-terminal end of the SEX4 protein exhibiting, in fact, all relevant CBM20 sequence features (Fig. 5). This protein is chloroplastic, can bind to starch and regulates the initial steps of starch degradation [85, 87].

Laforin

Laforins are dual-specificity protein phosphatases of animal origin that contain a CBM20 motif (Fig. 5) at the N-terminus [88]. Interestingly, mammalian laforins are the proteins most closely related to the SEX4-like proteins (Fig. 6), which control the above-mentioned starch metabolism in plants [85]. In humans, laforin is a product of the *EPM2A* gene, and it was shown that mutations of this gene cause the Lafora form of epilepsy [89]. The CBM20 motif of laforin was demonstrated to be critical for association with glycogen both *in vitro* and *in vivo* [90]; however, it does indeed preferentially bind starch versus glycogen [91–93]. The presence of a functional SBD in laforin is necessary due to accumulation of Lafora polyglucosan bodies consisting almost exclusively of glucose molecules. Laforin has to detect the appearance of Lafora bodies and control mechanisms to preclude their further formation or to initiate their elimination [88].

Genethonin-1

Genethonin-1 [38] is another human protein containing an unambiguous C-terminally positioned CBM20 motif (Fig. 5). This skeletal muscle protein could play a structural or regulatory role [94]; however, it deserves to be studied in a more detail because its exact function remains undiscovered. Moreover, the CBM20 of genethonin-1 displays a remarkably high ~40% sequence similarity to those of amylases [38].

Glycogen-binding domain of the β-subunit of AMPK

AMPK is an $\alpha\beta\gamma$ heterotrimer AMP-activated protein kinase that co-ordinates cellular metabolism in response to energy demand and other stimuli [95]. The glycogen-binding domain (Fig. 5) forms a middle part of the β-subunit (β-GBD) and is responsible for localising AMPK to glycogen [96, 97]. From the evolutionary point of view, Polekhina et al. [97] indicated the relatedness of β-GBD to representatives of both the CBM20 and CBM21 families and additionally pointed out its more pronounced similarity to domain N of GH13 isoamylase [98] and branching enzyme [99]. With regard to the recently proposed CBM clan of CBM20 and CBM21 [5], the β-GBD together with the SBD-like motifs of GH13 amylopullulanases could represent the CBMs intermediary between the two respective families (Fig. 6). The structure of rat β-GBD was recently solved [100] and confirmed a fold of antiparallel β-sandwich (Fig. 7) that is related, as predicted from sequence similarities, to those of CBM20 and CBM21 (Fig. 4). The complex of β-GBD with β-cyclodextrin revealed [100] the essential roles of two conserved tryptophans, Trp^{100} and Trp^{133} (rat β-GBD numbering [101]), thus suggesting a similar situation in binding both the glycogen in β-GBD and the starch in classical SBDs of the CBM20 and CBM21 families.

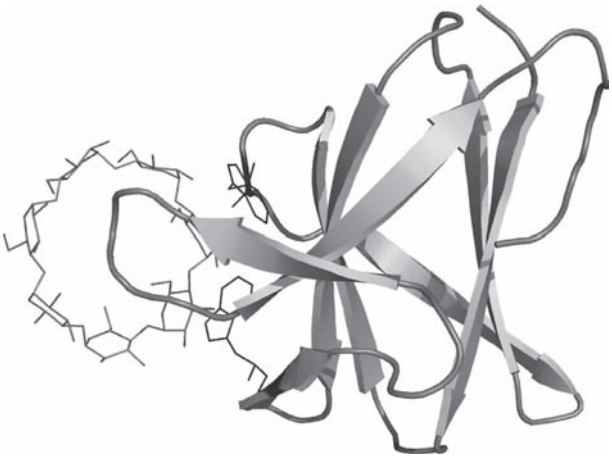

Figure 7. The structural features of the GBD of the β-subunit of AMPK. The structure is, so far as possible, oriented similarly to those illustrated in Figure 4. The two tryptophans involved in stacking interactions with the substrate are shown. PDB code: 1z0m4 [100]. The structure was displayed using the program Pymol [113].

Examples of practical applications of SBDs

SBDs have found practical utilization in various aspects of science and technology. The most important examples are applications of SBDs related to (i) purification of recombinant biologically active proteins, e.g. affinity purification of β-galactosidase-SBD fusion protein [102]; (ii) starch bioengineering, e.g. SBDs used as tools to anchor proteins (that do not have affinity for starch granules) inside starch granules during their biosynthesis [103, 104]; (iii) probiotic food technologies, e.g. expression of SBDs at the cell surface of bifidobacteria to make them susceptible to raw starch binding [105]; (iv) biomedical applications ranging from bone replacement to engineering of tissue scaffolds and drug delivery systems, e.g. a cellulose/starch cross-bridging protein composed of a cellulose-binding domain and an SBD fused in-frame via a synthetic elastin linker [106]; (v) improving the properties of non-amylolytic enzymes, e.g. leucine aminopeptidase II fused with SBD to gain higher thermostability and catalytic efficiency [107]; and (vi) enabling amylases without the ability to bind and degrade raw starch to perform these functions, e.g. α-amylase from *Bacillus subtilis* X-23 and SBD from *Bacillus* sp. A2-5a CGTase [108], glucoamylase from *Saccharomyces cerevisiae* and SBD from *Aspergillus niger* glucoamylase [109], and barley α-amylase AMY1 and SBD from *Aspergillus niger* glucoamylase [110].

It is worth mentioning that although various types of SBDs have already been recognised and classified into seven different CBM families, most of the practical applications have been done with the 'classical' SBD of the CBM20 family.

Acknowlegements. This work was financially supported by the VEGA grant no. 2/5067/26 from the Slovak Grant Agency for Science. We thank Dr. Margaret D.T. Chang (National Tsing Hua University, Taiwan) and Dr. Bunzo Mikami (Kyoto University, Japan) for providing us with the co-ordinates of CBM21 from *Rhizopus oryzae* glucoamylase and CBM41 from *Klebsiella pneumoniae* pullulanase, respectively.

1 Svensson, B., Pedersen, T. G., Svendsen, I. B., Sakai, T. and Ottesen, M. (1982) Characterization of two forms of glucoamylase from *Aspergillus niger*. Carlsberg Res. Commun. 47, 55–69.
2 Flor, P. Q. and Hayashida, S. (1983) Production and characteristics of raw starch-digesting glucoamylase O from a protease-negative, glycosidase-negative *Aspergillus awamori var. kawachi* mutant. Appl. Environ. Microbiol. 45, 905–912.
3 Takahashi, T., Kato, K., Ikegami, Y. and Irie, M. (1985) Different behavior towards raw starch of three forms of glucoamylase from a *Rhizopus* sp. J. Biochem. 98, 663–671.
4 Tanaka, Y., Ashikari, T., Nakamura, N., Kiuchi, N., Shibano, Y., Amachi, T. and Yoshizumi, H. (1986) Comparison of amino acid sequences of three glucoamylases and their structure-function relationships. Agric. Biol. Chem. 50, 965–969.
5 Machovic, M., Svensson, B., MacGregor, E. A. and Janecek, S. (2005) A new clan of CBM families based on bioinformatics of starch-binding domains from families CBM20 and CBM21. FEBS J. 272, 5497–5513.
6 Rodriguez-Sanoja, R., Oviedo, N. and Sanchez, S. (2005) Microbial starch-binding domain. Curr. Opin. Microbiol. 8, 260–267.
7 Sogaard, M., Kadziola, A., Haser, R. and 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.
8 Tibbot, B. K., Wong, D. W. S. and 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.
9 Hostinova, E., Solovicova, A., Dvorsky, R. and Gasperik J (2003) Molecular cloning and 3D structure prediction of the first raw-starch-degrading glucoamylase without a separate starch-binding domain. Arch. Biochem. Biophys. 411, 189–195.
10 Gibson, R. M. and Svensson, B. (1987) Identification of tryptophanyl residues involved in binding of carbohydrate ligands to barley α-amylase 2. Carlsberg Res. Commun. 52, 373–379.
11 Bozonnet, S., Bønsager, B. C., Kramhøft, B., Mori, H., Abou Hachem, M., Willemoës, M., Jensen, M. T., Fukuda, K., Nielsen, P. K. et al. (2005) Binding of carbohydrates and protein inhibitors to the surface of α-amylases. Biologia, Bratislava 60 (Suppl. 16), 27–36.
12 Robert, X., Haser, R., Gottschalk, T. E., Ratajczak, F., Driguez, H., Svensson, B. and Aghajari, N. (2003) The structure of barley α-amylase isozyme 1 reveals a novel role of domain C in substrate recognition and binding: a pair of sugar tongs. Structure 11, 973–984.
13 Tranier, S., Deville, K., Robert, X., Bozonnet, S., Haser, R., Svensson, B. and Aghajari, N. (2005) Insights into the 'pair of sugar tongs' surface binding site in barley α-amylase isozymes and crystallization of appropriate sugar tongs mutants. Biologia, Bratislava 60 (Suppl. 16), 37–46.
14 Robert, X., Haser, R., Mori, H., Svensson, B. and Aghajari, N. (2005) Oligosaccharide binding to barley α-amylase 1. J. Biol. Chem. 280, 32968–32978.
15 Gasperik, J., Hostinova, E. and Sevcik, J. (2005) Acarbose binding at the surface of *Saccharomycopsis fibuligera* glu-

coamylase suggests the presence of a raw starch binding site. Biologia, Bratislava 60 (Suppl. 16), 167–170.
16 Sevcik, J., Hostinova, E., Solovicova, A., Gasperik, J., Dauter, Z. and Wilson, K.S. (2006) Structure of the complex of a yeast glucoamylase with acarbose reveals the presence of a raw starch binding site on the catalytic domain. FEBS J. 273, 2161–2171.
17 Svensson, B., Jespersen, H., Sierks, M. R. and MacGregor, E. A. (1989) Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem. J. 264, 309–311.
18 Ashikari, T., Nakamura, N., Tanaka, Y., Kiuchi, N., Shibano, Y., Tanaka, T., Amachi, T. and Yoshizumi, H. (1986) *Rhizopus* raw-starch-degrading glucoamylase: its cloning and expression in yeast. Agric. Biol. Chem. 50, 957–964.
19 Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochem. J. 280, 309–316.
20 Coutinho, P. M. and 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, pp. 15–23, Ohmiya, K., Hayashi, K., Sakka, K., Kobayashi, Y., Karita, S. and Kimura, T. (eds.), Uni Publishers Company, Tokyo.
21 Janecek, S. and Sevcik, J. (1999) The evolution of starch-binding domain. FEBS Lett. 456, 119–125.
22 Steyn, A. J., Marmur, J. and Pretorius, I. S. (1995) Cloning, sequence analysis and expression in yeasts of a cDNA containing a *Lipomyces kononenkoae* α-amylase-encoding gene. Gene 166, 65–71.
23 Bui, D. M., Kunze, I., Förster, S., Wartmann, T., Horstmann, C., Manteuffel, R. and Kunze, G. (1996) Cloning and expression of an *Arxula adeninivorans* glucoamylase gene in *Saccharomyces cerevisiae*. Appl. Microbiol. Biotechnol. 44, 610–619.
24 Sumitani, J., Tottori, T., Kawaguchi, T. and Arai, M. (2000) New type of starch-binding domain: the direct repeat motif in the C-terminal region of *Bacillus* sp. no. 195 α-amylase contributes to starch binding and raw starch degrading. Biochem. J. 350, 477–484.
25 Boraston, A. B., Bolam, D. N., Gilbert, H. J. and Davies, G. J. (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. 382, 769–781.
26 Sorimachi, K., Jacks, A. J., Le Gal-Coeffet, M. F., Williamson, G., Archer, D. B. and Williamson, M. P. (1996) Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy. J. Mol. Biol. 259, 970–987.
27 Sorimachi, K., Le Gal-Coeffet, M. F., Williamson, G., Archer, D. B. and Williamson, M. P. (1997) Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin. Structure 5, 647–661.
28 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) 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.
29 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) The raw starch binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. J. Biol. Chem. 271, 32777–32784.
30 Dauter, Z., Dauter, M., Brzozowski, A. M., Christensen, S., Borchert, T. V., Beier, L., Wilson, K. S. and 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.
31 Mikami, B., Adachi, M., Kage, T., Sarikaya, E., Nanmori, T., Shinke, R. and Utsumi, S. (1999) Structure of raw starch-digesting *Bacillus cereus* β-amylase complexed with maltose. Biochemistry 38, 7050–7061.
32 Oyama, T., Kusunoki, M., Kishimoto, Y., Takasaki, Y. and Nitta, Y. (1999) Crystal structure of β-amylase from *Bacillus cereus* var. *mycoides* at 2.2 Å resolution. J. Biochem. 125, 1120–1130.
33 MacGregor, E. A., Janecek, S. and Svensson, B. (2001) Relationship of sequence and structure to specificity in the α-amylase family of enzymes. Biochim. Biophys. Acta 1546, 1–20.
34 Janecek, S., Svensson, B. and MacGregor, E. A. (2003) Relation between domain evolution, specificity, and taxonomy of the alpha-amylase family members containing a C-terminal starch-binding domain. Eur. J. Biochem. 270, 635–645.
35 Mukai, K., Maruta, K., Satouchi, K., Kubota, M., Fukuda, S., Kurimoto, M. and Tsujisaka, Y. (2004) Cyclic tetrasaccharide-synthesizing enzymes from *Arthrobacter globiformis* A19. Biosci. Biotechnol. Biochem. 68, 2529–2540.
36 Baunsgaard, L., Lutken, H., Mikkelsen, R., Glaring, M. A., Pham, T. T. and Blennow, A. (2005) A novel isoform of glucan, water dikinase phosphorylates pre-phosphorylated α-glucans and is involved in starch degradation in *Arabidopsis*. Plant J. 41, 595–605.
37 Minassian, B. A., Ianzano, L., Meloche, M., Andermann, E., Rouleau, G. A., Delgado-Escueta, A. V. and Scherer SW (2000) Mutation spectrum and predicted function of laforin in Lafora's progressive myoclonus epilepsy. Neurology 55, 341–346.
38 Janecek, S. (2002) A motif of microbial starch-binding domain found in human genethonin. Bioinformatics 18, 1534–1537.
39 Southall, S. M., Simpson, P. J., Gilbert, H. J., Williamson, G. and Williamson, M. P. (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. FEBS Lett. 447, 58–60.
40 Kubota, M., Matsuura, Y., Sakai, S. and Katsube, Y. (1991) Molecular structure of *B. stearothermophilus* cyclodextrin glucanotransferase and analysis of substrate binding site. Denpun Kagaku 38, 141–146.
41 Klein, C. and Schulz, G. E. (1991) Structure of cyclodextrin glycosyltransferase refined at 2.0 Å resolution. J. Mol. Biol. 217, 737–750.
42 Harata, K., Haga, K., Nakamura, A., Aoyagi, M. and 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. D Biol. Crystallogr. 52, 1136–1145.
43 Knegtel, R. M., Wind, R. D., Rozeboom, H. J., Kalk, K. H., Buitelaar, R. M., Dijkhuizen, L. and 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.
44 Jacks, A. J., Sorimachi, K., Le Gal-Coëffet, M. F., Williamson, G., Archer, D. B. and Williamson, M. P. (1995) ^{1}H and ^{15}N assignment and secondary structure of the starch-binding domain of glucoamylase from *Aspergillus niger*. Eur. J. Biochem. 233, 568–578.
45 Williamson, M. P., Le Gal-Coëffet, M. F., Sorimachi, K., Furniss, C. S., Archer, D. B. and Williamson, G. (1997) Function of conserved tryptophans in the *Aspergillus niger* glucoamylase 1 starch binding domain. Biochemistry 36, 7535–7539.
46 Ye, Z., Miyake, H., Tatsumi, M., Nishimura, S. and Nitta, Y. (2004) Two additional carbohydrate-binding sites of β-amylase from *Bacillus cereus* var. *mycoides* are involved in hydrolysis and raw starch-binding. J. Biochem. 135, 355–363.
47 Yoon, H. J., Hirata, A., Adachi, M., Sekine, A., Utsumi, S. and Mikami, B. (1999) Structure of the starch-binding domain of

Bacillus cereus β-amylase. J. Microbiol. Biotechnol. 9, 619–623.
48 Giardina, T., Gunning, A. P., Juge, N., Faulds, C. B., Furniss, C. S., Svensson, B., Morris, V. J. and 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.
49 Goto, M., Seminaru, T., Fukukawa, K. and Hayashida, S. (1994) Analysis of the raw starch-binding domain by mutation of a glucoamylase from *Aspergillus awamori* var. *kawachi* expressed in *Saccharomyces cerevisiae*. Appl. Environ. Microbiol. 60, 3926–3930.
50 Chen, L., Coutinho, P. M., Nikolov, Z. and Ford, C. (1995) Deletion analysis of the starch-binding domain of *Aspergillus* glucoamylase. Protein Eng. 8, 1049–1055.
51 Lo, H. F., Lin, L. L., Chiang, W. Y., Chie, M. C., Hsu, W. H. and 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.
52 Lo, H. F., Chiang, W. Y., Chi, M. C., Hu, H. Y. and Lin, L. L. (2004) Site-directed mutagenesis of the conserved threonine, tryptophan, and lysine residues in the starch-binding domain of *Bacillus* sp. strain TS-23 α-amylase. Curr. Microbiol. 48, 280–284.
53 Takahashi, T., Tsuchida, Y. and Irie, M. (1982) Isolation of two inactive fragments of a *Rhizopus* sp. glucoamylase: relationship among three forms of the enzyme and the isolated fragments. J. Biochem. 92, 1623–1633.
54 Tanaka, Y., Ashikari, T., Nakamura, N., Kiuchi, N., Shibano, Y., Amachi, T. and Yoshizumi, H. (1986) Comparison of amino acid sequences of three glucoamylases and their structure-function relationships. Agric. Biol. Chem. 50, 965–969.
55 Kang, H. K., Lee, J. H., Kim, D., Day, D. F., Robyt, J. F., Park, K. H. and Moon, T. W. (2004) Cloning and expression of *Lipomyces starkeyi* α-amylase in *Escherichia coli* and determination of some of its properties. FEMS Microbiol. Lett. 233, 53–64.
56 Houghton-Larsen, J. and Pedersen, P. A. (2003) Cloning and characterisation of a glucoamylase gene (GlaM) from the dimorphic zygomycete *Mucor circinelloides*. Appl. Microbiol. Biotechnol. 62, 210–217.
57 Ramachandran, N., Pretorius, I. S. and Cordero Otero, R. R. (2005) Amylolytic enzymes from the yeast *Lipomyces kononenkoae*. Biologia, Bratislava 60 (Suppl. 16), 103–110.
58 Ceulemans, H., Stalmans, W. and Bollen, M. (2002) Regulator-driven functional diversification of protein phosphatase-1 in eukaryotic evolution. BioEssays 24, 371–381.
59 Bork, P., Dandekar, T., Eisenhaber, F. and Huynen, M. (1998) Characterization of targeting domains by sequence analysis: glycogen-binding domains in protein phosphatases. J. Mol. Med. 76, 77–79.
60 Chou, W. I., Pai, T. W., Liu, S. H., Hsiung, B. K. and Chang, M. D. (2006) The family 21 carbohydrate binding module of glucoamylase from *Rhizopus oryzae* consists of two sites playing distinct roles in ligand binding. Biochem. J. 396, 469–477.
61 Svensson, B., Larsen, K., Svendsen, I. and Boel, E. (1983) The complete amino acid sequence of the glycoprotein, glucoamylase G1, from *Aspergillus niger*. Carlsberg Res. Commun. 48, 529–544.
62 Nitschke, L., Heeger, K., Bender, H. and Schulz, G. E. (1990) Molecular cloning, nucleotide sequence and expression in *Escherichia coli* of the β-cyclodextrin glycosyltransferase gene from *Bacillus circulans* strain no. 8. Appl. Microbiol. Biotechnol. 33, 542–546.
63 Henrissat, B. and Bairoch, A. (1996) Updating the sequence-based classification of glycosyl hydrolases. Biochem. J. 316, 695–696.
64 Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A. and Law, V. (2006) A structural and functional analysis of α-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. J. Biol. Chem. 281, 587–598.
65 Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S. and Horikoshi, K. (2000) Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. Nucleic Acids Res. 28, 4317–4331.
66 Siggens, K. W. (1987) Molecular cloning and characterization of the β-amylase gene from *Bacillus circulans*. Mol. Microbiol. 1, 86–91.
67 Santiago, M., Linares L, Sanchez, S. and Rodriguez-Sanoja, R. (2005) Functional characteristics of the starch-binding domain of *Lactobacillus amylovorus* α-amylase. Biologia, Bratislava 60 (Suppl. 16), 111–114.
68 Satoh, E., Uchimura, T., Kudo, T. and Komagata, K. (1997) Purification, characterization, and nucleotide sequence of an intracellular maltotriose-producing α-amylase from *Streptococcus bovis* 148. Appl. Environ. Microbiol. 63, 4941–4944.
69 Giraud, E. and Cuny, G. (1997) Molecular characterization of the α-amylase genes of *Lactobacillus plantarum* A6 and *Lactobacillus amylovorus* reveals an unusual 3′ end structure with direct tandem repeats and suggests a common evolutionary origin. Gene 198, 149–157.
70 Morlon-Guyot, J., Mucciolo-Roux, F., Rodriguez-Sanoja, R. and Guyot, J. P. (2001) Characterization of the *L. manihotivorans* α-amylase gene. DNA Seq. 12, 27–37.
71 Rodriguez-Sanoja, R., Ruiz, B., Guyot, J. P. and Sanchez, S. (2005) Starch-binding domain affects catalysis in two *Lactobacillus* α-amylases. Appl. Environ. Microbiol. 71, 297–302.
72 Park, K. H., Kim, T. J., Cheong, T. K., Kim, J. W., Oh, B. H. and Svensson, B. (2000) Structure, specificity and function of cyclomaltodextrinase, a multispecific enzyme of the α-amylase family. Biochim. Biophys. Acta 1478, 165–185.
73 Oslancova, A. and Janecek, S. (2002) Oligo-1,6-glucosidase and neopullulanase enzyme subfamilies from the α-amylase family defined by the fifth conserved sequence region. Cell. Mol. Life Sci. 59, 1945–1959.
74 Abe, A., Tonozuka, T., Sakano, Y. and Kamitori, S. (2004) Complex structures of *Thermoactinomyces vulgaris* R-47 α-amylase 1 with maltooligosaccharides demonstrate the role of domain N acting as a starch-binding domain. J. Mol. Biol. 335, 811–822.
75 Tonozuka, T., Mogi, S., Shimura, Y., Ibuka, A., Sakai, H., Matsuzawa, H., Sakano, Y. and Ohta, T. (1995) Comparison of primary structures and substrate specificities of two pullulan-hydrolyzing α-amylases, TVA I and TVA II, from *Thermoactinomyces vulgaris* R-47. Biochim. Biophys. Acta 1252, 35–42.
76 Tonozuka, T., Yokota, T., Ichikawa, K., Mizuno, M., Kondo, S., Nishikawa, A., Kamitori, S. and Sakano, Y. (2002) Crystal structures and substrate specificities of two α-amylases hydrolyzing cyclodextrins and pullulan from *Thermoactinomyces vulgaris* R-47. Biologia, Bratislava 57 (Suppl. 11), 71–76.
77 Abe, A., Yoshida, H., Tonozuka, T., Sakano, Y. and Kamitori, S. (2005) Complexes of *Thermoactinomyces vulgaris* R-47 α-amylase 1 and pullulan model oligossacharides provide new insight into the mechanism for recognizing substrates with α-(1,6) glycosidic linkages. FEBS J. 272, 6145–6153.
78 Kamitori, S., Abe, A., Ohtaki, A., Kaji, A., Tonozuka, T. and Sakano, Y. (2002) Crystal structures and structural comparison of *Thermoactinomyces vulgaris* R-47 α-amylase 1 (TVAI) at 1.6 Å resolution and α-amylase 2 (TVAII) at 2.3 Å resolution. J. Mol. Biol. 318, 443–453.
79 Lammerts van Bueren, A., Finn, R., Ausio, J. and Boraston, A. B. (2004) α-Glucan recognition by a new family of carbo-

hydrate-binding modules found primarily in bacterial pathogens. Biochemistry 43, 15633–15642.
80 Bibel, M., Brettl, C., Gosslar, U., Kriegshaeuser, G. and Liebl, W. (1998) Isolation and analysis of genes for amylolytic enzymes of the hyperthermophilic bacterium *Thermotoga maritima*. FEMS Microbiol. Lett. 158, 9–15.
81 Mikami, B., Iwamoto, H., Malle, D., Yoon, H. J., Demirkan-Sarikaya, E., Mezaki, Y. and Katsuya, Y. (2006) Crystal structure of pullulanase: evidence for parallel binding of oligosaccharides in the active site. J. Mol. Biol. 359, 690–707.
82 Katsuragi, N., Takizawa, N. and Murooka Y (1987) Entire nucleotide sequence of the pullulanase gene of *Klebsiella aerogenes* W70. J. Bacteriol. 169, 2301–2306.
83 Mikkelsen, R., Suszkiewicz, K. and Blennow, A. (2006) A novel type carbohydrate-binding module identified in α-glucan, water dikinases is specific for regulated plastidial starch metabolism. Biochemistry 45, 4674–4682.
84 Kerk, D., Bulgrien, J., Smith, D. W., Barsam, B., Veretnik, S. and Gribskov, M. (2002) The complement of protein phosphatase catalytic subunits encoded in the genome of *Arabidopsis*. Plant Physiol. 129, 908–925.
85 Niittyla, T., Comparot-Moss, S., Lue, W. L., Messerli, G., Trevisan, M., Seymour, M. D., Gatehouse, J. A., Villadsen, D., Smith, S. M., Chen, J., Zeeman, S. C. and Smith, A. M. (2006) Similar protein phosphatases control starch metabolism in plants and glycogen metabolism in mammals. J. Biol. Chem. 281, 11815–11818.
86 Fordham-Skelton, A. P., Chilley, P., Lumbreras, V., Reignoux, S., Fenton, T. R., Dahm, C. C., Pages, M. and Gatehouse, J. A. (2002) A novel higher plant protein tyrosine phosphatase interacts with SNF1-related protein kinases via a KIS (kinase interaction sequence) domain. Plant J. 29, 705–715.
87 Kerk, D., Conley, T. R., Rodriguez, F. A., Tran, H. T., Nimick, M., Muench, D. G. and Moorhead, G. B. (2006) A chloroplast-localized dual-specificity protein phosphatase in *Arabidopsis* contains a phylogenetically dispersed and ancient carbohydrate-binding domain, which binds the polysaccharide starch. Plant J. 46, 400–413.
88 Lohi, H. T. and Minassian, B. A. (2005) Starch-like polyglucosan formation in neuronal dendrites in the Lafora form of human epilepsy: a theory of pathogenesis. Biologia, Bratislava 60 (Suppl. 16), 123–129.
89 Minassian, B. A., Lee, J. R., Herbrick, J. A., Huizenga, J., Soder, S., Mungall, A. J., Dunham, I., Gardner, R., Fong C. Y., Carpenter, S. et al. (1998) Mutations in a gene encoding a novel protein tyrosine phosphatase cause progressive myoclonus epilepsy. Nat. Genet. 20, 171–174.
90 Wang, J., Stuckey, J. A., Wishart, M. J. and Dixon, J. E. (2002) A unique carbohydrate binding domain targets the lafora disease phosphatase to glycogen. J. Biol. Chem. 277, 2377–2380.
91 Chan, E. M., Ackerley, C. A., Lohi, H., Ianzano, L., Cortez, M. A., Shannon, P., Scherer, S. W. and Minassian, B.A. (2004) Laforin preferentially binds the neurotoxic starch-like polyglucosans, which form in its absence in progressive myoclonus epilepsy. Hum. Mol. Genet. 13, 1117–1129.
92 Ganesh, S., Tsurutani, N., Suzuki, T., Hoshii, Y., Ishihara, T., Delgado-Escueta, A. V. and Yamakawa, K. (2004) The carbohydrate-binding domain of Lafora disease protein targets Lafora polyglucosan bodies. Biochem. Biophys. Res. Commun. 313, 1101–1109.
93 Wang, W. and Roach, P. J. (2004) Glycogen and related polysaccharides inhibit the laforin dual-specificity protein phosphatase. Biochem. Biophys. Res. Commun. 325, 726–730.
94 Bouju, S., Lignon, M. F., Pietu, G., Le Cunff, M., Leger, J. J., Auffray, C. and Dechesne, S. A. (1998) Molecular cloning and functional expression of a novel human gene encoding two 41–43 kDa skeletal muscle internal membrane proteins. Biochem. J. 335, 549–556.
95 Manning, G., Whyte, D. B., Martinez, R., Hunter, T. and Sudarsanam, S. (2002) The protein kinase complement of the human genome. Science 298, 1912–1934.
96 Hudson, E. R., Pan, D. A., James, J., Lucocq, J. M., Hawley, S. A., Green, K. A., Baba, O., Terashima, T. and Hardie, D. G. (2003) A novel domain in AMP-activated protein kinase causes glycogen storage bodies similar to those seen in hereditary cardiac arrhythmias. Curr. Biol. 13, 861–866.
97 Polekhina, G., Gupta, A., Michell, B. J., van Denderen, B., Murthy, S., Feil, S. C., Jennings, I. G., Campbell, D. J., Witters, L. A., Parker, M. W., Kemp, B. E. and Stapleton, D. (2003) AMPK β subunit targets metabolic stress sensing to glycogen. Curr. Biol. 13, 867–871.
98 Katsuya, Y., Mezaki, Y., Kubota, M. and Matsuura, Y. (1998) Three-dimensional structure of *Pseudomonas* isoamylase at 2.2 Å resolution. J. Mol. Biol. 281, 885–897.
99 Abad, M. C., Binderup, K., Rios-Steiner, J., Arni, R. K., Preiss, J. and Geiger, J. H. (2002) The X-ray crystallographic structure of *Escherichia coli* branching enzyme. J. Biol. Chem. 277, 42164–42170.
100 Polekhina, G., Gupta, A., van Denderen, B. J., Feil, S. C., Kemp, B. E., Stapleton, D. and Parker, M. W. (2005) Structural basis for glycogen recognition by AMP-activated protein kinase. Structure 13, 1453–1462.
101 Stapleton, D., Gao, G., Michell, B. J., Widmer, J., Mitchelhill, K., Teh, T., House, C. M., Witters, L. A. and Kemp, B. E. (1994) Mammalian 5′-AMP-activated protein kinase non-catalytic subunits are homologs of proteins that interact with yeast Snf1 protein kinase. J. Biol. Chem. 269, 29343–29346.
102 Dalmia, B. K. and Nikolov, Z. L. (1994) Characterization of a β-galactosidase fusion protein containing the starch-binding domain of *Aspergillus* glucoamylase. Enzyme Microb. Technol. 16, 18–23.
103 Ji, Q., Vicken, J. P., Suurs, L. C. J. M. and Visser, R. G. F. (2003) Microbial starch-binding domain as a tool for targeting proteins to granule during starch biosynthesis. Plant Mol. Biol. 51, 789–801.
104 Ji, Q., Oomen, R. J. F. J., Vicken, J. P., Bolam, N. D., Gilbert, H. J., Suurs, L. C. J. M. and Visser, R. G. F. (2004) Reduction of starch granule size by expression of an engineered tandem starch-binding domain in potato plants. Plant Biotechnol. J. 2, 251–260.
105 Crittenden, R., Laitila, A., Forssell, P., Matto, J., Saarela, M., Mattila-Sandholm, T. and Myllarinen, P. (2001) Adhesion of *Bifidobacteria* to granular starch and its implications in probiotic technologies. Appl. Environ. Microbiol. 67, 3469–3475.
106 Levy, I., Paldi, T. and Shoseyov, O. (2004) Engineering a bifunctional starch-cellulose cross-bridge protein. Biomaterials 25, 1841–1849.
107 Hua, Y. W., Chi, M. CH., Lo, H. F., Hsu, W. H. and Lin, L. L. (2004) Fusion of *Bacillus stearothermophilus* leucine aminopeptidase II with the raw-starch-binding domain of *Bacillus* sp. strain TS-23 α-amylase generates a chimeric enzyme with enhanced thermostability and catalytic activity. J. Ind. Microbiol. Biotechnol. 31, 273–277.
108 Ohdan, K., Kuriki, T., Takata, H., Kaneko, H. and 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.
109 Latorre-Garcia, L., Adam, A. C., Manzanares, P. and Polaina, J. (2005) Improving the amylolytic activity of *Saccharomyces cerevisiae* glucoamylase by the addition of a starch binding domain. J. Biotechnol. 118, 167–176.
110 Juge, N., Nohr, J., Le Gal-Coeffet, M. F., Kramhoft, B., Furniss, C. S. M., Planchot, V., Archer, D. B., Williamson, G. and Svensson, B. (2006) The activity of barley α-amylase

on starch granules is enhanced by fusion of a starch binding domain from *Aspergillus niger* glucoamylase. Biochim. Biophys. Acta 1764, 275–284.

111 Bairoch, A., Apweiler, R., Wu, C. H., Barker, W. C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M. et al. (2005) The Universal Protein Resource (UniProt). Nucleic Acids Res. 33 (Database issue), D154–D159.

112 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242.

113 DeLano, W. L. (2006) The PyMOL Molecular Graphics System. DeLano Scientific LLC, San Carlos, CA, USA. http://www.pymol.org.

114 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22, 4673–4680.

115 Page, R. D. (1996) TreeView: an application to display phylogenetic trees on personal computers. Comput. Applic. Biosci. 12, 357–358.