The starch-binding domain family CBM41 - an in silico analysis of evolutionary relationships

Within the CAZy database, there are 81 carbohydrate-binding module (CBM) families. A CBM represents a non-catalytic domain in a modular arrangement of glycoside hydrolases (GHs). The present in silico study has been focused on starch-binding domains from the family CBM41 that are usually part of pullulanases from the α-amylase family GH13. Currently there are more than 1,600 sequences classified in the family CBM41, almost exclusively from Bacteria, and so a study was undertaken in an effort to divide the members into relevant groups (subfamilies) and also to contribute to the evolutionary picture of family CBM41. The CBM41 members adopt a β-sandwich fold (∼100 residues) with one carbohydrate-binding site formed by the side-chains of three aromatic residues that interact with carbohydrate. The family CBM41 can be divided into two basic subdivisions, distinguished from each other by a characteristic sequence pattern or motif of the three essential aromatics as follows: (i) "W-W-∼10aa-W" (the so-called Streptococcus/Klebsiella-type); and (ii) "W-W-∼30aa-W" (Thermotoga-type). Based on our bioinformatics analysis it is clear that the first and second positions of the motif can be occupied by aromatic residues (Phe, Tyr, His) other than tryptophan, resulting in the existence of six different carbohydrate-binding CBM41 groups, that reflect mostly differences in taxonomy, but which should retain the ability to bind an α-glucan. In addition, three more groups have been proposed that, although lacking the crucial aromatic motif, could possibly employ other residues from remaining parts of their sequence for binding carbohydrate. This article is protected by copyright. All rights reserved.