A motif of a microbial starch-binding domain found in human genethonin

Štefan Janeček

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

ABSTRACT

Summary: The sequence of the starch-binding domain present in 10% of amylolytic enzymes of microbial origin and classified as the carbohydrate-binding module family 20, was identified in the equivalent part of sequence of human genethonin, a skeletal muscle protein of unknown function. The sequence identity between the starch-binding domain from *Bacillus* sp. strain 1011 cyclodextrin glucanotransferase and the corresponding segment of human genethonin was higher than 28%. The amino acid residues known to be involved in the raw starch binding were found to be conserved in the genethonin sequence. The three-dimensional structure of the genethonin 'starch-binding domain' was modelled and its eventual function briefly discussed.

**Contact:** stefan.janecek@savba.sk

INTRODUCTION

About 10% of amylases and related enzymes contain a characteristic sequence motif at the C-terminus, the so-called starch-binding domain (SBD). There is at least one remote SBD positioned N-terminally in *Rhisopus oryzae* glucoamylase (Ashikari *et al.*, 1986). The SBD module, consisting of several β-strands forming an open-sided, distorted β-barrel structure, is responsible for the ability to bind and degrade raw starch by certain amylolytic enzymes (Lawson *et al.*, 1994; Penninga *et al.*, 1996; Sorimachi *et al.*, 1997; Giardina *et al.*, 2001). Recently it has been recognised that this property may be connected to other structural elements, such as some C-terminal repeats (Rodriguez-Sanoja *et al.*, 2000; Sumitani *et al.*, 2000), a part of the catalytic $(\beta/\alpha)_8$-barrel domain in barley α-amylase (Tibbot *et al.*, 2000), or even as a starch granule binding site in this enzyme (Søgaard *et al.*, 1993). The 'classical' C-terminal SBD sequence is common for α-amylases, β-amylases and glucoamylases (Svensson *et al.*, 1989) that differ in catalytic-domain structure (Matsuura *et al.*, 1984; Aleshin *et al.*, 1992; Mikami *et al.*, 1993) and reaction mechanism (Henrissat and Davies, 1997; Uitdehaag *et al.*, 1999). It was found that the evolution of SBD in the three amylolytic families reflects the evolution of species (taxonomy) rather than the evolution of the individual amylase specificities (Janeček and Ševčík, 1999). Recently the SBD, consisting of several β-strands forming an open-sided, distorted β-barrel structure (Lawson *et al.*, 1994; Sorimachi *et al.*, 1997), has been defined as the family 20 (CBM20) in the sequence-based classification system of carbohydrate-binding modules (Coutinho and Henrissat, 1999).

For many years this SBD was thought to exist exclusively in microbial amylolytic enzymes. Two basic sequence types of SBD were recognised forming a bacterial and a fungal group, the SBDs originated from actinomycetes being found closer to the fungal type (Janeček and Ševčík, 1999). The current situation with a huge amount of sequence data stored in the molecular-biology databases (especially from the sequencing of complete genomes) has evoked the interest to find out whether or not an SBD-related sequence can be revealed in a protein other than amylase and of non-microbial origin. This 'database hunting' has been accelerated since the year 2000 when laforin, a protein involved in the Lafora type of epilepsy in man, was discovered to contain some sequence features of SBD (Minassian *et al.*, 2000). In the present paper another non-microbial protein, genethonin from human skeletal muscle, is presented to contain not only some SBD-like sequence features, but to exhibit unambiguous sequence similarity (identity higher than 25%) with the entire length of the SBD from CBM20 as well as to retain the position at the C-terminal end of the polypeptide chain.

BACKGROUND

BLAST (Altschul *et al.*, 1997) was used for performing the searches in the molecular-biology databases (using the default parameters). As query the entire sequences of SBD from *Bacillus circulans* strain 251 cyclodextrin glucanotransferase (CGTase; SwissProt: P43379; Lawson *et al.*, 1994) and *Aspergillus niger* glucoamylase (SwissProt: P04064; Svensson *et al.*, 1983) were applied.
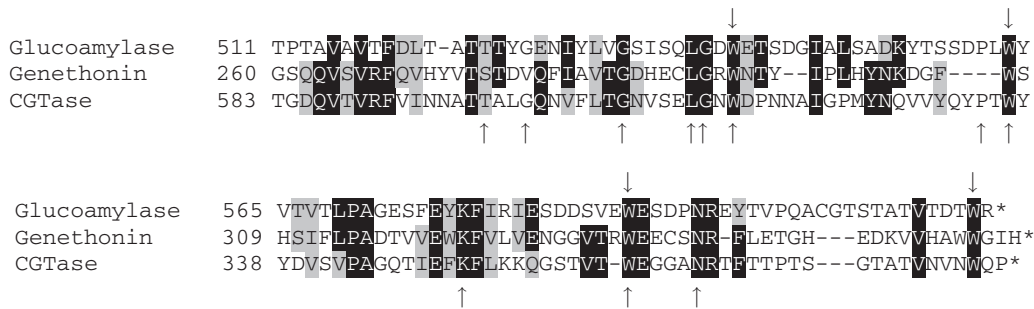
Fig. 1. Alignment of starch-binding domain sequences from *Aspergillus niger* glucoamylase (SwissProt: P04064; Svensson *et al.*, 1983) and *Bacillus* sp. strain 1011 cyclodextrin glucanotransferase (SwissProt: P05618; Kimura *et al.*, 1987) with the equivalent sequence of human genethonin (SwissProt: O95210; Bouju *et al.*, 1998). The arrows above the alignment signify the tryptophan positions important for structure and function of starch-binding domain, while the arrows below the alignment mark the residues identified by Svensson *et al.* (1989) as *consensus* starch-binding domain positions. The C-terminal end of each protein is denoted by an asterisk.

The relevant amino acid sequences and coordinates of protein structures were retrieved from SwissProt sequence database (Bairoch and Apweiler, 2000) and Protein Data Bank (Berman *et al.*, 2000), respectively. The alignments were done using the program CLUSTALW (Thompson *et al.*, 1994) utilizing the alignment of 43 sequences of SBD made previously (Janeček and Ševčík, 1999) as template. The Pfam database (Bateman *et al.*, 2000) and the Prodom tool (Corpet *et al.*, 2000) integrated in the InterPro server (Apweiler *et al.*, 2000) were explored in order to gain as much information as possible about the SBD module concerning its sequence, structure and evolution, and possible presence among the available sequences. The three-dimensional modelling was performed using the 3D-PSSM automatic fold recognition method (Kelley *et al.*, 2000) via the SwissModel server (Guex and Peitsch, 1997). As a template for homology modelling the SBD from *Bacillus* sp. strain 1011 CGTase (PDB code: 1PAM; Harata *et al.*, 1996) was used.

RESULTS AND DISCUSSION

BLAST searches focused on catching the proteins with a sequence similarity to SBD sequence have yielded other proteins, besides a lot of SBDs of known amylolytic enzymes, which are either non-amylolytic enzymes or putative proteins. Two human proteins, laforin and genethonin, deserve immediate attention. Laforin is a protein responsible for the Lafora type of epilepsy in man and it has already been recognised as having a limited sequence similarity to SBD concerning only a few best conserved amino acid residues (Minassian *et al.*, 2000). Remarkably, Wang *et al.* (2002) have very recently shown that laforin may indeed contain a functional starch/carbohydrate-binding domain that could target the Lafora disease phosphatase to glycogen. Laforin, however, has the relevant sequence seg-

ment localized at its N-terminal end in comparison with the position of microbial-type SBD in amylases.

The similarity of a segment of genethonin corresponding to SBD is even more evident (Figure 1). Genethonin is a protein from human skeletal muscle with as yet unknown function (Bouju *et al.*, 1998). The sequence identity between this genethonin part and SBD from *A. niger* glucoamylase and *Bacillus* sp. strain 1011 CGTase is 25.7 and 28.3%, respectively, as distributed over the entire length of SBD. The similarity reaches 37.6 and 42.5%, respectively. As can be seen from Figure 1 genethonin contains its 'starch-binding domain' at the C-terminal end of the protein, i.e. in the position equivalent to that of classical SBD from microbial amylases. It is worth mentioning that all four tryptophan residues (Trp616, Trp636, Trp662 and Trp684; *Bacillus* CGTase numbering) important for SBD structure and function (Penninga *et al.*, 1996; Sorimachi *et al.*, 1997; Giardina *et al.*, 2001) are perfectly conserved in the genethonin sequence (Figure 1). With regard to the so-called *consensus* SBD sequence identified by Svensson *et al.* (1989) eight of the total eleven positions are conserved in the genethonin sequence while one conservative substitution (Thr598), one non-conservative substitution (Gly601) and one deletion (Pro634) are seen for the three remaining positions (Figure 1). It should be taken into account, however, that as the number of available SBD sequences increased, up to 2% of these *consensus* positions were found to be substituted and/or deleted even among SBDs originating from amylolytic enzymes (Janeček and Ševčík, 1999). All these observations indicate that the C-terminal part of genethonin may exhibit a carbohydrate-binding function.

The overall high sequence identity and similarity between SBD and an equivalent segment of genethonin over a stretch of more than 100 amino acid residues
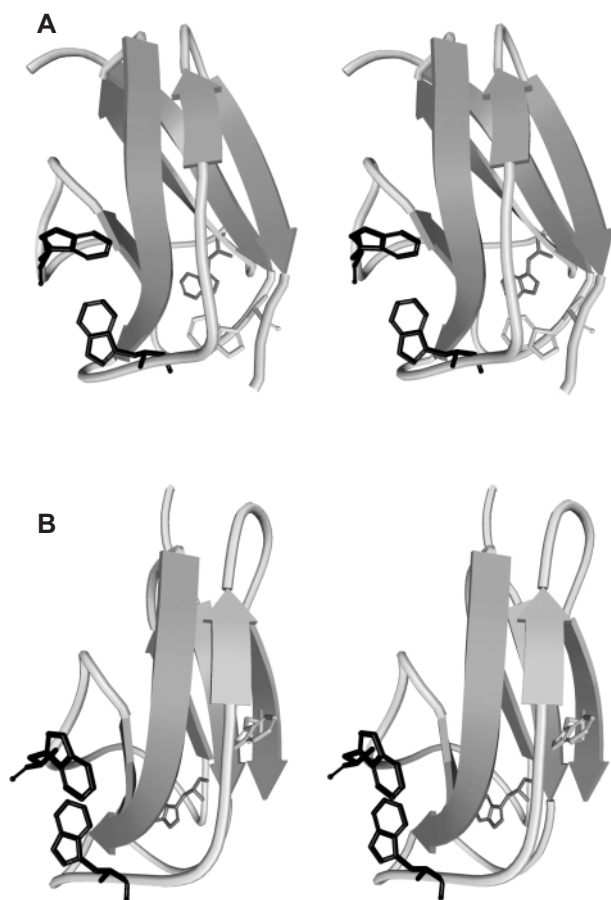
Fig. 2. The three-dimensional ribbon diagrams of (a) the X-ray structure of starch-binding domain from *Bacillus* sp. strain 1011 cyclodextrin glucanotransferase (PDB code: 1PAM; Harata *et al.*, 1996) and (b) the modelled structure of genethonin 'starch-binding domain'. The side-chains of the four important tryptophan residues of the starch-binding domain forming the starch-binding sites 1 (Trp616 and Trp662) and 2 (Trp636), and playing the primary structural role (Trp684), and their equivalents in genethonin (W293 and W334, W307 and W355), are highlighted in black, dark grey and light grey, respectively, in both proteins. Figure in stereo was prepared using the program WebLab ViewerLite 4.0, Molecular Simulations Inc.

(Figure 1) enabled homology modelling of three-dimensional structure of the C-terminal part of genethonin. Figure 2 shows the basic features of the genethonin 'starch-binding domain' which, in the number of β-strands and their mutual orientation, strongly resemble the template structure of SBD from *Bacillus* sp. strain 1011 CGTase. Interestingly, also the side-chain orientation of the three tryptophans involved in starch-binding sites 1 and 2 mentioned above seem to be conserved in a way which could allow them to exert their binding functions.

The fourth tryptophan residue (Trp684) was determined to play a primarily structural role (Sorimachi *et al.*, 1997) and exhibits different orientation. This may also be due to the eventual shift in the last β-strand segment (Figure 2).

It is thus very probable that genethonin has at its C-terminal end a domain structurally homologous to SBD, i.e. a motif consisting of several β-strands forming an open-sided, distorted β-barrel structure (Lawson *et al.*, 1994; Sorimachi *et al.*, 1997). Whether also the starch binding function has been preserved in genethonin, remains unclear and cannot be deciphered from this study. The evolutionary, structural and functional independence of SBD, however, has already been demonstrated (Dalmia *et al.*, 1995; Janeček and Ševčík, 1999; Ohdan *et al.*, 2000). The main aim of this work was to draw the attention to genethonin since the function for the whole protein is still unknown. The results concerning the laforin SBD-like features (Minassian *et al.*, 2000) and the glycogen-binding function of this laforin segment (Wang *et al.*, 2002) together with the clear sequence and/or structural similarity between the C-terminal part of genethonin and SBD (presented here) may indicate that human genethonin does contain a carbohydrate-binding domain at its C-terminus. The human skeletal muscle genethonin is thus a suitable candidate for future protein engineering studies.

ACKNOWLEDGEMENTS

REFERENCES

Aleshin,A., Golubev,A., Firsov,L.M. and Honzatko,R.B. (1992) Crystal structure of glucoamylase from *Aspergillus awamori* var. X100 to 2.2 Å resolution. *J. Biol. Chem.*, **267**, 19291–19298.

Altschul,S.F., Stephen,F., Madden,T.L., Schäffer,A.A., Zhang,J., Zhang,Z., Miller,W. and Lipman,D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.*, **25**, 3389–3402.

Apweiler,R., Attwood,T.K., Bairoch,A., Bateman,A., Birney,E., Biswas,M., Bucher,P., Cerutti,L., Corpet,F., Croning,M.D. *et al.* (2000) InterPro—an integrated documentation resource for protein families, domains and functional sites. *Bioinformatics*, **16**, 1145–1150.

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.

Bairoch,A. and Apweiler,R. (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, **28**, 45–48.

Bateman,A., Birney,E., Durbin,R., Eddy,S.R., Howe,K.L. and Sonnhammer,E.L. (2000) The Pfam protein families database. *Nucleic Acids Res.*, **28**, 263–266.

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.

Bouju,S., Lignon,M.F., Pietu,G., Le Cunff,M., Leger,J.J., Auffray,C. and Dechesne,C.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.

Corpet,F., Servant,F., Gouzy,J. and Kahn,D. (2000) ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Res.*, **28**, 267–269.

Coutinho,P.M. and Henrissat,B. (1999) Carbohydrate-active enzymes: an integrated database approach. In Gilbert,H.J., Davies,G., Henrissat,B. and Svensson,B. (eds), *Recent Advances in Carbohydrate Bioengineering*. The Royal Society of Chemistry, Cambridge, pp. 3–12.

Dalmia,B.K., Schütte,K. and Nikolov,Z.L. (1995) Domain E of *Bacillus macerans* cyclodextrin glucanotransferase: an independent starch-binding domain. *Biotechnol. Bioeng.*, **47**, 575–584.

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.

Guex,N. and Peitsch,M.C. (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modelling. *Electrophoresis*, **18**, 2714–2723.

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.*, **D52**, 1136–1145.

Henrissat,B. and Davies,G. (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.*, **7**, 637–644.

Janeček,Š. and Ševčík,J. (1999) The evolution of starch-binding domain. *FEBS Lett.*, **456**, 119–125.

Kelley,L.A., MacCallum,R.M. and Sternberg,M.J.E. (2000) Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.*, **299**, 499–520.

Kimura,K., Kataoka,S., Ishii,Y., Takano,T. and Yamane,K. (1987) Nucleotide sequence of the β-cyclodextrin glucanotransferase gene of alkalophilic *Bacillus* sp. strain 1011 and similarity of its amino acid sequence to those of α-amylases. *J. Bacteriol.*, **169**, 4399–4402.

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.

Matsuura,Y., Kusunoki,M., Harada,W. and Kakudo,M. (1984) Structure and possible catalytic residues of Taka-amylase A. *J. Biochem.*, **95**, 697–702.

Mikami,B., Hehre,E.J., Sato,M., Katsube,Y., Hirose,M., Morita,Y. and Sacchettini,J.C. (1993) The 2.0 Å resolution structure of soybean β-amylase complexed with α-cyclodextrin. *Biochemistry*, **32**, 6836–6845.

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

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.

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

Rodriguez-Sanoja,R., Morlon-Guyot,J., Jore,J., Pintado,J., Juge,N. and 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.

Sorimachi,K., Le Gal-Coëffet,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.

Søgaard,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.

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.

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.

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.

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

Tibbot,B.K., Wong,D.W. and Robertson,G.H. (2000) A functional raw starch-binding domain of barley α-amylase expressed in *Escherichia coli*. *J. Protein Chem.*, **19**, 663–669.

Uitdehaag,J.C.M., Mosi,R.M., Kalk,K.H., van der Veen,B.A., Dijkhuizen,L., Withers,S.G. and Dijkstra,B.W. (1999) X-ray structures along the reaction pathway of cyclodextrin glycosyltransferase elucidate catalysis in the α-amylase family. *Nature Struct. Biol.*, **6**, 432–436.

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.