VERSITA

Biologia **69**/9: 1087—1096, 2014
Section Cellular and Molecular Biology
DOI: 10.2478/s11756-014-0415-3

Review

Two structurally related starch-binding domain families CBM25 and CBM26*

Katarína Majzlová[1] & Štefan Janeček[1,2]**

[1] *Laboratory of Protein Evolution, Institute of Molecular Biology, Slovak Academy of Sciences, SK-84551 Bratislava, Slovakia; e-mail: Stefan.Janecek@savba.sk*
[2] *Department of Biology, Faculty of Natural Sciences, University of SS Cyril and Methodius, SK-91701 Trnava, Slovakia*

Abstract: Starch binding domains (SBDs) are able to bind to and facilitate the degradation of raw starch and starchy substrates. In general, in the CAZy database they have been classified among the carbohydrate-binding module (CBM) families. The two families CBM25 and CBM26 together with families CBM20, 21, 34, 41, 45, 48, 53, 58, 68 and 69 belong to twelve SBD CAZy families. They represent a group of closely related modules exhibiting some sequence similarity, although each of the two families possesses its own features. Both CBM25 and CBM26 adopt a typical SBD fold of distorted β-barrel as recognized in the modules present in the maltohexaose-producing amylase from *Bacillus halodurans*. With regard to catalytic domains, most members are α-amylases and maltooligosaccharide-producing amylases from the α-amylase glycoside hydrolase (GH) family GH13, but also some β-amylases (GH14) and hypothetical proteins (e.g. from the family GH31) are known. The main goal of this review was to compare the available amino acid sequences of SBDs from both families CBM25 and CBM26 and to reveal, if possible, SBD(s) with the character "intermediary" between the CBM25 and CBM26. Emphasis was also given on a structural comparison of the identified intermediary SBD with the CBM25 and CBM26 representatives and a detailed evolutionary division of both CBM families that can be utilized for defining the future subfamilies.

Key words: starch-binding domain; families CBM25 and CBM26; alpha-amylase; amylolytic enzymes; carbohydrate-binding module; sequence comparison; evolutionary relatedness.

Abbreviations: CAZy, Carbohydrate-Active enzymes; CBM, carbohydrate-binding module; GH, glycoside hydrolase; PDB, Protein Data Bank; SBD, starch-binding domain.

Introduction

A starch-binding domain (SBD) has been proposed as a continuous segment of approximately 100 amino acid residues present in various amylases and related starch hydrolases containing a few consensus residues (Svensson et al. 1989). Its main function is to help the catalytic domain to bind and degrade raw, i.e non-gelatinized starch (Southal et al. 1999). The SBD has been found in roughly 10% of amylolytic enzymes (Janecek & Sevcik 1999). Of the consensus residues, especially the aromatic tryptophans and tyrosines (or phenylalanines) have later been confirmed experimentally to be responsible for binding the various carbohydrates (Penninga et al. 1996; Sorimachi et al. 1997; Abe et al. 2004; Boraston et al. 2006; van Bueren & Boraston 2007; Tung et al. 2008; Rodriguez-Sanoja et al. 2009; Koropatkin & Smith 2010; Wayllace et al. 2010). Originally, only two types of SBD were known, i.e. a more frequent usually C-terminally positioned domain (Janecek et al. 2003) and a rather rare typically N-terminal one (Ashikari et al. 1986). When the carbohydrate-binding module (CBM) classification was introduced in the Carbohydrate-Active enzymes (CAZy) database (http://www.cazy.org/; Cantarel et al. 2009), the former SBD type has been defined as the family CBM20, whereas the latter one as the family CBM21. Currently (Lombard et al. 2014), SBDs – involving also glycogen-binding domains – are present in 12 CBM families: in CBM25, 26, 34, 41, 45, 48, 53, 58, 68 and 69, in addition to CBM20 and CBM21 mentioned above (Janecek et al. 2011; Peng et al. 2014; Xu et al. 2014). An insightful analysis of surface-binding sites in amylolytic enzymes that are not part of distinct CBMs but interact with carbohydrates has been published recently (Cockburn et al. 2014).

The present review concerns the pair of closely related SBDs, the families CBM25 and CBM26. The family CBM25 was based on a discovery of two copies of, at that time, novel-type SBD positioned C-terminally in

* Electronic supplementary material. The online version of this article (DOI:10.2478/s11756-014-0415-3) contains supplementary material, which is available to authorized users.
** Corresponding author

Springer

the raw-starch degrading GH13 α-amylase from *Bacillus* sp. No. 195 (Sumitani et al. 2000), whereas the family CBM26 was established as a family of SBDs present in 4-5 tandem repeats of the raw-starch degrading GH13 α-amylases from lactobacilli (Rodriguez-Sanoja et al. 2000, 2005b). The modules from both families CBM25 and CBM26 are associated mostly with amylolytic enzymes from the main α-amylase family GH13 (α-amylases and maltooligosaccharide-producing amylases dominate among them), but there are CBM25 and CBM26 examples found in the family GH14 β-amylases, hypothetical proteins from the α-glucosidase family GH31 and even the α-amylase from the family GH119 (Lombard et al. 2014). Thus the increasing spectrum of amylolytic enzyme specificities (i.e. catalytic domains) possessing in their protein molecule either a CBM25 or a CBM26 SBD (or even both of them) is of special interest.

Although the members of both families CBM25 and CBM26 possess in general the ability to bind to starch, they may differ from each other by their contributions to overall efficiency of the amylolytic enzyme as well as its substrate preference even if the SBD is a part of the α-amylase specificity. Thus, for example, in the case of *Bacillus* sp. No. 195 α-amylase containing two copies of CBM25 at its C-terminus (Fig. 1; protein No. 2) the digestion rate for raw starches decreased in the order for rice, maize, wheat, sweet potato and potato (Sumitani et al. 2000). A halophilic α-amylase from *Kocuria varians* having a homologous domain arrangement (i.e. two C-terminal CBM25) exhibited a higher percentage of digestion for wheat raw starch in comparison with maize raw starch (Yamaguchi et al. 2011). With regard to the maltohexaose-producing amylase from *Bacillus halodurans* with both CBM26 and CBM25 (present in the sequence in that order; see Figure 1; protein No. 1), each SBD was shown to be able to bind granular starch individually, however, the affinity for the both CBM together was approximately 50 times higher (Boraston et al. 2006). In addition, the CBM25 revealed a strong affinity dependence for a ligand length up to 7 glucose units, whereas the CBM26 affinity was observed to be maximum for 5 glucose units (Boraston et al. 2006). For the two SBD copies from the family CBM26 positioned C-terminally in the maltotriose-forming amylase from *Streptococcus bovis* (Fig. 1; protein No. 10), it was demonstrated that they are important not only for adsorption onto raw starch but also for enzymatic properties, i.e. efficient hydrolysis of raw starch (Matsui et al. 2007). Concerning the CBM26 SBDs from lactobacilli (Fig. 1; proteins No. 7 and 8), the α-amylase from *Lactobacillus amylovorus* with 5 copies was found to be roughly 10 times efficient in hydrolysis of various starches in comparison with the counterpart from *Lactobacillus plantarum* possessing 4 copies (Rodriguez-Sanoja et al. 2005b). Moreover in optimizing the binding to raw starch each CBM26 unit of the five SBDs present in the *L. amylovorus* α-amylase may act in an additive or synergic way (Guillen et al. 2007).

Some others of the above-mentioned twelve CBM families of SBDs were revealed to may share a closer evolutionary history, *e.g.*, CBM20 and CBM21 (Machovic et al. 2005) together with CBM48 and CBM53 (Machovic & Janecek 2006a; Christiansen et al. 2009). Although typical SBDs with their main role to enable the amylolytic enzyme to cope with raw starch are associated mainly with microbial amylases (Rodriguez-Sanoja et al. 2005a; Machovic & Janecek 2006b), in plants and animals, remarkably, their homologues have adopted various related functions, e.g. in phosphoglucan, water dikinase-3 (Mikkelsen et al. 2006; Glaring et al. 2011, Orzechowski et al. 2013), starch-excess protein-4 (Vander Kooi et al. 2010), laforin (Minassian et al. 2000; Gentry et al. 2013), genethonin-1 (Janecek 2002; Jiang et al. 2010) and β-subunit of AMP-activated protein kinase (Hudson et al. 2003; Polekhina et al. 2003, 2005). Nevertheless, all these non-microbial SBDs (or glycogen-binding domains) have always something to do with α-glucans (Janecek et al. 2011).

As far as the families CBM25 and CBM26 are concerned, currently they consist of almost 160 and more than 140 members from Bacteria, respectively (Lombard et al. 2014). Both families adopt a typical SBD fold (Boraston et al. 2004; Guillen et al. 2010) of distorted β-barrel (a β-sandwich) formed by 9 or 10 β-strands, as shown for solved three-dimensional structure of both modules present in the family GH13 maltohexaose-producing amylase from *Bacillus halodurans* (Boraston et al. 2006).

The main goal of this mini-review was to shed more light on mutual evolutionary relationships of SBDs from the two closely related families CBM25 and CBM26. Based on a comparison of available primary structures, emphasis was given on indicating the sequence features that are common for both families as well as those that are unique for a respective family. An attention was focused on indicating, if possible, the SBD(s) with a sequence-structural character intermediary between the two families as well as the features necessary to be present in real raw starch CBM25 and CBM26 binders.

Occurrence and sequence comparison

The present review delivers a comparison of amino acid sequences of 113 SBDs from families CBM25 (53) and CBM26 (60) originating from 57 various amylolytic enzymes and some hypothetical proteins not included in any GH family (Table 1). All details concerning each CBM copy used in this study can be found in Table S1. It is of interest that no CBM26 has been identified in any archaeal and eukaryotic enzyme (protein), i.e. it seems that – at least in the lights of our current knowledge – SBDs from the family CBM26 are only of bacterial origin. Moreover, almost exclusively they are present in amylolytic enzyme from the main α-amylase family GH13 (Table S1). With regard to the family CBM25, in addition to dominating bacterial enzymes from the α-amylase family GH13 (Janecek et al. 2014),

Table 1. Summary of proteins with CBM25 and CBM26 used in the present study.[a]

Enzyme/protein	EC	GH family	CBM25	CBM26
α-Amylase	3.2.1.1	GH13	11	18
Maltotriose-forming amylase	3.2.1.116	GH13	–	2
Maltopentaose-forming amylase	3.2.1.-	GH13	3	1
Maltohexaose-forming amylase	3.2.1.98	GH13	2	2
α-Amylase-pullulanase	3.2.1.1/3.2.1.41	GH13	1	–
α-1,6-Cyclomaltopentaose-forming glucanotransferase	2.4.1.-	GH13	2	–
Hypothetical protein		GH13	24	32
α-Amylase	3.2.1.1	GH119	2	–
Hypothetical protein		GH31	2	2
β-Amylase	3.2.1.2	GH14	1	–
Multidomain β-amylase/α-amylase	3.2.1.2/3.2.1.1	GH14/GH13	2	–
Hypothetical protein			3	3
Total No. of enzymes and proteins[b]			32	30
Total No. of CBMs			53	60

[a] Sequences (for details, see Table S1) were collected based on the information in the CAZy database for families CBM25 and CBM26 (Cantarel et al. 2009), and completed by the BLAST (Altschul et al. 1990) search using as queries the modules from *Bacillus halodurans* maltohexaose-producing amylase (UniProt: Q9KFR4; CBM25 and CBM26), *Bacillus* sp. No. 195 α-amylase (UniProt: O24781; CBM25 copy 1), *Lactobacillus amylovorus* α-amylase (UniProt: Q48502; CBM26 copy 1) and hypothetical family GH31 protein (UniProt: Q97F62; CBM26 both copies). The sequences were selected in an effort to include mainly all the CBM25 and CBM26 members possessing tryptophans to be essential for binding as revealed in three-dimensional structures of SBDs from both families in the *Bacillus halodurans* maltohexaose-producing amylase (Boraston et al. 2006); although some examples lacking one or even both functional tryptophans were also included for comparison. The above set of total 57 proteins (mostly amylolytic enzymes from GH13 and GH14 as well as putative amylases) yielding 113 CBM25s and CBM26s was obtained after several rounds of analysis from ~330 BLAST results in combination with CAZy server.

[b] The total number of proteins was 57 (i.e. lower than 62) since some proteins possess both CBM25 and CBM26.

there are examples of other amylolytic families, such as the α-amylase family GH119 (Watanabe et al. 2006; Janecek & Kuchtova 2012) and β-amylase family GH14 (Siggens 1987). Interestingly, the representative of related α-glucosidase family GH31 originates from a eukaryotic fungus responsible for potato late blight (Haas et al. 2009). The Eucarya are completed by two non-amylolytic hypothetical proteins from *Physcomitrella patens* subsp. patens (moss) and *Volvox carteri* (green alga) and the CBM25 was found also in one more non-amylolytic hypothetical protein from *Arthrobacter chlorophenolicus* designated as polysaccharide deacetylase (Table S1).

Concerning the domain arrangement (Fig. 1), both CBM25 and CBM26 are positioned typically at the C-terminus of a protein or succeeding the catalytic domain (Machovic & Janecek 2006b). They may be present as a single SBD (Siggens 1987), but also as multiple copies (Sumitani et al. 2000; Yamaguchi et al. 2011) either connected via linkers or found as repeated units without any linker (Giraud & Cuny 1997; Rodriguez-Sanoja et al. 2005b; Guillen et al. 2007). Within an amylolytic enzyme, they may exist in various alternatives, such as a copy (or copies) of one CBM family only (i.e. either CBM25 or CBM26), both CBM25 and CBM26 copies together (Boraston et al. 2006) and even with additional CBM, like CBM41 and CBM48 (Ryan et al. 2006). Note that a common occurrence with a CBM not considered to be an SBD (i.e. for example a xylan or cellulose binding domain) was not identified (Fig. 1; Table S1).

The amino acid sequence alignment of SBDs from both families CBM25 and CBM26 (Fig. 2) support unambiguously their close relatedness indicated originally by Boraston et al. (2006) for the two modules from the family GH13 maltohexaose-producing amylase from *Bacillus halodurans*. Both families share two of the three aromatic positions essential for each family: (i) the first one corresponding with Trp34 and Trp35 (numbering for the isolated SBDs) of CBM25 and CBM26, respectively, involved in stacking interactions with glucose moieties; and (ii) the second one corresponding with His26 (CBM25) and Tyr22 (CBM26). The third aromatic position, i.e. Trp74 (CBM25) and Tyr24 (CBM26), has no conserved equivalent in the respective family (Fig. 2). With regard to other residues indicated as functionally important for the above-mentioned CBM25 and CBM26 SBDs from maltohexaose-producing amylase from *B. halodurans* (Boraston et al. 2006) – Asp75, Asn76 and Asp81 for CBM25 and Gln70, Gly75 and Glu76 for CBM26 – it is worth mentioning that while the positions of Asp75 (CBM25) and Gln70 (CBM26) are highly conserved, it is not the case of the remaining pairs (Asn76 and Asp81 in CBM25 and Gly75 and Glu76 in CBM26), i.e. they may be important only specifically for some of the CBMs (Fig. S1). This means that each individual SBD from both families may contain its own residues helping the aromatic residues responsible for direct binding (Rodriguez-Sanoja et al. 2009; Yamaguchi et al. 2012a,b). However, lack of the aromatic residues in the essential positions, e.g., in both CBM25 copies of a GH13 hypothetical protein from *Bifidobacterium adolescentis* (UniProt Acc. No.: A7A7M5) may indicate that those CBMs might not be real SBDs able to bind to and facilitate degradation of raw starch (Christiansen et al. 2009; Janecek et al. 2011).

Evolutionary relationships

Although the two SBD families CBM25 and CBM26

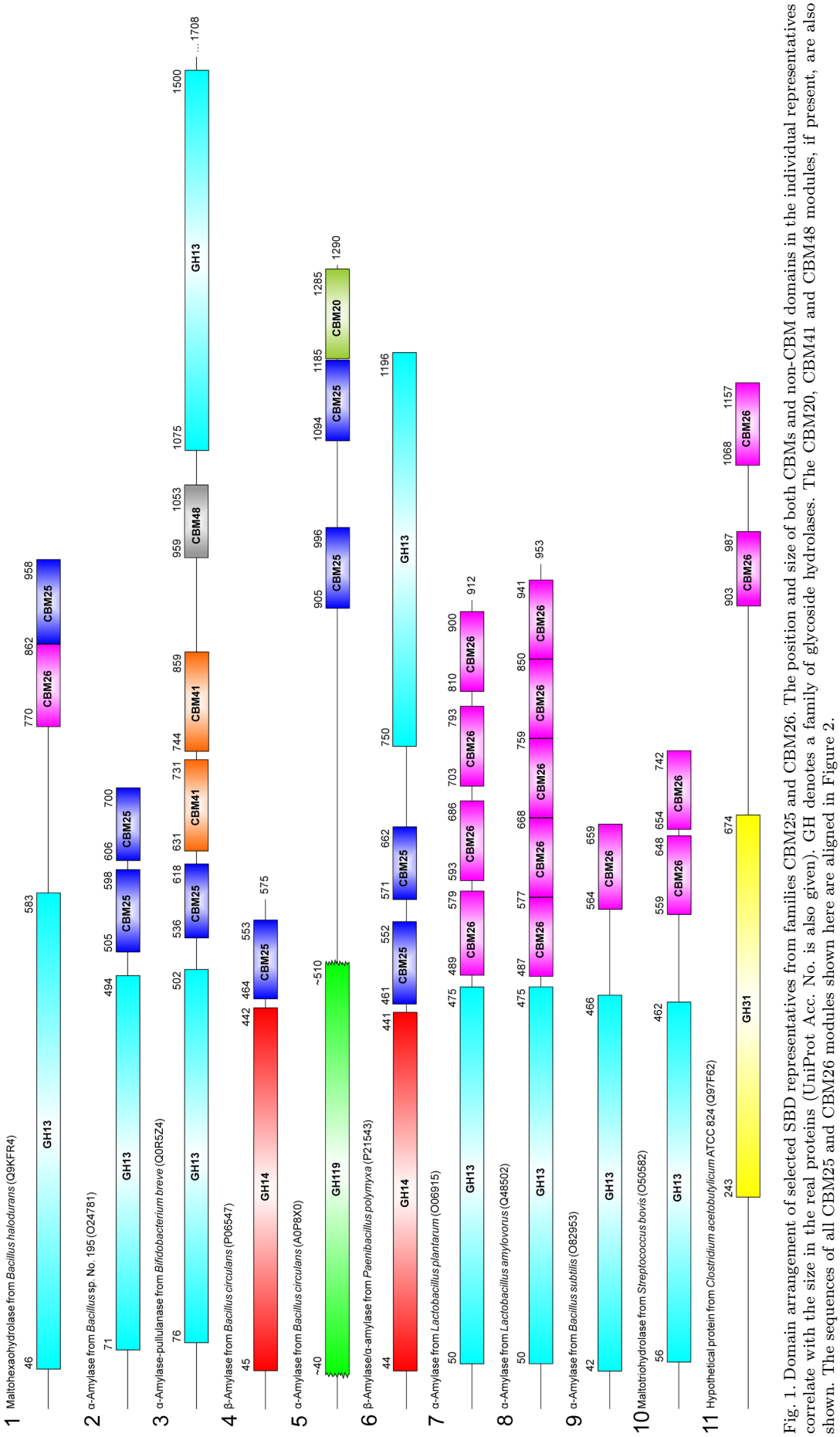

Fig. 1. Domain arrangement of selected SBD representatives from families CBM25 and CBM26. The position and size of both CBMs and non-CBM domains in the individual representatives correlate with the size in the real proteins (UniProt Acc. No. is also given). GH denotes a family of glycoside hydrolases. The CBM20, CBM41 and CBM48 modules, if present, are also shown. The sequences of all CBM25 and CBM26 modules shown here are aligned in Figure 2.

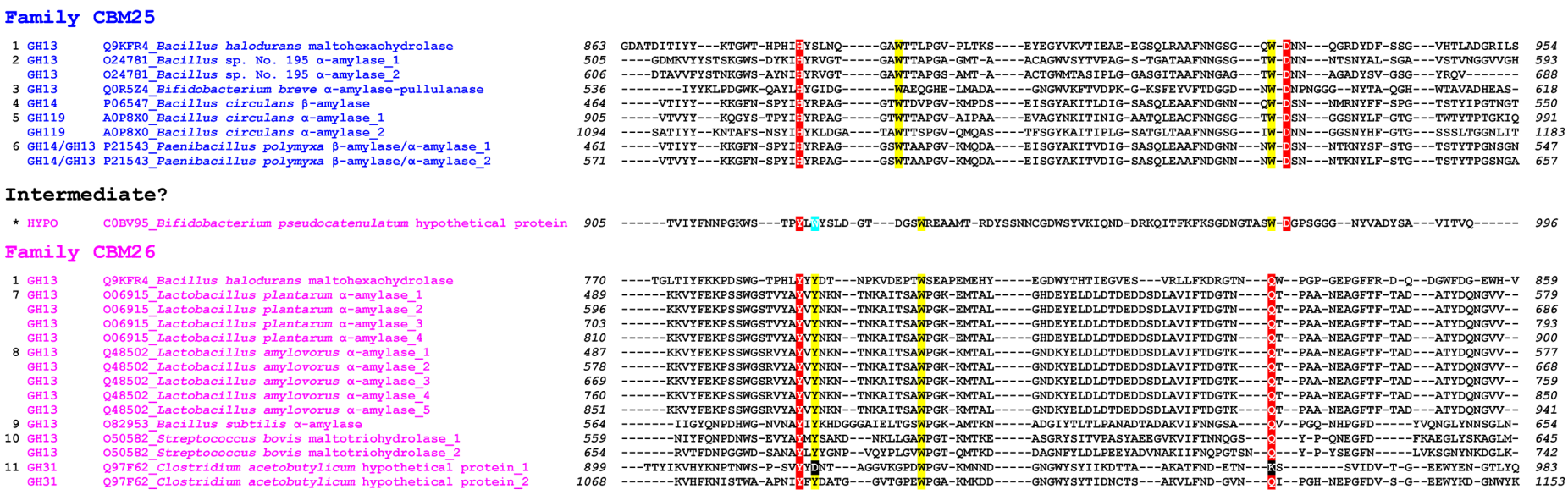

Fig. 2. Sequence alignment of SBD representatives from families CBM25 (blue) and CBM26 (magenta). The selection of the amylolytic enzymes and hypothetical proteins possessing the studied CBM modules corresponds with that shown in Figure 1. The residues important for carbohydrate binding in CBM25 and CBM26 modules of *Bacillus halodurans* GH13 maltohexaose-producing amylase (Boraston et al. 2006) are indicated by yellow (stacking interactions) and red highlighting. Note that the numbering for the isolated CBMs differs from the numbers shown at the beginning and the end of the aligned modules, i.e. the highlighted residues are as follows: His26, Trp34, Trp74 and Asp75 (*Bacillus halodurans* maltohexaohydrolase CBM25), Tyr15, Trp17, Trp27, Trp71 and Trp72 (*B pseudocatenulatum* hypothetical protein "intermediary" CBM26 SBD) and Tyr22, Tyr24, Trp35 and Gln70 (*Bacillus halodurans* maltohexaohydrolase CBM26). Substitutions in carbohydrate-binding positions are indicated by cyan and black highlighting for keeping and lost of aromatic character, respectively, in the position. More copies from the same CBM family in the same protein are distinguished from each other by a digit succeeding the UniProt accession number. The borders of SBDs in both families CBM25 and CBM26 (Table S1) were determined according to modules in *Lactobacillus* α-amylases (Rodriguez-Sanoja et at. 2005b, 2009) and *Bacillus halodurans* maltohexaose-producing amylase (Boraston et at. 2006) with support gained from the BLAST (Altschul et al. 1990) results. The alignment of all 113 sequences (Fig. S1) of CBM25 (53) and CBM26 (60) retrieved from the UniProt knowledge database (UniProt Consortium 2013) was performed using the program Clustal-W2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/; Larkin et al. 2007). In order to maximize similarities and reduce the background noise, the alignment was manually tuned taking into account mainly the previous structural (Boraston et al. 2006) and site-directed mutagenesis (Rodriguez-Sanoja et al. 2009) studies.

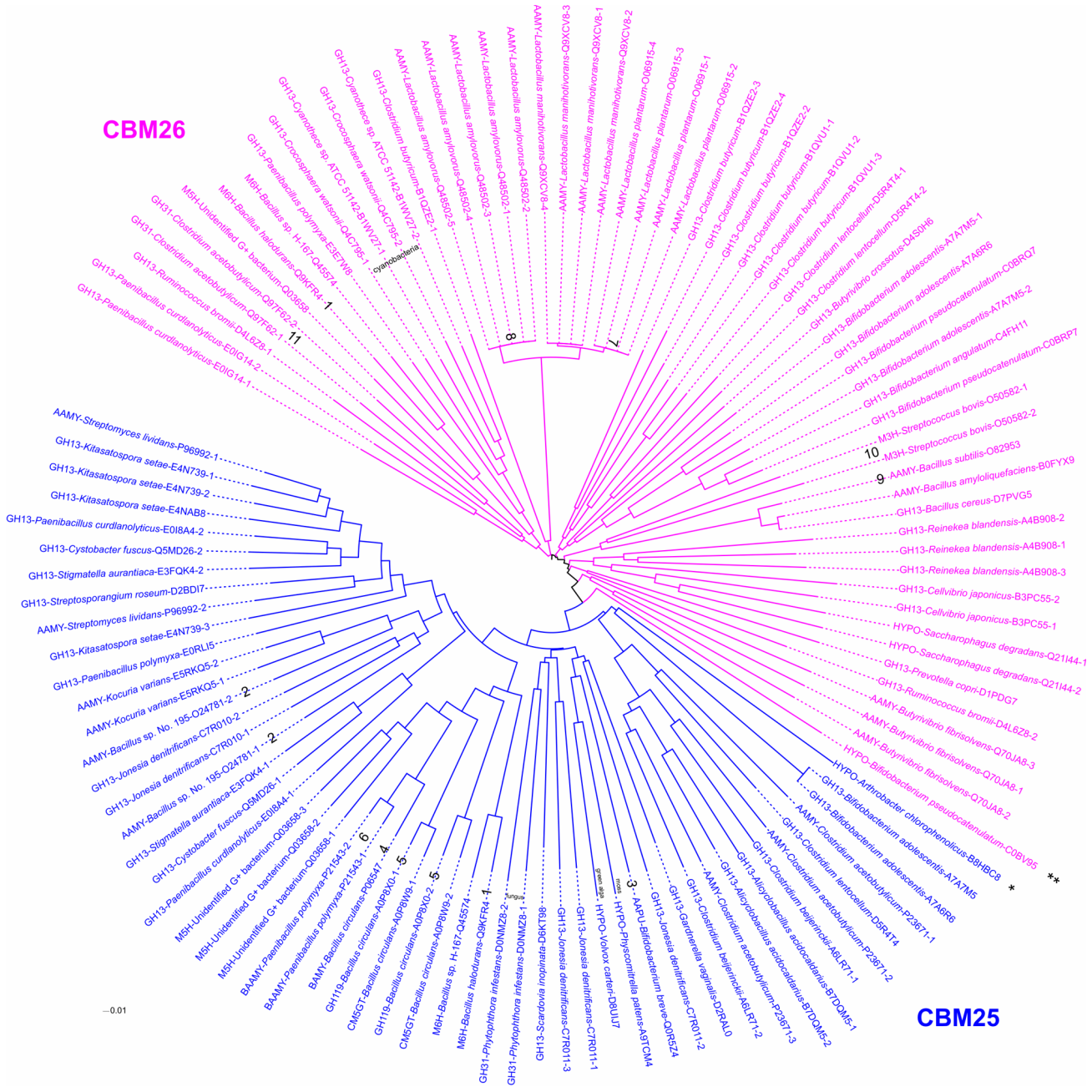

Fig. 3. Evolutionary tree of SBDs from the families CBM25 and CBM26. The tree is based on the alignment of complete SBD sequences including the gaps (Fig. S1) and calculated with the neighbour-joining method (Saitou & Nei 1987) implemented in the Clustal-W package (Larkin et al. 2007). The digits 1-11 represent the CBM25 and CBM26 modules depicted in details in Figures 1 and 2. An asterisk signifies the potential "intermediary" CBM26 copy from the *Bifidobacterium pseudocatenulatum* hypothetical protein. Remarkable (i.e. non-bacterial) CBM25 members from Eucarya – the one from fungus (*Phytophthora infestans*) and the two from plant kingdom from green alga (*Volvox carteri*) and moss (*Physcomitrella patens*) – are also indicated. The tree was displayed with the program iTOL (Letunic & Bork 2006).

are undoubtedly closely related, each one keeps its independence in the evolutionary tree (Fig. 3). It should be possible to trace some taxonomical and very probably enzyme specificity features in both parts of the tree, the phenomena typical also for other CBM families of SBD (Janecek & Sevcik 1999; Janecek et al. 2003). In the CBM25 part of the tree, there are groups formed by Firmicutes and Actinobacteria accompanied by some representatives of Proteobacteria, whereas the groups of Firmicutes and Actinobacteria are completed by some Bacteroidetes, Cyanobacteria and Proteobacteria representatives among the CBM26 copies (cf. Table S1). The presence and position of the three eukaryotic CBM25 SBDs (Fig. 3), currently not classified in the CAZy database (Lombard et al. 2014), should be of interest since they may indicate the future expansion of at least the family CBM25 outside the kingdom of Bacteria covering eventually non-amylolytic enzymes and proteins. Such phenomenon is typical for other SBDs from families CBM20 and CBM48 (Gentry et al. 2009; Janecek et al. 2011).

Putting the SBDs from both families CBM25 and

Table 2. Structural comparison of the "intermediary" CBM26 SBD from *B. pseudocatenulatum* hypothetical protein with representatives of the families CBM25 and CBM26.[a]

	Bifps_2C3H	CBM25 (2C3W)	CBM26 (2C3H)
Bifps_2C3W	60 (1.53)	62 (0.98)	58 (1.36)
Bifps_2C3H		58 (1.54)	68 (0.82)
CBM25 (2C3W)			79 (1.44)

[a] The CBM25 and CBM26 are the real tertiary structures originating from the *Bacillus halodurans* GH13 maltohexaose-forming amylase (Boraston et al. 2006) deposited in the Protein Data Bank (Deshpande et al. 2005) under the codes 2C3W (95 C_α-atoms) and 2C3H (93 C_α-atoms), respectively. The CBM26 from *Bifidobacterium pseudocatenulatum* hypothetical protein (Bifps; UniProt Acc. No.: C0BV95; Table S1) is the SBD potentially intermediate between the two families CBM25 and CBM26. Two putative structures of the CBM26 from *Bifidobacterium pseudocatenulatum* hypothetical protein were modelled using the Phyre-2 server (Kelley & Sternberg 2009): (i) Bifps_2C3W (78 C_α-atoms) using the real structure of CBM25 (PDB code: 2C3W) as a template with the highest confidence; and (ii) Bifps_2C3H (77 C_α-atoms) using the real structure of CBM26 (PDB code: 2C3H) as a template with a lower confidence. The structural models were superimposed using the program MultiProt (Shatsky et al. 2004). The values indicate the number of corresponding C_α-atoms with the root-mean-square deviation in parenthesis.

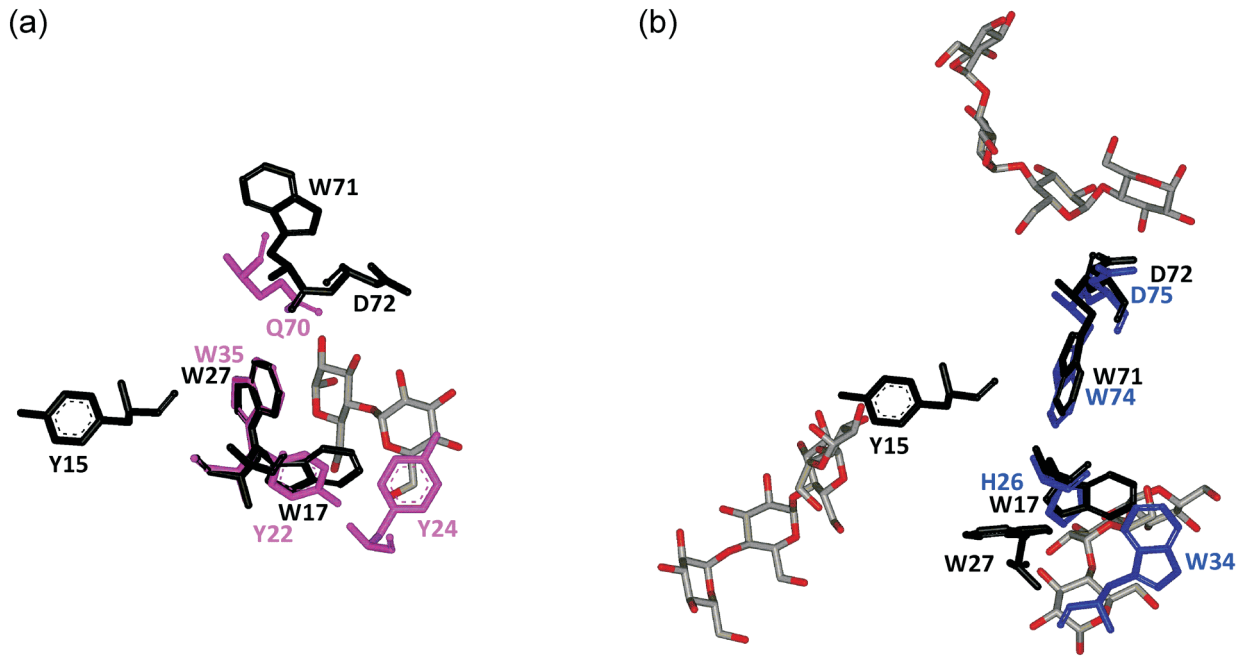

Fig. 4. Comparison of the model of the *Bifidobacterium pseudocatenulatum* "intermediary" SBD with the real CBM25 (a) and CBM26 (b) structures. The models shown in (a) and (b) of the *B. pseudocatenulatum* "intermediary" SBD (black) were generated by the Phyre-2 server (Kelley & Sternberg 2009) according to the real structure of CBM26 (magenta; PDB code: 2C3H) and CBM25 (blue; PDB code: 2C3W), respectively, both originating from the *B. halodurans* maltohexaose-producing amylase (Boraston et al. 2006). The side-chains of the functionally important residues: (i) Tyr15, Trp17, Trp27, Trp71 and Asp72 (intermediary SBD); (ii) Tyr22, Tyr24, Trp35 and Gln70 (real CBM26); and (iii) His26, Trp34, Trp74 and Asp75 (real CBM25) are shown in black, magenta and blue, respectively, and labelled (cf. Fig. 1 and Fig S1). The bound carbohydrates, i.e. a maltose (a) as well as two maltotetraose and one maltotriose (b) are also shown. Structures were retrieved from the PDB (Deshpande et al. 2005) and superimposed using the program MultiProt (Shatsky et al. 2004). Picture prepared with WebLabViewerLite (Molecular Simulations, Inc.).

CBM26 together into a common evolutionary tree offers also the possibility to try to find out the CBM(s) with an intermediary character. Currently, the best candidate for the "intermediary" SBD could be the CBM26 copy from a hypothetical protein from *Bifidobacterium pseudocatenulatum* (UniProt Acc. No.: C0BV95) located on its own independent branch just at the border between the two CBM families (Fig. 3). It possesses the tryptophan essential for the family CBM25 (Trp74 in the *B. halodurans* maltohexaohydrolase CBM25) succeeded even by the functionally important aspartic acid residue (Fig. 1). It is worth mentioning that the CBM from the hypothetical protein from *B. pseudocatenulatum* has still not been included into any of the two CBM25 and CBM26 families of the CAZy database (Lombard et al. 2014).

In order to get a structural support for the intermediary character of the CBM26 SBD mentioned above, its three-dimensional structure was modelled and compared with the real CBM25 and CBM26 structures of the family GH13 maltohexaose-producing amylase from *B. halodurans* (Fig. 4). It was possible to obtain two relevant structural models of the CBM26 from the *B. pseudocatenulatum* hypothetical protein, i.e. one model according to the maltohexaohydrolase CBM25 template and the other one according to the maltohexaohydro-

lase CBM26 template. Although the structure modelled using the CBM25 template was produced with a higher confidence (not shown), the overlays with respective templates are comparable to each other (Table 2). It is clear, however, that Tyr15 of the *B. pseudocatenulatum* hypothetical protein CBM26 corresponds to His26 and Tyr22 of the CBM25 and CBM26, respectively, from the *B. halodurans* maltohexaohydrolase in the sequence comparison (Fig. 1), it is the Trp17 that superimposes correctly with the two binding residues from the two CBM families in the structural comparison (Fig. 4). Remarkably, while the first (i.e. the N-terminal) part of the "intermediary" CBM seems to resemble more the CBM26 template (Fig. 4a; Trp35 from the template perfectly overlapped with Trp27 from the model), its remaining (i.e. the C-terminal) part is obviously more similar to the CBM25 template (Fig. 4b; both Trp74 and Asp75 from the template superimposed with Trp71 and Asp72 from the model). This observation may support, indeed, the intermediary character of the CBM26 SBD from the *B. pseudocatenulatum* hypothetical protein (Fig. 3).

Conclusions

The SBDs from both families CBM25 and CBM26 represent a closely related pair of SBD families. It is very probable that they are a product of divergent evolution from a common SBD ancestor. It is, however, clear, that despite sharing one of the essential aromatic position responsible for the stacking interaction in carbohydrate binding, each family contains its own specific sequence features. It is also possible to trace the SBDs, currently members of the family CBM26, that may exhibit the character intermediary between the two families CBM25 and CBM26. Remarkably, although both families originally contained only microbial amylolytic enzymes, a few eukaryotic proteins were identified as possessing the CBM25 motif. This study could therefore contribute to a detailed future division of both CBM families into subfamilies, revealing the relationships between enzyme specificity and taxonomy among these SBDs and indicating the real raw starch CBM25 and CBM26 binders.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract No. LPP-0417-09 and by the VEGA grant No. 2/0150/14.

References

Abe A., Tonozuka T., Sakano Y. & Kamitori S. 2004. Complex structures of *Thermoactinomyces vulgaris* R-47 α-amylase 1 with malto-oligosaccharides demonstrate the role of domain N acting as a starch-binding domain. J. Mol. Biol. **335:** 811–822.

Altschul S.F., Gish W., Miller W., Myers E.W. & Lipman D.J. 1990. Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Ashikari T., Nakamura N., Tanaka Y., Kiuchi N., Shibano Y., Tanaka T., Amachi T. & Yoshizumi H. 1986. *Rhizopus* raw-starch-degrading glucoamylase: its cloning and expression in yeast. Agric. Biol. Chem. **50:** 957–964.

Boraston A.B., Bolam D.N., Gilbert H.J. & Davies G.J. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. Biochem. J. **382:** 769–781.

Boraston A.B., Healey M., Klassen J., Ficko-Blean E., Lammerts van Bueren A. & 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.

Cantarel B.L., Coutinho P.M., Rancurel C., Bernard T., Lombard V. & Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Res. **37:** D233–D238.

Christiansen C., Abou Hachem M., Janecek S., Viksř-Nielsen A., Blennow A. & Svensson B. 2009. The carbohydrate-binding module family 20 – diversity, structure, and function. FEBS J. **276:** 5006–5029.

Cockburn D., Wilkens C., Ruzanski C., Andersen S., Nielsen J.W., Smith A.M., Field R.A., Willemoës M., Abou Hachem M. & Svensson B. 2014. Analysis of surface binding sites (SBSs) in carbohydrate active enzymes with focus on glycoside hydrolase families 13 and 77 – a mini-review. Biologia **69:** 705–712.

Deshpande N., Addess K.J., Bluhm W.F., Merino-Ott J.C., Townsend-Merino W., Zhang Q., Knezevich C., Xie L., Chen L., Feng Z., Green R.K., Flippen-Anderson J.L., Westbrook J., Berman H.M. & Bourne P.E. 2005. Nucleic Acids Res. **33:** D233–D237.

Gentry M.S., Dixon J.E. & Worby C.A. 2009. Lafora disease: insights into neurodegeneration from plant metabolism. Trends Biochem. Sci. **34:** 628–639.

Gentry M.S., Romá-Mateo C. & Sanz P. 2013. Laforin, a protein with many faces: glucan phosphatase, adapter protein, et alii. FEBS J. **280:** 525–537.

Giraud E. & 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.

Glaring M.A., Baumann M.J., Abou Hachem M., Nakai H., Nakai N., Santelia D., Sigurskjold B.W., Zeeman S.C., Blennow A. & Svensson B. 2011. Starch-binding domains in the CBM45 family – low-affinity domains from glucan, water dikinase and α-amylase involved in plastidial starch metabolism. FEBS J. **278:** 1175–1185.

Guillen D., Sanchez S. & Rodriguez-Sanoja R. 2010. Carbohydrate-binding domains: multiplicity of biological roles. Appl. Microbiol. Biotechnol. 85: 1241–1249.

Guillen D., Santiago M., Linares L., Perez R., Morlon J., Ruiz B., Sanchez S. & Rodriguez-Sanoja R. 2007. α-Amylase starch binding domains: cooperative effects of binding to starch granules of multiple tandemly arranged domains. Appl. Environ. Microbiol. **73:** 3833–3837.

Haas B.J., Kamoun S., Zody M.C., Jiang R.H., Handsaker R.E., Cano L.M., Grabherr M., ... & Nusbaum C. 2009. Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature **461:** 393–398.

Hudson E.R., Pan D.A., James J., Lucocq J.M., Hawley S.A., Green K.A., Baba O., Terashima T. & 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.

Janecek S. 2002. A motif of a microbial starch-binding domain found in human genethonin. Bioinformatics **18:** 1534–1537.

Janecek S. & Kuchtova A. 2012. *In silico* identification of catalytic residues and domain fold of the family GH119 sharing the catalytic machinery with the α-amylase family GH57. FEBS Lett. **586:** 3360–3366.

Janecek S. & Sevcik J. 1999. The evolution of starch-binding domain. FEBS Lett. **456:** 119–125.

Janecek S., Svensson B. & MacGregor E.A. 2003. Relation between domain evolution, specificity, and taxonomy of the

α-amylase family members containing a C-terminal starch-binding domain. Eur. J. Biochem. **270:** 635–645.

Janecek S., Svensson B. & MacGregor E.A. 2011. Structural and evolutionary aspects of two families of non-catalytic domains present in starch and glycogen binding proteins from microbes, plants and animals. Enzyme Microb. Technol. **49:** 429–440.

Janecek S., Svensson B. & MacGregor E.A. 2014. α-Amylase: an enzyme specificity found in various families of glycoside hydrolases. Cell. Mol. Life Sci. **71:** 1149–1170.

Jiang S., Heller B., Tagliabracci V.S., Zhai L., Irimia J.M., DePaoli-Roach A.A., Wells C.D., Skurat A.V. & Roach P.J. 2010. Starch binding domain-containing protein 1/genethonin 1 is a novel participant in glycogen metabolism. J. Biol. Chem. **285:** 34960–34971.

Kelley L.A. & Sternberg M.J. 2009. Protein structure prediction on the Web: a case study using the Phyre server. Nat. Protoc. **4:** 363–371.

Koropatkin N.M. & Smith T.J. 2010. SusG: a unique cell-membrane-associated α-amylase from a prominent human gut symbiont targets complex starch molecules. Structure **18:** 200–215.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. & Higgins D.G. 2007. Clustal W and Clustal X version 2.0. Bioinformatics **23:** 2947–2948.

Letunic I. & Bork P. 2007. Interactive Tree Of Life (iTOL): an online tool for phylogenetic tree display and annotation. Bioinformatics **23:** 127–128.

Lombard V., Golaconda Ramulu H., Drula E., Coutinho P.M. & Henrissat B. 2014. The Carbohydrate-active enzymes database (CAZy) in 2013. Nucleic Acids Res. **42:** D490–D495.

Machovic M. & Janecek S. 2006a. The evolution of putative starch-binding domains. FEBS Lett. **580:** 6349–6356.

Machovic M. & Janecek S. 2006b. Starch-binding domains in the post-genome era. Cell. Mol. Life Sci. **63:** 2710–2724.

Machovic M., Svensson B., MacGregor E.A. & 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.

Matsui Y., Okada S., Uchimura T., Kondo A. & Satoh E. 2007. Determination and analysis of the starch binding domain of *Streptococcus bovis* 148 raw-starch-hydrolyzing α-amylase. J. Appl. Glycosci. **54:** 217–222.

Mikkelsen R., Suszkiewicz K. & 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.

Minassian B.A., Ianzano L., Meloche M., Andermann E., Rouleau G.A., Delgado-Escueta A.V. & Scherer S.W. 2000. Mutation spectrum and predicted function of laforin in Lafora's progressive myoclonus epilepsy. Neurology **55:** 341–346.

Orzechowski S., Grabowska A., Sitnicka D., Siminska J., Felus M., Dudkiewicz M., Fudali S. & Sobczak M. 2013. Analysis of the expression, subcellular and tissue localisation of phosphoglucan, water dikinase (PWD/GWD3) in *Solanum tuberosum* L.: a bioinformatics approach for the comparative analysis of two α-glucan, water dikinases (GWDs) from *Solanum tuberosum* L. Acta Physiol. Plant. **35:** 483–500.

Peng H., Zheng Y., Chen M., Wang Y., Xiao Y. & Gao Y. 2014. A starch-binding domain identified in α-amylase (AmyP) represents a new family of carbohydrate-binding modules that contribute to enzymatic hydrolysis of soluble starch. FEBS Lett. **588:** 1161–1167.

Penninga D., van der Veen B.A., Knegtel R.M., van Hijum S.A., Rozeboom H.J., Kalk K.H., Dijkstra B.W. & Dijkhuizen L. 1996. The raw starch binding domain of cyclodextrin glycosyltransferase from *Bacillus circulans* strain 251. J. Biol. Chem. **271:** 32777–32784.

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. & Stapleton D. 2003. AMPK β subunit targets metabolic stress sensing to glycogen. Curr. Biol. **13:** 867–871.

Polekhina G., Gupta A., van Denderen B.J., Feil S.C., Kemp B.E., Stapleton D. & Parker M.W. Structural basis for glycogen recognition by AMP-activated protein kinase. Structure **13:** 1453–1462.

Rodriguez Sanoja R., Morlon-Guyot J., Jore J., Pintado J., Juge N. & Guyot J.P. 2000. Comparative characterization of complete and truncated forms of *Lactobacillus amylovorus* α-amylase and role of the C-terminal direct repeats in raw-starch binding. Appl. Environ. Microbiol. **66:** 3350–3356.

Rodriguez-Sanoja R., Oviedo N., Escalante L., Ruiz B. & Sanchez S. 2009. A single residue mutation abolishes attachment of the CBM26 starch-binding domain from *Lactobacillus amylovorus* α-amylase. J. Ind. Microbiol. Biotechnol. **36:** 341–346.

Rodriguez-Sanoja R., Oviedo N. & Sanchez S. 2005a. Microbial starch-binding domain. Curr. Opin. Microbiol. **8:** 260–267.

Rodriguez-Sanoja R., Ruiz B., Guyot J.P. & Sanchez S. 2005b. Starch-binding domain affects catalysis in two *Lactobacillus* α-amylases. Appl. Environ. Microbiol. **71:** 297–302.

Ryan S.M., Fitzgerald G.F. & van Sinderen D. 2006. Screening for and identification of starch-, amylopectin-, and pullulan-degrading activities in bifidobacterial strains. Appl. Environ. Microbiol. **72:** 5289–5296.

Saitou N. & Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4:** 406–425.

Shatsky M., Nussinov R. & Wolfson H.J. 2004. A method for simultaneous alignment of multiple protein structures. Proteins **56:** 143–156.

Siggens K.W. 1987. Molecular cloning and characterization of the β-amylase gene from *Bacillus circulans*. Mol. Microbiol. **1:** 86–91.

Sorimachi K., Le Gal-Coëffet M.F., Williamson G., Archer D.B. & Williamson M.P. 1997. Solution structure of the granular starch binding domain of *Aspergillus niger* glucoamylase bound to β-cyclodextrin. Structure **5:** 647–661.

Southall S.M., Simpson P.J., Gilbert H.J., Williamson G. & Williamson M.P. 1999. The starch-binding domain from glucoamylase disrupts the structure of starch. FEBS Lett. **447:** 58–60.

Sumitani J., Tottori T., Kawaguchi T. & Arai M. 2000. New type of starch-binding domain: the direct repeat motif in the C-terminal region of *Bacillus* sp. no. 195 α-amylase contributes to starch binding and raw starch degrading. Biochem. J. **350:** 477–484.

Svensson B., Jespersen H., Sierks M.R. & MacGregor E.A. 1989. Sequence homology between putative raw-starch binding domains from different starch-degrading enzymes. Biochem. J. **264:** 309–311.

Tung J.Y., Chang M.D., Chou W.I., Liu Y.Y., Yeh Y.H., Chang F.Y., Lin S.C., Qiu Z.L. & Sun Y.J. 2008. Crystal structures of the starch-binding domain from *Rhizopus oryzae* glucoamylase reveal a polysaccharide-binding path. Biochem. J. **416:** 27–36.

UniProt Consortium. 2013. Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res. **41:** D43–D47.

van Bueren A.L. & Boraston A.B. 2007. The structural basis of α-glucan recognition by a family 41 carbohydrate-binding module from *Thermotoga maritima*. J. Mol. Biol. **365:** 555–560.

Vander Kooi C.W., Taylor A.O., Pace R.M., Meekins D.A., Guo H.F., Kim Y. & Gentry M.S. 2010. Structural basis for the glucan phosphatase activity of Starch Excess4. Proc. Natl. Acad. Sci. USA **107:** 15379–15384.

Watanabe H., Nishimoto T., Kubota M., Chaen H. & Fukuda S. 2006. Cloning, sequencing, and expression of the genes encoding an isocyclomaltooligosaccharide glucanotransferase and an α-amylase from a *Bacillus circulans* strain. Biosci. Biotechnol. Biochem. **70:** 2690–2702.

Wayllace N.Z., Valdez H.A., Ugalde R.A., Busi M.V. & Gomez-Casati D.F. 2010. The starch-binding capacity of the noncatalytic SBD2 region and the interaction between the N- and

C-terminal domains are involved in the modulation of the activity of starch synthase III from *Arabidopsis thaliana*. FEBS J. **277:** 428–440.

Xu J., Ren F., Huang C.H., Zheng Y., Zhen J., Sun H., Ko T.P., He M., Chen C.C., Chan H.C., Guo R.T., Song H. & Ma Y. 2014. Functional and structural studies of pullulanase from *Anoxybacillus* sp. LM18-11. Proteins (in press); doi: 10.1002/prot.24498.

Yamaguchi R., Arakawa T., Tokunaga H., Ishibashi M. & Tokunaga M. 2012a. Distinct characteristics of single starch-binding domain SBD1 derived from tandem domains SBD1-SBD2 of halophilic *Kocuria varians* α-amylase. Protein J. **31:** 250–258.

Yamaguchi R., Inoue Y., Tokunaga H., Ishibashi M., Arakawa T., Sumitani J., Kawaguchi T. & Tokunaga M. 2012b. Halophilic characterization of starch-binding domain from *Kocuria varians* α-amylase. Int. J. Biol. Macromol. **50:** 95–102.

Yamaguchi R., Tokunaga H., Ishibashi M., Arakawa T. & Tokunaga M. 2011. Salt-dependent thermo-reversible α-amylase: cloning and characterization of halophilic α-amylase from moderately halophilic bacterium, *Kocuria varians*. Appl. Microbiol. Biotechnol. **89:** 673–684.

Received March 5, 2014
Accepted May 12, 2014