Starch-hydrolyzing enzymes lacking α-glucan-specific carbohydrate-binding modules (CBMs) typically have lowered activity on granular starch relative to their counterparts with CBMs. Thus, consideration of starch recognition by CBMs is a key factor in understanding granular starch hydrolysis. To this end, we have dissected the modular structure of the maltohexaose-forming amylase from Bacillus halodurans (C-125). This five-module protein comprises an N-terminal family 13 catalytic module followed in order by two modules of unknown function, a family 26 CBM (BhCBM26), and a family 25 CBM (BhCBM25). Here we present a comprehensive structure-function analysis of starch and α-glucooligosaccharide recognition by BhCBM25 and BhCBM26 using UV methods, isothermal titration calorimetry, and x-ray crystallography. The results reveal that the two CBMs bind α-glucooligosaccharides, particularly those containing α-1,6 linkages, with different affinities but have similar abilities to bind granular starch. Notably, these CBMs appear to recognize the same binding sites in granular starch. The enhanced affinity of the tandem CBMs for granular starch is suggested to be the main biological advantage for this enzyme to contain two CBMs. Structural studies of the native and ligand-bound forms of BhCBM25 and BhCBM26 show a structurally conserved mode of ligand recognition but through non-sequence-conserved residues. Comparison of these CBM structures with other starch-specific CBM structures reveals a generally conserved mode of starch recognition.

It is well established that carbohydrates play vital roles in numerous biological settings. They can act as the carriers of biological information, such as in cell development, carcinogenesis, immune response, and cell trafficking; as structural macromolecules, such as the cellulose of plant cell walls or chitin of insect exoskeletons; or as an energy source. Poly- saccharides, which are highly polymerized molecules, can perform any of these functions. As mentioned, cellulose and chitin, the first and second most abundant biopolymers on earth, respectively, are the premier structural carbohydrates. Glycosaminoglycans are a highly complex class of polysaccharides that make up the extracellular glue of mammalian tissues and function both in structural and information content roles. Glycogen and starch are related polysaccharides that function as the primary storage carbohydrates in animals and plants, respectively.

Glycogen is a polymer of glucose comprising linear α-1,4-linked glucose with α-1,6 branch points occurring approximately every 8–12 glucose residues. Amylopectin, a component of starch, is a glycogen-like molecule but with α-1,6 branch points occurring approximately every 24–30 glucose residues. Amylose, the other component of starch, is a polymer of pure linear α-1,4-linked glucose. The α-1,4 linkages in these polysaccharides make them fold into tight helical structures, resulting in dense granules that function as highly effective storage systems. The release of small metabolizable sugars from these polysaccharides is achieved by the action of glycoside hydrolases falling primarily into families 13 and 15 of the glycoside hydrolase classification (on the World Wide Web, see afmb.cnr-s.mrs.fr/CAZY/). A remarkable feature of these enzymes is their modularity. Many of these enzymes have complex multimodular architectures comprising at least one catalytic module and often several ancillary modules. The most common class of ancillary module is the carbohydrate-binding modules (CBMs)2 (1), which are 50–200-amino acid modules found within the modular structures of carbohydrate-active enzymes. CBMs exert their biological activity by maintaining the interaction of the enzyme with its insoluble substrate (2). This proximity/targeting effect raises the local concentration of enzyme, thereby increasing the rate of substrate hydrolysis. In the case of insoluble polysaccharide hydrolysis, including granular starch hydrolysis (3), this function of CBMs has often been observed to be critical (4–6).

CBMs are currently classified into 43 primary structure-based families (on the World Wide Web, see afmb.cnr-s.mrs.fr/CAZY/), of which there are currently only six classified families of starch-binding CBMs: CBM families 20, 21, 25, 26, 34, and 41. The starch binding properties have only been thoroughly characterized for two starch-specific CBMs. The stoichiometry and thermodynamics of α-glucan binding as well as the influence of binding on the tertiary structure of starch have been examined for the family 20 CBM from the Aspergillus niger glucoamylase. NMR and thermodynamic studies by isothermal titration calorimetry (ITC) indicated that this module has two binding sites with approximately similar affinities for maltohexaosaccharides (ranging from $6 \times 10^2$ M$^{-1}$ for a triose to $3 \times 10^4$ M$^{-1}$ for a pentose) (7, 8). Remarkably, studies have shown that the concerted action of these two binding sites results in the disruption of amylase structure (9, 10). In contrast, the family 41 CBM from Thermotoga maritima Pul13 had only

[1] Holder of a Canada Research Chair in Molecular Interactions. To whom correspondence should be addressed: Biochemistry & Microbiology, University of Victoria, P.O. Box 3055 STN CSC, Victoria, British Columbia V8W 3P6, Canada. Tel.: 250-721-8855; E-mail: boraston@uvic.ca.

2 The abbreviations used are: CBM, carbohydrate-binding module; BhCBM25 and BhCBM26, family 25 and 26 CBMs, respectively, from a B. halodurans maltohexaose-forming amylase; ITC, isothermal titration calorimetry; OG-CBM, Oregon Green$^{1}$-labeled CBM; MES, 4-morpholineethanesulfonic acid; SAD, single-anomalous dispersion; GM3, 6'-α-o-glucosyl-maltotriose; GM3M3, 6'-α-o-glucosyl-maltotriose-maltotriose.
a single carbohydrate binding site on the protein, and the affinity of this module for maltooligosaccharides was substantially higher, in the range of 10^6 M^-1 (11).

Currently, of the starch-binding CBMs, only families 20 and 34 have representative structures. The solution NMR structure of the A. niger CBM20 was the first to be determined (12). Subsequently, numerous CBM20s, many of these from Bacillus sp., have been solved by x-ray crystallography as part of the complete enzyme structures. The first family 34 CBM to have its structure solved by x-ray crystallography was the N-terminal CBM from Thermotoga maritima with its structure solved by x-ray crystallography in complex with soluble maltooligosaccharides. Occasionally, in these structures, sugars have been observed to be associated with modules not classified as CBMs. Although these may indeed be starch-specific CBMs, they remain unclassified until additional methods are used to demonstrate carbohydrate binding.

The complex architecture of plant cells is elegantly reflected in the diversity of CBMs that recognize the individual polysaccharides and even subcomponents of the polysaccharides. Although starch does not have the diversity of composition that plant cell walls do, it is, nonetheless, a structurally complex polysaccharide whose recognition by CBMs in a broad sense is poorly studied. To better understand this phenomenon, we have engaged in structure-function studies of additional starch-specific CBMs from families 25 and 26, respectively, from the Bacillus halodurans (C-125) maltodextrinase-forming amylase with granular starch and soluble oligosaccharides. The three-dimensional structures of these CBMs solved by x-ray crystallography in complex with soluble maltooligosaccharides provide insight into the molecular determinants of these CBM-starch interactions.

**MATERIALS AND METHODS**

**Carbohydrates and Polysaccharides—**All carbohydrates and polysaccharides used in this study were purchased from Sigma. The granular starch discussed throughout was from corn.

**Cloning—**The DNA fragments encoding the desired modules (see Fig. 1) of the maltodextrinase-forming amylase gene (see GenBank accession number AP001508; open reading frame BH0413) were amplified by PCR from B. halodurans (C-125) genomic DNA (ATCC BAA-125) using previously described methods (15). The oligonucleotide primers used are shown in Table 1. The catalytic module, GH13CM, was amplified with the oligonucleotide pair GH13F and GH13R; the X23-1 module was amplified with X23AF and X23AR; the X23-2 module was amplified with X23BF and X23BR; the BhCBM26 module was amplified with CBM26F and CBM26R; and the BhCBM25 module was amplified with CBM25F and CBM25R. The BhCBM26/25 tandem was amplified with CBM26F and CBM25R. The PCR-amplified products were ligated into the pET-150 TOPO Directional Cloning kit (Invitrogen) to generate pCM, pX1, pX2, pCBM26, pCBM25, and pCBM26/25, respectively. The encoded polypeptides contained an N-terminal His6/V5 epitope tag and a TEV protease cleavage site.

The BhCBM26 module and the BhCBM25 module were subsequently reamplified with the primer pairs CBM26NF/CBM26HR and CBM25NF/CBM25HR and cloned into NheI- and HindIII-digested PET28a to give pET28-BhCBM26 and pET28-BhCBM25, respectively. The encoded polypeptide contained an N-terminal His6 fused to the polypeptide via a thrombin cleavage sequence. These constructs were made in order to obtain recombinant protein production levels sufficient to purify protein for calorimetric and structural studies.

**Protein Production and Purification—**Protein production and purification procedures were identical for pET-150 and pET-28 constructs, except that growth medium used for the latter was supplemented with 50 mg/liter kanamycin rather than 100 mg/liter ampicillin. All protein production was done in Escherichia coli BL21 STAR (DE3). Cultures were grown in LB medium supplemented with antibiotic to an approximate A_600 of 0.8, and expression was induced by the addition of isopropyl-1-thio-β-D-galactopyranoside to 0.3 mM. Cells harvested by centrifugation were ruptured by passage through a French pressure cell. Recombinant protein was purified from clarified extracts by immobilized metal ion affinity chromatography as described previously (16). Purified polypeptides were concentrated and exchanged into the appropriate buffer in a stirred ultrafiltration unit (Amicon, Beverly, MA) on a
5,000 molecular weight cut-off membrane (Filtron, Northborough, MA). Purity, assessed by SDS-PAGE, was greater than 95%.

**Protein Concentration Determination**—The concentrations of purified proteins were determined by UV absorbance (280 nm) using calculated molar extinction coefficients (17).

**Granular Starch Binding Studies**—Adsorption isotherms using granular corn starch (at 10 mg/ml) were obtained and analyzed using methods described previously for cellulose (19) with the following modifications. A constant concentration of Oregon Green labeled CBM (total OG-CBM was measured in control samples, where the volume of starch suspension added was replaced by buffer). Samples assessing competition-displacement contained OG-CBM and one of the following competitors: 250 μM bovine serum albumin, 250 μM BhCBM25 (unlabeled), 300 μM BhCBM26 (unlabeled), or 500 μM maltotriose. Displacement and, therefore, binding competition was determined by the relative amount of bound OG-CBM in the absence of competitor with the relative amount of bound OG-CBM in the presence of competitor. All experiments were performed in duplicate.

**UV Difference Titrations**—Automated UV difference titrations were performed as described previously (11). Difference spectra were examined for peak and trough wavelengths, and values at the appropriate wavelengths were extracted for further analysis. The peak-to-trough heights at the wavelength pairs 292.8/288.4 nm, 288.4/285 nm, and 285/277.6 nm for BhCBM25 and 294/288.8 nm, 288.8/285.8 nm, and 285.8/279 nm for BhCBM26 were calculated by subtraction of the trough values from the peak values and the dilution-corrected data plotted against total carbohydrate concentration. Data for the three wavelength pairs were analyzed simultaneously with MicroCal Origin (version 7.0) using a one-site binding model accounting for ligand depletion (20). Experiments were performed at 20 °C in 50 mM Tris, pH 7.5. The data reported are the averages and S.D. values of three independent titrations.

**Isothermal Titrination Calorimetry**—ITC was performed as described previously (16), using a VP-ITC (MicroCal, Northampton, MA) in 50 mM Tris, pH 7.5, at 25 °C using 100–450 μM BhCBM25 or BhCBM26 in the reaction cell, which gave C values (21) of >5, and 1–5 μM oligosaccharide in the syringe. Protein samples were extensively dialyzed against the buffer. Sugar solutions were prepared by mass in buffer saved from the final protein dialysis step. Both protein and sugar solutions were filtered and degassed immediately prior to use. Protein concentrations were determined by UV absorbance as described above. All data show the average and S.D. of three independent titrations.

**Crystallography**—

-BhCBM25 and BhCBM26—5,000 molecular weight cut-off membrane. All crystallizations were done by the hanging drop vapor diffusion method at 18 °C. All computing was done using the CCP4 suite (22) unless otherwise stated. All crystals were frozen at 113 K after a short soak in artificial mother liquor supplemented with ethylene glycol at 20–25% (v/v). Diffraction data were collected with a Rigaku R-AXIS IV++ area detector coupled to an MM-002 x-ray generator with Osmic “blue” optics and an Oxford Cryostream 700. Data were processed using the Crystal Clear/d’trek (23) software provided with the instrument. Data collection and processing statistics are given in Table 2.

-BhCBM25 (30 mg/ml) was iodinated by overnight treatment with 10 mM N-iodosuccinimide in 100 mM sodium acetate, pH 3.75, at room temperature. The reaction was centrifuged to eliminate precipitated material, and excess N-iodosuccinimide was removed by gel filtration on Sephadex G-25 pre-equilibrated with water. Iodinated BhCBM25 at 24 mg/ml was crystallized in 0.1 M sodium citrate, pH 5.6, 10% isopropanol alcohol, and 20% polyethylene glycol 4000. Unmodified BhCBM25 at 30 mg/ml, preincubated with a 20-fold molar excess of maltotetraose, was crystallized in 0.1 M sodium acetate, pH 5.5, and 30% polyethylene glycol 4000. BhCBM25 at 15 mg/ml preincubated with an equimolar amount of maltotetraose crystallized in 0.1 M MES, pH 6.5, 0.2 M ammonium sulfate, and 30% polyethylene glycol 5000 monomethyl ether.

-BhCBM26 at 25 mg/ml preincubated with a 20-fold molar excess of maltose crystallized in 0.1 M sodium acetate, pH 4.6, 0.2 mM ammonium sulfate, and 30% polyethylene glycol 1500. A BhCBM26 complex with cadmium was obtained in 0.1 M HEPES, pH 7.5, 0.01 M CdSO₄ and 0.6 mM sodium acetate using protein at 5 mg/ml.

**Phasing, Model Building, and Refinement**—The structures of BhCBM25 and BhCBM26 were solved by single-anomalous dispersion (SAD) using iodinated and Cd²⁺ derivatives, respectively. Heavy atom sites in these derivatives were found using ShelixD (24) with data extending to 3.0 Å. Two strong iodine sites and a third weaker site were found in the BhCBM25 data. Refinement of heavy atom parameters and initial phasing with SHARP (25) using these three sites yielded an overall phasing power of 2.0 and a figure of merit of 0.35 for the full 1.39 Å resolution range. In the case of the BhCBM26 Cd²⁺ derivative, a highly redundant data set was collected that covered a sweep angle of 300 °C. This full set data was processed to 2.5 Å (Cd²⁺ redundant data set; Table 2) and was used to find four Cd²⁺ sites. A subset of the data between the oscillation angles of 47 and 128° was processed to 2.0 Å (Cd²⁺ 2.0 Å subset; Table 2) and was used for phasing, model building, and refinement. Refinement of heavy atom parameters and initial phasing with SHARP using these four Cd²⁺ sites yielded an overall phasing power of 0.9 and a figure of merit of 0.36 for the 2.0 Å resolution range. Density improvement was performed with DM (26), which yielded a final correlation coefficient of 0.61 for BhCBM25 and 0.73 for BhCBM26. For both BhCBM25 and BhCBM26, the resulting electron density maps were readily interpretable, and initial models of ~70–80% completeness were built by ARP/wARP (27). These initial models were corrected and completed manually by successive rounds of building using COOT (28) and refinement with REFMAC (29). The coordinates of these refined BhCBM25 and BhCBM26 structures were used as molecular replacement models to solve the structures of their respective ligand-complexed forms. The models of the complexed forms were also corrected by successive rounds of building using COOT and refinement with REFMAC. Water molecules were added using the ARP/wARP option within REFMAC and inspected visually prior to deposition. In all data sets, 5% of the

---

**JOURNAL OF BIOLOGICAL CHEMISTRY**

**VOLUME 281 • NUMBER 1**

**JANUARY 6, 2006**

**589**
Structural and Functional Analysis of BhCBM25 and BhCBM26

**TABLE 2**

| Module   | Amylose | Amylopentin | Dextran | Pullulan |
|----------|---------|-------------|---------|----------|
| BhCBM25  | +++     | +++         | -       | +++      |
| BhCBM26  | ++      | +           | -       | ++       |
| BhCBM25/26 | +++    | +           | -       | ++       |
| BhX23-1  | -       | -           | -       | -        |
| BhX23-2  | -       | -           | -       | -        |
| BhGH13CM | -       | -           | -       | -        |

**RESULTS AND DISCUSSION**

Modular Dissektion of the B. halodurans C-125 Maltooliose-forming Amylase—The B. halodurans C-125 protein encoded by open reading frame Bh0413 has ~97% amino acid sequence identity with the recently characterized maltooliose-forming amylase from B. halodurans LBK 34 (AMY34) (31) and ~75% identity with the Bacillus sp. H-167 G6-amylase (32). Amino acid sequence alignments of the B. halodurans C-125 amylase using PSI-BLAST indicate that this protein comprises five modules: an N-terminal family 13 13-glycose hydrolase module followed by a tandem repeat of modules with unknown function, a module having amino acid sequence similarity to family 26 CBMs, and a C-terminal module having amino acid sequence similarity to family 25 CBMs (Fig. 1). This modular architecture is shared by AMY34 and G6-amylase. In order to assess the contributions of its cognate modules to carbohydrate binding, each individual module and the BhCBM26/CBM25 tandem (see Fig. 1) were cloned, produced independently in E. coli, and assessed for binding function by affinity gel electrophoresis. BhCBM25, BhCBM26, and the BhCBM26/CBM25 tandem bound tightly to amylose, amylpectin, and pullulan (Table 3). BhCBM25 and BhCBM26/CBM25 were able to bind weakly to the α-1,6-glucan, dextran. The X23 modules did not display any affinity for the α-glucans tested.

Starch Binding by BhCBM25 and BhCBM26—BhCBM25 and BhCBM26 were both able to bind granular starch as assessed by depletion binding isotherms (Fig. 2). BhCBM25 bound with an association constant ($K_a$) of $3.3 \pm 0.3 \times 10^4$ M$^{-1}$ and a saturating level (or binding capacity, $N_a$) of $5.3 \pm 0.3$ μmol of CBM/10 g of starch. The corresponding values for BhCBM26 were $3.7 \pm 0.2 \times 10^4$ M$^{-1}$ and $9.9 \pm 0.3$ μmol of CBM/10 g of starch. The BhCBM26/CBM25 tandem bound with a $K_a$ of $1.6 \pm 0.2 \times 10^4$ M$^{-1}$ and an $N_a$ of $3.4 \pm 0.0$ μmol of CBM/10 g of starch.

The roughly 50-fold affinity enhancement for BhCBM26/CBM25 relative to the single CBMs is common with tandem CBMs and is thought to result from the simultaneous interaction of the two CBMs, with the insoluble substrate giving an increase in affinity through this avidity effect (19, 33, 34). The ability of the BhCBM26/CBM25 tandem to strongly agglutinate soluble amylpectin and pullulan (not shown) suggests that the individual modules bind to sites on separate glucan chains in starch, a “trans” effect, rather than adjacent sites on the same chain, a “cis” effect. Thus, this avidity effect is probably only relevant to binding aggregated α-glucan chains, such as in granular starch, where the individual glucan chains are effectively tethered in proximity to one another as an insoluble mass.
Structural and Functional Analysis of BhCBM25 and BhCBM26

The difference in $N_j$ between BhCBM25 and BhCBM26 is $\approx 2$-fold, and the $N_j$ of the tandem is significantly lower than both of the individual modules. This is somewhat surprising, given the similarities in physical dimensions and location of the sugar binding site on the BhCBM25 and BhCBM26 polypeptides (see below), which would predict these two CBMs to have similar binding capacities on granular starch. This is reminiscent of CBM17 and CBM28 modules, which frequently occur in tandem. Examples of CBMs from these two families had about 2-fold different binding capacities on noncrystalline insoluble cellulose despite having the same fold and physical dimensions (19, 35, 36). Furthermore, it was shown that CBM17 and CBM28 did not share binding sites on this substrate, and the CBM17/CBM28 tandem bound more tightly than the individual CBMs to a subset of the cellulose binding sites (19). The similarities between this cellulose binding system and the BhCBM26/CBM25 starch binding system suggested the possibility that BhCBM25 and BhCBM26 may recognize separate binding sites on granular starch. We tested this by competition-displacement binding studies using granular starch and Oregon Green® 514-labeled CBMs (see “Materials and Methods”). An excess of maltotriose was able to completely prevent binding of both CBMs to granular starch (Fig. 3), whereas bovine serum albumin was entirely unable to compete for binding (not shown). For each CBM, the unlabeled CBM was able to effectively compete for binding to granular starch with the OG-labeled version of itself, as would be expected (Fig. 3). Likewise, unlabeled BhCBM25 was able to compete for binding with OG-labeled BhCBM26, and unlabeled BhCBM26 was able to compete for binding with OG-labeled BhCBM25 (Fig. 3). This suggests that BhCBM25 and BhCBM26 bind predominantly to the same class of binding sites on granular starch but does not exclude the possibility of differential recognition of minor classes of binding sites. Thus, we are unable to confidently explain the differences in granular starch binding capacity for the two CBMs. However, one possibility is that the CBMs have a tendency to self-associate, resulting in artificial estimates of the binding capacities due to protein-protein interactions at the starch surface.

Both BhCBM25 and BhCBM26 contain several lysine residues, which are labeled by Oregon Green® 514 succinimidyl ester, making the modification of these residues with label a potential concern with respect to starch binding. However, based on the structures of these CBMs (see below), the lysine residues and N termini are clearly removed from the binding sites, making it very unlikely that the labeling would have a substantial impact on starch binding. Furthermore, the performance of the OG-CBMs in the binding assays suggests that starch binding was largely unaffected by the labeling.

Quantitative Characterization of α-Glucosaccharide Binding by BhCBM25 and BhCBM26—To gain better insight into the size of the binding site required by BhCBM25 and BhCBM26, their soluble α-glucosaccharide binding properties were studied by UV difference titrations and ITC. The addition of α-glucosaccharides to solutions of BhCBM25 and BhCBM26 induced large perturbations in their UV spectra. Use of this signal in a quantitative manner gave hyperbolic binding curves consistent with a simple binding mechanism (Fig. 4). Conversion of the UV difference titrations to a Scatchard form assuming a 1:1 binding mechanism yielded straight lines, supporting this assumption (Fig. 4, inset). Furthermore, ITC binding isotherms generated under conditions appropriate for experimental determination of stoichiometries indicated a 1:1 binding mechanism (Table 4). This differentiates these CBMs from the fungal family 20 CBMs, which have two maltotrioside binding sites per protein (7). Analysis of the UV difference and ITC isotherms with a one-binding site model gave affinities in the range of $10^3$ to $10^5$ M$^{-1}$ (Table 4), similar to the affinities determined for granular starch binding. BhCBM25 showed a strong dependence of affinity on ligand length up to a maltotrioside binding length of seven glucose units, whereas the affinity of BhCBM26 peaked at a maltotrioside binding length of five glucose units.

The enhanced binding of BhCBM25 and BhCBM26 to longer maltotriosaccharides (degree of polymerization $> 3$) is somewhat surprising, given that their x-ray crystal structures (see below) suggest that these CBMs only have two, or at most three, binding subsites. A similar phenomenon, where the optimal ligand length exceeds that of the number of subsites, has been observed with a number of other CBMs (e.g. see Refs. 37 and 38). We can postulate four possible explanations for this. First, the additional sugars of the longer maltotriosaccharides and their intramolecular interactions may influence the tertiary conformation of the sugars such that they adopt an overall structure more appropriate for recognition by CBMs. Second, an alternative but related explanation is that the additional sugars flank the binding site on the sugar restrict the freedom of the binding sugar residues and reduce the loss of configurational entropy upon binding, thus making the overall change in free energy more favorable. Third, if the minimal binding site in the sugar is a fraction of the overall oligosaccharide and is a repeating unit, then the oligosaccharide can be considered a linear array of overlapping binding sites. For example, if the required binding motif is three adjacent sugar residues, then maltopentaose can be thought of as a linear array of three overlapping binding sites. Although overall the observed interaction is observed to be 1:1 due to steric exclusion at the overlapping sites, there are three ways to form the bound state, giving rise to a statistical advantage to binding longer sugars. This would be reflected as an increase in the apparent association constant when modeling such an interaction based on a simple bimolecular interaction model. Last, binding subsites are typically identified in x-ray crystal structures of protein-carbohydrate complexes by the observation of direct interactions between sugar residues and the protein. Although they do not necessarily make significant direct interactions with the protein, the sugar residues within a bound oligosaccharide that flank the sugars occupying obvious binding subsites on the protein must displace waters from the hydration shells surrounding the protein. It is possible that this process contributes favorably to the overall free energy of binding. It is currently not possible to comment on which of these possibilities is the most likely...
contributor, but it is probable that all play a role to some degree in how oligosaccharides of a length exceeding the size of the binding site on the protein are bound more tightly than expected.

BhCBM25 and BhCBM26 were clearly able to recognize the α-1,4-linked glucose portions of starch, as indicated by their ability to bind tightly to maltooligosaccharides. We assessed their potential to recognize the α-1–6 branch points of the amylopectin component of starch by studying the binding of these two CBMs to isomaltose (α-1,6-linked glucobiose), isomaltotriose (α-1,6-linked glucotriose), 63-α-D-glucosyl-maltotriose (GM3; α-1,4-linked glucotriose with a glucose linked α-1–6 to the nonreducing end), and 63-α-D-glucosyl-maltotriose-maltotriose (GM3M3; GM3 with an α-1,4-linked glucotriose linked α-1–6 to the reducing end). Neither CBM bound significantly to isomaltose or isomaltotriose (not shown). BhCBM25 bound to GM3 with an affinity roughly similar to its affinity for maltotriose and to GM3M3 with an affinity roughly similar to its affinity for maltotetraose. Relative to maltotriosaccharide binding, BhCBM26 bound quite poorly to GM3 (~4-fold lower affinity relative to maltotriose) and bound to GM3M3 with an affinity approximating that of maltotriose. Overall, these results indicate that the α-1–6 branch points in starch are not a recognition determinant for these two CBMs. However, they do tolerate α-1–6 branches in the context of their α-1,4-linked glucose recognition sequences, with BhCBM25 being able to tolerate an adjacent branch slightly better.

All of the binding reactions between BhCBM25 and BhCBM26 and α-glucooligosaccharide ligands were favored by a negative change in enthalpy (ΔH) (Table 4). In contrast, all changes in entropy (ΔS) were unfavorable. This signature is common to protein-carbohydrate interactions, CBMs in particular (1).

Structures of BhCBM25 and BhCBM26—The x-ray crystal structure of BhCBM25 was solved by the SAD method using a protein derivative obtained by the iodination of tyrosines with N-iodosuccinimide. Excellent initial phases were obtained using three iodine sites. Anomalous difference peaks, presumably corresponding to the iodine positions in the asymmetric unit (designated chain A; Fig. 5). Anomalous difference peaks, presumably corresponding to the iodine positions in the asymmetric unit (designated chain A; Fig. 5).
Structural and Functional Analysis of BhCBM25 and BhCBM26

**TABLE 4**

Parameters of BhCBM25 and BhCBM26 binding to α-glucans as determined by isothermal titration calorimetry and UV difference

|                | UV, $K_a$ | ITC |      |      |
|----------------|-----------|-----|------|------|
|                | $\times 10^{-4}$ M$^{-1}$ | $\times 10^{-4}$ M$^{-1}$ | $\Delta G$ kcal/mol | $\Delta H$ kcal/mol | $\Delta S$ cal/mol/K | $n$ |
| BhCBM25        |           |     |      |      |
| Maltotriose    | 0.49 ± 0.00 | 0.33 ± 0.03 | 4.79 ± 0.00 | 10.64 ± 0.11 | 19.6 ± 0.4 | 0.97 ± 0.00 |
| Maltotetraose  | 1.21 ± 0.02 | 0.70 ± 0.21 | 5.24 ± 0.02 | 11.01 ± 0.17 | 19.3 ± 0.6 | 0.98 ± 0.00 |
| Maltopentaose  | 2.25 ± 0.03 | 1.17 ± 0.00 | 5.55 ± 0.00 | 12.94 ± 0.11 | 24.8 ± 0.6 | 0.92 ± 0.01 |
| Maltotriose    | 3.29 ± 0.05 | 1.48 ± 0.03 | 5.69 ± 0.01 | 16.68 ± 0.37 | 36.8 ± 1.3 | 0.73 ± 0.02 |
| Maltotetraose  | 4.67 ± 0.06 | 1.84 ± 0.21 | 5.82 ± 0.00 | 20.34 ± 0.41 | 48.7 ± 0.0 | 0.61 ± 0.01 |
| GM3            | ND$^a$     | 0.21 ± 0.00 | 4.54 ± 0.02 | 15.31 ± 0.35 | 19.4 ± 1.2 | 1.01 ± 0.02 |
| GM3M33        | 1.28 ± 0.02 | 0.65 ± 0.03 | 4.79 ± 0.00 | 14.57 ± 0.42 | 31.4 ± 1.6 | 0.77 ± 0.02 |
| BhCBM26        |           |     |      |      |
| Maltotriose    | 5.14 ± 0.22 | 3.42 ± 0.05 | 6.18 ± 0.00 | 14.12 ± 0.00 | 26.7 ± 0.46 | 1.18 ± 0.01 |
| Maltotetraose  | 8.37 ± 0.27 | 6.46 ± 0.08 | 6.56 ± 0.00 | 13.81 ± 0.00 | 24.3 ± 0.35 | 1.21 ± 0.01 |
| Maltopentaose  | 20.75 ± 0.94 | 8.23 ± 0.05 | 6.71 ± 0.00 | 13.76 ± 0.00 | 23.7 ± 0.26 | 1.19 ± 0.03 |
| Maltotriose    | 22.99 ± 0.89 | 8.08 ± 0.09 | 6.69 ± 0.00 | 17.04 ± 0.02 | 34.7 ± 0.56 | 1.00 ± 0.02 |
| Maltotetraose  | ND$^a$     | 8.08 ± 0.26 | 6.69 ± 0.02 | 18.91 ± 0.02 | 41.0 ± 0.74 | 0.88 ± 0.00 |
| GM3            | ND$^a$     | 0.77 ± 0.04 | 5.53 ± 0.00 | 10.19 ± 0.02 | 16.5 ± 0.90 | 1.36 ± 0.02 |
| GM3M33        | 5.33 ± 0.28 | 2.69 ± 0.10 | 6.04 ± 0.00 | 14.77 ± 0.02 | 29.3 ± 0.64 | 1.09 ± 0.02 |

$^a$ ND, not determined.

quite diffuse and suggested multiple conformations. Indeed, the loop comprising residues 46–49 in chain B, which contains the iodinated tyrosines, was extremely disordered and could not be modeled, although the iodines were tentatively modeled on the basis of the anomalous difference maps as partially occupied atoms in multiple conformations. In addition to this missing loop, most of the atoms in chain B refined with higher B-factors than those in chain A, reflecting the greater disorder of this molecule. These factors are probably contributors to the somewhat higher R-factors than would be expected for a structure of this resolution (Table 2).

The crystals of BhCBM26 obtained in the presence of cadmium were small plates and diffracted substantially better in certain orientations in the beam. A highly redundant data set to 2.5 Å (Table 2) was required to locate four cadmium atoms using only the anomalous signal (Fig. 5). Using these heavy atom positions, a subset of the data that extended to 2.0 Å could be phased using only the anomalous differences. The resulting electron density maps were considerably better than those generated with the full 2.5 Å data set, making map interpretation straightforward and automatic model building possible.

Both CBMs adopted extremely similar β-sandwich folds with immunoglobulin-like topologies (Fig. 5). The structures of BhCBM25 and BhCBM26 overlapped with a root mean square deviation of 1.12 Å over 72 (of a possible 92) matched Cα atoms (structural overlap performed with COOT (28) and root mean square deviation calculated with DeepView (39)). Structural comparisons with DALI (40) showed that the closest structural relatives are a number of CBMs falling into CBM families 20 and 34.

**BhCBM25 in Complex with Maltotetraose—BhCBM25 was initially crystallized in the space group P21, when using a large molar excess of maltotetraose. Clear electron density was found for two molecules of maltotetraose bound to each of the four CBM monomers in the asymmetric unit, suggesting two functional binding sites (here called α and β) per CBM (Fig. 6). The base of the α-site is formed by Trp34 and Trp24. These stack against the pyranose rings of two adjacent sugar residues. Only these two sugars appear to make substantial interactions with the protein, suggesting only two binding subsites. His26, which is bracketed by Trp34 and Trp24, hydrogen-bonds with the O6 of the sugar stacking against Trp24 (Fig. 7). Asp43 hydrogen bonds with O2 and O3 of the same sugar. N76 completes the complement of hydrogen bonds by interacting with O2 and O3 of the sugar stacking against Trp24. Although there is an extensive water network surrounding the protein and sugar, there appear to be no protein-carbohydrate hydrogen bonds mediated by only a single water molecule.

In the β-site, only a single tryptophan side chain, Trp20, stacks against the sugar. Although the sugar does make 3–4 potential direct hydrogen bonds with the protein, it also makes at least one with a neighboring molecule and three with the sugar in the β-site of the same neighboring molecule.

This presence of two bound sugar molecules per BhCBM25 monomer is somewhat at odds with solution binding studies, which indicate a single binding site on the protein. It is possible that the β-site has an affinity too low to be detectable in the solution binding studies, whereas the combination of 20-fold molar excess of maltotetraose used in the crystalization conditions and crystal contacts between the sugar in the β-site and the neighboring protein molecule(s) promoted binding to this site.

In order to gain further insight into this phenomenon, we co-crystallized BhCBM25 with a ~1:1 molar ratio of maltotetraose to protein to improve the probability of obtaining a crystal form with the predominating binding site occupied. Crystals belonging to the space group P4_2_2_2 were obtained that had a very high solvent content (~70%), probably explaining the poor diffraction of these crystals. Despite the relatively low resolution of these data, they clearly showed that only the α-site was occupied (Fig. 6) despite full accessibility of the β-site. This, taken with the solution binding studies, suggests that occupation of the β-site in the P21 crystal form was an effect of the crystalization conditions and that BhCBM25 has only one predominant maltoligosaccharide binding site.

In the P4_2_2_2 crystal form, the bound sugar spanned the α-site of one CBM monomer in the asymmetric unit and the α-site of a symmetry-related molecule (Fig. 6). Furthermore, five monosaccharide units in the bound sugar could be modeled, rather than the expected four of maltotetraose, suggesting at least two bound conformations. Again, the pyranose rings of two adjacent sugars stack against Trp34 and Trp24. These two sugars overlap almost perfectly with the analogous sugars in the P21 maltotetraose complex and make an identical range of interactions with the protein (not shown). This somewhat surprising result demonstrates how closely packed this CBM can be on even a relatively small ligand. Whether this occurs in solution is unknown. However, the decrease in stoichiometry with increasing ligand length observed by ITC suggests that a species of complex with two bound CBMs per oli-
gosaccharide is possible, as was observed with the *T. maritima* CBM41 (11).

Occupation of the α-site by maltotetraose in the P2₁ complex results in the burial of 178 Å² of polar surface area and 269 Å² of apolar surface area. The sugar residues other than those that stack with the tryptophan side chains make a negligible contribution to these surface area values. With five direct hydrogen bonds, this equates to a hydrogen bonding density of ~2.8 hydrogen bonds/100 Å² of buried polar surface area.

Lectins have a hydrogen bonding density of 3.45 ± 0.52 bonds/100 Å² of buried polar surface area, and Type B CBMs have hydrogen bonding densities of 2.11 ± 0.61 bonds/100 Å² of buried polar surface area (1). The hydrogen bonding density of *Bh*CBM25 is intermediate, suggesting properties of both types of CBMs.

*Bh*CBM26 in Complex with Maltose—*Bh*CBM26 crystallized in the space group P3₁2₁ in the presence of a large excess of maltose with eight molecules in the asymmetric unit. In contrast to *Bh*CBM25, each of the

![Image of Phasing and structures of BhCBM25 and BhCBM26.](image-url)

**A**, anomalous difference peaks (red) and representative electron density (blue; 0.39 electrons/Å³) contoured around the iodotyrosine heavy atom sites used for SAD phasing of *Bh*CBM25. **B**, anomalous difference peaks (red) and representative electron density (blue; 0.37 electrons/Å³) contoured around the cadmium sites used for SAD phasing of *Bh*CBM26. **C**, the overall secondary structure of *Bh*CBM25 as representative of the fold and topology of both *Bh*CBM25 and *Bh*CBM26. Selected amino acid side chains are shown in a “licorice” representation. Electron density maps are maximum likelihood (29)/σ_a (44) weighted 2Fo – Fc electron density maps.
molecules in the asymmetric unit was found to bind only a single molecule of maltose (Fig. 6). The architecture of the BhCBM26 binding site is similar to that of BhCBM25. The base of the binding is formed by Trp36, Tyr23, and Tyr25. Trp36 and Tyr25 stack against the pyranose rings of the two sugar residues. Tyr23 is flanked by Trp36 and Tyr25 and contributes a hydrogen bond, with the O6 of the sugar stacking against Tyr25 (Fig. 7). Gln71 hydrogen bonds with O2 and O3 of the sugar stacking against Trp36. The backbone carbonyls of Gly76 and Glu77 hydrogen bond with O3 and O2, respectively, of the sugar stacking against Tyr25. As with BhCBM25, there is an extensive water network surrounding the protein and sugar, but there appear to be no water-mediated hydrogen bonds. Occupation of the BhCBM26 binding site results in the burial of 172 Å² of polar surface area and 292 Å² of apolar surface area. Again, with only five direct hydrogen bonds, this equates to a hydrogen bonding density of ~2.9 hydrogen bonds/100 Å² of buried polar surface area, placing BhCBM26, like BhCBM5, as an intermediate between Type B and C CBMs with respect to this particular property (1).

The binding of maltose to BhCBM26 appears to cause a small structural change in the protein. The loop comprising residues Pro73–Gly79 that flanks the binding site moves ~3 Å (measured at the Cα of Gly79) to close in
on the sugar (Fig. 8). This new conformation allows the hydrogen bonding between the backbone carbonyls of Gly76 and Glu77 and the sugar, which would not otherwise be possible in the open loop conformation. In contrast, no substantial conformational changes upon ligand binding by BhCBM25 were observed. Of the numerous CBM-ligand interactions studied by structural methods, notable conformational changes upon binding have only been observed in PpCBM36 (41) and CjCBM35 (42).

Structural Conservation of the Main Binding Platform—An overlap of the BhCBM25 and BhCBM26 structures reveals the overall similarities in their folds despite only having ~15% amino acid sequence identity based on this structural alignment. The functional residues are quite poorly conserved as judged by the sequence alignment generated from the structural overlap (Fig. 9). Only the central amino acid side chain in the main binding platform, which is His26 in BhCBM25 and Tyr23 in BhCBM26, is conserved at the structural and sequence level. Trp34 and Trp36 of BhCBM25 are structurally conserved with Trp36 and Tyr25 of BhCBM26, since their rings approximately overlap in space, but are not conserved at the sequence level (Fig. 10). Thus, this apolar frame that provides the main protein-carbohydrate interactions and accommodates the curvature in a maltooligosaccharide is structurally well conserved in these two CBMs despite their lack of significant sequence identity.

A Conserved Mode of α-Glucan Recognition—The three-dimensional structures in ligand-complexed forms are now available for examples from CBM families 20, 25, 26, and 34, four of the six known starch-specific CBM families. These CBMs are remarkably similar in how they recognize α-glucans. The general location of the primary binding site in families 25 and 26 is well conserved with one of the binding sites in family 20. Like the family 20 CBMs, the family 34 CBM from T. vulgaris also appears to have two functional maltooligosaccharide binding sites (14); however, the location of these binding sites is not conserved with the other CBM families. Despite the differences in the family 34 CBMs, the binding site architectures of the CBMs in all of these families are all similarly tailored to recognize the curvature of maltooligosaccharides formed by the α-faces of the pyranose rings (Fig. 10). The relative spacing between the aromatic side chains making up the binding platforms does vary somewhat but, with the exception of TvCBM34, which appears to have an additional subsite, does not stretch beyond a distance required to accommodate a disaccharide. This conserved mechanism of recognizing a structural motif in α-glucans can be rationalized by assuming that the natural biological ligand of these proteins is granular starch. In this context, α-glucan chains form double helices where only the α-faces of the pyranose rings are substantially exposed. These helices are then packed into crystalline arrays (43). Thus, as a class of proteins, these CBMs are well designed for binding to the exposed surfaces formed by the α-faces of the glucose rings making up the α-glucan chains in granular starch.
CONCLUSION

Although BhCBM25 and BhCBM26 recognize maltooligosaccharides with differing affinities (5–10-fold different) they appear to recognize predominantly the same binding site on granular starch with similar affinities. Thus, the biological significance of the tandem arrangement of these CBMs is probably its conference of an avidity effect, which enhances the affinity of the enzyme for granular starch. It is also conceivable that this tandem arrangement is suited to the disruption of starch structure much like the two binding sites of a single fungal CBM20; however, this remains to be determined.

The arrangement of the α-glucan chains in granular starch did not appear to confer any advantage in the binding of BhCBM25 and BhCBM26 compared with their binding to soluble maltooligosaccharides. Given the conformation of maltooligosaccharides in the x-ray crystal structures of these CBMs, it seems likely that BhCBM25 and BhCBM26 only require a single α-glucan chain in an amylase helix for tight binding. The significance of this, together with their apparent tolerance for adjacent α-1,6 branch points and small "footprint," is that these CBMs are unlikely to discriminate between crystalline and non-crystalline regions of starch. Thus, in the biological setting, the enzyme may be able to efficiently bind to most, if not all, regions of the starch granule to maximize the potential for substrate hydrolysis.

REFERENCES

1. Boraston, A. B., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2004) Biochem. J. 382, 769–782
2. Bolam, D. N., Ciruela, A., Mcqueen-Mason, S., Simpson, P., Williamson, M. P., Rixon,
