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

Received on February 5, 2002; revised on April 3, 2002; accepted on April 25, 2002

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 (β/α)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 (Henrisat 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.
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 (Bairoch and Apweiler, 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; Kimura et al., 1987) with the equivalent sequence of human genethonin (Svensson et al., 1998) was used.

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.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucoamylase</td>
<td>511 TPTAANSFTGTVKGDHYSISGGDTRNLDGQSSVTSSDGPLL</td>
</tr>
<tr>
<td>Genethonin</td>
<td>260 GSLSQVRVQVYVGYVANCTGGHGNYRLYDSRAMHOG-------AG</td>
</tr>
<tr>
<td>CGTase</td>
<td>583 TC00TF7PFNVNALTALCGNFTQGSVSHCQFPNRAQPPOVYTPY</td>
</tr>
<tr>
<td>Glucoamylase</td>
<td>565 VTVLPEGSEPSFATFBRSSDSSVBGSDDPSRNYTTVPACGTSTATDTDR*</td>
</tr>
<tr>
<td>Genethonin</td>
<td>309 HSVFDPITVVRKLVNGQTNEGCRGERASLETGHC----EDKVGNUMGTH*</td>
</tr>
<tr>
<td>CGTase</td>
<td>338 YLQGSEPQTSFTFRKKQSTTVPPSAGRRGQTTPTS----GTAQVNTGDP*</td>
</tr>
</tbody>
</table>

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 segment 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 corresponding consensus 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...
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 $\beta$-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 $\beta$-strands forming an open-sided, distorted $\beta$-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

This work was supported in part by the VEGA Grant No. 2/2057/22 from the Slovak Grant Agency.

REFERENCES


