# α-Glucan Recognition by a New Family of Carbohydrate-Binding Modules Found Primarily in Bacterial Pathogens†

Alicia Lammerts van Bueren, Ron Finn, Juan Ausió, and Alisdair B. Boraston*

*Department of Biochemistry and Microbiology, University of Victoria, P.O. Box 3055 STN CSC, Victoria, British Columbia V8W 3P6, Canada*

*Received August 17, 2004; Revised Manuscript Received September 20, 2004*

ABSTRACT: *Tm*Pul13, a family 13 glycoside hydrolase from *Thermotoga maritima*, is a four-module protein having pullulanase activity; the three N-terminal modules are of unknown function while the large C-terminal module is likely the catalytic module. Dissection of the functions of the three unknown modules revealed that the 100 amino acid module at the extreme N-terminus of *Tm*Pul13 comprises a new family of carbohydrate-binding modules (CBM) that a bioinformatic analysis shows are most frequently found in pullulanase-like sequences from bacterial pathogens. Detailed binding studies of this isolated CBM, here called *Tm*CBM41, reveals a preference for α-(1,4)-linked glucans, but occasional α-(1,6)-linked glucose residues, such as those found in pullulan, are tolerated. UV difference, isothermal titration calorimetry, and analytical ultracentrifugation binding studies suggest that maltooligosaccharides longer than four glucose residues are able to bind two *Tm*CBM41 molecules per oligosaccharide when sugar concentrations are below the CBM concentration. This is explained in terms of an equilibrium expression involving the formation of both a 1 to 1 sugar to CBM complex and a 1 to 2 sugar to CBM complex (i.e., a CBM dimer ligated by an oligosaccharide). The presence of an α-(1−6) linkage in the oligosaccharide appears to prevent this phenomenon.

α-Glucans are a structurally diverse group of polysaccharides found in a number of biological settings. Starch and glycogen are the prevalent storage polysaccharides in plants and animals, respectively. Starch is a composite of amylose [linear α-(1,4)-linked glucose] and amylopectin [linear α-(1,4)-linked glucose with α-(1,6) branch points occurring approximately every 24−30 glucose residues]. Glycogen is similar in structure to starch but is more extensively branched, with α-(1,6) branch points occurring approximately every 8−12 residues. The α-(1,4) linkages cause these polysaccharides to form tight helical structures, resulting in dense granules which function as highly effective storage systems. Pullulan is a structural polysaccharide, the main source of which is the fungus *Aureobasidium pullulans* (*1*). This polysaccharide is composed of repeating α-(1,6)-linked maltotriose [trisaccharide of α-(1,4)-linked glucose] units with occasional maltotetraose [tetrasaccharide of α-(1,4)-linked glucose] units present within the structure (*2*). Purified pullulan, like the other α-glucans, has many practical applications from dietary supplements to the production of biodegradable plastics (*3*). Dextran is a polysaccharide entirely comprising α-(1,6)-linked glucose found mainly as a bacterial extracellular polysaccharide. Bacteria can employ this polysaccharide coat in biofilm formation (*4*) or protection, in the case of pathogenic bacteria, from host phagocytes (*5*).

Glycoside hydrolase family 13 is, based on the number of amino acid sequence entries, one of the largest glycoside hydrolase families (http://afmb.cnrs-mrs.fr/CAZY/index.html). This family comprises a number of α-glucan active enzymes, including amylases, pullulanases, cyclomaltodextrin transferases, dextranases, and α-glucosidases, which are found in a number of biological locales where they are involved in the depolymerization, modification, or synthesis of α-glucans. These enzymes are usually highly modular, comprising a catalytic module along with accessory modules, most frequently noncatalytic carbohydrate-binding modules (CBM)[1] that mediate the tight association of the enzymes with their substrates. Currently, there are 40 families of CBMs defined on the basis of sequence similarity. Members of four of these families have been observed to bind to granular starch and/or α-glucooligosaccharides; however, only one fungal member of one family has been studied in detail (*6*, *7*).

*Thermotoga maritima* is a hyperthermophilic eubacteria first discovered in geothermal heated marine sediment (*8*). It has an optimal growth temperature of 80 °C and is one of the most thermophilic bacteria known. It produces a number of thermostable enzymes for polysaccharide depolymerization. One of these is a pullulanase, *Tm*Pul13, encoded by *pulA*, whose enzymatic activity has previously been studied (*9*, *10*). This enzyme appears to be specific for the α-(1,6)-glycosidic linkage in pullulan; however, it is not able to

† This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. A.B.B. is a Canada Research Chair in Molecular Interactions.

* To whom correspondence may be addressed. Tel: 250-472-4168. Fax: 250-721-8855. E-mail: boraston@uvic.ca.

[1] Abbreviations: CBM, carbohydrate-binding module; IMAC, immobilized metal ion affinity chromatography; IPTG, isopropyl β-D-thiogalactopyranoside; ITC, isothermal titration calorimetry; $K_a$, association constant.

10.1021/bi048215z CCC: $27.50 
Published on Web 11/12/2004

cleave the α-(1−6) bonds in amylopectin or α-(1−4) bonds in α-glucans. *Tm*Pul13 is a relatively large enzyme (843 amino acids) and likely comprises of number of modules. In this study, we dissect the modular structure of *Tm*Pul13 by heterologous production of its individual modular components in *Escherichia coli*. We report the α-glucan binding affinity and specificity of these modules and, on the basis of this information, propose a new CBM family, family 41. In addition, we dissect the kinetic and thermodynamic mechanism(s) of ligand recognition by this new CBM. This is the first study of a CBM from a bacterial pullulanase, which is made more unique by its hyperthermophilic source. Furthermore, examination of this new family of CBMs reveals its distribution in primarily pathogenic bacteria and suggests a carbohydrate-binding function for these modules in pathogenic microbes.

## MATERIALS AND METHODS

*Carbohydrates and Polysaccharides.* $6^3$-α-D-Glucosylmaltotriose (GM3) and $6^3$-α-D-glucosylmaltotriosylmaltotriose (GM3M3) were purchased from Megazyme Ltd. (Bray, C. Wicklow, Ireland). Maltose (M2), maltotriose (M3), maltotetraose (M4), maltopentaose (M5), maltohexaose (M6), isomaltose, isomaltotriose, panose, pullulan, and amylopectin (starch) were from Sigma (St. Louis, MO).

*Cloning, Expression, and Purification of CBMs.* The DNA fragments encoding the modules X28, X45, X20, X28/45/20, and *Tm*Pul13 (see Figure 1) were amplified by PCR from *T. maritima* genomic DNA (strain MS8B, ATCC 43589D) by a procedure described previously (*11*). The 5′ oligonucleotide primers were 5′-CACCGAAACCACCATCGTAGTC-3′ (X28, X28/45/20, and Pul13), 5′-CACCGACACATCTCCCAGAATC-3′ (X45), and 5′-CACCGGAGAGCTCGGAGCCGTA-3′ (X20). The 3′ oligonucleotide primers were 5′-CTT**TTA**TGGTTTTTCGTAGAAAAA-3′ (X28), 5′-CTT**TTA**ATCGTAATAGTAGTCGTC-3′ (X45), 5′-CTT-**TTA**TTCGTATCCTTCGATTTT-3′ (X20 and X28/45/20), and 5′-CTT**TTA**CTCTCTGTACAGAACGTA-3′ (Pul13) (the stop codon is in bold). These allowed for the amplification of nucleotide sequences from *pulA* encoding amino acids 20−120 (X28), 121−222 (X45), 223−339 (X20), 20−339 (X28/45/20), and 20−843 (Pul13) of *Tm*Pul13. The 5′ CACC in the 5′ oligonucleotide primers was essential for inserting purified fragments into pET Directional TOPO Expression Kits (Invitrogen, Carlsbad, CA) using pET150 to give pET150-X28 and pET150-Pul13 and using pET100 to give pET100-X45, pET100-X20, and pET100-X28/45/20. They were transformed into *E. coli* BL21(DE3) for polypeptide production. All polypeptides comprised a $His_6$ tag fused to the N-terminus by an enterokinase cleavage site. The fidelity of the cloned inserts was confirmed by bidirectional DNA sequencing.

Six liters of LB medium inoculated with *E. coli* BL21/DE3* harboring each expression vector was incubated at 37 °C in shaking flasks to an $OD_{600nm}$ of ∼0.6. Isopropyl β-D-thiogalactopyranoside (IPTG) was added to a final concentration of 0.1 mM for pET150-Pul13 and incubation continued for ∼4 h. For pET150-X28, pET100-X28/45/20, pET100-X45, and pET100-X20, the flasks were incubated at 37 °C with shaking to an $OD_{600nm}$ of ∼0.6, at which point the temperature was cooled to room temperature (∼21 °C) and incubation continued for ∼16 h. The addition of IPTG was not necessary. The cells were harvested by centrifugation at 5000 rpm for 10 min, resuspended in 90 mL of 20 mM Tris, pH 8.0, containing 0.5 M NaCl, and lysed by French press. The pET150-Pul13 cells were resuspended in Bug-Buster protein extraction reagent (Novagen, Madison, WI) and lysed according to manufacturer's protocols. After centrifugation at 15000 rpm for 45 min, the supernatant was collected, and the polypeptides were purified by immobilized metal affinity chromatography (IMAC) with HIS-Select nickel affinity gel (Sigma, St. Louis, MO) according to manufacturer's protocols. Purified polypeptides were concentrated in a stirred ultrafiltration unit on a 5K molecular weight cutoff filter and dialyzed extensively against 50 mM Tris, pH 7.5, using regenerated cellulose dialysis tubing with a 3K MWCO. Purity was greater than 95%, as assessed by SDS−PAGE. Yields were typically 50 mg/L of culture or more.

*Determination of Protein Concentration.* The concentration of purified proteins was determined by UV absorbance at 280 nm using the following calculated extinction coefficients (*12*): 34850 $M^{-1}$ $cm^{-1}$ for *Tm*CBM41 (X28), 7680 $M^{-1}$ $cm^{-1}$ for X45, 35560 $M^{-1}$ $cm^{-1}$ for X20, 76810 $M^{-1}$ $cm^{-1}$ for X28/45/20, and 155730 $M^{-1}$ $cm^{-1}$ for *Tm*Pul13.

*Affinity Electrophoresis.* The binding of all polypeptides was assessed by affinity electrophoresis (*13*) in 10% native polyacrylamide gels polymerized without polysaccharide or in the presence of 0.5% amylopectin, pullulan, amylose, or dextran. Electrophoresis was performed for 2 h at room temperature with native running buffer (25 mM Tris−base, 0.2 M glycine) in an XCell SureLock Mini-Cell system (Invitrogen, Carlsbad, CA) at a constant voltage of 100.

*Macroarrays.* BSA, *Cc*CBM17, *Tm*CBM4-2, and *Tm*-CBM41 were labeled with Alexa Fluor 680 succinimidyl ester (Molecular Probes, Eugene, OR) according to the manufacturer's recommendations. Free label removal and buffer exhange were achieved by gel filtration using Sephadex G-25 (Amersham Biosciences). Macroarrays were prepared by spotting 1 μL of a 10 or 1 μg/μL solution of polysaccharide onto a nitrocellulose membrane. The membranes were allowed to dry for at least 2 h. Membranes were then blocked for 1 h at 10 °C with a solution of PBS containing 1% BSA and 0.05% Tween 20. These were probed by incubation at 10 °C overnight with 10 mL of blocking buffer containing ∼10 μg/mL Alexa Fluor 680 labeled protein. Probed membranes were briefly washed twice with PBS and then laser scanned with an excitation wavelength of 700 nm using a LICOR Odyssey infrared scanner (LICOR, Lincoln, NE). Binding was visualized by the presence of fluorescence.

*Isothermal Titration Calorimetry.* Isothermal titration calorimetry (ITC) was performed as described previously (*11*) using an VP-ITC (MicroCal, Northampton, MA). All polypeptides were dialyzed into 50 mM Tris buffer (pH 7.5). Carbohydrate solutions were prepared by mass using dialysate buffer saved from the protein dialysis. Protein and carbohydrate solutions were filtered (0.2 μm) and degassed prior to use. Experiments were performed at 25 °C. Titrations were performed by injecting 10 μL samples of oligosaccharide solutions at 1−5 mM into the ITC sample cell containing 50−150 μM *Tm*CBM41. The concentrations of protein were chosen such they were in 5-fold or greater excess of the

dissociation constants. Heats of dilution upon titration of buffer into carbohydrate or buffer into buffer were negligible. Reverse titrations were performed by injecting 4−10 μL samples of concentrated *Tm*CBM41 (1.385−2.22 mM) into solutions containing M4, M5, M6, pullulan, or amylopectin. These binding data were corrected for the heats of dilution determined by titrating *Tm*CBM41 into buffer. Binding stoichiometries, enthalpies, and equilibrium association constants were determined by fitting the corrected data to the appropriate binding model (see text) with MicroCal Origin 7. In the cases of pullulan and amylopectin, the carbohydrate concentrations were expressed as equivalents of a tetrasaccharide in all of the calculations. Using a method adapted from Sigurskjold et al. (*14*), the acceptor concentration (i.e., the concentration of binding sites in the polysaccharides in terms of tetrasaccharide equivalents) was a regressed parameter. All data show the average and standard deviation of three independent titrations.

*UV Difference Titrations*. High-precision, automated UV difference titrations were performed using equipment essentially the same as that described previously (*15*, *16*). Protein and carbohydrate samples were prepared identically to those for ITC, filtered, and degassed prior to use. All experiments were held at 25 °C in a Peltier thermostated cuvette holder. Samples of carbohydrate (1 mM) were added to 2 mL of protein (~33 μM) and allowed to equilibrate for 80 s with stirring. One thousand scans (10 ms integration time) collected from 225 to 380 nm were averaged for each carbohydrate addition. The protein concentration used for each titration was calculated from the absorbance at 280 nm extracted from the spectrum at zero carbohydrate concentration. Each spectrum was corrected for dilution due to the addition of ligand, and the difference spectra were calculated by subtraction of the absorbance spectrum at zero carbohydrate concentration. Difference spectra were examined for peak and trough wavelengths and values at the appropriate wavelengths extracted for further analysis. The peak-to-trough heights at the wavelength pairs 292.72/288.82 nm, 285.06/288.82 nm, and 285.06/278.24 nm were calculated by subtraction of the trough values from the peak values and the data plotted against total carbohydrate concentration. The resulting isotherms were analyzed with DynaFit (*17*) using the model equilibria as discussed in the text. Data for the three wavelength pairs were fit simultaneously to improve the precision of the regressed parameters. The parameters determined were the $K_a$ and molar difference absorbance response ($\Delta A$). Though the CBM concentration ([M]) was measured during the experiment, because the experiments were performed under pseudo-first-order conditions (acceptor concentration >10-fold in excess of $K_d$), this parameter was also allowed to float in the analysis to give a regressed value of [M] (called [$M_{fit}$]). Under these conditions, [$M_{fit}$] actually represents an experimental estimate of the concentration of total binding sites, rather than just the macromolecule concentration. Assuming that the ligand concentration is known with accuracy as it is prepared by mass from pure lyophilized material, the ratio [$M_{fit}$]/[M] (i.e., the experimentally determined concentration of total binding sites divided by the known concentration of CBM) gives an estimate of the number of ligand binding sites on M, i.e., the stoichiometry or *n* value. The data reported are the averages and standard deviations of four independent titrations.

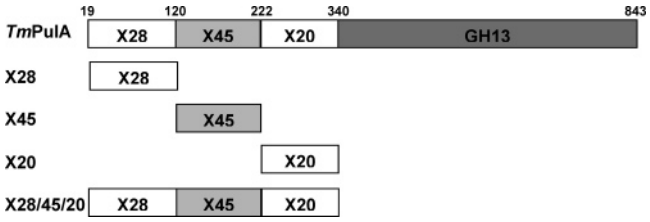


FIGURE 1: Modular organization of *Tm*Pul13. Amino acid numbers corresponding to the module boundaries are shown above the schematic. The individual module constructs used in this study are also shown.

*Analytical Ultracentrifugation.* Sedimentation equilibrium experiments were carried out using a Beckman XL-I analytical ultracentrifuge using an An-60 Ti (titanium) rotor. The samples were loaded on six-hole charcoal-filled Epon 12 mm cells. All runs were carried out at 20 °C. The scans were analyzed using XL-A UltraScan II version 6.0 sedimentation data analysis software (Borries Demeler, Missoula, MT) using a global nonlinear, least squares, curve fitting (*18*, *19*). The protein samples at different concentrations ranging from an $OD_{230}$ of 0.3−0.8 to an $OD_{280}$ of 0.3−0.8 were extensively dialyzed against 50 mM Tris-HCl (pH 7.5) buffer and were analyzed for equilibrium conditions achieved at different rotor speeds (see legend to Figure 6 for more details).The partial specific volume of the *Tm*CBM41 protein (0.740 $cm^3 \cdot g^{-1}$) was calculated from its amino acid composition according to ref *20* using the partial specific volumes provided in ref *21*.

## RESULTS AND DISCUSSION

*Modular Properties of TmPul13 and Identification of Its Carbohydrate-Binding Module.* Pullulanase activity has been demonstrated for *Tm*Pul13, but the modular arrangement and the functions of the accessory modules have not been reported. On the basis of PSI-BLAST amino acid sequence alignments (*22*) of the entire *Tm*Pul13 sequence, this protein appeared to comprise four modules (Figure 1). The three N-terminal modules have undetermined function and fall into the X-module families X28, X45, and X20 (*23*). The large C-terminal module shows sequence homology to glycoside hydrolases from family 13, so it is therefore highly likely that this is the catalytic module responsible for *Tm*Pul13's pullulanase activity. To determine if any of these accessory modules are CBMs, each individual module (called X28, X45, and X20, respectively), a triple module consisting of all the X domains (called X28/45/20), and the full-length enzyme (*Tm*Pul13) were cloned and produced independently in *E. coli* (see Figure 1). All of the polypeptides could be produced at high levels in the cytoplasm and purified to homogeneity by IMAC.

X28, X28/45/20, and *Tm*Pul13 bound to amylopectin, pullulan, and amylose but not to dextran [linear α-(1,6)-linked glucose] as assessed by affinity gel electrophoresis (AGE) binding experiments (*13*) (Table 1). Neither, X45 nor X20 bound to any of these polysaccharides. These results clearly indicated that X28 is a CBM having affinity for α-glucans with α-(1,4) or mixed α-(1,4)(1,6) linkages but has no affinity for α-glucans of α-(1,6)-linked glucose. On the basis of amino acid sequence comparisons, X28 and its

Table 1: Qualitative Assessment of Binding of *Tm*Pul13 and Its Modules to α-Glucans Determined by Affinity Electrophoresis

| protein | amylopectin | amylose | pullulan | dextran |
|---|---|---|---|---|
| *Tm*Pul13 | +[a] | + | + | −[b] |
| X28/X45/X20 | + | + | + | − |
| *Tm*CBM41 (X28) | + | + | + | − |
| X45 | − | − | − | − |
| X20 | − | − | − | − |

[a] (+) indicates binding. [b] (−) indicates no binding.

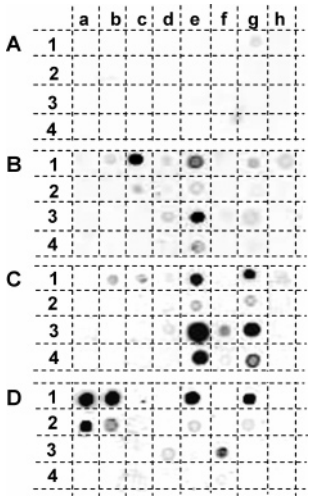


FIGURE 2: Polysaccharide macroarray binding analysis of Alexa Fluor 680 labeled *Tm*CBM41. (A) BSA (negative control), (B) *Cc*CBM17 (positive control), (C) *Tm*CBM4-2 (positive control), and (D) *Tm*CBM41. Proteins were fluorescently labeled, and blots were prepared and probed as described in Materials and Methods. Polysaccharides tested include (1a) 1% amylopectin, (1b) 1% pullulan, (1c) 1% glucomannan, (1d) 1% galactomannan, (1e) 1% pectic galactan, (1f) 1% arabinogalactan, (1g) 1% wheat arabinoxylan, (1h) 1% xyloglucan; (2a−2h) polysaccharides in the same order except at 0.1%; (3a) blank, (3b) 1% dextran, (3c) 1% pachyman, (3d) 1% laminarin, (3e) 1% oat *β*-glucan, (3f) 1% birchwood xylan, (3g) 1% methylcellulose, (3h) 0.1% glycerol; (4a−4g) polysaccharides in the same order except at 0.1%, (4h) blank. Both positive controls displayed the expected binding specificity based on previous studies (*15*, *24*) except that *Tm*CBM4-2 did not appear to bind laminarin, suggesting that this short and soluble polysaccharide was not effectively immobilized on the membrane.

homologues comprise a new family of CBMs now classified as CBM family 41. Thus, the X28 module from *Tm*Pul13 will be referred to hereafter as *Tm*CBM41. The functions of the X45 and X20 modules remain to be determined.

A more complete analysis of *Tm*CBM41's specificity to polysaccharides was performed by a macroarray binding assay (Figure 2). This confirmed binding to amylopectin and pullulan and the lack of binding to dextran. *Tm*CBM41 also displayed binding to pectin galactan, wheat arabinoxylan, and birchwood xylan, revealing cross-specificity for other polysaccharides. However, the binding to these polysaccharides was sufficiently weak to be unquantifiable by ITC. Thus, *Tm*CBM41 has a clear preference for α-glucans. Curiously, *Tm*CBM4-2 did not appear to bind to laminarin [a *β*-(1,3)-glucan], for which *Tm*CBM4-2 has a very high affinity, unfortunately suggesting that this relatively short and soluble polysaccharide was not effectively immobilized on the membrane.

*A Kinetic Model of Oligosaccharide Recognition by TmCBM41*. The addition of maltooligosaccharides to *Tm*CBM41 resulted in relatively large changes in the UV spectra as shown in the UV difference spectra (not shown). Large peaks at ∼293, 285, and 275 nm and troughs at ∼289 and 278 nm are distinctive of the involvement of tryptophan side chains in binding (*15*, *24*). The lack of the distinctive tyrosine peak at 286.5 nm suggested no involvement of tyrosine residues in binding; however, any tyrosine signal may have been masked by the large tryptophan signal.

The peak-to-trough heights at wavelength pairs 292.72/288.82 nm, 285.06/288.82 nm, and 285.06/278.24 nm were dependent on the concentration of added carbohydrate and were used to quantitatively assess the binding interactions (Figure 3). Preliminary fitting of the binding isotherms to a binding model assuming 1:1 interactions (equilibrium 1 in Figure 4) yielded two observations. First, the affinities for maltotriose (M3) up to maltoheptaose (M7) were in the vicinity of $5 \times 10^5$ to $2 \times 10^6$ $M^{-1}$. Second, the stoichiometry of binding appeared to steadily decrease from a 1:1 carbohydrate:protein ratio in the case of M3 and M4 to ∼0.5:1 for M7. Because the concentration of protein used for these experiments was 16−70-fold in excess of the dissociation constants, the binding isotherms at total ligand concentrations less than the CBM concentration (∼33 $\mu$M) should approximate pseudo-first-order reactions. Thus, the change in stoichiometry with ligand length was visually evident in the different initial slopes of the binding isotherms (Figure 3). This indicated that some of the interactions were not simple 1:1 interactions. The 1:2 carbohydrate:protein stoichiometry for M7 binding is reminiscent of observations made for the family 27 CBM from *T. maritima* binding to mannohexaose (*25*) and a mutant of the family 29 CBM from *Pyromyces equi* binding to cellohexaose (*26*). In both cases the hexasaccharides were able to function as divalent acceptors, as in equilibrium 2 (see Figure 4). Such a model could explain the interaction of *Tm*CBM41 with maltooligosaccharides. Another possibility is that *Tm*CBM41 preassociates to form a dimer prior to binding carbohydrate, resulting in a 1:2 stoichiometry (equilibrium 3 in Figure 4). The fits of the data to three binding models (equilibria 1−3 in Figure 4) were evaluated with the program DynaFit (*17*) in order to discriminate between the possible binding mechanisms.

The fits of all of the data using equilibrium 3 were extremely poor, and the model was immediately rejected on a statistical basis (large sum-of-squares values and clear systematic deviations between the model data and the experimental data; not shown). The data for $6^3$-α-D-glucosylmaltotriose [GM3; glucose linked α-(1−4) to maltotriose], $6^3$-α-D-glucosylmaltotriosylmaltotriose [GM3M3; maltotriose linked α-(1−4) to GM3], maltose (M2), maltotriose (M3), and M4 were adequately described by the bimolecular model of equilibrium 1. Trial of equilibrium 2 failed to reduce the sum of the squares of the fits. Furthermore, *P*-values of fit comparisons indicated a greater than 99% chance that the single binding site model (equilibrium 1) was more appropriate for these carbohydrate ligands. The near unitary stoichiometries obtained from the fits were consistent with this (Table 2). In contrast, the M5−M7 isotherms were not well modeled by equilibrium 1. The application of equilibrium 2 greatly reduced the sum of the squares of the fits. The residuals of the fits to equilibrium 2 (Figure 3C) and equilibrium 1 (Figure 3D) show random scatter for equilibrium 2 but large systematic deviations for equilibrium 1.

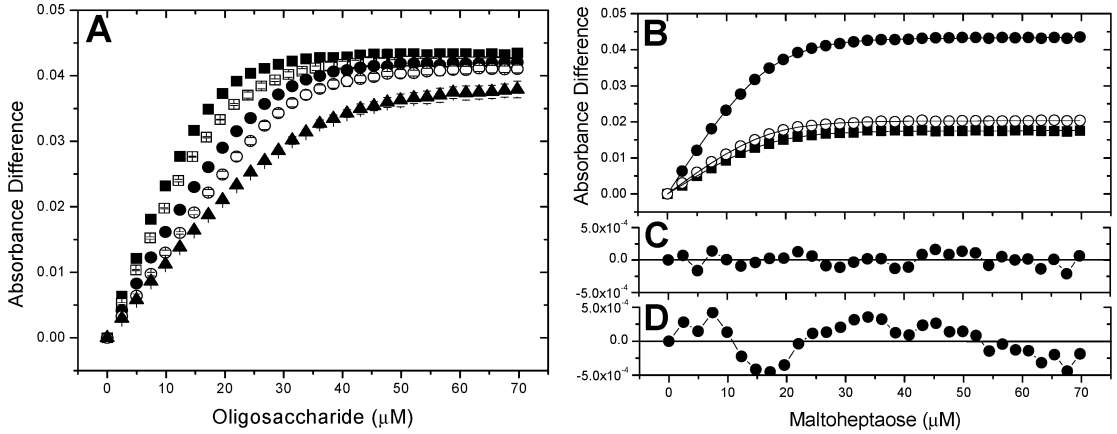


FIGURE 3: Quantitative UV difference analysis of *Tm*CBM41 binding to α-glucooligosaccharides. Panel A: Isotherms of M3 (closed triangles), M4 (open circles), M5 (closed circles), M6 (open squares), and M7 (closed squares) titrated into *Tm*CBM41. The absorbance difference shown is the peak to trough height at 292.72 and 288.82 nm, respectively. Panel B: Isotherm of M7 titrated into *Tm*CBM41. The curves show the data at the wavelength pairs of 292.72/288.82 nm (closed circles), 285.06/288.82 nm (open circles), and 285.06/278.24 nm (closed squares). Solid lines show the fits resulting from the application of equilibrium 2 (see Figure 4). Panels C and D show the residuals resulting from the fits of equilibrium 2 and equilibrium 1, respectively.

Equilibrium 1:

$$M + L \rightleftharpoons LM$$

Equilibrium 2:

$$M + L \rightleftharpoons LM$$

$$LM + M \rightleftharpoons LM_2$$

Equilibrium 3:

$$M + M \rightleftharpoons M_2$$

$$M_2 + L \rightleftharpoons LM_2$$

FIGURE 4: Equilibria used to model the interactions of *Tm*CBM41 with α-glucooligosaccharides. M represents *Tm*CBM41 and L the α-glucooligosaccharide. LM represents the 1:1 complex of *Tm*-CBM41 and the α-glucooligosaccharide while $LM_2$ represents the α-glucooligosaccharide with two bound *Tm*CBM41 molecules. $M_2$ represents a *Tm*CBM41 dimer.

Statistical comparison of equilibrium 1 vs equilibrium 2 indicated an insignificantly small probability (<1%) that equilibrium 1 was the more appropriate model. Thus, equilibrium 2 was deemed the best description of the interaction of *Tm*CBM41 with M5, M6, and M7.

Experiments performed by isothermal titration calorimetry (ITC) yielded the same conclusions. Titrations performed by titrating GM3, GM3M3, M2, M3, and M4 into *Tm*CBM41 gave isotherms of the expected sigmoidal shape (Figure 5A). These isotherms could be analyzed in a symmetrical manner. That is, when the same isotherm was fitted using a one-site binding model (equilibrium 1, Figure 4), the same results were obtained if the protein was treated as the macromolecule (M) and the carbohydrate as the ligand (L) and when reversing these assignments (results not shown). Furthermore, a reverse titration of *Tm*CBM41 into M4 gave essentially the same results as the titration of M4 into protein (Table 2, Figure 5B), confirming the symmetry of the system and the suitability of using a single binding site model (equilibrium 1). As would be expected, analysis of the experiments using these ligands gave stoichiometries near unity (Table 3). The isotherms of M5, M6, and M7 titrated into *Tm*CBM41 were not sigmoidal (Figure 5C), and analyses using equilibrium 1 gave extremely poor *P*-values (<0.005) from run test analysis of the residuals, indicating strong systematic deviations between the data and model. The reverse titrations of *Tm*CBM41 into M5 (Figure 5D) and M6 were essentially sigmoidal but also did not fit a single site binding model as judged by the same statistical tests. Analyses that treated *Tm*CBM41 as having a single binding site and the carbohydrate as having two different and independent binding sites (i.e., equilibrium 2) gave good agreement between the model and the data. Thus, the ITC results were in accord with the UV difference binding studies. The binding of $\beta$-cyclodextrin (cycloheptaamylose), however, was a more ambiguous case. $\beta$-Cyclodextrin clearly presented two binding sites, each accommodating a single CBM. However, these binding sites appeared to be so similar in affinity that the individual parameters for binding these sites could not be resolved in these experiments. Thus, the two binding sites were treated as independent but, unlike with M5−M7, identical in the analysis.

The recognition of all of the oligosaccharides tested, other than M5−M7, appears to adhere to equilibrium 1 whereas the recognition M5−M7 appears to go by equilibrium 2. Should this be the case, then at subsaturating concentrations of M5−M7 there should be a detectable concentration of dimeric species, but at excess M5−M7 concentrations, where the singly ligated high-affinity species is preferred, there should be no dimeric species. Likewise, in the absence of ligand or with any concentration of the other oligosaccharides (i.e., M2−M4, GM3, and GM3M3) no dimers should be formed. Sedimentation equilibrium analytical ultracentrifugation analysis of *Tm*CBM41 revealed only monomeric species in the absence of ligand, with a 1:2 molar ratio of M3 to protein, with a 5:1 molar ratio of M3 to protein, and with a 5:1 molar ratio of M6 to protein (Figure 6). The determined molecular masses were 15670, 15780, 16310, and 16700 Da for these species, respectively. These values are in reasonably good agreement with the expected molecular mass of 15391 Da for monomeric *Tm*CBM41, 15895 Da for monomeric *Tm*CBM41 + M3, and 16382 Da for monomeric *Tm*CBM41 + M6. In contrast, the experiments performed with a 1:2 molar ratio of M6 to protein yielded data that was inconsistent with the presence of a single monomeric molecular species. These data could be satisfactorily modeled using a two-component system representing a monomer−dimer equilibrium of a 16370 Da species (32740 Da for the dimer), suggesting an association constant of dimerization estimated to be $(2-4) \times 10^4\ M^{-1}$, in acceptable

Table 2: Parameters of *Tm*CBM41 Binding to Maltooligosaccharides Determined by UV Difference Titrations at 25 °C in 50 mM Tris, pH 7.5

| ligand[a] | species formed[b] | $K_a$ ($\times 10^5$ $M^{-1}$) | $\Delta G$ (kcal/mol) | $\Delta A^c$ ($\times 10^3$ units $\mu mol^{-1} L^{-1}$) | $n^d$ |
|---|---|---|---|---|---|
| M2 | LM | 0.53 (±0.01) | −6.45 (±0.03) | 1.15 (±0.02) | 1.00 |
| M3 | LM | 4.26 (±0.20) | −7.68 (±0.03) | 1.23 (±0.03) | 1.00 (±0.02) |
| M4 | LM | 11.18 (±1.36) | −8.25 (±0.07) | 1.24 (±0.05) | 0.92 (±0.02) |
| M5 | LM | 22.72 (±4.56) | −8.40 (±0.12) | 1.25 (±0.07) | 1.02 (±0.05) |
| | $LM_2$ | 0.25 (±0.13) | −6.02 (±0.33) | 1.26 (±0. 77) | |
| M6 | LM | 22.59 (±6.81) | −8.29 (±0.41) | 1.20 (±0.04) | 1.05 (±0.03) |
| | $LM_2$ | 0.49 (±0.21) | −6.35 (±0.40) | 1.68 (±0.51) | |
| M7 | LM | 14.58 (±12.99) | −8.01 (±0.47) | 1.25 (±0.04) | 1.02 (±0.05) |
| | $LM_2$ | 1.22 (±0.07) | −6.73 (±0.04) | 1.64 (±0.11) | |
| GM3 | LM | 0.27 (±0.01) | −6.04 (±0.01) | 1.25 (±0.03) | 1.00 |
| GM3M3 | LM | 1.79 (±0.14) | −7.16 (±0.05) | 1.20 (±0.02) | 0.85 (±0.01) |

[a] Abbreviations: M2, maltose; M3, maltotriose; M4, maltotetraose; M5, maltopentaose; M6, maltohexaose; M7, maltoheptaose; GM3, $6^3$-α-D-glucosylmaltotriose; GM3M3, $6^3$-α-D-glucosylmaltotriosyl maltotriose. [b] GM3, GM3M3, and M2−M4 were fit to a one-site binding model (see equilibrium 1 in Figure 4); M5−M7 were fit to a binding model treating the carbohydrate as divalent (see equilibrium 2 in Figure 4). Notation corresponds to the species shown in the schemes representing the respective equilibria. [c] $\Delta A$ = molar UV difference response using the peak-to-trough height at 292 and 289 nm. [d] Binding stoichiometry. Where no error is reported, the value was fixed as a constant during the nonlinear fitting. The values for M5−M7 represent the number of binding sites per CBM.

agreement with the association constants estimated by UV difference and ITC for the formation of the maltohexaose ligated dimeric species [i.e., (∼5−10) × $10^4$ $M^{-1}$; Tables 1 and 2]. Thus, these results are entirely consistent with the proposed kinetic model of maltooligosaccharide binding.

On the basis of the dependence of $K_a$ on the length of the oligosaccharides, *Tm*CBM41 requires four to five sugars for maximal binding. Thus, M5, M6, and M7 must comprise up to two, three, and four overlapping and identical binding sites, respectively. Initial occupation of the first of these overlapping binding sites is likely to be the high-affinity interaction. The second binding site must then be some fragment of the maximal four to five sugar binding site, most likely an overhang at a terminus of the oligosaccharide. In strict terms, the two binding sites on the oligosaccharides are nonidentical but dependent: formation of the low-affinity site first depends on occupation of the high-affinity site and possible rearrangement to accommodate binding of the second CBM. The discrepancy between the M5−M7 data obtained in both titration modes likely reflects the inability to account for negative cooperativity. In the reverse titration mode (i.e., protein titrated into sugar) the occupation of the binding sites is sequential due to the initial occupation of the highly favored high-affinity sites, and thus, the approach that we employed to analyze the reverse titration data, which assumed nonidentical but independent binding sites, still provides reasonable approximations. However, at low sugar to protein molar ratios when sugar was titrated into protein, both high-affinity and low-affinity interactions would occur simultaneously, resulting in possibly large negatively cooperative effects. Failure to account for this would cause inaccuracies in determining the binding parameters. This mechanism of binding maltooligosaccharides longer than M4 is the same as the mechanism proposed for *Tm*CBM27 binding to mannohexaose (*25*). A similar mechanism likely also explains the ligand-mediated dimerization of a mutant of *Pe*CBM29 (*26*). However, in the case of the *Pe*CBM29 mutant, the formation of the species with one CBM bound to one sugar occurs with a low affinity. Binding of the second CBM to this complex occurs with a high affinity and stabilizes the doubly ligated species of two CBMs per sugar. Unlike the *Tm*CBM41 scenario, this must involve positive cooperativity whereby binding of the first CBM creates a new binding site of greater affinity for another CBM, either by presentation of a sugar with an optimal binding conformation or by the formation of additional CBM−CBM interactions.

The potential biological significance of these dimerization events is unclear. *Tm*CBM41 does not appear to dimerize upon binding to complex sugars that more accurately represent the biological substrate of *Tm*Pul13 (see below), leaving the potential function of this phenomenon ambiguous. However, manipulation of this occurrence may lead to applications that require tunable dimerization.

*Recognition of Complex Oligosaccharides.* Pullulan is a repeating unit of α-(1,6)-linked M3. To assess the impact of the α-(1,6) linkages on the ability of *Tm*CBM41 to bind α-glucans, we studied the interaction of *Tm*CBM41 with GM3, GM3M3, isomaltose [disaccharide of α-(1,6)-linked glucose], isomaltotriose [trisaccharide of α-(1,6)-linked glucose], and panose [a trisaccharide of glucose with an α-(1,6) linkage followed by an α-(1,4) linkage]. *Tm*CBM41 did not bind to the latter three sugars. The affinity of *Tm*CBM41 for GM3 was similar to that of M2 and an order of magnitude weaker than for M3 and longer maltooligosaccharides (Tables 2 and 3). GM3M3 was only bound as tightly as M3 and comprised a single binding site, unlike similarly sized maltooligosaccharides. These results indicate that *Tm*CBM41 tolerates α-(1,6) linkages, but not particularly well, and only in the context of a sufficient number of α-(1,4)-linked glucose residues. Given the four to five sugar footprint of *Tm*CBM41, this CBM must likely accommodate an internal α-(1,6) linkage giving *Tm*CBM41 the ability to bind to pullulan. However, its preference for α-(1,4) linkages would target *Tm*CBM41 to regions rich in this linkage, possibly leaving the α-(1,6) linkages, the preferred substrate for *Tm*Pul13, exposed and susceptible to cleavage.

*Recognition of Polysaccharides*. Pullulan and amylopectin ITC binding isotherms were obtained by titration of *Tm*-CBM41 into the polysaccharide (not shown). The data for both polysaccharides could be suitably fit using a bimolecular interaction model (equilibrium 1 in Figure 4) where the concentration of the polysaccharide was expressed as equivalents of M4 (see Materials and Methods). Furthermore, Scatchard plots of the data were linear (not shown), further indicating a simple mode of interaction with a single class

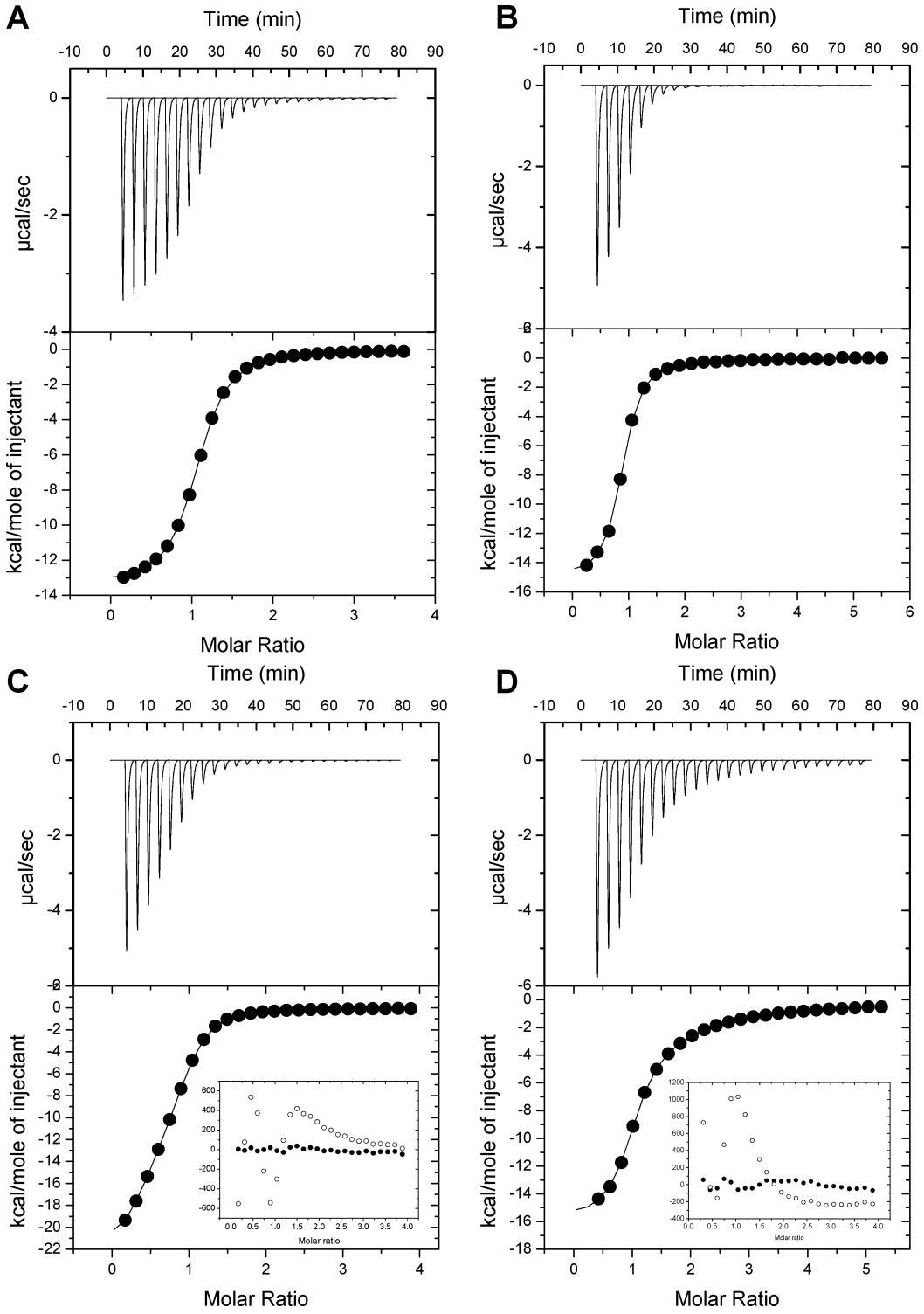


FIGURE 5: Isotherms of *Tm*CBM41 binding to α-glucooligosaccharides produced by ITC. Panels A and B show the isotherms of *Tm*CBM41 interacting with M4; panels C and D show the isotherms of *Tm*CBM41 interacting with M5. The isotherms in panels A and C were produced by titrating M4 or M5, respectively, into *Tm*CBM41. The isotherms in panels B and D were produced by titrating *Tm*CBM41 into M4 or M5, respectively. The solid lines in the lower panes of panels A and B show the fits of a one-site binding model (equilibrium 1). The solid lines in the lower panes of panels C and D show the fits of a two-site binding model where the carbohydrate was treated as the divalent species (equilibrium 2). Insets in panels C and D show plots of the residuals for the model of equilibrium 1 (open circles) and equilibrium 2 (closed circles).

of binding site. The stoichiometry obtained for pullulan (Table 3) equated to one *Tm*CBM41 module binding every ~17 glucose residues, supporting the idea of well-spaced, independent binding sites. This was curiously at odds with the results obtained with GM3M3. One GM3M3 molecule accommodates a single CBM (Tables 2 and 3), suggesting that approximately every seven sugars in pullulan should constitute a binding site. Since the tertiary structure of pullulan in solution is not known, perhaps the decreased number of binding sites in pullulan is due to a conformation adopted by the polysaccharide that prevents binding of *Tm*CBM41 to some of the theoretical binding sites on the

Table 3: Parameters of *Tm*CBM41 Binding α-Glucans Determined by Isothermal Titration Calorimetry at 25 °C in 50 mM Tris, pH 7.5

| ligand[a] | species formed[b] | $K_a$ ($\times 10^5$ $M^{-1}$) | $\Delta G$ (kcal/mol) | $\Delta H$ (kcal/mol) | $\Delta S$ (cal $mol^{-1}$ $K^{-1}$) | $n$[c] |
|---|---|---|---|---|---|---|
| M2 | LM | 0.33 (±0.00) | −5.97 (±0.01) | −13.07 (±0.03) | −23.15 (±0.07) | 1.06 (±0.01) |
| M3 | LM | 2.76 (±0.04) | −7.20 (±0.01) | −14.80 (±0.23) | −24.75 (±0.78) | 0.94 (±0.01) |
| M4 | LM | 5.42 (±0.08) | −7.59 (±0.01) | −13.73 (±0.44) | −19.80 (±1.41) | 1.00 (±0.05) |
| M4 (rev) | LM | 7.46 (±0.19) | −7.77 (±0.01) | −13.42 (±0.17) | −18.15 (±0.49) | 0.83 (±0.00) |
| M5 | LM | 15.75 (±2.62) | −8.20 (±0.10) | −14.76 (±0.18) | −21.15 (±0.92) | 1.06 (±0.10) |
| | $LM_2$ | 0.40 (±0.11) | −6.08 (±0.16) | −7.58 (±0.25) | −4.41 (±1.41) | 1.16 (±0.04) |
| M5 (rev) | LM | 5.32 (±0.32) | −7.57 (±0.03) | −15.99 (±0.32) | −27.40 (±0.57) | 0.98 (±0.04) |
| | $LM_2$ | 0.14 (±0.01) | −5.47 (±0.01) | −8.77 (±2.59) | −10.49 (±8.92) | 1.00 (±0.23) |
| M6 | LM | 10.80 (±0.73) | −7.98 (±0.06) | −13.70 (±0.05) | −18.20 (±0.26) | 1.07 (±0.05) |
| | $LM_2$ | 0.70 (±0.03) | −6.41 (±0.04) | −11.30 (±0.86) | −15.70 (±2.90) | 1.10 (±0.01) |
| M6 (rev) | LM | 20.70 (±1.84) | −8.35 (±0.05) | −14.39 (±0.03) | −19.35 (±0.07) | 0.76 (±0.00) |
| | $LM_2$ | 1.04 (±0.04) | −6.63 (±0.02) | −11.61 (±0.13) | −15.95 (±0.50) | 0.77 (±0.01) |
| M7 | LM | 9.24 (±2.50) | −7.87 (±0.23) | −14.00 (±0.16) | −19.60 (±0.94) | 1.15 (±0.08) |
| | $LM_2$ | 0.69 (±0.18) | −6.38 (±0.22) | −6.65 (±4.33) | −0.18 (±14.50) | 0.67 (±0.22) |
| GM3 | LM | 0.26 (±0.01) | −6.02 (±0.01) | −12.40 (±0.20) | −21.50 (±0.71) | 1.11 (±0.04) |
| GM3M3 | LM | 1.50 (±0.07) | −7.06 (±0.00) | −17.2 (±0.38) | −34.1 (±1.27) | 1.05 (±0.01) |
| β-CD | LM + $LM_2$ | 4.20 (±0.13) | −7.44 (±0.03) | −12.70 (±0.04) | −16.70 (±0.17) | 1.90 (±0.05) |
| β-CD (rev) | LM + $LM_2$ | 3.46 (±0.00) | −7.33 (±0.00) | −10.71 (±0.05) | −10.55 (±0.21) | 1.94 (±0.05) |
| AmPec (rev) | LM | 1.29 (±0.37) | −6.75 (±0.17) | −15.17 (±1.48) | −27.50 (±4.38) | 0.32[d] (±0.04) |
| pullulan (rev) | LM | 1.01 (±0.05) | −6.62 (±0.27) | −16.66 (±0.12) | −32.95 (±0.33) | 0.23[d] (±0.01) |

[a] Abbreviations: M2, maltose; M3, maltotriose; M4, maltotetraose; M5, maltopentaose; M6, maltohexaose; M7, maltoheptaose; GM3, $6^3$-α-D-glucosylmaltotriose; GM3M3, $6^3$-α-D-glucosylmaltotriosyl maltotriose; β-CD, β-cylodextrin; AmPec, amylopectin. Experiments performed by titrating protein into carbohydrate (reverse titrations) are indicated by (rev). [b] GM3, GM3M3, AmPec, pullulan, β-CD, and M2−M4 were fit to a one-site binding model (see equilibrium 1 in Figure 4); M5−M7 were fit to a binding model treating the carbohydrate as divalent (see equilibrium 2 in Figure 4). Notation corresponds to the species shown in the schemes representing the respective equilibria. [c] Binding stoichiometry. The values for M5 and M6 represent the number of binding sites per CBM. [d] Amylopectin and pullulan are polysaccharides of unknown length. These stoichiometries are expressed as moles of CBM per tetraose equivalent (see Materials and Methods).

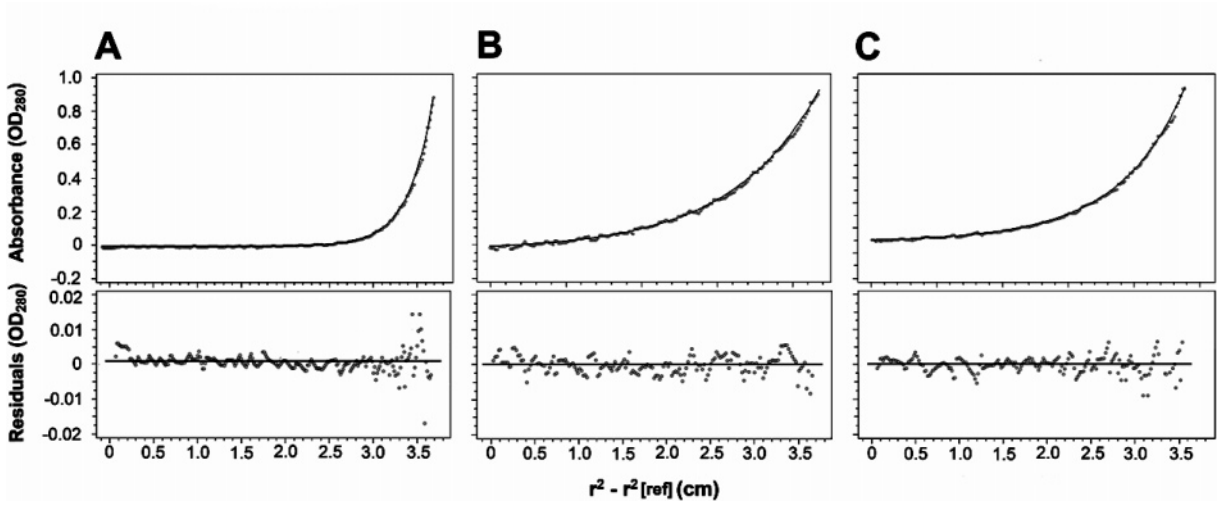


FIGURE 6: Sedimentation equilibrium analysis of *Tm*CBM41 (4.3 μM) in the absence (A) and in the presence of either 2.15 μM maltotriose (B) or 2.15 μM maltohexose (C). The upper plots show the absorbance at 280 nm as a function of the square of the radial distance of the sample at any position within the cell ($r$) minus the square of the radial position at a reference point ($r$[ref]) ($r^2 - r^2$[ref]). The continuous lines in the upper plots were obtained by fitting the experimental data (circles) to a single ideal component model of $M_r$ 15670 (A), to a single ideal component of $M_r$ 16310 (B), and to a monomer−dimer model of monomer $M_r$ 16370 (C). The lower plots show $\chi^2$ residuals as a function of $r^2 - r_0^2$ for the best fit (solid line). The rotor speed of the equilibrium plots shown in this figure was taken at 45000 rpm at 20 °C. The buffer composition was 50 mM Tris-HCl (pH 7.5). The $A_{280}$ of the starting sample was approximately 0.3.

polysaccharide. The stoichiometry for amylopectin equated to one *Tm*CBM41 molecule binding every ∼13 glucose residues (Table 3). The association constants for pullulan and amylopectin were $\sim 10^5$ $M^{-1}$ (Table 3), similar to that of M3 and GM3M3 binding. Considering the composition of pullulan, the similarity in affinities is, perhaps, expected. However, amylopectin contains infrequent α-(1,6) branches and, therefore, contains longer stretches of linear α-(1,4)-linked glucose, leading one to expect affinities upward of those found for M4−M7. Again, this decrease in affinity may be due to a conformation adopted by amylopectin in solution that is nonoptimal for *Tm*CBM41 binding.

*Thermodynamic Mechanism of α-Glucan Binding.* All binding reactions between *Tm*CBM41 and α-glucan substrates reported were favored by a negative change in enthalpy ($\Delta H$) (Table 3). In contrast, all changes in entropy ($\Delta S$) were unfavorable. This signature is common to protein−carbohydrate interactions, CBMs in particular. Though Gibb's free energy of binding ($\Delta G$) generally increased with oligosaccharide length, there was no correla-

Table 4: Proteins Containing Modules Similar to *Tm*CBM41

| organism | protein | GenPept accession number | no. of modules | % identity with *Tm*CBM41 |
|---|---|---|---|---|
| *Thermotoga maritima* | pullulanase | NP_229641 | 1 | 100 |
| *Fervidobacterium pennivorans* | pullulanase type I | AAD30387 | 1 | 72 |
| *Bacillus cereus* | pullulanase | NP_832487 | 1 | 49 |
| *Bacillus anthracis* | pullulanase | NP_845079 | 1 | 46 |
| *Bacillus* sp. | alkaline amylopullulanase | BAA11332 | 3 | 30, 21, 35 |
| *Streptomyces coelicolor* | α-amylase/dextrinase | NP_626477 | 2 | 39, 32 |
| *Micrococcus* sp. | α-amylase | A60999 | 2 | 23, 24 |
| *Bacillus halodurans* | alkaline amylopullulanase | NP_244564 | 2 | 33, 27 |
| *Streptomyces avermitilis* | pullulanase | NP_827159 | 2 | 28, 22 |
| *Streptococcus agalactiae* | pullulanase | NP_687867 | 1 | 31 |
| *Streptococcus agalactiae* | unknown | NP_735320 | 1 | 31 |
| *Actinoplanes* sp. | α-amylase | CAC02970 | 2 | 31, 32 |
| *Bacillus* sp. KSM-1876 | alkaline pullulanase | BAB47586 | 2 | 24, 22 |
| *Streptomyces lividans* | 1,4-α-D-glucan glucanohydrolase, α-amylase precursor | Q05884 | 2 | 31, 40 |
| *Streptococcus pneumoniae* | pullulanase | AAG33958 | 2 | 21, 23 |
| *Streptococcus pneumoniae* TIGR4 | alkaline amylopullulanase | NP_344806 | 2 | 21, 23 |
| *Streptococcus pneumoniae* R6 | alkaline amylopullulanase | NP_357841 | 2 | 21, 23 |
| *Streptococcus pyogenes* | pullulanase | NP_269940 | 2 | 15, 20 |
| *Streptococcus pyogenes* | pullulanase | CAD32942 | 2 | 15, 20 |
| *Streptococcus pyogenes* | pullulanase | NP_665498 | 2 | 15, 20 |
| *Streptococcus pyogenes* | pullulanase | NP_608014 | 2 | 15, 20 |
| *Streptococcus agalactiae* | unknown | NP_735732 | 2 | 12, 18 |
| *Streptococcus agalactiae* | pullulanase | NP_688225 | 2 | 11, 17 |
| *Microbulbifer degradans* | hypothetical protein | ZP_00065715 | 1 | 22 |
| *Vibrio parahaemolyticus* | pullulanase precursor | NP_801148 | 1 | 19 |
| *Streptococcus mutans* | pullulanase | NP_721884 | 1 | 30 |
| *Streptococcus pneumoniae* | thermostable pullulanase | NP_358619 | 1 | 23 |
| *Klebsiella pneumoniae* | α-dextrin endo-1,6-α-glucosidase | P07811 | 1 | 22 |
| *Klebsiella pneumoniae* | pullulanase precursor | P07206 | 1 | 19 |
| *Vibrio vulnificus* | type II secretory pathway protein | NP_763140 | 2 | 22, 13 |
| *Klebsiella aerogenes* | pullulanase precursor | AAA25124 | 1 | 21 |
| *Microbulbifer degradans* | hypothetical protein | ZP_00065687 | 1 | 16 |
| *Klebsiella pneumoniae* | pullulanase precursor | S38801 | 1 | 20 |
| *Deinococcus radiodurans* | α-dextran endo-1,6- α-glucosidase | NP_294128 | 1 | 24 |

tion between oligosaccharide length and changes in $\Delta H$ and $\Delta S$, suggesting different contributions to the thermodynamics from interactions at the different subsites in the CBM binding site.

The interaction of *Tm*CBM41 with GM3M3 was enthalpically more favorable than with other maltooligosaccharides (Table 3); however, the change in entropy for this interaction was comparably worse, resulting in a less favorable $\Delta G$. A similar observation can be made for the interactions with amylopectin and pullulan in comparison with maltooligosaccharides. This latter phenomenon, whereby soluble polysaccharides are bound more weakly than oligosaccharides due to entropic penalties, appears to be common with CBMs (see refs *27*−*30* for examples).

*A New Family of CBMs*. *Tm*CBM41 from *T. maritima* pullulanase Pul13 is a CBM that binds tightly to α-glucans. Modules with significant amino acid sequence identity are found in no fewer than 34 proteins from 22 different bacterial species (Table 4). These modules frequently occur as repeats, almost exclusively in pullulanases. Interestingly, a great number of these pullulanases are from human pathogens. Recently, a membrane-associated pullulanase (GenPept accession number CAD32942) from *Streptococcus pyogenes* strain NZ131rgg was identified as a glycoprotein-specific adhesin (*31*). In light of our study, it appears likely that the tandem N-terminal CBM41 modules of the *S. pyogenes* pullulanase are responsible for its glycoprotein binding activity. Thus, *Tm*CBM41 from *T. maritima* Pul13 is the first well-characterized member of a new, relatively large, and apparently diversely functioning family of carbohydrate-binding modules. This family has now been classified as CBM family 41 (carbohydrate-active enzyme families and CBM families can be accessed through the CAZy server at http://afmb.cnrs-mrs.fr/CAZY/index.html).

## ACKNOWLEDGMENT

The authors thank Jonathan Klassen for assistance.

## REFERENCES

1. Gibson, L. H., and Coughlin, R. W. (2002) Optimization of high molecular weight pullulan production by Aureobasidium pullulans in batch fermentations, *Biotechnol. Prog. 18*, 675−678.
2. Catley, B. J. (1970) Pullulan, a relationship between molecular weight and fine structure, *FEBS Lett. 10*, 190−193.
3. Leathers, T. D. (2003) Biotechnological production and applications of pullulan, *Appl. Microbiol. Biotechnol.*
4. Banas, J. A., and Vickerman, M. M. (2003) Glucan-binding proteins of the oral streptococci, *Crit. Rev. Oral Biol. Med. 14*, 89−99.
5. Meddens, M. J., Thompson, J., Leijh, P. C., and van Furth, R. (1984) Role of granulocytes in the induction of an experimental endocarditis with a dextran-producing *Streptococcus sanguis* and its dextran-negative mutant, *Br. J. Exp. Pathol. 65*, 257−265.
6. Sigurskjold, B. W., Svensson, B., Williamson, G., and Driguez, H. (1994) Thermodynamics of ligand binding to the starch-binding domain of glucoamylase from *Aspergillus niger*, *Eur. J. Biochem 225*, 133−141.
7. Sorimachi, K., Jacks, A. J., Le Gal-Coeffet, M. F., Williamson, G., Archer, D. B., and Williamson, M. P. (1996) Solution structure of the granular starch binding domain of glucoamylase from *Aspergillus niger* by nuclear magnetic resonance spectroscopy, *J. Mol. Biol. 259*, 970−987.

8. Huber, R., Langworthy, T. A., König, H., Thomm, M., Woese, C. R., Sleytr, U. B., and Stetter, K. O. (1986) *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90 °C, *Arch. Microbiol. 144*, 324−333.
9. Kriegshauser, G., and Liebl, W. (2000) Pullulanase from the hyperthermophilic bacterium *Thermotoga maritima*: purification by beta-cyclodextrin affinity chromatography, *J. Chromatogr., B: Biomed. Sci. Appl. 737*, 245−251.
10. Bibel, M., Brettl, C., Gosslar, U., Kriegshauser, G., and Liebl, W. (1998) Isolation and analysis of genes for amylolytic enzymes of the hyperthermophilic bacterium *Thermotoga maritima*, *FEMS Microbiol. Lett. 158*, 9−15.
11. Boraston, A. B., Creagh, A. L., Alam, M. M., Kormos, J. M., Tomme, P., Haynes, C. A., Warren, R. A., and Kilburn, D. G. (2001) Binding specificity and thermodynamics of a family 9 carbohydrate-binding module from *Thermotoga maritima* xylanase 10A, *Biochemistry 40*, 6240−6247.
12. Mach, H., Middaugh, C. R., and Lewis, R. V. (1992) Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins, *Anal. Biochem. 200*, 74−80.
13. Tomme, P., Boraston, A., Kormos, J. M., Warren, R. A., and Kilburn, D. G. (2000) Affinity electrophoresis for the identification and characterization of soluble sugar binding by carbohydrate-binding modules, *Enzyme Microb. Technol. 27*, 453−458.
14. Sigurskjold, B. W., Altman, E., and Bundle, D. R. (1991) Sensitive titration microcalorimetric study of the binding of *Salmonella* O-antigenic oligosaccharides by a monoclonal antibody, *Eur. J. Biochem. 197*, 239−46.
15. Boraston, A. B., Warren, R. A., and Kilburn, D. G. (2001) $\beta$-1,3-Glucan binding by a thermostable carbohydrate-binding module from *Thermotoga maritima*, *Biochemistry 40*, 14679−14685.
16. Boraston, A. B., Ghaffari, M., Warren, R. A., and Kilburn, D. G. (2002) Identification and glucan-binding properties of a new carbohydrate-binding module family, *Biochem J. 361*, 35−40.
17. Kuzmic, P. (1996) Program DYNAFIT for the analysis of enzyme kinetic data: application to HIV proteinase, *Anal. Biochem 237*, 260−273.
18. Straume, M., and Johnson, M. L. (1992) Analysis of residuals: criteria for determining goodness-of-fit, *Methods Enzymol. 210*, 87−105.
19. Straume, M., and Johnson, M. L. (1992) Monte Carlo method for determining complete confidence probability distributions of estimated model parameters, *Methods Enzymol. 210*, 117−129.
20. Cohn, E. J., and Edsall, J. T. (1942) *Proteins, Amino Acids and Peptides*, Reinhold, New York.
21. Perkins, S. J. (1986) Protein volumes and hydration effects. The calculations of partial specific volumes, neutron scattering match-points and 280-nm absorption coefficients for proteins and glycoproteins from amino acid sequences, *Eur. J. Biochem. 157*, 169−180.
22. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs, *Nucleic Acids Res. 25*, 3389−3402.
23. Coutinho, P. M., and Henrissat, B. (1999) in *Recent advances in carbohydrate bioengineering* (Gilbert, H. J., Davies, G. J., Henrissat, B., and Svensson, B., Eds.) pp 3−12, Royal Society of Chemistry, Cambridge.
24. Boraston, A. B., Chiu, P., Warren, R. A. J., and Kilburn, D. G. (2000) Specificity and affinity of substrate binding by a family 17 carbohydrate-binding module from *Clostridium cellulovorans* cellulase 5A, *Biochemistry 39*, 11129−11136.
25. Boraston, A. B., Revett, T. J., Boraston, C. M., Nurizzo, D., and Davies, G. J. (2003) Structural and thermodynamic dissection of specific mannan recognition by a carbohydrate binding module, TmCBM27, *Structure (Cambridge) 11*, 665−675.
26. Flint, J., Nurizzo, D., Harding, S. E., Longman, E., Davies, G. J., Gilbert, H. J., and Bolam, D. N. (2004) Ligand-mediated dimerization of a carbohydrate-binding molecule reveals a novel mechanism for protein-carbohydrate recognition, *J. Mol. Biol. 337*, 417−426.
27. Czjzek, M., Bolam, D. N., Mosbah, A., Allouch, J., Fontes, C. M., Ferreira, L. M., Bornet, O., Zamboni, V., Darbon, H., Smith, N. L., Black, G. W., Henrissat, B., and Gilbert, H. J. (2001) The location of the ligand-binding site of carbohydrate-binding modules that have evolved from a common sequence is not conserved, *J. Biol. Chem 276*, 48580−48587.
28. Szabo, L., Jamal, S., Xie, H., Charnock, S. J., Bolam, D. N., Gilbert, H. J., and Davies, G. J. (2001) Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation, *J. Biol. Chem. 276*, 49061−49065.
29. Charnock, S. J., Bolam, D. N., Nurizzo, D., Szabo, L., McKie, V. A., Gilbert, H. J., and Davies, G. J. (2002) Promiscuity in ligand-binding: The three-dimensional structure of a *Piromyces* carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose, *Proc. Natl. Acad. Sci. U.S.A. 99*, 14077−14082.
30. Charnock, S. J., Bolam, D. N., Turkenburg, J. P., Gilbert, H. J., Ferreira, L. M., Davies, G. J., and Fontes, C. M. (2000) The X6 "thermostabilizing" domains of xylanases are carbohydrate-binding modules: structure and biochemistry of the *Clostridium thermocellum* X6b domain, *Biochemistry 39*, 5013−5021.
31. Hytonen, J., Haataja, S., and Finne, J. (2003) *Streptococcus pyogenes* glycoprotein-binding strepadhesin activity is mediated by a surface-associated carbohydrate-degrading enzyme, pullulanase, *Infect. Immun. 71*, 784−793.

BI048215Z