Molecular Microbiology (2018) 107(2), 249–264 ■
doi:10.1111/mmi.13881
First published online 1 December 2017

Novel carbohydrate binding modules in the surface anchored α-amylase of *Eubacterium rectale* provide a molecular rationale for the range of starches used by this organism in the human gut

Darrell W. Cockburn,[1,2] Carolyn Suh,[1] Krizia Perez Medina,[1] Rebecca M. Duvall,[1] Zdzislaw Wawrzak,[3] Bernard Henrissat[4,5,6] and Nicole M. Koropatkin[1]*

[1]*Department of Microbiology and Immunology, University of Michigan Medical School, Ann Arbor, MI 48109, USA.*
[2]*Department of Food Science, Pennsylvania State University, University Park, PA 16802, USA.*
[3]*Life Sciences Collaborative Access Team (LS-CAT), Advanced Photon Source, Argonne National Laboratory, 9700 S. Cass Avenue, Argonne, IL, 60439, USA.*
[4]*Architecture et Fonction des Macromolécules Biologiques, CNRS, Aix-Marseille University, Marseille, F-13288, France.*
[5]*Institut National de la Recherche Agronomique, USC1408 Architecture et Fonction des Macromolécules Biologiques, Marseille, F-13288, France.*
[6]*Department of Biological Sciences, King Abdulaziz University, Jeddah, 21589, Saudi Arabia.*

Summary

Gut bacteria recognize accessible glycan substrates within a complex environment. Carbohydrate binding modules (CBMs) of cell surface glycoside hydrolases often drive binding to the target substrate. *Eubacterium rectale*, an important butyrate-producing organism in the gut, consumes a limited range of substrates, including starch. Host consumption of resistant starch increases the abundance of *E. rectale* in the intestine, likely because it successfully captures the products of resistant starch degradation by other bacteria. Here, we demonstrate that the cell wall anchored starch-degrading α-amylase, Amy13K of *E. rectale* harbors five CBMs that all target starch with differing specificities. Intriguingly these CBMs efficiently bind to both regular and high amylose corn starch (a type of resistant starch), but have almost no affinity for potato starch (another type of resistant starch). Removal of these CBMs from Amy13K reduces the activity level of the enzyme toward corn starches by ~40-fold, down to the level of activity toward potato starch, suggesting that the CBMs facilitate activity on corn starch and allow its utilization *in vivo*. The specificity of the Amy13K CBMs provides a molecular rationale for why *E. rectale* is able to only use certain starch types without the aid of other organisms.

Accepted 11 November, 2017. *For correspondence. E-mail nkoropat@umich.edu; Tel. (+1) 734 647 5718; Fax (+1) 734 764 3562.

Introduction

The human gut microbiota consists of trillions of individual bacteria and the interaction of this dense microbial population with our diet and other environmental factors is an important determinant of our health. A healthy microbiome is protective against a number of conditions including colon cancer (Zackular *et al.*, 2013), inflammatory bowel disease (Rajilic-Stojanovic *et al.*, 2013; De Cruz *et al.*, 2015), diabetes (Forslund *et al.*, 2015) and obesity (Ridaura *et al.*, 2013). Saccharolytic gut bacteria offer particular benefits, persisting in the host through the fermentation of fiber, carbohydrates that human enzymes are unable to process (Shanahan *et al.*, 2017). One prominent fiber in the human diet is resistant starch, starches that for a variety of reasons are indigestible by human enzymes, but are susceptible to attack by certain microorganisms (Birt *et al.*, 2013). Uncooked potato starch is one such resistant starch as it adopts an alternative crystal structure, known as the B-type structure, than that of the more easily digestible wheat and corn starches (Imberty *et al.*, 1991). In corn, certain mutations result in a higher relative abundance of amylose in the starch granules. This high amylose

corn starch also adopts the B-type crystalline structure and is a resistant starch (Gallant *et al.*, 1992). The end result of resistant starch and other carbohydrate fermentation in the gut is often organic acids, particularly the short chain fatty acids (SCFA) acetate, propionate and butyrate (Rios-Covian *et al.*, 2016).

While SCFAs have been shown to influence our physiology (Berggren *et al.*, 1996; Wong *et al.*, 2006; Boets *et al.*, 2017), butyrate has been particularly noted for its health promoting effects (Guilloteau *et al.*, 2010). This SCFA can provide as much as 10% of our daily caloric intake (McNeil, 1984) and it is the preferred energy source of colonocytes (Roediger, 1980). Butyrate increases the rate of proliferation of colonocytes and strengthens tight junctions (Wang *et al.*, 2012), improving gut barrier function. It increases the rate of apoptosis for malignant cells, protecting against colon cancer (Fung *et al.*, 2012). Butyrate also downregulates the expression of pro-inflammatory cytokines, leading to lower levels of inflammation in the gut (Nastasi *et al.*, 2015). When processed in the liver, butyrate shifts glucose metabolism toward storage as glycogen (Beauvieux *et al.*, 2008), thereby protecting against the development of diabetes. Thus, butyrate is clearly a critical regulator of health making it important to understand the unique physiology of the bacteria responsible for its production.

One of the most prominent groups of butyrate-producing organisms in the gut is the cluster XIVa clostridia, exemplified by one of their most abundant members, *Eubacterium rectale*. This Gram positive organism has long been recognized as a dominant species in the human gut (Gossling and Slack, 1974) and a core member of the healthy microbiome (Tap *et al.*, 2009). *E. rectale* decreases in abundance in a number of disease states, including obesity (Haro *et al.*, 2016), inflammatory bowel disease (Kang *et al.*, 2010; Rajilic-Stojanovic *et al.*, 2013), diabetes (Qin *et al.*, 2012) and cystic fibrosis (Bruzzese *et al.*, 2014). Intriguingly, *E. rectale* levels are found to increase, along with butyrate levels, in diets rich in resistant starch (Martínez *et al.*, 2010; Martínez *et al.*, 2013). However, *in vitro* studies indicate that it is unable to directly use resistant starch, although it grows robustly in the presence of a primary resistant starch degrader such as *Ruminococcus bromii* (Ze *et al.*, 2012). Our recent study of the cell wall and membrane proteome of *E. rectale* when grown on starch as compared to glucose, revealed that two ABC transporters that target different maltooligosaccharides, along with two amylases were strongly upregulated in the presence of starch. Thus, we proposed a model by which the larger cell surface amylase EUR_21100 plays a crucial role in the organism's growth on starch, cleaving starch molecules into maltotetraose and larger oligosaccharides, which is directly bound by the ABC transporter solute-binding protein EUR_01830 (Cockburn *et al.*, 2015b). Here, we present a structural and functional characterization of the cell surface amylase EUR_21100, which we have renamed Amy13K and demonstrate empirically that Amy13K contains five discrete starch-binding CBMs that establish two new CBM families and are critical for starch processing. These new CBM families exhibit an extremely narrow taxonomic distribution, suggesting that they are highly adapted to the niche of *E. rectale* in the human gut. These CBMs effectively target corn starch, including high amylose corn starch, but bind poorly to potato starch, explaining the weak activity of the enzyme against this substrate and why the organism cannot grow on resistant potato starch.

Results

Amy13K harbors CBMs that define novel families

We previously reported that *E. rectale* Amy13K (EUR_21100) was likely comprised of five CBMs at its N-terminus based on weak sequence homology to the starch-binding families CBM26 (BLAST *E*-value $2e^{-5}$) and CBM41 (BLAST *E*-value $1e^{-5}$ – $5e^{-7}$) (Cockburn *et al.*, 2015b). The biochemical and structural data presented here supports that there are five CBMs, labeled as CBMa-e and two warrant classification into new CBM families (Fig. 1) within the Carbohydrate Active enZymes database: www.cazy.org (Lombard *et al.*, 2014). CBMb and CBMc showed similarity to several CBM26 modules (Supporting Information Fig. S1) allowing them to be placed in this family. Similarly CBMd showed relatedness to several CBMs classified as CBM41 and, thus, was assigned to that family (Supporting Information Fig. S2). CBMa and CBMe did not show similarity to known CBM families or to each other. A BLAST search was then conducted against full length proteins in CAZy to identify similar domains. CBMa and its homologs (Supporting Information Fig. S3) were classified in a new CBM family called CBM82 while CBMe and its homologs define family CBM83 (Supporting Information Fig. S4).

The sequences of CBMb and CBMc identify them as members of the CBM26 family and their structures presented here (later in Fig. 3) point to their structural relationship to members of this family as well. The CBM26 domains are typically associated with α-amylases, including enzymes from bacteria related to *E. rectale* (Ramsay *et al.*, 2006). According to Pfam (pfam.xfam.org) CBM26s occur in tandem repeats approximately one third of the time. Conversely, CBMd can be placed within the CBM41 family, although it is a somewhat

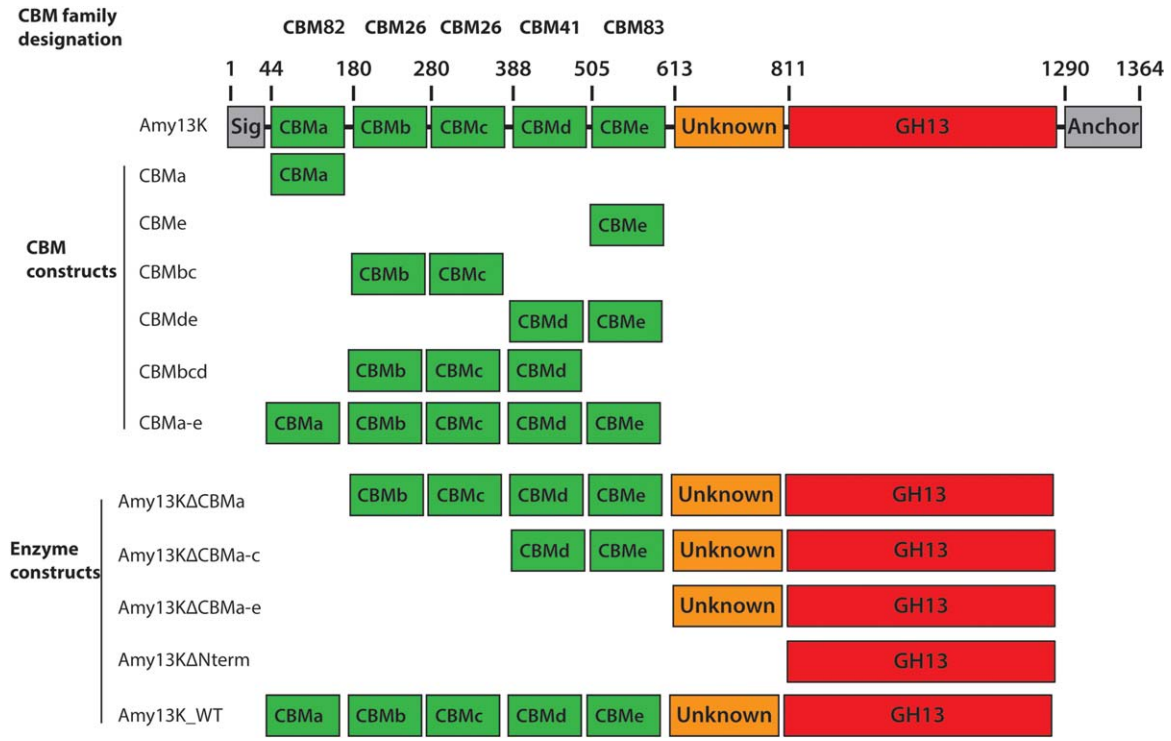

Fig. 1. Domain organization of Amy13K. The signal sequence and cell wall anchor are indicated in gray, CBMs are indicated in green and the catalytic domain is indicated in red. The numbers along the length of the protein represent the start points of the domains. CBM containing constructs and enzymatic constructs used within this study are shown below. At the top, the CAZy classification for each CBM, including two new families designated as CBM82 and CBM83, is noted.

distant relative, exhibiting only a 28% sequence identity with its closest relative within the family. Despite this, it can be placed within subfamily 5 of this group, which includes the *T. maritima* PulA CBM41, its closest structural homologue (Janecek *et al.*, 2017). This group is characterized by the pattern of its aromatic residues with W-X-W-∼30aa-W, with the first tryptophan acting as a hydrogen-bonding residue, while the second and third form the aromatic binding platform for starch recognition. These residues are W416, W418 and W469 in CBMd. While the 51 amino acid distance between the second and third tryptophan is larger than the typical distance seen in this subfamily, there is some variability in this distance (Janecek *et al.*, 2017). This atypical distance and the overall low sequence identity with its fellow family members may be due to the fact that there are no prior examples of CBM41s from organisms closely related to *E. rectale*, as the CBMs in this family cluster along taxonomic lines (Janecek *et al.*, 2017). Intriguingly Pfam indicates that CBM41s occur as tandem pairs a slight majority of the time. This suggests that the new CBM82 family may be evolutionarily related to CBM41 and has diverged over time. In general the CBM41 family of binding domains is typically associated with pullulanases, that is, α-1,6 specific enzymes. The catalytic domain of Amy13K is related to the pullulanase subfamilies (GH13_12 and GH13_14) (Cockburn *et al.*, 2015b; Møller *et al.*, 2016), although the enzyme itself is an α-amylase, that is, α-1,4 specific (Cockburn *et al.*, 2015b) and assigned to subfamily GH13_41. Interestingly the GH13_41 domains are typically found in multidomain proteins in conjunction with one of the pullulanase families mentioned above (Møller *et al.*, 2016) suggesting GH13_41 may have arisen from duplication of the pullulanase domains followed by further evolution or vice versa.

Intriguingly CBMa and CBMe each represent novel, previously uncharacterized CBM families. Their narrow distribution among similar gut bacteria within the Lachnospiraceae points to the highly specialized nature of these binding modules. Having a single large enzyme with a variety of adapted starch specific CBMs seems to be employed by a number of members of the Lachnospiraceae. Both *Roseburia inulinovorans* and *Butyrivibrio fibrosolvens* possess a large cell-associated amylase (Ramsay *et al.*, 2006). The catalytic domain of Amy13K and *R. inulinovorans* Amy13a both belong to the GH13_41 subfamily and the Amy13K_CBMa has a strong resemblance to the N-terminal R1 and R2 domain of Amy13a and together are part of the newly defined family CBM82 (Supporting Information Fig. S3). Additionally, the previously described PUD domain (now

CBM41) of Amy13a, is similar to Amy13K_CBMd and both also have CBMs that are part of the new CBM83 family (Supporting Information Fig. S4). While both enzymes have large N-terminal regions upstream of their catalytic domains, these regions have little sequence similarity other than the domains already mentioned. In contrast the *B. fibrosolvens* protein has an entirely different domain organization with the catalytic domain at the C-terminus, followed by a pair of CBM26 domains. Thus, members of this family of bacteria seem to have diverged over time with regards to their machinery for starch digestion, perhaps as part of their segregation into subtly different niches within the gut.

Crystal structures of Amy13K CBMd (CBM41) and CBMbc (CBM26)

Crystallization trials of CBMde yielded crystals of CBMd alone after several months, suggesting flexibility between the domains inhibited crystal formation until proteolysis occurred. Attempts to produce crystals of CBMe alone were unsuccessful. The structure of Amy13K_CBMd was solved to a resolution of 2.20 Å (R_{work} = 23.4%, R_{free} = 25.7%) Table 1 revealing a β-sandwich fold like other CBM41 structures (Fig. 2A). A search of the DALI server suggests that the closest structural relatives of CBMd are the CBM41 domain of *Thermotoga maritima* pullulanase PulA (PDB 2J73, *Z*-score = 10.6) (Lammerts van Bueren and Boraston, 2007), which also shares 22% sequence identity and the CBM41 domains of *Streptococcus pneumoniae* alkaline amylopullanase SpuA (PDB 2J44, *Z*-score = 9.3) (Lammerts Van Bueren *et al.*, 2004b). Overall the secondary structure elements of CBMd align well with those of the CBM41s from PulA and SpuA with the major differences confined to loop regions. An overlay of the structure of CBM41d with that of CBM41 in PulA with bound maltotriose (PDB 2J73) identified the putative starch-binding site by conservation with similar starch-binding residues in PulA (Fig. 2B). In PulA, the aromatic platform comprised of W29 and W73 overlays well with residues W418 and W469 of CBMd. Likewise in CBMd, additional hydrogen-bonding to the hydroxyl oxygens of adjacent glucose residues may be supported by W416, K460 and D477, which are present as W27, K76 and D81 in PulA. As seen in many starch-binding proteins, the aromatic binding platform forms the classic convex angle that matches the helical pitch of amylose and amylopectin chains (Imberty *et al.*, 1991). This orientation is seen in diverse starch-binding proteins ranging from dedicated starch-binding proteins like SusD (Koropatkin *et al.*, 2008), to CBMs (Boraston *et al.*, 2006) and surface binding sites on amylolytic enzymes (Cockburn and Svensson, 2016).

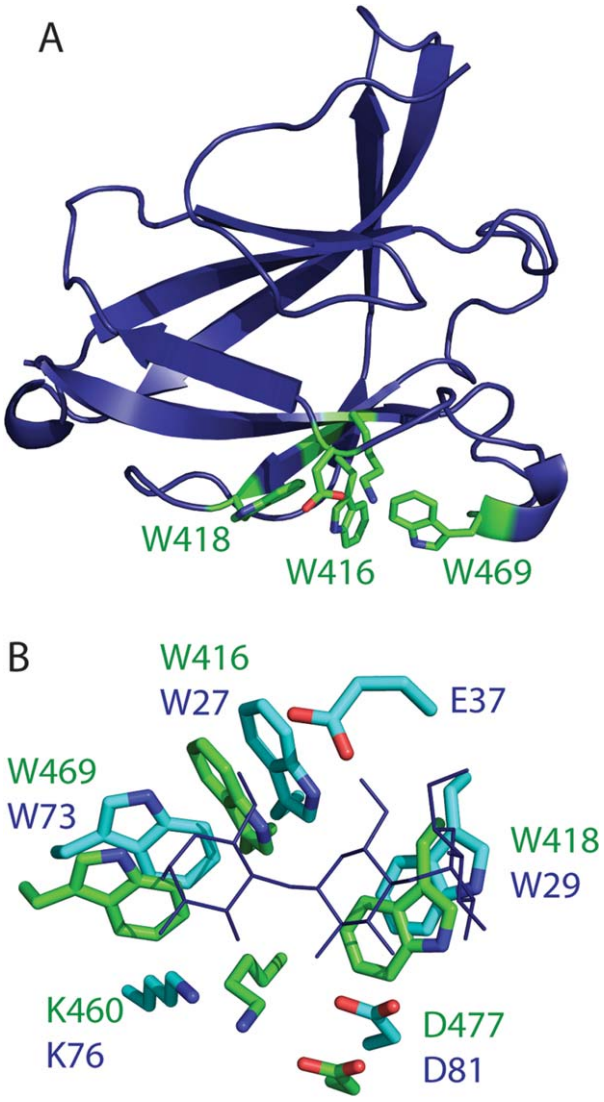

Fig. 2. Crystal structure of Amy13K CBM41d.
A. Cartoon diagram of Amy13K_CBMd (PDB 6AZ5). The putative binding site residues are shown in green and the tryptophans of the binding site are labeled for orientation purposes.
B. An overlay of the carbohydrate binding sites of Amy13K_CBMd and the CBM41 from *T. maritima* pullulanase PulA (PDB 2J73). CBMd binding residues and labels are shown in green, while PulA CBM41 binding residues, bound carbohydrate and labels are shown in blue.

We determined the crystal structure of CBMbc with maltoheptaose (2.01 Å, R_{work} = 18.9%, R_{free} = 20.0%) and without (2.10 Å, R_{work} = 20.8%, R_{free} = 23.2%; Table 1 Fig. 3A). In the substrate bound structure, the maltoheptaose molecule bound to CBMb spans across adjacent asymmetric units and likely facilitated crystallization. The structures of the free and maltoheptaose bound proteins overlay well with an RMSD <0.4 Å for all Cα atoms, and, thus, no structural change occurs on ligand binding. Both CBMb and CBMc display bound

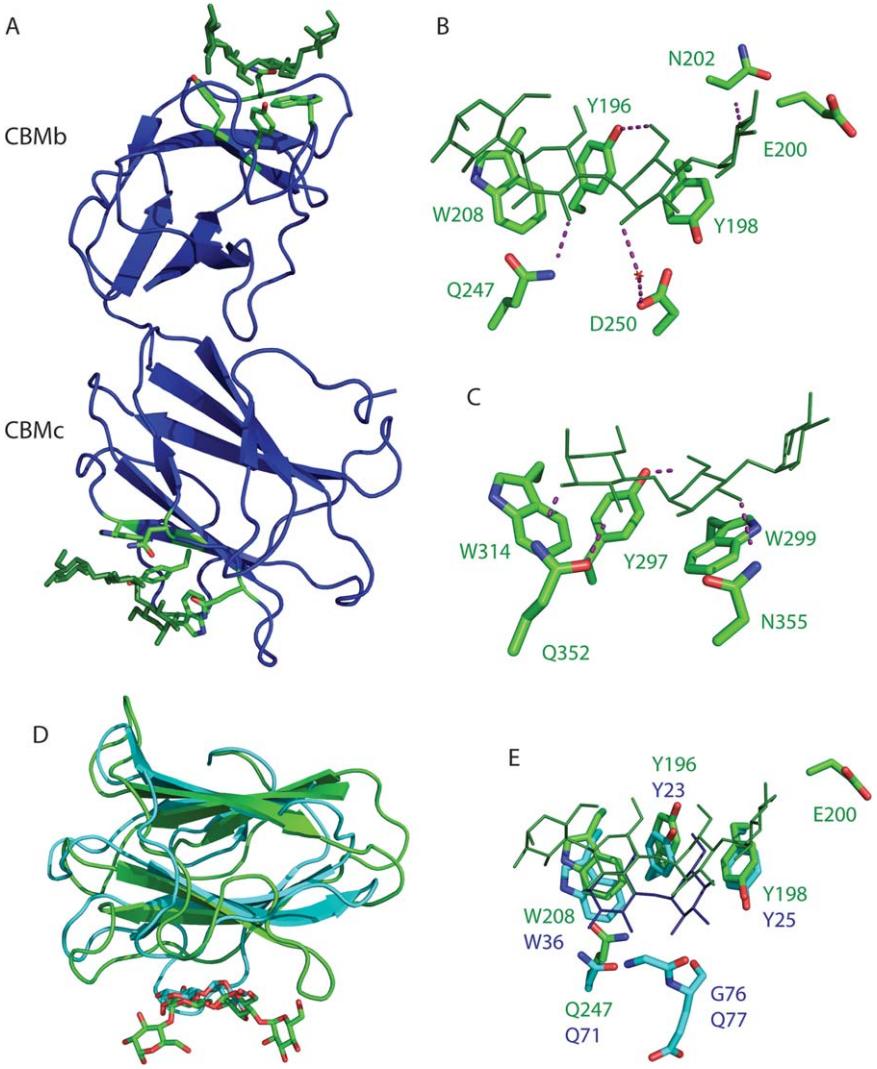

Fig. 3. Crystal structures of Amy13K_CBMbc.
A. Cartoon diagram of Amy13K_CBMbc in complex with maltoheptaose (PDB 6B3P) with the carbohydrate chain shown in green.
B. A close-up view of the CBMb oligosaccharide binding site. The sugar chain is shown in dark green and hydrogen bonds are indicated by dashed lines.
C. A close-up view of the CBMc oligosaccharide binding site. The sugar chain is shown in dark green and hydrogen bonds are indicated by dashed lines.
D. An overlay of Amy13K_CBMb and the CBM26 from *Bacillus halodurans* α-amylase G6 (PDB 2C3H). CBMb residues and bound carbohydrate are shown in green, while those for the *B. halodurans* CBM26 are shown in aqua.
E. An overlay of the carbohydrate binding sites of Amy13K_CBMb and the CBM26 from *B. halodurans* α-amylase G6 (PDB 2C3H). CBMb residues, carbohydrate and labels are shown in green, while those for CBM26 are shown in blue.

maltoheptaose with some minor differences. Both CBMs possess a pair of aromatic residues at the center of the binding interface, however in CBMb these are a tyrosine (Y198) and a tryptophan (W208), while in CBMc they are a pair of tryptophans (W299, W314) (Fig. 3B and C). Comparing the CBMs, the phenol ring of Y198 overlays with the indole ring of W299, while there is a 2.4 Å separation between the relative positions of W208 and W314. In addition to the aromatic binding platform there are several conserved hydrogen bonds between the two CBMs. The Y196 and Y297 hydroxyl groups form hydrogen bonds with the O6 of the glucose stacking on Y198 and W299, respectively, while Q247 and Q352, form hydrogen bonds with both the O2 and O3 of the glucose stacked on W208 and W314 respectively. In CBMc N355 forms a hydrogen bond with the O3 of the glucose stacked on W299, however, the equivalent residue in CBMb, D250, has a water-mediated contact with the O2 of the glucose stacked on Y198. A loop spanning from K199 to P204 in chain A of CBMb diverges in position from the equivalent loop in CBMc (A301–A308) and forms contacts with the glucose residues of the maltoheptaose molecule as it spans into the neighboring asymmetric unit into the binding site of CBMc. This places E200 and N202 of CBMb chain A in hydrogen bonding position with the maltoheptaose molecule as it extends out of the CBMb binding site, potentially expanding the CBMb binding site (Fig. 3B and Supporting Information Fig. S5). The equivalent loop in CBMc from A301 to A308 is composed of smaller sidechains, packing into the body of the CBM and does not appear to be capable of making additional contacts to a longer sugar.

Like CBMd, both CBMb and CBMc exhibit the typical β-sandwich fold seen in many CBMs (Fig. 3A). DALI searches reveal that the closest structural matches for CBMb/c are the CBM25 (2C3X, *Z*-score = 11.7/10.5)

and CBM26 (2C3H, *Z*-score = 10.7/10.1) from *Bacillus halodurans* α-amylase (Boraston *et al.*, 2006) along with the CBM25 from the *Paenibacillus polymyxa* β/α-amylase (PDB 2LAA, *Z*-score = 10.7/10.1). Somewhat weaker matches are found to the CBM41 family of domains, despite our previous suggestion that CBMbc might be members of this family (Cockburn *et al.*, 2015b). An overlay with the CBM26 from *B. halodurans* (2C3H) shows a high degree of overlap between the aromatic platform residues from both CBMb and c, as well as some of the hydrogen bonding residues of these CBMs. Of note, Y196/Y297 from CBMb/c with Y23 from *B. halodurans* CBM26 as well as Q247/352 from CBMb/c with Q71 from *B. halodurans* CBM26 (Fig. 3D and E) are superimposable. Thus, it appears that CBMb and CBMc are most structurally homologous to CBM26, supporting their placement within this family.

Amy13K CBMs bind soluble starch and oligosaccharides

A total of six recombinant protein constructs were used in this study to test the ability of the Amy13K CBMs to bind to starch (Fig. 1). The CBMs were expressed independently (e.g., CBMa) or in combination (e.g., CBMa-e) to examine how they might synergize to enhance starch-binding. Both affinity electrophoresis and isothermal titration calorimetry (ITC) were used to determine the binding of the constructs to polymers and smaller maltooligosaccharides. In affinity electrophoresis binding is monitored via migration of the protein through an acrylamide gel in the presence or absence of ligand and slower migration occurs as the protein interacts with polysaccharide (Abbott and Boraston, 2012; Cockburn *et al.*, 2017). By affinity electrophoresis, all of the CBM constructs except the CBMe construct displayed binding to both amylopectin and pullulan (Fig. 4). While amylopectin is one of the two components of starch along with amylose, pullulan is a linear fungal cell wall polysaccharide composed of α-1,6 linked maltotriose residues. Like amylopectin, pullulan contains both α-1,4 and α-1,6 linkages, however, it is linear rather than branched and has a much greater frequency of α-1,6 bonds, occurring every three linkages (Prajapati *et al.*, 2013). Thus, it serves as a model substrate for de-branching enzymes and can be useful for determining the tolerance for and importance of α-1,6 linkages for binding (Cockburn *et al.*, 2015a). While CBMa and CBMbc display slower migration in the presence of amylopectin, which has longer regions of α1,4-linked glucan, the CBMde construct is slowed much more by the presence of pullulan. Enhanced binding to pullulan could be driven by specific interactions with CBMd or due to enhanced avidity from the tandem CBM construct. Indeed, longer constructs such as CBMb-e and CBMa-e also displayed some enhanced binding to pullulan over amylopectin, either due to the presence of CBMde or by avidity. From the crystal structure of CBMd without substrate, it is difficult to speculate how this CBM may specifically accommodate α1,6 linkages, however, it may be the influence of the α1,6 bond on the surrounding structure that is recognized. The α1,6 bond introduces considerable structural flexibility relative to α1,4 bonds, resulting in pullulan behaving as a random coil in solution compared to the helices formed by amylose (Dais *et al.*, 2001). In starch, the branch points cause the creation of amorphous layers that alternate with the longer linear regions that make up the crystalline layers (Damager *et al.*, 2010). Thus, it takes 3–6 glucose residues after an α1,6 linkage before regular helices begin to form (Motawia *et al.*, 2005) and it is possible that CBMd and or CBMe preferentially recognize these less ordered regions. In total, these data suggest some differences in the relative affinity and tolerance of the various CBMs toward α1,6 branch points in starch and likely help the enzyme recognize a variety of starch particles in the gut environment. As expected, none of the CBMs demonstrated binding to dextran (data not shown), an all α-1,6 polymer of glucose on which Amy13K lacks activity.

ITC was also utilized to determine the affinity of the CBMs to glycogen, maltoheptaose and β-cyclodextrin. Glycogen is a starch-like storage molecule in many animals and bacteria, and it is structurally similar to amylopectin with an increased α1,6-branch frequency making it much more soluble and, thus, easily used in ITC. Maltoheptaose represents a stretch of α-1,4 linked glucose with flexible geometry and typically longer than the binding surface of CBMs, while β-cyclodextrin is identical in composition to maltoheptaose but circular, with a curvature matching that seen in amylose (Imberty *et al.*, 1991). All of the CBM constructs tested exhibited binding to each of these ligands (Table 2), with varying affinity. For the oligosaccharides most constructs displayed similar affinity for maltoheptaose and β-cyclodextrin with K_d in the 10^{-3} to 10^{-4} M range, somewhat weaker than has previously been observed in CBM26 at 10^{-5} M (Boraston *et al.*, 2006) and CBM41 at 10^{-6} M (Lammerts van Bueren *et al.*, 2004a), although this latter was derived from a thermophile which may explain the tighter binding at room temperature. Interestingly the CBMde construct had a K_d for maltoheptaose approximately an order of magnitude lower than that for β-cyclodextrin. This is consistent with increased affinity for binding near branch points where the helical structure of the starch is disrupted, and in line with our affinity electrophoresis data. The longer constructs, CBMb-e and CBMa-e each displayed higher affinity for the oligosaccharides than the smaller constructs. As we do not

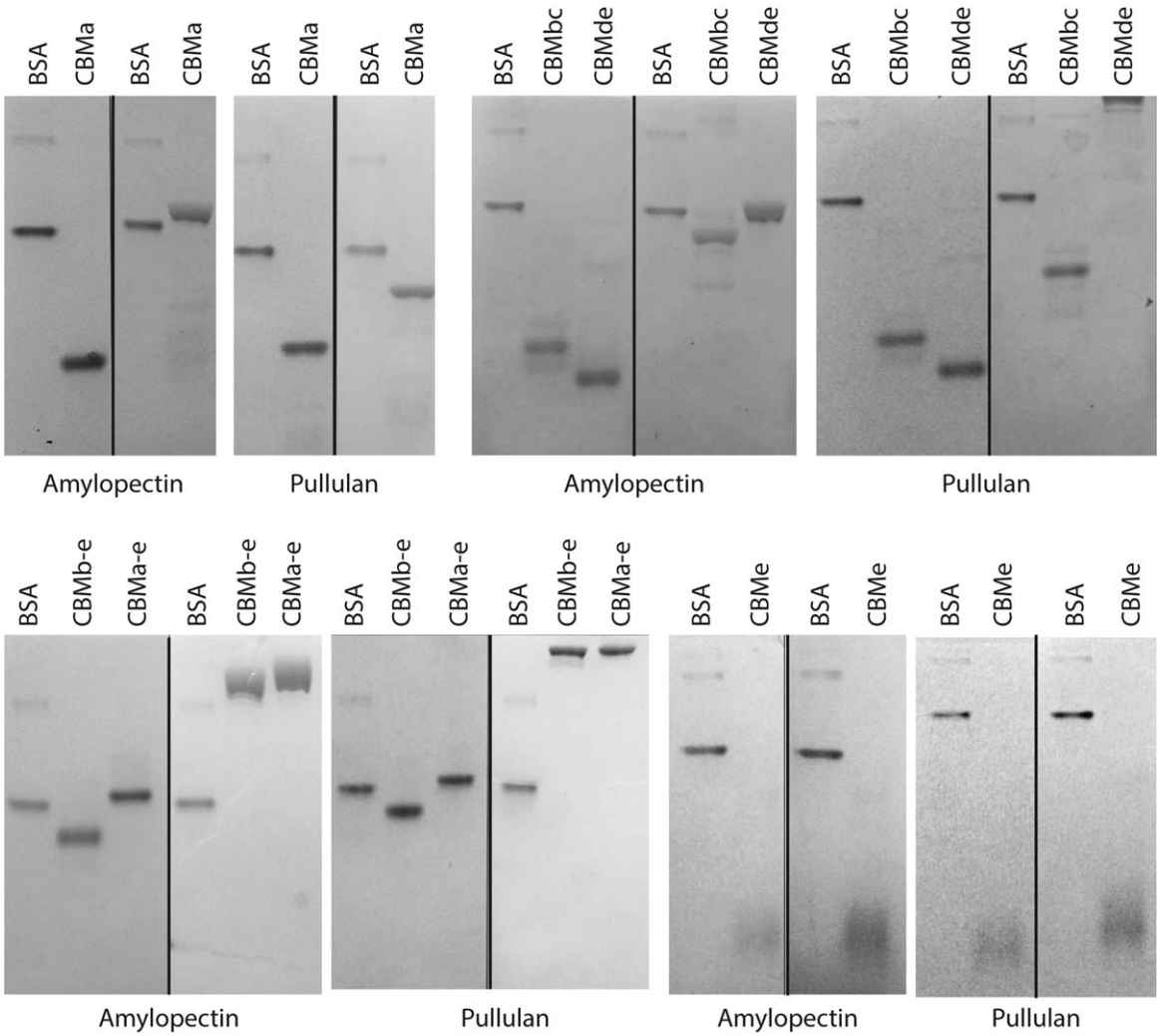

Fig. 4. Affinity electrophoresis of Amy13K CBMs. Proteins are separated with (right) or without (left) the indicated polysaccharide incorporated into the gel at 0.1% final concentration in native-PAGE. Bovine serum albumin (BSA) was loaded as a nonbinding control.

expect cooperativity in the binding of these substrates nor do we expect the substrates to span multiple CBMs, we speculate that improved structural stability in the larger constructs may be responsible for this enhanced affinity. With glycogen, the individual constructs CBMa and CBMe display binding with $K_d \sim 10^{-4}$ M, and this is enhanced an order of magnitude ($K_d \sim 10^{-5}$ M) when two CBMs such CBMbc or CBMde are expressed in tandem. Glycogen binding is further enhanced an additional two orders of magnitude ($K_d \sim 10^{-7}$ M) when four or five CBMs are expressed in tandem as in CBMa-e and CBMb-e, demonstrating that there is a significant avidity effect from having multiple CBMs. When considering the binding of CBMs to a polysaccharide it is not just the affinity, but also the frequency of binding sites that is important. Not surprisingly the single CBM constructs have a relatively high frequency of binding sites in glycogen (Table 2, values in parentheses) with CBMa displaying a binding site frequency of 2.3 mM/% glycogen, while CBMe is somewhat lower at 0.8 mM/% glycogen. Interestingly the dual CBM constructs diverge significantly in this regard with CBMbc displaying a binding site frequency of only 0.05 mM/% glycogen, while CBMde is at 2.4 mM/% glycogen, despite the similar affinities of these two constructs for glycogen. This higher frequency for binding sites for CBMde may suggest that these CBMs have a greater tolerance for the frequent α-1,6 branch points in glycogen in line with its greater affinity for pullulan in the AE gels (Fig. 4). The longer constructs CBMb-e and CBMa-e seem to be limited by the CBMbc binding restrictions as they display similar binding site frequencies at 0.06 mM/% glycogen

Table 1. Data collection and refinement statistics for Amy13K_CBMbc and Amy13K_CBMd.

	CBMbc native	CBMbc M7	CBMd
PDB code	6B15	6B3P	6AZ5
Wavelength	0.979	0.979	0.979
Resolution range	36.02 - 2.1 (2.175 - 2.1)	28.15 - 2.01 (2.082 - 2.01)	33.28 - 2.2 (2.279 - 2.2)
Space group	P 31 1 2	P 65 2 2	P 65 2 2
Unit cell	131.73 131.73 151.87 90 90 120	134.2 134.2 231.13 90 90 120	51 51 151.83 90 90 120
Total reflections	557,901 (55,398)	1,264,933 (124,942)	109,448 (10,683)
Unique reflections	86,483 (8664)	81,845 (7989)	6492 (608)
Multiplicity	6.5 (6.4)	15.5 (15.6)	16.9 (17.6)
Completeness (%)	98.46 (99.38)	99.65 (99.37)	99.94 (100.00)
Mean I/sigma(I)	7.86 (2.03)	10.90 (1.30)	15.08 (9.98)
Wilson B-factor	35.61	31.84	27.70
R-merge	0.09175 (0.5615)	0.1781 (3.103)	0.07451 (0.1759)
R-meas	0.09971 (0.6107)	0.1841 (3.206)	0.07692 (0.1812)
R-pim	0.03852 (0.237)	0.04639 (0.8003)	0.01875 (0.04297)
CC1/2	0.995 (0.832)	0.997 (0.373)	0.999 (0.996)
CC*	0.999 (0.953)	0.999 (0.737)	1 (0.999)
Reflections used in refinement	86,454 (8666)	81,781 (7983)	6492 (608)
Reflections used for R-free	1992 (199)	1998 (194)	650 (61)
R-work	0.2080 (0.2749)	0.1885 (0.2915)	0.2336 (0.3189)
R-free	0.2318 (0.3235)	0.2001 (0.2926)	0.2566 (0.3106)
CC(work)	0.949 (0.780)	0.960 (0.611)	0.892 (0.811)
CC(free)	0.932 (0.666)	0.949 (0.687)	0.939 (0.774)
Number of atoms	7136	4079	968
Macromolecules	6393	3248	920
Ligands	12	279	0
Solvent	731	552	48
Protein residues	836	418	116
RMS(bonds)	0.003	0.005	0.004
RMS(angles)	0.77	0.97	0.85
Ramachandran favored (%)	95.53	96.86	96.49
Ramachandran allowed (%)	4.11	3.14	3.51
Ramachandran outliers (%)	0.36	0.00	0.00
Rotamer outliers (%)	0.75	0.58	0.00
Clashscore	0.33	1.03	0.57
Average B-factor	44.18	38.81	26.18
Macromolecules	44.06	37.40	26.06
Ligands	52.53	45.06	
Solvent	45.08	43.93	28.46

Parentheses indicate statistics for the highest resolution shell.

and 0.1 mM/% glycogen, respectively, although with significantly better affinities. Thus, the combination of these CBMs provides high affinity binding, but perhaps at the cost of less frequent binding sites. Furthermore this appears to be driven not just by avidity effects and size of the construct, but also by differing binding specificities.

Table 2. Binding of CBMs to oligosaccharides and glycogen measured by ITC.

	K_d (M) ± SD		
Construct	β-cyclodextrin	Maltoheptaose	Glycogen (mM/%))[a]
CBMa ($n = 1$)[b]	$4.6 \times 10^{-4} \pm 7.4 \times 10^{-5}$	$1.0 \times 10^{-3} \pm 8.2 \times 10^{-5}$	$1.0 \times 10^{-4} \pm 8.2 \times 10^{-5}$ (2.3)
CBMbc ($n = 2$)[b]	$3.7 \times 10^{-4} \pm 6.1 \times 10^{-5}$	$5.0 \times 10^{-4} \pm 7.0 \times 10^{-5}$	$3.0 \times 10^{-5} \pm 5.2 \times 10^{-6}$ (0.05)
CBMde ($n = 2$)[b]	$1.0 \times 10^{-4} \pm 9.3 \times 10^{-5}$	$8.2 \times 10^{-6} \pm 2.3 \times 10^{-6}$	$2.4 \times 10^{-5} \pm 3.5 \times 10^{-6}$ (2.4)
CBMe ($n = 1$)[b]	$8.5 \times 10^{-4} \pm 3.2 \times 10^{-4}$	$3.1 \times 10^{-3} \pm 9.7 \times 10^{-4}$	$6.0 \times 10^{-4} \pm 3.8 \times 10^{-4}$ (0.8)
CBMb-e ($n = 4$)[b]	$7.8 \times 10^{-5} \pm 9.8 \times 10^{-6}$	8.4×10^{-5}	$6.8 \times 10^{-7} \pm 3.4 \times 10^{-7}$ (0.06)
CBMa-e ($n = 5$)[b]	$9.6 \times 10^{-5} \pm 2.9 \times 10^{-5}$	3.1×10^{-5}	$3.4 \times 10^{-7} \pm 5.4 \times 10^{-8}$ (0.1)

a. The mM/% values in parentheses represent the concentration of binding sites in 1% glycogen for this construct, see Experimental Procedures.
b. To obtain binding affinity it was necessary to fix the value of *n*. Values were chosen to reflect the number of CBMs present in the construct.

Table 3. Binding of CBMs to insoluble, intact starch granules measured by depletion assay.

	K_d (mg/ml) ± SD			
Construct	HiMaize 260	Corn starch	Potato starch	Fibersym
CBMa	44.6 ± 7.2	23.3 ± 9.3	NB[a]	NB[a]
CBMbc	32.5 ± 6.8	102.3 ± 35.8	NB[a]	NB[a]
CBMde	35.2 ± 26.2	31.0 ± 7.9	NB[a]	NB[a]
CBMb-e	4.6 ± 1.4	15.9 ± 2.9	NB[a]	81.9 ± 17.5
CBMa-e	14.2 ± 8.0	20.7 ± 4.4	> 100[b]	40.4 ± 18.6

a. No significant binding detected for the range of concentrations used in this assay.
b. Binding detected, but did not exhibit saturation within the concentration range tested.

Binding of CBMs to granular starch

In complement to binding studies with soluble substrates, the ability to interact with granular starch, which represents some of the starch that would be expected to traverse the distal gut, was investigated. In these adsorption assays, proteins and starch were incubated followed by centrifugation to determine the remaining concentration of unbound protein. Binding of the CBM constructs to standard cornstarch, whole grain corn starch, a high amylose corn starch (HiMaize 260), potato starch, and the chemically modified resistant starch Fibersym was tested (Table 3). No binding was detected for whole grain corn starch (data not shown) likely because the starch itself is inaccessible due to the presence of the bran, making it a type-1 resistant starch (Birt *et al.*, 2013). It has not been tested if whole grain starch serves as a growth substrate for *E. rectale*, but this result suggests it is unlikely. Significant binding toward Fibersym was only evident for the CBMa-e, and CBMb-e constructs, suggesting that no single CBM domain drives affinity, rather avidity from multiple domains is required. With regular corn starch, similar binding was observed among the CBMa and CBMde constructs, while CBMbc displays ∼ threefold lower affinity. The pairing of these CBMs together in the longer constructs CBMa-e and CBMb-e did not significantly enhance binding, suggesting avidity is not as important for access to corn starch. However, for the high-amylose starch HiMaize260, the longer CBMa-e and CBMb-e constructs displayed threefold–tenfold enhanced affinity over the smaller constructs. Surprisingly, the CBMa-e construct on average had lower affinity than the construct lacking only CBMa, despite the fact that CBMa binds HiMaize with similar affinity to the CBMbc and CBMde constructs. Thus, Amy13K CBMs seem to recognize high-amylose and mixed amylopectin/amylose corn starches with similar affinities. This was consistent across most of the CBM constructs tested, suggesting that these two forms of corn starch present similar binding surfaces to Amy13K.

Interestingly, the CBMs of Amy13K do not appreciably bind potato starch, with only slight binding observed for the largest CBM construct, and binding was not saturable. Relatively few studies have examined the difference in binding between corn starch and potato starch for amylolytic CBMs, although both the pig pancreatic amylase (Warren *et al.*, 2011) and the barley α-amylase AMY1 (Cockburn *et al.*, 2015a) have surface binding sites that seem to preferentially bind corn starch over potato starch. One interesting study found that for the *A. niger* CBM20 the affinity for potato starch and corn starch binding sites seemed to be the same, but there were far fewer of the binding sites available on the potato starch, resulting in a much lower apparent affinity (Paldi *et al.*, 2003). One point to note is that the approximately sevenfold difference in CBM20 binding sites in this study is significantly larger than the ∼threefold difference in specific surface area between these two starch types (Warren *et al.*, 2011). This indicates that it is not just the size of the surface available for binding that differs, but also the frequency of structural binding motifs. Binding to the surface of starch granules is an important barrier to enzyme action as the granule interiors seem to be readily attacked (Gallant *et al.*, 1992), even by enzymes lacking CBMs or surface binding sites (Cockburn *et al.*, 2015a). Removal of the CBM(s) of the *Microbacterium aurum* α-amylase MaAmyA removed the ability of the enzyme to form pores in starch granules (Valk *et al.*, 2015). Interestingly, it was discovered that this enzyme possesses a novel type of CBM, which was assigned to the new CBM74 family (Valk *et al.*, 2016). This CBM is of particular interest as it displays a tenfold better affinity for potato starch compared to corn starch and is enriched in gut bacteria, particularly resistant starch degraders such as a variety of *Bifidobacterium* species and *R. bromii*. Thus, acquiring novel binding functionalities may be a key adaptation of potato starch-degrading organisms.

Activity of Amy13K and CBM truncation mutants

To probe the role of these CBMs on enzyme activity, the activity of the full-length (WT) enzyme was compared to that of truncation mutants lacking one, three or all five of the identified CBMs (Fig. 1). In preliminary tests, constructs lacking three or all five CBMs displayed greatly reduced activity such that substrate saturation could not be attained, even when pushed to the maximum feasible levels. In addition, none of the enzyme constructs exhibited saturation kinetics for the potato starch. Therefore, we compared the catalytic efficiency (k_{cat}/K_M) of these

Table 4. Activity of Amy13K toward soluble, insoluble and resistant starch.

	k_{cat}/K_M (s^{-1} mg^{-1} ml) ± SD			
Construct	Amylopectin	Corn starch	HiMaize 260	Potato starch
WT	$3.1 \times 10^2 \pm 1.8 \times 10^1$	$3.8 \times 10^{-1} \pm 3.7 \times 10^{-2}$	$7.0 \times 10^{-1} \pm 2.8 \times 10^{-1}$	$7.7 \times 10^{-2} \pm 3.3 \times 10^{-3}$
ΔCBMa	$2.3 \times 10^2 \pm 9.5 \times 10^1$	$8.2 \times 10^{-2} \pm 1.0 \times 10^{-2}$	$1.6 \times 10^{-1} \pm 2.8 \times 10^{-2}$	$3.6 \times 10^{-2} \pm 1.0 \times 10^{-3}$
ΔCBMa-c	$1.2 \times 10^2 \pm 1.0 \times 10^1$	$1.4 \times 10^{-2} \pm 6.7 \times 10^{-4}$	$3.6 \times 10^{-2} \pm 1.5 \times 10^{-3}$	$1.9 \times 10^{-2} \pm 2.4 \times 10^{-3}$
ΔCBMa-e	$1.2 \times 10^2 \pm 4.0 \times 10^1$	$9.2 \times 10^{-3} \pm 1.5 \times 10^{-3}$	$1.8 \times 10^{-2} \pm 1.3 \times 10^{-3}$	$1.5 \times 10^{-2} \pm 4.2 \times 10^{-4}$

constructs to focus on the role of these CBMs on starch hydrolysis.

To examine if removal of the CBMs affected activity of the enzyme overall, we measured the catalytic efficiency of the WT and ΔCBMa-e on maltoheptaose, as hydrolysis of this small substrate would not be subject to an avidity affect via the CBMs. The catalytic efficiency of the WT and ΔCBMa-e on maltoheptaose are $2.8 \times 10^4 \pm 1.9 \times 10^4$ $s^{-1}M^{-1}$ and $1.2 \times 10^4 \pm 8.6 \times 10^3$ $s^{-1}M^{-1}$, respectively, suggesting that these truncated constructs have a slightly lower inherent activity as it is unclear how the CBMs could contribute to activity on a small substrate. However, with soluble starch amylopectin, the same twofold-threefold decrease in catalytic efficiency between the full-length enzyme and the constructs lacking various CBMs was observed, suggesting the CBMs are not required to enhance access to this substrate (Table 4). This is despite the efficient binding of these CBMs to the similar substrate glycogen (Table 2), suggesting a high affinity of the enzyme active site for amylopectin that is not further enhanced by the CBMs. Intriguingly, activity levels were similar for each construct toward both regular and high-amylose corn starch granules. This is in line with the binding assays for these substrates, and may suggest that the initial surface presented to the enzyme by these substrates is similar. For both regular and HiMaize corn starch there is a significantly higher dependence on the CBMs with the activity decreasing by an order of magnitude as the CBMs are removed. In contrast the potato starch, which has a low dependence on the CBMs, displayed an approximately threefold decrease in activity for the ΔCBMa-e enzyme compared to the WT enzyme. Most strikingly the enzymes lacking 3 or all 5 of the CBMs show similar activity toward the three insoluble starches, while the full length enzyme shows a much more dramatic decrease in catalytic efficiency between the corn starches and the potato starch.

Potato starch adopts the B-type crystalline form, as opposed to the A-type form seen in most corn and wheat starches (Imberty *et al.*, 1991). High amylose corn starch such as the HiMaize 260 used in this study also adopts the B-type crystalline form and this may explain its resistance to degradation (Gallant *et al.*, 1992). Interestingly the Amy13K CBMs bind similarly to both regular corn starch and HiMaize260 and the enzyme exhibits similar activity toward the two substrates. This indicates that surface binding does not represent the barrier to efficient degradation (and hence growth) in this case as it does with potato starch. It should be noted that the activities measured in this study only represent the initial stages of degradation that occurs on the granule surface. It is possible that the total amount of starch susceptible to degradation by Amy13K is much smaller for HiMaize260 compared to regular corn starch, once the granule surface has been degraded.

One final construct was tested where all the identified CBMs as well as an additional ~200 amino acids that occur between the CBMs and the predicted start of the GH13 catalytic domain, labeled as the unknown region in Fig. 1, were removed. While we were able to obtain large amounts of this recombinant protein in a soluble form during expression in *E. coli*, it had no detectable activity, even toward soluble substrates and oligosaccharides. Secondary structure predictions (JPred4, http://www.compbio.dundee.ac.uk/jpred/) (Drozdetskiy *et al.*, 2015) do not indicate β-strand rich regions as are typically found in CBMs and there are no domains with ascribed function that match this sequence. However, this region clearly plays an important structural role in Amy13K, perhaps directly impacting the active site.

Discussion

Importance of CBMs for activity

It has been demonstrated that CBMs are important for targeting substrates in complex environments such as the plant cell wall (Hervé *et al.*, 2010) and undoubtedly the breakdown of starch in the gut provides similar challenges. During *in vitro* studies the removal of a starch binding CBM20 abolished activity of the *Aspergillus niger* glucoamylase toward granular starch (Svensson *et al.*, 1982). Conversely, recombinantly fusing this CBM20 to the barley α-amylase AMY1 increased its

activity toward granular starch sixfold (Juge *et al*., 2006). This is despite the fact that AMY1, while lacking a CBM, has a pair of surface binding sites on its catalytic module, which have been shown to be important for its activity (Nielsen *et al*., 2009; Nielsen *et al*., 2012; Cockburn *et al*., 2015a). In Amy13K, we have identified five CBMs that provide the enzyme with affinity for starch granules. Removal of these CBMs has little impact on the activity of the enzyme toward maltoheptaose or soluble amylopectin, but has a more dramatic effect on the activity toward cornstarch granules. Notably this difference disappears when examining activity toward potato starch, for which the CBMs have apparently less affinity. The protein lacking all five CBMs displays little discrimination between the three types of insoluble starches tested, suggesting that the CBMs account for the differences in activity against these substrates. The lower activity and lack of dependence on the CBMs for potato starch corresponds to the lack of binding seen for the isolated CBMs.

Implications for relationships with other gut microorganisms

Resistant starch represents an important substrate for the gut microbiota, while nonresistant starch is processed in the small intestine and, thus, does not reach the microbial populations of the large intestine. Potato and high amylose corn starch consist of about 40–80% resistant starch (depending on the specific type and method of measurement used) (McCleary and Monaghan, 2002) and, thus, a large proportion of these starches reach the colon. *E. rectale* alone is unable to grow on resistant starches, but grows well in coculture with resistant starch degraders such as *Ruminococcus bromii* (Ze *et al*., 2012). Our results suggest that for potato starch it is the lack of efficient targeting by the CBMs of Amy13K that underpins the molecular basis for *E. rectale*'s inability to grow on this substrate. This inefficient targeting seems to be entirely due to the granular structure of potato starch as purified and autoclaved potato amylopectin readily supports growth of *E. rectale* (Desai *et al*., 2016) and is efficiently bound by the Amy13K CBMs (Fig. 4).

Despite its limited ability to grow on resistant starches, people who consume resistant starch often have increased levels of *E. rectale* in their large intestine (Martínez *et al*., 2010; Martínez *et al*., 2013; Venkataraman *et al*., 2016). While *E. rectale* has a suite of transporters specializing in the uptake of starch breakdown products (Cockburn *et al*., 2015b), it would clearly be advantageous for the organisms to localize to this food source. *E. rectale* lacks accessory starch binding proteins such as those found in the starch utilization system (Sus) of *Bacteroides thetaiotaomicron* (Cameron *et al*., 2012; Foley *et al*., 2016). Instead it is possible that the CBMs of the cell wall anchored Amy13K help localize the bacteria to resistant starches such as high amylose corn starch. Indeed *E. rectale* was found to colonize high amylose corn starch in an *in vitro* continuous flow system (Leitch *et al*., 2007). *E. rectale* levels are enriched on diet supplementation with potato starch (Venkataraman *et al*., 2016), but given the lack of binding to this starch by the Amy13K CBMs other means of localization may be needed or it is possible that enough soluble material is released by degraders to render binding to potato starch granules unnecessary. However, the affinity of these CBMs for potato amylopectin may indicate that on initial processing of potato starch, new binding sites are opened up for binding by Amy13K. The Amy13K CBMs do display weak binding to Fibersym, a Type IV, chemically modified resistant starch, however, a study with people consuming this starch did not find elevated levels of *E. rectale* (Martínez *et al*., 2010). It is also currently unknown if this starch can directly support *E. rectale* growth or indirectly through crossfeeding with a primary degrader.

Conclusion. We have identified and characterized the five CBMs of Amy13K, which allow the definition of two new CBM families. These CBMs bind efficiently to corn starch, including high amylose corn starch as well as amylopectin and maltooligosaccharides, but display little affinity for potato starch. The low affinity of these CBMs for granular potato starch may be a key factor in the low activity of Amy13K for this substrate and provides a molecular rationale for why this is a poor growth substrate for *E. rectale*. In contrast it seems that other factors are at play in limiting the ability of *E. rectale* to utilize high amylose corn starch as the surface binding and initial rates of degradation are similar to regular corn starch. It could be that following the initial surface erosion the binding motifs recognized by the Amy13K CBMs are eliminated decreasing affinity and activity to that seen with potato starch. Indeed, in previous work examining the ability of *E. rectale* to utilize corn starches, the bacterium can utilize less than 20% of high amylose corn starch when cultured with the raw granules that have not been heat treated (Ze *et al*., 2012). Our results presented here provide important insight into the potential roles of CBMs in determining substrate utilization profiles in the human gastrointestinal tract.

Experimental procedures

Reagents

Primers used for cloning were synthesized by IDT DNA Technologies and are listed in Supporting Information Table

S1. HiMaize260 starch and whole grain starch were kindly provided by Ingredion (Bridgewater, NJ, USA). FiberSym starch (MGP Ingredients) was a gift from Jens Walter (University of Alberta, Canada). All other chemicals were purchased from Sigma Aldrich, except where noted.

Cloning, protein expression and purification

All genes and gene fragments were amplified from *E. rectale* genomic DNA using the Phusion™ Flash polymerase (Thermo Fisher Scientific) according to the manufacturer's instructions and all primer sequences are listed in Supporting Information Table S1. All genetic constructs used in this study were created using the Expresso® T7 Cloning system (Lucigen Inc.) according to the manufacturer's instructions and are listed in Supporting Information Table S2. Expression plasmids were transformed into *E. coli* Rosetta (DE3) pLysS cells, expressed and purified as previously described (Cockburn *et al.*, 2015b). Selenomethionine substituted Amy13K_CBMbc was produced by first transforming the plasmid into *E. coli* Rosetta(DE3)/pLysS and plating onto LB supplemented with kanamycin (50 μg/ml) and chloramphenicol (20 μg/ml). The bacteria were grown for 16 h at 37°C and then colonies were harvested from the plate to inoculate 100 ml of M9 minimal medium supplemented with the same antibiotics. After 16 h of incubation at 37°C this starter culture was used to inoculate a 2 l baffled flask containing 1 l of Molecular Dimensions SelenoMet premade medium supplemented with 50 ml of the recommended sterile nutrient mix, chloramphenicol and kanamycin. Cultures were incubated at 37°C until an OD_{600} of 0.45 was reached. At this point the temperature was adjusted to 20°C and each flask was supplemented with 100 mg each of L-lysine, L-threonine and L-phenylalanine and 50 mg each of L-leucine, L-isoleucine, L-valine and L-selenomethionine (Van Duyne *et al.*, 1993). After 20 min of further incubation, protein expression was induced by the addition of 0.5 mM IPTG and cultures were allowed to grow for an additional 48 h before being harvested. Cells were then lysed and the protein purified as previously described via Ni^{2+} affinity chromatography (Cockburn *et al.*, 2015b).

Crystallization experiments

All proteins were subjected to a series of 96-well hanging drop sparse matrix screens to identify crystallization conditions. Selenomethionine-substituted crystals of Amy13k_CBMbc (54 mg/ml) were obtained via hanging drop vapor diffusion at room temperature against 1.5 M ammonium sulfate, 0.1 M Bis-Tris Propane, pH 7.0 (Hampton Research SaltRx). Native Amy13K_CBMbc crystals were obtained without (free) or with 14mM maltoheptaose via hanging drop experiments against 60% Tacsimate, 0.1 M Bis-Tris Propane, pH 7.0 (Hampton Research SaltRx), also at room temperature. Native Amy13K_CBMd crystals were obtained via hanging drop against 50% pentaerythritol propoxylate (5/4 PO/OH), 0.1 M Tris-HCl pH 8.0 in the Molecular Dimensions Midas screen using 20 mg/ml protein. All crystals used in this study were cryoprotected prior to freezing in liquid nitrogen by quickly swiping the crystal through a solution of 80% mother liquor supplemented with 20% ethylene glycol. X-ray data were collected at the Life Sciences Collaborative Access Team (LSCAT) beamline ID-D of the Advanced Photon Source at Argonne National Laboratory. Data were integrated using iMosFLM (Battye *et al.*, 2011) and then indexed and scaled using the program Aimless (Evans and Murshudov, 2013) from the CCP4 package (Winn *et al.*, 2011). For the selenomethionine substituted Amy13K_CBMbc, phases were solved by single anomalous dispersion (SAD) using the AutoSol program of the Phenix package (Adams *et al.*, 2010). This structure was then used to solve the native Amy13K_CBMbc with or without maltoheptaose by molecular replacement using Phaser-MR (McCoy *et al.*, 2007) within Phenix. In the substrate free structure, four molecules were found in the asymmetric unit, however, chain D exhibited higher mobility than the other chains and not all amino acid sidechains from Y297-I356 could be confidently fit to the electron density and were, thus, omitted. The Amy13K_CBMd structure was solved via sulfur SAD after merging seven datasets from three crystals with autoPROC (Vonrhein *et al.*, 2011) and phasing in AutoSol. The resulting structure was then used to solve the structure from a single dataset via molecular replacement with Phaser-MR. Structures were refined using Phenix.refine (Afonine *et al.*, 2012). In the maltoheptaose bound structure of Amy13K_CBMbc the conformation of bound carbohydrates was validated using Privateer (Agirre *et al.*, 2015) from the CCP4 package.

Enzyme activity assays

For activity assays with polysaccharide substrates the production of free reducing ends was monitored using the bicinchoninic acid (BCA) method (Waffenschmidt and Jaenicke, 1987) as previously described (Cockburn *et al.*, 2015b). All reactions included 10 mM HEPES pH 6.5, with 5 mM $CaCl_2$ and 0.02% Tween80. All granular starch substrates were washed 10x in pure water prior to activity assays. For activity toward amylopectin Amy13K_WT (2.2 nM), ΔCBMa (2.4 nM), ΔCBMa-c (3.0 nM) or ΔCBMa-e (4.0 nM) was incubated with six concentrations of potato amylopectin (0.003–0.5%). Reactions were monitored for 30 min. Initial velocities were calculated and fitted to a Michalis–Menten curve to calculate k_{cat} and K_M. For activity toward corn starch Amy13K_WT (3 nM), ΔCBMa (6 nM), ΔCBMa-c (15 nM) or ΔCBMa-e (26 nM) were incubated with six concentrations of granular corn starch (0.2–10%) and activity was monitored for 135 min. Initial velocities were plotted against substrate concentration and for WT and ΔCBMa the k_{cat} and K_M were derived through fitting to a Michalis–Menten curve. For ΔCBMa-c and ΔCBMa-e it was only possible to derive k_{cat}/K_M from the slope of the line. For activity toward HiMaize 260 starch, Amy13K_WT (4 nM), ΔCBMa (8 nM), ΔCBMa-c (20 nM) or ΔCBMa-e (40 nM) were incubated with six concentrations of granular HiMaize 260 high amylose corn starch (0.06–6%) and activity was monitored for 135 min. Initial velocities were plotted against substrate concentration and for WT and ΔCBMa the k_{cat} and K_M were derived through fitting to a Michalis–

Menten curve. For ΔCBMa-c and ΔCBMa-e it was only possible to derive k_{cat}/K_M from the slope of the line. For activity toward potato starch, Amy13K_WT (1.1 nM), ΔCBMa (1.2 nM), ΔCBMa-c (1.5 nM) or ΔCBMa-e (2.0 nM) were incubated with six concentrations of granular potato starch (Bob's RedMill; 0.5–20%). Initial velocities were plotted against substrate concentration and k_{cat}/K_M was calculated from the slope of the line. Activity toward oligosaccharides was monitored via isothermal titration calorimetry. Amy13K_WT (11 nM) or ΔCBMa-e (20 nM) were placed into the cell of a standard volume Nano ITC (TA Instruments, New Castle, Delaware). Forty mM maltoheptaose was serially injected into the cell while stirring at 350 RPM at a temperature of 37°C (see Supporting Information Table S3 for injection volumes and times). The molecular enthalpy of the reaction was calculated to be 4.41 kJ/mol by monitoring the complete conversion of 10 mM maltotetraose to maltose by 8 mg/ml Amy13B, formerly EUR_01860 (Cockburn *et al.*, 2015b) in duplicate, which agreed well with previous estimates (Goldberg *et al.*, 1991). Catalytic parameters were determined using the NanoAnalyze software (TA instruments).

Isothermal titration calorimetry CBM binding assays

CBM binding to maltoheptaose, β-cyclodextrin and glycogen was determined by isothermal titration calorimetry (ITC) using a TA Instruments low volume NanoITC. For CBMa, 63 μM protein was titrated with 10 and 13 mM β-cyclodextrin, 20 mM maltoheptaose or 1% and 2% glycogen (from rabbit liver). CBMbc was used at a concentration of 42 μM and titrated with 10 mM β-cyclodextrin, 10 and 25 mM maltoheptaose or 5% glycogen. The CBMde construct was at a concentration of 190 μM and titrated with 1 mM and 5 mM maltoheptaose, 2.5 mM β-cyclodextrin or 1%, 1.5% and 2% glycogen. CBMe was measured at 780 μM and titrated with 10 mM and 13 mM β-cyclodextrin, 20 mM maltoheptaose or 5% and 10% glycogen. CBMb-e was measured at 5 μM and titrated with 4 and 5 mM β-cyclodextrin, 20 mM maltoheptaose or 0.5% and 1% glycogen. CBMa-e was measured at 7.8 μM and titrated with 2 or 4 mM β-cyclodextrin, 5 mM maltoheptaose or 1% and 2% glycogen. All data were analyzed using the manufacturer's NanoAnalyze software, using a constant blank correction and an independent binding model unless otherwise noted. To obtain K_d values it was necessary to fix the value of n (number of binding sites) in these calculations. For maltoheptaose and β-cyclodextrin the value of n was fixed at the number of CBMs in the construct (e.g., 1 for CBMa, 2 for CBMbc). For glycogen the molar concentration of the ligand used was empirically set such that it produced a value for n of 1 when fitting the curve to the data. Thus, the concentration of glycogen used in this calculation represents the molar concentration of available binding sites on the polysaccharide ligand, according to the protocol of Abbott and Boraston (2012). The binding site frequency (mM/% glycogen) for each construct was calculated as the slope of the concentration of binding sites over the w/v% of glycogen used. For example, the CBMb-e construct was assayed for binding at 0.5% and 1% glycogen. The binding site concentration (the concentration found to give $n = 1$ during curve fitting) was 0.022 mM and 0.05 mM respectively. Calculating the slope (0.05–0.022)/(1–0.5) gives a binding site frequency of 0.056 mM/%glycogen (rounded to 0.06 in Table 2). This represents the concentration of binding sites for a particular construct, that is, how many copies can bind before reaching saturation for 1% w/v glycogen. Thus, we would expect lower numbers for larger constructs (each takes up more space) and for constructs that have a relatively infrequently occurring binding motif.

Starch binding assays

Binding of isolated CBMs to insoluble corn starch, whole grain starch, HiMaize 260 high amylose starch, potato starch or Fibersym chemically modified resistant starch was determined through protein depletion assays (Abbott and Boraston, 2012). Prior to protein addition, the starch was washed two times with 10 mM HEPES pH 6.5, 150 mM NaCl. CBMs (80 μg/ml) were then incubated with 1–100 mg/ml starch for 10 min at room temp with end-over-end rotation and insoluble material (including bound protein) was removed by two rounds of centrifugation at 20,000 × *g*. Protein concentration in the supernatant (unbound) was then determined by the Bradford assay (Bio-Rad) according to the manufacturer's protocol, using the CBM construct under study as the protein standard. The fraction of protein bound to starch was then plotted against starch concentration to determine binding constants using the following formula:

$$B = \frac{B_{max}[S]}{K_d + [S]}$$

Where B is the fraction of protein bound, B_{max} is the maximum proportion of protein bound, $[S]$ is the concentration of starch and K_d is the dissociation constant.

Affinity electrophoresis

To investigate binding of CBMs to amylose, amylopectin, glycogen, pullulan and dextran, affinity electrophoresis was used (Abbott and Boraston, 2012; Cockburn *et al.*, 2017). Native polyacrylamide gels with and without added polysaccharide were compared for each CBM construct. Binding was considered positive if the migration of the protein in the polysaccharide gel relative to a noninteracting protein (bovine serum albumin) was significantly slower (<0.85 relative mobility) compared to that in the control gel. All polysaccharides were used at 0.1% final concentration. Gels were made at 12% acrylamide with 0.375 M Tris-HCl pH 8.8. Gels were subjected to 100 V for 4 h and then stained for 2 h with 0.1% Coomassie Brilliant Blue R-250 in 10% acetic acid, 50% methanol, 40% water, before destaining with solution lacking Coomassie overnight with one change of solution.

Bioinformatic analysis

The boundaries of the CBMs were determined from examination of the 3-D structures. The sequence corresponding

to each CBM was compared to the sequences of the CBM families listed in the CAZy database using BLAST (Altschul *et al*., 1990).

Acknowledgements

This work was supported by funds from a pilot/feasibility grant from the University of Michigan Gastrointestinal Peptides Research Center (DK034933) awarded to N.M.K. and a scientist development grant from the American Heart Association (17SDG32770001) to D.W.C. This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02–06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor (grant 085P1000817).

Author contributions

The study was designed by D.W.C. and N.M.K. Data were collected by D.W.C., C.S., K.P.M. and B.M.D., and analyzed by D.W.C, N.M.K. and B.H. Sequence alignments and CBM classification was performed by B.H. X-ray data was collected and processed by D.W.C. and Z.W. The manuscript was written by D.W.C. and N.M.K.

References

Abbott, D.W., and Boraston, A.B. (2012) Quantitative approaches to the analysis of carbohydrate-binding module function. *Methods Enzymol* **510**: 211–231.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**: 213–221.

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., *et al.* (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr D Biol Crystallogr* **68**: 352–367.

Agirre, J., Iglesias-Fernandez, J., Rovira, C., Davies, G.J., Wilson, K.S., and Cowtan, K.D. (2015) Privateer: software for the conformational validation of carbohydrate structures. *Nat Struct Mol Biol* **22**: 833–834.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J Mol Biol* **215**: 403–410.

Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011) iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D* **67**: 271–281.

Beauvieux, M.C., Roumes, H., Robert, N., Gin, H., Rigalleau, V., and Gallis, J.L. (2008) Butyrate ingestion improves hepatic glycogen storage in the re-fed rat. *BMC Physiol* **8**: 19.

Berggren, A.M., Nyman, E.M., Lundquist, I., and Bjorck, I.M. (1996) Influence of orally and rectally administered propionate on cholesterol and glucose metabolism in obese rats. *Br J Nutr* **76**: 287–294.

Birt, D.F., Boylston, T., Hendrich, S., Jane, J.-L., Hollis, J., Li, L., *et al.* (2013) Resistant starch: promise for improving human health. *Adv Nutr* **4**: 587–601.

Boets, E., Gomand, S., Deroover, V., Preston, L., Vermeulen, T.K., Preter, V., *et al.* (2017) Systemic availability and metabolism of colonic-derived short-chain fatty acids in healthy subjects—a stable isotope study. *J Physiol* **595**: 541–555.

Boraston, A.B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A., and Law, V. (2006) A structural and functional analysis of α-glucan recognition by family 25 and 26 carbohydrate-binding modules reveals a conserved mode of starch recognition. *J Biol Chem* **281**: 587–598.

Bruzzese, E., Callegari, M.L., Raia, V., Viscovo, S., Scotto, R., Ferrari, S., *et al.* (2014) Disrupted intestinal microbiota and intestinal inflammation in children with cystic fibrosis and its restoration with *Lactobacillus* GG: a randomised clinical trial. *PLoS One* **9**: e87796.

Cameron, E.A., Maynard, M.A., Smith, C.J., Smith, T.J., Koropatkin, N.M., and Martens, E.C. (2012) Multidomain carbohydrate-binding proteins involved in Bacteroides thetaiotaomicron starch metabolism. *J Biol Chem* **287**: 34614–34625.

Cockburn, D., and Svensson, B. (2016) Structure and functional roles of surface binding sites in amylolytic enzymes. In *Understanding Enzymes: Function, Design, Engineering and Analysis*. Allan Svendsen (ed). Singapore: Pan Stanford Publishing Pte. Ltd, pp. 267–295.

Cockburn, D., Nielsen, M.M., Christiansen, C., Andersen, J.M., Rannes, J.B., Blennow, A., and Svensson, B. (2015a) Surface binding sites in amylase have distinct roles in recognition of starch structure motifs and degradation. *Int J Biol Macromol* **75**: 338–345.

Cockburn, D.W., Orlovsky, N.I., Foley, M.H., Kwiatkowski, K.J., Bahr, C.M., Maynard, M., *et al.* (2015b) Molecular details of a starch utilization pathway in the human gut symbiont *Eubacterium rectale*. *Mol Microbiol* **95**: 209–230.

Cockburn, D., Wilkens, C., and Svensson, B. (2017) Affinity electrophoresis for analysis of catalytic module-carbohydrate interactions. *Methods Mol Biol* **1588**: 119–127.

Dais, P., Vlachou, S., and Taravel, F.R. (2001) (13)C nuclear magnetic relaxation study of segmental dynamics of the heteropolysaccharide pullulan in dilute solutions. *Biomacromolecules* **2**: 1137–1147.

Damager, I., Engelsen, S.B., Blennow, A., Møller, B.L., and Motawia, M.S. (2010) First principles insight into the α-glucan structures of starch: their synthesis, conformation, and hydration. *Chem Rev* **110**: 2049–2080.

De Cruz, P., Kang, S., Wagner, J., Buckley, M., Sim, W.H., Prideaux, L., *et al.* (2015) Association between specific mucosa-associated microbiota in Crohn's disease at the time of resection and subsequent disease recurrence: a pilot study. *J Gastroenterol Hepatol* **30**: 268–278.

Desai, M.S., Seekatz, A.M., Koropatkin, N.M., Kamada, N., Hickey, C.A., Wolter, M., *et al.* (2016) A dietary fiber-

deprived gut microbiota degrades the colonic mucus barrier and enhances pathogen susceptibility. *Cell* **167**: 1339–1353.e1321.
Drozdetskiy, A., Cole, C., Procter, J., and Barton, G.J. (2015) JPred4: a protein secondary structure prediction server. *Nucleic Acids Res* **43**: W389–W394.
Evans, P.R., and Murshudov, G.N. (2013) How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**: 1204–1214.
Foley, M.H., Cockburn, D.W., and Koropatkin, N.M. (2016) The Sus operon: a model system for starch uptake by the human gut Bacteroidetes. *Cell Mol Life Sci* **73**: 2603–2617.
Forslund, K., Hildebrand, F., Nielsen, T., Falony, G., Le Chatelier, E., Sunagawa, S., *et al.* (2015) Disentangling type 2 diabetes and metformin treatment signatures in the human gut microbiota. *Nature* **528**: 262–266.
Fung, K.Y.C., Cosgrove, L., Lockett, T., Head, R., and Topping, D.L. (2012) A review of the potential mechanisms for the lowering of colorectal oncogenesis by butyrate. *Br J Nutr* **108**: 820–831.
Gallant, D.J., Bouchet, Buleon, B.A., and Perez, S. (1992) Physical characteristics of starch granules and susceptibility to enzymatic degradation. *Eur J Clin Nutr* **46**: S3–S16.
Goldberg, R.N., Bell, D., Tewari, Y.B., and McLaughlin, M.A. (1991) Thermodynamics of hydrolysis of oligosaccharides. *Biophys Chem* **40**: 69–76.
Gossling, J., and Slack, J.M. (1974) Predominant Gram-positive bacteria in human feces: numbers, variety, and persistence. *Infect Immun* **9**: 719–729.
Guilloteau, P., Martin, L., Eeckhaut, V., Ducatelle, R., Zabielski, R., and Van Immerseel, F. (2010) From the gut to the peripheral tissues: the multiple effects of butyrate. *Nutr Res Rev* **23**: 366–384.
Haro, C., Garcia-Carpintero, S., Alcala-Diaz, J.F., Gomez-Delgado, F., Delgado-Lista, J., Perez-Martinez, P., *et al.* (2016) The gut microbial community in metabolic syndrome patients is modified by diet. *J Nutr Biochem* **27**: 27–31.
Hervé, C., Rogowski, A., Blake, A.W., Marcus, S.E., Gilbert, H.J., and Knox, J.P. (2010) Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc Natl Acad Sci USA* **107**: 15293–15298.
Imberty, A., Buléon, A., Tran, V., and Péerez, S. (1991) Recent advances in knowledge of starch structure. *Starch-Stärke* **43**: 375–384.
Janecek, S., Majzlova, K., Svensson, B., and MacGregor, E.A. (2017) The starch-binding domain family CBM41-An in silico analysis of evolutionary relationships. *Proteins* **85**: 1480–1492.
Juge, N., Nøhr, J., Le Gal-Coëffet, M.-F., Kramhøft, B., Furniss, C.S.M., Planchot, V., *et al.* (2006) The activity of barley α-amylase on starch granules is enhanced by fusion of a starch binding domain from *Aspergillus niger* glucoamylase. *Biochim Biophys Acta* **1764**: 275–284.
Kang, S., Denman, S.E., Morrison, M., Yu, Z., Dore, J., Leclerc, M., and McSweeney, C.S. (2010) Dysbiosis of fecal microbiota in Crohn's disease patients as revealed by a custom phylogenetic microarray. *Inflamm Bowel Dis* **16**: 2034–2042.
Koropatkin, N.M., Martens, E.C., Gordon, J.I., and Smith, T.J. (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* **16**: 1105–1115.
Lammerts van Bueren, A., and Boraston, A.B. (2007) The structural basis of α-glucan recognition by a family 41 carbohydrate-binding module from *Thermotoga maritima*. *J Mol Biol* **365**: 555–560.
Lammerts van Bueren, A., Finn, R., Ausió, J., and Boraston, A.B. (2004a) Alpha-glucan recognition by a new family of carbohydrate-binding modules found primarily in bacterial pathogens. *Biochemistry* **43**: 15633–15642.
Lammerts Van Bueren, A., Finn, R., Ausió, J., and Boraston, A.B. (2004b) α-Glucan recognition by a new family of carbohydrate-binding modules found primarily in bacterial pathogens. *Biochemistry* **43**: 15633–15642.
Leitch, E.C.M., Walker, A.W., Duncan, S.H., Holtrop, G., and Flint, H.J. (2007) Selective colonization of insoluble substrates by human faecal bacteria. *Environ Microbiol* **9**: 667–679.
Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M., and Henrissat, B. (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* **42**: D490–D495.
Martínez, I., Kim, J., Duffy, P.R., Schlegel, V.L., Walter, J., and Heimesaat, M.M. (2010) Resistant starches types 2 and 4 have differential effects on the composition of the fecal microbiota in human subjects. *PLoS One* **5**: e15046.
Martínez, I., Lattimer, J.M., Hubach, K.L., Case, J.A., Yang, J., Weber, C.G., *et al.* (2013) Gut microbiome composition is linked to whole grain-induced immunological improvements. *ISME J* **7**: 269–280.
McCleary, B.V., and Monaghan, D.A. (2002) Measurement of resistant starch. *J AOAC Int* **85**: 665–675.
McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007) Phaser crystallographic software. *J Appl Crystallogr* **40**: 658–674.
McNeil, N.I. (1984) The contribution of the large intestine to energy supplies in man. *Am J Clin Nutr* **39**: 338–342.
Møller, M.S., Henriksen, A., and Svensson, B. (2016) Structure and function of α-glucan debranching enzymes. *Cell Mol Life Sci* **73**: 2619–2641.
Motawia, M.S., Damager, I., Olsen, C.E., Møller, B.L., Engelsen, S.B., Hansen, S., *et al.* (2005) Comparative study of small linear and branched alpha-glucans using size exclusion chromatography and static and dynamic light scattering. *Biomacromolecules* **6**: 143–151.
Nastasi, C., Candela, M., Bonefeld, C.M., Geisler, C., Hansen, M., Krejsgaard, T., *et al.* (2015) The effect of short-chain fatty acids on human monocyte-derived dendritic cells. *Sci Rep* **5**: 16148.
Nielsen, J.W., Kramhøft, B., Bozonnet, S., Abou Hachem, M., Stipp, S.L.S., Svensson, B., *et al.* (2012) Degradation of the starch components amylopectin and amylose by barley α-amylase 1: role of surface binding site 2. *Arch Biochem Biophys* **528**: 1–6.

Nielsen, M.M., Bozonnet, S., Seo, E.-S., Mótyán, J.A., Andersen, J.M., Dilokpimol, A., *et al.* (2009) Two secondary carbohydrate binding sites on the surface of barley α-amylase 1 have distinct functions and display synergy in hydrolysis of starch granules. *Biochemistry* **48**: 7686–7697.

Paldi, T., Levy, I., and Shoseyov, O. (2003) Glucoamylase starch-binding domain of *Aspergillus niger* B1: molecular cloning and functional characterization. *Biochem J* **372**: 905–910.

Prajapati, V.D., Jani, G.K., and Khanda, S.M. (2013) Pullulan: an exopolysaccharide and its various applications. *Carbohydr Polym* **95**: 540–549.

Qin, J., Li, Y., Cai, Z., Li, S., Zhu, J., Zhang, F., *et al.* (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**: 55–60.

Rajilic-Stojanovic, M., Shanahan, F., Guarner, F., and de Vos, W.M. (2013) Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. *Inflamm Bowel Dis* **19**: 481–488.

Ramsay, A.G., Scott, K.P., Martin, J.C., Rincon, M.T., and Flint, H.J. (2006) Cell-associated α-amylases of butyrate-producing Firmicute bacteria from the human colon. *Microbiology* **152**: 3281–3290.

Ridaura, V.K., Faith, J.J., Rey, F.E., Cheng, J., Duncan, A.E., Kau, A.L., *et al.* (2013) Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**: 1079–1090.

Rios-Covian, D., Ruas-Madiedo, Margolles, P., Gueimonde, A.M., de Los Reyes-Gavilan, C.G., and Salazar, N. (2016) Intestinal short chain fatty acids and their link with diet and human health. *Front Microbiol* **7**: 185.

Roediger, W.E. (1980) Role of anaerobic bacteria in the metabolic welfare of the colonic mucosa in man. *Gut* **21**: 793–798.

Shanahan, F., van Sinderen, D., O'toole, P.W., and Stanton, C. (2017) Feeding the microbiota: transducer of nutrient signals for the host. *Gut* **66**: 1709–1717.

Svensson, B., Svendsen, T.G., Svendsen, I., Sakai, T., and Ottesen, M. (1982) Characterization of two forms of glucoamylase from *Aspergillus niger*. *Carlsberg Res Comm* **47**: 55.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.P., *et al.* (2009) Towards the human intestinal microbiota phylogenetic core. *Environ Microbiol* **11**: 2574–2584.

Valk, V., Eeuwema, W., Sarian, F.D., van der Kaaij, R.M., Dijkhuizen, L., and Parales, R.E. (2015) Degradation of granular starch by the bacterium *Microbacterium aurum* Strain B8.A involves a modular α-amylase enzyme system with FNIII and CBM25 domains. *Appl Environ Microbiol* **81**: 6610–6620.

Valk, V., Lammerts van Bueren, A., van der Kaaij, R.M., and Dijkhuizen, L. (2016) Carbohydrate-binding module 74 is a novel starch-binding domain associated with large and multidomain α-amylase enzymes. *FEBS J* **283**: 2354–2368.

Van Duyne, G.D., Standaert, R.F., Karplus, P.A., Schreiber, S.L., and Clardy, J. (1993) Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J Mol Biol* **229**: 105–124.

Venkataraman, A., Sieber, J.R., Schmidt, A.W., Waldron, C., Theis, K.R., and Schmidt, T.M. (2016) Variable responses of human microbiomes to dietary supplementation with resistant starch. *Microbiome* **4**: 33.

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., and Bricogne, G. (2011) Data processing and analysis with the autoPROC toolbox. *Acta Crystallogr D Biol Crystallogr* **67**: 293–302.

Waffenschmidt, S., and Jaenicke, L. (1987) Assay of reducing sugars in the nanomole range with 2,2′-bicinchoninate. *Anal Biochem* **165**: 337–340.

Wang, H.B., Wang, P.Y., Wang, X., Wan, Y.L., and Liu, Y.C. (2012) Butyrate enhances intestinal epithelial barrier function via up-regulation of tight junction protein Claudin-1 transcription. *Dig Dis Sci* **57**: 3126–3135.

Warren, F.J., Royall, P.G., Gaisford, S., Butterworth, P.J., and Ellis, P.R. (2011) Binding interactions of α-amylase with starch granules: the influence of supramolecular structure and surface area. *Carbohydr Polym* **86**: 1038–1047.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr D Biol Crystallogr* **67**: 235–242.

Wong, J.M.W., de Souza, R., Kendall, C.W.C., Emam, A., and Jenkins, D.J. (2006) Colonic health: fermentation and short chain fatty acids. *J Clin Gastroenterol* **40**: 235–243.

Zackular, J.P., Baxter, N.T., Iverson, K.D., Sadler, W.D., Petrosino, J.F., Chen, G.Y., and Schloss, P.D. (2013) The gut microbiome modulates colon tumorigenesis. *mBio* **4**: e00692–e00613.

Ze, X., Duncan, S.H., Louis, P., and Flint, H.J. (2012) *Ruminococcus bromii* is a keystone species for the degradation of resistant starch in the human colon. *ISME J* **6**: 1535–1543.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.