Functional Roles of Starch Binding Domains and Surface Binding Sites in Enzymes Involved in Starch Biosynthesis

Wilkens, Casper; Svensson, Birte; Møller, Marie Sofie

Published in:
Frontiers in Plant Science

Link to article, DOI:
10.3389/fpls.2018.01652

Publication date:
2018

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):
Functional Roles of Starch Binding Domains and Surface Binding Sites in Enzymes Involved in Starch Biosynthesis

Casper Wilkens¹, Birte Svensson² and Marie Sofie Møller²*

¹ Enzyme Technology, Department of Bioengineering and Biomedicine, Technical University of Denmark, Lyngby, Denmark, ² Enzyme and Protein Chemistry, Department of Bioengineering and Biomedicine, Technical University of Denmark, Lyngby, Denmark

Biosynthesis of starch is catalyzed by a cascade of enzymes. The activity of a large number of these enzymes depends on interaction with polymeric substrates via carbohydrate binding sites, which are situated outside of the catalytic site and its immediate surroundings including the substrate-binding crevice. Such secondary binding sites can belong to distinct starch binding domains (SBDs), classified as carbohydrate binding modules (CBMs), or be surface binding sites (SBSs) exposed on the surface of catalytic domains. Currently in the Carbohydrate-Active enZYmes (CAZy) database SBDs are found in 13 CBM families. Four of these families; CBM20, CBM45, CBM48, and CBM53 are represented in enzymes involved in starch biosynthesis, namely starch synthases, branching enzymes, isoamylases, glucan, water dikinases, and α-glucan phosphatases. A critical role of the SBD in activity has not been demonstrated for any of these enzymes. Among the well-characterized SBDs important for starch biosynthesis are three CBM53s of Arabidopsis thaliana starch synthase III, which have modest affinity. SBSs, which are overall less widespread than SBDs, have been reported in some branching enzymes, isoamylases, synthases, phosphatases, and phosphorylases active in starch biosynthesis. SBSs appear to exert roles similar to CBMs. SBSs, however, have also been shown to modulate specificity for example by discriminating the length of chains transferred by branching enzymes. Notably, the difference in rate of occurrence between SBDs and SBSs may be due to lack of awareness of SBSs. Thus, SBSs as opposed to CBMs are not recognized at the protein sequence level, which hampers their identification. Moreover, only a few SBSs in enzymes involved in starch biosynthesis have been functionally characterized, typically by structure-guided site-directed mutagenesis. The glucan phosphatase Like SEX4 2 from A. thaliana has two SBSs with weak affinity for β-cyclodextrin, amylose and amylopectin, which were indicated by mutational analysis to be more important than the active site for initial substrate recognition. The present review provides an update on occurrence of functional SBDs and SBSs in enzymes involved in starch biosynthesis.

Keywords: carbohydrate binding module, surface binding site, starch synthesis, protein-carbohydrate interaction, starch binding domain, glycoside hydrolase, glycosyl transferase
INTRODUCTION

Starch is an insoluble polymer composed of two α-glucans: the branched amylopectin, containing α-1,4-, and α-1,6-linked glucose units; and amylose, an essentially linear α-1,4-glucan. During biosynthesis starch is deposited as supramolecular, semicrystalline granules (see Vamadevan and Bertolt, 2015) for a recent review on starch). Despite the chemical simplicity of the two polysaccharides, starch biosynthesis is a complicated process that involves several enzyme specificities from the initial precursor formation until completion of the starch granule: ADP-glucose pyrophosphorylase, starch synthase (soluble and granule bound), starch branching enzyme, starch debranching enzyme (isoamylase), and enzymes for phosphorylation and dephosphorylation. Most of the enzymes occur in more than one isoform (see Pfister and Zeeman, 2016; Tetlow and Emes, 2017; Goren et al., 2018) for recent reviews on the starch structure and biosynthesis) showing the same activity, but having different substrate preferences, and catalytic efficiency. In some cases, the variations may be explained by how the enzymes interact with their substrates, which might be nature’s way of fine-tuning the intricate starch granule formation processes. In plantae experiments have shown that a few mutations in one of the enzymes involved in starch biosynthesis can affect the structure and functional properties of the produced polysaccharides. These mutations can be situated far from the active site and may lead to loss of carbohydrate binding ability. Some enzymes possess CBMs, i.e., non-catalytic domains with carbohydrate binding sites which are connected to catalytic modules, sometimes through polypeptide linkers. The CBMs are grouped into families in the Carbohydrate-Active enZYmes (CAZy) database based on the amino acid sequence (Boraston et al., 2004; Lombard et al., 2014). Currently starch binding domains (SBDs) are found in 13 CBM families. Four of these families (CBM20, CBM45, CBM48, and CBM53) are represented in enzymes involved in starch biosynthesis, namely starch synthases, branching enzymes, isoamylases, glucan, water dikinases, and α-glucan phosphatases (Table 1). Surface binding sites (SBSs), capable of interacting with carbohydrates and exposed on the surface of catalytic domains at a certain distance from the active site or on modules intimately associated with catalytic domains (Cuyvers et al., 2011; Cockburn and Svensson, 2013; Cockburn et al., 2014) have also been identified in above mentioned enzymes (Table 1). The present review focuses on enzymes in starch biosynthesis, which have been shown experimentally to possess functional CBMs (SBDs) or SBSs.

STARCH SYNTHASES

Starch synthases (SSs; EC 2.4.1.21) catalyze transfer of glucose from the soluble precursor ADP-glucose to the non-reducing end of an α-1,4-glucan primer or growing chain. SSs belong to glycosyl transferase family 5 (GT5) and contain also a GT1 domain (Pfister and Zeeman, 2016). The GT1 domain is involved in interaction with other proteins (Seung et al., 2017). Five SSs exist: granule-bound SS (GBSS), solely responsible for biosynthesis of amylose (Tetlow and Emes, 2017), and four soluble SSs (SSI, SSII, SSIII, and SSIV) thought to be exclusively involved in amylopectin biosynthesis. In cereals the individual SSs have unique roles, each form predominantly synthesizing chains of different lengths. SSI specifically acts in synthesis of short chains in amylopectin (Delvalle et al., 2005), while SSII and SSIII have a major role in amylopectin synthesis producing chains of short to intermediate length (Zhang et al., 2005; Goren et al., 2018), and SSIV is involved in the initiation of the starch granule formation and control of the number of starch granules in the chloroplast (Roldán et al., 2007; Szydlowski et al., 2009; Crompton-Taylor et al., 2013) (see Sections “Granule Bound SSS - SSIV”below). The different SSs use individual strategies for interacting with starch and related carbohydrates. Some have CBMs, while others have distinct SBSs. Yet other enzymes contain none of these structural features, but interact with specific proteins, which in turn interact with starch.

Granule Bound SSSs

Plant

Two isoforms of GBSSs belonging to GT5 are found in cereals and have no identified CBM or SBS. Barley GBSSI did not bind to starches or polysaccharides sampled in a glucan microarray analysis; the only detected binding being to 3-phosphomaltose (Cuesta-Seijo et al., 2016). Recently, however, it was shown that the localization of GBSSI to starch granules and normal amylose synthesis depended on interaction with another protein, the so-called PROTEIN TARGETING TO STARCH (PTST) 1 that is a non-catalytic protein containing coiled-coils and a CBM48. PTST interacts with the C-terminal GT1 domain of GBSS (Seung et al., 2015). PTST1 was first identified in Arabidopsis, but appears to exist in all plant species (Lohmeier-Vogel et al., 2008; Seung et al., 2015). Two additional plastidial PTSTs (PTST2 and PTST3) were identified in Arabidopsis thaliana leaves. PTST2 is an ortholog to the CBM48-containing FLOURY ENDOSPERM6 (FLO6) from rice (Oryza sativa) endosperm that influences grain starch content, granule morphology, and starch physico-chemical properties. The CBM48 of FLO6 was shown to bind to starch as well as amylopectin and amylose. FLO6 also interacts with the rice debranching enzyme isoamylase 1 that does not bind starch directly (Peng et al., 2014). PTST2 and PTST3 are proposed to interact with SSIV in Arabidopsis leaves (Seung et al., 2017), see Section “SSIV” below.

Cyanobacterium sp. CLg1

Very recently, an SBS was identified in the crystal structure of the GT5 granule bound starch synthase from the Cyanobacterium sp. CLg1 (CLg1GBSS). CLg1GBSS crystallized as a trimer and on molecule B two planar electron densities corresponding to maltose was present out side the active site (Figure 1). At molecule A and C the putative SBS interacts with the His6 purification tag (Nielsen et al., 2018), which poses a steric hindrance for the maltose. However, a mutational
TABLE 1 | Binding data for full-length enzymes and interacting proteins involved in starch biosynthesis.

<table>
<thead>
<tr>
<th>Protein</th>
<th>Organism</th>
<th>Binding site</th>
<th>Substrate</th>
<th>$K_d$</th>
<th>Method</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>GBSSI</td>
<td>Barley</td>
<td>–</td>
<td>3-phosphomaltose*</td>
<td>–</td>
<td>Glucan microarray</td>
<td>Cuesta-Seijo et al., 2016</td>
</tr>
<tr>
<td>FLO6</td>
<td>O. sativa</td>
<td>CBM48</td>
<td>Starch*</td>
<td>–</td>
<td>Pull down assay</td>
<td>Peng et al., 2014</td>
</tr>
<tr>
<td>SSI</td>
<td>H. vulgare</td>
<td>SBS</td>
<td>Maltotriose</td>
<td>1.99 mM</td>
<td>SPR</td>
<td>Wilkens et al., 2014</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>β-cyclodextrin</td>
<td>0.94 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSIIIa</td>
<td>Z. mays L.</td>
<td>–</td>
<td>Amylopectin*</td>
<td>3.1 mg ml$^{-1}$</td>
<td>AGE</td>
<td>Liu et al., 2012</td>
</tr>
<tr>
<td>SSIII</td>
<td>P. vulgaris L.</td>
<td>3 CBM53s</td>
<td>Amylopectin*</td>
<td>0.53 mg ml$^{-1}$</td>
<td>AGE</td>
<td>Senoura et al., 2007</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amylose</td>
<td>0.30 mg ml$^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glycogen</td>
<td>4.04 mg ml$^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Pullulan</td>
<td>6.86 mg ml$^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSIV</td>
<td>H. vulgare</td>
<td>SBS</td>
<td>**</td>
<td>–</td>
<td>–</td>
<td>Cuesta-Seijo et al., 2016</td>
</tr>
<tr>
<td>PTST2</td>
<td>A. thaliana</td>
<td>CBM48</td>
<td>β -cyclodextrin</td>
<td>1–3.3 µM</td>
<td>ITC</td>
<td>Seung et al., 2017</td>
</tr>
<tr>
<td>StBE1</td>
<td>S. tuberosum</td>
<td>–</td>
<td>Maltose</td>
<td>&gt;50 mM</td>
<td>Fluorescence quenching</td>
<td>Mouille, 1996</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Maltotriose</td>
<td>11.7 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glucosyl maltotriose</td>
<td>5.8 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Maltotetraose</td>
<td>1.1 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Maltoctaose</td>
<td>0.75 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Maltohexaose</td>
<td>0.25 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Maltoheptaose</td>
<td>0.16 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DP10 malto-oligo.</td>
<td>0.14 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DP13 malto-oligo.</td>
<td>0.13 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DP21 malto-oligo.</td>
<td>0.18 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>α-cyclodextrin</td>
<td>6.0 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>β-cyclodextrin</td>
<td>0.25 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>γ-cyclodextrin</td>
<td>0.67 µM</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amylopectin</td>
<td>0.066 mg ml$^{-1}$</td>
<td>AGE</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amylose</td>
<td>0.018 mg ml$^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Glycogen</td>
<td>&gt;20 mg ml$^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEX4</td>
<td>A. thaliana</td>
<td>CBM48</td>
<td>Amylopectin*</td>
<td>0.03 mg ml$^{-1}$</td>
<td>AGE</td>
<td>Wilkens et al., 2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amylose</td>
<td>5.42 mg ml$^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>β-cyclodextrin</td>
<td>1.69 mM</td>
<td>SPR</td>
<td></td>
</tr>
<tr>
<td>LSF2</td>
<td>A. thaliana</td>
<td>2 SBSs</td>
<td>Amylopectin</td>
<td>1.59 mg ml$^{-1}$</td>
<td>AGE</td>
<td>Wilkens et al., 2016</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Amylose</td>
<td>0.68 mg ml$^{-1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>β-cyclodextrin</td>
<td>0.72 mM</td>
<td>SPR</td>
<td></td>
</tr>
</tbody>
</table>

$K_d$ values are given in mM. "**" No binding detected on a glucan microarray. "ND" means not detected. Analysis of adsorption is needed to confirm the SBSs impact on CLgGBSS activity.

**Soluble SSs**

SSI
The structure of barley SSI revealed a maltooligosaccharide-binding SBS situated 30 Å from the active site (Figure 2) and which was suggested based on mutational analysis to be involved in recognition of soluble starch and glycogen (Cuesta-Seijo et al., 2013). The SBS binds maltoptaose, -hexaose, -heptaose, and β-cyclodextrin (β-CD) as shown by surface plasmon resonance (SPR) analysis. However, $K_d$ was only measurable for maltoheptaose and β-CD being 1.99 mM and 0.94 mM, respectively. An SBS mutant lost ability to bind β-CD.
and starch binding (Commuri and Keeling, 2001). The native SSI from maize binds amylpectin with a $K_d$ of 2.5 mg/ml (Liu et al., 2012).

**SSII**

SSII is thought to be important for the formation of chains with intermediate lengths (DP12–25), since SSII deficiency in numerous species results in increased frequency of short glucan chains (DP6–DP11) within amylpectin, decreased abundance of DP12 to DP25 chains, and in most instances an elevated amyllose to amylpectin ratio (Zhang et al., 2008). Furthermore, SSIIa in maize plays a crucial role in trafficking SSI and starch branching enzyme IIb (SBEIIb) into the granule matrix (Liu et al., 2012). The sugary-2 mutation in maize (Zea mays L.), which causes major alterations to amylpectin architecture, is due to a catalytically inactive form of the endosperm-specific SSIIa having two amino acid substitutions (one outside the GT domains and one in the GT domains) and being unable to bind amylpectin. Thus, while wild type SSIIa has a $K_d$ of 3.1 mg/ml for amylpectin no binding was observed of the sugary-2 mutant using AGE (Liu et al., 2012).

**SSIII**

SSIIIs contain one or more CBM53s and are the best studied SSs with regard to carbohydrate binding. In the CAZy database (Lombard et al., 2014), 133 protein sequences are classified as CBM53. Most of these are found in SSIIIs, which usually have three CBM53s in tandem and therefore the number of unique proteins containing CBM53 is quite low. Three archaeal Thermococci GH enzymes each with one CBM53 are not classified in CAZy. No structure is available of CBM53 (Lombard et al., 2014). Among the SSIIIs with characterized CBM53s are SSIII from *A. thaliana* (AtSSIII) (Valdez et al., 2008, 2011), two SSIIIs from *Osterococcus tauri* (OtSSIII-A and OtSSIII-B) (Barchiesi et al., 2015), and SSIII from kidney bean (*Phaseolus vulgaris* L.; PSSIII).

AtSSIII has three N-terminal CBM53s in tandem, which preferentially bind to amylose, the first CBM53 being mainly responsible for this selective binding (Valdez et al., 2008, 2011). The second CBM53 has two binding sites, containing Y394 (binding site 1), and W366 (binding site 2), which act cooperatively with the first CBM53 in starch binding. Mutations in these sites affect kinetic parameters of AtSSIII toward a polysaccharide substrate (Wayllace et al., 2010; Valdez et al., 2011). Recently, *in vitro* experiments showed that SBDs of AtSSIII preferentially bind to cell wall polysaccharides over starch. Recombinantly produced AtSSIII comprising the three AtCBM53s thus binds to xylan and pectins with two-fold and to cellulose with 2.4-fold higher affinity than to starch (Grisolia et al., 2017). Amino acid residues at the second CBM53 were found to be important for the binding to plant cell wall polysaccharides, Y394 (binding site 1) being the most critical. Finally, transgenic plants overexpressing the three CBM53s in the cell wall are larger than wild type and have altered cell wall components (Grisolia et al., 2017).

The unicellular green alga, *O. tauri*, has three SSIII isoforms, OtSSIII-A, OtSSIII-B, and OtSSIII-C, which according to the CAZy database contain two, three, and no CBM53, respectively.
Phylogenetic analysis showed that OtSSIII-B is more closely related to higher plant SSIIIs, such as AtSSIII (Barchiesi et al., 2015). The CBM53s from OtSSIII-A and OtSSIII-B were recombinantly produced individually and in tandem and their ability to bind starch, amylose and amylopectin was evaluated. All individual OtCBM53s interacted at different levels with amylose and amylopectin in co-sedimentation assays, except for the first CBM53 from OtSSIII-A that lacked both residues equivalent to those important for binding in the second AtSSIII CBM53 (Barchiesi et al., 2015). The tandem CBM53s from OtSSIII-A were able to bind amylose, but not amylopectin. Furthermore, starch binding was only observed for the first CBM53 of OtSSIII-B and all three OtSSIII-B CBM53s together. The results from co-sedimentation was supported by an adsorption assay using starch, amylose, or amylopectin (Barchiesi et al., 2015).

The third CBM53 from barley (Hordeum vulgare) SSIII (HvSSIII) has been recombinantly produced and shown by glucan microarray analysis to bind to some starches, but not to any of the linear and branched oligosaccharides tested, while strong binding was observed to 3-phosphomaltose, which mimics the 3-phosphorylation of starch found in vivo (Cuesta-Seijo et al., 2016).

Recombinant full-length SSIII from kidney bean (Phaseolus vulgaris L.) (PvSSIII), as well as its N-terminal and catalytic domains on their own, have been characterized with regard to enzymatic activity and binding to amylose and amylopectin (Senoura et al., 2007). The N-terminal region of PvSSIII is predicted to contain three CBM53s like other characterized SSIIIs. These N-terminal CBM53s were not essential for catalysis and had moderate effect on thermostability and pH stability. However, deletion of the CBM53s drastically decreased affinity and catalytic efficiency on glucan primers (Senoura et al., 2007). AGE indicated that N-terminal CBM53s have high affinity for amylose and amylopectin. \( K_d \) values for binding amylose, amylopectin, and pullulan, were similar for full-length \( P\varepsilon\)SSIII and N-terminal CBM53s alone indicating the importance of CBM53s in polysaccharide binding. \( K_d \) was in the range from 0.30 to 6.88 mg/ml for the four polysaccharides (Senoura et al., 2007).

**SSIV**

Interestingly, barley SSIV did not bind to any of the starches and polysaccharides included in a glucan microarray (Cuesta-Seijo et al., 2016). However, the non-catalytic, CBM48-containing PTST2 and PTST3 proteins are proposed to interact with SSIV in Arabidopsis leaves, and play a critical role in starch granule initiation by delivering suitable glucan primers to SSIV. PTST2 was shown by isothermal titration calorimetry (ITC) analysis to interact with \( \beta\) -CD (\( K_d \) ranged from 1 to 3.3 \( \mu \)M), and the affinity of CBM48 alone with \( \beta\) -CD was very similar (\( K_d \) ranged from 1.7 to 4 \( \mu \)M). Injection of \( \beta\) -CD into an equimolar mixture of PTST2 and maltodecaose resulted in similar heat changes as when PTST2 was analyzed alone, indicating that maltodecaose and \( \beta\) -CD occupy different binding sites. On the other hand, when the experiment was repeated with maltodecaose together with PTST2, significant variations were observed, while the \( K_d \) was in the same range as for PTST2 alone. This suggested that the ability of maltodecaose to adopt a structure mimicking helical amylose chains is important for binding long maltooligosaccharides to CBM48 of PTST2 (Seung et al., 2017). Very recently a crystal structure was published of GT5 A. thaliana starch synthase IV (AtSSIV) in complex with the inhibitor acarbose, which showed an SBS (Figure 3) located in the same area as the SBS on the GT5 CLgIGBSS (Figure 1). AtSSIV crystallized as a dimer and the SBS was only occupied on molecule A and electron density was only present for a part of acarbose. Crystal contacts of molecule B blocked for glucan interactions (Nielsen et al., 2018). Unfortunately, a mutational analysis was not performed, so the role of the SBS remains enigmatic.

**STARCH BRANCHING ENZYMES**

Starch branching enzymes (SBEs; EC 2.4.1.18) catalyze formation of \( \alpha\)-1,6-glucosidic linkages creating branch points in amylopectin and are crucial in determining structural and physical properties of starch granules (see (Tetlow and Emes, 2014) for a review on SBEs and their role in starch biosynthesis). SBEs catalyze transglycosylation reactions involving cleavage of internal \( \alpha\)-1,4-glucosidic linkages, and transfer of products from the substrate chain toward the non-reducing end to C-6 hydroxyl groups. Multiple forms of SBEs are required for starch biosynthesis in plants and algae, unlike in glycogen synthesizing
prokaryotes and eukaryotes, which have a single BE. The SBE isoforms show wide variation in chain-length transfer pattern related to their glucan substrate preferences (Tetlow and Emes, 2014). Starch BEs belong to GH13 subfamilies 8 (eukaryotic BEs; GH13_8) and 9 (bacterial BEs; GH13_9) (Stam et al., 2006); glycogen BEs are also found in GH57 (Janeček et al., 2014). GH13 BEs according to CAZy are multimodular enzymes generally including an N-terminal CBM48 (Lombard et al., 2014). Three-dimensional structures of BEs revealed the presence of both CBM48 binding sites and SBSs.

**Potato Tuber Starch BE1**

A thorough analysis of StBE1 binding to a wide range of maltooligosaccharides by using tryptophan fluorescence quenching showed increasing affinity from maltose ($K_d > 50$ mM) to a maltooligosaccharide with degree of polymerization (DP) 13 ($K_d$ of 0.13 mM) (Table 1). Notably, $K_d$ was 0.18 mM for a maltodextrin of DP 21. It was demonstrated that maltooligosaccharides of DP below 20 were poor substrates. Binding of amylose and amylopectin yielded $K_d$ of 0.018 and 0.066 mg ml$^{-1}$, respectively, indicating that StBE1 has a substrate preference for unbranched chains. Ten-fold lower affinity for maltooligosaccharides was obtained by AGE in competition with amylose and amylopectin than by using tryptophan fluorescence quenching, which suggests the presence of a tryptophan-containing binding site different from the active site (Blennow et al., 1998).

AGE binding analysis of amylose and amylopectin to trypsin-digested StBE1 suggested that the C-terminal part is not important for starch binding (Blennow et al., 1998). This is surprising, because several SBSs as mentioned below are found in this region of other characterized BE1s.

**Cyanobacterial BE1**

Crystal structures of a BE from *Cyanothece* sp. ATCC 51142 (CyBE) in complex with maltohexaose and maltoheptaose were published recently (Hayashi et al., 2017). Unlike most cyanobacteria, *Cyanothece* sp. ATCC 51142 produces an amylopectin-like polysaccharide designated as cyanobacterial starch (Hayashi et al., 2017). Glycogen producing cyanobacteria usually have one GH13_9 BE and one GH57 BE. By contrast *Cyanothece* sp. ATCC 51142 contains three GH13_9 BEs and one GH57 BE (Suzuki et al., 2013; Suzuki and Suzuki, 2016). The three GH13_9 BE isoforms have been characterized and CyBE1 and CyBE2 transferred short glucans (DP 6–7), while CyBE3 transfers short as well as long glucan chains (DP 30) (Suzuki et al., 2015). The chain length preferences of CyBE1, CyBE2 and the SBE2b isoform of rice (*O. sativa* L.) were similar, while the CyBE3 specificity was similar to that of rice BE1 (Suzuki et al., 2015).

The crystal structure of CyBE1 with maltoheptaose at the active site revealed seven additional binding sites (Figure 4A; Hayashi et al., 2017). The architecture of CyBE1 was unprecedented among BEs and consists of domain N, CBM48, and the catalytic domain common for all GH13s known as domain A, and domain C (Figure 4A). Two binding sites were situated on CBM48 and the domain C contained three SBSs (Figure 4A). The functional role of these five binding sites was not investigated, but mutation of two SBSs on domain A (A1 and A2) (Figure 4A) resulted in up to 50% loss of activity with unchanged chain length profile for A1 mutants. By contrast A2 mutants showed changes in the relative proportions of the products and up to 90% decrease in activity. On the basis of these results, the maltoheptaose orientation at A1 and A2, as well as the A1 and A2 locations, the authors suggested that A1 functions as entrance for the $\alpha$-glucan chain to the acceptor binding site where the new branch is transferred to the chain coming in via A1. A2 functions as an exit for the chain coming in via A1 and is also suggested to be responsible for discriminating the length of the incoming chains (Hayashi et al., 2017).
Rice BE1
Excluding the N-domain, GH13_9 BE1 from rice (O. sativa; OsBE1) has a similar domain architecture to CyBE1 (Figure 4B; Hayashi et al., 2017). The OsBE1 complex structure with maltopentaose reveals two binding sites on the CBM48 (Figure 4B), which are also occupied in CyBE1 (Chaen et al., 2012; Hayashi et al., 2017). In addition, an SBS was observed on domain A (Figure 4B). A large number of residues interacting with maltopentaoses are conserved (Chaen et al., 2012), suggesting that this SBS has a functional role. However, mutational experiments are needed to describe the various functional roles.

Escherichia coli BE1
GH13_9 BE1 from Escherichia coli (EcBE1) resembles CyBE1 in domain architecture, but does not contain the N-domain (Figure 4C; Hayashi et al., 2017). EcBE1 has been crystallized in complex with α-, β-, and γ-cyclodextrins, maltohexaose and maltotetraose, but not all ligands are seen to bind to all binding sites in EcBE1. Further, EcBE1 crystallized as a tetramer and the binding sites are not occupied on all four EcBE1 monomers (Figure 4C). Of the seven EcBE1 binding sites, binding site 1 is at a distance of approximately 18 Å from and the closest to the active site (Feng et al., 2015, 2016).

A binding site corresponding to CBM48_1 of CyBE1 was observed for EcBE1 (binding site IV), while the remaining 6 sites are SBSs some of which (Figure 4C) are corresponding to A1, A2 and C1 on CyBE1 (Feng et al., 2015; Hayashi et al., 2017). In addition, binding site I, occupied by linear maltooligosaccharides (Feng et al., 2015), and binding site VII, only occupied by α- and γ-CDs (Feng et al., 2016), were only seen in EcBE1. For linear maltooligosaccharides to bind at binding site VII they would have to adopt a conformation similar to the CDs. Notably, CDs were only observed at binding sites IV–VI, which suggests that binding sites I–III are unable to accommodate the curved ligands (Feng et al., 2016). Binding sites I and II (Figure 4C) apparently accommodate maltotetraose and maltose, respectively, but their orientation suggests that they together represent a maltotetraose molecule. Binding sites III and VI are located close to where glucan chains typically exit GH13 enzymes. Mutations in binding sites I, II, VI resulted in up to 84, 24, and 86% reduced activity, respectively (Feng et al., 2015), which clearly demonstrate their importance for the function of EcBE1. Mutation of binding site VII led to loss of up to 92% activity (Feng et al., 2016).

EcBE1, similarly to potato (Solanum tuberosum) tuber starch BE1 (SBE1) described above (Blennow et al., 1998), transfers maltohexaose and larger maltooligosaccharides to polymeric substrates (Binderup et al., 2000). The EcBE1 active site is unoccupied in all published complex structures (Feng et al., 2015, 2016), and considered to have poor affinity for shorter linear and circular maltooligosaccharides. This probably prevents EcBE1 from transferring short branches onto relatively short polymers. If this is true, the SBSs and the CBM48 binding site seem to have distinct roles. Binding sites I, II, and IV (on CBM48) could bind the glucan chain entering the active site, and III and VI could bind the glucan chain exiting the active site (Feng et al., 2015).

STARCH DEBRANCHING ENZYMES
The starch debranching enzymes (DBEs) hydrolyze α-1,6-glycosidic linkages at branch points in amylpectin and oligosaccharides derived thereof. DBEs play a role during amylpectin biosynthesis and degradation. Two types of starch DBEs are found in plants, the isoamylase-type (EC 3.2.1.68) and the pullulanase-type (EC 3.2.1.41). While at least three isoforms exist of the isoamylase-type, the pullulanase-type occurs as a single form, known as limit dextrinase. Both belong to GH13 but are categorized into different subfamilies with distinct specificities; isoamylases are found in GH13 subfamily 11 (GH13_11), and pullulanase-type DBEs are found in GH13 subfamilies 12–14 (GH13_12–14) with the plant pullulanas being found solely in GH13 subfamily 13 (Møller et al., 2016). Isoamylases were hypothesized to trim misplaced branches in amylpectin, which otherwise will prevent adjacent linear chains from associating and crystallizing during the biosynthesis (Myers et al., 2000; Nakamura, 2002; Jeon et al., 2010; Gore et al., 2018). The function of pullulanase-type DBE during starch synthesis is less understood (Li et al., 2017). However, besides during germination, substantial pullulanase activity has been detected in developing rice and maize endosperms (Nakamura, 1996; Beatty et al., 1999), and the presence of mRNA for barley limit dextrinase was found in this stage of the plant lifecycle (Burton et al., 1999). A recent study has shown that variations in the gene encoding the pullulanase type DBE in Sorghum bicolor result in starch with better digestibility (Gilding et al., 2013).

Plant DBEs are multimodular possessing at least a CBM48 in addition to the catalytic domain A and the C-domain typical of GH13. In most CBM48s from GH13_11–14 DBEs, residues predicted to constitute a canonical SBS are not fully conserved. Carbohydrate binding to CBM48 has not been demonstrated experimentally for isoamylase- and pullulanase-type DBEs (Malle et al., 2006; Mikami et al., 2006; Gourlay et al., 2009; Vester-Christensen et al., 2010; Lamerts van Bueren et al., 2011; Møller et al., 2015), but CBM48s from these enzymes possess a conserved tryptophan corresponding to Trp563 in SBS2 of a closely related CBM20 in Aspergillus niger glucoamylase, suggesting that plant DBE CBM48s may be intermediates between CBM20 and CBM48 (Janeček et al., 2011; Møller et al., 2016).

Rice Isoamylase
Rice (O. sativa) isoamylase 1 (OsISA1) does not bind starch directly, but it can interact both in vivo and in vitro with the CBM48-containing FLO6 protein (see also Section “Plant”). Hence, FLO6 might assist binding of OsISA1 to starch (Peng et al., 2014). Recently, a point mutation in the barley gene Fra corresponding to FLO6 was shown to cause fractured starch granules (Saito et al., 2018). These CBM48 containing proteins could be a way of the isoamyloses to overcome the lack of a functional SBD.

Chlamydomonas reinhardtii Isoamylase
The green alga Chlamydomonas reinhardtii is a model for studying starch synthesis, and two C. reinhardtii isoamyloses have been shown to be important for the starch synthesis
Two enzyme classes, Starch phosphorylation can stimulate and hence perhaps be essential for starch biosynthesis (Hejazi et al., 2014; Santelia et al., 2011). However, much remains to be learned about the exact metabolic role of starch phosphorylation (Mahlow et al., 2016; Xu et al., 2017). Two enzyme classes, C. reinhardtii isoamylase 1 consists of the GH13 catalytic domain, a CBM48, and a characteristic GH13 C-terminal domain. The structure of the GH13 maltoheptaose-complexed C. reinhardtii isoamylase 1 showed carbohydrates occupying an SBS situated at the reducing-end binding area of the active site as well as a second SBS at the interface of the catalytic and C-terminal domains (Figure 5). However, the possible function of these SBSs has not been investigated (Sim et al., 2014).

**α-GLUCAN, WATER DIKINASES AND alpha-GLUCAN PHOSPHATASES**

Starch phosphorylation can stimulate and hence perhaps be essential for starch biosynthesis (Hejazi et al., 2014; Skeffington et al., 2014). However, much remains to be learned about the exact metabolic role of starch phosphorylation (Mahlow et al., 2016; Xu et al., 2017). Two enzyme classes, α-glucan, water dikinases (EC 2.7.9.4) and α-glucan phosphatases (EC 3.1.3.48) are responsible for starch phosphorylation and dephosphorylation, respectively (Zeeman et al., 2010). The α-glucan, water dikinase 1 (GWD1) and GWD3 phosphorylate starch at the C6 and C3 position, respectively, and the two phosphatases, Starch Excess 4 (SEX4) and Like SEX4 2 (LSF2) dephosphorylates starch at the C6 and C3 position, respectively (Hejazi et al., 2010; Santelia et al., 2011).

**GWD1**

GWD1 from *S. tuberosum* (StGWD1) has a C-terminal catalytic domain and two CBM45s: CBM45-1 located at the N-terminal next to the chloroplast transit peptide and CBM45-2 at the center of the sequence, and large areas of the sequence were unmapped (Glarling et al., 2011). Despite the plethora of sequence data available since 2011, a new search against the Conserved Domains Database (Marchler-Bauer et al., 2017) did not reveal additional domains in the unmapped regions.

Each of the StGWD1 CBM45-1 and CBM45-2 was produced recombinantly, but only CBM45-2 was stable in solution and only at pH 8. SPR and ITC showed low affinity of CBM45-2 for the starch model β-CD (SPR: $K_d = 0.38$ mM; ITC: $K_d = 0.68$ mM). Such low affinity may be a prerequisite for a dynamic interaction of GWD1 with the starch granule facilitating a necessary tight control of the starch metabolism (Glarling et al., 2011). StGWD1 CBM45-1 was prepared by trypsin hydrolysis of the full-length enzyme and showed relatively weak affinity, yielding $K_d$ of 7.2 mg ml$^{-1}$ for starch granules in an adsorption assay and $K_d$ of 1.2 mg ml$^{-1}$ for soluble starch by using AGE (Mikkelsen et al., 2006). Mutations in the binding site of CBM45-1 of full-length StGWD1 resulted in complete loss of binding for these starch substrates demonstrating its pivotal role in enzyme function (Mikkelsen et al., 2006).

**GWD3**

GWD3 from *A. thaliana* (AtGWD3) has a single CBM20 appended at the N-terminus and a C-terminal GWD catalytic domain, while the function of the ~700 amino acid residues long segment in between these two domains is unknown (Janeček et al., 2011). As for GWD1, search against the Conserved Domains Database (Marchler-Bauer et al., 2017) failed to identify any additional domains. Most CBM20s have two binding sites (Janeček et al., 2011), but presumably only binding site 1 is functional in AtGWD3 CBM20, which lacks residues typically for binding site 2 (Christiansen et al., 2009a).

Compared to CBM20s from amylolytic enzymes having $K_d$ for β-CD in the μM-range recombinant AtGWD3 CBM20 showed unusually low affinity for α-, β-, and γ-CDs ($K_d$ of 0.22–0.84 mM) as determined by SPR (Christiansen et al., 2009a). As for CBM45-2 of StGWD1, this may allow dynamic binding of AtGWD3 CBM20 facilitating GWD3 regulation (Christiansen et al., 2009b). The binding of α-, β-, and γ-CD by AtGWD3 CBM20 depended on pH – e.g., $K_d$ for γ-CD was 0.84 and 5.56 mM at pH 5.5 and pH 9, respectively. However, for β-CD the observed affinity increased from $K_d$ of 1.09 mM at pH 6 to 0.53 mM at pH 9, and this pH dependence was suggested to be related to the plants physiological needs as the binding would be stronger during the day when stromal pH increases (Christiansen et al., 2009a).

**Starch Excess 4**

The crystal structure of *A. thaliana* SEX4 contains a single maltoheptaose molecule that spans the extended binding pocket at the interface of the catalytic dual-specificity phosphatase (DSP) and CBM48 domains (Figure 6A; Meekins et al., 2014). $K_d$ of *A. thaliana* SEX4 for amylose and amylopectin was determined by AGE to be 5.42 and 0.03 mg ml$^{-1}$, respectively, in agreement with the preference of SEX4 for double helical α-1,4-glycosidic linkages (Santelia et al., 2011). SPR analysis gave $K_d$ for β-CD of 1.69 mM and a stoichiometry of 2.89, indicating three $K_d$ for α-, β-, and γ-CD binding sites on SEX4, although this remains to be confirmed by structural analysis (Wilkins et al., 2016). Mutation of residues responsible for binding in both domains resulted in 10–80% loss of activity toward para-nitrophenyl phosphate.
Like SEX4 2

Three distinct maltohexaose binding sites are seen in a crystal structure of A. thaliana Like SEX4 2 (LSF2) (Figure 6B). Of these, two SBSs, SBS1 and SBS2, are situated > 20 Å away from the active site which accommodates the third maltohexose molecule (Meekins et al., 2013). Using AGE A. thaliana LSF2 was shown to have $K_d$ of 0.68 and 1.59 mg ml$^{-1}$ for amylose and amylpectin, respectively, emphasizing its preference for the single chain helical conformation in amylose (Wilkens et al., 2016). LSF2 specifically hydrolyses C3 phosphate ester groups in starch (Santelia et al., 2011), which are hypothesized to disrupt the double helical conformation in amylpectin (Blennow and Engelsen, 2010). The higher affinity of LSF2 for amylose thus mimics its natural substrate preference (Wilkens et al., 2016). Unexpectedly, despite the long distance to the active site, SBS1 and SBS2 mutations resulted in 50 and 88% loss of activity for para-nitrophenyl phosphate, respectively (Meekins et al., 2013). Remarkably, however, molecular dynamics simulation showed that binding at the two SBSs affected the active site thus supporting the role of SBSs in activity. Mutation at all three binding sites reduced affinity for amylose below a measurable level. Still, comparison of the mobility of SBSs mutants in AGE pointed to their involvement in starch binding. As SBS2 seemed to have higher affinity for amylose than SBS1, it was suggested to be responsible for the initial starch recognition (Wilkens et al., 2016).

CONCLUSION AND LOOKING AHEAD

As the presence of either CBMs or SBSs in several cases has been crucial for correct starch biosynthesis, the recent focus on non-catalytic sites can potentially open new promising routes to obtain useful modified starches. Current insights on CBM53s from SSIIIs and their interaction with starch may be useful for manipulating catalytic efficiency of SSIII by altering the active site structure without modifying the catalytic module. Determination of the structure of a CBM53 would take the knowledge to the next level and improve understanding of the protein–carbohydrate interactions, maybe to deduce a mechanism for how the different CBM53s can alter activity of SSIII (Wayllace et al., 2010).

SBSs in BE1s appear to define the length of transferred branch chains (Blennow et al., 1998; Feng et al., 2015; Hayashi et al., 2017), which could potentially constitute a target for rational protein engineering. Recently, the non-catalytic proteins PTST2 and FLO6 gained attention as targets for obtaining modified starches. The PTST gene has been demonstrated to be essential for amylose synthesis in A. thaliana leaves and could be useful in obtaining amylose-free starches, which find extensive applications in food and non-food industrial products (Santelia and Zeeman, 2011; Seung et al., 2015). CBM48 of PTST2 (FLO6) was suggested as a potential target for biotechnological modification of starch, particularly for modifying granule size. A pst2 A. thaliana mutant had larger starch granules, but with similar morphology to wild type granules, while much smaller granules were observed in the pst2 overexpression lines (Seung et al., 2017). Furthermore, a knock-out of FLO6 in rice was shown to affect starch synthesis (Peng et al., 2014).

Notably, these various examples evidently demonstrate that properties of starch can be modified by altering the enzyme–substrate interaction during starch biosynthesis. Yet, the impact of especially SBSs on the function of the starch biosynthetic enzymes is still to be further explored. However, the challenge is that they are not easily recognized based on bioinformatics analysis, like the carbohydrate binding sites of CBMs.
AUTHOR CONTRIBUTIONS

MM, CW, and BS conceptualized the manuscript, which was written by CW and MM with invaluable input from BS.

FUNDING

This work was supported by Independent Research Fund Denmark | Natural Sciences (FNU) by grants to the projects ‘Discovery and Characterization of Carbohydrate SBS in Polysaccharide Converting Enzymes’ (Grant no. 09-027151) and ‘Roles and Structural Determinants of Low-affinity Carbohydrate-Protein Interactions’ (Grant No. 6108-00476B).

ACKNOWLEDGMENTS

Andreas Blennow was thanked for discussions on α-glucan, water dikinases, and α-glucan phosphatases in starch biosynthesis.

REFERENCES


Funding

This work was supported by Independent Research Fund Denmark | Natural Sciences (FNU) by grants to the projects ‘Discovery and Characterization of Carbohydrate SBS in Polysaccharide Converting Enzymes’ (Grant no. 09-027151) and ‘Roles and Structural Determinants of Low-affinity Carbohydrate-Protein Interactions’ (Grant No. 6108-00476B).


Suzuki, E., Onoda, M., Colleoni, C., Ball, S., Fujita, N., and Nakamura, Y. (2013). Physicochemical variation of cyanobacterial starch, the insoluble
Wilkens et al. Carbohydrate Binding in Starch Biosynthesis


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Wilkens, Svensson and Møller. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.